การเปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มที่มากกว่าสองกลุ่ม ด้วยวิธี IRT LRT วิธี MGCFA-ALIGNMENT และวิธี MIMIC

สาริศา คงมี

ดุษฎีนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปรัชญาดุษฎีบัณฑิต
สาขาวิชา วิจัย วัดผลและสถิติการศึกษา
คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา
สิงหาคม 2561
ลิขสิทธิ์เป็นของมหาวิทยาลัยบูรพา

คณะกรรมการควบคุมดุษฎีนิพนธ์และคณะกรรมการสอบดุษฎีนิพนธ์ ได้พิจารณาดุษฎีนิพนธ์ของ สาริศา คงมี ฉบับนี้แล้ว เห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปรัชญาดุษฎีบัณฑิต สาขาวิชาวิจัย วัดผลและสถิติการศึกษา ของมหาวิทยาลัยบูรพาได้

คณะกรรมการควบคุมดุษฎีนิพนธ์

.....อาจารย์ที่ปรึกษาหลัก
(รองศาสตราจารย์ ดร.ไพรัตน์ วงษ์นาม)

..... อาจารย์ที่ปรึกษาร่วม
(ดร.ณัฐกฤตา งามมีฤทธิ์)

คณะกรรมการสอบดุษฎีนิพนธ์

.....ประธาน
(ดร.อาวีพร ปานทอง)

.....กรรมการ
(รองศาสตราจารย์ ดร.ไพรัตน์ วงษ์นาม)

..... กรรมการ
(ดร.ณัฐกฤตา งามมีฤทธิ์)

..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.สมโภชน์ อเนกสุข)

คณะศึกษาศาสตร์อนุมัติให้รับดุษฎีนิพนธ์ฉบับนี้ เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปรัชญาดุษฎีบัณฑิต สาขาวิชาวิจัย วัดผลและสถิติการศึกษา ของมหาวิทยาลัยบูรพา

..... คณบดีคณะศึกษาศาสตร์
(รองศาสตราจารย์ ดร.วิชิต สุรัตน์เรืองชัย)
วันที่...๖.....เดือน...กันยายน.....พ.ศ. 2561

กิตติกรรมประกาศ

ดุษฎีนิพนธ์ฉบับนี้สำเร็จลงได้ด้วยความกรุณายิ่งจาก รองศาสตราจารย์ ดร. ไพรัตน์ วงษ์นาม อาจารย์ที่ปรึกษาหลัก อาจารย์ ดร.ณัฐกฤตา งามมีฤทธิ์ อาจารย์ที่ปรึกษาร่วม ที่กรุณาให้คำปรึกษา และแนะแนวทางที่ถูกต้อง ตลอดจนแก้ไขข้อบกพร่องต่าง ๆ ด้วยความละเอียดถี่ถ้วน และเอาใจใส่ด้วยดีเสมอมา ผู้วิจัยรู้สึกซาบซึ้งเป็นอย่างยิ่ง จึงขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

ขอกราบขอบพระคุณ อาจารย์ ดร.อาวีพร ปานทอง และ ผู้ช่วยศาสตราจารย์ ดร.สมโภชน์ เอนกสุข ประธานและกรรมการสอบปากเปล่าดุษฎีนิพนธ์ ที่กรุณาให้ข้อเสนอแนะต่าง ๆ จนทำให้ดุษฎีนิพนธ์ฉบับนี้มีความสมบูรณ์มากยิ่งขึ้น

ขอกราบขอบพระคุณ ผู้อำนวยการสถาบันทดสอบทางการศึกษา (องค์การมหาชน) และขอขอบคุณ เจ้าหน้าที่ ที่กรุณาให้ความอนุเคราะห์ข้อมูลผลการสอบของนักเรียนที่เป็นกลุ่มตัวอย่าง ทำให้ผู้วิจัยสามารถทำดุษฎีนิพนธ์ฉบับนี้ได้อย่างมีคุณภาพ

ขอขอบคุณผู้อำนวยการโรงเรียนพัทลุง เพื่อนครูที่คอยให้กำลังใจและสนับสนุน ช่วยเหลือจนทำให้ผู้วิจัยมีโอกาสได้พัฒนาตนเอง ช่วยส่งเสริมให้ผู้วิจัยสามารถทำดุษฎีนิพนธ์ได้อย่างมีคุณภาพ

ขอกราบขอบพระคุณ คุณพ่อจำเริญ คุณแม่พิน คงทวี นายสุธรรม คงมี ที่ให้การสนับสนุนผู้วิจัยในทุก ๆ ด้าน ขอบคุณเด็กชายเมธัส คงมี รวมไปถึงพี่ ๆ น้อง ๆ หลาน ๆ ที่ให้กำลังใจและสนับสนุนผู้วิจัยเสมอมา ขอขอบคุณพี่ ๆ เพื่อน ๆ น้อง ๆ ร่วมสาขา และเจ้าหน้าที่ผู้เกี่ยวข้องทุกคนที่ให้บริการ ให้การดูแลช่วยเหลือแนะนำผู้วิจัย จนทำให้งานวิจัยฉบับนี้สำเร็จลุล่วงไปด้วยดี คุณค่าและประโยชน์ของ ดุษฎีนิพนธ์ฉบับนี้ ผู้วิจัยขอมอบเป็นกตัญญูกตเวทีตาแด่บุพการี บูรพาจารย์ และผู้มีพระคุณทุกท่านทั้งในอดีตและปัจจุบัน ที่ทำให้ผู้วิจัยเป็นผู้มีการศึกษาและประสบความสำเร็จมาจนตราบเท่าทุกวันนี้

สาริศา คงมี

58810143: สาขาวิชา: วิจัย วัดผลและสถิติการศึกษา; ปร.ด. (วิจัย วัดผลและสถิติการศึกษา)
คำสำคัญ: การทำหน้าที่ต่างกันของข้อสอบ/หลายกลุ่ม/วิธี IRT LRT/วิธี MGCFA-ALIGNMENT/ วิธี MIMIC

สาริศา คงมี: การเปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มที่มากกว่าสองกลุ่ม ด้วยวิธี IRT LRT วิธี MGCFA-ALIGNMENT และวิธี MIMIC (A COMPARISONS DIF FOR MULTIPLE GROUPS BY USING IRT LRT, MGCFA-ALIGNMENT AND MIMIC) คณะกรรมการควบคุมดุษฎีนิพนธ์: ไพรัตน์ วงษ์นาม, ค.ด., ณัฐกฤตา งามมีฤทธิ์, ปร.ด. 286 หน้า. ปี พ.ศ. 2561.

การวิจัยครั้งนี้ มีวัตถุประสงค์เพื่อ 1) ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT วิธี MGCFA-ALIGNMENT และวิธี MIMIC และ 2) เปรียบเทียบผลของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT วิธี MGCFA-ALIGNMENT และวิธี MIMIC ภายใต้ปัจจัยที่ต่างกัน 2 ปัจจัย ได้แก่ สถานที่ตั้งทางภูมิศาสตร์ และขนาดโรงเรียน ข้อมูลที่ใช้ในการวิจัยครั้งนี้ เป็นข้อมูลผลการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2559 ที่ผ่านการตรวจให้คะแนนแบบ 1-0 วิชาคณิตศาสตร์ จำนวน 25 ข้อ วิชาภาษาไทย และวิชาวิทยาศาสตร์ จำนวนวิชาละ 40 ข้อ กลุ่มตัวอย่าง จำนวน 8,000 คน ได้มาโดยการสุ่มแบบหลายขั้นตอน (Multi-stage random sampling technique) จำแนกเป็นโรงเรียนขนาดใหญ่พิเศษ ใหญ่ กลาง และเล็กที่อยู่ในภาคกลาง เหนือ ตะวันออกเฉียงเหนือ และภาคใต้ สถิติที่ใช้ในการวิเคราะห์ข้อมูลพื้นฐาน ได้แก่ ค่าต่ำสุด ค่าสูงสุด ค่าเฉลี่ย ค่าส่วนเบี่ยงเบนมาตรฐาน ค่าความเบ้ และค่าความโด่ง

ข้อค้นพบที่สำคัญ มีดังนี้ 1) การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วย วิธี IRT LRT วิธี MGCFA-ALIGNMENT และวิธี MIMIC พบข้อสอบที่ทำหน้าที่ต่างกันจากทุกวิชา 2) การเปรียบเทียบผลการตรวจสอบ พบว่า เมื่อจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ ผลของวิธี IRT LRT สอดคล้องกับวิธี MIMIC ในวิชาภาษาไทยและวิชาวิทยาศาสตร์ เมื่อจำแนกตามขนาดโรงเรียน ผลของวิธี IRT LRT และวิธี MGCFA-ALIGNMENT สอดคล้องกับวิธี MIMIC ในวิชาภาษาไทย วิธี IRT LRT สอดคล้องกับวิธี MGCFA-ALIGNMENT ในวิชาวิทยาศาสตร์ 3) ขนาดโรงเรียนตรวจพบการทำหน้าที่ต่างกันของข้อสอบมากกว่าสถานที่ตั้งทางภูมิศาสตร์ 4) วิธี MIMIC ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน มากที่สุดรองลงมา ได้แก่ วิธี IRT LRT และวิธี MGCFA-ALIGNMENT

ผลการวิจัยครั้งนี้เสนอแนะให้ใช้วิธี MGCFA-ALIGNMENT สำหรับใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับหลายกลุ่ม วิธีนี้สามารถให้ค่าพารามิเตอร์จำแนกเป็นรายกลุ่มในการวิเคราะห์ขั้นตอนเดียว สำหรับ 2-4 กลุ่ม สามารถใช้วิธี IRTLRT หรือวิธี MIMIC เพราะผลการวิเคราะห์ไม่ซับซ้อนายต่อการทำความเข้าใจ นับเป็นอรรถประโยชน์ต่อการพัฒนาเทคนิควิธีด้านการตรวจสอบความตรงของเครื่องมือวัดผล

58810143: MAJOR: EDUCATIONAL RESEARCH, MEASUREMENT AND STATISTICS; Ph.D. (EDUCATIONAL RESEARCH, MEASUREMENT AND STATISTICS)

KEYWORDS: DIFFERENTIAL ITEM FUNCTIONING/ MULTIGROUPS/ IRT LRT/ MGCFA-ALIGNMENT/ MIMIC

SARISA KONGMEE: A COMPARISONS DIF FOR MULTIPLE GROUPS BY USING IRT LRT, MGCFA-ALIGNMENT AND MIMIC. ADVISORY COMMITTEE: PAIRATANA WONGNAM, Ph.D., NATKRITA NGAMMEERITH, Ph.D. 286 P. 2018.

This study aimed to: 1) investigate the differential item functioning for more than two groups using the IRT LRT, MGCFA-ALIGNMENT and MIMIC methods, and 2) to compare the result of implementing differential item functioning by the IRT LRT, MGCFA-ALIGNMENT and MIMIC methods with a variety of two factors: 4 levels of geographic location and 4 levels of school size. The data for the study were from The Ordinary National Educational Test (O-NET) scores of Mathayomsuksa 3 students in the academic year 2016, Mathematics, 25 dichotomous items and Thai, 40 dichotomous items from National Institute of Educational Testing Service (Public Organization). The sampls were 8,000 students, selected by using the multi-stage random sampling techniques, classified as extra-large, large, medium and small schools size in the Central, North, Northeast and South regions. The IRTPRO program was used to analyze IRT LRT method. The Mplus program was used to analyze MGCFA-ALIGNMENT and MIMIC methods.

The main findings were: 1) According to DIF detection, the results of IRT LRT, MGCFA-ALIGNMENT and MIMIC methods were found DIF in all subjects. 2) The IRT LRT method corresponds to MIMIC method in Thai and Science, the IRT LRT method and MGCFA-ALIGNMENT method corresponds to MIMIC method in Thai, when considered by school size it was found to be consistent at 55.55% in all conditions. 3) The school sizes were found DIF more than the geographic locations. 4) The MIMIC method was more sensitive to detect DIF while the MGCFA-ALIGNMENT method was hardly to find DIF. Finally, the result of this study suggested that using the MGCFA ALIGNMENT method for item functioning for multiple groups analysis and interpretation is not complicated because using the IRTLRT and the MIMIC methods for 2-4 group was not complicated.

สารบัญ

หน้า

บทคัดย่อภาษาไทย..... ง

บทคัดย่อภาษาอังกฤษ..... จ

สารบัญ..... ฉ

สารบัญตาราง..... ซ

สารบัญภาพ..... ฑ

บทที่

1 บทนำ..... 1

ความเป็นมาและความสำคัญของปัญหา..... 1

คำถามการวิจัย..... 12

วัตถุประสงค์ของการวิจัย..... 12

สมติฐานของการวิจัย..... 13

ขอบเขตของการวิจัย..... 13

นิยามศัพท์เฉพาะ..... 14

กรอบแนวคิดในการวิจัย..... 16

ประโยชน์ที่ได้รับจากการวิจัย..... 18

2 เอกสารและงานวิจัยที่เกี่ยวข้อง..... 19

ตอนที่ 1 มโนทัศน์เกี่ยวกับทฤษฎีการตอบสนองข้อสอบ 19

ตอนที่ 2 มโนทัศน์เกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบ..... 27

ตอนที่ 3 การทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน 62

ตอนที่ 4 งานวิจัยที่เกี่ยวข้อง..... 63

3 วิธีดำเนินการวิจัย..... 88

ตอนที่ 1 การกำหนดข้อมูล..... 88

ตอนที่ 2 การวิเคราะห์ข้อมูลการทำหน้าที่ต่างกันของข้อสอบ..... 97

4 ผลการวิเคราะห์ข้อมูล..... 109

ตอนที่ 1 การวิเคราะห์ค่าสถิติพื้นฐาน 112

สารบัญ (ต่อ)

บทที่ หน้า

ตอนที่ 2 ผลการวิเคราะห์การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สำหรับกลุ่ม ที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT วิธี MGCFA ALIGNMENT และวิธี MIMIC ภายใต้ปัจจัยที่ต่างกัน 2 ปัจจัย ได้แก่ สถานที่ตั้งทางภูมิศาสตร์ และขนาดโรงเรียน 120

ตอนที่ 3 การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT วิธี MGCFA-ALIGNMENT และวิธี MIMIC ภายใต้ปัจจัยที่ต่างกัน 2 ปัจจัย ได้แก่ สถานที่ตั้งทางภูมิศาสตร์ และขนาดโรงเรียน..... 219

5 สรุปผล อภิปราย และข้อเสนอแนะ..... 243

สรุปผลการวิจัย..... 243

อภิปรายผล..... 245

ข้อเสนอแนะ..... 253

บรรณานุกรม..... 255

ภาคผนวก..... 266

ภาคผนวก ก..... 267

ภาคผนวก ข..... 269

ประวัติย่อของผู้วิจัย..... 286

สารบัญตาราง

ตารางที่		หน้า
2-1	วิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจำแนกตามลักษณะข้อมูล.....	35
2-2	คุณลักษณะของวิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 3 วิธี.....	61
2-3	งานวิจัยที่เกี่ยวข้องกับการศึกษาปัจจัยที่มีผลต่อการทำหน้าที่ต่างกันของข้อสอบ.....	66
2-4	งานวิจัยที่เกี่ยวข้องกับการศึกษาประสิทธิภาพของวิธีและศึกษาเชิงเปรียบเทียบวิธีที่ใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ.....	77
3-1	การแบ่งกลุ่มจังหวัดตามศึกษาธิการภาค.....	90
3-2	ประชากรและกลุ่มตัวอย่าง.....	93
3-3	มาตรฐานการผลิตและการพัฒนา แบบทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ระดับชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559.....	95
3-4	จำนวนข้อสอบ O-NET ชั้นมัธยมศึกษาปีที่ 3 จำแนกตามรูปแบบข้อสอบประจำปีการศึกษา 2559.....	96
3-5	วิธี IRT LRT วิธี MGCFA-ALIGNMENT และวิธี MIMIC.....	98
4-1	ค่าสถิติพื้นฐานของข้อมูลผลการสอบรายข้อจากแบบทดสอบวิชาคณิตศาสตร์จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน.....	112
4-2	ค่าสถิติพื้นฐานของข้อมูลผลการสอบรายข้อจากแบบทดสอบวิชาภาษาไทยจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน.....	114
4-3	ค่าสถิติพื้นฐานของข้อมูลผลการสอบรายข้อจากแบบทดสอบวิชาวิทยาศาสตร์จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน.....	115
4-4	ค่าสถิติพื้นฐานของข้อมูลผลการสอบรายข้อจากแบบทดสอบวิชาคณิตศาสตร์จำแนกตามขนาดโรงเรียน.....	116
4-5	ค่าสถิติพื้นฐานของข้อมูลผลการสอบรายข้อจากแบบทดสอบวิชาภาษาไทยจำแนกตามขนาดโรงเรียน.....	118
4-6	ค่าสถิติพื้นฐานของข้อมูลผลการสอบรายข้อจากแบบทดสอบวิชาวิทยาศาสตร์จำแนกตามขนาดโรงเรียน.....	119
4-7	ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ ด้วยวิธี IRT LRT จำแนกตามสถานที่ตั้งทางภูมิศาสตร์	121

สารบัญตาราง (ต่อ)

ตารางที่ หน้า

4-8 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาภาษาไทย ด้วยวิธี IRT LRT จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ 123

4-9 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาวิทยาศาสตร์ ด้วยวิธี IRT LRT จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ 126

4-10 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ ด้วยวิธี IRT LRT จำแนกตามขนาดโรงเรียน 130

4-11 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาภาษาไทย ด้วยวิธี IRT LRT จำแนกตามขนาดโรงเรียน..... 132

4-12 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาวิทยาศาสตร์ ด้วยวิธี IRT LRT จำแนกตามขนาดโรงเรียน 136

4-13 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ ด้วยวิธี MGCFA-ALIGNMENT แบบ Uniform DIF และแบบ Nonuniform DIF จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ 140

4-14 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย ด้วยวิธี MGCFA-ALIGNMENT แบบ Uniform DIF และแบบ Nonuniform DIF จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ 141

4-15 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาวิทยาศาสตร์ ด้วยวิธี MGCFA ALIGNMENT แบบ Uniform DIF และแบบ Nonuniform DIF จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ 143

4-16 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ ด้วยวิธี MGCFA ALIGNMENT แบบ Uniform DIF และแบบ Nonuniform DIF จำแนกตามขนาดโรงเรียน 146

4-17 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย ด้วยวิธี MGCFA ALIGNMENT แบบ Uniform DIF และแบบ Nonuniform DIF จำแนกตามขนาดโรงเรียน 147

4-18 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาวิทยาศาสตร์ ด้วยวิธี MGCFA ALIGNMENT แบบ Uniform DIF และแบบ Nonuniform DIF จำแนกตามขนาดโรงเรียน 149

สารบัญตาราง (ต่อ)

ตารางที่ หน้า

4-19 ความกลมกลืนของโมเดล CFA MIMIC วิชาคณิตศาสตร์ จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ 151

4-20 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ ด้วยวิธี MIMIC แบบ Uniform DIF จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ 154

4-21 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ ด้วยวิธี MIMIC แบบ Nonuniform DIF จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ 157

4-22 ความกลมกลืนของโมเดล CFA MIMIC วิชาภาษาไทย จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ 161

4-23 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย ด้วยวิธี MIMIC แบบ Uniform DIF จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ 163

4-24 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย ด้วยวิธี MIMIC แบบ Nonuniform DIF จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ 168

4-25 ความกลมกลืนของโมเดล CFA MIMIC วิชาวิทยาศาสตร์ จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ 173

4-26 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาวิทยาศาสตร์ด้วยวิธี MIMIC แบบ Uniform DIF จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ 175

4-27 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิธี MIMIC แบบ Nonuniform DIF วิชาวิทยาศาสตร์ จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ 180

4-28 ความกลมกลืนของโมเดล CFA MIMIC วิชาคณิตศาสตร์ จำแนกตามขนาดโรงเรียน.. 186

4-29 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ วิธี MIMIC แบบ Uniform DIF จำแนกตามขนาดโรงเรียน 188

4-30 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ ด้วยวิธี MIMIC แบบ Nonuniform DIF จำแนกตามขนาดโรงเรียน 191

4-31 ความกลมกลืนของโมเดล CFA MIMIC วิชาภาษาไทย จำแนกตามขนาดโรงเรียน..... 195

4-32 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาภาษาไทย ด้วยวิธี MIMIC แบบ Uniform DIF จำแนกตามขนาดโรงเรียน 197

สารบัญตาราง (ต่อ)

ตารางที่ หน้า

4-34 ความกลมกลืนของโมเดล CFA MIMIC วิชาวิทยาศาสตร์ จำแนกตามขนาดโรงเรียน 207

4-35 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาวิทยาศาสตร์ ด้วยวิธี MIMIC แบบ Uniform DIF จำแนกตามขนาดโรงเรียน 209

4-36 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาวิทยาศาสตร์ วิธี MIMIC แบบ Nonuniform DIF จำแนกตามขนาดโรงเรียน 214

4-37 สรุปผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ ด้วย 3 วิธี จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ 220

4-38 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ วิธี IRT LRT และวิธี MGCFA-ALIGMENT จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ 221

4-39 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ วิธี IRT LRT และวิธี MIMIC จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ 222

4-40 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ วิธี MGCFA-ALIGMENT และวิธี MIMIC จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ 222

4-41 สรุปผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ ด้วย 3 วิธี จำแนกตามขนาดโรงเรียน 223

4-42 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ วิธี IRT LRT และวิธี MGCFA-ALIGMENT จำแนกตามขนาดโรงเรียน 224

4-43 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ วิธี IRT LRT และวิธี MIMIC จำแนกตามขนาดโรงเรียน 224

4-44 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ วิธี MGCFA-ALIGMENT และวิธี MIMIC จำแนกตามขนาดโรงเรียน 225

สารบัญตาราง (ต่อ)

ตารางที่ หน้า

4-45 สรุปผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาภาษาไทย ด้วย 3 วิธี จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ 226

4-46 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาภาษาไทย วิธี IRT LRT และวิธี MGCFA-ALIGMENT จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ 228

4-47 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาภาษาไทย วิธี IRT LRT และวิธี MIMIC จำแนกตามสถานที่ตั้งทางภูมิศาสตร์..... 228

4-48 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาภาษาไทย วิธี MGCFA-ALIGMENT และวิธี MIMIC จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ 229

4-49 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาภาษาไทย จำแนกตามขนาดโรงเรียน..... 229

4-50 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาภาษาไทย วิธี IRT LRT และวิธี MGCFA-ALIGMENT จำแนกตามขนาดโรงเรียน..... 231

4-51 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาภาษาไทย วิธี IRT LRT และวิธี MIMIC จำแนกตามขนาดโรงเรียน..... 232

4-52 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาภาษาไทย วิธี MGCFA-ALIGMENT และวิธี MIMIC จำแนกตามขนาดโรงเรียน..... 232

4-53 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาวิทยาศาสตร์ จำแนกตามสถานที่ตั้งทางภูมิศาสตร์..... 233

4-54 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาวิทยาศาสตร์ วิธี IRT LRT และวิธี MGCFA-ALIGMENT จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ 235

4-55 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาวิทยาศาสตร์ วิธี IRT LRT และวิธี MIMIC จำแนกตามสถานที่ตั้งทางภูมิศาสตร์.. 236

สารบัญตาราง (ต่อ)

ตารางที่ หน้า

4-56 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาภาษาไทย วิธี MGCFA-ALIGMENT และวิธี MIMIC จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ 236

4-57 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาวิทยาศาสตร์ จำแนกตามขนาดโรงเรียน..... 237

4-58 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาวิทยาศาสตร์ วิธี IRT LRT และวิธี MGCFA-ALIGMENT จำแนกตามขนาดโรงเรียน..... 239

4-59 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาวิทยาศาสตร์ วิธี IRT LRT และวิธี MIMIC จำแนกตามขนาดโรงเรียน..... 239

4-60 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาวิทยาศาสตร์ วิธี MGCFA-ALIGMENT และวิธี MIMIC จำแนกตามขนาดโรงเรียน..... 240

4-61 สรุปผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ วิชาภาษาไทย และ วิชาวิทยาศาสตร์ ด้วย 3 วิธี จำแนกตามสถานที่ตั้งทางภูมิศาสตร์และขนาดโรงเรียน..... 241

สารบัญภาพ

ภาพที่ หน้า

1-1 กรอบแนวคิดในการวิจัย..... 17

2-1 กรอบแนวคิดของการทำหน้าที่ต่างกันของข้อสอบ..... 30

2-2 รูปแบบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป..... 31

2-3 รูปแบบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป..... 32

2-4 (a) เมื่อข้อสอบข้อที่ i เป็นการศึกษาข้อสอบและข้อสอบอื่น ๆ ที่เป็นข้อสอบร่วมกัน และ (b) เมื่อข้อสอบข้อที่ i เป็นข้อสอบร่วมและข้อสอบอื่น ๆ ที่เป็นข้อสอบที่ศึกษา..... 52

2-5 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี MIMIC แบบเอกรูป..... 54

2-6 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี MIMIC แบบอเนกรูป..... 55

2-7 โมเดลการวิเคราะห์องค์ประกอบตามแนวคิด IRT..... 57

2-8 โมเดลการวิเคราะห์การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ MIMIC model โดยใช้ตัวแปรสาเหตุ 1 ตัว..... 58

2-9 โมเดลการวิเคราะห์การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ MIMIC model โดยใช้ตัวแปรสาเหตุมากกว่า 1 ตัว แบบ Uniform DIF..... 59

2-10 โมเดลการวิเคราะห์การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ MIMIC model โดยใช้ตัวแปรสาเหตุมากกว่า 1 ตัว แบบ Nonuniform DIF..... 60

3-1 โมเดล MIMIC แบบเอกรูปโดยมีข้อ Q_I เป็นข้อที่ทำหน้าที่ต่างกัน 103

3-2 Model CFA ไม่มีตัวแปรทำนาย..... 104

3-3 เพิ่มตัวแปรทำนาย (G) แต่ไม่มีอิทธิพลตรงต่อข้อสอบ..... 105

3-4 เพิ่มอิทธิพลทางตรงของ G ต่อข้อสอบแต่บังคับค่าให้เป็น 0..... 105

3-5 โมเดล MIMIC แบบเอกรูป และแบบอเนกกรูป..... 106

บทที่ 1
บทนำ

ความเป็นมาและความสำคัญของปัญหา

แนวคิดการวัดผลที่สำคัญยิ่งคือ การมุ่งหวังวัดได้ตรงกับคุณลักษณะที่ต้องการวัด วัดได้ตรงก็สามารถแปลผลได้ตรง สิ่งที่เกี่ยวข้องโดยตรงกับการทดสอบหรือการวัดคือ เครื่องมือวัด เครื่องมือวัดที่นำมาใช้ควรเป็นเครื่องมือที่มีคุณภาพ โดยพิจารณาจากค่าความตรง ค่าความเชื่อมั่น ค่าความยากง่าย และค่าอำนาจจำแนก สามารถจำแนกผู้สอบได้อย่างเหมาะสม ความสามารถของเครื่องมือในการวัดสิ่งที่ต้องการจะวัดได้ถูกต้อง เป็นคุณสมบัติด้านความเที่ยงตรงหรือความตรงของเครื่องมือวัด ครอนบาค (Cronbach, 1971 อ้างถึงใน ไพรัตน์ วงษ์นาม, 2546, หน้า 271) กล่าวว่า ความตรงเป็นคุณสมบัติของเครื่องมือที่ชี้ให้เห็นว่าค่าที่วัดได้เป็นตัวแทนของสิ่งที่ถูกวัดในประชากรได้อย่างถูกต้อง การพัฒนาการแบบวัดหรือแบบสอบที่ได้ศึกษากันตลอด คือ กระบวนการตรวจสอบความตรง ศิริชัย กาญจนวาสี (2555, หน้า 103) กล่าวว่า การสร้างและการตรวจสอบคุณภาพของแบบสอบจะต้องคำนึงถึงคุณภาพด้านความตรงเป็นสำคัญ เนื่องด้วยความตรงเป็นคุณสมบัติของแบบสอบที่สะท้อนถึงความสามารถในการวัดได้ถูกต้องแม่นยำ ถ้าผลการวัดมีค่าที่ใกล้เคียงกับค่าคุณลักษณะ ที่แท้จริงเท่าใด ก็ถือว่าการวัดมีความตรงมากขึ้นเท่านั้น Marie (2007, Abstract) กล่าวว่า ความตรงของเครื่องมือวัดขึ้นอยู่กับคุณภาพของข้อสอบที่อยู่ในเครื่องมือนั้น เช่นเดียวกันกับที่ อาวีพร ปานทอง (2560) ได้พัฒนาแบบวัดทักษะการคิดขั้นสูงขึ้น โดยเห็นว่า การวัดผลการจัดการเรียนการสอนพัฒนาทักษะการคิดขั้นสูงได้นั้น ต้องวัดจากเครื่องมือที่มีคุณภาพ จึงจะให้ผลการวัดที่มีคุณภาพ หากใช้เครื่องมือที่ไม่ได้ตรวจสอบคุณภาพถือเป็นจุดอ่อนที่สำคัญ คือ คุณสมบัติของข้อคำถามซึ่งไม่คงที่แปรเปลี่ยนไปตามกลุ่มของผู้สอบ ทำให้ผลการประเมินคลาดเคลื่อน ไม่สามารถนำข้อมูลที่ได้ไปใช้เป็นข้อมูลในการปรับปรุงแก้ไขอ้างอิงไปสู่กลุ่มอื่น ๆ หรือพัฒนาการเรียนรู้ของผู้เรียนและการจัดการเรียนการสอนได้อย่างมีประสิทธิภาพ วิธีตรวจสอบความตรงมีหลายวิธี การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (Differential item functioning: DIF) เป็นวิธีหนึ่งที่ใช้สำหรับตรวจสอบคุณภาพด้านความตรงในประเด็นของความยุติธรรมของข้อสอบและแบบสอบ (Item and test unfairness) ผลจากการวัดผลย่อมส่งผลต่อการตัดสินผล ความไม่ยุติธรรมที่เกิดจากข้อสอบเกิดขึ้นในกรณีที่ทำให้ผู้สอบระหว่างกลุ่มย่อยที่มีความสามารถหลักเท่ากันมีการได้เปรียบหรือเสียเปรียบกัน การทำหน้าที่ต่างกันของข้อสอบเป็นเรื่องที่ได้รับความสนใจของนักวิจัยและนักการศึกษา ได้มีการศึกษา

วิพากษ์วิจารณ์เป็นอย่างมาก เนื่องจากแบบสอบเป็นเครื่องมือสำคัญในการวัดผลการเรียนรู้ Zumbo กล่าวว่า การทดสอบที่ไม่มีความลำเอียงเป็นสิ่งที่นำมาพิจารณาและให้ความสำคัญสำหรับการคัดเลือกและการใช้แบบทดสอบวัดทางด้านจิตวิทยาหรือด้านการวัดและประเมินผลการศึกษา จำเป็นอย่างยิ่งที่จะต้องมีการทดสอบที่เป็นธรรม แก่ผู้สอบทั้งหมดว่าไม่ลำเอียงเข้าข้างกับกลุ่มผู้สอบกลุ่มย่อยใดกลุ่มย่อยหนึ่ง (Zumbo, 1999, p. 5) เช่นเดียวกับ Camilli and Shepard กล่าวว่า เมื่อความน่าจะเป็นของการตอบสนองต่อข้อสอบ (ตอบถูก) แตกต่างกันสำหรับสมาชิกในกลุ่มต่าง ๆ ในประชากร (เช่น เชื้อชาติ) หลังจากการควบคุมลักษณะแฝงที่ถูกวัดให้มีค่าเท่ากันในกลุ่มย่อยโดยเครื่องมือเดียวกัน เช่น ผลการอ่าน หากผลเกิดความแตกต่างแสดงว่า ข้อสอบเกิดการทำหน้าที่ต่างกัน (Differential item functioning: DIF) (Camilli & Shepard, 1994) ดังนั้น ผู้สร้างแบบสอบจึงจำเป็นต้องคำนึงถึงคุณภาพของแบบสอบโดยต้องสร้างข้อสอบที่เชื่อถือได้ วัดได้ตามคุณลักษณะที่ต้องการวัด เกิดความคลาดเคลื่อนในการวัดน้อยที่สุด และจะต้องระมัดระวังไม่ให้เกิดการได้ประโยชน์หรือเสียประโยชน์ของผู้สอบในแต่ละกลุ่มที่มีความสามารถเท่ากัน (Allen & Yen, 1979; Popham, 1981) การทำหน้าที่ต่างกันของข้อสอบ เดิมใช้คำว่า “ความลำเอียงของข้อสอบ” (Item bias) ซึ่งเป็นภาษาที่ใช้กันโดยทั่วไปและมีความหมายในเชิงลบ สำหรับการตัดสินว่าข้อสอบมีความลำเอียงหรือไม่นั้น มักจะพิจารณาอิทธิพลที่สังเกตได้ของผู้สอบกลุ่มย่อยที่นำมาศึกษา โดยไม่คำนึงถึงวิธีการทางสถิติ จึงทำให้เกิดความคลุมเครือเกี่ยวกับเกณฑ์ที่ใช้ในการตัดสินความลำเอียง ต่อมานักวิธีวิทยาการวิจัยได้พัฒนาวิธีเพื่อใช้วิเคราะห์ดัชนีความลำเอียงของข้อสอบ โดยมุ่งเน้นความแตกต่างระหว่างกลุ่มผู้สอบที่ตอบข้อสอบข้อเดียวกันและเลือกเกณฑ์ที่ใช้ในการจับคู่กลุ่มผู้สอบ (Matching criterion) แล้วนำสารสนเทศทางสถิติมาใช้เป็นเกณฑ์ในการตัดสินความลำเอียงของข้อสอบ ดังนั้น จึงเปลี่ยนมาใช้คำว่า “การทำหน้าที่ต่างกันของข้อสอบ” (Differential item functioning: DIF) แทน ซึ่งเป็นคำที่มีความหมายเป็นกลางและมีความเหมาะสมมากกว่า (Holland & Wainer, 1993; วลีมาศ แซ่อึ้ง, 2543)

สรุปความเชื่อมโยงของแนวคิดได้ว่าการทำหน้าที่ต่างกันของข้อสอบ (DIF) กับความสำเอียงของข้อสอบ (Item bias) มีแนวคิดที่ต่างกันในประเด็น การทำหน้าที่ต่างกันของข้อสอบ เป็นการดำเนินการที่เน้นการใช้วิธีการทางสถิติสำหรับการตรวจสอบ เพื่อให้ได้สารสนเทศเกี่ยวกับการทำหน้าที่ของข้อสอบสำหรับกลุ่มผู้สอบกลุ่มย่อยที่มีลักษณะเฉพาะบางอย่างแตกต่างกัน ส่วนความลำเอียงของข้อสอบ เป็นกระบวนการตัดสินความยุติธรรมของข้อสอบ โดยนำสารสนเทศการทำหน้าที่ต่างกันของข้อสอบมาวิเคราะห์ (Logical analysis) โดยผู้เชี่ยวชาญพิจารณาถึงการเขียนข้อสอบ เนื้อหาสาระของข้อสอบและจุดมุ่งหมายของการวัด เพื่อระบุว่าข้อสอบข้อนั้นลำเอียงเข้าข้างกลุ่มหนึ่งกลุ่มใดหรือไม่ (Camilli & Shepard, 1994)

การกำหนดความหมายหรือนิยามของ “การทำหน้าที่ต่างกันของข้อสอบ” ทั้งนักวิชาการและนักวิจัยส่วนมากได้ให้ความหมายในทำนองเดียวกัน สามารถสรุปได้ว่า ข้อสอบที่ทำหน้าที่ต่างกันจะเกิดความลำเอียง โดยข้อสอบหรือแบบทดสอบอาจเข้าข้างผู้สอบกลุ่มใดกลุ่มหนึ่ง ซึ่งมีผลทำให้ผู้สอบกลุ่มนั้นได้เปรียบในการตอบข้อสอบ ส่วนผู้สอบอีกกลุ่มหนึ่งเสียเปรียบ ทั้ง ๆ ที่ผู้สอบสองกลุ่มมีระดับความสามารถเท่ากัน การทำหน้าที่ต่างกันของข้อสอบหรือแบบทดสอบดังกล่าวถือเป็นความคลาดเคลื่อนอย่างเป็นระบบ (Systematic error) (Camilli & Shepard, 1994, p. 8)

โดยในการเปรียบเทียบผลการตอบข้อสอบระหว่างผู้สอบกลุ่มย่อยอย่างน้อยสองกลุ่ม ที่มีความสามารถระดับเดียวกันนั้น ผู้สอบกลุ่มที่ได้เปรียบ เรียกว่า กลุ่มอ้างอิง (Rreference group: R) และผู้สอบกลุ่มที่เสียเปรียบ เรียกว่า กลุ่มเปรียบเทียบ (Focal group: F) ซึ่งเป็นกลุ่มที่สนใจศึกษาการทำหน้าที่ต่างกันของข้อสอบ ข้อสอบทำหน้าที่ต่างกัน เมื่อผู้สอบที่มีความสามารถระดับเดียวกันแต่อยู่ต่างกลุ่มกันมีโอกาสของการตอบข้อสอบได้ถูกต้องไม่เท่ากัน (Li & Stout, 1996 อ้างถึงในวลีมาศ แซ่อึ้ง, 2543) ซึ่งขนาดและทิศทางของข้อสอบที่ทำหน้าที่ต่างกันจะแปรเปลี่ยนไปตามระดับความสามารถที่แตกต่างกัน เกณฑ์ที่ใช้ในการแบ่งกลุ่มอ้างอิงและกลุ่มเปรียบเทียบมีความหลากหลาย เช่น เพศ เชื้อชาติ ภูมิลำเนา ศาสนา ภาษา เป็นต้น

การทำหน้าที่ต่างกันของข้อสอบแบ่งเป็นสองประเภท คือ ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป (Uniform DIF) และข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป (Nonuniform DIF) (ศิริชัย กาญจนวาสี, 2555) ข้อสอบทำหน้าที่ต่างกันประเภทแรกเกิดขึ้นเมื่อไม่มีปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับการเป็นสมาชิกของกลุ่ม ส่วนข้อสอบทำหน้าที่ต่างกันประเภทหลังเกิดขึ้นเมื่อมีปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับการเป็นสมาชิกของกลุ่ม ซึ่งทฤษฎีการตอบสนองข้อสอบ (Item response theory: IRT) สามารถพิจารณาปฏิสัมพันธ์ได้จากความแตกต่างของค่าพารามิเตอร์อำนาจจำแนกของข้อสอบระหว่างผู้สอบกลุ่มย่อยสองกลุ่ม กล่าวคือ ถ้าข้อสอบระหว่างผู้สอบกลุ่มย่อยสองกลุ่ม มีค่าอำนาจจำแนกเท่ากันแล้ว โค้งลักษณะข้อสอบ (Iem characteristic curves: ICCs) ระหว่างกลุ่มผู้สอบจะขนานกัน แสดงว่า ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป (Uniform DIF) แต่ถ้าข้อสอบระหว่างผู้สอบกลุ่มย่อยสองกลุ่ม มีค่าอำนาจจำแนกไม่เท่ากันแล้ว โค้งลักษณะข้อสอบระหว่างกลุ่มผู้สอบจะไม่ขนานกัน แสดงว่าข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป (Nonuniform DIF) (Camilli & Shepard, 1994)

สำหรับวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (Differential item functioning methods) เมื่อใช้เกณฑ์จำแนกตามการตรวจให้คะแนนสามารถจำแนกได้เป็น 2 กลุ่ม คือ การตรวจให้คะแนนแบบทวิวิภาค (Dichotomous) และการตรวจให้คะแนนแบบพหุวิภาค (Polytomous)

สำหรับข้อสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค (Dichotomous) นั้นได้แบ่งตามมิติลักษณะของตัวแปรเกณฑ์ ได้เป็นสองกลุ่ม ดังนี้ (Potenza & Dorans, 1995; Feinstein, 1995 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555)

1. กลุ่มวิธีที่ใช้คะแนนสังเกตได้ (Observed score) ซึ่งวิเคราะห์ตามทฤษฎีแบบดั้งเดิม (Classical test theory: CTT) ใช้คะแนนรวมของผู้สอบเป็นเกณฑ์การจับคู่ของกลุ่มผู้สอบ วิธีการ ตรวจสอบที่สำคัญ ได้แก่ การวิเคราะห์ความแปรปรวน (ANOVA) วิธีการถดถอยโลจิสติก (Logistic regression: LR) วิธีแปลงค่าความยากของข้อสอบ (Transformed item difficulty: TID) วิธีแมนเทล-แฮนส์เซล (Mantel-haenszel; MH) และวิธีดัชนีมาตรฐาน (Standardization: STND)

2. กลุ่มวิธีที่ใช้คุณลักษณะแฝง (Latent variable) ซึ่งวิเคราะห์บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ (Item response theory: IRT) ใช้ค่าประมาณความสามารถของผู้สอบเป็น เกณฑ์การจับคู่กลุ่มผู้สอบ วิธีการตรวจสอบที่สำคัญ ได้แก่ วิธีวัดพื้นที่ความแตกต่างระหว่างโค้ง การตอบสนองข้อสอบ (IRT-D_2) วิธีไค-สแควร์ของลอร์ด (Lord's χ^2) วิธีอัตราส่วนไลค์ลิฮูดทั่วไป (General IRT likelihood ratio) วิธีอัตราส่วนไลค์ลิฮูด ลอกลิเนียร์ (Loglinear IRT likelihood ratio) และวิธีซิปเทสท์ (Simultaneous item bias test: SIBTEST)

วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่วิเคราะห์ด้วยทฤษฎี CTT มีข้อจำกัดเกี่ยวกับความคงที่ของค่าสถิติ โดยจะมีค่าแปรเปลี่ยนไปตามกลุ่มตัวอย่าง ส่งผลให้การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบไม่น่าเชื่อถือ นอกจากนี้ ยังใช้เกณฑ์การจับคู่กลุ่มผู้สอบภายใน (Internal matching criterion) โดยใช้คะแนนรวมของแบบสอบแทนระดับความสามารถของผู้สอบ ซึ่งอาจมีผลทำให้แบบสอบเกิดความลำเอียง เนื่องมาจากเกณฑ์ดังกล่าวจะใช้คะแนนรวมที่ได้มาจากข้อสอบที่ทำหน้าที่ต่างกันรวมอยู่ด้วย (Zieky, 1993) จากจุดอ่อนของทฤษฎี CTT จึงทำให้นักวิธีวิทยาการวิจัยนำทฤษฎี IRT ไปประยุกต์ใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยเชื่อว่าเป็นทฤษฎีที่มีความแกร่ง สามารถแก้ปัญหาของทฤษฎี CTT ได้ จุดเด่นของทฤษฎี IRT คือ ความไม่แปรเปลี่ยนของค่าพารามิเตอร์ของข้อสอบ โดยสามารถนำคุณสมบัติทางสถิติของข้อสอบไปอธิบายลักษณะของข้อสอบที่ทำหน้าที่ต่างกันได้แม่นยำกว่าทฤษฎี CTT (Camilli & Shepard, 1994) วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้ทฤษฎี IRT ที่นิยมนำมาใช้ เช่น วิธีการวัดพื้นที่ของราชู วิธีการวัดพื้นที่ของคิมและโคเฮน วิธีการทดสอบไค-สแควร์ของลอร์ด และวิธีการทดสอบอัตราส่วนไลค์ลิฮูด (วลีมาศ แซ่อึ้ง, 2543)

ในการศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (Differential item functioning methods) ที่ให้คะแนนแบบทวิวิภาค (Dichotomous) โดยตอบถูกให้ 1 และตอบผิดให้ 0 นั้น ยังคงเป็นประเด็นทางการศึกษาที่นักวัดผลให้ความสำคัญ มีการศึกษาค้นคว้ากันอย่างต่อเนื่อง

เพราะยังคงใช้เป็นเครื่องมือวัดผลอย่างกว้างขวาง ทำให้ยังมีการศึกษาในเชิงลึกและเชิงกว้าง เชิงลึกเป็นการศึกษาปัจจัยที่เกี่ยวข้องกับประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแต่ละวิธี ปัจจัยที่ศึกษา แบ่งเป็นปัจจัยที่เกี่ยวกับข้อสอบหรือแบบสอบ ได้แก่ ความยากของข้อสอบ ความยาวของแบบสอบ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน ขนาดกลุ่มตัวอย่าง ปัจจัยที่เกี่ยวกับกลุ่มย่อยของประชากร ได้แก่ เพศ ภูมิภาค เชื้อชาติ การศึกษาในเชิงกว้าง เป็นการศึกษาเชิงเปรียบเทียบวิธีที่มีประสิทธิภาพสำหรับข้อมูลที่มีความหลากหลาย ก่อให้เกิดสารสนเทศที่กว้างขึ้นในประเด็นข้อมูลที่ศึกษามีทั้งข้อมูลที่เป็นข้อมูลจำลองโดยใช้ทฤษฎีรองรับ และมีการศึกษากับข้อมูลเชิงประจักษ์ การศึกษาล้วนแล้วแต่ศึกษาผู้สอบกลุ่มที่ได้เปรียบหรือกลุ่มอ้างอิง (Reference group: R) และผู้สอบกลุ่มที่เสียเปรียบหรือกลุ่มเปรียบเทียบ (Focal group: F) แบบ 1: 1 ทั้งสิ้น แม้แต่เกณฑ์การแบ่งกลุ่มนั้นสามารถแบ่งได้หลายกลุ่ม จากการศึกษาของ Seokjoon (2014) ได้ศึกษาประสิทธิภาพของโมเดล MIMIC ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและเอนกรูป เมื่อมีกลุ่มตัวอย่างที่ศึกษาพร้อมกันหลายกลุ่ม ใช้วิธี Sequential-free baseline สำหรับวิเคราะห์โมเดล MIMIC ที่ได้ถูกพัฒนาขึ้น และเปรียบเทียบกับวิธี Free and constrained baseline model วิธี Sequential-free baseline ใช้ในการจำแนกข้อสอบที่ระบุไว้ว่ามี Non-DIF ในการทดสอบ Constrained baseline model เพื่ออ้างอิงสำหรับการเปรียบเทียบกับ Free baseline model นอกจากนี้ยังได้มีการตรวจสอบผลของการละเมิดข้อตกลงเบื้องต้นของวิธี MIMIC DIF เกี่ยวกับการเท่ากันของความแปรปรวนของตัวแปรแฝง โดยรวมแล้วการศึกษาแบบจำลอง วิธี Sequential-free baseline พบว่า ความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบคล้ายกับ Free baseline method ที่เกี่ยวข้องกับรูปแบบที่กำหนดไว้ และความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบดีกว่า Constrained baseline model ต่อมา Finch (2016) ได้ศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มที่มากกว่าสองกลุ่ม การศึกษาครั้งนี้ใช้วิธีการจำลองข้อมูลศึกษาการเปรียบเทียบวิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มากกว่า 2 กลุ่ม 4 วิธี ได้แก่ วิธี The generalized mantel-haenszel test, Generalized logistic regression, Lord’s chi-square test, and the multiple group alignment procedure ผลการศึกษาพบว่า วิธี The generalized mantel-haenszel test, and the multiple group alignment procedure เป็นวิธีที่ดีที่สุดควบคุมความคลาดเคลื่อนประเภทที่ 1 และอำนาจของการทดสอบ ข้อจำกัด และทิศทางการวิจัยในอนาคต หวังว่าการศึกษาในปัจจุบันได้ให้ข้อมูลเชิงลึกบางประการเกี่ยวกับประสิทธิภาพของหลายวิธีในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป (Uniform DIF) ซึ่งมีมากกว่าสองกลุ่ม จากแนวทางที่ศึกษานี้ผู้วิจัยจึงสนใจนำมโนทัศน์ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มที่มากกว่าสองกลุ่มมาศึกษาทั้งแบบเอกรูปและแบบอเนกรูปเพื่อขยายแนวคิดในการวิเคราะห์ต่อ แต่เป็นการศึกษากับ

ข้อมูลจริงที่เป็นการสอบที่มีความสำคัญสำหรับการศึกษาในประเทศไทย ที่เรียกว่า การประเมินคุณภาพการศึกษาระดับชาติ (O-NET) ปีการศึกษา ดำเนินการสอบโดยสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน)

Thissen, Steinberg and Wainer (1988 cited in De Ayala, 2009) ได้เสนอวิธีทดสอบอัตราส่วนความควรจะเป็น (Likelihood ratio test: LR) สำหรับใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบความแตกต่างของผลการตอบข้อสอบระหว่างผู้สอบกลุ่มย่อย 2 กลุ่ม เช่น เพศ ภูมิลำเนา เชื้อชาติ ศาสนา เป็นต้น วิธีทดสอบอัตราส่วนความควรจะเป็นจะใช้วิธีทดสอบอัตราส่วนความควรจะเป็นสูงสุดตามทฤษฎีการตอบสนองข้อสอบในรูปเข้าสู่ระบบเชิงเส้น (Loglinear: LR) เป็นการประมาณค่าพารามิเตอร์ด้วยวิธีความน่าจะเป็นสูงสุด (Maximum likelihood: ML) วิธีนี้ใช้ทดสอบอัตราส่วนความควรจะเป็นเพื่อทดสอบนัยสำคัญของการทำหน้าที่ต่างกันของข้อสอบ โดยใช้โปรแกรม IRTLRDIF v.2.0b (Thissen, 2001)

งานวิจัยของ Cohen, Kim and Wollack (1996) ได้ศึกษาอัตราความคลาดเคลื่อนประเภทที่ 1 สำหรับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการทดสอบอัตราส่วนความควรจะเป็น โดยใช้เทคนิคมอนติคาโลในการจำลองข้อมูล โมเดลทฤษฎีการตอบสนองข้อสอบ (IRT) แบบสองและสามพารามิเตอร์ จำลองข้อมูลโดยทำการวนซ้ำ 100 รอบ ข้อสอบ 50 ข้อ ขนาดกลุ่มตัวอย่าง 250 และ 1,000 คน สำหรับแบบจำลอง IRT การประมาณค่าพารามิเตอร์ของข้อสอบด้วยวิธีความน่าจะเป็นสูงสุด สำหรับโมเดล IRT สามรูปแบบ ได้แก่ โมเดลสามพารามิเตอร์ โมเดลสามพารามิเตอร์กำหนดค่าการเดาให้มีค่าคงที่ และโมเดลสองพารามิเตอร์ การเปรียบเทียบ DIF ทั้งหมดได้รับการจำลองโดยการจับคู่ตัวอย่างจากทุกขนาดกลุ่มตัวอย่าง และภายใต้ข้อตกลงเบื้องต้นของโมเดล IRT ได้กลุ่มอ้างอิงและกลุ่มโฟกัส 50 คู่ ผลการศึกษาพบว่าอัตราความคลาดเคลื่อนประเภทที่ 1 สำหรับโมเดลสองพารามิเตอร์ อยู่ในค่าที่คาดการณ์ไว้ในทางทฤษฎีในแต่ละระดับ α ที่กำหนดไว้ อัตราความคลาดเคลื่อนประเภทที่ 1 สำหรับโมเดลสามพารามิเตอร์และโมเดลสามพารามิเตอร์กำหนดค่าการเดาให้มีค่าคงแตกต่างจากค่าที่คาดการณ์ไว้ตามทฤษฎีในระดับα ที่พิจารณา

Kim (2000) ได้ศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการทดสอบอัตราส่วนไลด์ลิฮู้ด (Likelihood ratio test) วิธีแมนเทล (Mantel) และวิธีแมนเทล-แฮนเซลส์แบบทั่วไป (Generalized mantel-haenszel: GMH) ใช้ข้อมูลจากการประเมินโรงเรียนระดับอนุบาลของรัฐจอเจียร์ ปัจจัยที่ศึกษา ได้แก่ ขนาดของกลุ่มตัวอย่าง (105,731: 10,000: 1,000: 100) ข้อคำถามเป็นแบบให้คะแนนมากกว่า 2 ค่า ผลการศึกษาพบว่า วิธีตรวจสอบการทำหน้าที่ต่างกันทั้ง 3 วิธี ให้ผลการตรวจสอบได้ดีเมื่อกลุ่มตัวอย่างมีขนาด 100 คน

Bolt (2002) ได้ศึกษาเปรียบเทียบการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีอัตราส่วนความควรจะเป็น (LRT), วิธี DFIT และวิธี Poly-SIBTEST โดยใช้โมเดล Graded response (GRM; Samejima, 1969) ผลการวิจัยพบว่า วิธี LRT สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 และมีอำนาจการทดสอบดีกว่าวิธี DFIT และวิธี Poly-SIBTEST เมื่อใช้กับกลุ่มตัวอย่างที่มีขนาดเล็ก

Wu and Tsai (2010) ได้ศึกษาเปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีสมการถดถอยโลจิสติก (Logistic regression) การทดสอบอัตราส่วนความควรจะเป็น The likelihood ratio test) และการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบด้วยดีเอฟไอที (DFIT) ภายใต้โมเดล GRM ข้อสอบเป็นแบบ 4 ตัวเลือก ข้อมูลจำลองตามปัจจัยสามปัจจัย คือ ขนาดของกลุ่มตัวอย่าง 2 รูปแบบ ขนาด 300 และ 1,000 คน กลุ่มอ้างอิงและกลุ่มเปรียบเทียบเท่ากัน ปัจจัยที่สอง การแจกแจงความสามารถของกลุ่มย่อย 2 รูปแบบ คือ การแจกแจงความสามารถเท่ากัน ($d_\theta = 0$) และการแจกแจงความสามารถไม่เท่ากัน ($d_\theta = 1$) และปัจจัยที่สาม คือ จำนวนข้อสอบที่เกิดการทำหน้าที่ต่างกัน 4 ขนาด คือ 0%, 7%, 10%, และ 20% รวมทั้งหมด 16 เงื่อนไข การจำลองข้อมูลทำซ้ำ 100 รอบ ผลการศึกษาพบว่า LRT มีอำนาจการทดสอบมากที่สุดในทุกเงื่อนไข DFIT มีอำนาจการทดสอบน้อยกว่า LRT แต่สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในปัจจัยการแจกแจงความสามารถที่แตกต่างกัน และมีความสัมพันธ์กับขนาดของข้อสอบที่มีการทำหน้าที่ต่างกันที่มีจำนวนมาก LogR มีค่าเฉลี่ยของอำนาจการทดสอบต่ำกว่า 0.4 ในทุกเงื่อนไข และจะไวต่อขนาดของข้อสอบที่ทำหน้าที่ต่างกันเป็นจำนวนมาก ๆ LogR มีอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่า LRT และ DFIT

Finch and French (2007) ได้ศึกษาการทำหน้าที่ต่างกันของข้อสอบเปรียบเทียบสี่วิธี กล่าวว่า การทำหน้าที่ต่างกันของข้อสอบ (DIF) ยังคงได้รับความสนใจอย่างต่อเนื่องทั้งในด้านการศึกษาประยุกต์และระเบียบวิธีการ เนื่องจาก DIF สามารถบ่งชี้ความแปรปรวนที่ไม่เกี่ยวข้องซึ่งอาจมีผลต่อคะแนนการสอบ การประเมินและปรับปรุงความถูกต้องของวิธีการตรวจสอบเป็นขั้นตอนสำคัญในการสร้างความถูกต้องของคะแนน วิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป (Uniform) เป็นที่ยอมรับ ในขณะที่แบบอเนกรูป (Nonuniform หรือ Crossing DIF) ยังมีความชัดเจนน้อยกว่า การศึกษาครั้งนี้เป็นการเปรียบเทียบวิธีการสี่แบบ ได้แก่ SIBTEST วิธีถดถอยโลจิสติกส์ IRT-LRT และ CFA ปัจจัยที่มีอิทธิพล ได้แก่ ขนาดกลุ่มตัวอย่าง ความแตกต่างของความสามารถระหว่างกลุ่ม ร้อยละของ DIF และแบบจำลองที่ใช้ในการสร้างข้อมูล ผลการวิจัยชี้ให้เห็นว่าวิธีการทั้งหมดสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 แต่ SIBTEST มีประสิทธิภาพสูงสุด ปัญหาเกี่ยวกับอัตราการตรวจจับพบได้ชัดเจนในโมเดลต้นแบบที่แตกต่างกัน

อาวีพร ปานทอง (2558) ได้ศึกษาการเปรียบเทียบประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบให้คะแนนหลายค่าโดย วิธีทดสอบอัตราส่วนความควรจะเป็น วิธีเบส์เซียน และวิธีโพลี-ซิปเทสท์ ผลการศึกษา พบว่า

1. วิธีทดสอบอัตราส่วนความควรจะเป็น และวิธีเบส์เซียนมีอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ใกล้เคียงกัน ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนหลายค่าที่เป็นรูปแบบเดียวกัน ทั้งอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีโพลี-ซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน เมื่อความยาวของข้อสอบเพิ่มขึ้น วิธีทดสอบอัตราส่วนความควรจะเป็น และวิธีเบส์เซียน สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ดีกว่าวิธีโพลี-ซิปเทสท์ นอกจากนี้เมื่อขนาดตัวอย่างเพิ่มขึ้นมีผลทำให้ทุกวิธีมีอำนาจการทดสอบเพิ่มขึ้นภายใต้ทุกปัจจัย

2. โดยภาพรวมวิธีทดสอบอัตราส่วนความควรจะเป็น และวิธีเบส์เซียน มีอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในขอบเขตที่กำหนด ส่วนวิธีโพลี-ซิปเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าที่กำหนด ผลการศึกษาครั้งนี้เสนอแนะให้ใช้วิธีทดสอบอัตราส่วนความควรจะเป็น และวิธีเบส์เซียน เพื่อทำให้ควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้และมีอำนาจการทดสอบสูง ข้อดีของวิธีทดสอบอัตราส่วนความควรจะเป็น คือ 1) สามารถวัดความคลาดเคลื่อนของโมเดล 2) จำแนกพารามิเตอร์ของข้อสอบในการประมาณค่าแต่ละกลุ่มผู้สอบ 3) ข้อสอบมีลักษณะเป็นการให้คะแนน 2 ค่า แบบเรียงลำดับ หรือนามบัญญัติในการทดสอบ 4) สามารถใช้ทดสอบการทำหน้าที่ต่างกันของข้อสอบได้ทั้ง Uniform และ Nonuniform และ 5) ขนาดของอิทธิพลพารามิเตอร์ของข้อสอบเข้าใจได้ง่าย ข้อเสียของวิธีนี้คือ อยู่บนพื้นฐานของข้อตกลงเบื้องต้นที่เข้มงวด เช่น ฟังก์ชันการตอบสนองข้อสอบต้องเป็นสมการโลจิสติก และการแจกแจงของคุณลักษณะแฝง และคุณลักษณะหลักของผู้สอบ θ ต้องเป็นการแจกแจงปกติ ซึ่งถ้าฟังก์ชันการตอบสนองข้อสอบหรือข้อตกลงเบื้องต้นของคุณลักษณะหลักของผู้สอบ θ ของกลุ่มผู้สอบย่อยกลุ่มใดกลุ่มหนึ่ง หรือทั้งสองกลุ่มแจกแจงไม่ปกติ ความคลาดเคลื่อนประเภทที่ 1 จะเกิด 2-4 เท่าของระดับ α ที่กำหนดไว้ (Bolt, 2002)

นอกจากวิธี IRT-LRT แล้ว ยังมีวิธีหนึ่งที่มีการศึกษากรณีมากกว่าสองกลุ่มกับข้อมูลจำลองแล้วพบว่า สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 และมีอำนาจการทดสอบ เป็นวิธีที่ใช้แนวคิดทางสถิติการวิเคราะห์องค์ประกอบเชิงยืนยันแบบหลายกลุ่ม (MGCFA) จากการศึกษาของ Asparouhov and Muthén (2014) ได้นำเสนอวิธีการใหม่สำหรับการวิเคราะห์องค์ประกอบเชิงยืนยันแบบหลายกลุ่ม (CFA) เรียกว่า วิธี ALIGNMENT วิธีนี้สามารถใช้เพื่อประมาณค่า

ลักษณะเฉพาะของกลุ่ม อันได้แก่ ค่าเฉลี่ยองค์ประกอบและความแปรปรวนโดยไม่ต้องมีการวัดค่าที่แน่นอน และกล่าวว่า จุดแข็งของวิธีนี้คือความสามารถในการประมาณแบบจำลองได้อย่างง่ายดายสำหรับหลาย ๆ กลุ่มเช่นการเปรียบเทียบประเทศ การศึกษามุ่งเน้นไปที่การประยุกต์ใช้ทฤษฎี IRT การตรวจสอบกับข้อมูลเชิงประจักษ์ แบบวัดที่ให้คะแนนแบบ 2 ค่า และศึกษาประสิทธิภาพของวิธีโดยใช้ Monte Carlo และพบว่า เป็นวิธีที่มีคุณภาพ และ Finch (2016) ได้ศึกษาการตรวจสอบการ ทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มที่มากกว่าสองกลุ่ม การศึกษาครั้งนี้เพื่อ ใช้วิธีการจำลองข้อมูลในการเปรียบเทียบวิธีที่ใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มากกว่า 2 กลุ่ม 4 วิธี ได้แก่ The generalized mantel-haenszel test, Generalized logistic regression, Lord's chi-square test, and the multiple group alignment procedure ผลการศึกษาพบว่า The generalized mantel-haenszel test, and the multiple group alignment procedure เป็นวิธีที่ดีที่สุดควบคุมความคลาดเคลื่อนประเภทที่ 1 และอำนาจของการทดสอบ และจากขอบเขตของการวิจัยและข้อเสนอแนะสำหรับการทำวิจัยในอนาคต ที่ว่า การวิจัยครั้งนี้เป็นการศึกษาการตรวจสอบ การทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป (NonUniform) ซึ่งมีมากกว่าสองกลุ่ม การศึกษาในอนาคตควรเปรียบเทียบวิธีการเหล่านี้ สำหรับการทดสอบ Nonuniform DIF จากผลการวิจัยพบว่าวิธี The multiple group alignment เป็นวิธีที่ดีสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 และมีอำนาจของการทดสอบ และเพื่อเป็นการขยายขอบเขตขององค์ความรู้ในเชิงเปรียบเทียบวิธีภายใต้ทฤษฎีการตอบสนองข้อสอบให้กว้างขวางยิ่งขึ้น

นอกจาก 2 วิธีที่กล่าวมาแล้ว การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) เกี่ยวข้องกับเกณฑ์ที่ใช้ในการแบ่งกลุ่มผู้สอบ เกณฑ์ภายนอกที่คาดว่ามีผลต่อผู้สอบได้มีผู้ศึกษาในเชิงการประยุกต์ใช้การวิเคราะห์ด้วยสมการโครงสร้าง (A structural equation modeling analysis: SEM) เรียกว่า โมเดล MIMIC ในขณะที่ Muthén and Muthén (2007) ได้พัฒนาโปรแกรม Mplus เพื่อให้นักวิจัยมีเครื่องมือสำหรับวิเคราะห์ข้อมูลด้วยสถิติวิเคราะห์ขั้นสูง ที่ให้ผลการวิเคราะห์ที่มีความถูกต้องมากกว่าสถิติแบบเดิม โปรแกรม Mplus ได้รับการพัฒนาให้เป็นโปรแกรมที่ใช้งานง่ายและสะดวก และได้รับการปรับปรุงให้ดีขึ้น สามารถวิเคราะห์ข้อมูลได้หลายประเภท วิธี MIMIC เป็นวิธีการทางสถิติที่สามารถวิเคราะห์ข้อมูลได้ด้วยโปรแกรม Mplus นี้ตามทฤษฎีการตอบสนองข้อสอบ

วิธี MIMIC เป็นอีกวิธีหนึ่งที่สามารถใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิธี MIMIC เป็นวิธีหนึ่งที่ใช้สำหรับวิเคราะห์คุณภาพของข้อสอบตามทฤษฎีการตอบสนองข้อสอบเช่นเดียวกับวิธี IRT-LRT วิธี MIMIC เป็นโมเดลลิสเรลที่มีตัวแปรแฝงเพียงตัวเดียว โดยที่ตัวแปร

แฝงนั้นได้รับอิทธิพลจากตัวแปรภายนอกสังเกตได้หลายตัวแปรและส่งอิทธิพลไปยังตัวแปรภายในสังเกตได้หลายตัวแปร กล่าวอีกอย่างหนึ่งคือ เป็นโมเดลของคุณลักษณะแฝงที่มีหลายสาเหตุและวัดได้จากตัวบ่งชี้หลายตัว ลักษณะโมเดลจะเห็นว่าการวัดตัวแปรภายนอกสังเกตได้ต้องมีข้อตกลงข้างต้นว่าไม่มีความคลาดเคลื่อนในการวัด โมเดลมิมิคสามารถใช้สำหรับตรวจสอบความเป็นเอกมิติ (Unidimensionality) ในการวิจัยด้านการวัดผลการศึกษา สามารถวิเคราะห์ค่าพารามิเตอร์คุณลักษณะของข้อสอบ และค่าความสามารถของผู้สอบที่ไม่สามารถสังเกตได้โดยตรง จึงต้องประมาณจากการตอบข้อสอบ โดยประมาณค่าพารามิเตอร์ในทฤษฎีการตอบสนองข้อสอบ (สุพัฒนา หอมบุปผา, 2556, หน้า 7)

วิธี MIMIC มีข้อดีหลายประการเมื่อนำโมเดล MIMIC มาตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) (Muthén et al., 1991 อ้างถึงใน สุพัฒนา หอมบุปผา, 2556, หน้า 7) วิธีนี้แสดงขนาดของการทำหน้าที่ต่างกันของข้อสอบ (DIF) โดยใช้หลักทฤษฎีการตอบสนองข้อสอบ (IRT) ประมาณค่าการทำหน้าที่ต่างกันของข้อสอบ (DIF) จากค่าพารามิเตอร์ตามทฤษฎีการตอบสนองข้อสอบ (IRT) ซึ่งมีประโยชน์ต่อการวิเคราะห์ข้อมูล จากการศึกษาของ Finch (2005) ได้เปรียบเทียบโมเดล MIMIC กับวิธีซิบเทสท์ (SIBTEST) วิธีแมนเทล-แฮนส์เซล (Mantel-haenszel) และวิธีทดสอบอัตราส่วนไลด์ลิฮูดแบบ IRT (IRT Likelihood ratio test) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการจำลองข้อมูลจำนวนผู้สอบและจำนวนข้อสอบด้วยวิธีมอนติคาร์โล ผลการศึกษา พบว่า โมเดล MIMIC ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ดีในกรณีที่ข้อสอบมีจำนวน 20 ข้อ แบบ 3 พารามิเตอร์ และข้อสอบ 50 ข้อ แบบ 2 พารามิเตอร์

สำหรับการศึกษาการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี MIMIC ในประเทศไทยนั้นได้มีนักวิจัยได้ทำการศึกษาไว้ดังนี้

งานวิจัยของ สุพัฒนา หอมบุปผา (2556) ได้ศึกษา การเปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี HGLM วิธี MIMIC และวิธี BAYESIAN ผลการศึกษาพบว่า

1. ผลการวิเคราะห์ค่าพารามิเตอร์ความยากของข้อสอบ ค่าพารามิเตอร์ความสามารถของผู้สอบ และผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ในวิชาภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์ ด้วยวิธี HGLM-2L วิธี MIMIC และวิธี BAYESIAN พบว่า มีความสัมพันธ์กันในระดับสูงมาก อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

2. วิธีตรวจสอบที่พบการทำหน้าที่ต่างกันของข้อสอบมากที่สุด คือ วิธี HGLM-2L ส่วนวิธีที่ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบน้อยที่สุด คือ วิธี MIMIC

3. ผลการศึกษาลักษณะของข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ (DIF) วิชาภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์ ลักษณะของข้อสอบที่เกิดจากการทำหน้าที่ต่างกัน เมื่อ

จำแนกตามเพศ ส่วนใหญ่ข้อสอบที่เกิดการทำหน้าที่ต่างกันของข้อสอบ จะมีคำศัพท์ที่เกี่ยวข้องกับเพศนั้น จึงทำให้ข้อสอบเข้าข้างเพศนั้น และอาจเป็นเพราะความสามารถที่แตกต่างระหว่างเพศชายและเพศหญิงที่มีลักษณะความสามารถ ความถนัด และความสนใจในเรื่องนั้น ๆ ต่างกัน เมื่อจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน ส่วนใหญ่สาเหตุที่ทำให้ข้อสอบเกิดการทำหน้าที่ต่างกันอาจเป็นเพราะประสบการณ์ ความคุ้นเคยเกี่ยวกับเรื่องนั้น สภาพแวดล้อมและการฝึกปฏิบัติที่แตกต่างกันระหว่างนักเรียนในเขตกรุงเทพมหานครและปริมณฑลนักเรียนนอกเขตกรุงเทพมหานครและปริมณฑล และ สุพัฒนา หอมบุปผา (2556) ได้ให้ข้อเสนอแนะไว้ว่า การวิเคราะห์การประมาณค่าพารามิเตอร์ความยากของข้อสอบ พารามิเตอร์ความสามารถของผู้สอบและการทำหน้าที่ต่างกันของข้อสอบ ควรเลือกการวิเคราะห์ด้วยวิธี MIMIC จากโปรแกรม Mplus เพราะมีกระบวนการในการวิเคราะห์ ที่ง่าย ใช้งานสะดวก และไม่ซับซ้อน โดยการเขียนคำสั่งและวิเคราะห์ในขั้นตอนเดียว ซึ่งผลการวิเคราะห์ได้ค่าที่ใกล้เคียงและสอดคล้องกันกับวิธี HGLM-2L และวิธี BAYESIAN

จากอรรถประโยชน์ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่นักวัดผลและนักวิจัยหลาย ๆ คนได้ศึกษาและนำเสนอแนวทางมานั้น ผู้วิจัยจึงสนใจนำมาศึกษาเพื่อขยายแนวคิดต่อจากการแบ่งกลุ่มที่มากกว่าสองกลุ่มหรือหลายกลุ่มโดยเพิ่มขั้นตอนการเปรียบเทียบวิธีที่อยู่บนพื้นฐานทฤษฎีเดียวกัน คือ ทฤษฎีการตอบสนองข้อสอบ (IRT) ซึ่งเป็นที่ทราบข้างต้นแล้วว่าเป็นทฤษฎีที่มีความแกร่ง สามารถแก้ปัญหาของทฤษฎีการทดสอบแบบดั้งเดิม (CTT) ได้ การวิเคราะห์ข้อสอบด้วยทฤษฎีการตอบสนองข้อสอบสามารถให้ทั้งสารสนเทศค่าพารามิเตอร์รายข้อ และพารามิเตอร์รายบุคคล รวมทั้งความสามารถในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในการวิจัยนี้จึงมุ่งศึกษาการวิเคราะห์คุณภาพของข้อสอบตามทฤษฎีการตอบสนองข้อสอบ (IRT) 2-Parameter logistic measurement model (2PL) ทั้ง 3 วิธี คือ วิธี IRT LRT วิธี MGCFA ALIGNMENT และวิธี MIMIC สามารถวิเคราะห์ เพื่อให้สามารถเปรียบเทียบผลบนพื้นที่ที่เหมือนกันได้สมเหตุสมผล ประกอบกับทั้ง 3 วิธี สามารถตรวจสอบการทำหน้าที่ต่างกันได้ทั้ง 2 รูปแบบ คือ แบบเอกรูป (Uniform DIF) และแบบอเนกรูป (Nonuniform DIF) ได้เช่นเดียวกันภายใต้ปัจจัยที่ต่างกัน 2 ปัจจัย ได้แก่ สถานที่ตั้งทางภูมิศาสตร์ และขนาดโรงเรียน ศึกษากับข้อมูลคะแนนผลการสอบรายข้อจากแบบทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ระดับชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 วิชาคณิตศาสตร์ ภาษาไทย และวิชาวิทยาศาสตร์ ดำเนินการสอบโดยสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) เป็นการจัดสอบที่มีแนวทางให้การดำเนินการสอบเป็นไปในทิศทางและมาตรฐานเดียวกัน ผลการประเมินมีความน่าเชื่อถือและนำไปใช้ในการปรับปรุงและพัฒนาได้อย่างถูกต้องตามความสามารถของ

นักเรียน ทั้ง 3 วิชา ที่ผู้วิจัยได้ศึกษาครั้งนี้ เนื่องจากเป็นการประเมินเพื่อให้ต่อเนื่องกับการประเมินคุณภาพการศึกษาระดับชาติ (NT) ชั้นประถมศึกษาปีที่ 3 และการประเมินคุณภาพการศึกษาระดับชาติ(O-NET) ชั้นประถมศึกษาปีที่ 6 และเป็นกลุ่มวิชาที่เป็นวิชาที่เป็นแกนหลักตามโครงสร้างหลักสูตรแกนกลางการศึกษาขั้นพื้นฐานพุทธศักราช 2551 รวมไปถึงการประเมินระดับนานาชาติ PISA เพื่อเป็นแนวทางสำหรับผู้เกี่ยวข้องได้นำสารสนเทศจากผลการศึกษาดังกล่าวไปเป็นประโยชน์ เป็นแนวทางสำหรับการวางแผนพัฒนาการสร้างข้อสอบหรือแบบสอบให้มีความเป็นมาตรฐาน ส่งผลให้เกิดความยุติธรรมต่อผู้สอบ นำไปสู่การออกแบบการวัดการประเมินผลและการตัดสินผลตามความเหมาะสมกับสภาพการสอบและการใช้ผลการสอบในปัจจุบันปลายทางเพื่อพัฒนาคุณภาพของเครื่องมือที่ใช้ทางการศึกษาและทางจิตวิทยาต่อไป

คำถามการวิจัย

1. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT วิธี MGCFA ALIGNMENT และวิธี MIMIC กับข้อมูลเชิงประจักษ์ภายใต้ปัจจัยที่ต่างกัน ได้แก่ สถานที่ตั้งทางภูมิศาสตร์ และขนาดโรงเรียน พบข้อสอบที่ทำหน้าที่ต่างกันหรือไม่และมีความสอดคล้องหรือแตกต่างกันอย่างไร

2. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT วิธี MGCFA ALIGNMENT และวิธี MIMIC กับข้อมูลเชิงประจักษ์ภายใต้ปัจจัยที่ต่างกัน ได้แก่ สถานที่ตั้งทางภูมิศาสตร์ และขนาดโรงเรียน มีความสอดคล้องหรือแตกต่างกันหรือไม่

วัตถุประสงค์ของการวิจัย

1. เพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT วิธี MGCFA ALIGNMENT และวิธี MIMIC ภายใต้ปัจจัยที่ต่างกัน 2 ปัจจัย ได้แก่ สถานที่ตั้งทางภูมิศาสตร์ และขนาดโรงเรียน

2. เพื่อเปรียบเทียบผลของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT วิธี MGCFA ALIGNMENT และวิธี MIMIC ภายใต้ปัจจัยที่ต่างกัน 2 ปัจจัย ได้แก่ สถานที่ตั้งทางภูมิศาสตร์ และขนาดโรงเรียน

สมมติฐานของการวิจัย

ผู้วิจัยได้ตั้งสมมติฐานในการวิจัยครั้งนี้ไว้ดังนี้

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT วิธี MGCFA-ALIGNMENT และวิธี MIMIC มีความสอดคล้องกันภายใต้เงื่อนไขเดียวกัน

ขอบเขตของการวิจัย

การศึกษาครั้งนี้ได้กำหนดขอบเขตของการศึกษา ไว้ดังนี้

1. ขอบเขตด้านประชากรและกลุ่มตัวอย่าง

ประชากรที่ใช้ในการศึกษาครั้งนี้ เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ที่เข้ารับการประเมินคุณภาพการศึกษาระดับชาติ (O-NET) ปีการศึกษา 2559 ประกอบด้วยวิชา ภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์ ดำเนินการสอบโดยสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน)

กลุ่มตัวอย่างที่ใช้ในการศึกษาครั้งนี้ เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ที่เข้ารับการประเมินคุณภาพการศึกษาระดับชาติ (O-NET) ปีการศึกษา 2559 ประกอบด้วยวิชา ภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์ ดำเนินการสอบโดยสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) จำนวน 8,000 คน

2. ขอบเขตด้านเนื้อหา

การวิจัยครั้งนี้เป็นการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT วิธี MGCFA ALIGNMENT และวิธี MIMIC โดยศึกษาปัจจัยที่ต่างกัน 2 ปัจจัย ได้แก่ สถานที่ตั้งทางภูมิศาสตร์ และขนาดโรงเรียน

3. ตัวแปรที่ศึกษา ประกอบด้วย

3.1 ตัวแปรอิสระ แบ่งเป็น 3 กลุ่มตัวแปร

3.1.1 กลุ่มที่ใช้เป็นเกณฑ์ในการแบ่งกลุ่มได้แก่ 1) สถานที่ตั้งทางภูมิศาสตร์ และ 2) ขนาดของโรงเรียน

3.1.2 กลุ่มวิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ได้แก่ 1) วิธี IRT LRT 2) วิธี MGCFA-ALIGNMENT และ 3) วิธี MIMIC

3.1.3 กลุ่มข้อสอบที่ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ได้แก่ 1) วิชาคณิตศาสตร์ 2) วิชาภาษาไทย และ 3) วิชาวิทยาศาสตร์

3.2 ตัวแปรตาม ได้แก่ ข้อสอบที่ตรวจพบว่า ทำหน้าที่ต่างกัน (DIF)

4. การวิจัยครั้งนี้มุ่งตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT วิธี MGCFA ALIGNMENT และวิธี MIMIC โดยใช้โปรแกรม IRT PRO และโปรแกรม MPlus ในการวิเคราะห์ข้อมูล

นิยามศัพท์เฉพาะ

ข้อสอบที่มีรูปแบบการตรวจให้คะแนนแบบทวิวิภาค (Dichotomous scored items) หมายถึง การให้คะแนนการตอบข้อสอบแบบมี 2 ค่า คือ ตอบถูก ได้ 1 คะแนน, ตอบผิด ได้ 0 คะแนน

การทำหน้าที่ต่างกันของข้อสอบ (Differential item functioning: DIF) หมายถึง การที่ข้อสอบวิชาคณิตศาสตร์ วิชาภาษาไทย และวิชาวิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ในการประเมินคุณภาพการศึกษาระดับชาติ (O-NET) ปีการศึกษา 2559 ทำให้ผู้สอบจากต่างกลุ่มกัน (สถานที่ตั้งทางภูมิศาสตร์ ได้แก่ กลุ่มภาคกลาง กับ กลุ่มภาคเหนือ ภาคตะวันออกเฉียงเหนือ และภาคใต้,ขนาดโรงเรียนของผู้สอบ ได้แก่ กลุ่มโรงเรียนขนาดใหญ่พิเศษกับกลุ่มโรงเรียนเล็ก โรงเรียนขนาดกลาง และโรงเรียนขนาดใหญ่) ที่วัดคุณลักษณะผู้สอบที่มีความสามารถเท่ากันมาจากต่างกลุ่มกัน แต่มีโอกาสในการตอบข้อสอบข้อที่ตรวจสอบได้ถูกต้องต่างกัน

กลุ่มอ้างอิง (Reference group: R) หมายถึง กลุ่มผู้สอบวิชาคณิตศาสตร์ วิชาภาษาไทย และวิชาวิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ในการประเมินคุณภาพการศึกษาระดับชาติ (O-NET) ปีการศึกษา 2559 ที่คาดว่าจะได้เปรียบในการตอบข้อสอบที่ทำหน้าที่ต่างกัน คือ เป็นกลุ่มที่มีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องมากกว่าผู้สอบอีกกลุ่มหนึ่ง แม้ว่าผู้สอบทั้งสองกลุ่มจะมีระดับความสามารถเท่ากัน ในการวิจัยครั้งนี้ ตัวแปรสถานที่ตั้งทางภูมิศาสตร์ ได้แก่ กลุ่มภาคกลาง ตัวแปรขนาดโรงเรียนได้แก่ กลุ่มโรงเรียนขนาดใหญ่พิเศษ

กลุ่มเปรียบเทียบ (Focal group: F) หมายถึง กลุ่มผู้สอบวิชาคณิตศาสตร์ วิชาภาษาไทย และวิชาวิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ในการประเมินคุณภาพการศึกษาระดับชาติ (O-NET) ปีการศึกษา 2559 ที่คาดว่าจะเสียเปรียบในการตอบข้อสอบที่ทำหน้าที่ต่างกัน คือ เป็นกลุ่มที่มีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องน้อยกว่าผู้สอบอีกกลุ่มหนึ่ง แม้ว่าผู้สอบทั้งสองกลุ่มจะมีระดับความสามารถเท่ากัน ในการวิจัยครั้งนี้ ตัวแปรสถานที่ตั้งทางภูมิศาสตร์ของผู้สอบ ได้แก่ กลุ่มภาคเหนือ ภาคตะวันออกเฉียงเหนือ และภาคใต้ ตัวแปรขนาดโรงเรียนของผู้สอบ ได้แก่ กลุ่มโรงเรียนเล็ก โรงเรียนขนาดกลาง และโรงเรียนขนาดใหญ่

ประเภทของการทำหน้าที่ต่างกันของข้อสอบ หมายถึง รูปแบบการเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มผู้สอบ สามารถแบ่งได้เป็น 2 ประเภท ดังนี้

1. ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป (Uniform DIF) หมายถึง ข้อสอบที่ทำให้ผู้สอบกลุ่มหนึ่งมีโอกาสในการตอบข้อสอบถูกมากกว่าผู้สอบอีกกลุ่มอย่างสม่ำเสมอในทุกระดับความสามารถ และไม่มีปฏิสัมพันธ์ระหว่างการเป็นสมาชิกของกลุ่มกับระดับความสามารถของผู้สอบ

2. ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป (Nonuniform DIF) หมายถึง ข้อสอบที่ทำให้โอกาสในการตอบข้อสอบถูกของผู้สอบต่างกลุ่มกันแตกต่างกันอย่างไม่สม่ำเสมอในทุกระดับความสามารถ และรูปร่างของโค้งคุณลักษณะข้อสอบของผู้สอบ 2 กลุ่ม มีปฏิสัมพันธ์ร่วมกันระหว่างการเป็นสมาชิกของกลุ่มกับระดับความสามารถของผู้สอบ

กลุ่มที่มากกว่า 2 กลุ่ม หมายถึง จำนวนกลุ่มย่อยของผู้สอบที่จำแนกตามปัจจัยที่สามารถแบ่งเป็นกลุ่มย่อยได้มากกว่า 2 กลุ่ม ปัจจัยที่ศึกษาได้แก่ สถานที่ตั้งทางภูมิศาสตร์ และขนาดโรงเรียน

สถานที่ตั้งทางภูมิศาสตร์ หมายถึง ที่ตั้งทางภูมิศาสตร์ของโรงเรียนของผู้สอบ แบ่งเป็น 4 ภูมิภาค ได้แก่ เหนือ กลาง ตะวันออกเฉียงเหนือ และใต้

ขนาดของโรงเรียน หมายถึง ขนาดโรงเรียนของผู้สอบ โดยการแบ่งขนาดของโรงเรียนใช้จำนวนนักเรียนเป็นเกณฑ์ในการแบ่ง แบ่งได้ 4 ขนาด ได้แก่ ขนาดใหญ่พิเศษ ขนาดใหญ่ ขนาดกลาง และขนาดเล็ก

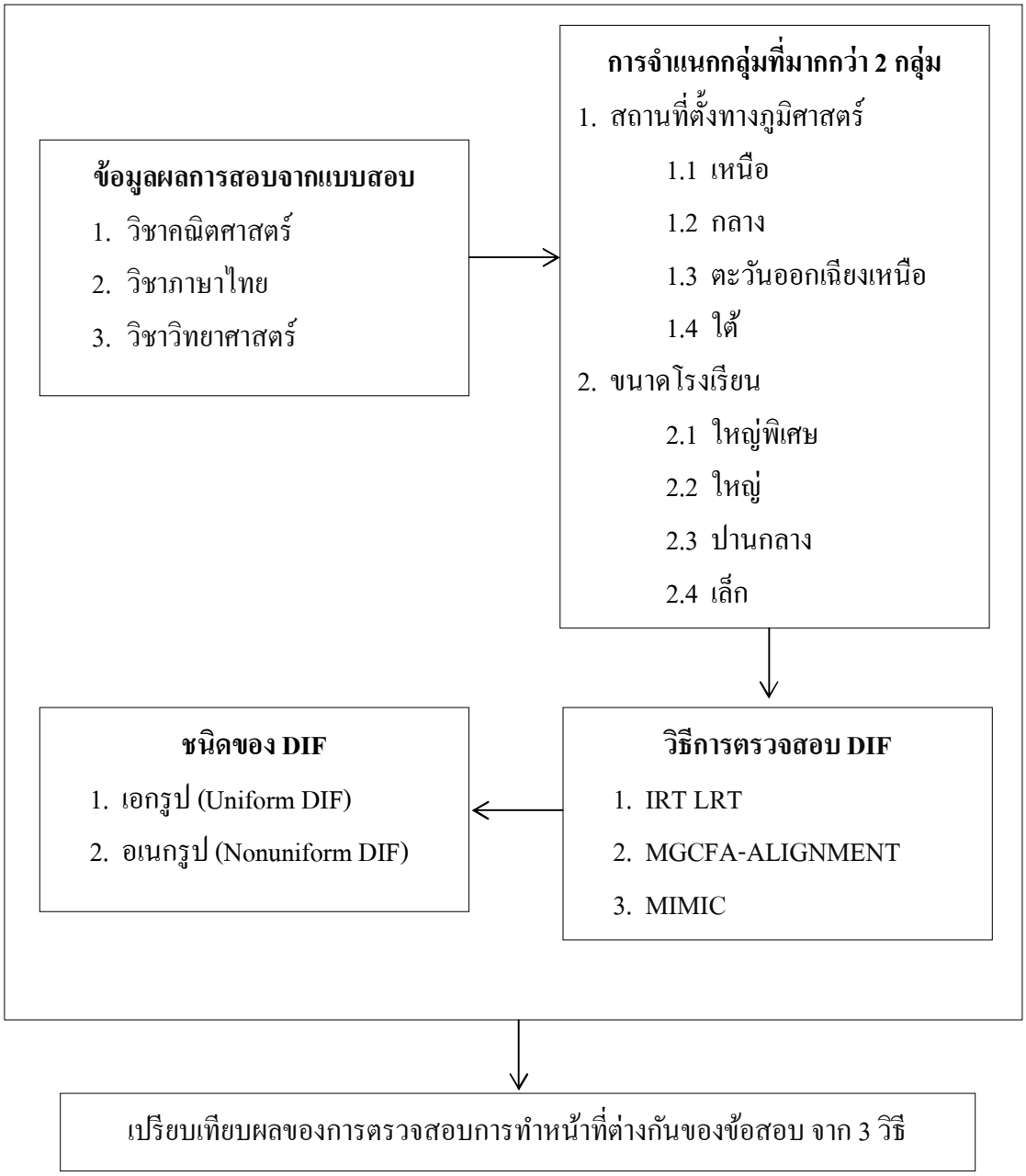
วิธี IRT LRT (IRT Likelihood ratio test) หมายถึง วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบความแตกต่างของค่าพารามิเตอร์ของโมเดลผลการตอบข้อสอบระหว่างผู้สอบกลุ่มย่อยครั้งละ 2 กลุ่ม (Contras group) หากทดสอบว่าคู่ใดแตกต่างอย่างมีนัยสำคัญบ่งชี้ว่าข้อสอบข้อนั้นทำหน้าที่ต่างกัน วิธีทดสอบอัตราส่วนความควรจะเป็นตามทฤษฎีการตอบสนองข้อสอบในรูปของล็อกเชิงเส้น (Loglinear: LR) เป็นการประมาณค่าพารามิเตอร์ด้วยวิธีความน่าจะเป็นสูงสุด (Maximum likelihood: ML) ใช้สถิติทดสอบไคว์-สแควร์ χ^2 ในการทดสอบนัยสำคัญ

วิธี MGCFA-ALIGNMENT หมายถึง วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแนวคิดการวิเคราะห์องค์ประกอบเชิงยืนยันแบบหลายกลุ่ม (MGCFA) เป็นการกำหนดโมเดลการเชื่อมโยงค่าพารามิเตอร์ของข้อสอบทั้งค่าความยากและค่าอำนาจจำแนกของข้อสอบ โดยค่าความยากของข้อสอบจาก IRT คือค่า Item thresholds และค่าอำนาจจำแนกของข้อสอบจาก IRT จะเชื่อมโยงกับค่า Factor loadings จากนั้นเมื่อมีการจัดเรียงรูปแบบเสร็จสมบูรณ์ จะมีการทดสอบการแปรเปลี่ยนของค่าพารามิเตอร์ข้ามกลุ่มทั้งค่าความยากและค่าอำนาจจำแนกของข้อสอบหากข้อใดมีการแปรเปลี่ยบ่งชี้ว่าข้อสอบนั้นทำหน้าที่ต่างกัน

วิธี MIMIC (Multiple indicators and multiple causes) หมายถึง วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ประยุกต์โมเดล MIMIC ที่มีตัวแปรแฝงเพียงตัวแปรเดียวโดยที่ตัวแปรแฝงนั้นได้รับอิทธิพลจากตัวแปรภายนอกสังเกตได้หลายตัวแปรและส่งอิทธิพลไปยังตัวแปรภายในสังเกตได้หลายตัวแปร โดยพิจารณาว่าถ้าข้อสอบได้รับอิทธิพลอย่างมีนัยสำคัญจากตัวแปรสาเหตุหรือตัวแปรกลุ่มแสดงให้เห็นว่าข้อสอบขึ้นอยู่กับตัวแปรสาเหตุหรือตัวแปรกลุ่ม โดยที่ไม่ได้อธิบายด้วยตัวแปรแฝงอย่างเดียว แสดงว่าข้อสอบข้อนั้นทำหน้าที่ต่างกันแบบเอกรูป และหากตัวแปรกลุ่มมีปฏิสัมพันธ์กับตัวแปรแฝงแล้วมีอิทธิพลต่อข้อสอบโดยตรง แสดงว่า ข้อสอบข้อนั้นทำหน้าที่ต่างกันแบบอเนกรูป

กรอบแนวคิดในการวิจัย

จากการศึกษาแนวคิดทฤษฎีและงานวิจัยต่าง ๆ ที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบ กรณีกลุ่มที่นำมาเปรียบเทียบมีมากกว่า 2 กลุ่ม แนวคิดที่สำคัญได้แก่ แนวคิดทฤษฎีตอบสนองข้อสอบ แนวคิดการทำหน้าที่ต่างกันของข้อสอบ วิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 3 วิธี ได้แก่ วิธี IRT-LRT (Thissen, Steinberg &Wainer, 1988), วิธี Alignment (Asparouhov & Muthén, 2014), MIMIC (Múthen, 1985; 1988; 1989; Múthen, Kao & Burstein, 1991) ผู้วิจัยได้กำหนดเป็นกรอบแนวคิดในการวิจัยครั้งนี้ ได้ดังนี้



ภาพที่ 1-1 กรอบแนวคิดในการวิจัย

ประโยชน์ที่ได้รับจากการวิจัย

การศึกษาครั้งนี้เป็นการเปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มที่มากกว่าสองกลุ่ม ด้วยวิธี IRT LRT วิธี MGCFA-ALIGNMENT และวิธี MIMIC ผลจากการศึกษานี้จะให้ประโยชน์ ดังนี้

1. ได้ทราบถึงวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่เหมาะสมสำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT วิธี MGCFA-ALIGNMENT และวิธี MIMIC ภายใต้ปัจจัยที่แปรเปลี่ยน 2 ปัจจัย

2. เป็นแนวทางในการเลือกใช้วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบให้เหมาะสมสำหรับกลุ่มที่มากกว่าสองกลุ่มโดยคำนึงถึงตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

3. เป็นแนวทางในการระบุลักษณะของข้อสอบที่ทำหน้าที่ต่างกันและเป็นการปรับปรุงข้อสอบให้มีความยุติธรรมสำหรับผู้สอบเมื่อมีการพิจารณาหรือจำแนกกลุ่มผู้สอบตามปัจจัยอื่น ๆ

4. ผลการวิจัยจะเป็นประโยชน์ในเชิงวิชาการในด้านการประยุกต์วิธีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) ตามทฤษฎีการตอบสนองข้อสอบ ซึ่งจะเป็นแนวทางในการนำไปใช้วิเคราะห์ข้อสอบระดับชาติ ที่จะต้องมีการตรวจสอบคุณภาพของเครื่องมือในทุกครั้งที่มีการสอบ เพื่อก่อให้เกิดความยุติธรรมต่อผู้สอบหรือผู้ที่มีส่วนได้ส่วนเสียกับการใช้ผลจากการทดสอบต่อไป

5. เป็นแนวทางในการศึกษาเรื่องวิธีการประมาณค่าพารามิเตอร์ และการทำหน้าที่ต่างกันของข้อสอบ โดยประยุกต์ใช้โปรแกรม Mplus และโปรแกรม IRTPRO

6. ทำให้ทราบลักษณะข้อสอบที่ทำหน้าที่ต่างกันสำหรับกลุ่มผู้สอบจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ และขนาดของโรงเรียน เพื่อให้หน่วยงาน บุคคลที่เกี่ยวข้องกับการสร้างแบบทดสอบระดับชาติ นำสารสนเทศที่ได้จากงานวิจัยไปใช้พัฒนาแบบทดสอบให้มีคุณภาพสมบูรณ์ต่อไป

บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

การวิจัยครั้งนี้มีวัตถุประสงค์ของการวิจัยเพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT วิธี MGCFA ALIGNMENT และวิธี MIMIC ภายใต้ปัจจัยที่ต่างกัน 2 ปัจจัย ได้แก่ ภูมิภาค และขนาดโรงเรียน และเพื่อเปรียบเทียบผลของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT วิธี MGCFA ALIGNMENT และวิธี MIMIC ภายใต้ปัจจัยที่ต่างกัน 2 ปัจจัย ได้แก่ ภูมิภาค และขนาดโรงเรียน ผู้วิจัยได้ศึกษากรอบแนวคิด ทฤษฎี หลักการ รูปแบบและวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ตลอดจนงานวิจัยที่เกี่ยวข้อง สามารถสรุปและนำเสนอไว้ ดังนี้

ตอนที่ 1 มโนทัศน์เกี่ยวกับทฤษฎีการตอบสนองข้อสอบ (IRT)

ตอนที่ 2 มโนทัศน์เกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบ (DIF)

ตอนที่ 3 การทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET)

ตอนที่ 4 งานวิจัยที่เกี่ยวข้อง

ตอนที่ 1 มโนทัศน์เกี่ยวกับทฤษฎีการตอบสนองข้อสอบ (IRT)

ด้วยข้อจำกัดที่สำคัญของทฤษฎีการทดสอบแบบดั้งเดิม (CTT) ได้แก่ การที่พารามิเตอร์ของข้อสอบแปรผันตามกลุ่มผู้สอบ และคะแนนที่สังเกตได้หรือค่าประมาณความสามารถของผู้สอบไม่เป็นอิสระ หรือขึ้นอยู่กับข้อสอบและแบบสอบที่นำมาใช้ นักทฤษฎีการทดสอบจึงได้พัฒนาระบบการวัดแนวใหม่ เรียกกันว่า ทฤษฎีการตอบสนองข้อสอบ (Item response theory: IRT) หรือทฤษฎีคุณลักษณะแฝง (Latent trait theory: IRT) เป็นทฤษฎีการวัดที่อธิบายความสัมพันธ์ระหว่างความสามารถที่มีอยู่ภายในของบุคคล (Latent trait or ability) กับผลการตอบข้อสอบหรือข้อคำถามโดยใช้โค้งลักษณะข้อสอบ (Item characteristic curve: ICC) ซึ่งมีการกำหนดลักษณะข้อสอบด้วยพารามิเตอร์ความยาก (b) อำนาจจำแนก (a) และโอกาสการเดาข้อสอบถูก (ศิริชัย กาญจนวาสี, 2545, หน้า 45; Linda & James, 2008, pp. 339-340)

พื้นฐานความคิดของทฤษฎีการตอบข้อสอบ

ทฤษฎีการตอบสนองข้อสอบตั้งอยู่บนพื้นฐานความคิด 2 ประการ คือ

1. การตอบข้อสอบของผู้สอบ สามารถพยากรณ์หรืออธิบายคุณลักษณะภายใน (Traits หรือ Latent traits หรือ Abilities) ได้

2. ความสัมพันธ์ระหว่างผลการตอบข้อสอบของผู้สอบและคุณลักษณะภายในอธิบายได้โดยโค้งลักษณะข้อสอบ (Item characteristic curve: ICC) เป็นฟังก์ชันทางคณิตศาสตร์ คือ ฟังก์ชันโลจิส (Logistic function) หรือใกล้เคียงกับฟังก์ชันปกติสะสม ซึ่งโค้ง ICC นี้มีลักษณะเฉพาะ คือ เมื่อผู้สอบมีความสามารถสูง ความน่าจะเป็นในการตอบข้อสอบถูกก็มากขึ้นด้วย

ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบมุ่งอธิบายความสัมพันธ์ระหว่างความสามารถที่แท้จริงของผู้สอบ (θ) กับพฤติกรรมการตอบสนองข้อสอบในแต่ละข้อว่ามีความน่าจะเป็นในการตอบข้อสอบถูกได้มากน้อยเพียงใด $P_i(\theta)$ ด้วยโค้งคุณลักษณะของข้อสอบ (Item characteristic curve: ICC) ซึ่งมีลักษณะเป็นฟังก์ชันคณิตศาสตร์ โดยมีข้อตกลงเบื้องต้นที่สำคัญ ได้แก่ (DeMars, 2010; Hambleton, Swaminathan & Roger, 1991; Fukuhara & Kamata, 2011; Le, 2013; ชนะศึก นิชานนท์, 2553; อริสพา เตหลิ่ม, 2559, หน้า 33-38)

ข้อที่ 1 ความเป็นเอกมิติ (Unidimensional)

ความเป็นเอกมิติเป็นข้อตกลงเบื้องต้นที่ใช้กันทั่วไป สำหรับ IRT หมายถึง แบบสอบต้องวัดความสามารถด้านเดียว (θ) สำหรับผู้สอบแต่ละคน และปัจจัยอื่นที่มีผลต่อการตอบสนองข้อสอบจะถูกกำหนด (Treat) ให้มีความคลาดเคลื่อนแบบสุ่ม (Random error) ซึ่งการละเลยข้อตกลงนี้จะนำไปสู่ความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ หรือ Standard error อย่างไรก็ตาม ข้อตกลงนี้สามารถผ่อนปรนได้โดยสามารถอธิบายแบบสอบที่มีการใช้ความสามารถพหุมิติ (Multidimensionality) ได้ เช่น Multidimensional rash model เป็นต้น

ข้อที่ 2 ความเป็นอิสระในการตอบข้อสอบ (Local independence)

ความเป็นอิสระในการตอบข้อสอบมีความเกี่ยวข้องและเชื่อมโยงมาจากความเป็นเอกมิติ หมายถึง เมื่อควบคุมความสามารถที่มีผลต่อการตอบข้อสอบให้คงที่แล้ว ความน่าจะเป็นในการตอบข้อสอบถูกในแต่ละข้อมีความเป็นอิสระต่อกัน หรือกล่าวอีกนัยหนึ่ง เมื่อควบคุมอิทธิพลของที่มีผลต่อการตอบข้อสอบให้คงที่แล้ว ผลการตอบข้อสอบรายข้อไม่มีความสัมพันธ์กัน โดยความเป็นอิสระสามารถจำแนกได้เป็น

2.1 ความเป็นอิสระระหว่างข้อสอบ หมายถึง ข้อสอบแต่ละข้อเป็นอิสระจากกัน กล่าวคือ การตอบข้อสอบข้อหนึ่งไม่มีผลกระทบต่อการตอบข้อสอบข้ออื่น ๆ ในแบบสอบฉบับนั้น

2.2 ความเป็นอิสระระหว่างผู้สอบ หมายถึง ผู้สอบแต่ละคนตอบข้อสอบแต่ละข้ออย่างเป็นอิสระกัน

ความเป็นอิสระในการตอบข้อสอบเป็นข้อตกลงเบื้องต้นที่สำคัญของทฤษฎีการวัด ซึ่งหากละเลยข้อตกลงนี้จะทำให้เกิดผลต่าง ๆ ดังนี้

1. ในมุมของทฤษฎีการวัดแบบดั้งเดิม (Classical test theory: CTT) จะทำให้ประมาณค่าความคาดเคลื่อนมาตรฐานของการวัด (Standard error of measurement) ได้ต่ำกว่าเกินจริง ส่งผลให้ความเที่ยง (Reliability) สูงเกินจริง

2. ในมุมของทฤษฎีการสอบสนองข้อสอบ (IRT) การประมาณค่าสารสนเทศ (Information) จะสูงเกินจริง นั่นหมายถึง ค่าความคาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถจะต่ำเกินจริง นอกจากนี้ อำนาจของอำนาจจำแนก (Item discrimination power) อาจจะลำเอียง ซึ่งทำให้การประมาณค่าความสามารถก็อาจจะลำเอียงไปด้วย เนื่องจากขึ้นอยู่กับอำนาจจำแนก (กรณีฟังก์ชัน 2 พารามิเตอร์และ 3 พารามิเตอร์)

3. โค้งคุณลักษณะของข้อสอบ (Item characteristic curve: ICC) ทฤษฎีการตอบสนองข้อสอบให้ความสำคัญกับฟังก์ชันคุณลักษณะของข้อสอบหรือโค้งคุณลักษณะข้อสอบ โดยเป็นฟังก์ชันทางคณิตศาสตร์ที่แสดงความสัมพันธ์ระหว่างโอกาสในการตอบข้อสอบถูกกับระดับความสามารถของผู้สอบมีลักษณะเป็นฟังก์ชันทางคณิตศาสตร์ เรียกว่า ฟังก์ชันโลจิสติก (Logistic function) หรือใกล้เคียงกับฟังก์ชันปกติสะสม (Normal ogive function) ดังนั้น จะเห็นได้ว่าโอกาสที่ผู้สอบตอบข้อสอบถูกจึงขึ้นอยู่กับโค้งลักษณะข้อสอบ ซึ่งเป็นอิสระจากการกระจายของความสามารถของผู้สอบหรือโอกาสที่ผู้ตอบตอบข้อสอบถูกไม่ขึ้นอยู่กับจำนวนของผู้สอบที่มีความสามารถเหมือนกัน ลักษณะของโค้งคุณลักษณะข้อสอบในแต่ละข้อมีคุณสมบัติไม่แปรเปลี่ยนไปตามกลุ่มของผู้สอบ จึงทำให้โอกาสในการตอบข้อสอบถูกในแต่ละข้อไม่แปรเปลี่ยน ซึ่งไม่เหมือนกับการวิเคราะห์ด้วยทฤษฎีการทดสอบแบบดั้งเดิม (CTT)

4. การทดสอบที่ไม่แข่งขันด้านเวลา (Nonspeed test administration)

ทฤษฎีการตอบสนองข้อสอบมุ่งอธิบายความสัมพันธ์ระหว่างความสามารถที่แท้จริงของผู้สอบกับพฤติกรรมการตอบสนองข้อสอบในแต่ละข้อ ดังนั้น ความสามารถที่แท้จริงของผู้สอบจึงเป็นสิ่งสำคัญสำหรับการวิเคราะห์ตามแนวทฤษฎีนี้ โดยแบบสอบประเภทใช้ความเร็วในการตอบ (Speed test) ผู้สอบต้องใช้ความสามารถอย่างน้อยสองมิติ คือ ความเร็วในการตอบ (Responsespeed) และความสามารถที่แบบสอบต้องการวัด ดังนั้น เพื่อให้สอดคล้องกับข้อตกลงเบื้องต้นเกี่ยวกับความเป็นมิติเดียว แบบสอบที่นำมาใช้จึงเป็นแบบสอบที่ไม่ใช้ความเร็วในการตอบ (Nonspeed test) ทำให้ทุกคนมีเวลาทำข้อสอบทุกข้อ และสามารถใช้คะแนนรวมจากการทำแบบทดสอบร่วมกับลักษณะข้อสอบเป็นตัวประมาณค่าความสามารถที่แท้จริงของผู้ตอบข้อสอบโดยไม่ต้องพิจารณาเรื่องความเร็ว

การตรวจสอบข้อตกลงเบื้องต้นข้อนี้มีวิธีการตรวจสอบได้หลายวิธี ดังนี้

1. การพิจารณาจากสัดส่วนหรือร้อยละของจำนวนผู้สอบที่สามารถทำข้อสอบได้ครบทุกข้อ โดยผู้สอบส่วนใหญ่ต้องสามารถตอบข้อสอบได้ครบหรือเกือบครบทุกข้อ

2. การเปรียบเทียบความแปรปรวนของจำนวนข้อที่เว้นกับความแปรปรวนของจำนวนข้อที่ตอบผิด ถ้าอัตราส่วนของความแปรปรวนเข้าใกล้ศูนย์ แสดงว่าการดำเนินการจัดการสอบไม่เน้นการแข่งขันด้านเวลา

โมเดลที่ใช้วิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบเป็นทฤษฎีการวัดที่เกิดจากการพยายามค้นหาวิธีแก้ไขข้อบกพร่องของทฤษฎีการทดสอบแบบดั้งเดิม (Classical test theory: CTT) ซึ่งทฤษฎีการตอบสนองข้อสอบนี้มีความเชื่อว่าพฤติกรรมการตอบสนองต่อข้อสอบของผู้สอบเป็นสิ่งที่สังเกตได้โดยตรงว่าถูกหรือผิด แต่คุณลักษณะภายในหรือความสามารถที่อยู่ภายในตัวบุคคลไม่สามารถสังเกตได้โดยตรง ดังนั้น โอกาสที่จะตอบข้อสอบถูกหรือผิดของผู้สอบ จึงขึ้นอยู่กับระดับความสามารถและคุณลักษณะของข้อสอบ ได้แก่ ค่าพารามิเตอร์ประจำข้อสอบแต่ละข้อประกอบด้วย ความยาก (b) อำนาจจำแนก (a) และโอกาสการเดา (c) โดยความสัมพันธ์ของค่าพารามิเตอร์ดังกล่าว อธิบายได้ในรูปของฟังก์ชันคณิตศาสตร์ เรียกว่า ฟังก์ชันโลจิส (Logistic function) ซึ่งมีความใกล้เคียงกับฟังก์ชันปกติสะสม (Normal ogive function) โดยฟังก์ชันโลจิสเป็นฟังก์ชันที่มีความทนต่อความคลาดเคลื่อนที่เกิดกับผู้สอบที่มีความสามารถสูง จึงทำให้ถูกใช้อย่างแพร่หลายและนิยมนำไปใช้จริงมากกว่าฟังก์ชันปกติสะสม โดยโมเดลการตอบสนองข้อสอบแบบสองค่า (Dichotomous item response theory) เป็นดังนี้ (Hambleton et al., 1991; ศิริชัย กาญจนวาสี, 2545; อิทธิฤทธิ์ พงษ์ปิยะรัตน์, 2551; อริสพา เตหลิ่ม, 2559)

1. โมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์

โมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ (One-parameter logistic model: 1PL Model) เป็นโมเดลที่มีลักษณะซับซ้อนน้อยที่สุดของโมเดลในกลุ่มของ IRT หรือ โมเดลราสช์ (Simple Rash model) พัฒนาขึ้นโดย George Rash นักคณิตศาสตร์ชาวเดนมาร์ก ในปี ค.ศ. 1960 มีความเชื่อว่า องค์ประกอบที่สำคัญที่ทำให้กระบวนการวัดผลเกิดประสิทธิภาพที่ดีที่สุด นั่นคือ การใช้เครื่องมือในการวัดผลต้องมีความเป็นอิสระในตัวเอง หรืออีกนัยหนึ่งคือ คุณภาพของการวัดผลต้องไม่ได้ขึ้นอยู่กับสิ่งที่ต้องการวัด ดังนั้น จึงได้เกิดแนวคิดขึ้นว่าควรมีการพัฒนาการวิเคราะห์ข้อสอบที่ให้ผลการวิเคราะห์เป็นอิสระ โดย Rash ได้พยายามคิดหาค่าความยากของข้อสอบโดยไม่ต้องไปสัมพันธ์กับผู้สอบและหาค่าความสามารถของผู้สอบโดยไม่ต้องไปสัมพันธ์กับระดับความยากง่ายของข้อสอบ ซึ่งก็คือ เครื่องมือวัดและสิ่งที่ถูกวัดเป็นอิสระต่อกันและกัน

ใน Rash model ตัวแปรตามจะเป็นการตอบแบบให้คะแนนสองค่า (Dichotomous response) ของผู้สอบแต่ละคน ส่วนตัวแปรอิสระเป็นคะแนนความสามารถของผู้สอบ (θ_s) และความยากง่ายของข้อสอบ (β_i) ทั้งนี้ ตัวแปรอิสระจะเป็นการรวมกันเชิงบวก (Combine additively) ความสัมพันธ์ดังกล่าวจะเป็นกรณีที่ตัวแปรตามมีค่าเป็น Log odds หรือความน่าจะเป็นสำหรับสมการ Log odd ของ Rash model ตัวแปรจะเป็นลอกธรรมชาติของโอกาสในการตอบข้อสอบได้ถูกต้อง Odd จะถูกกำหนดเป็นอัตราส่วนของจำนวนข้อสอบที่ตอบถูกต่อจำนวนสำหรับสมการ Log odd ของ Rash model ตัวแปรจะเป็น Log ธรรมชาติของโอกาสในการตอบข้อสอบได้ถูกต้อง Odd จะถูกกำหนดเป็นอัตราส่วนของจำนวนข้อสอบที่ตอบถูกต่อจำนวนข้อสอบที่ตอบผิด เช่น Odd ของผู้สอบคนหนึ่งเป็น 4/1 นั่นคือ ข้อสอบ 5 ข้อ ผู้สอบตอบถูก 4 ข้อ และตอบผิด 1 ข้อ นอกจากนี้ Odd ยังสามารถเป็นความน่าจะเป็นของการตอบข้อสอบถูกหารด้วยความน่าจะเป็นที่จะตอบข้อสอบผิดได้อีกด้วย สำหรับ Rash model ค่า Log_n ของอัตราส่วน Odd จะถูกทำให้เป็นโมเดลด้วยผลต่างระหว่าง θ_s และ β_i นั่นคือ อัตราส่วนของ ความน่าจะเป็นที่ผู้สอบจะตอบข้อสอบข้อที่ i ได้ถูกต้องต่ออัตราส่วนที่จะตอบข้อสอบข้อนี้ผิด เขียนเป็นสมการได้ดังนี้

$$\ln\left[\frac{p_{is}}{1-p_{is}}\right] = \theta_s - \beta_i \qquad [2\text{-}1]$$

หาก $\theta_s = \beta_i$ แสดงว่าค่าของ log odd ของการตอบถูกจะเท่ากับศูนย์เมื่อถอดค่า Log ออกมา จะเท่ากับ 1 สามารถแปลความหมายได้ว่าบุคคลมีโอกาสที่จะตอบข้อสอบได้ถูกเท่ากับการตอบข้อสอบผิด

ใน Rash model ที่นำเสนอสามารถประมาณค่าความสามารถของผู้สอบได้ เมื่อทราบค่าความยากของข้อสอบ โดยมีการประมาณค่าที่แยกออกจากกัน ซึ่งเป็นคุณสมบัติที่สำคัญของทฤษฎีการวัด

อีกหนึ่งลักษณะของ Rash model คือ ตัวแปรตามเป็นความน่าจะเป็นของผู้สอบที่จะตอบข้อสอบถูก โดยค่า θ_s และ β_i มีความสัมพันธ์กันในเชิงบวก แต่โมเดลนี้ความเกี่ยวข้องของตัวแปรตามและตัวแปรอิสระ ซึ่งมีลักษณะไม่เป็นเส้นตรง (Non linear function) เป็นฟังก์ชั่นโลจิสเขียนเป็นสมการได้ ดังนี้

$$P(x_{is}=1|\theta_s,\beta_i)=\frac{\exp(\theta_s-\beta_i)}{1+\exp(\theta_s-\beta_i)} \qquad [2\text{-}2]$$

เมื่อ θ หมายถึง ระดับความสามารถของผู้สอบที่ประมาณได้จากแบบจำลองตามทฤษฎีการตอบสนองข้อสอบ ปรับให้เป็นคะแนนมาตรฐานที่มีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 θ มีพิสัยระหว่าง $\pm\infty$ แต่ในทางปฏิบัติส่วนใหญ่ θ มีค่าระหว่าง ± 3

$P_i(\theta)$ หมายถึง ความน่าจะเป็นที่ผู้สอบที่มีระดับความสามารถที่ θ จะสามารถตอบข้อสอบข้อที่ i ได้ถูกต้อง

β_i หมายถึง ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i (Difficulty parameter) ซึ่งเป็นค่าที่แสดงตำแหน่งของ ICC ณ จุด θ ที่มีโอกาสตอบข้อสอบถูก 0.50

e หมายถึง ค่าคงที่ของลอการิทึมธรรมชาติ (Natural log) มีค่าประมาณ 2.71828

บนพื้นฐานของการแจกแจงแบบปกติสะสม Normal ogive function

$$P_i(\theta)=\int_{-\infty}^{\theta-b_i}\frac{1}{\sqrt{2\pi}}e^{-Z^2/2}dz$$

พัฒนามาเป็น Logistic function

$$P_i(\theta)=\frac{1}{1+e^{-(\theta-b_i)}} \qquad [2\text{-}3]$$

โมเดลที่ [2-3] จะต่างจากโมเดลที่ [2-1] เพราะตัวแปรตามจะถูกทำนายในลักษณะความน่าจะเป็นมากกว่า log odd สมการที่ [2-3] คือ สมการ One-parameter logistic model

2. โมเดลการตอบสนองข้อสอบแบบ 2 พารามิเตอร์

Lord (1952 cited in Hambleton et al., 1991) พัฒนาโมเดลการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ (Two-parameter logistic model: 2PL Model) โดยอยู่บนพื้นฐานของการแจกแจงแบบปกติสะสม (Cumulative normal distribution หรือ Normal ogive) ซึ่งมีสมการ ดังนี้ (Allen & Yen, 1979)

$$P_i(\theta) = \int_{-\infty}^{a_i(\theta - b_i)} \frac{1}{\sqrt{2\pi}} e^{-Z^2/2} dz \qquad [2\text{-}4]$$

เมื่อ b_i หมายถึง ความยากของแบบสอบ

a_i หมายถึง อำนาจจำแนกของแบบสอบ

z หมายถึง ตัวแปรโค้งปกติ (Standard normal variable)

ต่อมา Birnbaum (1968 cited in Hambleton et al., 1991) ได้พัฒนาฟังก์ชันโลจิสแบบ 2 พารามิเตอร์ เข้ามาแทนฟังก์ชันปกติสะสมแบบ 2 พารามิเตอร์ ฟังก์ชันโลจิสมีข้อได้เปรียบมากกว่าตรงที่สามารถคำนวณได้สะดวกกว่า โดยในเส้นโค้งโลจิสมีค่าพารามิเตอร์ 2 ตัว คือ ความยากของข้อสอบ (b) และอำนาจจำแนกของข้อสอบ (a)

โมเดลการตอบสนองข้อสอบแบบ 2 พารามิเตอร์มีโค้งลักษณะข้อสอบที่เขียนด้วย

$$P_i(\theta) = \frac{1}{1 + e^{-Da_i(\theta - b_i)}} \qquad [2\text{-}5]$$

เมื่อ θ หมายถึง ระดับความสามารถของผู้สอบที่ประมาณได้จากแบบจำลองตามทฤษฎีการตอบสนองข้อสอบ

$P_i(\theta)$ หมายถึง ความน่าจะเป็นที่ผู้สอบที่มีระดับความสามารถที่ θ จะสามารถตอบข้อสอบข้อที่ i ได้ถูกต้อง

b_i หมายถึง ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i ซึ่งเป็นค่าที่แสดงตำแหน่งของ ICC ณ จุด θ ที่มีโอกาสตอบข้อสอบถูก 0.5

a_i หมายถึง ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่ i ซึ่งเป็นค่าที่แสดงความชันของ ICC ณ ตำแหน่ง b_i

e หมายถึง ค่าคงที่ของลอการิทึมธรรมชาติมีค่าประมาณ 2.71828

D หมายถึง ค่าองค์ประกอบของการปรับสเกล (Scaling factor) ให้ Logistic function มีค่าใกล้เคียงกับ Normal ogive function หรือมีค่าประมาณ θ ต่างกันไม่เกิน 0.01 โดยที่ D มีค่าเท่ากับ 1.70

3. โมเดลการตอบสนองข้อสอบแบบ 3 พารามิเตอร์

โมเดลการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ (Three-parameter logistic model: 3PL model) นี้ ได้รับการพัฒนามาจากโมเดลแบบ 2 พารามิเตอร์เพื่อให้เหมาะสมกับ การนำในไปใช้

ในการทดสอบ ซึ่งมีอิทธิพลจากโอกาสในการเดา (c) หรือจุดต่ำที่สุดของโค้งการตอบข้อสอบ (Lower asymptote) หมายถึง ความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องของผู้สอบที่มีระดับความสามารถต่ำสุด เข้ามาแฝงอยู่ด้วย โดยมีโค้งลักษณะข้อสอบที่เขียนด้วยฟังก์ชันปกติสะสมแบบ 3 พารามิเตอร์ (Three-parameter normal ogive) ดังสมการ

$$P_i(\theta) = c_i + (1 - c_i) \int_{-\infty}^{a_i(\theta - b_i)} \frac{1}{\sqrt{2\pi}} e^{-Z^2/2} dZ \qquad [2\text{-}6]$$

เมื่อ b_i หมายถึง ความยากของแบบสอบ

a_i หมายถึง อำนาจจำแนกของแบบสอบ

c_i หมายถึง โอกาสการเดาของแบบสอบ

z หมายถึง ตัวแปรโค้งปกติ (Standard normal variable)

หรือโค้งลักษณะข้อสอบที่เขียนด้วยฟังก์ชันโลจิส ดังสมการ

$$P_i(\theta) = c_i + \frac{(1 - c_i)}{1 + e^{-Da_i(\theta - b_i)}} \qquad [2\text{-}7]$$

เมื่อ θ หมายถึง ระดับความสามารถของผู้สอบที่ประมาณได้จากแบบจำลองตามทฤษฎีการตอบข้อสอบ

$P_i(\theta)$ หมายถึง ความน่าจะเป็นที่ผู้สอบที่มีระดับความสามารถที่ θ จะสามารถตอบข้อสอบข้อที่ i ได้ถูกต้อง

b_i หมายถึง ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i ซึ่งเป็นค่าที่แสดงตำแหน่งของ ICC ณ จุด θ ที่มีโอกาสตอบข้อสอบถูก $\frac{1 + c_i}{2}$

a_i หมายถึง ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่ i ซึ่งเป็นค่าที่แสดงความชันของ ICC ณ ตำแหน่ง b_i

c_i หมายถึง ค่าพารามิเตอร์โอกาสในการเดาข้อสอบข้อที่ i ได้ถูก

e หมายถึง ค่าคงที่ของลอการิทึมธรรมชาติมีค่าประมาณ 2.71828

D หมายถึง ค่าการปรับสเกลให้ Logistic function มีค่าใกล้เคียงกับ Normal ogive function โดยที่ D มีค่าเท่ากับ 1.70

ตอนที่ 2 มโนทัศน์เกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบ

การศึกษาความลำเอียงของข้อสอบมีการศึกษาอย่างจริงจังในช่วงปี ค.ศ. 1960 เพื่อพัฒนาวิธีการสำหรับศึกษาความแตกต่างทางวัฒนธรรมและเพื่อรวบรวมสำหรับการยืนยัน เนื่องจากความไม่เท่าเทียมกันระหว่างนักเรียนผิวดำและลาตินอเมริกา กับนักเรียนผิวขาวในการทดสอบความสามารถทางสมอง ซึ่งข้อสอบเกี่ยวข้องกับเนื้อหาที่นักเรียนชนกลุ่มน้อย เช่น นักเรียนที่เป็นคนผิวดำหรือเชื้อชาติที่มาจากประเทศอื่น ๆ มีโอกาสที่จะเรียนรู้ได้น้อยกว่า และเป้าหมายที่เฉพาะคือการระบุข้อสอบที่มีความลำเอียงต่อนักเรียนชนกลุ่มน้อยและตัดข้อสอบเหล่านั้นออกจากแบบสอบ ในระยะเวลาเดียวกันก็ได้มีการศึกษาเพื่อรวบรวมความสมบูรณ์ของแบบสอบ โดยเฉพาะแบบสอบที่ใช้เพื่อการคัดเลือก เนื่องมาจากความลำเอียงต่อชนกลุ่มน้อยทำให้คะแนนของพวกเขาต่ำกว่าตำแหน่งบนเกณฑ์การวัดที่ควรจะเป็นการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (Differential item functioning: DIF) มีขึ้นกับข้อสอบที่ใช้ในการประเมินทางด้านจิตวิทยาหรือการประเมินทางการศึกษา เมื่อความน่าจะเป็นของการตอบสนองต่อข้อสอบ (ตอบถูก) แตกต่างกันสำหรับสมาชิกในกลุ่มต่าง ๆ ในประชากร (เช่น เชื้อชาติ) หลังจากการควบคุมลักษณะแฝงที่ถูกวัดโดยเครื่องมือ เช่น ผลการอ่าน (Camilli & Shepard, 1994) เกี่ยวข้องโดยตรงกับการวัดผล เมื่อมีการวัดผลมุ่งหวังวัดได้ตรงกับคุณลักษณะที่ต้องการวัด วัดได้ตรง ก็สามารถแปลผลได้ตรง สิ่งที่เกี่ยวข้องโดยตรงกับการทดสอบหรือการวัด คือ เครื่องมือวัด ความสามารถของเครื่องมือในการวัดสิ่งที่ต้องการจะวัดได้ถูกต้อง เป็นคุณสมบัติด้านความเที่ยงตรงหรือความตรงของเครื่องมือวัด ได้มีนักการศึกษานักวิชาการได้ให้ความหมายและความสำคัญของความตรงไว้ไปในแนวทางเดียวกัน สรุปความได้ว่า ความตรงเป็นคุณสมบัติของเครื่องมือที่ชี้ให้เห็นว่าค่าที่วัดได้เป็นตัวแทนของสิ่งที่ถูกวัดในประชากรได้อย่างถูกต้อง ถ้าผลการวัดมีค่าที่ใกล้เคียงกับค่าคุณลักษณะที่แท้จริงเท่าใด ก็ถือว่าการวัดมีความตรงมากขึ้นเท่านั้น (Cronbach, 1971 อ้างถึงใน ไพรัตน์ วงษ์นาม, 2546, หน้า 271; ศิริชัย กาญจนวาสี, 2545, หน้า 103; ไพรัตน์ วงษ์นาม, 2546, หน้า 271; Marie, 2007, Abstract) วิธีการตรวจสอบความตรงมีหลายวิธีด้วยกัน การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ (Differential item functioning: DIF) เป็นวิธีหนึ่งที่นำมาใช้ โดยเป็นการตรวจสอบในประเด็นของความอยุติธรรมของข้อสอบและแบบสอบ (Item and test unfairness)

ความยุติธรรมต่อผู้ถูกวัดเป็นหัวใจสำคัญ เพราะผลจากการวัดส่งผลต่อการประเมินผลหรือการตัดสินผล ความไม่ยุติธรรมที่เกิดจากข้อสอบเกิดขึ้นในกรณีที่ทำให้ผู้สอบระหว่างกลุ่มย่อยมีความได้เปรียบหรือเสียเปรียบกัน เดิมใช้คำว่า “ความลำเอียงของข้อสอบ” (Item bias) ซึ่งเป็นภาษาที่ใช้กันโดยทั่วไปและมีความหมายในเชิงลบ สำหรับการตัดสินว่าข้อสอบมีความลำเอียง

หรือไม่นั้น มักจะพิจารณาอิทธิพลที่สังเกตได้ของผู้สอบกลุ่มย่อยที่นำมาศึกษา โดยไม่คำนึงถึงวิธีการทางสถิติ จึงทำให้เกิดความคลุมเครือเกี่ยวกับเกณฑ์ที่ใช้ในการตัดสินความลำเอียง ต่อมานักวิธีวิทยาการวิจัยได้พัฒนาวิธีเพื่อใช้วิเคราะห์ดัชนีความลำเอียงของข้อสอบ โดยมุ่งเน้นความแตกต่างระหว่างกลุ่มผู้สอบที่ตอบข้อสอบข้อเดียวกัน และเลือกเกณฑ์ที่ใช้ในการจับคู่กลุ่มผู้สอบ (Matching criterion) แล้วนำสารสนเทศทางสถิติมาใช้เป็นเกณฑ์ในการตัดสินความลำเอียงของข้อสอบ ดังนั้น จึงเปลี่ยนมาใช้คำว่า “การทำหน้าที่ต่างกันของข้อสอบ” (Differential item functioning: DIF) แทน ซึ่งเป็นคำที่มีความหมายเป็นกลางและมีความเหมาะสมมากกว่า (Holland & Wainer, 1993)

การทำหน้าที่ต่างกันของข้อสอบ (DIF) กับความสำเอียงของข้อสอบ (Item bias) มีแนวคิดที่ต่างกัน ในประเด็น การทำหน้าที่ต่างกันของข้อสอบ เป็นการดำเนินการที่เน้นการใช้วิธีการทางสถิติสำหรับการตรวจสอบ เพื่อให้ได้สารสนเทศเกี่ยวกับการทำหน้าที่ของข้อสอบสำหรับกลุ่มผู้สอบกลุ่มย่อยที่มีลักษณะเฉพาะบางอย่างแตกต่างกัน ส่วนความลำเอียงของข้อสอบเป็นกระบวนการตัดสินความยุติธรรมของข้อสอบ โดยนำสารสนเทศการทำหน้าที่ต่างกันของข้อสอบมาวิเคราะห์ (Logical analysis) โดยผู้เชี่ยวชาญพิจารณาถึงการเขียนข้อสอบ เนื้อหาสาระของข้อสอบและจุดมุ่งหมายของการวัด เพื่อระบุว่าข้อสอบข้อนั้นลำเอียงเข้าข้างกลุ่มหนึ่งกลุ่มใดหรือไม่ (Camilli & Shepard, 1994)

ความหมายของการทำหน้าที่ต่างกันของข้อสอบ

จากความเป็นมาข้างต้นเมื่อกล่าวถึงการทำหน้าที่ต่างกันของข้อสอบ ซึ่งมีความเชื่อมโยงกับความลำเอียงของข้อสอบ ดังนั้น ในการให้ความหมายของการหน้าที่ต่างกันของข้อสอบ จึงต้องนำเสนอในส่วนของความลำเอียงของข้อสอบร่วมด้วย ดังนี้

ความลำเอียงของข้อสอบ หมายถึง สัดส่วนของผู้สอบที่ตอบข้อสอบได้ถูกต้องไม่เท่ากันในแต่ละกลุ่มประชากรที่ใช้ในการศึกษา เมื่อกลุ่มผู้สอบมีคะแนนเท่ากันและข้อสอบมีความเป็นเอกพันธ์ (Scheuneman,1979 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555)

ความลำเอียงของข้อสอบ หมายถึง โอกาสในการตอบข้อสอบได้ถูกต้องของผู้สอบกลุ่มหนึ่ง มีค่าต่ำกว่าหรือสูงกว่า ผู้สอบอีกกลุ่มหนึ่งที่มีระดับความสามารถเดียวกัน (Dorans and Kulick, 1986 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555)

ความลำเอียงของข้อสอบ หมายถึง คะแนนข้อสอบของกลุ่มผู้สอบที่มีความสามารถเท่ากัน แต่มาจากต่างกลุ่มกัน มีความแตกต่างกันอย่างเป็นระบบ (Kederman, 1990 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555)

การทำหน้าที่ต่างกันของข้อสอบ หมายถึง การสังเกตได้ว่าข้อสอบแสดงคุณสมบัติทางสถิติที่แตกต่างกัน ในผลการตอบข้อสอบจากกลุ่มผู้สอบแตกต่างกัน ภายหลังการควบคุมความแตกต่างในความสามารถของกลุ่มผู้สอบ (Angoff, 1993, pp. 4-5)

การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ข้อสอบที่มีความลำเอียงเป็นความไม่เที่ยงตรงหรือเป็นความคลาดเคลื่อนอย่างเป็นระบบของการวัด สำหรับสมาชิกกลุ่มใดกลุ่มหนึ่ง (เช่น เชื้อชาติ, เพศ ที่ตั้งตามภูมิศาสตร์ (Camilli & Shepard, 1994, p. 8)

การทำหน้าที่ต่างกันของข้อสอบ หมายถึง โอกาสของการประสบความสำเร็จในการตอบข้อสอบที่ต่างกัน ของผู้สอบที่มาจากต่างกลุ่มกัน เมื่อได้จับคู่ความสามารถที่สนใจของผู้สอบ (Borsboom, Mellenbergh & Heerden, 2002)

การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ระดับความสำเร็จในการตอบข้อสอบที่ต่างกันของผู้สอบต่างกลุ่มที่มีความสามารถเท่ากันในเป้าหมายที่ต้องการวัดจากความหมายของการทำหน้าที่ต่างกันของข้อสอบ สรุปได้ว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง โอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกัน สาหรับผู้สอบที่มีความสามารถหรือคุณลักษณะในระดับเท่ากันที่มาจากกลุ่มประชากรที่แตกต่างกัน (Brennan, 2006)

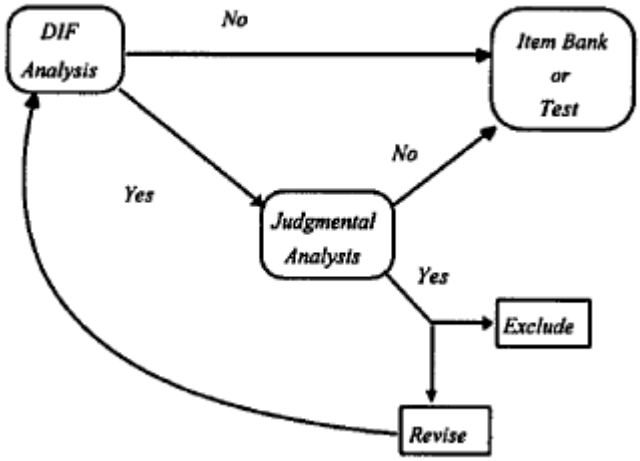
การทำหน้าที่ต่างกันของข้อสอบ หมายถึง การที่ข้อสอบทำให้ผู้สอบจากต่างกลุ่มกันที่มีความสามารถหรือคุณลักษณะที่มุ่งวัดเท่ากัน มีโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกัน หรือมีฟังก์ชันการตอบสนองข้อสอบแตกต่างกัน (ศิริชัย กาญจนวาสี, 2555, หน้า 105)

การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ข้อสอบที่ทำให้ผลการตอบของผู้สอบที่มีความสามารถเท่ากันในสิ่งที่ต้องการวัด มีโอกาสตอบข้อสอบข้อนั้นได้ถูกต้องไม่เท่ากันเนื่องจากอยู่ในกลุ่มย่อยต่างกัน เช่นเพศ หรือสถานที่ตั้งทางภูมิศาสตร์ (สุพัฒนา หอมบุปผา, 2556, หน้า 13)

การทำหน้าที่ต่างกันของข้อสอบ หมายถึง การตรวจสอบข้อสอบที่วัดคุณลักษณะของผู้สอบจากกลุ่มที่แตกต่างกันที่นำมาจับคู่ความสามารถที่เท่ากัน แล้วเปรียบเทียบความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง ถ้าข้อสอบที่มีการทำหน้าที่ต่างกันจะให้ค่าความน่าจะเป็นในการตอบถูกของข้อสอบข้อนั้นไม่เท่ากัน ทั้ง ๆ ที่มีความสามารถเท่ากัน แสดงว่าข้อสอบข้อนั้นทำหน้าที่ต่างกัน (อาวีพร ปานทอง, 2558, หน้า 18)

การทำหน้าที่ต่างกันของข้อสอบ (DIF) หมายถึง ข้อสอบที่ทำให้ผู้สอบซึ่งมีลักษณะต่างกัน แต่มีความสามารถระดับเดียวกัน มีโอกาสตอบถูกไม่เท่ากัน (อริสพา เตหลิ่ม, 2559, หน้า 15)

จากความหมายของการทำหน้าที่ต่างกันของข้อสอบได้แสดงให้เห็นกระบวนการของของการสร้างข้อสอบให้มีคุณภาพ เมื่อมีการตรวจสอบคุณภาพโดยใช้หลักการทำหน้าที่ต่างกันของข้อสอบ แสดงดังภาพที่ 2-1



ภาพที่ 2-1 กรอบแนวคิดของการทำหน้าที่ต่างกันของข้อสอบ (Park, 1995)

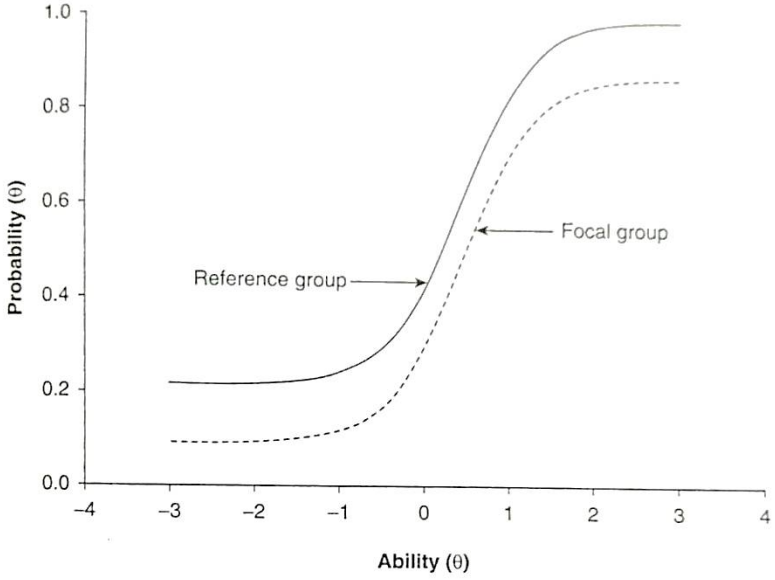
จากความเป็นมาและความหมายข้างต้น สรุปได้ว่า การทำหน้าที่ต่างกันของข้อสอบ (Differential item functioning: DIF) หมายถึง การตรวจสอบข้อสอบที่วัดคุณลักษณะผู้สอบที่มีความสามารถหรือคุณลักษณะที่มุ่งวัดเท่ากัน แต่มาจากต่างกลุ่มกัน ถ้าข้อสอบทำหน้าที่ต่างกันจะทำให้ค่าความน่าจะเป็นหรือโอกาสในการตอบข้อสอบข้อที่ตรวจสอบได้ของแต่ละกลุ่มถูกต้องแตกต่างกัน

ประเภทของการทำหน้าที่ต่างกันของข้อสอบ

การทำหน้าที่ต่างกันของข้อสอบจะเปรียบเทียบผลการตอบข้อสอบระหว่างผู้สอบกลุ่มย่อยอย่างน้อยสองกลุ่ม ที่มีความสามารถระดับเดียวกัน โดยผู้สอบกลุ่มที่ได้เปรียบ เรียกว่า กลุ่มอ้างอิง (Reference group: R) และผู้สอบกลุ่มที่เสียเปรียบ เรียกว่า กลุ่มเปรียบเทียบ (Focal group: F) ซึ่งเป็นกลุ่มที่สนใจศึกษาการทำหน้าที่ต่างกันของข้อสอบ ข้อสอบทำหน้าที่ต่างกัน เมื่อผู้สอบที่มีความสามารถระดับเดียวกันแต่อยู่ต่างกลุ่มกันมีโอกาสของการตอบข้อสอบได้ถูกต้องไม่เท่ากัน (Li & Stout, 1996 อ้างถึงใน วลีมาศ แซ่อึ้ง, 2543) ซึ่งขนาดและทิศทางของข้อสอบที่ทำหน้าที่ต่างกันจะแปรเปลี่ยนไปตามระดับความสามารถที่แตกต่างกัน โดยข้อสอบที่ทำหน้าที่ต่างกันแบ่งเป็นสองประเภท (Osterlind & Everson, 2009; Mellenbergh, 1982 อ้างถึงใน วลีมาศ แซ่อึ้ง,

2543) คือ ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป (Uniform DIF) และข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป (Nonuniform DIF)

ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป (Uniform DIF) เกิดขึ้นเมื่อไม่มีปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับการเป็นสมาชิกของกลุ่ม ส่วนข้อสอบทำหน้าที่ต่างกันประเภทหลังเกิดขึ้นเมื่อมีปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับการเป็นสมาชิกของกลุ่ม ซึ่งทฤษฎีการตอบสนองข้อสอบ (Item response theory: IRT) สามารถพิจารณาปฏิสัมพันธ์ได้จากความแตกต่างของค่าพารามิเตอร์อำนาจจำแนกของข้อสอบระหว่างผู้สอบกลุ่มย่อยสองกลุ่ม กล่าวคือ ถ้าข้อสอบระหว่างผู้สอบกลุ่มย่อยสองกลุ่ม มีค่าอำนาจจำแนกเท่ากันแล้ว โค้งลักษณะข้อสอบ (Item characteristic curves: ICCs) ระหว่างกลุ่มผู้สอบจะขนานกัน แสดงว่า ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป แต่ถ้าข้อสอบระหว่างผู้สอบกลุ่มย่อยสองกลุ่ม มีค่าอำนาจจำแนกไม่เท่ากันแล้ว โค้งลักษณะข้อสอบระหว่างกลุ่มผู้สอบจะไม่ขนานกัน แสดงว่า ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป (Camilli & Shepard, 1994)



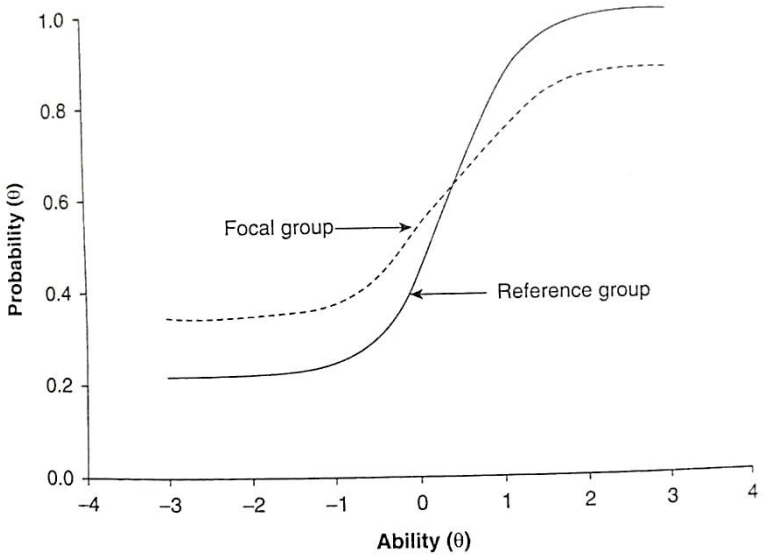
ภาพที่ 2-2 รูปแบบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป (Uniform DIF)

ข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป (Nonuniform DIF) สามารถจำแนกได้เป็นสองลักษณะ (Swaminathan & Rogers, 1990 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555) ดังนี้

1. ข้อสอบทำหน้าที่ต่างกันแบบมีทิศทางเดียว (Unidirectional DIF) คือ ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูปที่มีปฏิสัมพันธ์เป็นลำดับ (Ordinal interaction) เกิดขึ้นเมื่อ

โค้งลักษณะข้อสอบต่างกันอย่างไม่สม่ำเสมอแต่ไม่ตัดกันหรือเป็นลักษณะตัดกันนอกช่วงความสามารถของผู้สอบตรงปลายสุดของช่วงความสามารถต่ำหรือสูง

2. ข้อสอบทำหน้าที่ต่างกันแบบไม่มีทิศทาง (Non-unidirectional DIF) คือ ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูปที่มีปฏิสัมพันธ์ไม่เป็นลำดับ (Dis ordinal interaction) เกิดขึ้นเมื่อโค้งลักษณะข้อสอบตัดกันระหว่างช่วงความสามารถของผู้สอบ



ภาพที่ 2-3 รูปแบบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป (Nonuniform DIF)

การวิจัยครั้งนี้ ได้กำหนดกรอบวิธีที่นำมาตรวจสอบไว้ก่อนเป็นเบื้องต้นแล้วว่าสามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ทั้ง 2 รูปแบบ เนื่องจากว่าการศึกษากับข้อมูลจริงนั้นเราไม่สามารถกำหนดเงื่อนไขได้เหมือนการจำลองข้อมูลได้ จึงวางแผนในการศึกษาไว้ทั้ง 2 รูปแบบ คือ ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป (Uniform DIF) และข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป (Nonuniform DIF)

วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

สำหรับวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเมื่อใช้เกณฑ์จำแนกตามการตรวจให้คะแนนสามารถจำแนกได้เป็น 2 กลุ่ม คือ การตรวจให้คะแนนแบบทวิวิภาค (Dichotomous) และ การตรวจให้คะแนนแบบพหุวิภาค (Polytomous) สำหรับข้อสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค (Dichotomous) นั้นได้แบ่งตามมิติลักษณะของตัวแปรเกณฑ์ ได้เป็นสองกลุ่ม ดังนี้ (Potenza & Dorans, 1995; Feinstein, 1995 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555)

1. กลุ่มวิธีที่ใช้คะแนนสังเกตได้ (Observed score) ซึ่งวิเคราะห์ตามทฤษฎีแบบดั้งเดิม (Classical test theory: CTT) ใช้คะแนนรวมของผู้สอบเป็นเกณฑ์การจับคู่ของกลุ่มผู้สอบ วิธีการตรวจสอบที่สำคัญ ได้แก่ การวิเคราะห์ความแปรปรวน (ANOVA) วิธีการถดถอยโลจิสติก (Logistic regression: LR) วิธีแปลงค่าความยากของข้อสอบ (Transformed item difficulty: TID) วิธีแมนเทล-แฮนส์เซล (Mantel-haenszel; MH) และวิธีดัชนีมาตรฐาน (Standardization: STND)

2. กลุ่มวิธีที่ใช้คุณลักษณะแฝง (Latent variable) ซึ่งวิเคราะห์บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ (Item response theory: IRT) ใช้ค่าประมาณความสามารถของผู้สอบเป็นเกณฑ์การจับคู่กลุ่มผู้สอบ วิธีการตรวจสอบที่สำคัญ ได้แก่ วิธีวัดพื้นที่ความแตกต่างระหว่างโค้งการตอบสนองข้อสอบ (IRT-D^2) วิธีไค-สแควร์ของลอร์ด (Lord's χ^2) วิธีอัตราส่วนไลค์ลิฮูดทั่วไป (General IRT likelihood ratio) วิธีอัตราส่วนไลค์ลิฮูด ลอกลิเนียร์ (Log linear IRT likelihood ratio) และวิธีซิปเทสท์ (Simultaneous item bias test: SIBTEST)

วิธีที่ใช้หลักการของตารางการณ์จร (Contingency table) ดังเช่น วิธีแมนเทล-แฮนส์เซล และวิธีในกลุ่ม IRT ทั้งสองวิธีมีแนวคิดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบคล้ายกัน โดยใช้หลักการวิเคราะห์ความแตกต่างของผลการตอบข้อสอบจากผู้สอบสองกลุ่มที่มีความสามารถระดับเดียวกัน ในวิธีตารางการณ์จรจะใช้คะแนนรวมที่สังเกตได้แทนระดับความสามารถ ส่วนวิธีในกลุ่ม IRTจะประมาณค่าความสามารถโดยใช้ฟังก์ชันทางคณิตศาสตร์ตามลักษณะโมเดล IRT จุดเด่นของวิธีตารางการณ์จร คือ สามารถนาไปประยุกต์ใช้กับกลุ่มตัวอย่างที่มีขนาดเล็ก มีวิธีดำเนินการตรวจสอบง่าย และไม่จำเป็นต้องใช้โปรแกรมคอมพิวเตอร์ที่ซับซ้อนเหมือนกับวิธีในกลุ่ม IRT (Camilli & Shepard, 1994) ในการวิเคราะห์ด้วยวิธีนี้จะนำข้อมูลมาสร้างตารางการณ์จรแบบ 3 ทิศทาง คือ กลุ่มผู้สอบ (กลุ่มอ้างอิง/กลุ่มเปรียบเทียบ) ผลการตอบข้อสอบ (ถูก/ ผิด) และคะแนนรวม (k ระดับ) ซึ่งวิธีที่ใช้หลักการของตารางการณ์จรที่นิยมใช้กันโดยทั่วไป คือ วิธีแมนเทล-แฮนส์เซล (MH) พัฒนา โดย Holland and Thayer (1988 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555) เป็นที่ยอมรับจากนักวิธีวิทยาการวิจัยอย่างกว้างขวางว่าเป็นวิธีที่ใช้ง่าย สะดวก และประหยัด มีขั้นตอนการคำนวณที่ไม่สลับซับซ้อน มีการทดสอบทางสถิติแบบนันพาราเมตริก ซึ่งไม่จำเป็นต้องใช้โมเดลประมาณค่า

วิธีที่ใช้คะแนนสังเกตได้ (Observed score) ซึ่งวิเคราะห์ตามทฤษฎีแบบดั้งเดิม (CTT) ดังเช่น วิธีการถดถอยโลจิสติก (LR) เป็นวิธีที่พัฒนามาจากวิธีลอกลิเนียร์ ของ Mellenberg ซึ่งเสนอโดย Swaminathan andRogers (1990 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555) วิธีนี้อยู่บนพื้นฐานของโมเดลการวิเคราะห์สมการถดถอย

โลจิสติก ซึ่งเป็นโมเดลที่มีพื้นฐานเป็นแบบจำลองที่สามารถเพิ่มตัวแปรความสามารถและปฏิสัมพันธ์เข้าไปในสมการได้ และสอดคล้องกับธรรมชาติของการทำหน้าที่ต่างกันของข้อสอบได้ดี จึงสามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ทั้งแบบเอกรูปและอเนกรูป โมเดลถดถอยโลจิสติก เป็นโมเดลที่มีความยืดหยุ่นและใช้ได้ง่าย ในการวิเคราะห์สมการถดถอยโลจิสติก หรือเรียกว่าการวิเคราะห์โลจิท (Logit analysis) เป็นการวิเคราะห์สมการทำนาย เมื่อ ต้องการศึกษาผลของตัวแปรทำนายที่มีต่อตัวแปรเกณฑ์ซึ่งมีลักษณะเป็นทวิวิภาค (Dichotomous variable) โดยใช้ฟังก์ชันโลจิสติก (Logistic function) ในการแสดงความสัมพันธ์ระหว่างค่าของตัวแปรทำนายกับค่าความน่าจะเป็นของการเกิดเหตุการณ์ตามตัวแปรเกณฑ์วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่วิเคราะห์ด้วยทฤษฎี CTT มีข้อจำกัดเกี่ยวกับความคงที่ของค่าสถิติ โดยจะมีค่าแปรเปลี่ยนไปตามกลุ่มตัวอย่าง ส่งผลให้การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบไม่น่าเชื่อถือ นอกจากนี้ยังใช้เกณฑ์การจับคู่กลุ่มผู้สอบภายใน (Internal matching criterion) โดยใช้คะแนนรวมของแบบสอบแทนระดับความสามารถของผู้สอบ ซึ่งอาจมีผลทำให้แบบสอบเกิดความลำเอียง เนื่องมาจากเกณฑ์ดังกล่าวจะใช้คะแนนรวมที่ได้มาจากข้อสอบที่ทำหน้าที่ต่างกันรวมอยู่ด้วย (Zieky, 1993) จากจุดอ่อนของทฤษฎี CTT จึงทำให้นักวิธีวิทยาการวิจัยนำทฤษฎี IRT ไปประยุกต์ใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยเชื่อว่าเป็นทฤษฎีที่มีความแกร่งสามารถแก้ปัญหาของทฤษฎี CTT ได้ จุดเด่นของทฤษฎี IRT คือ ความไม่แปรเปลี่ยนของค่าพารามิเตอร์ของข้อสอบ โดยสามารถนำคุณสมบัติทางสถิติของข้อสอบไปอธิบายลักษณะของข้อสอบที่ทำหน้าที่ต่างกันได้แม่นยากว่าทฤษฎี CTT (Camilli & Shepard, 1994) วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้ทฤษฎี IRT ที่นิยมนำมาใช้ เช่น วิธีการวัดพื้นที่ของราชู วิธีการวัดพื้นที่ของคิมและโคเฮน วิธีการทดสอบไค-สแควร์ของลอร์ด และวิธีการทดสอบอัตราส่วนไลค์ลิฮูด (วลีมาศ แซ่อึ้ง, 2543)

ตารางที่ 2-1 วิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจำแนกตามลักษณะข้อมูล

Id	Method	Par/ Non-p	Obs/ Latent	Item scores	T/M	U/N
1	Mantel-Haenszel	Non-p	Obs	D/P	T/M	U
2	Standardization	Non-p	Obs	D	M	U
3	Chi-square methods	Non-p	Obs	D	T	U
4	SIBTEST	Non-p	Lat	D/P	T/M	U/N
5	Logistic regression	Par/	Obs	D/P	T/M	U/N
6	Likelihood ratio test	Par/	Obs /Lat	D/P	T/M	U/N
7	Prob. diff. indices	Par/	Lat	D	M	U/N
8	b parameter indices	Par/	Lat	D	M	U/N
9	General IRT-LR	Par/	Lat	D/P	T/M	U/N
10	IRT LRT	Par/	Lat	D/P	T	U/N
11	IRT methods	Par/	Lat	D/P	T/M	U/N
12	Lord's chi-squared test	Par/	Lat	D	T	U/N
13	Log linear models	Par/	Obs	D/P	T	U/N
14	Mixed effect models	Par/	Lat	D/P	T	U/N

หมายเหตุ 1. (Par) ametric or (non-p) arametric. 2. Matching variable; (Obs) erved or (Lat) ent 3. D) ichotomusly or (P) olytomously scored items. 4. Whether one can (T) est and/or (M) easure DIF. 5. Able to handle (U) niform and (N) onUniform DIF.

จากตารางที่ 2-1 Marie (2007) ได้จำแนกวิธีการทำหน้าที่ต่างกันของข้อสอบตามเกณฑ์ 5 เกณฑ์ ดังนี้

เกณฑ์ที่ 1 ชนิดของสถิติ

1. (non-p) arametric ได้แก่วิธี Mantel-haenszel, Standardization, Chi-square methods และ SIBTEST

2. (Par) ametric ได้แก่วิธี Logistic regression, Likelihood ratio test, Prob. diff. indices, b parameter indices, General IRT-LR, IRT LRT, IRT methods, Lord's chi-squared test, Log linear models และ Mixed effect models

เกณฑ์ที่ 2 ตัวแปรที่ใช้ในการจับคู่ (Matching variable)

1. (Obs) erved ได้แก่ วิธี Mantel-haenszel, Standardization, Chi-square methods Logistic regression, Likelihood ratio test และ Log linear models

2. (Lat) ent ได้แก่ วิธี SIBTEST, Likelihood ratio test, General IRT-LR, IRT LRT, IRT methods, Lord's chi-squared test, และ Mixed effect models

เกณฑ์ที่ 3 การตรวจให้คะแนนของข้อสอบ

1. (D)ichotomusly scored items ได้แก่ วิธีทั้ง 14 วิธี

2. (P)olytomously scored items ได้แก่ วิธี Mantel-Haenszel, SIBTEST, Logistic Regression, Likelihood ratio test, General IRT-LR, IRT LRT, IRT methods, Log linear models และ Mixed effect models

เกณฑ์ที่ 4 ใช้ตรวจสอบ DIF (Whether one can (T) est) หรือใช้วัดขนาดของ DIF ((M) easure DIF)

1. (T) est DIF ได้แก่ วิธี Mantel-haenszel, Chi-square methods, SIBTEST, Logistic Regression, Likelihood ratio test, General IRT-LR และ IRT methods

2. (M) easure DIF ได้แก่ วิธี Mantel-haenszel, Standardization, SIBTEST, Logistic Regression, Likelihood ratio test, Prob. diff. indices, b parameter indices, General IRT-LR และ IRT methods

เกณฑ์ที่ 5 รูปแบบของการทำหน้าที่ต่างกัน (U)niform and (N)onUniform DIF

1. (U) niform DIF ได้แก่ วิธี Mantel-haenszel, Chi-square methods, SIBTEST, Logistic Regression, Likelihood ratio test, General IRT-LR และ IRT methods

2. (N) onuniform DIF ได้แก่ วิธี Mantel-haenszel, Standardization, SIBTEST, Logistic regression, Likelihood ratio test, Prob. diff. indices, B parameter indices, General IRT-LR และ IRT methods

สุพัฒนา หอมบุปผา (2556, หน้า 42) ได้สรุปงานเขียนของ Good review by Teresi (2006) Medecal care Vol. 44 ในประเด็นวิธีการกำหนดการทำหน้าที่ต่างกันของข้อสอบเมื่อใช้ประเภทของสถิติเป็นเกณฑ์ในการจำแนก แบ่งได้เป็น 2 ประเภท

1. นันพาราเมตริก ได้แก่ วิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel: MH) (Holland & Thayer, 1988)

2. พาราเมตริก ได้แก่

2.1 วิธีถดถอยโลจิสติก (Polytomous logistic regression) (Zumbo, 1999)

2.2 Ordinal logistic regression (Crane, 2006)

2.3 วิธี MIMIC models (Muthén, 1989)

2.4 โมเดลกลุ่มพหุ (Multiple group models)

2.5 IRT Based Method (Thissen, 2001)

สรุปได้ว่าวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจำแนกตามเกณฑ์การแบ่งได้ 5 เกณฑ์ ได้แก่ เกณฑ์ที่ 1 ชนิดของสถิติ เกณฑ์ที่ 2 ตัวแปรที่ใช้ในการจับคู่ (Matching variable) เกณฑ์ที่ 3 การตรวจให้คะแนนของข้อสอบ เกณฑ์ที่ 4 จุดประสงค์ของการใช้ คือ ใช้ตรวจสอบ DIF หรือใช้วัดขนาดของ DIF และเกณฑ์ที่ 5 รูปแบบของการทำหน้าที่ต่างกัน ผู้วิจัยได้ศึกษาวิธีและเกณฑ์ต่าง ๆ เพื่อให้สามารถเลือกใช้วิธีที่มีความถูกต้องเหมาะสมที่สุดกับลักษณะข้อมูลและวัตถุประสงค์ของการวิจัย รวมถึงปัจจัยที่ใช้ในการจำแนกกลุ่มผู้สอบ การวิจัยครั้งนี้ผู้วิจัยได้นำวิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 3 วิธี ประกอบด้วยวิธี IRT LRT วิธี MGCFA-ALIGNMENT และวิธี MIMIC มาตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มผู้สอบที่จำแนกได้มากกว่า 2 กลุ่ม โดยมีรายละเอียดของแต่ละวิธี ดังนี้

1. วิธีทดสอบอัตราส่วนความควรจะเป็น (Item response theory likelihood ratio test: IRT LRT)

Thissen, Steinberg and Wainer (1988; อาวีพร ปานทอง, 2558) ได้เสนอวิธีการทดสอบอัตราส่วนความควรจะเป็น สำหรับใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบความแตกต่างของผลการตอบข้อสอบระหว่างผู้สอบกลุ่มย่อย 2 กลุ่ม วิธีทดสอบ แบ่งออกเป็น 3 วิธีย่อย ๆ คือ 1) วิธีการทดสอบอัตราส่วนความเป็นไปได้ในทฤษฎีการตอบข้อสอบในรูปทั่วไป (General LR) เป็นวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ซึ่งใช้การประมาณค่าพารามิเตอร์ในโมเดลการตอบข้อสอบด้วยวิธีความเป็นไปได้สูงสุดแบบมาร์จินอล (Marginal maximum likelihood: MML) 2) วิธีการทดสอบอัตราส่วนความเป็นไปได้ในทฤษฎีการตอบข้อสอบในรูปลอกลิเนียร์ (Loglinear LR) เป็นการประมาณค่าพารามิเตอร์ด้วยวิธีความเป็นไปได้สูงสุด (Maximum likelihood: ML) และ 3) วิธีการทดสอบอัตราส่วนความเป็นไปได้ในทฤษฎีการตอบข้อสอบในรูปสารสนเทศที่มีขอบเขตจำกัด (Limited information LR) สำหรับวิธีนี้จะใช้การประมาณค่าพารามิเตอร์ในโมเดลการตอบข้อสอบแบบ Normal ogive ด้วยวิธีกำลังสองน้อยที่สุดในรูปทั่วไป (Generalized least squares: GLS) ทั้ง 3 วิธีดังกล่าวจะใช้การทดสอบอัตราส่วน

ความเป็นไปได้ เพื่อทดสอบนัยสำคัญของการทำหน้าที่ต่างกันของข้อสอบ ซึ่งผลความกลมกลืนของโมเดลเป็นที่รู้จักโดยทั่วไปในความกลมกลืนของฟังก์ชันในทฤษฎีการตอบข้อสอบ (IRT) ค่าความกลมกลืนของฟังก์ชันเป็นดัชนีบอกว่าโมเดลความเหมาะสมกับข้อมูลที่จะใช้กระบวนการประมาณค่าความเป็นไปได้สูงสุดในการประมาณค่าพารามิเตอร์ข้อสอบ (Camilli & Shepard, 1994)

หลักการตรวจสอบความเท่าเทียมกันของการวัดด้วยวิธีการทดสอบอัตราส่วนความเป็นไปได้ จะเปรียบเทียบระหว่าง 2 โมเดล คือ โมเดลพื้นฐาน (Compact) และโมเดลเปรียบเทียบ (Augmented) ในโมเดลแรกสมมติให้มีกลุ่มผู้สอบที่แตกต่างกัน ดังนั้น จึงบังคับให้พารามิเตอร์ของข้อสอบระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบมีค่าเท่ากัน (Group l-Item 1 และ Group 2-Item 1) (Thissen et al., 1988; 1993) โมเดลพื้นฐานนี้เป็นการจัดเตรียมค่าความเป็นไปได้พื้นฐานสำหรับความกลมกลืนของพารามิเตอร์ข้อสอบในโมเดล จากนั้นจะกระทำการทดสอบแต่ละข้อที่ทำหน้าที่ต่างกัน โดยแยกข้อมูลวิเคราะห์ในแต่ละข้อซึ่งคล้ายกับการประมาณค่าพารามิเตอร์ข้อสอบที่บังคับให้เท่ากันข้ามกลุ่ม ยกเว้นค่าพารามิเตอร์ของข้อสอบเริ่มต้นการทดสอบการทำหน้าที่ต่างกันของข้อสอบ โมเดลเปรียบเทียบจัดเตรียมค่าความเป็นไปได้ที่เกี่ยวข้องกับการประมาณค่าพารามิเตอร์ข้อสอบข้อที่ i ในแต่ละกลุ่มประชากรค่าความเป็นไปได้นี้สามารถเปรียบเทียบทุกพารามิเตอร์ของข้อสอบทั้งหมดที่บังคับให้เท่ากันข้ามกลุ่มในโมเดลพื้นฐาน

สมมติฐานในการทดสอบสำหรับโมเดลของพารามิเตอร์ของข้อสอบโดยใช้สูตร ดังนี้

H_0: $a_{jR} = a_{jF}$ และ $b_{jR} = b_{jF}$ สำหรับทุก j

H_A: พารามิเตอร์ของข้อสอบในข้อที่ j ของทั้งสองกลุ่มไม่เท่ากันอย่างน้อย 1 พารามิเตอร์

สูตรการทดสอบอัตราส่วนความเป็นไปได้ของสองโมเดล แสดงได้ดังนี้

$$LR = \frac{L^{*}(Model_C)}{L^{*}(Model_A)} \qquad [2\text{-}8]$$

เมื่อ $L^{*}(Model_C)$ แทน ฟังก์ชันความเป็นไปได้ของโมเดลพื้นฐาน (จะเป็นค่าพารามิเตอร์ที่น้อยกว่า)

$L^{*}(Model_A)$ แทน ฟังก์ชันความเป็นไปได้ของโมเดลเปรียบเทียบที่ยอมให้พารามิเตอร์ข้อสอบของข้อสอบข้อที่ j ข้ามกลุ่มผู้สอบมีความหลากหลาย

นั่นคือ ในโมเดลพื้นฐาน (Compact) จะประกอบด้วย ข้อสอบที่ทำหน้าที่ไม่ต่างกัน สำหรับในโมเดลเปรียบเทียบจะประกอบด้วย ข้อสอบที่มีค่าพารามิเตอร์ระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีค่าแปรเปลี่ยนไปตามกลุ่มโมเดลเปรียบเทียบ (Augmented) อาจมีข้อสอบจำนวน 1 ข้อ หรือมากกว่าที่ทำหน้าที่ต่างกัน นอกจากนี้ ระหว่าง 2 โมเดล จะต้องมีข้อสอบร่วมซึ่งเป็นข้อสอบที่สมมติว่าทำหน้าที่ไม่ต่างกัน จากนั้นจึงทำการเปรียบเทียบระหว่าง 2 โมเดล ด้วยสถิติอัตราส่วนความเป็นไปได้ ดังนี้

$$G_j^2 = -2\mathrm{Ln}(LR)$$
$$G^2 = -2\mathrm{Ln}(L^*(\mathrm{Model}_C) - (-2\mathrm{Ln}(L^*(\mathrm{Model}_A)) \qquad [2\text{-}9]$$

เมื่อ G_j^2 แทน สถิติอัตราส่วนความเป็นไปได้ของข้อสอบข้อที่ i

$L^*(\mathrm{Model}_C)$ แทน ฟังก์ชันความเป็นไปได้ของโมเดลพื้นฐาน (Compact)

$L^*(\mathrm{Model}_A)$ แทน ฟังก์ชันความเป็นไปได้ของโมเดลเปรียบเทียบ (Augmented)

โดยทั่วไปแล้ว $L^*(\mathrm{Model}_c) < L^*(\mathrm{Model}_A)$ และสถิติ $G_j^2 > 0$ มีการแจกแจงแบบ ไค-สแควร์ ซึ่งระดับของความเป็นอิสระเท่ากับจำนวนพารามิเตอร์ในโมเดล $L^*(\mathrm{Model}_c)$ และโมเดล $L^*(\mathrm{Model}_A)$ ดัชนีความไม่กลมกลืนในระดับนัยสำคัญทางสถิติของผลการทดสอบจะชี้บอกว่าโมเดลพื้นฐานกลมกลืนน้อยกว่าโมเดลเปรียบเทียบอย่างมีนัยสำคัญทางสถิติในการใช้การทดสอบอัตราส่วนความเป็นไปได้ ค่าไค-สแควร์ของข้อสอบเหล่านั้นมีนัยสำคัญทางสถิติ แสดงว่ามีการทำหน้าที่ต่างกันของข้อสอบ

ในการเปรียบเทียบค่าพารามิเตอร์ในโมเดลที่ตั้งสมมติฐานให้มีค่าเท่ากันระหว่างกลุ่มเปรียบเทียบกับกลุ่มอ้างอิง สถิติที่ใช้ทดสอบอัตราส่วนความควรจะเป็น คือ -2 log likelihood (-2LL) และนำค่าที่ได้ไปเทียบการแจกแจง χ^2 กับ df เท่ากับจำนวนค่าพารามิเตอร์ที่ทดสอบ ถ้าผลปรากฏว่าปฏิเสธสมมติฐานว่างแสดงว่าข้อสอบทำหน้าที่ต่างกัน นั่นคือ ค่าพารามิเตอร์อำนาจจำแนก (a_i) หรือ ค่าพารามิเตอร์ลำดับขั้น (b_{jk}) หรือทั้งสองค่า มีค่าไม่เท่ากันระหว่างกลุ่มอ้างอิงกับกลุ่มสนใจ (Woods, 2009) ค่า Chi-squared (χ^2) จะมี Degree of Freedom (df) เท่ากับจำนวนตัวแปรอิสระหรือจำนวนพารามิเตอร์ของข้อสอบ ค่า χ^2 สามารถคำนวณ ได้ดังนี้

$$\chi^2 = -2[LL(C)-LL(A)] \qquad [2\text{-}10]$$

โดยที่ LL (C) คือ ค่า Log-likelihood เมื่อ โมเดลไม่มีข้อจำกัด (Unrestricted model) และ LL (A) คือ ค่า Log-likelihood เมื่อโมเดลมีข้อจำกัด (Restricted model) เนื่องจากค่า LL (C) มากกว่า LL (A) ดังนั้น ค่า χ^2จึงมีค่าเป็นบวกเสมอ การทดสอบอัตราความควรจะเป็นมีลักษณะคล้ายกับการทดสอบแบบ SIBTEST ในส่วนของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สถิติที่ใช้ของ Log-Likelihood เขียนสูตรได้ดังนี้ (Finch & French, 2007)

$$LL_{equal} = \sum_{G=1}^{2}\sum_{p=1}^{N} \ln\left[\sum_{1}^{q}\prod_{i=1}^{nitems}\left(T_{iG}\left(\mu_{ipG}\right)\phi_G(\theta)d\theta\right)\right] \qquad [2\text{-}11]$$

เมื่อ $T_{iG}\left(\mu_{ipG}\right)$ แทน ค่าพารามิเตอร์ของกลุ่ม G ที่ถูกกำหนดให้มีค่าพารามิเตอร์เท่ากันสำหรับทั้งสองกลุ่ม

$\phi_G(\theta)$ แทน การแจกแจกของพารามิเตอร์คุณลักษณะของผู้ตอบ (θ) ของกลุ่ม G

ให้ U_{jki} แทน เมตริกซ์ของการตอบสนองข้อสอบ (Response pattern) คนที่ i U_{jki} = 1 เมื่อตอบตัวเลือก k ในข้อที่ j และตอบตัวเลือกอื่น ๆ ในข้อที่ j ให้ U_{jki} = 0 โดยให้ผู้ตอบเป็นอิสระกัน ความน่าจะเป็นแบบมีเงื่อนไขของการตอบสนองข้อสอบคนที่ i กำหนดให้เป็น θ การตอบตัวเลือก m และจำนวนข้อทั้งหมด n กำหนดให้หา ความน่าจะเป็นร่วมของเมตริกซ์ $U_i = (U_{jki})_i$ (Joint probability) ดังนี้

$$P(U_i|\theta) = P\left[\left(U_{jk}\right)\theta\right] = \prod_{j=1}^{n}\prod_{k=1}^{m}\left[P_{jk}(\theta)\right]^{jki} \qquad [2\text{-}12]$$

เมื่อ U_{jki} = 1 ถ้า เลือกตอบตัวเลือกที่ k ในข้อที่ j และมีค่าเป็น 0 เมื่อเลือกตัวเลือก อื่น ๆ สำหรับผู้สอบที่ทำการสุ่มจากประชากรในการแจกแจงปกติของคุณลักษณะแฝง และ $\phi(\theta)$ แทน ความน่าจะเป็นที่จะเกิดเหตุการณ์ในการตอบข้อที่ j (Marginal probability) คือ

$$P_i\left[\left(U_{jk}\right)\right] = \int_{-\infty}^{\infty} P_i\left[\left(U_{jk}\right)\theta\right]\phi(\theta)d(\theta) \qquad [2\text{-}13]$$

ถ้าผู้ตอบตอบข้อคำถามทั้งหมด n ข้อที่มีตัวเลือกทั้งสิ้น m ตัวเลือก รูปแบบการตอบสนองข้อสอบของคนที่ i สามารถระบุเป็นหนึ่งใน m^n ของรูปแบบทั้งหมด ดังนั้น กำหนดให้ r_i แทนจำนวนการตอบสนองข้อสอบ ในรูปแบบที่ i และ ให้ N แทน จำนวนตัวอย่างผู้ตอบทั้งหมด ที่สุ่มมาจากประชากร ดังนั้น r_i แทน การแจกแจงแบบพหุนาม (Multinomial distribution) กับพารามิเตอร์ N และ $P_i[U_{jk})]$ ดังสมการต่อไปนี้

$$L = \frac{N!}{\prod_{i=1}^{m^n} r!} \prod_{i=1}^{m^n} \left\{ P_i \left[\left(U_{jk} \right) \right] \right\}^{\eta} \qquad [2\text{-}14]$$

$$\ln L = \ln N! - \sum_{i=1}^{m^n} \ln r_i! + \sum_{i=1}^{m^n} \ln r_i P_i \left[\left(U_{jk} \right) \right] \qquad [2\text{-}15]$$

สมการไลด์ลิฮูดสำหรับ $\hat{a}_j$, $\hat{b}_j$ และ $\hat{d}_k$ สามารถหาค่าได้จากอนุพันธ์ย่อยของสมการข้างต้น โดยการแทนค่าและกำหนดให้สมการมีค่าเป็นศูนย์

การประมาณค่าพารามิเตอร์ กำหนดให้ υ_j แทน พารามิเตอร์ a_j หรือ b_j ถ้าแทนค่า υ_k หมายความว่า ค่าพารามิเตอร์ υ_j สำหรับข้อที่ j มีค่าเท่ากับ h สามารถหาค่าอนุพันธ์ของความควรจะเป็นในสมการข้างต้น ได้ดังนี้

$$\frac{\partial \ln L}{\partial \upsilon_h} = \sum_{i=1}^{m^n} \frac{r_i}{P_i(U_{jk})} \int_{-\infty}^{\infty} P_i \left[\left(U_{jk} \right) \theta \right] \sum_{k=1}^{m} \frac{\partial \left[P_i(\theta) \right]^{U_{hki}}}{\partial \upsilon_h} \frac{\varphi(\theta) d(\theta)}{\left[P_{hk}(\theta) \right]^{U_{hki}}} \qquad [2\text{-}16]$$

ผลของการตอบสนองข้อสอบ โดยกำหนด l = 1, 2, …, S เมื่อ $S \leq \min(N, m^n)$

ถ้าจำนวนการตอบของผู้ตอบตัวเลือก l กำหนดให้เป็น r_l แล้ว $\sum_{i=1}^{s} r_l = N$

ในการอนุพันธ์อันดับหนึ่งของความควรจะเป็นในสมการ [2-16] สามารถประมาณค่าโดยใช้วิธีของเกาส์เฮอร์มิทควอดาร์เจอร์ (Gauss-hermite quadrature) ได้ดังนี้

$$\frac{\partial \ln L}{\partial \upsilon_h} \approx \sum_{l=1}^{S} \sum_{f=1}^{F} \frac{r_l L_l(X_f) A(X_f)}{\widetilde{P}_l} \sum_{k=1}^{m} \frac{\partial [P_i(\theta)]^{U_{hkl}}}{\partial \upsilon_h} \frac{1}{[P_{hk}(X_f)]^{U_{hkl}}} \quad [2\text{-}17]$$

เมื่อ $P = \sum_{f=1}^{F} L_1(X_f) A(X_f)$ [2-18]

และ $L_1(X_f) = \prod_{j=1}^{n} \prod_{k=1}^{m} [P_{jk}(X_f)]^{U_{jkl}}$ [2-19]

ในสมการ [17] A (X_f) แทนค่าน้ำหนักในการหาพื้นที่โดยวิธีเกาส์เฮอร์มิท และ X_f แทนจุดของการหาพื้นที่ (Quadrature point) (Stroud & Secrest, 1996) น้ำหนักของการหาพื้นที่ A (X_f) มีการแจกแจงเป็นปกติมาตรฐานที่จุด X_f จะได้ว่า $\sum_{f=1}^{F} A(X_f) = 1$

เมื่อ F แทนจำนวนจุดในการหาพื้นที่ทั้งหมด เพราะ U_{hkl} มีค่าที่เป็นไปได้สองค่าคือ 1 และ 0 จากสมการ สามารถเขียนใหม่ได้ว่า

$$\sum_{f=1}^{F} \sum_{k=1}^{m} \frac{\bar{r}_{hkf}}{P_{hk}(X_f)} \frac{\partial P_{hk}(X_f)}{\partial \upsilon_h} \quad [2\text{-}20]$$

เมื่อ

$$\bar{r}_{hkf} = \sum_{f=1}^{S} \frac{r_l L_l(X_f) A(X_f) U_{hkl}}{\widetilde{P}_l} \quad [2\text{-}21]$$

และ $\bar{r}_{hkf}$ แทน จำนวนของการตอบตัวเลือก k ของข้อที่ h ณ จุดของการหาพื้นที่ที่ f การหาพารามิเตอร์จุดตัดของการเลือกตอบ (Threshold parameter estimate) ค่าจุดตัดของการเลือกตอบตัวเลือกในข้อที่ h แทนสัญลักษณ์คือ d_g ซึ่งก็คือ พารามิเตอร์ d_k สำหรับแต่ละค่าของตัวเลือกที่ k มีค่าเป็น g ดังนั้น $P_{jk}(\theta)$ (k = 1, 2, …, m) แสดงดังสมการต่อไปนี้ในการหาค่าอนุพันธ์อันดับที่หนึ่งของฟังก์ชันความควรจะเป็นในสมการ (9) หาค่าเป็น d_g

$$\frac{\partial L_1(X_f)}{\partial d_g} = L_1(X_f)\sum_{j=1}^{n}\sum_{k=1}^{m}\frac{U_{jkl}}{P_{jk}(X_f)}\frac{\partial P_{jk}(X_f)}{\partial d_g} \qquad [2\text{-}22]$$

จากที่กล่าวมาสรุปได้ว่า แนวคิดของวิธี IRT LRT คือ การทดสอบความแตกต่างของโมเดลผลการตอบข้อสอบระหว่างผู้สอบกลุ่มย่อยครั้งละ 2 กลุ่ม (Contras group) หากทดสอบว่าคู่ใดแตกต่างอย่างมีนัยสำคัญ บ่งชี้ว่าข้อสอบข้อนั้นทำหน้าที่ต่างกัน วิธีทดสอบอัตราส่วนความควรจะเป็นตามทฤษฎีการตอบสนองข้อสอบในรูปของล็อกเชิงเส้น (Loglinear: LR) เป็นการประมาณค่าพารามิเตอร์ด้วยวิธีความน่าจะเป็นสูงสุด (Maximum likelihood: ML) ใช้สถิติทดสอบไคว์-สแควร์ χ^2ในการทดสอบ

2. วิธี MGCFA-ALIGNMENT

วิธี Multiple group alignment หรือในการวิจัยครั้งนี้ใช้คำว่า วิธี MGCFA-ALIGNMENT เป็นวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยประยุกต์มาจากการวิเคราะห์องค์ประกอบเชิงยืนยันแบบหลายกลุ่ม (MGCFA) นำเสนอโดย Asparouhov and Muthén (2014) วิธี MGCFA นำมาประยุกต์ใช้สำหรับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ภายใต้ทฤษฎีการตอบสนองข้อสอบ (IRT) ที่มีการเชื่อมโยงโดยตรงระหว่างค่าพารามิเตอร์จากการวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) โดยค่าความยากของข้อสอบตามทฤษฎีการตอบสนองข้อสอบ (IRT) จะเชื่อมโยงค่า CFA Item thresholds และค่าอำนาจจำแนกตามทฤษฎีการตอบสนองข้อสอบ (IRT) จะเชื่อมโยงกับ CFA Factor loadings

รูปแบบ CFA สำหรับการตอบสนองข้อสอบที่ให้คะแนนแบบ 2 ค่า ซึ่งเป็นตัวแปรสังเกตได้สามารถแสดงเป็นสมการ

$$y_{kig} = \nu_{kg} + \lambda_{kg}\eta_{ig} + \varepsilon_{kig} \qquad [2\text{-}23]$$

เมื่อ y_{kig} = การตอบสนองต่อข้อที่ k ของผู้สอบคนที่ i ในกลุ่ม g

ν_{kg} = ค่า Threshold สำหรับข้อที่ k ในกลุ่ม g (ค่าความยาก)

λ_{kg} = ค่า Factor loading สำหรับข้อที่ k ในกลุ่ม g (ค่าอำนาจจำแนก)

η_{ig} = คุณลักษณะแฝงของผู้สอบคนที่ i ในกลุ่ม g

ε_{kg} = ความคลาดเคลื่อนจากการสุ่มของผู้สอบคนที่ i ของข้อที่ k ในกลุ่ม g

จากนั้น วิธี Alignment มีขั้นตอนต่อไปนี้

1. ประมาณค่า The configural factor model ให้ λ_{kg} และ V_{kg} แตกต่างกันในกลุ่ม แต่กำหนดโครงสร้างขององค์ประกอบภายในกลุ่มให้เหมือนกัน เรียกว่าเป็นโมเดล M0

2. แปลงค่าเฉลี่ยขององค์ประกอบ α_g เป็น 0 และความแปรปรวนขององค์ประกอบ ψ_g เป็น 1 สำหรับแต่ละกลุ่ม

3. ค่าความแปรปรวนรายข้อ (V) และค่าเฉลี่ย ($E(y_{kg})$) ในเทอมของความแปรปรวนขององค์ประกอบ ค่าเฉลี่ย และค่าพารามิเตอร์รายข้อสำหรับโมเดล M0
ค่าพารามิเตอร์สำหรับ M0:

$$V(y_{kig}) = \lambda_{kg}^{2}\psi_g = \lambda_{kg0}^{2} \qquad [2\text{-}24]$$

$$E(y_{kig}) = \nu_{ik} + \lambda_{kg}\alpha_g = \nu_{kg0} \qquad [2\text{-}25]$$

ค่า Factor loading และค่า Threshold ของโมเดล M0 คือ

$$\lambda_{kg0} = \lambda_{kg}\sqrt{\psi_g} \qquad [2\text{-}26]$$

$$\nu_{kg0} = \nu_{kg} + \frac{\lambda_{kg0}}{\sqrt{\psi_g}} \qquad [2\text{-}27]$$

4. สำหรับทุก α_g และ ψ_g มีจำนวนอนันต์ของชุดของค่า Threshold และ ค่า Factor loading ที่ทำให้ข้อมูลสอดคล้องกับโมเดล M0 โดยไม่ต้องกำหนดให้ค่า Factor loading หรือค่า Threshold เท่ากันระหว่างกลุ่ม เป้าหมายของวิธี Alignment คือการหาโมเดลทางเลือก M1 ที่ค่าเหมือนกัน (หรือใกล้เคียง) สอดคล้องกับโมเดล M0 ดังนั้น การใช้ชุดขององค์ประกอบ และ thresholds ก็คล้ายกันระหว่างกลุ่มก็เป็นไปได้ ซึ่งแตกต่างกับกรณี M0 Factor loading และ thresholds สำหรับโมเดล M1 สามารถแสดงในแบบของ Factor loading M0 และ Thresholds

$$\lambda_{kg1} = \frac{\lambda_{kg0}}{\sqrt{\psi_g}} \quad [2\text{-}28]$$

$$\nu_{kg1} = \nu_{kg0} - \frac{\lambda_{kg0}}{\sqrt{\psi_g}}\alpha_g \quad [2\text{-}29]$$

5. ค่าของ α_g และ ψ_g สำหรับขั้นตอนที่ 4 ถูกคัดเลือกเพื่อลดปริมาณของความแตกต่างระหว่างกลุ่มในค่าพารามิเตอร์รายข้อ (เช่น λ_{kg1} และ ν_{kg1}) ซึ่งจะทำโดยการลดการสูญเสียการทำงานของฟังก์ชัน F

$$F = \sum_{p} \sum_{g1<g2} w_{g1.g2} f(\lambda_{pg1.g1} - \lambda_{pg1.g2}) + \sum_{p} \sum_{g1<g2} w_{g1.g2} f(\nu_{pg1.g1} - \nu_{pg1.g2}) \quad [2\text{-}30]$$

เมื่อ

$$f(x) = \sqrt{\sqrt{x^2 + 0.0001}}$$

$w_{g1.g2} = \sqrt{N_{g1} N_{g2}}$; การถ่วงขนาดน้ำหนักของกลุ่ม

สำหรับทุกคู่ของกลุ่ม (เช่น g_1, g_2) F จะเพิ่มขึ้นจากความแตกต่างในการประมาณค่าพารามิเตอร์ วิธี Alignment ถูกออกแบบเพื่อให้ได้รูปแบบที่มีจำนวนของพารามิเตอร์เพียงเล็กน้อย แสดงให้เห็นถึงความแตกต่างกันในแต่กลุ่มและส่วนใหญ่ค่าพารามิเตอร์จะมีค่าเท่ากันในแต่ละกลุ่ม เพื่อให้มั่นใจว่าการระบุรูปแบบคำสั่งที่ไม่ต้องการใช้ Unwanted command used in TEX: Pleaseremove $\prod^{\psi_g} = 1$. ค่า α_g สามารถกำหนดให้เป็นอิสระ (Free alignment) หรือคงที่ เท่ากับ 0 สำหรับทุกกลุ่ม (Fix alignment)

เมื่อการจัดเรียงรูปแบบเสร็จสมบูรณ์ ค่าพารามิเตอร์ของรายบุคคลจะถูกทดสอบการแปรเปลี่ยน (Lack of DIF) ข้ามกลุ่ม การประเมิน DIF จะกระทำโดยใช้วิธีต่อไปนี้:

1. สำหรับพารามิเตอร์ที่กำหนด เช่น i.e., Loading/ Discrimination, Threshold/ Difficulty กับข้อสอบที่ต้องการตรวจสอบเปรียบเทียบค่าประมาณสำหรับคู่ของกลุ่มแต่ละคนด้วยสมมติฐานหลัก กลุ่มที่มีการเชื่อมต่อกับพารามิเตอร์ ถ้า $p > .01$ สำหรับสถิติการทดสอบ

2. การตั้งค่าการเชื่อมต่อที่ใหญ่ที่สุดของกลุ่มจะถูกระบุสำหรับพารามิเตอร์ข้อสอบที่ต้องการตรวจสอบ

3. ค่าเฉลี่ยของพารามิเตอร์ข้อสอบที่ต้องการตรวจสอบที่มีการคำนวณสำหรับการตั้งค่าการเชื่อมต่อที่ระบุไว้ในขั้นตอนที่ 2

4. พารามิเตอร์ข้อสอบที่ต้องการตรวจสอบสำหรับแต่ละกลุ่มจะถูกเปรียบด้วยค่าเฉลี่ยสำหรับค่าการเชื่อมต่อโดยใช้สถิติทดสอบ ถ้า $p > 001$ กลุ่มที่จะถูกเพิ่มในการตั้งค่าการเชื่อมต่อและถ้าไม่มีกลุ่มใดถูกลบออกจากชุดที่เกี่ยวโยงกัน

5. ทำซ้ำขั้นตอนที่ 3 และ 4 จนถึงการเชื่อมโยงไม่เปลี่ยนแปลง

6. กลุ่มคู่ใดคู่หนึ่งที่ไม่ได้อยู่ในชุดที่เกี่ยวโยงกันและอื่น ๆ ที่มีการกล่าวถึง บ่งชี้ว่าข้อสอบที่ต้องการตรวจสอบทำหน้าที่ต่างกัน (DIF) เป็นรายข้อ

การประมาณค่าพารามิเตอร์สำหรับ The alignment สามารถประมาณค่าด้วยวิธี Maximum likelihood หรือ วิธี Bayessian ตามกรอบแนวคิดของ Markov chain monte carlo (MCMC) วิธี Bayessian เป็นวิธีการที่ดูเป็นที่นิยมและถูกนำมาใช้ในงานวิจัย การวิเคราะห์ให้ทั้งค่าพารามิเตอร์และการทดสอบความเท่าเทียมโดยอัตโนมัติ ให้ผลการเปรียบเทียบจากจำนวนของกลุ่มแต่ละกลุ่มด้วย จึงได้ดำเนินตามขั้นตอนสำหรับกลุ่มเปรียบเทียบและกลุ่มอ้างอิง

วิธี MGCFA-ALIGNMENT เป็นวิธีที่สามารถวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแบบหลายกลุ่มได้ ความแกร่งของวิธีนี้คือสามารถประมาณค่าโมเดลได้อย่างสะดวกสำหรับหลาย ๆ กลุ่ม เช่น การเปรียบเทียบระหว่างหลายประเทศ พอสรุปได้ว่าวิธีนี้มีจุดเด่นอยู่ 4 ประการ คือ 1) เป็นวิธีที่เปรียบเทียบแต่ละกลุ่มโดยอัตโนมัติในขั้นตอนเดียว 2) ให้ค่าพารามิเตอร์ทั้งค่าความยาก และค่าอำนาจจำแนกของข้อสอบในการตรวจหา DIF ในหลายกลุ่ม และ 3) สามารถตรวจสอบได้ทั้ง Uniform และ Nonuniform DIF 4) การเตรียมข้อมูลและการวิเคราะห์ข้อมูลไม่ซับซ้อน

เพื่อให้สามารถเชื่อมต่อความรู้ ความเข้าใจเกี่ยวกับการประยุกต์ใช้วิธี MGCFA สำหรับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ผู้วิจัยขอนำเสนอความรู้เกี่ยวกับ การวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory factory analysis: CFA) และ MGCFA ดังนี้

1. การวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory factory analysis: CFA) เป็นเทคนิคการทดสอบความสอดคล้องระหว่างข้อมูลกับโมเดล ที่กำหนดให้มีตัวแปรคุณลักษณะแฝงที่อยู่เบื้องหลัง ตัวแปรที่สังเกตได้ทั้งชุดมีวัตถุประสงค์เช่นเดียวกับ EFA คือ 1) เพื่อตรวจสอบทฤษฎีที่ใช้เป็นพื้นฐานใน การวิเคราะห์องค์ประกอบ 2) เพื่อสำรวจและระบุองค์ประกอบ และ 3) ใช้เป็นเครื่องมือในการสร้างตัวแปรใหม่ โดยใช้โมเดลการวัด (Measurement model) ของโมเดล

โครงสร้างความแปรปรวนร่วม (Covariance structure analysis) (สุวิมล ติรกานนท์, 2551, หน้า 9-63) CFA มีการปรับปรุงจุดอ่อนของ EFA ทำให้ผลการวิเคราะห์ที่ได้มีความสมเหตุสมผลตรงตามความเป็นจริงมากขึ้น เช่น ความคลาดเคลื่อนอาจสัมพันธ์กันได้ เป็นต้น และต้องมีทฤษฎีสนับสนุน ในการกำหนดเงื่อนไขบังคับในการวิเคราะห์หาค่าน้ำหนักองค์ประกอบ (Factor loading) ทั้งยังมีการตรวจสอบโครงสร้างของโมเดลว่ามีความแตกต่างกันระหว่างกลุ่มตัวอย่างหลาย ๆ กลุ่มหรือไม่ (Joreskog & Sorbom, 1996, pp. 409-426; นงลักษณ์ วิรัชชัย, 2542, หน้า 40)

CFA มีข้อตกลงเบื้องต้นที่ควรตรวจสอบก่อนการ วิเคราะห์ CFA ได้แก่ 1) ข้อมูลควรมีลักษณะการแจกแจงปกติ (Normal distributions) 2) มีความเป็นเอกพันธ์ของการกระจาย (Homoscedasticity) 3) มีความสัมพันธ์ระหว่างตัวแปรแต่ละคู่เป็นแบบเส้นตรง (Linear relationships) 4) ความคลาดเคลื่อน (Error terms) ต้องไม่สัมพันธ์กับตัวแปรแฝงใด ๆ ในโมเดล

ขั้นตอนการวิเคราะห์ CFA

1. การเตรียมเมทริกซ์สหสัมพันธ์ การกำหนดข้อมูลจำเพาะของโมเดล และระบุความเป็นไปได้ค่าเดียวของโมเดล

การระบุความเป็นไปได้ค่าเดียวของโมเดล CFA มีความสำคัญต่อการประมาณค่าพารามิเตอร์ของโมเดลมีความสำคัญต่อการประมาณค่าพารามิเตอร์ของโมเดล เพราะการประมาณค่าพารามิเตอร์จะทำได้ต่อเมื่อโมเดลระบุความเป็นได้ค่าเดียวพอดี ซึ่งหมายความว่าการแก้สมการหาค่าตัวไม่ทราบค่าจะได้รากของสมการที่เป็นได้ค่าเดียว การกำหนดเงื่อนไขบังคับ (Constraints) ในการวิเคราะห์ทำได้ 2 แบบ คือ การตั้งเงื่อนไขให้พารามิเตอร์เป็นพารามิเตอร์กำหนด และการตั้งเงื่อนไขให้พารามิเตอร์เท่ากัน เงื่อนไขบังคับจะทำให้จำนวนพารามิเตอร์หรือตัวไม่ทราบค่าลดลงและโมเดลมีโอกาสระบุได้พอดีมากขึ้น (นงลักษณ์ วิรัชชัย, 2542, หน้า 153)

2. การสกัดองค์ประกอบขั้นต้น

3. การหมุนแกน

4. การสร้างสเกลองค์ประกอบ

2. Multiple-group CFA

ธีรเดช ฉายอรุณ (2560) ได้อธิบายไว้ว่า กลุ่ม (Group) ในที่นี้หมายถึง ตัวแปรจำแนกประเภท (Categorical variable) ที่ไม่ใช่ตัว แปรตามหรือตัวแปรภายใน หากแต่เป็นตัวแปรที่ทำให้กลุ่มตัวอย่างของเราถูกแบ่งออกเป็นกลุ่มย่อย ๆ ตามตัวแปรเหล่านั้นแบ่งได้เป็น

กลุ่มที่วัดได้โดยตรง (Observed groups) ได้แก่ เพศ เชื้อชาติ อายุ ภูมิลำเนา ประเทศ ฯลฯ

กลุ่มแฝงหรือกลุ่มที่วัดไม่ได้โดยตรง (Unobserved groups) เช่น ชนชั้นทางสังคม (Social class) กลุ่มนักดื่ม หรือ กลุ่มผู้ใช้สารเสพติด เป็นต้น

การวิเคราะห์ Multi groups SEM จะทำกับกลุ่มที่วัดได้โดยตรงเท่านั้นสำหรับกลุ่มแฝงต้องใช้ Latent mixture modeling

Multiple-group CFA คือการทดสอบว่าแบบจำลองการวัดที่เราพัฒนาขึ้นนั้นจะไม่แปรเปลี่ยน (Invariance/ equivalence) ไปตามกลุ่มย่อย

การไม่แปรเปลี่ยนไปตามกลุ่มย่อยพิจารณาจากดัชนีต่อไปนี้

เมื่อใช้ในกรณีของการพัฒนาเครื่องมือวัด (แบบวัด) การทดสอบดังกล่าวจะทดสอบว่าโครงสร้างของแบบวัดที่พัฒนาขึ้นมาจากทฤษฎีนั้น จะไม่แปรเปลี่ยนไปตามกลุ่มย่อย เป็นการทดสอบที่บังคับให้น้ำหนักตัวประกอบ (Λx) และ item intercepts ไม่ผันแปรในทุกกลุ่มย่อย หาก Item intercepts แปรเปลี่ยนไปในแต่ละกลุ่มย่อยแล้ว เป็นไปได้ว่ามีอย่างน้อย 1 กลุ่มที่จะมีโอกาสได้คะแนนแต่ละข้อของแบบวัดนี้สูงหรือต่ำกว่ากลุ่มอื่น (ในทางการวัดเรียกว่าเกิด Systematic error หรือ Bias) แม้ว่าน้ำหนักตัวประกอบจะไม่แปรเปลี่ยนก็ตาม มีผลให้ระบบการให้คะแนนไม่สามารถใช้ในทุกกลุ่มย่อยได้และการวิเคราะห์ที่ซับซ้อนกว่านี้จะมีปัญหาตามมา หาก Item intercepts ไม่แปรเปลี่ยน ระบบการให้คะแนนและการตีความสามารถใช้เหมือนกันได้ทุกกลุ่มย่อย

จากที่กล่าวมาสรุปได้ว่า วิธี MGCFA-ALIGNMENT เป็นวิธีที่ประยุกต์มาจาก MGCFA ที่มีข้อจำกัดคือไม่สามารถวิเคราะห์ได้เมื่อมีหลายกลุ่มและมีดัชนีปรับแต่งจำนวนมากเกินไป โมเดลใช้สำหรับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ มีการกำหนดโมเดลการเชื่อมโยงค่าพารามิเตอร์ของข้อสอบทั้งค่าความยากและค่าอำนาจจำแนกของข้อสอบ โดยค่าความยากของข้อสอบจาก IRT คือค่า Item thresholds และค่าอำนาจจำแนกของข้อสอบจาก IRT จะเชื่อมโยงกับค่า Factor loadings จากนั้นมีการกำหนดโมเดลเพื่อปรับระดับ เมื่อมีการจัดเรียงรูปแบบเสร็จสมบูรณ์ จะมีการทดสอบการไม่แปรเปลี่ยนของค่าพารามิเตอร์ข้ามกลุ่ม หากข้อใดมีการแปรเปลี่ยนบ่งชี้ว่าข้อสอบนั้นทำหน้าที่ต่างกัน

วิธี MIMIC

โมเดล MIMIC เป็นโมเดลลิสเรลที่มีตัวแปรแฝงเพียงตัวแปรเดียวโดยที่ตัวแปรแฝงนั้นได้รับอิทธิพลจากตัวแปรภายนอกสังเกตได้หลายตัวแปรและส่งอิทธิพลไปยังตัวแปรภายในสังเกตได้หลายตัวแปร Latent mixture modeling Joreskog and Goldberger (1975) ได้นำเสนอ แบบจำลอง MIMIC สำหรับข้อมูลตัวแปรต่อเนื่อง

การประยุกต์ใช้แบบจำลอง MIMIC ในโมเดล IRT สำหรับการทดสอบ DIF ถูกนำเสนอโดย Múthen (1985; 1988; 1989) และ Múthen et al. (1991) โมเดล MIMIC ช่วยให้สามารถรวมตัว

แปรร่วมหลายแบบเพื่อทดสอบ DIF ได้หลายกลุ่มและในเวลาเดียวกันสามารถควบคุมการทำงานของกลุ่มอื่น ๆ และ/ หรือเพิ่มขึ้นของตัวแปรร่วมอย่างต่อเนื่อง (เช่น อายุ) (Harpole, 2015)

โมเดล MIMIC สำหรับการตรวจสอบ DIF ในโมเดล IRT เป็นไปตามข้อตกลงแบบเบื้องต้นทั้งหมดของโมเดลการวัด IRT โดยกำหนดเบื้องต้นว่าความแปรปรวนของตัวแปรแฝงระหว่างกลุ่มเหมือนกัน (โดยปกติจะกำหนดความแปรปรวนของตัวแปรแฝงให้เป็น 1)

นอกจากนี้เมื่อทดสอบ DIF แบบ Uniform DIF การไม่แปรเปลี่ยนของพารามิเตอร์อำนาจจำแนกต่างกลุ่ม (Across) อย่างไรก็ตาม ข้อตกลงเบื้องต้นนี้ไม่สามารถทำให้การทดสอบ DIF แบบ Nonuniform DIF possible เป็นไปได้ (Woods & Grimm, 2011) ในการทดสอบ DIF จะต้องมีการสร้างสเกลร่วมสำหรับพารามิเตอร์ของข้อสอบระหว่างกลุ่มต่าง ๆ โดยปกติแล้วสิ่งนี้สามารถทำได้โดยใช้เทคนิคข้อสอบร่วม

วิธีมิมิค (MIMIC) เป็นวิธีการทางสถิติวิธีหนึ่งที่นำมาใช้สำหรับการวิเคราะห์คุณภาพของข้อสอบ เป็นวิธีตามทฤษฎี IRT ซึ่งวิธี MIMIC เป็นโมเดลลิสเรลที่มีตัวแปรแฝงเพียงตัวแปรเดียว โดยที่ตัวแปรแฝงนั้นได้รับอิทธิพลจากตัวแปรภายนอกสังเกตได้หลายตัวแปรและส่งอิทธิพลไปยังตัวแปรภายในสังเกตได้หลายตัวแปร นั่นคือเป็นโมเดลลิสเรลของคุณลักษณะแฝงที่มีหลายสาเหตุและวัดได้จากตัวบ่งชี้หลายตัว โดยมีข้อตกเบื้องต้นว่าการวัดตัวแปรภายนอกสังเกตได้ต้องไม่มีความคลาดเคลื่อนในการวัด โมเดลมิมิคนี้เป็นประโยชน์มากในการตรวจสอบความเป็นเอกมิติ (Unidimensionality) ในการวิจัยสาขาในการวัดผลการศึกษา สามารถวิเคราะห์ ค่าพารามิเตอร์คุณลักษณะข้อสอบ และค่าความสามารถของผู้สอบซึ่งไม่สามารถสังเกตโดยตรงจึงต้องประมาณจากการตอบข้อสอบ แล้วใช้การประมาณค่าพารามิเตอร์ในทฤษฎีการตอบสนองข้อสอบ การวิเคราะห์ MIMIC เป็นวิธีการทางสถิติที่ใช้เทคนิคการวิเคราะห์การถดถอยพหุคูณและการวิเคราะห์องค์ประกอบเข้าด้วยกัน โดยมีวัตถุประสงค์ของการวิเคราะห์เพื่อสร้างสมการทำนายตัวแปรเกณฑ์ที่เป็นตัวแปรแฝงด้วยตัวแปรสังเกตได้ที่เป็นตัวแปรต่อเนื่อง

โมเดลความสัมพันธ์โครงสร้างเชิงสาเหตุ (Causal structural models) (สุพัฒนา หอมบุปผา, 2556, หน้า 60)

แบบจำลองกลุ่มนี้ประกอบด้วยแบบจำลองความสัมพันธ์ทั้งแบบที่มีและไม่มีความคลาดเคลื่อนในการวัด แบบจำลองที่ไม่มีความคลาดเคลื่อนในการวัดจะประกอบขึ้นด้วยตัวแปรสังเกตได้ทั้งหมด โดยไม่มีตัวแปรแฝง เขียนรูปสมการได้ดังต่อไปนี้

$$Y = \beta Y + \Gamma X + \varsigma \qquad [2\text{-}31]$$

หรือเขียนในรูปเมทริกซ์ได้ดังนี้

$$[Y] = [BE][Y] + [GA][X] + [z]$$

จากลักษณะสมการโมเดลผู้อ่านจะเห็นได้ว่าโมเดลกลุ่มนี้ไม่มีตัวแปรแฝงทั้งที่เป็นตัวแปรภายในและตัวแปรภายนอก ดังนั้น ในการวิเคราะห์ข้อมูล เมทริกซ์พารามิเตอร์ LY, LX, TD และ TE จึงมีค่าเป็นศูนย์ทั้งหมดการกำหนดข้อมูลจำเพาะของโมเดลกำหนดรูปแบบและสถานะของเมทริกซ์ GA, BE, PH และ PH เท่านั้น โมเดลความสัมพันธ์โครงสร้างเชิงสาเหตุที่มีความคลาดเคลื่อนในการวัดจะมีตัวแปรครบทุกประเภทได้ตามโมเดลใหญ่ในโปรแกรมลิสเรล เมื่อเขียนในรูปสมการจะประกอบด้วยสมการการวัดสองสมการ และสมการโมเดลโครงสร้างหนึ่งสมการ ดังนี้

$$[X] = [LX][K] + [d]$$
$$[Y] = [LY][E] + [e]$$
$$[E] = [BE][E] + [GA][K] + [z]$$

แบบจำลองกลุ่มนี้ยังแบ่งออกได้เป็น 3 แบบ ได้แก่

1. Regression models and ANOVA models
2. Path analysis
3. Multiple indicators and multiple causes models หรือ MIMIC Models

โมเดลมิมิค (MIMIC Model)

MIMIC เป็นคำที่ย่อมาจาก Multiple indicators and multiple causes ซึ่งหมายถึงโมเดลลิสเรลที่มีตัวแปรแฝงเพียงตัวแปรเดียว โดยที่ตัวแปรแฝงนั้น ได้รับอิทธิพลจากตัวแปรภายนอกสังเกตได้หลายตัวแปร และส่งอิทธิพลไปยังตัวแปรภายในสังเกตได้หลายตัวแปร กล่าวอีกอย่างหนึ่งคือ เป็นโมเดลลิสเรลของคุณลักษณะแฝงที่มีหลายสาเหตุและวัดได้จากตัวบ่งชี้หลายตัว ดังแสดงดังภาพที่ 2-14 ในที่นี้มีตัวบ่งชี้ 3 ตัวแปร และมีตัวแปรสาเหตุ 3 ตัวแปร ตามลักษณะโมเดลจะเห็นว่าการวัดตัวแปรภายนอกสังเกตได้ ต้องมีข้อตกลงข้างต้นว่า ไม่มีความคลาดเคลื่อนในการวัดและในการวิเคราะห์ข้อมูลจะกำหนดข้อมูลจำเพาะ เฉพาะรูปแบบและสถานะของเมทริกซ์ PH, BE, GA, PS, LY และ TE เท่านั้นส่วนเมทริกซ์ TD และ LX มีค่าเป็นศูนย์ทั้งหมด โมเดลมิมิคนี้เป็น

ประโยชน์มากในการตรวจสอบความเป็นเอกมิติ (Unidimensionality) ในการวิจัยสาขาในการวัดผลการศึกษา (นงลักษณ์ วิรัชชัย, 2542)

การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธี MIMIC

วิธี MIMIC ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) สามารถแบ่งออกเป็นการวัดองค์ประกอบและโครงสร้างองค์ประกอบ ในองค์ประกอบการวัด γ_i^* ของข้อที่ i ลักษณะของตัวแปรแฝง Y ที่ทดสอบเป็นการออกแบบการวัด และกลุ่มของตัวแปร z (ในที่นี้เป็นการศึกษาเพียง 1 กลุ่มตัวแปร) ที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบ (DIF) ในการวิเคราะห์องค์ประกอบของโมเดล ดังนี้

สูตรสำหรับวิธี MIMIC ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) (Muthén, Kao, Chih-Fen & Burstein, 1991) คือ

$$y_i^* = \lambda_i \theta + \beta_i' z + \varepsilon_i \qquad [2\text{-}32]$$

เมื่อ y_i^* คือ ข้อที่ i

θ คือ องค์ประกอบ

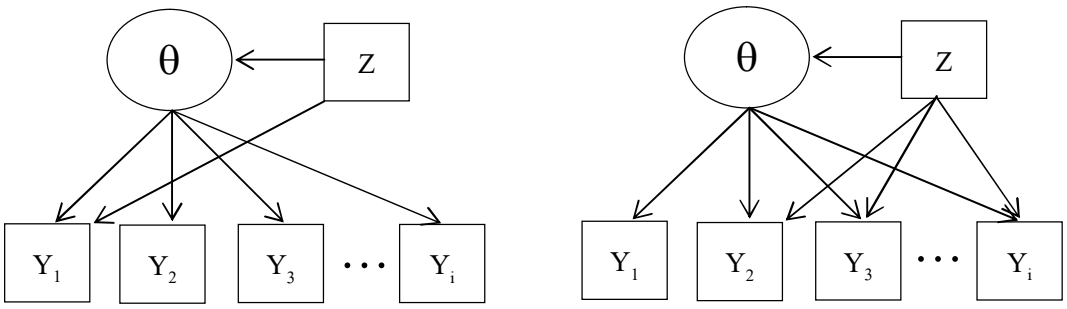
β_i' คือ สัมประสิทธิ์ของตัวแปรสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน/ขนาดโรงเรียน

z คือ กลุ่มสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน และขนาดโรงเรียน

λ_i คือ น้ำหนักองค์ประกอบ

ε_i คือ ค่าความแปรปรวน

เมื่อ λ_i เป็นน้ำหนักองค์ประกอบและเกี่ยวข้องกับความชันของพารามิเตอร์ของข้อสอบข้อที่ i ในบริบทของทฤษฎีการตอบสนองข้อสอบ (IRT) แล้ว ε_i มีการแจกแจงแบบปกติสำหรับ Ordinal probit และการแจกแจงแบบโลจิสติก สำหรับ Ordinal logit และ β_i คือ อิทธิพลของกลุ่มตัวแปร z ต่อ y_i^* ถ้า $\beta_i = 0$ แล้วข้อสอบข้อที่ i มีค่าเท่ากันในทุก ๆ กลุ่ม ตรงกันข้าม ถ้า $\beta_i \neq 0$ จะเกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF) แบบอเนกรูป หากสมการไม่มีเทอมปฏิสัมพันธ์เป็นตัวทำนาย (θz) ดังนั้น สมการ MIMIC จึงเป็นแบบเอกรูปดังแสดง ในภาพที่ 2-4



[a] [b]

ภาพที่ 2-4 (a) เมื่อข้อสอบข้อที่ i เป็นการศึกษาข้อสอบและข้อสอบอื่น ๆ ที่เป็นข้อสอบร่วมกัน และ (b) เมื่อข้อสอบข้อที่ i เป็นข้อสอบร่วมและข้อสอบอื่น ๆ ที่เป็นข้อสอบที่ศึกษา

ซึ่งวิธีหลายตัวชี้วัดหลายสาเหตุในรูปแบบองค์ประกอบเชิงยืนยัน (MIMIC) เป็นหลักการของ CFA กับตัวแปร (Muthén, 1988) แล้ววิธี MIMIC ยังสามารถนำไปใช้สำหรับการวิเคราะห์ DIF ได้ด้วย (Muthén et al., 1991) ซึ่งผลที่ได้ต้องมีค่าเป็นแบบ 2 ค่า (Dichotomous) ค่าพารามิเตอร์ของตัวชี้วัดไม่ต่อเนื่องเป็นสิ่งที่จำเป็น ในความเป็นจริงแล้ว มีหลายวิธีที่ตัวชี้วัดของค่าพารามิเตอร์เป็นแบบ 2 ค่า (Dichotomous) โดยใช้ฟังก์ชั่นเชื่อมโยงที่เหมาะสม (เช่น การเชื่อมโยงแบบโลจิตหรือโพรบิต) ข้อตกลงเบื้องต้นคือ ตัวแปรแฝงเป็นตัวแปรต่อเนื่องและตัวแปรสังเกตได้เป็นการตอบแบบไบนารี (Binary) เมื่อ $y^{*}_{i}j$ เป็นตัวแปรแฝงแบบต่อเนื่องและตัวแปรสังเกตได้เป็นการตอบแบบไบนารี (Binary) ของข้อสอบ y_{ij} แล้วสามารถเขียนสมการได้ดังนี้

$$y_{ij} = \begin{cases} 1 & \text{if } y^{*}_{ij} > 0 \\ 0 & \text{if } y^{*}_{ij} \leq 0 \end{cases}$$

สูตรสำหรับวิธี MIMIC ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) (Muthén et al., 1991) คือ

$$y^{*}_{ij} = \lambda_{i}\theta_{j} + \beta'_{i}G_{j} + \varepsilon_{ij} \qquad [2\text{-}33]$$

เมื่อ λ_i เป็นน้ำหนักองค์ประกอบของข้อที่ i และ θ_j เป็นลักษณะของตัวแปรแฝง ส่วน β_i เป็นสัมประสิทธิ์ความชันสำหรับความแปรปรวนร่วม G_i ซึ่งเป็นกลุ่มตัวชี้วัดของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) และ ε_{ij} เป็นเศษเหลือ นอกจากนี้ โมเดลการถดถอยเป็นสิ่งจำเป็นสำหรับการพยากรณ์ตัวแปรแฝง θ โดยกลุ่มของตัวชี้วัด G_j เพื่อควบคุมความแตกต่างในลักษณะตัวแปรแฝงข้ามกลุ่มย่อย

$$\eta_j = yz_j + \zeta_j \qquad [2\text{-}34]$$

เมื่อ γ เป็นความชันของกลุ่มตัวแปร G_j และ ζ_j เป็นความคลาดเคลื่อนของสมการถดถอย β_i เป็นการทำหน้าที่ต่างกันของข้อสอบ (DIF) แบบเอกรูป เมื่อ γ เป็นผลต่างของค่าเฉลี่ยคุณลักษณะแฝงของกลุ่มเปรียบเทียบกับกลุ่มอ้างอิงและมีเกณฑ์การจับคู่ ตามตัวแปรคงที่ในสมการข้างต้นมีข้อตกลงกำหนดให้เป็น 0 ซึ่งจะไม่ปรากฏในสมการข้างต้น

$$a_i = \frac{\lambda_i \sqrt{\sigma_\eta^2}}{\sqrt{1 - \lambda_i^2 \sigma_\zeta^2}} \qquad [2\text{-}35]$$

$$b_i = \frac{\left[\left(\tau_i - \beta_i z\right)\right]\lambda_i^1 - \mu_\eta}{\left(\sigma_\eta^2\right)^{1/2}} \qquad [2\text{-}36]$$

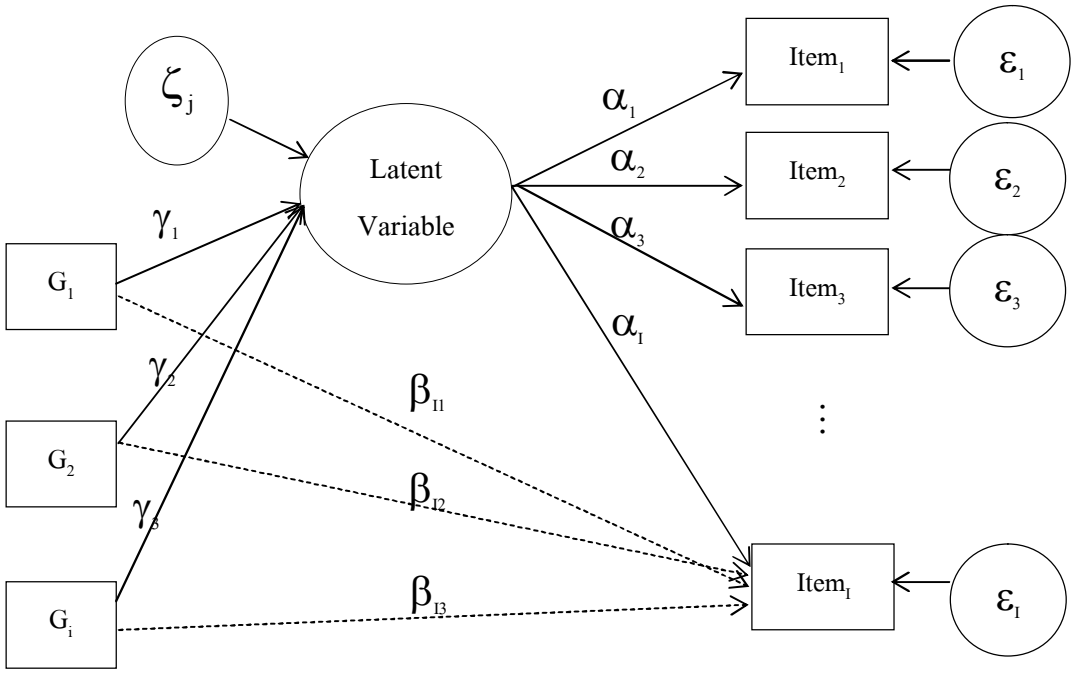
เมื่อ σ_η^2 เป็นตัวแปรสำหรับองค์ประกอบ θ_j และ σ_ζ^2 เป็นตัวแปรของความคลาดเคลื่อนของสมการถดถอยเชิงเส้นตรง ζ_j สำหรับการทำนายองค์ประกอบทั่วไป τ_i เป็นความยากของข้อสอบข้อที่ i และ μ_η เป็นค่าเฉลี่ยขององค์ประกอบทั่วไป θ_j

มีข้อดีหลายประการของการใช้โมเดล MIMIC ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ที่แสดงข้างต้น (Muthén et al., 1991) แสดงขนาดของการทำหน้าที่ต่างกันของข้อสอบ (DIF) โดยใช้หลักทฤษฎีการตอบสนองข้อสอบ (IRT) ประมาณค่าการทำหน้าที่ต่างกันของข้อสอบ (DIF) จากค่าพารามิเตอร์ตามทฤษฎีการตอบสนองข้อสอบ (IRT) ซึ่งมีประโยชน์กับผู้ปฏิบัติ

โมเดล MIMIC ในการประยุกต์ใช้สำหรับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสามารถตรวจสอบได้ทั้งแบบเอกรูป และแบบอเนกรูป (Seokjoon, 2014)

รูปแบบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี MIMIC แสดงได้ดังภาพที่ 2-16 และภาพที่ 2-17 ต่อไปนี้

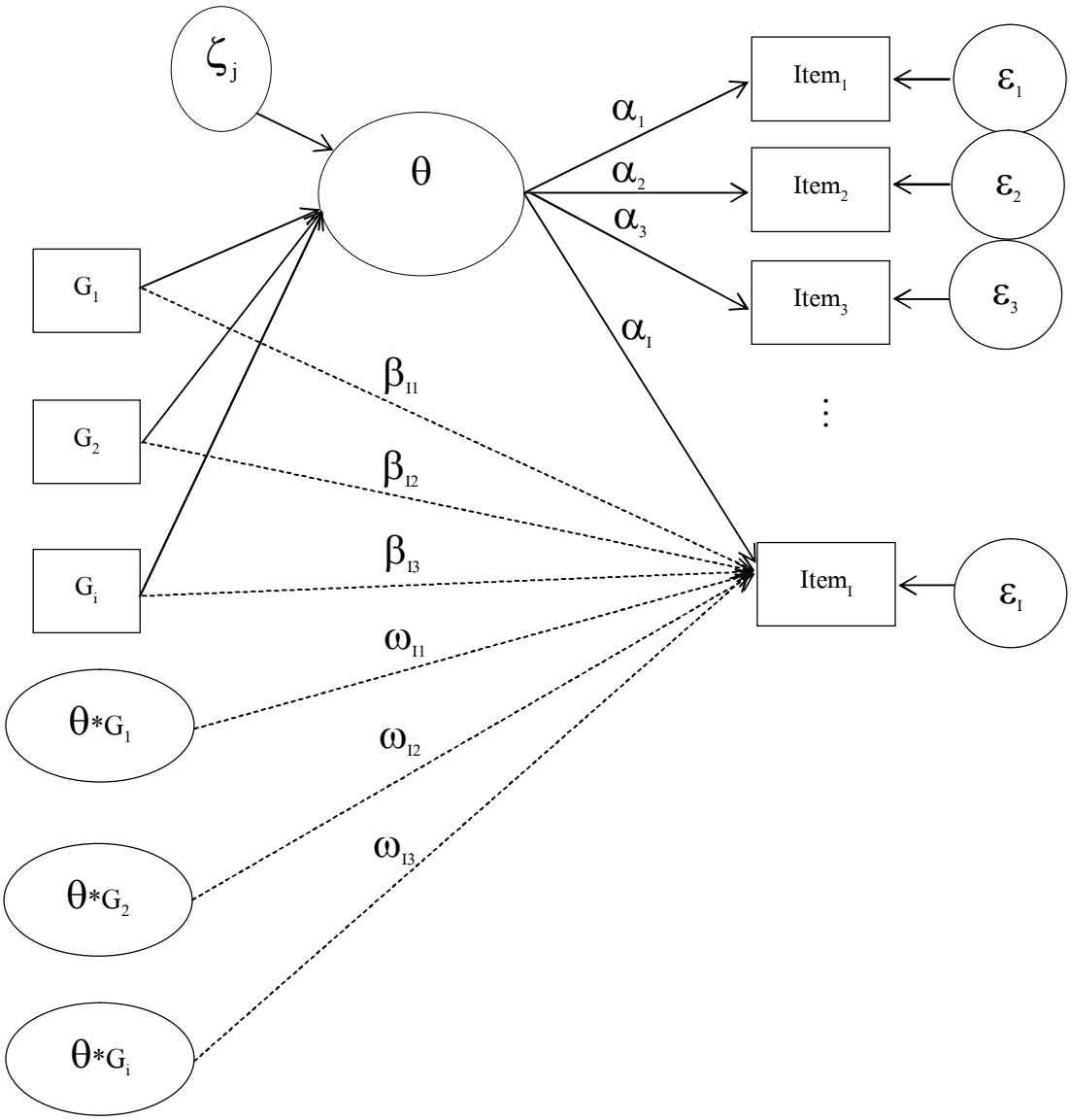
1. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี MIMIC แบบเอกรูป (MIMIC Uniform DIF path diagram)



ภาพที่ 2-5 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี MIMIC แบบเอกรูป

เมื่อกำหนดให้ G_1 และ G_2 เป็นความแปรปรวนร่วมของกลุ่ม 1 และกลุ่ม 2 กับ G_0 ซึ่งเป็นกลุ่มอ้างอิง θ เป็นตัวแปรแฝง ζ เป็นความแปรปรวนความคลาดเคลื่อนของตัวแปรแฝง γ เป็นความแตกต่างของค่าเฉลี่ยบนตัวแปรแฝงระหว่างกลุ่มเปรียบเทียบกับกลุ่มอ้างอิง β เป็นการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป α เป็นค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ ε เป็นความคลาดเคลื่อนของข้อสอบ

2. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป (MIMIC Nonuniform DIF path diagram)



ภาพที่ 2-6 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี MIMIC แบบอเนกรูป

เมื่อกำหนดให้ G1 และ G2 เป็นความแปรปรวนร่วมของกลุ่ม 1 และกลุ่ม 2 กับ G_0 ซึ่งเป็นกลุ่มอ้างอิง $\theta * G$ เป็นปฏิสัมพันธ์ระหว่างกลุ่มและตัวแปรแฝง θ เป็นตัวแปรแฝง ζ เป็นความแปรปรวนความคลาดเคลื่อนของตัวแปรแฝง γ เป็นความแตกต่างของค่าเฉลี่ยบนตัวแปรแฝงระหว่างกลุ่มเปรียบเทียบกับกลุ่มอ้างอิง β เป็นการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป ω

เป็นการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป α เป็นค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ ε เป็นความคลาดเคลื่อนของข้อสอบ

Finch (2005) เปรียบเทียบประสิทธิภาพของโมเดล MIMIC กับการทดสอบโดยวิแมนเทลเฮนเซล (Mantel & Haenszel, 1959) และวิธี SIBTEST (Shealy & Stout, 1993) และวิธีการทดสอบ IRT likelihood ratio (Thissen et al., 1988) กับความคลาดเคลื่อนประเภทที่ 1 และอำนาจการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ได้แสดงให้เห็นว่าวิธี MIMIC มีค่าสูงขึ้นและความคลาดเคลื่อนประเภทที่ 1 มีค่าลดลงกับจำนวนข้อสอบ 50 ข้อ นอกจากนี้วิธี MIMIC ยังสามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบ Uniform DIF ได้ด้วย และ Fukahara and Kamata (2007) ได้มีการประเมินประสิทธิภาพการทำงานของวิธี MIMIC แบบละเมิดข้อตกลงเบื้องต้น โดยการทำชุดข้อสอบ พบว่า การทำหน้าที่ต่างกันของข้อสอบ (DIF) มีแนวโน้มที่จะ Underestimated เมื่อข้อสอบไม่เป็นอิสระ

ซึ่งจากงานวิจัยครั้งนี้ ผู้วิจัยได้แทนตัวแปร คือ β_i แทนค่าพารามิเตอร์ความยากของข้อสอบ และค่าพารามิเตอร์ความสามารถผู้วิจัยใช้ตัวแปร θ_j ดังนี้

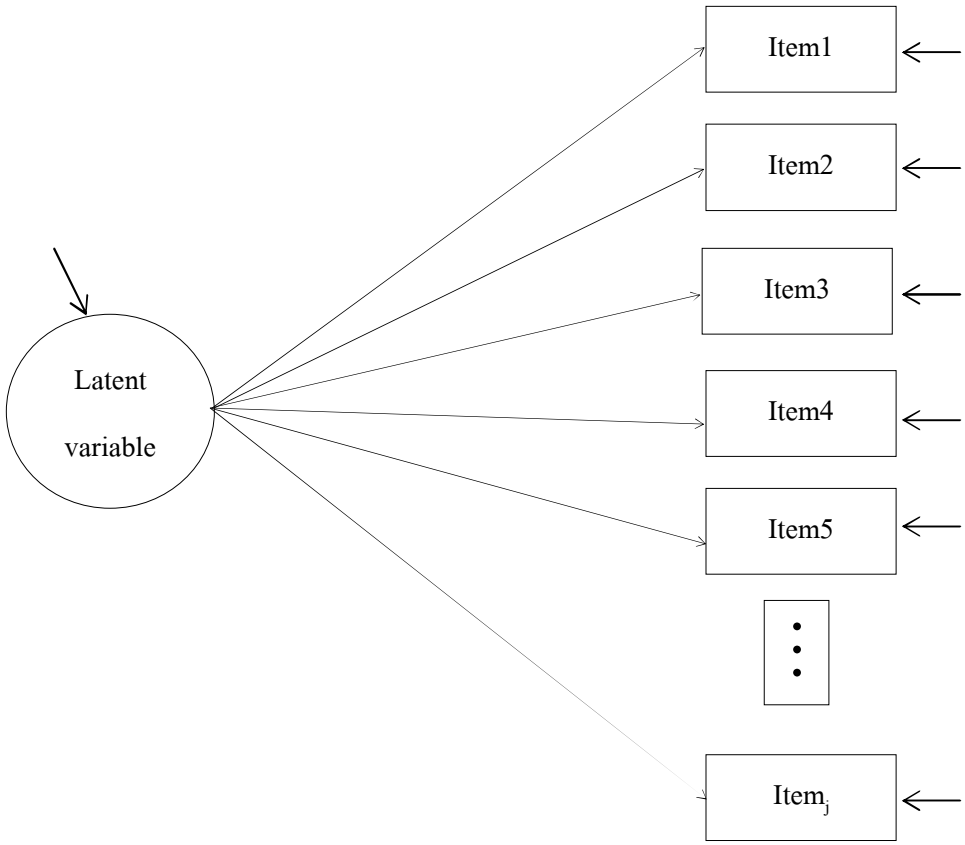
$$\theta = \gamma' z + \zeta \qquad [2\text{-}37]$$

เมื่อ γ คือ เวกเตอร์ของสัมประสิทธิ์การถดถอยของกลุ่มตัวแปร z ที่จะอธิบายความต่างความต่างระหว่างกลุ่มใน θ ซึ่งมักจะอ้างถึงในการวิเคราะห์ DIF และ ζ คือ Normally distributed กับค่าเฉลี่ย 0 และตัวแปรอิสระของ z ซึ่งการรวมกันของสมการ 1, 2 และ 3 Reveals ว่า β_i คือเงื่อนไขการทดสอบของ θ ซึ่งเป็นไปตามเงื่อนไขของการทำหน้าที่ต่างกันของข้อสอบ (DIF) เช่นเดียวกันกับที่ได้กล่าวมาแล้ว

นอกจากนี้ ในการวัดสิ่งต่าง (Measurement) สิ่งที่น่าสนใจคือความแตกต่างของกลุ่มในตัวแปรแฝง (Latent variable) ในการศึกษาความแตกต่างของค่าเฉลี่ยของตัวแปรแฝง เป็นการศึกษาความไม่แปรเปลี่ยนของกลุ่ม (Invariant) ในขณะที่การศึกษาความแตกต่างของตัวแปรสังเกตได้ เช่น ค่าเฉลี่ยของข้อคำถามในแต่ละกลุ่ม ซึ่งการศึกษาการทำหน้าที่ต่างกันของข้อคำถามเป็นการศึกษาความแตกต่างของตัวแปรสังเกตได้หรือตัวชี้วัด

โดยปกติรูปแบบการวิเคราะห์ข้อมูลตามทฤษฎีการตอบสนองข้อสอบ (IRT) เป็นการศึกษาคุณสมบัติอยู่บนพื้นฐานข้อตกลงความเป็นเอกมิติ (Unidimensional) ของตัวแปรแฝงซึ่งสังเกตไม่ได้โดยตรง สำหรับตัวแปรแฝงในโมเดล IRT จะดูจากค่าเซต้า (θ) ซึ่งสามารถประมาณค่า

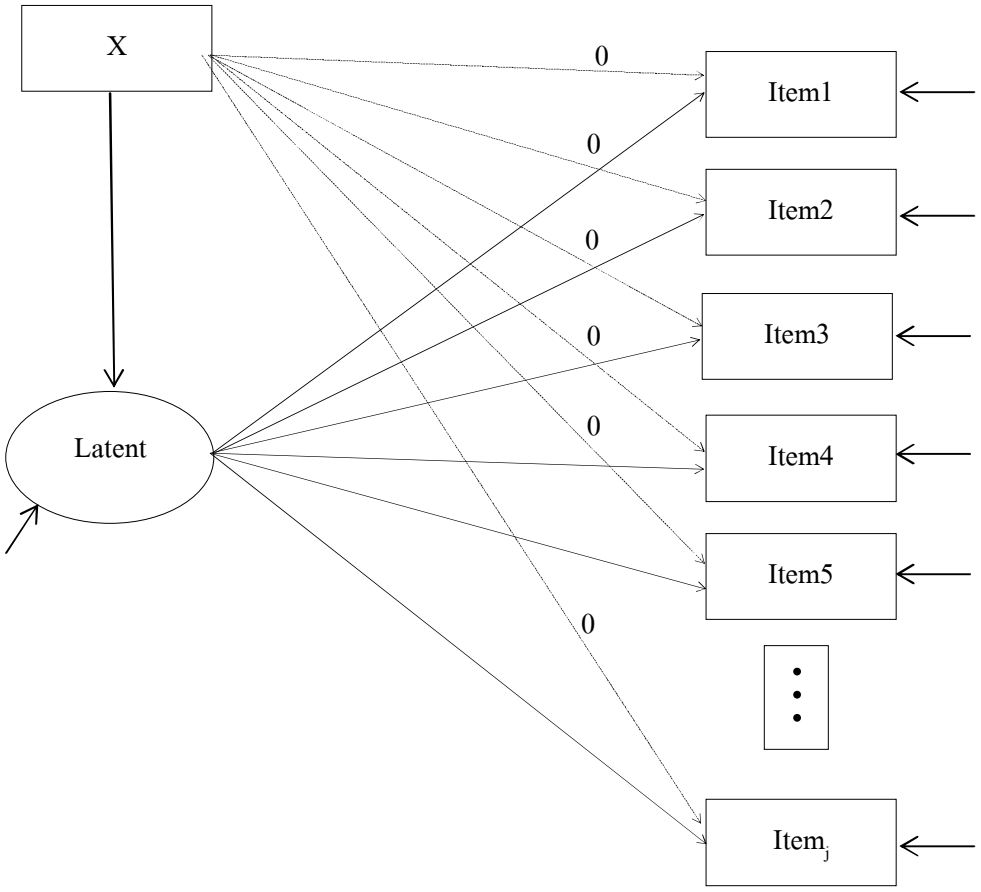
ได้โดยตรงซึ่งมีอิทธิพลตรงต่อตัวชี้วัดหรือข้อคำถามที่สังเกตได้ ซึ่งเราสามารถอธิบายในรูปของโมเดลการวิเคราะห์องค์ประกอบ แสดงดังภาพที่ 2-7



ภาพที่ 2-7 โมเดลการวิเคราะห์องค์ประกอบตามแนวคิด IRT

จากภาพถ้าข้อคำถามหรือตัวชี้วัดเป็นตัวแปรจัดกลุ่ม (Dichotomous) และตัวแปรแฝงมีการแจกแจงแบบปกติ (Normal distribution) ซึ่งจะมีลักษณะเช่นเดียวกับการแจกแจงโค้งความถี่สะสมในโมเดล IRT น้ำหนักองค์ประกอบที่เกิดขึ้นบนตัวชี้วัดจะหมายถึง ค่าดัชนีประมาณค่าอำนาจจำแนกตามทฤษฎีการตอบสนองข้อสอบ ในขณะที่ค่าเฉลี่ย (Intercepts) ของแต่ละข้อคำถามคือค่าประมาณความยาก (Difficulty) ตามทฤษฎีการตอบสนองในกรณีที่มีตัวแปรแฝงมากกว่าหนึ่งตัวโปรแกรมที่พัฒนามาใช้ตามทฤษฎี IRT โดยปกติจะอนุมาน (Assumes) ว่ามีข้อมูลเป็นลักษณะมีความเป็นเอกมิติ (Fleishman, 2003, p. 6) ในการนำมาประยุกต์ใช้ในการวิจัยจึงทำได้กว้างยิ่งขึ้น ดังนั้น จึงง่ายต่อการนำแนวคิดมาประยุกต์ใช้ในกรณีที่ต้องการนำแนวคิดของโมเดล MIMIC มาใช้ในกรณีที่ตัวแปรแฝงมีหลายมิติ (Multi-dimensional) หลายองค์ประกอบ (Multi-factor)

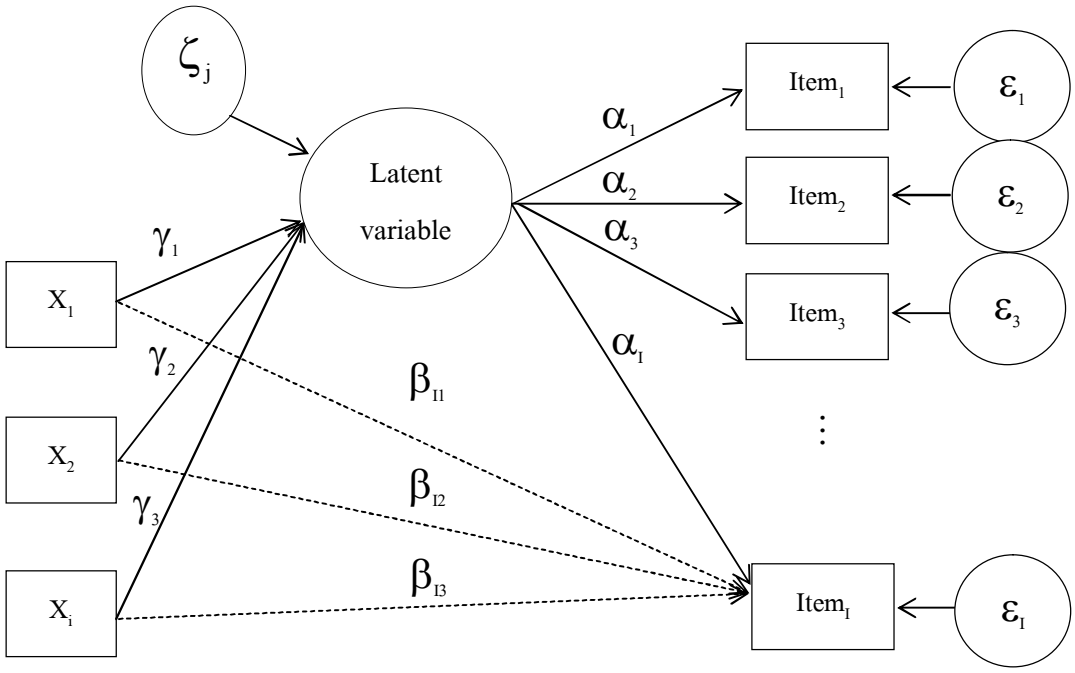
การนำโมเดล MIMIC มาใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อคำถามแสดงได้ตามโดยใช้ตัวแปรสาเหตุ (Causes) เพียงตัวเดียว แสดงได้ดังภาพที่ 2-8



ภาพที่ 2-8 โมเดลการวิเคราะห์การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ MIMIC model โดยใช้ตัวแปรสาเหตุ 1 ตัว

จากภาพที่ 2-19 เป็นการวิเคราะห์องค์ประกอบโดยใช้ตัวแปรทำนาย (X) จำนวนหนึ่งตัวแปร ในการทำนายตัวแปรแฝงที่ประกอบด้วยตัวแปรสังเกตได้ที่เป็นข้อสอบหรือข้อคำถาม (Item) จำนวน i ตัว โดยการจำกัดความคลาดเคลื่อนจากการวัดของตัวแปรแฝง และให้อิสระกับความคลาดเคลื่อนของตัวแปรวัดที่สอดคล้องกับโมเดลมากกว่าในการประมาณค่าระหว่าง X กับตัวแปรแฝง (Latent variable) อิทธิพลตรงของตัวแปร X ที่ทำนาย Item หลังจากที่มีอิทธิพลตรงไปยังตัวแปรแฝงแสดงทิศทางเดียว (Uniform) ในการทำหน้าที่ต่างกันของข้อคำถาม (DIF) ซึ่งเป็นสิ่งที่แสดงความลำเอียง (Biased) ที่เกิดจากข้อสอบหรือข้อคำถามหรืออธิบายได้ว่า

ถ้าข้อสอบหรือข้อคำถาม (Item) ได้รับอิทธิพลอย่างมีนัยสำคัญจากตัวแปรสาเหตุ X แสดงให้เห็นว่าข้อสอบหรือข้อคำถาม (Item) ขึ้นอยู่กับตัวแปรสาเหตุ X ไม่ได้อธิบายด้วยตัวแปรแฝงอย่างเดียว แสดงว่าข้อสอบหรือข้อคำถาม (Item) ข้อนั้นทำหน้าที่ต่างกันหรือมีความลำเอียง (Biased) แสดงได้ดังภาพที่ 2-9



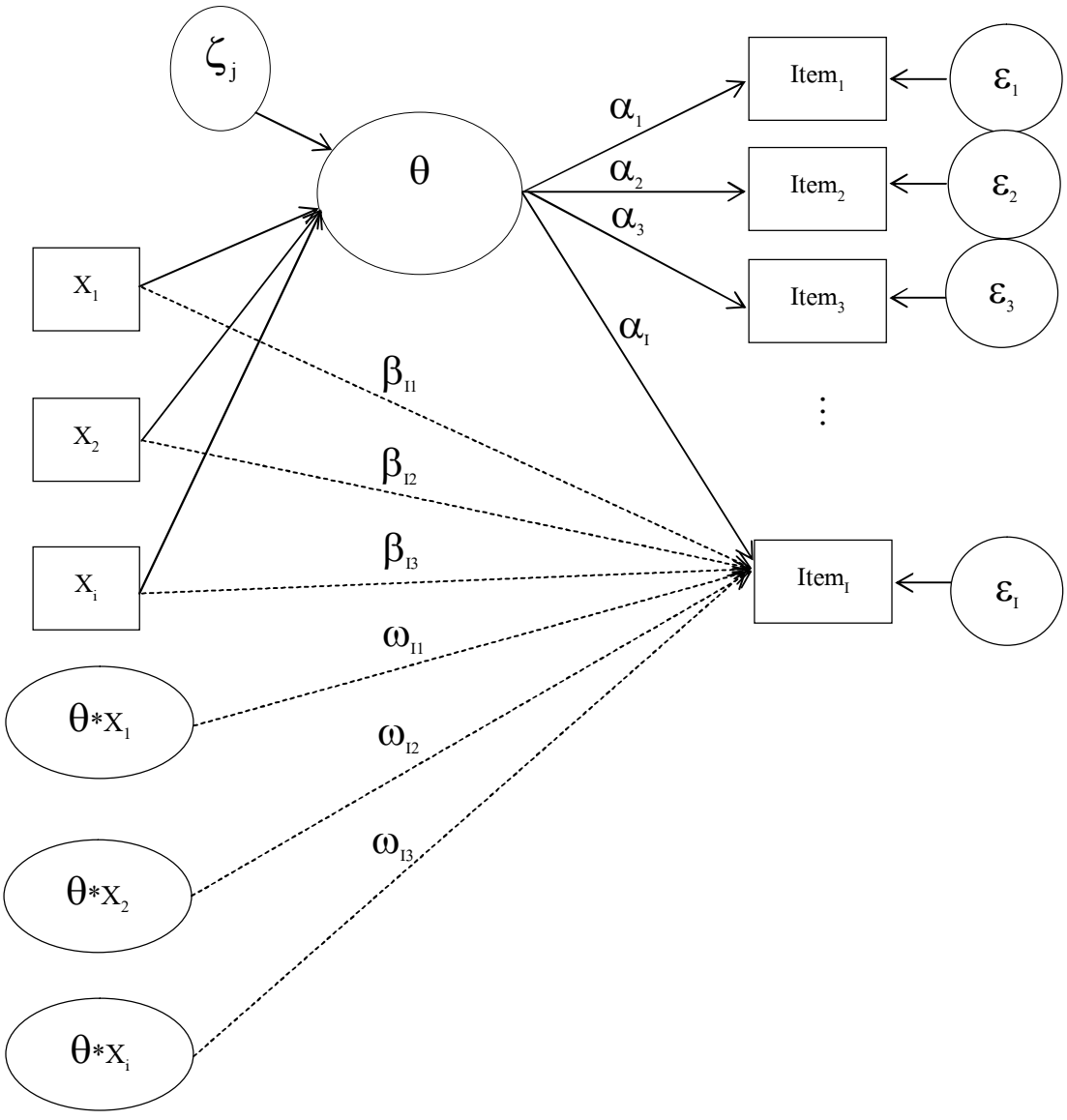
ภาพที่ 2-9 โมเดลการวิเคราะห์การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ MIMIC model โดยใช้ตัวแปรสาเหตุมากกว่า 1 ตัว แบบ Uniform DIF

จากภาพข้างต้น เป็นการวิเคราะห์องค์ประกอบโดยใช้ตัวแปรทำนาย $X_1 - X_i$ ในการทำนายตัวแปรแฝงที่ประกอบด้วยตัวแปรสังเกตได้ที่เป็นข้อสอบหรือข้อคำถาม (Item) จำนวน i ตัว ในการตรวจสอบการทำหน้าที่ต่างกันของข้อคำถาม (DIF) สามารถดำเนินการตามขั้นตอน ดังนี้

1. วิเคราะห์องค์ประกอบในโมเดลวัดที่ประกอบด้วยตัวแปรแฝงและข้อคำถาม
2. เพิ่มความแปรปรวนร่วมในการทดสอบโมเดล
3. เพิ่มอิทธิพลตรงกับตัวแปรแฝง (γ) อิทธิพลตรง (a) และกำหนดให้มีค่าเท่ากับศูนย์
4. ตรวจสอบโมเดลดัชนีปรับแก้ (Modification indices)
5. เพิ่มอิทธิพลตรงจากความแปรปรวนร่วมกับข้อคำถามที่มีค่าดัชนีปรับแก้สูงสุด

6. ดำเนินในขั้นตอนที่ 4-5 จนกว่าจะไม่พบค่าดัชนีปรับแก้ (M.I.) ที่ไม่มีนัยสำคัญ ประเมินความสอดคล้องของโมเดลและอิทธิพลตรง (β_i)

7. ดำเนินการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป Nonuniform DIF เขียนคำสั่งวิเคราะห์ต่อเนื่องจาก ข้อที่ 6 เมื่อความสามารถของผู้สอบมีปฏิสัมพันธ์กับการเป็นสมาชิกของกลุ่มพิจารณาค่าประมาณที่มีนัยสำคัญทางสถิติ $p < .05$ บ่งชี้ว่าข้อสอบทำหน้าที่ต่างกัน



ภาพที่ 2-10 โมเดลการวิเคราะห์การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ MIMIC model โดยใช้ตัวแปรสาเหตุมากกว่า 1 ตัว แบบ Nonuniform DIF

จากที่กล่าวมาสรุปได้ว่า วิธี MIMIC เป็นวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ประยุกต์โมเดล MIMIC ที่มีตัวแปรแฝงเพียงตัวแปรเดียวโดยที่ตัวแปรแฝงนั้นได้รับอิทธิพลจากตัวแปรภายนอกสังเกตได้หลายตัวแปรและส่งอิทธิพลไปยังตัวแปรภายในสังเกตได้หลายตัวแปร โดยพิจารณาว่าถ้าข้อสอบได้รับอิทธิพลอย่างมีนัยสำคัญจากตัวแปรสาเหตุหรือตัวแปรกลุ่มแสดงให้เห็นว่าข้อสอบขึ้นอยู่กับตัวแปรสาเหตุหรือตัวแปรกลุ่ม โดยที่ไม่ได้อธิบายด้วยตัวแปรแฝงอย่างเดียว แสดงว่าข้อสอบข้อนั้นทำหน้าที่ต่างกันแบบเอกรูป หากตัวแปรกลุ่มมีปฏิสัมพันธ์กับตัวแปรแฝงแล้วมีอิทธิพลต่อข้อสอบโดยตรง แสดงว่า ข้อสอบข้อนั้นทำหน้าที่ต่างกันแบบอเนกรูป

ผู้วิจัยได้สรุปคุณลักษณะของวิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 3 วิธี ที่ใช้ในการวิจัยครั้งนี้ได้ดังนี้

ตารางที่ 2-2 คุณลักษณะของวิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 3 วิธี

ประเด็น	IRT-LRT	MGCFA-ALIGNMENT	MIMIC
1. พื้นฐานทฤษฎี	IRT/2PL	IRT/2PL	IRT/2PL
2. การวิเคราะห์	ความแตกต่างของฟังก์ชันการตอบข้อสอบของกลุ่มผู้สอบที่มีความสามารถระดับเดียวกัน	Factor loading (Λx) และ Item intercepts ไม่ผันแปรในทุกกลุ่มย่อย	สัมประสิทธิ์ของกลุ่มตัวแปรที่มีอิทธิพลต่อผลการตอบข้อสอบ
3. การแจกแจงของพารามิเตอร์ข้อสอบ	ปกติ	ปกติ	ปกติ
4. สถิติที่พิจารณาในการตรวจสอบ DIF	ค่า G^2 และความมีนัยสำคัญทางสถิติของ χ^2	Factor Loading (Λx) และ item intercepts และความมีนัยสำคัญทางสถิติของ Z	ค่า β' ของตัวแปรทำนายและความมีนัยสำคัญทางสถิติของ Z
5. ประเภท DIF ตรวจสอบได้	Uniform, Nonuniform	Uniform, Nonuniform	Uniform, Nonuniform

ตารางที่ 2-2 (ต่อ)

ประเด็น	IRT-LRT	MGCFA-ALIGNMENT	MIMIC
6. ข้อดี	คำนวณง่าย ประหยัด	รูปแบบที่มีจำนวนของพารามิเตอร์เพียงเล็กน้อยแต่แสดงให้เห็นถึงความแตกต่างกันในแต่กลุ่ม, เปรียบเทียบแต่ละกลุ่มในขั้นตอนเดียว, เหมาะสำหรับการตรวจหา DIF หลายกลุ่ม, ให้ค่าพารามิเตอร์เป็นรายกลุ่ม	ประมาณค่าการทำหน้าที่ต่างกันของข้อสอบจากค่าพารามิเตอร์ตามทฤษฎี IRT
7. ข้อจำกัด	กลุ่มตัวอย่างขนาดใหญ่,มีความแกร่งในเรื่องข้อตกลงเบื้องต้น	กลุ่มตัวอย่างขนาดใหญ่	กลุ่มตัวอย่างขนาดใหญ่
8. โปรแกรมที่ใช้วิเคราะห์ข้อมูล	IRTPRO 2.1 (Cai, Thissen, & Toit, 2011)	Mplus 7.11 (Muthén & Muthén, 2007)	Mplus 7.11 (Muthén & Muthén, 2007)

ตอนที่ 3 การทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET)

การทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน O-NET (Ordinary national educational test) คือ การทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน เป็นการทดสอบเพื่อวัดความรู้และความคิดของนักเรียนชั้นประถมศึกษาปีที่ 6 ชั้นมัธยมศึกษาปีที่ 3 และชั้นมัธยมศึกษาปีที่ 6 ประเมินตามมาตรฐานการเรียนรู้ในหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 จำนวน 51 มาตรฐานการเรียนรู้ ครอบคลุม 5 กลุ่มสาระการเรียนรู้ ได้แก่

1. ภาษาไทย
2. คณิตศาสตร์
3. วิทยาศาสตร์
4. สังคมศึกษา ศาสนา และวัฒนธรรม
5. ภาษาอังกฤษ

วัตถุประสงค์ของการทดสอบ O-NET

1. เพื่อทดสอบความรู้และความคิดของนักเรียนชั้นประถมศึกษาปีที่ 6 ชั้นมัธยมศึกษาปีที่ 3 และชั้นมัธยมศึกษาปีที่ 6

2. เพื่อนำผลการทดสอบไปใช้เป็นองค์ประกอบหนึ่งในการจบการศึกษา ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551

3. เพื่อนำผลการทดสอบไปใช้ในการปรับปรุงคุณภาพการเรียนการสอนของโรงเรียน

4. เพื่อนำผลการทดสอบไปใช้ในการประเมินผลการเรียนรู้ของนักเรียนระดับชาติ

5. เพื่อนำผลการทดสอบไปใช้ในวัตถุประสงค์อื่น (สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน), 2560)

ตอนที่ 4 งานวิจัยที่เกี่ยวข้อง

เอกสารและงานวิจัยที่เกี่ยวข้องกับการศึกษาการทำหน้าที่ต่างกันของข้อสอบที่มีการตรวจให้คะแนนแบบทววิภาคผู้วิจัยขอนำเสนอเป็น 2 ประเด็น

1. งานวิจัยที่เกี่ยวข้องกับการศึกษาปัจจัยที่มีผลต่อการทำหน้าที่ต่างกันของข้อสอบ

2. งานวิจัยที่เกี่ยวข้องกับการศึกษาประสิทธิภาพของวิธีและศึกษาเชิงเปรียบเทียบวิธีที่ใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ มีรายละเอียดมีดังนี้

3. งานวิจัยที่เกี่ยวข้องกับปัจจัยที่ใช้จำแนกกลุ่ม

งานวิจัยที่เกี่ยวข้องกับการศึกษาปัจจัยที่มีผลต่อการทำหน้าที่ต่างกันของข้อสอบ

กาญจนา วัธนสุนธร (2537) ได้ศึกษาพัฒนาเกณฑ์ตัดสินข้อสอบลำเอียงทางเพศด้วยข้อมูลเชิงประจักษ์ด้วยวิธีแมนเทล-แฮนส์เซล (Mantel-haenszel) และวิธีซิบเทสท์ (SIBTEST) ข้อมูลที่ใช้เป็นผลการตอบข้อสอบคัดเลือกศึกษาต่อในระดับอุดมศึกษา ปีการศึกษา 2535 ปัจจัยที่ศึกษาเป็นความยาวของแบบทดสอบ และขนาดกลุ่มตัวอย่าง ความยาวของแบบทดสอบวิชาภาษาอังกฤษ (50: 60: 70: 80) และวิชาคณิตศาสตร์ (20: 30: 40) และขนาดกลุ่มตัวอย่าง (100: 200: 400: 600: 800: 1,000) ผลการวิจัยพบว่า เกณฑ์ตัดสินข้อสอบลำเอียงทางเพศเมื่อนำมาใช้มีความไม่คงที่ตามขนาดกลุ่มตัวอย่างและความยาวแบบสอบ ความสอดคล้องวิธีเดียวกันขนาดกลุ่มตัวอย่างต่างกันค่อนข้างต่ำ และมีความสอดคล้องสูงขึ้นเมื่อขนาดกลุ่มตัวอย่าง 600 คนขึ้นไป ส่วนการวิเคราะห์ความลำเอียงของข้อสอบที่มีต่อเพศพบว่า ข้อสอบวิชาภาษาอังกฤษมีความลำเอียงเข้าข้างเพศหญิง ส่วนข้อสอบวิชาคณิตศาสตร์มีความลำเอียงเข้าข้างเพศชาย

เกษร หว่างจิตร (2539) ได้ศึกษาการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย และวิชาภาษาอังกฤษ ของศูนย์ทดสอบทางการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ด้วยวิธีแมน

เทล-เฮนส์เซล เมื่อจำแนกกลุ่มผู้สอบตามเพศ ภูมิลำเนา ประสบการณ์ในการสอบ และสังกัด ของสถานศึกษา โดยมีกลุ่มตัวอย่างเป็นผู้สอบวิชาภาษาไทย จำนวน 506 คน และผู้สอบวิชาภาษาอังกฤษ จำนวน 501 คน ผลการวิจัย พบว่า ข้อสอบที่พบ DIF ส่วนใหญ่เป็นแบบอเนกรูป ข้อสอบที่ DIF ทั้งแบบเอกรูปและอเนกรูปเกิดขึ้นเมื่อการวิเคราะห์จำแนกกลุ่มผู้สอบตามเพศมากที่สุด ลักษณะของข้อสอบที่พบ DIF ส่วนใหญ่เป็นข้อสอบที่มีค่าอำนาจจำแนกค่อนข้างต่ำ ค่าความเที่ยง และค่าความตรงตามทฤษฎีของแบบสอบก่อนและหลังตัดข้อสอบที่พบ DIF ออกจากแบบสอบส่วนใหญ่ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ

รักชนก ยี่สุ่นศรี (2544) ได้ศึกษาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบด้วยกระบวนการ ดี เอฟ ไอ ที กับแบบสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาของทบวงมหาวิทยาลัย ปีการศึกษา 2543 ข้อมูลที่นำมาศึกษาเป็นวิชาภาษาอังกฤษและวิชาคณิตศาสตร์ ขนาดกลุ่มตัวอย่างจำนวน 3,600 คน และ 4,000 คน ผลการวิจัย พบว่า ปัจจัยเพศข้อสอบวิชาภาษาอังกฤษและวิชาคณิตศาสตร์ทำหน้าที่ต่างกัน ปัจจัยสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษาของผู้สอบข้อสอบวิชาภาษาอังกฤษทำหน้าที่ต่างกัน ผลการเปรียบเทียบค่าความตรงเชิงโครงสร้างของแบบสอบฉบับก่อนและหลังตัดข้อสอบที่พบการทำหน้าที่ต่างกัน พบว่าไม่แตกต่างกัน ผลการเปรียบเทียบค่าความเที่ยงพบว่า แบบสอบฉบับหลังตัดข้อสอบที่พบการทำหน้าที่ต่างกันมีค่าความเที่ยงลดลง เมื่อพิจารณาค่าฟังก์ชันสารสนเทศของแบบสอบพบว่ามีค่ามากขึ้น

สุมาลี แกล้วทนงค์ (2547) ได้ศึกษาสาเหตุของการทำหน้าที่ต่างกันของข้อสอบสาระการเรียนรู้ภาษาไทย และสาระการเรียนรู้สังคมศึกษา ศาสนาและวัฒนธรรม กลุ่มตัวอย่างเป็นนักเรียนระดับชั้นมัธยมศึกษาปีที่ 1 สำนักงานเขตพื้นที่การศึกษาสงขลา พัทลุง ตรัง และสตูล จำนวน 1,320 คน ผลการวิจัย พบว่า ข้อสอบที่ทำหน้าที่ต่างกันด้านเพศ ด้านภาษาพูด และด้านเชื้อชาติของแบบสอบสาระการเรียนรู้ภาษาไทย มีสาเหตุมาจากเนื้อเรื่องที่สนใจและภาษาที่ใช้ในแบบสอบที่เป็นสำนวนและคำศัพท์ คำศัพท์เฉพาะ และคำราชาศัพท์หรือคำศัพท์ยาก ตามลำดับ ส่วนแบบสอบสาระการเรียนรู้สังคมศึกษา ศาสนาและวัฒนธรรม มีสาเหตุมาจากเรื่องที่สนใจและเนื้อเรื่องเกี่ยวกับวัฒนธรรม

อิทธิฤทธิ์ พงษ์ปิยะรัตน์ (2551) ได้ศึกษาการวิเคราะห์ข้อสอบและการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ: การวิเคราะห์พหุระดับ ผลการวิจัยพบว่า 1) ผลการประมาณค่าพารามิเตอร์ข้อสอบด้วยโมเดล HGLM-2L และ HGLM-3L ด้วยสถิติ Empirical bayesian มีความสัมพันธ์อย่างสัมบูรณ์กับผลการประมาณค่าด้วยโปรแกรม BILOG-MG ส่วนโมเดล HGLM-3L มีระดับของค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.793 ซึ่งน้อยกว่าค่าสัมประสิทธิ์สหสัมพันธ์

ระหว่างโมเดล HGLM-2L กับ โปรแกรม BILOG-MG 2) ผลการวิเคราะห์ระดับนักเรียน (Level 2) พบว่า ตัวแปรผลการเรียนคณิตศาสตร์ในภาคเรียนที่ผ่านมาส่งผลต่อค่าเฉลี่ยของโอกาสในการตอบข้อสอบได้ถูกต้องในแต่ละโรงเรียนอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ผลการวิเคราะห์ระดับโรงเรียน (Level 3) พบว่า ตัวแปรขนาดของโรงเรียนและตัวแปรความเป็นผู้นำทางวิชาการของผู้บริหารส่งผลต่อค่าเฉลี่ยของโอกาสในการตอบข้อสอบได้ถูกต้องในโรงเรียน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และ 0.05 ตามลำดับ และ 3) การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยโมเดล HGLM สามารถตรวจสอบพบข้อสอบที่ทำหน้าที่ต่างกันสอดคล้องตรงกับผลการวิเคราะห์จากโปรแกรม BILOG-MG

Sheppard, Han, Colarelli, Dai and King (2006) ได้ศึกษาการทำหน้าที่ต่างกันของข้อสอบจำแนกตามเพศและเชื้อชาติในแบบประเมินบุคลิกภาพของ Hogan (Hogan personality inventory: HPI) ที่ประกอบด้วย 13 สเกล จานวน 206 ข้อ กลุ่มตัวอย่างเป็นคนผิวขาว จำนวน 1,579 คน และคนผิวดำ จำนวน 523 คน แบ่งเป็นเพศชาย จำนวน 1,344 คน และเพศหญิง 758 คน ใช้วิธีแมนเทล-แฮนส์เซลในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ผลการวิจัยพบว่า หนึ่งในสามของข้อคำถามแสดงข้อที่พบ DIF จำแนกตามเพศ ร้อยละ 38.4 และตามเชื้อชาติ ร้อยละ 37.3 เมื่อวิเคราะห์เนื้อหาของข้อคำถามที่มีความลำเอียง พบว่า สาระสำคัญของข้อคำถามที่พบ DIF เกาะเกี่ยวกับเพศมากกว่าเชื้อชาติ

Shin and Wang (2009) ได้ทำการศึกษาวิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้วิธี Multiple indicators multiple causes (MIMIC) กับข้อสอบ anchor แบบสั้นบริสุทธิ์ โดยใช้การจำลองข้อมูล และมีตัวแปรอิสระที่จัดกระทำ 6 ตัว ได้แก่ โมเดลการตอบสนองข้อสอบ ขนาดกลุ่มตัวอย่าง ความยาวของแบบสอบ ร้อยละของข้อสอบที่พบ DIF ในแบบสอบ รูปแบบของ DIF และความยาวของข้อสอบ Anchor ผลการวิจัย พบว่า การใช้วิธี MIMIC ที่ความยาวของข้อสอบ Anchor 1, 2, 4 และ 10 ข้อ (ไม่มี DIF) เมื่อข้อสอบที่พบ DIF มีร้อยละ 40 ในแบบสอบ สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้เป็นอย่างดี โดยทั่วไปการเพิ่มความยาวของข้อสอบ Anchor ทำให้เพิ่มอำนาจการตรวจสอบ DIF ความยาวของข้อสอบ Anchor 4 ข้อ ก็เพียงพอต่ออำนาจการตรวจสอบDIF ที่สูง กระบวนการ MIMIC ที่ทำซ้ำ เป็นการกำหนดข้อสอบที่ไม่มี DIF ในฐานะที่เป็นข้อสอบ Anchor บริสุทธิ์ได้อย่างเหมาะสม นอกจากนี้ยังให้อำนาจการทดสอบที่สมบูรณ์ ณ จุดที่มีข้อสอบที่ไม่มี DIF จำนวน 4 ข้อ

สรุปได้ดังตารางที่ 2-3

ตารางที่ 2-3 งานวิจัยที่เกี่ยวข้องกับการศึกษาปัจจัยที่มีผลต่อการทำหน้าที่ต่างกันของข้อสอบ

ผู้วิจัย	ปี	ประเด็นที่ศึกษา	ปัจจัย	องค์ความรู้ที่ได้จากการศึกษา
กาญจนา วัฒนสุนทร	2537	พัฒนาเกณฑ์ตัดสินข้อสอบลำเอียงทางเพศด้วยข้อมูลเชิงประจักษ์ด้วยวิธีแมนเทล-แฮนส์เซล และวิธีซิบเทสท์	เพศ	มีความไม่คงที่ข้ามขนาดผู้สอบและความยาวแบบสอบ ความสอดคล้องในการตรวจข้อสอบลำเอียงภายในวิธีเดียวกันข้ามขนาดผู้สอบค่อนข้างต่ำแต่จะสูงขึ้นเมื่อขนาดผู้สอบ 600 คนขึ้นไป การวิเคราะห์ความลำเอียงของข้อสอบที่มีต่อเพศ พบว่าข้อสอบวิชาภาษาอังกฤษลำเอียงเข้าข้างผู้หญิง ส่วนข้อสอบวิชาคณิตศาสตร์ลำเอียงเข้าข้างผู้ชาย
เกษร หว่างจิตร	2539	การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบสำหรับแบบสอบคัดเลือกระดับบัณฑิตศึกษาวิชาภาษาไทยและภาษาอังกฤษด้วยวิธีแมนเทล-แฮนส์เซล	1. เพศ 2. ภูมิลำเนา 3. ประสบการณ์ในการสอบ 4. สังกัดของสถานศึกษา	ข้อสอบที่พบ DIF ทั้งแบบเอกรูปและอเนกรูปเกิดขึ้นเมื่อวิเคราะห์จำแนกผู้สอบตามกลุ่มเพศมากที่สุด
รักชนก ยี่สุ่นศรี	2544	การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบด้วยกระบวนการ ดี เอฟ ไอ ที สาหรับแบบสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษา	1. เพศ 2. สถานที่ตั้งของโรงเรียนที่จบการศึกษา 3. ความตรงเชิงโครงสร้าง 4. ความเที่ยง 5. ฟังก์ชัน	แบบสอบวิชาภาษาอังกฤษทำหน้าที่ต่างกันตามเพศและสถานที่ตั้งของโรงเรียนที่จบการศึกษา ส่วนแบบสอบวิชาคณิตศาสตร์ทำหน้าที่ต่างกันตามเพศ 1. ค่าความตรงเชิงโครงสร้างไม่แตกต่างกัน ค่าความเที่ยงลดลงและค่าฟังก์ชันสารสนเทศมากขึ้น

ตารางที่ 2-3 (ต่อ)

ผู้วิจัย	ปี	ประเด็นที่ศึกษา	ปัจจัย	องค์ความรู้ที่ได้จากการศึกษา
		วิชาภาษาอังกฤษและวิชาคณิตศาสตร์	สารสนเทศของแบบสอบ 6. ความสัมพันธ์ระหว่างตำแหน่งของคะแนนรวม	2. ความสัมพันธ์ในทางบวกอย่างมีนัยสำคัญระหว่างตำแหน่งของคะแนนรวมของผู้สอบก่อนและหลังตัดข้อสอบที่พบ DIF
สุมาลี แกล้วทนงค์	2547	ศึกษาสาเหตุของการทำหน้าที่ต่างกันของข้อสอบสาระการเรียนรู้ภาษาไทย และสาระการเรียนรู้สังคมศึกษาศาสนาและวัฒนธรรม กลุ่มตัวอย่างเป็นนักเรียนระดับชั้นมัธยมศึกษาปีที่ 1 สำนักงานเขตพื้นที่การศึกษาสงขลา พัทลุง ตรัง และสตูล	1. เพศ 2. ภาษาพูด 3. เชื้อชาติ	1. แบบสอบสาระการเรียนรู้ภาษาไทย ด้านเพศ ภาษาพูด และเชื้อชาติ มีสาเหตุมาจากเนื้อเรื่องที่สนใจ และภาษาที่ใช้ในแบบสอบ 2. แบบสอบสาระการเรียนรู้สังคมศึกษา ศาสนา และวัฒนธรรม ด้านเพศ ภาษาพูด และเชื้อชาติ มีสาเหตุมาจากเนื้อเรื่องที่สนใจ และเนื้อเรื่องเกี่ยวกับวัฒนธรรม
Penfield and Algina	2006	การประมาณค่าขนาดอิทธิพลของการทำหน้าที่ต่างกันในแบบสอบโดยไม่คิดเครื่องหมายในแบบสอบแบบผสม (Mixed format test)	ข้อสอบแบบผสม	แบบสอบที่มีข้อสอบที่ตรวจให้คะแนนแบบ 2 ค่าจำนวนมากจะส่งผลต่อความลำเอียงทางลบแต่แบบสอบที่มีข้อสอบที่ตรวจให้คะแนนแบบหลายค่าจำนวนมากจะส่งผลต่อความลำเอียงทางบวกเพียงเล็กน้อย
Sheppard, et al.	2006	Differential item functioning by sex and race in the hogan rersonality inventory	1. เพศ 2. เชื้อชาติ	1 ใน 3 ของข้อคาถามเป็นข้อที่พบ DIF โดยส่วนใหญ่ข้อที่พบ DIF จะเกี่ยวกับเพศมากกว่าเชื้อชาติ

ตารางที่ 2-3 (ต่อ)

ผู้วิจัย	ปี	ประเด็นที่ศึกษา	ปัจจัย	องค์ความรู้ที่ได้จากการศึกษา
Shin and Wang	2009	Differential item functioning detection using the multiple indicators, multiple causes method with a pure short anchor.	โมเดลการตอบสนองข้อสอบ ขนาดกลุ่มตัวอย่าง ความยาวของแบบสอบ ร้อยละของข้อสอบที่พบ DIF ในแบบสอบ รูปแบบของ DIF ความยาวของข้อสอบ Anchor	ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี Multiple indicators multiple causes (MIMIC) กับข้อสอบ Anchor แบบสั้น โดยการจำลองข้อมูล ผลการวิจัยพบว่า วิธี MIMIC ที่ความยาวของข้อสอบ Anchor 1, 2, 4 และ 10 ข้อ (ไม่มี DIF) เมื่อข้อสอบที่พบ DIF มีร้อยละ 40 ในแบบสอบ สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ดี โดยทั่วไปการเพิ่มความยาวของข้อสอบ anchor ทำให้เพิ่มอำนาจการตรวจสอบ DIF ความยาวของข้อสอบ Anchor 4 ข้อ มีอำนาจการตรวจสอบ DIF สูง MIMIC ที่ทำซ้ำ เป็นการกำหนดข้อสอบที่ไม่มี DIF ในฐานะที่เป็นข้อสอบ Anchor อย่างเดียวได้อย่างเหมาะสม นอกจากนี้ยังให้อำนาจการทดสอบที่สมบูรณ์ ณ จุดที่มีข้อสอบที่ไม่มี DIF จำนวน 4 ข้อ

งานวิจัยที่เกี่ยวข้องกับการศึกษาประสิทธิภาพของวิธีและศึกษาเชิงเปรียบเทียบวิธีที่ใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

เรวดี อินทะสะระ (2539) ได้ศึกษาความเที่ยงตรงเชิงพยากรณ์ของแบบสอบคัดเลือกที่วิเคราะห์ความลำเอียงต่อเพศด้วยวิธีทฤษฎีการตอบสนองข้อสอบ วิธีแมนเทล-แฮนส์เซล และวิธีซิบเทสท์ พร้อมทั้งศึกษาการตัดสินผลการสอบที่คิดคะแนนมาตรฐานทีปกติ และคะแนนน้ำหนักความสามารถและสาเหตุของความลำเอียงของข้อสอบโดยศึกษาความลำเอียงของข้อสอบกับข้อมูลผลการสอบคัดเลือกเข้าศึกษาในชั้นปี ที่ 1 ประเภทรับตรง ปีการศึกษา 2538 วิชาภาษาไทย ก วิชาสังคมศึกษา ก และวิชา ภาษาอังกฤษ กข ของมหาวิทยาลัยสงขลานครินทร์ ศึกษากลุ่มตัวอย่าง

วิชาละ 8,127 คน วิชาภาษาไทย กข วิชาสังคม ศึกษา กข และวิชาภาษาอังกฤษ กขค ศึกษา กลุ่มตัวอย่าง 5,415 คน ผลการวิจัย พบว่า วิธีตรวจสอบความลำเอียง ทั้ง 3 วิธีพบข้อสอบที่ลำเอียงแตกต่างกันในวิชาภาษาไทย ก ฉบับที่ 2 และวิชาสังคมศึกษา ก ฉบับที่ 1 ที่ระดับนัยสำคัญทางสถิติที่.05 นอกนั้นพบข้อสอบที่ทำที่ลำเอียงแตกต่างกันที่ ระดับนัยสำคัญ .01 โดยวิธีทฤษฎีการตอบสนองข้อสอบตัดสินจำนวนข้อสอบที่ลำเอียงได้มากที่สุด ความสัมพันธ์ของลำดับที่ของการสอบไม่ว่าจะคิดคะแนนมาตรฐานปกติทีหรือคิดคะแนนน้ำหนัก ความสามารถและใช้ข้อสอบทั้งหมดหรือใช้เฉพาะข้อสอบที่ปราศจากความลำเอียงต่างมีความสัมพันธ์กัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

จิติมา วรรณศรี (2539) ได้ทำการศึกษาเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซลกับวิธีซิปเทสท์ เมื่อความยาวของแบบสอบ ขนาดกลุ่มตัวอย่าง และอัตราส่วนของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบต่างกัน โดยใช้ข้อมูลจำลองจากโปรแกรม IRTDATA ผลการวิจัย พบว่า วิธีแมนเทล-แฮนส์เซลกับวิธีซิปเทสท์มีประสิทธิภาพเท่าเทียมกัน ที่ทุกขนาดกลุ่มตัวอย่างและทุกอัตราส่วน ภายใต้ความยาวแบบสอบเดียวกัน และวิธีแมนเทล-แฮนส์เซลกับวิธีซิปเทสท์มีประสิทธิภาพเท่าเทียมกัน ที่ทุกระดับความยาวแบบสอบ

เสรี ชัดแช้ม (2539) ศึกษาเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป ระหว่างวิธีแมนเทล-แฮนส์เซลแบบปกติกับวิธีแมนเทล-แฮนส์เซลแบบแบ่งกลุ่มความสามารถของผู้สอบและความยากของข้อสอบใช้วิธี IRT เป็นเกณฑ์ ศึกษาจากข้อมูลผลการตอบแบบสอบวัดความสามารถในการอ่านภาษาไทยของนักเรียนชั้นมัธยมศึกษาปีที่ 1 สังกัดกรมสามัญศึกษาจังหวัดชลบุรี จำนวน 1,200 คน จำแนกตามเพศ ผลการวิจัยพบว่า วิธีแมนเทล-แฮนส์เซลแบบแบ่งกลุ่มความสามารถของผู้สอบและความยากของข้อสอบสามารถตรวจพบข้อสอบทำหน้าที่ต่างกันแบบอเนกรูปได้สอดคล้องกับวิธี IRT และตรวจพบข้อสอบทำหน้าที่ต่างกันมากกว่าวิธีแมนเทล-แฮนส์เซลแบบปกติ

รัชรินทร์ มุคดา (2540) ได้ทำการศึกษาเปรียบเทียบประสิทธิภาพระหว่างวิธีแมนเทล-แฮนส์เซลกับวิธีถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป ในกรณีที่จัดกลุ่มความสามารถ ค่าความยากของข้อสอบ และค่าอำนาจจำแนกของข้อสอบต่างกัน โดยศึกษาจากข้อมูลที่จำลองขึ้นด้วยโปรแกรม IRTDATA ผลการวิจัย พบว่า โดยภาพรวมวิธีแมนเทล-แฮนส์เซลกับวิธีถดถอยโลจิสติกมีประสิทธิภาพเท่ากัน ในทุกกลุ่มความสามารถของผู้สอบ ข้อสอบที่พบ DIF แบบอเนกรูป ในกลุ่มผู้สอบที่มีความสามารถสูง ปานกลาง และต่ำ เป็นข้อสอบที่มีค่าความยากสูง ปานกลาง และต่ำ ตามลาดับ และข้อสอบที่พบ DIF แบบอเนกรูป ในทุกกลุ่มความสามารถของผู้สอบเป็นข้อสอบที่มีค่าอำนาจจำแนกสูง

นพมาศ พิพัฒนสุข (2541) ได้ทำการศึกษาเปรียบเทียบประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซลกับวิธีถดถอยโลจิสติก เครื่องมือเป็นแบบสอบความสามารถทางคณิตศาสตร์ ชนิดพหุมิติ กลุ่มตัวอย่างเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 จำนวน 1,076 คน ผลการวิจัย พบว่า เมื่อใช้เกณฑ์จับคู่คะแนนรวม วิธีแมนเทล-แฮนส์เซลมีประสิทธิภาพมากกว่าวิธีถดถอยโลจิสติก แต่เมื่อใช้เกณฑ์จับคู่คะแนนแบบสอบย่อยทั้ง 2 วิธีมีประสิทธิภาพไม่แตกต่างกัน เมื่อใช้เกณฑ์จับคู่คะแนนหลายแบบสอบย่อยวิธีถดถอยโลจิสติกมีความเหมาะสมในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับแบบสอบชนิดพหุมิติ

นิคม กีรติวรางกูร (2542) ได้ทำการศึกษาเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีการวิเคราะห์องค์ประกอบจำกัด วิธีแมนเทล-แฮนส์เซลและวิธีทฤษฎีการตอบสนองข้อสอบ ปัจจัยที่ศึกษาเป็นขนาดกลุ่มตัวอย่าง ความยาวของแบบสอบ ค่าความยากของข้อสอบ ค่าอำนาจจำแนกของข้อสอบ และขนาดความลำเอียงของข้อสอบแตกต่างกัน โดยใช้ข้อมูลจำลอง ผลการวิจัย พบว่า ภาพรวมวิธีการวิเคราะห์องค์ประกอบจำกัดมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูงที่สุด วิธีแมนเทล-แฮนส์เซลมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูง ภายใต้เงื่อนไขแบบสอบที่มีค่าความยากต่ำ ค่าอำนาจจำแนกสูง ขนาดความยาวแบบสอบ 75 ข้อ เมื่อใช้กลุ่มตัวอย่างขนาด 1,000 คน และวิธีทฤษฎีการตอบสนองข้อสอบ แบบ 2 พารามิเตอร์ มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูง ภายใต้เงื่อนไขแบบสอบที่มีค่าความยากต่ำ ความยาวแบบสอบ 75 ข้อ เมื่อใช้กลุ่มตัวอย่างขนาด 1,000 คน

ชัยยศ ชาวระนอง (2553) ได้ศึกษาประสิทธิภาพของโมเดลและการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบพหุมิติโดยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยันแฝงภายในการวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของโมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง องค์ประกอบเชิงยืนยันอันดับสอง องค์ประกอบเชิงยืนยันอันดับหนึ่งแฝงภายใน องค์ประกอบเชิงยืนยันอันดับสองแฝงภายใน และการทดสอบการทำหน้าที่ต่างกันของข้อสอบโดยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองแฝงภายใน ภายใต้เงื่อนไขกลุ่มตัวอย่างขนาด 100, 200, 400, 800, 1,200, 1,600 และ 2,000 คน จำนวนข้อสอบ 3, 5, 8, 10 และ15 ข้อต่อองค์ประกอบ และขนาดกลุ่มตัวอย่างในการทดสอบการทำงานที่ต่างกันของข้อสอบขนาด 1,600, 2,000, 2,400, 2,800, 3,200, 3,600 และ 4,000 คน ซึ่งเป็นนักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 โรงเรียนในเขตพื้นที่การศึกษาจังหวัดบุรีรัมย์ เขต 2 จากการทดสอบทางการศึกษาแห่งชาติ

ปี พ.ศ. 2548 การวิเคราะห์ข้อมูลเชิงบรรยายใช้โปรแกรม SPSS for windows version 16.0 การหาคุณภาพแบบทดสอบใช้โปรแกรม BILOG MG 3.0 การวิเคราะห์องค์ประกอบใช้โปรแกรม AMOS 16.0 และการทดสอบค่าความเชื่อมั่น G-Coefficient ใช้โปรแกรม GENOVA ผลการวิจัยพบว่า 1)โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองแฝงภายในมีประสิทธิภาพมากที่สุด โมเดลมีความสอดคล้องกับข้อมูลเชิงประจักษ์เมื่อไม่มีการปรับโมเดลอยู่ในระดับดี โดยพิจารณาจากค่าดัชนีวัดความสอดคล้องของโมเดล ซึ่งพบว่า = 1658.181, df = 1130, p = .000, GFI = .919, AGFI = .909, CFI = .643 และ RMSEA = .024 เมื่อใช้กลุ่มตัวอย่างตั้งแต่ 800 ตังอย่างขึ้นไป และเมื่อพิจารณาตามขนาดของกลุ่มตัวอย่างพบว่า ค่าดัชนีวัดความสอดคล้องไค-สแควร์ ไม่แปรเปลี่ยนตามขนาดกลุ่มตัวอย่างแต่แปรเปลี่ยนตามจำนวนข้อความต่อองค์ประกอบ และ 2) การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยวิธีวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองแฝงภายในจำแนกตามเพศโดยใช้เทคนิค Multi-group พบว่า ข้อสอบทำหน้าที่ต่างกัน 5 ข้อ และเมื่อตัดข้อสอบที่ทำหน้าที่ต่างกันออกพบว่า ค่าสัมประสิทธิ์ความเชื่อมั่นของข้อสอบสูงขึ้น

สุพัฒนา หอมบุปผา (2556) ได้ศึกษาการเปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี HGLM วิธี MIMIC และวิธี BAYESIAN 1) ผลการวิเคราะห์ค่าพารามิเตอร์ความยากของข้อสอบ ค่าพารามิเตอร์ความสามารถของผู้สอบ และผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ในวิชาภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์ ด้วยวิธี HGLM-2L วิธี MIMIC และวิธี BAYESIAN พบว่า มีความสัมพันธ์กันในระดับสูงมาก อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 2) วิธีตรวจสอบที่พบการทำหน้าที่ต่างกันของข้อสอบมากที่สุด คือ วิธี HGLM-2L ส่วนวิธีที่ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบน้อยที่สุด คือ วิธี MIMIC และ 3) ผลการศึกษาลักษณะของข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ (DIF) วิชาภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์ ลักษณะของข้อสอบที่เกิดจากการทำหน้าที่ต่างกัน เมื่อจำแนกตามเพศ ส่วนใหญ่ข้อสอบที่เกิดการทำหน้าที่ต่างกันของข้อสอบ จะมีคำศัพท์ที่เกี่ยวข้องกับเพศนั้น จึงทำให้ข้อสอบเข้าข้างเพศนั้น และอาจเป็นเพราะความสามารถที่แตกต่างระหว่างเพศชายและเพศหญิงที่มีลักษณะความสามารถ ความถนัด และความสนใจในเรื่องนั้น ๆ ต่างกัน เมื่อจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน ส่วนใหญ่สาเหตุที่ทำให้ข้อสอบเกิดการทำหน้าที่ต่างกันอาจเป็นเพราะประสบการณ์ ความคุ้นเคยเกี่ยวกับเรื่องนั้น สภาพแวดล้อมและการฝึกปฏิบัติที่แตกต่างกันระหว่างนักเรียนในเขตกรุงเทพมหานครและปริมณฑลนักเรียนนอกเขตกรุงเทพมหานครและปริมณฑล

อาวีพร ปานทอง (2558) ได้ศึกษาการเปรียบเทียบประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบให้คะแนนหลายค่าโดย วิธีทดสอบอัตราส่วนความควรจะเป็น

วิธีเบส์เซียน และวิธีโพลี-ซิปเทสท์ ผลการศึกษาพบว่า 1) วิธีทดสอบอัตราส่วนความควรจะเป็น และวิธีเบส์เซียนมีอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ใกล้เคียงกัน ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนหลายค่าที่เป็นรูปแบบเดียวกัน ทั้งอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีโพลี-ซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน เมื่อความยาวของข้อสอบเพิ่มขึ้น วิธีทดสอบอัตราส่วนความควรจะเป็น และวิธีเบส์เซียน สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ดีกว่าวิธีโพลี-ซิปเทสท์ นอกจากนี้เมื่อขนาดตัวอย่างเพิ่มขึ้นมีผลทำให้ทุกวิธีมีอำนาจการทดสอบเพิ่มขึ้นภายใต้ทุกปัจจัย 2) โดยภาพรวมวิธีทดสอบอัตราส่วนความควรจะเป็น และวิธีเบส์เซียนมีอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในขอบเขตที่กำหนด ส่วนวิธีโพลี-ซิปเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าที่กำหนด ผลการศึกษาครั้งนี้เสนอแนะให้ใช้วิธีทดสอบอัตราส่วนความควรจะเป็น และวิธีเบส์เซียน เพื่อทำให้ควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้และมีอำนาจการทดสอบสูง

สุภะ อภิญญาภิบาล (2558) ได้ศึกษาการวิเคราะห์เปรียบเทียบประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค ระหว่างวิธีการถดถอยโลจิสติกวิธีซิปเทสท์และวิธีราสช์ทรี การศึกษาครั้งนี้ มีวัตถุประสงค์เพื่อเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี โดยการทดสอบระดับนัยสำคัญและการวัดขนาดอิทธิพล ข้อมูลที่ใช้ศึกษาจำลองภายใต้ทฤษฎีการตอบสนองข้อสอบ โมเดล 1 พารามิเตอร์ แล้วจัดกระทำข้อมูลตามปัจจัย 4 ปัจจัย คือ 1) ค่าความยากของข้อสอบ 3 ระดับ 2) ความยาวของแบบสอบ 2 ขนาด 3) สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด และ 4) ขนาดกลุ่มตัวอย่าง 3 ขนาด รวมข้อมูลที่ศึกษาทั้งหมด 36 เงื่อนไข ในแต่ละเงื่อนไขจำลองซ้ำ 50 ครั้ง และข้อมูลเชิงประจักษ์ศึกษาการทำหน้าที่ต่างกันของข้อสอบตามตัวแปร 3 ตัว คือ เพศ สังกัดของโรงเรียน และภูมิภาค แล้ววิเคราะห์ข้อมูลด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบทั้งหมด ใช้ระดับนัยสำคัญ .05 ผลการวิจัยสรุปได้ดังนี้ 1) อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญของวิธีซิปเทสท์มีค่าสูงกว่าวิธีการถดถอยโลจิสติกและวิธีราสช์ทรีภายใต้เกือบทุกเงื่อนไข และอำนาจการทดสอบของวิธีราสช์ทรีมีค่าสูงกว่าวิธีการถดถอยโลจิสติกภายใต้เกือบทุกเงื่อนไข ส่วนอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพลของวิธีซิปเทสท์มีค่าสูงกว่าวิธีการถดถอยโลจิสติกภายใต้เกือบทุกเงื่อนไข 2) อัตราความคลาด

เคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญของวิธีราสช์ทรีมีค่าต่ำกว่าวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ภายใต้เกือบทุกเงื่อนไข และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์มีค่าต่ากว่าวิธีการถดถอยโลจิสติกภายใต้เกือบทุกเงื่อนไข ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพลของวิธีการถดถอยโลจิสติกมีค่าต่ำกว่าวิธีซิปเทสท์ภายใต้เกือบทุกเงื่อนไข และ 3) ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับข้อมูลเชิงประจักษ์ด้วยวิธีราสช์ทรีมีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ แต่มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าเกณฑ์ที่กำหนดที่ระดับ 10%

อริสพา เตหลิ่ม (2559) ได้ศึกษาเปรียบเทียบประสิทธิผลการประมาณค่าพารามิเตอร์และการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมกซิมัมไลค์ลิฮู้ด วิธีของเบส์และวิธีของเบส์แบบมีอิทธิพลเทสต์เลท ผลการศึกษา พบว่า 1) ผลการประมาณค่าพารามิเตอร์ จำแนกเป็น (1) พารามิเตอร์ความยาก พบว่า วิธีของเบส์ แบบมีอิทธิพลเทสต์เลท (Bayesγ) ประมาณค่าได้ดีเมื่อข้อมูลมีการแจกแจงความสามารถเป็นแบบปกติส่วนวิธีของเบส์ (Bayes) ยังไม่มีแนวโน้มแน่นอน แต่จะประมาณค่าพารามิเตอร์ความยากได้ดีเป็นส่วนใหญ่เมื่อข้อมูลมีการแจกแจงความสามารถเป็นแบบเบ้ซ้าย และวิธีแมกซิมัมไลค์ลิฮู้ด (ML) จะประมาณค่าพารามิเตอร์ความยากได้ดี เมื่อข้อมูลมีการแจกแจงความสามารถเป็นแบบเบ้ขวา (2) พารามิเตอร์อำนาจจำแนก พบว่า ส่วนใหญ่วิธีการประมาณค่าด้วยวิธีของเบส์แบบมีอิทธิพลเทสต์เลท (Bayesγ) จะประมาณค่าพารามิเตอร์อำนาจจำแนกดีกว่าเมื่อข้อมูลความสามารถมีการแจกแจงแบบปกติ (3) พารามิเตอร์ความสามารถ พบว่า วิธีของเบส์แบบมีอิทธิพลเทสต์เลท (Bayesγ) ประมาณค่าได้ดีที่สุด 2) ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ พบว่า วิธีของเบส์แบบมีอิทธิพลเทสต์เลท (Bayesγ) และวิธีของเบส์ (Bayes) สามารถควบคุมอัตราความคาดเคลื่อนประเภทที่ 1 ได้ดี (ยกเว้นกรณีที่อิทธิพลเทสต์เลทมีค่าเป็น 0.25, 0.5, 1, 2 ร่วมกับการแจกแจงความสามารถที่เป็นแบบเบ้ซ้ายและเบ้ขวา การประมาณค่าด้วยวิธีของเบส์มีความคาดเคลื่อนประเภทที่ 1 สูง) และมีอำนาจการตรวจสอบสูงเมื่อมีการแจกแจงความสามารถแบบเบ้ซ้ายและจำนวนตัวอย่างมาก แต่ไม่มากถึงเกณฑ์ที่กำหนด ตรงข้ามกับวิธีแมกซิมัมไลค์ลิฮู้ด (ML) ซึ่งไม่สามารถควบคุมอัตราความคาดเคลื่อนประเภทที่ 1 แต่มีอำนาจการตรวจสอบสูง

Kim (2000) ได้ศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการทดสอบอัตราส่วนไลค์ลิฮู้ด (Likelihood ratio test) วิธีแมนเทล (Mantel) และวิธีแมนเทล-แฮนเซลส์แบบทั่วไป (Generalized mantel-haenszel: GMH) ใช้ข้อมูลจากการประเมินโรงเรียนระดับอนุบาลของรัฐจอเจียร์ ปัจจัยที่ศึกษา ได้แก่ ขนาดของกลุ่มตัวอย่าง (105,731: 10,000: 1,000: 100)

ข้อคำถามเป็นแบบให้คะแนนมากกว่า 2 ค่า ผลการศึกษาพบว่า วิธีตรวจสอบการทำหน้าที่ต่างกันทั้ง 3 วิธีให้ผลการตรวจสอบได้ดีเมื่อกลุ่มตัวอย่างมีขนาด 100 คน

Finch (2005) เปรียบเทียบประสิทธิภาพของโมเดล MIMIC กับการทดสอบโดยวิธีแมนเทลเฮนเซล (Mantel and Haenzel), 1959) วิธี SIBTEST (Shealy and Stout), 1993) และวิธีการทดสอบ IRT likehood ratio (Thissen et al., 1988) เกณฑ์การเปรียบเทียบพิจารณาจากความคลาดประเภท ที่ 1 และอำนาจการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ได้แสดงให้เห็นว่าวิธี MIMIC มีอำนาจการตรวจสอบสูงขึ้น และความคลาดประเภทที่ 1 มีค่าลดลงเมื่อข้อสอบ 50 ข้อ

Lei, Chen and Yu (2006) ได้ศึกษาเปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์จำลองข้อมูลการทำหน้าที่ต่างกันของข้อสอบทั้งแบบมีทิศทางและไม่มีทิศทาง ภายใต้เงื่อนไขที่ศึกษา คือ ขนาดกลุ่มตัวอย่างของกลุ่มอ้างอิงกับกลุ่มเปรียบเทียบ การแจกแจงความสามารถที่แตกต่างกัน ข้อสอบแบ่งเป็น 3 แบบ คือ ข้อสอบไม่เกิดการทำหน้าที่ต่างกัน ข้อสอบทำหน้าที่ต่างกันแบบมีทิศทาง ข้อสอบทำหน้าที่ต่างกันแบบไม่มีทิศทาง ใช้วิธีวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ 3 วิธี ได้แก่ วิธีถดถอยโลจิสติก วิธีการทดสอบอัตราส่วนไลด์ลิฮู้ดแบบ IRT (IRT likelihood ratio test) และวิธีแคทซิบ (CATSIB) ผลการศึกษาพบว่า วิธีถดถอยโลจิสติก (Logistic regression) และวิธีการทดสอบอัตราส่วนไลด์ลิฮู้ดแบบ IRT (IRT Likelihood ratio test) ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้งแบบมีทิศทางและแบบไม่มีทิศทางได้ดีเท่าเทียมกัน และดีกว่าวิธีแคทซิบ (CATSIB) ในขณะที่วิธีแคทซิบ (CATSIB) สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบมีทิศทางได้ดีกว่าแบบไม่มีทิศทาง

Stark, Chernyshenko and Drasgow (2006) ได้ศึกษาพัฒนาและทดสอบการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยการทดสอบอัตราส่วนไลค์ลิฮูด (Likelihood ratio test) ซึ่งสามารถใช้ได้ทั้งการวิเคราะห์องค์ประกอบเชิงยืนยันและทฤษฎีการตอบสนองข้อสอบ โดยใช้ข้อมูลจำลองในการตรวจสอบความตรงของทั้งสองวิธี เครื่องมือที่ใช้เป็นแบบวัดมิติเดียวจำนวน 15 ข้อ ศึกษาตัวแปร 8 ตัว วิเคราะห์วิธีวิเคราะห์องค์ประกอบเชิงยืนยัน โดยใช้โปรแกรมลิสเรล 8 และการทดสอบด้วยทฤษฎีตอบสนองข้อสอบด้วยวิธี Likelihood ratio test โดยใช้โปรแกรมคอมพิวเตอร์ MULTILOG ผลการวิจัย พบว่า ทฤษฎีการตอบสนองข้อสอบ วิธีทดสอบ Likelihood ratio test ให้ผลดีกว่าวิธีวิเคราะห์องค์ประกอบเชิงยืนยัน กรณีกลุ่มตัวอย่างมีขนาดใหญ่และข้อมูลเป็นแบบ Dichotomous มิติเดียว การตรวจสอบการทำหน้าที่ต่างกันของแบบสอบโดยใช้วิธีทฤษฎีการตอบสนองข้อสอบให้ผลดีกว่า และวิธีทฤษฎีการตอบสนองข้อสอบ หลาย ๆ วิธีจะมีความแกร่งในการทดสอบ (Robust) หากมีการฝ่าฝืนข้อตกลงเบื้องต้นในเรื่องความเป็นเอกมิติ

Burcu and Akihito (2011) ได้ศึกษาอัตราความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบของวิธี IRT likelihood ratio test และ วิธี Cumulative logit ordinal logistic regression ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ใช้การจำลอง Monte Carlo 54 เงื่อนไขการจำลอง (3 ขนาดกลุ่มตัวอย่าง 2 อัตราส่วนขนาดกลุ่มตัวอย่าง 3 ขนาดของ DIF และ 3 เงื่อนไขของ DIF) และแต่ละเงื่อนไขจำลองถูกทำซ้ำ 200 ครั้ง ผลการศึกษาพบว่า วิธี IRT likelihood ratio test และ วิธี Cumulative logit ordinal logistic regression สามารถควบคุมอัตราความคลาดเคลื่อนประเภท 1 ได้ดี ในทุกเงื่อนไขการจำลอง อำนาจการทดสอบของวิธี IRT likelihood ratio test มีค่าสูง สำหรับกลุ่มตัวอย่างขนาดกลางหรือใหญ่ และขนาด DIF ปานกลางหรือขนาดใหญ่ อำนาจการทดสอบเพิ่มขึ้นเมื่อขนาดของกลุ่มตัวอย่างหรือขนาดของ DIF เพิ่มขึ้น

Seokjoon (2014) ได้ศึกษาประสิทธิภาพของโมเดล MIMIC ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและเอนกรูป เมื่อมีกลุ่มตัวอย่างที่ศึกษาพร้อมกันหลายกลุ่ม ใช้วิธี Sequential-free baseline สำหรับวิเคราะห์โมเดล MIMIC ที่ได้ถูกพัฒนาขึ้น และเปรียบเทียบกับวิธี Free and constrained baseline model วิธี Sequential-free baseline ใช้ในการจำแนกข้อสอบที่ระบุไว้ว่ามี non-DIF ในการทดสอบ Constrained baseline model เพื่ออ้างอิงสำหรับการเปรียบเทียบกับ Free baseline model นอกจากนี้ยังได้มีการตรวจสอบผลของการละเมิดข้อตกลงเบื้องต้นของวิธี MIMIC DIF เกี่ยวกับการเท่ากันของความแปรปรวนของตัวแปรแฝง โดยรวมแล้วการศึกษาแบบจำลอง วิธี Sequential-free baseline พบว่า ความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบคล้ายกับ Free baseline method ที่เกี่ยวข้องกับรูปแบบที่กำหนดไว้ และความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบดีกว่า Constrained baseline model อย่างไรก็ตาม เมื่อมีการละเมิดข้อตกลงเบื้องต้นเกี่ยวกับความแปรปรวนของตัวแปรแฝง อัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้น แม้ว่าวิธีการของ MIMIC พบว่า มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ทั้งแบบเอกรูปและอเนกรูปก็ตาม ยังต้องมีการพัฒนาวิธีการเพิ่มเติมต่อไป

Finch (2016) ได้ศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มที่มากกว่าสองกลุ่ม การศึกษาครั้งนี้ใช้การจำลองข้อมูลในการเปรียบเทียบวิธีที่ใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มากกว่า 2 กลุ่ม 4 วิธี ได้แก่ The generalized mantel-haenszel test, Generalized logistic regression, Lord’s chi-square test, and the multiple group alignment procedure การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) เป็นวิธีการสำคัญในการพัฒนาการทดสอบ การพัฒนาเครื่องมือที่เหมาะสมกับหลายกลุ่มภายในประชากร เมื่อไม่ใช่กรณีนี้คะแนนที่ได้อาจจะไม่ถูกต้องในการสะท้อนให้เห็นถึงลักษณะที่สนใจสำหรับทุกคนในประชากร งานวิจัย เรื่อง DIF ส่วนใหญ่มุ่งเน้นไปที่กรณีที่สองกลุ่ม แต่ในทางปฏิบัตินักวิจัยอาจ

ต้องการที่จะตรวจสอบ DIF มากกว่าสองกลุ่ม นั่นคือ เชื้อชาติของผู้สอบ, ประเทศที่เกิดหรือสภาพอื่น ๆ ในกลุ่ม วิธีการตรวจสอบ DIF สำหรับกรณีดังกล่าวได้รับการเสนอ แต่ในการศึกษากับข้อมูลเชิงประจักษ์ยังมีน้อยที่ได้รับการดำเนินการเพื่อตรวจสอบประสิทธิภาพ ดังนั้น เป้าหมายของการศึกษาครั้งนี้คือการใช้วิธีการจำลองเพื่อเปรียบเทียบสี่วิธีในการประเมิน DIF ได้แก่ วิธี The generalized mantel-haenszel test, Generalized logistic regression, Lord's chi-square test, and the multiple group alignment procedure. ผลการศึกษาพบว่า The generalized mantel-haenszel test, and the multiple group alignment procedure เป็นวิธีที่ดีที่สุดสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 และอำนาจของการทดสอบ การศึกษานี้มีข้อจำกัดบางประการ ประการแรกมุ่งเน้นเฉพาะกรณี uniDIF ดังนั้น การศึกษาในอนาคตควรเปรียบเทียบวิธีการเหล่านี้โดยเฉพาะวิธี The multiple group alignment procedure GLR และ Lord สำหรับการทดสอบ Nonuniform DIF นอกจากนี้ควรใช้เงื่อนไขการจำลองแบบที่กว้างขึ้น

Chun, Stark, Kim and Chermvshenko (2016) ได้ศึกษาแบบจำลองเพื่อตรวจสอบประสิทธิภาพของโมเดล MIMIC สำหรับการตรวจสอบประสิทธิภาพการทำหน้าที่ต่างกันของข้อสอบ (DIF) ในหลายกลุ่มทั้งแบบเอกรูปและเอนกรูป DIF ได้ถูกจำลองโดยใช้ตัวแปรภูมิหลังสองตัว ได้แก่ เพศและชาติพันธุ์ เป็นการศึกษาเปรียบเทียบวิธี MIMIC 3 แบบ ได้แก่ Constrained baseline, free baseline, and a new sequential-free baseline เมื่อทดสอบข้อตกลงเบื้องต้นของ MIMIC ว่ามีความแปรปรวนของตัวแปรเท่ากันในแต่ละกลุ่มเปรียบเทียบ ผลการศึกษา พบว่า A new sequential-free baseline ให้ค่าความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบ ดีเยี่ยม ซึ่งให้ผลลัพธ์ที่คล้ายคลึงกับวิธี Free baseline method และให้ผลดีกว่าวิธี Constrained baseline ซึ่งใช้รายการทั้งหมดที่ไม่ใช่การศึกษา เป็นแบบสอบร่วม อย่างไรก็ตามเมื่อข้อตกลงเบื้องต้นเกี่ยวกับการเท่ากันของความแปรปรวนขององค์ประกอบถูกละเมิด วิธีการทั้งหมดเกิดความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้น

สรุปได้ดังตารางที่ 2-4

ตารางที่ 2-4 งานวิจัยที่เกี่ยวข้องกับการศึกษาประสิทธิภาพของวิธีและศึกษาเชิงเปรียบเทียบวิธีที่ใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

ผู้วิจัย	ปี	เรื่อง	ปัจจัย	องค์ความรู้ที่ได้จากการศึกษา
เรวดี อินทสระ	2539	ผลการตรวจสอบความลำเอียงของข้อสอบต่อการศึกษาความเที่ยงตรงเชิงพยากรณ์ของแบบทดสอบคัดเลือกที่คิดคะแนนต่างกัน	เพศ	วิธีการตรวจสอบความลำเอียงทั้ง 3 วิธี ตัดสินจำนวนข้อสอบที่ลำเอียงแตกต่างกัน โดยวิธี IRT ตัดสินจำนวนข้อสอบที่ลำเอียงได้มากที่สุด
จิติมา วรรณศรี	2539	การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซลกับวิธีซิบเทสท์ เมื่อความยาวของแบบสอบขนาดกลุ่มตัวอย่างและอัตราส่วนของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบต่างกัน	1. ความยาวของแบบสอบ 2. ขนาดกลุ่มตัวอย่าง	วิธี MH กับวิธี SIBTEST มีประสิทธิภาพเท่าเทียมกันทุกเงื่อนไขที่ศึกษา
รัชรินทร์ มุคดา	2540	การเปรียบเทียบระสิทธิภาพระหว่างวิธีแมนเทล-แฮนส์เซลกับวิธีถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันของแบบสอบแบบอเนกรูปในกรณีที่จัดกลุ่มความสามารถ ค่าความยากของข้อสอบ และค่าอำนาจจำแนกของข้อสอบต่างกัน	1. DIF แบบอเนกรูป 2. กลุ่มความสามารถ 3. ค่าความยาก 4. ค่าอำนาจจำแนก	วิธี MH กับวิธี LR ในการตรวจสอบการทำหน้าที่ต่างกันของแบบสอบอเนกรูปเท่ากันในทุกกลุ่มความสามารถของผู้สอบ

ตารางที่ 2-4 (ต่อ)

ผู้วิจัย	ปี	เรื่อง	ปัจจัย	องค์ความรู้ที่ได้จากการศึกษา
นพมาศ พิพัฒนสุข	2541	การเปรียบเทียบประสิทธิภาพวิธีแมนเทล-แฮนส์เซลกับวิธีถดถอยโลจิสติกในการตรวจสอบการทาหน้าที่ต่างกันของข้อสอบ เมื่อใช้เกณฑ์จับคู่เปรียบเทียบแตกต่างกันในแบบสอบชนิดพหุมิติ	เกณฑ์จับคู่เปรียบเทียบ	วิธีแมนเทล-แฮนส์เซลมีประสิทธิภาพมากกว่าวิธีถดถอยโลจิสติก เมื่อใช้เกณฑ์จับคู่คะแนนรวม วิธีถดถอยโลจิสติกเมื่อใช้เกณฑ์จับคู่คะแนนหลายแบบสอบย่อยมีความเหมาะสม
นิคม กีรติวรางกูร	2542	การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีการวิเคราะห์องค์ประกอบจำกัดแมนเทล-แฮนส์เซลและการตอบสนองข้อสอบ	1. ขนาดกลุ่มตัวอย่าง 2. ความยาวของแบบสอบ 3. ค่าความยาก 4. ค่าอำนาจจำแนก 5. ขนาดความลำเอียงของข้อสอบ	วิธี RFA มีประสิทธิภาพสูงที่สุด
ชัยยศ ชาวระนอง	2553	ประสิทธิภาพของโมเดลและการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบพหุมิติโดยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยันแฝงภายใน	เพศ ขนาดกลุ่มตัวอย่าง จำนวนข้อสอบ	โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองแฝงภายในมีประสิทธิภาพมากที่สุด โมเดลมีความสอดคล้องกับข้อมูลเชิงประจักษ์เมื่อไม่มีการปรับโมเดลอยู่ในระดับดี เมื่อใช้กลุ่มตัวอย่างตั้งแต่ 800 ตัวอย่างขึ้นไปวิธีวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองแฝงภายในจำแนกตามเพศโดยใช้เทคนิค Multi-group

ตารางที่ 2-4 (ต่อ)

ผู้วิจัย	ปี	เรื่อง	ปัจจัย	องค์ความรู้ที่ได้จากการศึกษา
				พบข้อสอบทำหน้าที่ต่างกัน 5 ข้อ และเมื่อตัดข้อสอบที่ทำหน้าที่ต่างกันออก พบว่า ค่าสัมประสิทธิ์ความเชื่อมั่นของข้อสอบสูงขึ้น
สุพัฒนา หอมบุปผา	2556	การเปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี HGLM วิธี MIMIC และวิธี BAYESIAN	วิชา เพศ สถานที่ตั้งทาง ภูมิศาสตร์ของโรงเรียน	วิธีตรวจสอบที่พบการทำหน้าที่ต่างกันของข้อสอบมากที่สุด คือ วิธี HGLM-2L ส่วนวิธีที่ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบน้อยที่สุด คือ วิธี MIMIC
พนัส จันทร์เปล่ง	2557	การวิเคราะห์เปรียบเทียบโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์: การประยุกต์ใช้โมเดลมูลค่าเพิ่มพหุระดับที่มีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ	เพศ เศรษฐานะ	ค่าความเที่ยงของแบบทดสอบในโมเดลที่ 2 ลดลงเมื่อข้อสอบที่ตรวจพบว่า ทำหน้าที่ต่างกันถูกตัดออก
อาวีพร ปานทอง	2558	การเปรียบเทียบประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบให้คะแนนหลายค่าโดยทดสอบอัตราส่วนความควรจะเป็น วิธีเบส์เซียน และวิธีโพลี-ซิปเทสท์	ความยาวแบบสอบ ขนาดกลุ่มตัวอย่าง	1.โดยภาพรวมวิธีทดสอบอัตราส่วนความควรจะเป็น และวิธีเบส์เซียนมีอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในขอบเขตที่กำหนด เสนอแนะให้ใช้วิธีทดสอบอัตราส่วนความควรจะเป็นและวิธีเบส์เซียน เพื่อทำให้ควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้และมีอำนาจการทดสอบสูง

ตารางที่ 2-4 (ต่อ)

ผู้วิจัย	ปี	เรื่อง	ปัจจัย	องค์ความรู้ที่ได้จากการศึกษา
สุภะ อภิญญาภิบาล	2558	การวิเคราะห์เปรียบเทียบประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี.	ค่าความยากของข้อสอบ, ความยาวของแบบสอบ, สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน, ขนาดกลุ่มตัวอย่าง	วิธีซิปเทสท์มีค่าสูงกว่าวิธีการถดถอยโลจิสติกและวิธีราสช์ทรีภายใต้เกือบทุกเงื่อนไข
อริสฟา เตหลิ่ม	2559	การเปรียบเทียบประสิทธิผลการประมาณค่าพารามิเตอร์และการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมกซิมัมไลค์ลิฮู้ด วิธีของเบส์และวิธีของเบส์แบบมีอิทธิพลเทสต์เล	เทสต์เลท	1. ผลการประมาณค่าพารามิเตอร์ จำแนกเป็น (1) พารามิเตอร์ความยาก พบว่า วิธีของเบส์แบบมีอิทธิพลเทสต์เลท (Bayesγ) ประมาณค่าได้ดีเมื่อข้อมูลมีการแจกแจงความสามารถเป็นแบบปกติและวิธีแมกซิมัมไลค์ลิฮู้ด (ML) จะประมาณค่าพารามิเตอร์ความยากได้ดี เมื่อข้อมูลมีการแจกแจงความสามารถเป็นแบบเบ้ขวา (2) พารามิเตอร์อำนาจจำแนก พบว่า ส่วนใหญ่วิธีการประมาณค่าด้วยวิธีของเบส์แบบมีอิทธิพลเทสต์เลท (Bayesγ) จะประมาณค่าพารามิเตอร์อำนาจจำแนกดีกว่าเมื่อข้อมูลความสามารถมีการแจกแจงแบบปกติ

ตารางที่ 2-4 (ต่อ)

ผู้วิจัย	ปี	เรื่อง	ปัจจัย	องค์ความรู้ที่ได้จากการศึกษา
				(3) พารามิเตอร์ความสามารถ พบว่า วิธีของเบส์แบบมีอิทธิพลเทสต์เลท (Bayesγ) ประมาณค่าได้ดีที่สุด 2. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ พบว่า วิธีของเบส์แบบมีอิทธิพลเทสต์เลท (Bayesγ) และวิธีของเบส์ (Bayes) สามารถควบคุมอัตราความคาดเคลื่อนประเภทที่ 1 ได้ดี (ยกเว้นกรณีที่อิทธิพลเทสต์เลทมีค่าเป็น 0.25, 0.5, 1, 2 ร่วมกับการแจกแจงความสามารถที่เป็นแบบเบ้ซ้ายและเบ้ขวา การประมาณค่าด้วยวิธีของเบส์มีความคาดเคลื่อนประเภทที่ 1 สูง) และมีอำนาจการตรวจสอบสูงเมื่อมีการแจกแจงความสามารถแบบเบ้ซ้ายและจำนวนตัวอย่างมาก แต่ไม่มากถึงเกณฑ์ที่กำหนด ตรงข้ามกับวิธีแมกซิมัมไลค์ลิฮู้ด (ML) ซึ่งไม่สามารถควบคุมอัตราความคาดเคลื่อนประเภทที่ 1 แต่มีอำนาจการตรวจสอบสูง
Kim	2000	การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการทดสอบอัตราส่วนไลค์ลิฮู้ด (Likelihood ratio test) วิธีแมนเทล Mantel และวิธีแมนเทล-แฮนเซลส์แบบทั่วไป (Generalized mantel-haenszel: GMH	ขนาดกลุ่มตัวอย่าง	ทั้ง 3 วิธีให้ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ดีเมื่อกลุ่มตัวอย่างมีขนาด 100 คน

ตารางที่ 2-4 (ต่อ)

ผู้วิจัย	ปี	เรื่อง	ปัจจัย	องค์ความรู้ที่ได้จากการศึกษา
Penfield	2001	การทำหน้าที่ต่างกันของข้อสอบในหลายกลุ่มด้วยวิธีแมนเทล-แฮนเซลส์ 3 แบบ	กลุ่มตัวอย่างที่ศึกษาพร้อมกันมีหลายกลุ่ม	วิธี แมนเทล-แฮนเซลส์แบบทั่วไป (Generalized mantel-haenszel: GMH) ดีที่สุดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในหลายกลุ่ม
Finch	2005	เปรียบเทียบประสิทธิภาพของโมเดล MIMIC กับการทดสอบโดยวิธีแมนเทลเฮนเซล (Mantel & Haenzel, 1959) วิธี SIBTEST (Shealy & Stout, 1993) และวิธีการทดสอบ IRT Likehood Ratio (Thissen et al., 1988)	เปรียบเทียบวิธีตรวจสอบ DIF	วิธี MIMIC มีอำนาจการตรวจสอบสูงขึ้น และความคลาดประเภทที่ 1 มีค่าลดลงเมื่อข้อสอบ 50 ข้อ
Lei et al.	(2006)	เปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์จำลองข้อมูลการทำหน้าที่ต่างกันของข้อสอบทั้งแบบมีทิศทางและไม่มีทิศทาง) วิเคราะห์ข้อมูลด้วยวิธีถดถอยโลจิสติก วิธีการทดสอบอัตราส่วนไลค์ลิฮู้ดแบบ IRT (IRT Likelihood ratio test) และวิธีแคทซิบ (CATSIB)	ขนาดกลุ่มตัวอย่าง	วิธีถดถอยโลจิสติก (Logistic regression) และวิธีการทดสอบอัตราส่วนไลค์ลิฮู้ดแบบ IRT (IRT Likelihood ratio test) ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้งแบบมีทิศทางและแบบไม่มีทิศทางได้ดีเท่าเทียมกัน และทั้งสองวิธีตรวจสอบ การทำหน้าที่ต่างกันของข้อสอบได้ดีกว่าวิธีแคทซิบ (CATSIB) ในขณะที่วิธีแคทซิบ (CATSIB) ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบมีทิศทางได้ดีกว่าแบบไม่มีทิศทาง

ตารางที่ 2-4 (ต่อ)

ผู้วิจัย	ปี	เรื่อง	ปัจจัย	องค์ความรู้ที่ได้จากการศึกษา
Oishi	(2006)	การตรวจสอบความเท่าเทียมของการวัดความพึงพอใจด้วยแบบวัดความพึงพอใจในชีวิตระหว่างกลุ่มตัวอย่างชาวอเมริกันและชาวจีนโดยใช้ Multigroup structural equation modeling (SEM), Multiple indicator multiple cause model (MIMIC) และทฤษฎีการตอบสนองข้อสอบ (IRT)	วัฒนธรรม	วิธี IRT แสดงให้เห็นความแตกต่างของค่าเฉลี่ยระหว่างกลุ่มชาวจีนและชาวอเมริกัน ข้อสอบที่มีความลำเอียงจะให้คะแนนโดยมีน้ำหนักน้อยกว่า
Stark et al.	2006	การระบุการทำหน้าที่ต่างกันของข้อสอบด้วยการทดสอบอัตราส่วนไลค์ลิฮูด (Likelihood ratio test) ซึ่งสามารถใช้ได้ทั้งการวิเคราะห์องค์ประกอบเชิงยืนยันและทฤษฎีการตอบสนองข้อสอบ	การทดสอบอัตราส่วนไลค์ลิฮูด (Likelihood ratio test) ด้วยวิธีที่ต่างกัน	IRT วิธี Likelihood ratio test ให้ผลดีกว่าวิธี MACS
Seokjoon	2014	ประสิทธิภาพของโมเดล MIMIC ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและเอนกรูป	การศึกษากลุ่มที่มากกว่า 2 กลุ่ม	วิธี MIMIC พบว่า มีประสิทธิภาพในการตรวจจับ DIF ทั้งแบบเอกรูปและแบบ เอนกรูป
Burcu and Akihito	2011	เปรียบเทียบวิธี วิธี IRT Likelihood ratio test และวิธี Cumulative logit ordinal logistic regression ในการตรวจสอบการทำ	ใช้การจำลอง Monte Carlo 54 เงื่อนไขการจำลอง (ขนาดกลุ่ม	วิธี IRT Likelihood ratio test และ วิธี Cumulative logit ordinal logistic regression สามารถควบคุมอัตรา

ตารางที่ 2-4 (ต่อ)

ผู้วิจัย	ปี	เรื่อง	ปัจจัย	องค์ความรู้ที่ได้จากการศึกษา
		หน้าที่ต่างกันของข้อสอบ (DIF)	ตัวอย่าง, อัตราส่วนขนาดกลุ่มตัวอย่าง, ขนาดของ DIF และเงื่อนไขของ DIF)	ความคลาดเคลื่อนประเภท 1 ได้ดี ในทุกเงื่อนไขการจำลอง
Finch	2016	ศึกษาวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มที่มากกว่าสองกลุ่ม	กลุ่มที่มากกว่า 2 กลุ่ม	วิธี GMH and MGCFA alignment procedure เป็นวิธีที่ดีที่สุดสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 และอำนาจของการทดสอบ
Chun et al.	2016	ตรวจสอบประสิทธิภาพของโมเดล MIMIC สำหรับการตรวจสอบประสิทธิภาพการทำหน้าที่ต่างกันของข้อสอบ (DIF) ในหลายกลุ่ม	กลุ่มที่มากกว่า 2 กลุ่ม	a new sequential-free baseline ให้ค่าความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบ ดีเยี่ยม

งานวิจัยที่เกี่ยวข้องกับปัจจัยที่ใช้จำแนกกลุ่ม

เกณฑ์หรือปัจจัยที่ใช้แบ่งกลุ่มเป็นตัวแปรหนึ่งที่มีผลต่อการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ การวิจัยครั้งนี้มีตัวแปรที่ใช้จำแนกกลุ่มที่ต้องจำแนกได้มากกว่า 2 กลุ่ม และสอดคล้องกับตัวแปรพื้นฐานของข้อมูลที่เป็นกลุ่มตัวอย่าง ซึ่งได้แก่ ข้อมูลผลการสอบรายข้อจากการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ระดับชั้นมัธยมศึกษาปีที่ 3 ได้แก่ สถานที่ตั้งทางภูมิศาสตร์และขนาดโรงเรียน บนพื้นฐานของงานวิจัยที่เกี่ยวข้องดังนี้

1. สถานที่ตั้งทางภูมิศาสตร์ การทำหน้าที่ต่างกันของของข้อสอบเป็นการเปรียบเทียบผลการตอบข้อสอบระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ดังนั้น สถานที่ตั้งทางภูมิศาสตร์ของกลุ่มตัวอย่างจึงมีความสำคัญในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในประเด็น

ภาษาที่ใช้ในข้อสอบเป็นภาษาไทยกลาง และทั้งสามวิชาเป็นข้อสอบที่เกี่ยวข้องกับสถานการณ์ สถานที่ ข้อสอบที่ดีย่อมไม่ทำหน้าที่ต่างกันเมื่อกลุ่มตัวอย่างสถานที่ตั้งทางภูมิศาสตร์ต่างกัน สุพัฒนา หอมบุปผา (2556) ได้ศึกษาเรื่อง การเปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี HGML วิธี MIMIC และวิธี BAYESIAN ศึกษากับข้อมูลคะแนนการสอบวัดผลสัมฤทธิ์ทางการเรียน เพื่อประเมินคุณภาพระดับชาติ ของนักเรียนชั้นประถมศึกษาปีที่ 3 ศึกษาปัจจัยเพศ และสถานที่ตั้งทางภูมิศาสตร์ ผลการศึกษาพบว่า เมื่อจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนส่วนใหญ่สาเหตุที่ทำให้ข้อสอบเกิดการทำหน้าที่ต่างกันอาจเป็นเพราะประสบการณ์ ความคุ้ยเคยเกี่ยวกับเรื่องนั้น สภาพแวดล้อม และการฝึกปฏิบัติที่แตกต่างกันระหว่างนักเรียนในเขตกรุงเทพมหานครและปริมณฑล นักเรียนนอกเขตกรุงเทพมหานครและปริมณฑล เกษร หว่างจิตร (2539) ได้ศึกษาการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย และวิชาภาษาอังกฤษ ของศูนย์ทดสอบทางการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ด้วยวิธีแมนเทล-เฮนส์เซล เมื่อจำแนกกลุ่มผู้สอบตามเพศ ภูมิลำเนา ประสบการณ์ในการสอบ และสังกัดของสถานศึกษา โดยมีกลุ่มตัวอย่างเป็นผู้สอบวิชาภาษาไทย จำนวน 506 คน และผู้สอบวิชา ภาษาอังกฤษ จำนวน 501 คน ผลการวิจัย พบว่า ข้อสอบที่พบ DIF ส่วนใหญ่เป็นแบบอเนกรูป ข้อสอบที่ DIF ทั้งแบบเอกรูปและอเนกรูปเกิดขึ้นเมื่อการวิเคราะห์จำแนกกลุ่มผู้สอบตามเพศมากที่สุด รองลงมาคือ จำแนกตามภูมิลำเนา สังกัดของสถานศึกษา ลักษณะของข้อสอบที่พบ DIF ส่วนใหญ่เป็นข้อสอบที่มีค่าอำนาจจำแนกค่อนข้างต่ำ ค่าความเที่ยงและค่าความตรงตามทฤษฎีของแบบสอบก่อนและหลังตัดข้อสอบที่พบ DIF ออกจากแบบสอบส่วนใหญ่ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ รักชนก ยี่สุ่นศรี (2544) ได้ศึกษาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบด้วยกระบวนการ ดี เอฟ ไอ ที กับแบบสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาของทบวงมหาวิทยาลัย ปีการศึกษา 2543 ครั้งที่ 1 วิชาภาษาอังกฤษและวิชาคณิตศาสตร์ จากกลุ่มตัวอย่าง จำนวน 4,000 คน และ 3,600 คน ตามลำดับ ทำการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบสำหรับกลุ่มผู้สอบเมื่อจำแนกตามเพศ และสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา ผลการวิจัย พบว่า แบบสอบวิชาภาษาอังกฤษทำหน้าที่ต่างกันตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษาของผู้สอบ ส่วนแบบสอบวิชาคณิตศาสตร์เมื่อจำแนกกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา มีข้อสอบที่พบ DIF แบบสอบฉบับก่อนและหลังตัดข้อสอบที่พบ DIF มีค่าความตรงเชิงโครงสร้างไม่แตกต่างกัน แบบสอบฉบับหลังตัดข้อสอบที่พบ DIF ส่วนใหญ่มีค่าความเที่ยงลดลง แต่ค่าฟังก์ชันสารสนเทศของแบบสอบมากขึ้น และความสัมพันธ์ระหว่างตำแหน่งของคะแนนรวมของผู้สอบก่อนและหลังตัดข้อสอบที่พบ DIF มีความสัมพันธ์กันในทางบวกอย่างมีนัยสำคัญ และ สุภะ อภิญญาภิบาล (2558) ได้ศึกษาการวิเคราะห์เปรียบเทียบ

ประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค ระหว่างวิธีการถดถอยโลจิสติกวิธีซิปเทสท์และวิธีราสช์ทรี การศึกษากับข้อมูลเชิงประจักษ์ศึกษาการทำหน้าที่ต่างกันของข้อสอบตามตัวแปร 3 ตัว คือ เพศ สังกัดของโรงเรียน และภูมิภาค ผลการวิจัยพบว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับข้อมูลเชิงประจักษ์ด้วยวิธีราสช์ทรีมีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ แต่มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าเกณฑ์ที่กำหนดที่ระดับ 10% จากงานวิจัยที่ได้ศึกษาภูมิภาคเป็นปัจจัยที่มีผลเกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบ ดังนั้น การศึกษาครั้งนี้ผู้วิจัยจึงได้ศึกษาปัจจัยสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนของกลุ่มตัวอย่าง โดยจำแนกเป็น 4 ภาค ได้แก่ ภาคกลาง ภาคเหนือ ภาคตะวันออกเฉียงเหนือ และภาคใต้

2. ขนาดโรงเรียน ขนาดโรงเรียนเกี่ยวข้องกับผู้เรียนในส่วนของความพร้อมด้านการจัดการศึกษาอันประกอบด้วย บุคคล งบประมาณ สภาพแวดล้อมภายในโรงเรียน โดยโรงเรียนที่มีขนาดใหญ่มักเป็นโรงเรียนที่ได้เปรียบในเชิงความพร้อมในการจัดการเรียนการสอนมากกว่าโรงเรียนขนาดเล็ก เช่น Kim (2000) ศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการทดสอบอัตราส่วนไลด์ลิฮู้ด (Likelihood ratio test) วิธีแมนเทล Mantel และวิธีแมนเทล-แฮนเซลส์แบบทั่วไป (Generalized Mantel-Haenszel) ข้อมูลจากการประเมินโรงเรียนระดับอนุบาลของรัฐจอเจียร์ ปัจจัยที่ศึกษาเป็นขนาดของกลุ่มตัวอย่าง (105,731: 10,000: 1,000: 100) ข้อคำถามเป็นแบบให้คะแนนมากกว่า 2 ค่า ผลการศึกษา พบว่า ทั้ง 3 วิธีสามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ดีเมื่อกลุ่มตัวอย่างมีขนาด 100 คน นิตยา เหมือดไชสง (2543) ได้ศึกษาการส่งอิทธิพลผ่านตัวกลางเชิงสาเหตุของปัจจัยด้านนักเรียน ครู และโรงเรียนไปยังผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ ซึ่งเป็นการวิเคราะห์อภิมานจากงานวิจัยเชิงทดลอง ผลการวิจัยพบว่า งานวิจัยทั้ง 197 เรื่อง มีค่าเฉลี่ยค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับของปัจจัยทั้ง 3 ด้านเท่ากับ .342 โดยแยกเป็นปัจจัยด้านนักเรียนมีค่าเฉลี่ยค่าสัมประสิทธิ์สหสัมพันธ์กับผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์สูงสุด (.376) รองลงมาได้แก่ปัจจัยด้านโรงเรียน (.318) และปัจจัยด้านครู (.303) เมื่อพิจารณาเป็นรายด้านพบว่า ด้านโรงเรียน ตัวแปรขนาดโรงเรียนมีค่าเฉลี่ยค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .307 ลักษณะเดียวกันกับที่ พิชิต ธรรมรักษ์ (2549) ได้ทำการศึกษาเรื่องปัจจัยที่มีผลต่อผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ของนักเรียนแผนการเรียนศิลป์ภาษาในกรุงเทพมหานคร ผลการวิจัยสรุปได้ว่า ปัจจัยที่สำคัญ 6 ตัวที่ส่งผลต่อผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ ของนักเรียนแผนศิลป์ภาษา ชั้นมัธยมศึกษาปีที่ 4 ได้แก่ ประสบการณ์สอนคณิตศาสตร์ ขนาดของโรงเรียน ความรู้พื้นฐานเดิม อาชีพผู้ปกครอง ความเป็นผู้นำทางวิชาการของผู้บริหาร เจตคติต่อวิชาคณิตศาสตร์ ซึ่งตัวแปรทั้ง 6 ตัวแปรร่วมกัน ทำนายผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ได้ร้อยละ

39.10 อิทธิฤทธิ์ พงษ์ปิยะรัตน์ (2551) ได้ศึกษาเกี่ยวกับการวิเคราะห์ข้อสอบและการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ: การวิเคราะห์พหุระดับ ผลการวิจัยพบว่า ปัจจัยผลการเรียนคณิตศาสตร์ในภาคเรียนที่ผ่านมาส่งผลต่อค่าเฉลี่ยของโอกาสในการตอบข้อสอบได้ถูกต้องในแต่ละโรงเรียนอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ผลการวิเคราะห์ระดับโรงเรียน (ระดับ 3) พบว่า ตัวแปรขนาดของโรงเรียนและตัวแปรความเป็นผู้นำทางวิชาการของผู้บริหารส่งผลต่อค่าเฉลี่ยของโอกาสในการตอบข้อสอบได้ถูกต้องในโรงเรียน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และ .05 ตามลำดับ การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยโมเดล HGLM สามารถตรวจสอบพบข้อสอบที่ทำหน้าที่ต่างกันสอดคล้องตรงกับผลการวิเคราะห์จากโปรแกรม BILOG-MG และได้ให้ข้อเสนอแนะไว้ว่า ควรศึกษาตัวแปรคุณลักษณะระดับโรงเรียนที่ส่งผลต่อโอกาสในการตอบข้อสอบได้ถูกต้องเพื่อทำนายการเกิด DIF ของข้อสอบร่วมกับการให้เหตุผลจากผู้เชี่ยวชาญ

จากรายงานของธนาคารโลกที่ได้มีการศึกษาปัญหาโรงเรียนขนาดเล็ก ดิลกะ ลัทธพิพัฒน์ (2559) กล่าวว่า จากการประเมินผลนักเรียนร่วมกับนานาชาติในด้านการรู้เรื่องการอ่านเมื่อปี พ.ศ. 2555 (2012 PISA Reading assessment) พบว่า เกือบ 1 ใน 3 ของนักเรียนไทยอายุ 15 ปี “รู้หนังสือไม่เพียงพอที่จะนำไปใช้งานได้” (Functionally illiterate) และยังขาดทักษะการวิเคราะห์ที่จำเป็นต่อการทำงานซึ่งต้องใช้ทักษะการอ่านที่สูงกว่าขั้นพื้นฐาน นอกจากนี้ โรงเรียนยังมีคุณภาพแตกต่างกันสูงขึ้นในช่วงที่ผ่านมา โดยพบว่า นักเรียนที่ด้อยโอกาสและมีผลการเรียนไม่ดีจะกระจุกตัวอยู่ตามโรงเรียนขนาดเล็กในหมู่บ้านในเขตชนบท เมื่อเปรียบเทียบกับมาตรฐานสากลแล้ว โรงเรียนมัธยมของไทยยังมีข้อจำกัดอย่างมากในด้านสื่อการเรียนรู้และด้านโครงสร้างพื้นฐาน ซึ่งเป็นอุปสรรคต่อการเรียนการสอนที่มีคุณภาพ นอกจากนี้ ระบบการศึกษาระดับมัธยมของไทยยังประสบปัญหาในด้านคุณภาพครูอีกด้วย โรงเรียนมัธยมในเขตชนบทนั้นมีจำนวนบุคลากรและทรัพยากรน้อยกว่าโรงเรียนที่อยู่ในเขตเมืองอย่างมาก การจัดสรรทรัพยากรของไทยนั้น มีความไม่เท่าเทียมสูงกว่าประเทศในกลุ่ม OECD มาก ซึ่งปัญหาดังกล่าว ทำให้ครูในโรงเรียนเหล่านี้ต้องสอนหลายวิชาและต้องสอนหลายระดับชั้นมากกว่าครูที่สอนในโรงเรียนขนาดใหญ่ นอกจากนี้ โรงเรียนขนาดเล็กในชนบทยังมีปัญหาจากการที่ครูมีคุณวุฒิต่ำและด้อยประสบการณ์ อีกด้วย ในการนำเสนอผลการประเมินคุณภาพการศึกษาระดับชาติ (O-NET) ได้นำเสนอจำแนกผลตามขนาดโรงเรียน แม้ไม่ได้บ่งชี้เชิงเปรียบเทียบก็ตาม แต่จากงานวิจัยข้างต้นขนาดโรงเรียนเป็นปัจจัยที่มีผลต่อความสามารถของผู้เรียน ดังนั้นการศึกษาครั้งนี้ จึงได้ศึกษาปัจจัยขนาดโรงเรียนในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยแบ่งขนาดโรงเรียนออกเป็น 4 ขนาด ตามข้อมูลผลการประเมินคุณภาพการศึกษาระดับชาติ (O-NET) ได้แก่ ขนาดใหญ่พิเศษ ขนาดใหญ่ ขนาด ปานกลาง และขนาดเล็ก

บทที่ 3
วิธีดำเนินการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT วิธี MGCFA ALIGNMENT และวิธี MIMIC และเพื่อเปรียบเทียบผลของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT วิธี MGCFA ALIGNMENT และวิธี MIMIC ภายใต้ปัจจัยที่ต่างกัน 2 ปัจจัย ได้แก่ สถานที่ตั้งทางภูมิศาสตร์ และขนาดโรงเรียน

โดยผู้วิจัยได้ออกแบบวิธีการดำเนินการวิจัยตามขั้นตอน ดังนี้

ตอนที่ 1 การกำหนดข้อมูล

ประชากรและกลุ่มตัวอย่าง

เครื่องมือที่ใช้ในการวิจัย

การเก็บรวบรวมข้อมูล

ตอนที่ 2 การวิเคราะห์ข้อมูลการทำหน้าที่ต่างกันของข้อสอบ

ตอนที่ 1 การกำหนดข้อมูล

ข้อมูลที่ใช้ในการศึกษาครั้งนี้ เป็นข้อมูลผลการสอบรายข้อจากการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ระดับชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 เพื่อนำข้อมูลมาวิเคราะห์หาข้อสรุปตามวัตถุประสงค์ของการวิจัย ข้อมูลที่นำมาศึกษา ได้แก่ สถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน ขนาดของโรงเรียน และคำตอบของนักเรียนรายบุคคลที่ผ่านการตรวจให้คะแนนแบบทวิวิภาค (1-0) วิชาคณิตศาสตร์ วิชาภาษาไทย และวิชาวิทยาศาสตร์ กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2559 สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ ครอบคลุมทุกพื้นที่ทั่วประเทศ วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ตามสภาพที่เกิดขึ้นจริง ถ้าวิธีตรวจสอบวิธีใด ตรวจพบข้อสอบที่ทำหน้าที่ต่างกันสอดคล้องกับเกณฑ์ของวิธี ถือว่า วิธีนั้นตรวจสอบการทำหน้าที่ต่างกันได้ถูกต้อง แนวทางในการวิเคราะห์ข้อมูลมีดังนี้

1. ศึกษาค้นคว้าเกี่ยวกับมโนทัศน์ของการทำหน้าที่ต่างกันของข้อสอบ และงานวิจัยที่เกี่ยวข้องจากเอกสาร วารสารวิชาการ และงานวิจัยทั้งในประเทศและต่างประเทศ

2. ศึกษารายละเอียดการวิเคราะห์ข้อมูลของแต่ละวิธีและศึกษาโปรแกรมคอมพิวเตอร์ที่ใช้วิเคราะห์

3. กำหนดแนวทางในการวิเคราะห์ข้อมูล ดังนี้

3.1 ประชากรและกลุ่มตัวอย่าง ศึกษากับข้อมูลเชิงประจักษ์ โดยพิจารณาความสมบูรณ์ครบถ้วนของข้อมูล จากคะแนนผลการสอบรายข้อจากการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ระดับชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 วิชาคณิตศาสตร์ วิชาภาษาไทย และวิชาวิทยาศาสตร์ คัดเลือกมาเฉพาะข้อสอบที่มีรูปแบบการตรวจให้คะแนน ทวิวิภาค (1-0) กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2559 สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ

3.2 วิเคราะห์สถิติพื้นฐาน

3.3 วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธี IRT LRT วิธี MGCFA-ALIGNMENT และวิธี MIMIC

3.4 เปรียบเทียบผลของการระบุการทำหน้าที่ต่างกันของข้อสอบจากการตรวจสอบระหว่างวิธี ด้วยวิธี IRT LRT วิธี MGCFA ALIGNMENT และวิธี MIMIC

4. สรุปผลการวิเคราะห์ และอภิปรายผล

5. เขียนรายงานการวิจัย

ประชากรและกลุ่มตัวอย่าง

ในปีการศึกษา 2559 สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) ดำเนินการจัดการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ชั้นมัธยมศึกษาปีที่ 3 เมื่อวันที่ 4-5 กุมภาพันธ์ พ.ศ. 2560 และประกาศผลสอบในวันที่ 25 มีนาคม พ.ศ. 2560 โดยได้ดำเนินการจัดสอบให้กับผู้เรียนที่กำลังศึกษาอยู่ในชั้นมัธยมศึกษาปีที่ 3 ทุกคนของโรงเรียนจากทุกสังกัด ในการศึกษาครั้งนี้กำหนดขอบเขตเฉพาะนักเรียนที่ศึกษาในโรงเรียน สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ ซึ่งมีจำนวนข้อมูลเหมาะสมกับวิธีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ โดยจำแนกตามตัวแปรที่ศึกษาครั้งนี้ ดังนี้

1. สถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน ได้แบ่งส่วนจังหวัดตามการกำหนดเขตจังหวัดรับผิดชอบของศึกษาธิการภาค (กระทวงศึกษาธิการ, 2560) รายละเอียดดังตารางที่ 3-1

ตารางที่ 3-1 การแบ่งกลุ่มจังหวัดตามศึกษาธิการภาค

ภาคที่	ภาค	ภูมิภาค	ที่ตั้งสำนักงานศึกษาธิการภาค	จังหวัด	รวมจังหวัด
1	กลาง	ภาคกลางตอนบน 1	ปทุมธานี	นนทบุรี ปทุมธานี พระนครศรีอยุธยา สระบุรี และกรุงเทพมหานคร	5
2		ภาคกลางตอนบน 2	ลพบุรี	ชัยนาท ลพบุรี สิงห์บุรี และอ่างทอง	4
3		ภาคกลางตอนกลาง	ฉะเชิงเทรา	ฉะเชิงเทรา นครนายก ปราจีนบุรี สมุทรปราการ และสระแก้ว	5
4		ภาคกลางตอนล่าง 1	ราชบุรี	กาญจนบุรี นครปฐม ราชบุรี และสุพรรณบุรี	4
5		ภาคกลางตอนล่าง 2	สมุทรสงคราม	ประจวบคีรีขันธ์ เพชรบุรี สมุทรสงครามและสมุทรสาคร	4
					22
6	ใต้	ภาคใต้ฝั่งอ่าวไทย	นครศรีธรรมราช	ชุมพร นครศรีธรรมราช พัทลุง และ สุราษฎร์ธานี	4
7		ภาคใต้ฝั่งอันดามัน	ภูเก็ต	กระบี่ ตรัง พังงา ภูเก็ต และระนอง	5
8		ภาคใต้ชายแดน	ยะลา	นราธิวาส ปัตตานี ยะลา สงขลา และสตูล	5
					14
9	ตะวันออก	ภาคตะวันออก	ชลบุรี	จันทบุรี ชลบุรี ตราด และระยอง	4
10	ตะวันออกเฉียงเหนือ	ภาคะวันออกเฉียงเหนือตอนบน 1	อุดรธานี	บึงกาฬ เลย หนองคาย หนองบัวลำภู และอุดรธานี	5
11		ภาคะวันออกเฉียงเหนือตอนบน 2	สกลนคร	นครพนม มุกดาหาร และสกลนคร	3
12		ภาคตะวันออกเฉียงเหนือตอนกลาง	ขอนแก่น	กาฬสินธุ์ ขอนแก่น มหาสารคาม และร้อยเอ็ด	4
13		ภาคตะวันออกเฉียงเหนือตอนล่าง 1	อุบลราชธานี	ยโสธร ศรีสะเกษ อำนาจเจริญ และอุบลราชธานี	4
14		ภาคตะวันออกเฉียงเหนือตอนล่าง 2	นครราชสีมา	ชัยภูมิ นครราชสีมา บุรีรัมย์ และสุรินทร์	4
					20

ตารางที่ 3-1 (ต่อ)

ภาคที่	ภาค	ภูมิภาค	ที่ตั้งสำนักงานศึกษาธิการภาค	จังหวัด	รวมจังหวัด
15	ภาคเหนือ	ภาคเหนือตอนบน 1	**เชียงใหม่**	เชียงใหม่ แม่ฮ่องสอน **ลำปาง** และลำพูน	4
16		ภาคเหนือตอนบน 2	เชียงราย	เชียงราย น่าน พะเยา และแพร่	4
17		ภาคเหนือตอนล่าง 1	พิษณุโลก	ตาก พิษณุโลก เพชรบูรณ์ สุโขทัยและอุตรดิตถ์	5
18		ภาคเหนือตอนล่าง 2	นครสวรรค์	กำแพงเพชร นครสวรรค์ พิจิตร และอุทัยธานี	4
					17

การแบ่งภูมิภาคของกระทรวงศึกษาธิการ (2560) แบ่งเป็น 5 ภูมิภาค ขณะที่สำนักงานสถิติแห่งชาติ (2553) แบ่งเป็น 4 ภูมิภาค และเมื่อพิจารณาเป็นรายจังหวัดในแต่ละภาคแล้ว งานวิจัยครั้งนี้จึงได้กำหนดภูมิภาคออกเป็น 4 ภูมิภาค โดยภาคกลางและภาคตะวันออกรวมเป็นภาคกลาง มีรายละเอียดแต่ละภาคดังนี้

ภาคกลาง ประกอบด้วย 26 จังหวัด
ภาคใต้ ประกอบด้วย 14 จังหวัด
ภาคตะวันออกเฉียงเหนือ ประกอบด้วย 20 จังหวัด
ภาคเหนือ ประกอบด้วย 17 จังหวัด

2. ขนาดโรงเรียน จำแนกเป็น 4 ขนาด ประกอบด้วย

2.1 ขนาดเล็ก มีจำนวนนักเรียนไม่เกิน 500 คน
2.2 ขนาดกลาง มีจำนวนนักเรียน ตั้งแต่ 501-1,500 คน
2.3 ขนาดใหญ่ มีจำนวนนักเรียน ตั้งแต่ 1,501-2,500 คน
2.4 ขนาดใหญ่พิเศษ มีจำนวนนักเรียน ตั้งแต่ 2,501 คนขึ้นไป

ประชากร

ประชากรที่ใช้ในการศึกษาครั้งนี้ เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ที่กำลังศึกษาอยู่ในโรงเรียน สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ ปีการศึกษา 2559 ที่เข้ารับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ชั้นมัธยมศึกษาปีที่ 3 วิชาคณิตศาสตร์ วิชาภาษาไทย และวิชาวิทยาศาสตร์ จำนวน 688,552 คน

กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการศึกษาครั้งนี้ เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ที่กำลังศึกษาอยู่ในโรงเรียน สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ ปีการศึกษา 2559 ที่เข้ารับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ชั้นมัธยมศึกษาปีที่ 3 วิชาภาษาไทย วิชาคณิตศาสตร์ และวิชาวิทยาศาสตร์ จำนวน 8,000 คน

ในการกำหนดขนาดกลุ่มตัวอย่างสำหรับการวิจัย ผู้วิจัยได้ศึกษาการกำหนดขนาดกลุ่มตัวอย่างจากประชากร ขนาดกลุ่มตัวอย่างที่เหมาะสมสำหรับการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ และการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ Hair, Black, Babin and Anderson (2010, pp. 661-662) เสนอไว้ว่า การตัดสินใจเกี่ยวกับขนาดของตัวอย่างต้องเป็นไปตามปัจจัยหลายประการ โดยพิจารณาจากความซับซ้อนของโมเดล และโมเดลการวัดที่ใช้ ปัจจัยที่นำมาพิจารณาได้แก่ ข้อตกลงเบื้องต้นเกี่ยวกับการแจกแจงของข้อมูล เทคนิคที่นำมาใช้ประมาณค่าความซับซ้อนของโมเดล การวิจัยครั้งนี้ศึกษามากกว่า 2 กลุ่ม ดังนั้น ต้องพิจารณาไปถึงกลุ่มย่อยที่นำมาศึกษา ข้อมูลที่ผิดพลาดหรือการสูญหายของข้อมูล (Missing data) และ ความแปรปรวนของความคลาดเคลื่อนของตัวแปร นักวิจัยต้องพิจารณาความซับซ้อนของโมเดลที่ใช้ด้วย ซึ่งตามเงื่อนไขแล้วกลุ่มตัวอย่างควรไม่ต่ำกว่า 10-20 เท่า ของจำนวนข้อคำถาม (Linda & James, 2008; นงลักษณ์ วิรัชชัย, 2542, หน้า 54; สุวิมล ติรกานันท์, 2553, หน้า 76) งานวิจัยนี้มีข้อคำถาม 105 ข้อ จำนวนกลุ่มตัวอย่างควรมีอย่างน้อย 1,050 คน ถึง 2,100 คน แต่เพื่อความมั่นใจในการวิเคราะห์ข้อมูล สมโภชน์ อเนกสุข (2553, หน้า 81) ให้แนวคิดว่าจำนวนกลุ่มตัวอย่างใกล้เคียงกับจำนวนประชากรมาก ความคลาดเคลื่อนของผลสรุปที่ได้จะมีค่าน้อย และข้อมูลที่ได้ศึกษามีจำนวนเพียงพอ รวมทั้งปัจจัยที่ศึกษาครั้งนี้ ได้แก่ สถานที่ตั้งทางภูมิศาสตร์ จำนวนกลุ่มย่อยของภูมิภาคมี 4 กลุ่ม (รวมจำนวนกลุ่มอ้างอิงกับกลุ่มเปรียบเทียบ) และขนาดโรงเรียน จำนวนกลุ่มย่อยของขนาดโรงเรียน มี 4 กลุ่ม เท่ากัน สรุปได้ว่า กลุ่มที่มากที่สุดมี 4 กลุ่ม ผู้วิจัยจึงได้ขยายจำนวนกลุ่มตัวอย่างจากกลุ่มย่อยให้ได้กลุ่มละ 500 คน 4 ขนาด 4 ภาค รวมทั้งสิ้น 8,000 คน จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ และขนาดโรงเรียน

โดยมีขั้นตอนการสุ่มตัวอย่างแบบหลายขั้นตอน (Multi-stage random sampling technique) ดังนี้

ขั้นที่ 1 ใช้การสุ่มแบบแบ่งชั้นภูมิผสมกับการสุ่มแบบกลุ่ม (Stratified cluster sampling) โดยการแบ่งจังหวัดทั้งหมดออกตามภูมิภาค แล้วสุ่มจังหวัดจากแต่ละภาคในขั้นนี้ภาคเป็นชั้นภูมิและจังหวัดเป็นกลุ่มของโรงเรียน สุ่มได้ 11 จังหวัด ดังนี้

ภาคกลาง	ประกอบด้วย 26 จังหวัด	สุ่มมา 4 จังหวัด
ภาคใต้	ประกอบด้วย 14 จังหวัด	สุ่มมา 2 จังหวัด
ภาคตะวันออกเฉียงเหนือ	ประกอบด้วย 20 จังหวัด	สุ่มมา 3 จังหวัด
ภาคเหนือ	ประกอบด้วย 17 จังหวัด	สุ่มมา 2 จังหวัด

ขั้นที่ 2 ในแต่ละจังหวัดใช้การสุ่มแบบแบ่งชั้นภูมิผสมกับการสุ่มแบบกลุ่ม (Stratified cluster sampling) โดยแบ่งโรงเรียนในแต่ละจังหวัดเป็น 3 ชั้นภูมิ ตามขนาดของโรงเรียน คือ ใหญ่ กลาง และเล็ก แล้วสุ่มโรงเรียนมาจากแต่ละชั้นภูมิ ในขั้นนี้ตัวแปรเกณฑ์ที่ใช้ในการแบ่งเป็นชั้นภูมิ คือขนาดของโรงเรียน และโรงเรียนคือกลุ่มของนักเรียน โดยกำหนดขนาดกลุ่มตัวอย่างขนาดละ 250 คน

ขั้นที่ 3 ในแต่ละโรงเรียนสุ่มนักเรียนชั้นมัธยมศึกษาปีที่ 3 โดยวิธีสุ่มแบบง่าย จนได้ตามจำนวนที่ต้องการ

ตารางที่ 3-2 ประชากรและกลุ่มตัวอย่าง

ภาค	จังหวัด	ขนาดโรงเรียน S		M		L		XL		รวมทั้งสิ้น	
		ประชากร	กลุ่มตัวอย่าง	ประชากร	กลุ่มตัวอย่าง	ประชากร	กลุ่มตัวอย่าง	ประชากร	กลุ่มตัวอย่าง	ประชากร	กลุ่มตัวอย่าง
กลาง	กทม.	147	35	2641	177	8287	305	29272	349	40347	866
	นครปฐม	543	131	1479	99	1972	73	3391	40	7385	343
	อยุธยา	1213	292	2079	139	1016	37	2569	31	6877	499
	สมุทรปราการ	171	42	1269	85	2310	85	6664	80	10414	292
	รวม	2074	500	7468	500	13585	500	41896	500	65023	2000
เหนือ	เชียงใหม่	2331	283	4787	273	2319	312	1927	226	11364	1094
	เชียงราย	1791	217	3988	227	1397	188	2344	274	9520	906
	รวม	4122	500	8775	500	3716	500	4271	500	20884	2000
ตะวันออก	อุดรธานี	3006	140	4963	168	1701	167	4328	130	13998	605
เฉียงเหนือ	ขอนแก่น	3609	168	4011	135	1247	122	5342	160	14209	586
	โคราช	4112	192	5834	197	2146	211	6995	210	19087	809
	รวม	10727	500	14808	500	5094	500	16665	500	47294	2000

ตารางที่ 3-2 (ต่อ)

ภาค	จังหวัด	ขนาดโรงเรียน S ประชากร	S กลุ่มตัวอย่าง	M ประชากร	M กลุ่มตัวอย่าง	L ประชากร	L กลุ่มตัวอย่าง	XL ประชากร	XL กลุ่มตัวอย่าง	รวมทั้งสิ้น ประชากร	รวมทั้งสิ้น กลุ่มตัวอย่าง
ใต้	นครศรี	2011	339	3705	336	2985	297	3259	209	11960	1181
	สงขลา	957	161	1814	164	2040	203	4532	291	9343	819
รวม		2968	500	5519	500	5025	500	7791	500	21303	2000
รวมทั้งสิ้น		19891	2000	36570	2000	27420	2000	70623	2000	154504	8000

หมายเหตุ ขนาดโรงเรียน S = ขนาดเล็ก, M = ขนาดกลาง, L = ขนาดใหญ่, XL = ขนาดใหญ่พิเศษ

เครื่องมือที่ใช้ในการวิจัย

การวิจัยครั้งนี้ใช้ผลการทดสอบของนักเรียน จากการสอบด้วยแบบทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ระดับชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 วิชาคณิตศาสตร์วิชาภาษาไทย และวิชาวิทยาศาสตร์ คัดเลือกมาเฉพาะข้อสอบที่มีรูปแบบการตรวจให้คะแนน แบบทวิวิภาค (1-0) ซึ่งประกอบด้วยจำนวนข้อสอบดังนี้

1. วิชาคณิตศาสตร์ จำนวน 25 ข้อ
2. วิชาภาษาไทย จำนวน 40 ข้อ
3. วิชาวิทยาศาสตร์ จำนวน 40 ข้อ

ขั้นตอนการสร้างและพัฒนาแบบสอบ

แบบทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ระดับชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 มีเนื้อหาสาระตามหลักสูตรการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 โดยมีมาตรฐานการผลิตและการพัฒนา ดังนี้ (สัมพันธ์ พันธุ์พฤกษ์, 2560)

ตารางที่ 3-3 มาตรฐานการผลิตและการพัฒนา แบบทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ระดับชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559

จัดทำ Test blueprint	ออกข้อสอบตาม จัดทำ Test blueprint	พิมพ์ต้นฉบับ
1. ผู้เชี่ยวชาญด้านหลักสูตรประชุมเพื่อวิเคราะห์หลักสูตร (ศึกษาแก่นและกรอบของความรู้ในกลุ่มสาระการเรียนรู้ทั้ง 8 กลุ่มสาระ ฯ ตามโครงสร้างหลักสูตร ฯ) จัดทำร่าง Test blueprint ประกอบด้วย สาระ มาตรฐานการเรียนรู้ มาตรฐานการเรียนรู้ช่วงชั้น หัวข้อเนื้อหา ระดับชั้นพฤติกรรม (จำ เข้าใจ นำไปใช้ วิเคราะห์ สังเคราะห์ ประเมินค่า) 2. ผู้เชี่ยวชาญด้านหลักสูตรสร้างตัวชี้วัดพฤติกรรม (วิเคราะห์มาตรฐานการเรียนรู้ช่วงชั้น เพื่อสร้างตัวชี้วัดพฤติกรรม) สำหรับการวัดและประเมินผลของแต่ละมาตรฐานการเรียนรู้ช่วงชั้น เลือกระดับชั้นพฤติกรรมที่เหมาะสมกับธรรมชาติวิชา 3. ผู้เชี่ยวชาญด้านหลักสูตร และคณะทำงานกลั่นกรองข้อสอบชุดที่ 1 ร่วมกันพิจารณาและปรับปรุงแก้ไข Test blueprint ตัวชี้วัดพฤติกรรม และน้ำหนักเนื้อหา	1. สทศ. ประสานงานเชิญครูทุกสังกัดมาออกข้อสอบ (ซึ่งจะแต่งตั้งเป็นคณะทำงานสร้างข้อสอบ) 2. สทศ. แต่งตั้งคณะทำงานสร้างและกลั่นกรองข้อสอบ 3. คณะทำงานสร้างข้อสอบตามมาตรฐานการเรียนรู้ 4. คณะทำงานกลั่นกรองข้อสอบ ชุดที่ 1 จัดฉบับตาม Test blueprint 5. คณะทำงานกลั่นกรองข้อสอบ ชุดที่ 1 กลั่นกรองข้อสอบที่จัดฉบับแล้ว โดยพิจารณาว่า 1) ข้อสอบตรงตามสิ่งที่ต้องการวัด 2) ข้อสอบตรงตามวัตถุประสงค์ในการวัดและสอดคล้องกับมาตรฐานการเรียนรู้ 3) ข้อสอบมีระดับความยาก-ง่าย ตามกำหนด 6. คณะทำงานกลั่นกรองข้อสอบ ชุดที่ 2 (เป็นผู้เชี่ยวชาญด้านเนื้อหาที่เป็น External review อิสระจากคณะทำงานกลั่นกรองข้อสอบชุดที่ 1 ซึ่งต้องมีคุณธรรม จริยธรรม ขั้นสูง) พิจารณา 1) ความถูกต้องของเนื้อหา 2) เฉลยถูกต้องหรือไม่ และปรับแก้ข้อสอบ 7. ผู้เชี่ยวชาญด้านภาษาและการสื่อสาร ตรวจสอบความถูกต้องของข้อสอบและความชัดเจนของภาษา	1. เจ้าหน้าที่ สทศ. พิมพ์ต้นฉบับข้อสอบลงกระดาษไข 2. คณะทำงานกลั่นกรองข้อสอบชุดที่ 2 ตรวจสอบความถูกต้องของต้นฉบับในกระดาษไข 3. คณะทำงานกลั่นกรองข้อสอบชุดที่ 2 ทำเฉลย

โครงสร้างข้อสอบ

ตารางที่ 3-4 จำนวนข้อสอบ O-NET ชั้นมัธยมศึกษาปีที่ 3 จำแนกตามรูปแบบข้อสอบ ประจำปีการศึกษา 2559

ที่	รูปแบบ	วิชา									
		ภาษาไทย		สังคมศึกษาฯ		ภาษาอังกฤษ		คณิตศาสตร์		วิทยาศาสตร์	
		จำนวนข้อ	คะแนน	จำนวนข้อ	คะแนน	จำนวนข้อ	คะแนน	จำนวนข้อ	คะแนน	จำนวนข้อ	คะแนน
1	ปรนัย										
	1.1 4 ตัวเลือก 1 คำตอบ	40	80	45	90	50	100	25	80	40	80
	1.2 4 ตัวเลือก 2 คำตอบ	-	-	-	-	-	-	-	-	-	-
	1.3 5 ตัวเลือก 1 คำตอบ	-	-	-	-	-	-	-	-	-	-
	1.4 5 ตัวเลือก 2 คำตอบ	10	20	-	-	-	-	-	-	-	-
2	ปรนัย หลายตัวเลือก 1 คำตอบ	-	-	5	10	-	-	-	-	-	-
3	ปรนัย หลายตัวเลือก มากกว่า 1 คำตอบ	-	-	-	-	-	-	-	-	5	20
4	เลือกคำตอบจากแต่ละหมวดที่สัมพันธ์กัน	-	-	-	-	-	-	-	-	-	-
5	ระบายคำตอบที่เป็นค่า/ ตัวเลข	-	-	-	-	-	-	5	20		
6	อัตนัย	-	-	-	-	-	-	-	-	-	-
	รวม	50	100	50	100	50	100	30	100	45	100
	จำนวนเวลาที่ใช้สอบ (นาที)	90 นาที		90 นาที		90 นาที		90 นาที		90 นาที	

การเก็บรวบรวมข้อมูล

ข้อมูลที่ใช้ในการวิจัยครั้งนี้เป็นคะแนนผลการสอบรายข้อจากแบบทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ระดับชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 โรงเรียน สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ โดยมีขั้นตอนการเก็บรวบรวมข้อมูลดังนี้

1. ผู้วิจัยได้ขอหนังสือจากภาควิชาวิจัยและจิตวิทยาประยุกต์ คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา เพื่อขอข้อมูลผลการตอบข้อสอบรายข้อของนักเรียน จากสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน)

2. ผู้วิจัยได้ขอความร่วมมือจากเจ้าหน้าที่ผู้รับผิดชอบข้อมูลของสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) ในการสุ่มข้อมูลผลการตอบข้อสอบรายข้อของกลุ่มตัวอย่างตามคุณลักษณะและขนาดกลุ่มตัวอย่างที่กำหนดในการวิจัย

3. ตรวจสอบความเรียบร้อยของข้อมูล จากขั้นตอนที่ 2 เป็นที่เรียบร้อย จึงได้นำข้อมูลมาศึกษาวิเคราะห์ตามวัตถุประสงค์ของการวิจัยต่อไป

ตอนที่ 2 การวิเคราะห์ข้อมูลการทำหน้าที่ต่างกันของข้อสอบ

ผู้วิจัยได้กำหนดขั้นตอนของการวิเคราะห์ข้อมูลการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT วิธี MGCFA-ALIGNMENT และวิธี MIMIC กับข้อมูลที่นำมาวิเคราะห์เป็นคะแนนผลการสอบรายข้อจากการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ระดับชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 โรงเรียน สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ ที่มีรูปแบบการตรวจให้คะแนนแบบทวิวิภาค (1-0) รายละเอียดของการวิเคราะห์มีดังนี้

1. การวิเคราะห์ข้อมูลค่าสถิติพื้นฐานของคะแนนผลการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ระดับชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 วิชาคณิตศาสตร์ วิชาภาษาไทย และวิชาวิทยาศาสตร์ สถิติที่ใช้ได้แก่ ค่าเฉลี่ย (Mean) ค่าส่วนเบี่ยงเบนมาตรฐาน ค่าความเบ้ (Skewness) ค่าความโด่ง (Kurtosis) เพื่อพิจารณาลักษณะการแจกแจงของคะแนนการสอบทั้ง 3 วิชา

2. การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF)

ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ผู้วิจัยนำคะแนนผลการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ระดับชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 วิชาคณิตศาสตร์ วิชาภาษาไทย และวิชาวิทยาศาสตร์ มาวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ

ด้วยวิธี 3 วิธี ได้แก่ วิธี IRT LRT วิธี MGCFA-ALIGNMENT และวิธี MIMIC โดยมีรายละเอียดของแต่ละวิธี ดังตารางที่ 3-5

ตารางที่ 3-5 วิธี IRT LRT วิธี MGCFA-ALIGNMENT และวิธี MIMIC

ที่	วิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF)	โปรแกรมที่ใช้วิเคราะห์ข้อมูล
1	IRT LRT	IRTPRO (Cai et al., 2011)
2	MGCFA-ALIGNMENT	Mplus 7.11 (Muthén & Muthén, 2007)
3	MIMIC	Mplus 7.11 (Muthén & Muthén, 2007)

วิธี IRT LRT

การวิเคราะห์ด้วยวิธี IRT LRT ในการศึกษาครั้งนี้ผู้วิจัยใช้โปรแกรม IRTPRO (Cai et al., 2011) ในการประมาณค่าพารามิเตอร์ โดยมีสมมติฐานในการทดสอบสำหรับโมเดลดังนี้ข้อสอบโดยใช้สูตร ดังนี้

$H_0 : a_{jR} = a_{jF}$ และ $b_{ijR} = b_{ijF}$ สำหรับทุก j

H_A : พารามิเตอร์ของข้อสอบในข้อที่ j ของทั้งสองกลุ่มไม่เท่ากัน อย่างน้อย 1 พารามิเตอร์

สูตรการทดสอบอัตราส่วนความเป็นไปได้ของสองโมเดล แสดงได้ดังนี้

$$LR = \frac{L^{*}(Model_C)}{L^{*}(Model_A)} \qquad [3\text{-}1]$$

เมื่อ $L^{*}(Model_C)$ คือ ฟังก์ชันความเป็นไปได้ของโมเดลพื้นฐานค่าพารามิเตอร์ที่น้อยกว่า

$L^{*}(Model_A)$ คือ ฟังก์ชันความเป็นไปได้ของโมเดลเปรียบเทียบที่ยอมให้พารามิเตอร์ข้อสอบของข้อสอบข้อที่ j ข้ามกลุ่มผู้สอบมีความหลากหลาย

กำหนดฟังก์ชันไลด์ลิฮูด (Likelihood function) ในการประมาณค่าพารามิเตอร์จากค่าที่มากที่สุดในฟังก์ชัน (Maximum likelihood function) ซึ่งสามารถแสดงได้ดังสูตรทางคณิตศาสตร์ดังนี้

$$L(X|\theta,a,b)=\prod_{i=1}^{N}\prod_{j=1}^{J}\prod_{h=1}^{k} P(\theta_i,a_j,b_j)^{x_{ij}}\left[1-P(\theta_i,a_j,b_j)^{1-x_{ij}}\right] \quad [3\text{-}2]$$

กำหนดให้ i แทน ผู้ตอบคำถามคนที่ 1, 2, …, N

J แทน ข้อคำถาม ข้อที่ 1, 2, …, J

h แทน ตัวเลือกในข้อคำถาม ตัวเลือกที่ 1, 2, …, k

นั่นคือ ในโมเดลพื้นฐาน (Compact) จะประกอบด้วยข้อสอบที่ทำหน้าที่ไม่ต่างกัน สำหรับในโมเดลเปรียบเทียบจะประกอบด้วยข้อสอบที่มีค่าพารามิเตอร์ระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีค่าแปรเปลี่ยนไปตามกลุ่มโมเดลเปรียบเทียบ (Augmented) อาจมีข้อสอบจำนวน 1 ข้อ หรือมากกว่าที่ทำหน้าที่ต่างกัน นอกจากนี้ ระหว่าง 2 โมเดล จะต้องมีข้อสอบร่วมซึ่งเป็นข้อสอบที่สมมติว่าทำหน้าที่ไม่ต่างกัน จากนั้นจึงทำการเปรียบเทียบระหว่าง 2 โมเดลด้วยสถิติอัตราส่วนความเป็นไปได้ ดังนี้

$$G_j^2 = -2\mathrm{Ln}(LR) \quad [3\text{-}3]$$

$$G^2 = -2\mathrm{Ln}(L^*(\mathrm{Model}_C) - (-2\mathrm{Ln}(L^*(\mathrm{Model}_A)) \quad [3\text{-}4]$$

เมื่อ G_j^2 แทน สถิติอัตราส่วนความเป็นไปได้ของข้อสอบข้อที่ i

$L^*(\mathrm{Model}_C)$ แทน ฟังก์ชันความเป็นไปได้ของโมเดลพื้นฐาน (Compact)

$L^*(\mathrm{Model}_A)$ แทน ฟังก์ชันความเป็นไปได้ของโมเดลเปรียบเทียบ (Augmented)

โดยทั่วไปแล้ว $L^*(\mathrm{Model}_C) < L^*(\mathrm{Model}_A)$ และสถิติ $G_j^2 > 0$ มีการแจกแจงแบบไค-สแควร์ ซึ่งระดับของความเป็นอิสระเท่ากับจำนวนพารามิเตอร์ในโมเดล $L^*(\mathrm{Model}_C)$ และโมเดล $L^*(\mathrm{Model}_A)$ (ATAR, 2011) ดัชนีความไม่กลมกลืนในระดับนัยสำคัญทางสถิติของผลการทดสอบจะชี้บอกว่าโมเดลพื้นฐานกลมกลืนน้อยกว่าโมเดลเปรียบเทียบอย่างมีนัยสำคัญทางสถิติในการใช้การทดสอบอัตราส่วนความเป็นไปได้ ค่าไค-สแควร์ที่องศาความเป็นอิสระเท่ากับจำนวนพารามิเตอร์ของข้อสอบเหล่านั้นมีนัยสำคัญทางสถิติ แสดงว่ามีการทำหน้าที่ต่างกันของข้อสอบ

วิธี MGCFA-ALIGNMENT

วิธี multiple group alignment หรือในการวิจัยครั้งนี้ใช้คำว่า วิธี MGCFA-ALIGNMENT เป็นวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยประยุกต์มาจากการวิเคราะห์องค์ประกอบเชิงยืนยันแบบหลายกลุ่ม (MGCFA) นำเสนอโดย Asparouhov and Muthén (2014) วิธี MGCFA นำมาประยุกต์ใช้สำหรับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ภายใต้ทฤษฎีการตอบสนองข้อสอบ (IRT) ที่มีการเชื่อมโยงโดยตรงระหว่างค่าพารามิเตอร์จากการวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) โดยค่าความยากของข้อสอบตามทฤษฎีการตอบสนองข้อสอบ (IRT) จะเชื่อมโยงค่า CFA item thresholds และค่าอำนาจจำแนกตามทฤษฎีการตอบสนองข้อสอบ (IRT) จะเชื่อมโยงกับ CFA factor loadings

รูปแบบ CFA สำหรับการตอบสนองข้อสอบที่ให้คะแนนแบบ 2 ค่า ซึ่งเป็นตัวแปรสังเกตได้สามารถแสดงเป็นสมการ

$$y_{kig} = \nu_{kg} + \lambda_{kg}\eta_{ig} + \varepsilon_{kig} \qquad [3\text{-}5]$$

เมื่อ y_{kig} = การตอบสนองต่อข้อที่ k ของผู้สอบคนที่ i ในกลุ่ม g

ν_{kg} = ค่า Threshold สำหรับข้อที่ k ในกลุ่ม g (ค่าความยาก)

λ_{kg} = ค่า Factor loading สำหรับข้อที่ k ในกลุ่ม g (ค่าอำนาจจำแนก)

η_{ig} = คุณลักษณะแฝงของผู้สอบคนที่ i ในกลุ่ม g

ε_{kig} = ความคลาดเคลื่อนจากการสุ่มของผู้สอบคนที่ i ของข้อที่ k ในกลุ่ม g

จากนั้น วิธี Alignment มีขั้นตอนต่อไปนี้:

1. ประมาณค่า the configural factor model ให้ λ_{kg} และ V_{kg} แตกต่างกันในกลุ่ม แต่กำหนดโครงสร้างขององค์ประกอบภายในกลุ่มให้เหมือนกัน เรียกว่าเป็นโมเดล M0

2. แปลงค่าเฉลี่ยขององค์ประกอบ α_g เป็น 0 และความแปรปรวนขององค์ประกอบเป็น ψ_g เป็น 1 สำหรับแต่ละกลุ่ม

3. ค่าความแปรปรวนรายข้อ (V) และค่าเฉลี่ย ($E(y_{kg})$) ในเทอมของความแปรปรวนขององค์ประกอบ ค่าเฉลี่ย และค่าพารามิเตอร์รายข้อสำหรับโมเดล M0

ค่าพารามิเตอร์สำหรับ M0:

$$V(y_{kig}) = \lambda_{kg}^2 \psi_g = \lambda_{kg0}^2 \quad [3\text{-}6]$$

$$E(y_{kig}) = \nu_{ik} + \lambda_{kg}\alpha_g = \nu_{kg0} \quad [3\text{-}7]$$

ค่า Factor loading และค่า Threshold ของโมเดล M0 คือ

$$\lambda_{kg0} = \lambda_{kg}\sqrt{\psi_g} \quad [3\text{-}8]$$

$$\nu_{kg0} = \nu_{kg} + \frac{\lambda_{kg0}}{\sqrt{\psi_g}} \quad [3\text{-}9]$$

(4) สำหรับทุก α_g และ ψ_g มีจำนวนอนันต์ของชุดของค่า Threshold และ ค่า Factor loading ที่ทำให้ข้อมูลสอดคล้องกับโมเดล M0 โดยไม่ต้องกำหนดให้ค่า Factor loading หรือค่า Threshold เท่ากันระหว่างกลุ่ม เป้าหมายของวิธี Alignment คือการหาโมเดลทางเลือก M1 ที่ค่าเหมือนกัน (หรือใกล้เคียง) สอดคล้องกับโมเดล M0 ดังนั้น การใช้ชุดขององค์ประกอบ และ Thresholds ก็คล้ายกันระหว่างกลุ่มก็เป็นไปได้ ซึ่งแตกต่างกับกรณี M0 Factor loading และ Thresholds สำหรับโมเดล M1 สามารถแสดงในแบบของ Factor loading M0 และ Thresholds

$$\lambda_{kg1} = \frac{\lambda_{kg0}}{\sqrt{\psi_g}} \quad [3\text{-}10]$$

$$\nu_{kg1} = \nu_{kg0} - \frac{\lambda_{kg0}}{\sqrt{\psi_g}}\alpha_g \quad [3\text{-}11]$$

5. ค่าของ α_g และ ψ_g สำหรับขั้นตอนที่ 4 ถูกคัดเลือกเพื่อลดปริมาณของความแตกต่างระหว่างกลุ่มในค่าพารามิเตอร์รายข้อ (เช่น λ_{kg1} และ ν_{kg1}) ซึ่งจะทำโดยการลดการสูญเสียการทำงานของฟังก์ชัน F

$$F = \sum_{p} \sum_{g1<g2} w_{g1.g2} f(\lambda_{pg1.g1} - \lambda_{pg1.g2}) + \sum_{p} \sum_{g1<g2} w_{g1.g2} f(\nu_{pg1.g1} - \nu_{pg1.g2}) \quad [3\text{-}12]$$

เมื่อ $f(x) = \sqrt{\sqrt{x^2 + 0.0001}}$, $w_{g1.g2} = \sqrt{N_{g1} N_{g2}}$; การถ่วงขนาดน้ำหนักของกลุ่ม

สำหรับทุกคู่ของกลุ่ม (เช่น g_1, g_2) F จะเพิ่มขึ้นจากความแตกต่างในการประมาณค่าพารามิเตอร์ วิธี Alignment ถูกออกแบบเพื่อให้ได้รูปแบบที่มีจำนวนของพารามิเตอร์เพียงเล็กน้อย แสดงให้เห็นถึงความแตกต่างกันในแต่กลุ่มและส่วนใหญ่ค่าพารามิเตอร์จะมีค่าเท่ากันในแต่ละกลุ่ม เพื่อให้มั่นใจว่าการระบุรูปแบบคำสั่งที่ไม่ต้องการใช้ Unwanted command used in TEX: Pleaseremove $\prod^{\psi}{}_{g} = 1$. ค่า α_g สามารถกำหนดให้เป็นอิสระ (Free alignment) หรือคงที่เท่ากับ 0 สำหรับทุกกลุ่ม (Fix alignment)

เมื่อการจัดเรียงรูปแบบเสร็จสมบูรณ์ ค่าพารามิเตอร์ของรายบุคคลจะถูกทดสอบการแปรเปลี่ยน (lack of DIF) ข้ามกลุ่ม การประเมิน DIF จะกระทำโดยใช้วิธีต่อไปนี้:

1. สำหรับพารามิเตอร์ที่กำหนด เช่น i.e., Loading/ Discrimination, Threshold/ Difficulty กับข้อสอบที่ต้องการตรวจสอบเปรียบเทียบค่าประมาณสำหรับคู่ของกลุ่มแต่ละคนด้วยสมมติฐานหลัก กลุ่มที่มีการเชื่อมต่อกับพารามิเตอร์ ถ้า $p > .01$ สำหรับสถิติการทดสอบ
2. การตั้งค่าการเชื่อมต่อที่ใหญ่ที่สุดของกลุ่มจะถูกระบุสำหรับพารามิเตอร์ข้อสอบที่ต้องการตรวจสอบ
3. ค่าเฉลี่ยของพารามิเตอร์ข้อสอบที่ต้องการตรวจสอบที่มีการคำนวณสำหรับการตั้งค่าการเชื่อมต่อที่ระบุไว้ในขั้นตอนที่ 2
4. พารามิเตอร์ข้อสอบที่ต้องการตรวจสอบสำหรับแต่ละกลุ่มจะถูกเปรียบด้วยค่าเฉลี่ยสำหรับค่าการเชื่อมต่อโดยใช้สถิติทดสอบ ถ้า $p > 001$ กลุ่มที่จะถูกเพิ่มในการตั้งค่าการเชื่อมต่อและถ้าไม่มีกลุ่มใดถูกลบออกจากชุดที่เกี่ยวโยงกัน
5. ทำซ้ำขั้นตอนที่ 3 และ 4 จนถึงการเชื่อมโยงไม่เปลี่ยนแปลง
6. กลุ่มคู่ใดคู่หนึ่งที่ไม่ได้อยู่ในชุดที่เกี่ยวโยงกันและอื่น ๆ ที่มีการกล่าวถึง บ่งชี้ว่าข้อสอบที่ต้องการตรวจสอบทำหน้าที่ต่างกัน (DIF) เป็นรายข้อ การประมาณค่าพารามิเตอร์สำหรับ The alignment สามารถประมาณค่าด้วยวิธี Maximum likelihood หรือ วิธี Bayessian ตามกรอบแนวคิดของ Markov Chain Monte Carlo (MCMC) วิธี Bayessian เป็นวิธีการที่ดูเป็นที่นิยมและถูกนำมาใช้ในงานวิจัย การวิเคราะห์ให้ทั้งค่าพารามิเตอร์และการทดสอบความเท่าเทียมโดย

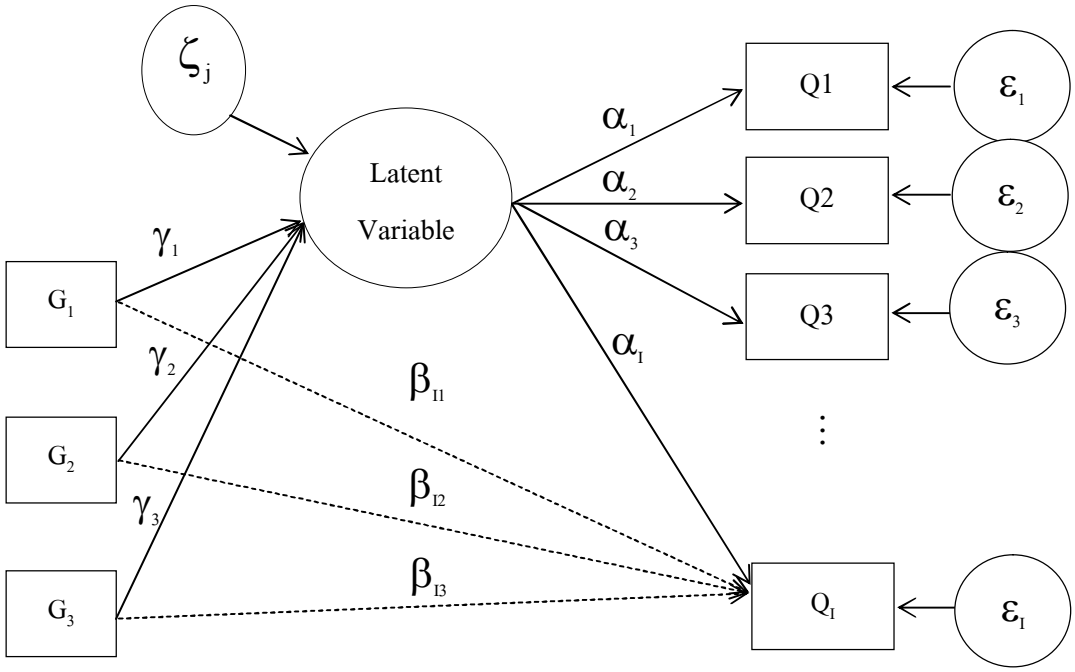
อัตโนมัติ ให้ผลการเปรียบเทียบจากจำนวนของกลุ่มแต่ละกลุ่มด้วย จึงได้ดำเนินตามขั้นตอนสำหรับกลุ่มเปรียบเทียบและกลุ่มอ้างอิง

วิธี MIMIC

การวิเคราะห์ด้วยโมเดลลิสเรลของคุณลักษณะแฝงที่มีหลายสาเหตุและวัดได้จากตัวบ่งชี้หลายตัว (MIMIC) เป็นการนำมาประยุกต์ใช้สำหรับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ในครั้งนี้ ได้ศึกษาทั้งการวิเคราะห์แบบ Uniform DIF และ Nonuniform DIF และได้กำหนดให้ตัวแปรจำแนกกลุ่มเป็นสถานที่ตั้งทางภูมิศาสตร์ และขนาดโรงเรียน ซึ่งดำเนินการวิเคราะห์ข้อมูลตามขั้นตอนหลัก ดังนี้

การวิเคราะห์ โมเดล MIMIC แบบเอกรูป (UUniform DIF)

ขั้นที่ 1 กำหนดโมเดลการวิเคราะห์ดังภาพที่ 3-1



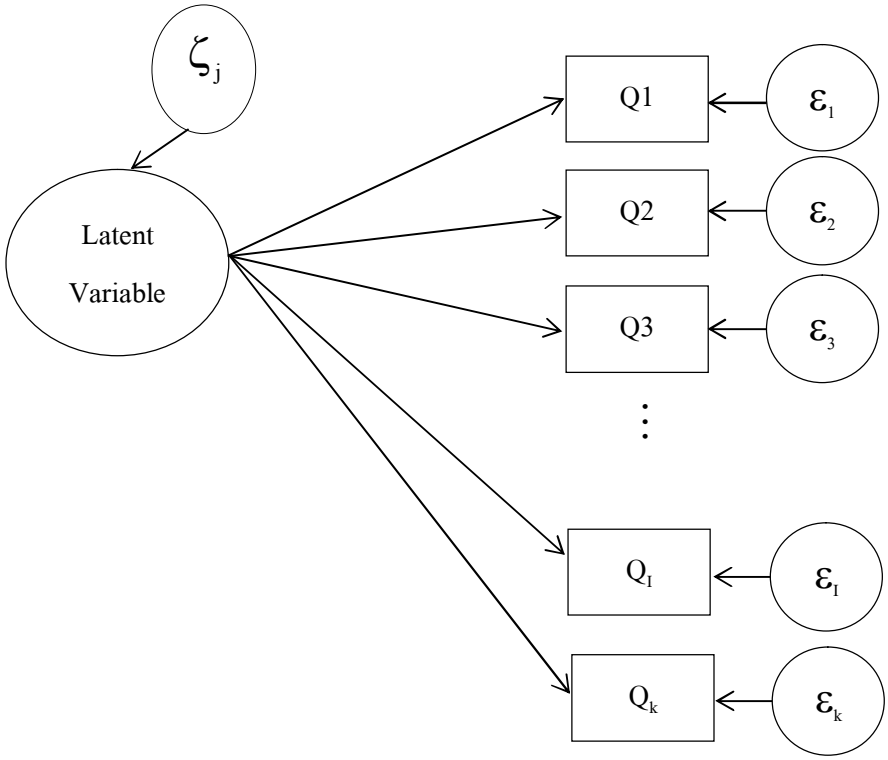
ภาพที่ 3-1 โมเดล MIMIC แบบเอกรูป (Uniform DIF) โดยมีข้อ Q_I เป็นข้อที่ทำหน้าที่ต่างกัน (DIF)

ขั้นที่ 2 การเตรียมไฟล์ข้อมูลสำหรับการวิเคราะห์

ในการวิเคราะห์ส่วนนี้ ผู้วิจัยได้เตรียมไฟล์ให้อยู่ในรูปไฟล์ข้อมูลนามสกุล .csv สำหรับการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ

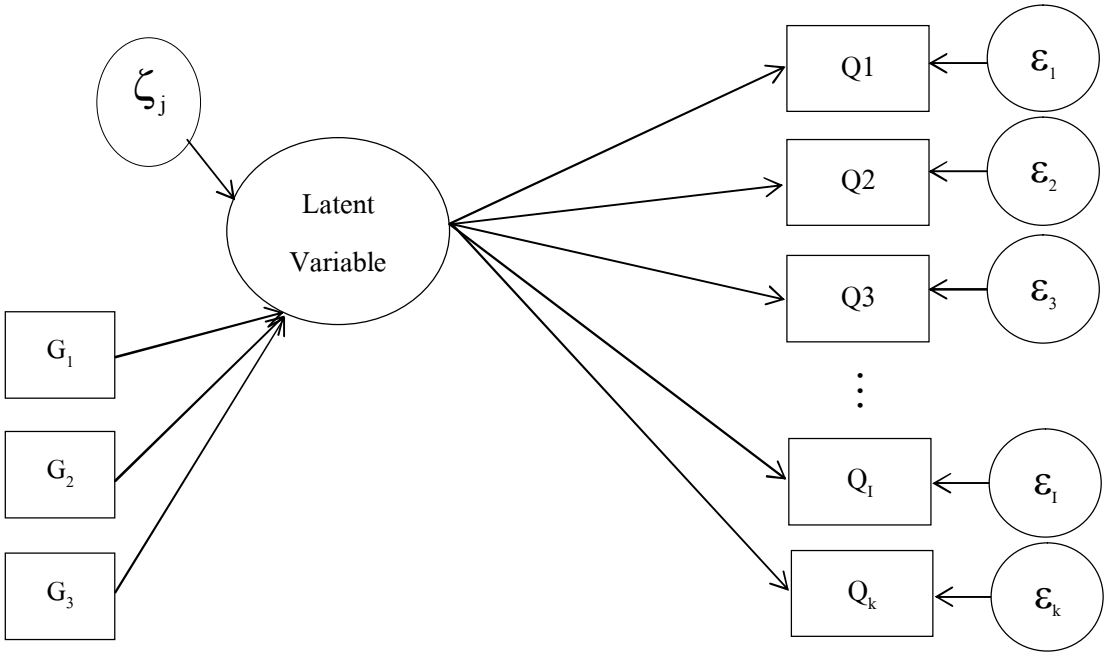
ขั้นที่ 3 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF)

1. วิเคราะห์ภายใต้ทฤษฎีการตอบสนองข้อสอบ (IRT) แบบ 2PL ผู้วิจัยเขียนคำสั่งวิเคราะห์โมเดล MGCFA โดยไม่มีตัวแปรทำนาย ในการวิเคราะห์ข้อมูลตามทฤษฎีการตอบสนองข้อสอบ (IRT) เป็นการศึกษาคุณสมบัติอยู่บนพื้นฐานข้อตกลงความเป็นเอกมิติ (Unidimensional) ของตัวแปรแฝง



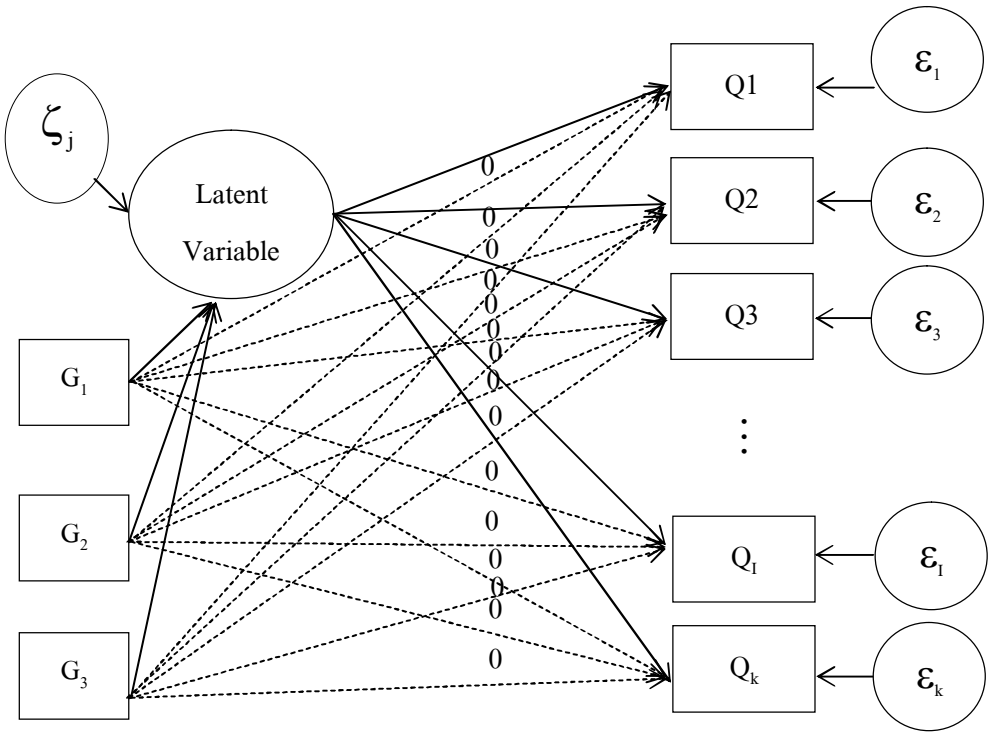
ภาพที่ 3-2 Model CFA ไม่มีตัวแปรทำนาย

2. การเพิ่มตัวแปรทำนาย แต่ไม่มีอิทธิพลตรงต่อข้อสอบ ซึ่งในการวิจัยครั้งนี้ผู้วิจัยได้มีตัวแปรทำนาย คือ สถานที่ตั้งทางภูมิศาสตร์ จำแนกเป็น 4 กลุ่ม และขนาดโรงเรียน จำแนกเป็น 4 กลุ่มเช่นกัน เขียนคำสั่งวิเคราะห์ทีละตัวแปรทำนาย ดังนั้น G คือ REGION หรือ SIZE



ภาพที่ 3-3 เพิ่มตัวแปรทำนาย (G) แต่ไม่มีอิทธิพลตรงต่อข้อสอบ

3. เพิ่มอิทธิพลทางตรงต่อข้อสอบและบังคับค่าให้เป็น 0 เพื่อกำหนดให้ตัวแปรทำนาย (G) ไม่มีผลทางตรงต่อข้อสอบ



ภาพที่ 3-4 เพิ่มอิทธิพลทางตรงของ G ต่อข้อสอบแต่บังคับค่าให้เป็น 0

4. ตรวจดูดัชนีปรับแต่ง (Modification Index: MI)

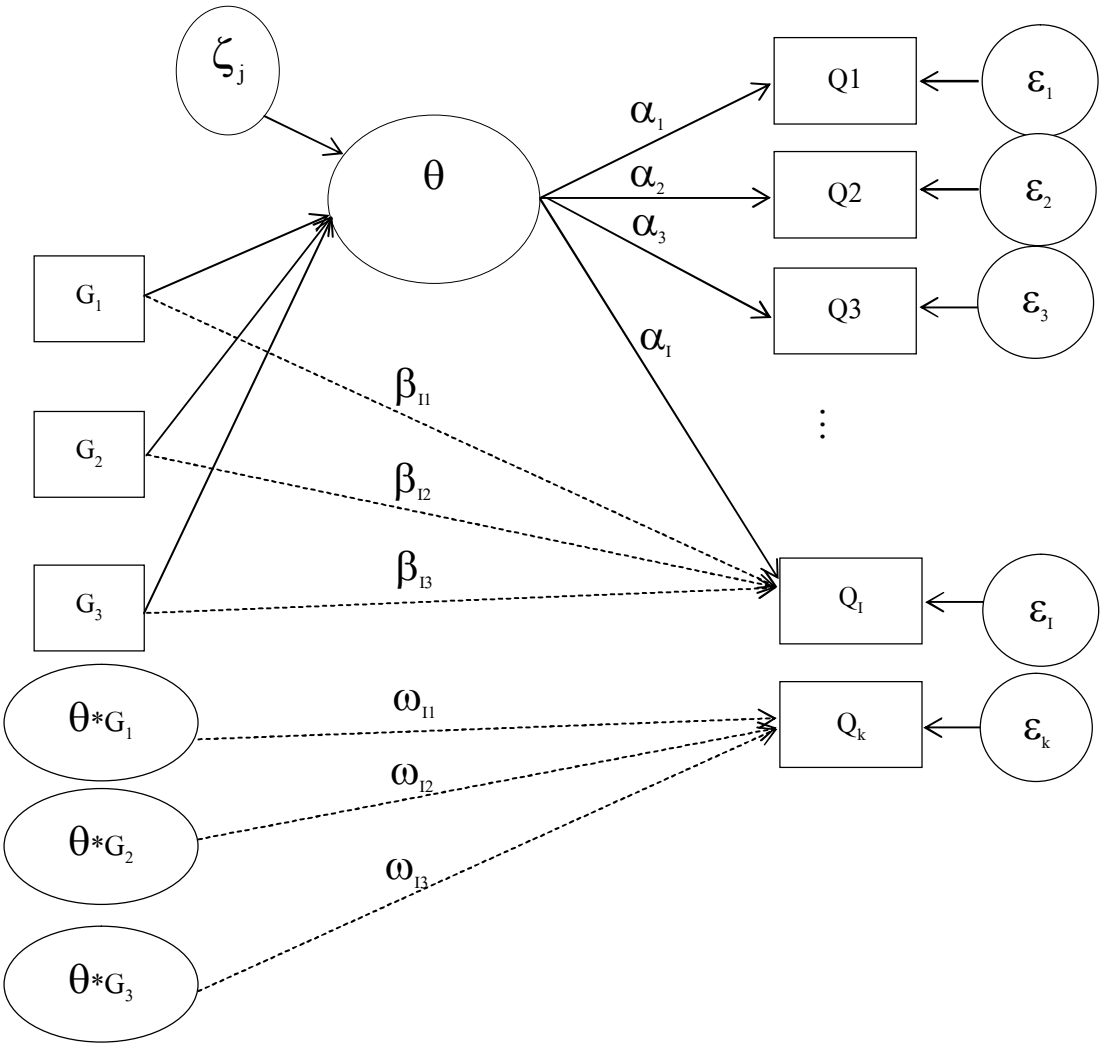
2. เพิ่มอิทธิพลทางตรงจากตัวทำนาย (G) ไปยังข้อสอบในข้อที่มีค่าดัชนีปรับแก้ (MI) เป็นรายข้อ

6. ทำซ้ำในขั้นที่ 4 และ 5 จนไม่มีดัชนีปรับแก้ที่มีนัยสำคัญทางสถิติ

7. ประเมินความกลมกลืนของโมเดล และตรวจผลทางตรงที่มีนัยสำคัญทางสถิติ ถ้าตรวจพบว่า มีผลทางตรง และมีนัยสำคัญทางสถิติ แสดงว่าข้อสอบข้อนั้น คือ ข้อที่เกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF) แบบเอกรูป (Uniform DIF)

การวิเคราะห์ โมเดล MIMIC แบบเอกรูป (Uniform DIF) ร่วมกับแบบอเนกรูป (Nonuniform DIF)

ขั้นที่ 1 กำหนดโมเดลการวิเคราะห์ดังภาพที่ 3-5 ให้



ภาพที่ 3-5 โมเดล MIMIC แบบเอกรูป (Uniform DIF) และแบบอเนกกรูป (Nonuniform DIF)

ขั้นที่ 2 ตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป (Uniform DIF) ว่ามีข้อใดบ้าง

ขั้นที่ 3 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF)

1. เขียนคำสั่งวิเคราะห์ภายใต้ทฤษฎีการตอบสนองข้อสอบ (IRT) แบบ 2PL วิเคราะห์โมเดล MGCFA ในการวิเคราะห์มีคำสั่งตรวจสอบ Uniform DIF และ Nonuniform DIF อยู่ร่วมกัน

2. ตรวจผลทางตรงที่มีนัยสำคัญทางสถิติทั้งแบบ Uniform DIF และ Nonuniform DIF ถ้าตรวจพบว่า มีผลทางตรงและมีนัยสำคัญทางสถิติ แสดงว่าข้อสอบข้อนั้นเป็นข้อที่เกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF)

สถิติที่ใช้ในการวิเคราะห์ข้อมูล

1. การวิเคราะห์ข้อมูลพื้นฐานของกลุ่มตัวอย่างได้แก่ ค่าต่ำสุด (Min) ค่าสูงสุด (Max) ค่าเฉลี่ย (Mean) ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard deviation) ค่าความเบ้ และค่าความโด่ง

2. การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มที่มากกว่าสองกลุ่มใช้วิธี IRT LRT วิธี MGCFA ALIGNMENT และวิธี MIMIC ภายใต้ปัจจัยที่ต่างกัน 2 ปัจจัย ได้แก่ สถานที่ตั้งทางภูมิศาสตร์ และขนาดโรงเรียน โดยใช้โปรแกรมสำเร็จรูป IRTPRO (Cai et al., 2011) และโปรแกรม Mplus 7.11 (Muthén & Muthén, 2007)

การแปลผลค่าสถิติข้อมูลพื้นฐาน

การพิจารณาค่าการแจกแจงของข้อมูล พิจารณาจากค่าความเบ้และความโด่งของของข้อมูล มีเกณฑ์ดังนี้ (Curran, West & Finch, 1997 อ้างถึงใน กัลยา วานิชย์บัญชา, 2556, หน้า 97-99)

1. ถ้า ดัชนี [SI] >3 หมายถึง ข้อมูลไม่สมมาตรหรือมีความเบ้มาก

2. ถ้า ดัชนี [KI] > 10 แสดงว่า ข้อมูลมีปัญหา คือข้อมูลไม่มีการแจกแจงแบบปกติ

เกณฑ์ที่ใช้ในการพิจารณาความสอดคล้องของโมเดล

1. Root mean square error of approximation (RMSEA): ถ้าค่า 0.05 แสดงว่าโมเดลมีความสอดคล้องมาก ยอมรับได้ $.05 \leq RMSEA \leq .05$ ถ้าค่าที่ได้สูงขึ้นถึง 0.08 แสดงว่าเกิดความคลาดเคลื่อนขึ้นในการประมาณค่าประชากร และค่า RMSEA เหมาะที่จะใช้ใน Confirmatory model เมื่อใช้กลุ่มตัวอย่างขนาดใหญ่ตั้งแต่ 500 ตัวอย่างขึ้นไป

2. ดัชนี Comparative fit index (CFI): ดัชนีมีค่าตั้งแต่ 0.90 ขึ้นไป โมเดลมีความสอดคล้อง

3. ดัชนี The tucker-lewis index (TLI) หรือ Non-normed fit index (NNFI): โมเดลมีความสอดคล้องดีที่สุดเมื่อ NNFI มีค่าเท่ากับ 1 ยอมรับได้ $.95 \leq CFI \leq .97$ (สุวิมล ติรกานันท์, 2555, หน้า 243-246)

การวิเคราะห์ความสอดคล้องของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

การวิเคราะห์ความสอดคล้องของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในครั้งใช้คำสั่ง Crosstabs ซึ่งเป็นการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรตั้งแต่ 2 ตัวแปรขึ้นไป ที่มีระดับการวัดของข้อมูลเป็นแบบกลุ่ม (Nominal scale) หรือแบบอันดับ (Ordinal scale) การวิจัยครั้งนี้เป็นการทดสอบความสอดคล้องของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของทั้ง 3 วิธี โดยทดสอบเป็นรายคู่โดยมีสมมติฐานของการทดสอบ ดังนี้

H_0: ผลการตรวจสอบ DIF จากวิธีทั้งสองไม่สอดคล้องกัน

H_1: ผลการตรวจสอบ DIF จากวิธีทั้งสองสอดคล้องกัน

เกณฑ์พิจารณา

1. ใช้สถิติ Chi- square ในการทดสอบ ถ้าค่า Pearson Chi- square มีค่า asymp.sig. (Asympototic significance) ของการทดสอบแบบ 2-sided (2 ข้าง) น้อยกว่า .05 จึงยอมรับ H_1 หรือปฏิเสธ H_0 หมายถึง ผลจากวิธีทั้งสองสอดคล้องกัน

2. ใช้สถิติ Chi- square ในการทดสอบ ถ้าค่า Pearson Chi- square มีค่า asymp.sig. (asympototic significance) ของการทดสอบแบบ 2-sided (2 ข้าง) มากกว่า .05 จึงยอมรับ H_0 หรือปฏิเสธ H_1 หมายถึง ผลจากวิธีทั้งสองไม่สอดคล้องกัน (ยุทธ ไกยวรรณ์, 2553)

3. ถ้าตารางขนาด 2 × 2 และ Expected count < 5 (Eij < 5) เกินกว่า 20% ของ Cell ทั้งหมด ให้ใช้ Fisher's exact test แทน Pearson Chi-square

บทที่ 4
ผลการวิเคราะห์ข้อมูล

การวิจัยครั้งนี้ มีวัตถุประสงค์เพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT วิธี MGCFA-ALIGNMENT และวิธี MIMIC และเพื่อเปรียบเทียบผลของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT วิธี MGCFA-ALIGNMENT และวิธี MIMIC ภายใต้ปัจจัยที่ต่างกัน 2 ปัจจัย ได้แก่ สถานที่ตั้งทางภูมิศาสตร์ และขนาดโรงเรียน

ข้อมูลที่ใช้ในการศึกษาครั้งนี้เป็นข้อมูลผลการสอบรายข้อจากการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ระดับชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 เพื่อนำข้อมูลมาวิเคราะห์หาข้อสรุปตามวัตถุประสงค์ของการวิจัย ข้อมูลที่นำมาศึกษา ได้แก่ ข้อมูลสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน ขนาดของโรงเรียน และคำตอบของนักเรียนรายบุคคลที่ผ่านการตรวจให้คะแนนแบบทวิวิภาค (1-0) วิชาคณิตศาสตร์ วิชาภาษาไทย และวิชาวิทยาศาสตร์ กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2559 จากโรงเรียน สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ แล้วนำข้อมูลมาตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยสามวิธี คือ วิธี IRT LRT วิธี MGCFA ALIGNMENT และวิธี MIMIC สำหรับผลการวิเคราะห์ข้อมูลจำแนกออกเป็น 3 ตอน ดังนี้

ตอนที่ 1 การวิเคราะห์ค่าสถิติพื้นฐาน

ตอนที่ 2 การวิเคราะห์การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT วิธี MGCFA ALIGNMENT และวิธี MIMIC ภายใต้ปัจจัยที่ต่างกัน 2 ปัจจัย ได้แก่ สถานที่ตั้งทางภูมิศาสตร์ และขนาดโรงเรียน

ตอนที่ 3 การเปรียบเทียบผลของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT วิธี MGCFA ALIGNMENT และวิธี MIMIC ภายใต้ปัจจัยที่ต่างกัน 2 ปัจจัย ได้แก่ สถานที่ตั้งทางภูมิศาสตร์ และขนาดโรงเรียน

ในการนำเสนอผลการวิเคราะห์ข้อมูล ผู้วิจัยได้กำหนดสัญลักษณ์และความหมายแทนตัวแปรที่ศึกษาต่าง ๆ ดังนี้

DIF	หมายถึง	การทำหน้าที่ต่างกันของข้อสอบ
IRT LRT	หมายถึง	วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีทดสอบอัตราส่วนความควรจะเป็น

MGCFA-ALIGNMENT หมายถึง วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยันแบบหลายกลุ่ม (Multiple group confirmatory factor analysis alignment)

MIMIC หมายถึง วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวิเคราะห์ข้อมูลด้วยโมเดลลิสเรลของคุณลักษณะแฝงที่มีหลายสาเหตุและวัดได้จากตัวบ่งชี้หลายตัว (Multiple indicators and multiple causes)

$\overline{X}$ หมายถึง ค่าเฉลี่ย (Mean)

SD หมายถึง ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard deviation)

Max หมายถึง ค่าสูงสุด (Maximum)

Min หมายถึง ค่าต่ำสุด (Minimum)

Sk หมายถึง ค่าความเบ้ (Skewness)

Ku หมายถึง ค่าความโด่ง (Kurtosis)

θ หมายถึง ค่าพารามิเตอร์ความสามารถของผู้สอบ (Ability parameter)

χ^2 หมายถึง ค่าสถิติไค-สแควร์ (Chi-square)

df หมายถึง องศาความเป็นอิสระ (Degree of freedom)

RMSEA หมายถึง ดัชนีวัดรากที่สองของความคลาดเคลื่อนในการประมาณค่าค่าพารามิเตอร์ที่เหมาะสม (Root mean square error of approximation)

CFI หมายถึง ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (Comparative fit index)

TLI หมายถึง ดัชนี The tucker lewis index

สัญลักษณ์ที่ใช้แทนตัวแปรแฝง

MATH หมายถึง แบบทดสอบวิชาคณิตศาสตร์

THAI หมายถึง แบบทดสอบวิชาภาษาไทย

SCIENCE หมายถึง แบบทดสอบวิชาวิทยาศาสตร์

สัญลักษณ์ที่ใช้แทนตัวแปรสังเกตได้

Q1-Q25 หมายถึง ข้อสอบวิชาคณิตศาสตร์ข้อที่ 1 ถึงข้อที่ 25

Q1-Q40 หมายถึง ข้อสอบวิชาภาษาไทยและวิชาวิทยาศาสตร์ข้อที่ 1 ถึงข้อที่ 40

สัญลักษณ์ที่ใช้แทนตัวแปรสถานที่ตั้งทางภูมิศาสตร์

REGION	หมายถึง	สถานที่ตั้งทางภูมิศาสตร์
CENTRAL	หมายถึง	ภาคกลาง
NORTH	หมายถึง	ภาคเหนือ
NORTHEST	หมายถึง	ภาคตะวันออกเฉียงเหนือ
SOUTH	หมายถึง	ภาคใต้

สัญลักษณ์ที่ใช้แทนตัวแปรกลุ่มขนาดโรงเรียน

SIZE	หมายถึง	ขนาดโรงเรียน
SMALL	หมายถึง	โรงเรียนขนาดเล็ก
MEDIUM	หมายถึง	โรงเรียนขนาดปานกลาง
LARGE	หมายถึง	โรงเรียนขนาดใหญ่
EXTRA LARGE	หมายถึง	โรงเรียนขนาดใหญ่พิเศษ

สัญลักษณ์ที่ใช้แทนปฏิสัมพันธ์ระหว่างความสามารถ (θ) กับสถานที่ตั้งทางภูมิศาสตร์

MATHNORTH	หมายถึง	ปฏิสัมพันธ์ระหว่าง θ คณิตศาสตร์กับภาคเหนือ
MATHNORTHEST	หมายถึง	ปฏิสัมพันธ์ระหว่าง θ คณิตศาสตร์กับภาคตะวันออกเฉียงเหนือ
MATHSOUTH	หมายถึง	ปฏิสัมพันธ์ระหว่าง θ คณิตศาสตร์กับภาคใต้
THAINORTH	หมายถึง	ปฏิสัมพันธ์ระหว่าง θ ภาษาไทยกับภาคเหนือ
THAINORTHEST	หมายถึง	ปฏิสัมพันธ์ระหว่าง θ ภาษาไทยกับภาคตะวันออกเฉียงเหนือ
THAISOUTH	หมายถึง	ปฏิสัมพันธ์ระหว่าง θ ภาษาไทยกับภาคใต้
SCIENCENORTH	หมายถึง	ปฏิสัมพันธ์ระหว่าง θ วิทยาศาสตร์กับภาคเหนือ
SCIENCENORTHEST	หมายถึง	ปฏิสัมพันธ์ระหว่าง θ วิทยาศาสตร์กับภาคตะวันออกเฉียงเหนือ
SCIENCESOUTH	หมายถึง	ปฏิสัมพันธ์ระหว่าง θ วิทยาศาสตร์กับภาคใต้

สัญลักษณ์ที่ใช้แทนปฏิสัมพันธ์ระหว่างความสามารถ (θ) กับขนาดโรงเรียน

MATHSMALL	หมายถึง	ปฏิสัมพันธ์ระหว่าง θ คณิตศาสตร์กับขนาดเล็ก
MATHMEDIUM	หมายถึง	ปฏิสัมพันธ์ระหว่าง θ คณิตศาสตร์กับขนาดกลาง
MATHLARGE	หมายถึง	ปฏิสัมพันธ์ระหว่าง θ คณิตศาสตร์กับขนาดใหญ่
THAISMALL	หมายถึง	ปฏิสัมพันธ์ระหว่าง θ ภาษาไทยกับขนาดเล็ก

THAIMEDIUM	หมายถึง	ปฏิสัมพันธ์ระหว่าง θ ภาษาไทยกับขนาดกลาง
THAILARGE	หมายถึง	ปฏิสัมพันธ์ระหว่าง θ ภาษาไทยกับขนาดใหญ่
SCIENCESMALL	หมายถึง	ปฏิสัมพันธ์ระหว่าง θ วิทยาศาสตร์กับขนาดเล็ก
SCIENCEMEDIUM	หมายถึง	ปฏิสัมพันธ์ระหว่าง θ วิทยาศาสตร์กับขนาดกลาง
SCIENCELARGE	หมายถึง	ปฏิสัมพันธ์ระหว่าง θ วิทยาศาสตร์กับขนาดใหญ่

ตอนที่ 1 การวิเคราะห์ค่าสถิติพื้นฐาน

ผลการวิเคราะห์ค่าสถิติพื้นฐานของข้อมูลผลการสอบรายข้อจากแบบทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ระดับชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 วิชาคณิตศาสตร์ วิชาภาษาไทย และวิชาวิทยาศาสตร์ ที่เป็นกลุ่มตัวอย่างในการศึกษาครั้งนี้ มีรายละเอียดของผลการวิเคราะห์ดังแสดงในตารางที่ 4-1 ถึง 4-6 ดังนี้

ตารางที่ 4-1 ค่าสถิติพื้นฐานของข้อมูลผลการสอบรายข้อจากแบบทดสอบวิชาคณิตศาสตร์ จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน

สถานที่ตั้งทางภูมิศาสตร์	ค่าสถิติ						
	n	Min	Max	$\bar{X}$	SD	Sk	Ku
กลาง	2000	1	25	8.76	4.58	1.331 (.055)	1.865 (.109)
เหนือ	2000	1	25	9.66	4.48	.741 (.055)	.394 (.109)
ตะวันออกเฉียงเหนือ	2000	1	25	8.42	3.71	.954 (.055)	1.461 (.109)
ใต้	2000	1	25	9.60	5.08	1.161 (.055)	.925 (.109)
รวม	8000	1	25	9.11	4.52	1.125 (.027)	1.323 (.055)

จากตารางที่ 4-1 ผู้สอบวิชาคณิตศาสตร์ทั้งหมดจำนวน 8,000 คน มีคะแนนต่ำสุดเท่ากับ 1 คะแนน และสูงสุดเท่ากับ 25 คะแนน มีค่าเฉลี่ยเท่ากับ 9.11 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 4.52 โค้งการแจกแจงมีความเบ้เท่ากับ 1.125 ค่าความคลาดเคลื่อนมาตรฐานของความเบ้เท่ากับ .027 และค่าความโด่งเท่ากับ 1.323 ค่าความคลาดเคลื่อนมาตรฐานของความโด่งเท่ากับ .055 ซึ่งเมื่อพิจารณาค่าความเบ้และค่าความโด่งกลุ่มตัวอย่างเป็นโค้งปกติ

เมื่อพิจารณาตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนผู้สอบแล้วพบว่า

1. ภาคกลาง ผู้สอบวิชาคณิตศาสตร์มีจำนวน 2,000 คน มีคะแนนต่ำสุดเท่ากับ 1 คะแนน และสูงสุดเท่ากับ 25 คะแนน มีค่าเฉลี่ยเท่ากับ 8.76 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 4.58 โค้งการแจกแจงมีความเบ้เท่ากับ 1.331 ค่าความคลาดเคลื่อนมาตรฐานของความเบ้เท่ากับ .055 และค่าความโด่งเท่ากับ 1.865 ค่าความคลาดเคลื่อนมาตรฐานของความโด่งเท่ากับ .109 ซึ่งเมื่อพิจารณาค่าความเบ้และค่าความโด่งกลุ่มตัวอย่างเป็นโค้งปกติ

2. ภาคเหนือ ผู้สอบวิชาคณิตศาสตร์มีจำนวน 2,000 คน มีคะแนนต่ำสุดเท่ากับ 1 คะแนน และสูงสุดเท่ากับ 25 คะแนน มีค่าเฉลี่ยเท่ากับ 9.66 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 4.48 โค้งการแจกแจงมีความเบ้เท่ากับ .741 ค่าความคลาดเคลื่อนมาตรฐานของความเบ้เท่ากับ .055 และค่าความโด่งเท่ากับ .394 ค่าความคลาดเคลื่อนมาตรฐานของความโด่งเท่ากับ .109 ซึ่งเมื่อพิจารณาค่าความเบ้และค่าความโด่งกลุ่มตัวอย่างเป็นโค้งปกติ

3. ภาคตะวันออกเฉียงเหนือ ผู้สอบวิชาคณิตศาสตร์มีจำนวน 2,000 คน มีคะแนนต่ำสุดเท่ากับ 1 คะแนน และสูงสุดเท่ากับ 25 คะแนน มีค่าเฉลี่ยเท่ากับ 8.42 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 3.71 โค้งการแจกแจงมีความเบ้เท่ากับ .954 ค่าความคลาดเคลื่อนมาตรฐานของความเบ้เท่ากับ .055 และค่าความโด่งเท่ากับ 1.461 ค่าความคลาดเคลื่อนมาตรฐานของความโด่งเท่ากับ .109 ซึ่งเมื่อพิจารณาค่าความเบ้และค่าความโด่งกลุ่มตัวอย่างเป็นโค้งปกติ

4. ภาคใต้ ผู้สอบวิชาคณิตศาสตร์มีจำนวน 2,000 คน มีคะแนนต่ำสุดเท่ากับ 1 คะแนน และสูงสุดเท่ากับ 25 คะแนน มีค่าเฉลี่ยเท่ากับ 9.60 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5.08 โค้งการแจกแจงมีความเบ้เท่ากับ 1.161 ค่าความคลาดเคลื่อนมาตรฐานของความเบ้เท่ากับ .055 และค่าความโด่งเท่ากับ .925 ค่าความคลาดเคลื่อนมาตรฐานของความโด่งเท่ากับ .109 ซึ่งเมื่อพิจารณาค่าความเบ้และค่าความโด่งกลุ่มตัวอย่างเป็นโค้งปกติ

ตารางที่ 4-2 ค่าสถิติพื้นฐานของข้อมูลผลการสอบรายข้อจากแบบทดสอบวิชาภาษาไทย จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน

สถานที่ตั้งทางภูมิศาสตร์	ค่าสถิติ						
	n	Min	Max	$\bar{X}$	SD	Sk	Ku
กลาง	2000	5	33.00	17.84	5.34	.072 (.055)	-.675 (.109)
เหนือ	2000	3	33.00	18.51	5.36	-.064 (.055)	-.538 (.109)
ตะวันออกเฉียงเหนือ	2000	5	34.00	17.64	5.36	.137 (.055)	-.718 (.109)
ใต้	2000	4	34.00	19.10	5.62	.-047.055)	-.601 (.109)
รวม	8000	3	34.00	18.27	5.45	.033 (.027)	-.643 (.055)

จากตารางที่ 4-2 ผู้สอบวิชาภาษาไทยทั้งหมดจำนวน 8,000 คน มีคะแนนต่ำสุดเท่ากับ 3 คะแนน และสูงสุดเท่ากับ 34 คะแนน จากคะแนนเต็ม 40 คะแนน มีค่าเฉลี่ยเท่ากับ 18.27 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5.45 โค้งการแจกแจงมีความเบ้เท่ากับ .033 ค่าความคลาดเคลื่อนมาตรฐานของความเบ้เท่ากับ .027 และค่าความโด่งเท่ากับ -.643 ค่าความคลาดเคลื่อนมาตรฐานของความโด่งเท่ากับ .055 ซึ่งเมื่อพิจารณาค่าความเบ้และค่าความโด่งกลุ่มตัวอย่างเป็นโค้งปกติ

เมื่อพิจารณาตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนผู้สอบแล้วพบว่า

1. ภาคกลาง ผู้สอบวิชาภาษาไทยมีจำนวน 2,000 คน มีคะแนนต่ำสุดเท่ากับ 5 คะแนน และสูงสุดเท่ากับ 33 คะแนน มีค่าเฉลี่ยเท่ากับ 17.84 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5.34 โค้งการแจกแจงมีความเบ้เท่ากับ .072 ค่าความคลาดเคลื่อนมาตรฐานของความเบ้เท่ากับ .055 และค่าความโด่งเท่ากับ -.675 ค่าความคลาดเคลื่อนมาตรฐานของความโด่งเท่ากับ .109 ซึ่งเมื่อพิจารณาค่าความเบ้และค่าความโด่งกลุ่มตัวอย่างเป็นโค้งปกติ

2. ภาคเหนือ ผู้สอบวิชาภาษาไทยมีจำนวน 2,000 คน มีคะแนนต่ำสุดเท่ากับ 3 คะแนน และสูงสุดเท่ากับ 33 คะแนน มีค่าเฉลี่ยเท่ากับ 18.51 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5.36 โค้งการแจกแจงมีความเบ้เท่ากับ -.064 ค่าความคลาดเคลื่อนมาตรฐานของความเบ้เท่ากับ .055 และค่าความโด่งเท่ากับ -.538 ค่าความคลาดเคลื่อนมาตรฐานของความโด่งเท่ากับ .109 ซึ่งเมื่อพิจารณาค่าความเบ้และค่าความโด่งกลุ่มตัวอย่างเป็นโค้งปกติ

3. ภาคตะวันออกเฉียงเหนือ ผู้สอบวิชาภาษาไทยมีจำนวน 2,000 คน มีคะแนนต่ำสุดเท่ากับ 5 คะแนน และสูงสุดเท่ากับ 34 คะแนน มีค่าเฉลี่ยเท่ากับ 17.64 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5.36 โค้งการแจกแจงมีความเบ้เท่ากับ .137 ค่าความคลาดเคลื่อนมาตรฐานของ

ความเบ้เท่ากับ .055 และค่าความโด่งเท่ากับ -.486 ค่าความคลาดเคลื่อนมาตรฐานของความโด่งเท่ากับ -.718 ซึ่งเมื่อพิจารณาค่าความเบ้และค่าความโด่งกลุ่มตัวอย่างเป็นโค้งปกติ

4. ภาคใต้ ผู้สอบวิชาภาษาไทยมีจำนวน 2,000 คน มีคะแนนต่ำสุดเท่ากับ 4 คะแนน และสูงสุดเท่ากับ 34 คะแนน มีค่าเฉลี่ยเท่ากับ 19.10 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5.62 โค้งการแจกแจงมีความเบ้เท่ากับ -.047 ค่าความคลาดเคลื่อนมาตรฐานของความเบ้เท่ากับ .055 และค่าความโด่งเท่ากับ -.618 ค่าความคลาดเคลื่อนมาตรฐานของความโด่งเท่ากับ -.601 ซึ่งเมื่อพิจารณาค่าความเบ้และค่าความโด่งกลุ่มตัวอย่างเป็นโค้งปกติ

ตารางที่ 4-3 ค่าสถิติพื้นฐานของข้อมูลผลการสอบรายข้อจากแบบทดสอบวิชาวิทยาศาสตร์ จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน

สถานที่ตั้งทางภูมิศาสตร์	**ค่าสถิติ**						
	n	**Min**	**Max**	$\bar{X}$	**SD**	**Sk**	**Ku**
กลาง	2000	4	37	13.94	4.72	0.942 (.055)	1.684 (.109)
เหนือ	2000	3	37	14.43	4.49	0.833 (.055)	1.396 (.109)
ตะวันออกเฉียงเหนือ	2000	4	35	12.98	3.92	0.654 (.055)	1.176 (.109)
ใต้	2000	4	37	14.84	5.58	1.118 (.055)	1.387 (.109)
รวม	8000	3	37	14.05	4.77	1.034 (.027)	1.894 (.055)

จากตารางที่ 4-3 ผู้สอบวิชาวิทยาศาสตร์ทั้งหมดจำนวน 8,000 คน มีคะแนนต่ำสุดเท่ากับ 3 คะแนน และสูงสุดเท่ากับ 37 คะแนน จากคะแนนเต็ม 40 คะแนน มีค่าเฉลี่ยเท่ากับ 14.05 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 4.77 โค้งการแจกแจงมีความเบ้เท่ากับ 1.034 ค่าความคลาดเคลื่อนมาตรฐานของความเบ้เท่ากับ .027 และค่าความโด่งเท่ากับ 1.894 ค่าความคลาดเคลื่อนมาตรฐานของความโด่งเท่ากับ .055 ซึ่งเมื่อพิจารณาค่าความเบ้และค่าความโด่งกลุ่มตัวอย่างเป็นโค้งปกติ

เมื่อพิจารณาตามปัจจัยสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนผู้สอบแล้วพบว่า

1. ภาคกลาง ผู้สอบวิชาวิทยาศาสตร์มีจำนวน 2,000 คน มีคะแนนต่ำสุดเท่ากับ 4 คะแนน และสูงสุดเท่ากับ 37 คะแนน มีค่าเฉลี่ยเท่ากับ 13.94 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 4.72 โค้งการแจกแจงมีความเบ้เท่ากับ 0.942 ค่าความคลาดเคลื่อนมาตรฐานของความเบ้เท่ากับ .055 และค่าความโด่งเท่ากับ 1.684 ค่าความคลาดเคลื่อนมาตรฐานของความโด่งเท่ากับ .109 ซึ่งเมื่อพิจารณาค่าความเบ้และค่าความโด่งกลุ่มตัวอย่างเป็นโค้งปกติ

2. ภาคเหนือ ผู้สอบวิชาวิทยาศาสตร์มีจำนวน 2,000 คน มีคะแนนต่ำสุดเท่ากับ 3 คะแนน และสูงสุดเท่ากับ 37 คะแนน มีค่าเฉลี่ยเท่ากับ 14.43 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 4.50 โค้งการแจกแจงมีความเบ้เท่ากับ .833 ค่าความคลาดเคลื่อนมาตรฐานของความเบ้เท่ากับ .055 และค่าความโด่งเท่ากับ 1.396 ค่าความคลาดเคลื่อนมาตรฐานของความโด่งเท่ากับ .109 ซึ่งเมื่อพิจารณาค่าความเบ้และค่าความโด่งกลุ่มตัวอย่างเป็นโค้งปกติ

3. ภาคตะวันออกเฉียงเหนือ ผู้สอบวิชาวิทยาศาสตร์มีจำนวน 2,000 คน มีคะแนนต่ำสุดเท่ากับ 4 คะแนน และสูงสุดเท่ากับ 35 คะแนน มีค่าเฉลี่ยเท่ากับ 12.98 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 3.92 โค้งการแจกแจงมีความเบ้เท่ากับ .654 ค่าความคลาดเคลื่อนมาตรฐานของความเบ้เท่ากับ .055 และค่าความโด่งเท่ากับ 1.176 ค่าความคลาดเคลื่อนมาตรฐานของความโด่งเท่ากับ .109 ซึ่งเมื่อพิจารณาค่าความเบ้และค่าความโด่งกลุ่มตัวอย่างเป็นโค้งปกติ

4. ภาคใต้ ผู้สอบวิชาวิทยาศาสตร์มีจำนวน 2,000 คน มีคะแนนต่ำสุดเท่ากับ 4 คะแนน และสูงสุดเท่ากับ 37 คะแนน มีค่าเฉลี่ยเท่ากับ 14.84 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5.58 โค้งการแจกแจงมีความเบ้เท่ากับ 1.118 ค่าความคลาดเคลื่อนมาตรฐานของความเบ้เท่ากับ .055 และค่าความโด่งเท่ากับ 1.387 ค่าความ คลาดเคลื่อนมาตรฐานของความโด่งเท่ากับ .109 ซึ่งเมื่อพิจารณาค่าความเบ้และค่าความโด่งกลุ่มตัวอย่างยังเป็นโค้งปกติ

ตารางที่ 4-4 ค่าสถิติพื้นฐานของข้อมูลผลการสอบรายข้อจากแบบทดสอบวิชาคณิตศาสตร์ จำแนกตามขนาดโรงเรียน

ขนาดโรงเรียน	ค่าสถิติ						
	n	Min	Max	$\bar{X}$	SD	Sk	Ku
ใหญ่พิเศษ	2000	1	25	11.22	5.43	.649 (.055)	-.405 (.109)
ใหญ่	2000	1	24	8.22	3.53	.896 (.055)	1.298 (.109)
กลาง	2000	1	25	9.70	4.75	.974 (.055)	1.085 (.109)
เล็ก	2000	1	20	7.30	2.90	.737 (.055)	1.050 (.109)
รวม	8000	1	25	9.11	4.52	1.125 (.027)	1.323 (.055)

จากตารางที่ 4-4 ผู้สอบวิชาคณิตศาสตร์ทั้งหมดจำนวน 8,000 คน มีคะแนนต่ำสุดเท่ากับ 1 คะแนน และสูงสุดเท่ากับ 25 คะแนน มีค่าเฉลี่ยเท่ากับ 9.11 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 4.52 โค้งการแจกแจงมีความเบ้เท่ากับ 1.125 ค่าความคลาดเคลื่อนมาตรฐานของความเบ้เท่ากับ .027

และค่าความโด่งเท่ากับ 1.323 ค่าความคลาดเคลื่อนมาตรฐานของความโด่งเท่ากับ .055 ซึ่งเมื่อพิจารณาค่าความเบ้และค่าความโด่งกลุ่มตัวอย่างเป็นโค้งปกติ

เมื่อพิจารณาตามปัจจัยขนาดโรงเรียนของผู้สอบแล้วพบว่า

1. โรงเรียนขนาดใหญ่พิเศษ ผู้สอบวิชาคณิตศาสตร์มีจำนวน 2,000 คน มีคะแนนต่ำสุดเท่ากับ 1 คะแนน และสูงสุดเท่ากับ 25 คะแนน จากคะแนนเต็ม 25 คะแนน มีค่าเฉลี่ยเท่ากับ 11.22 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5.43 โค้งการแจกแจงมีความเบ้เท่ากับ .649 ค่าความคลาดเคลื่อนมาตรฐานของความเบ้เท่ากับ .055 และค่าความโด่งเท่ากับ -.405 ค่าความคลาดเคลื่อนมาตรฐานของความโด่งเท่ากับ .109 ซึ่งเมื่อพิจารณาค่าความเบ้และค่าความโด่งกลุ่มตัวอย่างเป็นโค้งปกติ

2. โรงเรียนขนาดใหญ่ ผู้สอบวิชาคณิตศาสตร์มีจำนวน 2,000 คน มีคะแนนต่ำสุดเท่ากับ 1 คะแนน และสูงสุดเท่ากับ 24 คะแนน มีค่าเฉลี่ยเท่ากับ 8.22 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5.44 โค้งการแจกแจงมีความเบ้เท่ากับ .896 ค่าความคลาดเคลื่อนมาตรฐานของความเบ้เท่ากับ .055 และค่าความโด่งเท่ากับ 1.298 ค่าความคลาดเคลื่อนมาตรฐานของความโด่งเท่ากับ .109 ซึ่งเมื่อพิจารณาค่าความเบ้และค่าความโด่งกลุ่มตัวอย่างเป็นโค้งปกติ

3. โรงเรียนขนาดกลาง ผู้สอบวิชาคณิตศาสตร์มีจำนวน 2,000 คน มีคะแนนต่ำสุดเท่ากับ 1 คะแนน และสูงสุดเท่ากับ 25 คะแนน มีค่าเฉลี่ยเท่ากับ 9.70 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 4.75 โค้งการแจกแจงมีความเบ้เท่ากับ .974 ค่าความคลาดเคลื่อนมาตรฐานของความเบ้เท่ากับ .055 และค่าความโด่งเท่ากับ 1.085 ค่าความคลาดเคลื่อนมาตรฐานของความโด่งเท่ากับ .109 ซึ่งเมื่อพิจารณาค่าความเบ้และค่าความโด่งกลุ่มตัวอย่างเป็นโค้งปกติ

4. โรงเรียนขนาดเล็ก ผู้สอบวิชาคณิตศาสตร์มีจำนวน 2,000 คน มีคะแนนต่ำสุดเท่ากับ 1 คะแนน และสูงสุดเท่ากับ 20 คะแนน มีค่าเฉลี่ยเท่ากับ 7.30 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 2.90 โค้งการแจกแจงมีความเบ้เท่ากับ .729 ค่าความคลาดเคลื่อนมาตรฐานของความเบ้เท่ากับ .055 และค่าความโด่งเท่ากับ 1.059 ค่าความคลาดเคลื่อนมาตรฐานของความโด่งเท่ากับ .109 ซึ่งเมื่อพิจารณาค่าความเบ้และค่าความโด่งกลุ่มตัวอย่างเป็นโค้งปกติ

ตารางที่ 4-5 ค่าสถิติพื้นฐานของข้อมูลผลการสอบรายข้อจากแบบทดสอบวิชาภาษาไทย จำแนกตามขนาดโรงเรียน

ขนาดโรงเรียน	ค่าสถิติ						
	n	Min	Max	$\bar{X}$	SD	Sk	Ku
ใหญ่พิเศษ	2000	4	34	20.61	5.49	-.282 (.055)	-.497 (.109)
ใหญ่	2000	4	33	18.35	5.24	-.028 (.055)	-.629 (.109)
กลาง	2000	3	32	17.33	5.14	.084 (.055)	-.518 (.109)
เล็ก	2000	5	31	16.78	5.12	.184 (.055)	-.612 (.109)
รวม	8000	3	34.00	18.27	5.45	.033 (.027)	-.643 (.055)

จากตารางที่ 4-5 ผู้สอบวิชาภาษาไทยทั้งหมดจำนวน 8,000 คน มีคะแนนต่ำสุดเท่ากับ 0 คะแนน และสูงสุดเท่ากับ 35 คะแนน จากคะแนนเต็ม 40 คะแนน มีค่าเฉลี่ยเท่ากับ 18.11 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5.48 โค้งการแจกแจงมีความเบ้เท่ากับ .071 ค่าความคลาดเคลื่อนมาตรฐานของความเบ้เท่ากับ .027 และค่าความโด่งเท่ากับ -.559 ค่าความคลาดเคลื่อนมาตรฐานของความโด่งเท่ากับ .055 ซึ่งเมื่อพิจารณาค่าความเบ้และค่าความโด่งกลุ่มตัวอย่างเป็นโค้งปกติ

เมื่อพิจารณาตามปัจจัยขนาดโรงเรียนของผู้สอบแล้วพบว่า

1. โรงเรียนขนาดใหญ่พิเศษ ผู้สอบวิชาภาษาไทยมีจำนวน 2,000 คน มีคะแนนต่ำสุดเท่ากับ 3 คะแนน และสูงสุดเท่ากับ 35 คะแนน มีค่าเฉลี่ยเท่ากับ 20.70 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5.46 โค้งการแจกแจงมีความเบ้เท่ากับ -.194 ค่าความคลาดเคลื่อนมาตรฐานของความเบ้เท่ากับ .055 และค่าความโด่งเท่ากับ -.449 ค่าความคลาดเคลื่อนมาตรฐานของความโด่งเท่ากับ .109 ซึ่งเมื่อพิจารณาค่าความเบ้และค่าความโด่งกลุ่มตัวอย่างเป็นโค้งปกติ

2. โรงเรียนขนาดใหญ่ ผู้สอบวิชาภาษาไทยมีจำนวน 2,000 คน มีคะแนนต่ำสุดเท่ากับ 4 คะแนน และสูงสุดเท่ากับ 33 คะแนน มีค่าเฉลี่ยเท่ากับ 18.21 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5.18 โค้งการแจกแจงมีความเบ้เท่ากับ .002 ค่าความคลาดเคลื่อนมาตรฐานของความเบ้เท่ากับ .055 และค่าความโด่งเท่ากับ -.532 ค่าความคลาดเคลื่อนมาตรฐานของความโด่งเท่ากับ .109 ซึ่งเมื่อพิจารณาค่าความเบ้และค่าความโด่งกลุ่มตัวอย่างเป็นโค้งปกติ

3. โรงเรียนขนาดกลาง ผู้สอบวิชาภาษาไทยมีจำนวน 2,000 คน มีคะแนนต่ำสุด 3 คะแนน และสูงสุดเท่ากับ 31 คะแนน มีค่าเฉลี่ยเท่ากับ 16.82 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5.09 โค้งการแจกแจงมีความเบ้เท่ากับ .110 ค่าความคลาดเคลื่อนมาตรฐานของความเบ้เท่ากับ .055

และค่าความโด่งเท่ากับ -.536 ค่าความคลาดเคลื่อนมาตรฐานของความโด่งเท่ากับ .109 ซึ่งเมื่อพิจารณาค่าความเบ้และค่าความโด่งกลุ่มตัวอย่างยังโค้งปกติ

4. โรงเรียนขนาดเล็ก ผู้สอบวิชาภาษาไทยมีจำนวน 2,000 คน มีคะแนนต่ำสุดเท่ากับ 0 คะแนน และสูงสุดเท่ากับ 33 คะแนน มีค่าเฉลี่ยเท่ากับ 16.72 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5.23 โค้งการแจกแจงมีมีความเบ้เท่ากับ .188 ค่าความคลาดเคลื่อนมาตรฐานของความเบ้เท่ากับ .055 และค่าความโด่งเท่ากับ -.529 ค่าความคลาดเคลื่อนมาตรฐานของความโด่งเท่ากับ .109 ซึ่งเมื่อพิจารณาค่าความเบ้และค่าความโด่งกลุ่มตัวอย่างเป็นโค้งปกติ

ตารางที่ 4-6 ค่าสถิติพื้นฐานของข้อมูลผลการสอบรายข้อจากแบบทดสอบวิชาวิทยาศาสตร์ จำแนกตามขนาดโรงเรียน

ขนาดโรงเรียน	ค่าสถิติ						
	n	Min	Max	$\bar{X}$	SD	Sk	Ku
ใหญ่พิเศษ	2000	4	37.00	15.98	5.46	.809 (.055)	.892 (.109)
ใหญ่	2000	3	31.00	13.51	4.01	.647 (.055)	.843 (.109)
กลาง	2000	4	34.00	13.96	5.24	1.147 (.055)	1.788 (.109)
เล็ก	2000	4	27.00	12.74	3.45	.316 (.055)	.131 (.109)
รวม	8000	3	37	14.05	4.77	1.034 (.027)	1.894 (.055)

จากตารางที่ 4-6 ผู้สอบวิชาวิทยาศาสตร์ทั้งหมดจำนวน 8,000 คน มีคะแนนต่ำสุดเท่ากับ 3 คะแนน และสูงสุดเท่ากับ 37 คะแนน จากคะแนนเต็ม 40 คะแนน มีค่าเฉลี่ยเท่ากับ 14.05 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 4.77 โค้งการแจกแจงมีความเบ้เท่ากับ 1.034 ค่าความคลาดเคลื่อนมาตรฐานของความเบ้เท่ากับ .027 และค่าความโด่งเท่ากับ 1.894 ค่าความคลาดเคลื่อนมาตรฐานของความโด่งเท่ากับ .055 ซึ่งเมื่อพิจารณาค่าความเบ้และค่าความโด่งกลุ่มตัวอย่างเป็นโค้งปกติ

เมื่อพิจารณาตามปัจจัยขนาดโรงเรียนของผู้สอบแล้วพบว่า

1. โรงเรียนขนาดใหญ่พิเศษ ผู้สอบวิชาภาษาไทยมีจำนวน 2,000 คน มีคะแนนต่ำสุดเท่ากับ 4 คะแนน และสูงสุดเท่ากับ 37 คะแนน มีค่าเฉลี่ยเท่ากับ 15.98 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5.46 โค้งการแจกแจงมีความเบ้เท่ากับ .809 ค่าความคลาดเคลื่อนมาตรฐานของความเบ้เท่ากับ .055 และค่าความโด่งเท่ากับ .892 ค่าความคลาดเคลื่อนมาตรฐานของความโด่งเท่ากับ .109 ซึ่งเมื่อพิจารณาค่าความเบ้และค่าความโด่งกลุ่มตัวอย่างเป็นโค้งปกติ

2. โรงเรียนขนาดใหญ่ ผู้สอบวิชาวิทยาศาสตร์มีจำนวน 2,000 คน มีคะแนนต่ำสุดเท่ากับ 3 คะแนน และสูงสุดเท่ากับ 31 คะแนน มีค่าเฉลี่ยเท่ากับ 13.51 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 4.01โค้งการแจกแจงมีความเบ้เท่ากับ .647 ค่าความคลาดเคลื่อนมาตรฐานของความเบ้เท่ากับ .055 และค่าความโด่งเท่ากับ .843 ค่าความคลาดเคลื่อนมาตรฐานของความโด่งเท่ากับ .109 ซึ่งเมื่อพิจารณาค่าความเบ้และค่าความโด่งกลุ่มตัวอย่างเป็นโค้งปกติ

3. โรงเรียนขนาดกลาง ผู้สอบวิชาวิทยาศาสตร์มีจำนวน 2,000 คน มีคะแนนต่ำสุด 4 คะแนน และสูงสุดเท่ากับ 34 คะแนน มีค่าเฉลี่ยเท่ากับ 13.96 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5.24 โค้งการแจกแจงมีความเบ้เท่ากับ 1.147 ค่าความคลาดเคลื่อนมาตรฐานของความเบ้เท่ากับ .055 และค่าความโด่งเท่ากับ 1.788 ค่าความคลาดเคลื่อนมาตรฐานของความโด่งเท่ากับ .109 ซึ่งเมื่อพิจารณาค่าความเบ้และค่าความโด่งกลุ่มตัวอย่างเป็นโค้งปกติ

4. โรงเรียนขนาดเล็ก ผู้สอบวิชาวิทยาศาสตร์มีจำนวน 2,000 คน มีคะแนนต่ำสุดเท่ากับ 4คะแนน และสูงสุดเท่ากับ 27 คะแนน มีค่าเฉลี่ยเท่ากับ 12.74 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 3.45 โค้งการแจกแจงมีความเบ้เท่ากับ .316 ค่าความคลาดเคลื่อนมาตรฐานของความเบ้เท่ากับ .055 และค่าความโด่งเท่ากับ .131 ค่าความคลาดเคลื่อนมาตรฐานของความโด่งเท่ากับ .109 ซึ่งเมื่อพิจารณาค่าความเบ้และค่าความโด่งกลุ่มตัวอย่างเป็นโค้งปกติ

ตอนที่ 2 ผลการวิเคราะห์การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT วิธี MGCFA ALIGNMENT และวิธี MIMIC ภายใต้ปัจจัยที่ต่างกัน 2 ปัจจัย ได้แก่ สถานที่ตั้งทางภูมิศาสตร์ และขนาดโรงเรียน

การศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มที่มากกว่าสองกลุ่ม ในครั้งนี้ ศึกษากับผลการสอบวิชาคณิตศาสตร์ จำนวนข้อสอบ 25 ข้อ วิชาภาษาไทย และวิชาวิทยาศาสตร์ จำนวนข้อสอบวิชาละ 40 ข้อ วิธีที่ผู้วิจัยนำมาใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้แก่ วิธี IRT LRT โดยใช้โปรแกรม IRTPRO วิธี MGCFA ALIGNMENT และวิธี MIMIC โดยใช้โปรแกรม Mplus ภายใต้ปัจจัยที่ต่างกัน 2 ปัจจัย ได้แก่ สถานที่ตั้งทางภูมิศาสตร์ และขนาดโรงเรียน ซึ่งมีรายละเอียดของผลการวิเคราะห์ดังต่อไปนี้

การวิเคราะห์การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT โดยประยุกต์ใช้โปรแกรม IRTPRO

1. ผลการวิเคราะห์การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ วิชาภาษาไทย และวิชาวิทยาศาสตร์ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT จำแนกตาม

สถานที่ตั้งทางภูมิศาสตร์ โดยกำหนดให้กลุ่มอ้างอิง (R) เป็นภาคกลาง กลุ่มเปรียบเทียบ (F) เป็นภาคเหนือ ภาคตะวันออกเฉียงเหนือ และภาคใต้ โดยที่ผลการเปรียบเทียบเขียนในรูป Contrast group ดังนี้

Contrast group 1 เปรียบเทียบ ภาคกลางกับภาคเหนือ

Contrast group 2 เปรียบเทียบ ภาคกลางกับภาคตะวันออกเฉียงเหนือ

Contrast group 3 เปรียบเทียบ ภาคกลางกับภาคใต้

รายละเอียดแสดงในตารางที่ 4-7 ถึง ตารางที่ 4-9

ตารางที่ 4-7 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ ด้วยวิธี IRT LRT จำแนกตามสถานที่ตั้งทางภูมิศาสตร์

Item	Contrast groups	Total χ^2	df	p	Item	Contrast groups	Total χ^2	df	p
1	1	0.1	2	0.9279	8	1	3.5	2	0.1722
	2	0.5	2	0.7765		2	0.1	2	0.9665
	3	0.4	2	0.8175		3	5.1	2	0.0784
2	1	0.8	2	0.6809	9	1	3.1	2	0.2104
	2	0.2	2	0.9014		2	0.7	2	0.7043
	3	3.8	2	0.1482		3	0.5	2	0.7723
3	1	6.2*	2	0.0444	10	1	2.1	2	0.3455
	2	8.7*	2	0.0129		2	3.2	2	0.2051
	3	6.1*	2	0.0481		3	5.1	2	0.0777
4	1	6.2*	2	0.0453	11	1	4.4	2	0.1125
	2	15.1*	2	0.0005		2	4.9	2	0.0864
	3	2.4	2	0.2960		3	5.9	2	0.0522
5	1	3.9	2	0.1454	12	1	5.1	2	0.0762
	2	5.9	2	0.0521		2	0.8	2	0.6756
	3	0.2	2	0.9267		3	7.8*	2	0.0200

ตารางที่ 4-7 (ต่อ)

Item	Contrast groups	Total χ^2	df	p	Item	Contrast groups	Total χ^2	df	p
6	1	4.6	2	0.0980	13	1	0.4	2	0.8234
	2	13.2*	2	0.0013		2	0.5	2	0.7879
	3	1.7	2	0.4328		3	2.6	2	0.2716
7	1	2.7	2	0.2615	14	1	4.0	2	0.1332
	2	2.0	2	0.3610		2	2.2	2	0.3307
	3	2.0	2	0.3699		3	1.9	2	0.3943
15	1	0.0	2	0.9956	21	1	2.2	2	0.3281
	2	2.8	2	0.2492		2	0.6	2	0.7473
	3	0.5	2	0.7871		3	3.2	2	0.2048
16	1	1.8	2	0.4103	22	1	9.0*	2	0.0113
	2	3.0	2	0.2248		2	2.5	2	0.2871
	3	3.5	2	0.1730		3	3.9	2	0.1469
17	1	12.2*	2	0.0022	23	1	6.3*	2	0.0421
	2	3.3	2	0.1891		2	5.2	2	0.0741
	3	12.4*	2	0.0020		3	1.3	2	0.5276
18	1	0.5	2	0.7867	24	1	1.9	2	0.3837
	2	3.0	2	0.2241		2	1.2	2	0.5435
	3	1.0	2	0.6022		3	0.9	2	0.6516
19	1	2.1	2	0.3463	25	1	6.7*	2	0.0358
	2	14.4*	2	0.0007		2	0.3	2	0.8590
	3	2.3	2	0.3239		3	11.0*	2	0.0041
20	1	11.1*	2	0.0039					
	2	0.1	2	0.9662					
	3	4.2	2	0.1227					

หมายเหตุ * $p < .05$; ข้อสอบ DIF คือข้อที่อย่างน้อยมี 1 Contrast group ที่มีนัยสำคัญที่ .05

จากตารางที่ 4-7 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ เมื่อพิจารณาดัชนีรวมค่าไคว์-สแควร์ในการทดสอบความแตกต่างของ Contrast groups ข้อใดพบว่า แตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 บ่งชี้ว่าข้อสอบข้อนั้นทำหน้าที่ต่างกัน ผลการวิเคราะห์ พบว่า ข้อสอบที่ทำหน้าที่ต่างกันทั้ง 3 Contrast groups มีจำนวน 1 ข้อ ได้แก่ ข้อที่ 3 ข้อสอบที่ทำหน้าที่ต่างกัน 2 ใน 3 ของ Contrast groups มีจำนวน 3 ข้อ ได้แก่ข้อที่ 4, 17 และ ข้อที่ 25 และข้อสอบที่ทำหน้าที่ต่างกัน 1 ใน 3 ของ Contrast groups มีจำนวน 6 ข้อ ได้แก่ข้อที่ 6, 12, 19, 20, 22 และข้อที่ 23

ตารางที่ 4-8 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาภาษาไทย ด้วยวิธี IRT LRT จำแนกตามสถานที่ตั้งทางภูมิศาสตร์

Item	Contrast groups	Total χ^2	df	p	Item	Contrast groups	Total χ^2	df	p
1	1	15.7*	2	0.0004	9	1	4.0	2	0.1323
	2	10.0*	2	0.0067		2	0.5	2	0.7818
	3	15.7*	2	0.0004		3	3.3	2	0.1980
2	1	15.8*	2	0.0004	10	1	1.0	2	0.6172
	2	3.8	2	0.1545		2	0.5	2	0.7966
	3	13.0*	2	0.0015		3	0.8	2	0.6662
3	1	1.3	2	0.5221	11	1	1.0	2	0.6221
	2	0.1	2	0.9290		2	1.2	2	0.5486
	3	5.5	2	0.0639		3	3.2	2	0.2044
4	1	44.7*	2	0.0001	12	1	0.2	2	0.8955
	2	3.6	2	0.1625		2	0.9	2	0.6355
	3	32.6*	2	0.0001		3	1.9	2	0.3938
5	1	55.6*	2	0.0001	13	1	1.3	2	0.5169
	2	14.1*	2	0.0009		2	1.0	2	0.5977
	3	30.1*	2	0.0001		3	0.3	2	0.8430

ตารางที่ 4-8 (ต่อ)

Item	Contrast groups	Total χ^2	df	p	Item	Contrast groups	Total χ^2	df	p
6	1	2.6	2	0.2757	14	1	2.9	2	0.2371
	2	2.2	2	0.3342		2	0.8	2	0.6717
	3	3.7	2	0.1614		3	1.6	2	0.4593
7	1	2.7	2	0.2628	15	1	8.2*	2	0.0169
	2	3.3	2	0.1903		2	5.3	2	0.0722
	3	0.7	2	0.6882		3	2.9	2	0.2360
8	1	0.8	2	0.6666	16	1	0.3	2	0.8673
	2	0.1	2	0.9694		2	0.0	2	0.9797
	3	3.3	2	0.1966		3	4.6	2	0.1022
17	1	12.5*	2	0.0020	25	1	18.0*	2	0.0001
	2	2.3	2	0.3141		2	2.0	2	0.3653
	3	11.5*	2	0.0032		3	18.6*	2	0.0001
18	1	1.7	2	0.4328	26	1	44.9*	2	0.0001
	2	1.4	2	0.4909		2	2.9	2	0.2344
	3	1.8	2	0.4172		3	11.3*	2	0.0034
19	1	1.7	2	0.4264	27	1	12.3*	2	0.0022
	2	4.3	2	0.1139		2	1.8	2	0.4070
	3	7.3*	2	0.0265		3	18.0*	2	0.0001
20	1	25.6*	2	0.0001	28	1	2.4	2	0.3008
	2	7.3*	2	0.0259		2	6.1*	2	0.0471
	3	31.7*	2	0.0001		3	7.9*	2	0.0189
21	1	43.9*	2	0.0001	29	1	4.1	2	0.1280
	2	6.2*	2	0.0448		2	5.4	2	0.0661
	3	23.4*	2	0.0001		3	3.2	2	0.2043

ตารางที่ 4-8 (ต่อ)

Item	Contrast groups	Total χ^2	df	p	Item	Contrast groups	Total χ^2	df	p
22	1	49.8*	2	0.0001	30	1	24.1*	2	0.0001
	2	5.9	2	0.0533		2	16.0*	2	0.0003
	3	35.9*	2	0.0001		3	31.1*	2	0.0001
23	1	32.5*	2	0.0001	31	1	28.6*	2	0.0001
	2	4.4	2	0.1128		2	2.9	2	0.2386
	3	11.7*	2	0.0028		3	14.9*	2	0.0006
24	1	23.3*	2	0.0001	32	1	39.9*	2	0.0001
	2	3.1	2	0.2158		2	18.5*	2	0.0001
	3	12.1*	2	0.0024		3	31.3*	2	0.0001
33	1	59.0*	2	0.0001	37	1	8.4*	2	0.0149
	2	15.6*	2	0.0004		2	12.0*	2	0.0025
	3	28.3*	2	0.0001		3	13.0*	2	0.0015
34	1	2.0	2	0.3636	38	1	11.5*	2	0.0032
	2	2.9	2	0.2374		2	5.3	2	0.0705
	3	4.7	2	0.0938		3	5.1	2	0.0766
35	1	12.1*	2	0.0023	39	1	9.0*	2	0.0111
	2	0.2	2	0.8831		2	11.5*	2	0.0032
	3	1.2	2	0.5428		3	5.1	2	0.0790
36	1	6.7*	2	0.0344	40	1	6.8*	2	0.0328
	2	1.7	2	0.4293		2	1.4	2	0.4976
	3	3.9	2	0.1410		3	20.5*	2	0.0001

หมายเหตุ * $p < .05$; ข้อสอบ DIF คือข้อที่อย่างน้อยมี 1 Contrast group ที่มีนัยสำคัญที่ .05

จากตารางที่ 4-8 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาภาษาไทย สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ เมื่อพิจารณา

ดัชนีรวมค่าไควร์-สแควร์ในการทดสอบความแตกต่างของ Contrast groups ข้อใดพบว่า แตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 บ่งชี้ว่าข้อสอบข้อนั้นทำหน้าที่ต่างกัน ผลการวิเคราะห์ พบว่าข้อสอบที่ทำหน้าที่ต่างกันทั้ง 3 Contrast groups มีจำนวน 8 ข้อ ได้แก่ ข้อที่ 1, 5, 20, 21, 30, 32, 33 และข้อที่ 37 ข้อสอบที่ทำหน้าที่ต่างกัน 2 ใน 3 ของ Contrast groups มีจำนวน 12 ข้อ ได้แก่ ข้อที่ 2, 4, 15, 22, 23, 24, 25, 26, 27, 28, 31, 39 และข้อที่ 40 และข้อสอบที่ทำหน้าที่ต่างกัน 1 ใน 3 ของ Contrast groups มีจำนวน 4 ข้อ ได้แก่ข้อที่ 19, 35, 36 และข้อที่ 38

ตารางที่ 4-9 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาวิทยาศาสตร์ ด้วยวิธี IRT LRT จำแนกตามสถานที่ตั้งทางภูมิศาสตร์

Item	Contrast groups	Total χ^2	df	p	Item	Contrast groups	Total χ^2	df	p
1	1	1.7	2	0.4282	8	1	1.3	2	0.5125
	2	1.6	2	0.4470		2	2.7	2	0.2607
	3	0.7	2	0.7064		3	0.6	2	0.7427
2	1	0.3	2	0.8646	9	1	0.4	2	0.8214
	2	3.4	2	0.1855		2	3.3	2	0.1890
	3	2.7	2	0.2607		3	1.0	2	0.6071
3	1	7.2*	2	0.0273	10	1	1.6	2	0.4415
	2	0.8	2	0.6828		2	10.1*	2	0.0065
	3	1.8	2	0.4037		3	0.1	2	0.9443
4	1	1.8	2	0.4149	11	1	6.6*	2	0.0363
	2	5.8	2	0.0547		2	2.7	2	0.2650
	3	6.7*	2	0.0342		3	3.7	2	0.1597
5	1	5.0	2	0.0823	12	1	2.7	2	0.2580
	2	1.5	2	0.4749		2	14.1*	2	0.0009
	3	2.0	2	0.3754		3	1.5	2	0.4756

ตารางที่ 4-9 (ต่อ)

Item	Contrast groups	Total χ^2	df	p	Item	Contrast groups	Total χ^2	df	p
6	1	8.2*	2	0.0163	13	1	0.0	2	0.9933
	2	0.4	2	0.8395		2	0.1	2	0.9603
	3	4.6	2	0.1008		3	5.6	2	0.0608
7	1	0.5	2	0.7744	14	1	0.5	2	0.7861
	2	1.7	2	0.4272		2	0.7	2	0.7024
	3	0.3	2	0.8423		3	0.3	2	0.8420
15	1	0.8	2	0.6610	22	1	3.0	2	0.2255
	2	0.2	2	0.9088		2	2.7	2	0.2584
	3	0.2	2	0.8984		3	6.8*	2	0.0329
16	1	0.4	2	0.7995	23	1	8.9*	2	0.0117
	2	2.5	2	0.2807		2	29.4*	2	0.0001
	3	6.4*	2	0.0410		3	5.9	2	0.0519
17	1	4.2	2	0.1235	24	1	1.3	2	0.5241
	2	8.4*	2	0.0152		2	7.3*	2	0.0266
	3	0.6	2	0.7553		3	2.5	2	0.2919
18	1	0.9	2	0.6279	25	1	4.8	2	0.0902
	2	4.4	2	0.1093		2	2.1	2	0.3445
	3	12.7*	2	0.0017		3	13.1*	2	0.0014
19	1	4.6	2	0.0990	26	1	1.0	2	0.6132
	2	0.9	2	0.6432		2	2.4	2	0.3032
	3	2.4	2	0.2989		3	10.7*	2	0.0048

ตารางที่ 4-9 (ต่อ)

Item	Contrast groups	Total χ^2	df	p	Item	Contrast groups	Total χ^2	df	p
20	1	3.9	2	0.1410	27	1	12.8*	2	0.0017
	2	2.4	2	0.3010		2	14.5*	2	0.0007
	3	8.5*	2	0.0144		3	3.6	2	0.1648
21	1	0.8	2	0.6619	28	1	3.6	2	0.1686
	2	0.0	2	0.9840		2	5.7	2	0.0574
	3	0.2	2	0.9265		3	9.3*	2	0.0097
29	1	2.4	2	0.3088	35	1	1.8	2	0.4117
	2	3.2	2	0.2057		2	5.7	2	0.0578
	3	0.1	2	0.9469		3	14.5*	2	0.0007
30	1	2.5	2	0.2871	36	1	5.6	2	0.0604
	2	0.7	2	0.7224		2	9.8*	2	0.0074
	3	1.8	2	0.4080		3	2.4	2	0.3049
31	1	4.1	2	0.1318	37	1	6.2*	2	0.0450
	2	0.8	2	0.6703		2	4.0	2	0.1373
	3	0.1	2	0.9420		3	8.4*	2	0.0150
32	1	1.0	2	0.6103	38	1	2.2	2	0.3279
	2	5.3	2	0.0722		2	1.8	2	0.4024
	3	4.6	2	0.0994		3	9.1*	2	0.0108
33	1	1.1	2	0.5783	39	1	0.8	2	0.6868
	2	1.4	2	0.4978		2	1.2	2	0.5385
	3	3.5	2	0.1753		3	0.5	2	0.7838
34	1	13.8*	2	0.0010	40	1	1.5	2	0.4706
	2	12.6*	2	0.0018		2	1.3	2	0.5224
	3	6.5*	2	0.0389		3	2.5	2	0.2896

หมายเหตุ * $p < .05$; ข้อสอบ DIF คือข้อที่อย่างน้อยมี 1 Contrast group ที่มีนัยสำคัญที่ .05

จากตารางที่ 4-9 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาวิทยาศาสตร์ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ เมื่อพิจารณาดัชนีรวมค่าไควร์-สแควร์ในการทดสอบความแตกต่างของ Contrast groups ข้อใดพบว่า แตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 บ่งชี้ว่าข้อสอบข้อนั้นทำหน้าที่ต่างกัน ผลการวิเคราะห์ พบว่า ข้อสอบที่ทำหน้าที่ต่างกันทั้ง 3 Contrast groups มีจำนวน 2 ข้อ ได้แก่ ข้อ 23 และข้อที่ 34 ข้อสอบที่ทำหน้าที่ต่างกัน 2 ใน 3 ของ Contrast groups มีจำนวน 3 ข้อ ได้แก่ข้อที่ 4, 27 และ ข้อที่ 37 และข้อสอบที่ทำหน้าที่ต่างกัน 1 ใน 3 ของ Contrast groups มีจำนวน 17 ข้อ ได้แก่ข้อที่ 3, 6, 10, 11, 12, 16, 17, 18, 20, 22, 24, 25, 26, 28, 35, 36 และ ข้อที่ 38

2. ผลการวิเคราะห์การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ วิชาภาษาไทย และวิชาวิทยาศาสตร์ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT จำแนกตามขนาดโรงเรียน โดยกำหนดให้กลุ่มอ้างอิง (R) เป็นโรงเรียนขนาดใหญ่พิเศษ กลุ่มเปรียบเทียบ (F) เป็นโรงเรียนขนาดใหญ่ ขนาดกลาง และขนาดเล็ก โดยที่ผลการเปรียบเทียบเขียนในรูป Contrast group ดังนี้

Contrast group 1 เปรียบเทียบ ขนาดใหญ่พิเศษกับขนาดเล็ก

Contrast group 2 เปรียบเทียบ ขนาดใหญ่พิเศษกับขนาดกลาง

Contrast group 3 เปรียบเทียบ ขนาดใหญ่พิเศษกับขนาดใหญ่

รายละเอียดแสดงในตารางที่ 4-10 ถึง ตารางที่ 4-12 ดังนี้

ตารางที่ 4-10 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ ด้วยวิธี IRT LRT จำแนกตามขนาดโรงเรียน

Item	Contrast groups	Total χ^2	df	p	Item	Contrast groups	Total χ^2	df	p
1	1	1.3	2	0.5205	8	1	14.1*	2	0.0009
	2	8.9*	2	0.0118		2	16.1*	2	0.0003
	3	4.6	2	0.0983		3	7.0*	2	0.0296
2	1	6.6*	2	0.0377	9	1	13.6*	2	0.0011
	2	0.1	2	0.9667		2	6.1*	2	0.0484
	3	0.9	2	0.6511		3	3.4	2	0.1867
3	1	38.7*	2	0.0001	10	1	6.1*	2	0.0470
	2	2.3	2	0.3145		2	9.9*	2	0.0071
	3	6.7*	2	0.0342		3	2.7	2	0.2613
4	1	46.8*	2	0.0001	11	1	12.2*	2	0.0022
	2	12.5*	2	0.0020		2	1.0	2	0.6164
	3	4.4	2	0.1083		3	5.8	2	0.0539
5	1	33.5*	2	0.0001	12	1	11.8*	2	0.0027
	2	11.6*	2	0.0030		2	2.3	2	0.3234
	3	18.0*	2	0.0001		3	3.4	2	0.1833
6	1	11.8*	2	0.0027	13	1	2.5	2	0.2898
	2	4.3	2	0.1154		2	5.1	2	0.0779
	3	1.3	2	0.5182		3	0.2	2	0.9249
7	1	9.6*	2	0.0084	14	1	1.1	2	0.5758
	2	2.9	2	0.2409		2	3.9	2	0.1460
	3	0.1	2	0.9718		3	0.1	2	0.9502

ตารางที่ 4-10 (ต่อ)

Item	Contrast groups	Total χ^2	df	p	Item	Contrast groups	Total χ^2	df	p
15	1	2.3	2	0.3243	21	1	3.7	2	0.1562
	2	5.3	2	0.0704		2	6.6*	2	0.0363
	3	2.6	2	0.2751		3	2.2	2	0.3319
16	1	2.5	2	0.2817	22	1	5.8	2	0.0552
	2	1.9	2	0.3950		2	3.4	2	0.1803
	3	2.2	2	0.3297		3	5.2	2	0.0739
17	1	14.4*	2	0.0007	23	1	4.4	2	0.1126
	2	8.2*	2	0.0169		2	0.5	2	0.7894
	3	10.9*	2	0.0042		3	11.0*	2	0.0041
18	1	35.1*	2	0.0001	24	1	6.8*	2	0.0336
	2	7.5*	2	0.0237		2	3.6	2	0.1689
	3	15.1*	2	0.0005		3	1.6	2	0.4430
19	1	37.0*	2	0.0001	25	1	0.3	2	0.8495
	2	5.9	2	0.0528		2	2.0	2	0.3678
	3	18.1*	2	0.0001		3	3.8	2	0.1515
20	1	5.7	2	0.0578					
	2	4.7	2	0.0966					
	3	4.0	2	0.1344					

หมายเหตุ * $p < .05$; ข้อสอบ DIF คือข้อที่อย่างน้อยมี 1 Contrast group ที่มีนัยสำคัญที่ .05

จากตารางที่ 4-10 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT จำแนกตามขนาดโรงเรียน เมื่อพิจารณาดัชนีรวมค่าไควร์-สแควร์ในการทดสอบความแตกต่างของ Contrast groups ข้อใดพบว่า แตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 บ่งชี้ว่าข้อสอบข้อนั้นทำหน้าที่ต่างกัน ผลการวิเคราะห์ พบว่า ข้อสอบที่ทำหน้าที่ต่างกันทั้ง 3 Contrast groups มีจำนวน 4 ข้อ ได้แก่ 5, 8, 17 และข้อที่ 18 ข้อสอบที่ทำหน้าที่ต่างกัน 2 ใน 3 ของ Contrast groups มีจำนวน 5 ข้อ ได้แก่ ข้อที่ 3, 4, 9, 10 และ ข้อที่ 19 และข้อสอบที่ทำหน้าที่ต่างกัน 1 ใน 3 ของ Contrast groups มีจำนวน 9 ข้อ ได้แก่ ข้อที่ 1, 2, 6, 7, 11, 12, 21, 23 และ ข้อที่ 24

ตารางที่ 4-11 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาภาษาไทย ด้วยวิธี IRT LRT จำแนกตามขนาดโรงเรียน

Item	**Contrast groups**	**Total χ^2**	**df**	**p**	**Item**	**Contrast groups**	**Total χ^2**	**df**	**p**
1	1	3.3	2	0.1970	5	1	0.9	2	0.6393
	2	15.9*	2	0.0004		2	23.1*	2	0.0001
	3	74.6*	2	0.0001		3	125.9*	2	0.0001
2	1	3.0	2	0.2236	6	1	2.2	2	0.3326
	2	1.0	2	0.6062		2	7.5*	2	0.0239
	3	43.4*	2	0.0001		3	33.0*	2	0.0001
3	1	8.8*	2	0.0122	7	1	3.5	2	0.1776
	2	2.1	2	0.3452		2	4.2	2	0.1219
	3	39.9*	2	0.0001		3	7.6*	2	0.0222
4	1	2.4	2	0.3014	8	1	1.1	2	0.5780
	2	7.9*	2	0.0196		2	0.6	2	0.7510
	3	50.1*	2	0.0001		3	3.0	2	0.2209

ตารางที่ 4-11 (ต่อ)

Item	Contrast groups	Total χ^2	df	p	Item	Contrast groups	Total χ^2	df	p
9	1	3.7	2	0.1587	16	1	0.8	2	0.6814
	2	2.7	2	0.2653		2	2.8	2	0.2523
	3	18.2*	2	0.0001		3	1.7	2	0.4345
10	1	4.0	2	0.1398	17	1	1.8	2	0.3981
	2	1.9	2	0.3816		2	1.5	2	0.4734
	3	6.9*	2	0.0319		3	44.5*	2	0.0001
11	1	0.3	2	0.8432	18	1	2.8	2	0.2478
	2	1.7	2	0.4310		2	3.0	2	0.2197
	3	0.7	2	0.7222		3	6.3*	2	0.0429
12	1	2.2	2	0.3404	19	1	0.5	2	0.7670
	2	3.4	2	0.1881		2	1.3	2	0.5335
	3	13.3*	2	0.0013		3	26.4*	2	0.0001
13	1	1.3	2	0.5247	20	1	3.0	2	0.2255
	2	0.7	2	0.7064		2	12.7*	2	0.0018
	3	4.5	2	0.1040		3	90.1*	2	0.0001
14	1	2.1	2	0.3556	21	1	5.1	2	0.0792
	2	1.5	2	0.4773		2	10.6*	2	0.0051
	3	6.9*	2	0.0320		3	87.9*	2	0.0001
15	1	0.4	2	0.8295	22	1	1.1	2	0.5919
	2	2.7	2	0.2593		2	10.2*	2	0.0060
	3	6.7*	2	0.0347		3	74.5*	2	0.0001

ตารางที่ 4-11 (ต่อ)

Item	Contrast groups	Total χ^2	df	p	Item	Contrast groups	Total χ^2	df	p
23	1	0.8	2	0.6685	30	1	0.1	2	0.9696
	2	16.4*	2	0.0003		2	9.4*	2	0.0090
	3	29.0*	2	0.0001		3	78.7*	2	0.0001
24	1	2.3	2	0.3193	31	1	2.9	2	0.2320
	2	1.1	2	0.5787		2	12.6*	2	0.0018
	3	13.2*	2	0.0013		3	57.6*	2	0.0001
25	1	1.2	2	0.5615	32	1	0.8	2	0.6631
	2	19.4*	2	0.0001		2	10.9*	2	0.0043
	3	122.8*	2	0.0001		3	68.6*	2	0.0001
26	1	1.9	2	0.3955	33	1	0.4	2	0.8309
	2	0.6	2	0.7584		2	25.8*	2	0.0001
	3	14.9*	2	0.0006		3	83.8*	2	0.0001
27	1	0.2	2	0.9103	34	1	0.5	2	0.7933
	2	9.0*	2	0.0113		2	1.3	2	0.5131
	3	28.7*	2	0.0001		3	0.1	2	0.9336
28	1	0.3	2	0.8731	35	1	1.3	2	0.5222
	2	1.3	2	0.5120		2	3.2	2	0.2042
	3	6.6*	2	0.0363		3	29.6*	2	0.0001
29	1	2.7	2	0.2572	36	1	1.1	2	0.5792
	2	7.0*	2	0.0301		2	1.5	2	0.4666
	3	2.4	2	0.3013		3	1.6	2	0.4584

ตารางที่ 4-11 (ต่อ)

Item	Contrast groups	Total χ^2	df	p	Item	Contrast groups	Total χ^2	df	p
37	1	1.1	2	0.5914	39	1	6.5*	2	0.0397
	2	9.1*	2	0.0104		2	1.9	2	0.3802
	3	12.8*	2	0.0017		3	28.2*	2	0.0001
38	1	0.2	2	0.8973	40	1	1.0	2	0.6171
	2	4.9	2	0.0843		2	5.4	2	0.0657
	3	19.3*	2	0.0001		3	50.1*	2	0.0001

หมายเหตุ * $p < .05$; ข้อสอบ DIF คือข้อที่อย่างน้อยมี 1 Contrast group ที่มีนัยสำคัญที่ .05

จากตารางที่ 4-11 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาวิชาภาษาไทย สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT จำแนกตามขนาดโรงเรียน เมื่อพิจารณาดัชนีรวมค่าไคว์-สแควร์ในการทดสอบความแตกต่างของ Contrast groups ข้อใดพบว่า แตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 บ่งชี้ว่าข้อสอบข้อนั้นทำหน้าที่ต่างกัน ผลการวิเคราะห์ ไม่พบข้อสอบที่ทำหน้าที่ต่างกันทั้ง 3 Contrast groups ข้อสอบที่ทำหน้าที่ต่างกัน 2 ใน 3 ของ Contrast groups มีจำนวน 17 ข้อ ได้แก่ข้อที่ 1, 3, 4, 5, 6, 20, 21, 22, 23, 25, 27, 30, 31, 32, 33, 37 และ ข้อที่ 39 และข้อสอบที่ทำหน้าที่ต่างกัน 1 ใน 3 ของ Contrast groups มีจำนวน 16 ข้อ ได้แก่ ข้อที่ 2, 7, 9, 10, 12, 15, 17, 18, 19, 24, 26, 28, 29, 35, 38 และ ข้อที่ 40

ตารางที่ 4-12 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาวิทยาศาสตร์ ด้วยวิธี IRT LRT จำแนกตามขนาดโรงเรียน

Item	Contrast groups	Total χ^2	df	p	Item	Contrast groups	Total χ^2	df	p
1	1	12.4	2	0.0020	8	1	0.3	2	0.8495
	2	10.7	2	0.0047		2	13.4	2	0.0012
	3	9.3	2	0.0094		3	4.7	2	0.0943
2	1	3.4	2	0.1798	9	1	1.1	2	0.5779
	2	3.1	2	0.2173		2	0.1	2	0.9307
	3	1.6	2	0.4553		3	0.2	2	0.9126
3	1	16.1	2	0.0003	10	1	5.1	2	0.0785
	2	10.1	2	0.0065		2	0.5	2	0.7863
	3	1.3	2	0.5120		3	0.8	2	0.6586
4	1	1.6	2	0.4460	11	1	0.2	2	0.9031
	2	5.0	2	0.0819		2	6.9	2	0.0317
	3	7.7	2	0.0208		3	3.4	2	0.1807
5	1	8.0	2	0.0185	12	1	12.5	2	0.0019
	2	0.2	2	0.8853		2	1.2	2	0.5582
	3	0.7	2	0.7013		3	3.2	2	0.2064
6	1	1.8	2	0.4054	13	1	5.2	2	0.0743
	2	4.8	2	0.0902		2	4.9	2	0.0855
	3	2.1	2	0.3437		3	3.6	2	0.1705
7	1	4.6	2	0.1025	14	1	0.8	2	0.6743
	2	2.8	2	0.2471		2	0.3	2	0.8673
	3	0.0	2	0.9920		3	0.9	2	0.6369

ตารางที่ 4-12 (ต่อ)

Item	Contrast groups	Total χ^2	df	p	Item	Contrast groups	Total χ^2	df	p
15	1	4.2	2	0.1206	22	1	8.5	2	0.0145
	2	4.9	2	0.0865		2	2.6	2	0.2801
	3	5.0	2	0.0831		3	31.2	2	0.0001
16	1	5.9	2	0.0531	23	1	35.6	2	0.0001
	2	7.8	2	0.0199		2	12.1	2	0.0024
	3	5.9	2	0.0532		3	15.7	2	0.0004
17	1	16.9	2	0.0002	24	1	12.0	2	0.0024
	2	0.7	2	0.6970		2	6.6	2	0.0363
	3	2.2	2	0.3321		3	16.3	2	0.0003
18	1	10.1	2	0.0065	25	1	2.2	2	0.3344
	2	6.6	2	0.0371		2	0.1	2	0.9392
	3	4.5	2	0.1069		3	4.1	2	0.1279
19	1	3.4	2	0.1825	26	1	29.3	2	0.0001
	2	7.0	2	0.0306		2	7.8	2	0.0199
	3	3.7	2	0.1572		3	25.2	2	0.0001
20	1	4.1	2	0.1282	27	1	26.4	2	0.0001
	2	6.1	2	0.0479		2	2.8	2	0.2443
	3	2.8	2	0.2494		3	5.6	2	0.0596
21	1	0.1	2	0.9391	28	1	22.7	2	0.0001
	2	3.4	2	0.1856		2	8.4	2	0.0148
	3	1.8	2	0.3990		3	19.5	2	0.0001

ตารางที่ 4-12 (ต่อ)

Item	Contrast groups	Total χ^2	df	p	Item	Contrast groups	Total χ^2	df	p
29	1	3.6	2	0.1661	35	1	43.3	2	0.0001
	2	3.1	2	0.2081		2	8.1	2	0.0176
	3	1.6	2	0.4418		3	19.1	2	0.0001
30	1	11.2	2	0.0036	36	1	8.3	2	0.0160
	2	6.6	2	0.0369		2	5.5	2	0.0625
	3	4.7	2	0.0941		3	4.6	2	0.0994
31	1	6.8	2	0.0333	37	1	0.2	2	0.9021
	2	3.3	2	0.1903		2	0.3	2	0.8548
	3	4.4	2	0.1083		3	0.2	2	0.8840
32	1	3.7	2	0.1558	38	1	15.5	2	0.0004
	2	4.5	2	0.1058		2	8.2	2	0.0164
	3	0.3	2	0.8501		3	2.8	2	0.2465
33	1	2.9	2	0.2316	39	1	1.2	2	0.5420
	2	7.8	2	0.0198		2	0.1	2	0.9401
	3	4.5	2	0.1072		3	3.7	2	0.1609
34	1	17.8	2	0.0001	40	1	5.3	2	0.0708
	2	7.4	2	0.0250		2	1.8	2	0.4038
	3	5.8	2	0.0557		3	0.1	2	0.9373

หมายเหตุ * $p < .05$; ข้อสอบ DIF คือข้อที่อย่างน้อยมี 1 Contrast group ที่มีนัยสำคัญที่ .05

จากตารางที่ 4-12 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาวิชาวิทยาศาสตร์ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT จำแนกตามขนาดโรงเรียน เมื่อพิจารณาดัชนีรวมค่าไคว์-สแควร์ในการทดสอบความแตกต่างของ Contrast groups ข้อใดพบว่าแตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 บ่งชี้ว่าข้อสอบข้อนั้นทำหน้าที่ต่างกัน

ผลการวิเคราะห์ พบว่า ข้อสอบที่ทำหน้าที่ต่างกันทั้ง 3 Contrast groups มีจำนวน 6 ข้อ ได้แก่ ข้อที่ 1, 23, 24, 26, 28 และข้อที่ 35 ข้อสอบที่ทำหน้าที่ต่างกัน 2 ใน 3 ของ Contrast groups มีจำนวน 7 ข้อ ได้แก่ ข้อที่ 3, 4, 18, 22, 30, 34 และ ข้อที่ 38 และข้อสอบที่ทำหน้าที่ต่างกัน 1 ใน 3 ของ Contrast groups มีจำนวน 12 ข้อ ได้แก่ ข้อที่ 5, 8, 11, 12, 16, 17, 19,20, 27, 31, 33 และ ข้อที่ 36

วิเคราะห์การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี MGCFA-ALIGNMENT โดยประยุกต์ใช้โปรแกรม Mplus

การวิจัยครั้งนี้ผู้วิจัยได้ประยุกต์ใช้โปรแกรม Mplus ด้วยวิธี MGCFA-ALIGNMENT ในการศึกษาการทำหน้าที่ต่างกันของข้อสอบ ซึ่งประกอบด้วย วิชาคณิตศาสตร์ จำนวน 25 ข้อ วิชาภาษาไทย และวิทยาศาสตร์ จำนวนวิชาละ 40 ข้อ ผลการวิเคราะห์นำเสนอ ดังนี้

1. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ วิชาภาษาไทย และวิชาวิทยาศาสตร์ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี MGCFA-ALIGNMENT จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ โดยที่ได้กำหนดสัญลักษณ์แทนตัวแปรกลุ่ม ดังนี้

1 หมายถึง ภาคกลาง

2 หมายถึง ภาคเหนือ

3 หมายถึง ภาคตะวันออกเฉียงเหนือ

4 หมายถึง ภาคใต้

รายละเอียดแสดงในตารางที่ 4-13 ถึง ตารางที่ 4-15

ตารางที่ 4-13 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ ด้วยวิธี MGCFA-ALIGNMENT แบบ Uniform DIF และแบบ Nonuniform DIF จำแนกตามสถานที่ตั้งทางภูมิศาสตร์

Item	Uniform DIF	Nonuniform DIF
Q1	1 2 3 4	1 2 3 4
Q2	1 2 3 4	1 2 3 4
Q3	1 2 3 4	1 2 3 4
Q4	1 2 3 4	1 2 3 4
Q5	1 2 3 4	1 2 3 4
Q6	1 2 3 4	1 2 3 4
Q7	1 2 3 4	1 2 3 4
Q8	1 2 3 4	1 2 3 4
Q9	1 2 3 4	1 2 3 4
Q10	1 2 3 4	1 2 3 4
Q11	1 2 3 4	1 2 3 4
Q12	1 2 3 **(4)**	1 2 3 4
Q13	1 2 3 4	1 2 3 4
Q14	1 2 3 4	1 2 3 4
Q15	1 2 3 4	1 2 3 4
Q16	1 2 3 4	1 2 3 4
Q17	1 **(2)** 3 **(4)**	1 2 3 4
Q18	1 2 3 4	1 2 3 4
Q19	1 2 3 4	1 2 3 4
Q20	1 2 3 4	1 2 3 4
Q21	1 2 3 4	1 2 3 4
Q22	1 2 3 4	1 2 3 4
Q23	1 2 3 4	1 2 3 4
Q24	1 2 3 4	1 2 3 4
Q25	1 2 3 4	1 2 3 4

หมายเหตุ ตัวเลขในวงเล็บคือภาคที่แปรเปลี่ยนตามกลุ่มบ่งชี้ว่าข้อสอบทำหน้าที่ต่างกัน

จากตารางที่ 4-13 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี MGCFA ALIGNMENT จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ ผลการวิเคราะห์พบว่า

1. เมื่อปรับพารามิเตอร์ความยากของข้อสอบ ข้อสอบที่คงที่น้อยที่สุดหรือแปรเปลี่ยนตามกลุ่ม คือ ข้อที่ 12 และข้อที่ 17 ต่างจากศูนย์อย่างมีนัยสำคัญที่ระดับ .05 ซึ่งหมายความว่า ในข้อที่ 12 และข้อที่ 17 ตัวแปรความเป็นสถานที่ตั้งทางภูมิศาสตร์ ส่งผลให้เกิดโอกาสในการตอบข้อสอบแตกต่างกันระหว่างสถานที่ตั้งทางภูมิศาสตร์ สามารถบ่งชี้ได้ว่าข้อสอบนั้นทำหน้าที่ต่างกันแบบ Uniform DIF

2. เมื่อปรับพารามิเตอร์อำนาจจำแนกของข้อสอบ ไม่พบข้อสอบที่แปรเปลี่ยนตามกลุ่ม

ตารางที่ 4-14 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย ด้วยวิธี MGCFA-ALIGNMENT แบบ Uniform DIF และแบบ Nonuniform DIF จำแนกตามสถานที่ตั้งทางภูมิศาสตร์

Item	Uniform DIF	Nonuniform DIF
Q1	1 2 3 4	1 2 3 4
Q2	1 2 3 4	1 2 3 4
Q3	1 2 3 4	1 2 3 4
Q4	1 2 3 4	1 2 3 4
Q5	1 2 3 4	1 2 3 4
Q6	1 2 3 4	1 2 3 4
Q7	1 2 3 4	1 2 3 4
Q8	1 2 3 4	1 2 3 4
Q9	1 2 3 4	1 2 3 4
Q10	1 2 3 4	1 2 3 4
Q11	1 2 3 4	1 2 3 4
Q12	1 2 3 4	1 2 3 4
Q13	1 2 3 4	1 2 3 4
Q14	1 2 3 4	1 2 3 4
Q15	1 2 3 4	1 2 3 4

ตารางที่ 4-14 (ต่อ)

Item	Uniform DIF	Nonuniform DIF
Q16	1 2 3 4	1 2 3 4
Q17	1 2 3 4	1 2 3 4
Q18	1 2 3 4	1 2 3 4
Q19	1 2 3 4	1 2 3 4
Q20	1 2 3 4	1 2 3 4
Q21	1 2 3 4	1 2 3 4
Q22	1 2 3 4	1 2 3 4
Q23	1 2 3 4	1 2 3 4
Q24	1 2 3 4	1 2 3 4
Q25	1 2 3 4	1 2 3 4
Q26	1 **(2)** 3 **(4)**	1 2 3 4
Q27	1 2 3 4	1 2 3 4
Q28	1 2 3 4	1 2 3 4
Q29	1 2 3 4	1 2 3 4
Q30	1 2 3 4	1 2 3 4
Q31	1 2 3 4	1 2 3 4
Q32	1 2 3 4	1 2 3 4
Q33	1 **(2)** 3 4	1 2 3 4
Q34	1 2 3 4	1 2 3 4
Q35	1 2 3 4	1 2 3 4
Q36	1 2 3 4	1 2 3 4
Q37	**(1)** 2 3 4	1 2 3 4
Q38	1 2 3 4	1 2 3 4
Q39	1 2 3 4	1 2 3 4
Q40	1 2 3 4	1 2 3 4

<u>หมายเหตุ</u> ตัวเลขในวงเล็บคือภาคที่แปรเปลี่ยนตามกลุ่มบ่งชี้ว่าข้อสอบทำหน้าที่ต่างกัน

จากตารางที่ 4-14 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาภาษาไทย สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี MGCFA ALIGNMENT จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ ผลการวิเคราะห์พบว่า

1. เมื่อปรับพารามิเตอร์ความยากของข้อสอบ ข้อสอบที่คงที่น้อยที่สุดหรือแปรเปลี่ยนตามกลุ่ม คือ ข้อที่ 26, 33 และข้อที่ 37 แตกต่างจากศูนย์อย่างมีนัยสำคัญที่ระดับ .05 ซึ่งหมายความว่า ในข้อที่ 26, 33 และข้อที่ 37 ตัวแปรความเป็นสถานที่ตั้งทางภูมิศาสตร์ ส่งผลให้เกิดโอกาสในการตอบข้อสอบข้อ แตกต่างกันระหว่างสถานที่ตั้งทางภูมิศาสตร์ สามารถบ่งชี้ได้ว่าข้อสอบนั้นทำหน้าที่ต่างกัน แบบ Uniform DIF

2. เมื่อปรับพารามิเตอร์อำนาจจำแนกของข้อสอบ ไม่พบข้อสอบที่แปรเปลี่ยนตามกลุ่ม สามารถบ่งชี้ได้ว่าไม่มีข้อสอบที่ทำหน้าที่ต่างกัน

ตารางที่ 4-15 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาวิทยาศาสตร์ ด้วยวิธี MGCFA ALIGNMENT แบบ Uniform DIF และแบบ Nonuniform DIF จำแนกตามสถานที่ตั้งทางภูมิศาสตร์

Item	Uniform DIF	Nonuniform DIF
Q1	1 2 3 4	1 2 3 4
Q2	1 2 3 4	1 2 3 4
Q3	1 2 3 4	1 2 3 4
Q4	1 2 3 4	1 2 3 4
Q5	1 **(2)** 3 4	1 2 3 4
Q6	1 2 3 4	1 2 3 4
Q7	1 2 3 4	1 2 3 4
Q8	1 2 3 4	1 2 3 4
Q9	1 2 3 4	1 2 3 4
Q10	1 2 3 4	1 2 3 4
Q11	1 2 3 4	1 2 3 4
Q12	1 2 3 4	1 2 3 4
Q13	1 2 3 4	1 2 3 4
Q14	1 2 3 4	1 2 3 4

ตารางที่ 4-15 (ต่อ)

Item	Uniform DIF	Nonuniform DIF
Q15	1 2 3 4	1 2 3 4
Q16	1 2 3 4	1 2 3 4
Q18	1 2 3 **(4)**	1 2 3 4
Q19	1 2 3 4	1 2 3 4
Q20	1 2 3 4	1 2 3 4
Q21	1 2 3 4	1 2 3 4
Q22	1 2 3 4	1 2 3 4
Q23	1 2 3 4	1 2 **(3)** 4
Q24	1 2 3 4	1 2 3 4
Q25	1 2 3 4	1 2 3 4
Q26	1 2 3 4	1 2 3 **(4)**
Q27	1 2 3 4	1 2 3 **(4)**
Q28	1 2 3 4	1 2 3 **(4)**
Q29	1 2 3 4	1 2 3 4
Q30	1 2 3 4	1 2 3 4
Q31	1 2 3 4	1 2 3 4
Q32	1 2 3 4	1 2 3 4
Q33	1 2 3 4	1 2 3 4
Q34	1 2 3 4	**(1)** 2 3 **(4)**
Q35	1 2 3 4	1 2 3 **(4)**
Q36	1 2 3 4	1 2 3 4
Q37	1 2 3 4	1 2 3 4
Q38	1 2 3 4	1 2 3 4
Q39	1 2 3 4	1 2 3 4
Q40	1 2 3 4	1 2 3 4

หมายเหตุ ตัวเลขในวงเล็บคือภาคที่แปรเปลี่ยนตามกลุ่มบ่งชี้ว่าข้อสอบทำหน้าที่ต่างกัน

จากตารางที่ 4-15 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาวิทยาศาสตร์ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี MGCFA ALIGNMENT จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ ผลการวิเคราะห์พบว่า

1. เมื่อปรับพารามิเตอร์ความยากของข้อสอบ ข้อสอบที่คงที่น้อยที่สุดหรือแปรเปลี่ยนตามกลุ่ม คือ ข้อที่ 5 และ ข้อที่ 18 ต่างจากศูนย์อย่าง มีนัยสำคัญที่ระดับ .05 ซึ่งหมายความว่า ในข้อที่ 5 และ ข้อที่ 18 ตัวแปรความเป็นสถานที่ตั้งทางภูมิศาสตร์ ส่งผลให้เกิดโอกาสในการตอบข้อสอบแตกต่างกันระหว่างสถานที่ตั้งทางภูมิศาสตร์ สามารถบ่งชี้ได้ว่าข้อสอบนั้นทำหน้าที่ต่างกันแบบ Uniform DIF

2. เมื่อปรับพารามิเตอร์อำนาจจำแนกของข้อสอบ ข้อสอบที่คงที่น้อยที่สุดหรือแปรเปลี่ยนตามกลุ่ม คือ ข้อที่ 23, 26, 27, 28, 34 และข้อที่ 35 ต่างจากศูนย์อย่างมีนัยสำคัญที่ระดับ .05 ซึ่งหมายความว่า ในข้อที่ 23, 26, 27, 28, 34 และข้อที่ 35 ตัวแปรความเป็นสถานที่ตั้งทางภูมิศาสตร์ ส่งผลให้เกิดโอกาสในการตอบข้อสอบแตกต่างกันระหว่างสถานที่ตั้งทางภูมิศาสตร์ สามารถบ่งชี้ได้ว่าข้อสอบนั้นทำหน้าที่ต่างกัน แบบ Nonuniform DIF

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ วิชาภาษาไทย และวิชาวิทยาศาสตร์ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี MGCFA ALIGNMENT จำแนกตามขนาดโรงเรียนโดยที่ได้กำหนดสัญลักษณ์แทนตัวแปรกลุ่ม ดังนี้

1 หมายถึง ขนาดใหญ่พิเศษ

2 หมายถึง ขนาดใหญ่

3 หมายถึง ขนาดกลาง

4 หมายถึง ขนาดเล็ก

รายละเอียดแสดงในตารางที่ 4-16 ถึง ตารางที่ 4-18

ตารางที่ 4-16 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ ด้วยวิธี MGCFA ALIGNMENT แบบ Uniform DIF และแบบ Nonuniform DIF จำแนกตามขนาดโรงเรียน

Item	Uniform DIF	Nonuniform DIF
Q1	1 2 3 4	1 2 3 4
Q2	1 2 3 4	1 2 3 4
Q3	1 2 3 4	(1) 2 (3) 4
Q4	1 2 3 4	1 2 3 (4)
Q5	1 2 3 4	(1) 2 (3) 4
Q6	1 2 3 4	1 2 3 4
Q7	1 2 3 4	1 2 3 4
Q8	1 2 (3) 4	1 2 3 4
Q9	1 2 3 4	1 2 3 4
Q10	1 2 3 4	1 2 3 4
Q11	1 2 3 4	1 2 3 4
Q12	1 2 3 4	1 2 3 4
Q13	1 2 (3) 4	1 2 3 4
Q14	1 2 3 4	1 2 3 4
Q15	1 2 3 4	1 2 3 4
Q16	1 2 3 4	1 2 3 4
Q17	1 2 3 4	(1) 2 3 4
Q18	1 2 3 4	(1) 2 3 4
Q19	1 2 (3) 4	(1) 2 (3) 4
Q20	1 2 3 4	1 2 3 4
Q21	(1) 2 3 4	1 2 3 4
Q22	1 2 3 4	1 2 3 4
Q23	1 2 3 4	1 2 3 4
Q24	1 2 3 4	1 2 3 4
Q25	1 2 3 4	1 2 3 4

หมายเหตุ ตัวเลขในวงเล็บคือขนาดโรงเรียนที่แปรเปลี่ยนตามกลุ่มบ่งชี้ว่าข้อสอบทำหน้าที่ต่างกัน

จากตารางที่ 4-16 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี MGCFA ALIGNMENT จำแนกตามขนาดโรงเรียน ผลการวิเคราะห์ พบว่า

1. เมื่อปรับพารามิเตอร์ความยากของข้อสอบ ข้อสอบที่คงที่น้อยที่สุดหรือแปรเปลี่ยนตามกลุ่ม จำนวน 4 ข้อ ได้แก่ ข้อที่ 8, 13, 19 และข้อที่ 21 ต่างจากศูนย์อย่างมีนัยสำคัญที่ระดับ .05 ซึ่งหมายความว่า ในข้อที่ 8, 13, 19 และข้อที่ 21 ตัวแปรขนาดโรงเรียน ส่งผลให้เกิดโอกาสในการตอบข้อสอบแตกต่างกันระหว่างขนาดโรงเรียน สามารถบ่งชี้ได้ว่าข้อสอบนั้นทำหน้าที่ต่างกัน แบบ Uniform DIF

2. เมื่อปรับพารามิเตอร์อำนาจจำแนกของข้อสอบ ข้อสอบที่คงที่น้อยที่สุดหรือแปรเปลี่ยนตามกลุ่ม จำนวน 6 ข้อ ได้แก่ คือ ข้อที่ 3, 4, 5, 17, 18 และข้อที่ 19 ต่างจากศูนย์อย่างมีนัยสำคัญที่ระดับ .05 ซึ่งหมายความว่า ในข้อที่ 3, 4, 5, 17, 18 และข้อที่ 19 ตัวแปรขนาดโรงเรียน ส่งผลให้เกิดโอกาสในการตอบข้อสอบแตกต่างกันระหว่างขนาดโรงเรียน สามารถบ่งชี้ได้ว่าข้อสอบนั้นทำหน้าที่ต่างกัน แบบ Nonuniform DIF

ตารางที่ 4-17 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย ด้วยวิธี MGCFA ALIGNMENT แบบ Uniform DIF และแบบ Nonuniform DIF จำแนกตามขนาดโรงเรียน

Item	Uniform DIF	Nonuniform DIF
Q1	1 2 3 4	1 2 3 4
Q2	1 2 3 4	(1) 2 3 4
Q3	1 2 3 4	1 2 3 4
Q4	1 2 3 4	1 2 3 4
Q5	1 2 3 4	1 2 3 4
Q6	1 2 3 4	(1) 2 3 4
Q7	1 2 3 4	1 2 3 4
Q8	1 2 3 4	1 2 3 4
Q9	1 2 3 4	1 2 3 4
Q10	1 2 3 4	1 2 3 4
Q13	1 2 3 4	1 2 3 4

ตารางที่ 4-17 (ต่อ)

Item	Uniform DIF	Nonuniform DIF
Q14	1 2 3 4	1 2 3 4
Q15	1 2 3 4	1 2 3 4
Q16	1 2 3 4	1 2 3 4
Q17	1 2 3 4	(1) 2 3 4
Q18	1 2 3 4	1 2 3 4
Q19	1 2 3 4	1 2 3 4
Q20	1 2 3 4	1 2 3 4
Q21	1 2 3 4	1 2 3 4
Q22	1 2 3 4	1 2 3 4
Q23	1 2 3 4	1 2 3 4
Q24	1 2 3 4	1 2 3 4
Q25	1 2 3 4	(1) 2 3 4
Q27	1 2 3 4	1 2 3 4
Q28	1 2 3 4	1 2 3 4
Q29	1 2 3 4	1 2 3 4
Q30	1 2 3 4	1 2 3 4
Q31	1 2 3 4	1 2 3 4
Q32	1 2 3 4	1 2 3 4
Q33	1 2 3 4	1 2 3 4
Q34	1 2 3 4	1 2 3 4
Q35	1 2 3 4	1 2 3 4
Q36	1 2 3 4	1 2 3 4
Q37	1 2 3 4	1 2 3 4
Q38	1 2 3 4	1 2 3 4
Q39	1 2 3 4	1 2 3 4
Q40	1 2 3 4	(1) 2 3 4

หมายเหตุ ตัวเลขในวงเล็บคือขนาดโรงเรียนที่แปรเปลี่ยนตามกลุ่มบ่งชี้ว่าข้อสอบทำหน้าที่ต่างกัน

จากตารางที่ 4-17 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาภาษาไทย สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี MGCFA ALIGNMENT จำแนกตามขนาดโรงเรียน ผลการวิเคราะห์ พบว่า

1. เมื่อปรับพารามิเตอร์ความยากของข้อสอบ ไม่พบข้อสอบที่แปรเปลี่ยนตามกลุ่ม สามารถบ่งชี้ได้ว่าไม่มีข้อสอบที่ทำหน้าที่ต่างกันแบบ Uniform DIF

2. เมื่อปรับพารามิเตอร์อำนาจจำแนกของข้อสอบ ข้อสอบที่คงที่น้อยที่สุดหรือแปรเปลี่ยนตามกลุ่ม คือ ข้อที่ 2, 6, 17, 25และข้อที่ 40 ต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งหมายความว่า ในข้อที่ 2, 6, 17, 25 และข้อที่ 40 ตัวแปรขนาดโรงเรียน ส่งผลให้เกิดโอกาสในการตอบข้อสอบแตกต่างกันระหว่างขนาดโรงเรียน สามารถบ่งชี้ได้ว่าข้อสอบนั้นทำหน้าที่ต่างกัน แบบ Nonuniform DIF

ตารางที่ 4-18 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาวิทยาศาสตร์ ด้วยวิธี MGCFA ALIGNMENT แบบ Uniform DIF และแบบ Nonuniform DIF จำแนกตามขนาดโรงเรียน

Item	Uniform DIF	Nonuniform DIF
Q1	1 2 3 4	**(1)** 2 3 4
Q2	1 2 3 4	1 2 3 4
Q3	1 2 3 4	1 2 3 4
Q4	1 2 3 4	1 2 3 4
Q5	1 2 3 4	1 2 3 4
Q6	1 2 3 4	1 2 3 4
Q7	1 2 3 4	1 2 3 4
Q8	1 2 3 4	1 2 3 4
Q9	1 2 3 4	1 2 3 4
Q10	1 2 3 4	1 2 3 4
Q11	1 2 3 4	1 2 3 4
Q12	1 2 3 4	1 2 3 4
Q13	1 2 3 4	1 2 3 4
Q15	1 2 3 4	1 2 3 4
Q16	1 2 3 4	1 2 3 4
Q17	1 2 3 4	1 2 3 **(4)**
Q18	1 2 3 4	1 2 3 4
Q19	1 2 3 4	1 2 3 4

ตารางที่ 4-18 (ต่อ)

Item	Uniform DIF	Nonuniform DIF
Q20	1 2 3 4	1 2 3 4
Q21	1 2 3 4	1 2 3 4
Q22	1 2 3 4	**(1)** 2 **(3)** 4
Q23	1 2 3 4	**(1)** 2 **(3)** 4
Q24	1 2 **(3)** 4	1 2 3 4
Q25	1 2 3 4	1 2 3 4
Q26	1 2 3 4	1 **(2)** 3 **(4)**
Q27	1 2 3 4	**(1)** 2 3 4
Q28	1 2 3 4	**(1)** 2 3 4
Q29	1 2 3 4	1 2 3 4
Q30	1 2 3 4	1 2 3 4
Q31	1 2 3 4	1 2 3 4
Q32	1 2 3 4	1 2 3 4
Q33	1 2 3 4	1 2 3 4
Q34	1 2 3 **(4)**	**(1)** 2 3 4
Q35	1 2 3 4	1 **(2)** 3 **(4)**
Q36	1 2 3 4	1 2 3 4
Q37	1 2 3 4	1 2 3 4
Q38	1 2 3 4	1 2 3 4
Q39	1 2 3 4	1 2 3 4
Q40	1 2 3 4	1 2 3 4

หมายเหตุ ตัวเลขในวงเล็บคือขนาดโรงเรียนที่แปรเปลี่ยนตามกลุ่มบ่งชี้ว่าข้อสอบทำหน้าที่ต่างกัน

จากตารางที่ 4-18 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาวิทยาศาสตร์ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี MGCFA ALIGNMENT จำแนกตามขนาดโรงเรียน ผลการวิเคราะห์ พบว่า

1. เมื่อปรับพารามิเตอร์ความยากของข้อสอบ ข้อสอบที่คงที่น้อยที่สุดหรือแปรเปลี่ยนตามกลุ่ม คือ ข้อที่ 24 และข้อที่ 34 ต่างจากศูนย์อย่างมีนัยสำคัญที่ระดับ .05 ซึ่งหมายความว่า ในข้อที่ 24 และข้อที่ 34 ตัวแปรขนาดโรงเรียน ส่งผลให้เกิดโอกาสในการตอบข้อสอบข้อแตกต่างกันระหว่างขนาดโรงเรียน สามารถบ่งชี้ได้ว่าข้อสอบนั้นทำหน้าที่ต่างกัน แบบ Uniform DIF

2. เมื่อปรับพารามิเตอร์อำนาจจำแนกของข้อสอบ ข้อสอบที่คงที่น้อยที่สุดหรือแปรเปลี่ยนตามกลุ่ม คือ ข้อที่ 1, 17, 22, 23, 26, 27, 28, 34 และข้อที่ 35 ต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งหมายความว่า ในข้อที่ 1, 17, 22, 23, 26, 27, 28, 34 และข้อที่ 35 ตัวแปรขนาดโรงเรียน ส่งผลให้เกิดโอกาสในการตอบข้อสอบแตกต่างกันระหว่างขนาดโรงเรียน สามารถบ่งชี้ได้ว่าข้อสอบนั้นทำหน้าที่ต่างกัน แบบ nunUniform DIF

การวิเคราะห์การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี MIMIC โดยประยุกต์ใช้โปรแกรม Mplus

การวิจัยครั้งนี้ผู้วิจัยได้ประยุกต์ใช้โปรแกรม Mplus ด้วยวิธี MIMIC ในการศึกษาการทำหน้าที่ต่างกันของข้อสอบ ซึ่งประกอบด้วย วิชาคณิตศาสตร์ จำนวนข้อสอบ 25 ข้อ วิชาภาษาไทย และวิทยาศาสตร์ มีจำนวนข้อสอบวิชาละ 40 ข้อ ผลการวิเคราะห์นำเสนอดังนี้

1. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ วิชาภาษาไทย และวิชาวิทยาศาสตร์ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี วิธี MIMIC จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ โดยกำหนดให้กลุ่มอ้างอิง (R) เป็นภาคกลาง กลุ่มเปรียบเทียบ (F) เป็นภาคเหนือ ภาคตะวันออกเฉียงเหนือ และภาคใต้ รายละเอียดแสดงในตารางที่ 4-19 ถึง ตารางที่ 4-24 ดังนี้

ตารางที่ 4-19 ความกลมกลืนของโมเดล CFA MIMIC วิชาคณิตศาสตร์ จำแนกตามสถานที่ตั้งทางภูมิศาสตร์

Statistics	**CFA**	**MIMIC (no Direct Effect) @0**	**MIMIC (1 Direct Effect)**	**MIMIC (2 Direct Effect)**	**MIMIC (3 Direct Effect)**	**MIMIC (4 Direct Effect)**
χ^2	1833.638 (df = 275)	2053.207 (df = 347)	2030.927 (df = 346)	2010.619 (df = 345)	1991.850 (df = 344)	1975.833 (df = 343)
RMSEA	0.027	0.025	0.025	0.025	0.025	0.024
CFI	0.705	0.691	0.695	0.698	0.701	0.704
TLI	0.678	0.666	0.669	0.672	0.675	0.677

ตารางที่ 4-19 (ต่อ)

Statistics	CFA	MIMIC (no Direct Effect) @0	MIMIC (1 Direct Effect)	MIMIC (2 Direct Effect)	MIMIC (3 Direct Effect)	MIMIC (4 Direct Effect)
Statistics	MIMIC (5 Direct Effect)	MIMIC (6 Direct Effect)				
χ^2	1960.876 (df = 342)	1946.019 (df = 341)				
RMSEA	0.024	0.024				
CFI	0.707	0.709				
TLI	0.678	0.680				

จากตารางที่ 4-19 ผลการวิเคราะห์ความกลมกลืนของโมเดล CFA MIMIC วิชาคณิตศาสตร์ จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ จากการวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) ซึ่งพิจารณาจากค่าไค-สแควร์ (Chi-square) มีค่าเท่ากับ 1833.638 ที่องศาอิสระ (df) เท่ากับ 275 ดัชนีวัดรากที่สองของความคลาดเคลื่อนโมเดลในการประมาณค่าพารามิเตอร์ที่เหมาะสม (RMSEA) มีค่าเท่ากับ 0.027 ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (CFI) มีค่าเท่ากับ 0.705 ดัชนี The tucker lewis index (TLI) มีค่าเท่ากับ 0.678 เมื่อพิจารณาการทำนายตัวแปรข้อสอบจากตัวแปรสถานที่ตั้งทางภูมิศาสตร์ ที่มีดัชนีปรับแต่งมีข้อค้นพบ ดังนี้

1. เมื่อพิจารณาการทำนายตัวแปรข้อสอบจากตัวแปรสถานที่ตั้งทางภูมิศาสตร์ ที่มีดัชนีปรับแต่ง พบว่า สามารถทำนายตัวแปรข้อสอบที่มีดัชนีปรับแต่ง จำนวน 4 ข้อ โดยโมเดล MIMIC เริ่มต้น เป็นการเพิ่มตัวแปรทำนายที่มีอิทธิพลทางตรงต่อข้อสอบและบังคับค่าให้เป็น 0 เพื่อกำหนดให้ตัวแปรต้นไม่มีผลทางตรงต่อข้อสอบ มีค่าไค-สแควร์ (Chi-square) เท่ากับ 2053.207 ที่องศาอิสระ (df) เท่ากับ 347 ดัชนีวัดรากที่สองของความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ที่เหมาะสม (RMSEA) มีค่าเท่ากับ 0.025 ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (CFI) มีค่าเท่ากับ 0.691 ดัชนี The tucker lewis index (TLI) มีค่าเท่ากับ 0.666

2. เมื่อพิจารณาโมเดล MIMIC ที่เพิ่มอิทธิพลทางตรงจากตัวทำนายไปที่ข้อสอบที่มีดัชนีปรับแก้ 1 ข้อ โดยเลือกข้อที่มีดัชนีปรับแก้สูงสุด มีค่าไค-สแควร์ (Chi-square) เท่ากับ 2030.927 ที่องศาอิสระ (df) เท่ากับ 346 ดัชนีวัดรากที่สองของความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ที่เหมาะสม (RMSEA) มีค่าเท่ากับ 0.025 ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (CFI) มีค่าเท่ากับ 0.695 ดัชนี The tucker lewis index (TLI) มีค่าเท่ากับ 0.669

3. เมื่อพิจารณาโมเดล MIMIC ที่เพิ่มอิทธิพลทางตรงจากตัวทำนายไปที่ข้อสอบ ที่มีดัชนีปรับแก้ 2 ข้อมีค่าไค-สแควร์ (Chi-square) เท่ากับ 2010.619 ที่องศาอิสระ (df) เท่ากับ 345 ดัชนีวัดรากที่สองของความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ที่เหมาะสม (RMSEA) มีค่าเท่ากับ 0.025 ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (CFI) มีค่าเท่ากับ 0.698 ดัชนี The tucker lewis index (TLI) มีค่าเท่ากับ 0.672

4. เมื่อพิจารณาโมเดล MIMIC ที่เพิ่มอิทธิพลทางตรงจากตัวทำนายไปที่ข้อสอบ ที่มีดัชนีปรับแก้ 3 ข้อ มีค่าไค-สแควร์ (Chi-square) เท่ากับ 1991.850 ที่องศาอิสระ (df) เท่ากับ 344 ดัชนีวัดรากที่สองของความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ที่เหมาะสม (RMSEA) มีค่าเท่ากับ 0.025 ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (CFI) มีค่าเท่ากับ 0.701 ดัชนี The tucker lewis index (TLI) มีค่าเท่ากับ 0.675

5. เมื่อพิจารณาโมเดล MIMIC ที่เพิ่มอิทธิพลทางตรงจากตัวทำนายไปที่ข้อสอบ ที่มีดัชนีปรับแก้ 4 ข้อ มีค่าไค-สแควร์ (Chi-square) เท่ากับ 1975.833 ที่องศาอิสระ (df) เท่ากับ 343 ดัชนีวัดรากที่สองของความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ที่เหมาะสม (RMSEA) มีค่าเท่ากับ 0.024 ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (CFI) มีค่าเท่ากับ 0.704 ดัชนี The tucker lewis index (TLI) มีค่าเท่ากับ 0.677

6. เมื่อพิจารณาโมเดล MIMIC ที่เพิ่มอิทธิพลทางตรงจากตัวทำนายไปที่ข้อสอบ ที่มีดัชนีปรับแก้ 5 ข้อ มีค่าไค-สแควร์ (Chi-square) เท่ากับ 1960.876 ที่องศาอิสระ (df) เท่ากับ 342 ดัชนีวัดรากที่สองของความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ที่เหมาะสม (RMSEA) มีค่าเท่ากับ 0.024 ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (CFI) มีค่าเท่ากับ 0.707 ดัชนี The tucker lewis index (TLI) มีค่าเท่ากับ 0.678

7. เมื่อพิจารณาโมเดล MIMIC ที่เพิ่มอิทธิพลทางตรงจากตัวทำนายไปที่ข้อสอบ ที่มีดัชนีปรับแก้ 6 ข้อ มีค่าไค-สแควร์ (Chi-square) เท่ากับ 1946.019 ที่องศาอิสระ (df) เท่ากับ 341 ดัชนีวัดรากที่สองของความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ที่เหมาะสม (RMSEA) มีค่าเท่ากับ 0.024 ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (CFI) มีค่าเท่ากับ 0.709 ดัชนี The tucker lewis index (TLI) มีค่าเท่ากับ 0.680

ตารางที่ 4-20 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ ด้วยวิธี MIMIC แบบ Uniform DIF จำแนกตามสถานที่ตั้งทางภูมิศาสตร์

Item		Estimate	SE	Z	P-value
MATH ON	NORTH	0.114	0.023	4.940*	0.000
	NORTHEST	-0.054	0.024	-2.238*	0.025
	SOUTH	0.090	0.020	4.487*	0.000
Q01 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q02 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q03 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q04 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q05 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q06 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.161	0.058	2.791*	0.005
	SOUTH	0.000	0.000	999.000	999.000
Q07 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000

ตารางที่ 4-20 (ต่อ)

Item		Estimate	SE	Z	P-value
Q08 ON	NORTH	-0.148	0.061	-2.431*	0.015
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q09 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q10 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q11 ON	NORTH	0.195	0.062	3.147*	0.002
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q12 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	-0.245	0.061	-4.032*	0.000
Q13 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q14 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q15 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q16 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000

ตารางที่ 4-20 (ต่อ)

Item		Estimate	SE	Z	P-value
Q17 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	-0.228	0.060	-3.782*	0.000
	SOUTH	0.000	0.000	999.000	999.000
Q18 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q19 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q20 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q21 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q22 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q23 ON	NORTH	-0.153	0.064	-2.386*	0.017
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q24 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q25 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000

หมายเหตุ * p < .05; ข้อสอบ DIF คือข้อที่อย่างน้อย 1 ภูมิภาคมีอิทธิพลทางตรงต่อข้อสอบ
อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

จากตารางที่ 4-20 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มที่มากกว่าสองกลุ่ม วิชาคณิตศาสตร์ ด้วยวิธี MIMIC แบบเอกรูป (Uniform DIF) จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ ผลการวิเคราะห์ข้อสอบจำนวน 25 ข้อ ตัวแปรทำนายซึ่งได้แก่ สถานที่ตั้งทางภูมิศาสตร์ (ภาคกลาง ภาคเหนือ ภาคตะวันออกเฉียงเหนือ และภาคใต้) ที่มีอิทธิพลทางตรงต่อข้อสอบ ซึ่งข้อสอบที่มีดัชนีปรับแต่งสูงสุดคือ ข้อที่ 23 รองลงมาคือ ข้อที่ 12, 17, 11, 8 และ ข้อที่ 6 ตามลำดับ ที่ทำให้ตัวแปรสถานที่ตั้งทางภูมิศาสตร์มีอิทธิพลทางตรงต่อข้อสอบที่มีดัชนีปรับแต่งมีค่าต่างจากศูนย์อย่างมีนัยสำคัญที่ระดับ .05 ซึ่งหมายความว่า ในข้อที่ 23 รองลงมาคือ ข้อที่ 12, 17, 11, 8 และ ข้อที่ 6 ตัวแปรความเป็นสถานที่ตั้งทางภูมิศาสตร์ ส่งผลให้เกิดโอกาสในการตอบข้อสอบแตกต่างกันระหว่างภาค สามารถบ่งชี้ได้ว่าข้อสอบนั้นทำหน้าที่ต่างกัน แบบ เอกรูป (Uniform DIF) จากผลการวิเคราะห์ พบว่า ความเป็นภาคตะวันออกเฉียงเหนือ (NORTHEST) จะได้เปรียบในข้อสอบข้อที่ 6 จะเสียเปรียบในข้อสอบข้อที่ 17 ความเป็นภาคเหนือ (NORTH) จะได้เปรียบในข้อสอบข้อที่ 11 จะเสียเปรียบในข้อสอบข้อที่ 8 และข้อที่ 23 และความเป็นภาคใต้ (SOUTH) จะเสียเปรียบในข้อสอบข้อที่ 12 เมื่อพิจารณาในภาพรวม พบว่า ภาคตะวันออกเฉียงเหนือจะเสียเปรียบในการตอบวิชาคณิตศาสตร์

ตารางที่ 4-21 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ ด้วยวิธี MIMIC แบบ Nonuniform DIF จำแนกตามสถานที่ตั้งทางภูมิศาสตร์

Item	Interaction	Estimate	SE	Z	P-value
Q01 ON	MATHNORTH	-0.032	0.148	-0.214	0.830
	MATHNORTHEST	-0.284	0.141	-2.005*	0.045
	MATHSOUTH	0.119	0.161	0.737	0.461
Q02 ON	MATHNORTH	0.075	0.178	0.422	0.673
	MATHNORTHEST	-0.602	0.205	-2.935*	0.003
	MATHSOUTH	0.597	0.198	3.011*	0.003
Q03 ON	MATHNORTH	-0.559	0.194	-2.881*	0.004
	MATHNORTHEST	-1.141	0.236	-4.837*	0.000
	MATHSOUTH	-0.142	0.191	-0.742	0.458

ตารางที่ 4-21 (ต่อ)

Item	Interaction	Estimate	SE	Z	P-value
Q04 ON	MATHNORTH	-0.458	0.171	-2.687*	0.007
	MATHNORTHEST	-1.019	0.204	-4.986*	0.000
	MATHSOUTH	0.026	0.170	0.155	0.877
Q05 ON	MATHNORTH	-0.352	0.171	-2.061*	0.039
	MATHNORTHEST	-0.822	0.198	-4.164*	0.000
	MATHSOUTH	0.176	0.177	0.994	0.320
Q06 ON	MATHNORTH	-0.228	0.160	-1.428	0.153
	MATHNORTHEST	-0.670	0.179	-3.743*	0.000
	MATHSOUTH	-0.039	0.159	-0.247	0.805
Q07 ON	MATHNORTH	-0.084	0.162	-0.522	0.601
	MATHNORTHEST	-0.227	0.178	-1.278	0.201
	MATHSOUTH	0.461	0.180	2.569*	0.010
Q08 ON	MATHNORTH	-0.032	0.153	-0.207	0.836
	MATHNORTHEST	-0.183	0.159	-1.156	0.248
	MATHSOUTH	0.459	0.161	2.854*	0.004
Q09 ON	MATHNORTH	-0.205	0.149	-1.382	0.167
	MATHNORTHEST	-0.332	0.161	-2.059*	0.040
	MATHSOUTH	0.254	0.153	1.662	0.097
Q10 ON	MATHNORTH	0.089	0.149	0.599	0.549
	MATHNORTHEST	-0.221	0.161	-1.375	0.169
	MATHSOUTH	0.472	0.161	2.927*	0.003
Q11 ON	MATHNORTH	0.406	0.235	1.727	0.084
	MATHNORTHEST	-0.245	0.212	-1.156	0.248
	MATHSOUTH	0.734	0.261	2.818*	0.005
Q12 ON	MATHNORTH	-0.266	0.256	-1.040	0.298
	MATHNORTHEST	-0.657	0.253	-2.600*	0.009
	MATHSOUTH	-0.224	0.255	-0.878	0.380

ตารางที่ 4-21 (ต่อ)

Item	Interaction	Estimate	SE	Z	P-value
Q13 ON	MATHNORTH	-0.185	0.239	-0.773	0.439
	MATHNORTHEST	-0.811	0.245	-3.305*	0.001
	MATHSOUTH	0.026	0.250	0.106	0.916
Q14 ON	MATHNORTH	0.302	0.238	1.268	0.205
	MATHNORTHEST	-0.572	0.226	-2.529*	0.011
	MATHSOUTH	0.068	0.230	0.294	0.769
Q15 ON	MATHNORTH	-0.030	0.184	-0.161	0.872
	MATHNORTHEST	-0.554	0.192	-2.884*	0.004
	MATHSOUTH	0.250	0.195	1.279	0.201
Q16 ON	MATHNORTH	-0.277	0.188	-1.478	0.140
	MATHNORTHEST	-0.673	0.201	-3.354*	0.001
	MATHSOUTH	-0.131	0.190	-0.691	0.490
Q17 ON	MATHNORTH	-0.273	0.229	-1.194	0.233
	MATHNORTHEST	-0.387	0.232	-1.666	0.096
	MATHSOUTH	-0.170	0.233	-0.730	0.466
Q18 ON	MATHNORTH	0.091	0.244	0.373	0.709
	MATHNORTHEST	-0.284	0.237	-1.201	0.230
	MATHSOUTH	0.175	0.251	0.695	0.487
Q19 ON	MATHNORTH	0.191	0.253	0.754	0.451
	MATHNORTHEST	-0.035	0.254	-0.138	0.890
	MATHSOUTH	0.440	0.273	1.611	0.107
Q20 ON	MATHNORTH	0.155	0.196	0.790	0.429
	MATHNORTHEST	-0.511	0.203	-2.514*	0.012
	MATHSOUTH	0.399	0.209	1.914	0.056
Q21 ON	MATHNORTH	0.183	0.199	0.920	0.358
	MATHNORTHEST	-0.451	0.205	-2.205*	0.027
	MATHSOUTH	0.443	0.214	2.067*	0.039

ตารางที่ 4-21 (ต่อ)

Item	Interaction	Estimate	SE	Z	P-value
Q22 ON	MATHNORTH	-0.483	0.162	-2.979*	0.003
	MATHNORTHEST	-0.557	0.176	-3.154*	0.002
	MATHSOUTH	0.082	0.157	0.521	0.603
Q23 ON	MATHNORTH	-0.240	0.162	-1.486	0.137
	MATHNORTHEST	-0.619	0.176	-3.522*	0.000
	MATHSOUTH	0.314	0.162	1.943	0.052
Q24 ON	MATHNORTH	0.093	0.163	0.569	0.570
	MATHNORTHEST	-0.229	0.175	-1.309	0.191
	MATHSOUTH	0.250	0.167	1.496	0.135
Q25 ON	MATHNORTH	0.323	0.162	1.993*	0.046
	MATHNORTHEST	-0.251	0.169	-1.485	0.137
	MATHSOUTH	0.720	0.183	3.929*	0.000

หมายเหตุ * $p < .05$; ข้อสอบ DIF คือข้อที่อย่างน้อย 1 ภูมิภาคที่มีปฏิสัมพันธ์กับ θ แล้วมีอิทธิพลทางตรงต่อข้อสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

จากตารางที่ 4-21 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มที่มากกว่าสองกลุ่ม ด้วยวิธี MIMIC แบบ Nonuniform DIF วิชาคณิตศาสตร์ จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ ผลการวิเคราะห์ปฏิสัมพันธ์ระหว่างความสามารถด้านคณิตศาสตร์หรือตัวแปรแฝง (θ) กับความเป็นสมาชิกของกลุ่มสถานที่ตั้งทางภูมิศาสตร์ 4 ภูมิภาค ประกอบด้วย ภาคกลาง ภาคเหนือ ภาคตะวันออกเฉียงเหนือ และภาคใต้ ที่มีอิทธิพลทางตรงต่อข้อสอบอย่างมีนัยสำคัญที่ระดับ .05 ถ้าพบว่า มีปฏิสัมพันธ์ตั้งแต่ 1 ปฏิสัมพันธ์ สามารถบ่งชี้ได้ว่าข้อสอบนั้นทำหน้าที่ต่างกันแบบ Nonuniform DIF ได้แก่ข้อสอบข้อที่ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 20, 21, 22, 23 และข้อที่ 25 หมายความว่า ปฏิสัมพันธ์ระหว่างความสามารถด้านคณิตศาสตร์ (θ) กับความเป็นสมาชิกของกลุ่มสถานที่ตั้งทางภูมิศาสตร์ส่งผลให้เกิดโอกาสในการตอบข้อสอบแตกต่างกันระหว่างคู่ปฏิสัมพันธ์

ตารางที่ 4-22 ความกลมกลืนของโมเดล CFA MIMIC วิชาภาษาไทย จำแนกตามสถานที่ตั้งทางภูมิศาสตร์

Statistics	CFA	MIMIC (no Direct Effect) @0	MIMIC (1 Direct Effect)	MIMIC (2 Direct Effect)	MIMIC (3 Direct Effect)	MIMIC (4 Direct Effect)	MIMIC (5 Direct Effect)
χ^2	6272.816 (df = 740)	6756.189 (df = 857)	6620.706 (df = 856)	6606.359 (df = 855)	6592.546 (df = 854)	6579.156 (df = 853)	6568.859 (df = 852)
RMSEA	0.031	0.029	0.029	0.029	0.029	0.029	0.029
CFI	0.600	0.592	0.601	0.602	0.603	0.604	0.604
TLI	0.579	0.571	0.581	0.581	0.582	0.582	0.582

จากตารางที่ 4-22 ผลการวิเคราะห์ความกลมกลืนของโมเดล CFA MIMIC วิชาภาษาไทย จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ จากการวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) ซึ่งพิจารณาจากค่าไค-สแควร์ (Chi-square) มีค่าเท่ากับ 6272.816 ที่องศาอิสระ (df) เท่ากับ 740 ดัชนีวัดรากที่สองของความคลาดเคลื่อนในการประมาณค่า พารามิเตอร์ที่เหมาะสม (RMSEA) มีค่าเท่ากับ 0.031 ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (CFI) มีค่าเท่ากับ 0.600 ดัชนี The tucker lewis index (TLI) มีค่าเท่ากับ 0.579 เมื่อพิจารณาการทำนายตัวแปรข้อสอบจากตัวแปรสถานที่ตั้งทางภูมิศาสตร์ที่มีดัชนีปรับแต่งมีข้อค้นพบ ดังนี้

1. เมื่อพิจารณาการทำนายตัวแปรข้อสอบจากสถานที่ตั้งทางภูมิศาสตร์ ที่มีดัชนีปรับแต่ง พบว่า สามารถทำนายตัวแปรข้อสอบที่มีดัชนีปรับแต่ง จำนวน 5 ข้อ โดยโมเดล MIMIC เริ่มต้นเป็นการเพิ่มตัวแปรทำนายที่มีอิทธิพลทางตรงต่อข้อสอบและบังคับค่าให้เป็น 0 เพื่อกำหนดให้ตัวแปรต้นไม่มีผลทางตรงต่อข้อสอบ มีค่าไค-สแควร์ (Chi-square) เท่ากับ 6756.189 ที่องศาอิสระ (df) เท่ากับ 857 ดัชนีวัดรากที่สองของความคลาดเคลื่อนในการประมาณค่า พารามิเตอร์ที่เหมาะสม (RMSEA) มีค่าเท่ากับ 0.029 ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (CFI) มีค่าเท่ากับ 0.592 ดัชนี The tucker lewis index (TLI) มีค่าเท่ากับ 0.571

2. เมื่อพิจารณาโมเดล MIMIC ที่เพิ่มอิทธิพลทางตรงจากตัวทำนายไปที่ข้อสอบ ที่มีดัชนีปรับแก้ 1 ข้อ โดยเลือกข้อที่มีดัชนีปรับแก้สูงสุด มีค่าไค-สแควร์ (Chi-square) เท่ากับ 6620.706 ที่องศาอิสระ (df) เท่ากับ 856 ดัชนีวัดรากที่สองของความคลาดเคลื่อนในการประมาณค่า พารามิเตอร์

ที่เหมาะสม (RMSEA) มีค่าเท่ากับ 0.029 ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (CFI) มีค่าเท่ากับ 0.601ดัชนี The tucker lewis index (TLI) มีค่าเท่ากับ 0.581

3. เมื่อพิจารณาโมเดล MIMIC ที่เพิ่มอิทธิพลทางตรงจากตัวทำนายไปที่ข้อสอบ ที่มีดัชนีปรับแก้ 2 ข้อ โดยเลือกข้อที่มีดัชนีปรับแก้สูงสุด มีค่าไค-สแควร์ (Chi-square) เท่ากับ 6606.359 ที่องศาอิสระ (df) เท่ากับ 855 ดัชนีวัดรากที่สองของความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ที่เหมาะสม (RMSEA) มีค่าเท่ากับ 0.029 ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (CFI) มีค่าเท่ากับ 0.602 ดัชนี The tucker lewis index (TLI) มีค่าเท่ากับ 0.581

4. เมื่อพิจารณาโมเดล MIMIC ที่เพิ่มอิทธิพลทางตรงจากตัวทำนายไปที่ข้อสอบ ที่มีดัชนีปรับแก้ 3 ข้อ มีค่าไค-สแควร์ (Chi-square) เท่ากับ 6592.546 ที่องศาอิสระ (df) เท่ากับ 854 ดัชนีวัดรากที่สองของความคลาดเคลื่อนในการประมาณค่า พารามิเตอร์ที่เหมาะสม (RMSEA) มีค่าเท่ากับ 0.029 ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (CFI) มีค่าเท่ากับ 0.603 ดัชนี The tucker lewis index (TLI) มีค่าเท่ากับ 0.582

5. เมื่อพิจารณาโมเดล MIMIC ที่เพิ่มอิทธิพลทางตรงจากตัวทำนายไปที่ข้อสอบ ที่มีดัชนีปรับแก้5 ข้อมีค่าไค-สแควร์ (Chi-square) เท่ากับ 6568.859 ที่องศาอิสระ (df) เท่ากับ 854 ดัชนีวัดรากที่สองของความคลาดเคลื่อนในการประมาณค่า พารามิเตอร์ที่เหมาะสม (RMSEA) มีค่าเท่ากับ 0.029 ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (CFI) มีค่าเท่ากับ 0.604 ดัชนี The tucker lewis index (TLI) มีค่าเท่ากับ 0.582

ตารางที่ 4-23 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย ด้วยวิธี MIMIC แบบ Uniform DIF จำแนกตามสถานที่ตั้งทางภูมิศาสตร์

Item		Estimate	SE	Z	P-value
THAI ON	NORTH	0.028	0.010	2.886*	0.004
	NORTHEST	-0.013	0.008	-1.688*	0.091
	SOUTH	0.048	0.013	3.699*	0.000
Q01 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q02 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q03 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q04 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q05 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q06 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q07 ON	NORTH	0.169	0.060	2.797*	0.005
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000

ตารางที่ 4-23 (ต่อ)

Item		Estimate	SE	Z	P-value
Q08 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q09 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q10 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q11 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.112	0.055	2.032*	0.042
Q12 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q13 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q14 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q15 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q16 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000

ตารางที่ 4-23 (ต่อ)

Item		Estimate	SE	Z	P-value
Q17 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q18 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q19 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q20 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q21 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q22 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q23 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q24 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q25 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000

ตารางที่ 4-23 (ต่อ)

Item		Estimate	SE	Z	P-value
Q26 ON	NORTH	-0.724	0.098	-7.377	0.005
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q27 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q28 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q29 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q30 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q31 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q32 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q33 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q34 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000

ตารางที่ 4-23 (ต่อ)

Item		Estimate	SE	Z	P-value
Q35 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q36 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q37 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q37 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q38 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q39 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q40 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000

หมายเหตุ * p < .05; ข้อสอบ DIF คือข้อที่อย่างน้อย 1 ภูมิภาคมีอิทธิพลทางตรงต่อข้อสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

จากตารางที่ 4-23 ผลการวิเคราะห์การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย สำหรับกลุ่มที่มากกว่าสองกลุ่ม ด้วยวิธี MIMIC แบบเอกรูป (Uniform DIF) จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ ผลการวิเคราะห์ข้อสอบจำนวน 40 ข้อ ตัวแปรทำนายซึ่งได้แก่ สถานที่ตั้ง

ทางภูมิศาสตร์ (ภาคกลาง ภาคเหนือ ภาคตะวันออกเฉียงเหนือ และภาคใต้) ที่มีอิทธิพลทางตรงต่อข้อสอบ ซึ่งข้อสอบที่มีดัชนีปรับแต่งสูงสุดคือ ข้อที่ 26 รองลงมาคือ ข้อที่ 7 และ ข้อที่ 11 ตามลำดับ ที่ทำให้ตัวแปรสถานที่ตั้งทางภูมิศาสตร์มีอิทธิพลทางตรงต่อข้อสอบที่มีดัชนีปรับแต่งมีค่าต่างจากศูนย์อย่างมีนัยสำคัญที่ระดับ .05 ซึ่งหมายความว่า ในข้อที่ 26 รองลงมาคือ ข้อที่ 7 และ ข้อที่ 11 ตัวแปรความเป็นสถานที่ตั้งทางภูมิศาสตร์ ส่งผลให้เกิดโอกาสในการตอบข้อสอบข้อแตกต่างกันระหว่างภาค สามารถบ่งชี้ได้ว่าข้อสอบนั้นทำหน้าที่ต่างกันแบบเอกรูป (Uniform DIF) จากผลการวิเคราะห์ พบว่า ความเป็นภาคเหนือ (NORTH) จะเสียเปรียบในข้อสอบข้อที่ 26 และจะได้เปรียบในข้อสอบข้อที่ 7 ความเป็นภาคใต้ (SOUTH) จะได้เปรียบในข้อสอบข้อที่ 11 และในภาพรวมภาคเหนือและภาคใต้จะได้เปรียบในการตอบวิชาภาษาไทย ภาคตะวันออกเฉียงเหนือจะเสียเปรียบในการตอบวิชาภาษาไทย

ตารางที่ 4-24 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย ด้วยวิธี MIMIC แบบ Nonuniform DIF จำแนกตามสถานที่ตั้งทางภูมิศาสตร์

Item	interaction	Estimate	SE	Z	P-value
Q01 ON	THAINORTH	-1.454	0.299	-4.860*	0.000
	THAINORTHEST	-1.161	0.261	-4.454*	0.000
	THAISOUTH	-1.559	0.296	-5.261*	0.000
Q02 ON	THAINORTH	1.397	0.490	2.853*	0.004
	THAINORTHEST	0.691	0.407	1.697	0.090
	THAISOUTH	1.559	0.504	3.094*	0.002
Q03 ON	THAINORTH	-0.116	0.412	-0.282	0.778
	THAINORTHEST	0.215	0.422	0.510	0.610
	THAISOUTH	-0.342	0.409	-0.835	0.404
Q04 ON	THAINORTH	3.148	0.885	3.557*	0.000
	THAINORTHEST	0.580	0.449	1.292	0.196
	THAISOUTH	3.122	0.829	3.764*	0.000
Q05 ON	THAINORTH	-2.701	0.722	-3.739*	0.000
	THAINORTHEST	-1.406	0.520	-2.704*	0.007
	THAISOUTH	-1.999	0.588	-3.398*	0.001

ตารางที่ 4-24 (ต่อ)

Item	interaction	Estimate	SE	Z	P-value
Q06 ON	THAINORTH	-0.289	0.458	-0.631	0.528
	THAINORTHEST	0.612	0.478	1.281*	0.200
	THAISOUTH	0.989	0.525	1.885	0.059
Q07 ON	THAINORTH	0.352	0.499	0.706	0.480
	THAINORTHEST	0.705	0.510	1.383	0.167
	THAISOUTH	0.425	0.491	0.866	0.387
Q08 ON	THAINORTH	0.540	0.432	1.251	0.211
	THAINORTHEST	0.183	0.419	0.437	0.662
	THAISOUTH	1.152	0.488	2.359*	0.018
Q12 ON	THAINORTH	-0.053	0.364	-0.145	0.884
	THAINORTHEST	0.369	0.380	0.970	0.332
	THAISOUTH	-0.165	0.359	-0.460	0.645
Q13 ON	THAINORTH	-0.382	0.499	-0.766	0.443
	THAINORTHEST	-0.155	0.486	-0.320	0.749
	THAISOUTH	-0.114	0.484	-0.236	0.813
Q14 ON	THAINORTH	-0.473	0.396	-1.195	0.232
	THAINORTHEST	-0.098	0.384	-0.254	0.799
	THAISOUTH	-0.154	0.382	-0.403	0.687
Q15 ON	THAINORTH	-0.950	0.425	-2.235*	0.025
	THAINORTHEST	-0.811	0.410	-1.979	0.048
	THAISOUTH	0.026	0.379	0.069	0.945
Q16 ON	THAINORTH	0.275	0.433	0.636	0.525
	THAINORTHEST	0.103	0.425	0.242	0.809
	THAISOUTH	0.358	0.432	0.829	0.407

ตารางที่ 4-24 (ต่อ)

Item	interaction	Estimate	SE	Z	P-value
Q17 ON	THAINORTH	-0.980	0.497	-1.973*	0.048
	THAINORTHEST	-0.311	0.448	-0.695	0.487
	THAISOUTH	-0.563	0.455	-1.239	0.215
Q18 ON	THAINORTH	-0.115	0.414	-0.277	0.782
	THAINORTHEST	-0.414	0.420	-0.984	0.325
	THAISOUTH	0.537	0.432	1.243	0.214
Q19 ON	THAINORTH	-0.021	0.429	-0.048	0.961
	THAINORTHEST	-0.285	0.423	-0.673	0.501
	THAISOUTH	-0.618	0.425	-1.455	0.146
Q20 ON	THAINORTH	-1.620	0.526	-3.078*	0.002
	THAINORTHEST	-0.992	0.431	-2.302*	0.021
	THAISOUTH	-1.913	0.551	-3.470*	0.001
Q21 ON	THAINORTH	2.316	0.653	3.544*	0.000
	THAINORTHEST	0.939	0.434	2.163*	0.031
	THAISOUTH	1.977	0.567	3.486*	0.000
Q22 ON	THAINORTH	-3.141	0.855	-3.675*	0.000
	THAINORTHEST	-1.187	0.588	-2.017*	0.044
	THAISOUTH	-2.619	0.754	-3.473*	0.001
Q23 ON	THAINORTH	3.261	0.948	3.442*	0.001
	THAINORTHEST	0.978	0.529	1.851	0.064
	THAISOUTH	2.148	0.709	3.029*	0.002
Q24 ON	THAINORTH	2.213	0.669	3.305*	0.001
	THAINORTHEST	0.832	0.460	1.807	0.071
	THAISOUTH	1.855	0.597	3.107*	0.002

ตารางที่ 4-24 (ต่อ)

Item	interaction	Estimate	SE	Est./SE	P-value
Q25 ON	THAINORTH	-1.374	0.527	-2.608*	0.009
	THAINORTHEST	0.103	0.454	0.226	0.821
	THAISOUTH	-1.383	0.508	-2.723*	0.006
Q26 ON	THAINORTH	-0.373	0.619	-0.602	0.547
	THAINORTHEST	-0.620	0.538	-1.152	0.249
	THAISOUTH	-0.289	0.603	-0.478	0.632
Q27 ON	THAINORTH	-1.075	0.441	-2.439*	0.015
	THAINORTHEST	-0.453	0.386	-1.172	0.241
	THAISOUTH	-1.144	0.440	-2.600*	0.009
Q28 ON	THAINORTH	0.559	0.400	1.399	0.162
	THAINORTHEST	0.802	0.425	1.886	0.059
	THAISOUTH	1.231	0.470	2.620*	0.009
Q29 ON	THAINORTH	0.035	0.444	0.079	0.937
	THAINORTHEST	0.552	0.453	1.219	0.223
	THAISOUTH	-0.512	0.460	-1.113	0.266
Q30 ON	THAINORTH	1.710	0.547	3.125*	0.002
	THAINORTHEST	1.375	0.497	2.766*	0.006
	THAISOUTH	2.282	0.625	3.655*	0.000
Q31 ON	THAINORTH	-1.817	0.572	-3.178*	0.001
	THAINORTHEST	-0.669	0.433	-1.546	0.122
	THAISOUTH	-1.204	0.468	-2.575*	0.010
Q32 ON	THAINORTH	2.399	0.677	3.544*	0.000
	THAINORTHEST	1.516	0.510	2.972*	0.003
	THAISOUTH	2.235	0.630	3.549*	0.000

ตารางที่ 4-24 (ต่อ)

Item	interaction	Estimate	SE	Est./SE	P-value
Q33 ON	THAINORTH	-2.351	0.649	-3.626*	0.000
	THAINORTHEST	-1.511	0.508	-2.972*	0.003
	THAISOUTH	-1.884	0.556	-3.390*	0.001
Q34 ON	THAINORTH	0.692	0.451	1.534	0.125
	THAINORTHEST	0.723	0.456	1.587	0.113
	THAISOUTH	1.297	0.515	2.519*	0.012
Q35 ON	THAINORTH	-1.130	0.492	-2.297*	0.022
	THAINORTHEST	0.264	0.460	0.574	0.566
	THAISOUTH	-0.060	0.439	-0.136	0.892
Q36 ON	THAINORTH	0.970	0.424	2.286*	0.022
	THAINORTHEST	0.439	0.382	1.151	0.250
	THAISOUTH	0.616	0.383	1.606	0.108
Q37 ON	THAINORTH	-0.542	0.446	-1.214	0.225
	THAINORTHEST	0.126	0.434	0.291	0.771
	THAISOUTH	0.725	0.472	1.538	0.124
Q38 ON	THAINORTH	-1.067	0.441	-2.417*	0.016
	THAINORTHEST	-0.791	0.411	-1.923	0.054
	THAISOUTH	-0.399	0.382	-1.044	0.296
Q39 ON	THAINORTH	-1.012	0.461	-2.193*	0.028
	THAINORTHEST	-1.334	0.498	-2.678*	0.007
	THAISOUTH	-0.573	0.421	-1.360	0.174
Q40 ON	THAINORTH	0.530	0.419	1.265	0.206
	THAINORTHEST	0.039	0.410	0.095	0.924
	THAISOUTH	1.197	0.465	2.576*	0.010

หมายเหตุ * p < .05; ข้อสอบ DIF คือข้อที่อย่างน้อย 1 ภูมิภาคที่มีปฏิสัมพันธ์กับ θ แล้วมีอิทธิพลทางตรงต่อข้อสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ .0

จากตารางที่ 4-24 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย สำหรับกลุ่มที่มากกว่าสองกลุ่ม ด้วยวิธี MIMIC แบบอเนกรูป Nonuniform DIF จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ ผลการวิเคราะห์ปฏิสัมพันธ์ระหว่างความสามารถด้านภาษาไทยหรือตัวแปรแฝงภายใน (θ) กับความเป็นสมาชิกของกลุ่ม สถานที่ตั้งทางภูมิศาสตร์ 4 ภูมิภาค ประกอบด้วย ภาคกลาง ภาคเหนือ ภาคตะวันออกเฉียงเหนือ และภาคใต้ ที่มีอิทธิพลทางตรงต่อข้อสอบอย่างมีนัยสำคัญที่ระดับ .05 ถ้าพบว่า มีปฏิสัมพันธ์ตั้งแต่ 1 ปฏิสัมพันธ์ สามารถบ่งชี้ได้ว่าข้อสอบนั้น ทำหน้าที่ต่างกันแบบ Nonuniform DIF ได้แก่ข้อสอบ ข้อที่ 1, 2, 4, 5, 15, 17, 20, 21, 22, 23, 24, 25, 27, 28, 30, 31, 32, 33, 34, 35, 36, 38, 39 และข้อที่ 40 หมายความว่า ปฏิสัมพันธ์ระหว่างความสามารถด้านภาษาไทย (θ) กับความเป็นสมาชิกของกลุ่มสถานที่ตั้งทางภูมิศาสตร์ส่งผลให้เกิดโอกาสในการตอบข้อสอบแตกต่างกันระหว่างคู่ปฏิสัมพันธ์

ตารางที่ 4-25 ความกลมกลืนของโมเดล CFA MIMIC วิชาวิทยาศาสตร์ จำแนกตามสถานที่ตั้งทางภูมิศาสตร์

Statistics	**CFA**	**MIMIC (no Direct Effect) @0**	**MIMIC (1 Direct Effect)**	**MIMIC (2 Direct Effect)**
χ^2	3420.963 (df = 740)	3630.835 (df = 857)	3616.785 (df = 856)	3606.730 (df = 855)
RMSEA	0.021	0.020	0.020	0.020
CFI	0.481	0.488	0.490	0.492
TLI	0.453	0.462	0.464	0.465

จากตารางที่ 4-25 ผลการวิเคราะห์ความกลมกลืนของโมเดล CFA MIMIC วิชาวิทยาศาสตร์ จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ จากการวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) ซึ่งพิจารณาจากค่าไค-สแควร์ (Chi-square) มีค่าเท่ากับ 3420.963 ที่องศาอิสระ (df) เท่ากับ 740 ดัชนีวัดรากที่สองของความคลาดเคลื่อนในการประมาณค่า พารามิเตอร์ที่เหมาะสม (RMSEA) มีค่าเท่ากับ 0.021 ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (CFI) มีค่าเท่ากับ 0.481 ดัชนี The tucker lewis index (TLI) มีค่าเท่ากับ 0.453

เมื่อพิจารณาการทำนายตัวแปรข้อสอบจากตัวแปรสถานที่ตั้งทางภูมิศาสตร์ ที่มีดัชนีปรับแต่งมีข้อค้นพบดังนี้

1. เมื่อพิจารณาการทำนายตัวแปรข้อสอบจากสถานที่ตั้งทางภูมิศาสตร์ ที่มีดัชนีปรับแต่งสามารถทำนายตัวแปรข้อสอบที่มีดัชนีปรับแต่ง จำนวน 2 ข้อ โดยโมเดล MIMIC เริ่มต้น เป็นการเพิ่มตัวแปรทำนายที่มีอิทธิพลทางตรงต่อข้อสอบและบังคับค่าให้เป็น 0 เพื่อกำหนดให้ตัวแปรต้นไม่มีผลทางตรงต่อข้อสอบ มีค่าไค-สแควร์ (Chi-square) เท่ากับ 3630.835 ที่องศาอิสระ (df) เท่ากับ 857 ดัชนีวัดรากที่สองของความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ที่เหมาะสม (RMSEA) มีค่าเท่ากับ 0.020 ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (CFI) มีค่าเท่ากับ 0.488 ดัชนี The tucker lewis index (TLI) มีค่าเท่ากับ 0.462

2. เมื่อพิจารณาโมเดล MIMIC ที่เพิ่มอิทธิพลทางตรงจากตัวทำนายไปที่ข้อสอบที่มีดัชนีปรับแก้ 1 ข้อ โดยเลือกข้อที่มีดัชนีปรับแก้สูงสุด มีค่าไค-สแควร์ (Chi-square) เท่ากับ 3616.785 ที่องศาอิสระ (df) เท่ากับ 857 ดัชนีวัดรากที่สองของความคลาดเคลื่อนในการประมาณค่า พารามิเตอร์ที่เหมาะสม (RMSEA) มีค่าเท่ากับ 0.020 ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (CFI) มีค่าเท่ากับ 0.490 ดัชนี The tucker lewis index (TLI) มีค่าเท่ากับ 0.464

3. เมื่อพิจารณาโมเดล MIMIC ที่เพิ่มอิทธิพลทางตรงจากตัวทำนายไปที่ข้อสอบที่มีดัชนีปรับแก้ 2 ข้อ โดยเลือกข้อที่มีดัชนีปรับแก้สูงสุด มีค่าไค-สแควร์ (Chi-square) เท่ากับ 3606.730 ที่องศาอิสระ (df) เท่ากับ 857 ดัชนีวัดรากที่สองของความคลาดเคลื่อนในการประมาณค่า พารามิเตอร์ที่เหมาะสม (RMSEA) มีค่าเท่ากับ 0.020 ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (CFI) มีค่าเท่ากับ 0.492 ดัชนี The tucker lewis index (TLI) มีค่าเท่ากับ 0.465

ตารางที่ 4-26 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาวิทยาศาสตร์ด้วยวิธี MIMIC แบบ Uniform DIF จำแนกตามสถานที่ตั้งทางภูมิศาสตร์

Item		Estimate	SE	Z	P-value
SCIENCE ON	NORTH	0.098	0.024	4.079*	0.000
	NORTHEST	-0.164	0.030	-5.407*	0.000
	SOUTH	0.098	0.021	4.595*	0.000
Q01 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q02 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q03 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q04 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q05 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q06 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q07 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000

ตารางที่ 4-26 (ต่อ)

Item		Estimate	SE	Z	P-value
Q08 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q09 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q10 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q11 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q12 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q13 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q14 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q15 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q16 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000

ตารางที่ 4-26 (ต่อ)

Item		Estimate	SE	Z	P-value
Q17 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q18 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.219	0.053	4.107*	0.000
Q19 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q20 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q21 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q22 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q23 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q24 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q25 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000

ตารางที่ 4-26 (ต่อ)

Item		Estimate	SE	Z	P-value
Q26 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q27 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q28 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q29 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q30 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q31 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q32 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q33 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q34 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000

ตารางที่ 4-26 (ต่อ)

Item		Estimate	SE	Z	P-value
Q35 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q36 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q37 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q37 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q38 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q39 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000
Q40 ON	NORTH	0.000	0.000	999.000	999.000
	NORTHEST	0.000	0.000	999.000	999.000
	SOUTH	0.000	0.000	999.000	999.000

หมายเหตุ * $p < .05$; ข้อสอบ DIF คือข้อที่อย่างน้อย 1 ภูมิภาคมีอิทธิพลทางตรงต่อข้อสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

จากตารางที่ 4-26 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาวิทยาศาสตร์ สำหรับกลุ่มที่มากกว่าสองกลุ่ม ด้วยวิธี MIMIC แบบเอกรูป (Uniform DIF) จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ ผลการวิเคราะห์ข้อสอบจำนวน 40 ข้อ ตัวแปรทำนายซึ่งได้แก่ สถานที่ตั้งทาง

ภูมิศาสตร์ (ภาคกลาง ภาคเหนือ ภาคตะวันออกเฉียงเหนือ และภาคใต้) ที่มีอิทธิพลทางตรงต่อข้อสอบ ซึ่งข้อสอบที่มีดัชนีปรับแต่งมี 2 ข้อ สูงสุด คือ ข้อที่ 18 รองลงมาคือ ข้อที่ 1ที่ทำให้ตัวแปรสถานที่ตั้งทางภูมิศาสตร์มีอิทธิพลทางตรงต่อข้อสอบที่มีดัชนีปรับแต่งมีค่าต่างจากศูนย์อย่างมีนัยสำคัญที่ระดับ .05 ซึ่งหมายความว่า ในข้อที่ 18 และข้อที่ 1 ตัวแปรความเป็นสถานที่ตั้งทางภูมิศาสตร์ ส่งผลให้เกิดโอกาสในการตอบข้อสอบข้อแตกต่างกันระหว่างภาค สามารถบ่งชี้ได้ว่าข้อสอบนั้นทำหน้าที่ต่างกัน จากผลการวิเคราะห์ พบว่า ความเป็นภาคใต้ (SOUTH) จะได้เปรียบในข้อสอบข้อที่ 18 และความเป็นภาคเหนือ (NORTH) จะได้เปรียบในข้อสอบข้อที่ 1 และในภาพรวมภาคเหนือและภาคใต้จะได้เปรียบในการตอบวิชาวิทยาศาสตร์ ส่วนภาคตะวันออกเฉียงเหนือจะเสียเปรียบในการตอบวิชาวิทยาศาสตร์

ตารางที่ 4-27 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิธี MIMIC แบบ Nonuniform DIF วิชาวิทยาศาสตร์ จำแนกตามสถานที่ตั้งทางภูมิศาสตร์

Item	Interaction	Estimate	SE	Z	P-value
Q01 ON	SCIENCENORTH	0.066	0.168	0.393	0.694
	SCIENCENORTHEST	0.012	0.173	0.069	0.945
	SCIENCESOUTH	0.435	0.205	2.129	0.033
Q02 ON	SCIENCENORTH	-0.098	0.164	-0.600	0.549
	SCIENCENORTHEST	0.060	0.179	0.336	0.737
	SCIENCESOUTH	0.264	0.171	1.542	0.123
Q03 ON	SCIENCENORTH	-0.388	0.151	-2.571*	0.010
	SCIENCENORTHEST	-0.119	0.148	-0.806	0.420
	SCIENCESOUTH	-0.082	0.141	-0.584	0.559
Q04 ON	SCIENCENORTH	-0.145	0.159	-0.907	0.364
	SCIENCENORTHEST	-0.219	0.162	-1.350	0.177
	SCIENCESOUTH	0.060	0.159	0.377	0.706
Q05 ON	SCIENCENORTH	0.062	0.152	0.407	0.684
	SCIENCENORTHEST	0.023	0.154	0.152	0.879
	SCIENCESOUTH	0.297	0.155	1.921	0.055

ตารางที่ 4-27 (ต่อ)

Item	Interaction	Estimate	SE	Z	P-value
Q06 ON	SCIENCENORTH	0.395	0.163	2.430*	0.015
	SCIENCENORTHEST	-0.114	0.165	-0.692	0.489
	SCIENCESOUTH	0.526	0.164	3.203*	0.001
Q07 ON	SCIENCENORTH	0.049	0.158	0.312	0.755
	SCIENCENORTHEST	-0.403	0.172	-2.340*	0.019
	SCIENCESOUTH	0.335	0.162	2.064*	0.039
Q08 ON	SCIENCENORTH	0.125	0.172	0.729	0.466
	SCIENCENORTHEST	-0.024	0.184	-0.130	0.897
	SCIENCESOUTH	0.410	0.180	2.281*	0.023
Q09 ON	SCIENCENORTH	0.009	0.195	0.045	0.964
	SCIENCENORTHEST	-0.572	0.210	-2.724*	0.006
	SCIENCESOUTH	0.471	0.213	2.211*	0.027
Q10 ON	SCIENCENORTH	0.103	0.150	0.690	0.490
	SCIENCENORTHEST	0.189	0.168	1.123	0.262
	SCIENCESOUTH	0.081	0.144	0.563	0.573
Q11 ON	SCIENCENORTH	-0.412	0.161	-2.562*	0.010
	SCIENCENORTHEST	-0.081	0.173	-0.471	0.638
	SCIENCESOUTH	0.126	0.153	0.819	0.413
Q12 ON	SCIENCENORTH	0.188	0.160	1.174	0.241
	SCIENCENORTHEST	0.375	0.178	2.110*	0.035
	SCIENCESOUTH	0.439	0.168	2.617*	0.009
Q13 ON	SCIENCENORTH	-0.058	0.174	-0.332	0.740
	SCIENCENORTHEST	-0.249	0.177	-1.408	0.159
	SCIENCESOUTH	-0.149	0.168	-0.887	0.375

ตารางที่ 4-27 (ต่อ)

Item	Interaction	Estimate	SE	Z	P-value
Q14 ON	SCIENCENORTH	0.030	0.153	0.195	0.845
	SCIENCENORTHEST	-0.084	0.162	-0.518	0.605
	SCIENCESOUTH	0.130	0.150	0.863	0.388
Q15 ON	SCIENCENORTH	-0.162	0.153	-1.055	0.291
	SCIENCENORTHEST	-0.176	0.163	-1.078	0.281
	SCIENCESOUTH	0.193	0.154	1.253	0.210
Q16 ON	SCIENCENORTH	0.035	0.161	0.215	0.830
	SCIENCENORTHEST	-0.423	0.177	-2.384*	0.017
	SCIENCESOUTH	0.746	0.193	3.867*	0.000
Q17 ON	SCIENCENORTH	-0.363	0.177	-2.056*	0.040
	SCIENCENORTHEST	-0.841	0.208	-4.052*	0.000
	SCIENCESOUTH	0.253	0.182	1.388	0.165
Q18 ON	SCIENCENORTH	0.058	0.144	0.406	0.685
	SCIENCENORTHEST	0.211	0.152	1.387	0.166
	SCIENCESOUTH	0.069	0.141	0.488	0.626
Q19 ON	SCIENCENORTH	-0.071	0.150	-0.474	0.636
	SCIENCENORTHEST	-0.153	0.152	-1.009	0.313
	SCIENCESOUTH	-0.059	0.145	-0.409	0.682
Q20 ON	SCIENCENORTH	0.220	0.153	1.432	0.152
	SCIENCENORTHEST	0.147	0.162	0.911	0.363
	SCIENCESOUTH	0.002	0.146	0.011	0.991
Q21 ON	SCIENCENORTH	-0.123	0.151	-0.812	0.417
	SCIENCENORTHEST	-0.119	0.160	-0.744	0.457
	SCIENCESOUTH	0.060	0.145	0.417	0.677

ตารางที่ 4-27 (ต่อ)

Item	Interaction	Estimate	SE	Z	P-value
Q22 ON	SCIENCENORTH	-0.024	0.169	-0.139	0.889
	SCIENCENORTHEST	-0.275	0.179	-1.533	0.125
	SCIENCESOUTH	0.562	0.173	3.252*	0.001
Q23 ON	SCIENCENORTH	-0.457	0.177	-2.582*	0.010
	SCIENCENORTHEST	-1.071	0.217	-4.939*	0.000
	SCIENCESOUTH	0.186	0.159	1.170	0.242
Q24 ON	SCIENCENORTH	-0.159	0.179	-0.889	0.374
	SCIENCENORTHEST	-0.057	0.190	-0.300	0.764
	SCIENCESOUTH	0.034	0.179	0.191	0.849
Q25 ON	SCIENCENORTH	-0.321	0.195	-1.649	0.099
	SCIENCENORTHEST	-0.674	0.203	-3.315*	0.001
	SCIENCESOUTH	-0.339	0.190	-1.781	0.075
Q26 ON	SCIENCENORTH	-0.094	0.185	-0.507	0.612
	SCIENCENORTHEST	-0.096	0.191	-0.503	0.615
	SCIENCESOUTH	0.638	0.183	3.477*	0.001
Q27 ON	SCIENCENORTH	-0.281	0.181	-1.555	0.120
	SCIENCENORTHEST	-0.622	0.198	-3.148*	0.002
	SCIENCESOUTH	0.432	0.172	2.514*	0.012
Q28 ON	SCIENCENORTH	-0.337	0.207	-1.632	0.103
	SCIENCENORTHEST	-0.073	0.188	-0.389	0.697
	SCIENCESOUTH	0.496	0.191	2.593*	0.010
Q29 ON	SCIENCENORTH	-0.135	0.175	-0.771	0.440
	SCIENCENORTHEST	-0.662	0.201	-3.290*	0.001
	SCIENCESOUTH	0.388	0.187	2.070*	0.038

ตารางที่ 4-27 (ต่อ)

Item	Interaction	Estimate	SE	Z	P-value
Q30 ON	SCIENCENORTH	0.177	0.167	1.059	0.290
	SCIENCENORTHEST	-0.127	0.180	-0.705	0.481
	SCIENCESOUTH	0.534	0.180	2.974*	0.003
Q31 ON	SCIENCENORTH	0.215	0.184	1.167	0.243
	SCIENCENORTHEST	-0.156	0.185	-0.845	0.398
	SCIENCESOUTH	0.305	0.183	1.663	0.096
Q32 ON	SCIENCENORTH	-0.173	0.164	-1.053	0.292
	SCIENCENORTHEST	-0.617	0.187	-3.301*	0.001
	SCIENCESOUTH	-0.111	0.159	-0.697	0.486
Q33 ON	SCIENCENORTH	0.092	0.153	0.605	0.545
	SCIENCENORTHEST	-0.149	0.165	-0.905	0.366
	SCIENCESOUTH	0.337	0.155	2.177*	0.030
Q34 ON	SCIENCENORTH	-0.587	0.187	-3.147*	0.002
	SCIENCENORTHEST	-0.869	0.212	-4.096*	0.000
	SCIENCESOUTH	0.203	0.163	1.252	0.211
Q35 ON	SCIENCENORTH	-0.091	0.176	-0.514	0.607
	SCIENCENORTHEST	-0.466	0.194	-2.398*	0.016
	SCIENCESOUTH	0.735	0.186	3.947*	0.000
Q36 ON	SCIENCENORTH	-0.217	0.153	-1.420	0.156
	SCIENCENORTHEST	-0.546	0.170	-3.215*	0.001
	SCIENCESOUTH	-0.036	0.146	-0.248	0.804
Q37 ON	SCIENCENORTH	-0.087	0.148	-0.584	0.559
	SCIENCENORTHEST	-0.242	0.158	-1.531	0.126
	SCIENCESOUTH	0.468	0.158	2.968*	0.003

ตารางที่ 4-27 (ต่อ)

Item	Interaction	Estimate	SE	Z	P-value
Q38 ON	SCIENCENORTH	0.198	0.175	1.134	0.257
	SCIENCENORTHEST	-0.317	0.185	-1.707	0.088
	SCIENCESOUTH	0.444	0.172	2.581*	0.010
Q39 ON	SCIENCENORTH	-0.157	0.152	-1.034	0.301
	SCIENCENORTHEST	-0.008	0.167	-0.048	0.962
	SCIENCESOUTH	0.189	0.150	1.254	0.210
Q40 ON	SCIENCENORTH	-0.207	0.157	-1.324	0.185
	SCIENCENORTHEST	-0.101	0.173	-0.586	0.558
	SCIENCESOUTH	-0.050	0.151	-0.335	0.738

หมายเหตุ * $p < .05$; ข้อสอบ DIF คือข้อที่อย่างน้อย 1 ภูมิภาคที่มีปฏิสัมพันธ์กับ θ แล้วมีอิทธิพลทางตรงต่อข้อสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

จากตารางที่ 4-27 ผลการวิเคราะห์การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาวิทยาศาสตร์ สำหรับกลุ่มที่มากกว่าสองกลุ่ม ด้วยวิธี MIMIC แบบ Nonuniform DIF จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ ผลการวิเคราะห์ปฏิสัมพันธ์ระหว่างความสามารถด้านคณิตศาสตร์หรือตัวแปรแฝงภายใน (θ) กับความเป็นสมาชิกของกลุ่ม สถานที่ตั้งทางภูมิศาสตร์ 4 ภูมิภาค ประกอบด้วย ภาคกลาง ภาคเหนือ ภาคตะวันออกเฉียงเหนือ และภาคใต้ ที่มีอิทธิพลทางตรงต่อข้อสอบอย่างมีนัยสำคัญที่ระดับ .05 ถ้าพบว่า มีปฏิสัมพันธ์ตั้งแต่ 1 ปฏิสัมพันธ์ สามารถบ่งชี้ได้ว่าข้อสอบนั้นทำหน้าที่ต่างกันแบบ Nonuniform DIF ได้แก่ข้อสอบข้อที่ 1, 3, 6, 7, 8, 9, 11, 12, 16, 17, 22, 23, 25, 26, 27, 28, 29, 30, 32, 33, 34, 35, 36, 37 และข้อที่ 38 หมายความว่า ปฏิสัมพันธ์ระหว่างความสามารถด้านคณิตศาสตร์ (θ) กับความเป็นสมาชิกของกลุ่มสถานที่ตั้งทางภูมิศาสตร์ส่งผลให้เกิดโอกาสในการตอบข้อสอบแตกต่างกันระหว่างคู่ปฏิสัมพันธ์

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ วิชาภาษาไทย และวิชาวิทยาศาสตร์ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี วิธี MIMIC ภายใต้ขนาดโรงเรียน โดยกำหนดให้กลุ่มอ้างอิง (R) เป็นโรงเรียนขนาดใหญ่พิเศษ กลุ่มเปรียบเทียบ (F) เป็นโรงเรียนขนาดใหญ่ ขนาดกลาง และขนาดเล็ก รายละเอียดแสดงในตารางที่ 4-28 ถึง ตารางที่ 4-36 ดังนี้

ตารางที่ 4-28 ความกลมกลืนของโมเดล CFA MIMIC วิชาคณิตศาสตร์ จำแนกตามขนาดโรงเรียน

Statistics	CFA	MIMIC (no Direct Effect) @0	MIMIC (1 Direct Effect)	MIMIC (2 Direct Effect)	MIMIC (3 Direct Effect)
χ^2	1833.638 (df = 275)	1958.064 (df = 347)	1940.241 (df = 346)	1927.513 (df = 345)	1916.043 (df = 344)
RMSEA	0.027	0.024	0.024	0.024	0.024
CFI	0.705	0.714	0.717	0.719	0.721
TLI	0.678	0.691	0.694	0.695	0.696

จากตารางที่ 4-28 ผลการวิเคราะห์ความกลมกลืนของโมเดล CFA MIMIC วิชาคณิตศาสตร์ จำแนกตามขนาดโรงเรียน จากการวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) ซึ่งพิจารณาจากค่าไค-สแควร์ (Chi-square) มีค่าเท่ากับ 1833.638 ที่องศาอิสระ (df) เท่ากับ 275 ดัชนีวัดราก ที่สองของความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ที่เหมาะสม (RMSEA) มีค่าเท่ากับ 0.027 ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (CFI) มีค่าเท่ากับ 0.705 ดัชนี The tucker lewis index (TLI) มีค่าเท่ากับ 0.678

เมื่อพิจารณาการทำนายตัวแปรข้อสอบจากตัวแปรขนาดโรงเรียน ที่มีดัชนีปรับแต่งมีข้อค้นพบ ดังนี้

1. เมื่อพิจารณาการทำนายตัวแปรข้อสอบจากขนาดโรงเรียนที่มีดัชนีปรับแต่งพบว่าสามารถทำนายตัวแปรข้อสอบที่มีดัชนีปรับแต่ง จำนวน 3 ข้อ โดยโมเดล MIMIC เริ่มต้นเป็นการเพิ่มตัวแปรทำนายที่มีอิทธิพลทางตรงต่อข้อสอบและบังคับค่าให้เป็น 0 เพื่อกำหนดให้ตัวแปรต้นไม่มีผลทางตรงต่อข้อสอบ มีค่าไค-สแควร์ (Chi-square) เท่ากับ 1958.064 ที่องศาอิสระ (df) เท่ากับ 347 ดัชนีวัดรากที่สองของความคลาดเคลื่อนในการประมาณค่า พารามิเตอร์ที่เหมาะสม

(RMSEA) มีค่าเท่ากับ 0.024 ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (CFI) มีค่าเท่ากับ 0.714 ดัชนี The tucker lewis index (TLI) มีค่าเท่ากับ 0.691

2. เมื่อพิจารณาโมเดล MIMIC ที่เพิ่มอิทธิพลทางตรงจากตัวทำนายไปที่ข้อสอบ ที่มีดัชนีปรับแก้ 1 ข้อ โดยเลือกข้อที่มีดัชนีปรับแก้สูงสุด มีค่าไค-สแควร์ (Chi-square) เท่ากับ 1940.241 ที่องศาอิสระ (df) เท่ากับ 345 ดัชนีวัดรากที่สองของความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ที่เหมาะสม (RMSEA) มีค่าเท่ากับ 0.024 ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (CFI) มีค่าเท่ากับ 0.717 ดัชนี The tucker lewis index (TLI) มีค่าเท่ากับ 0.694

3. เมื่อพิจารณาโมเดล MIMIC ที่เพิ่มอิทธิพลทางตรงจากตัวทำนายไปที่ข้อสอบ ที่มีดัชนีปรับแก้ 2 ข้อมีค่าไค-สแควร์ (Chi-square) เท่ากับ 1927.513ที่องศาอิสระ (df) เท่ากับ 345 ดัชนีวัดรากที่สองของความคลาดเคลื่อนในการประมาณค่า พารามิเตอร์ที่เหมาะสม (RMSEA) มีค่าเท่ากับ 0.024 ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (CFI) มีค่าเท่ากับ 0.719 ดัชนี The tucker lewis index (TLI) มีค่าเท่ากับ 0.695

4. เมื่อพิจารณาโมเดล MIMIC ที่เพิ่มอิทธิพลทางตรงจากตัวทำนายไปที่ข้อสอบ ที่มีดัชนีปรับแก้ 3 ข้อ มีค่าไค-สแควร์ (Chi-square) เท่ากับ 1916.043ที่องศาอิสระ (df) เท่ากับ 344 ดัชนีวัดรากที่สองของความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ที่เหมาะสม (RMSEA) มีค่าเท่ากับ 0.024 ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (CFI) มีค่าเท่ากับ 0.721 ดัชนี The tucker lewis index (TLI) มีค่าเท่ากับ 0.696

ตารางที่ 4-29 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ วิธี MIMIC แบบ Uniform DIF จำแนกตามขนาดโรงเรียน

Item		Estimate	SE	Z	P-value
MATH ON	SMALL	-0.915	0.105	-8.720*	0.000
	MEDIUM	-0.175	0.028	-6.237*	0.000
	LARGE	-0.518	0.062	-8.330*	0.000
Q01 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.132	0.063	2.081*	0.037
	LARGE	0.000	0.000	999.000	999.000
Q02 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q03 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q04 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	-0.170	0.084	-2.021*	0.043
Q05 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q06 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q07 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000

ตารางที่ 4-29 (ต่อ)

Item		Estimate	SE	Z	P-value
Q08 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q09 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q10 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q11 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q12 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q13 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q14 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q15 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q16 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000

ตารางที่ 4-29 (ต่อ)

Item		Estimate	SE	Z	P-value
Q17 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q18 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q19 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q20 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q21 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q22 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q23 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q24 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q25 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000

หมายเหตุ * p < .05; ข้อสอบ DIF คือข้อที่อย่างน้อย 1 ขนาดโรงเรียนมีอิทธิพลทางตรงต่อ
ข้อสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

จากตารางที่ 4-29 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มที่มากกว่าสองกลุ่ม ด้วยวิธี MIMIC แบบเอกรูป (Uniform DIF) วิชาคณิตศาสตร์ จำแนกตามขนาดโรงเรียน ผลการวิเคราะห์ข้อสอบจำนวน 25 ข้อ ตัวแปรทำนายซึ่งได้แก่ ขนาดโรงเรียน (ขนาดเล็ก ขนาดกลาง ขนาดใหญ่) ที่มีอิทธิพลทางตรงต่อข้อสอบ ซึ่งข้อสอบที่มีดัชนีปรับแต่งสูงสุดคือ ข้อที่ 4 รองลงมาคือ ข้อที่ 1 ที่ทำให้ตัวแปรขนาดโรงเรียนมีอิทธิพลทางตรงต่อข้อสอบที่มีดัชนีปรับแต่งมีค่าต่างจากศูนย์อย่างมีนัยสำคัญที่ระดับ .05 ซึ่งหมายความว่า ในข้อที่ 4 และ ข้อที่ 1 ตัวแปรขนาดโรงเรียน ส่งผลให้เกิดโอกาสในการตอบข้อสอบแตกต่างกันระหว่างขนาดโรงเรียน สามารถบ่งชี้ได้ว่าข้อสอบนั้นทำหน้าที่ต่างกัน จากผลการวิเคราะห์ พบว่า โรงเรียนขนาดใหญ่ (LARGE) จะเสียเปรียบในข้อสอบข้อที่ 4 และโรงเรียนขนาดกลาง (MEDIUM) จะได้เปรียบในข้อสอบข้อที่ 1 และในภาพรวมโรงเรียนทุกขนาดจะเสียเปรียบในการตอบวิชาคณิตศาสตร์

ตารางที่ 4-30 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ ด้วยวิธี MIMIC แบบ Nonuniform DIF จำแนกตามขนาดโรงเรียน

Item	Interaction	Estimate	SE	Z	P-value
Q01 ON	MATHSMALL	-0.415	0.094	-4.401*	0.000
	MATHMEDIUM	0.008	0.149	0.054	0.957
	MATHLARGE	-0.329	0.119	-2.768*	0.006
Q02 ON	MATHSMALL	-1.682	0.249	-6.755*	0.000
	MATHMEDIUM	-0.561	0.210	-2.665*	0.008
	MATHLARGE	-1.155	0.241	-4.795*	0.000
Q03 ON	MATHSMALL	-2.361	0.299	-7.886*	0.000
	MATHMEDIUM	-0.831	0.230	-3.611*	0.000
	MATHLARGE	-1.770	0.275	-6.435*	0.000
Q04 ON	MATHSMALL	-1.724	0.227	-7.601*	0.000
	MATHMEDIUM	-0.938	0.196	-4.778*	0.000
	MATHLARGE	-1.303	0.223	-5.847*	0.000
Q05 ON	MATHSMALL	-1.752	0.233	-7.516*	0.000
	MATHMEDIUM	-0.790	0.200	-3.953*	0.000
	MATHLARGE	-1.495	0.226	-6.622*	0.000

ตารางที่ 4-30 (ต่อ)

Item	Interaction	Estimate	SE	Z	P-value
Q06 ON	MATHSMALL	-1.115	0.171	-6.506*	0.000
	MATHMEDIUM	-0.508	0.165	-3.085*	0.002
	MATHLARGE	-0.786	0.167	-4.696*	0.000
Q07 ON	MATHSMALL	-1.271	0.192	-6.610*	0.000
	MATHMEDIUM	-0.463	0.180	-2.569*	0.010
	MATHLARGE	-0.822	0.184	-4.464*	0.000
Q08 ON	MATHSMALL	-0.887	0.153	-5.792*	0.000
	MATHMEDIUM	-0.412	0.156	-2.636*	0.008
	MATHLARGE	-0.749	0.159	-4.721*	0.000
Q09 ON	MATHSMALL	-0.946	0.154	-6.153*	0.000
	MATHMEDIUM	-0.318	0.152	-2.092*	0.036
	MATHLARGE	-0.663	0.153	-4.337*	0.000
Q10 ON	MATHSMALL	-0.231	0.124	-1.860	0.063
	MATHMEDIUM	0.186	0.149	1.249	0.212
	MATHLARGE	-0.146	0.137	-1.068	0.285
Q11 ON	MATHSMALL	-1.272	0.213	-5.961*	0.000
	MATHMEDIUM	-0.254	0.207	-1.225	0.221
	MATHLARGE	-0.883	0.201	-4.394*	0.000
Q12 ON	MATHSMALL	-1.245	0.215	-5.793*	0.000
	MATHMEDIUM	-0.103	0.214	-0.481	0.630
	MATHLARGE	-0.992	0.207	-4.781*	0.000
Q13 ON	MATHSMALL	-1.970	0.273	-7.229*	0.000
	MATHMEDIUM	-0.392	0.243	-1.614	0.106
	MATHLARGE	-1.494	0.250	-5.976*	0.000
Q14 ON	MATHSMALL	-1.660	0.245	-6.773*	0.000
	MATHMEDIUM	-0.329	0.228	-1.443	0.149
	MATHLARGE	-1.372	0.234	-5.856*	0.000

ตารางที่ 4-30 (ต่อ)

Item	Interaction	Estimate	SE	Z	P-value
Q15 ON	MATHSMALL	-1.292	0.200	-6.470*	0.000
	MATHMEDIUM	-0.458	0.192	-2.387*	0.017
	MATHLARGE	-1.088	0.196	-5.538*	0.000
Q16 ON	MATHSMALL	-1.350	0.201	-6.728*	0.000
	MATHMEDIUM	-0.381	0.190	-2.006*	0.045
	MATHLARGE	-1.099	0.195	-5.634*	0.000
Q17 ON	MATHSMALL	-1.005	0.188	-5.343*	0.000
	MATHMEDIUM	0.211	0.203	1.043	0.297
	MATHLARGE	-0.761	0.184	-4.127*	0.000
Q18 ON	MATHSMALL	-1.145	0.208	-5.502*	0.000
	MATHMEDIUM	0.050	0.211	0.239	0.811
	MATHLARGE	-0.749	0.197	-3.795*	0.000
Q19 ON	MATHSMALL	-1.204	0.215	-5.590*	0.000
	MATHMEDIUM	0.108	0.218	0.496	0.620
	MATHLARGE	-0.724	0.202	-3.580*	0.000
Q20 ON	MATHSMALL	-1.895	0.267	-7.097*	0.000
	MATHMEDIUM	-0.953	0.240	-3.969*	0.000
	MATHLARGE	-1.665	0.261	-6.381*	0.000
Q21 ON	MATHSMALL	-1.978	0.275	-7.198*	0.000
	MATHMEDIUM	-0.905	0.245	-3.688*	0.000
	MATHLARGE	-1.567	0.260	-6.034*	0.000
Q22 ON	MATHSMALL	-0.890	0.157	-5.661*	0.000
	MATHMEDIUM	-0.344	0.159	-2.169*	0.030
	MATHLARGE	-0.765	0.164	-4.667*	0.000
Q23 ON	MATHSMALL	-0.903	0.159	-5.663*	0.000
	MATHMEDIUM	-0.210	0.160	-1.316	0.188
	MATHLARGE	-0.926	0.172	-5.380*	0.000

ตารางที่ 4-30 (ต่อ)

Item	Interaction	Estimate	SE	Z	P-value
Q24 ON	MATHSMALL	-1.290	0.195	-6.634*	0.000
	MATHMEDIUM	-0.633	0.183	-3.463*	0.001
	MATHLARGE	-1.044	0.193	-5.417*	0.000
Q25 ON	MATHSMALL	-0.873	0.163	-5.352*	0.000
	MATHMEDIUM	-0.092	0.169	-0.547	0.584
	MATHLARGE	-0.937	0.177	-5.310*	0.000

หมายเหตุ * $p < .05$; ข้อสอบ DIF คือข้อที่อย่างน้อย 1 ขนาดที่มีปฏิสัมพันธ์กับ θ แล้วมีอิทธิพลทางตรงต่อข้อสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

จากตารางที่ 4-30 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ สำหรับกลุ่มที่มากกว่าสองกลุ่ม ด้วยวิธี MIMIC แบบ Nonuniform DIF จำแนกตามขนาดโรงเรียน ผลการวิเคราะห์ปฏิสัมพันธ์ระหว่างความสามารถด้านคณิตศาสตร์หรือตัวแปรแฝงภายใน (θ) กับความเป็นสมาชิกของกลุ่ม ขนาดโรงเรียน 4 กลุ่ม ประกอบด้วย โรงเรียนขนาดใหญ่พิเศษ ขนาดใหญ่ ขนาดกลาง และขนาดเล็ก ที่มีอิทธิพลทางตรงต่อข้อสอบอย่างมีนัยสำคัญที่ระดับ .05 ถ้าพบว่ามีปฏิสัมพันธ์ตั้งแต่ 1 ปฏิสัมพันธ์ สามารถบ่งชี้ได้ว่าข้อสอบนั้นทำหน้าที่ต่างกันแบบ Nonuniform DIF ได้แก่ข้อสอบข้อที่ 1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 และข้อที่ 25 หมายความว่า ปฏิสัมพันธ์ระหว่างความสามารถด้านคณิตศาสตร์ (θ) กับความเป็นสมาชิกของกลุ่มขนาดโรงเรียนส่งผลให้เกิดโอกาสในการตอบข้อสอบแตกต่างกันระหว่างคู่ปฏิสัมพันธ์

ตารางที่ 4-31 ความกลมกลืนของโมเดล CFA MIMIC วิชาภาษาไทย จำแนกตามขนาดโรงเรียน

Statistics	CFA	MIMIC (no Direct Effect) @0	MIMIC (1 Direct Effect)	MIMIC (2 Direct Effect)	MIMIC (3 Direct Effect)	MIMIC (4 Direct Effect)
χ^2	6272.816 (df = 740)	6920.090 (df = 857)	6899.203 (df = 856)	6876.553 (df = 855)	6853.849 (df = 854)	6835.680 (df = 853)
RMSEA	0.031	0.030	0.030	0.030	0.030	0.030
CFI	0.600	0.586	0.587	0.589	0.590	0.591
TLI	0.579	0.565	0.566	0.567	0.568	0.569

Statistics	MIMIC (5 Direct Effect)	MIMIC (6 Direct Effect)	MIMIC (7 Direct Effect)	MIMIC (8 Direct Effect)	MIMIC (9 Direct Effect)
χ^2	6817.026 (df = 852)	6800.794 (df = 851)	6754.931 (df = 850)	6739.608 (df = 849)	6725.280 (df = 848)
RMSEA	0.030	0.030	0.029	0.029	0.029
CFI	0.593	0.594	0.597	0.598	0.599
TLI	0.570	0.570	0.573	0.573	0.574

จากตารางที่ 4-31 ผลการวิเคราะห์ความกลมกลืนของโมเดล CFA MIMIC วิชาภาษาไทย จำแนกตามขนาดโรงเรียน จากการวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) ซึ่งพิจารณาจากค่าไค-สแควร์ (Chi-square) มีค่าเท่ากับ 6272.816 ที่องศาอิสระ (df) เท่ากับ 740 ดัชนีวัดรากที่สองของความคลาดเคลื่อนในการประมาณค่า พารามิเตอร์ที่เหมาะสม (RMSEA) มีค่าเท่ากับ 0.031 ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (CFI) มีค่าเท่ากับ 0.600 ดัชนี The tucker lewis index (TLI) มีค่าเท่ากับ 0.579

เมื่อพิจารณาการทำนายตัวแปรข้อสอบจากตัวแปรขนาดโรงเรียน ที่มีดัชนีปรับแต่งมีข้อค้นพบดังนี้

1. เมื่อพิจารณาการทำนายตัวแปรข้อสอบจากขนาดโรงเรียนที่มีดัชนีปรับแต่งพบว่าสามารถทำนายตัวแปรข้อสอบที่มีดัชนีปรับแต่ง จำนวน 5 ข้อ โดยโมเดล MIMIC เริ่มต้น เป็นการเพิ่มตัวแปรทำนายที่มีอิทธิพลทางตรงต่อข้อสอบและบังคับค่าให้เป็น 0 เพื่อกำหนดให้ตัวแปรต้นไม่มีผลทางตรงต่อข้อสอบ มีค่าไค-สแควร์ (Chi-square) เท่ากับ 6920.090 ที่องศาอิสระ (df) เท่ากับ 857 ดัชนีวัดรากที่สองของความคลาดเคลื่อนในการประมาณค่า พารามิเตอร์ที่เหมาะสม (RMSEA)

มีค่าเท่ากับ 0.030 ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (CFI) มีค่าเท่ากับ 0.586 ดัชนี The tucker lewis index (TLI) มีค่าเท่ากับ 0.565

2. เมื่อพิจารณาโมเดล MIMIC ที่เพิ่มอิทธิพลทางตรงจากตัวทำนายไปที่ข้อสอบ ที่มีดัชนีปรับแก้ 1 ข้อ โดยเลือกข้อที่มีดัชนีปรับแก้สูงสุด มีค่าไค-สแควร์ (Chi-square) เท่ากับ 6899.203 ที่องศาอิสระ (df) เท่ากับ 856 ดัชนีวัดรากที่สองของความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ที่เหมาะสม (RMSEA) มีค่าเท่ากับ 0.030 ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (CFI) มีค่า เท่ากับ 0.587ดัชนี The tucker lewis index (TLI) มีค่าเท่ากับ 0.566

3. เมื่อพิจารณาโมเดลMIMIC ที่เพิ่มอิทธิพลทางตรงจากตัวทำนายไปที่ข้อสอบ ที่มีดัชนีปรับแก้ 2 ข้อ โดยเลือกข้อที่มีดัชนีปรับแก้สูงสุด มีค่าไค-สแควร์ (Chi-square) เท่ากับ 6876.553 ที่องศาอิสระ (df) เท่ากับ 855 ดัชนีวัดรากที่สองของความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ที่เหมาะสม (RMSEA) มีค่าเท่ากับ 0.030 ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (CFI) มีค่าเท่ากับ 0.590 ดัชนี The tucker lewis index (TLI) มีค่าเท่ากับ 0.568

4. เมื่อพิจารณาโมเดลMIMIC ที่เพิ่มอิทธิพลทางตรงจากตัวทำนายไปที่ข้อสอบที่มีดัชนีปรับแก้ 5 ข้อจากการปรับครั้งที่ 9 มีค่าไค-สแควร์ (Chi-square) เท่ากับ 6725.280 ที่องศาอิสระ (df) เท่ากับ 848 ดัชนีวัดรากที่สองของความคลาดเคลื่อนในการประมาณค่า พารามิเตอร์ที่เหมาะสม (RMSEA) มีค่าเท่ากับ 0.029 ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (CFI) มีค่าเท่ากับ 0.599 ดัชนี The tucker lewis index (TLI) มีค่าเท่ากับ 0.574

ตารางที่ 4-32 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาภาษาไทย ด้วยวิธี MIMIC แบบ Uniform DIF จำแนกตามขนาดโรงเรียน

Item		Estimate	SE	Z	P-value
THAI ON	SMALL	-0.019	0.008	-2.553*	0.011
	MEDIUM	-0.017	0.007	-2.550*	0.011
	LARGE	-0.011	0.004	-2.529*	0.011
Q01 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q02 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q03 ON	SMALL	-0.223	0.086	-2.591*	0.010
	MEDIUM	-0.142	0.081	-1.760	0.078
	LARGE	-0.186	0.073	-2.528*	0.011
Q04 ON	SMALL	-0.376	0.099	-3.809*	0.000
	MEDIUM	-0.353	0.096	-3.694*	0.000
	LARGE	-0.363	0.089	-4.077*	0.000
Q05 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q06 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q07 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000

ตารางที่ 4-32 (ต่อ)

Item		Estimate	SE	Z	P-value
Q08 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q09 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q10 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q11 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q12 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q13 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q14 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q15 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q16 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000

ตารางที่ 4-32 (ต่อ)

Item		Estimate	SE	Z	P-value
Q17 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q18 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q19 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q20 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q21 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q22 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q23 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q24 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q25 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000

ตารางที่ 4-32 (ต่อ)

Item		Estimate	SE	Z	P-value
Q26 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q27 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q28 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q29 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q30 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q31 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q32 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q33 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q34 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000

ตารางที่ 4-32 (ต่อ)

Item		Estimate	SE	Z	P-value
Q35 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q36 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q37 ON	SMALL	-0.235	0.078	-2.998*	0.003
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q38 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q39 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q40 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000

หมายเหตุ * p < .05; ข้อสอบ DIF คือข้อที่อย่างน้อย 1 ขนาดมีอิทธิพลทางตรงต่อข้อสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

จากตารางที่ 4-32 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาภาษาไทย สำหรับกลุ่มที่มากกว่าสองกลุ่ม ด้วยวิธี MIMIC แบบเอกรูป (Uniform DIF) จำแนกตามขนาดโรงเรียน ผลการวิเคราะห์ข้อสอบจำนวน 40 ข้อ ตัวแปรทำนาย ซึ่งได้แก่ ขนาดโรงเรียน (ขนาดเล็ก ขนาดกลาง ขนาดใหญ่ และขนาดใหญ่พิเศษ) ที่มีอิทธิพลทางตรงต่อข้อสอบ ซึ่งข้อสอบที่มีดัชนีปรับแต่งสูงสุดคือ ข้อที่ 4 รองลงมาคือ ข้อที่ 3 และ ข้อที่ 37 ตามลำดับ ที่ทำให้ตัวแปรขนาดโรงเรียนมีอิทธิพลทางตรงต่อข้อสอบที่มีดัชนีปรับแต่งมีค่าต่างจากศูนย์อย่างมีนัยสำคัญที่ระดับ .05

ซึ่งหมายความว่า ในข้อที่ 4, 3 และ ข้อที่ 37 ตัวแปรขนาดโรงเรียน ส่งผลให้เกิดโอกาสในการตอบข้อสอบข้อแตกต่างกันระหว่างขนาดโรงเรียน สามารถบ่งชี้ได้ว่าข้อสอบนั้นทำหน้าที่ต่างกัน จากผลการวิเคราะห์ พบว่า โรงเรียนทุกขนาดจะเสียเปรียบ ในข้อสอบข้อที่ 4 และข้อที่ 3 โรงเรียนขนาดเล็กจะเสียเปรียบในข้อสอบข้อที่ 37 และข้อที่ 2 และจะเสียเปรียบในข้อสอบข้อที่ 31 และในภาพรวมโรงเรียนขนาดเล็ก กลาง และใหญ่จะเสียเปรียบในการตอบวิชาภาษาไทย

ตารางที่ 4-33 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาภาษาไทย ด้วยวิธี MIMIC แบบ Nonuniform DIF จำแนกตามขนาดโรงเรียน

Item	Interaction	Estimate	SE	Z	P-value
Q01 ON	THAISMALL	3.750	2.700	1.389	0.165
	THAIMEDIUM	-1.855	2.099	-0.884	0.377
	THAILARGE	-1.740	2.383	-0.730	0.465
Q02 ON	THAISMALL	-26.924	11.064	-2.433*	0.015
	THAIMEDIUM	-30.333	12.287	-2.469*	0.014
	THAILARGE	-30.756	12.485	-2.463*	0.014
Q03 ON	THAISMALL	10.084	5.661	1.781	0.075
	THAIMEDIUM	4.427	4.323	1.024	0.306
	THAILARGE	8.661	5.214	1.661	0.097
Q04 ON	THAISMALL	-41.895	17.195	-2.436*	0.015
	THAIMEDIUM	-42.408	17.384	-2.439*	0.015
	THAILARGE	-39.401	16.363	-2.408*	0.016
Q05 ON	THAISMALL	28.298	11.449	2.472*	0.013
	THAIMEDIUM	25.647	10.508	2.441*	0.015
	THAILARGE	22.804	9.576	2.381*	0.017
Q06 ON	THAISMALL	9.353	5.238	1.786	0.074
	THAIMEDIUM	11.247	5.810	1.936	0.053
	THAILARGE	16.541	7.579	2.182*	0.029

ตารางที่ 4-33 (ต่อ)

Item	Interaction	Estimate	SE	Z	P-value
Q07 ON	THAISMALL	-3.476	4.200	-0.828	0.408
	THAIMEDIUM	1.154	4.185	0.276	0.783
	THAILARGE	3.753	4.602	0.816	0.415
Q08 ON	THAISMALL	-1.380	3.884	-0.355	0.722
	THAIMEDIUM	-3.639	4.102	-0.887	0.375
	THAILARGE	-1.289	4.092	-0.315	0.753
Q09 ON	THAISMALL	7.444	4.602	1.618	0.106
	THAIMEDIUM	1.555	3.594	0.433	0.665
	THAILARGE	7.265	4.725	1.538	0.124
Q10 ON	THAISMALL	1.775	3.633	0.489	0.625
	THAIMEDIUM	-0.854	3.604	-0.237	0.813
	THAILARGE	-0.575	3.730	-0.154	0.877
Q11 ON	THAISMALL	-2.172	3.478	-0.625	0.532
	THAIMEDIUM	-2.518	3.566	-0.706	0.480
	THAILARGE	3.500	3.946	0.887	0.375
Q12 ON	THAISMALL	-0.426	3.260	-0.131	0.896
	THAIMEDIUM	0.430	3.329	0.129	0.897
	THAILARGE	-1.738	3.511	-0.495	0.621
Q13 ON	THAISMALL	-9.747	5.689	-1.713	0.087
	THAIMEDIUM	-6.984	5.177	-1.349	0.177
	THAILARGE	-12.476	6.600	-1.890	0.059
Q14 ON	THAISMALL	3.243	3.519	0.922	0.357
	THAIMEDIUM	0.747	3.334	0.224	0.823
	THAILARGE	1.992	3.588	0.555	0.579

ตารางที่ 4-33 (ต่อ)

Item	Interaction	Estimate	SE	Z	P-value
Q15 ON	THAISMALL	3.197	3.455	0.925	0.355
	THAIMEDIUM	4.387	3.705	1.184	0.236
	THAILARGE	2.339	3.562	0.657	0.511
Q16 ON	THAISMALL	-9.167	5.188	-1.767	0.077
	THAIMEDIUM	-10.903	5.675	-1.921	0.055
	THAILARGE	-5.258	4.610	-1.141	0.254
Q17 ON	THAISMALL	16.955	7.478	2.267*	0.023
	THAIMEDIUM	17.444	7.672	2.274*	0.023
	THAILARGE	19.216	8.359	2.299	0.022
Q18 ON	THAISMALL	-9.694	5.248	-1.847	0.065
	THAIMEDIUM	-10.354	5.464	-1.895	0.058
	THAILARGE	-3.956	4.362	-0.907	0.364
Q19 ON	THAISMALL	7.448	4.647	1.603	0.109
	THAIMEDIUM	9.283	5.180	1.792	0.073
	THAILARGE	10.273	5.572	1.844	0.065
Q20 ON	THAISMALL	28.060	11.349	2.473*	0.013
	THAIMEDIUM	27.524	11.175	2.463*	0.014
	THAILARGE	26.487	10.873	2.436*	0.015
Q21 ON	THAISMALL	-31.447	12.731	-2.470*	0.014
	THAIMEDIUM	-28.519	11.714	-2.435*	0.015
	THAILARGE	-26.888	11.210	-2.399*	0.016
Q22 ON	THAISMALL	21.105	9.087	2.322*	0.020
	THAIMEDIUM	21.344	9.192	2.322*	0.020
	THAILARGE	23.558	10.059	2.342*	0.019
Q23 ON	THAISMALL	-25.622	10.716	-2.391*	0.017
	THAIMEDIUM	-22.488	9.691	-2.320*	0.020
	THAILARGE	-14.530	7.366	-1.973*	0.049

ตารางที่ 4-33 (ต่อ)

Item	Interaction	Estimate	SE	Z	P-value
Q24 ON	THAISMALL	-18.571	8.340	-2.227*	0.026
	THAIMEDIUM	-17.580	8.067	-2.179*	0.029
	THAILARGE	-16.873	7.958	-2.120*	0.034
Q25 ON	THAISMALL	28.938	11.665	2.481*	0.013
	THAIMEDIUM	29.817	11.998	2.485*	0.013
	THAILARGE	28.688	11.645	2.464*	0.014
Q26 ON	THAISMALL	9.518	6.391	1.489	0.136
	THAIMEDIUM	14.154	7.569	1.870	0.061
	THAILARGE	15.637	8.122	1.925	0.054
Q27 ON	THAISMALL	11.153	5.369	2.077*	0.038
	THAIMEDIUM	9.259	4.842	1.913	0.056
	THAILARGE	7.218	4.416	1.635	0.102
Q28 ON	THAISMALL	-11.640	5.693	-2.045*	0.041
	THAIMEDIUM	-10.020	5.295	-1.892	0.058
	THAILARGE	-7.952	4.903	-1.622	0.105
Q29 ON	THAISMALL	6.492	4.553	1.426	0.154
	THAIMEDIUM	2.666	3.999	0.667	0.505
	THAILARGE	-1.741	4.134	-0.421	0.674
Q30 ON	THAISMALL	-35.062	14.101	-2.487*	0.013
	THAIMEDIUM	-32.168	13.094	-2.457*	0.014
	THAILARGE	-30.451	12.547	-2.427*	0.015
Q31 ON	THAISMALL	19.430	8.363	2.323*	0.020
	THAIMEDIUM	21.718	9.195	2.362*	0.018
	THAILARGE	19.609	8.553	2.293*	0.022
Q32 ON	THAISMALL	-31.927	12.913	-2.472*	0.013
	THAIMEDIUM	-30.579	12.457	-2.455*	0.014
	THAILARGE	-25.829	10.874	-2.375*	0.018

ตารางที่ 4-33 (ต่อ)

Item	Interaction	Estimate	SE	Z	P-value
Q33 ON	THAISMALL	23.488	9.661	2.431*	0.015
	THAIMEDIUM	19.876	8.400	2.366*	0.018
	THAILARGE	17.786	7.774	2.288*	0.022
Q34 ON	THAISMALL	-4.766	4.200	-1.135	0.256
	THAIMEDIUM	-1.757	3.980	-0.441	0.659
	THAILARGE	0.358	4.144	0.086	0.931
Q35 ON	THAISMALL	10.908	5.617	1.942	0.052
	THAIMEDIUM	7.157	4.623	1.548	0.122
	THAILARGE	13.278	6.490	2.046*	0.041
Q36 ON	THAISMALL	-3.457	3.644	-0.949	0.343
	THAIMEDIUM	-2.655	3.615	-0.734	0.463
	THAILARGE	-6.566	4.362	-1.505	0.132
Q37 ON	THAISMALL	-4.450	4.559	-0.976	0.329
	THAIMEDIUM	2.941	4.053	0.726	0.468
	THAILARGE	1.396	4.055	0.344	0.731
Q38 ON	THAISMALL	8.581	4.696	1.827	0.068
	THAIMEDIUM	9.315	4.944	1.884	0.060
	THAILARGE	5.938	4.195	1.415	0.157
Q39 ON	THAISMALL	10.351	5.509	1.879	0.060
	THAIMEDIUM	9.443	5.308	1.779	0.075
	THAILARGE	10.453	5.711	1.830	0.067
Q40 ON	THAISMALL	-27.999	11.328	-2.472*	0.013
	THAIMEDIUM	-27.800	11.283	-2.464*	0.014
	THAILARGE	-24.538	10.201	-2.405*	0.016

หมายเหตุ * $p < .05$; ข้อสอบ DIF คือข้อที่อย่างน้อย 1 ขนาดที่มีปฏิสัมพันธ์กับ θ แล้วมีอิทธิพลทางตรงต่อข้อสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

จากตารางที่ 4-33 ผลการวิเคราะห์การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาภาษาไทย สำหรับกลุ่มที่มากกว่าสองกลุ่ม ด้วยวิธี MIMIC แบบ Nonuniform DIF จำแนกตามขนาดโรงเรียน ผลการวิเคราะห์ปฏิสัมพันธ์ระหว่างความสามารถด้านภาษาไทยหรือตัวแปรแฝงภายใน (θ) กับความเป็นสมาชิกของกลุ่ม กับความเป็นสมาชิกของกลุ่ม ขนาดโรงเรียน 4 กลุ่ม ประกอบด้วย โรงเรียนขนาดใหญ่พิเศษ ขนาดใหญ่ ขนาดกลาง และขนาดเล็ก ที่มีอิทธิพลทางตรงต่อข้อสอบอย่างมีนัยสำคัญที่ระดับ .05 ถ้าพบว่า มีปฏิสัมพันธ์ตั้งแต่ 1 ปฏิสัมพันธ์ สามารถบ่งชี้ได้ว่าข้อสอบนั้นทำหน้าที่ต่างกันแบบ Nonuniform DIF ได้แก่ ข้อสอบข้อที่ 2, 4, 5, 6, 17, 20, 21, 22, 23, 24, 25, 27, 28, 30, 31, 32, 33, 35 และข้อที่ 40 หมายความว่า ปฏิสัมพันธ์ระหว่างความสามารถด้านภาษาไทย (θ) กับความเป็นสมาชิกของกลุ่มขนาดโรงเรียนส่งผลให้เกิดโอกาสในการตอบข้อสอบแตกต่างกันระหว่างคู่ปฏิสัมพันธ์

ตารางที่ 4-34 ความกลมกลืนของโมเดล CFA MIMIC วิชาวิทยาศาสตร์ จำแนกตามขนาดโรงเรียน

Statistics	**CFA**	**MIMIC (no Direct Effect) @0**	**MIMIC (1 Direct Effect)**	**MIMIC (2 Direct Effect)**	**MIMIC (3 Direct Effect)**
χ^2	3420.963 (df = 740)	3653.028 (df = 857)	3638.280 (df = 856)	3626.787 (df = 855)	3615.852 (df = 854)
RMSEA	0.021	0.020	0.020	0.020	0.020
CFI	0.481	0.508	0.511	0.512	0.514
TLI	0.453	0.484	0.485	0.487	0.488

จากตารางที่ 4-34 ผลการวิเคราะห์ความกลมกลืนของโมเดล CFA MIMIC วิชาวิทยาศาสตร์ จำแนกตามขนาดโรงเรียน จากการวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) ซึ่งพิจารณาจากค่าไค-สแควร์ (Chi-square) มีค่าเท่ากับ 3420.963 ที่องศาอิสระ (df) เท่ากับ 740 ดัชนีวัดรากที่สองของความคลาดเคลื่อนในการประมาณค่า พารามิเตอร์ที่เหมาะสม (RMSEA) มีค่าเท่ากับ 0.021 ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (CFI) มีค่าเท่ากับ 0.481ดัชนี The tucker lewis index (TLI) มีค่าเท่ากับ 0.453

เมื่อพิจารณาการทำนายตัวแปรข้อสอบจากตัวแปรขนาดโรงเรียน ที่มีดัชนีปรับแต่งมีข้อค้นพบ ดังนี้

1. เมื่อพิจารณาการทำนายตัวแปรข้อสอบจากขนาดโรงเรียนที่มีดัชนีปรับแต่ง พบว่า สามารถทำนายตัวแปรข้อสอบที่มีดัชนีปรับแต่ง จำนวน 3 ข้อ โดยโมเดล MIMIC เริ่มต้นเป็นการเพิ่มตัวแปรทำนายที่มีอิทธิพลทางตรงต่อข้อสอบและบังคับค่าให้เป็น 0 เพื่อกำหนดให้ตัวแปรต้นไม่มีผลทางตรงต่อข้อสอบ มีค่าไค-สแควร์ (Chi-square) เท่ากับ 3653.028 ที่องศาอิสระ (df) เท่ากับ 857ดัชนีวัดรากที่สองของความคลาดเคลื่อนในการประมาณค่า พารามิเตอร์ที่เหมาะสม (RMSEA) มีค่าเท่ากับ 0.020 ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (CFI) มีค่าเท่ากับ 0.508 ดัชนี The tucker lewis index (TLI) มีค่าเท่ากับ 0.484

2. เมื่อพิจารณาโมเดลMIMIC ที่เพิ่มอิทธิพลทางตรงจากตัวทำนายไปที่ข้อสอบ ที่มีดัชนีปรับแก้ 1 ข้อ โดยเลือกข้อที่มีดัชนีปรับแก้สูงสุด มีค่าไค-สแควร์ (Chi-square) เท่ากับ 3638.280 ที่องศาอิสระ (df) เท่ากับ 857 ดัชนีวัดรากที่สองของความคลาดเคลื่อนในการประมาณค่า พารามิเตอร์ที่เหมาะสม (RMSEA) มีค่าเท่ากับ 0.020 ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (CFI) มีค่าเท่ากับ 0.511ดัชนี The tucker lewis index (TLI) มีค่าเท่ากับ 0.485

3. เมื่อพิจารณาโมเดลMIMIC ที่เพิ่มอิทธิพลทางตรงจากตัวทำนายไปที่ข้อสอบ ที่มีดัชนีปรับแก้ 2 ข้อ โดยเลือกข้อที่มีดัชนีปรับแก้สูงสุด มีค่าไค-สแควร์ (Chi-square) เท่ากับ 3626.787 ที่องศาอิสระ (df) เท่ากับ 857 ดัชนีวัดรากที่สองของความคลาดเคลื่อนในการประมาณค่า พารามิเตอร์ที่เหมาะสม (RMSEA) มีค่าเท่ากับ 0.020 ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (CFI) มีค่าเท่ากับ 0.512ดัชนี The tucker lewis index (TLI) มีค่าเท่ากับ 0.487

4. เมื่อพิจารณาโมเดลMIMIC ที่เพิ่มอิทธิพลทางตรงจากตัวทำนายไปที่ข้อสอบ ที่มีดัชนีปรับแก้ 3 ข้อ โดยเลือกข้อที่มีดัชนีปรับแก้สูงสุด มีค่าไค-สแควร์ (Chi-square) เท่ากับ 3615.852 ที่องศาอิสระ (df) เท่ากับ 857 ดัชนีวัดรากที่สองของความคลาดเคลื่อนในการประมาณค่า พารามิเตอร์ที่เหมาะสม (RMSEA) มีค่าเท่ากับ 0.020 ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (CFI) มีค่าเท่ากับ 0.514 ดัชนี The tucker lewis index (TLI) มีค่าเท่ากับ 0.488

ตารางที่ 4-35 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาวิทยาศาสตร์ ด้วยวิธี MIMIC แบบ Uniform DIF จำแนกตามขนาดโรงเรียน

Item		Estimate	SE	Z	P-value
SCIENCE ON	SMALL	-0.306	0.017	-17.986*	0.000
	MEDIUM	-0.216	0.017	-12.716*	0.000
	LARGE	-0.221	0.016	-13.882*	0.000
Q01 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q02 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q03 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q04 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q05 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q06 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q07 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000

ตารางที่ 4-35 (ต่อ)

Item		Estimate	SE	Z	P-value
Q08 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q09 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q10 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q11 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q12 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q13 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q14 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q15 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q16 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000

ตารางที่ 4-35 (ต่อ)

Item		Estimate	SE	Z	P-value
Q17 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q18 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q19 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q20 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q21 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q22 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q23 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q24 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q25 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000

ตารางที่ 4-35 (ต่อ)

Item		Estimate	SE	Z	P-value
Q26 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q27 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q28 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q29 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q30 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q31 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q32 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q33 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q34 ON	SMALL	-0.569	0.142	-4.000*	0.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000

ตารางที่ 4-35 (ต่อ)

Item		Estimate	SE	Z	P-value
Q35 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q36 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q37 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q37 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q38 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q39 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000
Q40 ON	SMALL	0.000	0.000	999.000	999.000
	MEDIUM	0.000	0.000	999.000	999.000
	LARGE	0.000	0.000	999.000	999.000

หมายเหตุ * p < .05; ข้อสอบ DIF คือข้อที่อย่างน้อย 1 ขนาดมีอิทธิพลทางตรงต่อข้อสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

จากตารางที่ 4-35 ผลการวิเคราะห์การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาวิทยาศาสตร์ สำหรับกลุ่มที่มากกว่าสองกลุ่ม ด้วยวิธี MIMIC แบบ Uniform DIF จำแนกตามขนาดโรงเรียน ผลการวิเคราะห์ข้อสอบจำนวน 40 ข้อ ตัวแปรทำนาย ซึ่งได้แก่ ขนาดโรงเรียน (ขนาดเล็ก

ขนาดกลาง ขนาดใหญ่ และขนาดใหญ่พิเศษ) ที่มีอิทธิพลทางตรงต่อข้อสอบ ซึ่งข้อสอบที่มีดัชนีปรับแต่งคือ ข้อที่ 34 ที่ทำให้ตัวแปรขนาดโรงเรียนมีอิทธิพลทางตรงต่อข้อสอบที่มีดัชนีปรับแต่งมีค่าต่างจากศูนย์อย่างมีนัยสำคัญที่ระดับ .05 ซึ่งหมายความว่า ในข้อที่ 34 ตัวแปรขนาดโรงเรียนส่งผลให้เกิดโอกาสในการตอบข้อสอบแตกต่างกันระหว่างขนาดโรงเรียน สามารถบ่งชี้ได้ว่าข้อสอบนั้นทำหน้าที่ต่างกัน จากผลการวิเคราะห์ พบว่า โรงเรียนขนาดเล็กจะเสียเปรียบในข้อสอบข้อที่ 34 และในภาพรวมโรงเรียนทุกขนาดจะเสียเปรียบในการตอบวิชาวิทยาศาสตร์

ตารางที่ 4-36 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาวิทยาศาสตร์ วิธี MIMIC แบบ Nonuniform DIF จำแนกตามขนาดโรงเรียน

Item	Interaction	Estimate	SE	Z	P-value
Q01 ON	SCIENCESMALL	-0.129	0.133	-0.969	0.333
	SCIENCEMEDIUM	0.565	0.219	2.580	0.010
	SCIENCELARGE	0.123	0.171	0.719	0.472
Q02 ON	SCIENCESMALL	-0.383	0.158	-2.425	0.015
	SCIENCEMEDIUM	0.274	0.180	1.517	0.129
	SCIENCELARGE	-0.408	0.169	-2.422	0.015
Q03 ON	SCIENCESMALL	-0.345	0.136	-2.543	0.011
	SCIENCEMEDIUM	-0.140	0.144	-0.974	0.330
	SCIENCELARGE	-0.030	0.141	-0.213	0.831
Q04 ON	SCIENCESMALL	-0.223	0.137	-1.626	0.104
	SCIENCEMEDIUM	0.345	0.163	2.113	0.035
	SCIENCELARGE	0.079	0.152	0.521	0.603
Q05 ON	SCIENCESMALL	-0.192	0.139	-1.382	0.167
	SCIENCEMEDIUM	0.031	0.153	0.203	0.839
	SCIENCELARGE	-0.161	0.150	-1.070	0.285
Q06 ON	SCIENCESMALL	-0.726	0.175	-4.149	0.000
	SCIENCEMEDIUM	-0.422	0.174	-2.431	0.015
	SCIENCELARGE	-0.611	0.182	-3.355	0.001

ตารางที่ 4-36 (ต่อ)

Item	Interaction	Estimate	SE	Z	P-value
Q07 ON	SCIENCESMALL	-0.824	0.179	-4.600	0.000
	SCIENCEMEDIUM	-0.331	0.174	-1.905	0.057
	SCIENCELARGE	-0.486	0.181	-2.685	0.007
Q08 ON	SCIENCESMALL	-1.085	0.217	-5.001	0.000
	SCIENCEMEDIUM	-0.459	0.206	-2.233	0.026
	SCIENCELARGE	-0.965	0.220	-4.378	0.000
Q09 ON	SCIENCESMALL	-1.174	0.233	-5.037	0.000
	SCIENCEMEDIUM	0.015	0.230	0.066	0.948
	SCIENCELARGE	-0.829	0.230	-3.607	0.000
Q10 ON	SCIENCESMALL	-0.312	0.145	-2.157	0.031
	SCIENCEMEDIUM	0.020	0.154	0.132	0.895
	SCIENCELARGE	-0.286	0.156	-1.825	0.068
Q11 ON	SCIENCESMALL	-0.263	0.145	-1.813	0.070
	SCIENCEMEDIUM	0.244	0.163	1.501	0.133
	SCIENCELARGE	0.006	0.161	0.037	0.970
Q12 ON	SCIENCESMALL	-0.452	0.156	-2.894	0.004
	SCIENCEMEDIUM	0.019	0.167	0.111	0.911
	SCIENCELARGE	-0.144	0.165	-0.872	0.383
Q13 ON	SCIENCESMALL	-0.271	0.151	-1.799	0.072
	SCIENCEMEDIUM	0.394	0.180	2.183	0.029
	SCIENCELARGE	-0.184	0.162	-1.134	0.257
Q14 ON	SCIENCESMALL	-0.518	0.153	-3.378	0.001
	SCIENCEMEDIUM	0.018	0.161	0.112	0.911
	SCIENCELARGE	-0.446	0.163	-2.736	0.006

ตารางที่ 4-36 (ต่อ)

Item	Interaction	Estimate	SE	Est./SE	P-value
Q15 ON	SCIENCESMALL	-0.469	0.151	-3.115	0.002
	SCIENCEMEDIUM	0.031	0.161	0.191	0.848
	SCIENCELARGE	-0.488	0.162	-3.011	0.003
Q16 ON	SCIENCESMALL	-1.243	0.222	-5.594	0.000
	SCIENCEMEDIUM	-0.567	0.200	-2.835	0.005
	SCIENCELARGE	-0.978	0.218	-4.493	0.000
Q17 ON	SCIENCESMALL	-1.126	0.208	-5.406	0.000
	SCIENCEMEDIUM	0.037	0.199	0.185	0.853
	SCIENCELARGE	-0.572	0.200	-2.861	0.004
Q18 ON	SCIENCESMALL	-0.066	0.127	-0.521	0.602
	SCIENCEMEDIUM	0.221	0.146	1.515	0.130
	SCIENCELARGE	0.025	0.140	0.175	0.861
Q19 ON	SCIENCESMALL	-0.129	0.130	-0.994	0.320
	SCIENCEMEDIUM	0.402	0.157	2.560	0.010
	SCIENCELARGE	-0.065	0.142	-0.460	0.646
Q20 ON	SCIENCESMALL	-0.339	0.147	-2.308	0.021
	SCIENCEMEDIUM	-0.300	0.157	-1.909	0.056
	SCIENCELARGE	-0.426	0.162	-2.629	0.009
Q21 ON	SCIENCESMALL	-0.150	0.137	-1.093	0.275
	SCIENCEMEDIUM	0.091	0.152	0.597	0.550
	SCIENCELARGE	-0.045	0.152	-0.296	0.767
Q22 ON	SCIENCESMALL	-0.796	0.181	-4.406	0.000
	SCIENCEMEDIUM	-0.388	0.178	-2.187	0.029
	SCIENCELARGE	-0.909	0.200	-4.551	0.000
Q23 ON	SCIENCESMALL	-0.855	0.181	-4.737	0.000
	SCIENCEMEDIUM	-0.249	0.176	-1.418	0.156
	SCIENCELARGE	-0.782	0.191	-4.090	0.000

ตารางที่ 4-36 (ต่อ)

Item	Interaction	Estimate	SE	Est./SE	P-value
Q24 ON	SCIENCESMALL	-0.356	0.161	-2.212	0.027
	SCIENCEMEDIUM	0.244	0.182	1.339	0.181
	SCIENCELARGE	-0.085	0.173	-0.492	0.623
Q25 ON	SCIENCESMALL	-0.526	0.170	-3.086	0.002
	SCIENCEMEDIUM	0.056	0.183	0.306	0.760
	SCIENCELARGE	-0.352	0.178	-1.969	0.049
Q26 ON	SCIENCESMALL	-0.949	0.192	-4.951	0.000
	SCIENCEMEDIUM	-0.477	0.190	-2.509	0.012
	SCIENCELARGE	-0.904	0.204	-4.424	0.000
Q27 ON	SCIENCESMALL	-0.741	0.178	-4.162	0.000
	SCIENCEMEDIUM	-0.328	0.180	-1.818	0.069
	SCIENCELARGE	-0.588	0.189	-3.109	0.002
Q28 ON	SCIENCESMALL	-0.655	0.183	-3.576	0.000
	SCIENCEMEDIUM	-0.645	0.204	-3.166	0.002
	SCIENCELARGE	-0.735	0.201	-3.660	0.000
Q29 ON	SCIENCESMALL	-0.876	0.205	-4.263	0.000
	SCIENCEMEDIUM	-0.143	0.201	-0.712	0.477
	SCIENCELARGE	-0.639	0.209	-3.054	0.002
Q30 ON	SCIENCESMALL	-1.235	0.222	-5.570	0.000
	SCIENCEMEDIUM	-0.541	0.202	-2.673	0.008
	SCIENCELARGE	-1.006	0.221	-4.555	0.000
Q31 ON	SCIENCESMALL	-0.485	0.176	-2.754	0.006
	SCIENCEMEDIUM	0.093	0.191	0.487	0.626
	SCIENCELARGE	-0.351	0.185	-1.896	0.058
Q32 ON	SCIENCESMALL	-0.602	0.164	-3.667	0.000
	SCIENCEMEDIUM	0.028	0.172	0.162	0.871
	SCIENCELARGE	-0.422	0.173	-2.446	0.014

ตารางที่ 4-36 (ต่อ)

Item	Interaction	Estimate	SE	Est./SE	P-value
Q33 ON	SCIENCESMALL	-0.707	0.169	-4.181	0.000
	SCIENCEMEDIUM	-0.419	0.169	-2.483	0.013
	SCIENCELARGE	-0.625	0.178	-3.518	0.000
Q34 ON	SCIENCESMALL	-1.194	0.256	-4.670	0.000
	SCIENCEMEDIUM	-0.159	0.175	-0.905	0.365
	SCIENCELARGE	-0.535	0.187	-2.861	0.004
Q35 ON	SCIENCESMALL	-1.063	0.201	-5.284	0.000
	SCIENCEMEDIUM	-0.161	0.183	-0.879	0.379
	SCIENCELARGE	-0.858	0.205	-4.181	0.000
Q36 ON	SCIENCESMALL	-0.651	0.157	-4.158	0.000
	SCIENCEMEDIUM	-0.235	0.157	-1.500	0.134
	SCIENCELARGE	-0.467	0.163	-2.869	0.004
Q37 ON	SCIENCESMALL	-0.433	0.146	-2.967	0.003
	SCIENCEMEDIUM	0.005	0.154	0.034	0.973
	SCIENCELARGE	-0.276	0.156	-1.769	0.077
Q38 ON	SCIENCESMALL	-0.551	0.166	-3.323	0.001
	SCIENCEMEDIUM	-0.204	0.175	-1.167	0.243
	SCIENCELARGE	-0.345	0.178	-1.938	0.053
Q39 ON	SCIENCESMALL	-0.299	0.147	-2.037	0.042
	SCIENCEMEDIUM	0.070	0.158	0.442	0.659
	SCIENCELARGE	-0.319	0.160	-1.990	0.047
Q40 ON	SCIENCESMALL	-0.174	0.149	-1.169	0.242
	SCIENCEMEDIUM	0.177	0.164	1.080	0.280
	SCIENCELARGE	-0.316	0.161	-1.961	0.050

หมายเหตุ * $p < .05$; ข้อสอบ DIF คือข้อที่อย่างน้อย 1 ขนาดที่มีปฏิสัมพันธ์กับ θ แล้วมีอิทธิพลทางตรงต่อข้อสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

จากตารางที่ 4-36 ผลการวิเคราะห์การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชา วิทยาศาสตร์ สำหรับกลุ่มที่มากกว่าสองกลุ่ม ด้วยวิธี MIMIC แบบ Nonuniform DIF จำแนกตาม ขนาดโรงเรียน ผลการวิเคราะห์ปฏิสัมพันธ์ระหว่างความสามารถด้านวิทยาศาสตร์หรือตัวแปรแฝง ภายใน (θ) กับความเป็นสมาชิกของกลุ่ม กับความเป็นสมาชิกของกลุ่ม ขนาดโรงเรียน 4 กลุ่ม ประกอบด้วย โรงเรียนขนาดใหญ่พิเศษ ขนาดใหญ่ ขนาดกลาง และขนาดเล็ก ที่มีอิทธิพลทางตรง ต่อข้อสอบอย่างมีนัยสำคัญที่ระดับ .05 ถ้าพบว่า มีปฏิสัมพันธ์ตั้งแต่ 1 ปฏิสัมพันธ์ สามารถบ่งชี้ได้ ว่าข้อสอบนั้นทำหน้าที่ต่างกันแบบ Nonuniform DIF ได้แก่ข้อสอบข้อที่ 1, 2, 3, 4, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17, 20, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38 และข้อที่ 40 หมายความว่า ปฏิสัมพันธ์ระหว่างความสามารถด้านวิทยาศาสตร์ (θ) กับความเป็นสมาชิก ของกลุ่มขนาดโรงเรียนส่งผลให้เกิดโอกาสในการตอบข้อสอบแตกต่างกันระหว่างคู่ปฏิสัมพันธ์

ตอนที่ 3 การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT วิธี MGCFA-ALIGNMENTและวิธี MIMIC ภายใต้ปัจจัยที่ต่างกัน 2 ปัจจัย ได้แก่ สถานที่ตั้งทางภูมิศาสตร์ และขนาดโรงเรียน

การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มที่มากกว่า สองกลุ่มด้วยวิธี IRT LRT วิธี MGCFA-ALIGNMENT และวิธี MIMIC ภายใต้ปัจจัยที่ต่างกัน 2 ปัจจัย ได้แก่ สถานที่ตั้งทางภูมิศาสตร์ และขนาดโรงเรียน มีรายละเอียดดังแสดงในตารางที่ 4-37 ถึง ตารางที่ 4-61 ดังนี้

ตารางที่ 4-37 สรุปผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ ด้วย 3 วิธี จำแนกตามสถานที่ตั้งทางภูมิศาสตร์

Item	วิธีตรวจสอบ DIF		
	IRT LRT	MGCFA-ALIGMENT	MIMIC
Q1			D
Q2			D
Q3	D		D
Q4	D		D
Q5			D
Q6	D		D
Q7			D
Q8			D
Q9			D
Q10			D
Q11			D
Q12	D	D	D
Q13			D
Q14			D
Q15			D
Q16			D
Q17	D	D	D
Q18			
Q19	D		
Q20	D		D
Q21			D
Q22	D		D
Q23	D		D
Q24			
Q25	D		D
รวม	10	2	22

หมายเหตุ อักษร D หมายถึง ข้อสอบทำหน้าที่ต่างกัน (DIF)

จากตารางที่ 4-37 สรุปผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ทั้งแบบ เอกรูป (Uniform DIF) และแบบอเนกรูป (Nonuniform DIF) วิชาคณิตศาสตร์ ด้วย วิธี IRT LRT วิธี MGCFA-ALIGMENT และวิธี MIMIC จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ พบว่า วิธี MIMIC พบข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด จำนวน 22 ข้อ รองลงมาได้แก่วิธี IRT LRT พบข้อสอบ ที่ทำหน้าที่ต่างกัน จำนวน 10 ข้อ และวิธี MGCFA-ALIGMENT พบข้อสอบที่ทำหน้าที่ต่างกันน้อยที่สุด จำนวน 2 ข้อ

ตารางที่ 4-38 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ วิธี IRT LRT และวิธี MGCFA-ALIGMENT จำแนกตามสถานที่ตั้งทางภูมิศาสตร์

ผลการตรวจสอบ DIF	**วิธี MGCFA-ALIGMENT**		**Chi Sq (χ^2)**
	NODIF (n = 25)	**DIF (n = 25)**	**(Sig.)**
วิธี IRT LRT			3.261
NODIF	60.0%	0.0%	(.150)
DIF	32.0%	8.0%	

จากตารางที่ 4-38 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ ระหว่างวิธี IRT LRT และวิธี MGCFA-ALIGMENT จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ พบว่า ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี IRT LRT ไม่สอดคล้องกับวิธี MGCFA-ALIGMENT มีค่าไควร์-สแควร์ (χ^2) เท่ากับ 3.261

ตารางที่ 4-39 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ วิธี IRT LRT และวิธี MIMIC จำแนกตามสถานที่ตั้งทางภูมิศาสตร์

ผลการตรวจสอบ DIF	วิธี MIMIC		Chi Sq (χ^2) (Sig.)
	NODIF (n = 25)	DIF (n = 25)	
วิธี IRT LRT			.063 (1.000)
NODIF	8.0%	52.0%	
DIF	4.0%	36.0%	

จากตารางที่ 4-39 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ ระหว่างวิธี IRT LRT และวิธี MIMIC จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ พบว่า ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี IRT LRT ไม่สอดคล้องกับวิธี MIMIC มีค่าไควร์-สแควร์ (χ^2) เท่ากับ .063

ตารางที่ 4-40 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ วิธี MGCFA-ALIGMENT และวิธี MIMIC จำแนกตามสถานที่ตั้งทางภูมิศาสตร์

ผลการตรวจสอบ DIF	วิธี MIMIC		Chi Sq (χ^2) (Sig.)
	NODIF (n = 25)	DIF (n = 25)	
วิธี ALIGMENT			.296 (1.000)
NODIF	12.0%	80.0%	
DIF	0.0%	8.0%	

จากตารางที่ 4-40 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ ระหว่างวิธี ALIGMENT และวิธี MIMIC จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ พบว่า ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี ALIGMENT ไม่สอดคล้องกับวิธี MIMIC มีค่าไควร์-สแควร์ (χ^2) เท่ากับ .296

ตารางที่ 4-41 สรุปผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ ด้วย 3 วิธี จำแนกตามขนาดโรงเรียน

Item	วิธีตรวจสอบ DIF		
	IRT LRT	MGCFA-ALIGMENT	MIMIC
Q1	D		D
Q2	D		D
Q3	D	D	D
Q4	D	D	D
Q5	D	D	D
Q6	D		D
Q7	D		D
Q8	D	D	D
Q9	D		D
Q10	D		
Q11	D		D
Q12	D		D
Q13		D	D
Q14			D
Q15			D
Q16			D
Q17	D	D	D
Q18	D	D	D
Q19	D	D	D
Q20			D
Q21	D	D	D
Q22			D
Q23	D		D
Q24	D		D
Q25			D
รวม	18	9	24

หมายเหตุ อักษร D หมายถึง ข้อสอบทำหน้าที่ต่างกัน (DIF)

จากตารางที่ 4-41 สรุปผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ทั้งแบบเอกรูป (Uniform DIF) และแบบอเนกรูป (Nonuniform DIF) วิชาคณิตศาสตร์ ด้วย วิธี IRT LRT วิธี MGCFA-ALIGMENT และวิธี MIMIC จำแนกตามขนาดโรงเรียน พบว่า วิธี MIMIC พบข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด จำนวน 24 ข้อ รองลงมาได้แก่วิธี IRT LRT พบข้อสอบที่ทำหน้าที่ต่างกัน จำนวน 18 ข้อ และวิธี MGCFA-ALIGMENT พบข้อสอบที่ทำหน้าที่ต่างกันน้อยที่สุด จำนวน 9 ข้อ

ตารางที่ 4-42 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ วิธี IRT LRT และวิธี MGCFA-ALIGMENT จำแนกตามขนาดโรงเรียน

ผลการตรวจสอบ DIF	**วิธี MGCFA-ALIGMENT**		**Chi Sq (χ^2)**
	NODIF (n = 25)	**DIF (n = 25)**	**(Sig.)**
วิธี IRT LRT			1.990
NODIF	24.0%	4.0%	(.355)
DIF	40.0%	32.0%	

จากตารางที่ 4-42 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ ระหว่างวิธี IRT LRT และวิธี MGCFA ALIGMENT จำแนกตามขนาดโรงเรียน พบว่า ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี IRT LRT ไม่สอดคล้องกับวิธี MGCFA ALIGMENT มีค่าไควร์-สแควร์ (χ^2) เท่ากับ 1.990

ตารางที่ 4-43 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ วิธี IRT LRT และวิธี MIMIC จำแนกตามขนาดโรงเรียน

ผลการตรวจสอบ DIF	**วิธี MIMIC**		**Chi Sq (χ^2)**
	NODIF (n = 25)	**DIF (n = 25)**	**(Sig.)**
วิธี IRT LRT			.405
NODIF	0.0%	28.0%	(1.000)
DIF	4.0%	68.0%	

จากตารางที่ 4-43 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ ระหว่างวิธี IRT LRT และวิธี MIMIC จำแนกตามขนาดโรงเรียน พบว่า ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี IRT LRT ไม่สอดคล้องกับวิธี MIMIC มีค่าไควร์-สแควร์ (χ^2) เท่ากับ .405

ตารางที่ 4-44 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ วิธี MGCFA-ALIGMENT และวิธี MIMIC จำแนกตามขนาดโรงเรียน

ผลการตรวจสอบ DIF	**วิธี MIMIC**		**Chi Sq (χ^2)**
	NODIF (n = 25)	**DIF (n = 25)**	**(Sig.)**
วิธี ALIGMENT			.586
NODIF	4.0%	60.0%	(1.000)
DIF	0.0%	36.0%	

จากตารางที่ 4-44 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ ระหว่างวิธี ALIGMENT และวิธี MIMIC จำแนกตามขนาดโรงเรียน พบว่า ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี ALIGMENT ไม่สอดคล้องกับวิธี MIMIC มีค่าไควร์-สแควร์ (χ^2) เท่ากับ .586

ตารางที่ 4-45 สรุปผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาภาษาไทย ด้วย 3 วิธี จำแนกตามสถานที่ตั้งทางภูมิศาสตร์

Item	วิธีตรวจสอบ DIF		
	IRT LRT	MGCFA-ALIGMENT	MIMIC
Q1	D		D
Q2	D		D
Q3			
Q4	D		D
Q5	D		D
Q6			
Q7			D
Q8			D
Q9			
Q10			
Q11			D
Q12			
Q13			
Q14			
Q15	D		D
Q16			
Q17			D
Q18			
Q19	D		
Q20	D		D
Q21	D		D
Q22	D		D
Q23	D		D
Q24	D		D

ตารางที่ 4-45 (ต่อ)

Item	วิธีตรวจสอบ DIF		
	IRT LRT	MGCFA-ALIGMENT	MIMIC
Q25	D		D
Q26	D	D	D
Q27	D		D
Q28	D		D
Q29			
Q30	D		D
Q31	D		D
Q32	D		D
Q33	D	D	D
Q34			D
Q35	D		D
Q36	D		D
Q37	D	D	
Q38	D		D
Q39	D		D
Q40	D		D
รวม	25	3	28

หมายเหตุ อักษร D หมายถึง ข้อสอบทำหน้าที่ต่างกัน (DIF)

จากตารางที่ 4-45 สรุปผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ทั้งแบบเอกรูป (Uniform DIF) และแบบอเนกรูป (Nonuniform DIF) วิชาภาษาไทย ด้วย ด้วย วิธี IRT LRT วิธี MGCFA-ALIGMENT และวิธี MIMIC จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ พบว่า วิธี MIMIC พบข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด จำนวน 28 ข้อ รองลงมาได้แก่วิธี IRT LRT พบข้อสอบที่ทำหน้าที่ต่างกัน จำนวน 25 ข้อ และวิธี MGCFA-ALIGMENT พบข้อสอบที่ทำหน้าที่ต่างกันน้อยที่สุด จำนวน 3 ข้อ

ตารางที่ 4-46 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาภาษาไทย วิธี IRT LRT และวิธี MGCFA-ALIGMENT จำแนกตามสถานที่ตั้งทางภูมิศาสตร์

ผลการตรวจสอบ DIF	วิธี MGCFA-ALIGMENT		Chi Sq (χ^2)
	NODIF (n = 40)	DIF (n = 40)	(Sig.)
วิธี IRT LRT			1.946
NODIF	37.5%	0.0%	(.279)
DIF	55.0%	7.5%	

จากตารางที่ 4-46 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย ระหว่างวิธี IRT LRT และวิธี MGCFA-ALIGMENT จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ พบว่า ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี IRT LRT ไม่สอดคล้องกับวิธี MGCFA ALIGMENT มีค่าไควร์-สแควร์ (χ^2) เท่ากับ 1.936

ตารางที่ 4-47 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาภาษาไทย วิธี IRT LRT และวิธี MIMIC จำแนกตามสถานที่ตั้งทางภูมิศาสตร์

ผลการตรวจสอบ DIF	วิธี MIMIC		Chi Sq (χ^2)
	NODIF (n = 40)	DIF (n = 40)	(Sig.)
วิธี IRT LRT			15.365
NODIF	25.0%	12.5%	(.000)
DIF	5.0%	57.5%	

จากตารางที่ 4-47 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย ระหว่างวิธี IRT LRT และวิธี MIMIC จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ พบว่า ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี IRT LRT สอดคล้องกับวิธี MIMIC มีค่าไควร์-สแควร์ (χ^2) เท่ากับ 15.365 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

ตารางที่ 4-48 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาภาษาไทย วิธี MGCFA-ALIGMENT และวิธี MIMIC จำแนกตามสถานที่ตั้งทางภูมิศาสตร์

ผลการตรวจสอบ DIF	วิธี MIMIC		Chi Sq (χ^2)
	NODIF (n = 25)	DIF (n = 25)	(Sig.)
วิธี ALIGMENT			.017
NODIF	27.5%	65.0%	(1.000)
DIF	2.5%	5.0%	

จากตารางที่ 4-48 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย ระหว่างวิธี MGCFA-ALIGMENT และวิธี MIMIC จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ พบว่า ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี MGCFA-ALIGMENT ไม่สอดคล้องกับวิธี MIMIC มีค่าไควร์-สแควร์ (χ^2) เท่ากับ .017

ตารางที่ 4-49 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาภาษาไทย จำแนกตามขนาดโรงเรียน

Item	วิธีตรวจสอบ DIF		
	IRT LRT	MGCFA-ALIGMENT	MIMIC
Q1	D		
Q2	D	D	D
Q3	D		D
Q4	D		D
Q5	D		D
Q6	D	D	D
Q7	D		
Q8			
Q9	D		

ตารางที่ 4-49 (ต่อ)

Item	วิธีตรวจสอบ DIF		
	IRT LRT	MGCFA-ALIGMENT	MIMIC
Q10	D		
Q11			
Q12	D		
Q13			
Q14			
Q15	D		
Q16			
Q17	D	D	D
Q18	D		
Q19	D		
Q20	D		D
Q21	D		D
Q22	D		D
Q23	D		D
Q24	D		D
Q25	D	D	D
Q26	D		
Q27	D		D
Q28	D		D
Q29	D		
Q30	D		D
Q31	D		D
Q32	D		D
Q33	D		D
Q34			

ตารางที่ 4-49 (ต่อ)

Item	วิธีตรวจสอบ DIF		
	IRT LRT	MGCFA-ALIGMENT	MIMIC
Q35	D		D
Q36			
Q37	D		D
Q38	D		
Q39	D		
Q40	D	D	D
รวม	33	3	17

หมายเหตุ อักษร D หมายถึง ข้อสอบทำหน้าที่ต่างกัน (DIF)

จากตารางที่ 4-49 สรุปผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ทั้งแบบเอกรูป (Uniform DIF) และแบบอเนกรูป (Nonuniform DIF) วิชาภาษาไทย ด้วย วิธี IRT LRT วิธี MGCFA-ALIGMENT และวิธี MIMIC จำแนกตามขนาดโรงเรียน พบว่า วิธี IRT LRT พบข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด จำนวน 33 ข้อ รองลงมาได้แก่วิธี วิธี MIMIC พบข้อสอบที่ทำหน้าที่ต่างกัน จำนวน 17 ข้อ และวิธี MGCFA-ALIGMENT พบข้อสอบที่ทำหน้าที่ต่างกันน้อยที่สุด จำนวน 3 ข้อ

ตารางที่ 4-50 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ
วิชาภาษาไทย วิธี IRT LRT และวิธี MGCFA-ALIGMENT
จำแนกตามขนาดโรงเรียน

ผลการตรวจสอบ DIF	วิธี MGCFA-ALIGMENT		Chi Sq (χ^2)
	NODIF (n = 40)	DIF (n = 40)	(Sig.)
วิธี IRT LRT			1.212
NODIF	17.5%	0.0%	(.565)
DIF	70.0%	12.5%	

จากตารางที่ 4-50 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย ระหว่างวิธี IRT LRT และวิธี MGCFA-ALIGMENT จำแนกตามขนาดโรงเรียน พบว่า ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี IRT LRT ไม่สอดคล้องกับวิธี MGCFA-ALIGMENT มีค่าไควร์-สแควร์ (χ^2) เท่ากับ 1.212

ตารางที่ 4-51 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย วิธี IRT LRT และวิธี MIMIC จำแนกตามขนาดโรงเรียน

ผลการตรวจสอบ DIF	**วิธี MIMIC**		**Chi Sq (χ^2)**
	NODIF (n = 40)	**DIF (n = 40)**	**(Sig.)**
วิธี IRT LRT			9.378
NODIF	17.5%	0.0%	(.003)
DIF	30.0%	52.5%	

จากตารางที่ 4-51 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย ระหว่างวิธี IRT LRT และวิธี MIMIC จำแนกตามขนาดโรงเรียน พบว่า ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี IRT LRT สอดคล้องกับวิธี MIMIC มีค่าไควร์-สแควร์ (χ^2) เท่ากับ 9.378 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

ตารางที่ 4-52 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย วิธี MGCFA-ALIGMENT และวิธี MIMIC จำแนกตามขนาดโรงเรียน

ผลการตรวจสอบ DIF	**วิธี MIMIC**		**Chi Sq (χ^2)**
	NODIF (n = 40)	**DIF (n = 40)**	**(Sig.)**
วิธี ALIGMENT			5.170
NODIF	47.5%	40.0%	(.049)
DIF	0.0%	12.5%	

จากตารางที่ 4-52 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย ระหว่างวิธี MGCFA-ALIGMENT และวิธี MIMIC จำแนกตามขนาดโรงเรียน พบว่าผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี MGCFA-ALIGMENT ไม่สอดคล้องกับวิธี MIMIC มีค่าไควร์-สแควร์ (χ^2) เท่ากับ 5.170 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

ตารางที่ 4-53 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาวิทยาศาสตร์ จำแนกตามสถานที่ตั้งทางภูมิศาสตร์

Item	**วิธีตรวจสอบ DIF**		
	IRT LRT	**MGCFA-ALIGMENT**	**MIMIC**
Q1			D
Q2			
Q3	D		D
Q4	D		
Q5		D	
Q6	D		D
Q7			D
Q8			D
Q9			D
Q10	D		
Q11	D		D
Q12	D		D
Q13			
Q14			
Q15			
Q16	D		D
Q17	D		D
Q18	D	D	D

ตารางที่ 4-53 (ต่อ)

Item	วิธีตรวจสอบ DIF		
	IRT LRT	MGCFA-ALIGMENT	MIMIC
Q19			
Q20	D		
Q21			
Q22	D		D
Q23	D	D	D
Q24	D		
Q25	D		D
Q26	D	D	D
Q27	D	D	D
Q28	D	D	D
Q29			D
Q30			D
Q31			
Q32			D
Q33			D
Q34	D	D	D
Q35	D	D	D
Q36	D		D
Q37	D		D
Q38	D		D
Q39			
Q40			
รวม	22	8	26

หมายเหตุ อักษร D หมายถึง ข้อสอบทำหน้าที่ต่างกัน (DIF)

จากตารางที่ 4-53 สรุปผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ทั้งแบบเอกรูป (Uniform DIF) และแบบอเนกรูป (Nonuniform DIF) วิชาวิทยาศาสตร์ ด้วย วิธี IRT LRT วิธี MGCFA-ALIGMENT และวิธี MIMIC จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ พบว่า วิธี MIMIC พบข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด จำนวน 26 ข้อ รองลงมาได้แก่วิธี วิธี IRT LRT พบข้อสอบที่ทำหน้าที่ต่างกัน จำนวน 22 ข้อ และวิธี MGCFA-ALIGMENT พบข้อสอบที่ทำหน้าที่ต่างกันน้อยที่สุด จำนวน 8 ข้อ

ตารางที่ 4-54 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาวิทยาศาสตร์ วิธี IRT LRT และวิธี MGCFA-ALIGMENT จำแนกตามสถานที่ตั้งทางภูมิศาสตร์

ผลการตรวจสอบ DIF	**วิธี MGCFA-ALIGMENT**		**Chi Sq (χ^2)**
	NODIF (n = 40)	**DIF (n = 40)**	**(Sig.)**
วิธี IRT LRT			4.268 (.054)
NODIF	42.5%	2.5%	
DIF	37.5%	17.5%	

จากตารางที่ 4-54 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาวิทยาศาสตร์ ระหว่างวิธี IRT LRT และวิธี MGCFA ALIGMENT จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ พบว่า ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี IRT LRT ไม่สอดคล้องกับวิธี MGCFA ALIGMENT มีค่าไคว์-สแควร์ (χ^2) เท่ากับ 4.268

ตารางที่ 4-55 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาวิทยาศาสตร์ วิธี IRT LRT และวิธี MIMIC จำแนกตามสถานที่ตั้งทางภูมิศาสตร์

ผลการตรวจสอบ DIF	วิธี MIMIC		Chi Sq (χ^2)
	NODIF (n = 40)	DIF (n = 40)	(Sig.)
วิธี IRT LRT			6.078
NODIF	25.0%	20.0%	(.014)
DIF	10.0%	45.0%	

จากตารางที่ 4-55 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาวิทยาศาสตร์ ระหว่างวิธี IRT LRT และวิธี MIMIC จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ พบว่า ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี IRT LRT สอดคล้องกับ วิธี MIMIC มีค่าไควร์-สแควร์ (χ^2) เท่ากับ 6.078 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

ตารางที่ 4-56 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย วิธี MGCFA-ALIGMENT และวิธี MIMIC จำแนกตามสถานที่ตั้งทางภูมิศาสตร์

ผลการตรวจสอบ DIF	วิธี MIMIC		Chi Sq (χ^2)
	NODIF (n = 25)	DIF (n = 25)	(Sig.)
วิธี ALIGMENT			2.225
NODIF	32.5%	37.5%	(.222)
DIF	2.5%	17.5%	

จากตารางที่ 4-56 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาวิทยาศาสตร์ ระหว่างวิธี MGCFA-ALIGMENT และวิธี MIMIC จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ พบว่า ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี MGCFA-ALIGMENT ไม่สอดคล้องกับวิธี MIMIC มีค่าไควร์-สแควร์ (χ^2) เท่ากับ 2.225

ตารางที่ 4-57 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาวิทยาศาสตร์ จำแนกตามขนาดโรงเรียน

Item	วิธีตรวจสอบ DIF		
	IRT LRT	MGCFA-ALIGMENT	MIMIC
Q1	D	D	D
Q2			D
Q3	D		D
Q4	D		D
Q5	D		
Q6			D
Q7			D
Q8	D		D
Q9			D
Q10			D
Q11	D		
Q12	D		D
Q13			D
Q14			D
Q15			D
Q16	D		D
Q17	D	D	D
Q18	D		
Q19	D		D
Q20	D		D
Q21			
Q22	D	D	D
Q23	D	D	D
Q24	D	D	D

ตารางที่ 4-57 (ต่อ)

Item	วิธีตรวจสอบ DIF		
	IRT LRT	MGCFA ALIGMENT	MIMIC
Q25			D
Q26	D	D	D
Q27	D	D	D
Q28	D	D	D
Q29			D
Q30	D		D
Q31	D		D
Q32			D
Q33	D		D
Q34	D	D	D
Q35	D	D	D
Q36	D		D
Q37			D
Q38	D		D
Q39			D
Q40			
รวม	25	10	35

หมายเหตุ อักษร D หมายถึง ข้อสอบทำหน้าที่ต่างกัน (DIF)

จากตารางที่ 4-57 สรุปผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้งแบบเอกรูป (Uniform DIF) และแบบอเนกรูป (Nonuniform DIF) วิชาวิทยาศาสตร์ ด้วย วิธี IRT LRT วิธี MGCFA-ALIGMENT และวิธี MIMIC จำแนกตามขนาดโรงเรียน พบว่า วิธี MIMIC พบข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด จำนวน 35 ข้อ รองลงมาได้แก่วิธี วิธี IRT LRT พบข้อสอบที่ทำหน้าที่ต่างกัน จำนวน 25 ข้อ และวิธี MGCFA-ALIGMENT พบข้อสอบที่ทำหน้าที่ต่างกันน้อยที่สุด จำนวน 10 ข้อ

ตารางที่ 4-58 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาวิทยาศาสตร์ วิธี IRT LRT และวิธี MGCFA-ALIGMENT จำแนกตามขนาดโรงเรียน

ผลการตรวจสอบ DIF	วิธี MGCFA-ALIGMENT		Chi Sq (χ^2)
	NODIF (n = 40)	DIF (n = 40)	(Sig.)
วิธี IRT LRT			8.000
NODIF	37.5%	0.0%	(.006)
DIF	37.5%	25.0%	

จากตารางที่ 4-58 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาวิทยาศาสตร์ ระหว่างวิธี IRT LRT และวิธี MGCFA-ALIGMENT จำแนกตามขนาดโรงเรียน พบว่า ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี IRT LRT สอดคล้องกับวิธี MGCFA-ALIGMENT มีค่าไควร์-สแควร์ (χ^2) เท่ากับ 8.000 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

ตารางที่ 4-59 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาวิทยาศาสตร์ วิธี IRT LRT และวิธี MIMIC จำแนกตามขนาดโรงเรียน

ผลการตรวจสอบ DIF	วิธี MIMIC		Chi Sq (χ^2)
	NODIF (n = 40)	DIF (n = 40)	(Sig.)
วิธี IRT LRT			.015
NODIF	5.0%	32.5%	(1.000)
DIF	7.5%	55.0%	

จากตารางที่ 4-59 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาวิทยาศาสตร์ ระหว่างวิธี IRT LRT และวิธี MIMIC จำแนกตามขนาดโรงเรียน พบว่า ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี IRT LRT ไม่สอดคล้องกับวิธี MIMIC มีค่าไควร์-สแควร์ (χ^2) เท่ากับ .015

ตารางที่ 4-60 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาวิทยาศาสตร์ วิธี MGCFA-ALIGMENT และวิธี MIMIC จำแนกตามขนาดโรงเรียน

ผลการตรวจสอบ DIF	**วิธี MIMIC**		**Chi Sq (χ^2)**
	NODIF (n = 40)	**DIF (n = 40)**	**(Sig.)**
วิธี ALIGMENT			1.905
NODIF	12.5%	62.5%	(.306)
DIF	0.0%	25.0%	

จากตารางที่ 4-60 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาวิทยาศาสตร์ ระหว่างวิธี MGCFA-ALIGMENT และวิธี MIMIC จำแนกตามขนาดโรงเรียน พบว่า ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี MGCFA-ALIGMENT ไม่สอดคล้องกับวิธี MIMIC มีค่าไคว์-สแควร์ (χ^2) เท่ากับ 1.905

ตารางที่ 4-61 สรุปผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาคณิตศาสตร์ วิชาภาษาไทย และ วิชาวิทยาศาสตร์ ด้วย 3 วิธี จำแนกตามสถานที่ตั้งทางภูมิศาสตร์และขนาดโรงเรียน

Subjects/ Methods	IRT LRT	ALIGNMENT	MIMIC
MATH_REGION			
IRT LRT			
ALIGNMENT	×		
MIMIC	×	×	
THAI_REGION			
IRT LRT			
ALIGNMENT	×		
MIMIC	✓	×	
SCIENCE_REGION			
IRT LRT			
ALIGNMENT	×		
MIMIC	✓	×	
MATH_SIZE			
IRT LRT			
ALIGNMENT	×		
MIMIC	×	×	
THAI _SIZE			
IRT LRT			
ALIGNMENT	×		
MIMIC	✓	✓	
SCIENCE_SIZE			
IRT LRT			
ALIGNMENT	✓		
MIMIC	×	×	

หมายเหตุ เครื่องหมาย ✓ = ผลสอดคล้องกัน; × = ผลไม่สอดคล้องกัน

จากตารางที่ 4-61 สรุปผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ วิชาภาษาไทย และ วิชาวิทยาศาสตร์ ด้วย 3 วิธี จำแนกตามสถานที่ตั้งทางภูมิศาสตร์และขนาดโรงเรียน พบว่า

1. ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ และขนาดโรงเรียน ผลการตรวจสอบการทำหน้าที่ต่างกันของทั้ง 3 วิธีไม่สอดคล้องกัน

2. ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ วิธี IRT LRT สอดคล้องกับวิธี MIMIC ส่วน วิธี IRT กับวิธี MGCFA-ALIGMENT และ วิธี MGCFA-ALIGMENT กับวิธี MIMIC ไม่สอดคล้องกัน เมื่อจำแนกตามขนาดโรงเรียน วิธี IRT LRT และวิธี MGCFA-ALIGMENT สอดคล้องกับวิธี MIMIC ส่วนวิธี IRT กับวิธี MGCFA-ALIGMENT ไม่สอดคล้องกัน

3. ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาวิทยาศาสตร์ จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ วิธี IRT LRT สอดคล้องกับวิธี MIMIC แต่ไม่สอดคล้องกับวิธี MGCFA-ALIGMENT ส่วนวิธี MGCFA-ALIGMENT ไม่สอดคล้องกันกับวิธี MIMIC เมื่อจำแนกตามขนาดโรงเรียน วิธี IRT LRT สอดคล้องกับวิธี MGCFA-ALIGMENT ส่วน วิธี IRT กับวิธี MIMIC และ วิธี MGCFA-ALIGMENT กับวิธี MIMIC ไม่สอดคล้องกัน

สรุปได้ว่า ความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ 3 วิธี 3 วิชา ทำการทดสอบ 9 คู่ พบว่าสอดคล้อง 2 คู่ (ร้อยละ 22.22) เมื่อจำแนกตามขนาดโรงเรียน 3 วิธี 3 วิชา ทำการทดสอบ 9 คู่ พบว่าสอดคล้อง 3 คู่ (ร้อยละ 33.33)

บทที่ 5
สรุปผล อภิปราย และข้อเสนอแนะ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT วิธี MGCFA-ALIGNMENT และวิธี MIMIC และเพื่อเปรียบเทียบผลของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT วิธี MGCFA-ALIGNMENTและวิธี MIMIC ภายใต้ปัจจัยที่ต่างกัน 2 ปัจจัย ได้แก่ สถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนและขนาดโรงเรียน ข้อมูลที่ใช้ในการศึกษาครั้งนี้เป็นข้อมูลผลการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ระดับชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2559 ซึ่งผู้วิจัยได้นำข้อมูลมาวิเคราะห์หาข้อสรุปตามวัตถุประสงค์ของการวิจัย ข้อมูลที่นำมาศึกษาได้แก่ สถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน ขนาดของโรงเรียน และคำตอบของนักเรียนรายบุคคลที่ผ่านการตรวจให้คะแนนแบบทวิวิภาค (1-0) วิชาคณิตศาสตร์ วิชาภาษาไทย และวิชาวิทยาศาสตร์ กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2559 สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี IRT LRT วิธี MGCFA-ALIGNMENT และวิธี MIMIC ถ้าวิธีตรวจสอบวิธีใด ตรวจพบข้อสอบที่ทำหน้าที่ต่างกันสอดคล้องกับเกณฑ์ ถือว่าวิธีนั้นตรวจสอบการทำหน้าที่ต่างกันได้ถูกต้อง

สรุปผลการวิจัย

ผลการวิจัยสรุปได้ดังนี้

1. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT วิธี MGCFA ALIGNMENT และวิธี MIMIC เป็นการตรวจสอบการทำหน้าที่ต่างกัน ของข้อสอบ ภายใต้ทฤษฎีการตอบสนองข้อสอบ (IRT) แบบ 2 พารามิเตอร์ ตรวจสอบทั้ง 2 รูปแบบ คือ แบบเอกรูป (Uniform DIF) และแบบอเนกรูป (Nonuniform DIF) ภายใต้ปัจจัย ที่ต่างกัน 2 ปัจจัย ได้แก่ สถานที่ตั้งทางภูมิศาสตร์ และขนาดโรงเรียน การศึกษาครั้งนี้ ศึกษากับผลการตอบวิชาคณิตศาสตร์ จำนวน 25 ข้อ วิชาภาษาไทย และวิชาวิทยาศาสตร์ จำนวนข้อสอบวิชาละ 40 ข้อ เป็นข้อสอบที่ตรวจให้คะแนนแบบทวิวิภาค (1-0)

1.1 วิชาคณิตศาสตร์ ข้อสอบจำนวน 25 ข้อ

ปัจจัยสถานที่ตั้งทางภูมิศาสตร์ วิธี MIMIC ตรวจพบข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด จำนวน 22 ข้อ ส่วนวิธี MGCFA-ALIGMENTMIMIC พบข้อสอบที่ทำหน้าที่ต่างกันน้อยที่สุด

จำนวน 2 ข้อ ข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ จากทั้ง 3 วิธี ได้แก่ข้อที่ 12 และข้อที่ 17

ปัจจัยขนาดโรงเรียน วิธี MIMICพบข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด จำนวน 24 ข้อ ส่วนวิธี วิธี MGCFA-ALIGMENT พบข้อสอบที่ทำหน้าที่ต่างกันน้อยที่สุด จำนวน 9 ข้อ ข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ จากทั้ง 3 วิธี ได้แก่ ข้อที่ 3, 4, 5, 8, 17, 18, 19 และข้อที่ 21

1.2 วิชาภาษาไทย ข้อสอบจำนวน 40 ข้อ

ปัจจัยสถานที่ตั้งทางภูมิศาสตร์ วิธี MIMIC ตรวจพบข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด จำนวน 28 ข้อ ส่วนวิธี MGCFA-ALIGMENTMIMIC พบข้อสอบที่ทำหน้าที่ต่างกันน้อยที่สุด จำนวน 3 ข้อ ข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ จากทั้ง 3 วิธี ได้แก่ข้อที่ 26 และข้อที่ 33

ปัจจัยขนาดโรงเรียน วิธี IRT LRT พบข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด จำนวน 33 ข้อ ส่วนวิธี วิธี MGCFA-ALIGMENT พบข้อสอบที่ทำหน้าที่ต่างกันน้อยที่สุด จำนวน 3 ข้อ ข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ จากทั้ง 3 วิธี ได้แก่ ข้อที่ 17, 25 และข้อที่ 40

1.3 วิชาวิทยาศาสตร์ ข้อสอบจำนวน 40 ข้อ

ปัจจัยสถานที่ตั้งทางภูมิศาสตร์ วิธี MIMIC ตรวจพบข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด จำนวน 26 ข้อ ส่วนวิธี MGCFA-ALIGMENTMIMIC พบข้อสอบที่ทำหน้าที่ต่างกันน้อยที่สุด จำนวน 8 ข้อ ข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ จากทั้ง 3 วิธี ได้แก่ ข้อที่ 18, 23, 26, 27, 28, 34 และข้อที่ 35

ปัจจัยขนาดโรงเรียน วิธี MIMIC พบข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด จำนวน 35 ข้อ ส่วนวิธี วิธี MGCFA-ALIGMENT พบข้อสอบที่ทำหน้าที่ต่างกันน้อยที่สุด จำนวน 10 ข้อ ข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ จากทั้ง 3 วิธี ได้แก่ ข้อที่ 1, 17, 22, 23, 24, 26, 27, 28, 34 และข้อที่ 35

2. การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT วิธี MGCFA-ALIGNMENTและวิธี MIMIC ภายใต้ปัจจัยที่ต่างกัน 2 ปัจจัย ได้แก่ สถานที่ตั้งทางภูมิศาสตร์ และขนาดโรงเรียน พิจารณาจากความสอดคล้องของทั้ง 3 วิธีในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ผลการวิเคราะห์ พบว่า ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ วิชาภาษาไทย และวิชาวิทยาศาสตร์ ด้วย 3 วิธี จำแนกตามสถานที่ตั้งทางภูมิศาสตร์และขนาดโรงเรียน

2.1 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ และขนาดโรงเรียน ผลการตรวจสอบการทำหน้าที่ต่างกันของทั้ง 3 วิธีไม่สอดคล้องกัน

2.2 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ วิธี IRT LRT สอดคล้องกับวิธี MIMIC ส่วน วิธี IRT กับวิธี ALIGNMENT และวิธี ALIGNMENT กับวิธี MIMIC ไม่สอดคล้องกัน เมื่อจำแนกตามขนาดโรงเรียน วิธี IRT LRT และวิธี ALIGNMENT สอดคล้องกับวิธี MIMIC ส่วน วิธี IRT กับวิธี ALIGNMENT ไม่สอดคล้องกัน

2.3 ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาวิทยาศาสตร์ จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ วิธี IRT LRT สอดคล้องกับวิธี MIMIC แต่ไม่สอดคล้องกับวิธี MGCFA-ALIGMENT ส่วนวิธี MGCFA-ALIGMENT ไม่สอดคล้องกันกับวิธี MIMIC เมื่อจำแนกตามขนาดโรงเรียน วิธี IRT LRT สอดคล้องกับวิธี ALIGNMENT ส่วน วิธี IRT กับวิธี MIMIC และ วิธีALIGNMENT กับวิธี MIMIC ไม่สอดคล้องกัน

สรุปได้ว่า ความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ 3 วิธี 3 วิชา ทำการทดสอบ 9 คู่ พบว่า สอดคล้อง 2 คู่ (ร้อยละ 22.22) เมื่อจำแนกตามขนาดโรงเรียน 3 วิธี 3 วิชา ทำการทดสอบ 9 คู่ พบว่า สอดคล้อง 3 คู่ (ร้อยละ 33.33)

2.4 ขนาดโรงเรียนตรวจสอบพบการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มที่มากกว่าสองกลุ่มมากกว่าสถานที่ตั้งทางภูมิศาสตร์

2.5 วิธี MIMIC ตรวจสอบพบข้อสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มที่มากกว่าสองกลุ่มมากที่สุด รองลงมาได้แก่วิธี IRT LRT และวิธี MGCFA-ALIGNMENT

อภิปรายผล

การวิจัยครั้งนี้มุ่งประสงค์เพื่อศึกษาผลของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สำหรับกลุ่มที่มากกว่าสองกลุ่ม 3 วิธี ได้แก่ วิธี IRT LRT วิธี MGCFA-ALIGNMENT และวิธี MIMIC ภายใต้ปัจจัยที่ต่างกัน 2 ปัจจัย ได้แก่ สถานที่ตั้งทางภูมิศาสตร์ 4 กลุ่ม และขนาดโรงเรียน 4 กลุ่ม จากผลการวิจัยที่สรุปได้ข้างต้นมีประเด็นสำคัญที่นำมาอภิปรายผลดังนี้

1. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT วิธี MGCFA-ALIGNMENT และวิธี MIMIC พบว่า วิธี MIMIC และวิธี IRT LRT พบข้อสอบที่ทำหน้าที่ต่างกันมากกว่าวิธี MGCFA-ALIGMENT อาจเนื่องมากจากเงื่อนไขของการตรวจสอบ DIF และวิธีการทดสอบ กล่าวคือ วิธี MIMIC เป็นการตรวจสอบการทำหน้าที่ต่างกัน

ของข้อสอบโดยมีการทดสอบเป็นรายคู่อิทธิพลระหว่างตัวแปรกลุ่มกับข้อสอบที่เป็นข้อเป้าหมาย เมื่อมี 4 กลุ่ม ต้องทำการทดสอบทั้ง 3 กลุ่มเปรียบเทียบ หากพบว่า มีนัยสำคัญทางสถิติก็สามารถบ่งชี้ได้ว่าข้อสอบข้อนั้นทำหน้าที่ต่างกัน ข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูปพิจารณาจากดัชนีปรับแต่งที่ได้จากโปรแกรมและระดับนัยสำคัญสำหรับข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปพิจารณาปฏิสัมพันธ์ระหว่างระดับความสามารถกับการเป็นสมาชิกของกลุ่มที่ส่งอิทธิพลไปยังข้อสอบที่เป็นเป้าหมาย หากพบว่า มีนัยสำคัญทางสถิติก็สามารถบ่งชี้ได้ว่าข้อสอบข้อนั้นทำหน้าที่ต่างกัน จากงานวิจัยที่ศึกษาโมเดล MIMIC ไม่สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ด้วยข้อจำกัดของเครื่องมือที่ใช้ในการวิเคราะห์ แต่งานวิจัยครั้งนี้ได้นำแนวคิดและวิธีการพร้อมมีเครื่องมือที่ใช้วิเคราะห์รองรับ ทำให้ได้สารสนเทศเพิ่มขึ้นที่เป็นประโยชน์ต่อการเรียนรู้และสังเคราะห์ผลเพื่อนำไปใช้ให้เกิดประโยชน์ได้ ด้วยแนวคิดของโมเดล MIMIC เป็นโมเดลหนึ่งของ SEM ที่มีแบบจำลองผสมระหว่างการวิเคราะห์การถดถอยและการวิเคราะห์องค์ประกอบแบบจำลองกลมกลืนของโมเดลง่ายเพราะความแตกต่างระหว่างกลุ่มถูกกลบรวมเป็นเมทริกซ์ความแปรปรวนร่วมเดียวโดยที่การกำหนดแบบจำลองแบบ MIMIC คือ การกำหนดแบบจำลองที่ให้ตัวแปรจำแนกกลุ่ม ซึ่งการศึกษาครั้งนี้ได้แก่ สถานที่ตั้งทางภูมิศาสตร์ และขนาดโรงเรียนปรากฏเป็นตัวแปรหนึ่งในแบบจำลองที่ชัดเจน และกำหนดให้ข้อสอบเป็นตัวแปรสังเกตที่ใช้วัดตัวแปรแฝงซึ่งการศึกษาครั้งนี้คือ คะแนนสอบวิชาคณิตศาสตร์ ภาษาไทย และวิทยาศาสตร์ บทบาทของตัวแปรที่ใช้จำแนกกลุ่มนี้คือ ตัวแปรสาเหตุที่ส่งผลต่อตัวแปรแฝง มีระดับการวัดเป็นนามบัญญัติ (Nominal scale) การกำหนดแบบจำลองแบบ MIMIC เหมือนกันกับการใช้การวิเคราะห์การถดถอยที่ต้องแปลงตัวแปรกลุ่มให้เป็นตัวแปรหุ่นและใส่ตัวแปรนี้เข้าไปในสมการถดถอย และตรวจสอบความแตกต่างของค่าเฉลี่ยโดยดูจากค่าสัมประสิทธิ์ถดถอยของตัวแปรหุ่นว่ามีนัยสำคัญหรือไม่ ค่าสัมประสิทธิ์นี้ก็คือผลต่างของค่าเฉลี่ยระหว่างกลุ่มที่ให้ค่าเป็นเลข 1 และ 0 ในตัวแปรหุ่น การกำหนดแบบจำลองแบบ MIMIC จะง่ายและสะดวกเพราะเพียงแค่เพิ่มตัวแปรหุ่นเข้าไปในแบบจำลองก็สามารถวิเคราะห์ได้แล้วการอ่านแปลผลก็ตรงไปตรงมาและการที่มีกลุ่มที่ต้องการเปรียบเทียบมากขึ้นความยุ่งยากและซับซ้อนของการวิเคราะห์และการตีความหมายก็จะยิ่งมากตามไปด้วยโดยเฉพาะถ้าวิเคราะห์แล้วพบว่า ค่าสัมประสิทธิ์ที่เกี่ยวข้องกับการวัดตัวแปรแฝงมีความแตกต่างกันระหว่างกลุ่มเช่นค่าน้ำหนักองค์ประกอบ (Factor loadings) ก็ต้องเพิ่มการวิเคราะห์เพื่อหาให้ได้ว่าค่าน้ำหนักองค์ประกอบของตัวแปรสังเกตใดบ้างที่แตกต่างกันและแตกต่างกันระหว่างกลุ่มใด ดังนั้น ในกรณีที่มีหลายกลุ่มการวิเคราะห์ด้วยแบบจำลอง MIMIC จะง่ายและต่อการวิเคราะห์แต่ต้องระวังไม่ละเมิดข้อตกลงเบื้องของ แบบจำลอง MIMIC ข้อตกลงเบื้องต้นที่เพิ่มขึ้นคือการวัดตัวแปรแฝงผลการทดลองที่จะเปรียบเทียบนั้นจะต้องมีความเท่ากัน

ระหว่างกลุ่มอย่างเคร่งครัดทุกค่าสัมประสิทธิ์ที่เกี่ยวข้องกับการวัดตัวแปรแฝง (Strictmeasurement invariance) สอดคล้องกับผลการวิจัยของผลการวิจัยของ Seokjoon (2014) ได้ศึกษาประสิทธิภาพของโมเดล MIMIC ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและเอนกรูปได้เมื่อมีกลุ่มตัวอย่างที่ศึกษาพร้อมกันหลายกลุ่ม และพบว่า เมื่อมีการละเมิดข้อตกลงเบื้องต้นเกี่ยวกับความแปรปรวนของตัวแปรแฝง อัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้น อย่างไรก็ตาม วูดส์ (Woods, 2009) ได้ตั้งข้อสังเกตว่าหากไม่มีปฏิสัมพันธ์กันในโมเดล โมเดล MIMIC มีประสิทธิภาพดีกว่า IRT-LR-DIF ยกเว้นข้อสอบเป็นแบบ Nonuniform DIF

และสอดคล้องกับแนวคิดการตรวจสอบเป็นรายคู่ของวิธี IRT LRT วิธีนี้ให้ผลการวิเคราะห์ที่แสดงให้เห็นถึงผลการเปรียบเทียบระหว่างกลุ่มอ้างอิงกับกลุ่มเปรียบเทียบเป็นรายคู่ (Contrast groups) โดยได้กำหนดเกณฑ์สำหรับพิจารณาว่าข้อสอบข้อใดที่มีความต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ระหว่างกลุ่มแม้เพียงคู่เดียวก็ถือว่าข้อสอบข้อนั้นทำหน้าที่ต่างกัน ซึ่งมีทั้งหมด 3 (Contrast groups) ที่ได้ทดสอบ ดังนั้น ผลการวิจัยครั้งนี้ผู้วิจัยได้นำเสนอทุกข้อที่พบว่าผลการวิเคราะห์รายข้อที่มี 1 กลุ่มอ้างอิง 3 กลุ่มเปรียบเทียบ มีข้อสอบที่ทำหน้าที่ต่างกันทุกคู่, ต่างกัน 2 ใน 3 และ ต่างกัน 1ใน 3 พบในทั้ง 3 วิชา และที่เป็นเช่นนี้วิธี IRT LRT มีข้อตกลงเบื้องต้นที่เข้มแข็งและมากมาย เช่น ฟังก์ชันการตอบสนองข้อสอบต้องเป็นสมการโลจิสติค และการแจกแจงของคุณลักษณะแฝง และคุณลักษณะหลักของผู้สอบ (θ) ต้องแจกแจงเป็นปกติ ซึ่งถ้าฟังก์ชันการตอบสนองข้อสอบหรือข้อตกลงเบื้องต้นของคุณลักษณะหลักของผู้สอบ (θ) ของกลุ่มผู้สอบย่อยกลุ่มใดกลุ่มหนึ่ง หรือทั้งสองกลุ่มมีการแจกแจงไม่ปกติความคลาดเคลื่อนประเภทที่ 1 จะเกิด 2-4 เท่าของระดับ α ที่กำหนดไว้ (Bolt, 2002) ดังนั้น ข้อควรปฏิบัติก่อนใช้วิธี IRT LRT จะต้องมีการตรวจสอบให้เป็นไปตามข้อตกลงเบื้องต้น หากไม่เป็นไปตามข้อตกลงจะได้หาแนวทางแก้ไขต่อไปโดยอาจสุ่มตัวอย่างใหม่หรือเก็บข้อมูลเพิ่มเติม วิธีนี้มีข้อดีตรงที่โปรแกรม IRTPRO ที่ใช้ในการวิเคราะห์ไม่ต้องเขียนคำสั่งด้วยตนเอง ไม่ยุ่งยาก เวลาที่ใช้ในการวิเคราะห์ข้อมูลรวดเร็ว ผลการวิเคราะห์จากโปรแกรมทำความเข้าใจได้ง่ายไม่ซับซ้อน แต่ต้องเตรียมข้อมูลให้มีความถูกต้องชัดเจน ซึ่งสอดคล้องกับที่ อาวีพร ปานทอง (2556) ได้ให้ข้อเสนอแนะไว้ว่าให้ใช้วิธีทดสอบอัตราส่วนความควรจะเป็น (IRT LRT) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เพื่อทำให้ควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ และมีอำนาจการทดสอบสูง

แต่เมื่อมาพิจารณาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สำหรับกลุ่มที่มากกว่าสองกลุ่ม ด้วย วิธี MGCFA ALIGNMENT พบว่า สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มที่มากกว่าสองกลุ่มได้ สอดคล้องกับการศึกษาของ Asparouhov and Muthén (2014) ที่พบว่า วิธี MGCFA ALIGNMENT เป็นวิธีที่สามารถวิเคราะห์การทำหน้าที่ต่างกันของ

ข้อสอบแบบหลายกลุ่มได้ ความแกร่งของวิธีนี้คือสามารถประมาณค่าโมเดลได้อย่างสะดวกสำหรับหลาย ๆ กลุ่ม เช่น การเปรียบเทียบระหว่างหลายประเทศ วิธีการนี้เป็นวิธีที่พัฒนามาจากแนวคิดวิธี MGCFA ที่มีข้อจำกัดคือไม่สามารถวิเคราะห์ได้เมื่อมีหลายกลุ่มและมีดัชนีปรับแต่งจำนวนมากเกินไป วิธี ALIGNMENT อยู่บนพื้นฐานโมเดลที่มีการกำหนดค่า (Configural model) เป็นหลักโดยอัตโนมัติและช่วยลดความยุ่งยากในการวัดการวิเคราะห์การไม่แปรเปลี่ยน และประการสำคัญวิธีนี้ยังให้รายละเอียดของค่าพารามิเตอร์ที่ไม่แปรเปลี่ยนทั้งค่าความยากและค่าอำนาจจำแนก สำหรับพารามิเตอร์ในโมเดลของทุกกลุ่ม สอดคล้องกับ Finch (2016) ได้ศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มที่มากกว่าสองกลุ่มผลการศึกษาพบว่า The generalized Mantel-Haenszel test, and the multiple group alignment procedure เป็นวิธีที่ดีที่สุดสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 และมีอำนาจการทดสอบจากผลการศึกษาสรุปได้ว่าการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี ALIGNMENT มีจุดเด่นอยู่ 4 ประการ คือ 1) เป็นวิธีที่เปรียบเทียบแต่ละกลุ่มโดยอัตโนมัติในขั้นตอนเดียว 2) ให้ค่าพารามิเตอร์ทั้งค่าความยาก และค่าอำนาจจำแนกของข้อสอบในการตรวจหา DIF ในหลายกลุ่ม และ3) สามารถตรวจสอบได้ทั้ง Uniform และ Nonuniform DIF 4) การเตรียมข้อมูลและการวิเคราะห์ข้อมูลไม่ซับซ้อน ใช้เวลาน้อยกว่าวิธี MIMIC แต่ใช้เวลามากกว่าวิธี IRT LRT แต่สามารถวิเคราะห์ในขั้นตอนเดียว เช่นเดียวกับวิธี IRT LRT นักวิจัยหรือนักวิชาการสามารถใช้วิธีนี้ในการศึกษาข้อมูลเฉพาะของกลุ่มศึกษาข้อมูลเชิงลึกเกี่ยวกับคุณภาพของการประมาณค่าพารามิเตอร์เหล่านี้ ผลของการศึกษาครั้งนี้จะทำให้ผู้เชี่ยวชาญด้านการวัดผล ได้มองเห็นประโยชน์ในแง่ของการตัดสินใจว่าจะดำเนินการตรวจสอบ DIF สำหรับหลายกลุ่มได้อย่างไร ผลการตรวจสอบพบข้อสอบที่ทำหน้าที่ต่างกันน้อยแตกต่างจาก 2 วิธีที่กล่าวมา อาจเนื่องมาจากวิธีนี้เดิมใช้สำหรับทดสอบการทำหน้าที่ต่างกันของข้อสอบกับกลุ่มที่มีหลายกลุ่ม เช่น การเปรียบเทียบระหว่างประเทศ ซึ่งอาจมีความแตกต่างมากกว่าภายในประเทศเดียวกันแต่แบ่งตามสถานที่ตั้งทางภูมิศาสตร์ และขนาดโรงเรียน การตรวจสอบจึงทำให้พบน้อยกว่าวิธีอื่น

ทั้งสามวิธีต่างอยู่บนทฤษฎีเดียวกัน คือ ทฤษฎี IRT แบบ 2 พารามิเตอร์ มีความสามารถในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มที่มากกว่าสองกลุ่มหรือหลายกลุ่มต่างมีข้อดีและข้อจำกัดตามสภาพของวิธี เป็นหน้าที่ของนักวิจัยนักการศึกษาหรือนักวัดผลที่ต้องศึกษาต้องเลือกใช้วิธีที่มีความเหมาะสมกับปัญหาและความสนใจของตนเองเพื่อก่อเกิดประโยชน์สูงสุด

2. การเปรียบเทียบผลของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สำหรับกลุ่มที่มากกว่าสองกลุ่มด้วยวิธี IRT LRT วิธี MGCFA ALIGNMENTและวิธี MIMIC ภายใต้ปัจจัยที่ต่างกัน 2 ปัจจัย ได้แก่ สถานที่ตั้งทางภูมิศาสตร์ และขนาดโรงเรียน การศึกษาครั้งนี้จำแนกกลุ่มตาม

ปัจจัยสถานที่ตั้งทางภูมิศาสตร์ได้แบ่งออกเป็น 4 ภาค ประกอบด้วย ภาคกลาง ซึ่งในการวิจัยครั้งนี้ได้กำหนดให้เป็นกลุ่มอ้างอิง (R) หรือกลุ่มที่คาดว่าได้เปรียบในการตอบข้อสอบ และภาคเหนือ ภาคตะวันออกเฉียงเหนือ และภาคใต้ เป็นกลุ่มเปรียบเทียบ (F) หรือกลุ่มที่คาดว่าจะเสียเปรียบในการตอบข้อสอบ ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สำหรับกลุ่มที่มากกว่าสองกลุ่ม โดยใช้แนวคิดทฤษฎี IRT แบบ 2 พารามิเตอร์ด้วยวิธี IRT LRT วิธี MGCFA-ALIGNMENTและวิธี MIMIC พบว่า ผลความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ วิชาภาษาไทย และ วิชาวิทยาศาสตร์ ด้วย 3 วิธี จำแนกตามสถานที่ตั้งทางภูมิศาสตร์และขนาดโรงเรียน ผลการวิจัยพบว่า ความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ และขนาดโรงเรียน ผลการตรวจสอบการทำหน้าที่ต่างกันของทั้ง 3 วิธีไม่สอดคล้องกัน ในขณะที่วิชาภาษาไทย จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ วิธี IRT LRT สอดคล้องกับวิธี MIMIC ส่วนวิธี IRT กับวิธี ALIGNMENT และวิธี ALIGNMENT กับวิธี MIMIC ไม่สอดคล้องกัน เมื่อจำแนกตามขนาดโรงเรียน วิธี IRT LRT และวิธี ALIGNMENT สอดคล้องกับวิธี MIMIC ส่วน วิธี IRT กับวิธี ALIGNMENT ไม่สอดคล้องกัน ในขณะที่วิชาวิทยาศาสตร์ จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ วิธี IRT LRT สอดคล้องกับวิธี LIGNMENT และวิธี MIMIC ส่วน วิธี ALIGNMENT ไม่สอดคล้องกันกับวิธี MIMIC เมื่อจำแนกตามขนาดโรงเรียน วิธี IRT LRT สอดคล้องกับวิธี ALIGNMENT ส่วน วิธี IRT กับวิธี MIMIC และวิธีALIGNMENT กับวิธี MIMIC ไม่สอดคล้องกัน

จะเห็นได้ว่าเมื่อพิจารณาโดยรวมผลจากทั้ง 3 วิธี พบว่า สถานที่ตั้งทางภูมิศาสตร์ 3 วิธี 3 วิชา ทำการทดสอบ 9 คู่ พบว่า สอดคล้อง 3 คู่ (ร้อยละ 33.33) เมื่อจำแนกตามขนาดโรงเรียน 3 วิธี 3 วิชา ทำการทดสอบ 9 คู่ พบว่า สอดคล้อง 3 คู่ (ร้อยละ 33.33) เป็นผลที่ทำให้เห็นว่า 3 วิชา มีความต่างกันตามธรรมชาติของวิชา ต่างมาตรฐานการเรียนรู้ ต่างเนื้อหา และต่างตัวชี้วัด ดังนั้น ควรพิจารณาแยกรายวิชา สำหรับในรายวิชาเดียวกันเมื่อพิจารณาความสอดคล้องของผลที่เกิดขึ้นไม่สอดคล้องกันนั้น อาจเนื่องมาจากเกณฑ์พิจารณาแตกต่างกัน สิ่งที่ได้จากการตรวจสอบต้องพิจารณาลงไปถึงรายละเอียดของข้อมูลแต่ละข้อ ว่าแตกต่างในลักษณะใด แบบเอกรูป หรือแบบอเนกรูป DIF กลุ่มทุกกลุ่ม หรือ กลุ่มใดกลุ่มหนึ่ง จึงสามารถนำสารสนเทศไปใช้ให้เกิดประโยชน์ได้ ในส่วนของวิชาภาษาไทย เป็นวิชาที่เกี่ยวกับการใช้ภาษาเพื่อการสื่อสาร กฎเกณฑ์ของภาษาและการใช้ในชีวิตประจำวัน จะเห็นได้จากการวิเคราะห์ด้วยวิธี IRT LRT กับวิธี MIMIC ว่ามีความสอดคล้องในประเด็นการเปรียบเทียบเป็นรายคู่ ไม่ได้มีภาคใดภาคหนึ่งเสียเปรียบตลอด เป็นข้อมูลที่นำไปพิจารณาต่อลงไปถึงเนื้อหาสาระของวิชาที่ใช้สอบ

ดังนั้น เมื่อพบว่า ร้อยละของความสอดคล้องแตกต่างกันให้มาพิจารณาความสามารถของทั้ง 3 วิธี พบว่า ทั้ง 3 วิธีสามารถตรวจสอบพบการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มที่มากกว่าสองกลุ่มได้ สอดคล้องกับการศึกษาของ มิวเธน และ แอสพารอฟ (Asparouhov & Muthén, 2014, 2013b) ที่ได้ศึกษาทฤษฎีการตอบสนองข้อสอบ (IRT) กับกลุ่มหลายกลุ่ม ได้แก่ วิธี ALIGNMENTวิธีการดำเนินการวิจัยเพื่อการศึกษาประสิทธิภาพของวิธีศึกษากับข้อมูลการสอบ CIVED1999 (14 กลุ่ม) และ ICSS2009 (14 กลุ่ม) อีกทั้งจำลองข้อมูลเพื่อชี้ให้เห็นว่าวิธีที่นำมาใช้มีประสิทธิภาพ ผลการศึกษาพบว่า วิธี ALIGNMENT มีประสิทธิภาพสามารถวิเคราะห์ได้และสอดคล้องกับผลการวิจัยของ Finch (2016) ที่ได้ศึกษาเชิงเปรียบเทียบวิธีALIGNMENT กับวิธีอื่นในการประยุกต์ใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มที่มากกว่าสองกลุ่ม โดยใช้ข้อมูลจำลอง ใช้อัตราความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบ บ่งบอกประสิทธิภาพของวิธี ปัจจัยที่ศึกษาได้แก่ ขนาดของกลุ่มตัวอย่าง และขนาดของการทำหน้าที่ต่างกันของข้อสอบ และกล่าวว่าปัจจัยที่เป็นจริงที่นักวิจัยสนใจอาจได้แก่ เชื้อชาติ หรือเศรษฐานะทางสังคม (เป็นกลุ่มที่แบ่งได้มากกว่า 2 กลุ่ม) ผลการศึกษาพบว่า The generalized Mantel-Haenszel test, and the multiple group alignment procedure เป็นวิธีที่ดีที่สุดสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 และมีอำนาจการทดสอบ เดิมมีข้อจำกัดในเชิงวิธีการทางสถิติและโปรแกรมที่ใช้วิเคราะห์ ชี้ให้เห็นว่าปัจจุบันได้มีพัฒนาการการศึกษาวิจัยเพื่อค้นพบวิธีการแก้ข้อจำกัดต่าง ๆ อย่างต่อเนื่องสอดคล้องกับงานวิจัยของมัคและคณะ (Munck et al., 2014 cited in Muthén & Asparouhov, 2014) ที่ได้ศึกษากับประเทศ 92 ประเทศ ปัจจัยความแตกต่างทางวัฒนธรรมและปัจจัยระดับประเทศ ซึ่งมีหลายกลุ่มมาก วิธีนี้มีประสิทธิภาพสามารถวิเคราะห์ข้อมูลดังกล่าวได้ตามวัตถุประสงค์ของการศึกษา ดังนั้น นักวิจัยสามารถใช้วิธีนี้ในการศึกษาข้อมูลเฉพาะของกลุ่มทั้งค่าความยาก และค่าอำนาจจำแนก ของข้อสอบ ดังนั้น จึงเป็นเรื่องสำคัญที่จะต้องมีข้อมูลเชิงลึกเกี่ยวกับคุณภาพของการประมาณค่าพารามิเตอร์เหล่านี้ในการศึกษาเกี่ยวกับความลำเอียงของข้อสอบประสิทธิภาพและค่าพารามิเตอร์ จากผลการวิจัยชี้ให้เห็นว่าวิธี Alignment อาจเหมาะสมที่สุดสำหรับการประเมิน Uniform DIF เมื่อมีมากกว่าสองกลุ่ม สรุปได้ว่าจากผลการวิจัยและงานวิจัยที่เกี่ยวข้อง วิธี Alignment มีข้อดี สามประการ 1) เป็นวิธีที่เปรียบเทียบแต่ละกลุ่มโดยอัตโนมัติในขั้นตอนเดียว 2) ให้ค่าพารามิเตอร์ทั้งค่าความยาก และค่าอำนาจจำแนกของข้อสอบในขณะที่วิธีการของ Lord สามารถใช้เพื่อให้ได้สิ่งเหล่านี้ได้เช่นกันพลังงานที่ต่ำเมื่อเทียบกับการจัดแนวทำให้น้อยกว่าที่เหมาะสมสำหรับการตรวจหาDIF หลายกลุ่ม 3) สามารถวัดได้ทั้ง Uniform and Nonuniform DIF และถึงแม้ตามเจตนารมณ์เดิมของวิธีใหม่นี้เป็นการศึกษากับข้อมูลที่มาหลายกลุ่ม

ซึ่งครั้งนี้ศึกษาสูงสุดแค่ 4 กลุ่มเท่านั้น แต่ผู้วิจัยนำมาประยุกต์ใช้กับข้อมูลที่เป็นจริงเพื่อนำไปใช้ประโยชน์ได้จริงหากมีประสิทธิภาพ

เมื่อพิจารณาปัจจัยขนาดโรงเรียน การศึกษาครั้งนี้ตรวจสอบพบว่า เมื่อจำแนกตามขนาดโรงเรียนผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มที่มากกว่าสองกลุ่มพบข้อสอบที่ทำหน้าที่ต่างกันมากกว่าสถานที่ตั้งทางภูมิศาสตร์ สอดคล้องกับผลการวิจัยของ Kim (2000) ศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการทดสอบอัตราส่วนไลด์ลิฮู้ด (Likelihood ratio test) วิธีแมนเทล Mantel และวิธีแมนเทล-แฮนเซลส์แบบทั่วไป (Generalized Mantel-Haenszel) ข้อมูลจากการประเมินโรงเรียนระดับอนุบาลของรัฐจอเจียร์ ปัจจัยที่ศึกษาเป็นขนาดของกลุ่มตัวอย่าง ผลการศึกษา พบว่า ทั้ง 3 วิธี สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ดีเมื่อกลุ่มตัวอย่างมีขนาด 100 คน เช่นเดียวกับงานวิจัยของ อิทธิฤทธิ์ พงษ์ปิยะรัตน์ (2550) ได้ศึกษาเกี่ยวกับการวิเคราะห์ข้อสอบและการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ: การวิเคราะห์พหุระดับ ผลการวิจัยพบว่า ปัจจัยผลการเรียนคณิตศาสตร์ในภาคเรียนที่ผ่านมาส่งผลต่อค่าเฉลี่ยของโอกาสในการตอบข้อสอบได้ถูกต้องในแต่ละโรงเรียนอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ผลการวิเคราะห์ระดับโรงเรียน (ระดับ 3) พบว่า ตัวแปรขนาดของโรงเรียนและตัวแปรความเป็นผู้นำทางวิชาการของผู้บริหารส่งผลต่อค่าเฉลี่ยของโอกาสในการตอบข้อสอบได้ถูกต้องในโรงเรียน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และ .05 ตามลำดับ การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยโมเดล HGLM สามารถตรวจสอบพบข้อสอบที่ทำหน้าที่ต่างกันสอดคล้องตรงกับผลการวิเคราะห์จากโปรแกรม BILOG-MG ทั้งนี้อาจเนื่องมาจาก ขนาดของโรงเรียนเกี่ยวข้องกับผู้เรียนในส่วนของความพร้อมด้านการจัดการศึกษาอันประกอบด้วย บุคคล งบประมาณ สภาพแวดล้อมภายในโรงเรียน โดยโรงเรียนที่มีขนาดใหญ่มักเป็นโรงเรียนที่ได้เปรียบในเชิงความพร้อมในการจัดการเรียนการสอนมากกว่าโรงเรียนขนาดเล็ก การพิจารณางบประมาณและจำนวนครูผู้สอนในแต่ละโรงเรียนจะใช้ฐานการคำนวณจากจำนวนนักเรียนต่อครูเป็นหลัก ดังนั้น การได้รับงบประมาณสนับสนุน และการมีครูจำนวนเพียงพอจะพบว่า โรงเรียนขนาดใหญ่นั้นจะมีความพอดีมากกว่าโรงเรียนขนาดเล็ก

เมื่อพิจารณาสถานที่ตั้งทางภูมิศาสตร์ตรวจพบข้อสอบทำหน้าที่ต่างกันสถานที่ตั้งทางภูมิศาสตร์ของกลุ่มตัวอย่างจึงมีความสำคัญในการทดสอบการทำหน้าที่ต่างกันของข้อสอบ เพราะภาษาที่ใช้ในข้อสอบเป็นภาษาไทยกลาง และทั้งสามวิชาเป็นข้อสอบที่เกี่ยวข้องกับสถานการณ์สถานที่ ข้อสอบที่ดีย่อมไม่ทำหน้าที่ต่างกันเมื่อกลุ่มตัวอย่างสถานที่ตั้งทางภูมิศาสตร์ต่างกัน สอดคล้องกับผลการศึกษาของ สุพัฒนา หอมบุปผา (2556) ได้ศึกษาเรื่อง การเปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี HGML วิธี MIMIC และวิธี BAYESIAN ศึกษากับข้อมูลคะแนน

การสอบวัดผลสัมฤทธิ์ทางการเรียน เพื่อประเมินคุณภาพระดับชาติ ของนักเรียนชั้นประถมศึกษาปีที่ 3 ศึกษาปัจจัยเพศ และสถานที่ตั้งทางภูมิศาสตร์ ผลการศึกษาพบว่า เมื่อจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน ส่วนใหญ่สาเหตุที่ทำให้ข้อสอบเกิดการทำหน้าที่ต่างกันอาจเป็นเพราะประสบการณ์ ความคุ้ยเคยเกี่ยวกับเรื่องนั้น สภาพแวดล้อม และการฝึกปฏิบัติที่แตกต่างกันระหว่างนักเรียนในเขตกรุงเทพมหานครและปริมณฑล นักเรียนนอกเขตกรุงเทพมหานครและปริมณฑล เกษร หว่างจิตร (2539) ได้ศึกษาการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทยและวิชาภาษาอังกฤษของศูนย์ทดสอบทางการศึกษาคณะครุศาสตร์จุฬาลงกรณ์มหาวิทยาลัยด้วยวิธีแมนเทล-เฮนส์เซลเมื่อจำแนกกลุ่มผู้สอบตามเพศภูมิลำเนาประสบการณ์ในการสอบและสังกัดของสถานศึกษาโดยมีกลุ่มตัวอย่างเป็นผู้สอบวิชาภาษาไทยจำนวน 506 คน และผู้สอบวิชาภาษาอังกฤษจำนวน 501 คน ผลการวิจัยพบว่า ข้อสอบที่พบ DIF ส่วนใหญ่เป็นแบบอเนกรูปข้อสอบที่ DIF ทั้งแบบเอกรูปและอเนกรูปเกิดขึ้นเมื่อการวิเคราะห์จำแนกกลุ่มผู้สอบตามเพศมากที่สุดรองลงมาคือจำแนกตามภูมิลำเนา สังกัดของสถานศึกษา รักชนก ยี่สุ่นศรี (2544) ได้ศึกษาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบด้วยกระบวนการ ดี เอฟ ไอ ที กับแบบสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาของทบวงมหาวิทยาลัย ปีการศึกษา 2543 ครั้งที่ 1 วิชาภาษาอังกฤษและวิชาคณิตศาสตร์ จากกลุ่มตัวอย่าง จำนวน 4,000 คน และ 3,600 คน ตามลำดับ ทำการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบสำหรับกลุ่มผู้สอบเมื่อจำแนกตามเพศ และสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา ผลการวิจัย พบว่า แบบสอบวิชาภาษาอังกฤษทำหน้าที่ต่างกันตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษาของผู้สอบ ส่วนแบบสอบวิชาคณิตศาสตร์เมื่อจำแนกกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษามีข้อสอบที่พบ DIF

สิ่งที่เป็นเป้าหมายสำคัญในการศึกษาครั้งนี้ มีสองประเด็นหลัก ประการแรก เป็นการนำเสนอวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 3 วิธีโดยใช้ทฤษฎีเดียวกันคือทฤษฎี การตอบสนองข้อสอบ แบบ 2PL ประการที่สอง นำเสนอการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มที่มีมากกว่าสองกลุ่มหรือหลายกลุ่ม ซึ่งเป็นจริงตามภาวะของสังคมโลก ประการที่สาม เป็นการศึกษาเชื่อมโยงการศึกษากับข้อมูลจำลองที่เกิดผลดีมาใช้กับข้อมูลจริง ซึ่งมีข้อจำกัดในการกำหนดเงื่อนไขของการศึกษา เช่น ขนาดของการทำหน้าที่ต่างกัน การแจกแจงความสามารถของกลุ่ม ขนาดกลุ่มตัวอย่าง และขนาดความยาวของข้อสอบ ทั้งสิ้นนั้นเพื่อคุณค่าของการศึกษาการพัฒนาข้อสอบหรือแบบสอบให้มีความยุติธรรมอย่างแท้จริง

ข้อเสนอแนะ

ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1. จากผลการตรวจสอบพบว่า วิธี IRT LRT สอดคล้องกับวิธี MIMIC สามารถนำสารสนเทศที่เกิดจากการตรวจพบไปศึกษาลักษณะข้อสอบที่ทำหน้าที่ต่างกัน เพื่อเป็นแนวทางในการพัฒนาคุณข้อสอบต่อไป

2. จากผลการศึกษาสามารถเป็นแนวทางสำหรับการศึกษาเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบกรณีเกณฑ์ในการจำแนกกลุ่มผู้สอบมีมากกว่า 2 กลุ่ม เช่น การพัฒนาเครื่องมือในการสอบระดับชาติ หรือการพัฒนาเครื่องมือมาตรฐาน บนพื้นฐานทฤษฎีการตอบสนองข้อสอบ (IRT)

3. จากผลการศึกษาครั้งนี้มีแนวทางในการเลือกใช้วิธีภายใต้ทฤษฎีการตอบสนองข้อสอบ (IRT) แบบ 2 พารามิเตอร์ 2PL ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่ตรวจให้คะแนนแบบ 1-0 ดังนี้ ประการแรกจำนวนกลุ่ม วิธี IRT LRT และวิธี MIMIC มีความเหมาะสมสำหรับกลุ่มที่มีน้อยกลุ่ม เช่น 2 หรือ 3 กลุ่ม ในขณะที่วิธี MGCFA-ALIGNMENT เป็นวิธีที่พัฒนามาจากข้อจำกัดของวิธี MGCFA ปกติ ที่มีข้อจำกัดในการไม่สามารถวิเคราะห์ได้เมื่อมีหลายกลุ่ม และมีดัชนีปรับแต่งจำนวนมากเกินไป จึงเป็นวิธีที่มีความเหมาะสมสำหรับกลุ่มที่มีมากเพราะเนื่องจากการวิเคราะห์จะสรุปมาให้อย่างชัดเจนว่ากลุ่มใดไม่เข้าพวก โดยผู้วิเคราะห์มีความสะดวกในการสรุปผล เพื่อนำไปใช้ต่อไป ประการที่สอง ชนิดของข้อสอบที่ทำหน้าที่ต่างกัน เสนอแนะว่าทั้ง 3 วิธี สามารถตรวจสอบได้ทั้ง 2 รูปแบบ คือ Uniform DIF และ Nonuniform DIF

ข้อเสนอแนะในการวิจัยครั้งต่อไป

1. ควรทำการศึกษาลักษณะเดียวกับการศึกษาครั้งนี้ โดยศึกษากับตัวแปรอื่นในการเปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบ เพิ่มตัวแปรที่เป็นเกณฑ์ในการจำแนกกลุ่มที่คาดว่ามีผลต่อการทำหน้าที่ต่างกันของข้อสอบที่มีมากกว่าสองกลุ่ม เช่น สังกัดโรงเรียน ประเภทโรงเรียน ภาษา และเกณฑ์อื่น ๆ พิจารณาผลการตรวจสอบกับผลการศึกษาครั้งนี้ต่อไป

2. ควรทำการศึกษาลักษณะเดียวกับการศึกษาในครั้งนี้ ซึ่งได้ศึกษากับขนาดกลุ่มตัวอย่างระหว่างกลุ่มอ้างอิงกับกลุ่มเปรียบเทียบมีจำนวนเท่ากัน ดังนั้น ควรทำการศึกษาสัดส่วนขนาดของกลุ่มตัวอย่าง จำนวนข้อสอบ

3. ควรศึกษาเปรียบเทียบประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี IRT LRT วิธี MGCFA-ALIGNMENT และวิธี MIMIC โดยใช้ผลการตอบข้อสอบ V-NET, I-NET, B-NET และ N-NET ของนักเรียน หรือข้อมูลผลการทดสอบระดับนานาชาติของประเทศไทย วิเคราะห์ตามตัวแปรที่ใช้ในการศึกษาครั้งนี้หรือเพิ่มตัวแปรอื่น ๆ เช่น สังกัด

4. ควรมีการศึกษาพัฒนาเกณฑ์การเปรียบเทียบประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี IRT LRT วิธี MGCFA-ALIGNMENT และวิธี MIMIC ควบคู่กันระหว่างการใช้การจำลองข้อมูลกับใช้ข้อมูลจริงในเงื่อนไขที่สามารถกำหนดให้ศึกษาร่วมกันได้ เช่น สัดส่วนของกลุ่มตัวอย่าง ประเภทของการทำหน้าที่ต่างกัน

บรรณานุกรม

กระทรวงศึกษาธิการ. (2551). *หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551.* กรุงเทพฯ: โรงพิมพ์ชุมนุมสหกรณ์การเกษตรแห่งประเทศไทย.

กระทรวงศึกษาธิการ. (2560). *ประกาศกระทรวงศึกษาธิการ เรื่อง กำหนดสถานที่ตั้งของสำนักงานศึกษาธิการภาค.* กรุงเทพฯ: กระทรวงศึกษาธิการ.

กัลยา วานิชย์บัญชา. (2556). *การวิเคราะห์สมการโครงสร้าง (SEM) ด้วย AMOS.* กรุงเทพฯ: สามลดา.

กาญจนา วัธนสุนทร. (2537). *การพัฒนาเกณฑ์ตัดสินข้อสอบลำเอียงทางเพศ.* วิทยานิพนธ์ครุศาสตรดุษฎีบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา, คณะครุศาสตร์, จุฬาลงกรณ์มหาวิทยาลัย.

เกษร หว่างจิตร. (2539). *การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบสาหรับแบบสอบคัดเลือกระดับบัณฑิตศึกษาวิชาภาษาไทยและภาษาอังกฤษด้วยวิธีแมนเทล-แฮนส์เซล.* วิทยานิพนธ์ครุศาสตรมหาบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา, คณะครุศาสตร์, จุฬาลงกรณ์มหาวิทยาลัย.

จิติมา วรรณศรี. (2539). *การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซลกับวิธีซิบเทสท์ เมื่อความยาวของแบบสอบขนาดกลุ่มตัวอย่างและอัตราส่วนของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบต่างกัน.* วิทยานิพนธ์ครุศาสตรมหาบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา, คณะครุศาสตร์, จุฬาลงกรณ์มหาวิทยาลัย.

ชนะศึก นิชานนท์. (2553). *ประสิทธิภาพของการประมาณค่าพารามิเตอร์แบบเบส์ โดยใช้การสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ.* วิทยานิพนธ์ครุศาสตรดุษฎีบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา, คณะครุศาสตร์, จุฬาลงกรณ์มหาวิทยาลัย.

ชัยยศ ชาวระนอง. (2553). *ประสิทธิภาพของโมเดลและการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบพหุมิติโดยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยันแฝงภายใน.* ดุษฎีนิพนธ์ปรัชญาดุษฎีบัณฑิต, สาขาวิชาวิจัย วัดผลและสถิติการศึกษา, คณะศึกษาศาสตร์, มหาวิทยาลัยบูรพา.

ดิลกะ ลัทธพิพัฒน์. (2559). *ประเด็นท้าทายเกี่ยวกับโรงเรียนขนาดเล็ก และทางเลือกในการพัฒนาคุณภาพการศึกษาของประเทศไทย.* เข้าถึงได้จาก http://blogs.worldbank.org/eastasiapacific/th/node/3406.

ธีรเดช ฉายอรุณ. (2560). *Structural equation models*. เข้าถึงได้จาก https: //www.ncbi.nlm.nih.gov/pmc/articles/PMC4162377

นงลักษณ์ วิรัชชัย. (2542). *โมเดลลิสเรล สถิติวิเคราะห์สำหรับการวิจัย* (พิมพ์ครั้งที่ 3). กรุงเทพฯ: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

นพมาศ พิพัฒนสุข. (2541). *การเปรียบเทียบประสิทธิภาพระหว่างวิธีแมนเทล-แฮนส์เซลกับวิธีถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อใช้เกณฑ์จับคู่เปรียบเทียบแตกต่างกันในแบบสอบชนิดพหุมิติ.* วิทยานิพนธ์ครุศาสตรมหาบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา, คณะครุศาสตร์, จุฬาลงกรณ์มหาวิทยาลัย.

นิคม กีรติวรางกูร. (2542). *การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีการวิเคราะห์องค์ประกอบจำกัด แมนเทล-แฮนส์เซลและการตอบสนองข้อสอบ.* วิทยานิพนธ์ครุศาสตรมหาบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา, คณะครุศาสตร์, จุฬาลงกรณ์มหาวิทยาลัย.

นิตยา เหมือดไชสง. (2543). *การส่งอิทธิพลผ่านตัวกลางเชิงสาเหตุของปัจจัยด้านนักเรียน ครู และโรงเรียนไปยังผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์: การวิเคราะห์อภิมานงานวิจัย.* วิทยานิพนธ์ครุศาสตรมหาบัณฑิต, สาขาวิชาวิจัยและจิตวิทยาการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.

พนัส จันทร์เปล่ง. (2557). *การวิเคราะห์เปรียบเทียบโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์: การประยุกต์ใช้โมเดลมูลค่าเพิ่มพหุระดับที่มีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ.* วิทยานิพนธ์ครุศาสตรดุษฎีบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา, คณะครุศาสตร์, จุฬาลงกรณ์มหาวิทยาลัย.

พิชิต ธรรมรักษ์. (2549). *ปัจจัยที่มีผลต่อผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ของนักเรียนแผนการเรียนศิลป์ภาษา ในกรุงเทพมหานคร*. วิทยานิพนธ์ครุศาสตรมหาบัณฑิต, สาขาวิชาวิจัยและจิตวิทยาการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.

ไพรัตน์ วงษ์นาม. (2546). *หลักการวิจัยทางการศึกษา*. ชลบุรี: คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา.

ยุทธ ไกยวรรณ์. (2553). *หลักสถิติวิจัยและการใช้โปรแกรม SPSS*. กรุงเทพฯ: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

รักชนก ยี่สุ่นศรี. (2544). *การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบด้วยกระบวนการ ดี เอฟ ไอ ที สำหรับแบบสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษา วิชาภาษาอังกฤษและวิชาคณิตศาสตร์.* วิทยานิพนธ์ครุศาสตรมหาบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา, คณะครุศาสตร์, จุฬาลงกรณ์มหาวิทยาลัย.

รัชรินทร์ มุคดา. (2540). *การเปรียบเทียบประสิทธิภาพระหว่างวิธีแมนเทล-แฮนส์เซลกับวิธีถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันของแบบสอบแบบอเนกรูปในกรณีที่จัดกลุ่มความสามารถ ค่าความยากของข้อสอบ และค่าอำนาจจำแนกของข้อสอบต่างกัน.* วิทยานิพนธ์ครุศาสตรมหาบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา, คณะครุศาสตร์, จุฬาลงกรณ์มหาวิทยาลัย.

เรวดี อินทะสะระ. (2539). *ผลการตรวจสอบความลำเอียงของข้อสอบต่อการศึกษาความเที่ยงตรงเชิงพยากรณ์ของแบบทดสอบคัดเลือกที่คิดคะแนนต่างกัน.* ปริญญานิพนธ์การศึกษาดุษฎีบัณฑิต, สาขาวิชาการทดสอบและวัดผลการศึกษา, บัณฑิตวิทยาลัย, มหาวิทยาลัยศรีนครินทรวิโรฒ.

วลีมาศ แซ่อึ้ง. (2543). *การเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป ระหว่างวิธีซิปเทสท์ปรับใหม่วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก.* วิทยานิพนธ์ครุศาสตรดุษฎีบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา, คณะครุศาสตร์, จุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนวาสี. (2545). *ทฤษฎีการทดสอบแนวใหม่ (Modern test theories)* (พิมพ์ครั้งที่ 2). กรุงเทพฯ: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนวาสี. (2555). *ทฤษฎีการทดสอบแนวใหม่ (Modern test theories)* (พิมพ์ครั้งที่ 4). กรุงเทพฯ: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน). (2560). *การทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน O-NET (Ordinary national educational test)*. เข้าถึงได้จาก http://www.niets.or.th/th/

สมโภชน์ อเนกสุข. (2553). *วิธีการทางสถิติสำหรับการวิจัย* (พิมพ์ครั้งที่ 4). ชลบุรี: คณะศึกษาศาสตร์, มหาวิทยาลัยบูรพา

สัมพันธ์ พันธุ์พฤกษ์. (2560). *การนำผลการทดสอบ O-NET ไปใช้พัฒนาการเรียนการสอน และยกระดับผลสัมฤทธิ์ทางการเรียน*. เข้าถึงได้จาก http://www.acn.ac.th/2012/onet/pdf/Development.pdf

สำนักงานสถิติแห่งชาติ. (2553). *มองเมืองไทยด้วยสถิติ*. เข้าถึงได้จาก http://service.nso.go.th/nso/thailand/thailand.jsp

สุพัฒนา หอมบุปผา. (2556). *การเปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี HGLM วิธี MIMIC และวิธี BAYESIAN*. ดุษฎีนิพนธ์ปรัชญาดุษฎีบัณฑิต, สาขาวิชาวิจัย วัดผลและสถิติการศึกษา, คณะศึกษาศาสตร์, มหาวิทยาลัยบูรพา.

สุภะ อภิญญาภิบาล. (2558). *การวิเคราะห์เปรียบเทียบประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี*. วิทยานิพนธ์ครุศาสตรดุษฎีบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา, คณะครุศาสตร์, จุฬาลงกรณ์มหาวิทยาลัย.

สุมาลี แกล้วทนงค์. (2547). *สาเหตุของการทำหน้าที่ต่างกันของข้อสอบสาระการเรียนรู้ภาษาไทยและสาระการเรียนรู้สังคมศึกษา ศาสนาและวัฒนธรรม*. วิทยานิพนธ์ครุศาสตรมหาบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา, คณะครุศาสตร์, จุฬาลงกรณ์มหาวิทยาลัย.

สุวิมล ติรกานันท์. (2551). *การสร้างเครื่องมือวัดตัวแปรในการวิจัยทางสังคมศาสตร์*: แนวทางสู่การปฏิบัติ (พิมพ์ครั้งที่ 2). กรุงเทพฯ: สำนักพิมพ์พิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

สุวิมล ติรกานันท์. (2553). *การวิเคราะห์ตัวแปรพหุในงานวิจัยทางสังคมศาสตร์*. กรุงเทพฯ: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

สุวิมล ติรกานันท์. (2555). *การวิเคราะห์ตัวแปรพหุในงานวิจัยทางสังคมศาสตร์* (พิมพ์ครั้งที่ 2). กรุงเทพฯ: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

เสรี ชัดแช้ม. (2539). *การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันแบบไม่สม่ำเสมอของข้อสอบระหว่างวิธีแมนเทล-แฮนส์เซล แบบปกติ กับ วิธีแมนเทล-แฮนส์เซลแบบแบ่งกลุ่มความสามารถของผู้สอบและความยากของข้อสอบ*. วิทยานิพนธ์ครุศาสตรดุษฎีบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา, คณะครุศาสตร์, จุฬาลงกรณ์มหาวิทยาลัย.

อริสพา เตหลิ่ม. (2559). *การเปรียบเทียบประสิทธิผลการประมาณค่าพารามิเตอร์และการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมกซิมัมไลค์ลิฮู้ด วิธีของเบส์และวิธีของเบส์แบบมีอิทธิพลเทสต์เลท*. ดุษฎีนิพนธ์ปรัชญาดุษฎีบัณฑิต, สาขาวิชาวิจัย วัดผลและสถิติการศึกษา, คณะศึกษาศาสตร์, มหาวิทยาลัยบูรพา.

อาวีพร ปานทอง. (2558). *การเปรียบเทียบประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบให้คะแนนหลายค่า โดยทดสอบอัตราส่วนความควรจะเป็น วิธีเบส์เซียนและวิธีโพลี-ซิปเทสท์*. ดุษฎีนิพนธ์ปรัชญาดุษฎีบัณฑิต, สาขาวิชาวิจัย วัดผลและสถิติการศึกษา, คณะศึกษาศาสตร์, มหาวิทยาลัยบูรพา.

อาวีพร ปานทอง. (2560). การพัฒนาทักษะการคิดขั้นสูงเพื่อการประยุกต์ใช้ในการดำเนินชีวิตสำหรับนักศึกษา คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครสวรรค์. *วารสารพัฒนาเทคนิคศึกษา, 29*(101), 74-81.

อิทธิฤทธิ์ พงษ์ปิยะรัตน์. (2551). *การวิเคราะห์ข้อสอบและการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ: การวิเคราะห์พหุระดับ*. วิทยานิพนธ์ครุศาสตรดุษฎีบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา, คณะครุศาสตร์, จุฬาลงกรณ์มหาวิทยาลัย.

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Belmont, CA: Brooks/ Cole.

Angoff, W. H. (1993). Perspectives on differential item functioning methodology. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 3-23). Hillsdale, NJ: Lawrence Erlbaum Associates.

Asparouhov, T., & Muthén, B. (2014). IRT studies of many groups: The alignment meth*od. Frontiers in Psychology, 5*, 1-7.

Atar, B., & Kamata, A. (2011). Comparison of IRT likelihood ratio test and logistic regression DIF detection procedures. *H.U. Journal of education, 41*, 36-47.

Bolt, D. (2002). A monte carlo comparison of paramrtric and nonparamrtric polytomous DIFdetection methods. *Applied Measurement in Education, 15*, 113-141.

Borsboom, D., Mellenbergh, G. J., & Heerden, J. V. (2002). Different kinds of DIF: A distinction between absolute and relative forms of measurement invaiance and bias. *Applied psychological measurement, 26*(4), 433-450.

Brennan, R. L. (2006). *Educational measurement* (4th ed.). Connecticut: Praeger.

Burcu, A., & Akihito, K. (2011). Comparison of IRT likelihood ratio test and logistic regression

DIF detection procedures. *H.U. Journal of Educational, 41*, 36-47.

Cai, L., Thissen, D., & Toit, S. H. C. (2011). *IRTPRO for windows*. Lincolnwood, IL: Scientific Software International.

Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. Thousand Oaks, California: Sage.

Chun, S., Stark, S., Kim, E. S., & Chermvshenko, O. S. (2016). MIMIC Methods for detecting DIF among multiple groups exploring a new sequential-free baseline procedure. *Applied Psychological Measurement, 40*(7), 486-499.

Cohen, A. S., Kim, S. H., & Wollack, J. A. (1996). An investigation of the likelihood ratio test for detection of differential item functioning. *Applied Psychological Measurement, 20*, 15-26.

Crane, P. K. (2006). Differential item functioning analysis with ordinal logistic regression techniques DIFdetect and difwithpar. *Medical Care, 44*(11), 115-123.

De Ayala, R. J. (2009). *The theory and practice of item response theory*. New York: Adivision of Guilford.

DeMars, C. E. (2009). Modification of the mantel-haenszel and logistic regression DIF procedures to incorporate the SIBTEST regression correction. *Journal of Educational and Behavioral Statistics, 34*(2), 149-170.

DeMars, C. (2010). *Item response theory*. New York: Oxford University Press.

DeMars, C. E., & Lau, A. (2011). Differential item functioning detection with latent classes: How accurately can we detect who is responding differentially? *Educational and Psychological Measurement, 71*(4), 597-616.

Finch, W. H. (2005). The MIMIC model as a method for detecting DIF: Comparison with mantel-haenszel, SIBTEST and the IRT likelihood ratio. *Applied Psychological Measurement, 29*(5), 278-295.

Finch, W. H. (2016). Detection of differential item functioning for more than two Groups: A monte carlo comparison of methods. *Applied Measurement in Education, 29*(1), 30-45.

Finch, W. H., & French, B. F. (2007). Detection of crossing differential item functioning: A comparison of four methods. *Educational and Psychological Measurement, 67*(4), 565-582.

Fleishman, J. (2003). Using MIMIC models to assess the influence of differential item functioning. *Center for Financing, Access, and Cost Trends Agency for Healthcare Research and Quality, 1*(2), 6.

Fukuhara, H., & Kamata, A. (2007). *DIF detection in a presence of locally dependent items*. Florida: Florida Educational Research Association.

Fukuhara, H., & Kamata, A. (2011). A differential item functioning model for testlet-Based items using a bi-factor multidimensional item response theory model: A bayesian approach. *Applied Psychological Measurement, 35*(8), 604-622.

Hair, Jr, J. F., Black, W.C., Babin, B.J., Anderson, R.E. (2010). *Multivariate data analysis* (7th ed.). New Jersey: Pearson prentice Hall.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item Response theory*. California: Sage.

Harpole, J. K. (2015). *A Bayesian MIMIC model for testing non-uniform DIF in two and three groups*. Graduate Theses and Dissertations, The Graduate Faculty of the University of Kansas.

Holland, P. W., & Thayer, D. T. (1988). Differential Item Performance and the Mantel-Haenszel Procedure. In H. Wainer, & H. I. Brown (Eds.), *Test validity* (pp. 129-145). Hillsdale, NJ: Lawrence Erlbaum Associates.

Holland, P. W., & Wainer, H. (1993). *Differential item functioning*. New Jersey: Lawrence Erlbaum Associates.

Joreskog, K., & Goldberger, A. S. (1975). Estimation of a model with a multiple indicators and multiple causes of a single latent variable. *Journal of American Statistical Association, 70*, 631-639.

Joreskog, K. G., & Sorbom, D. (1996). *LISREL8 User's reference guide* (2nd ed.). Chicago: Scientific Software International.

Kim, S. H. (2000). An investication of the likelihood ratio test, the mantel test and the Generalized mantel-haenzel test of DIF. *Applied Measurement in Education, 13*(1), 155-169.

Kim, S. H., Chosen, A. S., & Kim, S. (2007). DIF Detection and effect size measures for Polytomously scored items. *Journal of Educational Measurement, Summer, 44*(2), 93-116.

Le, D. T. (2013). *Applying item response theory modeling in educational research. A dissertation submitted to the graduate faculty in partial fullment of the requirements for the degree of doctor of philosophy, Co-majors: Statistics; education (Curriculum and instructional technology)*. Iowa: Iowa State University Press.

Lei, P. W., Chen, S.Y., & Yu, L. (2006). Comparing methods of assessing differential item functioning in a computerized adaptive testing environment. *Journal of Educational Measurement, 43*(3), 245-264.

Linda, C., & James, A. (2008). *Introduction to classical & modern test theory*. Mason, Ohio: Cengage Learning.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22*, 719-748.

Marie, W. (2007). *Measuring and detecting differential item functioning in criterion-referenced licensing test, A theoretic comparison of methods.* Postadress: Umea University Press.

Muthén, L. K., & Muthén, B. O. (2007). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.

Muthén, B. (1985). A method for studying the homogeneity of test items with respect to other relevant variables. *Journal of Educational Statistics, 10*, 121-132.

Muthén, B. (1988). Some uses of structural equation modeling in validity studies: Extending IRT to external variables. In H. Wainer, & H. Braun (Eds.), *Test validity* (pp. 213-238). Hillsdale, NJ: Erlbaum Associates.

Muthén, B. (1989). Factor structure in groups selected on observed scores. *British Journal of Mathematical and Statistical Psychology, 42*, 81-90.

Muthén, B. O., Kao, C. F., & Burstein, L. (1991). Instructionally sensitive psychometrics: Application of a new IRT-based detection technique to mathematics Achievement test items. *Journal of Educational Measurement, 28*(1), 1-22.

Oishi, S. (2006). The concept of life satisfaction across culture: An IRT Analysis. *Journal of Research in Personality, 40*, 411-423.

Osterlind, S. J., & Everson, H. T. (2009). *Differential item functioning.* Los Angeles: Sage.

Park, H. (1995). *Differential item functioning in performance assessments: A comparison of three procedures.* Retrieved from https://digitalcommons.lsu.edu/gradschool_ disstheses/6124

Penfield, R. D. (2001). Assessing differential item functioning across multiple groups: A comparison of three mantel-haenszel procedures. *Applied Measurement in Education, 14*, 235-259.

Penfield, R. D., & Algina, J. (2006). A generalized DIF effect variance estimator for Measuring global differential test functioning in mixed format tests. *Journal of Educational Measurement, 43*, 295-312.

Popham, W. J. (1981). *Modern educational measurement.* Englewood Cliffs, NJ: Prentice-Hall.

Potenza, M. T., & Dorans, N. J. (1995). DIF assessment for polytomously scored items: A framework for classification and evaluation. *Applied Psychological Measurement, 19*, 23-37.

Rogers, H. J., & Swaminathan, H. (1993). A comparison of logistic regression and mantel-haenszel procedures for detecting differential item functioning. *Applied Psychological Measurement, 17*(2), 105-116.

Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores.* Richmond: Psychometric Society.

Seokjoon, C. (2014). *Using MIMIC methods to detect and identify sources of DIF among multiple groups*. Tampa, FL: University of South Florida Press.

Shealy, R., & Stout, W. (1993). A model-based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DTF as well as item bias/DIF. *Psychometrika, 58*(2), 159-194.

Sheppard, R., Han, K. T., Colarelli, S. M., Dai, G., & King, D. W. (2006). Differential item functioning by sex and race in the hogan personality inventory. *Assessment, 13*(4), 442-453.

Shin, C. L., & Wang, W. C. (2009). Differential item functioning detection using the multiple indicators, Multiple causes method with a pure short anchor. *Applied Psychological Measurement, 33*(3), 184-199.

Stark, S., Chernyshenko, O. S., & Drasgow, F. (2006). Detecting differential item functioning with comfirmatory factor analysis and item response theory: Toward a unified strategy. *Journal of Applied Psychology, 91*, 1292-1306.

Stroud, A. H., & Secrest, D. (1996). *Gaussian quadrature formulas*. Englewood Cliffs, NJ: Prentice-Hall.

Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement, 27*(4), 361-370.

Thissen, D. (2001). *IRTLRDIF v.2.0b (Computer program).* Chapel Hall, LL: University of North Carolina Press.

Thissen, D., Steinberg, L., & Wainer, H. (1988). Use of item response theory in the study of group differences in trace lines. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 147-172). Hillsdale, NJ: Lawrence Erlbaum Associates.

Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 67-113). Hillsdale, NJ: Lawrence Erlbaum Associates.

Woods, C. M. (2009). Evaluation of MIMIC-model methods for DIF testing with comparison to two-group analysis. *Multivariate Behavioral Research, 44*, 1-27.

Woods, C. M., & Grimm, K. J. (2011). Testing for nonuniform differential item functioning with multiple indicator multiple cause models. *Applied Psychological Measurement, 35*(5), 339-361.

Woods, C. M., & Oltmanns.T. F., & Turkheimer, E. (2009). Illustration of MIMIC-model DIF testing with the schedule for nonadaptive and adaptive personality. *The Journal of Psychopathology and Behavioral Assessment, 31*(4), 320-330.

Wu, L., & Tsai, R. (2010). *A comparison of tree polytomous DIF detection methods.* Taiwan: The National Science Council of Taiwan.

Zieky, M. (1993). Practical questions in the use of DIF statistics in test development. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 337-347). Hillsdale, NJ: Erlbaum.

Zumbo, B. D. (1999). *A Handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and likert-type (Ordinal) item scores.* Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.

ภาคผนวก

ภาคผนวก ก

หนังสือขอข้อมูลจากสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน)

(สำเนา)

ที่ ศธ ๖๒๑๘/๒๒๓๙

คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา
๑๖๙ ถ.ลงหาดบางแสน ต.แสนสุข
อ.เมือง จ.ชลบุรี ๒๐๑๓๑

๖ ธันวาคม ๒๕๖๐

เรื่อง ขอความอนุเคราะห์ผลสอบรายข้อ

เรียน ผู้อำนวยการสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน)

สิ่งที่ส่งมาด้วย โครงร่างงานวิจัย จำนวน ๑ ฉบับ

ด้วย นางสาริศา คงมี นิสิตปริญญาเอก หลักสูตรปรัชญาดุษฎีบัณฑิต สาขาวิชาวิจัย วัดผลและสถิติการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา ได้รับอนุมัติให้ทำดุษฎีนิพนธ์ เรื่อง "การเปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มที่มากกว่าสองกลุ่ม ด้วยวิธี IRT LRT วิธี MGCFA - ALIGNMENT และวิธี MIMIC" โดยมี รองศาสตราจารย์ ดร.ไพรัตน์ วงษ์นาม เป็นอาจารย์ที่ปรึกษาหลัก และ ดร.ณัฐกฤตา งามมีฤทธิ์ เป็นอาจารย์ที่ปรึกษาร่วม ในการทำวิจัยครั้งนี้จำเป็นต้องใช้ข้อมูลผลการสอบรายข้อจากการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ระดับชั้นมัธยมศึกษาปีที่ ๓ ประจำปีการศึกษา ๒๕๕๙ เพื่อนำข้อมูลมาวิเคราะห์หาข้อสรุปตามวัตถุประสงค์ของการวิจัย ข้อมูลที่ผู้วิจัยต้องการขอความอนุเคราะห์จากสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) ได้แก่ ข้อมูลเพศผู้สอบ สถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน ขนาดของโรงเรียน และคำตอบของนักเรียนรายบุคคลที่ผ่านการตรวจให้คะแนนแบบ ๑-๐ วิชาภาษาไทย วิชาคณิตศาสตร์ และวิชาวิทยาศาสตร์ ประจำปีการศึกษา ๒๕๕๙ โดยต้องการขอข้อมูลเฉพาะโรงเรียนมัธยมศึกษา สังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ และข้อสอบฉบับจริงวิชาดังกล่าว ซึ่งผลการวิจัยครั้งนี้จะเป็นประโยชน์ในเชิงวิชาการในด้านการประยุกต์ใช้วิธีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) ตามทฤษฎีการตอบสนองข้อสอบ ระหว่างการวิเคราะห์ข้อมูลผู้วิจัยจะเก็บรักษาข้อมูลดังกล่าวไว้เป็นความลับ และเข้าถึงข้อมูลได้เฉพาะข้าพเจ้าและผู้ร่วมวิเคราะห์ข้อมูลเท่านั้น เมื่อผู้วิจัยทำการวิจัย และสร้างข้อสรุปในภาพรวมเรียบร้อยแล้วจะดำเนินการลบข้อมูลดังกล่าวทิ้งทันที

จึงเรียนมาเพื่อโปรดพิจารณา คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา หวังเป็นอย่างยิ่งว่าคงจะได้รับความอนุเคราะห์จากท่านด้วยดี และขอขอบคุณอย่างสูงมา ณ โอกาสนี้

ขอแสดงความนับถือ

(ผู้ช่วยศาสตราจารย์ ดร.เชษฐ์ ศิริสวัสดิ์)
รองคณบดีฝ่ายบัณฑิตศึกษา ปฏิบัติการแทน
คณบดีคณะศึกษาศาสตร์ ปฏิบัติการแทน
ผู้ปฏิบัติหน้าที่อธิการบดีมหาวิทยาลัยบูรพา

ภาควิชาวิจัยและจิตวิทยาประยุกต์
โทรศัพท์ ๐-๓๘๑๐-๒๐๗๖
โทรสาร ๐-๓๘๓๙-๓๒๕๑
ผู้วิจัยโทร. ๐๘-๑๙๖๙-๐๖๓๕

ภาคผนวก ข

ตัวอย่างผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ

ด้วยโปรแกรม IRTPRO และโปรแกรม Mplus

โปรแกรม IRTPRO

วิธี IRT LRT

IRTPRO Version 2.1

Output generated by IRTPRO estimation engine Version 4.54 (32-bit)

Project:	DIF REGION (CENTRAL NORTH NORTHEST SOUTH)
Description:	CENTRAL=1 NORTH=2 NORTHEST=3 SOUTH=4
Date:	22 April 2018
Time:	10:54 PM

Table of Contents

2PL Model Item Parameter Estimates for Group 1, logit: $a\theta + c$ or $a(\theta - b)$

Summed-Score Based Item Diagnostic Tables and X^2s for Group 1

2PL Model Item Parameter Estimates for Group 2, logit: $a\theta + c$ or $a(\theta - b)$

Summed-Score Based Item Diagnostic Tables and X^2s for Group 2

2PL Model Item Parameter Estimates for Group 3, logit: $a\theta + c$ or $a(\theta - b)$

Summed-Score Based Item Diagnostic Tables and X^2s for Group 3

2PL Model Item Parameter Estimates for Group 4, logit: $a\theta + c$ or $a(\theta - b)$

Summed-Score Based Item Diagnostic Tables and X^2s for Group 4

Group Parameter Estimates

DIF Statistics for Graded Items

DIF Contrast among Groups:

Marginal fit (X^2) and Standardized LD X^2 Statistics for Group 1

Marginal fit (X^2) and Standardized LD X^2 Statistics for Group 2

Marginal fit (X^2) and Standardized LD X^2 Statistics for Group 3

Marginal fit (X^2) and Standardized LD X^2 Statistics for Group 4

Item Information Function Values for Group 1 at 15 Values of θ from -2.8 to 2.8

Item Information Function Values for Group 2 at 15 Values of θ from -2.8 to 2.8

Item Information Function Values for Group 3 at 15 Values of θ from -2.8 to 2.8

Item Information Function Values for Group 4 at 15 Values of θ from -2.8 to 2.8

Likelihood-based Values and Goodness of Fit Statistics

Summary of the Data and Control Parameters

Group Parameter Estimates (Back to TOC)

Group	Label	μ	*s.e.*	σ^2	*s.e.*	σ	*s.e.*
1	G1	0.00	-----	1.00	-----	1.00	-----
2	G2	0.22	-----	0.91	-----	0.95	-----
3	G3	-0.08	-----	0.52	-----	0.72	-----
4	G4	0.21	-----	1.33	-----	1.15	-----

DIF Statistics for Graded Items (Back to TOC)

Item numbers in:													
Group 1	Group 2	Group 3	Group 4	Contrast	Total X^2	*d.f.*	*p*	X^2_a	*d.f.*	*p*	$X^2_{c\|a}$	*d.f.*	*p*
1	1	1	1	1	0.1	2	0.9279	0.0	1	0.9037	0.1	1	0.7135
				2	0.5	2	0.7765	0.0	1	0.8633	0.5	1	0.4904
				3	0.4	2	0.8175	0.2	1	0.6578	0.2	1	0.6496
2	2	2	2	1	0.8	2	0.6809	0.3	1	0.5796	0.5	1	0.4971
				2	0.2	2	0.9014	0.0	1	0.8728	0.2	1	0.6699
				3	3.8	2	0.1482	3.8	1	0.0505	0.0	1	0.9154
3	3	3	3	1	6.2	2	0.0444	4.7	1	0.0310	1.6	1	0.2101
				2	8.7	2	0.0129	8.7	1	0.0032	0.0	1	0.9656
				3	6.1	2	0.0481	5.6	1	0.0183	0.5	1	0.4806
4	4	4	4	1	6.2	2	0.0453	5.4	1	0.0197	0.8	1	0.3867
				2	15.1	2	0.0005	15.1	1	0.0001	0.0	1	0.9791
				3	2.4	2	0.2960	0.5	1	0.4683	1.9	1	0.1674
5	5	5	5	1	3.9	2	0.1454	1.6	1	0.2109	2.3	1	0.1301

โปรแกรม MPLUS

วิธร MGCFA ALIGNMENT

Mplus VERSION 7.31

MUTHEN & MUTHEN

TITLE: ALIGNMENT MODEL FOR MATH REGION (1=CENTRAL 2=NORTH 3=NORTHEST 4=SOUTH)

MODEL FIT INFORMATION

MODEL RESULTS

			Two-Tailed	
	Estimate	S.E.	Est./S.E.	P-Value
Latent Class 1 (1)				
MATH BY				
Q1	0.534	0.060	8.868	0.000
Q2	0.970	0.083	11.705	0.000
Q3	1.241	0.094	13.259	0.000
Q4	0.804	0.069	11.635	0.000
Q5	0.821	0.069	11.976	0.000
Q6	0.677	0.063	10.777	0.000
Q7	0.700	0.066	10.674	0.000
Q8	0.410	0.057	7.149	0.000
Q9	0.429	0.058	7.381	0.000
Q10	0.418	0.057	7.342	0.000
Thresholds				
Q1$1	0.844	0.052	16.222	0.000
Q2$1	1.919	0.081	23.596	0.000
Q3$1	1.841	0.085	21.602	0.000
Q4$1	1.013	0.058	17.418	0.000
Q5$1	1.028	0.058	17.636	0.000
Q6$1	0.959	0.055	17.351	0.000
Q7$1	1.026	0.057	18.103	0.000

ALIGNMENT OUTPUT

Intercepts/Thresholds

Threshold Q1$1

Group	Group	Value	Value	Difference	SE	P-value
2	1	0.803	0.844	-0.041	0.070	0.556
3	1	0.803	0.844	-0.040	0.068	0.553
3	2	0.803	0.803	0.001	0.065	0.989
4	1	0.800	0.844	-0.043	0.069	0.531
4	2	0.800	0.803	-0.002	0.066	0.970
4	3	0.800	0.803	-0.003	0.065	0.958

Approximate Measurement Invariance Holds For Groups:

1 2 3 4

Weighted Average Value Across Invariant Groups: 0.812

R-square/Explained variance/Invariance index: 0.941

Invariant Group Values, Difference to Average and Significance

Group	Value	Difference	SE	P-value
1	0.844	0.031	0.043	0.471
2	0.803	-0.010	0.041	0.810
3	0.803	-0.009	0.039	0.821
4	0.800	-0.012	0.041	0.761

Threshold Q2$1

Group	Group	Value	Value	Difference	SE	P-value
2	1	1.866	1.919	-0.054	0.114	0.639
3	1	1.973	1.919	0.054	0.109	0.621
3	2	1.973	1.866	0.108	0.114	0.346
4	1	2.046	1.919	0.126	0.121	0.295
4	2	2.046	1.866	0.180	0.124	0.148
4	3	2.046	1.973	0.072	0.118	0.539

Approximate Measurement Invariance Holds For Groups:

1 2 3 4

Weighted Average Value Across Invariant Groups: 1.951

R-square/Explained variance/Invariance index: 0.839

Invariant Group Values, Difference to Average and Significance

Group	Value	Difference	SE	P-value
1	1.919	-0.032	0.069	0.646
2	1.866	-0.085	0.073	0.240
3	1.973	0.022	0.068	0.742
4	2.046	0.095	0.076	0.216

Loadings

Loadings for Q1

Group	Group	Value	Value	Difference	SE	P-value
2	1	0.517	0.534	-0.017	0.078	0.825
3	1	0.529	0.534	-0.005	0.095	0.954
3	2	0.529	0.517	0.012	0.095	0.902
4	1	0.498	0.534	-0.036	0.075	0.632
4	2	0.498	0.517	-0.018	0.071	0.795
4	3	0.498	0.529	-0.030	0.091	0.741

Approximate Measurement Invariance Holds For Groups:

1 2 3 4

Weighted Average Value Across Invariant Groups: 0.520

R-square/Explained variance/Invariance index: 0.964

Invariant Group Values, Difference to Average and Significance

Group	Value	Difference	SE	P-value
1	0.534	0.015	0.049	0.768
2	0.517	-0.003	0.048	0.957
3	0.529	0.009	0.063	0.884
4	0.498	-0.021	0.045	0.641

Loadings for Q2

Group	Group	Value	Value	Difference	SE	P-value
2	1	1.018	0.970	0.047	0.112	0.673
3	1	0.971	0.970	0.001	0.140	0.996
3	2	0.971	1.018	-0.047	0.150	0.755
4	1	1.174	0.970	0.204	0.113	0.072
4	2	1.174	1.018	0.156	0.121	0.196
4	3	1.174	0.971	0.203	0.151	0.177

Approximate Measurement Invariance Holds For Groups:

1 2 3 4

Weighted Average Value Across Invariant Groups: 1.033

R-square/Explained variance/Invariance index: 0.861

Invariant Group Values, Difference to Average and Significance

Group	Value	Difference	SE	P-value
1	0.970	-0.063	0.069	0.360
2	1.018	-0.016	0.077	0.838
3	0.971	-0.062	0.098	0.526
4	1.174	0.141	0.077	0.069

Copyright (c) 1998-2015 Muthen & Muthen

โปรแกรม MPLUS
วิธี MIMIC

ตัวอย่าง

ผลการวิเคราะห์โปรแกรม Mplus แบบ uniform DIF

Mplus VERSION 7.31

MUTHEN & MUTHEN

TITLE: MIMIC MODEL FOR MATH REGION CENTRAL REGION IS REFERENCE

MODEL FIT INFORMATION

Number of Free Parameters 57

Chi-Square Test of Model Fit

Value	1838.641
Degrees of Freedom	343
P-Value	0.0000

RMSEA (Root Mean Square Error Of Approximation)

Estimate	0.023	
90 Percent C.I.	0.022	0.024
Probability RMSEA <= .05	1.000	

CFI/TLI

CFI	0.678
TLI	0.648

Chi-Square Test of Model Fit for the Baseline Model

Value	5014.161
Degrees of Freedom	375
P-Value	0.0000

MODEL RESULTS

			Two-Tailed	
	Estimate	S.E.	Est./S.E.	P-Value
MATH BY				
Q1	1.000	0.000	999.000	999.000
Q2	1.956	0.114	17.151	0.000
Q3	1.961	0.107	18.276	0.000
Q4	1.292	0.077	16.834	0.000
Q5	1.391	0.083	16.695	0.000
Q6	1.127	0.072	15.713	0.000
Q7	1.449	0.086	16.835	0.000
Q8	0.801	0.063	12.719	0.000
Q9	0.810	0.062	13.140	0.000
Q10	0.848	0.067	12.614	0.000
Q11	1.524	0.083	18.437	0.000
Q12	1.623	0.086	18.817	0.000
Q13	1.691	0.089	18.989	0.000
…				
Q25	1.209	0.077	15.634	0.000
MATH ON				
NORTH	0.008	0.011	0.723	0.470
NORTHEST	-0.075	0.011	-6.836	0.000
SOUTH	-0.012	0.011	-1.105	0.269
Q1 ON				
NORTH	0.000	0.000	999.000	999.000
NORTHEST	0.000	0.000	999.000	999.000
SOUTH	0.000	0.000	999.000	999.000

Q2	ON			
NORTH	0.000	0.000	999.000	999.000
NORTHEST	0.000	0.000	999.000	999.000
SOUTH	0.000	0.000	999.000	999.000
Q3	ON			
NORTH	0.000	0.000	999.000	999.000
NORTHEST	0.000	0.000	999.000	999.000
SOUTH	0.000	0.000	999.000	999.000
Q4	ON			
NORTH	0.000	0.000	999.000	999.000
NORTHEST	0.000	0.000	999.000	999.000
SOUTH	0.000	0.000	999.000	999.000
Q5	ON			
NORTH	0.000	0.000	999.000	999.000
NORTHEST	0.000	0.000	999.000	999.000
SOUTH	0.000	0.000	999.000	999.000

STANDARDIZED MODEL RESULTS

STDYX Standardization

				Two-Tailed
	Estimate	S.E.	Est./S.E.	P-Value
MATH	BY			
Q1	0.316	0.016	20.295	0.000
Q2	0.618	0.018	33.681	0.000
Q3	0.619	0.017	35.960	0.000
Q4	0.408	0.015	26.344	0.000
Q5	0.440	0.015	28.616	0.000
Q6	0.357	0.015	23.223	0.000

Q7	0.458	0.015	30.657	0.000
Q8	0.253	0.016	16.037	0.000
Q9	0.256	0.016	16.253	0.000
Q10	0.268	0.016	16.970	0.000
…				
Q25	0.382	0.016	23.953	0.000
MATH ON				
NORTH	0.011	0.015	0.723	0.470
NORTHEST	-0.103	0.014	-7.284	0.000
SOUTH	-0.017	0.015	-1.106	0.269
Q1 ON				
NORTH	0.000	0.000	999.000	999.000
NORTHEST	0.000	0.000	999.000	999.000
SOUTH	0.000	0.000	999.000	999.000
Q2 ON				
NORTH	0.000	0.000	999.000	999.000
NORTHEST	0.000	0.000	999.000	999.000
SOUTH	0.000	0.000	999.000	999.000
Q3 ON				
NORTH	0.000	0.000	999.000	999.000
NORTHEST	0.000	0.000	999.000	999.000
SOUTH	0.000	0.000	999.000	999.000
Q4 ON				
NORTH	0.000	0.000	999.000	999.000
NORTHEST	0.000	0.000	999.000	999.000
SOUTH	0.000	0.000	999.000	999.000

Copyright (c) 1998-2015 Muthen & Muthen

ตัวอย่าง

ผลการวิเคราะห์โปรแกรม Mplus แบบ nonuniform DIF

Mplus VERSION 7.31

MUTHEN & MUTHEN

INPUT INSTRUCTIONS

```
TITLE:    MIMIC MODEL FOR MATH REGION CENTRAL REGION IS REFERENCE
DATA:     FILE IS C:\DIFDATASAZA\MEP6\MATH_R_EP6.csv;
VARIABLE:     NAMES ARE REGION NORTH NORTHEST SOUTH q1-q25;
              USEVARIABLE NORTH NORTHEST SOUTH q1-q25;
              CATEGORICAL q1-q25;
```

SUMMARY OF ANALYSIS

Number of groups	1
Number of observations	8000

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
MATH BY				
Q1	1.000	0.000	999.000	999.000
Q2	1.904	0.245	7.773	0.000
Q3	2.342	0.291	8.037	0.000
Q4	1.556	0.208	7.468	0.000
Q5	1.600	0.215	7.446	0.000
Q6	1.284	0.180	7.121	0.000
Q7	1.351	0.188	7.203	0.000
Q8	0.787	0.136	5.794	0.000
Q9	0.833	0.140	5.960	0.000

Q10	0.790	0.136	5.792	0.000
Q11	1.920	0.257	7.474	0.000
Q12	2.558	0.333	7.689	0.000
Q13	2.468	0.320	7.723	0.000
Q14	2.186	0.287	7.612	0.000
Q15	1.655	0.223	7.419	0.000
Q16	1.789	0.237	7.532	0.000
Q17	2.186	0.290	7.534	0.000
Q18	2.286	0.300	7.613	0.000
Q19	2.340	0.306	7.647	0.000
Q20	1.925	0.250	7.688	0.000
Q21	1.998	0.256	7.791	0.000
Q22	1.153	0.170	6.792	0.000
Q23	1.070	0.160	6.677	0.000
Q24	1.298	0.183	7.091	0.000
Q25	1.027	0.157	6.542	0.000
MATH ON				
NORTH	0.114	0.023	4.940	0.000
NORTHEST	-0.054	0.024	-2.238	0.025
SOUTH	0.090	0.020	4.487	0.000
Q1 ON				
MATHNORTH	-0.032	0.148	-0.214	0.830
MATHNORTHE	-0.284	0.141	-2.005	0.045
MATHSOUTH	0.119	0.161	0.737	0.461
Q2 ON				
MATHNORTH	0.075	0.178	0.422	0.673
MATHNORTHE	-0.602	0.205	-2.935	0.003
MATHSOUTH	0.597	0.198	3.011	0.003

Q3	ON				
	MATHNORTH	-0.559	0.194	-2.881	0.004
	MATHNORTHE	-1.141	0.236	-4.837	0.000
	MATHSOUTH	-0.142	0.191	-0.742	0.458
Q4	ON				
	MATHNORTH	-0.458	0.171	-2.687	0.007
	MATHNORTHE	-1.019	0.204	-4.986	0.000
	MATHSOUTH	0.026	0.170	0.155	0.877
Q5	ON				
	MATHNORTH	-0.352	0.171	-2.061	0.039
	MATHNORTHE	-0.822	0.198	-4.164	0.000
	MATHSOUTH	0.176	0.177	0.994	0.320