การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาคระหว่างวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ยกับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก

นางสาววริษชา ชะม้อย

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาครุศาสตรมหาบัณฑิต
สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา
คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2550

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

COMPARISONS OF THE EFFICIENCY IN DETECTING DIFFERENTIAL ITEM FUNCTIONING OF POLYTOMOUS ITEMS BETWEEN MEAN AND COVARIANCE STRUCTURES METHOD AND LOGISTIC DISCRIMINANT FUNCTION ANALYSIS METHOD

Miss Waritsa Chamoy

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Education Program in Educational Measurement and Evaluation
Department of Educational Research and Psychology
Faculty of Education
Chulalongkorn University
Academic Year 2007
Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาคระหว่างวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ยกับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก
โดย	นางสาววริษษา ชะม้อย
สาขาวิชา	การวัดและประเมินผลการศึกษา
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ ดร.ณัฏฐภรณ์ หลาวทอง

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

....คณบดีคณะครุศาสตร์
(รองศาสตราจารย์ ดร.พฤทธิ์ ศิริบรรณพิทักษ์)

คณะกรรมการสอบวิทยานิพนธ์

....ประธานกรรมการ
(ศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี)

....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(ผู้ช่วยศาสตราจารย์ ดร.ณัฏฐภรณ์ หลาวทอง)

....กรรมการภายนอกมหาวิทยาลัย
(รองศาสตราจารย์ ดร.กาญจนา วัธนสุนทร)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วริษษา ชะม้อย : การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาคระหว่างวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ยกับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก. (COMPARISONS OF THE EFFICIENCY IN DETECTING DIFFERENTIAL ITEM FUNCTIONING OF POLYTOMOUS ITEMS BETWEEN MEAN AND COVARIANCE STRUCTURES METHOD AND LOGISTIC DISCRIMINANT FUNCTION ANALYSIS METHOD) อ.ที่ปรึกษาวิทยานิพนธ์หลัก : ผศ.ดร.ณัฏฐภรณ์ หลาวทอง, 187 หน้า.

การวิจัยเรื่องนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาค ระหว่างวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย (MACS) กับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก (LDFA) กลุ่มตัวอย่างที่ใช้ในการวิจัย คือ นักเรียนชั้นมัธยมศึกษาปีที่ 2 ปีการศึกษา 2550 โรงเรียนมัธยมศึกษา สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กรุงเทพมหานคร เครื่องมือที่ใช้ในการวิจัย คือ แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ 5 ตัวเลือกที่มีคำตอบถูกหลายคำตอบ และให้ตัดสินทุกคำตอบ (Multiple True-false Answer) เงื่อนไขที่ศึกษา ได้แก่ (1) ความยาวของแบบสอบ มี 2 ระดับ คือ แบบสอบที่ตรวจให้คะแนนหลายค่าที่มีจำนวน 30 ข้อ และ 40 ข้อ และ (2) ขนาดกลุ่มตัวอย่าง มี 3 ระดับ คือ 200 คน, 500 คน และ 1,000 คน

ผลการวิจัยพบว่า

1. เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่าทั้ง 6 เงื่อนไขของการทดสอบ วิธี LDFA มีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในช่วงตั้งแต่ร้อยละ 0 ถึงร้อยละ 15.6 และวิธี MACS มีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในช่วงตั้งแต่ร้อยละ 8.3 ถึงร้อยละ 21.9 โดยที่วิธี LDFA เกือบทุกเงื่อนไขมีค่าอยู่ในเกณฑ์ที่กำหนด ส่วนวิธี MACS พบว่าส่วนใหญ่มีค่าเกินกว่าเกณฑ์ที่กำหนด และพบว่าทั้งวิธี LDFA และวิธี MACS จะมีอัตราความคลาดเคลื่อนประเภทที่ 1 ในเงื่อนไขความยาวแบบสอบ 30 ข้อต่ำกว่าเงื่อนไขความยาวแบบสอบ 40 ข้อทุกเงื่อนไข

2. เมื่อเปรียบเทียบอำนาจการทดสอบ พบว่า ทั้ง 6 เงื่อนไขของการทดสอบ วิธี LDFA มีอำนาจการทดสอบอยู่ในช่วงตั้งแต่ร้อยละ 62.5 ถึงร้อยละ 100 และวิธี MACS มีค่าอำนาจการทดสอบอยู่ในช่วงตั้งแต่ร้อยละ 37.5 ถึงร้อยละ 100 และพบว่าทั้งวิธี LDFA และวิธี MACS จะมีอำนาจการทดสอบในเงื่อนไขความยาวแบบสอบ 30 ข้อสูงกว่าเงื่อนไขความยาวแบบสอบ 40 ข้อทุกเงื่อนไข

ภาควิชา.....วิจัยและจิตวิทยาการศึกษา.......... ลายมือชื่อนิสิต......วริษษา ชะม้อย......

สาขาวิชา...การวัดและประเมินผลการศึกษา... ลายมือชื่ออาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก..........

ปีการศึกษา.....2550.....................................

498 37455 27: MAJOR EDUCATIONAL MEASUREMENT AND EVALUATION

KEY WORD: POLYTOMOUS/ DIFFERENTIAL ITEM FUNCTIONING/ MACS / LDFA

WARITSA CHAMOY : COMPARISONS OF THE EFFICIENCY IN DETECTING DIFFERENTIAL ITEM FUNCTIONING OF POLYTOMOUS ITEMS BETWEEN MEAN AND COVARIANCE STRUCTURES METHOD AND LOGISTIC DISCRIMINANT FUNCTION ANALYSIS METHOD. THESIS PRINCIPAL ADVISOR : NUTTAPORN LAWTHONG, Ph.D., 187 pp.

The purpose of this research was to compare the efficiency in detecting Differential Item Functioning of Polytomous items between Mean and Covariance Structures method (MACS) and Logistic Discriminant Function Analysis method (LDFA). The sample group of study was secondary students in academic year 2007 under the Office of the Basic Education Commission, Bangkok. The instruments were two Multiple True-False Answer Mathematic achievement tests. The conditions were, firstly, the length of tests: 30 and 40 items and, secondly, the group sizes: 200, 500 and 1,000 students.

The results of experiment found that:

1. Comparison with 6 conditions, LDFA method had type I error rates around 0% - 15.6% and MACS method had type I error rates around 8.3% - 21.9%. Almost all conditions of LDFA method were lower than criteria of type I error rates but almost all conditions of MACS method were higher than criteria of type I error rates. In addition, type I error rates of both methods in conditions of the length of tests 30 items were lower than type I error rates in conditions of the length of tests 40 items.

2. Comparison with 6 conditions, LDFA method had powers of test around 62.5% - 100% and MACS method had powers of test around 37.5% - 100%. In addition, the powers of test of both methods in conditions of the length of tests 30 items were higher than the powers of test in conditions of the length of tests 40 items.

Department.....Educational Research and Psychology..... Student's signature..Waritsa Chamoy....

Field of study...Educational Measurement and Evaluation.. Principal Advisor's signature..Nuttaporn Lawthong

Academic year........2007............

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปด้วยดี เพราะความกรุณา และเอื้ออาทรของผู้ช่วยศาสตราจารย์ ดร.ณัฏฐภรณ์ หลาวทอง อาจารย์ที่ปรึกษาอีกทั้งยังเป็นบุคคลที่สำคัญยิ่งที่ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูง อาจารย์คอยดูแล เอาใจใส่ และให้ความเมตตา ทั้งด้านวิชาการและด้านจิตใจ ตลอดจนให้คำชี้แนะ คำปรึกษาที่มีประโยชน์แก่ผู้วิจัยเสมอมา กรุณาช่วยแก้ไขข้อบกพร่องต่าง ๆ ในการทำวิทยานิพนธ์จนมีความสมบูรณ์มากขึ้น

ขอกราบขอบพระคุณ ศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี ประธานกรรมการสอบวิทยานิพนธ์ และรองศาสตราจารย์ ดร.กาญจนา วัธนสุนทร กรรมการสอบวิทยานิพนธ์ ที่ช่วยให้ข้อเสนอแนะที่เป็นประโยชน์ในการปรับปรุงแก้ไขวิทยานิพนธ์ให้มีความสมบูรณ์มากยิ่งขึ้น

ในรายวิชาสัมมนาการออกแบบการวิจัย ผู้วิจัยได้รับคำแนะนำและข้อเสนอแนะที่เป็นประโยชน์อย่างยิ่งจากรองศาสตราจารย์ ดร.เอมอร จังศิริพรปกรณ์ ที่ได้ช่วยกรุณาพัฒนาหัวข้อวิจัยจนมีความชัดเจน อีกทั้งให้กำลังใจผู้วิจัยเสมอมา

ขอกราบขอบพระคุณเป็นอย่างสูง สำหรับ Assistant Prof. Dr.Stephen Stark แห่ง University of South Florida ที่กรุณาให้คำปรึกษา และแนะนำในการเขียนคำสั่งวิเคราะห์ผล จนผู้วิจัยสามารถดำเนินการสำเร็จลุล่วงไปด้วยดี

ขอขอบพระคุณ คณะครูอาจารย์ทั้ง 13 โรงเรียนในเขตพื้นที่การศึกษาเขต 1 กรุงเทพมหานครทุกท่าน ที่อนุเคราะห์ให้ผู้วิจัยมาเก็บข้อมูล ช่วยประสานงานเป็นอย่างดี และเสียสละเวลาให้ผู้วิจัย ตลอดจนนักเรียนชั้นมัธยมศึกษาปีที่ 2 ทุกคน ที่ตั้งใจทำข้อสอบเป็นอย่างดี ผู้วิจัยขอขอบคุณด้วยใจจริง อีกทั้งขอขอบคุณบัณฑิตวิทยาลัยที่ให้ทุนอุดหนุนวิทยานิพนธ์ในครั้งนี้

ขอขอบคุณเพื่อน ๆ ในสาขาวิชาการวัดและประเมินผลการศึกษาทั้ง 3 คน คุณศรัญญา รณศิริ คุณน้ำทิพย์ องอาจวาณิชย์ และคุณกฤชรัตน์ วิทยาเวช เพื่อนที่คอยช่วยเหลือผู้วิจัย ช่วยให้คำปรึกษา เพื่อนที่มากด้วยน้ำใจ ขอบคุณที่เป็นเพื่อนร่วมทุกข์ร่วมสุขตลอดระยะเวลา 2 ปี

ขอขอบคุณเพื่อน ๆ ที่น่ารักของผู้วิจัย ที่ช่วยเหลือผู้วิจัยเสมอมา อาสาช่วยพิมพ์งาน และช่วยแปลภาษาอังกฤษ และที่สำคัญคำปลอบใจ คำให้กำลังใจ และความเชื่อมั่นในตัวผู้วิจัย เพื่อนที่ผู้วิจัยรักมาก คุณอภิชยา คุณพนิดา คุณอัญรัตน์ คุณศิริวรรณ คุณเยาวลักษณ์และคุณนนทรัฐ

ท้ายสุดนี้ สิ่งที่สำคัญที่สุดในชีวิต ผู้วิจัยรักครอบครัวสุภาดี และชะม้อยมากที่สุดในโลก ครอบครัวที่มีให้แต่ความรัก ความเมตตา เป็นกำลังใจและแรงผลักดันในชีวิต ครอบครัวที่เป็นแบบอย่างให้ผู้วิจัยเดินตาม ให้การสนับสนุนทางการศึกษาแก่ผู้วิจัยอย่างเต็มกำลัง ความดีทั้งหมดที่ลูกได้ทำ ขอมอบแด่แม่นันทิชา สุภาดี ที่ลูกรักยิ่งชีวิต

สารบัญ

หน้า

บทคัดย่อภาษาไทย.. ง
บทคัดย่อภาษาอังกฤษ.. จ
กิตติกรรมประกาศ.. ฉ
สารบัญ.. ช
สารบัญตาราง.. ฌ
สารบัญภาพ.. ฏ
บทที่
1. บทนำ
ความเป็นมาและความสำคัญของปัญหา.. 1
คำถามวิจัย.. 7
วัตถุประสงค์ของการวิจัย.. 7
สมมุติฐานการวิจัย.. 8
ขอบเขตของการวิจัย.. 8
ข้อจำกัดของการวิจัย.. 9
คำจำกัดความที่ใช้ในการวิจัย.. 10
ประโยชน์ที่คาดว่าจะได้รับ.. 13
2. เอกสารและงานวิจัยที่เกี่ยวข้อง
ตอนที่ 1 องค์ความรู้เชิงทฤษฎี โดยแบ่งออกเป็น 6 หัวข้อ
ความหมายของการทำหน้าที่ต่างกันของข้อสอบ.. 15
ความเป็นมาของการทำหน้าที่ต่างกันของข้อสอบ.. 17
ประเภทของการทำหน้าที่ต่างกันของข้อสอบ.. 19
หลักการและวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ.. 22
แนวคิดเกี่ยวกับการวิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบ
ที่ตรวจให้คะแนนแบบพหุวิภาค (Polytomous IRT Model).. 30
การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีโครงสร้างความแปรปรวนร่วม
และค่าเฉลี่ย กับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก.. 34
ตอนที่ 2 งานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบ.. 38

บทที่ หน้า

ตอนที่ 3 กรอบแนวคิดในการวิจัย... 60

3. วิธีดำเนินการวิจัย

ประชากร กลุ่มตัวอย่าง... 62

เครื่องมือที่ใช้.. 65

การเก็บรวบรวมข้อมูล.. 68

การวิเคราะห์ข้อมูล... 68

4. ผลการวิเคราะห์ข้อมูล

ตอนที่ 1 ผลการวิเคราะห์ค่าสถิติพื้นฐานของผลการตอบข้อสอบ........................ 76

ตอนที่ 2 ผลการวิเคราะห์คุณภาพของแบบสอบที่ใช้ในการวิจัย.......................... 78

ตอนที่ 3 ผลการวิเคราะห์องค์ประกอบเพื่อตรวจสอบความเป็นเอกมิติของข้อมูล... 79

ตอนที่ 4 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GPCM

โดยเสนอค่าพารามิเตอร์ของข้อสอบและผลของการตรวจสอบ............... 82

ตอนที่ 5 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 2 วิธีได้แก่

วิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ยกับวิธีการวิเคราะห์

ฟังก์ชันเชิงจำแนกแบบโลจิสติก.. 95

5. สรุป อภิปรายผลและข้อเสนอแนะ

สรุปผลการวิจัย.. 118

อภิปรายผล.. 122

ข้อเสนอแนะ.. 128

รายการอ้างอิง.. 130

ภาคผนวก

ภาคผนวก ก แบบสอบและกระดาษคำตอบที่ใช้ในการวิจัย................................ 137

ภาคผนวก ข ตัวอย่าง print out หาค่าความสามารถของผู้สอบ........................... 146

ภาคผนวก ค ตัวอย่างคำสั่งที่วิเคราะห์ด้วยโปรแกรม LISREL............................ 175

ภาคผนวก ง ตัวอย่างคำสั่งที่วิเคราะห์ด้วยโปรแกรม MULTILOG...................... 178

ภาคผนวก จ ตัวอย่างคำสั่งที่วิเคราะห์ด้วยโปรแกรม PARSCALE...................... 180

ภาคผนวก ฉ จดหมายขอความร่วมมือในการวิจัย... 183

ประวัติผู้เขียนวิทยานิพนธ์... 187

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญตาราง

ตารางที่ หน้า

1 สรุปงานวิจัยที่ได้ศึกษาเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบทวิวิภาค (Dichotomous).. 45
2 สรุปงานวิจัยที่ได้ศึกษาเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาค (Polytomous).. 51
3 สรุปงานวิจัยที่ได้ศึกษาเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนทั้งแบบทวิวิภาคและพหุวิภาค.. 54
4 ขนาดของกลุ่มประชากร และกลุ่มตัวอย่างจำแนกตามขนาดของโรงเรียน............ 63
5 รายชื่อโรงเรียน จำนวนห้องเรียน และจำนวนนักเรียนที่สุ่มได้............................. 63
6 รายชื่อโรงเรียน จำนวนห้องเรียน และจำนวนนักเรียนที่ใช้ในการเก็บข้อมูลจริง...... 65
7 ค่าสถิติพื้นฐานของแบบทดสอบวิชาคณิตศาสตร์ ขนาด 40 ข้อ และ 30 ข้อ......... 76
8 เปรียบเทียบค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของคะแนนระหว่างกลุ่มผู้สอบ... 77
9 ค่าความเที่ยงของแบบสอบ.. 78
10 ค่าไอเกน และร้อยละของความแปรปรวนขององค์ประกอบ ของแบบทดสอบขนาด 40 ข้อ...... 79
11 ค่าไอเกน และร้อยละของความแปรปรวนขององค์ประกอบ ของแบบทดสอบขนาด 30 ข้อ.. 80
12 ค่าพารามิเตอร์ของข้อสอบ แยกตามกลุ่มเพศชาย-หญิง เมื่อกลุ่มตัวอย่าง 1,000 คนและความยาวของแบบทดสอบ 40 ข้อ... 83
13 ค่าพารามิเตอร์ของข้อสอบ แยกตามกลุ่มเพศชาย-หญิง เมื่อกลุ่มตัวอย่าง 1,000 คนและความยาวของแบบทดสอบ 30 ข้อ... 85
14 ค่าสัมประสิทธิ์การเชื่อมโยงของแบบทดสอบขนาด 40 ข้อและ 30 ข้อ.................. 86
15 ค่าพารามิเตอร์ของข้อสอบที่ปรับเทียบสเกลแล้ว แยกตามกลุ่มเพศชาย-หญิง เมื่อกลุ่มตัวอย่าง 1,000 คนและความยาวของแบบทดสอบ 40 ข้อ....................... 88
16 ค่าพารามิเตอร์ของข้อสอบที่ปรับเทียบสเกลแล้ว แยกตามกลุ่มเพศชาย-หญิง เมื่อกลุ่มตัวอย่าง 1,000 คนและความยาวของแบบทดสอบ 30 ข้อ........................ 90
17 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GPCM ตามแนวคิดของทฤษฎีการตอบข้อสอบด้วยวิธีของลอร์ด (Lord's Chi-square) เมื่อกลุ่มตัวอย่าง 1,000 คนและความยาวแบบสอบ 40 ข้อ... 92

ตารางที่ หน้า

18 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GPCM ตามแนวคิดของทฤษฎีการตอบข้อสอบด้วยวิธีของลอร์ด (Lord’s Chi-square) เมื่อกลุ่มตัวอย่าง 1,000 คนและความยาวแบบสอบ 30 ข้อ.. 93
19 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ยเมื่อแบบสอบมีความยาว 40 ข้อและกลุ่มตัวอย่างจำนวน 1,000 คน.. 97
20 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ยเมื่อแบบสอบมีความยาว 40 ข้อและกลุ่มตัวอย่างจำนวน 500 คน.. 98
21 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ยเมื่อแบบสอบมีความยาว 40 ข้อและกลุ่มตัวอย่างจำนวน 200 คน.. 99
22 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ยเมื่อแบบสอบมีความยาว 30 ข้อและกลุ่มตัวอย่างจำนวน 1,000 คน.. 100
23 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ยเมื่อแบบสอบมีความยาว 30 ข้อและกลุ่มตัวอย่างจำนวน 500 คน.. 101
24 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ยเมื่อแบบสอบมีความยาว 30 ข้อและกลุ่มตัวอย่างจำนวน 200 คน.. 102
25 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติกเมื่อแบบสอบมีความยาว 40 ข้อ และกลุ่มตัวอย่างจำนวน 1,000 คน... 103

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ หน้า

26 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติกเมื่อแบบสอบมีความยาว 40 ข้อ และกลุ่มตัวอย่างจำนวน 500 คน........ 104

27 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติกเมื่อแบบสอบมีความยาว 40 ข้อ และกลุ่มตัวอย่างจำนวน 200 คน........ 105

28 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติกเมื่อแบบสอบมีความยาว 30 ข้อ และกลุ่มตัวอย่างจำนวน 1000 คน........ 106

29 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติกเมื่อแบบสอบมีความยาว 30 ข้อ และกลุ่มตัวอย่างจำนวน 500 คน........ 107

30 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติกเมื่อแบบสอบมีความยาว 30 ข้อ และกลุ่มตัวอย่างจำนวน 200 คน........ 108

31 สรุปข้อคำถามที่ปรากฏอยู่ในแบบสอบที่มีความยาว 40 ข้อและ 30 ข้อ........ 109

32 สรุปผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในทุกเงื่อนไข........ 110

33 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย กับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก........ 112

34 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย กับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก........ 113

สารบัญภาพ

ภาพที่ หน้า

1 ข้อสอบทำหน้าที่ต่างกันแบบมีทิศทางเดียว (Unidirectional DIF).......................... 20

2 ข้อสอบทำหน้าที่ต่างกันแบบไม่มีทิศทาง (Nondirectional DIF)............................. 21

3 โค้งฟังก์ชันสารสนเทศของ GPCM.. 32

4 กรอบแนวคิดในการวิจัย.. 60

5 ขั้นตอนดำเนินการวิจัย.. 72

6 ผลการตรวจสอบความเป็นมิติเดียวของแบบทดสอบความยาว 40 ข้อ........................ 80

7 ผลการตรวจสอบความเป็นมิติเดียวของแบบทดสอบความยาว 30 ข้อ......................... 81

8 โมเดลการวิเคราะห์สำหรับแบบสอบ 40 ข้อ... 96

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1
บทนำ

ความเป็นมาและความสำคัญของปัญหา

แบบสอบแบบหลายตัวเลือก (Multiple-choice Test) เป็นแบบสอบเลือกตอบที่นิยมใช้กันอย่างกว้างขวาง เพราะสามารถใช้วัดผลการเรียนรู้ ทั้งความรู้ ความเข้าใจ การนำไปใช้ และผลการเรียนรู้ขั้นสูงได้ สามารถสร้างให้วัดได้ครอบคลุมเนื้อเรื่องตามโครงสร้างอย่างมีประสิทธิภาพ และนำไปพัฒนาเป็นแบบสอบมาตรฐานได้ อีกทั้งมีความเป็นปรนัยในการตรวจให้คะแนน และสะดวกในการดำเนินการสอบ (ศิริชัย กาญจนวาสี, 2548) แต่แบบสอบแบบหลายตัวเลือกที่ให้นักเรียนเลือกตอบเฉพาะตัวเลือกถูกเพียงตัวเลือกเดียวและมีการให้คะแนนแบบสองค่า (Dichotomously Scored Items) คือ ตอบถูกให้ 1 คะแนน ตอบผิดให้ 0 คะแนนนั้น นักวัดผลกลุ่มหนึ่งเกิดข้อสงสัยว่านักเรียนที่ตอบถูกนั้นเป็นผู้ที่มีความรู้ในเรื่องนั้นจริงหรือเกิดจากการเดาและนักเรียนที่ตอบผิดนั้นเป็นเพราะไม่มีความรู้ในเรื่องนั้นเลยหรือเกิดจากปัญหาอื่น ๆ จึงได้พยายามพัฒนาแบบทดสอบที่สามารถบ่งบอกความรู้แท้จริงของผู้เรียนให้มากที่สุดเท่าที่จะทำได้ (อุทัยวรรณ สายพัฒนะ, 2547)

กระบวนการพัฒนาแบบสอบแบบหลายตัวเลือกได้ดำเนินมาอย่างต่อเนื่อง โดยเน้นที่วิธีการตอบและการตรวจให้คะแนน ซึ่งมีเป้าหมายเพื่อแก้จุดอ่อนของแบบสอบแบบหลายตัวเลือกเกี่ยวกับการเดาให้สามารถนำคะแนนผลการสอบไปประมาณค่าความเที่ยง และความตรงของแบบทดสอบให้มีความคลาดเคลื่อนน้อยที่สุด (อุทัยวรรณ สายพัฒนะ, 2547) แนวทางหนึ่งที่ใช้ คือ การใช้สูตรปรับคะแนนผลการสอบ ซึ่งเรียกว่า การปรับแก้คะแนนการเดา วิธีการดังกล่าวนี้จะเน้นที่คะแนนรวมทั้งฉบับ (Total Score) จึงไม่ให้รายละเอียดในส่วนที่เป็นรายข้อเท่าใดนัก ต่อมาจึงมีแนวทางใหม่เกิดขึ้นที่เหมาะสมกว่าด้วยการปรับวิธีการตอบและการให้คะแนนรายข้อใหม่ด้วยการปรับคะแนนแต่ละข้อเป็นการให้คะแนนที่เน้นการแสดงความรู้บางส่วน (Partial Knowledge) (Coombs et al., 1956 อ้างถึงใน อุทัยวรรณ สายพัฒนะ, 2547) และให้คะแนนแบบหลายค่า (Polytomous Score)

ข้อสอบชนิดให้คะแนนหลายค่า (Polytomously Scored Items) มีหลายประเภท เช่น ข้อสอบความเรียง (Essay Items) การประเมินงานปฏิบัติ (Performance Tasks Assessment) อาทิ การตัดสินคุณภาพของแฟ้มสะสมงาน (Portfolio) การพิสูจน์ทางคณิตศาสตร์ (Mathematical Proof) การทดลองทางวิทยาศาสตร์ (Scientific Experiment) เป็นต้น (Downing et al., 1995; Albanese & Sabers, 1988 อ้างถึงใน อุทัยวรรณ สายพัฒนะ, 2547)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

จะเห็นได้ว่า การพัฒนาวิธีการตอบและการให้คะแนนแบบทดสอบประเภทต่าง ๆ จำเป็นต้องคำนึงถึงคุณภาพของแบบทดสอบ ซึ่งเป็นหัวใจของการวัดและการประเมินผล คุณภาพที่สำคัญที่สุดของแบบทดสอบ คือ “ความตรง” ในการสร้างและการตรวจสอบคุณภาพของแบบสอบจะต้องคำนึงถึงคุณภาพด้านความตรงเป็นสำคัญ ทั้งนี้เพราะว่าความตรงเป็นคุณสมบัติของแบบสอบที่แสดงถึงความสามารถในการวัดได้ถูกต้องแม่นยำ ถ้าผลการวัดได้ค่าที่ใกล้เคียงกับค่าคุณลักษณะที่แท้จริงเพียงใดก็ถือว่ามีความตรงมากขึ้นเพียงนั้น โดยเป็นการตรวจสอบในประเด็นของความยุติธรรมของข้อสอบและแบบสอบ (Item and Test Unfairness) (ศิริชัย กาญจนวาสี, 2545) วิธีการตรวจสอบความตรงของข้อสอบสามารถทำได้หลายวิธี วิธีการหนึ่งคือ การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ แบบสอบหรือข้อสอบที่ไม่ได้วัดเฉพาะคุณลักษณะแฝงเป้าหมายตามที่ต้องการวัดแต่วัดคุณลักษณะแฝงแทรกซ้อนที่ไม่ต้องการวัดของผู้สอบ แสดงว่าแบบสอบหรือข้อสอบนั้นขาดความตรง (Green, 1994; Mazor et al., 1995 อ้างถึงใน รัชรินทร์ มุคดา, 2540) เพราะหากผู้สอบกลุ่มย่อยกลุ่มใดมีคุณลักษณะแฝงแทรกซ้อนสูงกว่าย่อมมีโอกาสที่จะตอบข้อสอบได้ถูกต้องมากกว่า ทั้ง ๆ ที่คุณลักษณะแฝงเป้าหมายเท่ากันกับกลุ่มผู้สอบย่อยกลุ่มอื่น จึงทำให้เกิดการได้เปรียบเสียเปรียบกันระหว่างกลุ่มผู้สอบย่อย แต่เดิมใช้คำว่า ความลำเอียงของข้อสอบ (Item Bias) หรือ ความลำเอียงของแบบสอบ (Test Bias) ต่อมาระยะหลังเกิดความคลุมเครือในการที่จะใช้เกณฑ์ในการตัดสินใจเรื่อง ความลำเอียง จึงนิยมใช้สารสนเทศทางสถิติมาเป็นเกณฑ์ในการตัดสินใจ และใช้คำว่า การทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning: DIF) หรือ การทำหน้าที่ต่างกันของแบบสอบ (Differential Test Functioning: DTF) เนื่องจากเห็นว่าเป็นคำที่มีความหมายกลาง ๆ มีความเหมาะสมในเชิงวิชาการมากกว่าคำว่า ความลำเอียง (Holland & Thayer, 1988; Green, 1994 อ้างถึงใน จิติมา วรรณศรี, 2539) หรืออาจกล่าวได้ว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning: DIF) เป็นวิธีการหนึ่งที่สามารถจำแนกการขาดความเสมอภาคทางการวัด (Measurement Equivalence) ได้ ซึ่งความเสมอภาคทางการวัดนี้เป็นข้อตกลงเบื้องต้นที่จะส่งผลให้การเปรียบเทียบระหว่างกลุ่มมีความหมาย (Drasgow, 1984 อ้างถึงใน Stark, Chernyshenko, & Drasgow, 2006) ดังเช่นจะเห็นว่า นักวิจัยส่วนใหญ่จะให้ความสนใจในการพัฒนาหรือทดสอบสมมุติฐานเกี่ยวกับความแตกต่างทางวัฒนธรรม โดยต้องการจำแนกระหว่างความแตกต่างทางวัฒนธรรมที่แท้จริงกับสิ่งที่เกิดขึ้นจากมนุษย์ ซึ่งอาจจะเป็นปัญหามาจากเครื่องมือ ผลกระทบ หรือความแตกต่างที่แท้จริงในการกระจายของคุณลักษณะ (Stark et al., 2006)

การตรวจสอบเกี่ยวกับการได้เปรียบเสียเปรียบในการสอบระหว่างผู้สอบกลุ่มย่อยตั้งแต่ 2 กลุ่มขึ้นไป เริ่มมีตั้งแต่ปี ค.ศ.1951 โดยเป็นการศึกษาเปรียบเทียบระหว่างผู้สอบที่มีความต่างกัน

ทางด้านเศรษฐกิจ สังคม เพศ วัฒนธรรม เชื้อชาติ ระดับสติปัญญา วิธีสอน (กาญจนา วัธนสุนทร, 2537) รวมทั้งสถาบันการศึกษา ประสบการณ์ (Holland & Thayer, 1988) นอกจากจะศึกษาปัจจัยอันเกิดจากผู้สอบซึ่งส่งผลให้เกิดการได้เปรียบเสียเปรียบระหว่างกลุ่มผู้สอบ ในระยะหลังได้มีการศึกษาเปรียบเทียบวิธีการต่าง ๆ ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ทั้งนี้เพราะมีวิธีการตรวจสอบหลายวิธีที่ถูกคิดค้นและพัฒนา ปรับปรุงเพื่อให้สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้อย่างมีประสิทธิภาพมากที่สุด (จิติมา วรรณศรี, 2539)

การทำหน้าที่ต่างกันของข้อสอบนั้น ขนาดและทิศทางของการทำหน้าที่ต่างกันของข้อสอบจะแปรเปลี่ยนไปตามระดับความสามารถที่แตกต่างกันของผู้สอบ โดยแบ่งลักษณะของข้อสอบที่ทำหน้าที่ต่างกันเป็น 2 ประเภท (Mellenbergh, 1982) ได้แก่ ข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป หรือเรียกว่า ข้อสอบที่ทำหน้าที่ต่างกันแบบสม่ำเสมอ (Uniform DIF) ในที่นี้ผู้วิจัยขอใช้คำว่า ข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป ซึ่งจะเกิดขึ้นเมื่อไม่มีปฏิสัมพันธ์ (Interaction) ระหว่างระดับความ สามารถของผู้สอบกับการเป็นสมาชิกของกลุ่มย่อย นั่นคือ โอกาสของการตอบข้อสอบได้ถูกต้องของผู้สอบกลุ่มย่อยกลุ่มหนึ่งสูงกว่าผู้สอบกลุ่มย่อยอีกกลุ่มหนึ่งตลอดช่วงความสามารถ และ ข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป หรือเรียกว่า ข้อสอบที่ทำหน้าที่ต่างกันแบบไม่สม่ำเสมอ (Nonuniform DIF) ในที่นี้ผู้วิจัยขอใช้คำว่า ข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป เช่นเดียวกันซึ่งจะเกิดขึ้นเมื่อมีปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับการเป็นสมาชิกของกลุ่มย่อย นั่นคือโอกาสของการตอบข้อสอบได้ถูกต้องของผู้สอบกลุ่มย่อยกลุ่มหนึ่งสูงกว่าผู้สอบกลุ่มย่อยอีกกลุ่มหนึ่งไม่ตลอดช่วงความสามารถ ซึ่งตามทฤษฎีการตอบสนองข้อสอบ (IRT) สามารถพิจารณาปฏิสัมพันธ์ได้จากความแตกต่างของค่าพารามิเตอร์อำนาจจำแนกของข้อสอบระหว่างผู้สอบกลุ่มย่อย 2 กลุ่ม ถ้าข้อสอบระหว่างผู้สอบ 2 กลุ่มย่อยมีค่าพารามิเตอร์อำนาจจำแนกเท่ากันแล้ว โค้งคุณลักษณะข้อสอบ (Item Characteristic Curves: ICC) ของผู้สอบ 2 กลุ่มจะขนานกัน แสดงว่าเป็น ข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป แต่ถ้าข้อสอบระหว่างผู้สอบ 2 กลุ่มย่อย มีค่าพารามิเตอร์อำนาจจำแนกไม่เท่ากันแล้ว โค้งคุณลักษณะข้อสอบ (Item Characteristic Curves: ICC) ของผู้สอบ 2 กลุ่มจะไม่ขนานกัน แสดงว่าเป็น ข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป (อุทัยวรรณ สายพัฒนะ, 2547)

หลักการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจะเปรียบเทียบผลการตอบข้อสอบระหว่างผู้สอบกลุ่มย่อยสองกลุ่มที่มีความสามารถระดับเดียวกัน โดยที่ผู้สอบกลุ่มหนึ่งเป็นตัวแทนกลุ่มหลักในประชากรเรียกว่า “กลุ่มอ้างอิง” (Reference Group: R) ซึ่งเป็นกลุ่มพื้นฐาน ส่วนอีกกลุ่มหนึ่งเป็นตัวแทนกลุ่มรองในประชากรเรียกว่า “กลุ่มเปรียบเทียบ” (Focal Group: F) ซึ่งตามปกติเป็นกลุ่มผู้สอบที่สนใจจะทำการศึกษาการทำหน้าที่ต่างกันของข้อสอบ (Angoff, 1993

อ้างถึงใน วลีมาศ แซ่อึ้ง, 2543) ถ้าข้อสอบทำหน้าที่ต่างกันแล้วโอกาสในการตอบข้อสอบถูกของผู้สอบแต่ละกลุ่มจะไม่เท่ากัน โดยคาดว่าผู้สอบกลุ่มอ้างอิงจะได้เปรียบในการตอบข้อสอบ ส่วนผู้สอบกลุ่มเปรียบเทียบคาดว่าจะเสียเปรียบในการตอบข้อสอบ

ในช่วงปลายทศวรรษ 1960 มีการศึกษาเรื่องการทำหน้าที่ต่างกันของข้อสอบอย่างจริงจัง โดยเฉพาะกับข้อสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค (นิคม กีรติวรางกูร, 2542) สำหรับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบทวิวิภาค (Dichotomous) ด้วยวิธีนอนพาราเมตริก (Nonparametric) วิธีแมนเทิล-แฮนส์เซล (Mantel-Haenszel: MH) (Holland & Thayer, 1988; Mantel & Haenszel, 1959) และวิธีถดถอยโลจิสติก (Logistic Regression) (Roger & Swaminathan, 1993; Swaminathan & Rogers, 1990) เป็น 2 วิธีที่ได้รับความนิยมมากที่สุด (Su & Wang, 2005)

สำหรับวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาค ได้รับความสนใจในไม่นานมานี้ เนื่องจากมีการเพิ่มขึ้นของการใช้แบบสอบที่มีการตรวจให้คะแนนแบบพหุวิภาค ทำให้ขั้นตอนการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาคเริ่มเป็นที่สนใจ และในบริบทของประเทศไทยการศึกษาการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาคยังไม่แพร่หลาย ผู้วิจัยจึงสนใจศึกษาในกรณีของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาค ซึ่งปัจจุบันมีวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบประเภทนี้อยู่หลายวิธี ทั้งนี้เพราะ มีการศึกษาและคิดค้นวิธีการต่างๆเพื่อให้สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในเงื่อนไขต่าง ๆ ได้อย่างมีประสิทธิภาพมากขึ้น (Su & Wang, 2005) ซึ่งแต่ละวิธีมีวิธีการพอสังเขป ดังนี้

กลุ่มวิธีที่ใช้คะแนนที่สังเกตได้ วิธีในกลุ่มนี้มักวิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory: CTT) หรือเรียกว่ากลุ่มที่ไม่ใช้ทฤษฎีการตอบสนองข้อสอบ (Non-IRT Approach) โดยใช้คะแนนรวมของผู้สอบเป็นเกณฑ์การจับคู่ของกลุ่มผู้สอบ วิธีการตรวจสอบที่สำคัญในกลุ่มนี้ ได้แก่ การวิเคราะห์ความแปรปรวน (ANOVA) วิธีวิเคราะห์การถดถอยโลจิสติกพหุวิภาค (Polytomous Logistic Regression) วิธีแมนเทล-แฮนส์เซลทั่วไป (General Mantel-Haenszel: MH) และวิธีดัชนีมาตรฐานพหุวิภาค (Polytomous Standardization) (ศิริชัย กาญจนวาสี, 2545)

กลุ่มวิธีที่ใช้คุณลักษณะแฝง วิธีในกลุ่มนี้ใช้คุณลักษณะหรือตัวแปรแฝง ซึ่งวิเคราะห์บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) สำหรับใช้เป็นเกณฑ์การจับคู่กลุ่มผู้สอบ วิธีการตรวจสอบที่สำคัญในกลุ่มนี้ ได้แก่ วิธีการให้คะแนนบางส่วนทั่วไป (Generalized Partial Credit Model: GPCM) วิธีอัตราส่วนไลค์ลิฮูดในรูปทั่วไป (General IRT

Likelihood Ratio) วิธีการให้คะแนนบางส่วน (Partial Credit Model: PCM) และวิธีซิปเทสท์พหุวิภาค (Polytomous SIBTEST) (ศิริชัย กาญจนวาสี, 2545)

ทั้งสองกลุ่มวิธีมีความแตกต่างกัน โดยที่กลุ่มวิธีที่ใช้คะแนนที่สังเกตได้ จะมีค่าพารามิเตอร์แปรเปลี่ยนไปตามกลุ่มผู้สอบ แต่กลุ่มวิธีที่ใช้คุณลักษณะแฝงจะมีค่าพารามิเตอร์คงที่ไม่ว่าจะใช้กลุ่มผู้สอบใด ๆ (ศิริชัย กาญจนวาสี, 2545)

นอกจากกลุ่มวิธีที่ใช้คุณลักษณะแฝง ซึ่งวิเคราะห์บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) ดังที่กล่าวมาแล้ว ปัจจุบันนี้มีวิธีการในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่หลากหลายมากขึ้น Stark et al. (2006) กล่าวไว้ว่า ปัจจุบันมีวิธีการ 2 วิธีที่มีความสามารถในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาคและมีกระบวนการพื้นฐานสำหรับระบุการทำหน้าที่ต่างกันของข้อสอบที่คล้ายกัน นั่นคือ วิธีวิเคราะห์บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) และวิธีวิเคราะห์บนพื้นฐานของการวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis: CFA)

นอกจากนั้นยังพบว่าวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย (Mean and Covariance Structures Method: MACS) วิธีวิเคราะห์บนพื้นฐานของการวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) เป็นวิธีที่เหมาะสมในการนำมาใช้กับเงื่อนไขการทดลองที่มีการจำลองข้อมูล (Simulate) และนำมาใช้ได้ดีกับกรณีที่เป็นข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาค (Polytomous) ในขณะที่กลุ่มตัวอย่างมีจำนวนน้อยและจำลองข้อมูลบนค่าความยาก (thresholds) ของข้อสอบ (Stark et al., 2006)

ในส่วนของวิธีวิเคราะห์บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ (IRT) Su & Wang (2005) กล่าวไว้ว่า วิธีถดถอยโลจิสติกแบบพหุนาม (Multinomial Logistic Regression: MLR) เป็นวิธีที่ผ่านการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาค (Polytomous) ว่ามีนัยสำคัญแล้ว แต่การแปลผลค่อนข้างยาก ดังนั้น วิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก (Logistic Discriminant Function Analysis: LDFA) จึงได้นำมาใช้ในการตรวจสอบแทนที่ เนื่องจากมีประโยชน์มากสำหรับกลุ่มตัวอย่างขนาดเล็กและสะดวกในการแปลผลได้ดีกว่าวิธี MLR

จากการศึกษางานวิจัยที่ผ่านมา จึงมีประเด็นปัญหาที่น่าสนใจ คือ ยังไม่มีการศึกษาเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาคด้วยวิธี MACS และ วิธี LDFA ว่าวิธีใดมีประสิทธิภาพดีกว่า โดยประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบพิจารณาได้จากค่าอำนาจการทดสอบ (Power of Test)

อัตราความคลาดเคลื่อนประเภทที่ 1 (Type I Error Rates) และอัตราความคลาดเคลื่อนประเภทที่ 2 (Type II Error Rates) จากการศึกษางานวิจัยที่ผ่านมาพบว่ามีปัจจัยบางตัวที่มีผลต่อความถูกต้องแม่นยำในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เช่น ขนาดกลุ่มตัวอย่าง ความยาวของแบบสอบ ลักษณะของข้อสอบ การแจกแจงค่าความสามารถ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันเป็นต้น (กาญจนา วัธนสุนทร, 2537; เกษร หว่างจิตร, 2539; จิติมา วรรณศรี, 2539; รัชรินทร์ มุคดา, 2540; Swaminathan & Rogers,1990; Cohen & Kim, 1993; Rogers & Swaminathan, 1993; Uttara & Milsap, 1994; Narayanan & Swaminathan, 1994, 1996; Flowers, Claudia et al., 1997; Stark et al., 2006)

โดยเฉพาะการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย พบว่า อำนาจการทดสอบจะมีค่าเพิ่มขึ้นและอัตราความคลาดเคลื่อนประเภทที่ 1 จะมีค่าลดลง เมื่อแบบวัดเป็นแบบวัดมิติเดียว ขนาดกลุ่มตัวอย่างเล็ก และจำลองข้อมูลบนค่าความยาก (thresholds) ส่วนวิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก พบว่าอำนาจการทดสอบจะมีค่าเพิ่มขึ้นและอัตราความคลาดเคลื่อนประเภทที่ 1 จะมีค่าลดลง เมื่อเป็นการตรวจสอบการทำหน้าที่ต่างกันแบบอเนกรูป (Nonuniform DIF) สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันไม่เกิน 20% ขนาดกลุ่มตัวอย่างเล็ก และเมื่อความยาวแบบสอบเป็น 25 ข้อ

ดังนั้น จากประเด็นปัญหาข้างต้น ผู้วิจัยจึงสนใจศึกษาเปรียบเทียบประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี MACS และวิธี LDFA โดยพิจารณาประสิทธิภาพจากอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบโดยใช้แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ 5 ตัวเลือกที่มีคำตอบถูกหลายคำตอบ และให้ตัดสินทุกคำตอบ (Multiple True-false Answer) คะแนนเต็มข้อละ 5 คะแนนของสุนิสา จุ้ยม่วงศรี (2546) ตามเงื่อนไขด้านขนาดกลุ่มตัวอย่างและความยาวของแบบสอบซึ่งเป็นปัจจัยที่ทดสอบแล้วว่ามีนัยสำคัญในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่วิเคราะห์บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ และการวิเคราะห์องค์ประกอบเชิงยืนยัน โดยกลุ่มที่จะทำการเปรียบเทียบการทำหน้าที่ต่างกัน ได้แก่ เพศ เนื่องจากงานวิจัยของรักชนก ยี่สุ่นศรี (2544) ได้วิเคราะห์สาเหตุการทำหน้าที่ต่างกันของแบบสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาวิชาภาษาอังกฤษและคณิตศาสตร์ พบว่าแบบสอบวิชาคณิตศาสตร์ทำหน้าที่ต่างกันตามเพศ ซึ่งสอดคล้องกับผลการศึกษาของ นพมาศ พิพัฒนสุข (2541) ได้ศึกษาเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยเก็บข้อมูลจากแบบสอบวัดความสามารถทางคณิตศาสตร์ที่ผู้วิจัยสร้างขึ้น พบว่ามีการทำหน้าที่ต่างกันตามตัวแปร เพศ ต่อมา Henderson

(2001) ศึกษาความลำเอียงของข้อสอบระดับมัธยมศึกษาตอนปลาย พบว่าวิชาภาษาอังกฤษ สังคมศึกษา คณิตศาสตร์ และชีววิทยา มีความลำเอียงตามตัวแปร เพศ และจากการศึกษาของ Azen (2002) พบว่าข้อสอบสอบเข้ามหาวิทยาลัยของประเทศอิสราเอลมีความลำเอียงตามตัวแปรเพศดังนั้น ผู้วิจัยจึงเลือกตัวแปรด้านเพศมาทำการเปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบตามทั้ง 2 วิธีที่ได้กล่าวมาแล้ว

การวิจัยในครั้งนี้จะเป็นแนวทางในการศึกษา ค้นคว้า วิจัย เกี่ยวกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาค อีกทั้งยังเป็นการขยายขอบเขตของการศึกษาค้นคว้าเกี่ยวกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและเป็นการขยายองค์ความรู้ทางด้านวัดผลการศึกษาอีกด้วย

คำถามวิจัย

1. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาค วิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย (Mean and Covariance Structures Method: MACS) และวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก (Logistic Discriminant Function Analysis: LDFA) เมื่อเงื่อนไขความยาวของแบบสอบและขนาดกลุ่มตัวอย่างที่แตกต่างกัน จะส่งผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบอย่างไร

2. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาค วิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย (Mean and Covariance Structures Method: MACS) และวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก (Logistic Discriminant Function Analysis: LDFA) เมื่อเงื่อนไขความยาวของแบบสอบและขนาดกลุ่มตัวอย่างที่แตกต่างกัน จะส่งผลต่ออำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบอย่างไร

วัตถุประสงค์ของการวิจัย

การวิจัยนี้มีวัตถุประสงค์สำคัญเพื่อเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาค ระหว่างวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ยกับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก โดยกำหนดวัตถุประสงค์เฉพาะในการศึกษาด้านต่าง ๆ ดังนี้

1. เพื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 2 วิธี ภายใต้เงื่อนไขของปัจจัยที่แตกต่างกันทางด้าน

1.1 ความยาวของแบบสอบ

1.2 ขนาดกลุ่มตัวอย่าง

2. เพื่อเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 2 วิธี ภายใต้เงื่อนไขของปัจจัยที่แตกต่างกันทางด้าน

2.1 ความยาวของแบบสอบ

2.2 ขนาดกลุ่มตัวอย่าง

สมมติฐานการวิจัย

1. สำหรับเงื่อนไขด้านความยาวของแบบสอบ พบว่า กรณีที่วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเป็นวิธีที่ใช้พื้นฐานของการวิเคราะห์องค์ประกอบ ระดับความยาวของแบบสอบจะไม่มีอิทธิพลต่ออำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 แต่ความยาวของแบบสอบจะมีอิทธิพลต่ออำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่เป็นวิธีที่ใช้พื้นฐานของทฤษฎีการตอบสนองข้อสอบ โดยข้อสอบที่มีความยาวปานกลางขึ้นไปจะส่งผลให้อำนาจการทดสอบมีค่ามากขึ้น และอัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าน้อยลง

2. สำหรับเงื่อนไขด้านขนาดกลุ่มตัวอย่าง พบว่า ผลการวิจัยเป็นไปในสองทิศทาง คือในกรณีที่วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเป็นวิธีที่ใช้พื้นฐานของทฤษฎีการตอบสนองข้อสอบ จะพบว่าเมื่อขนาดกลุ่มตัวอย่างมีค่าเพิ่มมากขึ้นจะส่งผลให้อำนาจการทดสอบมีค่ามากขึ้น และอัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าน้อยลง และกรณีที่วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเป็นวิธีที่ใช้พื้นฐานของการวิเคราะห์องค์ประกอบเชิงยืนยัน จะพบว่าเมื่อขนาดกลุ่มตัวอย่างมีค่าน้อยลงจะส่งผลให้อำนาจการทดสอบมีค่ามากขึ้น และอัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าน้อยลง

ขอบเขตของการวิจัย

1. ประชากรที่ใช้ในการวิจัยครั้งนี้ ได้แก่ นักเรียนชั้นมัธยมศึกษาปีที่ 2 ปีการศึกษา 2550 โรงเรียนมัธยมศึกษา กรุงเทพมหานคร (สาเหตุที่เลือกกรุงเทพมหานครเท่านั้น เนื่องจากเพื่อป้องกันการทำหน้าที่ต่างกันที่มาจากตัวแปรสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน) สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (สพฐ.) จำนวน 116 โรงเรียน รวมนักเรียน 54,990 คน (สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ, 2550) โดยแบ่งเป็น นักเรียนชาย 27,954 คน และนักเรียนหญิง 27,036 คน

2. การวิจัยครั้งนี้ ใช้วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 2 วิธี ได้แก่ วิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย กับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติกซึ่งทั้งสองวิธีมีกระบวนการพื้นฐานสำหรับระบุการทำหน้าที่ต่างกันของข้อสอบที่คล้ายคลึงกัน และเป็นสองวิธีที่มีความสามารถในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบพหุวิภาค

3. ตัวแปรที่ศึกษา มีดังนี้

3.1 ตัวแปรอิสระมีทั้งหมด 2 ตัว คือ ความยาวของแบบสอบ และขนาดกลุ่มตัวอย่าง

3.1.1 ความยาวของแบบสอบ มี 2 ระดับ

(1) แบบสอบที่ตรวจให้คะแนนหลายค่าที่มีจำนวน 30 ข้อ

(2) แบบสอบที่ตรวจให้คะแนนหลายค่าที่มีจำนวน 40 ข้อ

สาเหตุที่เลือกแบบสอบที่มีจำนวน 30 และ 40 ข้อ เนื่องจากการศึกษาที่ผ่านมาพบว่าข้อสอบที่มีความยาวปานกลางขึ้นไปจะส่งผลต่อประสิทธิภาพในการตรวจสอบมากที่สุด อีกทั้งเป็นระดับความยาวที่เหมาะสมกับการนำไปใช้เก็บข้อมูลจริง (จิติมา วรรณศรี, 2539; ญาณภัทร สีหะมงคล, 2540; ปิยะทิพย์ ตินวร, 2549; Kim & Cohen, 1998)

3.1.2 ขนาดกลุ่มตัวอย่างมี 3 ระดับ

(1) 200 คน

(2) 500 คน

(3) 1,000 คน

3.2 ตัวแปรตาม คือ อำนาจการทดสอบ และอัตราความคลาดเคลื่อนประเภทที่ 1

4. ข้อมูลที่ใช้ในการศึกษา เป็นข้อมูลจริงภายใต้การใช้คะแนนที่ได้จากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ด้านเหตุผล 5 ตัวเลือกที่มีคำตอบถูกหลายคำตอบ และให้ตัดสินทุกคำตอบ (Multiple True-false Answer) เรื่อง จำนวนเต็ม เศษส่วน และทศนิยม ระดับชั้นมัธยมศึกษาตอนต้น คะแนนเต็มข้อละ 5 คะแนน จำนวน 50 ข้อ ของ สุนิสา จุ้ยม่วงศรี (2546)

5. กลุ่มที่จะทำการเปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบทั้ง 2 วิธี ได้แก่ เพศ แบ่งเป็นเพศชายและเพศหญิง

ข้อจำกัดของการวิจัย

ในการวิจัยครั้งนี้ พบว่าในการตรวจสอบความเป็นเอกมิติของแบบสอบที่มีความยาว 40 ข้อ ไม่สามารถบ่งบอกได้ชัดเจนว่าแบบสอบมีความเป็นมิติเดียว ซึ่งการที่จะวิเคราะห์โดยใช้ทฤษฎีการตอบสนองข้อสอบต่อไปนั้น ถือว่าเป็นการละเมิดข้อตกลงเบื้องต้น แต่จากการศึกษาเอกสารที่เกี่ยวข้อง พบว่า พรรณี จินตมาศ (2540) ได้ตรวจสอบความเป็นเอกมิติของแบบสอบวิชา

คณิตศาสตร์ เพื่อนำผลมาวิเคราะห์ความลำเอียง 3 วิธี ได้แก่ วิธีแปลงค่าความยาก วิธีแมนเทล-แฮนส์เซล และวิธีซิปเทสท์ และพบว่าไม่เป็นเอกมิติ แต่มีความแกร่งพอที่จะสามารถนำไปวิเคราะห์หาค่าความลำเอียงต่อไปได้ ซึ่งสอดคล้องกับงานวิจัยของ อ้อมใจ คงเสาร์ (2540) ที่ตรวจสอบความเป็นเอกมิติของแบบสอบวิชาคณิตศาสตร์จำนวน 17 ข้อและพบว่าไม่เป็นเอกมิติเช่นเดียวกัน อีกทั้งยังพบว่า ญาณภัทร สีหะมงคล (2540) กล่าวว่า การตรวจสอบความเป็นเอกมิติ หากพิจารณาตามเกณฑ์ของ Reckase อ้างถึงใน Raju, 1993 ที่เสนอว่าค่าร้อยละของความแปรปรวนควรมีค่าไม่น้อยกว่าร้อยละ 20 จึงจะถือว่าแบบสอบมีความเป็นเอกมิติ ซึ่งในทางปฏิบัติหาพบได้ยากเนื่องจากแบบสอบประกอบด้วยเนื้อหาหลายอย่างและวัดหลายสมรรถภาพ

คำจำกัดความที่ใช้ในการวิจัย

การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ข้อสอบที่วัดคุณลักษณะแฝงอื่นนอกเหนือจากลักษณะแฝงที่ต้องการ จะส่งผลให้ผู้สอบต่างกลุ่มที่นำมาจับคู่เปรียบเทียบกัน มีโอกาสในการตอบข้อสอบถูกแตกต่างกัน ทั้ง ๆ ที่มีความสามารถที่ต้องการวัดเท่ากัน แสดงว่าข้อสอบข้อนั้นทำหน้าที่ต่างกัน ในกรณีนี้ข้อสอบรายวิชาคณิตศาสตร์ ของ สุนิสา จุ้ยม่วงศรี (2546) ทำหน้าที่ต่างกันตามตัวแปรเพศ

ข้อสอบที่ทำหน้าที่ต่างกัน หมายถึง ข้อสอบที่ทำให้ผลการตอบของผู้สอบที่มีความสามารถเท่ากันในสิ่งที่ต้องการวัด มีโอกาสตอบข้อสอบข้อนั้นได้ถูกต้องไม่เท่ากันเนื่องจากอยู่ในกลุ่มย่อยต่างกัน ซึ่งในการวิจัยครั้งนี้ ศึกษาการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มผู้สอบจำแนกตาม เพศ

วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ หมายถึง วิธีการวิเคราะห์ทางสถิติที่ใช้เพื่อบ่งบอกว่ามีการทำหน้าที่ต่างกันของข้อสอบระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ซึ่งวิธีการในงานวิจัยนี้เป็นวิธีการตรวจสอบที่ใช้กับข้อสอบที่มีการตรวจให้คะแนนแบบหลายค่า ได้แก่

วิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย หมายถึง วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตั้งอยู่บนพื้นฐานของการวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) โดยมีหลักการสำคัญคือ การทดสอบความไม่แปรเปลี่ยนของโมเดลลิสเรลระหว่างกลุ่มเพศชายและเพศหญิง โดยเป็นการทดสอบสมมติฐานทางสถิติว่า โมเดลลิสเรลมีเมทริกซ์น้ำหนักองค์ประกอบ (Factor Loading) ไม่แปรเปลี่ยนระหว่างกลุ่มประชากรหรือไม่ ในงานวิจัยนี้ตรวจสอบโดยใช้โปรแกรมคอมพิวเตอร์ลิสเรล (LISREL)

วิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก หมายถึง วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบ (IRT) เนื่องจาก IRT เป็นทฤษฎีการทดสอบที่ใช้โมเดลโลจิสติก โดยมีหลักสำคัญของการวิเคราะห์ คือ การ

เปรียบเทียบค่าไค-สแควร์ของผลการตอบข้อสอบของกลุ่มเพศชายและเพศหญิง เพื่อแสดงการเปรียบเทียบข้อมูลผลการตอบข้อสอบของทั้ง 2 กลุ่ม กับข้อมูลจากทฤษฎี เมื่อเทียบกับค่าวิกฤติในตารางแล้ว หากพบว่ามีค่าเกินกว่าค่าวิกฤติแสดงว่าข้อมูลทั้ง 2 กลุ่มไม่มีความสอดคล้องกันในทางตรงกันข้ามหากมีค่าต่ำกว่าค่าวิกฤติแสดงว่าข้อมูลทั้ง 2 กลุ่มมีความสอดคล้องกันอย่างมีนัยสำคัญทางสถิติ ในงานวิจัยนี้ตรวจสอบโดยใช้โปรแกรมคอมพิวเตอร์ MULTILOG

วิธีให้คะแนนบางส่วนทั่วไป (Generalized Partial Credit Model: GPCM) หมายถึง วิธีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ตามแนวทฤษฎีการตอบสนองข้อสอบ (IRT) ที่ใช้กับข้อสอบที่ให้คะแนนแบบหลายค่า และใช้เป็นวิธีมาตรฐานในการระบุว่าข้อสอบข้อใดทำหน้าที่ต่างกัน ซึ่งสามารถตรวจการทำหน้าที่ต่างกันของข้อสอบได้ 3 วิธี คือ วิธีการทดสอบไค-สแควร์ของลอร์ด (Lord's Chi-square, 1980) วิธีการวัดพื้นที่สัมบูรณ์ของราชู (Exact Area Measures) (Raju, 1990) และวิธีการวัดพื้นที่ช่วงปิด ในการวิจัยครั้งนี้ผู้วิจัยเลือกใช้วิธีการทดสอบไค-สแควร์ของลอร์ด (Lord's Chi-square) โดยพิจารณาค่าไค-สแควร์เทียบกับค่าวิกฤติในตารางการแจกจงแบบไค-สแควร์

ข้อสอบที่มีการให้คะแนนแบบหลายค่า หมายถึง ข้อสอบที่มีการกำหนดลำดับขั้นของการให้คะแนนผลการตอบข้อสอบไว้เป็นจำนวนเต็มหลายระดับคะแนน ซึ่งในการวิจัยนี้ใช้แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ด้านเหตุผล 5 ตัวเลือกที่มีคำตอบถูกหลายคำตอบ และให้ตัดสินทุกคำตอบ (Multiple True-false Answer) เรื่องจำนวนเต็ม เศษส่วน และทศนิยม ระดับชั้นมัธยมศึกษาตอนต้น คะแนนเต็มข้อละ 5 คะแนน จำนวน 50 ข้อ ของ สุนิสา จุ้ยม่วงศรี (2546)

ข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป หมายถึง ข้อสอบที่ให้ผู้สอบกลุ่มหนึ่งมีโอกาสในการตอบข้อสอบถูกมากกว่าผู้สอบอีกกลุ่มหนึ่งอย่างสม่ำเสมอในทุกระดับความสามารถ เมื่อพิจารณาโค้งคุณลักษณะข้อสอบของผู้สอบ 2 กลุ่ม จะพบว่าไม่มีปฏิสัมพันธ์ระหว่างโค้งลักษณะข้อสอบในทุกระดับความสามารถ

ข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป หมายถึง ข้อสอบที่ให้โอกาสในการตอบข้อสอบถูกของผู้สอบระหว่างกลุ่มไม่สม่ำเสมอกันในทุกระดับความสามารถ เมื่อพิจารณาโค้งคุณลักษณะข้อสอบของผู้สอบ 2 กลุ่ม พบว่าในบางระดับความสามารถจะมีปฏิสัมพันธ์ระหว่างโค้งลักษณะข้อสอบ

กลุ่มอ้างอิง หมายถึง กลุ่มผู้สอบที่คาดว่าจะได้รับประโยชน์จากการตอบข้อสอบที่ทำหน้าที่ต่างกัน คือเป็นกลุ่มที่มีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องสูงกว่าผู้สอบอีกกลุ่มหนึ่งทั้ง ๆ ที่มีความสามารถเท่ากัน สำหรับรายวิชาคณิตศาสตร์ กลุ่มอ้างอิงในงานวิจัยนี้ได้แก่ เพศชาย

กลุ่มเปรียบเทียบ หมายถึง กลุ่มผู้สอบที่คาดว่าจะเสียประโยชน์จากการตอบข้อสอบที่ทำหน้าที่ต่างกัน คือเป็นกลุ่มที่มีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องต่ำกว่าผู้สอบอีกกลุ่มหนึ่ง ทั้ง ๆ ที่มีความสามารถเท่ากัน สำหรับรายวิชาคณิตศาสตร์ กลุ่มเปรียบเทียบในงานวิจัยนี้ได้แก่ เพศหญิง

ความเที่ยง หมายถึง คุณภาพของแบบสอบซึ่งพิจารณาได้จากความสอดคล้องภายใน (Internal Consistency) ซึ่งคำนวณค่าความเที่ยงโดยหาค่าสัมประสิทธิ์แอลฟาของครอนบาค

ค่าฟังก์ชันสารสนเทศของแบบสอบ หมายถึง ผลรวมของค่าสารสนเทศของข้อสอบทุก ๆ ข้อในแบบสอบ สำหรับความสามารถของผู้สอบต่าง ๆ กัน ถ้ามีค่าสูงที่ความสามารถระดับใด แสดงว่า แบบสอบนั้นสามารถจำแนกความสามารถผู้สอบได้ดี ณ ความสามารถนั้น โดยในงานวิจัยนี้จะวัดจากแบบสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ด้านเหตุผล 5 ตัวเลือกที่มีคำตอบถูกหลายคำตอบ และให้ตัดสินทุกคำตอบ (Multiple True-false Answer) เรื่องจำนวนเต็ม เศษส่วน และทศนิยม ระดับชั้นมัธยมศึกษาตอนต้น คะแนนเต็มข้อละ 5 คะแนน จำนวน 50 ข้อ ของ สุนิสา จุ้ยม่วงศรี (2546)

ประสิทธิภาพในการตรวจสอบ หมายถึง ความถูกต้องของการระบุการทำหน้าที่ต่างกันของข้อสอบ จากการตรวจสอบด้วยวิธี MACS และ LDFA ซึ่งพิจารณาได้จากอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1

อำนาจการทดสอบ หมายถึง ร้อยละของจำนวนข้อสอบที่ระบุถูกต้องว่าทำหน้าที่ต่างกัน โดยคำนวณได้จากจำนวนข้อสอบที่ระบุถูกต้องว่าทำหน้าที่ต่างกันซึ่งตรวจสอบด้วยวิธีที่ศึกษา ได้แก่ วิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย (Mean and Covariance Structures Method: MACS) และวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก (Logistic Discriminant Function Analysis: LDFA) ต่อจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งหมดในแบบสอบซึ่งตรวจสอบด้วยวิธีเกณฑ์ได้แก่ วิธีให้คะแนนบางส่วนทั่วไป (Generalized Partial Credit Model: GPCM) ซึ่งเป็นวิธีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ตามแนวทฤษฎีการตอบสนองข้อสอบ (IRT) ที่ใช้กับข้อสอบที่ให้คะแนนแบบหลายค่า

อัตราความคลาดเคลื่อนประเภทที่ 1 หมายถึง ร้อยละของจำนวนข้อสอบที่ระบุผิดพลาดว่าทำหน้าที่ต่างกัน ทั้งที่ความเป็นจริงข้อสอบทำหน้าที่ไม่ต่างกัน โดยคำนวณได้จากจำนวนข้อสอบที่ระบุผิดพลาดว่าทำหน้าที่ต่างกันซึ่งตรวจสอบโดยวิธีที่ศึกษา วิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย (Mean and Covariance Structures Method: MACS) และวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก (Logistic Discriminant Function Analysis: LDFA) ต่อจำนวนข้อสอบที่ทำหน้าที่ไม่ต่างกันทั้งหมดในแบบสอบที่ตรวจสอบด้วยวิธีเกณฑ์

ขนาดกลุ่มตัวอย่าง หมายถึง จำนวนผู้สอบแต่ละขนาดที่ใช้ในการศึกษา ซึ่งการศึกษาครั้งนี้กลุ่มตัวอย่างที่ศึกษามี 3 ขนาดคือ 200 คน 500 คน และ 1,000 คน

ความยาวแบบสอบ หมายถึง จำนวนข้อสอบในแบบสอบ ในการวิจัยครั้งนี้ศึกษาความยาวแบบสอบ 2 ขนาด ได้แก่ 30 ข้อ และ 40 ข้อ

ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อเป็นแนวทางในการปรับปรุงข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาค ให้มีความยุติธรรมต่อกลุ่มผู้สอบ ซึ่งเป็นอีกทางเลือกหนึ่งในการพัฒนาแบบสอบให้มีคุณภาพ
2. เพื่อเป็นแนวทางในการเลือกวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาคระหว่างวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย (Mean and Covariance Structures Method: MACS) และวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก (Logistic Discriminant Function Analysis: LDFA) ภายใต้เงื่อนไขของปัจจัยที่ศึกษา 2 ตัว คือ ความยาวของแบบสอบ และขนาดกลุ่มตัวอย่าง
3. เพื่อเป็นแนวทางในการศึกษา ค้นคว้า วิจัย เกี่ยวกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาค โดยใช้วิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย (Mean and Covariance Structures Method: MACS) และวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก (Logistic Discriminant Function Analysis: LDFA)
4. เป็นการขยายขอบเขตของการศึกษาค้นคว้าเกี่ยวกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ และเป็นการขยายองค์ความรู้ทางด้านวัดผลการศึกษา

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

การศึกษาเอกสารที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบ ผู้วิจัยจะนำเสนอองค์ความรู้เชิงทฤษฎี และงานวิจัยต่าง ๆ ที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบ โดยได้แบ่งการนำเสนอออกเป็น 3 ตอน ดังนี้

<u>ตอนที่ 1</u> องค์ความรู้เชิงทฤษฎี โดยแบ่งออกเป็น 6 หัวข้อ ดังนี้

1. ความหมายของการทำหน้าที่ต่างกันของข้อสอบ
2. ความเป็นมาของการทำหน้าที่ต่างกันของข้อสอบ
3. ประเภทของการทำหน้าที่ต่างกันของข้อสอบ
4. หลักการและวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ
5. แนวคิดเกี่ยวกับการวิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาค (Polytomous IRT Model)
6. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย กับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก

<u>ตอนที่ 2</u> งานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบ

<u>ตอนที่ 3</u> กรอบแนวคิดในการวิจัย

โดยแต่ละตอนมีรายละเอียด ดังต่อไปนี้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตอนที่ 1 องค์ความรู้เชิงทฤษฎีเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบ

ในส่วนนี้ผู้วิจัยขอนำเสนอองค์ความรู้เชิงทฤษฎีที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบโดยนำเสนอตามลำดับ ดังนี้ ความหมาย ความเป็นมา ประเภท หลักการ วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ แนวคิดเกี่ยวกับการวิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาค (Polytomous IRT Model) และการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย กับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก

1. ความหมายของการทำหน้าที่ต่างกันของข้อสอบ

การทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning) เดิมเรียกว่าความลำเอียงของข้อสอบ (Item Bias) แต่ในระยะหลังเปลี่ยนมาใช้คำว่าการทำหน้าที่ต่างกันของข้อสอบเพราะเป็นคำที่เป็นกลางและมีความเหมาะสมมากกว่า (Holland & Thayer, 1988) มีนักวิจัยทางการวัดผลหลายท่านได้ให้ความหมายของความลำเอียงของข้อสอบและการทำหน้าที่ต่างกันของข้อสอบ โดยสามารถจัดเป็นประเภท ได้ดังนี้

ความหมายของการทำหน้าที่ต่างกันของข้อสอบที่เกี่ยวข้องกับความแตกต่างของสัดส่วนของจำนวนผู้ตอบข้อสอบ ได้แก่

Scheuneman (1979) กล่าวว่า ความลำเอียงของข้อสอบ หมายถึง สัดส่วนของผู้สอบที่ตอบข้อสอบได้ถูกต้องไม่เท่ากันในแต่ละกลุ่มประชากรที่ใช้ในการศึกษา เมื่อกลุ่มผู้สอบมีคะแนนเท่ากันและข้อสอบมีความเป็นเอกพันธ์

Rudner, Getson, & Knight (1980) กล่าวว่า ความลำเอียงของข้อสอบ หมายถึง ข้อสอบที่มีค่าความยากสัมพัทธ์สำหรับสมาชิกของผู้สอบกลุ่มหนึ่งมากกว่าสมาชิกของผู้สอบอีกกลุ่มหนึ่ง

ความหมายของการทำหน้าที่ต่างกันของข้อสอบที่เกี่ยวข้องกับโอกาสการตอบข้อสอบ ได้แก่

Hulin, Drasgow, & Parson (1983) กล่าวว่า ความลำเอียงของข้อสอบ หมายถึง โอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกันสำหรับการวัดความสามารถ หรือโอกาสในการตอบข้อสอบในทางบวกแตกต่างกันสำหรับการวัดเจตคติ เมื่อผู้สอบที่มีคุณลักษณะของการวัดในปริมาณเท่ากัน แต่มาจากกลุ่มประชากรย่อยที่แตกต่างกัน

Dorans & Kulick (1986) กล่าวว่า ความลำเอียงของข้อสอบ หมายถึง โอกาสในการตอบข้อสอบได้ถูกต้องของผู้สอบกลุ่มหนึ่ง มีค่าต่ำกว่าหรือสูงกว่าผู้สอบอีกกลุ่มหนึ่งที่มีระดับความสามารถเดียวกัน

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Holland & Wainer (1993) กล่าวว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง สารสนเทศทางสถิติของข้อสอบที่ได้จากผลการตอบของผู้สอบต่างกลุ่มกัน และมีความสามารถเท่ากันแต่มีโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกัน

Mazor & Clauser (1995) กล่าวว่า ข้อสอบทำหน้าที่ต่างกันเมื่อผู้สอบที่มีความสามารถระดับเดียวกัน แต่เป็นสมาชิกของกลุ่มย่อยต่างกัน มีโอกาสในการตอบข้อสอบถูกต้องแตกต่างกัน

ความหมายของการทำหน้าที่ต่างกันของข้อสอบที่เกี่ยวข้องกับความสามารถหลักหรือความสามารถรองของผู้สอบ ได้แก่

Kederman (1990) กล่าวว่า ความลำเอียงของข้อสอบ หมายถึง คะแนนข้อสอบของกลุ่มผู้สอบที่มีความสามารถเท่ากัน แต่มาจากต่างกลุ่มกัน มีความแตกต่างกันอย่างเป็นระบบ

Millsap & Everson (1993) กล่าวว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ความแตกต่างในการทำหน้าที่ของแบบสอบหรือข้อสอบ ระหว่างกลุ่มผู้สอบซึ่งถูกจับคู่ตามคุณลักษณะที่วัดโดยแบบสอบหรือข้อสอบนั้น

Camilli & Shepard (1994) กล่าวว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ความเป็นพหุมิติในการวัดของข้อสอบ ซึ่งแสดงได้จากการแจกแจงความสามารถหลัก (Primary Ability) ของผู้สอบตั้งแต่ 2 กลุ่มขึ้นไปมีความสามารถเท่ากัน แต่มีการแจกแจงความสามารถรอง (Secondary Ability) แตกต่างกัน

ศิริชัย กาญจนวาสี (2545) กล่าวว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง การที่ข้อสอบทำให้ผู้สอบจากต่างกลุ่มกันที่มีความสามารถหรือคุณลักษณะที่มุ่งวัดเท่ากัน มีโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกัน หรือมีฟังก์ชันการตอบสนองข้อสอบแตกต่างกัน การทำหน้าที่ต่างกันของข้อสอบเกิดขึ้นเมื่อนำข้อสอบไปทดสอบกับผู้สอบกลุ่มย่อยต่างกัน ที่มีความสามารถหลัก (Primary Ability) ระดับเดียวกันหรือมีคุณลักษณะแฝง (Latent Trait) ที่ต้องการวัดเท่ากัน แต่มีความสามารถรอง (Secondary Ability) แตกต่างกัน ทำให้ผู้สอบต่างกลุ่มที่นำมาเปรียบเทียบมีโอกาสตอบข้อสอบถูกแตกต่างกัน

ความหมายของการทำหน้าที่ต่างกันของข้อสอบที่เกี่ยวข้องกับความยุติธรรมของการตอบข้อสอบ ได้แก่

Popham (1981) กล่าวว่า ความลำเอียงของข้อสอบ หมายถึง ความโน้มเอียงของข้อสอบที่เมื่อใช้คะแนนจากข้อสอบนั้นแล้ว ทำให้การตัดสินผลเป็นไปอย่างไม่ยุติธรรม

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

ดังนั้นกล่าวโดยสรุป การทำหน้าที่ต่างกันของข้อสอบ เป็นกรณีที่เกิดขึ้นเมื่อข้อสอบวัดคุณลักษณะแฝงอื่น นอกเหนือจากลักษณะแฝงที่ต้องการ จะส่งผลให้ผู้สอบต่างกลุ่มที่นำมาจับคู่เปรียบเทียบกัน มีโอกาสในการตอบข้อสอบถูกแตกต่างกัน ทั้งๆที่มีความสามารถที่ต้องการวัดเท่ากัน แสดงว่าข้อสอบข้อนั้นทำหน้าที่ต่างกัน เมื่อได้ทราบความหมายแล้ว ในลำดับต่อไปจะให้ความสนใจในเรื่องของความเป็นมาของการทำหน้าที่ต่างกัน

2. ความเป็นมาของการทำหน้าที่ต่างกันของข้อสอบ

ในการทดสอบแต่ละครั้ง ผู้สอบระหว่างกลุ่มย่อยอาจมีลักษณะแตกต่างกัน เช่น เชื้อชาติ ศาสนา วัฒนธรรม ภูมิลำเนา สังคม เพศ ภาษา อายุ และประสบการณ์ เป็นต้น ผู้สอบกลุ่มย่อยดังกล่าวอาจไม่ได้รับความยุติธรรมในการทำข้อสอบ โดยข้อสอบบางข้ออาจมีความลำเอียงเข้าข้างผู้สอบกลุ่มย่อยบางกลุ่มของผู้สอบทั้งหมด ซึ่งทำให้เกิดการได้เปรียบเสียเปรียบระหว่างผู้สอบกลุ่มย่อยด้วยกัน ทั้งๆที่สอบด้วยข้อสอบข้อเดียวกันหรือแบบสอบฉบับเดียวกัน สาเหตุดังกล่าวอาจเนื่องมาจากแบบสอบไม่ได้วัดความสามารถเป้าหมายที่ต้องการวัด (Target Ability: θ) เพียงอย่างเดียว แต่ยังวัดความสามารถแทรกซ้อนที่ไม่ต้องการวัด (Nuisance Ability: η) อีกด้วย ตัวอย่างเช่น แบบสอบวัดคำศัพท์ในวิชาภาษาอังกฤษฉบับหนึ่ง ข้อสอบบางข้ออาจถามความรู้สำหรับผู้ชายเป็นพิเศษ เช่น ความรู้เรื่องกีฬา ในขณะที่ข้อสอบบางข้ออาจถามความรู้สำหรับผู้หญิงโดยเฉพาะ เช่น ความรู้เกี่ยวกับงานในบ้าน จากสถานการณ์ดังกล่าว ทักษะวัดคำศัพท์ในวิชาภาษาอังกฤษเป็นความสามารถเป้าหมาย (θ) ส่วนทักษะวัดความรู้ด้านกีฬา (η_1) และงานในบ้าน (η_2) เป็นความสามารถแทรกซ้อน ข้อสอบทุกข้อในแบบสอบจะวัดความสามารถเป้าหมาย ส่วนข้อสอบบางข้อที่ทำหน้าที่ต่างกันจะวัดทั้งความสามารถเป้าหมายและความสามารถแทรกซ้อน (Nandakumar, 1993 อ้างถึงใน วลีมาศ แซ่อึ้ง, 2543) นั่นคือ ถ้าผู้สอบกลุ่มย่อยกลุ่มใดมีความสามารถแทรกซ้อนสูงกว่าก็มีโอกาสในการตอบข้อสอบได้ถูกต้องมากกว่า ทั้ง ๆ ที่ระดับความสามารถเป้าหมายที่ต้องการวัดเท่ากัน จึงมีผลทำให้ข้อสอบทำหน้าที่ต่างกัน (วลีมาศ แซ่อึ้ง, 2543)

การศึกษาถึงคุณภาพของข้อสอบจากผลการตรวจข้อสอบของผู้สอบกลุ่มต่าง ๆ ในประชากรมีมานานแล้ว แต่การศึกษาคุณภาพด้านความยุติธรรมของข้อสอบหรือแบบสอบระหว่างผู้สอบกลุ่มต่าง ๆ เริ่มศึกษากันอย่างจริงจังในช่วงปลายทศวรรษของปี ค.ศ.1960 มีการเสนอวิธีการต่าง ๆ เพื่อตรวจสอบความลำเอียงของข้อสอบ (Item Bias) ความลำเอียงของแบบสอบ (Test Bias) และความลำเอียงในการคัดเลือก (Selection Bias) โดยนิยามความลำเอียงว่าเป็นความคลาดเคลื่อนอย่างเป็นระบบ (Systematic Error) ที่เกิดขึ้นจากการวัดความพยายามของการ

ตรวจสอบความลำเอียงดังกล่าวดำเนินไปเพื่อจำแนกข้อสอบที่ทำหน้าที่ไม่เหมาะสมหรือไม่ยุติธรรมสำหรับปรับปรุง หรือตัดข้อสอบข้อนั้นออกจากแบบสอบ เป็นการขจัดข้อสอบที่ทำให้เกิดปัญหาความยุติธรรมระหว่างกลุ่มข้อสอบกลุ่มต่าง ๆ ที่มีลักษณะบางอย่างแตกต่างกัน เช่น เชื้อชาติ วัฒนธรรม ภูมิลำเนา สังคม เพศ ภาษา อายุ ประสบการณ์ เป็นต้น เพื่อพัฒนาแบบสอบให้มีคุณภาพเหมาะสมสำหรับนำไปใช้ทดสอบต่อไป (ศิริชัย กาญจนวาสี, 2545)

ในเวลาต่อมา นักวัดผลการศึกษาได้ทำการศึกษาความลำเอียงของข้อสอบ (Item Bias) กันอย่างกว้างขวาง ทำให้เกิดความสับสนของการใช้คำและความหมายมีประเด็นโต้แย้งกันว่า ความลำเอียงของข้อสอบ เป็นผลการตัดสินว่าข้อสอบมีความยุติธรรมหรือไม่ อันส่งผลต่อการบรรลุจุดมุ่งหมายของการใช้แบบสอบหรือความลำเอียงถึงข้อสอบ เป็นสารสนเทศทางสถิติที่ได้จากข้อสอบเกี่ยวกับความสัมพันธ์ระหว่างคุณลักษณะที่ข้อสอบมุ่งวัดกับประสบการณ์ของผู้สอบกลุ่มต่าง ๆ ที่ทำการสอบ เมื่อกลุ่มผู้สอบต่างกลุ่มกันตอบข้อสอบข้อเดียวกัน ความแตกต่างที่เกิดขึ้นอาจมาจากความไม่เหมาะสมของข้อคำถาม ซึ่งสามารถเกิดขึ้นได้หลายลักษณะ หรือประสบการณ์ของผู้สอบซึ่งอาจมีลักษณะพื้นฐานเดิมแตกต่างกันในหลายสถานการณ์จึงไม่เหมาะสมที่จะใช้คำว่า ข้อสอบลำเอียง (Biased Item) เนื่องจากเป็นภาษาที่มีความหมายในเชิงลบ ประกอบกับเกณฑ์ที่ใช้สำหรับตัดสินความลำเอียงยังมีความคลุมเครือและค่อนข้างสับสน ดังนั้น จึงควรเปลี่ยนมาใช้คำว่า การทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning: DIF) ซึ่งเป็นคำที่มีความหมายเป็นกลางและเหมาะสมกว่า (Holland & Thayer, 1988; Holland & Wainer, 1993)

การทำหน้าที่ต่างกันของข้อสอบ (DIF) กับความลำเอียงของข้อสอบ (Item Bias) มีแนวคิดที่แตกต่างกัน สำหรับการทำหน้าที่ต่างกันของข้อสอบ เป็นกระบวนการที่เน้นการใช้วิธีการทางสถิติสำหรับการตรวจสอบเพื่อให้ได้สารสนเทศเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มผู้สอบกลุ่มย่อยที่มีลักษณะเฉพาะบางอย่างแตกต่างกัน ส่วนความลำเอียงของข้อสอบ เป็นกระบวนการตัดสินความยุติธรรมของข้อสอบ โดยนำสารสนเทศการทำหน้าที่ต่างกันของข้อสอบมาวิเคราะห์เชิงตรรกะ (Logical Analysis) โดยผู้เชี่ยวชาญพิจารณาถึงการเขียนข้อสอบ เนื้อหาสาระของข้อสอบและจุดมุ่งหมายของการวัด เพื่อระบุว่าข้อสอบข้อนั้นลำเอียงเข้าข้างกลุ่มใดหรือไม่เพราะเหตุใดจึงเป็นการตัดสินความลำเอียงของข้อสอบ (Camilli & Shapard, 1994)

จากการศึกษาเอกสารและงานวิจัยข้างต้นผู้วิจัยสามารถสรุปได้ว่า การทำหน้าที่ต่างกันของข้อสอบ แต่เดิมใช้คำว่า “ความลำเอียงของข้อสอบ” (Item Bias) ซึ่งเป็นภาษาที่ใช้กันในทางสังคมและมีความหมายไปในทางลบ ระยะหลังนักวิจัยได้เปลี่ยนไปใช้คำใหม่ว่า “การทำหน้าที่ต่างกันของข้อสอบ” (Differential Item Functioning: DIF) แต่อย่างไรก็ตาม คำสองคำนี้มีจุดเน้นที่แตกต่างกัน คำว่า “ความลำเอียงของข้อสอบ” เน้นที่อิทธิพลที่สังเกตได้ของกลุ่มผู้สอบย่อยที่มุ่ง

ศึกษา ส่วนคำว่า "ข้อสอบทำหน้าที่ต่างกัน" นั้นเน้นที่คุณลักษณะทางสถิติของข้อสอบที่ตรวจสอบได้ด้วยวิธีการวิเคราะห์ทางสถิติ สำหรับสิ่งที่ต้องให้ความสำคัญลำดับต่อมา ได้แก่ ประเภทของการทำหน้าที่ต่างกันของข้อสอบ

3. ประเภทของการทำหน้าที่ต่างกันของข้อสอบ

การทำหน้าที่ต่างกันของข้อสอบ เป็นการเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มผู้สอบอย่างน้อย 2 กลุ่มขึ้นไป ปกตินิยมทำการเปรียบเทียบ 2 กลุ่ม ประกอบด้วยกลุ่มแรกเรียกว่า กลุ่มเปรียบเทียบ (Focal Group หรือกลุ่ม F) เป็นกลุ่มที่สนใจศึกษาและคาดว่าจะเป็นกลุ่มที่เสียเปรียบในการตอบข้อสอบ และกลุ่มที่สอง เรียกว่า กลุ่มอ้างอิง (Reference Group หรือกลุ่ม R) เป็นกลุ่มที่คาดว่าจะได้เปรียบในการตอบข้อสอบได้ถูกต้อง (ศิริชัย กาญจนวาสี, 2545)

ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ จะพบว่า ข้อสอบสามารถทำหน้าที่แตกต่างกันได้ 2 ประเภท (Mellenbergh, 1982) ได้แก่ การทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป (Uniform DIF) และแบบอเนกรูป (Nonuniform DIF)

1. ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป (Uniform DIF) หมายถึงข้อสอบที่ทำให้ผู้สอบกลุ่มหนึ่ง มีโอกาสในการตอบข้อสอบถูกมากกว่าผู้สอบอีกกลุ่มหนึ่งอย่างสม่ำเสมอกัน ในทุกระดับความสามารถ เมื่อพิจารณาโค้งคุณลักษณะข้อสอบของผู้สอบ 2 กลุ่ม จะพบว่าไม่มีปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับการเป็นสมาชิกของกลุ่ม (Group Membership)

2. ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป (Nonuniform DIF) หมายถึง ข้อสอบที่ทำให้โอกาสในการตอบข้อสอบถูกของผู้สอบระหว่างกลุ่มแตกต่างกันอย่างไม่สม่ำเสมอกันในทุกระดับความสามารถ เมื่อพิจารณาโค้งคุณลักษณะข้อสอบของผู้สอบ 2 กลุ่ม พบว่ามีปฏิสัมพันธ์ร่วมกันระหว่างระดับความสามารถของผู้สอบ กับการเป็นสมาชิกของกลุ่ม เช่น ที่ระดับความสามารถหนึ่งกลุ่มผู้สอบกลุ่มอ้างอิง (R) มีโอกาสในการตอบข้อสอบถูกมากกว่ากลุ่มผู้สอบเปรียบเทียบ (F) แต่ที่ระดับความสามารถอีกระดับหนึ่งกลุ่มผู้สอบเปรียบเทียบ (F) มีโอกาสในการตอบข้อสอบถูกมากกว่ากลุ่มผู้สอบกลุ่มอ้างอิง (R)

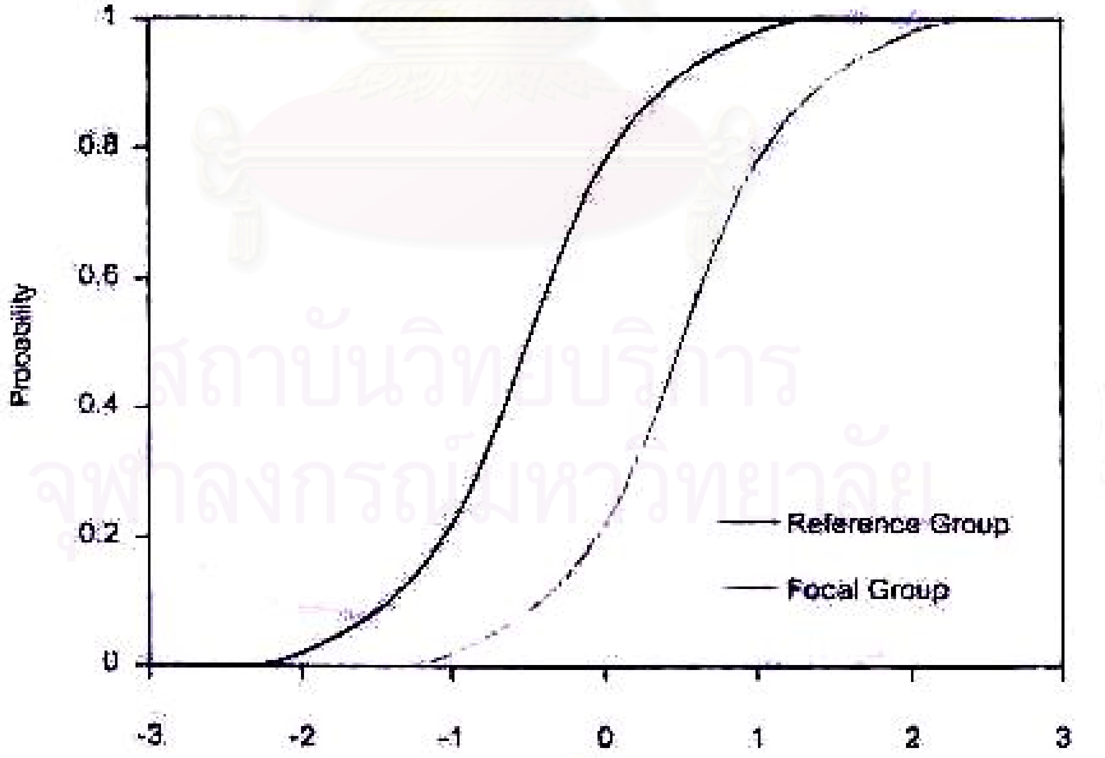
ตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) สามารถพิจารณา "ปฏิสัมพันธ์" ดังกล่าวได้จากความแตกต่างของค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ ระหว่างผู้สอบกลุ่มย่อยสองกลุ่ม กล่าวคือ ถ้าข้อสอบทำหน้าที่ต่างกันแบบเอกรูป แล้วโค้งลักษณะข้อสอบ (Item Characteristic Curve: ICCs) ระหว่างผู้สอบกลุ่มย่อยสองกลุ่มจะขนานกัน หรือมีฟังก์ชันการตอบสนองข้อสอบ (Item Response Functions: IRF) เหมือนกัน แต่ถ้าข้อสอบทำหน้าที่ต่างกันแบบอเนกรูปแล้วโค้งลักษณะข้อสอบระหว่างผู้สอบกลุ่มย่อยสองกลุ่มจะไม่ขนานกัน หรือมี

ฟังก์ชันการตอบสนองข้อสอบต่างกัน ดังนั้นความแตกต่างระหว่างโค้งคุณลักษณะข้อสอบทั้งสองแบบจะบ่งบอกถึงขนาดและทิศทางของข้อสอบที่ทำหน้าที่ต่างกัน ซึ่งสามารถคำนวณได้โดยใช้สูตรการคำนวณพื้นที่ของ Raju (1990)

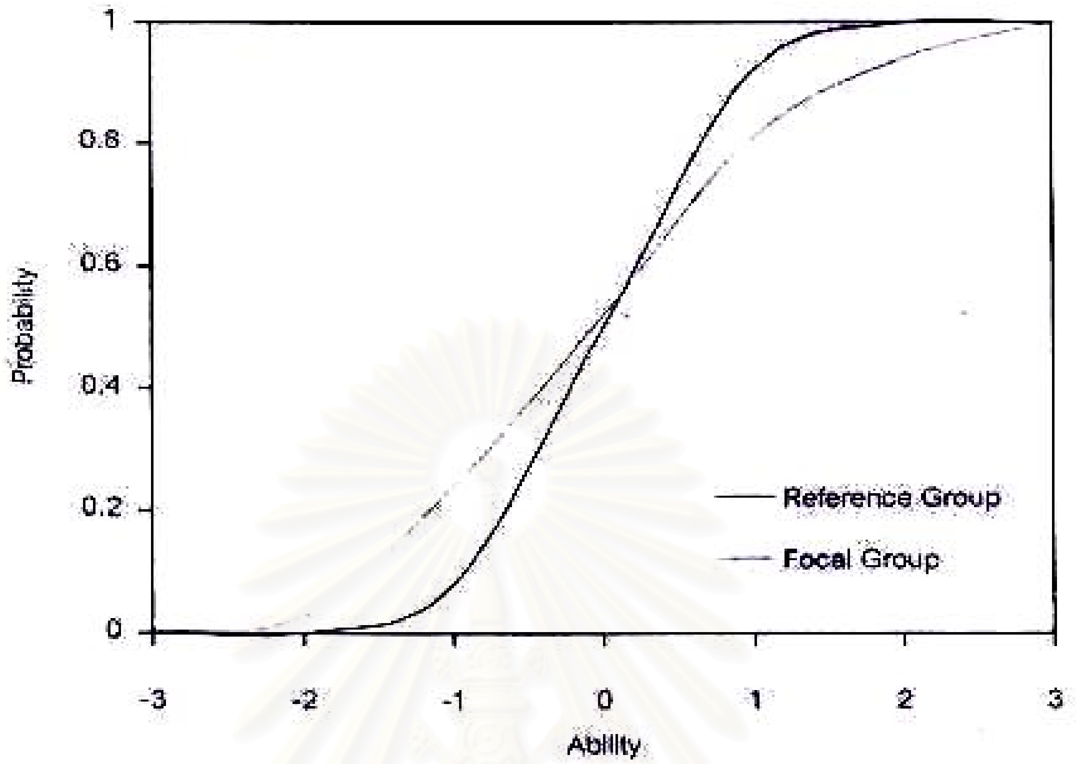
ข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป สามารถจำแนกได้เป็น 2 ลักษณะ (Swaminathan & Rogers, 1990) ดังนี้

1) ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูปโดยมีปฏิสัมพันธ์ไม่เป็นลำดับ (Disordinal Interaction) เป็นการทำหน้าที่ต่างกันสำหรับกลุ่มผู้สอบซึ่งเกิดขึ้น เมื่อโค้งลักษณะข้อสอบตัดกันระหว่างช่วงความสามารถของผู้สอบหรือเรียกว่าข้อสอบทำหน้าที่ต่างกันแบบไม่มีทิศทาง (Nondirectional DIF)

2) ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป โดยมีปฏิสัมพันธ์เป็นลำดับ (Ordinal Interaction) เป็นการทำหน้าที่ต่างกันสำหรับผู้สอบซึ่งเกิดขึ้นเมื่อโค้งลักษณะข้อสอบต่างกันอย่างไม่สม่ำเสมอ แต่ไม่ตัดกันหรืออาจตัดกันนอกช่วง ความสามารถของผู้สอบตรงปลายสุดของช่วงความสามารถต่ำหรือสูง อาจเรียกข้อสอบลักษณะนี้ว่า ข้อสอบทำหน้าที่ต่างกันแบบมีทิศทางเดียว (Unidirectional DIF) สำหรับข้อสอบที่ทำหน้าที่ต่างกันทั้งสองประเภทแสดงดังแผนภาพที่ 1 และ 2 ดังต่อไปนี้



ภาพที่ 1 ข้อสอบทำหน้าที่ต่างกันแบบมีทิศทางเดียว (Unidirectional DIF)



ภาพที่ 2 ข้อสอบทำหน้าที่ต่างกันแบบไม่มีทิศทาง (Nondirectional DIF)

โดยทั่ว ๆ ไป ในแบบสอบมาตรฐานมักจะมีข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูปมากกว่าข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป แต่ในข้อมูลจริงจะมีข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปได้มากกว่า

จะเห็นได้ว่าประเภทของการทำหน้าที่ต่างกันของข้อสอบแบ่งเป็น 2 ประเภท ได้แก่ การทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป (Uniform DIF) เกิดขึ้นเมื่อผู้สอบกลุ่มหนึ่งมีโอกาสในการตอบข้อสอบถูกมากกว่าผู้สอบอีกกลุ่มหนึ่งในทุกระดับความสามารถ และการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป (Nonuniform DIF) เกิดขึ้นเมื่อโอกาสในการตอบข้อสอบถูกของผู้สอบระหว่างกลุ่มย่อย 2 กลุ่มไม่สม่ำเสมอ ต่อไปจะเป็นสิ่งที่สำคัญมากในเรื่อง การทำหน้าที่ต่างกันของข้อสอบ นั่นคือ หลักการและวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ซึ่งจะเกี่ยวข้องกับวิธีการทางสถิติ

4. หลักการและวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

4.1 หลักการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF detection) เป็นการเปรียบเทียบผลการตอบข้อสอบเป็นรายข้อระหว่างกลุ่มผู้สอบอย่างน้อย 2 กลุ่ม ที่มีความสามารถหลัก (Primary Ability) ที่มุ่งวัดเท่ากัน แต่คาดว่าจะมีความได้เปรียบหรือเสียเปรียบกัน โดยกลุ่มหนึ่งถือเป็น กลุ่มอ้างอิง (Reference Group) ซึ่งคาดว่าน่าจะได้เปรียบในการตอบข้อสอบข้อนั้น หรือมีโอกาสตอบข้อสอบได้ถูกต้องมากกว่า ส่วนอีกกลุ่มคือ กลุ่มเปรียบเทียบ (Focal Group) ซึ่งเป็นกลุ่มที่สนใจศึกษา และคาดว่าจะเป็นกลุ่มที่เสียเปรียบ ถ้าข้อสอบทำหน้าที่ต่างกันแล้วโอกาสในการตอบข้อสอบถูกของผู้สอบแต่ละกลุ่มจะไม่เท่ากัน สำหรับเกณฑ์ที่ใช้ในการจำแนกผู้สอบเป็นกลุ่มเปรียบเทียบและกลุ่มอ้างอิงมีหลายลักษณะ เช่น เพศ สีผิว เชื้อชาติ ภาษา วัฒนธรรม ภูมิลำเนา เป็นต้น

ในการเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบจำเป็นต้องจับคู่ (Matching) ผู้สอบตามความสามารถ ซึ่งเป็นเงื่อนไขสำคัญของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เกณฑ์การจับคู่ (Matching Criteria) ที่นิยมใช้กันมี 2 วิธี ดังนี้

1) เกณฑ์ภายนอก (External Criterion)

การวิเคราะห์การทำหน้าที่ต่างกัน โดยใช้เกณฑ์ภายนอกนี้สามารถนำไปใช้ได้ทั้งข้อสอบรายข้อและแบบสอบทั้งฉบับ โดยการใช้คะแนนจากแบบสอบอื่นเป็นเกณฑ์ภายนอกแล้วใช้เทคนิคการวิเคราะห์การถดถอย (Regression Analysis) เพื่อทำการเปรียบเทียบเส้นกราฟความสัมพันธ์ระหว่างตัวแปรเกณฑ์กับตัวแปรทำนายระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ

หลักการมีจุดมุ่งหมาย เพื่อสร้างสมการทำนายตัวแปรเกณฑ์ ซึ่งเป็นคะแนนของแบบสอบอื่นจากตัวแปรทำนายซึ่งเป็นคะแนนรายข้อ หรือคะแนนแบบสอบระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบในการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบ จะใช้คะแนนรวมของแบบสอบทั้งฉบับเป็นตัวแปรทำนาย สำหรับตัวแปรเกณฑ์ภายนอก อาจใช้คะแนนรวมทั้งฉบับหรือเกรดเฉลี่ยหรือผลสัมฤทธิ์ในงานที่เกี่ยวข้องของผู้สอบ (Cronbach, 1970 อ้างถึงใน ศิริชัย กาญจนวาสี, 2545)

สมการทำนายสำหรับกลุ่มอ้างอิงและกลุ่มเปรียบเทียบแสดงได้ดังนี้

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

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

เมื่อ Y_i = คะแนนของตัวแปรเกณฑ์ภายนอก

X_i = คะแนนของตัวแปรทำนาย

A = ค่าคงที่หรือค่าตัดแกน (intercept)

B = ค่าความชัน (slope)

จากฟังก์ชันการทำนายทั้ง 2 สมการ สามารถเปรียบเทียบค่าตัดแกน (A) และค่าความชัน (B) ของเส้นกราฟระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบได้ ถ้าเส้นกราฟดังกล่าวมีค่าความชันหรือค่าตัดแกน แตกต่างกันสำหรับข้อสอบใด หรือแบบสอบใดแสดงว่าข้อสอบหรือแบบสอบนั้นมีการทำหน้าที่ต่างกัน โดยเข้าข้างกลุ่มผู้สอบที่มีค่าตัดแกนหรือค่าความชันที่สูงกว่า

การใช้เกณฑ์ภายนอกมีข้อดี คือเกณฑ์ที่ใช้มีอิสระจากข้อสอบ และแบบสอบที่ต้องการตรวจสอบ แต่มีจุดอ่อนตรงที่ความเหมาะสมของเกณฑ์ที่จะนำมาใช้ ในทางปฏิบัติเป็นทางยากที่จะหาตัวแปรเกณฑ์ภายนอกจากแบบสอบฉบับอื่นที่มีความตรงเชิงทำนาย และมีความยุติธรรมสำหรับกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ถ้าตัวแปรเกณฑ์ภายนอกขาดคุณสมบัติดังกล่าวจะทำให้ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบหรือแบบสอบขาดความแม่นยำ และความสมบูรณ์

2) เกณฑ์ภายใน (Internal Criterion)

การวิเคราะห์การทำหน้าที่ต่างกัน โดยใช้เกณฑ์ภายในเป็นการนำวิธีการทางสถิติมาตรวจสอบการทำหน้าที่ต่างกันของข้อสอบหรือแบบสอบ โดยเน้นการพิจารณาจากโครงสร้างภายในของแบบสอบเป็นหลัก ด้วยการวิเคราะห์ผลจากการตอบข้อสอบและความสามารถหรือคะแนนจริงของผู้สอบที่ได้จากแบบสอบฉบับนั้น เพื่อนำมาเปรียบเทียบระหว่างผู้สอบจากกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีความสามารถหรือคะแนนจริงเท่ากัน ว่าจะมีผลการตอบหรือโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกันหรือไม่ เพื่อบ่งชี้ถึงการทำหน้าที่ต่างกันของข้อสอบ

การวิเคราะห์ในลักษณะนี้นิยมใช้ค่าสถิติต่างๆเป็นตัวบ่งชี้ถึงการทำหน้าที่ต่างกันของข้อสอบ ค่าสถิติทดสอบที่นิยมนำมาใช้พอสรุปได้ดังนี้

2.1) การทดสอบปฏิสัมพันธ์ (Interaction)

ในระยะเริ่มแรกของการศึกษาความลำเอียงของข้อสอบ มีการใช้สถิติทดสอบเอฟ (F-test) จากการวิเคราะห์ความแปรปรวน (ANOVA) เพื่อทดสอบปฏิสัมพันธ์ระหว่างกลุ่มผู้สอบกับข้อสอบ ถ้าการทดสอบมีนัยสำคัญเป็นสัญญาณของการทำหน้าที่ต่างกันของข้อสอบ (Cleary & Hilton, 1968; Jensen, 1974) จากนั้นจึงทำการวิเคราะห์ต่อด้วยวิธีการ Post Hoc เพื่อระบุข้อสอบที่มีผลต่อการเกิดปฏิสัมพันธ์ ซึ่งเป็นข้อที่ทำหน้าที่ต่างกัน

วิธีการนี้มีข้อดีที่สามารถศึกษาผู้สอบหลาย ๆ กลุ่มได้สะดวก แต่มีจุดอ่อนในเรื่องการควบคุมกลุ่มต่าง ๆ ให้มีความสามารถที่ทัดเทียมกัน ขนาดกลุ่มตัวอย่างของกลุ่มต่าง ๆ และอัตราความคลาดเคลื่อนประเภทที่ 1 จะสูงขึ้น ถ้าจำนวนข้อสอบเพิ่มมากขึ้น

2.2) การวัดความเบี่ยงเบนสัมพัทธ์ (Relative Deviation)

การคำนวณค่าความยากของข้อสอบ เช่น p, b เป็นต้น เมื่อคำนวณแยกระหว่างกลุ่ม และแปลงให้เป็นค่าความยากมาตรฐาน (Δ) สามารถนำมาพล็อตเปรียบเทียบเป็นรายข้อ ถ้าข้อใดเบี่ยงเบนไปจากแกนหลักที่คาดหมาย หรือเบี่ยงเบนเกินจากความคลาดเคลื่อนมาตรฐานของค่าความยากที่กำหนด ย่อมแสดงถึงการทำหน้าที่ต่างกันของข้อสอบ (Cleary & Hilton, 1968; Angoff & Ford, 1973) รวมทั้งสามารถคำนวณค่าสหสัมพันธ์ระหว่างค่าความยากรายข้อระหว่างกลุ่ม เพื่อแสดงถึงการทำหน้าที่ต่างกันของข้อสอบ ถ้าสหสัมพันธ์เข้าใกล้ 1.00 แสดงว่าค่าความยากสัมพัทธ์ของข้อสอบมีค่าใกล้เคียงกันระหว่างกลุ่ม ดังนั้นแบบสอบวัดคุณลักษณะคล้ายกันระหว่างกลุ่ม

วิธีการนี้มีข้อดีและข้อเสียคล้ายการทดสอบปฏิสัมพันธ์ นอกจากนี้ค่าความยากของข้อสอบ (p) มิใช่ตัวแทนของค่าความยากจริงของข้อสอบ และได้รับอิทธิพลจากค่าแทรกซ้อนอื่น ได้แก่ ค่าอำนาจจำแนกและความสามารถของผู้สอบ

2.3) การเปรียบเทียบน้ำหนักองค์ประกอบ

การวิเคราะห์องค์ประกอบ (Factor Analysis) เป็นเทคนิคทางสถิติที่นิยมใช้ในการตรวจสอบความตรงเชิงทฤษฎีหรือโครงสร้าง (Construct Validity) เมื่อนำการวิเคราะห์ตัวประกอบมาใช้ในการวิเคราะห์โครงสร้างของแบบสอบแยกตามกลุ่มผู้สอบ ความไม่สอดคล้องกันระหว่างน้ำหนักตัวประกอบบนคุณลักษณะสำคัญที่มุ่งวัด หรือ ความแตกต่างของค่าเฉลี่ยคะแนนตัวประกอบ (Factor Scores) ระหว่างกลุ่มผู้สอบย่อมสะท้อนการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ

การใช้เทคนิคการวิเคราะห์องค์ประกอบเชิงสำรวจ (Exporatory Factor Analysis: EFA) สำหรับศึกษาการทำหน้าที่ต่างกัน จะมีจุดอ่อนในเรื่องความไม่สอดคล้องระหว่างน้ำหนักตัวประกอบอาจเกิดจากความแตกต่างของความสามารถระหว่างกลุ่มก็ได้ แนวทางที่เหมาะสมจึงควรใช้การวิเคราะห์ตัวประกอบเชิงยืนยัน (Confirmatory Factor Analysis: CFA) นอกจากนี้ยังสามารถใช้ CFA สำหรับตรวจสอบความแตกต่างระหว่างกลุ่มในด้านคุณลักษณะหรือความสามารถหลัก และความสามารถรองได้อีกด้วย (Camilli & Shepard, 1994)

2.4) การเปรียบเทียบโอกาสตอบข้อสอบถูก

การวิเคราะห์โอกาสตอบข้อสอบถูกของผู้สอบจากกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีความสามารถเท่ากัน เป็นแนวทางสำคัญที่นิยมใช้กันและเป็นที่ยอมรับในปัจจุบัน สำหรับบ่งชี้การทำหน้าที่ต่างกันของข้อสอบ มีการคำนวณค่าสถิติ 2 แนวทางดังนี้

เปรียบเทียบค่าสัดส่วนหรือความน่าจะเป็นในการตอบข้อสอบถูกของผู้สอบต่างกลุ่มที่มีความสามารถเท่ากัน เช่น วิธีแมนเทล-แฮนส์เซล (MH) เป็นต้น

เปรียบเทียบค่าฟังก์ชันการตอบสนองข้อสอบ หรือโค้งลักษณะข้อสอบระหว่างกลุ่มที่มีระดับความสามารถเท่ากัน เป็นวิธีที่อยู่บนพื้นฐานของทฤษฎี IRT เช่น วิธีวัดความแตกต่างของพื้นที่ วิธีวัดความแตกต่างของค่าพารามิเตอร์ความยาก วิธีการทดสอบไค-สแควร์ของลอร์ด (Lord's χ^2 - test) เป็นต้น

วิธีการนี้มีข้อดีที่สำคัญ ได้แก่ การคำนวณค่าสถิติของข้อสอบมีความน่าเชื่อถือ มีกลไกควบคุมความสามารถของผู้สอบโดยการจับคู่กลุ่มความสามารถ เพื่อทำการเปรียบเทียบ ณ ตำแหน่งต่าง ๆ ที่มีความสามารถเท่ากัน จึงมีวิธีการที่ยอมรับกันทั่วไป แต่มีข้อจำกัดในด้านความสลับซับซ้อนของแนวคิดพื้นฐาน และการวิเคราะห์มีความจำเป็นต้องใช้โปรแกรมคอมพิวเตอร์โดยเฉพาะ

จะพบว่า หลักการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจะเป็นการเปรียบเทียบผลการตอบข้อสอบระหว่างผู้สอบสองกลุ่มที่มีระดับความสามารถเดียวกัน โดยกำหนดให้ผู้สอบกลุ่มหนึ่งเป็นกลุ่มอ้างอิง และผู้สอบอีกกลุ่มหนึ่งเป็นกลุ่มเปรียบเทียบ ถ้าข้อสอบทำหน้าที่ต่างกันแล้วโอกาสในการตอบข้อสอบถูกของผู้สอบแต่ละกลุ่มจะไม่เท่ากัน ต่อไปจะให้ความสำคัญในเรื่องขั้นตอนทางสถิติ นั่นคือ วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

4.2 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF Detection) จำแนกตามลักษณะการตรวจให้คะแนนได้เป็น 2 ประเภท คือ ข้อสอบที่มีการให้คะแนนแบบทวิวิภาค หรือสองค่า (Dichotomous Scoring) และข้อสอบที่มีการให้คะแนนแบบพหุวิภาค หรือหลายค่า (Polytomous Scoring) วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแต่ละประเภท ยังสามารถจำแนกได้อีก 2 มิติ ได้แก่ มิติลักษณะของตัวแปรเกณฑ์ ซึ่งแบ่งเป็นกลุ่มวิธีที่ใช้คะแนนสังเกตได้ (Observed Score) และกลุ่มวิธีที่ใช้คะแนนสังเกตไม่ได้หรือคะแนนของตัวแปรแฝง (Latent Variable) และมิติลักษณะของสถิติวิเคราะห์ ซึ่งแบ่งเป็นกลุ่มวิธีที่ใช้สถิติพาราเมตริก (Parametric Approach) และกลุ่มวิธีที่ใช้สถิตินันพาราเมตริก (Nonparametric Approach) รายชื่อวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่สำคัญๆ มีดังนี้ (Potenza & Dorans, 1995; Feinstein, 1995)

1) วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนแบบทวิวิภาค

1.1) กลุ่มวิธีที่ใช้คะแนนที่สังเกตได้

วิธีในกลุ่มนี้มักวิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิม (CTT) หรือกลุ่มที่ไม่ใช้ทฤษฎีการตอบสนองข้อสอบ (Non-IRT Approach) โดยใช้คะแนนรวมของผู้สอบเป็นเกณฑ์การจับคู่ของกลุ่มผู้สอบ วิธีการตรวจสอบที่สำคัญในกลุ่มนี้ ได้แก่

1.1.1 การวิเคราะห์ความแปรปรวน (ANOVA) (Cleary & Hilton, 1968)

1.1.2 วิธีแปลงค่าความยากของข้อสอบ (Transformed Item Difficulty: TID) (Cleary & Hilton, 1968; Angoff & Ford, 1973)

1.1.3 วิธีตารางการณ์จร (Contingency Table: CT) วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ใช้พื้นฐานวิธีตารางการณ์จรที่สำคัญ มีดังนี้

- วิธีไค-สแควร์ (Chi-square: χ^2) (Scheuneman, 1979)
- วิธีลอก-ลิเนียร์ (Log-linear: LL)
- วิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel: MH) (Holland & Thayer, 1988, 1989)

1.1.4 วิธีดัชนีมาตรฐาน (Standardization: STND) การปรับให้เป็นมาตรฐานด้วยน้ำหนักตัวประกอบ (Dorans & Kulick, 1986)

1.1.5 วิธีการวิเคราะห์การถดถอยโลจิสติก (Logistic Regression: LR) (Swaminathan & Rogers, 1990)

1.2) กลุ่มวิธีที่ใช้คุณลักษณะแฝง

วิธีในกลุ่มนี้ใช้คุณลักษณะหรือตัวแปรแฝง ซึ่งวิเคราะห์บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ (IRT) สำหรับใช้เป็นเกณฑ์การจับคู่กลุ่มผู้สอบ ซึ่งแตกต่างจากวิธีการตรวจสอบในกลุ่ม non-IRT ในการวิเคราะห์โดยใช้ทฤษฎี IRT จะต้องเป็นไปตามข้อตกลงเบื้องต้นที่สำคัญ 3 ประการ ดังนี้ (Hambleton & Swaminathan, 1985; Lord, 1980 อ้างถึงใน วลีมาศ แซ่อึ้ง, 2543)

ประการแรก ความเป็นเอกมิติของแบบสอบ (Unidimensionality Test) หมายถึง แบบสอบจะต้องวัดความสามารถหรือคุณลักษณะเพียงลักษณะเดียว โดยทั่วไปแล้วเป็นการยากที่จะสร้างแบบสอบให้เป็นไปตามข้อตกลงเบื้องต้นนี้ ลักษณะที่เป็นไปได้ก็คือพยายามเลี่ยงไปใช้ข้อตกลงเบื้องต้นที่ผ่อนคลายลง โดยพิจารณามิติเด่น (Dominant Dimension) เพียงมิติเดียวก็น่าจะเพียงพอแล้ว นั่นคือ ทำให้แบบสอบสามารถวัดคุณลักษณะที่ต้องการวัดเป็นส่วนใหญ่ ถึงแม้ว่าอาจจะมีคุณลักษณะอื่นเข้ามาเกี่ยวข้องบ้างก็เป็นสิ่งที่หลีกเลี่ยงไม่ได้ ความเป็นเอกมิติในลักษณะนี้เรียกว่า เอกมิติสำคัญ (Essential Unidimension) สำหรับการตรวจสอบแบบสอบว่ามีความเป็นเอกมิติหรือไม่นั้นนิยมใช้วิธีวิเคราะห์องค์ประกอบ (Factor Analysis) โดยวิเคราะห์องค์ประกอบสำคัญ เพื่อพิจารณาความแปรปรวนสูงสุดจากองค์ประกอบหลักตัวแรกว่ามีสัดส่วนมากกว่าองค์ประกอบอื่น ๆ หรือไม่ ถ้ามีมากกว่าก็แสดงว่าแบบสอบนั้นมีแนวโน้มเป็นเอกมิติ

ประการที่สอง ความเป็นอิสระต่อกันในการตอบข้อสอบ (Local Independence) หมายถึง โอกาสของการตอบข้อสอบในแต่ละข้อได้ถูกต้องมีความเป็นอิสระทางสถิติ (Statistical

Independence) และมีความเป็นอิสระจากตำแหน่ง (Uncorrelated Independence) กล่าวคือ การตอบข้อสอบข้อใดข้อหนึ่งไม่ว่าถูกหรือผิดจะไม่มีผลต่อการตอบข้อสอบข้ออื่น ๆ ในแบบสอบ และลำดับในการตอบข้อสอบจะไม่มีผลต่อการตอบข้อสอบของผู้สอบ ดังนั้นข้อสอบจะปรากฏอยู่ในตำแหน่งใดของแบบสอบก็ได้ การตรวจสอบความเป็นอิสระต่อกันในการตอบข้อสอบสามารถทำได้โดยการทดสอบไค-สแควร์ของข้อสอบครั้งละคู่ที่ระดับความสามารถเดียวกัน นอกจากนี้ยังสามารถพิจารณาได้จากความเป็นเอกมิติของแบบสอบ กล่าวคือ ถ้าแบบสอบฉบับใดมีความเป็นเอกมิติแล้วแบบสอบฉบับนั้นจะมีความเป็นอิสระต่อกัน

ประการที่สาม ฟังก์ชันการตอบสนองข้อสอบ (Item Response Function: IRF) หรือโค้งลักษณะข้อสอบ (Item Characteristic Curve: ICC) เป็นโค้งแสดงความสัมพันธ์ระหว่างโอกาสของการตอบข้อสอบได้ถูกต้องกับระดับความสามารถของผู้สอบ โค้งลักษณะข้อสอบในยุคเริ่มต้นจะใช้ฟังก์ชันปกติสะสม (Normal Ogive Function) ซึ่งมีลักษณะเป็นโค้งตัว S ต่อมาได้เปลี่ยนมาใช้ฟังก์ชันโลจิสติก (Logistic Function) เนื่องจากมีสูตรคณิตศาสตร์ที่สามารถคำนวณได้ง่าย และให้ผลการประมาณค่าได้ใกล้เคียงกันมาก โค้งลักษณะข้อสอบมีหลายลักษณะ ทั้งนี้จะขึ้นอยู่กับโมเดลการตอบสนองข้อสอบ เช่น โมเดลปกติสะสม (Normal Ogive Model) โมเดลโลจิสติก (Logistic Model) โมเดลการตอบสนองแบบเกรด (Grade Response Model) โมเดลการให้คะแนนแบบบางส่วน (Partial Credit Model) โมเดลการตอบสนองแบบคะแนนต่อเนื่อง (Continuous Response Model) เป็นต้น วิธีการตรวจสอบที่สำคัญในกลุ่มนี้ได้แก่

1.2.1 วิธีวัดพื้นที่ความแตกต่างระหว่างโค้งการตอบสนองข้อสอบ (IRT-D^2) (Linn et al., 1981; Shepard et al., 1984; Raju, 1990; Kim & Cohen, 1991)

1.2.2 วิธีการเปรียบเทียบค่าพารามิเตอร์ของข้อสอบ

- วิธีเปลี่ยนค่าความยาก (Difficulty Shift)
- วิธีการทดสอบ F (Hulin, Drasgow, & Komocar, 1982)
- วิธีไค-สแควร์ของลอร์ด (Lord's χ^2) (Lord, 1980)
- วิธีการตอบสนองแบบเทียม (Pseudo-IRT) (Linn & Harnisch, 1981)
- วิธีการทดสอบอัตราส่วนไลค์ลิฮูด (Likelihood Ratio Test: LR) (Thissen, Steinberg, & Wainer, 1993)

1.2.3 วิธีซิปเทสท์ (SIBTEST) (Shealy & Stout, 1993)

2) วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนแบบพหุวิภาค

2.1) กลุ่มวิธีที่ใช้คะแนนที่สังเกตได้

2.1.1 การวิเคราะห์ความแปรปรวน (ANOVA) (Cleary & Hilton, 1968)

2.1.2 วิธีการวิเคราะห์การถดถอยโลจิสติกพหุวิภาค (Polytomous Logistic Regression) (Swaminathan & Rogers, 1990)

2.1.3 วิธีดัชนีมาตรฐานพหุวิภาค (Polytomous Standardization) (Dorans & Kulick, 1986)

2.1.4 วิธีแมนเทล-แฮนส์เซลทั่วไป (General Mantel-Haenszel: GMH) (Holland & Thayer, 1988, 1989)

2.2) กลุ่มวิธีที่ใช้คุณลักษณะแฝง

2.2.1 วิธีอัตราส่วนไลค์ลิฮูดในรูปทั่วไป (General IRT Likelihood Ratio) (Thissen, Steinberg, & Wainer, 1993)

2.2.2 วิธีซิปเทสท์พหุวิภาค (Polytomous SIBTEST) (Shealy & Stout, 1993)

2.2.3 วิธีการให้คะแนนบางส่วน (Partial Credit Model: PCM) (Master, 1982)

2.2.4 วิธีการให้คะแนนบางส่วนทั่วไป (Generalized Partial Credit Model: GPCM) (Muraki, 1992, 1993)

จะเห็นว่า การจัดประเภทของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สามารถจำแนกได้หลายลักษณะขึ้นอยู่กับเกณฑ์ที่ใช้จำแนก ได้แก่ หากแบ่งตามเกณฑ์การให้คะแนนของข้อสอบจะแบ่งออกได้เป็น 2 กลุ่มวิธี คือ กลุ่มวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบทวิวิภาค และกลุ่มวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบพหุวิภาค ถ้าใช้เกณฑ์ตามทฤษฎีของการวิเคราะห์ข้อมูลจะแบ่งออกเป็น 2 กลุ่มวิธีใหญ่ ๆ คือ กลุ่มวิธี non-IRT และกลุ่มวิธี IRT แต่ถ้าใช้ข้อตกลงเบื้องต้นของโมเดลเป็นเกณฑ์สามารถแบ่งออกเป็น 2 รูปแบบ คือ รูปแบบพาราเมทริก และรูปแบบนันพาราเมทริก ต่อไปจะให้ความสำคัญในเรื่องของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีทฤษฎีการตอบสนองข้อสอบ ซึ่งเป็นทฤษฎีที่ผู้วิจัยสนใจศึกษา

4.3 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีทฤษฎีการตอบสนองข้อสอบ (Item Response Theory Procedure: IRT)

ทฤษฎี IRT ถูกคิดค้นขึ้นมาเพื่อแก้ปัญหาเกี่ยวกับการแปรเปลี่ยนของค่าสถิติ ซึ่งเป็นจุดด้อยของทฤษฎีการวัดแบบดั้งเดิม (CTT) Ferguson (1942) & Lawleg (1943) เป็นผู้เสนอแนวคิดทฤษฎีการตอบสนองข้อสอบ มีหลักการว่า ผลการทดสอบของผู้สอบจากแบบสอบใด ๆ ขึ้นอยู่กับความสามารถของผู้สอบ และต่อมา Lord (1952) ได้เสนอทฤษฎีนี้ขึ้นใหม่ในรูปโค้งลักษณะข้อสอบ (Item Characteristic Curve: ICC) ต่อมาเรียกว่า Ogive Model ซึ่งจะกล่าวถึงพารามิเตอร์ 2 ตัว

คือ ค่าความยากและค่าอำนาจจำแนก แต่ในโมเดลดังกล่าวมีความยุ่งยากในการคำนวณและขาดแคลนโปรแกรมคอมพิวเตอร์ในการวิเคราะห์ข้อมูล จึงทำให้ Lord หยุดความสนใจในทฤษฎีนี้ไประยะหนึ่ง ในปี ค.ศ.1960 Rasch ได้ศึกษาเกี่ยวกับทฤษฎีดังกล่าวและได้เสนอแนวคิดในรูปพารามิเตอร์ตัวเดียว คือ ค่าความยาก ซึ่งบางครั้งเรียกแบบจำลองนี้ว่า Rasch Model ปี ค.ศ.1968 Bimbaum ได้เสนอแนวคิดใหม่เกี่ยวกับ Logistic Model ที่ใช้พารามิเตอร์ 2 ตัวคือ ค่าความยากและค่าอำนาจจำแนก ซึ่งเป็นแบบจำลองที่ง่ายกว่าของ Lord จึงทำให้ Logistic Model เป็นที่นิยมแพร่หลายและมีการพัฒนาขึ้นเรื่อย ๆ จนกระทั่งใช้ได้กับพารามิเตอร์ตัวเดียวและพารามิเตอร์ 3 ตัว (Warm, 1979 อ้างถึงใน สุรศักดิ์ อมรรัตนศักดิ์, 2531)

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วย IRT โมเดล ประกอบด้วย 2 ขั้นตอนที่สัมพันธ์กัน ได้แก่ ขั้นตอนแรก ทำการวัดขนาดของการทำหน้าที่ต่างกันของข้อสอบ (Measurement of DIF) และขั้นตอนที่สอง ทำการทดสอบทางสถิติของการทำหน้าที่ต่างกันของข้อสอบ (Statistical test of DIF) คำถามหลักของการตรวจสอบ DIF ด้วย IRT โมเดล คือ “โค้งลักษณะข้อสอบของประชากรต่างกลุ่มมีความแตกต่างกันหรือไม่” จึงต้องวัดขนาดของความแตกต่างและทดสอบทางสถิติว่ามีความแตกต่างอย่างมีนัยสำคัญระหว่างกลุ่มหรือไม่ (ศิริชัย กาญจนวาสี, 2545)

วิธีการตอบสนองข้อสอบเป็นวิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการพิจารณาเปรียบเทียบฟังก์ชั่นการตอบข้อสอบ ระหว่างกลุ่มผู้สอบที่มีต่อแบบสอบชุดเดียวกัน (Lord, 1980 อ้างถึงใน Green, 1994) หรือโดยการเปรียบเทียบความแตกต่างระหว่างโค้งคุณลักษณะข้อสอบ (ICC) ระหว่างกลุ่ม ซึ่งพื้นที่ระหว่างโค้งคุณลักษณะข้อสอบจะเป็นดัชนีบอกระดับของการทำหน้าที่ต่างกันของข้อสอบ

ในการวัดพื้นที่ระหว่างโค้งคุณลักษณะข้อสอบจะมี 2 ลักษณะคือ การวัดช่วงเปิด (Open Interval) จะเป็นการวัดในช่วงความสามารถทั้งหมดระหว่างโค้งทั้งสอง ซึ่งจะทำให้ได้พื้นที่ที่แน่นอนและการวัดช่วงปิด (Closed Interval) จะวัดในช่วงความสามารถตามที่กำหนดไว้ ซึ่งมีดัชนีที่ใช้บอกระดับการทำหน้าที่ต่างกันของข้อสอบได้แก่ พื้นที่ชนิดไม่มีเครื่องหมาย (Unsigned Areas) เป็นค่าสัมบูรณ์ของพื้นที่ระหว่างโค้งคุณลักษณะข้อสอบ พื้นที่ชนิดมีเครื่องหมาย (Signed Areas) เหมือนกับพื้นที่ชนิดไม่มีเครื่องหมายแต่จะมีเครื่องหมายแสดงให้ทราบว่ากลุ่มใดได้ประโยชน์ กลุ่มใดเสียประโยชน์ การทดสอบด้วยค่าสถิติ Z และไคสแควร์ (c^2) เป็นการทดสอบนัยสำคัญของความแตกต่างของค่าพารามิเตอร์ a และ b ระหว่างกลุ่ม ในคราวเดียวกัน ซึ่งเป็นวิธีของ Lord (Shepard et al., 1994)

5. แนวคิดเกี่ยวกับการวิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาค (Polytomous IRT Model)

คำว่า polytomous มาจากภาษากรีกว่า polus (many) + tomus (a cut) หมายถึงการแบ่งออกเป็นหลาย ๆ ส่วน คำทั้งสองมีการนำมาใช้กับการตรวจให้คะแนนในการวัดผลทางการศึกษาและเรียกแตกต่างกันไป เช่น เรียกการตรวจให้คะแนนแบบ dichotomous ว่า “การตรวจให้คะแนนแบบดั้งเดิม”และ “การตรวจให้คะแนนแบบทวิวิภาค” ส่วนการตรวจให้คะแนนแบบ polytomous มีผู้เรียกว่า “การตรวจให้คะแนนแบบหลายค่า” และ “การตรวจให้คะแนนแบบพหุวิภาค” ในการวิจัยนี้จะเรียกการตรวจให้คะแนนแบบแรกและแบบที่สองว่า “การตรวจให้คะแนนแบบทวิวิภาค” และ “การตรวจให้คะแนนแบบพหุวิภาค” ตามลำดับ

สำหรับการตรวจให้คะแนนแบบพหุวิภาค (polytomous) เป็นวิธีที่นิยมใช้กันมาก วิธีนี้จะกำหนดคะแนนให้กับคำตอบของนักเรียนในแต่ละข้อไว้มากกว่า 2 ค่ารวมถึงการให้คะแนนในมาตรประมาณค่าที่ใช้วัดคุณลักษณะชนิดต่าง ๆ ด้วย ซึ่งในมาตรประมาณค่าก็จะกำหนดให้คะแนนตามลำดับขั้นของคะแนนในเครื่องมือนั้น ๆ โดยถ้าเป็นแบบวัดคุณลักษณะคะแนนแต่ละค่ามักแสดงค่าถึงระดับของคุณลักษณะ แต่ถ้าเป็นแบบสอบผลสัมฤทธิ์ทางการให้คะแนนแต่ละค่าแสดงถึงระดับความสามารถของผู้สอบที่ตอบข้อกระทงนั้น อาจมีวิธีให้คะแนนแตกต่างกันไป เช่น การให้คะแนนตามระดับความมั่นใจในการตอบ การให้คะแนนตามความสามารถในการตัดตัวลวงของข้อกระทงนั้น และการให้คะแนนความรู้บางส่วน (Partial Knowledge) เป็นต้น การให้คะแนนผู้มีความรู้บางส่วน เป็นวิธีที่มีผู้สนใจศึกษากันมาก (Smith, 1987, 1982; Frary, 1982; Hutchinson,1984, Wright & Masters,1982 อ้างถึงใน ธนวัฒน์ แสนสุข, 2538)

การวิเคราะห์ข้อสอบมีจุดประสงค์สำคัญ 2 ประการ คือ การวิเคราะห์เพื่อหาคุณภาพรายข้อแล้วนำผลที่ได้ไปปรับปรุงข้อสอบ และวิเคราะห์เพื่อหาคุณภาพแบบสอบทั้งฉบับด้านความตรงและความเที่ยง การวิเคราะห์ดังกล่าวแบ่งได้เป็น 2 แนวทาง คือ การวิเคราะห์ตามทฤษฎีการทดสอบดั้งเดิม (Classical Test Theory: CTT) วิธีนี้มีข้อจำกัดที่ค่าสถิติที่ได้จะแปรเปลี่ยนไปตามกลุ่มตัวอย่าง อีกแนวทางหนึ่ง คือ การวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) การวิเคราะห์ตามแนวทางหลังนี้ให้ค่าพารามิเตอร์ที่สามารถอ้างอิงได้กับกลุ่มตัวอย่างทั่วไปจึงมีการนำมาใช้กันมากในปัจจุบัน IRT Model ดังกล่าว จำแนกได้เป็น 2 ประเภทตามลักษณะการให้คะแนน คือ (1) Dichotomous IRT Model ซึ่งใช้ข้อสอบที่ตรวจให้คะแนนแบบทวิวิภาค และ (2) Polytomous IRT Model ซึ่งใช้กับข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาค การวิเคราะห์ตามโมเดลแรกมีการนำมาใช้และรู้จักกันดี ส่วนการวิเคราะห์ตามแนวทางที่สองนั้น มีการพัฒนามากว่า 25 ปี (ธนวัฒน์ แสนสุข, 2538 อ้างถึงใน อารี มากมณี, 2541) โดยกำลังเป็นที่รู้จักในการ

นำมาใช้ในวงการวัดผลทางการศึกษา โมเดลในการวิเคราะห์ตามแนวทางดังกล่าวมีด้วยกันหลายโมเดล เช่น Graded Response Model (GRM), Rating Scale Model (RSM) และ Generalized Partial Credit Model (GPCM) เป็นต้น โมเดลที่รู้จักกันดีในปัจจุบัน คือ GRM, RSM, GPCM เป็นต้น

ในขณะที่มีการพัฒนาโมเดลก็มีการพัฒนาโปรแกรมคอมพิวเตอร์ เพื่อใช้วิเคราะห์ตามโปรแกรมดังกล่าวมากขึ้น ตัวอย่างโปรแกรมคอมพิวเตอร์ที่มีการพัฒนาขึ้นและเป็นที่รู้จักกันดี ได้แก่ MULTILOG, BIGSTEPS, PARSCALE เป็นต้น

โมเดล IRT ที่พัฒนาในระยะแรกจะใช้กับแบบสอบที่ให้คะแนนแบบทวิวิภาค (dichotomous) เป็นส่วนใหญ่ ต่อมาได้มีผู้พัฒนาโมเดล IRT เพื่อใช้กับแบบสอบและแบบวัดเจตคติที่ตรวจให้คะแนนแบบพหุวิภาค (polytomous) หรือลักษณะมาตรประมาณค่า (rating data) โมเดลในแนวทฤษฎีนี้ทั้งหมดเรียกว่า polytomous Item Response Models (Muraki,1992, 1993; Donoghue,1994; Hambleton & Zaal, 1991) มีโมเดลที่สำคัญ คือ

1. Graded Response Model (GRM) พัฒนาโดย Samejima ในปี ค.ศ. 1960
2. Nominal Response Model (NRM) พัฒนาโดย Bock ในปี ค.ศ. 1972
3. Continuous Model (CM) พัฒนาโดย Samejima ในปี ค.ศ. 1973
4. Rating Scale Model (RSM) พัฒนาโดย Andrich ปี ค.ศ.1978
5. Partial Credit Model (PCM) พัฒนาโดย Masters ปี ค.ศ.1982
6. Sucessive Interval Model (SIM) พัฒนาโดย Rost ในปี ค.ศ.1988
7. Generalized Partial Credit Model (GPCM) พัฒนาโดย Muraki ในปี ค.ศ.1992

GPCM (Generalized Partial Credit Model)

GPCM (Generalized Partial Credit Model) พัฒนาจาก PCM ให้มีการตรวจให้คะแนนความรู้บางส่วนในการแก้โจทย์ปัญหาคณิตศาสตร์ โดยพิจารณาให้คะแนนตามละดับขั้นความสำเร็จของการแก้โจทย์ปัญหาเป็นคะแนนเรียงลำดับ 0, 1, 2, 3 (Masters, 1982) จากนั้นก็มีการนำไปใช้กับแบบวัดเจตคติที่มีคะแนนเรียงลำดับหลายค่า (Dodd & Koch, 1989 อ้างถึงใน ธนวัฒน์ แสนสุข, 2538)

Muraki ได้พัฒนาโมเดลดังกล่าวมาเป็น GPCM โดยแก้ไขข้อจำกัดเกี่ยวกับความเป็นเอกมิติของเครื่องมือใช้วัดคุณลักษณะ และใช้ค่าอำนาจจำแนกที่แปรเปลี่ยนไปในแต่ละข้อมาร่วมประมาณค่าพารามิเตอร์ด้วย โดยมีข้อตกลงว่าบุคคลที่มีความสามารถหรือมีคุณลักษณะที่ต้องการวัดสูงมีความน่าจะเป็นที่จะตอบคำตอบลำดับคะแนนที่ K มากกว่า K-1 สามารถแทนด้วยสมการทั่วไปได้ดังนี้

$$C_{jk}(q) = P_{jk/k-1}(q) = \frac{P_{jk}(q)}{P_{j.k-1}(q) + P_{jk}(q)} = \frac{\exp^{[a_j(q)-b_{jk}]}}{1+\exp^{[a_j(q-b_{jk})]}}$$

เมื่อ k คือ ลำดับขั้นคะแนนที่ 2, 3,…, mj

$$P_{jk}(q) = \frac{C_{jk}}{1-C_{jk}} P_{jk-1}(q)$$

$P_{jk}(q)$ คือ ความน่าจะเป็นของคนที่มีความสามารถจะตอบได้คะแนน k (กรณี k-1 ก็เช่นกัน)

$C_{jk}/(1-C_{jk})$ คือ อัตราส่วนความน่าจะเป็นของสองเงื่อนไข ซึ่งอาจแสดงในรูปสมการ $\exp^{[a_j(q-b_{jk})]}$

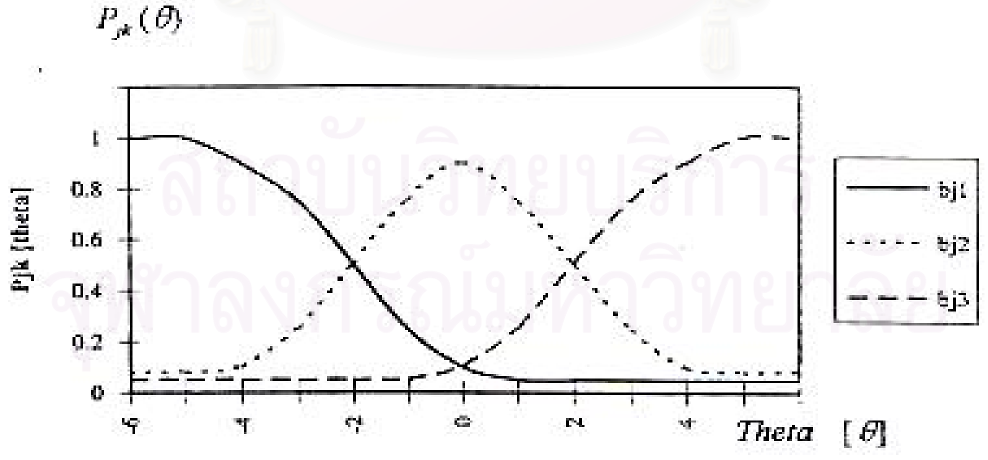
Masters เรียก b_{jk} ว่าเป็นค่าพารามิเตอร์ความยากในแต่ละลำดับขั้น เป็นจุดบนแกน θ ที่ฟังก์ชันของ $P_{j.k-1}(q)$ และ $P_{jk}(q)$ ตัดกัน ซึ่งโค้งทั้ง 2 เป็นฟังก์ชันของการตอบ (Item Category Response Function: ICRFs) ใน 2 ลำดับค่าคะแนนที่ตัดกันเพียงจุดเดียวบนค่า θ ใดๆ คือ

ถ้า $\theta = b_{jk}$ $P_{jk}(q) = P_{j.k-1}(q)$;

ถ้า $\theta > b_{jk}$ $P_{jk}(q) > P_{j.k-1}(q)$;

และถ้า $\theta < b_{jk}$ $P_{jk}(q) < P_{j.k-1}(q)$;

ซึ่งอยู่บนข้อตกลง $a_j > 0$ และ b_{jk} ไม่จำเป็นต้องเรียงกันในข้อสอบ j เพราะค่าพารามิเตอร์จะแสดงถึงขนาดของความน่าจะเป็นในการตอบ $P_{j.k-1}(q)$ และ $P_{jk}(q)$ ในลำดับที่อยู่ชิดกันดังแผนภาพที่ 3



ภาพที่ 3 โค้งฟังก์ชันสารสนเทศของ GPCM

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

การพิจารณาจุดตัดของโค้งฟังก์ชันการตอบแต่ละลำดับชั้น (Item Category Response Functions: ICRFs) ของโมเดล PCM แม้จะง่ายต่อการตีความหมายก็ตาม แต่จุดยอดของโค้งก็ไม่ใช่จะอยู่ตรงกลางของช่วงค่าคะแนนเสมอ ซึ่งที่มาของค่าฟังก์ชันการตอบสนองข้อกระทงแต่ละลำดับขั้น $P_{jk}(q)$ แสดงได้ดังนี้ (Muraki, 1992)

$$\frac{\partial}{\partial(q)} P_{jk}(q) = a_j P_{jk}(q)\left[k - \sum_{k=1}^{mj} k P_{jk}(q)\right]$$

โดยกำหนดสมการเป็น 0; $\sum_{k=1}^{mj} c P_j(q) = k$ สมการนี้จะแสดงโค้ง ICRFs ของ $P_{jk}(\theta)$
เมื่อ k = 1, 2, 3,…,k-1, k , k+1,…, mj

พารามิเตอร์ a_j คือ พารามิเตอร์อำนาจจำแนกของข้อ j โดยค่า a_j มีพิสัยที่เป็นไปได้อยู่ในช่วง -∞ ถึง + ∞ ซึ่งไม่เหมือนใน Dichotomous Model เพราะใน Polytomous Item Response Model อำนาจจำแนกในแต่ละลำดับขั้นคะแนนจะขึ้นอยู่กับผลรวมของพารามิเตอร์อำนาจจำแนกและค่าพารามิเตอร์เทรสโฮลด์ แต่ใน GPCM จะรวมค่าอำนาจจำแนกไว้ในโมเดลด้วย ซึ่งค่าอำนาจจำแนกนี้จะบอกถึงระดับการตอบในแต่ละลำดับค่าคะแนนที่เปลี่ยนไปในแต่ละข้อเช่นเดียวกับที่ค่า θ เปลี่ยนไปด้วย (Muraki , 1992)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

6. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย กับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก

แนวคิดของวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย (Mean and Covariance Structures Method: MACS)

สถิติวิเคราะห์ความแตกต่างของค่าเฉลี่ยและความแปรปรวน

โมเดลวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ยได้นำมาใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้งแบบเอกรูปและอเนกรูป และเป็นโมเดลที่เหมาะกับตัวแปรต่อเนื่อง (Continuous Variables) โดยทดสอบสมมุติฐานศูนย์ที่ว่า ค่าพารามิเตอร์น้ำหนักองค์ประกอบ (Factor Loading) ไม่แปรเปลี่ยนระหว่างกลุ่ม (ตัวอย่างเช่น $l_j^{(1)} = l_j^{(2)} = ... = l_j^{G}$)

ขั้นตอนของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย คือการวิเคราะห์ค่าความสอดคล้องของโมเดล ด้วยดัชนี GFI, AGFI, RMR, RMSEA และค่าไค-สแควร์ (c^2)

เมื่อพิจารณาความสอดคล้องของโมเดลด้วยค่าดังกล่าวแล้ว หากพบว่ามีความสอดคล้องกับข้อมูลแสดงว่าข้อสอบข้อนั้นไม่มีการทำหน้าที่ต่างกัน ในทางตรงกันข้ามหากพบว่าไม่มีความสอดคล้องกับข้อมูล แสดงว่าเกิดข้อสอบที่ทำหน้าที่ต่างกัน ในขั้นตอนต่อไป คือการเปรียบเทียบเมทริซ์พารามิเตอร์ LX (ตัวแปรสังเกตได้ภายนอก: ผลการตอบข้อสอบรายข้อ-ตัวแปรแฝงภายนอก: ค่าความสามารถ) (Joreskog & Sorbom, 1999)

จุดมุ่งหมายที่สำคัญของการวิเคราะห์ คือการตรวจสอบว่าโมเดลลิสเรลซึ่งเป็นกรอบความคิด (Conceptual Framework) ที่นักวิจัยสร้างขึ้นจากทฤษฎีนั้น สอดคล้องกับข้อมูลเชิงประจักษ์ของแต่ละกลุ่มหรือไม่อย่างไร ถ้ามีความสอดคล้องกลมกลืนกัน นักวิจัยจะสามารถนำโมเดลลิสเรลที่ตรวจสอบแล้วมาใช้อธิบายความสัมพันธ์ระหว่างตัวแปร และนำข้อค้นพบไปใช้ประโยชน์ได้ต่อไป หัวใจสำคัญของการวิเคราะห์ คือการวิเคราะห์ข้อมูลที่รวบรวมจากกลุ่มตัวอย่างทุกกลุ่ม โดยมีการกำหนดเงื่อนไขบังคับ (Constraints) ให้โมเดลลิสเรลที่เป็นกรอบความคิดในการวิจัยนั้น มีลักษณะเป็นแบบเดียวกันสำหรับการตรวจสอบความสอดคล้องกลมกลืนระหว่างโมเดลและข้อมูลเชิงประจักษ์ ถ้าผลการวิเคราะห์ให้ผลว่าค่า ไค-สแควร์ ในการทดสอบความกลมกลืนมีค่าต่ำกว่าค่าวิกฤติอย่างไม่มีนัยสำคัญทางสถิติ จะสรุปได้ว่า โมเดลลิสเรลที่สร้างขึ้นตามทฤษฎีสอดคล้องกับข้อมูลเชิงประจักษ์ทุกกลุ่ม และโมเดลมีลักษณะเป็นแบบเดียวกัน เรียกตามศัพท์สถิติวิเคราะห์ลิสเรลว่า โมเดลลิสเรลไม่แปรเปลี่ยน หรือมีความยืนยงระหว่างกลุ่ม (Invariance across groups) (นงลักษณ์ วิรัชชัย, 2542)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

การทดสอบความไม่แปรเปลี่ยนเริ่มต้นมาจากการทดสอบความเท่ากันของโครงสร้างความแปรปรวนร่วมของแต่ละกลุ่ม โดยมีการทดสอบสมมติฐานศูนย์ว่า (H_0) $E_1 = E_2 = ….. E_G$ เมื่อ E คือเมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมของประชากร และ G คือจำนวนกลุ่ม หากปฏิเสธสมมติฐานศูนย์แสดงว่ามีความเท่าเทียมกันระหว่างกลุ่ม แต่หากยอมรับสมมติฐานศูนย์แสดงว่าแต่ละกลุ่มมีความเท่าเทียมกัน

การทดสอบความไม่แปรเปลี่ยนของโมเดลลิสเรลระหว่างกลุ่ม เป็นการทดสอบสมมติฐานทางสถิติว่า โมเดลลิสเรลมีเมทริกซ์พารามิเตอร์ไม่แปรเปลี่ยนระหว่างกลุ่มประชากร โดยพิจารณาจากค่าสถิติทดสอบความกลมกลืน คือค่าไค-สแควร์ ซึ่งเป็นค่าสถิติทดสอบในภาพรวมทุกกลุ่ม ถ้าค่าไค-สแควร์รวมมีค่าต่ำกว่าค่าวิกฤติอย่างไม่มีนัยสำคัญทางสถิติ แสดงว่าโมเดลลิสเรลในภาพรวมของกลุ่มประชากรทุกกลุ่มหรือกลุ่มพหุสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ในกรณีตรงข้ามถ้าค่าไค-สแควร์รวมมีค่าสูงกว่าค่าวิกฤติอย่างมีนัยสำคัญทางสถิติ แสดงว่าโมเดลลิสเรลในภาพรวมของกลุ่มประชากรทุกกลุ่มหรือกลุ่มพหุไม่สอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ นอกจากการพิจารณาตีความหมายจากค่าไค-สแควร์รวมแล้ว นักวิจัยยังต้องพิจารณาตีความจากค่าสถิติที่เป็นดัชนีวัดความกลมกลืนของโมเดลด้วย ค่าดัชนีที่ควรนำมาพิจารณาร่วมด้วย ได้แก่ ดัชนี GFI ดัชนี RMSE ดัชนี RMR และดัชนี critical N เป็นต้น

เนื่องจากเมทริกซ์พารามิเตอร์ในโมเดลมีทั้งหมด 8 เมทริกซ์ การกำหนดสมมติฐานทางสถิติในการทดสอบความไม่แปรเปลี่ยนของโมเดลลิสเรลระหว่างกลุ่ม จึงสามารถตั้งสมมติฐานทางสถิติในการทดสอบได้หลายสมมติฐานตามจำนวนเมทริกซ์พารามิเตอร์ โดยตั้งสมมติฐานที่มีเงื่อนไขบังคับน้อยที่สุดไปจนถึงสมมติฐานที่มีเงื่อนไขบังคับมากที่สุด (Raju et al, 2002)

การนำเข้าข้อมูลสำหรับการวิเคราะห์

ในการนำเข้าข้อมูลสำหรับแต่ละกลุ่มนั้น ก็เหมือนกับการนำเข้าข้อมูลกลุ่มเดียว แต่ต้องเพิ่มคำสั่งดังนี้

- NG ต้องนิยามในบรรทัด DA สำหรับกลุ่มแรก
- สำหรับแต่ละกลุ่ม g กลุ่ม g = 2, 3,….,G
- ในรูปแบบเมทริกซ์แต่ละเมทริกซ์และค่าของพารามิเตอร์กำหนดที่ไม่เป็นศูนย์ สมาชิกในเมทริกซ์จะนำเสนอสมาชิกในกลุ่มปัจจุบัน โดยระบุเป็นพารามิเตอร์อิสระทั้งหมด

- การนิยามค่าคงที่ที่เท่าเทียมกันระหว่างกลุ่ม มีลักษณะเฉพาะอันหนึ่งที่สมาชิกคงที่เป็นอิสระสำหรับกลุ่มเพราะจะเป็นค่าคงที่เท่าเทียมในแต่ละกลุ่มอื่นๆ
- ในหลักการ NY, NX, NE และ NK ต้องเหมือนกันในทุก ๆ กลุ่ม อย่างไรก็ตามถ้าจำนวนของตัวแปรมีความแตกต่างกันในแต่ละกลุ่มเป็นไปได้ว่าจะเป็นตัวแปรเทียม (ตัวแปรสังเกตหรือแฝง) เป็นการทำจำนวนตัวแปรให้เท่าเทียมกันในแต่ละกลุ่ม (Du Toit, 2001)

แนวคิดของวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก (Logistic Discriminant Function Analysis: LDFA)

ในทฤษฎีการตอบสนองข้อสอบ ข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปจะแสดงให้เห็นโดยโค้งลักษณะข้อสอบจะไม่ขนานกัน การถดถอยโลจิสติกมาตรฐานเป็นวิธีที่มีประสิทธิภาพในการตรวจสอบทั้งข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูปและแบบอเนกรูปที่มีการตรวจให้คะแนนแบบ 2 ค่า (Roger & Swaminathan, 1993; Swaminathan & Rogers, 1990 อ้างถึงใน Su & Wang, 2005) สำหรับข้อสอบที่ตรวจให้คะแนนแบบหลายค่าจะมีฟังก์ชันการถดถอยแบบโลจิสติกเท่ากับ T-1 เมื่อ T คือจำนวนลำดับขั้นการตอบสำหรับข้อสอบที่ศึกษา โครงสร้างโลจิทประกอบไปด้วย 3 โครงสร้างหลัก ๆ ได้แก่ โลจิทสะสม โลจิทอัตราส่วนต่อเนื่อง และโลจิทลำดับขั้นที่ติดกัน การใช้วิธีการถดถอยโลจิสติกพหุนาม (MLR) ในการตอบสนองแบบพหุวิภาคต้องใช้ข้อมูลระดับนามบัญญัติในการตรวจสอบ DIF ของข้อสอบที่ตรวจให้คะแนนหลายค่า และต้องใช้กลุ่มตัวอย่างขนาดใหญ่ รวมถึงการแปลความหมายยาก เพราะหลาย ๆ พารามิเตอร์มีการทดสอบความมีนัยสำคัญทางสถิติ (French & Miller, 1996) ในทางตรงกันข้ามมีเพียงฟังก์ชันถดถอยโลจิสติก 1 อันที่จำเป็นต้องนำมาใช้กับวิธี LDFA คะแนนแต่ละข้อ (U) และตัวระบุกลุ่ม (G) ดังนั้นโมเดลจะกล่าวถึงความน่าจะเป็นของตัวระบุกลุ่ม (G) เงื่อนไขคะแนนรวม (X) ตัวเลือกตอบสนองที่สังเกตได้ (U = 0, 1, ….. , T-1) และปฏิสัมพันธ์ระหว่าง X และ U ฟังก์ชันแสดงดังต่อไปนี้

$$_Logit \equiv In\frac{[P(G \approx 1)]}{P(G \approx 0)} = a_0 + a_1x + a_2u + a_3xu$$ สมการที่ 1

อย่างไรก็ตาม ตัวแปรตอบสนอง (U) ไม่จำเป็นต้องถูกกำหนดให้อยู่ในเฉพาะ 2 กลุ่ม แต่สามารถใช้ค่าจำนวนตัวเลือกตอบสนอง (J) ไปสัมพันธ์กับแต่ละตัวเลือกได้ มีบางข้อมูลนำการสุ่มตัวอย่างไปใช้ในสมการที่ 1 เมื่อนำไปใช้ในการอ้างอิงในสมการที่ 1 เรียกว่า a Cohort study ข้อมูลจะถูกประมวลผลออกมาในรูปการวิเคราะห์ข้อมูลจากกรณีควบคุมผ่านทางการถดถอยโลจิสติก ซึ่ง

ค่าที่ได้จากการถดถอยโลจิสติก จะออกมามีทิศทางในทางเดียวกับการใช้คอมพิวเตอร์ ซึ่งผ่านทาง cohort study การประเมินค่าตัวแปรทางสถิติจะประกอบไปด้วยค่าการประมาณค่าไลค์ลิฮูด การทดสอบการมีนัยสำคัญและการทดสอบความเชื่อมั่น จะใช้วิธีพื้นฐานของการทดสอบอัตราส่วนไลค์ลิฮูดซึ่งจะทดสอบความมีนัยสำคัญซึ่งค่าที่ได้จะประมาณค่า 2 ค่า

LDFA เป็นวิธีที่เหมาะสมในการใช้วิเคราะห์ความคลาดเคลื่อนของข้อมูลได้อย่างมีประสิทธิภาพ เช่น การอธิบายความแตกต่างโดยทั่วไปของข้อมูล ซึ่งมีความแตกต่างที่ชัดเจน สูตรจะสามารถประมวลผลให้ค่าที่ใกล้เคียง ส่วนสำคัญในการวิเคราะห์ LDFA คือการ DIF แต่ละตัวออกมาและร่วมกับการใช้การตอบสนองแบบพหุวิภาค อย่างไรก็ตาม วิธี LDFA จะแสดงผลได้ดีก็ต่อเมื่อแบบสอบประกอบไปด้วยข้อสอบที่มี Dif หลายข้อและไม่ทราบค่า หากมี Dif ปริมาณสูง ๆ ไม่จำเป็นที่จะส่งผลให้ค่าอัตราความคลาดเคลื่อนประเภทที่ 1 สูงขึ้นสำหรับวิธีที่มีพื้นฐานมาจากทฤษฎีการตอบสนองข้อสอบ (Miller & Spray, 1993)

หลักสำคัญของการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก คือ การเปรียบเทียบค่าไค-สแควร์ระหว่าง 2 กลุ่ม เมื่อเทียบกับค่าวิกฤติในตารางแล้ว หากพบว่ามีค่าเกินกว่าค่าวิกฤติแสดงว่าข้อมูลทั้ง 2 กลุ่มมีความสอดคล้องกันอย่างไม่มีนัยสำคัญทางสถิติ ในทางตรงกันข้ามหากมีค่าต่ำกว่าค่าวิกฤติแสดงว่าข้อมูลทั้ง 2 กลุ่มมีความสอดคล้องกันอย่างมีนัยสำคัญทางสถิติ (Su & Wang, 2005)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

งานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของแบบสอบ

งานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบแบ่งออกเป็น 3 กลุ่มใหญ่ ได้แก่ งานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบทวิวิภาค (Dichotomous) งานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาค (Polytomous) และงานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนทั้งแบบทวิวิภาคและพหุวิภาค โดยรายละเอียดของงานวิจัยแต่ละกลุ่ม เป็นดังนี้

กลุ่มแรก งานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบทวิวิภาค (Dichotomous) ประกอบไปด้วยงานวิจัยจำนวน 16 เรื่อง ดังนี้

สุรศักดิ์ อมรรัตนศักดิ์ (2531) ได้ทำการศึกษาเปรียบเทียบผลของวิธีวิเคราะห์หาความลำเอียงของข้อสอบ 4 วิธี คือ 1) วิธีวิเคราะห์ความแปรปรวน 2) วิธีแปลงค่าความยากง่ายของข้อสอบ 3) วิธีโค้งลักษณะของข้อทดสอบที่มีพารามิเตอร์ 1 ตัว และ 4) วิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 3 ตัว เมื่อใช้วิเคราะห์หาความลำเอียงต่อเพศของข้อสอบที่ใช้สอบแข่งขันเพื่อบรรจุเป็นข้าราชการ 4 ฉบับ และใช้ผลวิจัยเพศชายและเพศหญิงอย่างละ 1,170 คน ผู้วิจัยวิเคราะห์หาดัชนีความลำเอียงของข้อสอบแล้วหาสัมประสิทธิ์สหสัมพันธ์ระหว่างวิธีการวิเคราะห์ทั้ง 4 วิธี และเปรียบเทียบความแตกต่างของผลการคัดเลือกก่อนและหลังการศึกษาความลำเอียงของข้อสอบตามวิธีการคิดคะแนนรวมที่แตกต่างกัน 6 วิธี ในด้านจำนวนผู้ที่ได้รับการคัดเลือกสัดส่วนของชายต่อหญิงที่ได้รับการคัดเลือกและความเที่ยงของแบบสอบ

ผลการวิจัยพบว่าวิธีวิเคราะห์หาความลำเอียงแต่ละวิธีพบข้อสอบที่มีความลำเอียงต่างกัน โดยวิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 3 ตัว พบข้อสอบที่มีความลำเอียงจำนวนมากที่สุด และยังพบว่าค่าสัมประสิทธิ์สหสัมพันธ์ของดัชนีที่บ่งบอกความลำเอียงของข้อทดสอบทั้ง 4 วิธีสูงมาก คือมีค่า r_{xy} ระหว่าง 0.754-0.992 สำหรับการใช้คะแนนดิบและคะแนนรวมแบบต่าง ๆ กัน 5 วิธี มีจำนวนผู้ได้รับการคัดเลือกแตกต่างกันประมาณร้อยละ 4-24 ส่วนการให้คะแนนมาตรฐานทีปกติ รวมกับคะแนนแปลงแบบอื่น ๆ 4 วิธี มีจำนวนผู้ได้รับการคัดเลือกแตกต่างกันร้อยละ 4-23 และเมื่อตัดข้อสอบที่มีความลำเอียงออกแล้ว พบว่า สัดส่วนหญิงและชายที่ได้รับการคัดเลือกมีความใกล้เคียงกัน และค่าความเที่ยงของแบบสอบลดลงเล็กน้อย

เกษร หว่างจิตร (2539) ได้วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซล โดยกลุ่มอ้างอิงและกลุ่มเปรียบเทียบจำแนกตามเพศ ภูมิลำเนา ประสบการณ์ในการสอบและสังกัดของสถานศึกษา ข้อมูลที่ใช้เป็นผลการตอบข้อสอบวิชาภาษาไทยของผู้สอบจำนวน 506 คน

และผลการตอบข้อสอบวิชาภาษาอังกฤษของผู้สอบจำนวน 501 คน ในส่วนที่เป็นข้อสอบแบบเลือกตอบของศูนย์ทดสอบทางการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ผลการศึกษาพบว่าข้อสอบที่ทำหน้าที่ต่างกันส่วนมากจะเป็นข้อสอบที่มีค่าอำนาจจำแนกค่อนข้างต่ำทั้งสองวิชา เมื่อพิจารณาในด้านค่าความยากพบว่าข้อสอบที่ทำหน้าที่ต่างกันส่วนมากเป็นข้อสอบที่ง่ายมากสำหรับวิชาภาษาไทย ส่วนวิชาภาษาอังกฤษข้อสอบที่ทำหน้าที่ต่างกันส่วนมากเป็นข้อสอบที่ยากมาก อีกทั้งพบว่าส่วนมากเป็นข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป เมื่อจำแนกกลุ่มอ้างอิงและกลุ่มเปรียบเทียบตามเพศ จะพบข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนมากที่สุด รองลงมาคือการจำแนกตามภูมิลำเนา สังกัดของสถานศึกษาและประสบการณ์ในการสอบตามลำดับ

จิติมา วรรณศรี (2539) ได้เปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันระหว่างวิธีแมนเทล-แฮนส์เซลกับวิธีซิปเทสท์ โดยศึกษาจากข้อมูลจำลอง ปัจจัยที่ศึกษาได้แก่ความยาวแบบสอบ 3 ขนาด คือ 30, 60 และ 90 ข้อ ขนาดกลุ่มตัวอย่าง 3 ขนาดคือ 200, 600 และ 1000 คน โดยแต่ละขนาดมีอัตราส่วนระหว่างผู้สอบกลุ่มอ้างอิงกับกลุ่มเปรียบเทียบต่างกัน คือ 1:1, 1:0.9, 1:0.75 และ 1:0.5

ผลการศึกษาพบว่า วิธีแมนเทล-แฮนส์เซล กับวิธีซิปเทสท์ มีประสิทธิภาพเท่าเทียมกันในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ทุกขนาดกลุ่มตัวอย่างและทุกอัตราส่วนภายใต้ความยาวแบบสอบเดียวกัน และเมื่อใช้แบบสอบที่มีความยาวปานกลาง (60 ข้อ) ทั้งสองวิธีสามารถตรวจสอบได้อย่างมีประสิทธิภาพที่สุด นอกจากนี้เมื่อใช้ขนาดกลุ่มตัวอย่างมากขึ้นจะสามารถตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันได้ถูกต้องมากขึ้น โดยส่วนมากวิธีซิปเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 มากกว่าวิธีแมนเทล-แฮนส์เซลเล็กน้อย

เรวดี อินทะสะระ (2539) ได้ศึกษาความเที่ยงตรงเชิงพยากรณ์ของแบบทดสอบคัดเลือกที่วิเคราะห์ความลำเอียงต่อเพศด้วยวิธีทฤษฎีการตอบสนองข้อสอบ (IRT) วิธีแมนเทล-แฮนส์เซล (MH) และวิธีซิปเทสท์ (SIBTEST) พร้อมทั้งศึกษาการตัดสินผลการสอบที่คิดคะแนนมาตรฐานทีปกติและคะแนนน้ำหนักความสามารถและสาเหตุของความลำเอียงของข้อสอบโดยศึกษาความลำเอียงของข้อสอบคัดเลือกเข้าศึกษาในชั้นปีที่ 1 ประเภทรับตรง ปีการศึกษา 2538 ของมหาวิทยาลัยสงขลานครินทร์ ในวิชาเอกภาษาไทย ก วิชาสังคมศึกษา ก และวิชาภาษาอังกฤษ กข วิชาละ 8,127 คน (เป็นชาย 2,722 คน หญิง 5,405 คน) วิชาภาษาไทย กข วิชาสังคมศึกษา กข และวิชาภาษาอังกฤษ กขค วิชาละ 5,415 คน (เป็นชาย 1,451 คน หญิง 3,961 คน)

ผลการศึกษา พบว่าวิธีการตรวจสอบความลำเอียงทั้ง 3 วิธีตัดสินจำนวนข้อสอบที่ลำเอียงแตกต่างกันในวิชาภาษาไทย ก ฉบับที่ 2 และวิชาสังคมศึกษา ก ฉบับที่ 1 ที่ระดับนัยสำคัญทางสถิติที่ .05 นอกนั้นแตกต่างกันที่ระดับนัยสำคัญ .01โดยวิธีทฤษฏีการตอบสนองข้อสอบตัดสินจำนวนข้อสอบที่ลำเอียงได้มากที่สุด ความสัมพันธ์ของลำดับที่ของการสอบไม่ว่าจะคิดคะแนนมาตรฐานปกติทีหรือคิดคะแนนน้ำหนักความสามารถและใช้ข้อสอบทั้งหมดหรือใช้เฉพาะข้อสอบที่ปราศจากความลำเอียงต่างมีความสัมพนธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

ญาณภัทร สีหะมงคล (2540) เปรียบเทียบความสอดคล้องของผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันระหว่างวิธี Lord's χ^2 วิธี Raju's Area Measures และวิธี Closed Interval Area เมื่อขนาดของกลุ่มตัวอย่าง ความยาวของแบบทดสอบ และสัดส่วนจำนวนข้อสอบที่ทำหน้าที่ต่างกันในแบบทดสอบต่างกัน ข้อมูลที่ใช้ในการศึกษาเป็นผลการสอบประเมินคุณภาพและความก้าวหน้าทางการศึกษา วิชาคณิตศาสตร์ของนักเรียนชั้นประถมศึกษาปีที่ 4 ปีการศึกษา 2536 ของสำนักงานการประถมศึกษาแห่งชาติจำนวน 11,404 คน เครื่องมือที่ใช้เป็นแบบทดสอบแบบเลือกตอบจำนวน 80 ข้อ

ผลการศึกษาพบว่า จำนวนข้อสอบที่ทำหน้าที่ต่างกันจากการตรวจสอบด้วยวิธีการทั้งสามแตกต่างกันเมื่อขนาดของกลุ่มตัวอย่างและความยาวของแบบทดสอบต่างกัน ส่วนความสัมพันธ์ระหว่างวิธีการทั้งสามมีค่าสัมประสิทธิ์สหสัมพันธ์ค่อนข้างสูงมากและมีนัยสำคัญทางสถิติเกือบทุกเงื่อนไขของการศึกษา และสำหรับความสอดคล้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบส่วนมากจะมีค่าปานกลางถึงต่ำเกือบทุกเงื่อนไขของการศึกษา

พรรณี จิตมาศ (2540) ได้วิเคราะห์ความลำเอียงต่อเพศของแบบทดสอบคณิตศาสตร์โจทย์ปัญหาที่ผู้วิจัยสร้างขึ้นเอง ด้วยวิธีวิเคราะห์ความลำเอียง 3 วิธี คือ วิธีแปลงค่าความยาก วิธีแมนเทล-แฮนส์เซล และวิธี SIBTEST ในแต่ละขนาดของกลุ่มผู้สอบ 500 และ 1,000 คน โดยเปรียบเทียบจำนวนข้อที่มีความลำเอียงและเปรียบเทียบ ค่าความเชื่อมั่นแบบครึ่งฉบับของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2539 ของโรงเรียนสังกัดกรมสามัญศึกษาส่วนกลาง จำนวน 2,200 คน ซึ่งเลือกมาโดยการสุ่มแบบแบ่งชั้น มีขนาดของโรงเรียนเป็นชั้นและโรงเรียนเป็นหน่วยการสุ่ม เครื่องมือที่ใช้คือ แบบทดสอบคณิตศาสตร์โจทย์ปัญหาเรื่องสมการ อัตราส่วน และร้อยละเป็นแบบเลือกตอบชนิด 5 ตัวเลือก จำนวน 40 ข้อ ที่ผู้วิจัยสร้างขึ้นเอง จำนวน 1 ฉบับ

ผลการวิจัยพบว่า เมื่อวิเคราะห์จากกลุ่มตัวอย่างขนาด 500 คน วิธี SIBTEST พบข้อสอบที่มีความลำเอียงมากที่สุดและวิธีแปลงค่าความยากพบข้อสอบที่มีความลำเอียงน้อยที่สุด โดย

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

จำนวนข้อสอบที่มีความลำเอียงแตกต่างกันอย่างไม่มีนัยสาคัญทางสถิติทุกวิธีวิเคราะห์และเมื่อวิเคราะห์จากกลุ่มผู้สอบขนาด 1,000 คน วิธีแมนเทล-แฮนส์เซล พบข้อสอบมีความลำเอียงมากที่สุด วิธีแปลงค่าความยากไม่พบข้อสอบที่ลำเอียง โดยจำนวนข้อสอบที่ลำเอียงจากการวิเคราะห์ด้วยวิธีแปลงค่าความยาก กับ วิธีแมนเทล-แฮนส์เซล และวิธีแปลงค่าความยากกับวิธี SIBTEST แตกต่าง กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 นอกนั้นมีค่าไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ

รัชรินทร์ มุคดา (2540) ศึกษาการเปรียบเทียบประสิทธิภาพระหว่างวิธีแมนเทล-แฮนส์เซลกับวิธีถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันของแบบสอบอเนกรูปในกรณีที่จัดกลุ่มความสามารถ ค่าความยากของข้อสอบ ค่าอำนาจจำแนกของข้อสอบต่างกัน โดยศึกษาจากข้อมูลที่จำลองขึ้นด้วยโปรแกรม IRTDATA และเงื่อนไขที่ศึกษา คือ กลุ่มความสามารถผู้สอบ 3 ระดับ ได้แก่ กลุ่มผู้สอบที่มีความสามารถสูง ปานกลางและต่ำ ค่าความยากของข้อสอบ 3 ระดับ ได้แก่ กลุ่มข้อสอบที่มีความยากสูง ปานกลาง และต่ำ ค่าอำนาจจำแนกของข้อสอบ 3 ระดับ ได้แก่ กลุ่มข้อสอบที่มีอำนาจจำแนกสูง ปานกลาง และต่ำ

ผลการวิจัยพบว่า วิธีแมนเทล-แฮนส์เซลกับวิธีถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันของแบบสอบอเนกรูปเท่ากันในทุกกลุ่มผู้สอบ และในกลุ่มผู้สอบที่มีความสามารถสูง ข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันแบบอเนกรูปมากที่สุดเป็นข้อสอบที่มีความยากสูง อำนาจจำแนกสูง ในกลุ่มผู้สอบที่มีความสามารถปานกลาง เป็นข้อสอบที่มีค่าความยากปานกลาง ค่าอำนาจจำแนกสูง ในกลุ่มผู้สอบที่มีความสามารถต่ำ เป็นข้อสอบที่มีความยากต่ำ ค่าอำนาจจำแนกสูง

เสรี ชัดแช้ม (2540) ศึกษาเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีแมนเทล-แฮนส์เซลแบบปกติกับวิธีแมนเทล-แฮนส์เซลแบบแบ่งกลุ่มความสามารถของผู้สอบและความยากของข้อสอบโดยใช้วิธี IRT เป็นเกณฑ์โดยศึกษาจากข้อมูลผลการตอบแบบสอบวัดความสามารถในการอ่านภาษาไทยของนักเรียนชั้นมัธยมศึกษาปีที่ 1 สังกัดกรมสามัญศึกษาจังหวัดชลบุรี จำนวน 1,200 คน โดยกลุ่มผู้สอบจำแนกตามเพศ

ผลการศึกษาพบว่า วิธี MH แบบแบ่งกลุ่มความสามารถของผู้สอบและความยากของข้อสอบ สามารถตรวจพบข้อสอบทำหน้าที่ต่างกันแบบอเนกรูปได้สอดคล้องกับวิธี IRT และตรวจพบข้อสอบทำหน้าที่ต่างกันมากกว่าวิธี MH แบบปกติ ข้อสอบที่ตรวจพบส่วนใหญ่เป็นข้อสอบยากปานกลางและข้อสอบง่าย ซึ่งมีโค้งลักษณะข้อสอบของกลุ่มผู้สอบสองกลุ่มตัดกันบริเวณใกล้ ๆ จุดกลางของช่วงความสามารถ

นพมาศ พิพัฒนสุข (2541) ได้ศึกษาเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีแมนเทล-แฮนส์เซลกับวิธีถดถอยโลจิสติกในแบบสอบพหุมิติ เมื่อใช้เกณฑ์จับคู่เปรียบเทียบแตกต่างกัน 3 เกณฑ์ ได้แก่ คะแนนรวม คะแนนแบบสอบย่อย และคะแนนหลายแบบสอบย่อย โดยเก็บข้อมูลจากแบบสอบวัดความสามารถทางคณิตศาสตร์ที่ผู้วิจัยสร้างขึ้น

ผลการวิจัยพบว่าวิธีแมนเทล-แฮนส์เซลมีประสิทธิภาพมากกว่าวิธีถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบชนิดพหุมิติเมื่อใช้เกณฑ์จับคู่คะแนนรวม และมีประสิทธิภาพไม่แตกต่างกันเมื่อใช้เกณฑ์จับคู่คะแนนแบบสอบย่อยและวิธีถดถอยโลจิสติก เมื่อใช้เกณฑ์การจับคู่เปรียบเทียบคะแนนหลายแบบสอบย่อยมีความเหมาะสมในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบชนิดพหุมิติ

นิคม กีรติวากูร (2542) ได้เปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีการวิเคราะห์องค์ประกอบจำกัด (RFA) วิธีแมนเทล-แฮนส์เซล (MH) และวิธีการตอบสนองข้อสอบ (IRT) แบบ 2 พารามิเตอร์ โดยเทียบกับเกณฑ์ที่กำหนด โดยศึกษาจากข้อมูลจำลอง ปัจจัยที่ศึกษาได้แก่ ขนาดกลุ่มตัวอย่าง 2 ขนาด คือ ขนาดเล็ก (300 คน) และขนาดใหญ่ (1000 คน) ค่าความยาวแบบสอบ 2 ขนาด คือ แบบสอบสั้น (25 ข้อ) และแบบสอบยาว (75 ข้อ) ค่าความยากของข้อสอบแบ่งออกเป็น 3 ระดับคือ กลุ่มข้อสอบที่มีความยากสูง ปานกลาง และต่ำ ขนาดความลำเอียงของข้อสอบแบ่งออกเป็น 2 ขนาดคือ กลุ่มข้อสอบที่มีความลำเอียงสูงและต่ำ

ผลการวิจัยพบว่า วิธี RFA มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูงที่สุดโดย วิธี MH มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูงภายใต้เงื่อนไขแบบสอบที่มีความยากต่ำ อำนาจจำแนกสูง ส่วนวิธี IRT แบบ 2 พารามิเตอร์ มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูง ภายใต้เงื่อนไขแบบสอบที่มีความยากต่ำ และวิธี IRT มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธี MH และวิธี RFA ตามลำดับ

อารี วัชรโสตถิกุล (2543) ได้ศึกษาเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้รูปแบบต่างกัน คือ รูปแบบคะแนนรวมทั้งฉบับแยกตามเนื้อหา และแยกตามระดับพฤติกรรมด้วยวิธีการตรวจสอบต่างกัน คือ วิธี SIBTEST และวิธีการถดถอยโลจิสติก แล้วทำการคัดเลือกข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ เพื่อเปรียบเทียบค่าความเชื่อมั่น

ผลการศึกษาพบว่า จำนวนข้อสอบที่ทำหน้าที่ต่างกันโดยใช้วิธีการตรวจสอบต่างกัน แตกต่างกันในรูปแบบรวมทั้งฉบับ ส่วนรูปแบบแยกตามเนื้อหาและแยกตามระดับพฤติกรรมไม่แตกต่างกัน

ทองอยู่ สาระ (2543) ได้ศึกษาเปรียบเทียบอำนาจการตรวจสอบและจำแนกผิดพลาดในการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบสม่ำเสมอและแบบไม่สม่ำเสมอ ระหว่างวิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก โดยใช้ความยาวแบบทดสอบและขนาดกลุ่มตัวอย่างต่างกัน

ผลการศึกษาพบว่าอำนาจการตรวจสอบและการจำแนกผิดพลาดในการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันทั้งแบบสม่ำเสมอและแบบไม่สม่ำเสมอ ระหว่างวิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกภายใต้ความยาวแบบทดสอบและขนาดกลุ่มตัวอย่างสูง ปานกลาง และต่ำที่ศึกษาเกือบทุกเงื่อนไขไม่แตกต่างกัน ส่วนความยาวของแบบทดสอบไม่มีผลต่ออำนาจการตรวจสอบและการจำแนกผิดพลาดในทั้ง 2 วิธี แต่ขนาดของกลุ่มตัวอย่างมีผลต่ออำนาจการตรวจสอบด้วยวิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกเกือบทุกเงื่อนไขของการศึกษา คือเมื่อขนาดกลุ่มตัวอย่างเพิ่มขึ้น อำนาจการตรวจสอบจะเพิ่มขึ้น แต่ขนาดของกลุ่มตัวอย่างไม่มีผลต่อการจำแนกผิดพลาดในเกือบทุกเงื่อนไข

วลีมาศ แซ่อึ้ง (2543) เปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ข้อมูลที่ใช้ในการศึกษาจำลองภายใต้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ชนิดกำหนดค่าการเดา (c) คงที่ แล้วจัดกระทำข้อมูลตามปัจจัย 4 ตัว คือ (1) ลักษณะของข้อสอบที่มีค่าความยาก (b) และอำนาจจำแนก (a) ระดับต่ำ ปานกลาง และสูง จำนวน 9 ลักษณะ (2) ความยาวของแบบสอบ 2 ระดับ (3) สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบ 3 ระดับและ (4) ขนาดกลุ่มตัวอย่าง 6 ระดับ รวมข้อมูลที่ศึกษาทั้งหมดจำนวน 324 เงื่อนไข แล้วนำข้อมูลของแต่ละเงื่อนไขมาคำนวณค่าอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป

ผลการวิจัยพบว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป ของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าเท่าเทียมกันภายใต้เกือบทุกเงื่อนไข และทั้งสองวิธีดังกล่าว มีอำนาจการทดสอบสูงกว่าวิธีซิปเทสท์ และวิธีแมนเทล-แฮนส์เซลภายใต้เกือบทุกเงื่อนไข ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปของวิธี ซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก มีค่าอยู่ภายในเกณฑ์ของอัตราความคลาดเคลื่อนประเภทที่ 1 ที่ระดับ 10% ภายใต้เกือบทุกเงื่อนไข

รักชนก ยี่สุ่นศรี (2544) ทำการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบสำหรับกลุ่มผู้สอบเมื่อจำแนกตามเพศและสถานที่ตั้งทางภูมิศาสตร์และโรงเรียนที่จบการศึกษาเพื่อเปรียบเทียบความเที่ยง ความตรง และฟังก์ชันสารสนเทศของแบบสอบระหว่างแบบสอบฉบับก่อนและหลังตัดข้อสอบที่พบ DIF โดยใช้ข้อมูลจากการตอบแบบสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษา วิชาภาษาอังกฤษและคณิตศาสตร์ ปีการศึกษา 2543 ครั้งที่ 1 และเลือกศึกษาในส่วนที่เป็นข้อสอบแบบหลายตัวเลือกจากกลุ่มตัวอย่างที่เป็นผู้เข้าสอบจำนวน 4,000 คน และ 3,600 คน ตามลำดับ

ผลวิจัยพบว่า แบบสอบวิชาภาษาอังกฤษทำหน้าที่ต่างกันตามเพศและสถานที่ตั้งตามภูมิศาสตร์ ส่วนแบบสอบวิชาคณิตศาสตร์ทำหน้าที่ต่างกันตามเพศของผู้สอบ และข้อสอบทำหน้าที่ต่างกันตามเพศของผู้สอบมากที่สุดทั้งสองวิชา นอกจากนี้เมื่อเปรียบเทียบคุณภาพของแบบสอบฉบับก่อนและหลังการตัดข้อสอบที่พบว่าทำหน้าที่ต่างกัน พบว่า ในด้านความตรงจะไม่ต่างกัน แต่แบบสอบฉบับหลังจากตัดข้อสอบที่พบว่าทำหน้าที่ต่างกัน ส่วนใหญ่จะมีค่าความเที่ยงลดลง และมีค่าฟังก์ชันสารสนเทศมากขึ้น ส่วนความสัมพันธ์ระหว่างตำแหน่งของคะแนนรวมของผู้สอบก่อนและหลังจากการตัดข้อสอบที่พบว่าทำหน้าที่ต่างกันทั้งในกรณีที่ตัดทุกข้อและตัดในบางข้อ พบว่าทุกกรณีมีความสัมพันธ์ในทางบวกซึ่งกันและกันอย่างมีนัยสำคัญ

ปิยะทิพย์ ตินวร (2549) ได้เปรียบเทียบประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบพหุมิติ ระหว่างวิธีการวิเคราะห์องค์ประกอบจำกัดกับวิธีถดถอยโลจิสติก ภายใต้เงื่อนไข 18 เงื่อนไข คือ ขนาดกลุ่มตัวอย่าง 3 ขนาด (2000, 1000 และ 300 คน) ความยาวของแบบทดสอบ 3 ขนาด (40, 30 และ 20 ข้อ) และเกณฑ์การจับคู่ 2 เกณฑ์ (คะแนนรวมทั้งฉบับ และคะแนนแบบทดสอบย่อย) เมื่อใช้ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยวิธีซิปเทสท์เป็นเกณฑ์สำหรับการเปรียบเทียบประสิทธิภาพ กลุ่มตัวอย่างเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ปีการศึกษา 2546 ที่เข้าสอบวัดผลสัมฤทธิ์ทางการเรียนระดับชาติ วิชาภาษาไทย จำนวน 2000 คน

ผลการวิจัย พบว่า วิธีถดถอยโลจิสติกมีประสิทธิภาพไม่แตกต่างกันกับวิธีวิเคราะห์องค์ประกอบจำกัด ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบพหุมิติจาก 2 เงื่อนไข และวิธีถดถอยโลจิสติกมีประสิทธิภาพมากกว่าวิธีวิเคราะห์องค์ประกอบจำกัด ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบพหุมิติ 16 เงื่อนไข

Mazor et al. (1992) ได้ศึกษาผลกระทบของขนาดกลุ่มตัวอย่างที่มีต่อการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซล โดยศึกษาจากข้อมูลจำลอง กลุ่มตัวอย่างที่

ใช้มี 5 ขนาด คือ 100 200 500 1,000 และ 2,000 คน ความยาวของแบบสอบ 75 ข้อ พบว่าเมื่อขนาดกลุ่มตัวอย่างเท่ากับ 500 คนหรือน้อยกว่า จะสามารถระบุข้อสอบที่ทำหน้าที่ต่างกันได้ถูกต้องน้อยกว่าร้อยละ 50 และเมื่อขนาดกลุ่มตัวอย่างเท่ากับ 2,000 คน จะสามารถระบุข้อสอบที่ทำหน้าที่ต่างกันได้ถูกต้อง ร้อยละ 70 ถึงร้อยละ 75 และได้กล่าวว่า ข้อสอบที่ไม่สามารถตรวจสอบพบหรือระบุว่าทำหน้าที่ต่างกันได้ เนื่องมาจากข้อสอบเหล่านั้นมีความยากมากหรือมีความยากต่างกันเพียงเล็กน้อยระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ อีกทั้งเป็นข้อที่มีค่าอำนาจจำแนกต่ำ

จากการศึกษางานวิจัยในกลุ่มแรกสามารถสรุปวิธีที่ได้ศึกษามาแล้ว ดังตารางที่ 1

ตารางที่ 1 สรุปงานวิจัยที่ได้ศึกษาเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบทวิวิภาค (Dichotomous)

ผู้ที่ศึกษา	ประเด็นที่ศึกษา	ผลการศึกษา
สุรศักดิ์ อมรรัตนศักดิ์ (2531)	เปรียบเทียบผลของวิธีวิเคราะห์หาความลำเอียงของข้อสอบ 4 วิธี คือ 1) วิธีวิเคราะห์ความแปรปรวน 2) วิธีแปลงค่าความยากง่ายของข้อสอบ 3) วิธีโค้งลักษณะของข้อทดสอบที่มีพารามิเตอร์ 1 ตัว และ 4) วิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 3 ตัว	วิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 3 ตัว พบข้อทดสอบที่มีความลำเอียงจำนวนมากที่สุด และยังพบว่าค่าสัมประสิทธิ์สหสัมพันธ์ของดัชนีที่บ่งบอกความลำเอียงของข้อทดสอบทั้ง 4 วิธีสูงมาก
เกษร หว่างจิตร (2539)	วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธี MH โดยกลุ่มอ้างอิงและกลุ่มเปรียบเทียบจำแนกตามเพศ ภูมิลำเนา ประสบการณ์ในการสอบ และสังกัดของสถานศึกษา	ข้อสอบที่ทำหน้าที่ต่างกันส่วนมากจะเป็นข้อสอบที่มีค่าอำนาจจำแนกค่อนข้างต่ำ และเมื่อจำแนกตามเพศจะพบข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนมากที่สุด
จิติมา วรรณศรี (2539)	เปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันระหว่างวิธี MHกับวิธีซิปเทสท์	แบบสอบที่มีความยาวปานกลาง ทั้ง 2 วิธีสามารถตรวจสอบได้อย่างมีประสิทธิภาพที่สุด เมื่อใช้ขนาดกลุ่มตัวอย่างมากขึ้นจะสามารถตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันได้ถูกต้องมากขึ้น

ตารางที่ 1 (ต่อ)

ผู้ที่ศึกษา	ประเด็นที่ศึกษา	ผลการศึกษา
เรวดี อินทะสะระ (2539)	ศึกษาความเที่ยงตรงเชิงพยากรณ์ของแบบทดสอบคัดเลือกที่วิเคราะห์ความลำเอียงต่อเพศด้วยวิธี IRT วิธี MH และวิธีซิปเทสท์	วิธีการตรวจสอบความลำเอียงทั้ง 3 วิธีตัดสินจำนวนข้อสอบที่ลำเอียงแตกต่างกัน โดยวิธี IRT ตัดสินจำนวนข้อสอบที่ลำเอียงได้มากที่สุด
ญาณภัทร สีหะมงคล (2540)	เปรียบเทียบความสอดคล้องของผลการตรวจสอบขอสอบที่ทำหน้าที่ต่างกันระหว่างวิธี Lord's χ^2 วิธี Raju's Area Measures และวิธี Closed Interval Area	จำนวนข้อสอบที่ทำหน้าที่ต่างกันจากการตรวจสอบด้วยวิธีการทั้งสามวิธีแตกต่างกันเมื่อขนาดของกลุ่มตัวอย่างและความยาวของแบบทดสอบต่างกัน
พรรณี จิตมาศ (2540)	วิเคราะห์ความลำเอียงต่อเพศของแบบทดสอบคณิตศาสตร์ ด้วยวิธีวิเคราะห์ความลำเอียง 3 วิธี คือ วิธีแปลงค่าความยาก วิธี MH และวิธี SIBTEST	เมื่อวิเคราะห์จากกลุ่มตัวอย่างขนาด 500 คน วิธี SIBTEST พบข้อสอบที่มีความลำเอียงมากที่สุดและเมื่อวิเคราะห์จากกลุ่มผู้สอบขนาด1,000 คน วิธี MH พบข้อสอบมีความลำเอียงมากที่สุด
รัชรินทร์ มุคดา (2540)	ศึกษาการเปรียบเทียบประสิทธิภาพระหว่างวิธี MHกับวิธีถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันของแบบสอบอเนกรูปในกรณีที่จัดกลุ่มความสามารถ ค่าความยากของข้อสอบ ค่าอำนาจจำแนกของข้อสอบต่างกัน	วิธี MH กับวิธีถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันของแบบสอบอเนกรูปเท่ากันในทุกกลุ่มผู้สอบ
เสรี ชัดแช้ม (2540)	ศึกษาเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธี MH แบบปกติกับวิธี MHแบบแบ่งกลุ่มความสามารถของผู้สอบและความยากของข้อสอบ	วิธี MH แบบแบ่งกลุ่มความสามารถของผู้สอบและความยากของข้อสอบสามารถตรวจพบข้อสอบทำหน้าที่ต่างกันแบบอเนกรูปได้สอดคล้องกับวิธี IRT และตรวจพบข้อสอบทำหน้าที่ต่างกันมากกว่าวิธี MH แบบปกติ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 1 (ต่อ)

ผู้ที่ศึกษา	ประเด็นที่ศึกษา	ผลการศึกษา
นพมาศ พิพัฒนสุข (2541)	เปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี MHกับวิธีถดถอยโลจิสติกในแบบสอบพหุมิติ เมื่อใช้เกณฑ์จับคู่เปรียบเทียบแตกต่าง	วิธี MH มีประสิทธิภาพมากกว่าวิธีถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบชนิดพหุมิติเมื่อใช้เกณฑ์จับคู่คะแนนรวม และมีประสิทธิภาพไม่แตกต่างกันเมื่อใช้เกณฑ์จับคู่คะแนนแบบสอบย่อย
นิคม กีรติวาภูร (2542)	เปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี RFA วิธี MH และวิธี IRT แบบ 2 พารามิเตอร์	วิธี RFA มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูงที่สุด และวิธี IRT มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธี MH และวิธี RFA ตามลำดับ
อารี วัชรโสตถิกุล (2543)	เปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้รูปแบบต่างกัน และวิธีการตรวจสอบต่างกัน คือ วิธี SIBTEST และวิธีการถดถอยโลจิสติก	จำนวนข้อสอบที่ทำหน้าที่ต่างกันโดยใช้วิธีการตรวจสอบต่างกัน แตกต่างกันในรูปแบบรวมทั้งฉบับ ส่วนรูปแบบแยกตามเนื้อหาและแยกตามระดับพฤติกรรมไม่แตกต่างกัน
ทองอยู่ สาระ (2543)	เปรียบเทียบอำนาจการตรวจสอบและจำแนกผิดพลาดในการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบสม่ำเสมอและแบบไม่สม่ำเสมอ ระหว่างวิธี MH และวิธีการถดถอยโลจิสติก	ความยาวของแบบทดสอบไม่มีผลต่ออำนาจการตรวจสอบและการจำแนกผิดพลาดในทั้ง 2 วิธี แต่เมื่อขนาดกลุ่มตัวอย่างเพิ่มขึ้น อำนาจการตรวจสอบจะเพิ่มขึ้น
วลีมาศ แซ่อึ้ง (2543)	เปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธี MH และวิธีการถดถอยโลจิสติก	อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป ของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าเท่าเทียมกันภายใต้เกือบทุกเงื่อนไข และทั้งสองวิธีมีอำนาจการทดสอบสูงกว่าวิธีซิปเทสท์ และวิธี MH ภายใต้เกือบทุกเงื่อนไข ส่วน อัตราความคลาดเคลื่อนประเภทที่ 1 ทั้ง 4 วิธี มีค่าอยู่ภายในเกณฑ์

ตารางที่ 1 (ต่อ)

ผู้ที่ศึกษา	ประเด็นที่ศึกษา	ผลการศึกษา
รักชนก ยี่สุ่นศรี (2544)	วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ สำหรับกลุ่มผู้สอบเมื่อจำแนกตามเพศและสถานที่ตั้งทางภูมิศาสตร์และโรงเรียนที่จบการศึกษา	แบบสอบวิชาภาษาอังกฤษทำหน้าที่ต่างกันตามเพศและสถานที่ตั้งตามภูมิศาสตร์ ส่วนแบบสอบวิชาคณิตศาสตร์ทำหน้าที่ต่างกันตามเพศของผู้สอบ และข้อสอบทำหน้าที่ต่างกันตามเพศของผู้สอบมากที่สุดทั้งสองวิชา
ปิยะทิพย์ ตินวร (2549)	เปรียบเทียบประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบพหุมิติ ระหว่างวิธีการวิเคราะห์องค์ประกอบจำกัดกับวิธีถดถอยโลจิสติก	วิธีถดถอยโลจิสติกมีประสิทธิภาพไม่แตกต่างกันกับวิธีวิเคราะห์องค์ประกอบจำกัดและวิธีถดถอยโลจิสติกมีประสิทธิภาพมากกว่าวิธีวิเคราะห์องค์ประกอบจำกัด
Mazor et al. (1992)	ผลกระทบของขนาดกลุ่มตัวอย่างที่มีต่อการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซล	เมื่อขนาดกลุ่มตัวอย่างเท่ากับ 500 คนหรือน้อยกว่า จะสามารถระบุข้อสอบที่ทำหน้าที่ต่างกันได้ถูกต้องน้อยกว่าร้อยละ 50 และเมื่อขนาดกลุ่มตัวอย่างเท่ากับ 2,000 คน จะสามารถระบุข้อสอบที่ทำหน้าที่ต่างกันได้ถูกต้องร้อยละ 70 ถึงร้อยละ 75

จะเห็นว่างานวิจัยเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบ 2 ค่า (แบบ 0-1) ตัวแปรที่ศึกษา ได้แก่ ขนาดของแบบทดสอบ ขนาดของกลุ่มตัวอย่าง อัตราส่วนของกลุ่มประชากรย่อย อัตราส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบทดสอบ และวิธีการที่ใช้ในการตรวจสอบ ส่วนการแบ่งกลุ่มประชากรย่อยส่วนมากจำแนกตามเพศ ภูมิลำเนา สังกัดของสถานศึกษา เป็นต้น

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

กลุ่มที่สอง งานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาค (Polytomous) ประกอบไปด้วยงานวิจัยจำนวน 4 เรื่อง ดังนี้

French & Miller (1996) ได้ศึกษาความเป็นไปได้ของการใช้วิธีการถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่าโดยศึกษาจากแบบทดสอบที่จำลองขึ้น จำนวน 25 ข้อ แต่ละข้อมี 4 ลำดับขั้นคะแนน คือ ตั้งแต่ 0 ถึง 3 คะแนน ซึ่งในแบบทดสอบจะมีข้อสอบข้อเดียวที่จำลองพารามิเตอร์ของข้อสอบให้แตกต่างกันเพื่อสร้างเงื่อนไข 3 เงื่อนไขสำหรับการทำหน้าที่ต่างกันแบบอเนกรูปและมี 1เงื่อนไขสำหรับการทำหน้าที่ต่างกันแบบเอกรูป ส่วนคะแนนสอบถูกจำลองขึ้นโดยใช้โมเดล GPCM (generalized partial credit model) ของ Muraki (1992) ข้อมูลที่ได้จะถูกนำมากำหนดรหัสใหม่ (recoded) ให้เป็น multiple dichotomies โดยใช้โมเดลตามแนวคิดของ Agresti (1990) 3 โมเดล คือโมเดลโลจิทของอัตราส่วนที่ต่อเนื่องกัน โมเดลโลจิทแบบสะสม และโมเดลโลจิทของลำดับขั้นที่ติดกัน เพื่อใช้กับวิธีการถดถอยโลจิสติกในการกำหนดรหัสให้ใช้กับรูปแบบการตอบแบบหลายค่า (multinomial response models)

ผลการศึกษาพบว่า การเปลี่ยนแปลงขนาดของกลุ่มตัวอย่างมีผลต่ออำนาจในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อกลุ่มตัวอย่างมีขนาดเล็กลง อำนาจในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบลดลง วิธีโมเดลโลจิทของอัตราส่วนที่ต่อเนื่องกันและวิธีโมเดลโลจิทแบบสะสมมีอำนาจในการตรวจสอบสูงสุดในทุกๆครั้งของการถดถอยแต่ต่างกันที่วิธีโมเดลโลจิทของอัตราส่วนที่ต่อเนื่องกันมีการสูญเสียข้อมูลในการกำหนดรหัสครั้งที่ 2 และ 3 ในขณะที่วิธีโมเดลโลจิทแบบสะสม ยังคงเก็บข้อมูลทั้งหมด ไว้ในทุกครั้งของการกำหนดรหัสส่วนวิธีโมเดลโลจิทของลำดับขั้นที่ติดกัน มีอำนาจการตรวจสอบของการถดถอยครั้งแรกต่ำ แต่มีอำนาจเพียงพอในการตรวจสอบของการถดถอยครั้งที่ 2 และ 3 วิธีการนี้มีการสูญเสียข้อมูลในทุกครั้งของการถดถอยและมีอำนาจในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบต่ำกว่าอีก 2 วิธี แต่มีประสิทธิภาพในการเปรียบเทียบระหว่างแต่ละลำดับขั้นคะแนนโดยตรง ซึ่งมีประโยชน์ในการค้นหาตำแหน่งของการทำหน้าที่ต่างกันของข้อสอบภายในข้อสอบ

นอกจากนั้นพบว่า เมื่อพารามิเตอร์อำนาจจำแนกของข้อสอบยิ่งแตกต่างกันมาก อำนาจในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปยิ่งเพิ่มขึ้น และพบว่าการทำให้ข้อมูลแบบหลายค่ากลายเป็นข้อมูลแบบ 2 ค่า ตามหลักการกำหนดรหัสของทั้ง 3 วิธี มีผลต่ออำนาจในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเนื่องจากเกิดการสูญหายของข้อมูล

Flowers et al. (1997) ได้ศึกษาการบรรยาย DFIT ของข้อสอบที่ให้คะแนนแบบ Polytomous และประเมินรวมถึงเปรียบเทียบการทำ DFIT ในการแผ่ขยายของขั้นตอน SIBTEST และ Lord’s chi-square โดยใช้ข้อมูลที่จำลองขึ้นมา กลุ่มตัวอย่างจำนวน 500 และ 1,000 คน การกระจายของกลุ่มเปรียบเทียบ (N(0,1) และ N(-1,1)) จำนวนข้อสอบที่ทำหน้าที่ต่างกัน (0%, 10%, และ 20%) ค่า DIF ที่มากที่สุดและค่าพารามิเตอร์ a ได้รับการประเมิน ค่าความคลาดเคลื่อนประเภทที่ 1 จะได้รับผลกระทบจากจำนวนข้อสอบที่ทำหน้าที่ต่างกัน ค่า DIF ที่มากที่สุดและค่าของพารามิเตอร์ a การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้รับผลกระทบจากทุกปัจจัยในการศึกษาครั้งนี้ การคำนวณ DFIT และ Lord’s chi-square จำเป็นต้องใช้การประมาณค่าความสามารถและข้อสอบคือขั้นตอนปรับเทียบ และใช้โปรแกรมคอมพิวเตอร์ PARSCALE 2 วิธี Maximum marginal likelihood และ EM algorithm การประมาณค่าสัมประสิทธิ์การปรับเทียบใช้โปรแกรม EQUATE 2.0 การคำนวณ DFIT ใช้โปรแกรม FORTRAN ที่เขียนโดย Raju (1995) และการคำนวณ Lord’s chi-square ใช้โปรแกรม FORTRAN ที่เขียนโดย Kim (1993) ส่วนการตรวจสอบ DIF โดยวิธี SIBTEST ใช้โปรแกรม PSIBTEST

สรุปผลการวิจัยได้ว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ใกล้เคียงกับระดับแอลฟา ยกเว้นเมื่อจำนวนข้อสอบที่มี DIF 20% และจำนวน DIF ที่มากที่สุดมีค่าสูง ปัจจัยที่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 คือค่าของพารามิเตอร์ นั่นคือหากค่า a พารามิเตอร์ต่ำอัตราความคลาดเคลื่อนประเภทที่ 1 จะสูง การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้รับผลกระทบมาจากทุกปัจจัยในการศึกษาครั้งนี้

Su & Wang (2005) ได้จำลองข้อมูลในการสืบสอบปัจจัยที่มีผลต่อการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธีแมนเทล (Mantel) วิธีแมนเทล-แฮนส์เซลทั่วไป (Generalized Mantel-Haenzel: GMH) และวิธี Logistic Discriminant Function Analysis (LDFA) ได้นำมาใช้ในการประเมินการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาค ผลแสดงให้เห็นว่าความสำคัญของคะแนนจับคู่ คือการวัดพื้นที่เฉลี่ยระหว่างโค้งลักษณะข้อสอบ 2 โค้งของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ซึ่งมีความสำคัญมากกว่าจำนวนของข้อสอบที่ทำหน้าที่ต่างกัน ทั้ง 3 วิธีมีการควบคุมค่าอัตราความคลาดเคลื่อนประเภทที่ 1 ได้เป็นอย่างดี วิธีแมนเทล (Mantel) และวิธี Logistic Discriminant Function Analysis (LDFA) มีอำนาจการตรวจสอบสูงกว่า วิธี Logistic Discriminant Function Analysis (LDFA)

Oishi (2006) ได้ตรวจสอบความเท่าเทียมของการวัดความพึงพอใจด้วยแบบวัดความพึงพอใจในชีวิตระหว่างกลุ่มตัวอย่างชาวอเมริกันและชาวจีนโดยใช้ Multigroup Structural Equation Modeling (SEM), Multiple indicator multiple cause model (MIMIC) และทฤษฎีการตอบสนองข้อสอบ (IRT) กลุ่มตัวอย่างเป็นนักเรียน 556 คน ในสถาบันเทคโนโลยี Zhejiang ประเทศจีน โดยผู้วิจัยให้ทำแบบสอบถามในชั้นเรียน และนักเรียน 442 คน ใน University of Illinois ที่ได้ลงทะเบียนวิชาจิตวิทยาเบื้องต้น โดยใช้แบบสอบถามทำในชั้นเรียน การวัดผลทำโดยใช้แบบวัด SWLS ที่ใช้ในการประเมินความพึงพอใจในชีวิตของคนทั่วโลก แบบวัดนี้ประกอบด้วย 5 ข้อคำถาม ผู้เข้าร่วมตอบสนองแต่ละข้อโดยใช้สเกล 7 ระดับเพื่อจัดลำดับจาก 1 (ไม่เห็นด้วยมากที่สุด) ถึง 7 (เห็นด้วยมากที่สุด)

สรุปผลวิจัยได้ว่า การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบจัดให้เห็นแง่มุมความแตกต่างของวิธีแบบดั้งเดิมในการวัดประเด็นการวิจัยทางวัฒนธรรมและความเป็นอยู่ที่ดี การวิเคราะห์ IRT แสดงให้เห็นความแตกต่างของค่าเฉลี่ยระหว่างกลุ่มชาวจีนและชาวอเมริกัน ข้อสอบที่มีความลำเอียงจะให้คะแนนโดยมีน้ำหนักน้อยกว่า ดังนั้นความแตกต่างของค่าเฉลี่ยที่ค้นพบก่อนหน้าระหว่างกลุ่มชาวจีนและชาวอเมริกันอาจจะไม่ค้นพบได้ง่ายในข้อสอบที่มีความลำเอียง สุดท้ายการวิเคราะห์ IRT จัดหาข้อมูลที่มากกว่าที่เกี่ยวข้องกับแนวคิดของความพึงพอใจในชีวิต การวิจัยครั้งนี้แสดงให้เห็นความสำคัญและประโยชน์ของการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบในโครงสร้างอื่นๆ ผู้วิจัยหวังว่า การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและโมเดล IRT อื่น ๆ จะเป็นประโยชน์ในอนาคตในหัวข้อการวิจัยอื่นๆที่เกี่ยวข้องในเรื่องวัฒนธรรมและบุคลิกภาพ

จากการศึกษางานวิจัยในกลุ่มที่สองสามารถสรุปวิธีที่ได้ศึกษามาแล้ว ดังตารางที่ 2

ตารางที่ 2 สรุปงานวิจัยที่ได้ศึกษาเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาค (Polytomous)

ผู้ที่ศึกษา	ประเด็นที่ศึกษา	ผลการศึกษา
French & Miller (1996)	ได้ศึกษาความเป็นไปได้ของการใช้วิธีการถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่า	เมื่อกลุ่มตัวอย่างมีขนาดเล็กลง อำนาจในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบลดลง และเมื่อพารามิเตอร์อำนาจจำแนกของข้อสอบยิ่งแตกต่างกันมาก อำนาจในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปยิ่งเพิ่มขึ้น

ตารางที่ 2 (ต่อ)

ผู้ที่ศึกษา	ประเด็นที่ศึกษา	ผลการศึกษา
Flowers et al. (1997)	ได้ศึกษาการบรรยาย DFIT ของข้อสอบที่ให้คะแนนแบบ Polytomous และประเมินรวมถึงเปรียบเทียบการทำ DFIT ในการแผ่ขยายของขั้นตอน SIBTEST และ Lord's chi-square	อัตราความคลาดเคลื่อนประเภทที่ 1 ใกล้เคียงกับระดับแอลฟายกเว้นเมื่อจำนวนข้อสอบที่มี DIF 20% และจำนวน DIF ที่มากที่สุดมีค่าสูง ปัจจัยที่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 คือค่าของพารามิเตอร์
Su & Wang (2005)	ได้จำลองข้อมูลในการสืบสอบปัจจัยที่มีผลต่อการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธีแมนเทล (Mantel) วิธีแมนเทล-แฮนส์เซลทั่วไป (Generalized Mantel-Haenzel: GMH) และวิธี Logistic Discriminant Function Analysis (LDFA)	ทั้ง 3 วิธีมีการควบคุมค่าอัตราความคลาดเคลื่อนประเภทที่ 1 ได้เป็นอย่างดี วิธี Mantel และวิธี LDFA มีอำนาจการตรวจสอบสูงกว่า วิธี LDFA
Oishi (2006)	ตรวจสอบความเท่าเทียมของการวัดความพึงพอใจด้วยแบบวัดความพึงพอใจในชีวิตระหว่างกลุ่มตัวอย่างชาวอเมริกันและชาวจีนโดยใช้ Multigroup Structural Equation Modeling (SEM), Multiple indicator multiple cause model (MIMIC) และทฤษฎีการตอบสนองข้อสอบ (IRT)	การวิเคราะห์ IRT แสดงให้เห็นความแตกต่างของค่าเฉลี่ยระหว่างกลุ่มชาวจีนและชาวอเมริกัน

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

กลุ่มที่สาม งานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนทั้งแบบทวิวิภาคและพหุวิภาค ประกอบด้วยงานวิจัย จำนวน 2 เรื่อง

Chang et al. (1996) ได้เปรียบเทียบประสิทธิภาพของวิธีซิปเทสท์ปรับใหม่กับวิธี Mantel และวิธี Standardized Mean Difference (SMD) โดยใช้การจำลองข้อมูล ในการศึกษา 2 ครั้ง การศึกษาครั้งแรกใช้การจำลองข้อมูลจากการสอบ NAEP ซึ่ง Zwick et al. (1993) ได้ศึกษาไว้ประกอบด้วยข้อสอบ 24 ข้อ เป็นข้อสอบที่ให้คะแนนแบบ 2 ค่า จำนวน 20 ข้อ และข้อสอบที่ให้คะแนนแบบหลายค่าแบบ 4 ลำดับขั้นคะแนน คือ 0, 1, 2, 3 จำนวน 4 ข้อ ส่วนวิธี Mantel และ SMD จะมีทั้งหมด 25 ข้อ ในการจำลองการตอบสนองข้อสอบใช้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ เงื่อนไขที่ศึกษาทั้งหมดมี 54 เงื่อนไข เป็นการแจกแจงความสามารถของกลุ่มสนใจ 2 แบบ คือ N(0,1) กับ N(-1,1) ทำการศึกษากับข้อสอบ 27 ข้อ [(4x2x3)+3] ได้จากประเภทของ DIF 4 ประเภท คือ Constant DIF, Low-shift DIF, High-shift DIF และ Balanced DIF, ขนาดของ DIF 2 ขนาด คือ .1 และ .25, พารามิเตอร์ความยาก 3 ค่าและข้อสอบที่เป็น null DIF อีก 3 ข้อ แต่ละเงื่อนไขกระทำซ้ำ (replications) 600 ครั้ง ผลการศึกษาพบว่า วิธี SIBTEST ตรวจสอบ DIF ได้ดี แต่วิธี Mantel และ SMD ดีกว่าวิธี SIBTEST เล็กน้อย ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของ วิธี Mantel และ SMD เป็น .049 และ .046 ตามลำดับ ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธี SIBTEST สูงกว่าเล็กน้อยเป็น .063 วิธีการทั้ง 3 สามารถตรวจสอบ DIF แบบทิศทางเดียวได้ดี

การศึกษาครั้งที่สองเป็นการศึกษาที่เพิ่มเติมจากการศึกษาครั้งแรก โดยรวมข้อสอบที่ศึกษากับอำนาจจำแนกที่แตกต่างกัน ส่วนระดับการเปลี่ยนแปลงค่าความยากเป็น .25 ปริมาณ DIFในข้อสอบที่ศึกษาขึ้นอยู่กับค่าพารามิเตอร์ความยากและอำนาจจำแนกผลการศึกษาพบว่าเมื่อพารามิเตอร์อำนาจจำแนกมีค่ามากขึ้น อัตราการปฏิเสธสมมติฐานที่เป็นกลางหรืออำนาจการทดสอบสูงขึ้นปริมาณ DIF ขึ้นอยู่กับพารามิเตอร์อำนาจจำแนกโดยตรงซึ่งเป็นไปตามที่คาดหวังเมื่อกำหนดการเปลี่ยนแปลงพารามิเตอร์ความยากไว้คงที่ วิธี SIBTEST สามารถควบคุมผลกระทบที่ก่อให้เกิดอัตราความคลาดเคลื่อนประเภทที่ 1 ได้เหนือกว่าวิธี Mantel และวิธี SMD ภายใต้เงื่อนไขทั่ว ๆ ไปของการตรวจสอบ DIF เมื่อขนาดของกลุ่มตัวอย่างเพิ่มขึ้น อำนาจในการตรวจสอบของวิธี Mantel และ SMD จะเพิ่มขึ้นอย่างรวดเร็ว ในขณะที่วิธี SIBTEST มีความคงที่พอควร

Stark et al. (2006) ได้พัฒนาและทดสอบแผนการร่วมที่ใช้ในการระบุการทำหน้าที่ต่างกันของข้อสอบรวมถึงการทดสอบอัตราส่วนไลค์ลิฮูด (LR) ซึ่งสามารถใช้ได้ทั้งการวิเคราะห์องค์ประกอบเชิงยืนยันและทฤษฎีการตอบสนองข้อสอบ โดยใช้ข้อมูลจำลองในการตรวจสอบความเที่ยงตรงของทั้งสองวิธี IRT ตั้งบนพื้นฐานของวิธี LR และ CFA ตั้งบนพื้นฐานของวิธี mean and covariance structures (MACS) โดยใช้แบบวัดมิติเดียวจำนวน 15 ข้อคำถาม มีตัวแปรที่เกี่ยวข้อง

8 ตัว จำนวนเงื่อนไขที่ใช้จำลองข้อมูลคือ 320 ($2^6 + 2^8$) แต่ละเงื่อนไขจะมีการทำซ้ำ 50 ครั้ง วิเคราะห์วิธี MACS โดยใช้โปรแกรมลิสเรล 8 และการทดสอบด้วยทฤษฎีตอบสนองข้อสอบ วิธี LR ใช้พื้นฐานของโมเดล graded response โดยใช้โปรแกรมคอมพิวเตอร์ MULTILOG ผลวิจัย พบว่า IRT วิธี LR ให้ผลดีกว่าวิธี MACS ในขณะที่กลุ่มตัวอย่างมีขนาดเล็ก ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบวัดมิติเดียวเมื่อใช้กลุ่มตัวอย่างขนาดเล็ก การวิเคราะห์ MACS ให้ผลดีกว่า หากในกรณีกลุ่มตัวอย่างมีขนาดใหญ่ และข้อมูลเป็นแบบ dichotomous มิติเดียว การตรวจสอบการทำหน้าที่ต่างกันของแบบสอบโดยใช้วิธี IRT ให้ผลดีกว่า อย่างไรก็ตามวิธี IRT หลายๆวิธีจะมีความแกร่งในการทดสอบ (robust) หากมีการฝ่าฝืนข้อตกลงเบื้องต้นในเรื่องการเป็นมิติเดียว

จากการศึกษางานวิจัยในกลุ่มที่สามสามารถสรุปวิธีที่ได้ศึกษามาแล้ว ดังตารางที่ 3

ตารางที่ 3 สรุปงานวิจัยที่ได้ศึกษาเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนทั้งแบบทวิวิภาคและพหุวิภาค

ผู้ที่ศึกษา	ประเด็นที่ศึกษา	ผลการศึกษา
Chang et al. (1996)	ได้เปรียบเทียบประสิทธิภาพของวิธีซิปเทสท์ปรับใหม่กับวิธี Mantel และวิธี Standardized Mean Difference (SMD)	วิธี SIBTEST ตรวจสอบ DIF ได้ดี แต่วิธี Mantel และ SMD ดีกว่าวิธี SIBTEST เล็กน้อย และวิธี SIBTEST สามารถควบคุมผลกระทบที่ก่อให้เกิดอัตราความคลาดเคลื่อนประเภทที่ 1 ได้เหนือกว่าวิธี Mantel และวิธี SMD ภายใต้เงื่อนไขทั่วๆ ไปของการตรวจสอบ DIF เมื่อขนาดของกลุ่มตัวอย่างเพิ่มขึ้น อำนาจในการตรวจสอบของวิธี Mantel และ SMD จะเพิ่มขึ้นอย่างรวดเร็ว
Stark et al. (2006)	ได้พัฒนาและทดสอบแผนการร่วมที่ใช้ในการระบุการทำหน้าที่ต่างกันของข้อสอบ โดยใช้วิธี MACS และวิธี LR โดยใช้ข้อมูลจำลองในการตรวจสอบความเที่ยงตรงของทั้งสองวิธี	IRT วิธี LR ให้ผลดีกว่าวิธี MACS ในขณะที่กลุ่มตัวอย่างมีขนาดเล็ก ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบวัดมิติเดียวเมื่อใช้กลุ่มตัวอย่างขนาดเล็ก การวิเคราะห์ MACS ให้ผลดีกว่า

จะเห็นว่างานวิจัยเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบในระดับบัณฑิตศึกษาในประเทศไทย เป็นการศึกษาการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบ 2 ค่า(แบบ 0-1) ยังไม่พบว่ามีผู้ใดศึกษาประเด็นของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่าเหมือนดังงานวิจัยในต่างประเทศ ดังนั้น ผู้วิจัยจึงสนใจศึกษาเกี่ยวกับวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่าระหว่างวิธี MACS และวิธี LDFA ในเงื่อนไขต่าง ๆ ซึ่งจากการศึกษาเงื่อนไข ปัจจัยที่ส่งผลต่อประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สามารถสรุปเป็นข้อ ๆ ได้ดังนี้

1) ขนาดกลุ่มตัวอย่าง (sample size)

ผลการศึกษาของ Swaminathan & Rogers (1990) พบว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและแบบอเนกรูป เมื่อเพิ่มขนาดกลุ่มตัวอย่างจะมีผลทำให้อำนาจการทดสอบของวิธีการถดถอยโลจิสติกมีค่าเพิ่มมากขึ้นเกือบทุกเงื่อนไข ส่วนอำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลจะมีค่าเพิ่มมากขึ้นเฉพาะกรณีแบบเอกรูปเท่านั้น ในขณะที่อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซลมีค่าลดลงเกือบทุกเงื่อนไข ต่อมาทั้ง 2 คนได้ศึกษาใหม่อีกครั้ง (Rogers & Swaminathan, 1993) ผลปรากฏว่าขนาดกลุ่มตัวอย่างมีผลต่ออำนาจการทดสอบของวิธีการถดถอยโลจิสติกและวิธีแมนเทล-แฮนส์เซล เมื่อขนาดกลุ่มตัวอย่างเพิ่มมากขึ้นจะทำให้อำนาจการทดสอบของทั้ง 2 วิธีมีค่าเพิ่มมากขึ้น สำหรับ Narayanan & Swaminathan (1994) ได้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป ผลการศึกษาพบว่า ขนาดกลุ่มตัวอย่างและอัตราส่วนของกลุ่มอ้างอิงต่อกลุ่มเปรียบเทียบมีผลต่ออำนาจการทดสอบของวิธีการตรวจสอบ กล่าวคือ เมื่อเพิ่มขนาดกลุ่มตัวอย่างจะทำให้อำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลและวิธีซิปเทสท์มีค่าเพิ่มมากขึ้น โดยเฉพาะเมื่อเพิ่มขนาดกลุ่มเปรียบเทียบจะมีผลทำให้อำนาจการทดสอบของทั้ง 2 วิธีมีค่าเพิ่มขึ้นมากกว่าเพิ่มขนาดกลุ่มอ้างอิง ต่อมาทั้ง 2 คนได้ศึกษาในกรณีแบบอเนกรูปอีกครั้งหนึ่ง (Narayanan & Swaminathan, 1996) พบว่าขนาดกลุ่มตัวอย่างมีผลต่ออำนาจการทดสอบของวิธีโคร-ซิป วิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติก ผลการศึกษาดังกล่าวสอดคล้องกับผลการศึกษาในครั้งแรก สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า เมื่อใช้กลุ่มตัวอย่างขนาดเล็ก อัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้ง 3 วิธีมีค่าต่ำกว่าใช้กลุ่มตัวอย่างขนาดใหญ่ ภายใต้เกือบทุกเงื่อนไขของการตรวจสอบ ต่อมา จิติมา วรรณศรี (2539) ได้เปรียบเทียบประสิทธิภาพของวิธีแมนเทล-แฮนส์เซลและวิธีซิปเทสท์ พบว่า เมื่อขนาดกลุ่มตัวอย่าง 200 และ 600 คน ทั้งสองวิธีสามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ถูกต้อง 50% แต่ถ้าขนาดกลุ่มตัวอย่าง 1,000 คน สามารถตรวจสอบได้

ถูกต้อง 100% ต่อมา Flowers et al. (1997) ได้ศึกษาการบรรยาย DFIT ของข้อสอบที่ให้คะแนนแบบ Polytomous และประเมินรวมถึงเปรียบเทียบการทำ DFIT ในการแผ่ขยายของขั้นตอน SIBTEST และ Lord's chi-square โดยใช้ข้อมูลที่จำลองขึ้นมา กลุ่มตัวอย่างจำนวน 500 และ 1,000 คน สรุปผลการวิจัยได้ว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้รับผลกระทบมาจากทุกปัจจัยในการศึกษาครั้งนี้ สำหรับ Stark et al. (2006) พบว่าการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ภายใต้การวิเคราะห์องค์ประกอบเชิงยืนยัน วิธี MACS ให้ผลดีกว่าทฤษฎีการตอบสนองข้อสอบ วิธี LR ในกรณีที่เป็นข้อสอบที่ตรวจให้คะแนนแบบหลายค่า และกลุ่มตัวอย่างขนาดเล็ก และในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบวัดมิติเดียวเมื่อใช้กลุ่มตัวอย่างขนาดเล็ก การวิเคราะห์ MACS ให้ผลดีกว่า หากในกรณีกลุ่มตัวอย่างมีขนาดใหญ่ และข้อมูลเป็นแบบ dichotomous มิติเดียว การตรวจสอบการทำหน้าที่ต่างกันของแบบสอบโดยใช้วิธี IRT ให้ผลดีกว่า

*2) **ความยาวของแบบสอบ** (test length)*

ผลการศึกษาของ Swaminathan & Rogers (1990) พบว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและแบบอเนกรูป เมื่อใช้แบบสอบที่มีความยาวมากขึ้นจะมีผลทำให้อำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติกมีค่ามากขึ้น ยกเว้นในกรณีแบบอเนกรูปของวิธีแมนเทล-แฮนส์เซล ต่อมาทั้ง 2 คนได้ทำการศึกษาใหม่อีกครั้งหนึ่ง (Rogers & Swaminathan, 1993) พบว่า ให้ผลขัดแย้งกับครั้งแรก กล่าวคือ ความยาวของแบบสอบไม่มีผลต่ออำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติกยกเว้นในกรณีแบบอเนกรูปของวิธีการถดถอยโลจิสติก ส่วนผลการศึกษาของกาญจนา วัธนสุนทร (2537) ปรากฏว่า ความยาวของแบบสอบไม่มีผลกระทบต่ออัตราการตรวจสอบของวิธีแมนเทล-แฮนส์เซล และวิธีซิปเทสท์ ซึ่งขัดแย้งกับผลการศึกษาของจิติมา วรรณศรี (2539) ที่พบว่าเมื่อใช้แบบสอบขนาด 60 ข้อ จะมีผลทำให้วิธีแมนเทล-แฮนส์เซลและวิธีซิปเทสท์มีประสิทธิภาพในการตรวจสอบดีที่สุด ส่วนผลการศึกษาของ Uttara & Milsap (1994) พบว่าเมื่อใช้แบบสอบที่มีความยาวมากขึ้นแล้วอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทส-แฮนส์เซลจะมีค่าลดลง ในขณะที่ผลการศึกษาของ Cohen & Kim (1993) พบว่า เมื่อใช้แบบสอบที่มีความยาวมากขึ้นแล้วอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ที่ตรวจสอบด้วยวิธี การวัดพื้นที่ของ Raju และวิธีการทดสอบไค-สแควร์ของ Lord จะมีค่ามากขึ้นด้วย

3) ลักษณะของข้อสอบ (type of item)

ผลการศึกษาของ Rogers & Swaminathan (1993) พบว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและอเนกรูป ลักษณะของข้อสอบมีผลต่ออำนาจการทดสอบของวิธีการถดถอยโลจิสติกและวิธีแมนเทล-แฮนส์เซล ในกรณีของเอกรูปเมื่อข้อสอบมีค่า b ปานกลางและ a สูงจะทำให้อำนาจการทดสอบทั้ง 2 วิธีมีค่าสูงสุด แต่ในกรณีแบบอเนกรูป พบว่า อำนาจการทดสอบของทั้ง 2 วิธีมีค่าสูงสุด เมื่อข้อสอบมีค่า b ปานกลางและ a สูงชนิดผสม (ข้อสอบมีค่า a และ b แตกต่างกันระหว่างผู้สอบ 2 กลุ่ม) สำหรับ Narayanan & Swaminathan (1994) ได้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป พบว่า ภายใต้เกือบทุกเงื่อนไขลักษณะของข้อสอบที่มีค่าค่า b ปานกลางและ a สูง จะทำให้อำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลและวิธีซิปเทสท์มีค่าสูงใกล้เคียงกัน ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่าลักษณะของข้อสอบไม่มีผลต่อวิธีการตรวจสอบทั้งสอง ต่อมาทั้งสองคนได้ตรวจสอบในกรณีอเนกรูป (Narayanan & Swaminathan,1996) ผลการศึกษาพบว่า เมื่อข้อสอบมีค่า a เพิ่มขึ้นจะมีผลทำให้อำนาจการทดสอบของวิธีโคร-ซิปและวิธีการถดถอยโลจิสติกมีค่าเพิ่มขึ้น และเมื่อข้อสอบมีค่า b เพิ่มขึ้นหรือลดลงจะมีผลทำให้อำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลมีค่าเพิ่มขึ้น ส่วนอัตราความคลาดเคลื่อนประเภทที่1 พบว่าเมื่อข้อสอบมีค่า a เพิ่มขึ้นจะมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่1 ของทั้ง 3 วีมีค่าเพิ่มขึ้น โดยวิธีแมนเทล-แฮนส์เซลมีค่าต่ำสุด ต่อมาเกษรหว่างจิตร (2539) ได้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซล ผลการศึกษาพบว่า ข้อสอบที่ตรวจพบว่าการทำหน้าที่ต่างกันแบบเอกรูปและแบบอเนกรูปส่วนมากเป็นข้อสอบที่มีค่า a ค่อนข้างต่ำ ทั้งวิชาภาษาไทยและวิชาภาษาอังกฤษซึ่งไม่สอดคล้องกับการศึกษาของรัชรินทร์ มุคดา (2540) ที่พบว่าข้อสอบที่ตรวจพบว่าทำหน้าที่ต่างกันแบบอเนกรูปส่วนมากเป็นข้อสอบที่มีค่า a สูง ภายใต้ลักษณะของข้อสอบ a สูงกับ b ต่ำ, a สูงกับ b ปานกลาง, a สูงกับ b สูง และกลุ่มผู้สอบที่มีความสามารถสูง ปานกลางและต่ำ และ Flowers et al. (1997) ได้ศึกษาการบรรยาย DFIT ของข้อสอบที่ให้คะแนนแบบหลายค่า โดยใช้ข้อมูลที่จำลองขึ้นมา โดยพิจารณาค่าพารามิเตอร์ a ได้รับการประเมิน สรุปผลการวิจัยได้ ปัจจัยที่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 คือค่าของพารามิเตอร์ นั่นคือหากค่า a พารามิเตอร์ต่ำอัตราความคลาดเคลื่อนประเภทที่ 1 จะสูง

4) การแจกแจงค่าความสามารถ (ability distribution)

ผลการศึกษาของ Roger & Swaminathan (1993) พบว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและอเนกรูป ความแตกต่างของการแจกแจงคะแนนระหว่างกลุ่ม

อ้างอิงและกลุ่มเปรียบเทียบไม่มีผลต่ออำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติก สำหรับ Narayanan & Swaminathan (1994) ได้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปปรากฏว่าความแตกต่างของการแจกแจงค่าความสามารถระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบไม่มีผลต่ออำนาจการทดสอบของวิธีซิปเทสท์ แต่มีผลต่ออำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซล เมื่อความแตกต่างของการแจกแจงค่าความสามารถมีค่าเพิ่มขึ้นจะมีผลให้อำนาจการทดสอบมีค่าลดลง สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า เมื่อความแตกต่างของการแจกแจงค่าความสามารถมีค่าเพิ่มขึ้นจะมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่1 ของทั้ง 2 วิธีมีค่าเพิ่มขึ้น ต่อมา Narayanan & Swaminathan (1996) ได้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป พบว่า ความแตกต่างของการแจกแจงค่าความสามารถไม่มีผลต่ออำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซล แต่จะมีผลต่ออำนาจการทดสอบของวิธีโคร-ซิปและวิธีการถดถอยโลจิสติก กล่าวคือ เมื่อความแตกต่างของการแจกแจงค่าความสามารถมีค่าเพิ่มขึ้นจะมีผลให้อำนาจการทดสอบทั้ง 2 วิธีดังกล่าวมีค่าลดลงส่วนอัตราการคลาดเคลื่อนประเภทที่1 ปรากฏว่า ภายใต้เงื่อนของความแตกต่างของการแจกแจงค่าความสามารถแบบไม่เท่ากัน อัตราความคลาดเคลื่อนประเภทที่1 ของทั้ง 3 วิธีมีค่าสูงกว่าภายใต้เงื่อนไขแบบเท่ากันโดยวิธีโคร-ซิปมีค่าสูงสุดทั้งสองเงื่อนไข

5) สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน (propotion of DIF items)

ผลการศึกษาของ Rogers & Swaminathan (1993) พบว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและแบบอเนกรูป สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันไม่มีผลต่ออำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติก ยกเว้นในกรณีแบบเอกรูปของวิธีการถดถอยโลจิสติก เมื่อสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบมีจำนวนลดลงจะมีผลทำให้อำนาจการทดสอบของวิธีการถดถอยโลจิสติกมีค่าเพิ่มมากขึ้น ต่อมา Narayanan & Swaminathan (1994) ได้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบกรณีแบบเอกรูป ผลการศึกษาพบว่า เมื่อสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบมีจำนวนลดลงจะมีผลทำให้อำนาจการทดสอบของวิธีซิปเทสท์และวิธีการถดถอยโลจิสติกมีค่าเพิ่มมากขึ้น ทั้งยังส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้ง 2 วิธีมีค่าเพิ่มมากขึ้นด้วย ต่อมา Narayanan & Swaminathan (1996) ได้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบอีกครั้งหนึ่ง โดยตรวจสอบในกรณีแบบอเนกรูป ผลการศึกษาพบว่า เมื่อสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบมีจำนวนลดลงจะมีผลทำให้อำนาจการทดสอบของวิธีการถดถอยโลจิสติกมีค่าเพิ่มขึ้น แต่จะไม่มีผลต่ออำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลและวิธีโคร-ซิป สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า เมื่อสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบมีจำนวนเพิ่มมากขึ้นจะส่งผลให้อัตรา

ความคลาดเคลื่อนประเภทที่ 1 ของทั้ง 3 วิธีจะมีค่าเพิ่มมากขึ้นด้วย โดยที่วิธีแมนเทล-แฮนส์เซล มีค่าต่ำที่สุด และ Flowers et al. (1997) ได้ศึกษาการบรรยาย DFIT ของข้อสอบที่ให้คะแนนแบบหลายค่าโดยจำลองข้อมูลในปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน (0%, 10%, และ 20%) สรุปผลการวิจัยได้ว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ใกล้เคียงกับระดับแอลฟายกเว้นเมื่อจำนวนข้อสอบที่มี DIF 20% และจำนวน DIF ที่มากที่สุดมีค่าสูง

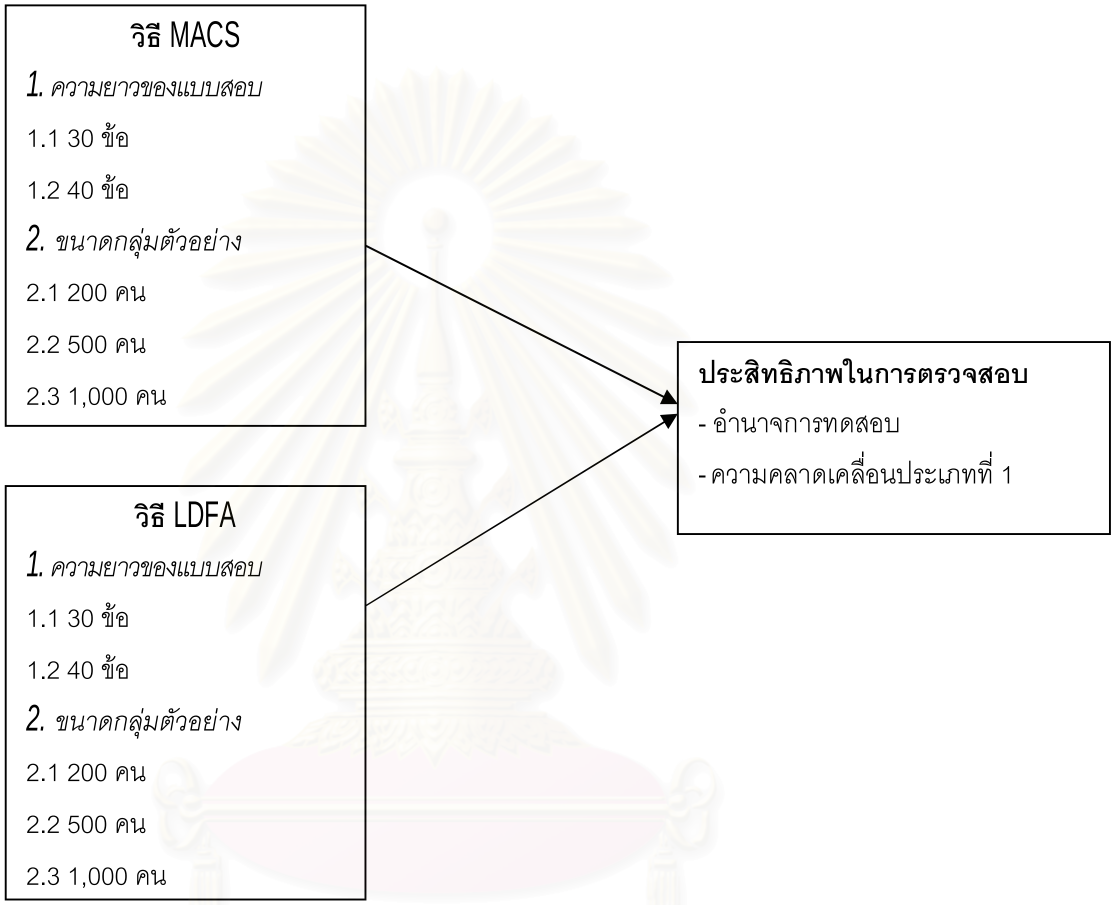
6) ขนาดอิทธิพลของข้อสอบที่ทำหน้าที่ต่างกัน (DIF effect size)

ผลการศึกษาของ Rogers & Swaminathan (1993) พบว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและแบบอเนกรูป ขนาดอิทธิพลของข้อสอบที่ทำหน้าที่ต่างกัน (ค่าพื้นที่ระหว่างโค้งลักษณะข้อสอบของผู้สอบ 2 กลุ่ม) มีผลต่ออำนาจการทดสอบของวิธีการถดถอยโลจิสติกและวิธีแมนเทล-แฮนส์เซล เมื่อขนาดอิทธิพลของข้อสอบที่ทำหน้าที่ต่างกันมีค่าเพิ่มขึ้นจะส่งผลให้อำนาจการทดสอบของทั้ง 2 วิธีมีค่าเพิ่มมากขึ้น สำหรับ Narayanan & Swaminathan (1994) ได้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป ปรากฏว่า ขนาดอิทธิพลของข้อสอบที่ทำหน้าที่ต่างกันมีผลต่ออำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซล เมื่อขนาดอิทธิพลของข้อสอบที่ทำหน้าที่ต่างกันมีค่าเพิ่มขึ้น อำนาจการทดสอบของทั้ง 2 วิธีมีค่าเพิ่มขึ้น ต่อมา Narayanan & Swaminathan (1996) ได้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบอีกครั้งหนึ่ง โดยศึกษาในกรณีแบบอเนกรูป ปรากฏว่า ขนาดอิทธิพลของข้อสอบที่ทำหน้าที่ต่างกันมีผลต่ออำนาจการทดสอบของวิธีการถดถอยโลจิสติก วิธีแมนเทล-แฮนส์เซล และวิธีโคร-ซิป เมื่อขนาดอิทธิพลของข้อสอบที่ทำหน้าที่ต่างกันมีค่าเพิ่มขึ้น อำนาจการทดสอบของทั้ง 3 วิธีจะมีค่าเพิ่มขึ้น ซึ่งสอดคล้องกับผลการศึกษาในครั้งแรก

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตอนที่ 3 กรอบแนวคิดในการวิจัย

แผนภาพที่ 4 กรอบแนวคิดในการวิจัย



กรอบแนวคิดในการวิจัย แสดงให้เห็นว่าการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี MACS ซึ่งตั้งอยู่บนพื้นฐานของการวิเคราะห์องค์ประกอบเชิงยืนยันและวิธี LDFA ซึ่งตั้งอยู่บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ ส่งผลต่อประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาค ได้แก่ อำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 โดยทั้งสองวิธีมีความสามารถในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบพหุวิภาค

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 3
วิธีดำเนินการวิจัย

วิธีดำเนินการวิจัย

การวิจัยเรื่อง การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาค ด้วยวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ยกับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก เป็นการวิจัยเชิงบรรยาย โดยมีวัตถุประสงค์ของการวิจัยเพื่อเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาค ระหว่างวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ยกับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก โดยกำหนดวัตถุประสงค์เฉพาะในการศึกษาด้านต่าง ๆ ดังนี้ 1) เพื่อเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 2 วิธี ภายใต้เงื่อนไขของปัจจัยที่แตกต่างกันทางด้านความยาวของแบบสอบและขนาดกลุ่มตัวอย่าง 2) เพื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 2 วิธี ภายใต้เงื่อนไขของปัจจัยที่แตกต่างกันทางด้านความยาวของแบบสอบและขนาดกลุ่มตัวอย่าง

ผู้วิจัยใช้ข้อมูลในการศึกษา เป็นข้อมูลจริงภายใต้การใช้คะแนนที่ได้จากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ด้านเหตุผล 5 ตัวเลือกที่มีคำตอบถูกหลายคำตอบ และให้ตัดสินทุกคำตอบ (Multiple True-false Answer) เรื่องจำนวนเต็ม เศษส่วน และทศนิยม ระดับชั้นมัธยมศึกษาตอนต้น คะแนนเต็มข้อละ 5 คะแนน จำนวน 50 ข้อ ของ สุนิสา จุ้ยม่วงศรี (2546) มาจัดกระทำ เมื่อแบ่งกลุ่มผู้สอบตามตัวแปรเพศ และวิเคราะห์ข้อมูลด้วยโปรแกรมคอมพิวเตอร์ภายใต้ความแตกต่างตามเงื่อนไขตัวแปรต้นในด้านความยาวของแบบสอบและขนาดกลุ่มตัวอย่าง โดยมีขั้นตอนดำเนินการวิจัย ดังนี้

1. การกำหนดประชากรและกลุ่มตัวอย่าง
2. แบบสอบที่ใช้ในการวิเคราะห์
3. การเก็บรวบรวมข้อมูล
4. การวิเคราะห์ข้อมูล

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

1. ประชากรและกลุ่มตัวอย่าง

ประชากร

ประชากรที่ใช้ในการวิจัยครั้งนี้ ได้แก่ นักเรียนชั้นมัธยมศึกษาปีที่ 2 เนื่องจากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ เรื่องจำนวนเต็ม เศษส่วน และทศนิยม ของ สุนิสา จุ้ยม่วงศรี (2546) เป็นเนื้อหาที่อยู่ในระดับชั้นมัธยมศึกษาตอนต้น และจากการศึกษาหลักสูตรของโรงเรียนในกรุงเทพมหานคร พบว่า เนื้อหาในส่วนนี้จะทำการสอนกันในระดับชั้นมัธยมศึกษาปีที่ 1 ภาคการศึกษาปลาย และเนื่องจากผู้วิจัยจะลงเก็บข้อมูลในช่วงภาคปลาย จึงมีความจำเป็นต้องเลือกกลุ่มประชากรเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ที่ผ่านการเรียนการสอนในเนื้อหาดังกล่าวมาแล้ว ดังนั้นกลุ่มประชากรจึงเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ปีการศึกษา 2550 โรงเรียนมัธยมศึกษา กรุงเทพมหานคร (สาเหตุที่เลือกกรุงเทพมหานครเท่านั้น เนื่องจากเพื่อป้องกันการทำหน้าที่ต่างกันที่มาจากตัวแปรสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน) สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (สพฐ.) จำนวน 116 โรงเรียน รวมนักเรียน 54,990 คน (สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ, 2550) โดยแบ่งเป็น นักเรียนชาย 27,954 คน และนักเรียนหญิง 27,036 คน

กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ คือ นักเรียนชั้นมัธยมศึกษาปีที่ 2 ปีการศึกษา 2550 โรงเรียนมัธยมศึกษา สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กรุงเทพมหานคร จำนวน 2,000 คน แบ่งเป็นนักเรียนชาย 1,000 คน และนักเรียนหญิง 1,000 คน ตามเงื่อนไขขนาดกลุ่มตัวอย่างที่ได้กำหนดไว้เพื่อใช้เป็นประชากรเทียมในการสุ่มกลุ่มตัวอย่างขนาดต่าง ๆ

กลุ่มตัวอย่างได้มาจากประชากรที่มีการสุ่มแบบหลายขั้นตอน (Multi-stage Random Sampling) โดยมีขั้นตอนการสุ่ม ดังนี้

ขั้นตอนแรก สุ่มโดยใช้สำนักงานเขตพื้นที่การศึกษาซึ่งมีทั้งหมด 3 เขต คือเขต 1, เขต 2 และเขต 3 เป็นหน่วยในการสุ่ม โดยใช้การสุ่มอย่างง่าย (Simple Random Sampling) โดยการจับสลาก ได้สำนักงานเขตพื้นที่การศึกษากรุงเทพมหานคร เขต 1 ซึ่งมีโรงเรียนที่มีชั้นมัธยมศึกษาปีที่ 2 จำนวน 39 โรงเรียน รวมนักเรียนจำนวน 15,982 คน

ขั้นตอนที่สอง สุ่มโรงเรียนในเขตพื้นที่การศึกษากรุงเทพมหานครเขต 1 ผู้วิจัยใช้การสุ่มแบบแบ่งชั้น (Stratified Random Sampling) โดยใช้ขนาดของโรงเรียนเป็นชั้น (strata) มีเกณฑ์จำแนกขนาดของโรงเรียนดังนี้ (1) โรงเรียนขนาดใหญ่ มีจำนวนนักเรียน 2,500 คนขึ้นไป (2) โรงเรียนขนาดกลาง มีจำนวนนักเรียน 1,500-2,500 คน และ (3) โรงเรียนขนาดเล็ก มีจำนวน

นักเรียนน้อยกว่า 1,500 คน (ศูนย์ปฏิบัติการสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน, 2550) การวิจัยครั้งนี้ผู้วิจัยใช้การสุ่มกลุ่มตัวอย่างแบบเป็นสัดส่วน รายละเอียดดังตารางที่ 4

ตารางที่ 4 ขนาดของกลุ่มประชากร และกลุ่มตัวอย่างจำแนกตามขนาดของโรงเรียน

เขตพื้นที่	ขนาดของประชากรโรงเรียน				ขนาดของกลุ่มตัวอย่างโรงเรียน			
	ใหญ่	กลาง	เล็ก	รวม	ใหญ่	กลาง	เล็ก	รวม
เขตพื้นที่กรุงเทพมหานครเขต1	14	14	11	36	4	4	3	11
รวม	14	14	11	36	4	4	3	11

ขั้นที่สาม สุ่มโรงเรียนตามจำนวนที่กำหนดในขั้นตอนที่สอง โดยใช้วิธีสุ่มอย่างง่าย (Simple Random Sampling) โดยการจับสลาก รวมทั้งสิ้น 11 โรงเรียน

ขั้นที่สี่ สุ่มนักเรียนตามเพศจากแต่ละโรงเรียน โดยสุ่มนักเรียนภายในโรงเรียนที่ได้รับการสุ่มมาแล้วในขั้นที่สาม เพื่อให้ได้กลุ่มตัวอย่างตามจำนวนที่ต้องการ นั่นคือ สุ่มนักเรียนชายและหญิงในโรงเรียนเป็นจำนวนกึ่งหนึ่งของจำนวนนักเรียนทั้งหมด รายชื่อโรงเรียน จำนวนห้องเรียน และจำนวนนักเรียน แสดงดังตารางที่ 5

ตารางที่ 5 รายชื่อโรงเรียน จำนวนห้องเรียน และจำนวนนักเรียนที่สุ่มได้

ชื่อโรงเรียน	จำนวนห้องเรียน	จำนวนนักเรียนชั้นมัธยมศึกษาปีที่ 2 (คน)		
		ชาย	หญิง	รวม
ขนาดใหญ่				
1. สันติราษฎร์วิทยาลัย	12	250	249	499
2. เทพศิรินทร์	10	532	0	532
3. ศรีอยุธยา	10	256	338	594
4. นนทรีวิทยา	12	346	267	613
ขนาดกลาง				
1. เบญจมราชาลัย	8	0	429	429
2. วัดราชาธิวาส	10	162	229	391
3. สตรีมหาพฤฒาราม ในพระบรมราชินูปถัมภ์	10	0	338	338
4. ยานนาเวศวิทยาคม	12	269	207	476

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 5 (ต่อ)

ชื่อโรงเรียน	จำนวนห้องเรียน	จำนวนนักเรียนชั้นมัธยมศึกษาปีที่ 2 (คน)		
		ชาย	หญิง	รวม
ขนาดเล็ก				
1. ไตรมิตรวิทยาลัย	6	259	0	259
2. พุทธจักรวิทยา	7	115	132	247
3. มัธยมวัดเบญจมบพิตร	6	193	0	193
รวม	103	2,382	2,189	4,571

จากการเก็บข้อมูลจริงผู้วิจัยได้ทำการสุ่มห้องเรียนภายในโรงเรียนที่ได้รับการสุ่มมาแล้ว ดังรายชื่อตามตารางด้านบน เพื่อให้ได้กลุ่มตัวอย่างตามจำนวนที่ต้องการ นั่นคือสุ่มห้องเรียนในแต่ละโรงเรียนเป็นจำนวนกึ่งหนึ่งของจำนวนห้องเรียนทั้งหมด แต่ปรากฏว่าโรงเรียนศรีอยุธยาไม่สามารถให้ผู้วิจัยทำการเก็บรวบรวมข้อมูลได้ ผู้วิจัยจึงสุ่มโรงเรียนสายปัญญาในพระบรมราชูปถัมภ์ และโรงเรียนสายน้ำผึ้งในพระบรมราชินูปถัมภ์มาเป็นกลุ่มตัวอย่างแทน รวมเป็น 12 โรงเรียน รวมกลุ่มตัวอย่างทั้งหมด 2,098 คน โดยแบ่งเป็นชาย 1,045 คน และหญิง 1,053 คน จากนั้นผู้วิจัยได้นำมาคัดเลือกโดยเลือกกระดาษคำตอบของนักเรียนที่มีความบกพร่องและไม่สมบูรณ์ในเรื่องต่อไปนี้ออก ได้แก่ ทำข้อสอบไม่ครบทุกข้อ เห็นได้ชัดว่ามีการเดาข้อสอบ เช่น กาเครื่องหมาย 1 ทุกข้อ หรือกาเครื่องหมาย 1 สลับกับ 0 อย่างมีระบบ

ดังนั้น จากการคัดเลือกจะได้กระดาษคำตอบของนักเรียนที่มีความสมบูรณ์ สามารถนำมาวิเคราะห์ได้ทั้งหมด 1,806 คน โดยแบ่งเป็นชาย 892 คน และหญิง 914 คน คิดเป็นร้อยละ 86.08

ดังนั้นรายชื่อโรงเรียน จำนวนห้องเรียน และจำนวนนักเรียนที่เก็บข้อมูลจริงเป็นดังตารางด้านล่าง

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 6 รายชื่อโรงเรียน จำนวนห้องเรียน และจำนวนนักเรียนที่ใช้ในการเก็บข้อมูลจริง

ชื่อโรงเรียน	จำนวนห้องเรียน	จำนวนนักเรียนชั้นมัธยมศึกษาปีที่ 2 (คน)		
		ชาย	หญิง	รวม
ขนาดใหญ่				
สันติราษฎร์วิทยาลัย	5	78	105	183
เทพศิรินทร์	5	251	-	251
นนทรีวิทยา	5	124	115	239
สายน้ำผึ้งในพระบรมราชินูปถัมภ์	2	-	87	87
ขนาดกลาง				
เบญจมราชาลัย	4	-	213	213
วัดราชาธิวาส	4	59	59	118
สตรีมหาพฤฒารามในพระบรมราชินูปถัมภ์	3	-	91	91
ยานนาเวศวิทยาคม	5	86	84	170
สายปัญญาในพระบรมราชูปถัมภ์	2	-	88	88
ขนาดเล็ก				
ไตรมิตรวิทยาลัย	3	99	-	99
พุทธจักรวิทยา	3	30	72	102
มัธยมวัดเบญจมบพิตร	6	165	-	165
รวม	47	892	914	1,806

2. แบบสอบที่ใช้ในการวิเคราะห์

แบบสอบที่นำมาใช้ในการศึกษาครั้งนี้ คือ แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ด้านเหตุผล 5 ตัวเลือกที่มีคำตอบถูกหลายคำตอบ และให้ตัดสินทุกคำตอบ (Multiple True-false Answer) เรื่องจำนวนเต็ม เศษส่วน และทศนิยม ระดับชั้นมัธยมศึกษาตอนต้น คะแนนเต็มข้อละ 5 คะแนน จำนวน 50 ข้อ ของ สุนิสา จุ้ยม่วงศรี (2546) โดยผู้วิจัยใช้การสุ่มเลือกตามเนื้อหา โดยเลือกให้มีเนื้อหาครอบคลุมทุกเรื่อง และตัดข้อที่มีเนื้อหาใกล้เคียงกันออกไป จนมีจำนวนข้อสอบ 40 ข้อ และทำเป็นแบบทดสอบฉบับสมบูรณ์ไปทดสอบกับนักเรียนที่เป็นกลุ่มตัวอย่าง

สำหรับแบบทดสอบขนาด 30 ข้อ ที่ใช้ในการศึกษา ใช้ข้อมูลผลการตอบข้อสอบจากแบบทดสอบที่ประกอบด้วยข้อสอบ 40 ข้อ ของนักเรียนกลุ่มตัวอย่างที่กำหนดเป็นประชากรเทียมมาคัดเลือกข้อสอบโดยการสุ่มอย่างง่าย

ลักษณะของแบบสอบที่ใช้ในการวิเคราะห์

เครื่องมือที่ใช้ในการวิจัยครั้งนี้ เป็นแบบทดสอบเลือกตอบที่มีคำตอบถูกหลายคำตอบ และให้ตัดสินทุกคำตอบ (Multiple True-false Answer) วัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ด้านเหตุผล ซึ่งวัดความสามารถด้านความคิดรวบยอดทางคณิตศาสตร์ที่เกี่ยวข้องกับหลักการ วิธีการ และมโนภาพในการคิดแบบนามธรรม ในเรื่องจำนวนเต็ม เศษส่วน และทศนิยม ระดับชั้นมัธยมศึกษาตอนต้น จำนวน 1 ฉบับ ประกอบด้วยข้อสอบชนิด 5 ตัวเลือก คะแนนเต็มข้อละ 5 คะแนน จำนวน 50 ข้อ ของ สุนิสา จุ้ยม่วงศรี (2546) ซึ่งมีลักษณะดังนี้

<u>ตัวอย่างข้อสอบ</u>

ข้อ (A) ให้ a และ b เป็นจำนวนเต็ม ค่าของ a และ b ในข้อใด ทำให้ผลลัพธ์ของ a-b เป็น จำนวนเต็มลบ

ก. a เป็นจำนวนเต็มลบและ b เป็นศูนย์

ข. a และ b เป็นจำนวนเต็มลบ โดยที่ a อยู่ห่างจากศูนย์มากกว่า b

ค. a และ b เป็นจำนวนเต็มบวก โดยที่ a อยู่ห่างจากศูนย์มากกว่า b

ง. a เป็นจำนวนเต็มลบและ b เป็นจำนวนเต็มบวก

จ. a เป็นศูนย์และ b เป็นจำนวนเต็มบวก

(คำตอบถูก คือ ก, ข, ง และ จ)

<u>ตัวอย่างการตอบ</u>

ข้อ	ก	ข	ค	ง	จ
(A)	**P**	**û**	**û**	**P**	**P**

ข้อ (A) ได้ 3 คะแนน

<u>เกณฑ์การตรวจให้คะแนน</u>

การตรวจให้คะแนนของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ เรื่องจำนวนเต็ม เศษส่วน และทศนิยม ถ้าตอบตรงเฉลยทุกตัวเลือก ให้ 5 คะแนน ถ้าตอบตรงกับเฉลยเพียง 4 ตัวเลือก ให้ 4 คะแนน ถ้าตอบตรงกับเฉลยเพียง 3 ตัวเลือก ให้ 3 คะแนน ถ้าตอบตรงกับเฉลยเพียง 2 ตัวเลือก ให้ 2 คะแนน ถ้าตอบตรงกับเฉลยเพียง 1 ตัวเลือก ให้ 1 คะแนน และถ้าตอบไม่ตรงกับเฉลยทุกตัวเลือกให้ 0 คะแนน

การตรวจสอบคุณภาพของแบบสอบ

แบบสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ของ สุนิสา จุ้ยม่วงศรี (2546) ผ่านการตรวจสอบคุณภาพด้านตรงเชิงเนื้อหา ค่าความยาก ค่าอำนาจจำแนกและค่าความเที่ยง ดังนี้

การตรวจสอบความตรงเชิงเนื้อหา

ผู้วิจัยนำข้อสอบที่สร้างขึ้นไปให้ผู้เชี่ยวชาญด้านการสอนคณิตศาสตร์ จำนวน 6 ท่าน ตรวจสอบความสอดคล้องระหว่างข้อสอบกับจุดประสงค์เชิงพฤติกรรมและพฤติกรรมการวัดตามวิธีของ Rovinelly & Hamberton พร้อมทั้งตรวจสอบความถูกต้องของภาษาที่ใช้ และเฉลยการตอบข้อสอบ

จากการนำข้อสอบไปให้ผู้เชี่ยวชาญตรวจสอบ แล้วนำคะแนนความคิดเห็นมาคำนวณค่า IOC จะได้ค่าคะแนนเฉลี่ยจากผลการตัดสินของผู้เชี่ยวชาญอยู่ในช่วง 0.905 ถึง 1.00 ข้อสอบที่มีค่าเฉลี่ยไม่ถึง 0.5 จำนวน 12 ข้อ ได้มีการปรับปรุงทั้งข้อคำถามและตัวเลือก แล้วนำไปให้ผู้เชี่ยวชาญตรวจสอบอีกครั้งหนึ่ง โดยค่า IOC เฉลี่ยทั้งฉบับของแบบทดสอบฉบับ X และ Y ของแบบทดสอบมีค่า 0.905 และ 0.900 ตามลำดับ

การตรวจสอบค่าความยากและอำนาจจำแนก

1) คุณภาพของข้อสอบตามแนวทฤษฎีการทดสอบแบบมาตรฐานเดิม เพื่อหาค่าความยาก (p) และอำนาจจำแนก (r) โดยใช้โปรแกรม SPSS version 7.5 ได้ค่าความยากของข้อสอบในช่วง 0.450 ถึง 0.874 และช่วง 0.478 ถึง 0.805 ในแบบทดสอบฉบับ X และ Y ตามลำดับ ค่าอำนาจจำแนกในช่วง 0.005 ถึง 0.625 และช่วง -0.147 ถึง 0.618 ในแบบทดสอบฉบับ X และ Y ตามลำดับ

2) นำแบบทดสอบทั้งฉบับมาตรวจให้คะแนนตามเกณฑ์ที่กำหนด พร้อมทั้งวิเคราะห์คุณภาพของข้อสอบตามทฤษฎีการทดสอบแบบมาตรฐานเดิม ได้แก่ ค่าความยาก และอำนาจจำแนก (r) โดยใช้โปรแกรม SPSS version 7.5 ได้ค่าสัมประสิทธิ์ความเชื่อมั่นของแบบทดสอบทั้งฉบับ โดยใช้สูตรของแอลฟา ครอนบาค เท่ากับ .936 และ .918 เมื่อวิเคราะห์คุณภาพของข้อสอบตามทฤษฎีการทดสอบแบบมาตรฐานเดิม แบบทดสอบทั้งสองฉบับมีค่าความยากอยู่ในช่วง 0.465 ถึง 0.936 และ 0.471 ถึง 0.927 ตามลำดับ และมีค่าอำนาจจำแนกอยู่ในช่วง 0.102 ถึง 0.658 และช่วง 0.111 ถึง 0.631 และเมื่อวิเคราะห์ข้อสอบตามทฤษฎีการตอบข้อสอบในโมเดลGRM โดยใช้โปรแกรม MULTILOG version 7.03 ได้แก่ ค่าความยาก ค่าอำนาจจำแนก และค่าสารสนเทศของข้อสอบแต่ละข้อ แบบทดสอบทั้งสองฉบับมีค่าอำนาจจำแนกไม่ติดลบ ซึ่งถือว่าอยู่ในเกณฑ์ที่มีคุณภาพ คือ มีอำนาจจำแนกอยู่ในช่วง 0.380 ถึง 2.180 และอยู่ในช่วง 0.470 ถึง 2.520 ตามลำดับ และส่วนใหญ่สอดคล้องกับค่าอำนาจจำแนกที่วิเคราะห์ได้จากทฤษฎีการทดสอบแบบ

มาตรฐานเดิม สำหรับค่าความยากของข้อสอบทั้งสองฉบับเรียงตามลำดับขั้นความยากจากน้อยไปหามาก

3. การเก็บรวบรวมข้อมูล

ผู้วิจัยดำเนินการ ดังนี้

1. ทำหนังสือจากคณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย และติดต่อขอความร่วมมือจากผู้อำนวยการโรงเรียนมัธยมศึกษาในสำนักงานเขตพื้นที่การศึกษากรุงเทพมหานคร เขต1 ที่เป็นกลุ่มตัวอย่าง เพื่อกำหนดวันและเวลาสอบ

2. นำหนังสือขอความร่วมมือในการเก็บรวบรวมข้อมูลไปติดต่อ ประสานงานกับโรงเรียนที่เป็นกลุ่มตัวอย่าง โดยติดต่อกับผู้อำนวยการ และครูผู้สอนวิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 2 เพื่อชี้แจงจุดประสงค์ของการศึกษา ขอบเขตเนื้อหาที่ต้องการวัด และรูปแบบการวัดหรือลักษณะข้อสอบที่ใช้วัด

3. ดำเนินการสอบโดยนำแบบทดสอบที่จัดเตรียมไว้จำนวน 40 ข้อ ไปทดสอบกับกลุ่มตัวอย่าง ก่อนการสอบควรชี้แจงให้นักเรียนกลุ่มตัวอย่างทราบถึงวัตถุประสงค์ของการสอบและวิธีทำข้อสอบโดยโน้มน้าวให้นักเรียนตั้งใจทำข้อสอบ เพื่อให้ได้ผลที่เชื่อถือได้

ทั้งนี้ ผู้วิจัยเป็นผู้เก็บรวบรวมข้อมูลด้วยตนเอง

4. การวิเคราะห์ข้อมูล

ผู้วิจัยดำเนินการวิเคราะห์ข้อมูลตามลำดับ ดังนี้

1. คำนวณค่าสถิติพื้นฐาน เพื่อบรรยายลักษณะการแจกแจงของคะแนนจากแบบสอบวิชาคณิตศาสตร์ ด้วยโปรแกรมคอมพิวเตอร์ SPSS 13.0 for Windows

2. ตรวจสอบความเป็นเอกมิติ (Unidimensionality) ของแบบสอบ โดยการวิเคราะห์องค์ประกอบ (Factor Analysis) ด้วยการวิเคราะห์ตัวประกอบสำคัญ (Principal Component Analysis) และหมุนแกนด้วยวิธี Varimax ตามวิธีการของ Lord (1980) เพื่อตรวจสอบความเป็นมิติเดียวของแบบทดสอบ ซึ่งเป็นข้อตกลงเบื้องต้นของการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบตามวิธีการของทฤษฎีตอบสนองข้อสอบ (IRT) โดยใช้โปรแกรม SPSS 13.0 for Windows

3. วิเคราะห์คุณภาพของแบบสอบวิชาคณิตศาสตร์ ด้วยโปรแกรมคอมพิวเตอร์ SPSS 13.0 for Windows และโปรแกรม MULTILOG 7.03 โดยคำนวณค่าความเที่ยงแบบความสอดคล้องภายใน (Internal Consistency) ของแบบสอบแต่ละฉบับด้วยสูตรแอลฟาของครอนบาค (Cronbach’s Alpha) ดังนี้

$$\text{Alpha} = \left[\frac{n}{n-1}\right]\left[1-\frac{\sum_{i=1}^{n} S_i^2}{S_x^2}\right]$$

เมื่อ n แทน จำนวนข้อสอบ

S_x^2 แทน ความแปรปรวนของการตอบในกลุ่มผู้สอบ

S_i^2 แทน ความแปรปรวนรายข้อ

4. ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GPCM

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในการศึกษาครั้งนี้มี 3 วิธี คือ วิธี GPCM วิธี MACS และวิธี LDFA โดยจะใช้วิธี MACS และวิธี LDFA วิเคราะห์ในทุกเงื่อนไขของขนาดกลุ่มตัวอย่างและความยาวของแบบทดสอบ เพื่อเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) แต่วิธี GPCM จะใช้เป็นวิธีมาตรฐานในการระบุว่าข้อสอบข้อใดแสดงการทำหน้าที่ต่างกัน เพื่อใช้เป็นเกณฑ์ในการประเมินประสิทธิภาพการตรวจสอบของวิธีอื่น โดยวิธี GPCM จะใช้วิเคราะห์ในเงื่อนไขที่กลุ่มตัวอย่างเพศชายขนาด 1,000 คน ต่อกลุ่มตัวอย่างเพศหญิงขนาด 1,000 คน กับความยาวของแบบทดสอบ 2 ขนาด คือ 40 ข้อ และ 30 ข้อ เนื่องจากวิธี GPCM ซึ่งเป็นวิธีตามทฤษฎีการตอบสนอง (IRT) จะตรวจสอบได้อย่างมีประสิทธิภาพเมื่อใช้กลุ่มตัวอย่างขนาดใหญ่ ดังนั้น วิธีการนี้จึงวิเคราะห์เพียง 2 เงื่อนไขดังกล่าว มีขั้นตอนการวิเคราะห์ดังนี้

4.1. การประมาณค่าพารามิเตอร์ข้อสอบของกลุ่มตัวอย่างเพศชายและเพศหญิงกลุ่มละ 1,000 คน ที่สุ่มมาจากประชากรเทียมของแต่ละกลุ่มเพศ เมื่อแบบทดสอบมีขนาด 40 ข้อและ 30 ข้อ โดยใช้โปรแกรม PARSCALE Version 2.2 พัฒนาขึ้นโดย Muraki & Bock (1991) ซึ่งใช้วิเคราะห์ข้อสอบที่ตรวจให้คะแนนแบบหลายค่า เพื่อประมาณค่าพารามิเตอร์ของข้อสอบ ตามแนวทฤษฎีการตอบข้อสอบ (IRT) ผลการวิเคราะห์จะได้แฟ้มข้อมูลของค่าประมาณพารามิเตอร์ข้อสอบของผู้สอบกลุ่มเพศชายและหญิง เนื่องจากพารามิเตอร์ของข้อสอบดังกล่าวประมาณค่ามาจากการวิเคราะห์แยกกันระหว่างกลุ่มผู้สอบย่อย ดังนั้น ก่อนที่จะนำข้อมูลของทั้งสองกลุ่มมาตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจะต้องปรับเทียบสเกลพารามิเตอร์ของข้อสอบให้อยู่บนสเกลเมตริกซ์เดียวกัน

$$a^*_{jF} = a_{jF} / A$$

$$b^*_{jF} = Ab_{jF} + K$$

a^*_{jF} แทน ค่าอำนาจแจกแจงของข้อสอบข้อที่ j จากผู้สอบกลุ่มสนใจที่แปลงค่าแล้ว

b^*_{jF} แทน ค่าความยากของข้อสอบข้อที่ j จากผู้สอบกลุ่มสนใจที่แปลงค่าแล้ว

a_{jF} แทน ค่าอำนาจแจกแจงของข้อสอบข้อที่ j จากผู้สอบกลุ่มสนใจที่ยังไม่ได้แปลงค่า

b_{jF} แทน ค่าความยากของข้อสอบข้อที่ j จากผู้สอบกลุ่มสนใจที่ยังไม่ได้แปลงค่า

A แทน ค่าความชัน (Slope)

K แทน ค่าจุดตัด (Intercept)

4.2 ปรับเทียบสเกลพารามิเตอร์ข้อสอบของทั้งสองกลุ่ม ด้วยการแปลงค่าเชิงเส้น (Linear Transformation) โดยใช้สูตรในการปรับเทียบ (Tate, 1999)

4.3 นำผลการปรับเทียบสเกลพารามิเตอร์ในข้อ 4.1.2 มาตรวจสอบการหน้าที่ต่างกันของข้อสอบตามวิธี GPCM เพื่อใช้เป็นวิธีมาตรฐานในการระบุว่าข้อสอบข้อใดทำหน้าที่ต่างกัน ซึ่งสามารถตรวจการทำหน้าที่ต่างกันของข้อสอบได้ 3 วิธี คือ วิธีการทดสอบไค-สแควร์ของลอร์ด (Lord's Chi-square, 1980) วิธีการวัดพื้นที่สัมบูรณ์ของราชู (Exact Area Measures) (Raju, 1990) และวิธีการวัดพื้นที่ช่วงปิด ในการวิจัยครั้งนี้ผู้วิจัยเลือกใช้วิธีการทดสอบไค-สแควร์ของลอร์ด (Lord's Chi-square, 1980) โดยมีสูตรดังนี้

$$c^2 = \sum \frac{(O-E)^2}{E}$$

เมื่อ O คือ ค่าที่ได้จากการสังเกตการณ์

เมื่อ E คือ ค่าที่ควรจะเป็นตามทฤษฎี

5. ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ จำแนกตามตัวแปรเพศซึ่งดำเนินการดังนี้

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในการศึกษาครั้งนี้มี 2 วิธี คือวิธี MACS และวิธี LDFA โดยจะวิเคราะห์ในทุกเงื่อนไขของขนาดกลุ่มตัวอย่างและความยาวของแบบสอบเพื่อเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF)

5.1 นำข้อมูลผลการตอบข้อสอบรายข้อที่เตรียมไว้ มาประมาณค่าพารามิเตอร์ของข้อสอบรายข้อ ค่าความสามารถของผู้สอบแต่ละคนและค่าความแปรปรวนร่วมของข้อมูล ด้วยโปรแกรม MULTILOG version 7 โดยจะวิเคราะห์แยกกันระหว่างกลุ่มอ้างอิง และกลุ่มเปรียบเทียบ ผลการวิเคราะห์จะได้แฟ้มข้อมูลของค่าประมาณพารามิเตอร์ข้อสอบของผู้สอบกลุ่มเพศชายและเพศหญิง

5.2 ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วย วิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย (Mean and Covariance Structures Method: MACS)

นำข้อมูลผลการตอบกลุ่มเพศชายและเพศหญิงในเงื่อนไขความยาวแบบสอบและกลุ่มตัวอย่างขนาดเดียวกันแต่ละคู่มาดำเนินการวิเคราะห์เพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี MACS โดยจับคู่นักเรียนเพศชายและเพศหญิงที่มีความสามารถเท่ากัน ซึ่งพิจารณาจากคะแนนรวมของผลการตอบข้อสอบวิชาคณิตศาสตร์ที่เท่ากัน และใช้สถิติ Mean and Covariance Structures ทดสอบนัยสำคัญ โดยใช้โปรแกรมสำเร็จรูป LISREL ช่วยในการวิเคราะห์

5.3 ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก (Logistic Discriminant Function Analysis: LDFA)

นำข้อมูลของกลุ่มเพศชายและเพศหญิงในเงื่อนไขความยาวของแบบสอบและขนาดกลุ่มตัวอย่างขนาดเดียวกัน แต่ละคู่มาดำเนินการวิเคราะห์ เพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี LDFA โดยใช้โปรแกรมสำเร็จรูป MULTILOG ช่วยในการวิเคราะห์

6. คำนวณประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

หลังจากวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธี MACS และวิธี LDFA แล้วนำผลการวิเคราะห์ที่ได้มาหาความถูกต้องของการตรวจสอบและความผิดพลาดด้วยการเปรียบเทียบกับผลที่ได้จากการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GPCM ถือว่าเป็นข้อที่ทำหน้าที่ต่างกัน ถ้าวิธี MACS และวิธี LDFA ตรวจสอบพบข้อสอบที่ทำหน้าที่ต่างกันตรงกับข้อสอบที่ถูกตรวจสอบพบด้วยวิธี GPCM ถือว่าวิธีนั้นมีความถูกต้องในการตรวจสอบ แต่ถ้าตรวจสอบพบว่าข้อสอบทำหน้าที่ต่างกันไม่ตรงกับข้อสอบที่ตรวจพบด้วยวิธี GPCM ถือว่ามีความผิดพลาดในการตรวจสอบ โดยนำผลมาคำนวณหาอัตราความคลาดเคลื่อนประเภทที่1 (Type I error rate) และค่าอำนาจการทดสอบ (Power of the test) ของแต่ละวิธี เพื่อเปรียบเทียบประสิทธิภาพของแต่ละวิธีซึ่งสามารถคำนวณได้ดังนี้

อัตราความคลาดเคลื่อนประเภทที่ 1 (Type I error rate)

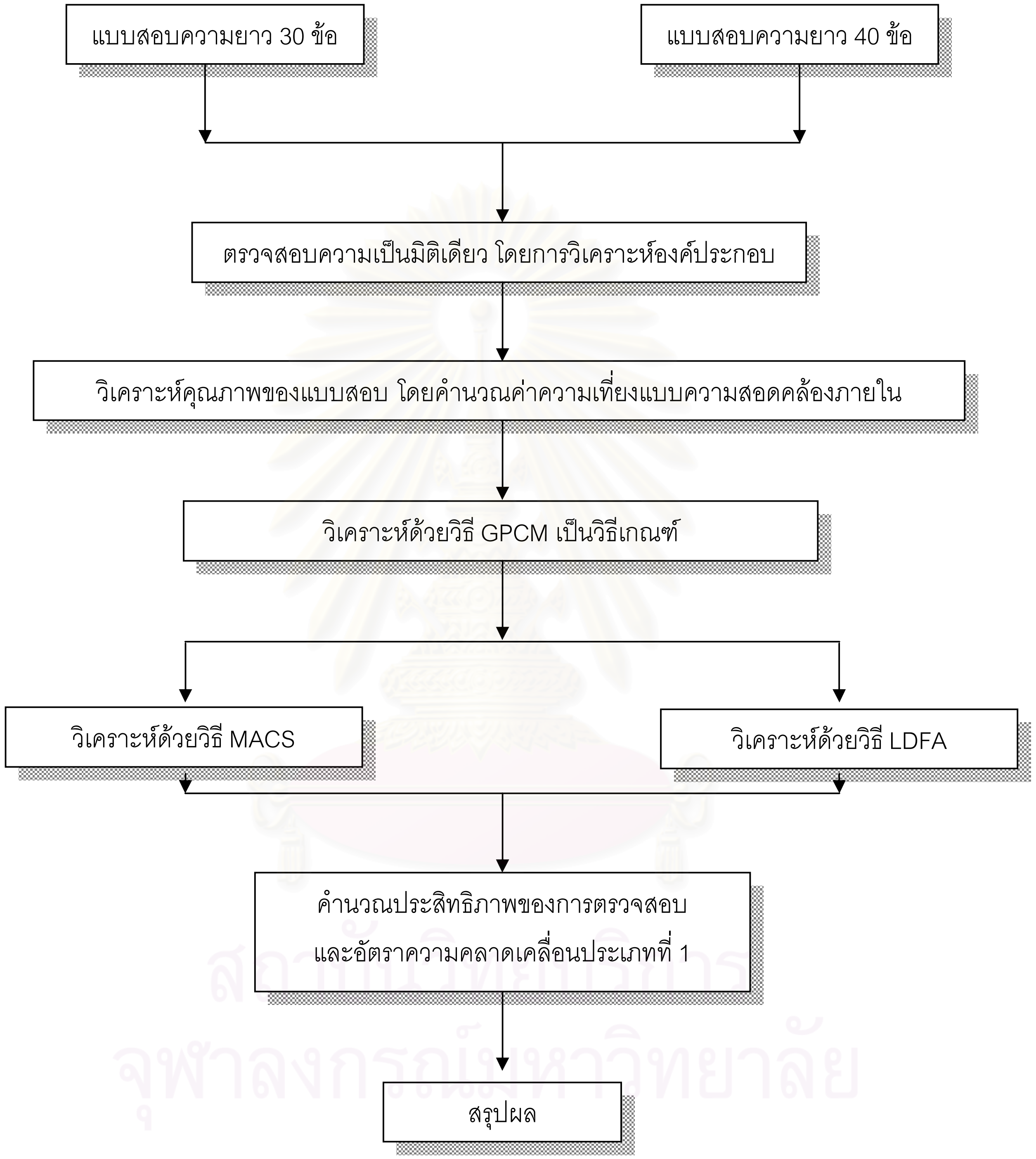
$$= \frac{\text{จำนวนข้อสอบที่ระบุผิดพลาดว่าทำหน้าที่ต่างกันทั้งที่ความจริงทำหน้าที่ไม่ต่างกัน}}{\text{จำนวนข้อสอบที่ทำหน้าที่ไม่ต่างกันทั้งหมดในแบบสอบ}}$$

อำนาจการทดสอบ (Power of the test)

$$= \frac{\text{จำนวนข้อสอบที่ตรวจพบว่าทำหน้าที่ต่างกันได้อย่างถูกต้อง}}{\text{จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งหมดในแบบสอบ}}$$

จากขั้นตอนดำเนินการวิจัยทั้งหมด สามารถสรุปเป็นแผนภาพได้ดังนี้

ขั้นตอนการดำเนินการวิจัย



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

แผนภาพที่ 5 ขั้นตอนดำเนินการวิจัย

บทที่ 4
ผลการวิเคราะห์ข้อมูล

การวิจัยนี้มีจุดประสงค์เพื่อเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาค ระหว่างวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย กับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก โดยใช้ข้อมูลจากผลการตอบข้อสอบ ของนักเรียนชั้นมัธยมศึกษาปีที่ 2 ในเขตพื้นที่การศึกษาเขต 1 กรุงเทพมหานคร ปีการศึกษา 2550 จำนวน 1,806 คน ซึ่งเครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ด้านเหตุผล 5 ตัวเลือกที่มีคำตอบถูกหลายคำตอบ และให้ตัดสินทุกคำตอบ (Multiple True-false Answer) เรื่องจำนวนเต็ม เศษส่วน และทศนิยม ระดับชั้นมัธยมศึกษาตอนต้นของสุนิสา จุ้ยม่วงศรี (2546)

ผลการวิเคราะห์ข้อมูลในบทนี้ ผู้วิจัยได้นำเสนอตามลำดับ ดังนี้

ตอนที่ 1 ผลการวิเคราะห์ค่าสถิติพื้นฐานของผลการตอบข้อสอบ

ตอนที่ 2 ผลการวิเคราะห์คุณภาพของแบบสอบที่ใช้ในการวิจัย

ตอนที่ 3 ผลการวิเคราะห์องค์ประกอบเพื่อตรวจสอบความเป็นเอกมิติของข้อมูล

ตอนที่ 4 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GPCM โดยเสนอค่าพารามิเตอร์ของข้อสอบ และผลของการตรวจสอบ

ตอนที่ 5 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 2 วิธี ได้แก่ วิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย กับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก โดยเรียงลำดับ ดังนี้

5.1 ผลการเปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย กับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก

5.2 ผลการเปรียบเทียบอำนาจการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย กับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก

5.3 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย กับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก

ในการวิเคราะห์ข้อมูลและการแปลผลการวิเคราะห์ข้อมูลในการวิจัยครั้งนี้ เพื่อให้การนำเสนอผลการวิเคราะห์ข้อมูลมีความสะดวกและเกิดความเข้าใจตรงกัน ผู้วิจัยจึงได้กำหนดสัญลักษณ์และความหมายที่ใช้แทนตัวแปรต่าง ๆ ดังต่อไปนี้

DIF หมายถึง การทำหน้าที่ต่างกันของข้อสอบ

GPCM หมายถึง การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีพาเชียลเครดิตโมเดลทั่วไป

MACS หมายถึง การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย

LDFA หมายถึง การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก

N_R หมายถึง ขนาดกลุ่มตัวอย่างกลุ่มอ้างอิง (เพศชาย)

N_F หมายถึง ขนาดกลุ่มตัวอย่างกลุ่มเปรียบเทียบ (เพศหญิง)

a_t หมายถึง ค่าพารามิเตอร์ความชัน (Slope Parameter) หรือค่าอำนาจจำแนกของข้อสอบรวม

b_t หมายถึง ค่าพารามิเตอร์ลำดับขั้นข้อสอบ (Item-Category Parameter) หรือค่าความยากของข้อสอบ

a_R หมายถึง ค่าพารามิเตอร์ความชัน (Slope Parameter) หรือค่าอำนาจจำแนกของข้อสอบของกลุ่มเพศชาย

b_{R1} หมายถึง ค่าพารามิเตอร์ลำดับขั้นข้อสอบ (Item-Category Parameter) หรือค่าความยากของข้อสอบของกลุ่มเพศชาย ลำดับขั้นที่ 1

b_{R2} หมายถึง ค่าพารามิเตอร์ลำดับขั้นข้อสอบ (Item-Category Parameter) หรือค่าความยากของข้อสอบของกลุ่มเพศชาย ลำดับขั้นที่ 2

b_{R3} หมายถึง ค่าพารามิเตอร์ลำดับขั้นข้อสอบ (Item-Category Parameter) หรือค่าความยากของข้อสอบของกลุ่มเพศชาย ลำดับขั้นที่ 3

b_{R4} หมายถึง ค่าพารามิเตอร์ลำดับขั้นข้อสอบ (Item-Category Parameter) หรือค่ายากของข้อสอบของกลุ่มเพศชาย ลำดับขั้นที่ 4

b_{R5} หมายถึง ค่าพารามิเตอร์ลำดับขั้นข้อสอบ (Item-Category Parameter) หรือค่าความยากของข้อสอบของกลุ่มเพศชาย ลำดับขั้นที่ 5

a_F หมายถึง ค่าพารามิเตอร์ความชัน (Slope Parameter) หรือค่าอำนาจจำแนก ของข้อสอบของกลุ่มเพศหญิง

b_{F1} หมายถึง ค่าพารามิเตอร์ลำดับขั้นข้อสอบ (Item-Category parameter) หรือค่าความยากของข้อสอบ ของกลุ่มเพศหญิง ลำดับขั้นที่ 1

b_{F2} หมายถึง ค่าพารามิเตอร์ลำดับขั้นข้อสอบ (Item-Category parameter) หรือค่าความยากของข้อสอบ ของกลุ่มเพศหญิง ลำดับขั้นที่ 2

b_{F3} หมายถึง ค่าพารามิเตอร์ลำดับขั้นข้อสอบ (Item-Category parameter) หรือค่าความยากของข้อสอบ ของกลุ่มเพศหญิง ลำดับขั้นที่ 3

b_{F4} หมายถึง ค่าพารามิเตอร์ลำดับขั้นข้อสอบ (Item-Category parameter) หรือค่าความยากของข้อสอบ ของกลุ่มเพศหญิง ลำดับขั้นที่ 4

b_{F5} หมายถึง ค่าพารามิเตอร์ลำดับขั้นข้อสอบ (Item-Category parameter) หรือค่าความยากของข้อสอบ ของกลุ่มเพศหญิง ลำดับขั้นที่ 5

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตอนที่ 1 การวิเคราะห์ค่าสถิติพื้นฐานของผลการตอบข้อสอบ

การวิเคราะห์ข้อมูลส่วนนี้ผู้วิจัยได้นำคะแนนจากการตอบข้อสอบของแบบทดสอบ วิชาคณิตศาสตร์ เรื่อง จำนวนเต็ม เศษส่วน และทศนิยม จำนวน 30 ข้อ และ 40 ข้อ มาคำนวณหาค่าสถิติพื้นฐาน โดยแบบทดสอบฉบับ 30 ข้อ ได้มาโดยการสุ่มอย่างง่าย (Simple Random Sampling) ซึ่งผลการตอบข้อสอบมีลักษณะการแจกแจงของคะแนนดังตารางที่ 7

ตารางที่ 7 ค่าสถิติพื้นฐานของแบบทดสอบวิชาคณิตศาสตร์ ขนาด 40 ข้อ และ 30 ข้อ

ค่าสถิติ	**ความยาวแบบทดสอบ**	
	40 ข้อ	**30 ข้อ**
คะแนนเต็ม	200	150
ค่าคะแนนสูงสุด (Max)	196	148
ค่าคะแนนต่ำสุด (Min)	81	57
ค่าคะแนนเฉลี่ย (M)	127.360	95.303
ค่าส่วนเบี่ยงเบนมาตรฐาน (SD)	26.870	19.937
พิสัย (Range)	115	91
มัธยฐาน (Mdn)	119	90
ฐานนิยม (Mo)	110	76
ความเบ้ (Sk)	.721	.661
ความโด่ง (Ku)	-.527	-.548
จำนวนผู้สอบ (N)	1806	1806

จากผลการวิเคราะห์ค่าสถิติพื้นฐานของแบบทดสอบทั้งฉบับ ซึ่งผู้วิจัยได้แยกวิเคราะห์ออกเป็น 2 ฉบับ ตามปัจจัยความยาวของแบบทดสอบ นั่นคือแบบทดสอบความยาว 40 ข้อและ 30 ข้อ ซึ่งได้มาโดยการสุ่มอย่างง่าย (Simple Random Sampling)

จากตารางที่ 7 พบว่า แบบทดสอบความยาว 40 ข้อ มีคะแนนเต็ม 200 คะแนน ค่าคะแนนสูงสุด 196 คะแนน ค่าคะแนนต่ำสุด 81 คะแนน ค่าเฉลี่ยของคะแนนเท่ากับ 127.36 คะแนน ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 26.870 พิสัยเท่ากับ 115 มัธยฐานเท่ากับ 119 ฐานนิยมเท่ากับ 110 ค่าความเบ้เท่ากับ .721 นั่นคือ ข้อมูลมีโค้งความถี่เบ้ขวา และค่าความโด่งเท่ากับ -.527 นั่นคือข้อมูลมีโค้งความถี่ลักษณะต่ำ (Platykurtic)

แบบทดสอบความยาว 30 ข้อ มีคะแนนเต็ม 150 คะแนน ค่าคะแนนสูงสุด 148 คะแนน ค่าคะแนนต่ำสุด 57 คะแนน ค่าเฉลี่ยของคะแนนเท่ากับ 95.30 คะแนน ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 19.937 พิสัยเท่ากับ 91 มัธยฐานเท่ากับ 90 ฐานนิยมเท่ากับ 76 ค่าความเบ้เท่ากับ .661 นั่นคือ ข้อมูลมีโค้งความถี่เบ้ขวา และค่าความโด่งเท่ากับ -.548 นั่นคือข้อมูลมีโค้งความถี่ลักษณะต่ำ (Platykurtic)

ตารางที่ 8 เปรียบเทียบค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของคะแนนระหว่างกลุ่มผู้สอบ

ความยาวแบบทดสอบ	เพศชาย	เพศหญิง
40 ข้อ (คะแนนเต็ม 200 คะแนน)		
ค่าคะแนนเฉลี่ย (M)	126.47	128.24
ค่าส่วนเบี่ยงเบนมาตรฐาน (SD)	28.768	24.864
30 ข้อ (คะแนนเต็ม 150 คะแนน)		
ค่าคะแนนเฉลี่ย (M)	94.226	96.316
ค่าส่วนเบี่ยงเบนมาตรฐาน (SD)	21.214	18.561

จากตารางที่ 8 เมื่อนำผลการตอบข้อสอบมาวิเคราะห์เพื่อเปรียบเทียบระหว่างกลุ่มผู้สอบโดยจำแนกตามเพศ ซึ่งเป็นเพศหญิงจำนวน 914 คน และเพศชายจำนวน 892 คน คิดเป็นร้อยละ 50.61 และ 49.39 ตามลำดับ พบว่าในแบบสอบความยาว 40 ข้อและ 30 ข้อ เพศหญิงมีค่าคะแนนเฉลี่ยสูงกว่าเพศชาย และมีค่าส่วนเบี่ยงเบนมาตรฐานน้อยกว่าเพศชาย

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตอนที่ 2 ผลการวิเคราะห์คุณภาพของแบบสอบที่ใช้ในการวิจัย

การวิเคราะห์คุณภาพของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ด้านเหตุผล 5 ตัวเลือกที่มีคำตอบถูกหลายคำตอบ และให้ตัดสินทุกคำตอบ (Multiple True-false Answer) เรื่องจำนวนเต็ม เศษส่วน และทศนิยม ระดับชั้นมัธยมศึกษาตอนต้น ของสุนิสา จุ้ยม่วงศรี (2546) วิเคราะห์โดยใช้โปรแกรม SPSS for windows version 13 และโปรแกรม Multilog 7.03 ซึ่งผู้วิจัยจะนำเสนอผลการวิเคราะห์ในด้านค่าความเที่ยงแบบความสอดคล้องภายในของแบบสอบ โดยใช้สูตรสัมประสิทธิ์แอลฟาของครอนบาค (Alpha's Cronbach) ดังตารางที่ 9

ตารางที่ 9 ค่าความเที่ยงของแบบสอบ

แบบสอบความยาว	ค่าความเที่ยงของแบบสอบ (Alpha's Cronbach)
40 ข้อ	0.958
30 ข้อ	0.938

จากตารางที่ 9 ผลการวิเคราะห์คุณภาพของแบบสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ด้านเหตุผล 5 ตัวเลือกที่มีคำตอบถูกหลายคำตอบ และให้ตัดสินทุกคำตอบ (Multiple True-false Answer) จำนวน 40 ข้อและ 30 ข้อนั้น พบว่า

แบบสอบจำนวน 40 ข้อ เมื่อพิจารณาผลการวิเคราะห์ค่าความเที่ยงของแบบสอบพบว่า ค่าความเที่ยงของแบบสอบมีค่าอยู่ในเกณฑ์ดี (0.958)

แบบสอบจำนวน 30 ข้อ เมื่อพิจารณาผลการวิเคราะห์ค่าความเที่ยงของแบบสอบพบว่า ค่าความเที่ยงของแบบสอบมีค่าอยู่ในเกณฑ์ดี (0.938)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตอนที่ 3 ผลการวิเคราะห์องค์ประกอบเพื่อตรวจสอบความเป็นเอกมิติของข้อมูล

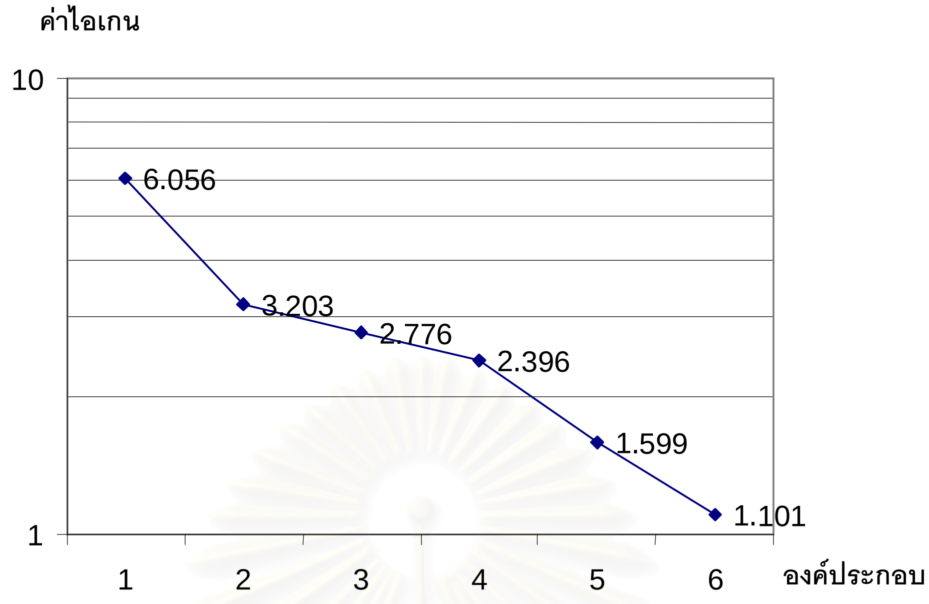
การตรวจสอบความเป็นมิติเดียวของแบบทดสอบ

การตรวจสอบความเป็นมิติเดียวของแบบทดสอบ เป็นข้อตกลงพื้นฐานที่สำคัญของทฤษฎีการตอบสนองข้อสอบ (IRT) เป็นการตรวจสอบว่าแบบทดสอบวัดคุณลักษณะแฝงเพียงคุณลักษณะเดียว ผู้วิจัยได้นำคะแนนที่ได้จากการตอบแบบสอบวิชา คณิตศาสตร์ เรื่อง จำนวนเต็ม เศษส่วน และทศนิยม จำนวน 40 ข้อและ 30 ข้อ มาตรวจสอบความเป็นมิติเดียว โดยใช้การวิเคราะห์องค์ประกอบ (Factor Analysis) ด้วยการวิเคราะห์องค์ประกอบหลัก (Principle Componant Analysis) หมุนแกนด้วยวิธีแวริแมกซ์ (Varimax) หาค่าไอเกน (Eigen Value) และร้อยละของความแปรปรวน ดังตารางที่ 10

ตารางที่ 10 ค่าไอเกน และร้อยละของความแปรปรวนขององค์ประกอบ ของแบบทดสอบขนาด 40 ข้อ

องค์ประกอบ	ค่าไอเกน	ร้อยละของความแปรปรวน
1	6.056	15.140
2	3.203	8.008
3	2.776	6.941
4	2.396	5.990
5	1.599	3.999
6	1.101	2.753

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



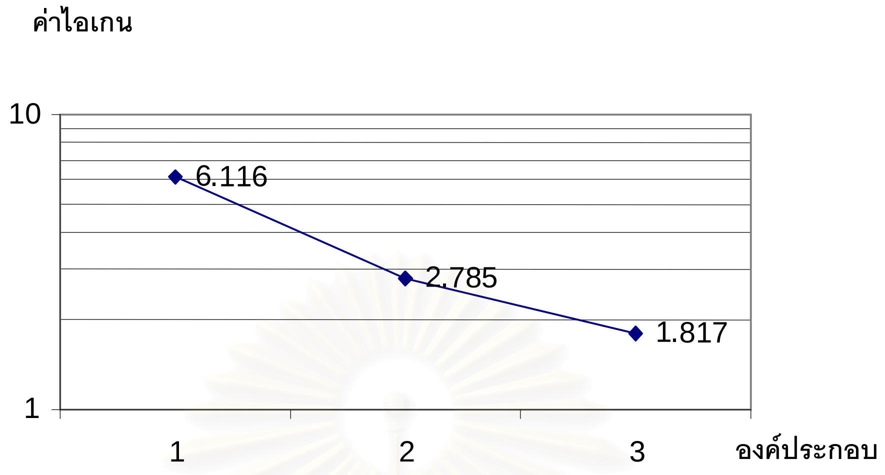
ภาพที่ 6 ผลการตรวจสอบความเป็นมิติเดียวของแบบทดสอบความยาว 40 ข้อ

ผลการวิเคราะห์จากตารางที่ 10 พบว่าแบบสอบความยาว 40 ข้อ ค่าไอเกนขององค์ประกอบที่ 1 มีค่าสูงสุด (6.056) ซึ่งมีค่าสูงกว่าค่าไอเกนขององค์ประกอบที่ 2 (3.203) ประมาณ 1.890 เท่า องค์ประกอบที่ 1 มีค่าไอเกนแตกต่างจากไอเกนขององค์ประกอบอื่นอย่างไม่เด่นชัด ส่วนค่าไอเกนขององค์ประกอบอื่น ๆ มีค่าใกล้เคียงกัน และองค์ประกอบที่ 1 มีค่าร้อยละของความแปรปรวนค่อนข้างสูง (15.140) ซึ่งใกล้เคียงกับเกณฑ์ของ Reckase (อ้างถึงใน Raju, 1993; อุทัยวรรรณ สายพัฒนะ, 2547) ที่เสนอว่าค่าร้อยละของความแปรปรวนควรมีค่าไม่น้อยกว่าร้อยละ 20 หรือค่าไอเกนขององค์ประกอบที่ 1 แตกต่างจากองค์ประกอบอื่นอย่างเด่นชัด จึงจะถือว่าแบบทดสอบมีความเป็นมิติเดียว เมื่อพิจารณาค่าไอเกนและค่าร้อยละของความแปรปรวนขององค์ประกอบที่ 1 และภาพที่ 6 ประกอบกันแล้ว ในทางปฏิบัติยังไม่สามารถถือได้ว่าแบบทดสอบฉบับนี้เป็นมิติเดียว

ดังนั้นจึงเป็นข้อจำกัดอย่างหนึ่งของการวิจัย ที่แบบสอบความยาว 40 ข้อ ยังไม่มีความเป็นเอกมิติอย่างสมบูรณ์ แต่ผู้วิจัยได้ทำการวิเคราะห์โดยนำข้อมูลไปวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบต่อไป

ตารางที่ 11 ค่าไอเกน และร้อยละของความแปรปรวนขององค์ประกอบ ของแบบทดสอบขนาด 30 ข้อ

องค์ประกอบ	ค่าไอเกน	ร้อยละของความแปรปรวน
1	6.116	20.385
2	2.785	9.283
3	1.817	6.058



ภาพที่ 7 ผลการตรวจสอบความเป็นมิติเดียวของแบบทดสอบความยาว 30 ข้อ

ผลการวิเคราะห์จากตารางที่ 11 พบว่า แบบทดสอบความยาว 30 ข้อค่าไอเกนขององค์ประกอบที่ 1 มีค่าสูงสุด (6.116) ซึ่งมีค่าสูงกว่าค่าไอเกนขององค์ประกอบที่ 2 (2.785) ประมาณ 2.196 เท่า องค์ประกอบที่ 1 มีค่าไอเกนแตกต่างจากไอเกนขององค์ประกอบอื่นอย่างเด่นชัด ส่วนค่าไอเกนขององค์ประกอบอื่น ๆ มีค่าใกล้เคียงกัน และองค์ประกอบที่ 1 มีค่าร้อยละของความแปรปรวนค่อนข้างสูง (20.385) เมื่อพิจารณาค่าไอเกนและค่าร้อยละของความแปรปรวนขององค์ประกอบที่ 1 และภาพที่ 7 ประกอบกันแล้ว ในทางปฏิบัติถือได้ว่าแบบทดสอบฉบับนี้เป็นมิติเดียว ดังนั้นจึงสามารถนำข้อมูลไปวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบได้

ตอนที่ 4 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GPCM โดยเสนอค่าพารามิเตอร์ของข้อสอบ และผลของการตรวจสอบ

หลังจากตรวจสอบความเป็นมิติเดียวของข้อมูล ซึ่งเป็นข้อตกลงเบื้องต้นในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบตามแนวคิดของทฤษฎีการตอบสนองข้อสอบ (IRT) แล้วพบว่า แบบทดสอบความยาว 40 ข้อและ 30 ข้อ มีความเป็นมิติเดียว จึงตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GPCM โดยการประมาณค่าพารามิเตอร์ของข้อสอบกลุ่มเพศชายและกลุ่มเพศหญิง ขนาดกลุ่มละ 1,000 คน เมื่อแบบทดสอบมีความยาว 40 ข้อและ 30 ข้อ โดยใช้โปรแกรม MULTILOG version 7.03 วิเคราะห์ค่าพารามิเตอร์ความชัน (Slope Parameter) หรือค่าอำนาจจำแนกของข้อสอบ (a) และค่าพารามิเตอร์ลำดับขั้นของข้อสอบ (Item-category Parameter) หรือค่าความยากลำดับขั้นของข้อสอบ (b) ดังตารางที่ 12

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 12 ค่าพารามิเตอร์ของข้อสอบ แยกตามกลุ่มเพศชาย-หญิง เมื่อกลุ่มตัวอย่าง 1,000 คนและความยาวของแบบทดสอบ 40 ข้อ

ข้อที่	ค่าพารามิเตอร์ของข้อสอบ											
	a_R	a_F	b_{R1}	b_{R2}	b_{R3}	b_{R4}	b_{R5}	b_{F1}	b_{F2}	b_{F3}	b_{F4}	b_{F5}
1	1.38	1.24	-3.33	-1.77	-0.92	0.23	6.16	-3.28	-1.91	-0.92	0.28	6.67
2	1.45	1.16	-2.21	-1.06	-0.34	0.70	6.70	-2.69	-1.31	-0.30	1.07	7.26
3	1.50	1.35	-2.39	-0.97	-0.18	1.17	6.87	-3.05	-1.41	-0.43	1.48	6.08
4	1.52	1.08	-1.82	-0.81	0.28	1.62	6.99	-2.54	-1.10	0.53	1.97	8.66
5	1.92	1.84	-2.06	-1.09	-0.21	0.77	6.07	-2.17	-1.11	-0.19	0.64	5.90
6	1.91	1.55	-1.56	-0.95	-0.11	0.54	5.67	-2.07	-1.33	-0.29	0.49	6.20
7	1.60	1.38	-1.46	-0.55	0.25	0.95	6.24	-1.60	-0.42	0.32	1.12	6.61
8	2.64	1.89	-1.50	-0.87	-0.07	0.45	4.88	-1.93	-1.02	-0.18	0.49	5.81
9	2.55	2.49	-1.47	-0.75	-0.25	0.19	4.67	-1.62	-0.83	-0.36	0.02	4.69
10	2.21	2.32	-1.30	-0.69	-0.15	0.61	5.01	-1.51	-0.87	-0.35	0.35	4.89
11	1.84	2.03	-1.88	-1.04	-0.29	0.52	6.15	-1.92	-1.17	-0.48	0.30	5.47
12	1.81	1.38	-1.42	-0.52	0.36	0.95	5.52	-1.72	-0.53	0.51	1.23	6.64
13	1.79	1.79	-1.76	-0.81	0.01	0.92	6.05	-1.93	-0.94	-0.01	0.94	6.15
14	1.25	1.03	-1.84	-0.74	0.50	1.88	7.93	-2.10	-0.66	0.82	2.60	8.39
15	1.87	1.46	-1.63	-0.79	0.01	1.07	5.88	-1.99	-0.90	0.11	1.52	6.93
16	1.84	1.64	-1.85	-0.89	-0.06	0.88	5.97	-2.13	-1.13	-0.20	0.84	6.32
17	1.85	1.42	-1.82	-0.87	-0.04	0.86	5.88	-2.44	-1.27	-0.05	1.18	7.04
18	1.62	1.19	-1.78	-0.77	0.05	1.01	6.60	-2.61	-1.10	0.06	1.38	7.45
19	1.40	1.50	-2.02	-0.67	0.21	1.22	6.51	-2.03	-0.91	-0.03	0.92	6.48
20	1.88	1.63	-2.78	-1.49	-0.28	0.56	6.65	-2.13	-1.11	-0.22	0.78	6.29
21	2.09	2.09	-1.62	-0.85	-0.37	0.22	5.18	-1.78	-1.14	-0.68	0.02	5.06
22	1.15	0.97	-2.39	-1.17	0.15	2.34	6.62	-2.52	-1.20	-0.16	2.19	7.14
23	1.93	1.43	-1.39	-0.63	0.12	1.09	5.53	-1.93	-1.01	-0.07	1.29	6.81
24	0.63	0.86	-3.26	-0.90	1.49	4.49	10.10	-2.60	-0.77	0.82	2.97	7.64
25	1.68	1.34	-1.40	-0.56	0.38	1.06	6.12	-1.83	-0.59	0.43	1.07	6.96
26	1.10	1.05	-2.24	-0.78	0.65	1.79	8.43	-2.26	-0.74	0.63	2.00	8.71

ตารางที่ 12 (ต่อ)

ข้อที่	ค่าพารามิเตอร์ของข้อสอบ											
	a_R	a_F	b_{R1}	b_{R2}	b_{R3}	b_{R4}	b_{R5}	b_{F1}	b_{F2}	b_{F3}	b_{F4}	b_{F5}
27	1.64	0.99	-1.52	-0.67	0.09	1.44	6.34	-2.58	-0.94	0.41	2.06	9.06
28	1.11	0.80	-2.21	-0.82	0.51	1.48	8.30	-2.45	-0.67	1.10	2.56	9.29
29	1.35	1.16	-2.09	-0.48	0.48	1.25	7.42	-2.13	-0.61	0.38	1.26	7.55
30	1.78	1.77	-1.62	-0.68	0.11	1.15	6.04	-1.72	-0.79	0.02	0.90	5.83
31	0.58	0.62	-2.11	1.07	4.50	11.52	14.97	-3.38	-0.37	2.01	5.87	11.10
32	1.55	1.30	-1.75	-0.80	0.02	1.12	6.51	-1.97	-0.88	-0.12	0.77	6.87
33	1.36	0.67	-1.78	-0.52	0.68	1.48	6.10	-3.28	-0.54	1.64	3.40	8.90
34	1.46	1.60	-1.66	-0.84	-0.13	1.14	6.85	-1.79	-0.86	-0.17	0.68	6.34
35	0.82	0.85	-2.57	-1.03	0.92	2.38	10.26	-2.13	-0.58	1.02	2.39	7.62
36	0.94	0.53	-2.47	-0.71	1.25	2.62	7.42	-3.93	-0.89	2.11	5.24	11.25
37	1.23	0.98	-2.03	-0.77	0.32	2.06	8.22	-2.73	-0.77	0.70	2.38	8.95
38	0.94	0.80	-2.58	-0.89	0.55	2.86	9.79	-3.07	-1.04	0.66	3.67	8.36
39	0.92	0.89	-2.15	-0.72	1.18	2.58	9.26	-2.44	-0.78	1.14	2.71	9.05
40	0.86	0.85	-2.72	-0.91	1.15	2.40	9.83	-2.93	-0.97	0.76	1.93	9.65
ค่าเฉลี่ย	1.52	1.32	-1.99	-0.79	0.32	1.59	7.04	-2.32	-0.93	0.27	1.62	7.30

จากตารางที่ 12 พบว่า ค่าพารามิเตอร์ของข้อสอบกลุ่มตัวอย่างเพศชายและเพศหญิงเมื่อกลุ่มตัวอย่างขนาด 1,000 คน และแบบทดสอบความยาว 40 ข้อ กลุ่มตัวอย่างเพศชายมีค่าอำนาจจำแนกเฉลี่ย 1.52 และค่าความยากเฉลี่ยในแต่ละขั้นเท่ากับ -1.99, -0.79, 0.32, 1.59 และ7.04 ตามลำดับ ส่วนในกลุ่มตัวอย่างเพศหญิง มีค่าอำนาจจำแนกเฉลี่ยเท่ากับ 1.32 และค่าความยากเฉลี่ยในแต่ละขั้นเท่ากับ -2.32, -0.93, 0.27, 1.62 และ 7.30 ตามลำดับ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 13 ค่าพารามิเตอร์ของข้อสอบ แยกตามกลุ่มเพศชาย-หญิง เมื่อกลุ่มตัวอย่าง 1,000 คนและความยาวของแบบทดสอบ 30 ข้อ

ข้อที่	ค่าพารามิเตอร์ของข้อสอบ											
	a_R	a_F	b_{R1}	b_{R2}	b_{R3}	b_{R4}	b_{R5}	b_{F1}	b_{F2}	b_{F3}	b_{F4}	b_{F5}
1	1.32	1.25	-3.51	-1.88	-1.00	0.17	6.34	-3.29	-1.92	-0.93	0.28	6.30
2	1.46	1.30	-2.52	-1.07	-0.26	1.13	7.20	-3.18	-1.47	-0.44	1.54	6.30
3	1.46	1.05	-1.97	-0.93	0.22	1.58	6.98	-2.62	-1.13	0.55	2.03	8.84
4	1.89	1.75	-2.16	-1.19	-0.29	0.72	6.21	-2.27	-1.16	-0.19	0.67	6.31
5	1.86	1.49	-1.69	-1.05	-0.18	0.49	5.84	-2.16	-1.38	-0.29	0.51	6.23
6	1.45	1.30	-1.66	-0.66	0.21	0.95	6.65	-1.69	-0.43	0.33	1.17	7.08
7	2.52	2.36	-1.61	-0.86	-0.35	0.11	4.87	-1.70	-0.86	-0.37	0.03	4.95
8	2.07	2.25	-1.45	-0.81	-0.23	0.57	5.29	-1.57	-0.90	-0.35	0.37	5.03
9	1.83	1.97	-2.00	-1.14	-0.38	0.46	6.03	-1.99	-1.22	-0.50	0.31	5.55
10	1.74	1.74	-1.90	-0.90	-0.08	0.86	6.13	-2.00	-0.97	-0.01	0.97	6.33
11	1.17	1.02	-2.04	-0.86	0.45	1.91	8.47	-2.13	-0.67	0.83	2.61	8.49
12	1.81	1.62	-1.98	-0.99	-0.15	0.83	6.18	-2.19	-1.16	-0.21	0.86	6.48
13	1.81	1.37	-1.95	-0.97	-0.11	0.80	6.10	-2.52	-1.31	-0.05	1.22	7.14
14	1.57	1.14	-1.91	-0.88	-0.02	0.97	6.73	-2.72	-1.14	0.06	1.43	7.83
15	1.34	1.40	-2.18	-0.77	0.15	1.18	6.89	-2.16	-0.96	-0.02	0.98	6.91
16	2.01	1.99	-1.76	-0.96	-0.45	0.16	5.57	-1.87	-1.19	-0.71	0.03	5.10
17	1.12	0.95	-2.53	-1.30	0.08	2.29	6.76	-2.60	-1.24	-0.16	2.25	7.52
18	1.86	1.36	-1.52	-0.72	0.05	1.05	5.78	-2.02	-1.05	-0.07	1.35	7.13
19	0.59	0.83	-3.58	-1.04	1.51	4.70	10.93	-2.70	-0.79	0.86	3.07	7.96
20	1.05	0.99	-2.44	-0.91	0.60	1.79	8.55	-2.40	-0.78	0.67	2.10	9.19
21	1.55	0.95	-1.69	-0.78	0.02	1.42	6.60	-2.67	-0.97	0.42	2.13	9.29
22	1.05	0.81	-2.41	-0.96	0.43	1.44	8.61	-2.45	-0.68	1.10	2.54	8.33
23	1.30	1.11	-2.23	-0.57	0.44	1.23	7.57	-2.23	-0.63	0.41	1.32	7.92
24	0.52	0.60	-2.40	1.13	4.97	35.36	1.61	-3.52	-0.38	2.09	6.09	11.45
25	1.45	1.25	-1.94	-0.92	-0.05	1.11	7.02	-2.04	-0.91	-0.13	0.81	7.16
26	0.75	0.80	-2.82	-1.16	0.93	2.47	10.71	-2.24	-0.61	1.08	2.50	7.87

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 13 (ต่อ)

ข้อที่	ค่าพารามิเตอร์ของข้อสอบ											
	a_R	a_F	b_{R1}	b_{R2}	b_{R3}	b_{R4}	b_{R5}	b_{F1}	b_{F2}	b_{F3}	b_{F4}	b_{F5}
27	0.90	0.50	-2.68	-0.84	1.21	2.64	7.62	-4.18	-0.95	2.24	5.57	11.90
28	1.17	0.95	-2.24	-0.89	0.26	2.07	8.67	-2.81	-0.79	0.72	2.44	9.10
29	0.88	0.89	-2.31	-0.83	1.12	2.60	9.60	-2.44	-0.78	1.13	2.69	7.73
30	0.83	0.85	-2.91	-1.02	1.12	2.40	10.18	-2.96	-0.98	0.76	1.94	9.11
ค่าเฉลี่ย	1.41	1.26	-2.20	-0.89	0.34	2.52	7.06	-2.44	-0.98	0.29	1.73	7.55

จากตารางที่ 13 พบว่า ค่าพารามิเตอร์ของข้อสอบกลุ่มตัวอย่างเพศชายและเพศหญิงเมื่อกลุ่มตัวอย่างขนาด 1,000 คน และแบบทดสอบความยาว 30 ข้อ กลุ่มตัวอย่างเพศชายมีค่าอำนาจจำแนกเฉลี่ย 1.41 และค่าความยากเฉลี่ยในแต่ละขั้นเท่ากับ -2.20, -0.89, 0.34, 2.52 และ 7.06 ตามลำดับ ส่วนในกลุ่มตัวอย่างเพศหญิง มีค่าอำนาจจำแนกเฉลี่ยเท่ากับ 1.26 และค่าความยากเฉลี่ยในแต่ละขั้นเท่ากับ -2.44, -0.98, 0.29, 1.73 และ 7.55 ตามลำดับ

การที่ค่าพารามิเตอร์ข้อสอบของกลุ่มเพศชายและเพศหญิงตามตางที่ 12-13 ทำการประมาณค่ามาจากการวิเคราะห์แยกกันระหว่างกลุ่มผู้สอบย่อย ดังนั้น ก่อนจะนำมาตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจะต้องปรับเทียบสเกลพารามิเตอร์ของข้อสอบให้อยู่บนสเกลเมตริกซ์เดียวกัน ผู้วิจัยได้นำค่าพารามิเตอร์ของข้อสอบทั้งสองกลุ่มมาคำนวณหา ค่าสัมประสิทธิ์การเชื่อมโยง (Linking Coefficients) คือค่าความชันของเส้นถดถอย (A) และค่าจุดตัดบนแกนตั้ง (K) โดยใช้สูตรความสัมพันธ์เชิงเส้น (Linear Relationship) ในการเชื่อมโยง (Linking) (Tate, 1999) โดยใช้โปรแกรม EQUATE version 2.0 ดังตาราง

ตารางที่ 14 ค่าสัมประสิทธิ์การเชื่อมโยงของแบบทดสอบขนาด 40 ข้อและ 30 ข้อ

แบบทดสอบความยาว	A	K
40 ข้อ	0.8014	-0.2911
30 ข้อ	0.7242	-0.2340

ผลการคำนวณค่าสัมประสิทธิ์การเชื่อมโยงตามตารางที่ 14 พบว่า ค่าความชันของเส้นถดถอย (A) และค่าจุดตัดบนแกนตั้ง (K) ของแบบทดสอบความยาว 40 ข้อ และ 30 ข้อ เป็นดังนี้

แบบทดสอบความยาว 40 ข้อ มีค่าความชันของเส้นถดถอย (A) เท่ากับ 0.8014 และค่าจุดตัดบนแกนตั้ง (K) เท่ากับ -0.2911 สำหรับแบบทดสอบความยาว 30 ข้อ มีค่าความชันของเส้นถดถอย (A) เท่ากับ 0.7242 และค่าจุดตัดบนแกนตั้ง (K) เท่ากับ -0.2340

ค่าสัมประสิทธิ์การเชื่อมโยงที่ได้ตามตารางที่ 14 ผู้วิจัยนำมาปรับเทียบสเกลพารามิเตอร์ของกลุ่มเพศหญิงซึ่งเป็นกลุ่มเปรียบเทียบ ให้อยู่บนสเกลเมตริกซ์เดียวกันกับสเกลพารามิเตอร์ของกลุ่มเพศชาย ซึ่งเป็นกลุ่มอ้างอิงของการศึกษาครั้งนี้โดยใช้สูตรการปรับเทียบ

ค่าอำนาจแจกแจงของข้อสอบข้อที่ j จากผู้สอบกลุ่มสนใจที่แปลงค่าแล้ว

= ค่าอำนาจแจกแจงของข้อสอบข้อที่ j จากผู้สอบกลุ่มสนใจที่ยังไม่ได้แปลงค่า / A

ค่าความยากของข้อสอบข้อที่ j จากผู้สอบกลุ่มสนใจที่แปลงค่าแล้ว

= A*ค่าความยากของข้อสอบข้อที่ j จากผู้สอบกลุ่มสนใจที่ยังไม่ได้แปลงค่า + K

จะได้สูตรคำนวณดังนี้

<u>กรณีแบบสอบความยาว 40 ข้อ</u>

ค่าอำนาจแจกแจงของข้อสอบข้อที่ j จากผู้สอบกลุ่มสนใจที่แปลงค่าแล้ว

= A_j / 0.8014

ค่าความยากของข้อสอบข้อที่ j จากผู้สอบกลุ่มสนใจที่แปลงค่าแล้ว

= 0.8014 * b_j + (-0.2911)

<u>กรณีแบบสอบความยาว 30 ข้อ</u>

ค่าอำนาจแจกแจงของข้อสอบข้อที่ j จากผู้สอบกลุ่มสนใจที่แปลงค่าแล้ว

= A_j / 0.7242

ค่าความยากของข้อสอบข้อที่ j จากผู้สอบกลุ่มสนใจที่แปลงค่าแล้ว

= 0.7242 * b_j + (-0.2340)

ค่าที่ได้แสดงดังตาราง

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 15 ค่าพารามิเตอร์ของข้อสอบที่ปรับเทียบสเกลแล้ว แยกตามกลุ่มเพศชาย-หญิง เมื่อกลุ่มตัวอย่าง 1,000 คนและความยาวของแบบทดสอบ 40 ข้อ

ข้อที่	ค่าพารามิเตอร์ของข้อสอบ											
	a_R	a_F	b_{R1}	b_{R2}	b_{R3}	b_{R4}	b_{R5}	b_{F1}	b_{F2}	b_{F3}	b_{F4}	b_{F5}
1	1.38	1.24	-3.33	-1.77	-0.92	0.23	6.16	-3.20	-1.81	-0.79	0.24	6.49
2	1.45	1.16	-2.21	-1.06	-0.34	0.70	6.70	-2.51	-1.22	-0.18	1.07	7.12
3	1.50	1.35	-2.39	-0.97	-0.18	1.17	6.87	-3.00	-1.35	-0.37	1.41	6.04
4	1.52	1.08	-1.82	-0.81	0.28	1.62	6.99	-2.45	-1.00	0.45	1.67	8.46
5	1.92	1.84	-2.06	-1.09	-0.21	0.77	6.07	-2.12	-1.03	-0.11	0.49	5.89
6	1.91	1.55	-1.56	-0.95	-0.11	0.54	5.67	-2.07	-1.20	-0.29	0.49	6.20
7	1.60	1.38	-1.46	-0.55	0.25	0.95	6.24	-1.52	-0.36	0.24	1.12	6.51
8	2.64	1.89	-1.50	-0.87	-0.07	0.45	4.88	-1.81	-0.95	-0.11	0.31	5.70
9	2.55	2.49	-1.47	-0.75	-0.25	0.19	4.67	-1.49	-0.63	-0.22	0.02	4.64
10	2.21	2.32	-1.30	-0.69	-0.15	0.61	5.01	-1.50	-0.77	-0.31	0.35	4.69
11	1.84	2.03	-1.88	-1.04	-0.29	0.52	6.15	-1.90	-1.11	-0.40	0.18	5.40
12	1.81	1.38	-1.42	-0.52	0.36	0.95	5.52	-1.65	-0.46	0.46	1.23	6.54
13	1.79	1.79	-1.76	-0.81	0.01	0.92	6.05	-1.78	-0.79	-0.01	0.79	6.15
14	1.25	1.03	-1.84	-0.74	0.50	1.88	7.93	-2.01	-0.52	0.62	2.01	8.39
15	1.87	1.46	-1.63	-0.79	0.01	1.07	5.88	-1.92	-0.76	0.11	1.52	6.93
16	1.84	1.64	-1.85	-0.89	-0.06	0.88	5.97	-2.10	-1.09	-0.10	0.70	6.32
17	1.85	1.42	-1.82	-0.87	-0.04	0.86	5.88	-2.32	-1.17	-0.05	1.18	7.04
18	1.62	1.19	-1.78	-0.77	0.05	1.01	6.60	-2.41	-1.10	0.06	1.38	7.31
19	1.40	1.50	-2.02	-0.67	0.21	1.22	6.51	-2.03	-0.84	-0.03	0.81	6.48
20	1.88	1.63	-2.78	-1.49	-0.28	0.56	6.65	-2.13	-1.11	-0.22	0.60	6.12
21	2.09	2.09	-1.62	-0.85	-0.37	0.22	5.18	-1.61	-1.14	-0.60	0.02	5.06
22	1.15	0.97	-2.39	-1.17	0.15	2.34	6.62	-2.40	-1.20	-0.16	2.11	7.14
23	1.93	1.43	-1.39	-0.63	0.12	1.09	5.53	-1.80	-1.01	-0.07	1.20	6.81
24	0.63	0.86	-3.26	-0.90	1.49	4.49	10.10	-2.45	-0.77	0.76	2.91	7.64
25	1.68	1.34	-1.40	-0.56	0.38	1.06	6.12	-1.83	-0.59	0.43	1.07	6.80

ตารางที่ 15 (ต่อ)

ข้อที่	ค่าพารามิเตอร์ของข้อสอบ											
	a_R	a_F	b_{R1}	b_{R2}	b_{R3}	b_{R4}	b_{R5}	b_{F1}	b_{F2}	b_{F3}	b_{F4}	b_{F5}
26	1.10	1.05	-2.24	-0.78	0.65	1.79	8.43	-2.26	-0.74	0.59	2.00	8.71
27	1.64	0.99	-1.52	-0.67	0.09	1.44	6.34	-2.58	-0.90	0.41	2.06	9.06
28	1.11	0.80	-2.21	-0.82	0.51	1.48	8.30	-2.45	-0.48	1.10	2.56	9.29
29	1.35	1.16	-2.09	-0.48	0.48	1.25	7.42	-2.13	-0.61	0.38	1.26	7.55
30	1.78	1.77	-1.62	-0.68	0.11	1.15	6.04	-1.72	-0.62	0.02	0.90	5.83
31	0.58	0.62	-2.11	1.07	4.50	11.52	14.97	-3.38	-0.37	2.01	5.62	10.92
32	1.55	1.30	-1.75	-0.80	0.02	1.12	6.51	-1.97	-0.88	-0.12	0.71	6.70
33	1.36	0.67	-1.78	-0.52	0.68	1.48	6.10	-3.28	-0.54	1.43	3.40	8.74
34	1.46	1.60	-1.66	-0.84	-0.13	1.14	6.85	-1.79	-0.86	-0.17	0.68	6.34
35	0.82	0.85	-2.57	-1.03	0.92	2.38	10.26	-2.13	-0.58	1.02	2.39	7.62
36	0.94	0.53	-2.47	-0.71	1.25	2.62	7.42	-3.93	-0.80	2.11	5.13	11.25
37	1.23	0.98	-2.03	-0.77	0.32	2.06	8.22	-2.64	-0.64	0.70	2.38	8.74
38	0.94	0.80	-2.58	-0.89	0.55	2.86	9.79	-3.07	-1.00	0.66	3.67	8.19
39	0.92	0.89	-2.15	-0.72	1.18	2.58	9.26	-2.44	-0.65	1.14	2.71	9.05
40	0.86	0.85	-2.72	-0.91	1.15	2.40	9.83	-2.79	-0.75	0.76	1.93	9.49
ค่าเฉลี่ย	1.52	1.32	-1.99	-0.79	0.32	1.59	7.04	-2.26	-0.86	0.28	1.56	7.23

จากตารางที่ 15 พบว่า ค่าพารามิเตอร์ของข้อสอบกลุ่มตัวอย่างเพศชายและเพศหญิงที่ปรับเทียบสเกลพารามิเตอร์แล้ว เมื่อกลุ่มตัวอย่างขนาด 1,000 คน และแบบทดสอบความยาว 40 ข้อ กลุ่มตัวอย่างเพศชายมีค่าอำนาจจำแนกเฉลี่ย 1.52 และค่าความยากเฉลี่ยในแต่ละขั้นเท่ากับ -1.99, -0.79, 0.32, 1.59 และ 7.04 ตามลำดับ ส่วนในกลุ่มตัวอย่างเพศหญิง มีค่าอำนาจจำแนกเฉลี่ยเท่ากับ 1.32 และค่าความยากเฉลี่ยในแต่ละขั้นเท่ากับ -2.26, -0.86, 0.28, 1.56 และ 7.23 ตามลำดับ

ตารางที่ 16 ค่าพารามิเตอร์ของข้อสอบที่ปรับเทียบสเกลแล้ว แยกตามกลุ่มเพศชาย-หญิง เมื่อกลุ่มตัวอย่าง 1,000 คนและความยาวของแบบทดสอบ 30 ข้อ

ข้อที่	ค่าพารามิเตอร์ของข้อสอบ											
	a_R	a_F	b_{R1}	b_{R2}	b_{R3}	b_{R4}	b_{R5}	b_{F1}	b_{F2}	b_{F3}	b_{F4}	b_{F5}
1	1.32	1.25	-3.51	-1.88	-1.00	0.17	6.34	-2.88	-1.77	-0.97	0.02	4.91
2	1.46	1.30	-2.52	-1.07	-0.26	1.13	7.20	-2.79	-1.40	-0.57	1.04	4.91
3	1.46	1.05	-1.97	-0.93	0.22	1.58	6.98	-2.34	-1.13	0.24	1.44	6.97
4	1.89	1.75	-2.16	-1.19	-0.29	0.72	6.21	-2.05	-1.15	-0.36	0.33	4.92
5	1.86	1.49	-1.69	-1.05	-0.18	0.49	5.84	-1.97	-1.33	-0.45	0.20	4.85
6	1.45	1.30	-1.66	-0.66	0.21	0.95	6.65	-1.58	-0.56	0.06	0.74	5.54
7	2.52	2.36	-1.61	-0.86	-0.35	0.11	4.87	-1.59	-0.91	-0.51	-0.19	3.81
8	2.07	2.25	-1.45	-0.81	-0.23	0.57	5.29	-1.49	-0.94	-0.49	0.09	3.88
9	1.83	1.97	-2.00	-1.14	-0.38	0.46	6.03	-1.83	-1.20	-0.62	0.04	4.30
10	1.74	1.74	-1.90	-0.90	-0.08	0.86	6.13	-1.84	-1.00	-0.22	0.58	4.93
11	1.17	1.02	-2.04	-0.86	0.45	1.91	8.47	-1.94	-0.75	0.46	1.91	6.69
12	1.81	1.62	-1.98	-0.99	-0.15	0.83	6.18	-1.99	-1.15	-0.38	0.49	5.05
13	1.81	1.37	-1.95	-0.97	-0.11	0.80	6.10	-2.26	-1.27	-0.25	0.78	5.59
14	1.57	1.14	-1.91	-0.88	-0.02	0.97	6.73	-2.42	-1.14	-0.16	0.95	6.15
15	1.34	1.40	-2.18	-0.77	0.15	1.18	6.89	-1.97	-0.99	-0.23	0.59	5.40
16	2.01	1.99	-1.76	-0.96	-0.45	0.16	5.57	-1.73	-1.18	-0.79	-0.19	3.93
17	1.12	0.95	-2.53	-1.30	0.08	2.29	6.76	-2.32	-1.22	-0.34	1.62	5.90
18	1.86	1.36	-1.52	-0.72	0.05	1.05	5.78	-1.85	-1.06	-0.27	0.89	5.58
19	0.59	0.83	-3.58	-1.04	1.51	4.70	10.93	-2.40	-0.85	0.49	2.28	6.26
20	1.05	0.99	-2.44	-0.91	0.60	1.79	8.55	-2.16	-0.84	0.33	1.50	7.26

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 16 (ต่อ)

ข้อที่	ค่าพารามิเตอร์ของข้อสอบ											
	a_R	a_F	b_{R1}	b_{R2}	b_{R3}	b_{R4}	b_{R5}	b_{F1}	b_{F2}	b_{F3}	b_{F4}	b_{F5}
21	1.55	0.95	-1.69	-0.78	0.02	1.42	6.60	-2.47	-0.84	0.31	2.05	9.02
22	1.05	0.81	-2.41	-0.96	0.43	1.44	8.61	-2.19	-0.41	1.00	2.39	8.14
23	1.30	1.11	-2.23	-0.57	0.44	1.23	7.57	-2.19	-0.47	0.28	1.20	7.78
24	0.52	0.60	-2.40	1.13	4.97	35.36	1.61	-3.37	-0.14	2.01	6.00	11.41
25	1.45	1.25	-1.94	-0.92	-0.05	1.11	7.02	-2.04	-0.74	-0.13	0.65	7.10
26	0.75	0.80	-2.82	-1.16	0.93	2.47	10.71	-2.24	-0.61	1.08	2.50	7.87
27	0.90	0.50	-2.68	-0.84	1.21	2.64	7.62	-4.18	-0.78	2.24	5.57	11.71
28	1.17	0.95	-2.24	-0.89	0.26	2.07	8.67	-2.69	-0.62	0.72	2.28	9.10
29	0.88	0.89	-2.31	-0.83	1.12	2.60	9.60	-2.44	-0.78	1.13	2.69	7.73
30	0.83	0.85	-2.91	-1.02	1.12	2.40	10.18	-2.73	-0.85	0.73	1.95	9.11
ค่าเฉลี่ย	1.41	1.26	-2.20	-0.89	0.34	2.52	7.06	-2.26	-0.94	0.14	1.41	6.53

จากตารางที่ 16 พบว่า ค่าพารามิเตอร์ของข้อสอบกลุ่มตัวอย่างเพศชายและเพศหญิงที่ปรับเทียบสเกลพารามิเตอร์แล้ว เมื่อกลุ่มตัวอย่างขนาด 1,000 คน และแบบทดสอบความยาว 30 ข้อ กลุ่มตัวอย่างเพศชายมีค่าอำนาจจำแนกเฉลี่ย 1.41 และค่าความยากเฉลี่ยในแต่ละขั้นเท่ากับ -2.20, -0.89, 0.34, 2.52 และ 7.06 ตามลำดับ ส่วนในกลุ่มตัวอย่างเพศหญิง มีค่าอำนาจจำแนกเฉลี่ยเท่ากับ 1.26 และค่าความยากเฉลี่ยในแต่ละขั้นเท่ากับ -2.26, -0.94, 0.14, 1.41 และ 6.53 ตามลำดับ

เมื่อปรับเทียบสเกลพารามิเตอร์ข้อสอบของกลุ่มผู้สอบให้อยู่บนสเกลเดียวกันตามตารางที่ 15-16 แล้ว จึงทำการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ตามวิธี GPCM โดยใช้วิธีการทดสอบไค-สแควร์ของลอร์ด (Lord's Chi-square) โดยใช้โปรแกรมสำเร็จรูป EXCEL และ PARSCALE version 4.1 ช่วยในการวิเคราะห์ ผลดังตาราง

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 17 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GPCM ตามแนวคิดของทฤษฎีการตอบข้อสอบด้วยวิธีของลอร์ด (Lord's Chi-square) เมื่อกลุ่มตัวอย่าง1,000 คนและความยาวแบบสอบ 40 ข้อ

ข้อที่	Lord's Chi-square	ข้อที่	Lord's Chi-square
1	5.2	21	6.7
2	6.8	22	10.4
3	8.5	23	15.6*
4	15.2*	24	10.5
5	8.7	25	8.0
6	6.4	26	7.1
7	11.2	27	5.4
8	26.0**	28	12.4
9	6.7	29	5.4
10	6.1	30	7.1
11	8.1	31	7.9
12	15.2*	32	7.1
13	9.0	33	7.6
14	7.5	34	6.3
15	14.9*	35	5.2
16	8.1	36	14.5*
17	15.2*	37	5.8
18	15.1*	38	4.9
19	8.7	39	4.5
20	5.3	40	6.5

* $p < .05$ ($\chi^2 = 12.6$)

** $p < .01$ ($\chi^2 = 16.8$)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ผลการตรวจสอบตามตารางที่ 17 พบว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี GPCM ตามแนวคิดของทฤษฎีการตอบข้อสอบด้วยวิธีการทดสอบไค-สแควร์ของลอร์ด (Lord's Chi-square) เมื่อกลุ่มตัวอย่าง 1,000 คน และแบบสอบความยาว 40 ข้อ พบข้อสอบที่ทำหน้าที่ต่างกัน (DIF) จำนวน 8 ข้อ ได้แก่ ข้อสอบที่ทำหน้าที่ต่างกันที่ระดับนัยสำคัญ .05 จำนวน 7 ข้อ คือ ข้อ 4, 12, 15, 17, 18, 23 และ 36 ข้อสอบที่ทำหน้าที่ต่างกันที่ระดับนัยสำคัญ .01 จำนวน 1ข้อ คือ ข้อ 8

ตารางที่ 18 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GPCM ตามแนวคิดของทฤษฎีการตอบข้อสอบด้วยวิธีของลอร์ด (Lord's Chi-square) เมื่อกลุ่มตัวอย่าง 1,000 คนและความยาวแบบสอบ 30 ข้อ

ข้อที่	Lord's Chi-square	ข้อที่	Lord's Chi-square
1	6.0	16	7.5
2	5.0	17	4.1
3	14.6*	18	15.5*
4	6.1	19	10.6
5	4.0	20	11.0
6	7.8	21	24.6**
7	6.7	22	3.7
8	6.0	23	9.6
9	8.0	24	3.4
10	4.3	25	4.9
11	7.1	26	6.7
12	8.0	27	14.3*
13	15.2*	28	6.8
14	15.2*	29	6.5
15	7.7	30	4.5

* $p < .05$ ($\chi^2 = 12.6$)

** $p < .01$ ($\chi^2 = 16.8$)

ผลการตรวจสอบตามตารางที่ 18 พบว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี GPCM ตามแนวคิดของทฤษฎีการตอบข้อสอบด้วยวิธีการทดสอบไค-สแควร์ของลอร์ด (Lord's Chi-square) เมื่อกลุ่มตัวอย่าง 1,000 คน และแบบสอบความยาว 30 ข้อ พบข้อสอบที่ทำหน้าที่ต่างกัน (DIF) จำนวน 6 ข้อ ได้แก่ ข้อสอบที่ทำหน้าที่ต่างกันที่ระดับนัยสำคัญ .05 จำนวน 5 ข้อ คือ ข้อ 3, 13, 14, 18 และ 27 ข้อสอบที่ทำหน้าที่ต่างกันที่ระดับนัยสำคัญ .01 จำนวน 1 ข้อ คือ ข้อ 21

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตอนที่ 5 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 2 วิธี ได้แก่ วิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย กับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก โดยเรียงลำดับ ดังนี้

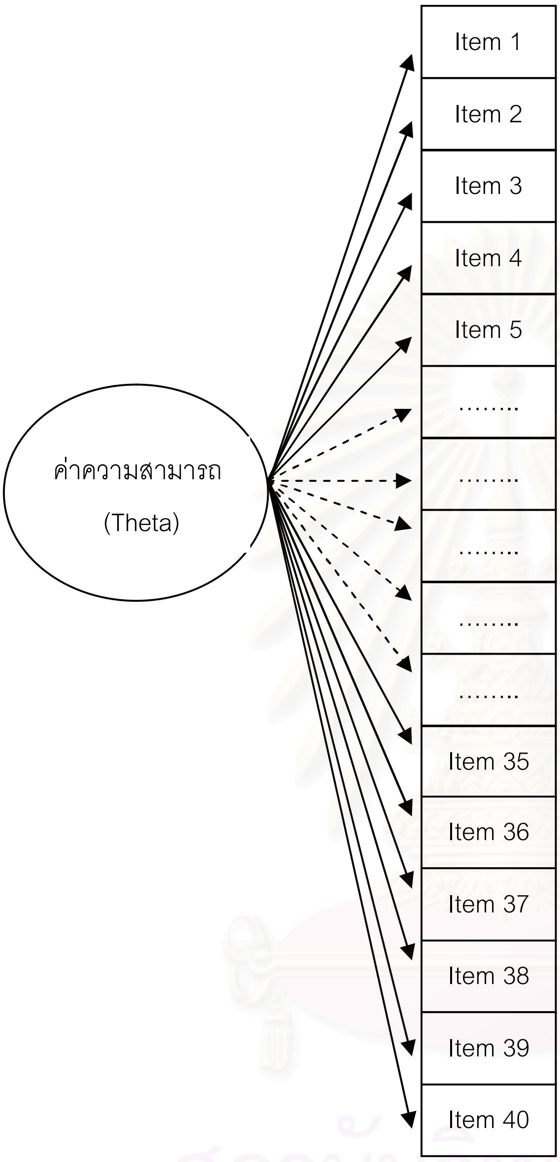
ผู้วิจัยนำข้อมูลผลการตอบแบบทดสอบมาจัดกระทำตามปัจจัยที่ศึกษา 2 ตัว คือ ความยาวของแบบสอบ 2 ระดับ และขนาดกลุ่มตัวอย่าง 3 ระดับ รวมทั้งหมด 6 เงื่อนไข (2x3) แล้ววิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ 2 วิธี คือ วิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ยกับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก

5.1 ผลการเปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย กับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก

5.1.1 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

นำเสนอโมเดลการวิเคราะห์สำหรับแบบสอบ 40 ข้อ ได้ดังนี้



ภาพที่ 8 โมเดลการวิเคราะห์สำหรับแบบสอบ 40 ข้อ

เมื่อพิจารณาโมเดลดังกล่าวแล้ว พบว่ามีค่าดัชนีวัดความกลมกลืนของโมเดลดังนี้

ค่าไค-สแควร์มีค่า 76.38 (P = 0.023) ดัชนี RMSEA = 0.143, ดัชนี RMR = 0.158, ดัชนี GFI = 0.811 และดัชนี AGFI = 0.793 ผลการทดสอบแสดงว่าโมเดลไม่สอดคล้องกับข้อมูลเชิงประจักษ์ ดังนั้นขั้นตอนถัดไป คือการเปรียบเทียบเมทริกซ์พารามิเตอร์ LX ของทั้ง 2 กลุ่ม ซึ่งเป็นเมทริกซ์ที่มีเงื่อนไขบังคับมากที่สุด หากพบว่าแตกต่างกันอย่างมีนัยสำคัญ แสดงว่าข้อสอบข้อนั้นทำหน้าที่ต่างกัน

ตารางที่ 19 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ยเมื่อแบบสอบมีความยาว 40 ข้อ และกลุ่มตัวอย่างจำนวน 1,000 คน

ข้อ	ค่าเมทริกซ์ของกลุ่ม		ข้อ	ค่าเมทริกซ์ของกลุ่ม	
	LX อ้างอิง	LXเปรียบเทียบ		LX อ้างอิง	LXเปรียบเทียบ
1	1.00	1.00	21	1.54	1.49
2	1.28	1.21	22	1.08	1.11
3	1.16	0.97	23*	1.56	1.30
4*	1.25	1.03	24*	0.57	0.85
5	1.34	1.44	25	1.47	1.37
6	1.40	1.27	26	1.07	1.10
7	1.51	1.54	27*	1.32	1.02
8	1.64	1.59	28*	1.12	0.75
9	1.67	1.78	29	1.23	1.30
10	1.71	1.73	30	1.42	1.52
11	1.45	1.52	31	0.53	0.57
12	1.57	1.42	32	1.33	1.42
13	1.50	1.56	33*	1.23	0.50
14	1.12	0.97	34*	1.32	1.54
15	1.43	1.27	35*	0.73	0.93
16	1.45	1.48	36*	0.86	0.44
17*	1.43	1.23	37	1.01	0.95
18*	1.45	1.22	38	0.87	0.71
19	1.37	1.46	39	0.83	0.90
20	1.22	1.41	40	0.78	0.94

* $p<.05$

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย เมื่อแบบสอบมีความยาว 40 ข้อ และกลุ่มตัวอย่างจำนวน 1,000 คน พิจารณาค่าเมทริกซ์พารามิเตอร์ LX ของทั้งสองกลุ่ม หากพบว่าค่าเมทริกซ์พารามิเตอร์ LX เท่ากันอย่างมีนัยสำคัญทั้งสองกลุ่ม แสดงว่าข้อมูลสอดคล้องกับข้อมูลเชิงประจักษ์ ดังนั้น ข้อสอบที่ถูกระบุว่าทำหน้าที่ต่างกันมีทั้งหมด 11 ข้อ ได้แก่ ข้อที่ 4, 17, 18, 23, 24, 27, 28, 33, 34, 35 และ 36

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 20 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ยเมื่อแบบสอบมีความยาว 40 ข้อ และกลุ่มตัวอย่างจำนวน 500 คน

ข้อ	ค่าเมทริกซ์ของกลุ่ม		ข้อ	ค่าเมทริกซ์ของกลุ่ม	
	LX อ้างอิง	LXเปรียบเทียบ		LX อ้างอิง	LXเปรียบเทียบ
1	1.00	1.00	21*	1.56	1.36
2	1.26	1.07	22	1.02	1.00
3*	1.13	0.87	23*	1.43	1.17
4	1.22	1.04	24	0.63	0.82
5	1.30	1.37	25	1.44	1.35
6	1.34	1.15	26	0.90	1.02
7	1.45	1.47	27*	1.26	0.96
8	1.63	1.49	28*	1.18	0.86
9	1.64	1.66	29	1.10	1.19
10	1.71	1.56	30	1.32	1.49
11	1.43	1.48	31	0.43	0.55
12	1.49	1.26	32	1.28	1.34
13	1.43	1.43	33*	1.19	0.55
14	1.03	1.00	34	1.27	1.37
15	1.33	1.27	35*	0.64	0.77
16	1.32	1.44	36*	0.76	0.50
17	1.32	1.27	37	0.76	0.81
18*	1.39	1.13	38*	0.86	0.63
19	1.22	1.29	39	0.78	0.84
20	1.22	1.36	40	0.84	0.73

* $p<.05$

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย เมื่อแบบสอบมีความยาว 40 ข้อ และกลุ่มตัวอย่างจำนวน 500 คน พิจารณาค่าเมทริกซ์พารามิเตอร์ LX ของทั้งสองกลุ่ม หากพบว่าค่าเมทริกซ์พารามิเตอร์ LX เท่ากันอย่างมีนัยสำคัญทั้งสองกลุ่ม แสดงว่าข้อมูลสอดคล้องกับข้อมูลเชิงประจักษ์ ดังนั้น ข้อสอบที่ถูกระบุว่าทำหน้าที่ต่างกันมีทั้งหมด 10 ข้อ ได้แก่ ข้อที่ 3, 18, 21, 23, 27, 28, 33, 35, 36 และ 38

ตารางที่ 21 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ยเมื่อแบบสอบมีความยาว 40 ข้อ และกลุ่มตัวอย่างจำนวน 200 คน

ข้อ	ค่าเมทริกซ์ของกลุ่ม		ข้อ	ค่าเมทริกซ์ของกลุ่ม	
	LX อ้างอิง	LXเปรียบเทียบ		LX อ้างอิง	LXเปรียบเทียบ
1	1.00	1.00	21	1.41	1.36
2	1.10	0.99	22	0.98	0.86
3	1.07	0.93	23*	1.43	1.09
4*	1.34	0.95	24	0.75	0.74
5	1.37	1.35	25	1.44	1.38
6	1.34	1.23	26	1.01	1.01
7	1.45	1.59	27*	1.36	0.84
8	1.43	1.60	28*	1.18	0.59
9*	1.52	1.76	29	1.02	1.11
10	1.63	1.65	30	1.26	1.37
11	1.34	1.58	31	0.44	0.55
12*	1.60	1.32	32	1.06	1.25
13	1.50	1.40	33*	1.03	0.63
14	0.97	1.15	34	1.19	1.22
15	1.11	1.22	35	0.72	0.74
16	1.26	1.15	36*	0.73	0.30
17*	1.12	1.34	37	1.06	0.81
18	1.39	1.20	38*	0.96	0.55
19	1.08	1.33	39*	0.72	0.09
20*	1.11	1.46	40	0.73	0.85

* $p<.05$

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย เมื่อแบบสอบมีความยาว 40 ข้อ และกลุ่มตัวอย่างจำนวน 200 คน พิจารณาค่าเมทริกซ์พารามิเตอร์ LX ของทั้งสองกลุ่ม หากพบว่าค่าเมทริกซ์พารามิเตอร์ LX เท่ากันอย่างมีนัยสำคัญทั้งสองกลุ่ม แสดงว่าข้อมูลสอดคล้องกับข้อมูลเชิงประจักษ์ ดังนั้น ข้อสอบที่ถูกระบุว่าทำหน้าที่ต่างกันมีทั้งหมด 12 ข้อ ได้แก่ ข้อที่ 4, 9, 12, 17, 20, 23, 27, 28, 33, 36, 38 และ 39

ตารางที่ 22 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ยเมื่อแบบสอบมีความยาว 30 ข้อ และกลุ่มตัวอย่างจำนวน 1,000 คน

ข้อ	ค่าเมทริกซ์ของกลุ่ม		ข้อ	ค่าเมทริกซ์ของกลุ่ม	
	LX อ้างอิง	LXเปรียบเทียบ		LX อ้างอิง	LXเปรียบเทียบ
1	1.00	1.00	16	1.56	1.47
2*	1.16	0.94	17	1.08	1.07
3*	1.25	1.00	18*	1.55	1.26
4	1.35	1.39	19*	0.57	0.83
5	1.40	1.24	20	1.05	1.05
6	1.47	1.45	21*	1.31	0.98
7	1.67	1.74	22*	1.11	0.72
8	1.69	1.70	23	1.23	1.25
9	1.46	1.49	24	0.51	0.56
10	1.51	1.53	25	1.31	1.38
11	1.10	0.96	26	0.70	0.87
12	1.46	1.46	27*	0.85	0.38
13*	1.44	1.21	28	1.00	0.93
14*	1.44	1.18	29	0.83	0.87
15	1.38	1.40	30	0.79	0.93

* $p<.05$

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย เมื่อแบบสอบมีความยาว 30 ข้อ และกลุ่มตัวอย่างจำนวน 1,000 คน พิจารณาค่าเมทริกซ์พารามิเตอร์ LX ของทั้งสองกลุ่ม หากพบว่าค่าเมทริกซ์พารามิเตอร์ LX เท่ากันอย่างมีนัยสำคัญทั้งสองกลุ่ม แสดงว่าข้อมูลสอดคล้องกับข้อมูลเชิงประจักษ์ ดังนั้น ข้อสอบที่ถูกระบุว่าทำหน้าที่ต่างกันมีทั้งหมด 9 ข้อ ได้แก่ ข้อที่ 2, 3, 13, 14, 18, 19, 21, 22, และ 27

ตารางที่ 23 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ยเมื่อแบบสอบมีความยาว 30 ข้อ และกลุ่มตัวอย่างจำนวน 500 คน

ข้อ	ค่าเมทริกซ์ของกลุ่ม		ข้อ	ค่าเมทริกซ์ของกลุ่ม	
	LX อ้างอิง	LXเปรียบเทียบ		LX อ้างอิง	LXเปรียบเทียบ
1	1.00	1.00	16*	1.54	1.33
2*	1.07	0.84	17	0.95	0.93
3*	1.23	0.97	18*	1.36	1.12
4	1.29	1.31	19	0.59	0.76
5	1.31	1.11	20	0.83	0.94
6	1.32	1.33	21*	1.21	0.92
7	1.61	1.59	22*	1.11	0.84
8	1.65	1.52	23	1.06	1.13
9	1.40	1.43	24	0.40	0.52
10	1.42	1.36	25	1.21	1.29
11	0.92	0.96	26	0.63	0.67
12	1.31	1.40	27*	0.74	0.43
13	1.28	1.23	28	0.83	0.80
14*	1.33	1.07	29	0.77	0.78
15	1.13	1.22	30	0.83	0.72

* $p<.05$

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย เมื่อแบบสอบมีความยาว 30 ข้อ และกลุ่มตัวอย่างจำนวน 500 คน พิจารณาค่าเมทริกซ์พารามิเตอร์ LX ของทั้งสองกลุ่ม หากพบว่าค่าเมทริกซ์พารามิเตอร์ LX เท่ากันอย่างมีนัยสำคัญทั้งสองกลุ่ม แสดงว่าข้อมูลสอดคล้องกับข้อมูลเชิงประจักษ์ ดังนั้น ข้อสอบที่ถูกระบุว่าทำหน้าที่ต่างกันมีทั้งหมด 8 ข้อ ได้แก่ ข้อที่ 2, 3, 14, 16, 18, 21, 22, และ 27

ตารางที่ 24 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ยเมื่อแบบสอบมีความยาว 30 ข้อ และกลุ่มตัวอย่างจำนวน 200 คน

ข้อ	ค่าเมทริกซ์ของกลุ่ม		ข้อ	ค่าเมทริกซ์ของกลุ่ม	
	LX อ้างอิง	LXเปรียบเทียบ		LX อ้างอิง	LXเปรียบเทียบ
1	1.00	1.00	16	1.41	1.35
2	1.10	0.87	17	1.01	0.91
3*	1.35	0.94	18*	1.43	1.09
4	1.42	1.31	19	0.71	0.74
5	1.38	1.26	20	0.98	0.96
6	1.39	1.48	21*	1.36	0.79
7	1.51	1.66	22*	1.23	0.70
8	1.55	1.56	23	1.10	1.13
9	1.31	1.56	24	0.44	0.48
10	1.48	1.41	25	1.04	1.19
11	0.97	1.12	26	0.67	0.60
12	1.37	1.47	27*	0.73	0.16
13	1.18	1.36	28	1.06	0.83
14	1.38	1.21	29*	0.73	1.02
15	1.09	1.34	30	0.78	0.81

* $p<.05$

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย เมื่อแบบสอบมีความยาว 30 ข้อ และกลุ่มตัวอย่างจำนวน 200 คน พิจารณาค่าเมทริกซ์พารามิเตอร์ LX ของทั้งสองกลุ่ม หากพบว่าค่าเมทริกซ์พารามิเตอร์ LX เท่ากันอย่างมีนัยสำคัญทั้งสองกลุ่ม แสดงว่าข้อมูลสอดคล้องกับข้อมูลเชิงประจักษ์ ดังนั้น ข้อสอบที่ถูกระบุว่าทำหน้าที่ต่างกันมีทั้งหมด 6 ข้อ ได้แก่ ข้อที่ 3, 18, 21, 22, 27และ 29

5.1.2 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก

ตารางที่ 25 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติกเมื่อแบบสอบมีความยาว 40 ข้อ และกลุ่มตัวอย่างจำนวน 1000 คน

ข้อ	chi-square	ข้อ	chi-square
1	1.17	21	3.44
2	0.24	22	1.50
3	6.23	23	9.26*
4	9.81*	24	6.24
5	4.81	25	2.27
6	5.07	26	0.58
7	1.85	27	13.26*
8	9.14*	28	0.47
9	0.63	29	2.61
10	3.35	30	1.73
11	2.18	31	0.85
12	7.24*	32	2.89
13	5.38	33	2.18
14	4.87	34	1.37
15	7.81*	35	3.65
16	1.94	36	7.71*
17	11.25*	37	2.57
18	12.21*	38	4.78
19	4.13	39	0.65
20	2.25	40	2.68

* $p<.05$

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ผลการตรวจสอบตามตารางที่ 25 พบว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี LDFA ตามแนวคิดของทฤษฎีการตอบข้อสอบด้วยการทดสอบค่าไค-สแควร์ เมื่อกลุ่มตัวอย่าง 1,000 คน และแบบสอบความยาว 40 ข้อ พบข้อสอบที่ทำหน้าที่ต่างกัน (DIF) จำนวน 9 ข้อ ได้แก่ ข้อ 4, 8, 12, 15, 17, 18, 23, 27 และ 36

ตารางที่ 26 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติกเมื่อแบบสอบมีความยาว 40 ข้อ และกลุ่มตัวอย่างจำนวน 500 คน

ข้อ	chi-square	ข้อ	chi-square
1	1.41	21	3.56
2	0.13	22	1.87
3	10.36*	23	8.56*
4	8.91*	24	6.10
5	4.73	25	2.34
6	5.19	26	0.81
7	1.98	27	12.14*
8	9.24*	28	0.73
9	0.46	29	2.91
10	3.46	30	1.38
11	2.57	31	0.93
12	7.35*	32	2.76
13	5.49	33	12.03*
14	4.99	34	1.20
15	3.02	35	9.18*
16	2.31	36	7.99*
17	10.22*	37	3.09
18	4.22	38	4.47
19	6.31	39	0.51
20	3.25	40	2.36

* $p<.05$

สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

ผลการตรวจสอบตามตารางที่ 26 พบว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี LDFA ตามแนวคิดของทฤษฎีการตอบข้อสอบด้วยการทดสอบค่าไค-สแควร์ เมื่อกลุ่มตัวอย่าง 500 คน และแบบสอบความยาว 40 ข้อ พบข้อสอบที่ทำหน้าที่ต่างกัน (DIF) จำนวน 10 ข้อ ได้แก่ ข้อ 3, 4, 8, 12, 17, 23, 27, 33, 35 และ 36

ตารางที่ 27 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติกเมื่อแบบสอบมีความยาว 40 ข้อ และกลุ่มตัวอย่างจำนวน 200 คน

ข้อ	chi-square	ข้อ	chi-square
1	1.26	21	3.55
2	0.54	22	1.56
3	9.34*	23	9.78*
4	10.12*	24	6.84
5	4.32	25	2.69
6	5.11	26	0.87
7	1.98	27	5.16
8	10.21*	28	8.87*
9	0.98	29	2.79
10	3.49	30	1.93
11	3.14	31	1.05
12	6.25	32	2.97
13	6.34	33	2.29
14	5.01	34	1.57
15	5.69	35	8.26*
16	2.03	36	7.99*
17	12.14*	37	3.16
18	4.58	38	5.29
19	11.01*	39	9.18*
20	2.10	40	2.97

* p<.05

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ผลการตรวจสอบตามตารางที่ 27 พบว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี LDFA ตามแนวคิดของทฤษฎีการตอบข้อสอบด้วยการทดสอบค่าไค-สแควร์ เมื่อกลุ่มตัวอย่าง 200 คน และแบบสอบความยาว 40 ข้อ พบข้อสอบที่ทำหน้าที่ต่างกัน (DIF) จำนวน 10 ข้อ ได้แก่ ข้อ 3, 4, 8, 17, 19, 23, 28, 35, 36 และ 39

ตารางที่ 28 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติกเมื่อแบบสอบมีความยาว 30 ข้อ และกลุ่มตัวอย่างจำนวน 1000 คน

ข้อ	chi-square	ข้อ	chi-square
1	1.68	16	5.62
2	1.25	17	0.42
3	11.74*	18	8.97*
4	4.47	19	4.32
5	4.54	20	5.18
6	6.36	21	7.68*
7	3.86	22	6.37
8	4.59	23	4.94
9	5.51	24	2.76
10	6.25	25	6.14
11	5.32	26	2.49
12	6.14	27	7.58*
13	9.23*	28	5.59
14	10.20*	29	5.63
15	5.09	30	3.29

* p<.05

ผลการตรวจสอบตามตารางที่ 28 พบว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี LDFA ตามแนวคิดของทฤษฎีการตอบข้อสอบด้วยการทดสอบค่าไค-สแควร์ เมื่อกลุ่มตัวอย่าง 1000 คน และแบบสอบความยาว 30 ข้อ พบข้อสอบที่ทำหน้าที่ต่างกัน (DIF) จำนวน 6 ข้อ ได้แก่ ข้อ 3, 13, 14, 18, 21และ 27

ตารางที่ 29 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติกเมื่อแบบสอบมีความยาว 30 ข้อ และกลุ่มตัวอย่างจำนวน 500 คน

ข้อ	chi-square	ข้อ	chi-square
1	1.52	16	5.25
2	1.15	17	0.98
3	10.50*	18	8.95*
4	4.31	19	4.82
5	4.08	20	5.13
6	6.13	21	7.41*
7	3.42	22	6.39
8	4.59	23	4.61
9	5.72	24	2.69
10	6.75	25	6.01
11	5.12	26	2.75
12	6.51	27	7.49*
13	8.54*	28	5.47
14	9.14*	29	5.49
15	5.07	30	3.21

* $p<.05$

ผลการตรวจสอบตามตารางที่ 29 พบว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี LDFA ตามแนวคิดของทฤษฎีการตอบข้อสอบด้วยการทดสอบค่าไค-สแควร์ เมื่อกลุ่มตัวอย่าง 500 คน และแบบสอบความยาว 30 ข้อ พบข้อสอบที่ทำหน้าที่ต่างกัน (DIF) จำนวน 6 ข้อ ได้แก่ ข้อ 3, 13, 14, 18, 21 และ 27

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 30 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติกเมื่อแบบสอบมีความยาว 30 ข้อ และกลุ่มตัวอย่างจำนวน 200 คน

ข้อ	chi-square	ข้อ	chi-square
1	1.48	16	5.69
2	1.26	17	0.78
3	9.38*	18	9.12*
4	4.87	19	8.65*
5	4.02	20	5.11
6	6.02	21	7.23*
7	3.24	22	6.57
8	4.87	23	4.87
9	5.65	24	2.74
10	6.32	25	6.00
11	5.04	26	2.63
12	6.49	27	7.74*
13	6.14	28	5.85
14	9.89*	29	5.03
15	5.25	30	3.14

* p<.05

ผลการตรวจสอบตามตารางที่ 30 พบว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี LDFA ตามแนวคิดของทฤษฎีการตอบข้อสอบด้วยการทดสอบค่าไค-สแควร์ เมื่อกลุ่มตัวอย่าง 200 คน และแบบสอบความยาว 30 ข้อ พบข้อสอบที่ทำหน้าที่ต่างกัน (DIF) จำนวน 6 ข้อ ได้แก่ ข้อ 3, 14, 18, 19, 21และ 27

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

จะพบว่าการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้ง 3 วิธี ในเงื่อนไขต่าง ๆ จะพบข้อสอบที่ทำหน้าที่ต่างกันเป็นจำนวนข้อที่ต่างกัน ดังนั้นก่อนที่จะทำการเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ผู้วิจัยขอสรุปถึงข้อคำถามที่ปรากฏอยู่ในแบบสอบที่มีความยาว 40 ข้อและ 30 ข้อ ในกรณีที่ว่าแต่ละข้อคำถามเป็นข้อเดียวกันหรือไม่ ดังนี้

ตารางที่ 31 สรุปข้อคำถามที่ปรากฏอยู่ในแบบสอบที่มีความยาว 40 ข้อและ 30 ข้อ

แบบสอบความยาว 40 ข้อ	**แบบสอบความยาว 30 ข้อ**	**แบบสอบความยาว 40 ข้อ**	**แบบสอบความยาว 30 ข้อ**
ข้อ 1	ข้อ 1	ข้อ 21	ข้อ 16
ข้อ 2	-	ข้อ 22	ข้อ 17
ข้อ 3	ข้อ 2	ข้อ 23*	ข้อ 18*
ข้อ 4*	ข้อ 3*	ข้อ 24	ข้อ 19
ข้อ 5	ข้อ 4	ข้อ 25	-
ข้อ 6	ข้อ 5	ข้อ 26	ข้อ 20
ข้อ 7	ข้อ 6	ข้อ 27	ข้อ 21*
ข้อ 8*	-	ข้อ 28	ข้อ 22
ข้อ 9	ข้อ 7	ข้อ 29	ข้อ 23
ข้อ 10	ข้อ 8	ข้อ 30	-
ข้อ 11	ข้อ 9	ข้อ 31	ข้อ 24
ข้อ 12*	-	ข้อ 32	ข้อ 25
ข้อ 13	ข้อ 10	ข้อ 33	-
ข้อ 14	ข้อ 11	ข้อ 34	-
ข้อ 15*	-	ข้อ 35	ข้อ 26
ข้อ 16	ข้อ 12	ข้อ 36*	ข้อ 27*
ข้อ 17*	ข้อ 13*	ข้อ 37	ข้อ 28
ข้อ 18*	ข้อ 14*	ข้อ 38	-
ข้อ 19	ข้อ 15	ข้อ 39	ข้อ 29
ข้อ 20	-	ข้อ 40	ข้อ 30

* ข้อที่มีการทำหน้าที่ต่างกันเมื่อตรวจสอบด้วยวิธี GPCM

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 32 สรุปผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในทุกเงื่อนไข

วิธีที่ใช้ในการตรวจสอบ	เงื่อนไข		ข้อสอบที่ทำหน้าที่ต่างกัน		จำนวนข้อทั้งหมด	ร้อยละ
	ขนาดกลุ่มตัวอย่าง (คน)	ความยาวแบบสอบ (ข้อ)	สอดคล้องกับวิธีเกณฑ์ (GPCM) (เทียบกับ 40 ข้อ)	ไม่สอดคล้องกับวิธีเกณฑ์ (GPCM)		
GPCM	1000	40	4, 8,12, 15, 17, 18, 23, 36		8	20.00
	1000	30	3 (4), 13(17), 14(18), 18(23), 21(27), 27(36)		6	20.00
MACS	1000	40	4, 17, 18, 23, 36	24, 27, 28, 33, 34, 35	11	27.50
	1000	30	3(4), 13(17), 14(18), 18(23), 21(27), 27(36)	2(3), 19(24), 22(28)	9	30.00
	500	40	18, 23, 36	3, 21, 27, 28, 33, 35, 38	10	25.00
	500	30	3(4), 14(18), 18(23), 21(27), 27(36)	2(3), 16(21), 22(27)	8	26.67
	200	40	4, 12, 17, 23, 36	9, 20, 27, 28, 33, 38, 39	12	30.00
	200	30	3(4), 21(27), 27(36)	18(23), 22(28), 29(39)	6	20.00
LDFA	1000	40	4, 8, 12, 15, 17, 18, 23, 36	27	9	22.50
	1000	30	3(4), 13(17), 14(18), 18(23), 21(27), 27(36)		6	20.00
	500	40	4, 8, 12, 17, 23, 36	3, 27, 33, 35	10	25.00
	500	30	3(4), 13(17), 14(18), 18(23), 21(27), 27(36)		6	20.00
	200	40	4, 8, 17, 23, 36	3, 19, 28, 35, 39	10	25.00
	200	30	3(4), 14(18), 18(23), 21(27), 27(36)	19(24)	6	20.00

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

จากตารางที่ 32 เป็นการเปรียบเทียบจำนวนข้อสอบที่ถูกระบุว่าทำหน้าที่ต่างกันซึ่งตรวจสอบด้วยวิธีต่าง ๆ ในเงื่อนไขที่กำหนด ปรากฏว่า วิธี GPCM ซึ่งเป็นวิธีที่ใช้เป็นเกณฑ์ในการตรวจสอบประสิทธิภาพของวิธีที่นำมาเปรียบเทียบ ในเงื่อนไขขนาดกลุ่มตัวอย่าง 1,000 คน ความยาวแบบสอบ 40 ข้อ ตรวจพบข้อสอบทำหน้าที่ต่างกัน จำนวน 8 ข้อ (20.00%) และในเงื่อนไขขนาดกลุ่มตัวอย่าง 1,000 คน ความยาวแบบสอบ 30 ข้อ ตรวจพบข้อสอบทำหน้าที่ต่างกัน จำนวน 6 ข้อ (20.00%)

สำหรับวิธี MACS เมื่อขนาดกลุ่มตัวอย่าง 1,000 คน ความยาวแบบสอบ 40 ข้อ ตรวจพบข้อสอบทำหน้าที่ต่างกัน จำนวน 11 ข้อ (27.50%) เมื่อขนาดกลุ่มตัวอย่าง 1,000 คน ความยาวแบบสอบ 30 ข้อ ตรวจพบข้อสอบทำหน้าที่ต่างกัน จำนวน 9 ข้อ (30.00%) เมื่อขนาดกลุ่มตัวอย่าง 500 คน ความยาวแบบสอบ 40 ข้อ ตรวจพบข้อสอบทำหน้าที่ต่างกัน จำนวน 10 ข้อ (25.00%) เมื่อขนาดกลุ่มตัวอย่าง 500 คน ความยาวแบบสอบ 30 ข้อ ตรวจพบข้อสอบทำหน้าที่ต่างกัน จำนวน 8 ข้อ (26.67%) เมื่อขนาดกลุ่มตัวอย่าง 200 คน ความยาวแบบสอบ 40 ข้อ ตรวจพบข้อสอบทำหน้าที่ต่างกัน จำนวน 12 ข้อ (30.00%) และเมื่อขนาดกลุ่มตัวอย่าง 200 คน ความยาวแบบสอบ 30 ข้อ ตรวจพบข้อสอบทำหน้าที่ต่างกัน จำนวน 6 ข้อ (20.00%)

สำหรับวิธี LDFA เมื่อขนาดกลุ่มตัวอย่าง 1,000 คน ความยาวแบบสอบ 40 ข้อ ตรวจพบข้อสอบทำหน้าที่ต่างกัน จำนวน 9 ข้อ (22.50%) เมื่อขนาดกลุ่มตัวอย่าง 1,000 คน ความยาวแบบสอบ 30 ข้อ ตรวจพบข้อสอบทำหน้าที่ต่างกัน จำนวน 6 ข้อ (20.00%) เมื่อขนาดกลุ่มตัวอย่าง 500 คน ความยาวแบบสอบ 40 ข้อ ตรวจพบข้อสอบทำหน้าที่ต่างกัน จำนวน 10 ข้อ (25.00%) เมื่อขนาดกลุ่มตัวอย่าง 500 คน ความยาวแบบสอบ 30 ข้อ ตรวจพบข้อสอบทำหน้าที่ต่างกัน จำนวน 6 ข้อ (20.00%) เมื่อขนาดกลุ่มตัวอย่าง 200 คน ความยาวแบบสอบ 40 ข้อ ตรวจพบข้อสอบทำหน้าที่ต่างกัน จำนวน 10 ข้อ (25.00%) และเมื่อขนาดกลุ่มตัวอย่าง 200 คน ความยาวแบบสอบ 30 ข้อ ตรวจพบข้อสอบทำหน้าที่ต่างกัน จำนวน 6 ข้อ (20.00%)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

5.2 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย กับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก

หลังจากวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีทั้งสองแล้ว นำผลการวิเคราะห์ที่ได้มาคำนวณหาค่าอัตราความคลาดเคลื่อนประเภทที่ 1 (Type I Error Rate) ของแต่ละวิธี เพื่อเปรียบเทียบประสิทธิภาพของแต่ละวิธี โดยที่ผู้วิจัยตรวจสอบทั้ง 2 วิธีกับเกณฑ์ที่ระดับ 10% ภายใต้เงื่อนไขปัจจัยที่ศึกษา หากมีค่าอยู่ในเกณฑ์ถือว่าวิธีดังกล่าวสามารถควบคุมค่าอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ ซึ่งแสดงดังนี้

ตารางที่ 33 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย กับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก

เงื่อนไข		วิธี MACS	วิธี LDFA
ขนาดกลุ่มตัวอย่าง (คน)	ความยาวแบบสอบ (ข้อ)		
1000	40	0.188	0.031
1000	30	0.125	0.000
500	40	0.219	0.125
500	30	0.125	0.000
200	40	0.219	0.156
200	30	0.083	0.042

จากตารางที่ 33 เป็นผลการวิเคราะห์และเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย กับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติกตามเงื่อนไขทั้ง 6 เงื่อนไข พบว่า

ภายใต้เงื่อนไขของขนาดกลุ่มตัวอย่างและความยาวแบบสอบ พบว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธี MACS มีค่าอยู่ในช่วง 8.3% ถึง 21.9% จะเห็นว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธี MACS มีค่าอยู่ในเกณฑ์ที่ยอมรับได้เพียง 1 เงื่อนไข นั่นคือที่เงื่อนไขขนาดกลุ่มตัวอย่าง 200 คนและความยาวแบบสอบ 30 ข้อ

ภายใต้เงื่อนไขของขนาดกลุ่มตัวอย่างและความยาวแบบสอบ พบว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธี LDFA มีค่าอยู่ในช่วง 0% ถึง 15.6% จะเห็นว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธี LDFA เกือบทุกเงื่อนไขมีค่าอยู่ในเกณฑ์ที่ยอมรับได้ ยกเว้น

เงื่อนไขขนาดกลุ่มตัวอย่าง 500 คนและความยาวแบบสอบ 40 ข้อ และเงื่อนไขขนาดกลุ่มตัวอย่าง 200 คนและความยาวแบบสอบ 40 ข้อ

5.3 ผลการเปรียบเทียบอำนาจการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย กับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก

หลังจากวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีทั้งสองแล้ว นำผลการวิเคราะห์ที่ได้มาคำนวณหาค่าอำนาจการทดสอบ (Power of test) ของแต่ละวิธี เพื่อเปรียบเทียบประสิทธิภาพของแต่ละวิธีซึ่งแสดงดังนี้

ตารางที่ 34 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย กับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก

เงื่อนไข		วิธี MACS	วิธี LDFA
ขนาดกลุ่มตัวอย่าง (คน)	ความยาวแบบสอบ (ข้อ)		
1000	40	0.625	1.000
1000	30	1.000	1.000
500	40	0.375	0.750
500	30	0.833	1.000
200	40	0.625	0.625
200	30	0.667	0.833

จากตารางที่ 34 เป็นผลการวิเคราะห์และเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย กับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติกตามเงื่อนไขทั้ง 6 เงื่อนไข พบว่า

วิธี LDFA เมื่อขนาดกลุ่มตัวอย่าง 1,000 คน ความยาวแบบสอบ 40 และ 30 ข้อ และเมื่อขนาดกลุ่มตัวอย่าง 500 คน ความยาวแบบสอบ 30 ข้อ รวมทั้งวิธี MACS เมื่อขนาดกลุ่มตัวอย่าง 1,000 คน ความยาวแบบสอบ 30 ข้อ มีค่าอำนาจการทดสอบสูงที่สุด เท่ากับ 1.000 นั่นคือ มีการตรวจสอบพบว่าข้อสอบทำหน้าที่ต่างกันถูกต้องตรงกับวิธี GPCM ซึ่งเป็นวิธีเกณฑ์ร้อยละ 100 รองลงมาคือ วิธี MACS เมื่อขนาดกลุ่มตัวอย่าง 500 คน ความยาวแบบสอบ 30 ข้อ และวิธี LDFA เมื่อขนาดกลุ่มตัวอย่าง 200 คน ความยาวแบบสอบ 30 ข้อ มีค่าอำนาจการทดสอบเท่ากับ 0.833

ส่วนวิธีที่มีค่าอำนาจการทดสอบต่ำที่สุด คือ วิธี MACS เมื่อขนาดกลุ่มตัวอย่าง 500 คน ความยาวแบบสอบ 40 ข้อ

หลังจากการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย กับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติกตามเงื่อนไขทั้ง 6 เงื่อนไขแล้วนำผลการตรวจสอบของแต่ละวิธีมาหาประสิทธิภาพในการตรวจสอบ ซึ่งดูจากค่าอำนาจการทดสอบ (Power of the test) และอัตราความคลาดเคลื่อนประเภทที่ 1 (Type I Error Rate) แล้วสามารถสรุปผลการเปรียบเทียบประสิทธิภาพได้ดังนี้

กรณีที่แบบสอบมีความยาว 40 ข้อ วิธี MACS ให้ค่าอำนาจการทดสอบหรือความถูกต้องในการตรวจสอบมีค่าเฉลี่ยสูงสุดเมื่อขนาดกลุ่มตัวอย่างเป็น 1,000 และ 200 คน คิดเป็นร้อยละ 62.5 แต่ให้ค่าความถูกต้องในการตรวจสอบน้อยที่สุดเมื่อขนาดกลุ่มตัวอย่างเป็น 500 คน คิดเป็นร้อยละ 37.5 สำหรับค่าอัตราความคลาดเคลื่อนประเภทที่ 1 หรือความผิดพลาดในการตรวจสอบมีค่าเฉลี่ยสูงสุดเมื่อขนาดกลุ่มตัวอย่างเป็น 1,000 คน คิดเป็นร้อยละ 18.8 และให้ค่าความผิดพลาดในการตรวจสอบเท่ากันเมื่อขนาดกลุ่มตัวอย่างเป็น 500 และ 200 คน คิดเป็นร้อยละ 21.9 สำหรับวิธี LDFA ให้ค่าอำนาจการทดสอบหรือความถูกต้องในการตรวจสอบมีค่าเฉลี่ยเมื่อขนาดกลุ่มตัวอย่างเป็น 1,000, 500 และ 200 คน คิดเป็นร้อยละ 100, 75 และ 62.5 ตามลำดับ และให้ค่าอัตราความคลาดเคลื่อนประเภทที่ 1 หรือความผิดพลาดในการตรวจสอบมีค่าเฉลี่ยเมื่อขนาดกลุ่มตัวอย่างเป็น 1,000, 500 และ 200 คน คิดเป็นร้อยละ 3.1, 12.5 และ 15.6 ตามลำดับเช่นกัน

กรณีที่แบบสอบมีความยาว 30 ข้อ วิธี MACS ให้ค่าอำนาจการทดสอบหรือความถูกต้องในการตรวจสอบมีค่าเฉลี่ยเมื่อขนาดกลุ่มตัวอย่างเป็น 1,000, 500 และ 200 คน คิดเป็นร้อยละ 100, 83.3 และ 66.7 ตามลำดับ และให้อัตราความคลาดเคลื่อนประเภทที่ 1 หรือความผิดพลาดในการตรวจสอบน้อยที่สุดเมื่อขนาดกลุ่มตัวอย่างเป็น 200 คิดเป็นร้อยละ 8.3 รองลงมาได้แก่ขนาดกลุ่มตัวอย่าง 1,000 และ 500 คนให้ค่าความผิดพลาดในการตรวจสอบเท่ากัน คิดเป็นร้อยละ 12.5 สำหรับวิธี LDFA ให้ค่าอำนาจการทดสอบหรือความถูกต้องในการตรวจสอบมีค่าเฉลี่ยสูงสุดเมื่อขนาดกลุ่มตัวอย่างเป็น 1,000 และ 500 คน คิดเป็นร้อยละ 100 และเมื่อขนาดกลุ่มตัวอย่างเป็น 200 คน ให้ความถูกต้องในการตรวจสอบน้อยที่สุด คิดเป็นร้อยละ 83.3 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 หรือความผิดพลาดในการตรวจสอบมีค่าน้อยที่สุดเมื่อขนาดกลุ่มตัวอย่างเป็น 1,000 และ 500 คน คิดเป็นร้อยละ 0 และคิดเป็นร้อยละ 4.2 เมื่อขนาดกลุ่มตัวอย่างเป็น 200 ค

บทที่ 5
สรุป อภิปรายผลและข้อเสนอแนะ

การวิจัยเรื่องนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาค ระหว่างวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ยกับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก โดยกำหนดวัตถุประสงค์เฉพาะในการศึกษาด้านต่าง ๆ ดังนี้ 1) เพื่อเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 2 วิธี ภายใต้เงื่อนไขของปัจจัยที่แตกต่างกันทางด้านความยาวของแบบสอบและขนาดกลุ่มตัวอย่างและ 2) เพื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 2 วิธี ภายใต้เงื่อนไขของปัจจัยที่แตกต่างกันทางด้านความยาวของแบบสอบและขนาดกลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการวิจัย คือ นักเรียนชั้นมัธยมศึกษาปีที่ 2 ปีการศึกษา 2550 โรงเรียนมัธยมศึกษา สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กรุงเทพมหานคร จำนวน 2,000 คน แบ่งเป็นนักเรียนชาย 1,000 คน และนักเรียนหญิง 1,000 คน ซึ่งได้มาจากการสุ่มแบบหลายขั้นตอน (Multi-stage Random Sampling) ตามเงื่อนไขขนาดกลุ่มตัวอย่างที่ได้กำหนดไว้เพื่อใช้เป็นประชากรเทียมในการสุ่มกลุ่มตัวอย่างขนาดต่าง ๆ โดยในขั้นแรกสุ่มได้จำนวนนักเรียนทั้งหมดทั้งหมด 2,098 คน โดยแบ่งเป็นชาย 1,045 คน และหญิง 1,053 คน จากนั้นผู้วิจัยได้นำมาคัดเลือกโดยเลือกกระดาษคำตอบของนักเรียนที่มีความบกพร่องและไม่สมบูรณ์ออก จากการคัดเลือกจะได้กระดาษคำตอบของนักเรียนที่มีความสมบูรณ์ สามารถนำมาวิเคราะห์ได้ทั้งหมด 1,806 คน โดยแบ่งเป็นชาย 892 คน และหญิง 914 คน

เครื่องมือที่ใช้ในการวิจัย คือ แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ด้านเหตุผล 5 ตัวเลือกที่มีคำตอบถูกหลายคำตอบ และให้ตัดสินทุกคำตอบ (Multiple True-false Answer) เรื่องจำนวนเต็ม เศษส่วน และทศนิยม ระดับชั้นมัธยมศึกษาตอนต้น คะแนนเต็มข้อละ 5 คะแนน จำนวน 50 ข้อ ของ สุนิสา จุ้ยม่วงศรี (2546) โดยผ่านการตรวจสอบคุณภาพด้านความตรงเชิงเนื้อหา ค่าความยาก ค่าอำนาจจำแนกและค่าความเที่ยง ได้ค่าดังนี้ ความตรงเชิงเนื้อหา ได้ค่า IOC เฉลี่ยทั้งฉบับของแบบทดสอบฉบับ X (30 ข้อ: มีข้อสอบร่วม 10 ข้อ) และฉบับ Y (30 ข้อ: มีข้อสอบร่วม 10 ข้อ) มีค่า 0.905 และ 0.900 ตามลำดับ ส่วนการตรวจสอบค่าความยากและอำนาจจำแนกได้ค่าความยากของข้อสอบอยู่ในช่วง 0.450 ถึง 0.874 และช่วง 0.478 ถึง 0.805 ในแบบทดสอบฉบับ X และ Y ตามลำดับ ค่าอำนาจจำแนกในช่วง 0.005 ถึง 0.625 และช่วง -0.147

ถึง 0.618 ในแบบทดสอบฉบับ X และ Y ตามลำดับ และได้ค่าสัมประสิทธิ์ความเชื่อมั่นของแบบทดสอบทั้งฉบับ โดยใช้สูตรของแอลฟา ครอนบาค เท่ากับ .936 และ .918 ตามลำดับ

ในการวิจัยครั้งนี้ดำเนินการวิเคราะห์ข้อมูล ดังนี้

1. คำนวณค่าสถิติพื้นฐาน เพื่อบรรยายลักษณะการแจกแจงของคะแนนจากแบบสอบวิชาคณิตศาสตร์ ด้วยโปรแกรมคอมพิวเตอร์ SPSS 13.0 for Windows

2. ตรวจสอบความเป็นเอกมิติ (Unidimensionality) ของแบบสอบ โดยการวิเคราะห์องค์ประกอบ (Factor Analysis) ด้วยการวิเคราะห์ตัวประกอบสำคัญ (Principal Component Analysis) และหมุนแกนด้วยวิธี Varimax ตามวิธีการของ Lord (1980) เพื่อตรวจสอบความเป็นมิติเดียวของแบบทดสอบ

3. วิเคราะห์คุณภาพของแบบสอบวิชาคณิตศาสตร์ ด้วยโปรแกรมคอมพิวเตอร์ SPSS 13.0 for Windows และโปรแกรม MULTILOG 7.03 โดยคำนวณค่าความเที่ยงแบบความสอดคล้องภายใน (Internal Consistency) ของแบบสอบแต่ละฉบับด้วยสูตรแอลฟาของครอนบาค (Cronbach's Alpha)

4. ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GPCM

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในการศึกษาครั้งนี้มี 3 วิธี คือ วิธี GPCM วิธี MACS และวิธี LDFA โดยจะใช้วิธี MACS และวิธี LDFA วิเคราะห์ในทุกเงื่อนไขของขนาดกลุ่มตัวอย่างและความยาวของแบบทดสอบ เพื่อเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) แต่วิธี GPCM จะใช้เป็นวิธีมาตรฐานในการระบุว่าข้อสอบข้อใดแสดงการทำหน้าที่ต่างกัน เพื่อใช้เป็นเกณฑ์ในการประเมินประสิทธิภาพการตรวจสอบของวิธีอื่น โดยวิธี GPCM จะใช้วิเคราะห์ในเงื่อนไขที่กลุ่มตัวอย่างเพศชายขนาด 1,000 คน ต่อกลุ่มตัวอย่างเพศหญิงขนาด 1,000 คน กับความยาวของแบบทดสอบ 2 ขนาด คือ 40 ข้อ และ 30 ข้อ เนื่องจากวิธี GPCM ซึ่งเป็นวิธีตามทฤษฎีการตอบสนอง (IRT) จะตรวจสอบได้อย่างมีประสิทธิภาพเมื่อใช้กลุ่มตัวอย่างขนาดใหญ่ ดังนั้น วิธีการนี้จึงวิเคราะห์เพียง 2 เงื่อนไขดังกล่าว มีขั้นตอนการวิเคราะห์ดังนี้

4.1. การประมาณค่าพารามิเตอร์ข้อสอบของกลุ่มตัวอย่างเพศชายและเพศหญิงกลุ่มละ 1,000 คน ที่สุ่มมาจากประชากรเทียมของแต่ละกลุ่มเพศ เมื่อแบบทดสอบมีขนาด 40 ข้อ และ 30 ข้อ โดยใช้โปรแกรม PARSCALE Version 2.2 พัฒนาขึ้นโดย Muraki & Bock (1991) ซึ่งใช้วิเคราะห์ข้อสอบที่ตรวจให้คะแนนแบบหลายค่า เพื่อประมาณค่าพารามิเตอร์ของข้อสอบ ตามแนวทฤษฎีการตอบข้อสอบ (IRT) ผลการวิเคราะห์จะได้แฟ้มข้อมูลของค่าประมาณพารามิเตอร์

ข้อสอบของผู้สอบกลุ่มเพศชายและหญิง เนื่องจากพารามิเตอร์ของข้อสอบดังกล่าวประมาณค่ามาจากการวิเคราะห์แยกกันระหว่างกลุ่มผู้สอบย่อย ดังนั้น ก่อนที่จะนำข้อมูลของทั้งสองกลุ่มมาตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจะต้องปรับเทียบสเกลพารามิเตอร์ของข้อสอบให้อยู่บนสเกลเมตริกซ์เดียวกัน

4.2 ปรับเทียบสเกลพารามิเตอร์ข้อสอบของทั้งสองกลุ่ม ด้วยการแปลงค่าเชิงเส้น (Linear Transformation) โดยใช้สูตรในการปรับเทียบ (Tate, 1999)

4.3 นำผลการปรับเทียบสเกลพารามิเตอร์ในข้อ มาตรวจสอบการหน้าที่ต่างกันของข้อสอบตามวิธี GPCM เพื่อใช้เป็นวิธีมาตรฐานในการระบุว่าข้อสอบข้อใดทำหน้าที่ต่างกัน ซึ่งสามารถตรวจการทำหน้าที่ต่างกันของข้อสอบได้ 3 วิธี คือ วิธีการทดสอบไค-สแควร์ของลอร์ด (Lord’s Chi-square, 1980) วิธีการวัดพื้นที่สัมบูรณ์ของราชู (Exact Area Measures) (Raju, 1990) และวิธีการวัดพื้นที่ช่วงปิด ในการวิจัยครั้งนี้ผู้วิจัยเลือกใช้วิธีการทดสอบไค-สแควร์ของลอร์ด (Lord’s Chi-square, 1980)

5. ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ จำแนกตามตัวแปรเพศซึ่งดำเนินการดังนี้

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในการศึกษาครั้งนี้มี 2 วิธี คือวิธี MACS และวิธี LDFA โดยจะวิเคราะห์ในทุกเงื่อนไขของขนาดกลุ่มตัวอย่างและความยาวของแบบสอบเพื่อเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF)

5.1 นำข้อมูลผลการตอบข้อสอบรายข้อที่เตรียมไว้ มาประมาณค่าพารามิเตอร์ของข้อสอบรายข้อ ค่าความสามารถของผู้สอบแต่ละคนและค่าความแปรปรวนร่วมของข้อมูล ด้วยโปรแกรม MULTILOG version 7 โดยจะวิเคราะห์แยกกันระหว่างกลุ่มอ้างอิง และกลุ่มเปรียบเทียบ ผลการวิเคราะห์จะได้แฟ้มข้อมูลของค่าประมาณพารามิเตอร์ข้อสอบของผู้สอบกลุ่มเพศชายและเพศหญิง

5.2 ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วย วิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย (Mean and Covariance Structures Method: MACS)

นำข้อมูลผลการตอบกลุ่มเพศชายและเพศหญิงในเงื่อนไขความยาวแบบสอบและกลุ่มตัวอย่างขนาดเดียวกันแต่ละคู่มาดำเนินการวิเคราะห์เพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี MACS โดยจับคู่นักเรียนเพศชายและเพศหญิงที่มีความสามารถเท่ากัน ซึ่งพิจารณาจากคะแนนรวมของผลการตอบข้อสอบวิชาคณิตศาสตร์ที่เท่ากัน และใช้สถิติ Mean and Covariance Structures ทดสอบนัยสำคัญ โดยใช้โปรแกรมสำเร็จรูป LISREL ช่วยในการวิเคราะห์

5.3 ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก (Logistic Discriminant Function Analysis: LDFA)

นำข้อมูลของกลุ่มเพศชายและเพศหญิงในเงื่อนไขความยาวของแบบสอบและขนาดกลุ่มตัวอย่างขนาดเดียวกัน แต่ละคู่มาดำเนินการวิเคราะห์ เพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี LDFA โดยใช้โปรแกรมสำเร็จรูป MULTILOG ช่วยในการวิเคราะห์

6. คำนวณประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

หลังจากวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธี MACS และวิธี LDFA แล้วนำผลการวิเคราะห์ที่ได้มาหาความถูกต้องของการตรวจสอบและความผิดพลาดด้วยการเปรียบเทียบกับผลที่ได้จากการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GPCM ถือว่าเป็นข้อที่ทำหน้าที่ต่างกัน ถ้าวิธี MACS และวิธี LDFA ตรวจสอบพบข้อสอบที่ทำหน้าที่ต่างกันตรงกับข้อสอบที่ถูกตรวจสอบพบด้วยวิธี GPCM ถือว่าวิธีนั้นมีความถูกต้องในการตรวจสอบ แต่ถ้าตรวจสอบพบว่าข้อสอบทำหน้าที่ต่างกันไม่ตรงกับข้อสอบที่ตรวจพบด้วยวิธี GPCM ถือว่ามีความผิดพลาดในการตรวจสอบ โดยนำผลมาคำนวณหาอัตราความคลาดเคลื่อนประเภทที่1 (Type I error rate) และค่าอำนาจการทดสอบ (Power of test) ของแต่ละวิธี เพื่อเปรียบเทียบประสิทธิภาพของแต่ละวิธี

สรุปผลการวิจัย

ผลการวิจัยแบ่งออกเป็น 2 ส่วน คือ ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 2 วิธี ภายใต้เงื่อนไขของปัจจัยที่แตกต่างกันทางด้านความยาวของแบบสอบและขนาดกลุ่มตัวอย่าง และผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 2 วิธี ภายใต้เงื่อนไขของปัจจัยที่แตกต่างกันทางด้านความยาวของแบบสอบและขนาดกลุ่มตัวอย่าง

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

1. ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 2 วิธี ภายใต้เงื่อนไขของปัจจัยที่แตกต่างกันทางด้านความยาวของแบบสอบและขนาดกลุ่มตัวอย่าง

1.1 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี MACS

กรณีที่แบบสอบมีความยาว 40 ข้อ ค่าอัตราความคลาดเคลื่อนประเภทที่ 1 หรือความผิดพลาดในการตรวจสอบมีค่าอยู่ในช่วงตั้งแต่ร้อยละ 8.3 ถึงร้อยละ 21.9 ซึ่งอยู่ในเกณฑ์ที่ยอมรับได้เพียงแค่ 1 เงื่อนไข โดยอัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าเฉลี่ยน้อยที่สุดเมื่อขนาดกลุ่มตัวอย่างขนาดใหญ่ คือ 1,000 คน คิดเป็นร้อยละ 18.8 และให้ค่าความผิดพลาดในการตรวจสอบเท่ากันเมื่อขนาดกลุ่มตัวอย่างขนาดกลางและขนาดเล็ก คือ 500 และ 200 คน คิดเป็นร้อยละ 21.9

กรณีที่แบบสอบมีความยาว 30 ข้อ อัตราความคลาดเคลื่อนประเภทที่ 1 หรือความผิดพลาดในการตรวจสอบมีค่าอยู่ในช่วงตั้งแต่ร้อยละ 8.3 ถึงร้อยละ 12.5 ซึ่งส่วนใหญ่อยู่ในเกณฑ์ที่ยอมรับได้ ยกเว้นเพียง 2 เงื่อนไขที่มีค่าเกินกว่าเกณฑ์ที่กำหนด โดยอัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าน้อยที่สุดเมื่อขนาดกลุ่มตัวอย่างขนาดเล็ก คือ 200 คน คิดเป็นร้อยละ 8.3 รองลงมาได้แก่ ขนาดกลุ่มตัวอย่างขนาดใหญ่ คือ 1,000 คน และกลุ่มตัวอย่างขนาดกลาง คือ 500 คน ให้ค่าความผิดพลาดในการตรวจสอบเท่ากัน คิดเป็นร้อยละ 12.5

1.2 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี LDFA

กรณีที่แบบสอบมีความยาว 40 ข้อ ค่าอัตราความคลาดเคลื่อนประเภทที่ 1 หรือความผิดพลาดในการตรวจสอบมีค่าอยู่ในช่วงตั้งแต่ร้อยละ 3.1 ถึงร้อยละ 15.6 และมีค่าเฉลี่ยเรียงตามขนาดกลุ่มตัวอย่างใหญ่ กลาง และเล็ก คือ 1,000, 500 และ 200 คน คิดเป็นร้อยละ 3.1, 12.5 และ 15.6 ตามลำดับ

กรณีที่แบบสอบมีความยาว 30 ข้อ อัตราความคลาดเคลื่อนประเภทที่ 1 หรือความผิดพลาดในการตรวจสอบมีค่าอยู่ในช่วงตั้งแต่ร้อยละ 0 ถึงร้อยละ 4.2 ซึ่งอยู่ในเกณฑ์ที่ยอมรับได้ทั้งหมด และมีค่าน้อยที่สุดเมื่อขนาดกลุ่มตัวอย่างขนาดใหญ่ และขนาดกลาง คือ 1,000 และ 500 คน คิดเป็นร้อยละ 0 และคิดเป็นร้อยละ 4.2 เมื่อขนาดกลุ่มตัวอย่างเป็นขนาดเล็ก คือ 200 คน

จะเห็นว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของแบบสอบทั้ง 2 ฉบับ มีค่าสูงกว่าเกณฑ์ที่ยอมรับได้คือ ร้อยละ 10 ดังนั้นจึงควรมีการศึกษาต่อไปในวิธีดังกล่าวนี้ โดยใช้เงื่อนไข

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

อื่น ๆ ที่มีนัยสำคัญ เพราะวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีประสิทธิภาพนั้น ควรจะสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในเกณฑ์ที่ยอมรับได้

1.3 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี MACS และวิธี LDFA

การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี MACS และ LDFA ในเงื่อนไขกลุ่มตัวอย่าง 1,000, 500 และ 200 คน และความยาวของแบบสอบ 40 ข้อ และ 30 ข้อ สรุปได้ว่า

กรณีกลุ่มตัวอย่าง 1,000 คน และแบบสอบความยาวทุกขนาด คือ 40 ข้อและ 30 ข้อ วิธี LDFA มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าวิธี MACS ทุกเงื่อนไข และมีค่าอยู่ในเกณฑ์ที่ยอมรับได้ทั้งหมด

กรณีกลุ่มตัวอย่าง 500 คน และแบบสอบความยาวทุกขนาด คือ 40 ข้อและ 30 ข้อ พบว่าวิธี LDFA มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าวิธี MACS ทุกเงื่อนไข โดยมีเพียง 1 เงื่อนไขเท่านั้นที่ค่าอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในเกณฑ์ที่ยอมรับได้ คือเงื่อนไขกลุ่มตัวอย่าง 500 คน ความยาวแบบสอบ 30 ข้อของวิธี LDFA

กรณีกลุ่มตัวอย่าง 200 คน และแบบสอบความยาวทุกขนาด คือ 40 ข้อและ 30 ข้อ พบว่าวิธี LDFA มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าวิธี MACS ทุกเงื่อนไข และมีเงื่อนไขแบบสอบความยาว 30 ข้อเท่านั้นที่มีค่าอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในเกณฑ์ที่กำหนด

2. ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 2 วิธี ภายใต้เงื่อนไขของปัจจัยที่แตกต่างกันทางด้านความยาวของแบบสอบและขนาดกลุ่มตัวอย่าง

2.1 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี MACS

กรณีที่แบบสอบมีความยาว 40 ข้อ พบว่าวิธี MACS มีค่าอัตราความคลาดเคลื่อนประเภทที่ 1 เกินกว่าเกณฑ์ที่กำหนดทุกเงื่อนไข และเมื่อพิจารณาค่าอำนาจการทดสอบ พบว่าวิธี MACS ให้ค่าอำนาจการทดสอบหรือความถูกต้องในการตรวจสอบมีค่าเฉลี่ยสูงสุดเมื่อขนาดกลุ่มตัวอย่างขนาดใหญ่ คือ 1,000 คน และกลุ่มตัวอย่างขนาดเล็ก คือ 200 คน คิดเป็นร้อยละ 62.5 แต่ให้ค่าความถูกต้องในการตรวจสอบน้อยที่สุดเมื่อขนาดกลุ่มตัวอย่างขนาดกลาง คือ 500 คน คิดเป็นร้อยละ 37.5

กรณีที่แบบสอบมีความยาว 30 ข้อ พบว่าวิธี MACS มีค่าอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในเกณฑ์ที่ยอมรับได้ 1 เงื่อนไข คือเงื่อนไขขนาดกลุ่มตัวอย่าง 200 คน และพบว่ามีค่าอำนาจการทดสอบอยู่ในเกณฑ์สูงเช่นเดียวกัน คือคิดเป็นร้อยละ 66.7 และเมื่อพิจารณาค่าอำนาจการทดสอบ วิธี MACS ให้ค่าอำนาจการทดสอบหรือความถูกต้องในการตรวจสอบเรียงตามขนาดกลุ่มตัวอย่าง โดยมีเมื่อขนาดกลุ่มตัวอย่างขนาดใหญ่ คือ 1,000 คน ให้ค่าความถูกต้องในการตรวจสอบมากที่สุด คิดเป็นร้อยละ 100 รองลงมาได้แก่ กลุ่มตัวอย่างขนาดกลาง และขนาดเล็ก คิดเป็นร้อยละ 83.3 และ 66.7 ตามลำดับ

2.2 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี LDFA

กรณีที่แบบสอบมีความยาว 40 ข้อ พบว่าวิธี LDFA มีค่าอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในเกณฑ์ที่ยอมรับได้ โดยเฉพาะกับเงื่อนไขกลุ่มตัวอย่างขนาดใหญ่ 1,000 คน และเมื่อพิจารณาค่าอำนาจการทดสอบพบว่าให้ค่าอำนาจการทดสอบหรือความถูกต้องในการตรวจสอบสูงถึงร้อยละ 100 รองลงมาได้แก่ กลุ่มตัวอย่างขนาดกลาง และขนาดเล็ก คิดเป็นร้อยละ 75 และ 62.5 ตามลำดับ

กรณีที่แบบสอบมีความยาว 30 ข้อ พบว่าวิธี LDFA มีค่าอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในเกณฑ์ที่ยอมรับได้ทุกเงื่อนไข โดยที่วิธี LDFA ให้ค่าอำนาจการทดสอบหรือความถูกต้องในการตรวจสอบมีค่าเฉลี่ยสูงสุดเมื่อขนาดกลุ่มตัวอย่างเป็นขนาดใหญ่ คือ 1,000 และขนาดกลาง คือ 500 คน คิดเป็นร้อยละ 100 และเมื่อขนาดกลุ่มตัวอย่างเป็นขนาดเล็ก คือ 200 คน ให้ความถูกต้องในการตรวจสอบน้อยที่สุด คิดเป็นร้อยละ 83.3

2.3 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี MACS และวิธี LDFA

การเปรียบเทียบผลอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี MACS และ LDFA ในเงื่อนไขกลุ่มตัวอย่าง 1,000, 500 และ 200 คน และความยาวของแบบสอบ 40 ข้อ และ 30 ข้อ สรุปได้ว่า

กรณีกลุ่มตัวอย่าง 1,000 คน และแบบสอบความยาว 40 ข้อ วิธี LDFA มีอำนาจการทดสอบสูงกว่าวิธี MACS ส่วนในกรณีแบบสอบความยาว 30 ข้อ วิธี LDFA และ MACS มีอำนาจการทดสอบเท่ากัน

กรณีกลุ่มตัวอย่าง 500 คน และแบบสอบความยาวทุกขนาด คือ 40 ข้อและ 30 ข้อ พบว่าวิธี LDFA มีอำนาจการทดสอบสูงกว่าวิธี MACS

กรณีกลุ่มตัวอย่าง 200 คน และแบบสอบความยาว 40 ข้อ วิธี LDFA และ MACS มีอำนาจการทดสอบเท่ากัน ส่วนในกรณีแบบสอบความยาว 30 ข้อ วิธี LDFA มีอำนาจการทดสอบสูงกว่าวิธี MACS

สรุปผลการวิจัยในภาพรวมของการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ผลดังนี้

ทั้ง 6 เงื่อนไขของการทดสอบ พบว่า วิธี LDFA มีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในช่วงตั้งแต่ร้อยละ 0 ถึงร้อยละ 15.6 และวิธี MACS มีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในช่วงตั้งแต่ร้อยละ 8.3 ถึงร้อยละ 21.9 โดยที่วิธี LDFA เกือบทุกเงื่อนไขมีค่าอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในเกณฑ์ที่ยอมรับได้ ยกเว้นเงื่อนไขขนาดกลุ่มตัวอย่าง 500 คนและ 200 คน ความยาวแบบสอบ 40 ข้อ ส่วนวิธี MACS พบว่าส่วนใหญ่อยู่เกินกว่าเกณฑ์ที่กำหนด ยกเว้นเงื่อนไขขนาดกลุ่มตัวอย่าง 200 คน ความยาวแบบสอบ 30 ข้อ ที่อัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในเกณฑ์ที่กำหนด

สรุปผลการวิจัยในภาพรวมของการเปรียบเทียบอำนาจการทดสอบ ได้ผลดังนี้

ทั้ง 6 เงื่อนไขของการทดสอบ พบว่า วิธี LDFA มีอำนาจการทดสอบอยู่ในช่วงตั้งแต่ร้อยละ 62.5 ถึงร้อยละ 100 และจะเห็นว่าเงื่อนไขที่มีค่าอำนาจการทดสอบต่ำสุดคือเงื่อนไขที่มีค่าอัตราความคลาดเคลื่อนประเภทที่ 1 เกินกว่าเกณฑ์ที่ยอมรับได้เช่นกัน สำหรับวิธี MACS พบว่ามีค่าอำนาจการทดสอบอยู่ในช่วงตั้งแต่ร้อยละ 37.5 ถึงร้อยละ 100 และพบว่าทั้งวิธี LDFA และวิธี MACS จะมีอำนาจการทดสอบในเงื่อนไขความยาวแบบสอบ 30 ข้อสูงกว่าเงื่อนไขความยาวแบบสอบ 40 ข้อทุกเงื่อนไข

อภิปรายผลการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่เหมาะสมกับข้อมูลที่มีการให้คะแนนแบบหลายค่า และมีวัตถุประสงค์เฉพาะเพื่อเปรียบเทียบประสิทธิภาพของผลการตรวจสอบของวิธีการตรวจสอบ 2 วิธี คือ วิธี MACS และวิธี LDFA ไม่ได้มุ่งนำผลการตรวจสอบไปใช้ในการปรับปรุงคุณภาพของแบบสอบ และจากการศึกษาค้นคว้ายังไม่พบว่ามีการศึกษาเพื่อเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจากข้อมูลผลการตอบข้อสอบที่ต้องตอบทุกตัวเลือกและมีการให้คะแนนแบบหลายค่า ด้วยวิธี MACS และวิธี LDFA แต่มีการเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ จากข้อมูลผลการตอบข้อสอบที่มีการให้คะแนนแบบสองค่า ดังนั้น การอ้างอิงส่วนใหญ่จึงเปรียบเทียบกับผล

การศึกษาที่เป็นข้อมูลผลการตอบข้อสอบที่มีการให้คะแนนแบบสองค่า ผลการวิจัย พบประเด็นสำคัญสรุปได้ดังนี้

1. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีเดียวกัน

1.1 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี MACS

ในกรณีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี MACS นั้นใช้หลักพื้นฐานมาจากการวิเคราะห์องค์ประกอบเชิงยืนยัน ดังนั้น การอ้างอิงส่วนใหญ่จึงเปรียบเทียบกับผลการศึกษาของวิธีการที่มีพื้นฐานมาจากการวิเคราะห์องค์ประกอบเช่นเดียวกัน และจากผลการวิจัยครั้งนี้ พบว่าความถูกต้องของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี MACS ระหว่างกลุ่มตัวอย่างขนาดต่างกัน คือ 1,000, 500 และ 200 คน พบว่าเมื่อกลุ่มตัวอย่างขนาด 1,000 คน และ 200 คน ให้ผลการตรวจสอบความถูกต้องดีที่สุด สอดคล้องกับผลการศึกษาบางส่วนของ Stark et al. (2006) ที่พบว่าในการระบุการทำหน้าที่ต่างกันของข้อสอบ โดยใช้วิธี MACS และวิธีถดถอยโลจิสติก (LR)โดยใช้ข้อมูลจำลองนั้น เมื่อใช้กลุ่มตัวอย่างขนาดเล็ก การวิเคราะห์ MACS ให้ผลดีกว่าวิธี LR และยังสอดคล้องกับข้อค้นพบของ นิคม กีรติวาภูร (2542) ที่พบว่าในกรณีที่กลุ่มตัวอย่างขนาดเล็ก (300 คน) วิธีการวิเคราะห์องค์ประกอบจำกัดสามารถตรวจสอบการทำหน้าที่ต่างกันได้ดีที่สุด และในกรณีที่กลุ่มตัวอย่างขนาดใหญ่ (1,000 คน) วิธีการวิเคราะห์องค์ประกอบจำกัด ยังมีอำนาจการทดสอบสูงที่สุดเช่นเดียวกัน

นอกจากนี้ยังได้ข้อค้นพบว่า อำนาจการทดสอบเมื่อตรวจสอบด้วยวิธี MACS จะมีค่ามากขึ้นเมื่อแบบสอบความยาวลดลง นั่นคือแบบสอบความยาว 30 ข้อให้อำนาจการทดสอบมากกว่าแบบสอบความยาว 40 ข้อ ในทุกเงื่อนไข และสอดคล้องกับข้อค้นพบที่ว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบด้วยวิธี MACS จะมีค่าน้อยลงเมื่อแบบสอบความยาวน้อยลง นั่นคือแบบสอบความยาว 30 ข้อให้ค่าอัตราความคลาดเคลื่อนประเภทที่ 1 น้อยกว่าแบบสอบความยาว 40 ข้อ ในทุกเงื่อนไข ซึ่งไม่สอดคล้องกับผลการศึกษาของวิธีอื่น ๆ เช่น วิธี MH และวิธีถดถอยโลจิสติก ที่พบว่าเมื่อใช้แบบสอบที่มีความยาวมากขึ้น อำนาจการทดสอบจะมีค่ามากขึ้นและอัตราความคลาดเคลื่อนประเภทที่ 1 จะมีค่าลดลง (Swaminathan & Rogers,1990; Uttaro & Millsap, 1994) แต่พบว่ามีการศึกษาเป็นจำนวนมากที่ได้ข้อค้นพบว่าความยาวของแบบสอบไม่ส่งผลต่ออำนาจการทดสอบ อาทิเช่น กาญจนา วัธนสุนทร (2538) พบว่า ความยาวของแบบสอบไม่มีผลกระทบต่อความถูกต้องในการตรวจสอบและความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี SIBTEST และวิธี MH และจากการศึกษาของสิริรัตน์ วิภาสศิลป์ (2545) ที่พบว่า ความยาวของแบบสอบไม่ส่งผลต่อความถูกต้องในการตรวจสอบและความผิดพลาดใน

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี SIBTEST และวิธี DFIT ซึ่งอาจเป็นไปได้ว่าความยาวของแบบสอบไม่ส่งผลต่อความถูกต้องและความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการตรวจให้คะแนนแบบหลายค่าโดยตรง แต่อาจร่วมกันกับขนาดกลุ่มตัวอย่าง

สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่าเมื่อตรวจสอบด้วยวิธี MACS สามารถควบคุมค่าอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในเกณฑ์ที่กำหนดร้อยละ 10 ได้เพียง 1 เงื่อนไข จึงแนะนำให้ใช้ผลเฉพาะเงื่อนไขที่สามารถควบคุมค่าอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ สำหรับเงื่อนไขอื่น ๆ มีค่าไม่ต่างกันในทุกเงื่อนไขของขนาดกลุ่มตัวอย่าง ซึ่งสอดคล้องกับผลการศึกษาของนิคม กีรติวาฏูร (2542) ที่พบว่าวิธีการวิเคราะห์องค์ประกอบจำกัดไม่ขึ้นอยู่กับขนาดกลุ่มตัวอย่าง และยังสอดคล้องกับผลการศึกษาของ Oort (1998) ที่พบว่าอำนาจการตรวจสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการวิเคราะห์องค์ประกอบจำกัดไม่ขึ้นกับขนาดกลุ่มตัวอย่าง

1.2. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี LDFA

ความถูกต้องของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี LDFA ระหว่างกลุ่มตัวอย่างขนาดต่างกัน คือ 1,000, 500 และ 200 คน พบว่าเมื่อขนาดกลุ่มตัวอย่างมากขึ้น ส่งผลให้ความถูกต้องของผลการตรวจสอบมากขึ้นเช่นกัน ซึ่งวิธี LDFA นั้นถือเป็นวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีพื้นฐานมาจากทฤษฎีการตอบสนองข้อสอบ และพัฒนามาจากวิธีการถดถอยโลจิสติก (Miller & Spray, 1993) ซึ่งถือว่าเป็นวิธีการที่ได้รับความนิยมมากที่สุดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบทวิวิภาค (Su & Wang, 2005) ดังนั้น การอ้างอิงส่วนใหญ่จึงเปรียบเทียบกับผลการศึกษาของวิธีการถดถอยโลจิสติก จากผลการศึกษาของ Miller & Spray (1993) และ Welch & Miller (1995) ได้ให้ความเห็นว่าควรใช้การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจสอบด้วยวิธี LDFA ในการทดสอบสถานการณ์จริง และจากผลการศึกษาของ Swaminathan & Rogers (1990) แสดงให้เห็นว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและแบบอเนกรูป เมื่อเพิ่มขนาดกลุ่มตัวอย่างจะมีผลทำให้อำนาจการทดสอบของวิธีการถดถอยโลจิสติกมีค่าเพิ่มมากขึ้นเกือบทุกเงื่อนไข ซึ่งสอดคล้องกับผลการศึกษาในครั้งนี้ นอกจากนี้ยังสอดคล้องกับผลการศึกษาของ Montesinos & Benito (2003) ที่พบว่าวิธีการถดถอยโลจิสติกจะมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ดีที่สุดเมื่อกลุ่มตัวอย่างมีขนาดใหญ่ และยังสอดคล้องกับผลการศึกษาของ French & Miller (1996) ที่พบว่าเมื่อกลุ่มตัวอย่างมีขนาดเล็กลง อำนาจในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบลดลง อีกทั้งพบว่า วิธีที่มีพื้นฐานมาจากการใช้ทฤษฎี

การตอบสนองข้อสอบนั้น ยิ่งมีขนาดกลุ่มตัวอย่างใหญ่ขึ้นเพียงใด จะส่งผลให้มีการประมาณค่าพารามิเตอร์ของข้อสอบได้ดีขึ้น ซึ่งศิริชัย กาญจนวาสี (2545) แนะนำว่าควรใช้กลุ่มตัวอย่างขนาด 1,000 คนขึ้นไป

นอกจากได้ข้อค้นพบว่า อำนาจการทดสอบจะมีค่ามากขึ้นเมื่อกลุ่มตัวอย่างขนาดเพิ่มขึ้นแล้ว ยังพบว่าอำนาจการทดสอบเมื่อตรวจสอบด้วยวิธี LDFA จะมีค่ามากขึ้นเมื่อแบบสอบความยาวลดลง นั่นคือแบบสอบความยาว 30 ข้อให้อำนาจการทดสอบมากกว่าแบบสอบความยาว 40 ข้อ ในทุกเงื่อนไข ซึ่งสอดคล้องกับผลการศึกษาของ จิติมา วรรณศรี (2539) ที่พบว่าเมื่อใช้แบบสอบความยาวขนาดกลาง จะมีผลทำให้วิธีแมนเทล-แฮนส์เซลและวิธีซิปเทสท์มีประสิทธิภาพในการตรวจสอบดีที่สุด แต่ขัดแย้งกับผลการศึกษาของทองอยู่ สาระ (2543) ที่เปรียบเทียบอำนาจการตรวจสอบและจำแนกผิดพลาดในการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบสม่ำเสมอและแบบไม่สม่ำเสมอระหว่างวิธี MH และวิธีการถดถอยโลจิสติก พบว่าความยาวของแบบทดสอบไม่มีผลต่ออำนาจการตรวจสอบและการจำแนกผิดพลาดในทั้ง 2 วิธี และ Rogers & Swaminathan (1993) ยังพบว่า ความยาวของแบบสอบไม่มีผลต่ออำนาจการทดสอบของวิธี MH และวิธีการถดถอยโลจิสติกยกเว้นในกรณีแบบอเนกรูปของวิธีการถดถอยโลจิสติก

ด้านอัตราความคลาดเคลื่อนประเภทที่ 1พบว่าวิธี LDFA สามารถควบคุมค่าอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในเกณฑ์ที่กำหนดร้อยละ 10 ได้เกือบทุกเงื่อน ยกเว้นเงื่อนไขกลุ่มตัวอย่างขนาดเล็ก และแบบสอบความยาว 40 ข้อ ซึ่งจะเห็นว่าทั้ง 2 เงื่อนไขนี้ มีค่าอำนาจการทดสอบต่ำเช่นเดียวกัน ซึ่งสอดคล้องกับหลักการของทฤษฎีการตอบสนองข้อสอบ ที่เมื่อกลุ่มตัวอย่างมีขนาดใหญ่ขึ้น ส่งผลให้สามารถประมาณค่าพารามิเตอร์ได้ดีขึ้นเช่นเดียวกัน และพบว่าอัตราความคลาดเคลื่อนประเภทที่ 1 จะมีค่ามากขึ้นเมื่อแบบสอบมีความยาวมากขึ้นในทุกเงื่อนไข ซึ่งสอดคล้องกับผลการศึกษาของอุทัยวรรณ สายพัฒนะ (2547) ที่พบว่า ความผิดพลาดของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาคด้วยวิธี Polytomous SIBTEST เมื่อกลุ่มตัวอย่างมีขนาดใหญ่ขึ้นความผิดพลาดในการตรวจสอบจะมีค่ามากขึ้นด้วย และยังสอดคล้องกับผลการศึกษาของ Cohen & Kim (1993) ที่พบว่า เมื่อใช้แบบสอบที่มีความยาวมากขึ้นแล้วอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ที่ตรวจสอบด้วยวิธี การวัดพื้นที่ของ Raju และวิธีการทดสอบไค-สแควร์ของ Lord จะมีค่ามากขึ้นด้วย ในทางตรงกันข้ามเมื่อขนาดกลุ่มตัวอย่างเพิ่มมากขึ้น อัตราความคลาดเคลื่อนประเภทที่ 1จะมีค่าน้อยลง ซึ่งสอดคล้องกับผลการศึกษาของจิติมา วรรรศรี (2539) ที่พบว่าเมื่อกลุ่มตัวอย่างมีขนาด 200 คน วิธี MH มีความผิดพลาดในการ

ตรวจสอบอยู่ในระดับหนึ่ง (1%-9%) แต่เมื่อเพิ่มขนาดกลุ่มตัวอย่างเป็น 1,000 คน ความผิดพลาดในการตรวจสอบลดลงเหลือร้อยละ 100

2. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีต่างกัน

เมื่อแบบสอบความยาว 40 ข้อ กลุ่มตัวอย่างขนาด 1,000, 500 และ 200 คน วิธี MACS มีอำนาจการทดสอบหรือความถูกต้องในการตรวจสอบประมาณร้อยละ 37.5 ถึง 62.5 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 หรือความผิดพลาดในการตรวจสอบมีค่าประมาณร้อยละ 18.8 ถึง 21.9 และมีแนวโน้มมีความผิดพลาดน้อยลงเมื่อกลุ่มตัวอย่างมีขนาดเพิ่มขึ้น วิธี LDFA มีอำนาจการทดสอบหรือความถูกต้องในการตรวจสอบประมาณร้อยละ 62.5 ถึง100.0 และมีแนวโน้มมีความถูกต้องมากขึ้นเมื่อกลุ่มตัวอย่างมีขนาดเพิ่มมากขึ้น ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 หรือความผิดพลาดในการตรวจสอบมีค่าประมาณร้อยละ 3.1 ถึง 15.6 และมีแนวโน้มมีความผิดพลาดน้อยลงเมื่อกลุ่มตัวอย่างมีขนาดเพิ่มขึ้น

เมื่อแบบสอบความยาว 30 ข้อ กลุ่มตัวอย่างขนาด 1,000, 500 และ 200 คน วิธี MACS มีอำนาจการทดสอบหรือความถูกต้องในการตรวจสอบประมาณร้อยละ 66.7 ถึง 100.0 และมีแนวโน้มมีความถูกต้องมากขึ้นเมื่อกลุ่มตัวอย่างมีขนาดเพิ่มมากขึ้น ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 หรือความผิดพลาดในการตรวจสอบมีค่าประมาณร้อยละ 8.3 ถึง 12.5 และมีแนวโน้มมีความผิดพลาดน้อยลงเมื่อกลุ่มตัวอย่างมีขนาดน้อยลง วิธี LDFA มีอำนาจการทดสอบหรือความถูกต้องในการตรวจสอบประมาณร้อยละ 83.3 ถึง 100.0 และมีแนวโน้มมีความถูกต้องมากขึ้นเมื่อกลุ่มตัวอย่างมีขนาดเพิ่มมากขึ้น ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 หรือความผิดพลาดในการตรวจสอบมีค่าประมาณร้อยละ 0.0 ถึง 4.2 และมีแนวโน้มมีความผิดพลาดน้อยลงเมื่อกลุ่มตัวอย่างมีขนาดมากขึ้น

จะเห็นได้ว่า อำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบด้วยวิธี MACS และวิธี LDFA มีความผันแปรค่อนข้างเป็นระบบในเงื่อนไขของขนาดกลุ่มตัวอย่างและความยาวแบบสอบ ยกเว้นกรณีของอำนาจการทดสอบของวิธี MACS เมื่อแบบสอบความยาว 40 ข้อ นอกจากนั้นยังพบว่าการตรวจสอบทั้ง 2 วิธีเป็นไปในทางเดียวกันในกรณีที่แบบสอบความยาว 30 ข้อ นั่นคืออำนาจการทดสอบมีแนวโน้มมีความถูกต้องมากขึ้นเมื่อกลุ่มตัวอย่างมีขนาดเพิ่มมากขึ้น แต่ไม่เป็นไปในทางเดียวกันในส่วนของอัตราความคลาดเคลื่อนประเภทที่ 1 ที่วิธี MACS มีแนวโน้มมีความผิดพลาดน้อยลงเมื่อกลุ่มตัวอย่างมีขนาดน้อยลง แต่วิธี LDFA มีแนวโน้มมีความผิดพลาดน้อยลงเมื่อกลุ่มตัวอย่างมีขนาดมากขึ้น แต่ในกรณีแบบสอบความยาว 40 ข้อ นั้น อัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้ง 2 วิธีมีแนวโน้มไปในทางเดียวกัน นั่นคือมีแนวโน้มมีความ

ผิดพลาดน้อยลงเมื่อกลุ่มตัวอย่างมีขนาดเพิ่มขึ้น แต่อย่างไรก็ตาม พบว่าการตรวจสอบด้วยวิธี LDFA มีประสิทธิภาพดีกว่าในทุกเงื่อนไขของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาค ซึ่งสอดคล้องกับผลการศึกษาของ Stark et al. (2006) ที่พบว่าวิธีที่มีพื้นฐานมาจากทฤษฎีการตอบสนองข้อสอบจะมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบดีกว่าวิธี MACS อีกทั้งในการศึกษาครั้งนี้ได้เลือกใช้วิธีเกณฑ์ คือวิธี GPCM ซึ่งเป็นวิธีที่มีพื้นฐานมาจากทฤษฎีการตอบสนองข้อสอบ ส่งผลให้วิธี LDFA ซึ่งมีพื้นฐานมาจากทฤษฎีการตอบสนองข้อสอบเช่นเดียวกัน มีประสิทธิภาพดีกว่าวิธี MACS ที่มีพื้นฐานมาจากการวิเคราะห์องค์ประกอบเชิงยืนยัน แต่อย่างไรก็ตามผู้วิจัยเน้นที่ค่าอำนาจการทดสอบที่เหมาะสมในเงื่อนไขต่าง ๆ และความสามารถในการควบคุมค่าอัตราความคลาดเคลื่อนประเภทที่ 1

ต่อไปนี้ผู้วิจัยขอยกตัวอย่างข้อคำถามที่มีการทำหน้าที่ต่างกัน ดังเช่น ข้อ 8 และ ข้อ 12 เป็นข้อที่ทำหน้าที่ต่างกัน เมื่อลองพิจารณาคำถามและคำตอบจะเห็นว่า หากนักเรียนดูที่ผลลัพธ์ก็จะได้ผลถูกต้องตรงตามเฉลย แต่หากนักเรียนตีความจากกระบวนการจะพบว่า ตัวเลือกมีปัญหาไม่ตรงตามเฉลยทุกขั้นตอน ทำให้นักเรียนอาจเกิดความเข้าใจผิด ดังเช่นตัวเลือก ง. ของข้อที่ 8 และตัวเลือก ข. ของข้อที่ 12 ซึ่งด้วยสาเหตุดังกล่าว ผู้วิจัยคิดว่าเป็นเพราะลักษณะของข้อคำถามที่ส่งผลให้การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในการวิจัยครั้งนี้ มีผลจำเพาะกับข้อมูลชุดนี้

ข้อ 8. $[(-10) + 8] - [(-9) + 2]$ มีค่าเท่ากับข้อใด

ก. $(-2) + 7$

ข. $(-17) + 10$

ค. $2 + (-7)$

ง. $5 + 0$

จ. $2 + 11$

ข้อ 12. กำหนด $(a \times b) = -15$ และ $(c \times d) = 3$ ดังนั้นผลลัพธ์ของ $(a \times b) \div (c \times d)$ จะมีค่าเท่ากับข้อใด

ก. $(-15) \div 3$

ข. $15 \div (-3)$

ค. $3 \div (-15)$

ง. $(-15) \div [(-1) \times (-3)]$

จ. $[(-15) \times (-1)] \div (-3)$

สรุปภาพรวมของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีต่างกัน

1. เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1

ทั้ง 6 เงื่อนไขของการทดสอบ พบว่า วิธี LDFA มีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในช่วงตั้งแต่ร้อยละ 0 ถึงร้อยละ 15.6 และวิธี MACS มีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในช่วงตั้งแต่ร้อยละ 8.3 ถึงร้อยละ 21.9 โดยที่วิธี LDFA เกือบทุกเงื่อนไขอยู่ในเกณฑ์ที่ยอมรับได้ ยกเว้นเงื่อนไขขนาดกลุ่มตัวอย่าง 500 คนและ 200 คน ความยาวแบบสอบ 40 ข้อ ส่วนวิธี MACS พบว่าส่วนใหญ่อยู่เกินกว่าเกณฑ์ที่กำหนด ยกเว้นเงื่อนไขขนาดกลุ่มตัวอย่าง 200 คน ความยาวแบบสอบ 30 ข้อ ที่อัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในเกณฑ์ที่กำหนด

2. เปรียบเทียบอำนาจการทดสอบ

ทั้ง 6 เงื่อนไขของการทดสอบ พบว่า วิธี LDFA มีอำนาจการทดสอบอยู่ในช่วงตั้งแต่ร้อยละ 62.5 ถึงร้อยละ 100 และจะเห็นว่าเงื่อนไขที่มีค่าอำนาจการทดสอบต่ำสุดคือเงื่อนไขที่มีค่าอัตราความคลาดเคลื่อนประเภทที่ 1 เกินกว่าเกณฑ์ที่ยอมรับได้เช่นกัน สำหรับวิธี MACS พบว่ามีค่าอำนาจการทดสอบอยู่ในช่วงตั้งแต่ร้อยละ 37.5 ถึงร้อยละ 100 และพบว่าทั้งวิธี LDFA และวิธี MACS จะมีอำนาจการทดสอบในเงื่อนไขความยาวแบบสอบ 30 ข้อสูงกว่าเงื่อนไขความยาวแบบสอบ 40 ข้อทุกเงื่อนไข

ข้อเสนอแนะ

ข้อเสนอแนะในการนำผลวิจัยไปใช้

1. การศึกษาครั้งนี้พบว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบคณิตศาสตร์ เรื่อง จำนวนเต็ม เศษส่วน และทศนิยม เมื่อแบบสอบประกอบด้วยข้อสอบ 40 ข้อ และ 30 ข้อ ด้วยวิธี MACS และวิธี LDFA ให้ผลในการตรวจสอบที่แตกต่างกันในเงื่อนไขของขนาดกลุ่มตัวอย่างที่ต่างกัน นั่นคือ ขนาดของกลุ่มตัวอย่างมีผลต่อประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ดังนั้น ในการศึกษาวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่านั้น ควรใช้กลุ่มตัวอย่างขนาดกลางขึ้นไป (ไม่ควรต่ำกว่า 500 คน) หรือกลุ่มตัวอย่างขนาดใหญ่ จะทำให้ผลการตรวจสอบการทำหน้าที่ต่างกันด้วยวิธี LDFA มีความถูกต้องในการตรวจสอบสูง

2. ตามทฤษฎีทางการทดสอบนั้น พบว่าหากแบบสอบมีความยาวมากเพียงใด จะส่งผลให้การประมาณค่าพารามิเตอร์ได้ผลดีขึ้น แต่ในการวิจัยครั้งนี้พบว่าแบบสอบที่มีความยาว 30 ข้อ มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบดีกว่าแบบสอบที่มีความยาว 40 ข้อ

ซึ่งเป็นข้อควรระวังในการศึกษาครั้งต่อไป เนื่องจากวิธีที่นำมาตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้ง 2 วิธีนั้น เป็นวิธีที่ใช้พื้นฐานมาจากทฤษฎีการตอบสนองข้อสอบและการวิเคราะห์องค์ประกอบเชิงยืนยัน ซึ่งเมื่อนำมาเทียบกับวิธีเกณฑ์คือวิธีที่มีพื้นฐานมาจากทฤษฎีการตอบสนองข้อสอบ ย่อมทำให้วิธีที่มีพื้นฐานเดียวกันได้เปรียบในการทดสอบมากกว่า

3. การศึกษาในครั้งนี้ได้พบว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่าด้วยวิธี GPCM ซึ่งเป็นวิธีตามแนวคิดของทฤษฎี IRT นั้น แม้จะมีข้อจำกัดในเรื่องต้องใช้กลุ่มตัวอย่างขนาดใหญ่ มีความยุ่งยากหลายขั้นตอนในการวิเคราะห์ เช่นเดียวกับวิธี LDFA แต่เมื่อพิจารณาจุดเด่นของวิธีการนี้ ในด้านความไม่แปรเปลี่ยนของค่าพารามิเตอร์ของข้อสอบไปตามกลุ่มผู้สอบและด้านความไว และความเฉียบคมในการตรวจสอบ อีกทั้งสามารถให้รายละเอียดของผลการตรวจสอบได้มาก ผู้วิจัยเห็นว่าเป็นวิธีการหนึ่งที่น่าจะนำมาใช้ได้ดีโดยเฉพาะกับหน่วยงานที่เกี่ยวข้องกับการสอบวัดทางการศึกษา

ข้อเสนอแนะในการวิจัยครั้งต่อไป

1. ควรมีการศึกษาเพิ่มเติมเกี่ยวกับประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเมื่อมีผู้สอบหลายกลุ่ม โดยใช้วิธีในกลุ่ม IRT และ non-IRT โดยเลือกใช้วิธีเกณฑ์ทั้งวิธีในกลุ่ม IRT และ non-IRT เพื่อส่งผลต่อการเปรียบเทียบได้ครอบคลุม และน่าเชื่อถือมากขึ้น

2. ควรศึกษาวิธีการตรวจสอบเดิม หรือใช้เครื่องมือที่มีรูปแบบในการตอบในลักษณะอื่นๆ เช่น การทดสอบการปฏิบัติ การเขียนเรียงความ ข้อสอบเขียนตอบ แบบวัดเจตคติ แบบวัดความพึงพอใจ เป็นต้น ซึ่งข้อสอบประเภทนี้มีการให้คะแนนแบบหลายค่า และที่สำคัญที่สุด คือ มีคุณค่าของสารสนเทศของผลการตอบข้อสอบของผู้สอบที่น่าสนใจเป็นอย่างมาก

3. ควรศึกษาเพื่อเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ของแบบทดสอบ (Differential Test Functioning: DTF) ด้วยวิธี MACS และวิธี LDFA ในเงื่อนไขอื่น ๆ เพิ่มเติม เช่น ขนาดกลุ่มตัวอย่าง เป็นต้น จะช่วยขยายขอบเขตทางวัดผลการศึกษาให้กว้างขึ้น

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รายการอ้างอิง

<u>*ภาษาไทย*</u>

กาญจนา วัธนสุนทร. (2537). *การพัฒนาเกณฑ์ตัดสินข้อสอบลำเอียงทางเพศ.* วิทยานิพนธ์ปริญญามหาบัณฑิต. ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

เกสร หว่างจิตร. (2539). *การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบสำหรับแบบสอบคัดเลือกระดับบัณฑิตศึกษาวิชาภาษาไทยและภาษาอังกฤษด้วยวิธีแมนเทล-เฮลส์เซล.* วิทยานิพนธ์ปริญญามหาบัณฑิต. ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

จิติมา วรรณศรี. (2539). *การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซลกับวิธีซิปเทสท์ เมื่อความยาวของแบบทดสอบ ขนาดกลุ่มตัวอย่างและอัตราส่วนของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบต่างกัน.* วิทยานิพนธ์ปริญญามหาบัณฑิต. ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

ญาณภัทร สีหะมงคล. (2540). *การเปรียบเทียบความสอดคล้องของผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันระหว่างวิธี Lord's^2 วิธี Raju's Area Measures และวิธี Closed Interval Area.* วิทยานิพนธ์ดุษฎีบัณฑิต. ภาควิชาการทดสอบและวัดผลทางการศึกษา บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ.

ธนวัฒน์ แสนสุข. (2538). *การใช้จีอาร์เอ็ม จีพีซีเอ็ม และโมเดลโ,จิสติกในการเปรียบเทียบฟังก์ชันสารสนเทศของแบบวัดที่มีวิธีการให้คะแนนต่างกัน. 2526.* วิทยานิพนธ์ปริญญามหาบัณฑิต. ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

ทัศนีย์ พีระมนตรี. (2536). *การวิเคราะห์ความลำเอียงของแบบสอบวิชาคณิตศาสตร์โครงการตรวจสอบคุณภาพการศึกษาชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526.* วิทยานิพนธ์ปริญญามหาบัณฑิต. ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

นพมาศ พิพัฒนสุข. (2541). *การเปรียบเทียบประสิทธิภาพระหว่างวิธีแมนเทล-เฮลส์เซลกับวิธีถดถอยโลจิสติก ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อใช้เกณฑ์จับคู่เปรียบเทียบ แตกต่างกันในแบบสอบชนิดพหุมิติ.* วิทยานิพนธ์ปริญญามหาบัณฑิต. ภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

นิคม กีรติวรางกูร. (2542). *การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีการวิเคราะห์องค์ประกอบจำกัด แมนเทล-แฮนส์เซล และการตอบสนองข้อสอบ.* วิทยานิพนธ์ปริญญามหาบัณฑิต. ภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

ปิยะทิพย์ ตินวร. (2549). *การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบพหุมิติ : การเปรียบเทียบประสิทธิภาพระหว่างวิธีการวิเคราะห์องค์ประกอบจำกัดกับวิธีถดถอยโลจิสติก.* วิทยานิพนธ์ปริญญามหาบัณฑิต. ภาควิชาวิจัยและวัดผลการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา.

ผจงจิต อินทสุวรรณ. (2545). *การวิเคราะห์ความแปรปรวนหลายตัวแปร.* กรุงเทพมหานคร: บริษัทธนธัชการพิมพ์ จำกัด.

พรรณี จินตมาศ. (2540). *การเปรียบเทียบผลการวิเคราะห์ความลำเอียงของข้อสอบ โดยใช้ขนาดกลุ่มผู้สอบและวิธีวิเคราะห์ต่างกัน.* วิทยานิพนธ์ปริญญามหาบัณฑิต. สาขาวิชาการวัดผลการศึกษา บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ.

พิมพ์สิริ เธียรนรเศรษฐ์. (2549). *การเปรียบเทียบค่าความสามารถ ลักษณะของการเปลี่ยนคำตอบและเวลาที่ใช้ในการทำข้อสอบในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ เมื่อเงื่อนไขการทดสอบและระดับความสามารถของผู้สอบแตกต่างกัน.* วิทยานิพนธ์ปริญญามหาบัณฑิต. ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

รัตนา ไชยตรี. (2549). *การเปรียบเทียบคุณภาพและอัตราความคลาดเคลื่อนของการกำหนดเกรดแบบอิงกลุ่ม และอิงเกณฑ์ เมื่อใช้แบบสอบเลือกตอบที่มีการตอบ และการตรวจให้คะแนนความรู้บางส่วนด้วยวิธีแตกต่างกัน.* วิทยานิพนธ์ปริญญามหาบัณฑิต. ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

รักชนก ยี่สุ่นศรี. (2544). *การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบด้วยกระบวนการ ดี เอฟ ไอ ที สำหรับแบบสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษา วิชาภาษาอังกฤษและวิชาคณิตศาสตร์.* วิทยานิพนธ์ปริญญามหาบัณฑิต. ภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

รัชรินทร์ มุคดา. (2540). *การเปรียบเทียบประสิทธิภาพระหว่างวิธีแมนเทล-แฮนส์เซลกับวิธีถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป ในกรณีการจัดกลุ่มความสามารถ ค่าความยากของข้อสอบ และค่าอำนาจจำแนกของข้อสอบต่างกัน.* วิทยานิพนธ์ปริญญามหาบัณฑิต. ภาควิชาวิจัยและจิตวิทยาการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

วลีมาศ แซ่อึ้ง. (2543). *การเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก.* วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต. ภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนวาสี. (2545). *ทฤษฎีการทดสอบแนวใหม่.* พิมพ์ครั้งที่ 2. กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ศุภวัฒน์ มะลิเผือก. (2548). *อิทธิพลของการทำหน้าที่ต่างกันของข้อสอบที่ส่งผลต่อคุณภาพของแบบสอบถามวัดผลสัมฤทธิ์ทางการเรียนระดับชาติ ชั้นประถมศึกษาปีที่ 6.* วิทยานิพนธ์ปริญญามหาบัณฑิต. ภาควิชาวิจัยและวัดผลการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา.

สุนิสา จุ้ยม่วงศรี. (2546). *การศึกษาผลของการเทียบคะแนนข้อสอบที่ตรวจให้คะแนนแบบหลายค่า.* วิทยานิพนธ์ดุษฎีบัณฑิต. ภาควิชาการทดสอบและวัดผลทางการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.

สุมาลี แกล้วทนงค์. (2547). *สาเหตุการทำหน้าที่ต่างกันของข้อสอบสาระการเรียนรู้ภาษาไทยและสาระการเรียนรู้สังคมศึกษา ศาสนาและวัฒนธรรม.* วิทยานิพนธ์ปริญญามหาบัณฑิต. ภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

สุรศักดิ์ อมรรัตนศักดิ์. (2530). *การศึกษาเปรียบเทียบผลของวิธีวิเคราะห์ความลำเอียงของข้อสอบที่แตกต่างกัน 4 วิธี.* วิทยานิพนธ์ปริญญามหาบัณฑิต. ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

เสรี ชัดแช้ม. (2539). *การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันแบบไม่สม่ำเสมอของข้อสอบระหว่างวิธีแมนเทล-แฮนส์เซลแบบปกติ กับวิธีแมนเทล-แฮนส์เซลแบบแบ่งกลุ่มความสามารถผู้สอบและความยาวของข้อสอบ.* วิทยานิพนธ์ดุษฎีบัณฑิต. ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

อวยพร เรืองตระกูล. (2544). *การพัฒนาและวิเคราะห์คุณภาพของวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิม และทฤษฎีการตอบสนองข้อสอบ.* วิทยานิพนธ์ดุษฎีบัณฑิต. ภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

อ้อมใจ คงเสาร์. (2540). *การศึกษาค่าพารามิเตอร์ของข้อสอบและค่าฟังก์ชันอินฟอร์เมชัน จากแบบทดสอบเลือกตอบกับแบบทดสอบโคลซที่วัดมโนทัศน์ทางคณิตศาสตร์ชั้นประถมศึกษาปีที่ 6.* วิทยานิพนธ์ปริญญามหาบัณฑิต. ภาควิชาการวัดผลการศึกษา บัณฑิตวิทยาลัย มหาวิทยาลัยนเรศวร.

อุทัยวรรณ สายพัฒนะ. (2547). *การเปรียบเทียบประสิทธิภาพของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่า วิธี GMH และวิธีPolytomous SIBTEST*. วิทยานิพนธ์ดุษฎีบัณฑิต. ภาควิชาการทดสอบและวัดผลทางการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.

อารี มากมณี. (2541). *การพัฒนาแบบสอบความเข้าใจในการอ่านภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6*. วิทยานิพนธ์ปริญญามหาบัณฑิต. ภาควิชาวิจัยการศึกษา คณะศึกษาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาษาอังกฤษ

Azen, R. (2002). Examination of Gender Bias in University Admissions. *Applied Measurement in Education* 15: 75-94.

Bolt, D. M. (2002). A monte Carlo Comparison of Parametric Polytomous DIF Detection Methods. *Journal of Applied Psychology* 15: 113-141.

Byrne, B. M. (1998). *Structural Equation Modeling with LISREL, PRELIS, and SIMPLIS: Basic Concepts, Applications, and Programming*. Mahwah, NJ: Erlbaum.

Camilli, G., & Shepard, L. A. (1994). *Methods for Identifying Biased Test Items*. Califonia: Sage Publications, Inc.

Chang, H., & Mazzeo, J. (1994). The Unigue Correspondence of the Item Response Function and Item Category Response Fundtions in Polytomously Scored Item Response Model. *Psychometrika* 59: 391-404.

Dorans, N. J., & Kulick, E. (1986). Demonstrating the Utility of the Standardization Approach to Assessing Unexpected Differential Item Performance on the Scholastic Aptitude Test. *Journal of Educational Measurement* 23: 355-368.

Flowers, C. P., et al. (1997). The Relationship between Polytomous DFIT and Other Polytomous DIF Procedures. *Paper presented at the Annual Meeting of the National Council on Measurement in Education,* 25-27. Chicago.

French, A. W., & Miller, T. R. (1996). Logistic Regresstion and Its Use in Detection Differential Item Functioning in Polytomous Items. *Journal of Education Measurement* 33: 315-332.

Holland, P. W., & Wainer, H. (1993). *Differential Item Functioning*. Hillsdale, New Jersey: Lowrence Erlbaum Associates.

Mazor, F. M., Clauser, B. E., & Hambleton, R. K. (1992). The Effect of Sample Size on the Functioning of the Mantel-Haenszel Statistic. *Education and Psychological Measurement* 52: 443-451.

Miller, T. R., & Spray, J. A. (1993). Logistic Discriminant Function Analysis for DIF Identification of Polytomous Scored Items. *Journal of Educational Measurement* 30: 107-122.

Millsap, R. E., & Everson, H. T. (1993). Methodology Review: Statistical Approaches for Assessing Measurement Bias. *Applied Psychological Measurement* 17: 297-334.

Narayanan, P. & Swaminathan, H. (1994). Performance of the Mantel-Haenzel and Simultaneous Item Bias Procedures for Detecting Differential Item Functioning. *Applied Psychological Measurement,* 18: 315-328.

Narayanan, P. & Swaminathan, H. (1996). Identification of Items that Show Nonuniform DIF. *Applied Psychological Measurement,* 20: 257-274.

Oishi, S. (2006). The Concept of Life Satisfaction Across Culture: An IRT Analysis. *Journal of Research in Personality* 40: 411-423.

Roussos, L. & Stout, W. (1996). Simulation Studies of Effect of Small Sample Size and Studied Item Parameter on SIBTEST and Mantel-Haenzel Type I Error Performance. *Journal of Educational Measurement*, 33: 215-230.

Scheuneman, J. D. (1979). A Method of Assessing Bias in Test Items. *Journal of Educational Measurement* 16: 143-152.

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.

Ya-Hui, S. & Wen-Chung, W. (2005). Efficiency of the Mantel, Generalized Mantel-Haenszel, and Logistic Discriminant Function Analysis Methods in Detectioning for Polytomous Items. *Journal of Applied Measurement in Education* 18: 313-350.

Zwick, R., Thayer, D. T., & Mazzeo, J. (1997). Descriptive and Inferential Procedures for Assessing Differential Item functioning in Polytomous Items. *Journal of Applied Measurement in Education* 10: 321-344.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ก

แบบสอบและกระดาษคำตอบที่ใช้ในการวิจัย

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน วิชาคณิตศาสตร์
เรื่อง จำนวนเต็ม เศษส่วน และทศนิยม

<u>คำชี้แจงในการตอบแบบทดสอบ</u>

1. แบบทดสอบฉบับนี้เป็นแบบทดสอบเลือกตอบ**ที่มีคำตอบถูกหลายคำตอบ** โดยในหนึ่งข้อจะมีตัวเลือกที่เป็นคำตอบถูกตั้งแต่สองตัวเลือกจนถึงสี่ตัวเลือก ในการตอบนักเรียนจะต้องพิจารณา**ตอบทุกตัวเลือก โดยทำเครื่องหมายถูก (ü) หน้าตัวเลือกที่เห็นว่าถูก และทำเครื่องหมายผิด (O) หน้าตัวเลือกที่เห็นว่าผิด** ลงในกระดาษคำตอบ
2. ข้อสอบมีทั้งหมด 40 ข้อ จำนวน 6 หน้า เวลาในการทำข้อสอบ 90 นาที (1 ชั่วโมง 30 นาที)
3. **ให้นักเรียนกรอกรายละเอียดลงบนกระดาษคำตอบ** แต่กรุณาอย่าทำเครื่องหมายใด ๆ ลงในแบบทดสอบ ถ้าต้องการทดให้ทดในกระดาษที่สำรองไว้
4. ถ้านักเรียนมีข้อสงสัยประการใดให้ถามผู้ดำเนินการสอบก่อนลงมือทำ
5. ให้นักเรียนตรวจทานก่อนส่งกระดาษคำตอบว่าได้ทำเครื่องหมายลงในกระดาษคำตอบครบทุกตัวเลือก

<u>ตัวอย่างการตอบแบบทดสอบ</u>

<u>คำถาม</u>

ข้อ 0 ข้อสรุปเกี่ยวกับจำนวนในข้อใดเป็นจริง
- ก. 1 เป็นจำนวนนับที่มีค่าน้อยที่สุด
- ข. 0 เป็นจำนวนนับ
- ค. 1 เป็นจำนวนตรงข้ามของ -1
- ง. จำนวนนับก็คือจำนวนเต็ม
- จ. จำนวนเต็มแบ่งเป็นจำนวนเต็มบวกและจำนวนเต็มลบ

<u>กระดาษคำตอบ</u>

ข้อ	ก	ข	ค	ง	จ
0	ü	O	ü	O	O
00					

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

1. ข้อสรุปเกี่ยวกับจำนวนเต็มในข้อใดเป็นจริง

ก. ถ้า a เป็นจำนวนเต็มบวกแล้ว a มากกว่าศูนย์

ข. ถ้า a เป็นจำนวนเต็มลบ และ b เป็นจำนวนเต็มบวก แล้ว a น้อยกว่า b

ค. ถ้า a เป็นจำนวนเต็มลบ แล้วจำนวนตรงข้ามของ a มีค่ามากกว่าศูนย์

ง. ถ้า b เป็นจำนวนตรงข้ามของ a แล้ว a อยู่ห่างจากศูนย์น้อยกว่า b

จ. ถ้า a และ b เป็นจำนวนเต็มลบ และ a น้อยกว่า b แล้ว a อยู่ห่างจากศูนย์น้อยกว่า b

2. ข้อสรุปเกี่ยวกับจำนวนเต็มในข้อใดเป็นจริง

ก. 0 เป็นจำนวนเต็มที่น้อยที่สุด

ข. จำนวนตรงข้ามของศูนย์เท่ากับศูนย์

ค. 1 เป็นจำนวนเต็มบวกที่มีค่าน้อยที่สุด

ง. -1 เป็นจำนวนเต็มลบที่มีค่าน้อยที่สุด

จ. -1 อยู่ห่างจากศูนย์เท่ากับ 1

3. ข้อใดเป็นจำนวนเต็มทุกจำนวน

ก. 1, 0, -1, -2

ข. 0, 2, 3, 4

ค. -1, -2, -3, -4

ง. 0.0, -1.5, -2.0, 3.0

จ. $\frac{-4}{2}, \frac{0}{2}, \frac{4}{2}, \frac{6}{2}$

4. ให้ a และ b เป็นจำนวนเต็มใด ๆ ค่าของ a, b ในข้อใดทำให้ผลลัพธ์ของ a-b เป็นจำนวนเต็มลบ

ก. a เป็นจำนวนเต็มลบ และ b เป็นศูนย์

ข. a และ b เป็นจำนวนเต็มลบ โดยที่ a อยู่ห่างจากศูนย์มากกว่า b

ค. a และ b เป็นจำนวนเต็มบวก โดยที่ a อยู่ห่างจากศูนย์มากกว่า b

ง. a เป็นจำนวนเต็มลบ และ b เป็นจำนวนเต็มบวก

จ. a เป็นศูนย์ และ b เป็นจำนวนเต็มบวก

5. กำหนด x และ y เป็นจำนวนเต็มใด ๆ จงพิจารณาว่าข้อใดเป็นจริง

ก. x + y = y + x

ข. x + y = - (-x - y)

ค. x + y = x - y

ง. x + y = - (x - y)

จ. x + y = x + (-y)

6. กำหนด P + Q = 10 และ R + S = -12 ดังนั้น (P + Q) - (R + S) มีค่าเท่ากับข้อใด

ก. 10 - 12

ข. 10 - (-12)

ค. 10 + 12

ง. (-12) + 10

จ. -10 -12

7. $[(-a) + b] - [(-c) + d]$ มีค่าเท่าใด เมื่อ a = 5, b = 4, c = 3 และ d = -2

ก. a + (-d)

ข. a - b + c

ค. b + c + d

ง. $1 \times b$

จ. 2c + d

8. $[(-10) + 8] - [(-9) + 2]$ มีค่าเท่ากับข้อใด

ก. (-2) + 7

ข. (-17) + 10

ค. 2 + (-7)

ง. 5 + 0

จ. 2 + 11

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

9. ให้ a, b, c เป็นจำนวนเต็มใด ๆ ค่าของ a, b, c ในข้อใดทำให้ผลลัพธ์ของ $a \times b \times c$ เป็นจำนวนเต็มบวก

ก. a และ b เป็นจำนวนเต็มบวก c เป็นจำนวนเต็มลบ

ข. a เป็นจำนวนเต็มบวก b และ c เป็นจำนวนเต็มลบ

ค. a และ b เป็นจำนวนเต็มลบ c เป็นจำนวนเต็มบวก

ง. a, b และ c เป็นจำนวนเต็มบวก

จ. a, b และ c เป็นจำนวนเต็มลบ

10. ข้อใดสรุปผลการบวกหรือลบจำนวนเต็มได้ถูกต้อง

ก. (-7) - (-8) = 1 เป็นจริง เพราะว่า (-7) - (-8) = -7 + (8) = 1

ข. 8 - (-7) = 1 เป็นจริง เพราะว่า 8 - (-7) = 8 - 7 = 1

ค. 7 - 8 = -1 เป็นจริง เพราะว่า 7 - 8 = 7 + -8 = -1

ง. -8 + 7 = 7 + (-8) เป็นจริง เพราะว่าจำนวนเต็มมีสมบัติการสลับที่สำหรับการบวก

จ. -7 - 8 = 1 ไม่จริง เพราะว่า -7 - 8 = (-7) + (-8) = -15

11. ข้อใดสรุปผลการคูณ หรือหารจำนวนเต็มได้ถูกต้อง

ก. $(-10) \times 5 = -50$ เป็นจริง เพราะว่า จำนวนเต็มลบคูณกับจำนวนเต็มบวกให้ผลลัพธ์เป็นจำนวนเต็มลบ

ข. $10 \times (-5) = -50$ เป็นจริง เพราะว่า จำนวนเต็มบวกคูณกับจำนวนเต็มลบให้ผลลัพธ์เป็นจำนวนเต็มลบ

ค. $(-10) \times (-5) = -50$ เป็นจริง เพราะว่า จำนวนเต็มลบคูณกับจำนวนเต็มลบให้ผลลัพธ์เป็นจำนวนเต็มลบ

ง. $(-10) \div (-5) = -2$ เป็นจริง เพราะว่า จำนวนเต็มลบหารกับจำนวนเต็มลบให้ผลลัพธ์เป็นจำนวนเต็มลบ

จ. $10 \div -5 = 2$ ไม่จริง เพราะว่า จำนวนเต็มบวกหารกับจำนวนเต็มลบให้ผลลัพธ์เป็นจำนวนเต็มลบ

12. กำหนด $(a \times b) = -15$ และ $(c \times d) = 3$ ดังนั้น ผลลัพธ์ของ $(a \times b) \div (c \times d)$ จะมีค่าเท่ากับข้อใด

ก. $(-15) \div 3$

ข. $15 \div (-3)$

ค. $3 \div (-15)$

ง. $(-15) \div [(-1) \times (-3)]$

จ. $[(-15) \times (-1)] \div (-3)$

13. ให้ a = -8, b = 4, c = -2 ดังนั้น $(a \div b) \times c$ มีค่าเท่ากับข้อใด

ก. $(-2) \times c$

ข. $b \times c$

ค. $a \div c$

ง. $(2 \times b) \div c$

จ. $(a \times c) \div b$

14. กำหนดให้ a และ b เป็นจำนวนเต็มลบ โดยที่ a มากกว่า b และ c เป็นจำนวนเต็มบวก ข้อใดสรุปผลลัพธ์ได้ถูกต้อง

ก. a - b + c เป็นจำนวนเต็มบวก

ข. a + b - c เป็นจำนวนเต็มบวก

ค. c -b - a เป็นจำนวนเต็มบวก

ง. b - a - c เป็นจำนวนเต็มลบ

จ. a - c + b เป็นจำนวนเต็มลบ

15. กำหนด a, b, c เป็นจำนวนเต็มลบใด ๆ และ a = b ข้อใดสรุปผลลัพธ์ได้ถูกต้อง

ก. ab = bc

ข. ab มากกว่า 0

ค. abc มากกว่า 0

ง. $a^2 = b^2$

จ. a + c = b + c

16. กำหนด a เป็นจำนวนเต็มใด ๆ ประโยคในข้อใดเป็นจริง

ก. a + 1 = 1 + a

ข. a - 1 = 1 - a

ค. a × 1 = 1 × a

ง. $\frac{1}{a} = \frac{a}{1}$

จ. $\frac{a}{1} = a \times 1$

17. ให้ a, b และ c เป็นจำนวนเต็มใด ๆ ข้อกำหนดในข้อใดทำให้ (a + b) × c = 0

ก. c มีค่าเป็น 0

ข. a เป็นจำนวนตรงข้ามของ b

ค. a + b มีค่าเท่ากับ 0

ง. a + c มีค่าเท่ากับ 0

จ. b ÷ c มีค่าเท่ากับ 1

18. กำหนด a เป็นจำนวนเต็มใด ๆ ข้อใดสรุปได้ถูกต้องตามสมบัติของศูนย์และหนึ่ง

ก. a × [(-5) + 5] = 0 เป็นจริง เพราะว่า (-5) + 5 = 0 และ a × 0 = 0

ข. a + [(-3) + 3] = a เป็นจริง เพราะว่า (-3) + 3 = 0 และ a + 0 = a

ค. a ÷ [(-1) × 1] = a เป็นจริง เพราะว่า (-1) × 1 = 1 และ a ÷ 1 = a

ง. a ÷ [(-11) + 11] = 0 ไม่จริง เพราะว่า (-11) + 11 = 0 และ a ÷ 0 = หาค่าไม่ได้

จ. a × [(-5) + 6] = 0 ไม่จริง เพราะว่า (-5) + 6 = 1 และ a × 1 = a

19. ถ้า $(a \times 10^2) + (b \times 10^2) = c \times 10^2$ โดยที่ a และ b เป็นจำนวนเต็มบวก แล้วข้อใดกล่าวถูกต้อง

ก. b น้อยกว่า c

ข. a น้อยกว่า c

ค. a มากกว่า b

ง. ab = c

จ. a +b = c

20. กำหนด x, y, z เป็นจำนวนเต็มใด ๆ ข้อใดกล่าวถึงสมบัติของจำนวนเต็มได้ถูกต้อง

ก. x(yz) = x(zy) ตามสมบัติการสลับที่ของการคูณ

ข. (yz)x = x(yz) ตามสมบัติการสลับที่ของการคูณ

ค. x(y + z) = xy + xz ตามสมบัติการแจกแจง

ง. x + y = y + x ตามสมบัติการสลับที่ของการบวก

จ. (y + z)x = xy + xz ตามสมบัติการเปลี่ยนกลุ่มของการบวก

21. $\frac{a}{b}$ จะเป็นเศษส่วนจำนวนคละเมื่อใด

ก. a = 10, b = 7

ข. a = 8, b = 0

ค. a = 5, b = 7

ง. a = 7, b = 4

จ. a = 0, b = 5

22. เศษส่วนในข้อใดมีค่าเท่ากันทุกจำนวน

ก. $\frac{1}{2}, \frac{2}{4}, \frac{3}{6}$

ข. $\frac{1}{3}, \frac{2}{6}, \frac{3}{9}$

ค. $\frac{1}{4}, \frac{2}{8}, \frac{4}{12}$

ง. $\frac{2}{3}, \frac{4}{6}, \frac{6}{9}$

จ. $\frac{16}{24}, \frac{24}{36}, \frac{32}{48}$

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

23. ผลลบของ $7\frac{2}{3}$ กับ $2\frac{1}{3}$ มีค่าตรงกับข้อใด

ก. (7 - 2) + $\left(-\frac{2}{3}-\frac{1}{3}\right)$

ข. (7 - 2) + $\left(\frac{2}{3}-\frac{1}{3}\right)$

ค. (7 - 2) - $\left(\frac{2}{3}+\frac{1}{3}\right)$

ง. $\frac{23}{3}+\left(-\frac{7}{3}\right)$

จ. $\frac{23}{3}-\frac{7}{3}$

24. ในการคูณเศษส่วนเกี่ยวข้องกับวิธีการในข้อใด

ก. การหา ค.ร.น.

ข. การทอนเศษส่วน

ค. กลับตัวเศษและตัวส่วนของเศษส่วนที่เป็นตัวหาร

ง. การคูณเศษกับเศษ และส่วนกับส่วน

จ. การทำเศษส่วนจำนวนคละให้เป็นเศษเกิน

25. ให้ b = $\frac{1}{2}$ - $\frac{1}{4}$ ค่าของ 12b ตรงกับข้อใด

ก. $\frac{1}{2}\times 6$

ข. $\frac{15}{3}\times\frac{3}{5}$

ค. $\frac{3}{5}\times\frac{5}{3}$

ง. $1\div\frac{1}{3}$

จ. $\frac{5}{6}\div\frac{2}{5}$

26. ข้อใดสรุปผลการคูณหรือหารเศษส่วนได้ถูกต้อง

ก. $\frac{2}{8}\times\frac{16}{5}=\frac{8}{40}$ ไม่จริง เพราะว่าทอนตัวเศษกับตัวเศษ

ข. $\frac{2}{5}\times\frac{14}{3}=\frac{16}{8}$ ไม่จริง เพราะว่านำตัวเศษรวมกับตัวเศษและตัวส่วนรวมกับตัวส่วน

ค. $\frac{2}{3}\div\frac{4}{15}=\frac{5}{2}$ ไม่จริง เพราะว่า $\frac{2}{3}\div\frac{4}{15}=\frac{2}{3}\times\frac{4}{15}=\frac{8}{45}$

ง. $\frac{2}{7}\div\frac{5}{9}=\frac{18}{35}$ ไม่จริง เพราะว่า $\frac{2}{7}\div\frac{5}{9}=\frac{7}{2}\times\frac{5}{9}=\frac{35}{18}$

จ. $\frac{3}{7}\div\frac{21}{5}=\frac{21}{5}\div\frac{3}{7}$ ไม่จริง เพราะว่าจำนวนเศษส่วนไม่มีสมบัติการสลับที่สำหรับการหาร

27. นายสุชาติแบ่งเงิน x บาท ให้บุตร 4 คน โดยบุตรคนโตได้รับส่วนแบ่งเป็นครึ่งหนึ่งของเงินทั้งหมด ที่เหลือแบ่งให้น้องอีก 3 คน ๆ ละ เท่ากันจงพิจารณาว่าข้อใดสรุปได้ถูกต้อง

ก. จำนวนเงินที่บุตรคนโตได้รับเป็น $\frac{x}{2}$ บาท

ข. จำนวนเงินที่น้องทั้งสามคนได้รับรวมเป็น $\frac{x}{2}$ บาท

ค. น้อง ๆ ได้รับเงินคนละ $\frac{x}{2}\div 3$ บาท

ง. จำนวนเงินที่เหลือจากให้บุตรคนโตเป็น x-$\frac{1}{2}$ บาท

จ. จำนวนเงินที่บุตรทั้ง 4 ได้รับรวมเป็น $\frac{x}{2}+\frac{x}{2}$ บาท

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

28. ให้ p, q, และ r เป็นจำนวนเต็มบวก โดยที่ p<q และ q<r ข้อใดสรุปค่าเศษส่วนได้ถูกต้อง

ก. $\frac{r}{q}$ มากกว่า 1

ข. $\frac{p}{q}$ มากกว่า 1

ค. $\frac{q}{r}$ น้อยกว่า 1

ง. $\frac{r+q}{p}$ น้อยกว่า 1

จ. $\frac{r}{q+r}$ น้อยกว่า 1

29. ค่าของเศษส่วนในข้อใด เท่ากับ $\frac{a}{b}$

ก. $\frac{a}{b}\times\frac{5}{5}$

ข. $\frac{a}{b}\div\frac{5}{5}$

ค. $\frac{a}{b}\div\frac{a}{b}$

ง. $\frac{a}{b}+\frac{5}{5}$

จ. $\frac{a}{b}-\frac{5}{5}$

30. ข้อใดมีผลลัพธ์เท่ากับ $\frac{4}{7}$

ก. $\frac{2}{7}\times 2$

ข. $\frac{2}{7}\div\frac{1}{2}$

ค. $\frac{1}{7}\div\frac{4}{28}$

ง. $\frac{4}{7}\times\frac{7}{8}\times\frac{8}{7}$

จ. $\left(\frac{14}{3}\div\frac{7}{3}\right)\times\frac{2}{7}$

31. จำนวนทศนิยมในข้อใดเรียงลำดับจากน้อยไปหามาก ได้ถูกต้อง

ก. $2.57\dot{8}$ $2.5\dot{7}\dot{8}$ $2.\dot{5}7\dot{8}$

ข. 2.578 $2.\dot{5}7\dot{8}$ $2.5\dot{7}\dot{8}$

ค. 2.578 $2.5\dot{7}\dot{8}$ $2.57\dot{8}$

ง. 2.5708 2.578 $2.57\dot{8}$

จ. 2.579 $2.5\dot{7}\dot{8}$ $2.57\dot{8}$

32. ให้ a = 20.8759, b = 20.5879, c = 20.7589 และ d = 20.9578 ข้อใดเรียงลำดับจำนวนทศนิยมจากมากไปน้อยได้ถูกต้อง

ก. a > c > b

ข. a > b > c

ค. d > a > c

ง. d > a > b

จ. d > c > b

33. ข้อใดเรียงลำดับจำนวนจากน้อยไปมากได้ถูกต้อง

ก. $\frac{1}{2}\langle 0.56\langle\frac{1}{4}$

ข. $\frac{7}{8}\langle 1.2\langle 1.3$

ค. $\frac{1}{10}\langle 0.2\langle\frac{3}{5}$

ง. $0.34\langle\frac{1}{4}\langle\frac{2}{3}$

จ. $0.78\langle\frac{9}{10}\langle 1.1$

34. ผลลัพธ์ที่ได้จากการบวกหรือลบจำนวนทศนิยม ในข้อใด มีทศนิยม 4 ตำแหน่ง

ก. 130.2 + 1.4568

ข. 10.78 + 1.304

ค. 15.347 + 10.1249

ง. 1.7856 - 0.49

จ. 30.18 - 15.2784

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

35. A และ B เป็นจำนวนที่อยู่ในรูปทศนิยม ถ้าผลคูณของ A กับ B มีทศนิยม 6 ตำแหน่ง แล้วจงพิจารณาว่าข้อสรุปใดถูกต้อง

ก.A มีทศนิยม 1 ตำแหน่ง และ B มีทศนิยม 6 ตำแหน่ง โดยที่ผลคูณตัวสุดท้ายไม่เป็น 0

ข. A มีทศนิยม 2 ตำแหน่ง และ B มีทศนิยม 4 ตำแหน่ง โดยที่ผลคูณตัวสุดท้ายไม่เป็น 0

ค. A มีทศนิยม 2 ตำแหน่ง และ B มีทศนิยม 3 ตำแหน่ง โดยที่ผลคูณตัวสุดท้ายไม่เป็น 0

ง. A มีทศนิยม 3 ตำแหน่ง และ B มีทศนิยม 3 ตำแหน่ง โดยที่ผลคูณตัวสุดท้ายไม่เป็น 0

จ. A มีทศนิยม 5 ตำแหน่ง และ B มีทศนิยม 1 ตำแหน่ง โดยที่ผลคูณตัวสุดท้ายไม่เป็น 0

36. ข้อใดสรุปผลลัพธ์ที่ได้จากการคูณหรือหารจำนวนทศนิยมได้ถูกต้อง

ก. 25.3×0.4 มีค่ามากกว่า 25.3

ข. 0.8×0.9 มีค่ามากกว่า 1

ค. $4.6 \div 0.6$ มีค่ามากกว่า 4.6

ง. $3.7 \div 4.7$ มีค่าน้อยกว่า 1

จ. $5.2 \div 2.2$ มีค่ามากกว่า 1

37. 0.01 หารด้วย 0.1 มีค่าเท่ากับข้อใด

ก. $\dfrac{\frac{1}{100}}{\frac{1}{10}}$

ข. $\dfrac{1}{100} \times \dfrac{10}{1}$

ค. $\dfrac{0.01}{\frac{1}{10}}$

ง. 0.01×10

จ. $\dfrac{1}{10} \times \dfrac{100}{1}$

38. ข้อใดสรุปผลการคูณหรือหารทศนิยมได้ถูกต้อง

ก. $2.35 \times 4.2 = 98.7$ ไม่จริง เพราะว่า ผลคูณของ 2.35 ซึ่งมีค่าน้อยกว่า 3 กับ 4.2 ซึ่งมีค่าน้อยกว่า 5 ต้องมีค่าน้อยกว่า 15

ข. $10.2 \times 0.12 = 1.224$ เป็นจริง เพราะว่า จำนวนที่มีทศนิยม 1 ตำแหน่งคูณด้วยจำนวนที่มีทศนิยม 2 ตำแหน่ง จะได้ผลคูณเป็นจำนวนที่มีทศนิยม 3 ตำแหน่ง

ค. $10.5 \div 0.4 = 26.25$ เป็นจริง เพราะว่า ตัวหารมีค่าน้อยกว่า 1 เมื่อหาร 10.5 จะทำให้ผลลัพธ์ที่ได้มีค่ามากกว่า 10.5

ง. $20.4 \div 1.2 = 204 \div 12$ เป็นจริง เพราะว่าการหารด้วยจำนวนทศนิยมจะเลื่อนตำแหน่งของตัวหารจนเป็นจำนวนเต็มก่อนหาผลหาร

จ. $14.5 \div 2.4 = 2.4 \div 14.5$ เป็นจริง เพราะว่าจำนวนทศนิยมมีสมบัติการสลับที่สำหรับการหาร

39. เชือกเส้นหนึ่งยาว y เมตร ตัดแบ่งออกเป็น 4 เส้น สองเส้นแรกยาวเส้นละ 1.25 เมตร เส้นที่สามยาว 1.50 เมตร ส่วนที่เหลือยาวเท่าใด

ก. $[(2 \times 1.25) + 1.50] - y$

ข. $y - [(2 \times 1.25) + 1.50]$

ค. $y - (1.25 + 1.50)$

ง. $y - 2.50 - 1.50$

จ. $y - 4.00$

40. ให้ a เป็นจำนวนทศนิยมที่มีค่ามากกว่า 1 ข้อใดสรุปผลคูณหรือผลหารของ a ได้ถูกต้อง

ก. $\dfrac{1}{a} > 1$

ข. $\dfrac{1}{a} < a$

ค. $a \times a > 1$

ง. $a \times a < a$

จ. $a \times a > \dfrac{1}{a}$

ขอขอบคุณนักเรียนที่ตั้งใจทำ ขอให้โชคดีค่ะ

กระดาษคำตอบ

ชื่อ-สกุล..เพศ ชาย หญิง เลขที่..................
โรงเรียน.. ห้อง.......................................

ข้อ	ก	ข	ค	ง	จ	ข้อ	ก	ข	ค	ง	จ
1						21					
2						22					
3						23					
4						24					
5						25					
6						26					
7						27					
8						28					
9						29					
10						30					
11						31					
12						32					
13						33					
14						34					
15						35					
16						36					
17						37					
18						38					
19						39					
20						40					

J ขอขอบคุณนักเรียนทุกคนนะคะที่ตั้งใจทำ ขอให้นักเรียนทุกคนโชคดี **J**

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ข

ตัวอย่าง print out หาค่าความสามารถของผู้สอบ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตัวอย่าง print out หาค่าความสามารถของผู้สอบ

```
MULTILOG--FOR MULTIPLE CATEGORICAL ITEM RESPONSE DATA--VERSION 7.0.3
MULTILOG for Windows 7.00.2327.2
Created on: 6 March 2008, 00:37:29

 >PROBLEM SCORE,
         INDIVIDUAL,
         DATA = 'D:\boy.DAT',
         NITEMS = 40,
         NGROUPS = 1,
         NEXAMINEES = 892,
         NCHARS = 3;
   DATA FILE NAME IS
         D:\BOY.DAT

 TYPE OF INPUT:
    INDIVIDUAL RESPONSE VECTORS
 >TEST ALL,
      GRADED,
      NC = (6(0)40);
 NUMBER OF CODES  7
0123456
  VECTOR OF CATEGORIES FOR CODE=0
0000000000000000000000000000000000000000
  VECTOR OF CATEGORIES FOR CODE=1
1111111111111111111111111111111111111111
  VECTOR OF CATEGORIES FOR CODE=2
2222222222222222222222222222222222222222
  VECTOR OF CATEGORIES FOR CODE=3
3333333333333333333333333333333333333333
  VECTOR OF CATEGORIES FOR CODE=4
4444444444444444444444444444444444444444
  VECTOR OF CATEGORIES FOR CODE=5
5555555555555555555555555555555555555555
```

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

VECTOR OF CATEGORIES FOR CODE=6

6666666666666666666666666666666666666666

(3A1,3X,40A1)

MULTILOG--FOR MULTIPLE CATEGORICAL ITEM RESPONSE DATA--VERSION 7.0.3

MULTILOG for Windows 7.00.2327.2

Created on: 6 March 2008, 00:37:29

DATA PARAMETERS:

NUMBER OF LINES IN THE DATA FILE: 892

NUMBER OF CATEGORICAL-RESPONSE ITEMS: 40

NUMBER OF CONTINUOUS-RESPONSE ITEMS, AND/OR GROUPS: 1

TOTAL NUMBER OF "ITEMS" (INCLUDING GROUPS): 41

NUMBER OF CHARACTERS IN ID FIELDS: 3

MAXIMUM NUMBER OF RESPONSE-CODES FOR ANY ITEM: 7

THE MISSING VALUE CODE FOR CONTINUOUS DATA: 9.0000

THE DATA WILL BE STORED IN MEMORY

ESTIMATION PARAMETERS:

MODAL THETAS WILL BE ESTIMATED

OUTPUT-CONTROL PARAMETERS

THE MODAL THETAS WILL BE PRINTED

MULTILOG for Windows 7.00.2327.2

READING DATA...

KEY-

CODE CATEGORY

0 0000000000000000000000000000000000000000

1 1111111111111111111111111111111111111111

2 2222222222222222222222222222222222222222

3 3333333333333333333333333333333333333333

4 4444444444444444444444444444444444444444

5 5555555555555555555555555555555555555555

6 6666666666666666666666666666666666666666

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

FORMAT FOR DATA-

(3A1,3X,40A1)

FIRST OBSERVATION AS READ-

ID 001

ITEMS 2332233314422343442334421422431321033351

NORML 0.000

SCORING DATA...

MULTILOG for Windows 7.00.2327.2

THETAHAT	S.E.	ITER	ID FIELD
-0.562	0.243	4	001
0.098	0.242	4	002
-0.463	0.245	4	003
-1.023	0.254	3	004
-1.095	0.252	4	005
0.096	0.240	4	006
-0.525	0.245	4	007
-0.560	0.235	4	008
-0.979	0.246	4	009
-0.584	0.249	4	010
-0.555	0.250	3	011
-0.799	0.256	3	012
-0.392	0.245	4	013
-0.797	0.245	4	014
-0.038	0.240	4	015
0.252	0.245	4	016
-0.624	0.249	4	017
-0.738	0.257	3	018
-0.501	0.248	3	019
-0.789	0.243	4	020
-0.029	0.243	4	021
0.133	0.239	4	022
-0.346	0.240	4	023

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

-0.521	0.246	3	024
-0.691	0.249	3	025
-0.390	0.249	3	026
-0.095	0.239	2	027
-0.524	0.236	4	028
-0.552	0.248	3	029
-0.603	0.246	4	030
0.535	0.232	4	031
-0.528	0.251	2	032
-0.439	0.242	4	033
-0.695	0.250	4	034
-0.651	0.246	4	035
-0.860	0.246	4	036
-0.683	0.243	3	037
-0.839	0.247	4	038
-0.351	0.238	4	039
-0.469	0.241	4	040
-0.659	0.244	4	041
-0.642	0.254	3	042
-0.903	0.248	4	043
-0.740	0.248	4	044
-0.741	0.250	3	045
-0.749	0.253	4	046
-0.640	0.251	3	047
-0.411	0.236	4	048
-0.621	0.236	4	049
-0.361	0.235	4	050
-0.756	0.249	4	051
-0.638	0.245	4	052
-0.731	0.248	4	053
0.158	0.242	4	054
-0.590	0.245	4	055
-0.222	0.252	2	056
-0.541	0.235	4	057
-0.625	0.247	4	058

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

-0.519	0.242	4	059
-0.568	0.251	3	060
-0.619	0.256	3	061
-0.305	0.251	2	062
-0.783	0.252	4	063
-0.606	0.247	4	064
-0.267	0.249	2	065
-0.389	0.244	4	066
-0.452	0.237	4	067
-0.847	0.263	4	068
0.019	0.240	4	069
0.312	0.240	3	070
-0.814	0.255	3	071
-0.785	0.264	4	072
-0.378	0.238	4	073
-0.605	0.249	4	074
-0.449	0.241	4	075
-0.200	0.245	2	076
-0.751	0.245	4	077
-0.676	0.250	3	078
-0.516	0.246	4	079
-0.319	0.245	2	080
-0.474	0.243	3	081
-0.796	0.250	4	082
0.234	0.241	4	083
-0.388	0.237	4	084
-0.562	0.253	3	085
-0.280	0.245	2	086
-0.354	0.236	4	087
0.946	0.226	5	088
-0.111	0.258	4	089
-0.380	0.245	2	090
-0.662	0.253	3	091
-0.607	0.243	4	092
-0.290	0.237	4	093

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

-0.038	0.240	4	094
-0.511	0.252	4	095
-0.561	0.235	4	096
-0.731	0.238	4	097
-0.246	0.248	2	098
-0.412	0.251	4	099
-0.605	0.245	4	100
-0.559	0.246	4	101
-0.981	0.263	3	102
-0.632	0.253	4	103
-0.572	0.242	4	104
-0.472	0.234	4	105
-0.593	0.249	4	106
-0.531	0.252	4	107
-0.854	0.243	3	108
-0.706	0.255	3	109
-0.487	0.246	4	110
0.666	0.232	5	111
-0.356	0.244	3	112
-0.132	0.238	2	113
-0.171	0.246	2	114
-0.662	0.242	4	115
-0.637	0.245	4	116
-0.266	0.239	2	117
-0.455	0.237	4	118
-0.287	0.240	2	119
-0.590	0.242	4	120
-0.563	0.236	4	121
-0.708	0.256	4	122
-0.493	0.245	4	123
-0.562	0.251	4	124
-0.759	0.244	4	125
-0.357	0.243	4	126
-0.475	0.246	3	127
-0.641	0.246	4	128

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

-0.761	0.245	4	129
-0.630	0.234	4	130
-0.385	0.240	4	131
-0.489	0.244	4	132
-0.379	0.248	2	133
-0.623	0.237	4	134
-0.631	0.244	4	135
-0.422	0.242	4	136
-0.517	0.252	3	137
-0.403	0.246	2	138
-0.249	0.240	2	139
-0.705	0.247	4	140
-0.346	0.240	3	141
-0.229	0.239	2	142
-0.740	0.245	4	143
-0.474	0.248	3	144
-0.173	0.240	2	145
-0.613	0.244	4	146
-0.298	0.240	4	147
-0.711	0.247	4	148
-0.608	0.243	4	149
-0.575	0.248	3	150
-0.860	0.252	4	151
-0.413	0.239	4	152
-0.762	0.244	4	153
-0.747	0.255	3	154
-0.599	0.249	4	155
-0.793	0.251	2	156
-0.612	0.249	3	157
-0.658	0.246	4	158
-0.526	0.244	3	159
-0.408	0.238	4	160
-0.719	0.244	4	161
-1.026	0.257	3	162
-0.765	0.250	4	163

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

-0.874	0.249	3	164
-0.476	0.238	4	165
-0.834	0.249	4	166
-0.508	0.237	4	167
-0.681	0.250	3	168
-0.753	0.248	4	169
-0.582	0.237	4	170
-0.869	0.246	3	171
-0.920	0.251	3	172
-0.581	0.244	4	173
-0.587	0.247	4	174
-0.434	0.243	3	175
-0.888	0.255	4	176
-0.481	0.244	4	177
-0.523	0.244	4	178
-0.920	0.249	3	179
-0.711	0.251	4	180
0.112	0.236	4	181
-0.324	0.232	4	182
-0.624	0.250	3	183
-0.584	0.244	4	184
-0.567	0.241	4	185
-0.517	0.239	4	186
-0.591	0.242	4	187
-0.820	0.250	4	188
-0.906	0.239	3	189
-0.684	0.249	4	190
-0.509	0.233	4	191
-0.891	0.245	3	192
-0.716	0.240	4	193
-0.502	0.242	4	194
-0.709	0.244	4	195
-0.735	0.254	3	196
-0.705	0.262	4	197
-0.827	0.250	2	198

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

-0.676	0.239	4	199
-0.543	0.250	3	200
-0.349	0.241	3	201
-0.421	0.248	2	202
-0.485	0.238	4	203
-0.580	0.240	4	204
-0.132	0.247	3	205
-0.797	0.253	4	206
-0.436	0.238	4	207
-0.442	0.240	4	208
-0.805	0.243	3	209
-0.974	0.263	2	210
-0.821	0.237	3	211
-0.641	0.252	4	212
-0.843	0.240	3	213
-0.019	0.242	4	214
-0.207	0.237	4	215
-0.403	0.246	2	216
-0.109	0.244	2	217
0.168	0.243	4	218
-0.155	0.241	2	219
-0.806	0.251	4	220
-0.737	0.245	4	221
-0.400	0.248	4	222
-0.056	0.237	2	223
-0.448	0.247	2	224
-0.664	0.247	4	225
-0.446	0.240	4	226
-0.826	0.246	4	227
-0.507	0.238	4	228
-0.584	0.245	2	229
-0.686	0.242	4	230
-0.149	0.245	2	231
0.079	0.244	4	232
0.543	0.238	5	233

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

-0.587	0.250	3	234
-0.253	0.233	4	235
0.106	0.241	4	236
-0.390	0.245	3	237
-0.104	0.250	2	238
-0.560	0.247	3	239
-0.532	0.245	2	240
-0.429	0.251	3	241
0.059	0.244	4	242
-0.443	0.251	2	243
-0.476	0.238	4	244
-0.790	0.245	4	245
0.151	0.238	4	246
0.095	0.239	4	247
-0.451	0.245	3	248
-0.261	0.246	2	249
-0.359	0.241	4	250
0.095	0.238	4	251
-0.077	0.241	3	252
-0.185	0.242	2	253
0.509	0.232	3	254
0.344	0.238	4	255
0.532	0.236	4	256
0.122	0.247	4	257
-0.228	0.250	2	258
-0.803	0.239	3	259
-0.868	0.245	3	260
-0.077	0.241	3	261
0.000	0.237	2	262
-0.737	0.243	4	263
-0.657	0.250	3	264
-0.327	0.240	2	265
-0.324	0.239	4	266
-0.295	0.243	2	267
0.349	0.234	4	268

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

0.357	0.236	4	269
0.703	0.234	5	270
0.410	0.235	4	271
0.167	0.236	4	272
0.457	0.237	3	273
0.217	0.238	4	274
-0.449	0.248	3	275
0.072	0.248	4	276
-0.031	0.250	4	277
0.243	0.245	4	278
-0.244	0.242	4	279
-0.423	0.248	4	280
-0.115	0.239	2	281
-0.255	0.248	2	282
-0.594	0.260	3	283
0.542	0.236	3	284
-0.223	0.247	3	285
0.114	0.239	4	286
-0.171	0.238	2	287
-0.094	0.251	4	288
0.069	0.243	4	289
0.100	0.245	4	290
-0.715	0.245	3	291
-0.348	0.248	2	292
0.731	0.234	5	293
-0.587	0.241	4	294
-0.719	0.250	4	295
-0.523	0.251	3	296
-0.531	0.237	4	297
-0.379	0.238	4	298
-0.770	0.245	4	299
0.000	0.243	4	300
-0.536	0.246	4	301
0.292	0.244	4	302
0.074	0.235	4	303

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

0.236	0.233	4	304
-0.542	0.240	4	305
-0.214	0.261	4	306
0.490	0.240	3	307
-0.799	0.244	4	308
0.203	0.235	4	309
-0.659	0.249	4	310
0.199	0.233	4	311
-0.433	0.246	3	312
-0.151	0.237	3	313
0.210	0.239	4	314
0.375	0.236	4	315
0.748	0.229	5	316
0.387	0.234	4	317
-0.165	0.234	3	318
0.177	0.236	4	319
-0.234	0.232	4	320
0.258	0.239	4	321
0.000	0.234	2	322
-0.321	0.238	2	323
0.163	0.233	4	324
-0.198	0.233	2	325
-0.038	0.239	2	326
-0.038	0.240	4	327
0.288	0.232	4	328
0.489	0.235	4	329
0.074	0.235	4	330
-0.161	0.232	3	331
0.106	0.231	3	332
0.186	0.236	4	333
0.092	0.235	4	334
-0.180	0.232	4	335
0.315	0.235	4	336
0.420	0.233	4	337
-0.261	0.236	4	338

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

0.579	0.234	4	339
-0.035	0.229	2	340
0.407	0.234	4	341
-0.269	0.240	2	342
0.457	0.232	4	343
-0.096	0.240	4	344
-0.175	0.229	4	345
-0.236	0.233	2	346
-0.179	0.232	4	347
-0.433	0.232	4	348
0.199	0.233	4	349
-0.261	0.237	2	350
-0.018	0.229	2	351
-0.106	0.230	3	352
-0.074	0.236	2	353
-0.129	0.235	2	354
0.150	0.237	4	355
0.173	0.234	4	356
0.070	0.229	2	357
-0.225	0.237	4	358
-0.222	0.235	2	359
0.313	0.234	4	360
-0.204	0.242	2	361
-0.198	0.232	4	362
0.295	0.235	4	363
0.733	0.230	5	364
0.082	0.234	2	365
-0.351	0.235	2	366
0.110	0.235	3	367
-0.428	0.238	4	368
0.784	0.229	5	369
0.588	0.235	5	370
0.650	0.232	5	371
0.691	0.232	5	372
0.558	0.237	5	373

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

0.373	0.235	4	374
0.443	0.233	4	375
0.132	0.237	4	376
0.677	0.236	5	377
0.349	0.240	4	378
0.837	0.230	5	379
0.530	0.242	5	380
-0.019	0.242	4	381
0.722	0.234	5	382
0.604	0.234	5	383
0.551	0.239	5	384
0.692	0.230	5	385
0.864	0.229	5	386
0.736	0.232	5	387
0.584	0.234	5	388
0.460	0.238	3	389
0.583	0.234	5	390
0.666	0.232	5	391
0.570	0.242	5	392
0.334	0.241	4	393
0.872	0.229	5	394
0.663	0.237	5	395
0.247	0.247	4	396
0.691	0.230	5	397
0.637	0.235	5	398
0.727	0.236	5	399
-0.086	0.254	4	400
-0.404	0.243	4	401
0.395	0.241	4	402
0.529	0.237	4	403
0.300	0.236	4	404
0.756	0.230	5	405
0.375	0.242	4	406
0.663	0.231	5	407
0.569	0.232	3	408

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

0.292	0.241	4	409
0.508	0.235	3	410
0.721	0.231	5	411
0.550	0.235	3	412
0.484	0.239	4	413
0.663	0.231	5	414
0.746	0.231	5	415
0.711	0.230	5	416
0.707	0.232	5	417
0.458	0.237	4	418
0.605	0.231	5	419
0.799	0.228	5	420
0.957	0.227	5	421
0.918	0.228	5	422
0.847	0.229	5	423
0.501	0.238	3	424
0.721	0.236	5	425
0.918	0.229	5	426
0.893	0.227	5	427
0.669	0.237	5	428
0.577	0.240	5	429
0.499	0.237	3	430
0.907	0.228	5	431
0.351	0.240	4	432
0.927	0.228	5	433
0.913	0.227	5	434
0.798	0.234	5	435
0.708	0.232	5	436
0.627	0.235	5	437
0.971	0.226	5	438
0.601	0.234	5	439
0.883	0.230	5	440
0.829	0.231	5	441
0.933	0.228	5	442
0.747	0.229	5	443

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

0.846	0.230	5	444
0.941	0.229	5	445
0.887	0.229	5	446
0.847	0.229	5	447
0.748	0.229	5	448
0.985	0.226	5	449
0.650	0.232	5	450
0.378	0.248	5	451
0.759	0.233	5	452
0.947	0.228	5	453
0.916	0.227	5	454
1.010	0.226	4	455
0.727	0.229	5	456
0.519	0.249	5	457
0.898	0.229	5	458
0.985	0.226	5	459
0.632	0.235	5	460
0.802	0.234	5	461
0.429	0.235	4	462
0.797	0.230	5	463
0.725	0.232	5	464
0.946	0.226	5	465
0.663	0.231	5	466
0.511	0.236	3	467
0.250	0.244	4	468
0.932	0.227	5	469
0.589	0.232	4	470
0.772	0.232	5	471
0.703	0.238	5	472
0.884	0.230	5	473
0.634	0.233	5	474
0.941	0.228	5	475
0.659	0.231	5	476
0.847	0.229	5	477
0.252	0.240	4	478

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

0.846	0.230	5	479
0.182	0.246	4	480
0.852	0.229	5	481
0.806	0.231	5	482
0.869	0.232	5	483
0.674	0.235	5	484
0.871	0.231	5	485
0.610	0.232	5	486
0.921	0.228	5	487
0.180	0.232	4	488
0.887	0.229	5	489
0.198	0.243	4	490
0.801	0.233	5	491
0.704	0.229	5	492
0.445	0.238	5	493
0.476	0.232	3	494
0.693	0.235	5	495
0.682	0.234	5	496
-0.147	0.251	2	497
0.736	0.230	5	498
0.480	0.242	5	499
0.697	0.235	5	500
0.814	0.228	5	501
0.892	0.229	5	502
1.004	0.226	4	503
0.450	0.235	4	504
0.757	0.230	5	505
1.024	0.225	4	506
-0.273	0.251	2	507
0.867	0.229	5	508
0.802	0.228	5	509
0.041	0.247	4	510
0.715	0.234	5	511
0.395	0.231	4	512
0.667	0.232	5	513

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

0.586	0.232	4	514
0.725	0.234	5	515
0.721	0.231	5	516
-0.643	0.247	4	517
-0.133	0.239	2	518
-0.445	0.240	4	519
-0.475	0.249	3	520
-0.848	0.240	3	521
-0.357	0.251	2	522
-0.205	0.248	2	523
-0.714	0.250	4	524
-0.476	0.244	4	525
-0.221	0.251	2	526
-0.354	0.239	4	527
-0.454	0.238	4	528
-0.287	0.244	4	529
-0.210	0.251	2	530
-0.544	0.253	3	531
-0.019	0.241	2	532
-0.539	0.244	4	533
-0.608	0.243	4	534
0.780	0.231	5	535
-0.614	0.243	4	536
-0.093	0.249	4	537
-0.578	0.235	4	538
-0.512	0.235	4	539
-0.167	0.236	3	540
0.295	0.238	4	541
-0.019	0.240	3	542
-0.944	0.246	4	543
0.152	0.246	4	544
-0.507	0.249	4	545
-0.638	0.241	4	546
-0.599	0.242	4	547
-0.360	0.257	2	548

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

-0.378	0.244	2	549
-0.698	0.237	4	550
-0.841	0.248	4	551
-0.619	0.243	4	552
-0.921	0.249	4	553
-0.558	0.249	3	554
-0.576	0.247	4	555
-0.722	0.248	4	556
-0.469	0.241	4	557
-0.827	0.245	4	558
-0.620	0.241	4	559
-0.693	0.249	4	560
-0.545	0.248	3	561
-0.484	0.246	3	562
-0.730	0.247	4	563
-0.981	0.249	3	564
-0.718	0.244	4	565
-0.837	0.246	4	566
-0.542	0.249	4	567
-0.932	0.250	3	568
-0.686	0.242	4	569
-0.992	0.255	3	570
-0.590	0.249	4	571
-0.748	0.243	4	572
-0.691	0.241	4	573
-0.618	0.251	3	574
-0.613	0.245	4	575
-0.487	0.238	4	576
-0.839	0.249	3	577
-0.594	0.245	4	578
-0.558	0.239	4	579
-0.787	0.253	3	580
-0.665	0.246	3	581
-0.719	0.240	4	582
-0.790	0.249	4	583

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

-0.567	0.243	4	584
-0.575	0.239	4	585
-0.674	0.251	2	586
-0.670	0.259	3	587
-0.330	0.234	4	588
-0.385	0.237	4	589
-0.462	0.245	4	590
-0.604	0.247	4	591
-0.650	0.248	4	592
-0.506	0.245	4	593
-0.644	0.249	4	594
-0.621	0.243	4	595
-0.543	0.236	4	596
-0.620	0.242	4	597
-0.500	0.240	3	598
-0.784	0.242	4	599
-0.608	0.254	4	600
-0.804	0.245	3	601
-0.558	0.244	3	602
-0.487	0.240	4	603
-0.546	0.239	4	604
-0.610	0.245	4	605
-0.489	0.236	4	606
-0.693	0.253	4	607
-0.756	0.256	2	608
-0.186	0.237	2	609
-0.709	0.240	4	610
-0.794	0.257	4	611
-0.721	0.249	4	612
-0.369	0.242	3	613
-1.006	0.245	3	614
-0.645	0.241	4	615
-0.422	0.242	4	616
-0.380	0.242	2	617
-0.436	0.238	4	618

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

-0.279	0.236	2	619
-0.575	0.246	3	620
0.196	0.242	4	621
-0.294	0.238	4	622
-0.019	0.237	2	623
-0.265	0.243	4	624
-0.077	0.241	2	625
0.213	0.246	4	626
-0.316	0.252	2	627
-0.373	0.249	2	628
-0.382	0.253	2	629
-0.020	0.245	4	630
0.113	0.248	4	631
-0.122	0.247	4	632
-0.571	0.244	4	633
-0.502	0.246	3	634
0.293	0.237	4	635
-0.594	0.238	4	636
-0.272	0.246	2	637
-0.637	0.248	3	638
-0.217	0.244	2	639
-0.400	0.239	3	640
-0.232	0.241	2	641
0.071	0.247	4	642
0.297	0.242	4	643
0.100	0.245	4	644
0.367	0.239	4	645
0.126	0.240	4	646
0.367	0.239	4	647
-0.558	0.244	4	648
0.271	0.240	4	649
-0.240	0.240	2	650
-0.322	0.246	2	651
-0.226	0.243	2	652
-0.292	0.255	2	653

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

-0.556	0.246	3	654
-0.485	0.246	2	655
-0.644	0.245	4	656
-0.507	0.241	4	657
-0.867	0.242	3	658
-0.799	0.244	3	659
-0.420	0.247	4	660
-0.712	0.250	4	661
-0.920	0.252	4	662
-0.551	0.249	4	663
-0.575	0.241	4	664
-0.237	0.254	2	665
-0.752	0.244	4	666
-0.552	0.250	3	667
-0.820	0.249	4	668
-0.584	0.243	4	669
-0.626	0.244	4	670
-0.429	0.239	4	671
0.619	0.235	5	672
0.118	0.242	4	673
0.180	0.244	4	674
0.581	0.234	4	675
-0.124	0.239	2	676
0.410	0.235	4	677
0.371	0.235	4	678
0.543	0.234	3	679
0.535	0.237	4	680
0.569	0.235	5	681
0.000	0.244	4	682
0.581	0.234	4	683
0.314	0.235	4	684
0.226	0.247	4	685
0.613	0.233	5	686
0.321	0.241	4	687
-0.209	0.244	2	688

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

-0.403	0.245	4	689
0.053	0.231	2	690
-0.528	0.243	4	691
-0.147	0.235	2	692
0.000	0.234	2	693
0.036	0.231	2	694
-0.147	0.235	2	695
-0.135	0.241	2	696
-0.135	0.241	2	697
-0.470	0.239	4	698
0.182	0.234	4	699
0.223	0.236	4	700
0.091	0.233	2	701
0.260	0.236	4	702
-0.010	0.239	3	703
0.089	0.231	2	704
0.204	0.235	4	705
-0.546	0.246	4	706
-0.439	0.239	4	707
-0.775	0.244	4	708
-0.608	0.243	4	709
-0.386	0.234	4	710
-0.639	0.246	4	711
-0.502	0.239	4	712
-0.483	0.243	3	713
-0.754	0.247	4	714
-0.550	0.242	3	715
-0.620	0.241	4	716
-0.786	0.257	4	717
-0.511	0.237	4	718
-0.660	0.241	4	719
-0.774	0.248	4	720
-0.454	0.238	4	721
-0.854	0.249	3	722
-0.605	0.244	4	723

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

-0.413	0.242	4	724
-0.836	0.247	4	725
-0.429	0.247	4	726
-0.705	0.261	4	727
-0.676	0.234	3	728
-0.651	0.246	4	729
-0.435	0.240	4	730
-0.578	0.246	4	731
-0.566	0.245	4	732
-0.693	0.241	3	733
-0.641	0.248	4	734
-0.320	0.237	4	735
-0.658	0.253	3	736
-0.314	0.242	4	737
-0.658	0.245	4	738
-0.477	0.246	4	739
-0.602	0.236	4	740
-0.748	0.243	3	741
-0.586	0.242	4	742
-0.656	0.249	4	743
-0.697	0.251	4	744
-0.874	0.242	3	745
-0.715	0.248	4	746
-0.438	0.244	4	747
-0.597	0.251	2	748
-0.435	0.246	4	749
-0.632	0.244	4	750
-0.695	0.239	4	751
-0.780	0.245	4	752
-0.569	0.238	4	753
-0.367	0.234	4	754
-0.458	0.239	2	755
-0.628	0.244	4	756
-0.602	0.236	4	757
-0.646	0.246	3	758

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

-0.372	0.242	4	759
-0.836	0.250	3	760
-0.373	0.249	4	761
-0.721	0.248	4	762
-0.429	0.244	4	763
-0.703	0.247	4	764
-0.199	0.233	2	765
-0.603	0.242	4	766
-0.865	0.256	4	767
-0.432	0.237	4	768
-0.416	0.237	4	769
-0.651	0.251	2	770
-0.673	0.251	4	771
-0.660	0.239	3	772
-0.733	0.253	3	773
-0.160	0.231	3	774
-0.715	0.252	4	775
-0.850	0.256	3	776
-0.578	0.240	4	777
-0.674	0.236	3	778
-0.673	0.251	4	779
-0.555	0.248	4	780
-0.319	0.233	4	781
-0.621	0.245	4	782
-0.681	0.241	4	783
-0.624	0.248	4	784
-0.677	0.250	3	785
-0.686	0.241	4	786
-0.493	0.237	4	787
-0.531	0.244	4	788
-0.571	0.238	4	789
-0.472	0.245	3	790
-0.728	0.261	3	791
-0.256	0.234	4	792
-0.665	0.245	4	793

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

-0.783	0.254	3	794
-0.557	0.244	4	795
-0.672	0.242	4	796
-0.479	0.242	4	797
-0.645	0.250	4	798
0.336	0.242	4	799
0.427	0.242	3	800
-0.574	0.238	4	801
0.535	0.236	3	802
0.421	0.238	4	803
-0.484	0.252	3	804
0.122	0.237	3	805
-0.372	0.242	3	806
-0.248	0.239	3	807
-0.122	0.247	2	808
0.467	0.239	4	809
0.554	0.236	5	810
-0.320	0.249	2	811
-0.168	0.251	4	812
-0.956	0.254	3	813
-0.173	0.240	4	814
-0.296	0.243	3	815
-0.599	0.246	4	816
-0.217	0.243	4	817
-0.393	0.243	4	818
-0.607	0.245	4	819
0.000	0.232	2	820
-0.455	0.246	2	821
-0.441	0.248	4	822
-0.262	0.246	2	823
-0.111	0.246	2	824
-0.372	0.242	3	825
-0.910	0.245	4	826
-0.622	0.253	4	827
-0.727	0.246	4	828

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

-0.873	0.251	3	829
-0.551	0.242	4	830
-0.570	0.245	4	831
-0.817	0.246	3	832
-0.659	0.244	4	833
-0.674	0.239	4	834
-0.937	0.270	3	835
-0.388	0.234	4	836
-0.927	0.253	4	837
-0.597	0.240	4	838
-0.820	0.255	4	839
-0.720	0.247	4	840
-0.800	0.256	4	841
-0.617	0.250	3	842
-0.899	0.247	4	843
-0.659	0.244	4	844
-0.660	0.243	4	845
-0.783	0.256	4	846
-0.728	0.247	4	847
-1.026	0.249	4	848
-0.738	0.247	4	849
-0.648	0.248	4	850
-0.822	0.247	3	851
-0.761	0.245	4	852
-0.604	0.240	4	853
-0.726	0.247	4	854
-0.736	0.247	4	855
-0.639	0.251	4	856
-0.667	0.243	4	857
-0.632	0.250	4	858
-0.532	0.251	3	859
-0.441	0.242	4	860
-0.821	0.259	3	861
-0.786	0.244	3	862
-0.678	0.259	3	863

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

-0.326	0.243	4	864
-0.745	0.246	4	865
-0.361	0.245	2	866
-0.638	0.249	3	867
-0.668	0.241	4	868
-0.465	0.231	4	869
-0.821	0.240	3	870
-0.596	0.239	4	871
-0.793	0.247	4	872
-0.402	0.252	2	873
-0.913	0.254	4	874
-0.705	0.245	4	875
-0.735	0.244	4	876
-0.402	0.254	4	877
-0.770	0.248	4	878
-0.763	0.240	4	879
-0.607	0.243	4	880
-0.806	0.240	3	881
-0.831	0.253	4	882
-0.658	0.248	4	883
-0.843	0.268	3	884
-0.603	0.259	3	885
-0.979	0.249	3	886
-0.777	0.247	3	887
-0.710	0.252	2	888
-0.998	0.253	3	889
-0.523	0.236	4	890
-0.503	0.241	3	891
-0.720	0.252	3	892

NORMAL PROGRAM TERMINATION

START DATE: 03-06-2008

START TIME: 00:37:41

END TIME: 00:37:41

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ค
ตัวอย่างคำสั่งที่วิเคราะห์ด้วยโปรแกรม LISREL

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตัวอย่างคำสั่งที่วิเคราะห์ด้วยโปรแกรม LISREL

```
GROUP 1: reference
DA NG=2 NI=40 NO=892 MA=CM
LA
v1 v2 v3 v4 v5 v6 v7 v8 v9 v10 v11 v12 v13 v14 v15 v16 v17 v18 v19 v20 v21
v22 v23 v24 v25 v26 v27 v28 v29 v30 v31 v32 v33 v34 v35 v36 v37 v38 v39
v40
RA=D:\boys.txt
KM FR=D:\coboy40.txt
ME
3.99 3.54 3.41 3.13 3.48 3.38 2.95 3.41 3.50 3.30 3.53 3.03 3.25 2.99 3.25
3.34 3.32 3.23 3.14 3.73 3.51 3.18 3.09 2.79 2.97 2.98 3.08 3.03 2.98 3.14
2.08 3.19 2.91 3.25 2.91 2.80 3.02 2.98 2.77 2.90
SD
1.12 1.31 1.17 1.19 1.24 1.36 1.50 1.30 1.41 1.42 1.33 1.36 1.35 1.23 1.28
1.30 1.31 1.34 1.34 1.04 1.46 1.19 1.37 1.17 1.39 1.31 1.29 1.37 1.32 1.30
1.02 1.34 1.31 1.37 1.30 1.21 1.21 1.17 1.26 1.21
MO NX=40 NK=1 TX = FR LX = FR KA = FR TD=DI
LK
F1
FI LX (1, 1)
VA 1 LX (1, 1)
OU TV SC ND=2
```

GROUP 2: focal

DA NI=40 NO=914 MA=CM

LA

v1 v2 v3 v4 v5 v6 v7 v8 v9 v10 v11 v12 v13 v14 v15 v16 v17 v18 v19 v20 v21
v22 v23 v24 v25 v26 v27 v28 v29 v30 v31 v32 v33 v34 v35 v36 v37 v38 v39
v40

RA=D:\girls.txt

KM FR=D:\cogirl40.txt

ME

3.96 3.50 3.54 3.11 3.54 3.57 2.91 3.48 3.64 3.54 3.73 2.98 3.30 2.86 3.18
3.44 3.39 3.31 3.32 3.47 3.81 3.23 3.25 2.89 2.97 2.91 3.09 2.77 2.98 3.24
2.59 3.36 2.72 3.40 2.77 2.72 2.98 2.98 2.80 2.99

SD

1.11 1.24 1.02 1.14 1.19 1.30 1.47 1.32 1.39 1.34 1.26 1.31 1.27 1.16 1.21
1.29 1.18 1.21 1.30 1.24 1.36 1.26 1.28 1.19 1.40 1.29 1.19 1.31 1.41 1.32
1.11 1.38 1.20 1.35 1.34 1.13 1.14 1.11 1.23 1.31

MO LX=FR TX=FR KA=FI TD=DI

EQ TX (1,1) TX (1)

FI LX (1,1)

VA 1 LX (1,1)

OU

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ง

ตัวอย่างคำสั่งที่วิเคราะห์ด้วยโปรแกรม MULTILOG

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตัวอย่างคำสั่งที่วิเคราะห์ด้วยโปรแกรม MULTILOG

```
DIFOUM.MLG - MALE VS FEMALE DIFFERENTIAL ITEM FUNCTIONING
             Mathematics test
>PROBLEM     RANDOM,INDIVIDUAL, NITEMS=40, NGROUPS=2,
             NEXAMINEES=1806, NCHARS=4,
             DATA='D:\poly40.txt';
>INPUT       NFMT=1, NITEMS=40, NGROUPS=2, NEXAMINEES=1806,
             NCHARS=4;
>TEST        ALL, L1;
>EQUAL       BJ,
ITEMS=(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,
27,28,29,30,31,32,33,34,35,36,37,38,39,40),
WITH=(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,2
7,28,29,30,31,32,33,34,35,36,37,38,39,40);
>END;
0001
1
5443342423314233413335433533343133312221
(4A1,3X,1A1,6X,40A1)
```

ภาคผนวก จ
ตัวอย่างคำสั่งที่วิเคราะห์ด้วยโปรแกรม PARSCALE

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตัวอย่างคำสั่งที่วิเคราะห์ด้วยโปรแกรม PARSCALE

```
DIFOUM.PSL - DIF ANALYSIS USING GENERAL PARTIAL CREDIT MODEL
          ARTIFICAL DATA, TWO SAMPLES, 40 ITEMS
>COMMENTS

  This example illustrates differential item functioning (DIF) analysis
  of multiple category item responses.

  The DIF keyword of the MGROUP command requests DIF analysis of the
item  threshold parameters.

  The data file contains the examinee ID and sample group code (1,2),
  then the responses on the 40 items, and finally the generating trait
  value for each examinee.

>FILES   DFNAME='D:\poly.txt';
>INPUT   NID=4, MGROUP=2, NTOTAL=40
(4A1,3X,1A1,6X,40A1)
>TEST    TNAME='MATH',
ITEM=(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,2
7,28,29,30,31,32,33,34,35,36,37,38,39,40),

INAME=('I001','I002','I003','I004','I005','I006','I007','I008','I009','I010','I011','I01
2','I013','I014','I015','I016','I017','I018','I019','I020','I021','I022','I023','I024','I02
5','I026','I027','I028','I029','I030','I031','I032','I033','I034','I035','I036','I037','I03
8','I039','I040'), NBLOCK=40;
>BLOCK1  REPEAT=40, NIT=1, NCAT=6, ORIGINAL=(0,1,2,3,4,5)
```

```
>MGROUP  DIF=(0,1,0,0), GNAME=('MALE','FEMALE'), GCODE=('1','2');
>CALIB   LOGISTIC, PARTIAL, NQPT=30, CYCLES=(10,1,1,1,1,1),
NEWTON=0,
         CRIT=0.001, POSTERIOR;
>SCORE   ;
```

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ฉ

จดหมายขอความร่วมมือในการวิจัย

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ที่ ศธ ๐๕๑๒.๖ (๒๗๕๕)/

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ถนนพญาไท เขตปทุมวัน กรุงเทพฯ 10330

พฤศจิกายน 2550

เรื่อง ขอความร่วมมือในการเก็บข้อมูล

เรียน ผู้อำนวยการ โรงเรียนนนทรีวิทยา

สิ่งที่แนบมาด้วย แบบทดสอบ

ด้วย นางสาววริษา ชะม้อย นิสิตระดับปริญญามหาบัณฑิต ชั้นปีที่ 2 สาขาวิชาการวัดและประเมินผลทางการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย กำลังอยู่ระหว่างดำเนินการเสนอโครงร่างวิทยานิพนธ์เรื่อง “การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาคระหว่างวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ยกับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก” โดยมี ผู้ช่วยศาสตราจารย์ ดร.ณัฏฐภรณ์ หลาวทอง เป็นอาจารย์ที่ปรึกษาวิทยานิพนธ์ ในการทำวิทยานิพนธ์ครั้งนี้ นิสิตมีความจำเป็นต้องใช้แบบสอบวิชาคณิตศาสตร์ เรื่อง จำนวนเต็ม เศษส่วนและทศนิยม ในการเก็บข้อมูลกับนักเรียนระดับชั้นมัธยมศึกษาปีที่ 2 จำนวน 6 ห้องเรียน โดยใช้เวลาในการทำแบบสอบ 1 ชั่วโมง 30 นาที ในการนี้ผู้วิจัยขอความอนุเคราะห์มาเก็บข้อมูลภายในเดือนพฤศจิกายน โดยวัน เวลาแล้วแต่ความสะดวก

จึงเรียนมาเพื่อขอความอนุเคราะห์ขอเก็บข้อมูลดังกล่าว เพื่อใช้ในการวิเคราะห์ข้อมูลในงานวิจัยซึ่งจะเป็นประโยชน์ในทางวิชาการต่อไป จึงขอขอบพระคุณล่วงหน้ามา ณ โอกาสนี้ด้วย

ขอแสดงความนับถือ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

(รองศาสตราจารย์ ดร.ศิริเดช สุชีวะ)
หัวหน้าภาควิชาวิจัยและจิตวิทยาการศึกษา

ภาควิชาวิจัยและจิตวิทยาการศึกษา
๐-๘๖๕๗-๖๓๙๓๓ (นิสิต)
โทรสาร ๐-๒๒๑๘-๒๕๗๘

ที่ ศธ ๐๕๑๒.๖ (๒๗๕๕) /

ภาควิชาวิจัยและจิตวิทยาทางการศึกษา
คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ถนนพญาไท เขตปทุมวัน กรุงเทพฯ ๑๐๓๓๐

๑๑ มกราคม ๒๕๕๑

เรื่อง ขอขอบพระคุณในการอนุเคราะห์ให้เก็บข้อมูล

เรียน อาจารย์สุภา เหร็นเส็บ

เนื่องด้วย นางสาววริษชา ชะม้อย นิสิตระดับปริญญามหาบัณฑิต ชั้นปีที่ ๒ สาขาวิชาการวัดและประเมินผลทางการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย กำลังดำเนินการทำวิทยานิพนธ์เรื่อง “ การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาคระหว่างวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ยกับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก” โดยมี ผู้ช่วยศาสตราจารย์ ดร. ณัฏฐภรณ์ หลาวทอง เป็นอาจารย์ที่ปรึกษาวิทยานิพนธ์ ในการทำวิทยานิพนธ์ครั้งนี้ นิสิตได้ใช้แบบสอบวิชาคณิตศาสตร์ เรื่อง จำนวนเต็ม เศษส่วนและทศนิยม ในการเก็บข้อมูลกับนักเรียนระดับชั้นมัธยมศึกษาปีที่ ๒ โดยได้รับความอนุเคราะห์จากอาจารย์สุภา เหร็นเส็บ อาจารย์ประจำกลุ่มสาระการเรียนรู้คณิตศาสตร์ โรงเรียนสายปัญญา ในพระบรมราชินูปถัมภ์ เป็นผู้อนุเคราะห์ให้มาเก็บข้อมูลในคาบวิชาคณิตศาสตร์ ของนักเรียนชั้นมัธยมศึกษาปีที่ ๒/๒ ในวันที่ ๒๑ ธันวาคม ๒๕๕๐ เวลา ๐๙.๐๐ – ๐๙.๕๐ น. ตลอดจนช่วยอำนวยความสะดวกให้ผู้วิจัยเป็นอย่างดี

จึงเรียนมาเพื่อขอขอบพระคุณในความกรุณา และให้ความช่วยเหลือในงานวิจัยดังกล่าวซึ่งจะเป็นประโยชน์ในทางวิชาการต่อไป จึงขอขอบพระคุณอีกครั้งมา ณ โอกาสนี้ด้วย

ขอแสดงความนับถือ

(รองศาสตราจารย์ ดร.ศิริเดช สุชีวะ)
หัวหน้าภาควิชาวิจัยและจิตวิทยาการศึกษา

ภาควิชาวิจัยและจิตวิทยาการศึกษา
๐-๘๖๕๗-๖๓๙๓๓ (นิสิต)
โทรสาร ๐-๒๒๑๘-๒๕๗๘

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ที่ ศธ ๐๕๑๒.๖ (๒๗๕๕) / ว.

ภาควิชาวิจัยและจิตวิทยาทางการศึกษา
คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ถนนพญาไท เขตปทุมวัน กรุงเทพฯ ๑๐๓๓๐

ธันวาคม ๒๕๕๐

เรื่อง ขออนุญาตใช้เครื่องมือวิจัย

เรียน คณบดี บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร

ด้วย นางสาววริษษา ชะม้อย นิสิตระดับปริญญาโท ชั้นปีที่ ๒ สาขาวิชาการวัดและประเมินผลทางการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย กำลังอยู่ระหว่างดำเนินการเสนอโครงร่างวิทยานิพนธ์เรื่อง “การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาคระหว่างวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ยกับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก” โดยมี ผู้ช่วยศาสตราจารย์ ดร.ณัฏฐภรณ์ หลาวทอง เป็นอาจารย์ที่ปรึกษาวิทยานิพนธ์ ในการทำวิทยานิพนธ์ครั้งนี้ นิสิตมีความจำเป็นต้องใช้แบบสอบเพื่อเก็บข้อมูลในการวิจัย โดยใช้แบบทดสอบวัดผลสัมฤทธิ์วิชาคณิตศาสตร์ เรื่อง จำนวนเต็ม เศษส่วนและทศนิยม ที่มีลักษณะเป็นแบบทดสอบเลือกตอบที่มีคำตอบถูกหลายคำตอบ ซึ่งเป็นแบบสอบที่ สุนิสา จุ้ยม่วงศรี สร้างขึ้นในการทำปริญญานิพนธ์ระดับปริญญาเอก คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ เรื่อง “การศึกษาผลของการเทียบคะแนนข้อสอบที่ตรวจให้คะแนนแบบหลายค่า” ปีการศึกษา ๒๕๔๖

จึงเรียนมาเพื่อขออนุญาตใช้เครื่องมือวิจัยดังกล่าวซึ่งจะเป็นประโยชน์ในทางวิชาการต่อไป จึงขอขอบพระคุณล่วงหน้ามา ณ โอกาสนี้ด้วย

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ขอแสดงความนับถือ

(รองศาสตราจารย์ ดร.ศิริเดช สุชีวะ)
หัวหน้าภาควิชาวิจัยและจิตวิทยาการศึกษา

ภาควิชาวิจัยและจิตวิทยาการศึกษา
๐-๘๖๕๗-๖๓๙๓๓ (นิสิต)
โทรสาร ๐-๒๒๑๘-๒๕๗๘

ประวัติผู้เขียนวิทยานิพนธ์

นางสาววริษษา ชะม้อย เกิดเมื่อวันอังคาร ที่ 3 เดือนกรกฎาคม พ.ศ. 2527 ที่จังหวัดสิงห์บุรี สำเร็จการศึกษาระดับปริญญาบัณฑิต ในปีการศึกษา 2548 สาขาวิชามัธยมศึกษา-วิทยาศาสตร์ วิชาเอกเคมี-วิทยาศาสตร์ เกียรตินิยมอันดับสอง ภาควิชาหลักสูตรและเทคโนโลยีทางการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย และในปีการศึกษา 2549 ได้เข้าศึกษาต่อในระดับปริญญามหาบัณฑิต สาขาวิชาการวัดและประเมินผลทางการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย