การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบด้วยกระบวนการ ดี เอฟ ไอ ที สำหรับแบบสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษา วิชาภาษาอังกฤษและวิชาคณิตศาสตร์

นางสาวรักชนก ยี่สุ่นศรี

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาครุศาสตรมหาบัณฑิต
สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยการศึกษา
คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2544

ISBN 974-030-568-7

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

AN ANALYSIS OF DIFFERENTIAL FUNCTIONING OF ITEMS AND TESTS BASED ON DFIT PROCEDURES IN ENGLISH AND MATHEMATICS FOR UNIVERSITY ENTRANCE EXAMINATION

Miss Rakchanok Yeesunesri

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Education in Educational Measurement and Evaluation

Department of Educational Research
Faculty of Education
Chulalongkorn University

Academic Year 2001
ISBN 974-030-568-7

หัวข้อวิทยานิพนธ์	การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบด้วยกระบวนการ ดี เอฟ ไอ ที สำหรับแบบสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษา วิชาภาษาอังกฤษและวิชาคณิตศาสตร์
โดย	นางสาวรักชนก ยี่สุ่นศรี
สาขาวิชา	การวัดและประเมินผลการศึกษา
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ดร. ศิริชัย กาญจนวาสี

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

..... คณบดีคณะครุศาสตร์
(รองศาสตราจารย์ ดร. ไพฑูรย์ สินลารัตน์)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(รองศาสตราจารย์ พวงแก้ว ปุณยกนก)

..... อาจารย์ที่ปรึกษา
(รองศาสตราจารย์ ดร. ศิริชัย กาญจนวาสี)

..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร. สุชาดา บวรกิติวงศ์)

..... กรรมการ
(รองศาสตราจารย์ ดร. ศิริเดช สุชีวะ)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รักชนก ยี่สุ่นศรี : การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบด้วยกระบวนการ ดี เอฟ ไอ ที สำหรับแบบสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษา วิชาภาษาอังกฤษและวิชาคณิตศาสตร์ (AN ANALYSIS OF DIFFERENTIAL FUNCTIONING OF ITEMS AND TESTS BASED ON DFIT PROCEDURES IN ENGLISH AND MATHEMATICS FOR UNIVERSITY ENTRANCE EXAMINATION) อาจารย์ที่ปรึกษา : รองศาสตราจารย์ ดร. ศิริชัย กาญจนวาสี , 260 หน้า
ISBN 974-030-568-7

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) และแบบสอบ (DTF) วิชาภาษาอังกฤษและวิชาคณิตศาสตร์ ด้วยกระบวนการ ดี เอฟ ไอ ที สำหรับกลุ่มผู้สอบเมื่อจำแนกตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา เพื่อเปรียบเทียบค่าความเที่ยง ความตรง และค่าฟังก์ชันสารสนเทศของแบบสอบ ระหว่างแบบสอบฉบับก่อนและหลังตัดข้อสอบที่พบ DIF และเพื่อวิเคราะห์ความสัมพันธ์ระหว่างตำแหน่งของคะแนนรวมของผู้สอบก่อนและหลังตัดข้อสอบที่พบ DIF โดยใช้ข้อมูลการตอบข้อสอบจากแบบสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาของทบวงมหาวิทยาลัย ปีการศึกษา 2543 ครั้งที่ 1 วิชาภาษาอังกฤษและวิชาคณิตศาสตร์ โดยทำการศึกษาในส่วนที่เป็นข้อสอบแบบหลายตัวเลือก จากกลุ่มตัวอย่างจำนวน 4,000 คน และ 3,600 คน ตามลำดับ

ผลการวิจัยพบว่า

1. แบบสอบวิชาภาษาอังกฤษทำหน้าที่ต่างกันตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษาของผู้สอบ ส่วนแบบสอบวิชาคณิตศาสตร์ทำหน้าที่ต่างกันตามเพศของผู้สอบ และผลการวิเคราะห์การทำหน้าที่ต่างกันที่ระดับข้อสอบ พบว่าข้อสอบทำหน้าที่ต่างกันตามเพศของผู้สอบมากที่สุด สอดคล้องกันทั้งสองวิชา

2. ผลการเปรียบเทียบคุณภาพของแบบสอบฉบับก่อนและหลังตัดข้อสอบที่พบ DIF ทั้งในกรณีตัดทุกข้อและบางข้อทั้ง 2 วิชาได้ผลสอดคล้องกันดังนี้

2.1 แบบสอบฉบับก่อนและหลังตัดข้อสอบมีค่าความตรงเชิงโครงสร้างของแบบสอบ ไม่แตกต่างกัน

2.2 แบบสอบฉบับหลังตัดข้อสอบที่พบ DIF ส่วนใหญ่มีค่าความเที่ยงลดลง

2.3 แบบสอบฉบับหลังตัดข้อสอบที่ DIF ส่วนใหญ่มีค่าฟังก์ชันสารสนเทศของแบบสอบมีมากขึ้น

3. ผลการวิเคราะห์ความสัมพันธ์ระหว่างตำแหน่งของคะแนนรวมของผู้สอบก่อนและหลังตัดข้อสอบที่พบ DIF ทั้งในกรณีตัดทุกข้อและบางข้อ พบว่าในทุกกรณีมีความสัมพันธ์ในทางบวกซึ่งกันและกันอย่างมีนัยสำคัญ แบบสอบวิชาภาษาอังกฤษมีค่าสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนแรงค์อยู่ระหว่าง 0.83 – 0.99 ส่วนในแบบสอบวิชาคณิตศาสตร์มีค่าสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนแรงค์อยู่ระหว่าง 0.87 – 0.98

ภาควิชา วิจัยการศึกษา **ลายมือชื่อนิสิต** .....
สาขาวิชา การวัดและประเมินผลการศึกษา **ลายมืออาจารย์ที่ปรึกษา** .....
ปีการศึกษา 2544 **ลายมืออาจารย์ที่ปรึกษาร่วม** —

##4283790027: MAJOR EDUCATIONAL MEASUREMENT AND EVALUATION
KEY WORD: DIFFERENTIAL FUNCTIONING OF ITEMS AND TESTS / DIFFERENTIAL ITEM FUNCTIONING / DIFFERENTIAL TEST FUNCTIONING / ENTRANCE EXAMINATION
RAKCHANOK YEESUNESRI: AN ANALYSIS OF DIFFERENTIAL FUNCTIONING OF ITEMS AND TESTS BASED ON DFIT PROCEDURES IN ENGLISH AND MATHEMATICS FOR UNIVERSITY ENTRANCE EXAMINATION. THESIS ADVISORS: ASSOC. PROF. SIRICHAI KANJANAWASI, Ph.D. 260 pp. ISBN 974-030-568-7.

This research has three main purposes. First, to analyze the differential item functioning (DIF) and differential test functioning (DTF) based on DFIT procedure for 2 subgroups ; gender and location of school. Second, to compare the discrepancy between the quality of test before and after DIF items are eliminated, in terms of reliability, validity and test information function. The last objective is to analyze the relationship between the rank of examinees‘ total score of the original test and the test after the DIF items are eliminated. The data are the examinees' item response to English and Mathematics university entrance examination tests in the part of multiple choice items, administered by the Ministry of the University Affairs in the first examination of the year 2000. The sample of study was 4,000 and 3,600 examinees in English and Mathematics tests, respectively.

The major findings are as follows:

1. English test is differential test functioning according to examinees‘ gender and school location while Mathematics test is differential test functioning according to examinees‘ gender subgroup. Detecting differential functioning at items level according to gender subgroup is the largest number of DIF items in both English and Mathematics tests.

2. The results from comparing the discrepancy between the quality of test before and after DIF items are eliminated, are as follows:

2.1 The construct validity of the test is not significantly different from that of the original test.

2.2 The reliability of the test is slightly lower than that of the original test.

2.3 The test information function has no difference by the test that after DIF items are eliminated is slightly greater than that of the original test.

3. The results from analyzing the relationship between the rank of examinee's total score before and after the DIF items are eliminated, have significant positive relationships. The Spearman rank correlation coefficient is in the range of 0.83 – 0.99 and 0.87 – 0.98 for the English and Mathematics tests, respectively.

Department Educational Research Student' s signature .....

Field of Study Educational Measurement and Evaluation Advisor' s signature .....

Academic Year 2001 Co-Advisor' s signature .....

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

กิตติกรรมประกาศ

วิทยานิพนธ์เล่มนี้สำเร็จได้ด้วยความช่วยเหลือจากบุคคลหลายท่าน โดยเฉพาะบุคคลแรกที่ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูง ณ ที่นี้ คือ รองศาสตราจารย์ ดร. ศิริชัย กาญจนวาสี อาจารย์ที่ปรึกษาวิทยานิพนธ์ ผู้ได้เสียสละเวลาอันมีค่ากรุณาให้คำปรึกษา แนะนำหัวข้อ ชี้แนะแนวทางที่สำคัญประกอบการเขียนวิทยานิพนธ์ แก้ไขข้อบกพร่องต่าง ๆ และให้กำลังใจแก่ผู้วิจัยอย่างดียิ่งตลอดมา จนงานวิจัยสำเร็จลุล่วงด้วยดี

ขอกราบขอบพระคุณ รองศาสตราจารย์ พวงแก้ว ปุณยกนก อาจารย์ที่ปรึกษาทางวิชาการและอาจารย์สอนวิชาสัมมนาการวิจัย ที่ให้ข้อเสนอแนะในการพัฒนาหัวข้อวิทยานิพนธ์และให้คำปรึกษาที่เป็นประโยชน์ อีกทั้งยังให้ความเอาใจใส่แก่ผู้วิจัยตลอดมา นอกจากนี้ขอกราบขอบพระคุณครู อาจารย์ทุกท่านที่ประสิทธิ์ประสาทวิชาความรู้แก่ผู้วิจัย จนผู้วิจัยสามารถในการทำวิทยานิพนธ์ฉบับนี้จนสำเร็จได้

ขอกราบขอบพระคุณ คุณวราภรณ์ สีหนาถ ผู้อำนวยสำนักทดสอบกลาง ทบวงมหาวิทยาลัย และเจ้าหน้าที่ของสำนักทดสอบกลางทุกท่าน ที่กรุณาอนุญาตให้ใช้ข้อมูลและช่วยเก็บรวบรวมข้อมูล เพื่อนำมาใช้ในการวิจัยครั้งนี้ ขอกราบขอบพระคุณบัณฑิตวิทยาลัย ที่กรุณาให้ทุนอุดหนุนการวิจัยบางส่วน และขอกราบขอบพระคุณ ดร. รังสรรค์ มณีเล็ก ที่กรุณาอนุเคราะห์โปรแกรมคอมพิวเตอร์ BILOG 3.04 และ EQUATE 2.0 อีกทั้งยังช่วยสอนการใช้โปรแกรม ให้คำแนะนำและคำปรึกษาอันเป็นประโยชน์แก่ผู้วิจัยมาโดยตลอด และขอกราบขอบพระคุณ อาจารย์สุพจน์ เกิดสุวรรณ์ ที่ช่วยแก้ไขปัญหาที่เกิดขึ้นในการใช้โปรแกรม BILOG

ขอขอบคุณพี่และเพื่อน ๆ ภาควิชาวิจัยการศึกษาทุกคน โดยเฉพาะพี่วรรณา เด่นขจรเกียรติ ที่ให้ความช่วยเหลือทุกครั้งที่มีปัญหา พี่กนกวรรณ รัตนธน พี่จุฬาลักษณ์ ขันธบุตร พี่ทิพวรรณ ประเสริฐอำไพสกุล พี่สุทิศา นามเหลา และคุณประทีป ปิ่นทอง ที่ให้กำลังใจและคำปรึกษาที่เป็นประโยชน์ในทุกเรื่อง อีกทั้งยังให้ความช่วยเหลือด้วยความเต็มใจมาโดยตลอด

ท้ายสุดนี้ผู้วิจัยขอกราบขอบพระคุณ พ่อ แม่ และพี่ชาย ที่ให้ความรัก ความห่วงใยตลอดมา และให้การสนับสนุน ส่งเสริมให้ผู้วิจัยได้มีโอกาสได้รับการศึกษามาจนถึงระดับนี้

รักชนก ยี่สุ่นศรี

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง
บทคัดย่อภาษาอังกฤษ ..... จ
กิตติกรรมประกาศ ..... ฉ
สารบัญตาราง ..... ฌ
สารบัญแผนภาพ ..... ญ
บทที่
1 บทนำ 1
ความเป็นมาและความสำคัญของปัญหา 1
วัตถุประสงค์ของการวิจัย 9
ขอบเขตของการวิจัย 9
ข้อจำกัดของการวิจัย 9
คำจำกัดความของการวิจัย 10
ประโยชน์ที่คาดว่าจะได้รับ 11
2 วรรณคดีที่เกี่ยวข้อง 12
ตอนที่ 1 ความเป็นมาของการสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษา 12
ตอนที่ 2 การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ 17
ตอนที่ 3 วิธีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบด้วย
กระบวนการ ดี เอฟ ไอ ที 20
ตอนที่ 4 เอกสารและงานวิจัยที่เกี่ยวข้อง 35
3 วิธีดำเนินการวิจัย 48
ประชากรและกลุ่มตัวอย่าง 48
เครื่องมือที่ใช้ในการวิจัย 49
การเก็บรวบรวมข้อมูล 49
การวิเคราะห์ข้อมูล 50
4 ผลการวิเคราะห์ข้อมูล 62
ตอนที่ 1 ผลการวิเคราะห์ข้อมูลเบื้องต้น 62
ตอนที่ 2 ผลการวิเคราะห์คุณภาพของแบบสอบก่อนการวิเคราะห์การวิเคราะห์
การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ 64

สารบัญ (ต่อ)

บทที่ หน้า

ตอนที่ 3 ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบด้วยกระบวนการ ดี เอฟ ไอ ที 92

ตอนที่ 4 ผลการเปรียบเทียบคุณภาพของแบบสอบก่อนและหลังการตัดข้อสอบที่ทำหน้าที่ต่างกัน 110

ตอนที่ 5 ผลการวิเคราะห์ความสัมพันธ์ระหว่างตำแหน่งของคะแนนรวมของผู้สอบก่อนและหลังตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบสอบ 193

5 สรุปผลการวิจัย อภิปรายผลและข้อเสนอแนะ 195

สรุปผลการวิจัย 196

อภิปรายผลการวิจัย 199

ข้อเสนอแนะ 203

รายการอ้างอิง 205

ภาคผนวก 211

ประวัติผู้เขียนวิทยานิพนธ์ 257

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญตาราง

ตารางที่ หน้า

1 สรุปงานวิจัยที่ได้ศึกษาเกี่ยวกับการตรวจสอบการทำหน้าที่ต่างกัน ของข้อสอบและแบบสอบ 4

2 เปรียบเทียบวิธีต่าง ๆ ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 19

3 ฟังก์ชันทางคณิตศาสตร์ของโมเดลการตอบสนองข้อสอบ 23

4 ลักษณะการแจกแจงของคะแนนสอบวิชาภาษาอังกฤษและวิชาคณิตศาสตร์ 63

5 ค่าไอเกนและร้อยละของความแปรปรวนของตัวประกอบวิชาภาษาอังกฤษ 64

6 ค่าไอเกนและร้อยละของความแปรปรวนของตัวประกอบวิชาคณิตศาสตร์ 65

7 ค่าความยาก อำนาจจำแนก และค่าความเที่ยงของแบบสอบวิชาภาษาอังกฤษ 67

8 ค่าความยาก อำนาจจำแนก และค่าความเที่ยงของแบบสอบวิชาคณิตศาสตร์ 70

9 ลักษณะของข้อสอบวิชาภาษาอังกฤษจำแนกตามโครงสร้างเนื้อหา 72

10 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับข้อเท็จจริง (FACT) ของแบบสอบวิชาภาษาอังกฤษ 74

11 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับตีความ (INTERP) ของแบบสอบวิชาภาษาอังกฤษ 77

12 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับประเมินค่า (EVAL) ของแบบสอบวิชาภาษาอังกฤษ 79

13 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความสามารถในการใช้ภาษา (LANG) ของแบบสอบวิชาภาษาอังกฤษ 82

14 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ (GRAM) ของแบบสอบวิชาภาษาอังกฤษ 85

15 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลองค์ประกอบความรู้วิชาภาษาอังกฤษ (ENGLISH) 87

16 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความรู้วิชาคณิตศาสตร์ (MATH) 90

17 ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามเพศ 92

18 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบ จำแนกตามเพศ เมื่อตัดข้อสอบข้อที่ 1 ออกจากแบบสอบ 94

สารบัญตาราง (ต่อ)

ตารางที่ หน้า

19 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามเพศ เมื่อตัดข้อสอบข้อที่ 1, 33, 9, 91, 17, 68, 79, 41, 47, 90, 46, 66 และ 20 ออกจากแบบสอบ 96

20 ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา 98

21 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา เมื่อตัดข้อสอบข้อที่ 33 ออกจากแบบสอบ 100

22 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา เมื่อตัดข้อสอบข้อที่ 33, 24, 37, 68, 1, 73 และ 66 ออกจากแบบสอบ 102

23 เปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบวิชาภาษาอังกฤษ เมื่อจำแนกผู้สอบตามตัวแปรเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา 104

24 ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบวิชาคณิตศาสตร์ สำหรับกลุ่มผู้สอบจำแนกตามเพศ 105

25 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาคณิตศาสตร์ สำหรับกลุ่มผู้สอบจำแนกตามเพศ เมื่อตัดข้อสอบข้อที่ 19 ออกจากแบบสอบ 106

26 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาคณิตศาสตร์ สำหรับกลุ่มผู้สอบจำแนกตามเพศ เมื่อตัดข้อสอบข้อที่ 19 และข้อที่ 25 ออกจากแบบสอบ 107

27 ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบวิชาคณิตศาสตร์ สำหรับกลุ่มผู้สอบจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา 108

28 เปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบวิชาคณิตศาสตร์ เมื่อจำแนกผู้สอบตามตัวแปรเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา 109

29 ผลการเปรียบเทียบค่าความเที่ยงของแบบสอบวิชาภาษาอังกฤษก่อนและหลังการตัดข้อสอบที่พบ DIF จากตัวแปรเพศ และสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา 111

สารบัญตาราง (ต่อ)

ตารางที่ หน้า

30 ผลการเปรียบเทียบค่าความเที่ยงของแบบสอบวิชาคณิตศาสตร์ก่อนและหลังการตัดข้อสอบที่พบ DIF จากตัวแปรเพศ และสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา 112

31 จำนวนข้อสอบที่ตัดออกจากแบบสอบวิชาภาษาอังกฤษ จำแนกตามโครงสร้างเนื้อหา ตัวแปรที่ใช้แบ่งกลุ่มผู้สอบและดัชนีที่ใช้ในการตัดข้อสอบออกจากแบบสอบ 113

32 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่านระดับข้อเท็จจริง (FACT) ของแบบสอบวิชาภาษาอังกฤษ หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ 115

33 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่านระดับตีความ (INTERP) ของแบบสอบวิชาภาษาอังกฤษ หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ 118

34 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่านระดับประเมินค่า (EVAL) ของแบบสอบวิชาภาษาอังกฤษ หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ 120

35 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความสามารถในการใช้ภาษา (LANG) ของแบบสอบวิชาภาษาอังกฤษ หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ 122

36 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ (GRAM) ของแบบสอบวิชาภาษาอังกฤษหลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ 125

37 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลองค์ประกอบความรู้วิชาภาษาอังกฤษ (ENGLISH) หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ 127

38 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่านระดับข้อเท็จจริง (FACT) ของแบบสอบวิชาภาษาอังกฤษ หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ 130

39 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่านระดับตีความ (INTERP) ของแบบสอบวิชาภาษาอังกฤษ หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ 133

สารบัญตาราง (ต่อ)

ตารางที่ หน้า

40 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับประเมินค่า (EVAL) ของแบบสอบวิชาภาษาอังกฤษ หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ 135

41 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความสามารถในการใช้ภาษา (LANG) ของแบบสอบวิชาภาษาอังกฤษ หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ 138

42 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ (GRAM) ของแบบสอบวิชาภาษาอังกฤษหลังการวิเคราะห์DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ 141

43 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลองค์ประกอบความรู้วิชาภาษาอังกฤษ (ENGLISH) หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ 143

44 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับข้อเท็จจริง (FACT) ของแบบสอบวิชาภาษาอังกฤษ หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา 146

45 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับตีความ (INTERP) ของแบบสอบวิชาภาษาอังกฤษ หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา 149

46 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับประเมินค่า (EVAL) ของแบบสอบวิชาภาษาอังกฤษ หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา 152

47 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความสามารถในการใช้ภาษา (LANG) ของแบบสอบวิชาภาษาอังกฤษ หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา 154

48 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ (GRAM) ของแบบสอบวิชาภาษาอังกฤษหลังการวิเคราะห์DIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา 157

49 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลองค์ประกอบความรู้วิชาภาษาอังกฤษ (ENGLISH) หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา 159

สารบัญตาราง (ต่อ)

ตารางที่ หน้า

50 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่านระดับตีความ (INTERP) ของแบบสอบวิชาภาษาอังกฤษ หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา 162

51 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความสามารถในการใช้ภาษา (LANG) ของแบบสอบวิชาภาษาอังกฤษ หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา 165

52 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ (GRAM) ของแบบสอบวิชาภาษาอังกฤษหลังการวิเคราะห์DTF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา 168

53 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลองค์ประกอบความรู้วิชาภาษาอังกฤษ (ENGLISH) หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา 170

54 ผลการเปรียบเทียบดัชนีความตรงจากการวิเคราะห์องค์ประกอบของแบบสอบวิชาภาษาอังกฤษก่อนและหลังการตัดข้อสอบที่พบ DIF จากตัวแปรเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา 172

55 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความรู้วิชาคณิตศาสตร์ (MATH) หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ 174

56 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความรู้วิชาคณิตศาสตร์ (MATH) หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ 177

57 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความรู้วิชาคณิตศาสตร์ (MATH) หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา 180

58 ผลการเปรียบเทียบดัชนีความตรงจากการวิเคราะห์องค์ประกอบของแบบสอบวิชาคณิตศาสตร์ ก่อนและหลังการตัดข้อสอบที่พบ DIF จากตัวแปรเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา 182

59 ค่าฟังก์ชันสารสนเทศของแบบสอบ (TIF) และค่าดัชนีประสิทธิภาพสัมพัทธ์ (RE) ของแบบสอบวิชาภาษาอังกฤษ ก่อนและหลังการตัดข้อสอบที่พบ DIF โดยใช้ดัชนี NCDIF และดัชนี CDIF/DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ ที่ระดับความสามารถต่าง ๆ 183

สารบัญตาราง (ต่อ)

ตารางที่ หน้า

60 ค่าฟังก์ชันสารสนเทศของแบบสอบ (TIF) และค่าดัชนีประสิทธิภาพสัมพัทธ์ (RE) ของแบบสอบวิชาภาษาอังกฤษ ก่อนและหลังการตัดข้อสอบที่พบ DIF โดยใช้ดัชนี NCDIF และดัชนี CDIF/DTF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา ที่ระดับความสามารถต่าง ๆ 186

61 ค่าฟังก์ชันสารสนเทศของแบบสอบ (TIF) และค่าดัชนีประสิทธิภาพสัมพัทธ์ (RE) ของแบบสอบวิชาคณิตศาสตร์ ก่อนและหลังการตัดข้อสอบที่พบ DIF โดยใช้ดัชนี NCDIF และดัชนี CDIF/DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ ที่ระดับความสามารถต่าง ๆ 189

62 ค่าฟังก์ชันสารสนเทศของแบบสอบ (TIF) และค่าดัชนีประสิทธิภาพสัมพัทธ์ (RE) ของแบบสอบวิชาคณิตศาสตร์ ก่อนและหลังการตัดข้อสอบที่พบ DIF โดยใช้ดัชนี NCDIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา ที่ระดับความสามารถต่าง ๆ 191

63 ผลการวิเคราะห์ความสัมพันธ์ระหว่างตำแหน่งของคะแนนรวมของผู้สอบก่อนและหลังการตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบสอบวิชาภาษาอังกฤษ 193

64 ผลการวิเคราะห์ความสัมพันธ์ระหว่างตำแหน่งของคะแนนรวมของผู้สอบก่อนและหลังการตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบสอบวิชาคณิตศาสตร์ 194

65 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามเพศ เมื่อตัดข้อสอบข้อที่ 1 และ 33 ออกจากแบบสอบ 230

66 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามเพศ เมื่อตัดข้อสอบข้อที่ 1 33 และ 9 ออกจากแบบสอบ 231

67 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามเพศ เมื่อตัดข้อสอบข้อที่ 1, 33, 9 และ 91 ออกจากแบบสอบ 233

68 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามเพศ เมื่อตัดข้อสอบข้อที่ 1, 33, 9, 91 และ 17 ออกจากแบบสอบ 235

69 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามเพศ เมื่อตัดข้อสอบข้อที่ 1, 33, 9, 91,17 และ 68 ออกจากแบบสอบ 236

70 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามเพศ เมื่อตัดข้อสอบข้อที่ 1, 33, 9, 91, 17, 68 และ 79 ออกจากแบบสอบ ... 238

71 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามเพศ เมื่อตัดข้อสอบข้อที่ 1, 33, 9, 91, 17, 68, 79 และ 41 ออกจากแบบสอบ 239

สารบัญตาราง (ต่อ)

ตารางที่ หน้า

72 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามเพศ เมื่อตัดข้อสอบข้อที่ 1, 33, 9, 91, 17, 68, 79, 41 และ 47 ออกจากแบบสอบ 241

73 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามเพศ เมื่อตัดข้อสอบข้อที่ 1, 33, 9, 91, 17, 68, 79, 41, 47 และ 90 ออกจากแบบสอบ 243

74 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามเพศ เมื่อตัดข้อสอบข้อที่ 1, 33, 9, 91, 17, 68, 79, 41, 47, 90 และ 46 ออกจากแบบสอบ 244

75 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามเพศ เมื่อตัดข้อสอบข้อที่ 1, 33, 9, 91, 17, 68, 79, 41, 47, 90, 46 และ 66 ออกจากแบบสอบ 246

76 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา เมื่อตัดข้อสอบข้อที่ 33 และข้อที่ 24 ออกจากแบบสอบ 247

77 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา เมื่อตัดข้อสอบข้อที่ 33 , 24 และ 37 ออกจากแบบสอบ 249

78 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา เมื่อตัดข้อสอบข้อที่ 33 , 24 ,37 และ 68 ออกจากแบบสอบ 251

79 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา เมื่อตัดข้อสอบข้อที่ 33 , 24,37 ,68 และ 1 ออกจากแบบสอบ 252

80 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา เมื่อตัดข้อสอบข้อที่ 33, 24, 37, 68, 1 และ 73 ออกจากแบบสอบ 254

สารบัญแผนภาพ

แผนภาพที่ หน้า

1 โค้งฟังก์ชันสารสนเทศของข้อสอบของโมเดลโลจิส แบบ 1 พารามิเตอร์ 4 ข้อ 24
2 โค้งฟังก์ชันสารสนเทศของข้อสอบของโมเดลโลจิส แบบ 2 พารามิเตอร์ 4 ข้อ 24
3 โค้งฟังก์ชันสารสนเทศของข้อสอบของโมเดลโลจิส แบบ 3 พารามิเตอร์ 6 ข้อ 25
4 ขั้นตอนการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบ (DTF) 57
5 ค่าไอเกนของตัวประกอบที่ได้จากการสกัดตัวประกอบในแบบสอบวิชาภาษาอังกฤษ 65
6 ค่าไอเกนของตัวประกอบที่ได้จากการสกัดตัวประกอบในแบบสอบวิชาคณิตศาสตร์ 66
7 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับข้อเท็จจริง 75
8 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับตีความ 78
9 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับประเมินค่า 80
10 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความความสามารถ ในการใช้ภาษา 83
11 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความรู้เกี่ยวกับ โครงสร้างไวยากรณ์ 86
12 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลองค์ประกอบความรู้วิชาภาษาอังกฤษ 88
13 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลความรู้วิชาคณิตศาสตร์ 91
14 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันโมเดลด้านความเข้าใจในการอ่าน ระดับข้อเท็จจริง หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มตามเพศ 116
15 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับตีความ หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ 119
16 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับประเมินค่า หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ 121
17 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความสามารถในการใช้ภาษา หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ 123
18 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความรู้เกี่ยวกับ โครงสร้างไวยากรณ์ หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ 126

สารบัญแผนภาพ (ต่อ)

แผนภาพที่ หน้า

19 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลองค์ประกอบความรู้วิชาภาษาอังกฤษ หลังจากการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ 128

20 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับข้อเท็จจริง หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ 131

21 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับตีความ หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ 134

22 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับประเมินค่า หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ 136

23 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความสามารถในการใช้ภาษา หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ 139

24 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลความรู้เกี่ยวกับโครงสร้างไวยากรณ์ หลังจากการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ 142

25 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลองค์ประกอบความรู้วิชาภาษาอังกฤษ หลังจากการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามตัวแปรเพศ 144

26 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับข้อเท็จจริง หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา 147

27 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับตีความ หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา 150

28 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับประเมินค่า หลังการวิเคราะห์ DIF มื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา 152

29 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความสามารถในการใช้ภาษา หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา 155

30 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลความรู้เกี่ยวกับโครงสร้างไวยากรณ์ หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา 158

สารบัญแผนภาพ (ต่อ)

แผนภาพที่ หน้า

31 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลองค์ประกอบความรู้วิชาภาษาอังกฤษ หลังจากการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามตัวแปรสถานที่ตั้งของโรงเรียนที่จบการศึกษา 160

32 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่านระดับตีความ หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา 163

33 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความสามารถในการใช้ภาษา หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา 166

34 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลความรู้เกี่ยวกับโครงสร้างไวยากรณ์ หลังจากการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา 169

35 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลองค์ประกอบความรู้วิชาภาษาอังกฤษ หลังจากการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา 171

36 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันโมเดลความรู้วิชาคณิตศาสตร์ หลังการวิเคราะห์ DIFเมื่อแบ่งกลุ่มผู้สอบตามเพศ 175

37 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันโมเดลความรู้วิชาคณิตศาสตร์ หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ 178

38 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันโมเดลความรู้วิชาคณิตศาสตร์ หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา 181

39 โค้งฟังก์ชันสารสนเทศของแบบสอบวิชาภาษาอังกฤษ ฉบับก่อนและหลังการตัดข้อสอบที่พบ DIF ทุกข้อ โดยใช้ดัชนี NCDIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ 184

40 โค้งฟังก์ชันสารสนเทศของแบบสอบวิชาภาษาอังกฤษ ฉบับก่อนและหลังการตัดข้อสอบที่พบ DIF บางข้อ โดยใช้ดัชนี CDIF/DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ 185

41 โค้งฟังก์ชันสารสนเทศของแบบสอบวิชาภาษาอังกฤษ ฉบับก่อนและหลังการตัดข้อสอบที่พบ DIF ทุกข้อ โดยใช้ดัชนี NCDIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา 187

สารบัญแผนภาพ (ต่อ)

แผนภาพที่ หน้า

42 โค้งฟังก์ชันสารสนเทศของแบบสอบวิชาภาษาอังกฤษ ฉบับก่อนและหลังการตัดข้อสอบที่พบ DIF บางข้อ โดยใช้ดัชนี CDIF/DTF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา 187

43 โค้งฟังก์ชันสารสนเทศของแบบสอบวิชาคณิตศาสตร์ ฉบับก่อนและหลังการตัดข้อสอบที่พบ DIF ทุกข้อ โดยใช้ดัชนี NCDIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ 189

44 โค้งฟังก์ชันสารสนเทศของแบบสอบวิชาคณิตศาสตร์ ฉบับก่อนและหลังการตัดข้อสอบที่พบ DIF บางข้อ โดยใช้ดัชนี CDIF/DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ 190

45 โค้งฟังก์ชันสารสนเทศของแบบสอบวิชาคณิตศาสตร์ ฉบับก่อนและหลังการตัดข้อสอบที่พบ DIF ทุกข้อ โดยใช้ดัชนี NCDIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา 193

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

การทดสอบเป็นวิธีการวัดผลการศึกษาวิธีหนึ่งที่นิยมใช้กันมากที่สุด ซึ่งเครื่องมือที่ใช้ในการทดสอบที่สำคัญ ก็คือแบบสอบต่างๆ ในการวัดผลและประเมินผลทางการศึกษามีจุดประสงค์เพื่อตรวจสอบว่าผู้สอบมีความรู้หรือมีคุณลักษณะที่ต้องการวัดอยู่ในระดับใดซึ่งผลที่ได้จากการวัดนี้มีความสำคัญต่อการพัฒนาคุณภาพของการศึกษา ดังนั้นในการพัฒนาแบบสอบใดขึ้นมา จะต้องคำนึงถึงคุณภาพของแบบสอบเป็นสำคัญ แบบสอบนั้นจะต้องมีความเชื่อถือได้ สามารถวัดได้ตรงตามความสามารถที่ต้องการด้วยเหตุที่เครื่องมือที่ใช้ในการทดสอบมีความสำคัญเช่นเดียวกับวิธีการวัดผล เพราะถ้าเครื่องมือที่ใช้มีคุณภาพก็ทำให้ได้ข้อมูลหรือคะแนนที่ถูกต้อง คุณสมบัติที่สำคัญมากอีกข้อหนึ่งของเครื่องมือที่ใช้สำหรับวัดคือต้องมีความยุติธรรมกับผู้สอบทุกกลุ่ม ไม่เข้าข้างกลุ่มใดกลุ่มหนึ่งหรือทำให้กลุ่มใดกลุ่มหนึ่งเสียประโยชน์ ข้อสอบที่เอื้อต่อผู้สอบกลุ่มใดกลุ่มหนึ่งอาจทำให้ผู้สอบกลุ่มดังกล่าวมีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องกว่ากลุ่มอื่นที่ระดับความสามารถเดียวกัน ลักษณะการทำหน้าที่ของข้อสอบประเภทนี้เรียกว่า “ การทำหน้าที่ต่างกันของข้อสอบ (differential item functioning : DIF) “ ซึ่งเดิมเรียกว่าความลำเอียงของข้อสอบ (item bias) (Holland and Wainer , 1993) และแบบสอบที่ทำให้ผลการตอบของผู้สอบที่มีความสามารถเท่ากันในสิ่งที่ต้องการวัด ได้คะแนนจริงไม่เท่ากัน เนื่องจากผู้สอบอยู่ในกลุ่มย่อยต่างกัน เรียกลักษณะการทำหน้าที่ของแบบสอบประเภทนี้ว่า “การทำหน้าที่ต่างกันของแบบสอบ (differential test functioning : DTF) “ ตามลำดับ

การศึกษาเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบยังคงได้รับความสนใจจากนักวัดผลการศึกษา เห็นได้จากมีการศึกษาค้นคว้าเกี่ยวกับวิธีการตรวจสอบการทำหน้าที่ต่างกันจำนวนมาก มีผู้ศึกษาค้นคว้าและเสนอวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบหลายวิธีที่รู้จักกันทั่วไป เช่น วิธีหาพื้นที่ระหว่างโค้งของฟังก์ชันการตอบสนองข้อสอบ 2 ฟังก์ชัน (item response function : IRFs) วิธีทดสอบค่าไคสแคว์ของลอร์ด (Lord's χ^2 test) และวิธีซิบเทสส์ (sibtest) เป็นต้น ซึ่งเป็นวิธีที่อยู่บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ ส่วนวิธีที่ไม่อยู่บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ เช่น วิธีแมนเทล-เฮนส์เซล (Mantel-Haenszel technique) วิธีวิเคราะห์ความแปรปรวน (ANOVA) และ

วิธีกำหนดจุดค่าเดลต้า (delta method) เป็นต้น ซึ่งวิธีการเหล่านี้จะทำการตรวจสอบการทำหน้าที่ต่างกันที่ระดับข้อสอบเท่านั้น

Park และ Lautenschlager (1990 , อ้างถึงใน กาญจนา วัธนสุนทร , 2537) ได้ตั้งข้อสังเกตเกี่ยวกับข้อสอบที่ลำเอียงหรือทำหน้าที่ต่างกันอยู่ 2 ประการ ได้แก่

1. การตอบข้อสอบมีอิทธิพลจากแหล่งความแปรปรวนอื่น นอกเหนือจากความแตกต่างบนมิติที่สนใจอยู่ 2 แหล่งได้แก่ แหล่งความแปรปรวนจากข้อสอบและแหล่งความแปรปรวนจากผู้สอบ

2. แหล่งความแปรปรวนภายนอกนี้มีผลต่อการตอบข้อสอบในลักษณะที่ทำให้กลุ่มผู้สอบย่อยบางกลุ่มแตกต่างจากกลุ่มอื่นอย่างเป็นระบบ ทำให้เกิดการได้เปรียบเสียเปรียบกันอย่างไม่ยุติธรรม

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ผ่านมา จะพิจารณาการทำหน้าที่ต่างกันที่เกิดขึ้นระหว่างกลุ่มประชากรกลุ่มย่อย ที่มีเพศต่างกันเช่น งานวิจัยของ Doolittle และ Cleary (1987) Raju , Drasgow และ Slinde (1993) สุรศักดิ์ อมรรัตนศักดิ์ (2531) สุพัฒน์ สุกมลสันต์ (2534) และกาญจนา วัธนสุนทร (2537) เป็นต้น ที่มีสีผิวต่างกัน เช่น งานวิจัยของ Subkoviak , Mack และ Craig (1984) Ryan (1991) และ Raju , Drasgow และ Slinde (1993) เป็นต้น ที่ได้รับการสอนด้วยวิธีต่างกัน เช่น งานวิจัยของ Tatsuoka (1988 อ้างถึงใน กาญจนา วัธนสุนทร , 2537) หรือกลุ่มย่อยที่มีภูมิลำเนาต่างกัน เช่น ทัศนีย์ พีรมนตรี (2530) สุพัฒน์ สุกมลสันต์ (2534) และเกษรหว่างจิตร (2539) เป็นต้น นอกจากนี้ยังมีการแบ่งกลุ่มย่อยตามวัฒนธรรม เชื้อชาติ สภาพเศรษฐกิจและสังคม ศาสนา ภาษา อุปกรณ์ที่ใช้ในการสอบ สังกัดสถานศึกษา และประสบการณ์ในการสอบเป็นต้น

โดยทั่วไปเมื่อผู้พัฒนาแบบสอบตรวจพบว่าข้อสอบข้อใดทำหน้าที่ต่างกันแล้ว ก็อาจจะตัดข้อสอบข้อนั้นออกจากแบบสอบหรือนำไปปรับปรุง แล้วนำกลับไปใส่รวมกับแบบสอบฉบับนั้นเช่นเดิม โดยมีวัตถุประสงค์เพื่อให้แบบสอบฉบับนั้นมีความยุติธรรมกับกลุ่มผู้สอบย่อยทุกกลุ่มเท่าเทียมกัน แต่เมื่อมีการตรวจสอบการทำหน้าที่ต่างกันของแบบสอบทั้งฉบับ พบว่าแบบสอบดังกล่าวยังไม่เหมาะที่จะนำไปใช้ได้ (Shealy and Stout , 1993) ดังนั้นจึงได้มีการค้นหาวิธีที่เหมาะสมกับการตรวจสอบการทำหน้าที่ต่างกันของแบบสอบทั้งฉบับ ซึ่งสามารถที่จะตรวจสอบผลของการคัดข้อสอบออกหรือเพิ่มข้อสอบที่มีลักษณะของการทำหน้าที่ต่างกันเข้าไปว่ามีผลต่อการทำหน้าที่ต่างกันของแบบสอบทั้งฉบับ ในขั้นตอนสุดท้ายอย่างไร (Raju , et al., 1995) นอกจากนี้การตรวจสอบการทำหน้าที่ต่างกันของแบบสอบทั้งฉบับยังมีประโยชน์ในการใช้เป็นสารสนเทศในการตัดสินประสิทธิภาพของเครื่องมือ ใช้ในการจัดลำดับหรือใช้ในการคัดเลือก ในความเป็นจริงเป็นการยากที่จะตัดข้อสอบทุกข้อที่ทำหน้าที่ต่างกันออกจากแบบ

สอบ ถ้าในกรณีที่ผู้พัฒนาแบบสอบไม่สามารถตัดข้อสอบที่ตรวจพบว่าทำหน้าที่ต่างกันออกจากแบบสอบได้ สาเหตุอาจมาจากการตัดข้อสอบนั้นออกอาจทำให้แบบสอบขาดความตรงตามเนื้อหา หรืออาจเป็นเพราะเหตุผลอื่นและจำเป็นจะต้องรวมข้อสอบดังกล่าวเข้าไปในแบบสอบ ซึ่งข้อสอบที่ทำหน้าที่ต่างกันบางข้อเข้าข้างผู้สอบกลุ่มอ้างอิง บางข้อเข้าข้างผู้สอบกลุ่มเปรียบเทียบ ดัชนีการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบทั้งฉบับจะมีประโยชน์ในการใช้พิจารณาถึงดุลยภาพระหว่างข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบที่เข้าข้างกลุ่มใดกลุ่มหนึ่ง (Raju, et al., 1995) และเราไม่จำเป็นต้องตัดข้อสอบที่ทำหน้าที่ต่างกันทุกข้อออกจากแบบสอบ เนื่องจากว่าข้อสอบที่ลำเอียงไม่จำเป็นต้องเป็นข้อสอบที่ไม่ดีเสมอไป เพราะในบางสถานการณ์ข้อสอบที่ลำเอียง อาจให้ข้อสารสนเทศที่ดีที่สุด (most information) หรือเป็นตัวทำนายที่ดีที่สุด (best predictor) (Rudner et al, 1984 อ้างถึงใน สุพัฒน์ สุกมลสันต์ , 2534)

สำหรับงานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันของแบบสอบทั้งฉบับ ยังพบไม่มาก วิธีการตรวจสอบที่พบในปัจจุบันมี 2 วิธี คือ วิธีซิบเทสส์ (sibtest) ที่เสนอโดย Shealy และ Stout (1993) ซึ่งเป็นวิธีการที่มีแนวคิดบนการตรวจสอบการทำหน้าที่ต่างกันของแบบสอบชนิดพหุมิติ โดยมีพื้นฐานบนทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ แตกต่างจาก วิธี IRT ตรงที่เป็นการใช้การทดสอบค่าสถิติแบบนอนพารามิเตอร์ (non-parametric) มีการคำนวณง่ายไม่ซับซ้อน ใช้ได้ดีสำหรับการตรวจสอบการทำหน้าที่ต่างกันแบบสม่ำเสมอและมีทิศทางเดียว (unidirectional) เท่านั้น ข้อตกลงเบื้องต้นของการวิเคราะห์คือมี มิติการวัด 2 มิติ มิติหนึ่งเป็นความสามารถหรือลักษณะแฝงเป้าหมายที่ต้องการวัด วิธีซิบเทสส์ นี้มีประสิทธิภาพในการตรวจสอบ การทำหน้าที่ต่างกันได้ดีทั้งในกรณีที่แบบสอบมีข้อสอบลำเอียงข้อเดียวหรือมีข้อสอบที่ลำเอียงหลายข้อ

ต่อมา Raju และคณะ (1995) ได้เสนอกระบวนการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ (differential functioning of items and tests framework : DFIT) ซึ่งเป็นวิธีการที่อยู่บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบเป็นวิธีที่ละเอียด มีการคำนวณซับซ้อนแต่ให้ผลถูกต้อง วิธีการนี้เป็นวิธีที่ไม่ใช้เกณฑ์ภายนอก (absence of criterion) โดยการนำวิธีทางสถิติมาตรวจสอบการทำหน้าที่ต่างกันโดยพิจารณาโครงสร้างภายในแบบสอบ

กระบวนการ DFIT ประกอบด้วยดัชนีที่ใช้วัดการทำหน้าที่ต่างกัน 2 ระดับ ดังนี้

1. ดัชนีการทำหน้าที่ต่างกันระดับแบบสอบ (DTF index) เป็นดัชนีที่บอกขนาดของผลการทำหน้าที่ต่างกันสุทธิ (net effect) ในแบบสอบทั้งฉบับ
2. ดัชนีการทำหน้าที่ต่างกันระดับข้อสอบ (DIF index) มี 2 ดัชนี คือ
   2.1 ดัชนีการทำหน้าที่ต่างกันของข้อสอบแบบชดเชย (compensatory DIF index :

CDIF) ซึ่งเป็นดัชนีที่ใช้ในการหาค่าดัชนี DTF ของแบบสอบทั้งฉบับ ข้อแตกต่างของดัชนีนี้กับวิธีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบวิธีอื่นที่รู้จักกันดี กล่าวคือ ในการวิเคราะห์ดัชนี CDIF สำหรับข้อสอบข้อใดข้อหนึ่ง จะไม่สมมติว่าข้อสอบข้ออื่นในแบบสอบจะไม่มีความลำเอียง นั่นคือพิจารณาความลำเอียงของข้ออื่นร่วมด้วยกับข้อที่พิจารณาด้วย

2.2 ดัชนีการทำหน้าที่ต่างกันของข้อสอบแบบไม่ชดเชย (noncompensatory DIF index : NCDIF) ในการวิเคราะห์ดัชนี NCDIF จะสมมติว่าข้อสอบข้ออื่นในแบบสอบที่ไม่ได้พิจารณาไม่มีความลำเอียงอย่างสมบูรณ์ นั่นคือสนใจเฉพาะความลำเอียงที่เกิดขึ้นของข้อที่พิจารณาเท่านั้น

กระบวนการ DFIT นี้เป็นวิธีที่ได้รับความสนใจในการตรวจสอบการทำหน้าที่ต่างกันที่ระดับแบบสอบเนื่องจากต่อมาได้มีงานวิจัยอีกหลายเรื่องที่ทำการศึกษาเกี่ยวกับประสิทธิภาพในการวิเคราะห์ของกระบวนการ DFIT ซึ่งสามารถสรุปประเด็นที่ศึกษาและผลการศึกษา ได้ดังตารางที่ 1

ตารางที่ 1 สรุปงานวิจัยที่ได้ศึกษาเกี่ยวกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ

ผู้ศึกษา	ประเด็นที่ศึกษา	ผลการศึกษา
Raju, et al. (1995)	ศึกษาวิธีการตรวจสอบ DIF และ DTF ด้วยกระบวนการ DFIT โดยเปรียบเทียบดัชนี CDIF , NCDIF กับ Lord's χ^2 test , SA และ UA ใช้ข้อมูลจำลอง	เมื่อกลุ่มตัวอย่างขนาดเล็กและจำนวนข้อสอบที่ DIF สูง ดัชนี NCDIF มีประสิทธิภาพดีกว่าดัชนีอื่น แต่ดัชนี CDIF มีประสิทธิภาพน้อยที่สุด เมื่อกลุ่มตัวอย่างมีขนาดใหญ่และมีจำนวนข้อสอบที่ DIF สูง ดัชนี SA มีประสิทธิภาพน้อยที่สุด ส่วนดัชนีอื่นมีประสิทธิภาพใกล้เคียงกัน
Flowers, et al. (1997)	ศึกษากระบวนการ Polytomous DFIT และเปรียบเทียบประสิทธิภาพกับวิธี SIBTEST และ Lord's χ^2 test ใช้ข้อมูลจำลอง	โดยภาพรวม DFIT มีประสิทธิภาพดี Type I error ขึ้นอยู่กับจำนวนข้อที่ DIF ขนาด DIF และค่าความยากของข้อสอบ
Oshima, et al. (1997)	ศึกษากระบวนการ DFIT สำหรับตรวจสอบแบบสอบชนิดพหุมิติ (Multidimensional) โดยใช้ข้อมูลจำลองจากโมเดลโลจิสติกแบบ	ในทุกเงื่อนไข ดัชนี CDIF/DTF และดัชนี NCDIF สามารถระบุข้อสอบที่ DIF และ ไม่ DIF ได้ถูกต้อง นอก

	2 พารามิเตอร์ (M2PL) และข้อมูลเป็นแบบ 2 มิติ องค์ประกอบที่สนใจศึกษาคือ รูปแบบการทำหน้าที่ต่างกัน, ทิศทางการทำหน้าที่ต่างกัน และการกระจายของค่าความสามารถแตกต่างกันระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ	จากนี้ในเงื่อนไขที่ข้อสอบ DIF แบบอเนกรูป พบว่าเมื่อค่าความยากของทั้ง 2 มิติแตกต่างกันจะทำให้ค่า CDIF และ NCDIF มีค่ามากขึ้น

ตารางที่ 1 (ต่อ)

ผู้ศึกษา	ประเด็นที่ศึกษา	ผลการศึกษา
Oshima, et al. (1998)	ศึกษาวิธีตรวจสอบการทำหน้าที่ต่างกันของกลุ่มข้อสอบ (Differential Bundle Functioning : DBF) เมื่อแบ่งกลุ่มข้อสอบต่างกัน โดยใช้วิธีการที่มีพื้นฐานมาจากกระบวนการ DFIT ของ Raju, et al. (1995) ใช้ข้อมูลเชิงประจักษ์จากแบบสอบการอ่านถอดความ	เมื่อแบ่งกลุ่มข้อสอบตามบทความพบว่ากลุ่มข้อสอบกลุ่มที่ 5 มีค่าดัชนี bundle NCDIF สูงที่สุด เมื่อแบ่งกลุ่มผู้สอบตามเพศ แต่เมื่อแบ่งกลุ่มผู้สอบตามสภาพเศรษฐสังคม ค่าดัชนีดังกล่าวมีค่าไม่แตกต่างกัน ในทุก ๆ กลุ่มข้อสอบ
Price (1999)	เปรียบเทียบเทคนิคกระบวนการ DFIT กับวิธี MH ในการตรวจสอบ DIF ของแบบสอบแปลความ ใช้ข้อมูลเชิงประจักษ์ โดยเปรียบเทียบในด้านจำนวนข้อสอบที่ตรวจพบ DIF และความสอดคล้องของทั้งสองวิธี	กระบวนการ DFIT มีความไวในการตรวจสอบ DIF มากกว่า วิธี MH ทั้งสองวิธีระบุข้อสอบที่ DIF ตรงกันเพียงร้อยละ 20
Flowers et al. (1999)	ศึกษากระบวนการ DFIT สำหรับแบบสอบชนิดเอกมิติ (unidimensional) ที่ข้อสอบให้คะแบบแบบหลายค่า (polytomous) โดยใช้ข้อมูลจำลอง องค์ประกอบที่ศึกษาคือความยาวของแบบสอบ การกระจายค่าความสามารถ จำนวนข้อสอบที่ DIF และชนิดของ DIF	กระบวนการ Polytomous DFIT มีประสิทธิภาพในการตรวจสอบ DIF และ DTF องค์ประกอบที่ศึกษาทุกองค์ประกอบ ในทุกเงื่อนไข พบว่ามีผลกระทบเพียงเล็กน้อยต่ออัตราความถูกต้องและผิดพลาดในการตรวจสอบ DIF และ DTF
Ellis and Mead (2000)	ได้ทำการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ สำหรับแบบสอบ	ดัชนี CDIF/DTF ตรวจพบข้อสอบที่ DIF และควรตัดออกจากแบบสอบ

	มาตรฐาน 16PF ที่แปลเป็นภาษาสเปน ด้วยกระบวนการ DFIT โดยแบ่งกลุ่มผู้สอบตามเชื้อชาติ 3 กลุ่ม	น้อยกว่าดัชนี NCDIF ในทุกกรณี และเมื่อวิเคราะห์ DIF ระหว่างผู้สอบกลุ่ม Anglo และ Spanish พบข้อสอบที่ DIF มากที่สุด

จากตารางที่ 1 สามารถสรุปได้ว่า กระบวนการ DFIT ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยภาพรวมจะมีประสิทธิภาพสูงกว่า วิธีซิบเทสส์ และวิธี Lord's χ^2 test อีกทั้งกระบวนการ DFIT ยังมีความสามารถในการตรวจสอบการทำหน้าที่ต่างกันรวมทั้งฉบับอีกด้วย จากตารางจะเห็นได้ว่ากระบวนการ DFIT ได้รับการพัฒนาเรื่อยมาตั้งแต่เริ่มจาก Raju และคณะ (1995) ได้เสนอกระบวนการนี้ในการตรวจสอบข้อสอบที่มีการให้คะแนนแบบ 2 ค่า (dichotomous) และ แบบสอบเป็นแบบเอกมิติ (unidimensional) ต่อมา Oshima และคณะ (1997) ได้พัฒนากระบวนการนี้ขึ้นให้สามารถตรวจสอบแบบสอบชนิดพหุมิติ (multidimensional) และข้อสอบมีการให้คะแนนแบบ 2 ค่า หลังจากนั้น Flowers และคณะ (1999) ได้พัฒนากระบวนการนี้ให้สามารถตรวจสอบแบบสอบชนิดเอกมิติ ที่มีการให้คะแนนแบบหลายค่า (polytomous) จะเห็นได้ว่ากระบวนการ DFIT นี้ได้รับการพัฒนาให้มีประสิทธิภาพในการตรวจสอบแบบสอบได้หลายแบบ ดังนั้นกระบวนการ DFIT นี้จึงเป็นทางเลือกที่น่าสนใจอีกทางเลือกหนึ่งที่ใช้ในการตรวจสอบคุณภาพของข้อสอบและแบบสอบ

ในแต่ละปีมีนักเรียนชั้นมัธยมศึกษาตอนปลายที่สำเร็จการศึกษาและต้องการเข้าศึกษาต่อในสถาบันอุดมศึกษาทั้งของรัฐและเอกชนเป็นจำนวนประมาณ 200,000 คน และจะมีเพียงประมาณร้อยละ 25 เท่านั้นที่สามารถสอบเข้าศึกษาในสถาบันอุดมศึกษาจำกัดจำนวนรับได้ ในการสอบคัดเลือกดังกล่าวนี้แบบสอบเป็นเครื่องมือการตัดสินความรู้ความสามารถของผู้สอบที่มีบทบาทสำคัญที่สุด สถาบันอุดมศึกษาในระบบจำกัดจำนวนรับของประเทศไทยได้ใช้การสอบคัดเลือกบุคคลเข้าศึกษา 3 วิธี คือ การรับตรงหรือการรับนักศึกษาตามระบบโควต้าของมหาวิทยาลัยส่วนภูมิภาค การรับนักศึกษาตามโครงการพิเศษต่าง ๆ ที่มหาวิทยาลัยกำหนดขึ้น และการสอบคัดเลือกรวม การสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาของรัฐระบบใหม่ ซึ่งเริ่มในปี พ.ศ. 2542 สำนักทดสอบกลาง ทบวงมหาวิทยาลัยได้จัดการสอบคัดเลือกรวมขึ้นปีละ 2 ครั้ง ผลการสอบวัดความรู้พื้นฐานทางวิชาการหรือวิชาหลักและวิชาเฉพาะเป็นองค์ประกอบหนึ่งของการคัดเลือกที่มีน้ำหนักของคะแนนมากที่สุดคือ ร้อยละ 90 จำนวนวิชาหลักที่จัดสอบทั้งสิ้น 12 วิชา ได้แก่ วิชาคณิตศาสตร์ 1 คณิตศาสตร์ 2 เคมี ฟิสิกส์ ชีววิทยา วิทยาศาสตร์กายภาพชีวภาพ ภาษาไทย ภาษาอังกฤษ สังคมวิทยา ภาษาฝรั่งเศส ภาษาเยอรมัน และภาษา

บาลี แบบสอบที่ใช้สำหรับคัดเลือกบุคคลเพื่อเข้าศึกษาในระดับอุดมศึกษานี้เป็นเครื่องมือการวัดผลและตัดสินความรู้ความสามารถของผู้สอบที่มีความสำคัญมาก ผลหรือคะแนนที่ได้จากเครื่องมือดัง กล่าวจะต้องสามารถทำนายผลการเรียนในระดับอุดมศึกษาได้อย่างแม่นยำ ผลการสอบเป็นเสมือนเครื่องชี้และตัดสินอนาคตของผู้สอบ กลุ่มผู้สมัครสอบเป็นกลุ่มใหญ่มีจำนวนมากและมีความแตกต่างกัน เช่นในด้านฐานะทางเศรษฐกิจและสังคม วัฒนธรรม เป็นต้น การที่ผู้สอบไม่สามารถสอบผ่านข้อทดสอบวิชาต่าง ๆ ได้นั้นส่วนหนึ่งอาจเกิดมาจากการที่ผู้สอบนั้นมีความรู้ความสามารถไม่เพียงพอเมื่อเทียบกับผู้ที่สอบได้ และปัจจัยอีกส่วนหนึ่งก็คือเกิดจากคุณภาพของแบบสอบที่นำมาใช้ ดังนั้นแบบสอบจำเป็นจะต้องเป็นแบบสอบที่มีคุณภาพสูง มีความตรงตามจุดมุ่งหมายที่ต้องการวัด ผู้สอบที่มีความสามารถเท่ากันแต่อยู่ต่างกลุ่มกันมีโอกาสในการตอบข้อสอบได้ถูกต้องเท่ากัน ซึ่งลักษณะดังกล่าวเป็นลักษณะของข้อสอบหรือแบบสอบที่มีความยุติธรรม ถ้าแบบสอบขาดความยุติธรรมอาจทำให้ผลการวัดที่ได้นั้นจะเบี่ยงเบนไปจากความเป็นจริง

จากการวิเคราะห์คุณภาพของแบบสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษา วิชาภาษาอังกฤษ ชุด กข และ กขค ในปี พ.ศ. 2533 ของ Bradford (1990 , อ้างถึงใน สุพัฒน์ สุกมลสันต์ , 2534) ซึ่งเป็นผู้สอนภาษาอังกฤษในมหาวิทยาลัยของไทยมานาน พบว่า ข้อสอบชุดดังกล่าวมีข้อบกพร่องเกี่ยวกับคำตอบที่ถูกต้องรวมแล้วประมาณร้อยละ 20 และในปี 2534 สุพัฒน์ สุกมลสันต์ ได้ทำการวิเคราะห์คุณภาพของแบบสอบวิชาภาษาอังกฤษเข้ามหาวิทยาลัย ปี พ.ศ. 2531-2533 ในด้านความลำเอียงด้วยวิธีที่ต่างกัน พบว่าแบบสอบมีความลำเอียงต่อเพศโดยเฉลี่ยประมาณร้อยละ 4-41 มีแนวโน้มว่าลำเอียงเข้าข้างเพศหญิงมากกว่าเพศชายและมีความลำเอียงต่อภาคภูมิศาสตร์ของผู้สอบโดยเฉลี่ยประมาณร้อยละ 5-45 โดยมีความลำเอียงต่อผู้สอบภาคอื่นมากกว่าภาคกลาง และกาญจนา วัธนสุนทร (2537) ได้ทำการวิเคราะห์ความลำเอียงของข้อสอบจากแบบสอบวิชาภาษาอังกฤษเข้ามหาวิทยาลัย ปีการศึกษา 2535 เพื่อพัฒนาเกณฑ์การตัดสินความลำเอียงทางเพศพบว่า ที่ขนาดกลุ่มตัวอย่าง 1,000 คนและแบบสอบที่สุ่มมา 80 ข้อจาก 100 ข้อ พบข้อสอบที่ลำเอียงซึ่งตรวจสอบด้วยวิธี IRT ถึง 26 ข้อ (32.50 %) จะเห็นได้ว่าแบบสอบดังกล่าวที่ผ่านมายังมีข้อบกพร่องอยู่ ส่วนแบบสอบวิชาคณิตศาสตร์นั้นจากผลการศึกษาที่ผ่านมา พบว่ามักจะลำเอียงเข้าข้างผู้สอบเพศชายมากกว่าเพศหญิง เนื่องจากว่าผู้ชายมีทักษะในเชิงมิติ การใช้เหตุผลและทักษะการประยุกต์มากกว่าเพศหญิง ดังเช่นผลการวิจัยของ Doolittle และ Clearly (1987) พบว่าข้อสอบวิชาเรขาคณิตจะลำเอียงเข้าข้างเพศชายเนื่องจากว่าเพศชายจะมีทักษะการใช้เหตุผลดีกว่าเพศหญิง ส่วนเพศหญิงจะมีทักษะด้านการคำนวณมากกว่า นอกจากนั้น Harnisch (1991 อ้างถึงใน กาญจนา วัธนสุนทร , 2537) พบว่าข้อสอบที่ใช้ทักษะการคำนวณจะ

เข้าข้างเพศหญิงและข้อสอบที่ใช้การประยุกต์จะเข้าข้างเพศชาย และผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบคัดเลือกบุคคลเข้าศึกษาระดับอุดมศึกษา ปีการศึกษา 2535 วิชาคณิตศาสตร์พบข้อสอบที่ทำหน้าที่ต่างกัน 19 ข้อ จาก 40 ข้อ คิดเป็นร้อยละ 47.50 (กาญจนา วัธนสุนทร , 2537) ซึ่งเป็นสัดส่วนที่ค่อนข้างสูง

ดังนั้นจากข้อค้นพบที่ผ่านมาจึงควรมีการตรวจสอบคุณภาพของแบบสอบในด้านการทำหน้าที่ต่างกันของข้อสอบและแบบสอบสำหรับแบบสอบคัดเลือกบุคคลเข้าศึกษาระดับอุดมศึกษา วิชาภาษาอังกฤษและวิชาคณิตศาสตร์ ซึ่งมีความสำคัญต่อคุณภาพของแบบสอบและผลการสอบของผู้สอบแต่ละคน เพื่อเป็นแนวทางในการพัฒนาคุณภาพของแบบสอบต่อไป

ส่วนตัวแปรที่ผู้วิจัยเลือกมาพิจารณาคือตัวแปรเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่ผู้สอบจบการศึกษาเนื่องจากเป็นตัวแปรที่คาดว่ามีผลในบริบทของประเทศไทย ก็เพราะมีผลการวิจัยหลายเรื่องที่พบว่า แบบสอบที่ใช้ทำหน้าที่ต่างกันต่อกลุ่มเพศและภูมิศาสตร์ของผู้สอบ ดังเช่น ผลการวิจัยของ ชัชชัย เผ่าพงษ์ (2526 อ้างถึงใน ทัศนีย์ พีรมนตรี , 2530) ซึ่งทำการศึกษาความลำเอียงของข้อสอบวัดความถนัดทางคณิตศาสตร์และภาษา ของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ทั่วประเทศ พบว่าข้อสอบบางข้อมีความลำเอียงต่อเพศใดเพศหนึ่งโดยเฉพาะ สุรศักดิ์ อมรรัตนศักดิ์ (2531) ทำการศึกษาความลำเอียงของแบบสอบแข่งขันเพื่อบรรจุเข้ารับราชการครู ระหว่างผู้สอบเพศชายและเพศหญิง ทัศนีย์ พีรมนตรี (2530) ทำการศึกษาความลำเอียงระหว่างกลุ่มนักเรียนในกรุงเทพมหานครกับกลุ่มนักเรียนในภาคภูมิศาสตร์ทั้ง 5 ภาคของแบบสอบวิชาคณิตศาสตร์ของโครงการตรวจสอบคุณภาพการศึกษา นักเรียนชั้นมัธยมศึกษาปีที่ 6 พบว่าข้อสอบมีความลำเอียงต่อนักเรียนในภาคอื่นเมื่อเทียบกับกลุ่มนักเรียนในกรุงเทพมหานคร อีกทั้งตัวแปรเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษาของผู้สอบเป็นตัวแปรที่สามารถแบ่งแยกความแตกต่างได้เด่นชัดและง่ายต่อการนำผลการวิจัยที่ได้ไปประกอบการพิจารณาเพื่อพัฒนาคุณภาพแบบสอบ

ด้วยเหตุผลนี้ผู้วิจัยจึงมีความสนใจที่จะศึกษาถึงวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ ด้วยวิธีที่จะนำมาศึกษาคือกระบวนการ DFIT และแบบสอบที่ผู้วิจัยสนใจนำมาศึกษาคือแบบสอบคัดเลือกบุคคลเข้าศึกษาในระดับอุดมศึกษา วิชาภาษาอังกฤษและวิชาคณิตศาสตร์ ตัวแปรเพศ และตัวแปรสถานที่ตั้งทางภูมิศาสตร์โรงเรียนของของผู้สอบ เป็นตัวแปรสำหรับจำแนกกลุ่มผู้สอบที่นำมาศึกษา โดยมุ่งศึกษาในด้านศึกษาค่าความเที่ยง ความตรงของแบบสอบและค่าฟังก์ชันสารสนเทศของแบบสอบ รวมทั้งศึกษาความสัมพันธ์ระหว่างตำแหน่งของคะแนนรวมของผู้สอบ ทั้งก่อนและหลังจากที่ตัดข้อสอบที่ทำหน้าที่ต่างกันทุกข้อหรือเฉพาะบางข้อออกจากแบบสอบ ว่ามีความสอดคล้องกันหรือไม่ เพียงใด สำหรับตัวแปรต่าง ๆ และผลที่ได้จากการวิจัยสามารถนำกลับไปเป็นสารสนเทศใน

การปรับปรุงแบบสอบให้มี คุณภาพยิ่งขึ้นไป อีกทั้งยังสามารถใช้เป็นแนวทางในการพัฒนาและนำเทคนิคการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ (DFIT Procedure) มาใช้ในการตรวจสอบคุณภาพของแบบสอบอื่น ๆ ให้มีความยุติธรรม และเป็นที่น่าเชื่อถือต่อไป ตลอดจนเพื่อให้ผลการสอบในครั้งต่อไปได้รับการพิจารณาอย่างรอบคอบมากขึ้น

วัตถุประสงค์ของการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์ ดังต่อไปนี้

1. เพื่อวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบสำหรับแบบสอบคัดเลือกบุคคลเข้าศึกษาในระดับอุดมศึกษาวิชาภาษาอังกฤษและวิชาคณิตศาสตร์ สำหรับกลุ่มผู้สอบจำแนกตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

2. เพื่อเปรียบเทียบค่าความเที่ยง ความตรง และค่าฟังก์ชันสารสนเทศของแบบสอบ สำหรับแบบสอบก่อนและหลังจากตัดข้อสอบที่ทำหน้าที่ต่างกันในแต่ละกลุ่มผู้สอบออกจากแบบสอบ โดยเปรียบเทียบระหว่าง

2.1 แบบสอบก่อนและหลังจากตัดข้อสอบทุกข้อที่ทำหน้าที่ต่างกันออกจากแบบสอบ

2.2 แบบสอบก่อนและหลังจากตัดข้อสอบบางข้อที่ทำหน้าที่ต่างกันออกจากแบบสอบจนแบบสอบทั้งฉบับไม่ทำหน้าที่ต่างกัน

3. เพื่อวิเคราะห์ความสัมพันธ์ระหว่างตำแหน่งของคะแนนรวมของผู้สอบก่อนและหลังการตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบสอบ

ขอบเขตของการวิจัย

1. ข้อมูลที่นำมาใช้เป็นข้อมูลที่ได้จากการตอบแบบสอบวิชาภาษาอังกฤษและวิชาคณิตศาสตร์ 1 ของแบบสอบคัดเลือกบุคคลเข้าศึกษาในระดับอุดมศึกษา ปี 2543 ครั้งที่ 1/มีนาคม

2. การวิจัยครั้งนี้เป็นการศึกษาในบริบทของประเทศไทย โดยมุ่งศึกษาในกลุ่มตัวแปรเพศ และสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน โดยไม่ศึกษาในตัวแปรอื่น เช่น ศาสนา เชื้อชาติ

ข้อจำกัดของการวิจัย

1. กลุ่มผู้สอบที่ใช้ในการศึกษาครั้งนี้ เป็นกลุ่มตัวอย่างที่ได้จากการสุ่มจากผู้สมัครสอบเข้าศึกษาต่อในสถาบันอุดมศึกษารับจำกัดจำนวน ซึ่งไม่ได้นำกลุ่มผู้สมัครสอบทั้งหมดมาศึกษา

2. ข้อสอบที่นำมาวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบในครั้งนี้ จะนำมาศึกษาเฉพาะข้อสอบแบบหลายตัวเลือก (multiple choice)

3. ผู้วิจัยไม่มีแบบสอบและคำเฉลย (keys) ของข้อสอบแต่ละข้อ ที่เป็นทางการใช้ ผู้วิจัยได้แบบสอบที่นำมาวิเคราะห์ครั้งนี้ จากหนังสือเฉลยข้อสอบเข้ามหาวิทยาลัยหลายสำนักพิมพ์และได้ตรวจสอบแล้วว่ามีข้อความตรงกัน

คำจำกัดความที่ใช้ในการวิจัย

ผู้สอบ หมายถึง ผู้สมัครสอบเข้าศึกษาต่อในสถาบันอุดมศึกษารับจำกัดจำนวน ประจำปี 2543 ครั้งที่ 1/มีนาคม ทั้งผู้สอบที่เข้าศึกษาต่อได้และผู้ที่สอบเข้าศึกษาต่อไม่ได้

แบบวิชาภาษาอังกฤษ หมายถึง แบบสอบคัดเลือกบุคคลเข้าศึกษาในระดับอุดมศึกษา ปี 2543 ครั้งที่ 1/มีนาคม วิชาภาษาอังกฤษ ของทบวงมหาวิทยาลัย

แบบวิชาคณิตศาสตร์ หมายถึง แบบสอบคัดเลือกบุคคลเข้าศึกษาในระดับอุดมศึกษา ปี 2543 ครั้งที่ 1/มีนาคม วิชาคณิตศาสตร์ 1 ของทบวงมหาวิทยาลัย

กระบวนการ DFIT (DFIT procedure) หมายถึง วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ โดยจะพิจารณาการทำหน้าที่ต่างกันของข้อสอบจากความแตกต่างระหว่างโอกาสในการตอบข้อสอบได้ถูกต้อง และจะพิจารณาการทำหน้าที่ต่างกันของแบบสอบจากความแตกต่างระหว่างคะแนนจริงของผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ (Raju และคณะ, 1995)

เกณฑ์การตัดสินข้อสอบที่ทำหน้าที่ต่างกันตามกระบวนการ DFIT หมายถึง ค่าของดัชนี NCDIF ที่แสดงว่าข้อสอบทำหน้าที่ต่างกัน เมื่อดัชนี NCDIF มีค่ามากกว่า 0.006 (Fleer, 1993 อ้างถึงใน Raju และคณะ, 1995)

เกณฑ์การตัดสินแบบสอบที่ทำหน้าที่ต่างกันตามกระบวนการ DFIT หมายถึง ค่าของดัชนี DTF ที่แสดงว่าแบบสอบทำหน้าที่ต่างกัน เมื่อดัชนี DTF มีค่ามากกว่า 0.006 (Fleer, 1993 อ้างถึงใน Raju และคณะ, 1995)

ข้อสอบที่ทำหน้าที่ต่างกัน (DIF) หมายถึง ข้อสอบที่ทำให้ผลการตอบของผู้สอบที่มีความสามารถเท่ากันในสิ่งที่ต้องการวัด มีโอกาสในการตอบข้อสอบข้อนั้นได้ถูกต้องไม่เท่ากัน เนื่องจากอยู่ในกลุ่มย่อยต่างกัน

แบบสอบที่ทำหน้าที่ต่างกัน (DTF) หมายถึง แบบสอบที่ทำให้ผลการตอบของผู้สอบที่มีความสามารถเท่ากันในสิ่งที่ต้องการวัด ได้คะแนนจริงไม่เท่ากัน เนื่องจากผู้สอบอยู่ในกลุ่มย่อยต่างกัน

สถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน หมายถึง เขตพื้นที่ที่โรงเรียนตั้งอยู่ โดยจำแนกเป็น 2 กลุ่ม ได้แก่ 1) เขตกรุงเทพมหานครและปริมณฑล และ 2) นอกเขตกรุงเทพมหานครและปริมณฑล

กลุ่มอ้างอิง (reference group : R) หมายถึง กลุ่มผู้สอบที่คาดว่าจะได้รับประโยชน์จากการตอบข้อสอบที่ทำหน้าที่ต่างกัน คือเป็นกลุ่มที่มีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องสูงกว่าผู้สอบอีกกลุ่มหนึ่ง ทั้งๆ ที่มีความสามารถเท่ากัน ในที่นี้สำหรับวิชาภาษาอังกฤษ คือกลุ่มผู้สอบเพศหญิงและกลุ่มผู้สอบที่อยู่ในเขตกรุงเทพมหานครและปริมณฑล สำหรับวิชาคณิตศาสตร์ คือกลุ่มผู้สอบเพศชายและกลุ่มผู้สอบที่อยู่ในเขตกรุงเทพมหานครและปริมณฑล

กลุ่มเปรียบเทียบ (focal group : F) หมายถึง กลุ่มผู้สอบที่คาดว่าจะเสียประโยชน์จากการตอบข้อสอบที่ทำหน้าที่ต่างกัน คือเป็นกลุ่มที่มีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องต่ำกว่าผู้สอบอีกกลุ่มหนึ่ง ทั้งๆ ที่มีความสามารถเท่ากัน ในที่นี้สำหรับวิชาภาษาอังกฤษ คือกลุ่มผู้สอบเพศชายและกลุ่มผู้สอบที่อยู่นอกเขตกรุงเทพมหานครและปริมณฑล สำหรับวิชาคณิตศาสตร์ คือกลุ่มผู้สอบเพศหญิงและกลุ่มผู้สอบที่อยู่นอกเขตกรุงเทพมหานครและปริมณฑล

ความเที่ยง (reliability) หมายถึง คุณภาพของแบบสอบซึ่งพิจารณาได้จากความสอดคล้องภายใน (internal consistency) ซึ่งคำนวณค่าความเที่ยงโดยใช้สูตร KR-20 ของคูเดอร์-ริชาร์ดสัน สำหรับแบบสอบวิชาภาษาอังกฤษ และใช้สูตร Alpha ของครอนบาค สำหรับแบบสอบวิชาคณิตศาสตร์

ความตรง (validity) หมายถึง คุณภาพของแบบสอบ ซึ่งพิจารณาจากความตรงเชิงโครงสร้าง (construct validity) โดยตรวจสอบด้วยวิธีวิเคราะห์องค์ประกอบเชิงยืนยัน (confirmatory factor analysis : CFA)

ค่าฟังก์ชันสารสนเทศของแบบสอบ (test information function : TIF) หมายถึง ผลรวมของค่าสารสนเทศของข้อสอบทุก ๆ ข้อในแบบสอบ สำหรับความสามารถของผู้สอบต่าง ๆ กัน ถ้ามีค่าสูงที่ความสามารถระดับใด แสดงว่า แบบสอบนั้นสามารถจำแนกความสามารถของผู้สอบได้ดี ณ ความสามารถนั้น

ความสัมพันธ์ระหว่างตำแหน่งของคะแนน หมายถึง ค่าสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนแรงค์ (spearman rank correlation coefficient) แสดงความสัมพันธ์ระหว่างตำแหน่งของคะแนนผู้สอบแต่ละคนก่อนและหลังตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบสอบ

ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อเป็นข้อมูลสำหรับคณะกรรมการหรือผู้ที่รับผิดชอบในการสร้างแบบสอบคัดเลือกเข้ามหาวิทยาลัยในปีต่อ ๆ ไป เพื่อหาทางปรับปรุงแก้ไขข้อสอบที่มีการทำหน้าที่ต่างกันให้สามารถนำไปใช้ได้อย่างยุติธรรมสำหรับผู้สอบต่างกลุ่ม อันจะทำให้ผลการคัดเลือกมีความถูกต้องยิ่งขึ้น
2. เพื่อเป็นการตรวจสอบคุณภาพของรายข้อและรายฉบับของแบบสอบวิชาภาษาอังกฤษและวิชาคณิตศาสตร์ เพื่อคัดเลือกนักเรียนเข้าศึกษาในระดับอุดมศึกษา ว่ามีคุณภาพด้านความยุติธรรมอย่างไร
3. เพื่อเป็นแนวทางในการวิเคราะห์คุณภาพของแบบสอบอีกวิธีหนึ่งให้แก่นักวัดผลการศึกษา และเพื่อเป็นแนวทางสำหรับการตรวจสอบคุณภาพในวิชาอื่น ๆ ต่อไป

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 2

วรรณคดีที่เกี่ยวข้อง

ผู้วิจัยจะนำเสนอเอกสารและงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบสำหรับแบบสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาวิชาภาษาอังกฤษและวิชาคณิตศาสตร์ ด้วยกระบวนการ ดี เอฟ ไอ ที (DFIT Procedure) โดยจะนำเสนอเป็น 4 ตอน ดังนี้

ตอนที่ 1 ความเป็นมาของการสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษา

ตอนที่ 2 การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ

ตอนที่ 3 วิธีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบด้วย กระบวนการ ดี เอฟ ไอ ที (DFIT Procedure)

ตอนที่ 4 งานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ

ตอนที่ 1 ความเป็นมาของการสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษา

การสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษา หรือการสอบคัดเลือกรวมได้ดำเนินการมาตั้งแต่ปีการศึกษา 2504 โดยมหาวิทยาลัยเกษตรศาสตร์ และมหาวิทยาลัยแพทย์ศาสตร์ (มหาวิทยาลัยมหิดลในปัจจุบัน) จัดสอบร่วมกันเป็นครั้งแรก มีสำนักงานสภาการศึกษาเป็นผู้ประสานงาน ปีการศึกษา 2505 มหาวิทยาลัยทั้ง 5 แห่งที่มีอยู่ในขณะนั้นคือ จุฬาลงกรณ์มหาวิทยาลัย มหาวิทยาลัยธรรมศาสตร์ มหาวิทยาลัยแพทย์ศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ และมหาวิทยาลัย ศิลปากร เข้าร่วมในการสอบด้วย

ปีการศึกษา 2509 คณะรัฐมนตรีเห็นชอบตามข้อเสนอของจุฬาลงกรณ์มหาวิทยาลัย ให้มหาวิทยาลัยต่าง ๆ กลับไปสอบแยก แต่วิธีดำเนินการมีปัญหามาก

ปีการศึกษา 2510 คณะรัฐมนตรีเห็นชอบตามข้อเสนอของสำนักงานสภาการศึกษาแห่งชาติให้มหาวิทยาลัยต่าง ๆ กลับมาใช้วิธีสอบรวมอีก

ปีการศึกษา 2516 ทบวงมหาวิทยาลัย รับโอนงานการสอบคัดเลือกมาจากสำนักงานสภาการศึกษาแห่งชาติ

ปัจจุบันมีสถาบันอุดมศึกษาของรัฐ 20 สถาบัน สถาบันอุดมศึกษาของรัฐที่สังกัดหน่วยงานอื่น 8 สถาบัน รวมทั้งสิ้น 28 สถาบัน และสถาบันประเภทไม่จำกัดรับ (มหาวิทยาลัยเปิด) 2 สถาบัน นอกจากนี้ยังมีสถาบันอุดมศึกษาของเอกชนอีก 33 สถาบัน รวมทั้งสิ้น 63 สถาบัน

สถาบันอุดมศึกษาในระบบจำกัดจำนวนรับของประเทศไทย ได้ใช้การสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาในสังกัดและในกำกับของทบวงมหาวิทยาลัย มีวิธีการรับนักศึกษา 3 วิธี คือ 1. การรับตรง หรือการรับนักศึกษาตามระบบโควต้าของมหาวิทยาลัยส่วนภูมิภาค เป็นการเปิดโอกาสให้นักเรียนในส่วนภูมิภาคเข้าสู่ระบบอุดมศึกษาได้มากขึ้น

2. การรับนักศึกษาตามโครงการพิเศษต่าง ๆ ที่มหาวิทยาลัยกำหนดขึ้น เช่น โครงการพิเศษสำหรับผู้ที่มีความสามารถพิเศษทางด้านการกีฬา ด้านวิชาการและด้านศิลปะ เป็นต้น

3. การสอบคัดเลือกรวมที่ทบวงมหาวิทยาลัยเป็นผู้ดำเนินการ

ในการดำเนินการสอบคัดเลือกรวมในปัจจุบันมีรูปแบบการบริหารและการดำเนินงาน สรุปได้ ดังนี้

1. การบริหารงานสอบคัดเลือกฯ จัดทำในรูปแบบคณะกรรมการ โดยมีคณะอนุกรรมการประสานงานการสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาของรัฐ ทำหน้าที่กำหนดนโยบายเกี่ยวกับการสอบคัดเลือกรวม โดยมีผู้แทนมหาวิทยาลัยที่เกี่ยวข้องร่วมเป็นอนุกรรมการ

2. การดำเนินการสอบคัดเลือกฯ คณะอนุกรรมการประสานงานการสอบคัดเลือกฯ จะแต่งตั้งคณะกรมการดำเนินการสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษา ทบวงมหาวิทยาลัยขึ้นดำเนินงานสอบคัดเลือกฯ ในแต่ละปี โดยมีประธานอนุกรรมการฝ่ายต่าง ๆ ร่วมเป็นกรรมการ

จากการประเมินรวบรวมปัญหาเกี่ยวกับการดำเนินการสอบคัดเลือกรวม พบว่ามีปัญหาที่สำคัญ เช่น มีผลกระทบในทางลบต่อการจัดการศึกษาในระดับมัธยมศึกษาตอนปลาย นักศึกษาที่สอบผ่านการคัดเลือกมีคุณสมบัติไม่สอดคล้องกับสาขาที่เรียน จำนวนผู้สมัครสอบมีแนวโน้มว่าจะมีจำนวนมากขึ้น สร้างความเครียดแก่ผู้ที่เข้าสอบและผู้ปกครอง และมีข้อจำกัดในด้านเวลา สามารถดำเนินการได้เฉพาะในช่วงปิดภาคการศึกษา จากปัญหาดังกล่าว ทบวงมหาวิทยาลัยจึงได้แต่งตั้งคณะอนุกรรมการพิจารณาระบบการสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาขึ้นเมื่อปี พ.ศ. 2535 เพื่อทำหน้าที่ศึกษาและเสนอแนะแนวทางในการปรับปรุงระบบคัดเลือกฯ คณะอนุกรรมการพิจารณาปรับปรุงระบบการสอบคัดเลือกฯ ได้เสนอรูปแบบการคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาระบบใหม่ ซึ่งทบวงมหาวิทยาลัยได้ให้ความเห็นชอบและได้ประกาศให้นำรูปแบบและวิธีการคัดเลือกฯ ระบบใหม่มาใช้ตั้งแต่ปีการศึกษา 2542 เป็นต้นไป

ในการสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษา ระบบใหม่ มีวัตถุประสงค์และองค์ประกอบดังนี้

1. วัตถุประสงค์หลัก มี 2 ประการ คือ

1.1 เพื่อให้มหาวิทยาลัย / สถาบัน ได้ผู้เรียนที่มีความรู้ ความสามารถและความถนัดตรงตามสาขาวิชาที่เรียน

1.2 เพื่อส่งเสริมให้การเรียนการสอนในระดับมัธยมศึกษาตอนปลาย เป็นไปตามปรัชญาและวัตถุประสงค์ของหลักสูตร

2. องค์ประกอบของการคัดเลือก ประกอบด้วย

2.1 คุณสมบัติของผู้สมัคร

1) ผู้ที่สำเร็จชั้นมัธยมศึกษาตอนปลายสายสามัญ (ชั้นมัธยมศึกษาปีที่ 6) หรือผู้ที่กำลังศึกษาอยู่ในชั้นมัธยมศึกษาปีที่ 6

2) ผู้ที่สำเร็จสายอาชีวศึกษาระดับประกาศนียบัตรวิชาชีพ (ปวช.) หรือผู้ที่กำลังศึกษาอยู่ชั้นปีที่ 3

3) ผู้ที่สำเร็จปริยัติธรรมสายสามัญ หรือผู้ที่กำลังศึกษา ปธ. 5

4) ผู้ที่สำเร็จเกรด 12 จากโรงเรียนนานาชาติ หรือกำลังศึกษาเกรด 12

5) ผู้ที่สำเร็จหลักสูตรเทียบเท่ามัธยมศึกษาตอนปลายจากต่างประเทศ หรือกำลังศึกษาเกรด 12

6) ผู้ที่สำเร็จหลักสูตรมัธยมศึกษาตอนปลาย หรือเทียบเท่า

2.2 คุณสมบัติทางการศึกษา

การคัดเลือกฯ ระบบใหม่จะพิจารณาจากองค์ประกอบของผลสัมฤทธิ์ทางการเรียนหลาย ๆ ด้านแทนการใช้คะแนนจากการสอบคัดเลือกฯ เพียงอย่างเดียว โดยจะพิจารณาจาก 3 องค์ประกอบคือ

1) ผลการเรียนตลอดหลักสูตรระดับมัธยมศึกษาตอนปลายหรือเทียบเท่าที่นักเรียนแต่ละคนได้รับจากสถานศึกษาที่สำเร็จการศึกษา ให้ค่าน้ำหนักคะแนนร้อยละ 10

2) ผลการสอบวิชาความรู้พื้นฐานวิชาการ (วิชาหลัก) และวิชาพื้นฐานวิชาชีพ (วิชาเฉพาะ) ให้ค่าน้ำหนักคะแนนรวมกันร้อยละ 90

3) ผลการสอบสัมภาษณ์และตรวจร่างกาย มหาวิทยาลัย/สถาบัน จะทำการ

สอบสัมภาษณ์และตรวจร่างกายเพื่อหาข้อมูลประกอบการพิจารณาความพร้อมและความเหมาะสมในขั้นสุดท้ายก่อนการรับเข้าศึกษา โดยไม่คิดเป็นค่าน้ำหนักคะแนน (สำนักทดสอบกลาง , ทบวงมหาวิทยาลัย , 2541)

วิธีการและขั้นตอนการคัดเลือก

ขั้นตอนการคัดเลือก มี 2 ขั้นตอน ได้แก่

1. การสอบวัดความรู้

1.1 ผู้ที่ประสงค์จะสมัครเข้ารับการศึกษาในสถาบันอุดมศึกษาจะต้องสอบวัดความรู้วิชา

ต่าง ๆ ที่มหาวิทยาลัย/สถาบัน กำหนดไว้ให้ครบถ้วนทุกรายวิชา โดยที่ทบวงมหาวิทยาลัยจะจัดสอบปีละ 2 ครั้ง

ครั้งที่ 1 ในช่วงเดือนกุมภาพันธ์-เดือนมีนาคม

ครั้งที่ 2 ในช่วงเดือนกันยายน-เดือนตุลาคม

1.2 วิชาที่จัดสอบมีทั้งสิ้น 41 วิชาแบ่งเป็นวิชาหลัก 15 วิชา วิชาเฉพาะ 26 วิชา วิชาหลักจัดสอบปีละ 2 ครั้ง วิชาเฉพาะจัดสอบปีละ 2 ครั้งเช่นกัน ยกเว้นวิชาความถนัดทางสถาปัตยกรรมศาสตร์ ปฏิบัติดุริยางคศิลป์ (ไทย-สากล) และปฏิบัตินาฏศิลป์ (ไทย-สากล) ที่จัดสอบปีละ 1 ครั้งเฉพาะการสอบครั้งที่ 2 (เดือนกันยายน-เดือนตุลาคม) เท่านั้น

วิชาหลักที่จัดสอบ มีทั้งหมด 15 วิชา ได้แก่

1. ภาษาไทย
2. ภาษาอังกฤษ
3. สังคมศึกษา
4. คณิตศาสตร์ 1 (สายวิทย์)
5. คณิตศาสตร์ 2 (สายศิลป์)
6. เคมี
7. ชีววิทยา
8. ฟิสิกส์
9. วิทยาศาสตร์กายภาพชีวภาพ
10. ภาษาฝรั่งเศส
11. ภาษาเยอรมัน
12. ภาษาบาลี
13. ภาษาอาหรับ
14. ภาษาจีน
15. ภาษาญี่ปุ่น

สำหรับวิชาภาษาไทยและวิชาสังคมศึกษา ทุกคณะ/ประเภทวิชากำหนดให้สอบข้อสอบชุดเดียวกัน เนื่องจากพิจารณาเห็นว่าภาษาไทยเป็นภาษาประจำชาติ และเป็นสื่อที่สำคัญในการเรียนการสอน และวิชาสังคมศึกษาเป็นวิชาที่จะช่วยให้นักเรียนเป็นผู้มีความรู้ความเข้าใจในสังคม ชีวิต ทำให้มีการปรับตัวได้ดี อีกทั้งการกำหนดให้สอบวิชาทั้ง 2 วิชา ดังกล่าวจะส่งเสริมให้นักเรียนมีความสนใจในวิชาดังกล่าวในระดับมัธยมศึกษาตอนปลายหรือเทียบเท่าด้วย

ข้อสอบแต่ละวิชาจะครอบคลุมเนื้อหาวิชาตามหลักสูตรมัธยมศึกษาตอนปลายของกรมวิชาการ กระทรวงศึกษาธิการทั้งหมด และใช้เวลาสอบวิชาละ 2 ชั่วโมง

วิชาเฉพาะที่จัดสอบ มีทั้งหมด 26 วิชา ได้แก่

1. พื้นฐานทางวิศวกรรม
2. วัดแววความเป็นครู
3. ความถนัดทางสถาปัตยกรรมศาสตร์
4. ความรู้ความถนัดทางศิลป์
5. ความสามารถทางศิลปะ
6. ความรู้ทั่วไปทางศิลปวัฒนธรรม
7. ความถนัดทางนิเทศศิลป์
8. ทฤษฎีทัศนศิลป์
9. ปฏิบัติทัศนศิลป์
10. ทฤษฎีนฤมิตรศิลป์
11. ปฏิบัตินฤมิตรศิลป์
12. วาดเส้น
13. องค์ประกอบศิลป์
14. วาดเส้นมัณฑณศิลป์
15. ออกแบบภายใน
16. ออกแบบนิเทศศิลป์
17. ออกแบบผลิตภัณฑ์
18. ออกแบบประยุกต์ศิลป์
19. ออกแบบเครื่องเคลือบดินเผา
20. ทฤษฎีดุริยางคศิลป์
21. ปฏิบัติดุริยางคศิลป์ (ไทย)
22. ปฏิบัติดุริยางคศิลป์ (สากล)
23. ทฤษฎีนาฏยศิลป์
24. ปฏิบัตินาฏยศิลป์ (ไทย)
25. ปฏิบัตินาฏยศิลป์ (ตะวันตก)
26. พลศึกษาปฏิบัติ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

การให้น้ำหนักวิชาเฉพาะมีความหลากหลายไม่ยิ่งหย่อนไปกว่าวิชาหลัก เช่น คณะวิศวกรรมศาสตร์ วิชาพื้นฐานวิศวกรรมคิด 18-20% คณะดนตรีสากล วิชาทฤษฎีดุริยาค์คิด 10% ปฏิบัติดุริยางค์ คิด 50% คณะครุศาสตร์ ศึกษาศาสตร์ วิชาวัดแววความเป็นครู คิด 10% (สำนักทะเบียนและประมวลผล มหาวิทยาลัยเชียงใหม่, 2542 อ้างถึงใน วิไลลักษณ์ บุญเคลือบและคณะ, 2544)

ตอนที่ 2 การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning : DIF)

การทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning : DIF) เดิมเรียกว่าความลำเอียงของข้อสอบ (item bias) ซึ่งเป็นคำที่ใช้ในการศึกษาหรือทดสอบความยุติธรรมของข้อสอบ แต่ต่อมาในระยะหลังเปลี่ยนมาใช้คำว่าการทำหน้าที่ต่างกันของข้อสอบ เพราะเป็นคำที่เป็นกลางและมีความเหมาะสมกว่า (Holland and Thayer , 1988 อ้างถึงใน จิติมา วรรณศรี , 2539) การทำหน้าที่ต่างกันของข้อสอบ คือการที่ข้อสอบวัดความสามารถรอง (secondary abilities) หรือ คุณลักษณะแฝงอื่นนอกเหนือจากความสามารถหลัก (primary abilities) หรือ คุณลักษณะแฝงที่ต้องการวัด จะส่งผลให้ผู้สอบต่างกลุ่มที่นำเข้ามาจับคู่เปรียบเทียบกัน มีโอกาสในการตอบข้อสอบได้ถูกต้องต่างกัน ทั้งๆ ที่มีความสามารถหลักที่ต้องการวัดเท่ากัน

การทำหน้าที่ต่างกันของข้อสอบมี 2 แบบ คือ

1 การทำหน้าที่ต่างกันแบบเอกรูป (uniform DIF) หมายถึง การที่ข้อสอบทำให้ผู้สอบกลุ่มหนึ่งมีโอกาสในการตอบข้อสอบได้ถูกต้องมากกว่าอีกกลุ่มหนึ่งสม่ำเสมอในทุกระดับความสามารถ โดยเมื่อพิจารณาโค้งคุณลักษณะข้อสอบระหว่างกลุ่มจะพบว่าไม่มีปฏิสัมพันธ์ระหว่างโค้งคุณลักษณะข้อสอบในทุกระดับความสามารถ

2 การทำหน้าที่ต่างกันแบบอเนกรูป (nonuniform DIF) หมายถึง การที่ข้อสอบทำให้โอกาสในการตอบข้อสอบได้ถูกต้องของผู้สอบระหว่างกลุ่มไม่สม่ำเสมอกันในทุกระดับความสามารถ เมื่อพิจารณาโค้งคุณลักษณะข้อสอบของทั้งสองกลุ่ม จะพบว่ามีปฏิสัมพันธ์ระหว่างโค้งคุณลักษณะข้อสอบ เช่น ที่ความสามารถระดับหนึ่ง กลุ่ม A มีโอกาสในการตอบถูกมากกว่ากลุ่ม B แต่ที่ความสามารถอีกระดับหนึ่ง กลุ่ม B กลับมีโอกาสในการตอบได้ถูกต้องมากกว่ากลุ่ม A เป็นต้น

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF detection) เป็นการเปรียบเทียบผลการตอบข้อสอบระหว่างผู้สอบ 2 กลุ่มได้แก่ กลุ่มแรก เรียกว่า กลุ่มเปรียบเทียบ (focal groups) เป็นกลุ่มที่สนใจศึกษาและคาดว่าจะเป็นกลุ่มที่เสียประโยชน์ในการตอบข้อสอบ และกลุ่มที่สอง เรียกว่า กลุ่มอ้างอิง (reference groups) เป็นกลุ่มที่คาดว่าจะได้ประโยชน์จากการตอบข้อสอบได้ถูกต้อง

Dorans และ Potenza (1994) ได้จำแนกวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบทวิวิภาค (dichotomously score) และ ข้อสอบที่มีการให้คะแนนแบบพหุวิภาค (polytomous score) โดยแบ่งออกเป็น 2 มิติ มิติแรกแบ่งตามประเภทการวิเคราะห์ระหว่างกลุ่มกระบวนการที่ใช้คะแนนสังเกตได้ (observed score) และกลุ่มที่ใช้คะแนนสังเกตไม่ได้หรือตัวแปรแฝง (latent variable) ส่วนอีกมิติหนึ่งเป็นการแบ่งระหว่างวิธีการพาราเมทริกซ์ (parametric approaches) และแบบนอนพาราเมทริกซ์ (nonparametric approaches)

นอกจากนี้สามารถแบ่งประเภทของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบตามทฤษฎีพื้นฐาน ได้เป็น 2 กลุ่ม (คมศักดิ์ ชื่นชม, 2539) ดังนี้คือ

กลุ่มที่ 1 กลุ่มที่ใช้หลักการของทฤษฎีการตอบสนองข้อสอบ (item response theory : IRT) ได้แก่ วิธีโค้งลักษณะข้อสอบแบบ 3 พารามิเตอร์ (item characteristic curve-3 parameter : ICC-3) วิธีโค้งลักษณะข้อสอบแบบ 2 พารามิเตอร์ (item characteristic curve-2 parameter : ICC-2) วิธีโค้งลักษณะข้อสอบแบบ 1 พารามิเตอร์ (item characteristic curve-1 parameter : ICC-1) วิธีทดสอบอัตราส่วนความเป็นไปได้ (likelihood ratio test) วิธีซิบเทสส์ (sibtest) และ วิธีทดสอบ ไคสแคว์ของลอร์ด (Lord's χ^2-test)

กลุ่มที่ 2 กลุ่มที่ใช้หลักการของทฤษฎีการทดสอบแบบดั้งเดิม (classical test theory : CTT) ได้แก่ วิธีแปลงค่าความยากของข้อสอบ (transformed item difficulty : TID) วิธีวิเคราะห์ความแปรปรวน (analysis of variance) วิธีวิเคราะห์ด้วยไคสแคว์ (chi-square) วิธีวิเคราะห์องค์ประกอบ (factor analysis) วิธีวิเคราะห์การถดถอย (regression analysis) วิธีค่าอำนาจจำแนกของข้อสอบ (item discrimination indices) วิธีลอกลิเนียร์ (log-linear) วิธีแมนเทล-เฮนส์เซล (Mantel-Haenszel) วิธีทำให้เป็นมาตรฐาน (standardization) และวิธีถดถอยโลจิสติก (logistic regression)

จะเห็นว่ามีวิธีการหลายวิธีในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ แต่วิธีที่เป็นที่รู้จักและนิยมใช้ในการเปรียบเทียบที่ผ่านมา กาญจนา วัธนสุนทร (2537) ได้สรุปว่าวิธีที่นิยมใช้มี 6 วิธีได้แก่ วิธีแปลงค่าความยากของข้อสอบ (transformed item difficulty : TID) วิธีวิเคราะห์ความแปรปรวน (analysis of variance : ANOVA) วิธีวิเคราะห์ด้วยไคสแคว์ (chi-square : χ^2) วิธีทฤษฎีการตอบสนองข้อสอบ (item response theory : IRT) วิธีแมนเทล-เฮนส์เซล (Mantel-Haenszel : MH) และวิธีซิบเทสส์ (sibtest) ซึ่ง จิติมา วรรณศรี (2539) ได้สรุปวิธีการวิเคราะห์ ข้อดี ข้อจำกัดไว้ดังตารางที่ 2

ตารางที่ 2 เปรียบเทียบวิธีต่าง ๆ ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

ประเด็น	TID	ANOVA	γ^2	IRT	MH	SIBTEST
1.ข้อตกลงเบื้องต้น	ผลร่วมระหว่างกลุ่มกับข้อกระทงเป็นตัวบ่งชี้ความลำเอียง	ความแปรปรวนและความแปรปรวนร่วมของข้อสอบต้องเท่ากัน	คะแนนรวมจากแบบสอบเป็นตัวแทนความสามารถของผู้สอบ	แบบสอบเป็นเอกมิติและโค้งลักษณะข้อสอบสามารถแสดงฟังก์ชันของค่าความสามารถและโอกาสในการตอบข้อสอบถูก	คะแนนรวมจากแบบสอบเป็นตัวแทนความสามารถของผู้สอบ	คะแนนรวมจากแบบสอบเป็นตัวแทนความสามารถของผู้สอบและมีมิติการวัด 2 มิติคือคุณลักษณะแฝงเป้าหมายและคุณลักษณะแฝงแทรกซ้อน
2. สิ่งที่ทำการวิเคราะห์	ผลร่วมระหว่างการเป็นสมาชิกในกลุ่มกับการตอบถูก	ผลร่วมระหว่างการเป็นสมาชิกในกลุ่มกับการตอบถูก	ความแตกต่างของอัตราส่วนการตอบถูกตามระดับคะแนนรวม	ความแตกต่างของฟังก์ชันการตอบข้อสอบที่ระดับความสามารถเดียวกัน	ความแตกต่างของอัตราส่วนการตอบระหว่างผู้ที่มีความสามารถระดับเดียวกัน	ความแตกต่างระหว่างคะแนนเฉลี่ยและอัตราส่วนการตอบข้อสอบระหว่างผู้ที่มีความสามารถระดับเดียวกัน
3. สิ่งที่พิจารณาในการตัดสิน DIF	ระยะห่างของจุดเดลต้าจากเส้นแกนหลัก	ความมีนัยสำคัญทางสถิติของ F-test	ความมีนัยสำคัญทางสถิติของ χ^2	พื้นที่ระหว่างโค้งลักษณะข้อสอบ	ค่าของดัชนี α_{MH} และความมีนัยสำคัญทางสถิติ	ค่าของดัชนี β_{SIB} และความมีนัยสำคัญทางสถิติ
4. ทฤษฎีพื้นฐาน	CTT	CTT	CTT	IRT	CTT	IRT ชนิดพหุมิติ
5. ข้อดี	คำนวณง่าย ใช้กลุ่มตัวอย่างน้อย	ใช้กลุ่มตัวอย่างน้อย	คำนวณง่ายมีเกณฑ์ตายตัวในการแปลผล	ให้รายละเอียดมากและการไม่แปรเปลี่ยนของค่าพารามิเตอร์ตามกลุ่มผู้สอบ	คำนวณง่าย ใช้กลุ่มตัวอย่างน้อย ประหยัด	คำนวณง่าย ใช้กลุ่มตัวอย่างน้อย ตรวจสอบ DIF ได้หลายข้อในคราวเดียวกัน
6. ข้อจำกัด	มีความคลาดเคลื่อนเมื่อค่า a สูง และค่า b เปลี่ยนตามกลุ่มผู้สอบ	การคำนวณค่อนข้างยุ่งยากและไม่มีดัชนีบอกระดับความลำเอียง	ไม่มีเกณฑ์ตายตัวในการกำหนดช่วงคะแนนและค่า b เปลี่ยนตามกลุ่มผู้สอบ	มีการคำนวณซับซ้อนหลายรอบแปลผลยาก ใช้กลุ่มตัวอย่างมาก ค่าใช้จ่ายสูง	ไม่มีความไวในการตรวจสอบ DIF แบบอเนกรูป (nonuniform DIF)	อัตรความคลาดเคลื่อนชนิดที่ 1 เพิ่มสูงเมื่อคะแนนเฉลี่ยแตกต่างกันมาก

จากตารางที่ 2 พบว่าวิธีวิเคราะห์ความแปรปรวน (ANOVA) วิธีไค-สแควร์ (χ^2) และวิธีค่าความยากแปลง (TID) เป็นการวิเคราะห์ข้อสอบโดยอาศัยทฤษฎีการวัดแบบดั้งเดิม ซึ่งมีจุดด้อยคือค่าพารามิเตอร์ของข้อสอบจะแปรเปลี่ยนไปตามกลุ่มผู้สอบ นอกจากนี้วิธีวิเคราะห์ความแปรปรวนและวิธีไคว์สแควร์ไม่มีดัชนีบอกระดับการทำหน้าที่ต่างกันของข้อสอบ แต่เป็นวิธีที่ประหยัดและใช้กลุ่มตัวอย่างน้อย วิธีทฤษฎีการตอบสนองข้อสอบ (IRT) เป็นวิธีที่วิเคราะห์ความแตกต่างของฟังก์ชันการตอบข้อสอบระหว่างกลุ่มผู้สอบ มีดัชนีบอกระดับของการทำหน้าที่ต่างกันของข้อสอบและทดสอบความมีนัยสำคัญทางสถิติ ซึ่งเป็นวิธีที่ให้รายละเอียดมากและมีข้อดีคือการไม่แปรเปลี่ยนของค่าพารามิเตอร์ แต่มีข้อเสียคือค่อนข้างสิ้นเปลือง วิธีแมนเทล-เฮนส์เซลคล้ายกับวิธีไคสแควร์ คือใช้คะแนนรวมจากแบบสอบเป็นตัวแทนของความสามารถ แต่วิธีแมนเทล-เฮนส์เซลจะวิเคราะห์ทีละระดับความสามารถ และมีดัชนีบอกระดับการทำหน้าที่ต่างกันของข้อสอบและมีการทดสอบความนัยสำคัญทางสถิติ วิธีซิบเทสส์ (sibtest) ใช้คะแนนรวมจากแบบสอบเป็นตัวแทนความสามารถ มีข้อตกลงว่ามีมิติการวัด 2 มิติ ดังนั้นคะแนนจากแบบสอบจึงมี 2 ส่วนคือ คะแนนจากแบบสอบที่มีความตรง (valid subtest) ซึ่งวัดคุณลักษณะแฝงเป้าหมาย และคะแนนจากแบบสอบที่ศึกษา (studied subtest) ซึ่งวัดคุณลักษณะแฝงแทรกซ้อน มีดัชนีบอกระดับการทำหน้าที่ต่างกันของข้อสอบและทดสอบนัยสำคัญทางสถิติ ทั้งวิธีแมนเทล-เฮนส์เซลและซิบเทสส์เป็นวิธีที่ประหยัด ใช้กลุ่มตัวอย่างน้อย (จิติมา วรรณศรี, 2539)

ตอนที่ 3 วิธีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบด้วยกระบวนการ ดี เอฟ ไอ ที (Differential Functioning of Items and Tests Procedure : DFIT)

กระบวนการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ (differential functioning of items and test procedure : DFIT) เป็นวิธีการวิเคราะห์ที่อยู่บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ (item response theory : IRT) ซึ่ง Raju และคณะ ได้เสนอขึ้นในปี 1995 ดังนั้นในส่วนนี้ผู้วิจัยได้เสนอเป็น 2 ตอน ประกอบด้วย ตอนที่ 1 ทฤษฎีการตอบสนองข้อสอบ (IRT) ส่วนที่ 2 กระบวน DFIT

3.1 ทฤษฎีการตอบสนองข้อสอบ (item response theory : IRT)

ทฤษฎีการตอบสนองข้อสอบ หรือทฤษฎีคุณลักษณะแฝง (latent trait theory) หรือทฤษฎีโค้งคุณลักษณะเฉพาะของข้อสอบ (item characteristic curve theory : ICC) เป็นทฤษฎีการวัดที่เกิดจากการที่นักการศึกษาพยายามคิดค้นหาวิธีแก้ไขข้อบกพร่องของทฤษฎีการทดสอบแบบดั้งเดิม (classical test theory : CTT) ซึ่งทฤษฎีการตอบสนองข้อสอบนี้เป็นทฤษฎีที่อธิบายความ

สัมพันธ์ระหว่างคุณลักษณะภายในหรือความสามารถที่มีอยู่ภายในตัวบุคคลนั้นว่ามีโอกาสตอบข้อสอบได้ถูกมากน้อยเพียงไร โดยการวิเคราะห์ผ่านข้อสอบรายข้อ ทฤษฎีนี้มีพื้นฐานความเชื่อว่าพฤติกรรมการตอบสนองต่อข้อสอบของผู้สอบ ซึ่งเป็นสิ่งที่สังเกตได้โดยตรงว่าถูกหรือผิด จะถูกกำหนดโดยคุณลักษณะภายในหรือความสามารถที่อยู่ภายในตัวบุคคลซึ่งเป็นสิ่งที่ไม่สามารถสังเกตได้โดยตรง (ศิริชัย กาญจนวาสี , 2541) โอกาสที่จะตอบข้อสอบถูกหรือผิดจากแบบสอบใด ๆ ของผู้สอบคนหนึ่ง ๆ จะขึ้นอยู่กับระดับความสามารถและคุณลักษณะของข้อสอบ ซึ่งได้แก่ค่าพารามิเตอร์ประจำข้อ อันประกอบด้วยค่าความยาก (b) อำนาจจำแนก (a) และโอกาสในการเดา (c) ค่าต่าง ๆ ดังกล่าวจะอธิบายด้วยความสัมพันธ์กันในรูปของฟังก์ชันคณิตศาสตร์หรือโมเดลที่แสดงความสัมพันธ์ระหว่างค่าดังกล่าว ที่เรียกว่า ฟังก์ชันการตอบสนองข้อสอบ ซึ่งมีหลายรูปแบบ เช่น ฟังก์ชันปกติสะสม ฟังก์ชันโลจิส เป็นต้น

3.1.1 ประเภทของโมเดลการตอบสนองข้อสอบ

จากแนวคิดตามทฤษฎีการตอบสนองข้อสอบได้มีการพัฒนาโมเดลหรือแบบจำลองขึ้นมาหลายรูปแบบด้วยกัน โดยแต่ละโมเดลจะมีฟังก์ชันทางคณิตศาสตร์และจำนวนพารามิเตอร์ในฟังก์ชันแตกต่างกัน ซึ่งสามารถจำแนกเป็น 3 ประเภทใหญ่ ๆ (Hambleton and Swaminathan, 1985 อ้างถึงใน ศิริชัย กาญจนวาสี, 2541) ดังนี้

1) โมเดลการตอบสนองข้อสอบที่ใช้กับการตรวจคะแนนรายข้อแบบ 2 ค่า เป็นข้อสอบที่ตรวจให้คะแนนรายข้อแบบ 2 ค่า เป็นข้อสอบที่ตรวจให้คะแนนแบบ 0 , 1 (ตอบผิดได้ 0 คะแนน ตอบถูกได้ 1 คะแนน) โมเดลประเภทนี้ถูกพัฒนาขึ้นในระยะเริ่มแรก (ค.ศ. 1943 – 1968) เช่น สเกลกัตแมนสมบูรณ์ , โมเดลระยะห่างแฝง (latent distance model) เป็นต้น ในระยะต่อมา (ค.ศ. 1952 – 1982) ได้มีการพัฒนาโมเดลประเภทนี้เพิ่มขึ้นมาอีก เช่น โมเดลปกติสะสมแบบ 1, 2, 3 พารามิเตอร์ เป็นต้น

2) โมเดลการตอบสนองข้อสอบที่ใช้กับการตรวจคะแนนรายข้อแบบมากกว่า 2 ค่า (multichotamous) เช่น โมเดลการตอบสนองแบบนามบัญญัติ โมเดลการตอบสนองแบบเกรด โมเดลการให้คะแนนบางส่วน เป็นต้น

3) โมเดลการตอบสนองข้อสอบที่ใช้กับการตรวจคะแนนรายข้อแบบต่อเนื่อง เช่น โมเดลการตอบสนองแบบคะแนนต่อเนื่อง เป็นต้น

จากโมเดลการตอบสนองข้อสอบทั้ง 3 ประเภท โมเดลที่นิยมใช้กันทั่วไปคือโมเดลการตอบสนองข้อสอบที่ใช้กับการให้คะแนนรายข้อแบบ 2 ค่า โมเดลดังกล่าวเป็นที่รู้จักกันในปัจจุบัน ได้แก่ โมเดลปกติสะสม และโมเดลโลจิสแบบ 1, 2, และ 3 พารามิเตอร์ โมเดลทั้งสองแบบ

นี้ใช้ฟังก์ชันที่ให้ผลลัพธ์การประมาณค่าใกล้เคียงกันมาก แต่ฟังก์ชันโลจิสมีลักษณะของสูตรทางคณิตศาสตร์และวิธีการคำนวณที่ง่ายและสะดวกกว่า

3.1.2 พารามิเตอร์ของโมเดลการตอบสนองข้อสอบ โมเดลการตอบสนองข้อสอบประกอบด้วย พารามิเตอร์และค่าคงที่ ดังนี้

1) พารามิเตอร์ของผู้สอบ (ศิริชัย กาญจนวาสี, 2541)

θ_p = ระดับความสามารถของผู้สอบคนที่ p ซึ่งประมาณได้จากโมเดล

ตามทฤษฎีการตอบสนองข้อสอบ นิยมปรับให้เป็นคะแนนมาตรฐานที่มีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 ค่า θ มีพิสัยอยู่ระหว่าง $-\infty$ ถึง ∞ แต่ผลการวิเคราะห์ส่วนใหญ่มักให้ค่าอยู่ในช่วง –3 ถึง 3

2) พารามิเตอร์ของข้อสอบ (เยาวดี วิบูลศรี, 2539)

b_i = ค่าความยากของข้อสอบข้อที่ i ได้มาจากค่าความสามารถที่ตรงจุดเปลี่ยนโค้ง (inflexion point) ซึ่งเป็นจุดที่โค้งมีความชันมากที่สุด หรือมีความหมายอีกนัยหนึ่งก็คือผู้สอบที่มีความสามารถถึงระดับ ณ จุดเปลี่ยนโค้งข้อสอบนั้น จะมีโอกาสตอบข้อสอบข้อนี้ถูกอยู่ 0.5 หรือในทางปฏิบัติกล่าวได้ว่า จากจุดบนแกน y ที่แสดงถึงตำแหน่งโอกาสในการตอบข้อสอบข้อนี้ถูกมีอยู่ 0.5 ถ้าลากเส้นตั้งขนานกับแกน X จนพบกับเส้นโค้ง ซึ่งเป็นจุดเปลี่ยนโค้งด้วยนั้น ในทางตรงกันข้ามเมื่อลากเส้นตั้งฉากจากจุดดังกล่าวให้มาจดแกน X ค่าที่วัดได้ในแกน X คือค่าความยากของข้อสอบนั้นๆ ซึ่งค่าความยากที่ได้จะกระจายอยู่บนแกน X จากค่า $-\infty$ ถึง ∞ แต่ในทางปฏิบัตินิยมใช้ช่วง –3 ถึง 3 และแบบสอบทั่ว ๆ ไป มักจะมีค่า b_i อยู่ระหว่าง –2.5 ถึง 2.5 ถ้าค่า b_i อยู่ใกล้ –2.5 แสดงว่าเป็นข้อสอบที่ง่าย ตรงกันข้าม ถ้าค่า b_i อยู่ใกล้ 2.5 แสดงว่าเป็นข้อสอบที่ยาก ในกรณีที่เป็นโมเดลแบบ 3 พารามิเตอร์ หรือคำนึงถึงโอกาสในการเดาคำตอบ (c) ค่าความยากซึ่งเริ่มต้นจากจุดบนแกน y นั้น จะใช้จุดตั้งต้นตรงที่โอกาสในการตอบถูก กล่าวคือ ค่าความยากมีค่าเริ่มต้นที่ $\frac{1-c}{2}$

a_i = ค่าอำนาจจำแนก เป็นสัดส่วนกับค่าความชัน (slope) ของ $P(\theta)$ ที่จุดเปลี่ยนโค้งหรือที่จุด $\theta = b$ โดยทฤษฎีแล้ว ค่าอำนาจจำแนกจะมีค่าอยู่ในช่วง $-\infty$ ถึง ∞ แต่ในทางการนำมาใช้ประโยชน์นั้น ข้อสอบข้อใดที่มีค่า a ติดลบ ย่อมแสดงว่า ข้อสอบข้อนั้นไม่ดีและควรตัดออกไป ส่วนข้อสอบที่มี a สูงขึ้น ย่อมแสดงว่า ความน่าจะเป็นของการตอบข้อสอบข้อนั้น ๆ ถูก เพิ่มขึ้น เมื่อระดับความสามารถของผู้สอบสูงขึ้น ในทางปฏิบัติมักจะใช้ข้อกระทงที่มีค่า a_i อยู่ระหว่าง 0.5 ถึง 2.5

c_i = ค่าการเดา เป็นค่าที่อยู่ปลายโค้งด้านต่ำ (lower asymptote) ของข้อสอบ ค่านี้เป็นค่าแทนโอกาสที่คนซึ่งมีความสามารถต่ำ แต่สามารถตอบข้อสอบนั้นถูกโดยการเดา ค่าการเดาจะอยู่ในช่วง 0.00 ถึง 0.20 เมื่อแบบสอบเป็นแบบเลือกตอบที่มี 4 ตัวเลือก ดังนั้นข้อสอบข้อใดที่มีค่า c_i มากกว่า 0.30 ขึ้นไปแสดงว่าข้อสอบนั้นไม่ดี ควรพิจารณาตัดออก

3.) ค่าคงที่

e = ค่าคงที่ของลอการิทึมธรรมชาติ ซึ่งมีค่าประมาณ 2.71828

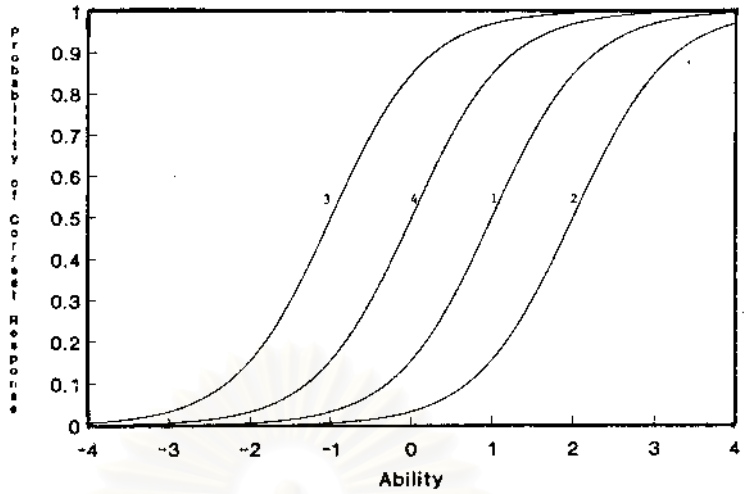
D = ค่าองค์ประกอบของการปรับสเกลซึ่งมีค่าเท่ากับ 1.70

ฟังก์ชันทางคณิตศาสตร์ของโมเดลการตอบสนองข้อสอบ สำหรับฟังก์ชันปกติสะสมและฟังก์ชันโลจิส สามารถแสดงได้ดังตารางที่ 3 (ศิริชัย กาญจนวาสี, 2541)

ตารางที่ 3 ฟังก์ชันทางคณิตศาสตร์ของโมเดลการตอบสนองข้อสอบ

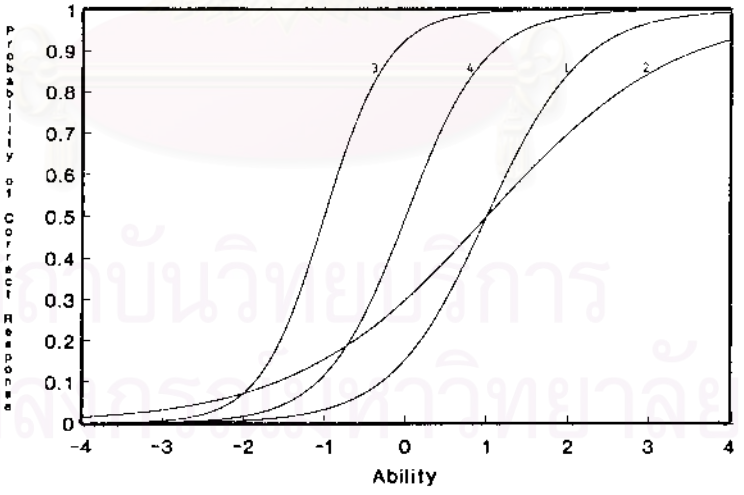
โมเดล	ฟังก์ชันปกติสะสม	ฟังก์ชันโลจิส
1 พารามิเตอร์	$P_i(\theta)=\int_{-\infty}^{e-b_i}\frac{1^{e-z^2/2}}{\sqrt{2\pi}}dz$	$P_i(\theta)=\frac{1}{1+e^{-D(\theta-b_i)}}$
2 พารามิเตอร์	$P_i(\theta)=\int_{-\infty}^{a_i(e-b_i)}\frac{1^{e-z^2/2}}{\sqrt{2\pi}}dz$	$P_i(\theta)=\frac{1}{1+e^{\{-Da_i(\theta-b_i)\}}}$
3 พารามิเตอร์	$P_i(\theta)=c_i+(1-c_i)\int_{-\infty}^{a_i(e-b_i)}\frac{1^{e-z^2/2}}{\sqrt{2\pi}}dz$	$P_i(\theta)=c_i\frac{1}{1+e^{\{-Da_i(\theta-b_i)\}}}$

โค้งคุณลักษณะข้อสอบ (Item Characteristic Curve, ICC) ของโมเดลโลจิสแบบ 1 , 2 และ 3 พารามิเตอร์ สามารถแสดงได้ดังภาพที่ 1-3 (Hambleton et al., 1991)



แผนภาพที่ 1 โค้งฟังก์ชันสารสนเทศของข้อสอบของโมเดลโลจิส แบบ 1 พารามิเตอร์ 4 ข้อ

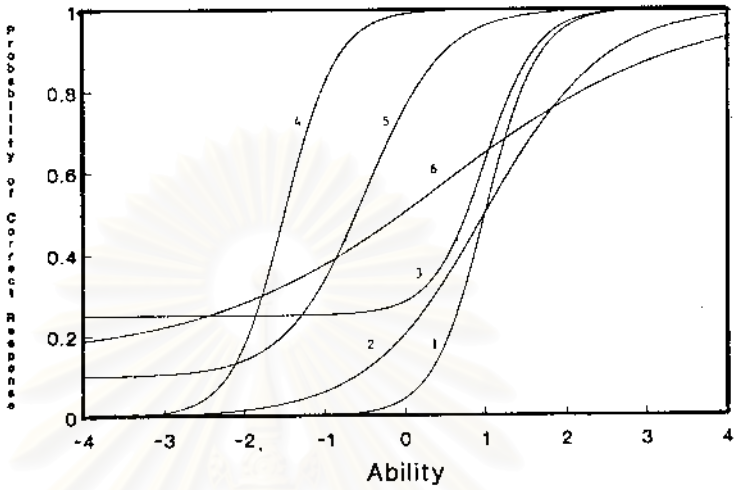
โมเดลโลจิสแบบ 1 พารามิเตอร์ เป็นโมเดลหนึ่งของโมเดล IRT ที่ใช้กันอย่างกว้างขวาง โมเดลนี้เป็นโมเดลที่รู้จักกันในชื่อของราชโมเดล (rash model) ในปัจจุบัน ได้รับการพัฒนาโดย Georg Rash นักคณิตศาสตร์ชาวเดนมาร์ก ในปี ค.ศ. 1960 Rash พัฒนาโมเดลนี้ขึ้นมาโดยไม่ได้เกี่ยวข้องกับทฤษฎีการตอบสนองข้อสอบ แต่โมเดลของ Rash มีลักษณะเป็นแนวคิดเดียวกันกับทฤษฎีการตอบสนองข้อสอบด้วยมีค่าพารามิเตอร์เพียงค่าเดียวคือค่าความยาก (b) โดยมีข้อกำหนดว่าข้อสอบแต่ละข้อ มีค่าอำนาจจำแนกเท่ากัน และไม่มีค่าการเดาคำตอบ



แผนภาพที่ 2 โค้งฟังก์ชันสารสนเทศของข้อสอบของโมเดลโลจิส แบบ 2 พารามิเตอร์ 4 ข้อ

Lord (1952 cited in Hambleton et al., 1991) ได้พัฒนาโมเดลการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ โดยอยู่บนพื้นฐานของการกระจายแบบปกติสะสม ต่อมา Birnbaum (1968 cited in Hambleton et al., 1991) ได้พัฒนาฟังก์ชันโลจิสแบบ 2 พารามิเตอร์เข้ามาแทนฟังก์ชันปกติสะสม

แบบ 2 พารามิเตอร์ ฟังก์ชันโลจิสมีข้อได้เปรียบมากกว่าตรงที่สามารถคำนวณได้สะดวกกว่า โดยในเส้นโค้งโลจิสมีค่าพารามิเตอร์ 2 ตัว คือ ค่าอำนาจจำแนก (a) และ ค่าความยาก (b)



แผนภาพที่ 3 โค้งฟังก์ชันสารสนเทศของข้อสอบของโมเดลโลจิส แบบ 3 พารามิเตอร์ 6 ข้อ

โมเดลแบบ 3 พารามิเตอร์นี้ ได้รับการพัฒนามาจากโมเดลแบบ 2 พารามิเตอร์เพื่อให้เหมาะสมกับการทดสอบ ซึ่งมีอิทธิพลจากการเดาเข้ามาแฝงอยู่ด้วย โดยคำนึงถึงโอกาสในการเดาคำตอบ (c) ด้วย

3.1.3 ฟังก์ชันสารสนเทศ (Information functions)

การวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ จะใช้แบบแผนการตอบสนองแบบสอบเป็นรายข้อในการประมาณค่าความสามารถของผู้สอบ ดังนั้นการประเมินคุณภาพของแบบสอบ จึงสามารถพิจารณาจากความถูกต้องแม่นยำในการประมาณค่าความสามารถของผู้สอบ ซึ่งมีดัชนีตัวหนึ่งสามารถใช้ชี้ถึงความถูกต้องแม่นยำดังกล่าว เรียกว่า สารสนเทศของแบบสอบ เกิดจากผลรวมพีชคณิตของค่าฟังก์ชันสารสนเทศของข้อสอบแต่ละข้อรวมเข้าด้วยกัน โดยค่าฟังก์ชันสารสนเทศของข้อสอบเป็นดัชนีผสมที่สร้างจากดัชนีคุณลักษณะของข้อสอบหลายลักษณะ ได้แก่ ค่าอำนาจจำแนก ค่าความยาก และค่าความแปรปรวนของคะแนนรายข้อ เพื่อบ่งชี้คุณภาพของข้อสอบ (Birnbaum, 1968 อ้างถึงใน
ศิริชัย กาญจนวาสี, 2541) เนื่องจากว่าคุณสมบัติความไม่แปรเปลี่ยนไปตามกลุ่มตัวอย่างของค่าพารามิเตอร์จากการวิเคราะห์ด้วยทฤษฎี IRT ทำให้ค่าฟังก์ชันสารสนเทศเหมาะสมที่จะใช้เป็นตัวบ่งชี้คุณภาพของแบบสอบแทนค่าความเที่ยงและความคลาดเคลื่อนมาตรฐานตามทฤษฎีการวัดแบบดั้งเดิม (Hambleton, 1977 อ้างถึงใน ศิริชัย กาญจนวาสี, 2541)

สูตรที่ใช้ในการคำนวณ มีรายละเอียดดังนี้

1) ความสามารถหรือคะแนนจริงของผู้สอบ

$$T_p = \sum_{i=1}^{k} P_i(\theta_p)$$

เมื่อ T_p = คะแนนจริงของผู้สอบคนที่ p ซึ่งมีความสามารถระดับ θ

k = จำนวนข้อสอบทั้งหมด

$P_i(\theta_p)$ = โอกาสของผู้สอบที่มีความสามารถระดับ θ จะทำข้อสอบข้อที่ i ได้ถูกต้อง

2) ค่าฟังก์ชันสารสนเทศของข้อสอบ (item information function : $I_i(\theta)$

$$I_i(\theta) = (P_i')^2 / P_i Q_i$$

เมื่อ P_i' = $P_i'(\theta_p)$ เป็นค่าอนุพันธ์ของโค้งคุณลักษณะของข้อสอบ (ICC) ข้อที่ i ที่ระดับความสามารถ θ

Q_i = $1-P_i(\theta_p)$

P_i' = ความชันของฟังก์ชันการตอบสนองข้อสอบ ณ ระดับความสามารถ θ

3) ค่าฟังก์ชันสารสนเทศของแบบสอบ (test information function : $I(\theta)$)

$$I(\theta) = \sum_{i=1}^{k} I_i(\theta)$$

4) ความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถ (standard error of estimates at ability θ : $SE(\theta)$)

$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}}$$

ถ้าฟังก์ชันสารสนเทศของแบบสอบมีค่าสูง ค่าความคลาดเคลื่อนมาตรฐานก็จะมีค่าต่ำ

5) ค่าดัชนีประสิทธิภาพสัมพัทธ์ (Relative Efficiency : $RE(\theta)$) ของแบบสอบ 2 ฉบับ ณ ระดับความสามรถต่างๆ ดัชนีนี้ใช้ในการเปรียบเทียบค่าฟังก์ชันสารสนเทศของแบบสอบ (Hambleton et al, 1991)

$$RE(\theta) = \frac{I_A(\theta)}{I_B(\theta)}$$

เมื่อ $I_A(\theta)$ = ค่าฟังก์ชันสารสนเทศของแบบสอบฉบับ A ผู้สอบความสามารถระดับ θ

$I_B(\theta)$ = ค่าฟังก์ชันสารสนเทศของแบบสอบฉบับ B ผู้สอบความสามารถระดับ θ

การแปลความหมายของค่าดัชนีประสิทธิภาพสัมพัทธ์

1. $RE(\theta) = 1$ แสดงว่าแบบสอบทั้งสองฉบับมีประสิทธิภาพเท่ากัน
2. $RE(\theta) > 1$ แสดงว่าแบบสอบฉบับ A มีประสิทธิภาพมากกว่าแบบสอบฉบับ B
3. $RE(\theta) < 1$ แสดงว่าแบบสอบฉบับ A มีประสิทธิภาพต่ำกว่าแบบสอบฉบับ B

3.1.4 ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบมีข้อตกลงเบื้องต้นที่สำคัญ ดังนี้ (ศิริชัย กาญจนวาสี , 2541)

1) แบบสอบวัดคุณลักษณะเด่นเพียงลักษณะเดียว (unidimensionality : one trait) คุณลักษณะหรือความสามารถที่เป็นตัวกำหนดพฤติกรรมการตอบข้อสอบแต่ละข้อ (ตอบถูกหรือตอบผิด) มีลักษณะเด่นหรือลักษณะสำคัญเพียงลักษณะเดียว นั่นคือ แบบสอบวัดคุณลักษณะสำคัญเพียงลักษณะเดียวหรือแบบสอบมีความเป็นเอกพันธ์ การตรวจสอบความเป็นไปตามข้อตกลงเบื้องต้นข้อนี้สามารถใช้ข้อมูลสนับสนุนที่ได้จากการวิเคราะห์ตัวประกอบ

2) การตอบข้อสอบเป็นไปอย่างอิสระ (independence : local independence) การตอบข้อสอบแต่ละข้อของผู้ตอบแต่ละคนมีความเป็นอิสระจากกัน ซึ่งประกอบด้วย

2.1) ความเป็นอิสระระหว่างข้อสอบ หมายความว่า ข้อสอบแต่ละข้อเป็นอิสระจากกัน กล่าวคือ การตอบสนองข้อสอบข้อหนึ่งไม่มีผลต่อการตอบข้อสอบข้ออื่น ๆ ในแบบสอบ เช่น เนื้อหาของคำถามข้อหนึ่งจะต้องไม่จบของข้ออื่นและตำแหน่งข้อสอบแต่ละข้อจะอยู่ที่ใดก็ได้โดยไม่มีผลต่อการตอบข้อสอบ

นั่นคือ ถ้า $P_i(\theta_A)$ เป็นอิสระจาก $P_j(\theta) \Rightarrow P_i(\theta) \cap P_j(\theta) = P_i(\theta) \cdot P_j(\theta)$

2.2) ความเป็นอิสระระหว่างผู้สอบ หมายความว่า ผู้สอบแต่ละคนตอบข้อสอบแต่ละข้ออย่างเป็นอิสระจากกัน กล่าวคือผลการตอบสนองข้อสอบของผู้สอบคนหนึ่งจะต้องไม่มีผลต่อการตอบสนองข้อสอบของผู้สอบคนอื่น

นั่นคือ ถ้า $P_i(\theta_A)$ เป็นอิสระจาก $P_i(\theta_B) \Rightarrow P_i(\theta_A) \cap P_i(\theta_B) = P_i(\theta_A) \cdot P_i(\theta_B)$

2.3) โค้งคุณลักษณะข้อสอบสามารถใช้อธิบายพฤติกรรมการตอบสนองข้อสอบ (Item characteristic curves : item response models) ความสัมพันธ์ระหว่างระดับความสามารถของผู้ตอบ (θ_p) กับโอกาสที่จะตอบข้อสอบถูก [$P_i(\theta)$] สามารถแสดงได้ด้วยโค้งคุณลักษณะข้อสอบ (ICC_s) โดย ICC_s เป็นฟังก์ชันคณิตศาสตร์ที่แสดงด้วยโมเดลการตอบสนองข้อสอบที่นิยมใช้กันมี 3 รูปแบบคือ 1, 2 หรือ 3 พารามิเตอร์ ในการเลือกใช้ขึ้นอยู่กับลักษณะข้อมูลและความเชื่อในโมเดลที่ใช้อธิบายความสัมพันธ์ดังกล่าว

2.4) ข้อสอบที่ใช้จะต้องไม่เป็นข้อสอบประเภทความเร็ว ผู้สอบทุกคนควรมีโอกาสทำข้อสอบทุกข้อ เพื่อให้ได้คะแนนรวมจากการสอบประกอบกับลักษณะของข้อสอบเป็นตัวประมาณค่าความสามารถของผู้สอบโดยไม่มีข้อจำกัดเกี่ยวกับเวลาที่ใช้ในการสอบ

3.2 กระบวนการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ (DFIT Procedure)

Raju และคณะ (1995) ได้เสนอกระบวนการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ (differential functioning of items and test procedure : DFIT) ซึ่งเป็นวิธีการที่อยู่บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบเป็นวิธีที่ละเอียด มีการคำนวณซับซ้อนแต่ให้ผลถูกต้อง วิธีการนี้เป็นวิธีที่ไม่ใช้เกณฑ์ภายนอก (absence of criterion) โดยการนำวิธีทางสถิติมาตรวจสอบการทำหน้าที่ต่างกันโดยพิจารณาโครงสร้างภายในแบบสอบ

กระบวนการ DFIT ประกอบด้วยดัชนีที่ใช้วัดการทำหน้าที่ต่างกัน 2 ระดับ ดังนี้

1. ดัชนีการทำหน้าที่ต่างกันระดับแบบสอบ (DTF index) เป็นดัชนีที่บอกขนาดของผลการ

ทำหน้าที่ต่างกันสุทธิ (net effect) ในแบบสอบทั้งฉบับ

2. ดัชนีการทำหน้าที่ต่างกันระดับข้อสอบ (DIF index) มี 2 ดัชนี คือ

1.1 ดัชนีการทำหน้าที่ต่างกันของข้อสอบแบบชดเชย (compensatory DIF index : CDIF) ซึ่งเป็นดัชนีที่ใช้ในการหาค่า DTF ของแบบสอบทั้งฉบับ ข้อแตกต่างของดัชนีนี้กับวิธีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบวิธีอื่นที่รู้จักกันดี กล่าวคือ ในการวิเคราะห์ดัชนี CDIF สำหรับข้อสอบข้อใดข้อหนึ่ง จะไม่สมมติว่าข้อสอบข้ออื่นในแบบสอบจะไม่มีความลำเอียง นั่นคือพิจารณาความลำเอียงของข้ออื่นร่วมด้วยกับข้อที่พิจารณาด้วย

1.2 ดัชนีการทำหน้าที่ต่างกันของข้อสอบแบบไม่ชดเชย (noncompensatory DIF index : NCDIF) ในการวิเคราะห์ดัชนี NCDIF จะสมมติว่าข้อสอบข้ออื่นในแบบสอบที่ไม่ได้พิจารณาไม่มีความลำเอียงอย่างสมบูรณ์ นั่นคือสนใจเฉพาะความลำเอียงที่เกิดขึ้นของ ข้อที่พิจารณาเท่านั้น

การทำหน้าที่ต่างกันของแบบสอบ (Differential Test Functioning : DTF)

ให้ $P_i(\theta_s)$ แทนความน่าจะเป็นของการตอบข้อสอบถูกของผู้สอบ s ซึ่งมีระดับความสามารถแฝง θ ที่ข้อสอบข้อที่ i

F แทน ผู้สอบกลุ่มเปรียบเทียบเป็นกลุ่มที่คาดว่าจะเสียประโยชน์ในการตอบข้อสอบหรือเป็นกลุ่มที่สนใจ

R แทน ผู้สอบกลุ่มอ้างอิงเป็นกลุ่มที่คาดว่าจะได้ประโยชน์ในการตอบข้อสอบ

$P_{iF}(\theta_s)$ และ $P_{iR}(\theta_s)$ แทน ความน่าจะเป็นของการตอบข้อสอบถูกของผู้สอบ s ซึ่งเป็นสมาชิกของกลุ่มเปรียบเทียบและกลุ่มอ้างอิง ที่ข้อสอบข้อที่ i ตามลำดับ

ถ้าข้อสอบทำหน้าที่ต่างกันในกลุ่ม 2 กลุ่ม P_{iF} และ P_{iR} ของผู้สอบแต่ละคนจะแตกต่างกัน สัดส่วนของการตอบข้อสอบถูกของผู้สอบ (an examinee's expected proportion : EPC) หรือคะแนนจริง (T) มีค่าเท่ากับ

$$T_s = \sum_{i=1}^{n} P_i(\theta_s)$$

ดังนั้นการวัดการทำหน้าที่ต่างกันของแบบสอบ (DTF) ที่ระดับผู้สอบ (examinee level) จะหาค่าได้จาก

$$DTF = \in (T_{sR} - T_{sF})^2$$

ถ้าให้ $D_s = T_{sR} - T_{sF}$ แล้วจะได้ว่า

$$DTF = \in D_s^2 = \int D_s^2 f_F(\theta) d\theta = \sigma_D^2 + (\mu_{TR} - \mu_{TF})^2 = \sigma_D^2 + \mu_D^2 \quad .....1)$$

โดยที่ $f_F(\theta)$ แทน ฟังก์ชันความหนาแน่น (density function) ของกลุ่มเปรียบเทียบ

μ_{TF} และ μ_{TR} แทน ค่าเฉลี่ยของคะแนนจริงของผู้สอบในกลุ่มเปรียบเทียบและกลุ่มอ้างอิงตามลำดับ

การทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning : DIF)

Compensatory DIF index (CDIF) เป็นดัชนีการทำหน้าที่ต่างกันของข้อสอบแบบชดเชย

$$CDIF_i = \in (d_i D) = Cov(d_i, D) + \mu_{di}\mu_D \quad ..... 2)$$

โดยที่ $CDIF_i$ แทน ดัชนีบ่งชี้การทำหน้าที่ต่างกันของข้อสอบแบบชดเชยของข้อสอบข้อที่ i

d_i แทน ความแตกต่างของความน่าจะเป็นของการตอบข้อสอบข้อที่ i ถูกของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบมีค่าเท่ากับ $d_i = P_{iR}(\theta) - P_{iF}(\theta)$

$Cov(d_i, D)$ แทน ความแปรปรวนร่วมระหว่างความแตกต่างของความน่าจะเป็นของการตอบข้อสอบข้อที่ i ถูก กับความแตกต่างระหว่างคะแนนจริงของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบจากสมการที่ 1) และ 2) จะได้ว่า

$$DTF = \in\left[\left(\sum_{i=1}^{n} d_{is}\right)^2\right] \quad , d_{is} = P_{iR}(\theta_s) - P_{iF}(\theta_s) \qquad \text{..... 3)}$$

ดังนั้น

$$DTF = \sum_{i=1}^{n} CDIF_i$$

จากสมการที่ 3) สามารถเขียนได้เป็น

$$DTF = \in\left[\left(\sum_{i=1}^{n} d_{is}\right)^2\right] = \in\left[\sum_{i=1}^{n}(P_{iF} - P_{iR})\right]^2 = \in\left[(P_{1F} - P_{1R}) + (P_{2F} - P_{2R}) + ... + (P_{nF} - P_{nF})\right]^2$$

จะเห็นได้ว่าดัชนีการทำหน้าที่ต่างกันของแบบสอบ คือผลรวมของความลำเอียงระหว่างข้อสอบที่ระดับผู้สอบ

ความแตกต่างของนิยาม CDIF ในสมการที่ 2) กับนิยามอื่นๆ คือ นิยาม DIF ที่รู้จักโดยทั่วไปจะพิจารณาเฉพาะดัชนีที่ระดับข้อสอบเท่านั้น ไม่ได้คำนึงถึงความสัมพันธ์ของ DIF หรือ ความลำเอียงที่อาจมีอยู่ในกลุ่มของข้อสอบ

Noncompensatory DIF index (NCDIF) เป็นดัชนีของการทำหน้าที่ต่างกันของข้อสอบแบบไม่ชดเชย

ถ้าสมมติว่าข้อกระทงทุกข้อในแบบสอบยกเว้นข้อที่ i ไม่มีความลำเอียงอย่างสมบูรณ์ แล้ว $d_j = 0$ สำหรับทุก j ที่ $j \neq i$ ดังนั้น

$$NCDIF_i = \sigma_{di}^2 + \mu_{di}^2 \qquad \text{.....4)}$$

จากสมการที่ 4) จะได้ว่า

1) NCDIF = 0 ก็ต่อเมื่อ ค่าพารามิเตอร์ของข้อสอบข้อที่ i มีค่าเท่ากันทั้งในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ

2) ถ้าให้ $f_F(\theta)$ แทน ค่าฟังก์ชันความสามารถในกลุ่มเปรียบเทียบ แล้ว

$$NCDIF_i = \int_{-\infty}^{\infty}\left[P_{iR}(\theta) - P_{iF}(\theta)\right]^2 f_F(\theta)d\theta \quad \text{..... 5)}$$

3) จากสมการที่ 5) เขียนได้เป็น

$$NCDIF_i = \int_{-\infty}^{\infty}\left|P_{iR}(\theta) - P_{iF}(\theta)\right|^2 f_F(\theta)d\theta \quad . \quad \text{..... 6)}$$

ซึ่งสามารถเขียนเป็นอสมการ Cauchy-Schwartz ได้ดังนี้

$$NCDIF_i \geq [\int_{-\infty}^{\infty}\left|P_{iR}(\theta) - P_{iF}(\theta)\right| f_F(\theta)d\theta]^2 \quad \text{.....7)}$$

ถ้า $f_F(\theta)$ เป็นสี่เหลี่ยม (rectangular) สมการที่ 7) คือค่าสัมบูรณ์ยกกำลังสองหรือพื้นที่แบบไม่คิดเครื่องหมายระหว่างโค้งการตอบข้อสอบ (IRFs) 2 โค้ง จะเห็นว่าดัชนี NCDIF เป็นดัชนีที่สัมพันธ์กับวิธีการเดิมที่รู้จักกันอย่างแพร่หลายในการประเมินการทำหน้าที่ต่างกันของข้อสอบ ที่อยู่บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ ซึ่งมีข้อตกลงว่าข้อสอบทุกข้อในแบบสอบยกเว้นข้อที่ทำการพิจารณา ทำหน้าที่ไม่ต่างกัน ซึ่งข้อตกลงนี้ยังไม่เป็นที่แน่ใจในการพัฒนาแบบสอบส่วนใหญ่

NCDIF เป็นดัชนีแบบไม่ชดเชยเพราะว่าค่าของ NCDIF มีค่ามากกว่าหรือเท่ากับ 0 เท่านั้น ดังนั้นจึงไม่สามารถหักล้างหรือชดเชยระหว่างข้อสอบข้ออื่นได้

ทั้ง DTF CDIF และ NCDIF มีประโยชน์ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบแตกต่างกัน ขึ้นอยู่กับวัตถุประสงค์ของการนำไปใช้ ดังนี้

1 ถ้าคะแนนรวมจากแบบสอบใช้ในการตัดสินประสิทธิภาพของเครื่องมือหรือใช้ในการจัดลำดับหรือ/และ คัดเลือก DTF จะมีประโยชน์มากกว่า CDIF และ NCDIF

2. ถ้าผู้พัฒนาแบบสอบจำเป็นจะต้องรวมข้อสอบที่ทำหน้าที่ต่างกันเข้าไปในแบบสอบ ซึ่งข้อสอบบางข้อเข้าข้างกลุ่มเปรียบเทียบและบางข้อเข้าข้างกลุ่มอ้างอิง ผลของการทำหน้าที่ต่างกันของแบบสอบ (DTF) จะต้องมีดุลยภาพระหว่างข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบที่เข้าข้างกลุ่มใดกลุ่มหนึ่ง เพราะว่าโดยทั่วๆ ไปแล้ว เป็นการยากที่จะคัดเลือกเอาข้อสอบที่ทำหน้าที่ต่างกันทั้งหมดออกจากแบบสอบ

การทดสอบความมีนัยสำคัญของ ดัชนี DTF (DTF index) และ ดัชนี NCDIF (NCDIF index)

ให้ $D_s = \sum_{i=1}^{n} d_{is}$ และ $d_{is} = P_{iR} - P_{iF}$

โดยที่ P_{iF}, P_{iR} เป็นความน่าจะเป็นในการตอบข้อสอบถูก ซึ่งประมาณมาจากค่าพารามิเตอร์ของข้อสอบและผู้สอบ

1. การทดสอบความมีนัยสำคัญ สำหรับดัชนี DTF

สมมติว่า D มีการแจกแจงเป็นโค้งปกติที่มีค่าเฉลี่ยเป็น μ_D และส่วนเบี่ยงเบนมาตรฐาน σ_D สำหรับผู้สอบ s ดังนั้นคะแนนซี (Z score) ของผู้สอบ s คือ

$$Z_s = \frac{D_s - \mu_D}{\sigma_D} \qquad .....8)$$

โดยที่ Z_s^2 มีการกระจายแบบไค-สแควร์ ที่มีองศาอิสระเท่ากับ 1 ผลรวมของค่า Z_s^2 ของผู้สอบแต่ละคนในกลุ่มเปรียบเทียบ ก็จะมีการกระจายแบบไค-สแควร์ ที่มีองศาอิสระเท่ากับ N_F โดยที่ N_F คือ จำนวนผู้สอบในกลุ่มเปรียบเทียบ สามารถแสดงได้ดังสมการ

$$\chi^2_{N_F} = \sum_{s=1}^{N_F} Z_s^2 = \frac{\sum_{s=1}^{N_F} (D_s - \mu_D)^2}{\sigma_D^2} \qquad .....9)$$

ซึ่งในการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบนี้ จะสนใจในกรณีที่ดัชนี DTF มีค่าน้อยที่สุด หรือมีค่าเป็น 0 ดังสมการ

$$\in (DTF) = \mu_D^2 = 0 \qquad .....10)$$

แล้วสมการที่ 9 สามารถเขียนได้เป็น

$$\chi^2_{N_F} = \frac{\sum_{s=1}^{N_F} D_s^2}{\sigma_D^2} \qquad .....11)$$

ที่มีองศาอิสระ เท่ากับ N_F

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

จากนิยามของ DTF จะได้ว่า

$$\chi^2_{N_F} = \frac{N_F(DTF)}{\sigma^2_D} \qquad .....12)$$

ในการประมาณค่าความแปรปรวนของค่า D จากกลุ่มตัวอย่าง จากสมการที่ 12 สามารถเขียนได้เป็น

$$\chi^2_{N_F} = \frac{N_F(DTF)}{\hat{\sigma}^2_D} \qquad .....13)$$

ถ้าค่าไค-สแควร์ มีนัยสำคัญแสดงว่า มีข้อสอบอย่างน้อย 1 ข้อ ที่ทำหน้าที่ต่างกัน นอกจากการทดสอบไค-สแควร์ แล้วยังมีสถิติทดสอบที (t-test) ที่สามารถใช้ในการทดสอบการทำหน้าที่ต่างกันของแบบสอบได้ สามารถแสดงได้ดังสมการที่ 14

$$t = \frac{(N_F)^{1/2}(\hat{\mu}_D - \mu_D)}{\hat{\sigma}_D} \qquad .....14)$$

ภายใต้สมมติฐานหลักที่ว่า $\mu_D = 0$ สมการที่ 14 สามารถเขียนได้เป็น

$$t = \frac{(N_F)^{1/2}(\hat{\mu}_D)}{\hat{\sigma}_D} \qquad .....15)$$

โดยมีองศาความเป็นอิสระเท่ากับ N_{F-1} เนื่องจากว่าขนาดกลุ่มตัวอย่างที่ใช้ในการวิเคราะห์ IRT มีขนาดใหญ่ สถิติทดสอบทีและไค-สแควร์ จะมีให้ผลสรุปในลักษณะที่คล้ายคลึงกันมาก

เมื่อผลการทดสอบดัชนี DTF พบว่ามีนัยสำคัญทางสถิติ ให้ตัดข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด โดยพิจารณาจากค่าดัชนี CDIF ที่มีค่าเป็นบวกสูงที่สุด หลังจากตัดข้อสอบข้อนั้นออกจากแบบสอบแล้วจึงทำการคำนวณค่าดัชนี DTF และค่าไค-สแควร์อีกครั้งจากข้อสอบที่เหลืออยู่ เนื่องจากค่าของ $Cov(d_i, D)$ ขึ้นอยู่กับข้อสอบทั้งหมดที่มีอยู่ในแบบสอบ ดำเนินการตามกระบวนการดังกล่าวจนค่าไค-สแควร์ที่คำนวณได้ไม่มีนัยสำคัญ ข้อสอบทั้งหมดที่ถูกตัดออก จะถูกระบุให้เป็นข้อสอบที่ทำหน้าที่ต่างกันและค่าดัชนี CDIF ของข้อสอบนั้น ๆ แต่ละข้อมีนัยสำคัญทางสถิติ จะไม่มีสถิติทดสอบความมีนัยสำคัญของดัชนี CDIF โดยเฉพาะ

2. การทดสอบความมีนัยสำคัญ สำหรับดัชนี NCDIF

สมมติว่า d_i มีการแจกแจงเป็นโค้งปกติที่มีค่าเฉลี่ยเป็น μ_{d_i} และมีค่าความแปรปรวนเป็น $\sigma^2_{d_i}$ สำหรับผู้สอบ s N_F คือ จำนวนผู้สอบในกลุ่มเปรียบเทียบ

$$\chi^2_{N_F} = \frac{N_F(NCDIF)}{\sigma^2_{d_i}} \qquad .....16)$$

โดยมีองศาความเป็นอิสระ เท่ากับ N_F

และการทดสอบที (t-test) สำหรับดัชนี NCDIF แสดงได้ดังสมการที่ 17

$$t = \frac{(N_F)^{1/2}(\hat{\mu}_{\hat{d}_i})}{\hat{\sigma}_{\hat{d}_i}} \qquad .....17)$$

โดยมีองศาความเป็นอิสระ เท่ากับ N_{F-1}

จากผลการวิจัยของ Fleer (1993, อ้างถึงใน Raju et al.,1995) ในการทดสอบความมีนัยสำคัญของดัชนี DTF และ NCDIF โดยใช้การทดสอบไค-แสควร์ พบว่าดัชนีเหล่านี้จะมีความไวมากเมื่อใช้กลุ่มตัวอย่างที่มีขนาดใหญ่ และเมื่อใช้จุดตัดมีค่าเท่ากับ 0.006 ทั้งดัชนี DTF และ NCDIF จะมีความผิดพลาดในการระบุข้อสอบที่ทำหน้าที่ต่างกันประมาณร้อยละ 1 เท่านั้น ดังนั้นเกณฑ์ที่ใช้ในการตัดสินว่าแบบสอบจะไม่ทำหน้าที่ต่างกัน เมื่อค่าดัชนี DTF $\leq$ 0.006 หรือไม่มีนัยสำคัญทางสถิติเมื่อทดสอบด้วยไค-แสควร์ และสำหรับดัชนี NCDIF ข้อสอบข้อใดที่มีค่าดัชนีนี้ มากกว่า 0.006 หรือมีนัยสำคัญทางสถิติเมื่อทดสอบด้วยไค-แสควร์ แสดงว่าข้อสอบข้อนั้นทำหน้าที่ต่างกัน

นอกจากกระบวนการ DFIT ที่สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ สำหรับแบบสอบชนิดเอกมิติ (unidimensional) ที่มีการให้คะแนนแบบทวิวิภาคหรือแบบสองค่า (dichotomous) ซึ่ง Raju และคณะ เสนอในปี ค.ศ. 1995 นี้แล้ว กระบวนการนี้ยังได้รับการพัฒนาให้มีความสามารถในการตรวจสอบข้อสอบและแบบสอบที่มีลักษณะอื่น ๆ อีก ดังนี้ ในปี ค.ศ. 1997 Oshima และคณะ ได้พัฒนากระบวนการนี้ให้สามารถตรวจสอบแบบสอบชนิดพหุมิติ (Multidimensional) ที่ข้อสอบมีการให้คะแนนแบบทวิวิภาค หลังจากนั้น Flowers และคณะ ได้พัฒนากระบวนการ DFIT นี้ให้สามารถวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบพหุวิภาคหรือให้คะแนนแบบหลายค่า (polytomous) ในแบบสอบชนิดเอกมิติอีกด้วย

ตอนที่ 4 เอกสารและงานวิจัยที่เกี่ยวข้อง

4.1 งานวิจัยในประเทศ

ในช่วงระยะเวลาที่ผ่านมา มีงานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบหลายงานวิจัยโดยใช้ข้อมูลจำลองและข้อมูลเชิงประจักษ์ ซึ่งใช้วิธีการตรวจค้นการทำหน้าที่ต่างกันของข้อสอบที่แตกต่างกัน แต่ยังไม่พบงานวิจัยในประเทศที่ทำการตรวจสอบการทำหน้าที่ต่างกันของแบบสอบ สำหรับงานวิจัยที่เกี่ยวข้องมีดังนี้

ชัชชัย เผ่าพงษ์ (2526 , อ้างถึงใน ทัศนีย์ พีรมนตรี , 2529) ได้ทำการศึกษาความลำเอียงของแบบสอบวัดความถนัดทางการเรียนด้านคณิตศาสตร์และภาษาในระดับมัธยมศึกษาตอนต้น ซึ่งพัฒนาขึ้นโดยสำนักทดสอบทางการศึกษาและจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร กลุ่มตัวอย่างที่ใช้เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2524 ทั่วทุกภาคของประเทศไทย ได้แก่ภาคเหนือ ภาคกลาง ภาคใต้ ภาคตะวันออก และภาคตะวันออกเฉียงเหนือ ทำการวิเคราะห์ระหว่างกลุ่มผู้สอบเพศชายและเพศหญิง โดยใช้วิธีวิเคราะห์ด้วยโค้งลักษณะของข้อสอบ 3 พารามิเตอร์

ผลการวิจัยพบว่าแบบทดสอบวัดความถนัดทางการเรียนด้านคณิตศาสตร์จะมีความลำเอียงในเรื่องร้อยละ การหาปริมาตรและการหาความยาวของเส้นด้านในของรูปสามเหลี่ยมและเนื้อหาเกี่ยวกับโจทย์ปัญหาทางคณิตศาสตร์ ซึ่งข้อสอบส่วนมากจะลำเอียงต่อเพศชาย ส่วนแบบสอบวัดความถนัดด้านภาษาเกี่ยวกับการอ่านเข้าใจ จะมีความลำเอียงในเรื่อง ความเข้าใจในการอ่านคำประพันธ์และบทร้อยกรอง อย่างละ 1 ข้อ นอกจากนั้นเป็นข้อสอบที่ลำเอียงที่มีเนื้อหาที่เกี่ยวกับความเข้าใจในการอ่าน ซึ่งข้อสอบส่วนใหญ่ที่มีความลำเอียง จะลำเอียงต่อกลุ่มนักเรียนหญิง

ทัศนีย์ พีรมนตรี (2529) ได้ทำการวิเคราะห์ความลำเอียงของแบบสอบวิชาคณิตศาสตร์ ในโครงการตรวจสอบคุณภาพการศึกษา ซึ่งพัฒนาแบบสอบโดยสำนักงานทดสอบทางการศึกษา กรมวิชาการ โดยใช้วิธีวิเคราะห์ 3 วิธี คือ วิธีกำหนดจุดค่าเดลต้า วิธีทดสอบความแตกต่างระหว่างกลุ่มด้วยสถิติไค-สแควร์ ในโมเดลลอกลิเนีย 2 โมเดล คือโมเดลที่ไม่มีพารามิเตอร์ผลร่วมระหว่างระดับคะแนนกับกลุ่ม และโมเดลที่ไม่มีพารามิเตอร์ของผลหลักที่เกิดจากกลุ่ม และวิธีการตอบสนองข้อสอบ 3 พารามิเตอร์ โดยกลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526 ทั่วทุกภาคของประเทศ ได้แก่ ภาคกลาง ภาคเหนือ ภาคใต้ ภาคตะวันออก และภาคตะวันออกเฉียงเหนือ

ผลการศึกษาพบว่าวิธีการตอบสนองข้อสอบ 3 พารามิเตอร์ พบข้อสอบที่มีความลำเอียงในแต่ละภาคจำนวนข้อมากที่สุด วิธีกำหนดจุดค่าเดลต้าและวิธีการตอบสนองข้อสอบ 3 พารามิเตอร์ มี

ข้อลำเอียงซ้ำกันมากที่สุดระหว่างกรุงเทพมหานครและภาคตะวันออกเฉียงเหนือ ส่วนวิธีไค-สแควร์มีจำนวนข้อที่ซ้ำกันมากที่สุดระหว่างกรุงเทพมหานครและภาคตะวันออก

สุรศักดิ์ อมรรัตนศักดิ์ (2531) ได้ทำการศึกษาเปรียบเทียบผลของวิธีวิเคราะห์ความลำเอียงของข้อสอบ 4 วิธี คือ วิธีวิเคราะห์ความแปรปรวน วิธีแปลงค่าความยากของข้อสอบและวิธีโค้งลักษณะข้อสอบ ที่มีพารามิเตอร์ 2 ตัวและ 3 ตัว โดยใช้ข้อมูลจากแบบสอบแข่งขันเพื่อบรรจุเข้ารับราชการเป็นข้าราชการครู สังกัดสำนักงานคณะกรรมการประถมศึกษาแห่งชาติ ปี พ.ศ. 2529 กลุ่มตัวอย่างคือผู้เข้าสอบแข่งขัน ตัวแปรที่ศึกษาคือ ตัวแปรเพศ โดยทำการเปรียบเทียบความแตกต่างของผลการคัดเลือกก่อนและหลังการศึกษาความลำเอียงของข้อสอบในด้าน จำนวนผู้ที่ได้รับการคัดเลือก สัดส่วนของชาย : หญิง ที่ได้รับการคัดเลือกและความเที่ยงของแบบสอบ

ผลการวิจัยพบว่า วิธีวิเคราะห์ความลำเอียงทั้ง 4 วิธี พบข้อสอบที่มีความลำเอียงแตกต่างกัน โดยวิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 3 ตัว พบข้อสอบที่ลำเอียงจำนวนมากที่สุด รองลงมาคือ วิธีวิเคราะห์ความแปรปรวน และวิธีที่พบข้อสอบที่ลำเอียงจำนวนน้อยที่สุดคือ วิธีแปลงค่าความยากของข้อสอบ วิธีทั้ง 4 วิธี มีความสัมพันธ์กันทางบวกอย่างมีนัยสำคัญที่ระดับ 0.001 และเมื่อใช้การคิดคะแนนดิบและคะแนนรวมที่แตกต่างกัน 5 วิธี พบว่ามีจำนวนผู้ที่ได้รับการคัดเลือกแตกต่างกันประมาณร้อยละ 4-24 และเมื่อใช้คะแนนมาตรฐานทีปกติรวมกับคะแนนแปลงแบบอื่น ๆ 4 วิธี พบว่าจำนวนผู้ที่ได้รับการคัดเลือกแตกต่างกันร้อยละ 4-23 และเมื่อตัดข้อสอบที่มีความลำเอียงออกแล้ว พบว่าจำนวนหญิงและชายที่ได้รับการคัดเลือกมีจำนวนใกล้เคียงกันมากขึ้นกว่าก่อนตัดข้อสอบที่มีความลำเอียง และค่าความเที่ยงของแบบสอบหลังการตัดข้อสอบที่ลำเอียงออกจากแบบสอบ ลดลงเล็กน้อย

สุพัฒน์ สุกมลสันต์ (2534) ได้ทำการศึกษาความลำเอียงของข้อสอบภาษาอังกฤษ ชุด กข และ กขค เข้ามหาวิทยาลัย ปี พ.ศ. 2531-2533 จำนวน 6 ฉบับ ฉบับละ 100 ข้อ โดยวิเคราะห์ความลำเอียงต่อเพศและต่อภาคภูมิศาสตร์ของผู้สมัครสอบ เปรียบเทียบวิธีวิเคราะห์ 3 วิธี คือ วิธีกำหนดจุดค่า เดลต้า วิธีไค-สแคว์ชนิดแบ่งความสามารถของผู้สอบเป็น 3 ระดับ ได้แก่ กลุ่มผู้ที่มีความสามารถระดับต่ำ (ผู้ที่ได้คะแนน 0 – 40 คะแนน) กลุ่มผู้ที่มีความสามารถระดับปานกลาง (ผู้ที่ได้คะแนนรวม 41 – 70 คะแนน) และระดับสูง (ผู้ที่ได้คะแนนรวม 71 – 100 คะแนน) และวิธีการวัดพื้นที่ความแตกต่างระหว่างโค้งลักษณะข้อสอบที่วิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์

กลุ่มตัวอย่างที่ใช้คือ ผู้สมัครสอบคัดเลือกเข้ามหาวิทยาลัย ซึ่งสอบแบบสอบภาษาอังกฤษเข้ามหาวิทยาลัย ชุด กข และ ชุด กขค ตัวแปรที่ศึกษาความลำเอียงคือ เพศ และภาคภูมิศาสตร์ของผู้

สอบ ซึ่งแยกออกออกเป็น 5 ภาคตามภูมิลำเนาของผู้สอบ ได้แก่ ภาคกลาง ภาคตะวันออก ภาคตะวันออกเฉียงเหนือ ภาคเหนือ และภาคใต้

ผลการวิจัยพบว่าแบบสอบวิชาภาษาอังกฤษ ชุด กข และชุด กขค ปี 2531-2533 มีความลำเอียงต่อเพศ โดยเฉลี่ยประมาณฉบับละ ร้อยละ 7-28และ 4-41 ตามลำดับและมีความลำเอียงต่อภาคภูมิศาสตร์โดยเฉลี่ยประมาณฉบับละ ร้อยละ 6-45และ 5-43 ตามลำดับ วิธีการวิเคราะห์ทั้ง 3 วิธี พบข้อสอบที่มีความลำเอียงต่อเพศและต่อภาคภูมิศาสตร์ของผู้สอบจำนวนแตกต่างกันอย่างมีนัยสำคัญ แต่จำนวนข้อที่ลำเอียงแต่ละวิธีมีความสัมพันธ์กันอย่างไม่มีนัยสำคัญ วิธีโลจิสติก 3 พารามิเตอร์จะพบข้อสอบที่ทำหน้าที่ต่างกันจำนวนมากที่สุด รองลงมาคือวิธีไค-สแควร์และวิธีกำหนดจุดเดลต้าพบข้อสอบที่ลำเอียงจำนวนน้อยที่สุด

กาญจนา วัธนสุนทร (2537) ได้พัฒนาเกณฑ์การตัดสินข้อสอบลำเอียงทางเพศด้วยข้อมูลเชิงประจักษ์ ด้วย 3 วิธีการตรวจสอบคือ วิธีทฤษฎีการตอบสนองข้อสอบโมเดล 2 พารามิเตอร์ วิธีแมนเทล-แฮนส์เซล ดัชนีที่พัฒนาเพื่อเกณฑ์ในการตัดสินข้อสอบลำเอียงคือ SA , UA , α_{MH} , β_{SIB} ตามลำดับ โดยใช้แบบสอบวิชาภาษาอังกฤษความยาว 50 60 70 และ 80 และแบบสอบวิชาคณิตศาสตร์ความยาว 20 30 และ 40 กับกลุ่มตัวอย่าง 6 ขนาด คือ 100 200 400 600 800 และ 1,000 คน

ผลการวิจัยพบว่าเกณฑ์ที่พัฒนาจากข้อมูลเชิงประจักษ์ เพื่อใช้เป็นเกณฑ์ในการตัดสินความลำเอียงของข้อสอบระหว่างเพศชายและหญิง คือ

1) $|SA| > .80$ และ $UA > .50$ กรณีความยาวแบบสอบต่ำกว่า 50 ข้อ และ
2) $|SA| > .40$ และ $UA > 1.20$ กรณีความยาวแบบสอบ 50 ข้อขึ้นไป
3) $\alpha_{MH} < .60$ และ $\alpha_{MH} > 1.40$ สำหรับทุกขนาดของผู้สอบและความยาวแบบสอบ
4) $\beta_{SIB} > .06$ สำหรับทุกขนาดของผู้สอบและความยาวแบบสอบ

การใช้ดัชนี SA หรือ UA ควรใช้ผู้สอบขนาด 800 คนขึ้นไป ส่วนดัชนี α_{MH} และ β_{SIB} ควรใช้ขนาดผู้สอบอย่างน้อย 600 คนขึ้นไป

จิติมา วรรณศรี (2539) ได้เปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีแมนเทล-แฮนส์เซลกับวิธีซิบเทสท์ โดยศึกษาจากข้อมูลจำลอง ปัจจัยที่ศึกษาได้แก่ความยาวของแบบสอบ 3 ขนาด คือ 30 60 และ 90 ข้อ ขนาดกลุ่มตัวอย่าง 3 ขนาด คือ 200 600 และ 1,000 คน โดยแต่ละขนาดมีอัตราส่วนระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบแตกต่างกัน คือ 1:1 1:0.9 1:0.75 และ 1:0.5

ผลการศึกษาพบว่าวิธีแมนเทล-แฮนส์เซลกับวิธีซิบเทสท์ มีประสิทธิภาพเท่าเทียมกันในการตรวจสอบการทำหน้าที่ต่างกันที่ทุกขนาดกลุ่มตัวอย่างและอัตราส่วนภายใต้ความยาวแบบสอบเดียวกันและเมื่อใช้แบบสอบที่มีความยาวปานกลาง (60 ข้อ) ทั้งสองวิธีมีประสิทธิภาพที่สุด นอกจากนี้เมื่อใช้ขนาดกลุ่มตัวอย่างมากขึ้นจะสามารถตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันได้ถูกต้องมากขึ้น โดยส่วนมากวิธีซิบเทสท์มีอัตราความคลาดเคลื่อนประเภทมี่ 1 มากกว่าวิธีแมนเทล-แฮนส์เซล เล็กน้อย

ปรมินทร์ อริเดช (2539) ได้ทำการศึกษาการใช้ GRM , GPCM และโมเดลโลจิสติก ในการเปรียบเทียบฟังก์ชันสารสนเทศของมาตรประมาณที่มีวิธีการตรวจให้คะแนนแบบทวิวิภาคและแบบพหุวิภาค โดยการตรวจให้คะแนนแบบทวิวิภาควิเคราะห์ตามโมเดลโลจิสติก 1 , 2 และ 3 พารามิเตอร์ส่วนการตรวจให้คะแนนแบบพหุวิภาควิเคราะห์ GRM และ GPCM โดยใช้มาตรประมาณค่าแบบลิเคิร์ทและมาตรประมาณค่าแบบตัวเลือกบังคับตอบ

ผลการศึกษาพบว่าในมาตรประมาณค่าแบบลิเคิร์ท การตรวจให้คะแนนแบบทวิวิภาคที่วิเคราะห์ตามโมเดลโลจิสติก ให้ค่าฟังก์ชันสารสนเทศสูงกว่าการตรวจให้คะแนนแบบพหุวิภาคที่วิเคราะห์ตาม GRM และ GPCM และในมาตรประมาณค่าแบบตัวเลือกบังคับตอบ การตรวจให้คะแนนแบบพหุวิภาคที่วิเคราะห์ตาม GRM ให้ค่าฟังก์ชันสารสนเทศสูงกว่าการตรวจให้คะแนนแบบทวิวิภาคและแบบพหุวิภาคที่วิเคราะห์ตาม GPCM ตามลำดับ

เกษร หว่างจิตร (2539) ได้ทำการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบสำหรับแบบสอบคัดเลือกระดับบัณฑิตศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ด้วยวิธีแมนเทล-เฮนส์เซล กลุ่มตัวอย่างที่ใช้คือผู้สมัครสอบคัดเลือกเข้าศึกษาระดับบัณฑิตศึกษา โดยกลุ่มผู้สอบจำแนกตามเพศ ภูมิลำเนา ประสบการณ์ในการสอบและสังกัดสถานศึกษา เครื่องมือที่ใช้เป็นแบบสอบคัดเลือกวิชาภาษาอังกฤษ จำนวน 80 ข้อและแบบสอบวิชาภาษาไทย จำนวน 95 ข้อ

ผลการวิจัยพบว่า ข้อสอบที่พบการทำหน้าที่ต่างกันส่วนใหญ่เป็นแบบอเนกรูป ข้อสอบที่มีค่าอำนาจจำแนกค่อนข้างต่ำ (0.20-0.39) สอดคล้องกันทั้งวิชาภาษาไทยและภาษาอังกฤษ สำหรับข้อสอบที่ทำหน้าที่ต่างกันในวิชาภาษาไทยส่วนใหญ่เป็นข้อสอบที่ง่ายมาก (0.80-1.00) แต่ในวิชาภาษาอังกฤษส่วนใหญ่เป็นข้อสอบที่ยาก (0.00-0.19) ลักษณะเนื้อหาของข้อสอบที่ทำหน้าที่ต่างกันในวิชาภาษาไทยเป็นเนื้อหาด้านการใช้ภาษาไทย 1 และวิชาภาษาอังกฤษเป็นข้อสอบที่มีลักษณะการให้รายละเอียดมากที่สุด ส่วนค่าความเที่ยงและความตรงตามทฤษฎีของแบบสอบกรณีก่อนและหลังตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบสอบ ไม่แตกต่างกันอย่างมีนัยสำคัญทั้งสองวิชา

นพมาศ พิพัฒนสุข (2541) ได้ศึกษาเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีแมนเทล-แฮนส์เซลกับวิธีถดถอยโลจิสติกในแบบสอบชนิดพหุมิติ เมื่อใช้เกณฑ์จับคู่เปรียบเทียบแตกต่างกัน 3 เกณฑ์ได้แก่ คะแนนรวม คะแนนแบบสอบย่อย และคะแนนหลายแบบสอบย่อย โดยข้อมูลเก็บจากแบบสอบวัดความสามารถทางคณิตศาสตร์ที่ผู้วิจัยสร้างขึ้น

ผลการวิจัยพบว่าวิธีแมนเทล-แฮนส์เซลมีประสิทธิภาพมากกว่าวิธีถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบชนิดพหุมิติเมื่อใช้เกณฑ์จับคู่คะแนนรวม และมีประสิทธิภาพไม่แตกต่างกันเมื่อใช้เกณฑ์จับคู่คะแนนแบบสอบย่อยและวิธีถดถอยโลจิสติกเมื่อใช้เกณฑ์จับคู่เปรียบเทียบคะแนนหลายแบบสอบย่อยมีความเหมาะสมในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบชนิดพหุมิติ

นิคม กีรติวรางกูร (2542) ได้ทำการศึกษาเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีการวิเคราะห์องค์ประกอบจำกัด (RFA) วิธีแมนเทล-แฮนส์เซล และวิธีการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ โดยเทียบกับเกณฑ์ที่กำหนด เมื่อขนาดกลุ่มตัวอย่างแบ่งออกเป็น 2 ขนาดคือ กลุ่มตัวอย่างขนาดเล็ก (300 คน) และขนาดใหญ่ (1,000 คน) ค่าความยาวของแบบสอบแบ่งออก เป็น 2 ขนาดคือ แบบสอบสั้น (25 ข้อ) และแบบสอบยาว (75 ข้อ) ค่าความยากของข้อสอบแบ่งออกเป็น 3 ระดับคือ กลุ่มข้อสอบที่มีค่าอำนาจจำแนกสูง ปานกลาง และต่ำ ขนาดความลำเอียงของข้อสอบแบ่งออกเป็น 2 ขนาดคือ กลุ่มข้อสอบที่มีความลำเอียงสูง และต่ำ

ผลการวิจัยพบว่า โดยภาพรวม วิธี RFA มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันสูงที่สุด รองลงมาคือ วิธี MH และวิธี IRT แบบ 2 พารามิเตอร์ ตามลำดับ วิธี IRT มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธี MH และ วิธี RFA ตามลำดับ วิธี MH มีประสิทธิภาพสูงในการตรวจสอบเมื่อแบบสอบมีค่าความยากต่ำ อำนาจจำแนกสูง ที่ขนาดความยาว 75 ข้อ เมื่อใช้กลุ่มตัวอย่างขนาด 1,000 คน และวิธี IRT แบบ 2 พารามิเตอร์ มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูง ภายใต้เงื่อนไขแบบสอบที่มีความยากต่ำ ที่ขนาดความยาว 75 ข้อ เมื่อใช้กลุ่มตัวอย่างขนาด 1,000 คน

4.2 งานวิจัยในต่างประเทศ

สำหรับงานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบและแบบสอบในต่างประเทศ ซึ่งผู้วิจัยได้ศึกษามีดังนี้

Zwick และคณะ (1993) ได้ทำการศึกษาผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่เป็นชิ้นงาน ด้วยวิธีแมนเทล (Mantel approach) และวิธี GMH (Generalized of the Mantel-Haenszel approach) ใช้ข้อมูลจำลองความยาวของแบบสอบ 25 ข้อ 24ข้อแรกใช้เป็นตัวแปรที่ใช้จับคู่และข้อที่ 25 เป็นข้อสอบที่ศึกษา กลุ่มเปรียบเทียบมีการกระจาย 2 ระดับ คือ N(0,1) และ N(-1,1) และลักษณะเฉพาะของข้อสอบที่ศึกษามี 27 ระดับ แตกต่างกันตามค่าความยาก รูปแบบของ DIF มี 4 รูปแบบคือ แบบคงที่ (constant) แบบสมดุล (balance) แบบมีผลเฉพาะคะแนนต่ำ (shift low) และแบบมีผลเฉพาะคะแนนสูง (shift high) และขนาดของ DIF มี 2 ขนาดคือ .10 และ .25 ข้อสอบข้อที่ 1-20 เป็นข้อสอบที่ให้คะแนนแบบสองค่าและข้อที่ 21-24 เป็นข้อสอบที่ให้คะแนนแบบ 4 ค่า ซึ่งข้อที่ 24 เป็นข้อสอบที่ทำหน้าที่ต่างกัน

ผลการศึกษาพบว่าทั้งสองวิธีมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในการวัดความสามารถในการปฏิบัติงานในกรณีที่มีการให้คะแนนแบบหลายค่า วิธี GMH จะมีประสิทธิภาพมากกว่าวิธีแมนเทล เมื่อการกระจายของข้อมูลในกลุ่มผู้สอบแตกต่างกัน ผลการศึกษาแสดงให้เห็นว่าควรรวมข้อสอบที่ศึกษาเข้าไปในตัวแปรที่ใช้ในการจับคู่ด้วยเหมือนในกรณีของการให้คะแนนแบบสองค่า ในเงื่อนไขของ DIF แบบคงที่ วิธีแมนเทลจะมีประสิทธิภาพมากกว่าวิธี GMH ส่วนในเงื่อนไข DIF แบบสมดุล วิธี GMH จะมีความสามารถเหนือกว่าวิธีแมนเทลอย่างเห็นได้ชัดที่ขนาดของ DIF เท่ากับ .25 ส่วนในกรณี DIF แบบมีผลเฉพาะที่ระดับคะแนนต่ำและระดับสูง ทั้งสองวิธีมีประสิทธิภาพการตรวจสอบใกล้เคียงกัน

Shealy และ Stout (1993) ได้เสนอวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบคือวิธีซิบเทสส์ (SIBTEST) ซึ่งเป็นวิธีการที่อยู่บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ เป็นวิธีการที่สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ทั้งกรณีที่แบบสอบมีความลำเอียงข้อเดียวและกรณีที่มีความลำเอียงหลายข้อ ได้ทำการศึกษาประสิทธิภาพในการตรวจสอบโดยการเปรียบเทียบกับวิธีแมนเทล-เฮนส์เซล ในกรณีที่มีข้อสอบลำเอียงข้อเดียว โดยใช้ข้อมูลผลการตอบจากแบบสอบวิชาคณิตศาสตร์ของ ACT และแบบสอบของ ASVAB ขนาดกลุ่มตัวอย่างของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบต่างกัน คือ (3000 , 3000) , (3000 , 1000) , (1500 ,1500) , (1000 , 500) , (500 , 500) และ (500 , 250)

ผลการศึกษาพบว่าในกรณีที่มีข้อสอบลำเอียงข้อเดียวทั้งวิธีแมนเทล-เฮนส์เซลและวิธีซิบเทสส์มีอำนาจในการตรวจสอบดีเทียบเท่ากันได้และวิธีซิบเทสส์มีประสิทธิภาพดีในการตรวจค้นกรณีที่มีข้อสอบที่ลำเอียงหลายๆ ข้อ แม้ว่ามีขนาดความลำเอียงค่อนข้างน้อยและผลการศึกษาพบว่าทั้งวิธีซิบเทสส์และแมนเล-เฮนส์เซลจะมีประสิทธิภาพดีเมื่อใช้กับแบบสอบที่ยาวพอควร ($\geq$ 25)

Cohen และ Kim (1993) ศึกษาเปรียบเทียบการใช้ทฤษฎีการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ โดยใช้สถิติทดสอบไค-สแควร์ ของลอร์ดและสถิติทดสอบพื้นที่ความแตกต่างระหว่างโค้งลักษณะข้อสอบชนิดคิดเครื่องหมายและไม่คิดเครื่องหมาย อันได้แก่ สถิติ Z ของ Raju โดยใช้ข้อมูลจำลองตามทฤษฎีการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ จำลองการตอบข้อสอบจากแบบสอบ 2 ขนาด คือ 20 ข้อ และ 60 ข้อ จำนวนผู้สอบ 100 และ 500 คน และมีจำนวนข้อสอบที่ลำเอียงในแบบสอบจำนวน ร้อยละ 0 10 และ 20

ผลการวิจัยพบว่า แบบสอบยิ่งมีความยาว จะทำให้การระบุข้อสอบลำเอียงผิดพลาด (FP) และข้อสอบที่ไม่ถูกระบุว่าลำเอียงผิดพลาด (FN) จะมากขึ้นด้วย และถ้ามีจำนวนข้อสอบที่ลำเอียงอยู่ในแบบสอบน้อย จำนวนข้อสอบที่ถูกระบุผิดพลาดก็จะยิ่งมากขึ้น องค์ประกอบที่มีอิทธิพลต่อการระบุข้อสอบลำเอียงผิดพลาด (FP) อีกประการคือ ระดับความมีนัยสำคัญ (α) ค่า α ยิ่งมาก FP จะเกิดมาก

กรณีที่จำนวนผู้สอบ 100 คน พบว่าความแตกต่างในการระบุว่าข้อสอบไม่ลำเอียงผิดพลาด สำหรับ Z และ ไคว์สแควร์ ไม่แตกต่างกันมากเท่ากับกรณีกลุ่มผู้สอบจำนวน 500 คน ซึ่งมีการระบุว่าไม่ลำเอียงผิดพลาด สำหรับไคว์สแควร์ น้อยกว่า Z โดยเฉพาะอย่างยิ่งในกรณีที่มีข้อสอบลำเอียงในแบบสอบมากที่สุดคือ ร้อยละ 20

Narayanan และ Swaminathan (1994) ได้ศึกษาผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-เฮนส์เซลและวิธีซิบเทสส์ ใช้ข้อมูลจำลองความยาวแบบสอบ 40 ข้อ ตัวแปรที่ศึกษาคือ (1) ขนาดกลุ่มตัวอย่าง โดยกลุ่มอ้างอิงมี 3 ขนาดได้แก่ 300 500 และ 1,000 คน (2) การกระจายของความสามารถ 2 แบบ (3) อัตราส่วนของข้อสอบที่ทำหน้าที่ต่างกันที่มีภายในแบบสอบ 2 ขนาด (4) ขนาดพื้นที่ระหว่างโค้งคุณลักษณะข้อสอบของผู้สอบสองกลุ่ม 4 ขนาด (5) ค่าความยากและค่าอำนาจจำแนกของแบบสอบ 6 ระดับ

ผลการศึกษาพบว่าขนาดของกลุ่มตัวอย่าง อัตราส่วนของข้อสอบที่ทำหน้าที่ต่างกันขนาดของพื้นที่ระหว่างโค้งคุณลักษณะ ค่าความยาก และค่าอำนาจจำแนกเป็นตัวแปรที่มีผลกระทบต่ออัตราการตรวจสอบของทั้งสองวิธีอย่างมีนัยสำคัญ วิธีแมนเทล-เฮนส์เซลและวิธีซิบเทสส์มีประสิทธิภาพเท่ากันในการตรวจสอบการทำหน้าที่ต่างกันแบบเอกรูปเมื่อการกระจายความสามารถเท่ากันระหว่างกลุ่ม แต่เมื่อการกระจายความสามารถไม่เท่ากันระหว่างกลุ่ม วิธีซิบเทสส์จะมีประสิทธิภาพมากกว่า จึงสรุปได้ว่าการกระจายความสามารถไม่มีผลกระทบต่อการตรวจสอบด้วยวิธีซิบเทสท์แต่มีผลกระทบต่อวิธีแมนเทล-เฮนส์เซล อย่างมีนัยสำคัญ ในกรณีการกระจายความสามารถต่างกันเพิ่มขึ้นระหว่างผู้สอบทั้งสองกลุ่มจะทำให้อัตราความคลาดเคลื่อนชนิดที่ 1 เพิ่มมากขึ้น

Welch และ Miller (1995) ได้ทำการศึกษาปัญหาของการตรวจสอบการทำหน้าที่ต่างกันของแบบสอบประเมินการเขียน โดยมีจุดมุ่งหมายเพื่อศึกษาว่าปัญหาใดบ้างที่เป็นอุปสรรคต่อการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ประเมินผลการปฏิบัติจริงและเพื่อค้นหาเกณฑ์การจับคู่ที่เหมาะสม โดยใช้สถิติ GMH (Generalized Mantel-Haenszel) และ การวิเคราะห์ฟังก์ชันโลจิสติก เครื่องมือที่ใช้ศึกษาเป็นแบบประเมินทักษะการเขียนโดยตรงและแบบสอบวัดทักษะการเขียนแบบเลือกตอบ จำนวน 40 ข้อ กลุ่มตัวอย่างเป็นนักเรียนเกรด 8 ข้อมูลที่นำมาใช้ได้จากแบบประเมินทักษะการเขียนโดยตรง มี 3 ฟอร์ม คือ 01A , 01B และ 01C ตัวแปรที่ใช้แบ่งกลุ่มผู้สอบคือตัวแปรเพศและตัวแปรสีผิว ผลการศึกษาโดยภาพรวมแล้วพบว่าเกณฑ์การจับคู่โดยใช้เกณฑ์ภายนอกไม่เหมาะสมและยังบ่งชี้อีกว่าเกณฑ์การจับคู่ที่ไม่เหมาะสมจะเป็นปัญหาในการตรวจสอบการทำหน้าที่ต่างกัน

Raju และคณะ (1995) ได้ศึกษาประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ ด้วยวิธี DFIT ซึ่งประกอบด้วยดัชนี CDIF และ NCDIF สำหรับตรวจสอบการทำหน้าที่ต่างกันที่ระดับข้อสอบ และดัชนี DTF สำหรับตรวจสอบการทำหน้าที่ต่างกันที่ระดับแบบสอบ โดยจะทำการศึกษาเปรียบเทียบกับวิธีวิเคราะห์อีก 3 วิธีคือวิธีทดสอบไค-สแควร์ของลอร์ด วิธีที่ใช้ทฤษฎีการตอบสนองข้อสอบแบบคิดเครื่องหมาย (ESA) และแบบไม่คิดเครื่องหมาย (EUA) ใช้ข้อมูลจำลองจากโมเดลโลจิสติกแบบ 2 พารามิเตอร์ (2PLM) เปรียบเทียบประสิทธิภาพวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อใช้ขนาดกลุ่มตัวอย่าง (500 , 1000) อัตราส่วนของข้อที่ทำหน้าที่ต่างกันในแบบสอบ (0% , 5% , 10% , 20%) และรูปแบบการทำหน้าที่ต่างกันของข้อสอบ (uniform , non-uniform)

ผลการวิจัยพบว่าในภาพรวมดัชนี LC ของวิธีทดสอบไค-สแควร์ของลอร์ด ดัชนี EUA ของวิธีที่ใช้ทฤษฎีการตอบสนองข้อสอบแบบไม่คิดเครื่องหมาย มีประสิทธิภาพเท่าเทียมกันในการตรวจสอบที่ระดับข้อสอบ โดยสามารถระบุข้อที่ทำหน้าที่ต่างกันได้ถูกต้องสูงในทุกเงื่อนไข ในขณะที่ดัชนี CDIF และดัชนี ESA มีประสิทธิภาพต่ำ เมื่อกลุ่มตัวอย่างมีขนาดเล็ก (N=500) และอัตราส่วนของข้อที่ทำหน้าที่ต่างกันสูง (20%) ดัชนี NCDIF จะมีประสิทธิภาพสูงที่สุดแต่ดัชนี CDIF มีประสิทธิภาพต่ำที่สุด เมื่อขนาดกลุ่มตัวอย่างมีขนาดใหญ่ (N=1000) และอัตราส่วนการทำหน้าที่ต่างกันสูง (20%) ดัชนี ESA มีประสิทธิภาพต่ำที่สุด

Giray (1995) ได้ทำการศึกษาวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ สำหรับแบบสอบคัดเลือกบุคคลเข้าศึกษาในมหาวิทยาลัย ซึ่งเป็นข้อสอบด้านการคำนวณ คำศัพท์ และเรขาคณิต โดยแบ่งกลุ่มผู้สอบตามตัวแปรที่ศึกษาคือ เพศและเศรษฐานะ กลุ่มตัวอย่างเป็นนักเรียนของโรงเรียนมัธยมศึกษาใน Ankara ประเทศตุรกี จากผลการวิจัยพบว่า ข้อสอบที่ทำหน้าที่ต่างกันเป็นข้อสอบใน

วิชาคณิตศาสตร์ ทั้งในกรณีแบ่งกลุ่มผู้สอบตามตัวแปรเพศและเศรษฐานะ สำหรับตัวแปรเพศ ผลการวิเคราะห์พบว่าผู้ชายจะมีข้อได้เปรียบในการทำข้อสอบด้านการคำนวณ ในขณะที่เพศหญิงจะมีข้อได้เปรียบในการทำข้อสอบด้านคำศัพท์และเรขาคณิต ซึ่งชี้ให้เห็นว่าผู้หญิงจะมีความสามารถในด้านการใช้ภาษาหรือเกี่ยวกับการใช้คำศัพท์ดีกว่าและเพศชายจะมีทักษะด้านการคำนวณดีกว่าเพศหญิง

Chang และคณะ (1996) ได้ทำการศึกษาผลของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่า โดยประยุกต์วิธีซิบเทสท์มาใช้ เปรียบเทียบกับวิธีแมนเทลและวิธีSMD แบ่งการศึกษาออกเป็น 2 ตอน ตอนที่ 1 ใช้ข้อมูลจำลองภายใต้เงื่อนไขเดียวกับงานวิจัยของ Zwick และคณะ (1993) เพื่อทำการเปรียบเทียบวิธีซิบเทสท์แบบประยุกต์กับวิธีแมนเทลและวิธี SMD ผลที่ได้ก็คือวิธีซิบเทสท์มีประสิทธิภาพในการตรวจสอบค่อนข้างดีแต่วิธีวิธีแมนเทลและวิธี SMD มีประสิทธิภาพค่อนข้างดีกว่า ในการศึกษาตอนที่ 2 ใช้ข้อมูลจำลองคือ ข้อสอบที่ศึกษามีอำนาจการจำแนกแตกต่างกัน 11 ค่า ตั้งแต่ .15 ถึง 2.00 ขนาดกลุ่มตัวอย่างต่างกัน คือ 500 1,000 และ 2,000 คน ข้อสอบมีความยาว 24 ข้อสำหรับวิธีซิบเทสท์ และ 25 ข้อสำหรับวิธีแมนเทลและวิธี SMD ผลการศึกษาสรุปว่าทั้งวิธีแมนเทลและวิธี SMD มีค่าความคลาดเคลื่อนชนิดที่ 1 ค่อนข้างสูงเมื่อพารามิเตอร์ของค่าอำนาจจำแนกของข้อสอบที่ศึกษามีค่าต่างจากค่าเฉลี่ยของอำนาจจำแนกของแบบสอบที่มีความตรง และอัตราการปฏิเสธ (rejection rates) ของทั้งสามวิธีการจะมีค่าสูงขึ้นเมื่อค่าอำนาจจำแนกสูงขึ้น

Katherine และ Meichu (1996) ได้ทำการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ สำหรับแบบสอบวิชาคณิตศาสตร์ ซึ่งเป็นแบบสอบแบบหลายตัวเลือก โดยแบ่งกลุ่มผู้สอบตามตัวแปรเพศ กลุ่มตัวอย่างเป็นนักเรียนชั้นปีที่ 3 ของโรงเรียนมัธยมศึกษา ข้อมูลที่ใช้ในการวิจัยรวบรวมมาจาก The Second International Mathematics Study (1985) ในการวิจัยครั้งนี้ได้ศึกษาความสัมพันธ์ระหว่างดัชนีที่ใช้วิเคราะห์การทำหน้าที่ต่างกันและลักษณะเฉพาะของแบบสอบแบบหลายตัวเลือก โดยมีสมมติฐานว่าข้อสอบด้านพีชคณิตและการคำนวณจะเข้าข้างผู้สอบเพศหญิง ส่วนข้อสอบด้านเรขาคณิตและเลขคณิตจะเข้าข้างเพศชาย ผลการวิจัยพบว่าข้อสอบด้านพีชคณิตและการคำนวณจะง่ายสำหรับผู้สอบเพศหญิงและข้อสอบด้านเรขาคณิตจะเข้าข้างเพศชาย ในขณะที่ข้อสอบด้านเลขคณิตจะเข้าข้างเพศหญิง ซึ่งเป็นสมมติฐานเดียวที่ถูกปฏิเสธ จากผลการวิจัยในภาพรวม พบว่าข้อสอบด้านการประยุกต์จะทำหน้าที่ต่างกันโดยจะเป็นข้อสอบที่ค่อนข้างยากสำหรับเพศหญิง

Ning และ Suzanne (1996) ได้ทำการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในแบบสอบการประเมินสมรรถภาพในวิชาคณิตศาสตร์ การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ โดยแบ่งกลุ่มผู้สอบตามเพศ และระบุองค์ประกอบที่อาจเกี่ยวข้องสัมพันธ์

กับการทำหน้าที่ต่างกัน เช่น เนื้อหา กระบวนการรับรู้ และความแตกต่างในการกระจายของค่าความสามารถ เป็นต้น โดยใช้แบบสอบ QUASAR (Quantitative Understanding : Amplifying Student Achievement and Reasoning) Cognitive Assessment Instrument (QCAI) ในการทดสอบทักษะการคิดและการใช้เหตุผลของนักเรียน ซึ่งประกอบด้วยข้อสอบแบบปลายเปิด โดยผู้สอบต้องแสดงวิธีทำและอธิบายคำตอบของเขาด้วย แบบสอบมี 4 ฟอร์ม 33 ข้อ ข้อสอบจะเป็นแบบให้คะแนนแบบหลายค่า (polytomously scored items) กลุ่มตัวอย่างเป็นนักเรียนเกรด 6 และเกรด 7 จาก 6 โรงเรียนที่ร่วมโครงการ QUASAR โดยใช้กลุ่มตัวอย่างจำนวน 1,782 คน ผลการวิจัยพบว่าข้อสอบจำนวน 31 ข้อ จากข้อสอบทั้งหมด 33 ข้อ คิดเป็นร้อยละ 94 ไม่ทำหน้าที่ต่างกันหรือไม่ทำหน้าที่ต่างกันในระดับที่มากนัก ส่วนข้อสอบอีก 2 ข้อที่เหลือ ทำหน้าที่ต่างกัน เมื่อแบ่งกลุ่มผู้สอบตามเพศ

Oshima และคณะ (1997) ได้ทำการศึกษาวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ (DFIT Framework) สำหรับแบบสอบชนิดพหุมิติ ซึ่งวิธีการนี้ได้ขยายมาจากวิธี DFIT แบบมิติเดียวที่ Raju และคณะเสนอไว้ในปี 1995 ในการศึกษาใช้ข้อมูลจำลองจากโมเดลโลจิสติกแบบ 2 พารามิเตอร์ (M2PL) จำนวน 40 ข้อ และข้อมูลเป็นแบบ 2 มิติ องค์ประกอบที่สนใจศึกษาคือ รูปแบบการทำหน้าที่ต่างกัน (เอกรูปและแบบอเนกรูป) ทิศทางการทำหน้าที่ต่างกัน (แบบทิศทางเดียว , แบบสองทิศทางที่สมดุลกัน) และการกระจายของค่าความสามารถแตกต่างกันระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ (N(0,1) , N(-1,1))

ผลการศึกษาพบว่าในทุกเงื่อนไขข้อสอบที่ถูกจำลองไม่ทำหน้าที่ต่างกัน ดัชนี CDIF และดัชนี NCDIF ที่คำนวณได้มีค่าเป็น 0.00 สำหรับเงื่อนไขที่ 1 (การทำหน้าที่ต่างกันแบบเอกรูปและมีทิศทางเดียว) ข้อสอบ 4 ข้อที่จำลองให้ทำหน้าที่ต่างกัน ดัชนี CDIF ของทั้ง 4 ข้อมีค่าประมาณ 0.40 และดัชนี DTF มีนัยสำคัญเมื่อทดสอบด้วยสถิติไค-สแควร์ และค่าดัชนี NCDIF มีค่าเป็น 0.10 ซึ่งระบุว่าทำหน้าที่ต่างกัน ในเงื่อนไขที่ 2 (เอกรูปและสองทิศทางแบบสมดุล) ค่าดัชนี CDIF ทั้ง 4 ข้อมีค่าเป็น 0.00 และ DTF มีค่าเป็น 0.00 แต่ค่าดัชนี NCDIF เท่ากับ 0.10 ซึ่งมีนัยสำคัญทางสถิติอีกเช่นกัน ส่วนในเงื่อนไขเกี่ยวกับการทำหน้าที่ต่างกันแบบอเนกรูป พบว่าเมื่อค่าความยากของทั้งสองมิติแตกต่างกันจะทำให้ค่า CDIF และ NCDIF มีค่ามากขึ้น ถ้าค่าความยากของทั้งสองมิติแตกต่างกันในทิศทางตรงข้าม ดัชนี CDIF จะมีค่าเท่ากันแต่ NCDIF จะมีค่ามากกว่า เมื่อค่าความยากทั้ง 2 มิติต่างกันแต่เป็นไปในทิศทางเดียวกัน

Flowers และคณะ (1997) ได้ศึกษาความสัมพันธ์ระหว่างวิธี DFIT แบบพหุวิภาคและวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบวิธีอื่นๆ โดยทำการเปรียบเทียบประ

สิทธิภาพในการตรวจสอบระหว่างวิธี DFIT วิธีซิบเทสส์ และวิธี ทดสอบค่าไค-สแควร์ของลอร์ด และใช้ข้อมูลจำลอง ศึกษาผลของขนาดกลุ่มตัวอย่างต่างกัน (500 และ 1000) ค่าการกระจายของค่าความสามารถของกลุ่มเปรียบเทียบ (N(0,1) และ N(-1,1)) จำนวนข้อสอบที่ทำหน้าที่ต่างกัน (0% , 10% , 20%) ผลการวิจัยพบว่าในภาพรวมวิธี DFIT มีประสิทธิภาพในการตรวจสอบดี จำนวนของข้อที่ทำหน้าที่ต่างกัน ขนาดของการทำหน้าที่ต่างกันและค่าความยากมีผลต่ออัตราความคลาดเคลื่อนชนิดที่ 1

Oshima และคณะ (1998) ได้ทำการศึกษาสาเหตุของการทำหน้าที่ต่างกันของกลุ่มข้อสอบ (Differential Bunble Functiong : DBF) โดยทำการแบ่งข้อสอบออกเป็นกลุ่ม ๆ แตกต่างกัน คือแบ่งตามลำดับขั้นของการเรียนรู้ วัตถุประสงค์ที่ต้องการวัดและแบ่งตามบทความที่อยู่ในแบบสอบ แล้วทำการคำนวณค่าดัชนี bundle-CDIF , bundle-NCDIF และ bundle-DTF ซึ่งมีพื้นฐานมาจากทฤษฎีของวิธี DFIT เครื่องมือที่ใช้เป็นแบบวัดทักษะการอ่านจาก Metropolitan Achievement Tests กลุ่มตัวอย่างที่ใช้เป็นนักเรียนเกรด 4 จำนวน 1,000 คน แบบสอบยาว 55 ข้อ ตัวแปรที่ใช้แบ่งกลุ่มตัวอย่างคือตัวแปรเพศและฐานะเศรษฐสังคม

ผลการศึกษาพบว่า สำหรับการเปรียบเทียบระหว่างเพศชายและเพศหญิง เมื่อตัดข้อสอบที่ตรวจพบว่าทำหน้าที่ต่างกันออกไป 2 ข้อ คือข้อที่ 22 และ 25 แบบสอบฉบับนั้นก็ไม่ทำหน้าที่ต่างกันอีกนั่นคือดัชนี DTF ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ สำหรับการเปรียบเทียบระหว่างฐานะเศรษฐสังคม ตรวจไม่พบข้อสอบที่ทำหน้าที่ต่างกัน และเมื่อแบ่งวิเคราะห์ตามกลุ่มของข้อสอบพบว่าบทความที่ 5 มีค่าดัชนี NCDIF สูงที่สุดโดยจะเข้าข้างเพศชายมากกว่าหญิงแต่ค่าดัชนี bundle NCDIF มีค่าไม่แตกต่างกันเมื่อแบ่งกลุ่มตามฐานะเศรษฐสังคม

Larry (1999) ได้ทำการศึกษาเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบการแปล โดยทำการเปรียบเทียบระหว่างกระบวนการ DFIT และวิธีแมนเทล-เฮนส์เซล (MH) ใช้ข้อมูลจากแบบสอบการแปล ซึ่งมีข้อสอบทั้งหมดจำนวน 50 ข้อ แต่ในการวิจัยครั้งนี้ใช้เฉพาะข้อสอบแบบเลือกตอบ จำนวน 30 ข้อ โดยมีวัตถุประสงค์เพื่อศึกษาจำนวนและรูปแบบที่สอดคล้องกันระหว่างทั้ง 2 วิธีดังกล่าว

ผลการศึกษาพบว่ากระบวนการ DFIT มีความไวในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ดีกว่า นั่นคือกระบวนการ DFIT สามารถตรวจพบข้อที่ทำหน้าที่ต่างกันอย่างมีนัยสำคัญจำนวน 10 ข้อ จากข้อสอบทั้งหมด จำนวน 30 ข้อ ในขณะที่วิธี แมนเทล-เฮนส์เซล (MH) ระบุข้อสอบที่ DIF อย่างมีนัยสำคัญจำนวน 2 ข้อ เท่านั้น ในภาพรวมพบว่าทั้งสองวิธีระบุข้อสอบที่ DIF สอดคล้องกันเพียงร้อยละ 20 ข้อสอบที่กระบวนการ DFIT ระบุว่าเกิด DIF เป็นแบบเอกรูป จำนวน

4 ข้อและเป็นแบบอเนกรูป จำนวน 6 ข้อ ในขณะที่วิธี MH ระบุว่าข้อสอบที่เกิด DIF เป็นแบบเอกรูปและอเนกรูปอย่างละ 1 ข้อ

Rozowski และ Reith (1999) ได้ทำการตรวจสอบคุณภาพในการวัดและความตรงของแบบสอบที่ประกอบด้วยข้อสอบที่ DIF และ ไม่ DIF โดยใช้วิธีแมนเทล-เฮนส์เซล (MH) ในการตรวจสอบ DIF และแบ่งกลุ่มผู้สอบตามตัวแปรเพศและเชื้อชาติ ในการตรวจสอบคุณภาพของแบบสอบนั้นได้ใช้ค่าสัมประสิทธิ์แอลฟ่า ตรวจสอบความตรงของแบบสอบโดยเทียบกับเกณฑ์ภายนอก คือ ผลการสอบ American College Testing (ACT) , Scholastic Aptitude Test (SAT) verbal (SATV) , quantitative (SATQ) , composite (SATC) และ แรงค์เปอร์เซนต์ไทล์ของเกรดเฉลี่ย โดยใช้ข้อมูลของผู้สอบจากกลุ่มข้อมูลของ High School and Beyond (HSB) ที่จัดทำโดย National Educational Longitudinal Studies Program

ผลการวิจัยพบว่าแบบสอบที่ประกอบด้วยข้อสอบที่ทำหน้าที่ต่างกันนั้น ไม่ได้ทำให้แบบสอบมีคุณภาพในการวัดหรือมีความตรงต่ำลงมากนัก และเมื่อพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์แบบแรงค์ระหว่างระหว่างคะแนนสอบของแบบสอบที่ไม่มีข้อสอบที่ DIF กับแบบสอบที่ประกอบด้วยข้อสอบที่ DIF เข้าข้างเฉพาะกลุ่มใดกลุ่มหนึ่ง และเข้าข้างกลุ่มใดกลุ่มหนึ่งในจำนวนที่เท่า ๆ กัน พบว่าค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้มีค่าสูง โดยแบบสอบที่มีข้อสอบที่เข้าข้างทั้งสองกลุ่มในจำนวนที่สมดุลกันจะมีความสัมพันธ์กับแบบสอบที่ไม่มีข้อสอบที่ DIF เลย สูงที่สุด

Flowers และคณะ (1999) ได้พัฒนากระบวนการ DFIT สำหรับแบบสอบเอกมิติ (unidimensional) ที่มีข้อสอบที่มีการให้คะแนนแบบหลายค่า (polytomous) โดยกระบวนการนี้ได้รับการพัฒนามาจากกระบวนการ DFIT สำหรับแบบสอบเอกมิติที่มีการให้คะแนนแบบ 2 ค่า (dichotomous) ของ Raju และคณะ (1995) ข้อมูลที่ใช้ในการศึกษาเป็นข้อมูลจำลอง ข้อสอบให้มีการให้คะแนนเป็น 0, 1, 2, 3 และ 4 องค์ประกอบที่สนใจศึกษาประกอบด้วยการกระจายค่าความสามารถของกลุ่มเปรียบเทียบ (ค่าความสามารถกระจายเหมือนกลุ่มอ้างอิง โดยทำการสุ่มกลุ่มตัวอย่างที่มีค่าความสามารถกระจายแบบ N(0,1) และค่าความสามารถกระจายไม่เหมือนกลุ่มอ้างอิง โดยทำการสุ่มกลุ่มตัวอย่างที่มีความสามารถกระจายแบบ N(-1,1)) ความยาวของแบบสอบ (20 ข้อ และ 40 ข้อ) จำนวนข้อสอบที่ DIF (0%, 5%, 10% และ 20%) และทิศทางการเกิด DIF (แบบทิศทางเดียว และแบบสมดุลสองทิศทาง)

ผลการวิจัยพบว่ากระบวนการ DFIT มีประสิทธิภาพในการระบุข้อสอบที่ DIF และแบบสอบที่ DTF สำหรับข้อมูลที่มีการให้คะแนนแบบหลายค่า ความยาวของแบบสอบ การกระจายของ

ค่าความสามารถของกลุ่มเปรียบเทียบ จำนวนข้อสอบที่ DIF และทิศทางของการเกิด DIF มีผลกระทบเพียงเล็กน้อยต่ออัตราการตรวจสอบที่ถูกต้องหรืออัตราความผิดพลาด ในทุก ๆ เงื่อนไข และข้อสอบที่มีค่า DIF มากจะถูกตรวจพบ ในขณะที่ข้อสอบที่มีค่า DIF น้อย ๆ จะไม่ถูกตรวจพบ ในภาพรวมดัชนี NCDIF มีความคงที่มากกว่าดัชนี CDIF มีอยู่ 2 เงื่อนไขที่ดัชนี CDIF ไม่เป็นไปตามที่คาดไว้ คือเงื่อนไขที่ 1 เมื่อมีแบบสอบที่มีความยาว 20 ข้อ และการกระจายของค่าความสามารถของกลุ่มอ้างอิงเหมือนกับกลุ่มเปรียบเทียบ ดัชนี CDIF ระบุข้อที่ไม่ DIF ว่าเกิด DIF ผิดพลาด ร้อยละ 18 เงื่อนไขที่ 2 เมื่อแบบสอบมีความยาว 40 ข้อ และการกระจายความสามารถของทั้งสองกลุ่มต่างกัน ดัชนี CDIF ระบุข้อที่ DIF ได้เพียงร้อยละ 50 จากข้อสอบที่ DIF ทั้งหมด

Ellis และ Mead (2000) ได้ทำการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ โดยใช้กระบวนการ DFIT สำหรับแบบสอบถาม Sixteen Personality Factor (16PF) ที่แปลจากภาษาอังกฤษเป็นภาษาเสปน โดยแบ่งกลุ่มผู้สอบเป็นกลุ่มย่อยตามเชื้อชาติ 3 กลุ่ม ประกอบด้วย กลุ่ม Anglo Americans , Spanish speakers และ Hispanic Americans

ผลการวิจัยพบว่า ดัชนี CDIF/DTF ตรวจพบข้อสอบที่ทำหน้าที่ต่างกันน้อยกว่า เมื่อใช้ดัชนี NCDIF ในการตรวจสอบ ในการวิเคราะห์พบว่า เมื่อแบ่งกลุ่มผู้สอบเปรียบเทียบระหว่างกลุ่ม Anglo Americans และ กลุ่ม Spanish speakers ตรวจพบข้อสอบที่ DIF มากที่สุด และเมื่อแบ่งกลุ่มผู้สอบเปรียบเทียบระหว่างกลุ่ม Anglo Americans และ กลุ่ม Hispanic Americans ตรวจพบข้อสอบที่ DIF จำนวนน้อยที่สุด

จากผลการศึกษางานวิจัยที่เกี่ยวข้องทั้งในประเทศและต่างประเทศสามารถสรุปได้ว่า มีการศึกษาโดยใช้ทั้งข้อมูลจำลองและข้อมูลเชิงประจักษ์ งานวิจัยส่วนใหญ่จะเน้นไปที่การใช้ข้อมูลจำลองเพื่อเปรียบเทียบประสิทธิภาพระหว่างวิธีการตรวจสอบที่แตกต่างกัน ภายใต้เงื่อนไขที่แตกต่างกัน คือความยาวแบบสอบ ขนาดกลุ่มตัวอย่าง สัดส่วนของขนาดกลุ่มเปรียบเทียบและกลุ่มอ้างอิง การกระจายของค่าความสามารถ รูปแบบการทำหน้าที่ต่างกัน ขนาดของการทำหน้าที่ต่างกันและสัดส่วนของข้อที่ทำหน้าที่ต่างกันในแบบสอบทั้งฉบับ ส่วนในงานวิจัยโดยใช้ข้อมูลเชิงประจักษ์ตัวแปรที่สนใจนำมาศึกษา เช่น เพศ สีผิว เชื้อชาติ วัฒนธรรม สภาพทางเศรษฐสังคม ประสบการณ์ในการสอบ ภูมิศาสตร์ ศาสนา ระดับสติปัญญา เป็นต้น ผลการศึกษาที่ผ่านมาระบุว่าข้อสอบที่ทำหน้าที่ต่างกันจะเป็นข้อสอบในวิชาคณิตศาสตร์และด้านภาษา ในงานวิจัยที่เกี่ยวข้องดังกล่าวส่วนมากจะศึกษาการทำหน้าที่ต่างกันที่ระดับข้อสอบ ส่วนที่ระดับกลุ่มข้อสอบและแบบสอบทั้งฉบับยังมีไม่มาก โดยเฉพาะในประเทศไทยยังไม่พบงานวิจัยในเรื่องดังกล่าว ที่ผ่านมาจะศึกษาเฉพาะที่ระดับข้อสอบเท่านั้น

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ ด้วยกระบวนการ DFIT สำหรับแบบสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษา วิชาภาษาอังกฤษและวิชาคณิตศาสตร์ เมื่อแบ่งกลุ่มผู้สอบตามตัวแปรเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา ซึ่งมีวิธีดำเนินการวิจัย ตามขั้นตอนดังต่อไปนี้

ประชากรและกลุ่มตัวอย่าง

ประชากร

ประชากรในการวิจัยครั้งนี้คือผู้สมัครสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษา ปี 2543 ครั้งที่ 1/มีนาคม ซึ่งเป็นผู้สมัครสอบวิชาภาษาอังกฤษ รหัส 03 จำนวน 160,356 คน และผู้สมัครสอบวิชาคณิตศาสตร์ 1 รหัส 04 จำนวน 124,718 คน

กลุ่มตัวอย่าง

กลุ่มตัวอย่างในการวิจัยครั้งนี้ มี 2 กลุ่ม ดังนี้คือ

กลุ่มที่ 1 ผู้สมัครสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษา ปี 2543 ครั้งที่ 1/มีนาคม ซึ่งเป็นผู้สมัครสอบวิชาภาษาอังกฤษ

กลุ่มที่ 2 ผู้สมัครสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษา ปี 2543 ครั้งที่ 1/มีนาคม ซึ่งเป็นผู้สมัครสอบวิชาคณิตศาสตร์

กลุ่มตัวอย่างได้มาจากการสุ่มของเจ้าหน้าที่สำนักทดสอบกลาง ทบวงมหาวิทยาลัย จำแนกตามตัวแปรที่ต้องการศึกษา กลุ่มย่อย กลุ่มละ 1,000 คน ซึ่งเป็นขนาดกลุ่มตัวอย่างที่มั่นใจได้ว่ามีขนาดพอเพียงสำหรับการประมาณค่าพารามิเตอร์ได้อย่างถูกต้องแม่นยำ ก่อนการวิเคราะห์ DIF/DTF (Muraki and Bock, 1993 cited in Flowers et al, 1999) มีรายละเอียดดังนี้

ตัวแปรเพศ ประกอบด้วย เพศชาย จำนวน 1,000 คน และเพศหญิง จำนวน 1,000 คน โดยทำการสุ่มจากเขตกรุงเทพมหานครและปริมณฑลเท่านั้นเพื่อป้องกันการทำหน้าที่ต่างกันที่อาจเกิดจากสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

ตัวแปรสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา ประกอบด้วยเขตกรุงเทพมหานครและปริมณฑล จำนวน 1,000 คน และนอกเขตกรุงเทพมหานครและปริมณฑล จำนวน 1,000 คน

รวมจำนวนกลุ่มตัวอย่างที่ใช้วิเคราะห์วิชาภาษาอังกฤษทั้งหมด จำนวน 4,000 คน และรวมจำนวนกลุ่มตัวอย่างที่ใช้วิเคราะห์วิชาคณิตศาสตร์ทั้งหมด จำนวน 4,000 คน

แบบสอบที่ใช้ในการวิเคราะห์

แบบสอบที่นำมาใช้ในการศึกษาครั้งนี้ ประกอบด้วย แบบสอบคัดเลือกบุคคลเพื่อเข้าศึกษาต่อในสถาบันอุดมศึกษา ปี 2543 ครั้งที่ 1/มีนาคม วิชาภาษาอังกฤษและวิชาคณิตศาสตร์ ซึ่งเป็นแบบสอบที่สร้างขึ้นโดยคณะกรรมการสร้างแบบสอบ ของสำนักทดสอบกลาง ทบวงมหาวิทยาลัย

แบบสอบวิชาภาษาอังกฤษ ประกอบด้วยข้อสอบแบบเลือกตอบ 4 ตัวเลือก จำนวน 100 ข้อ คะแนนเต็ม 100 คะแนน ส่วนแบบสอบวิชาคณิตศาสตร์ ประกอบด้วยข้อสอบแบบเลือกตอบ 4 ตัวเลือก จำนวน 28 ข้อ คะแนนเต็ม 88 คะแนน และข้อสอบแบบเติมคำ จำนวน 6 ข้อ คะแนนเต็ม 12 คะแนน รวม 100 คะแนน ซึ่งในการวิจัยครั้งนี้นำมาใช้เฉพาะข้อสอบแบบเลือกตอบ

สำหรับแบบสอบคัดเลือกทั้งสองฉบับนี้ สำนักทดสอบกลาง ทบวงมหาวิทยาลัย ได้วิเคราะห์คุณภาพของแบบสอบแล้ว ซึ่งผลการวิเคราะห์มีรายละเอียดดังนี้

แบบสอบวิชาภาษาอังกฤษ มีค่าความเที่ยงแบบสอดคล้องภายใน (internal consistency) ตามสูตร KR-20 เท่ากับ 0.87 มีค่าความยาก (p) เฉลี่ยทั้งฉบับเท่ากับ 0.28 และค่าอำนาจจำแนก (r)เฉลี่ยทั้งฉบับ เท่ากับ 0.37 คะแนนสูงสุดเท่ากับ 100 คะแนน คะแนนต่ำสุดเท่ากับ 1 คะแนน และคะแนนเฉลี่ยเท่ากับ 36.43 คะแนน ส่วนแบบสอบวิชาคณิตศาสตร์ 1 ทำการวิเคราะห์เฉพาะข้อสอบแบบเลือกตอบ จำนวน 28 ข้อ เท่านั้น โดยมีค่าความเที่ยงแบบสอดคล้องภายใน (internal consistency) ตามสูตร KR-20 เท่ากับ 0.48 มีค่าความยาก (p) เฉลี่ยทั้งฉบับเท่ากับ 0.31 และค่าอำนาจจำแนก (r) เฉลี่ยทั้งฉบับ เท่ากับ 0.26 คะแนนสูงสุดทั้งฉบับเท่ากับ 100 คะแนน คะแนนต่ำสุดเท่ากับ 0 คะแนน และคะแนนเฉลี่ยเท่ากับ 28.73 คะแนน

การเก็บรวบรวมข้อมูล

ข้อมูลที่ใช้ในการวิจัยครั้งนี้เป็นผลการตอบรายข้อของผู้สมัครสอบวิชาภาษาอังกฤษและวิชาคณิตศาสตร์ ปี 2543 ครั้งที่ 1/มีนาคม โดยแบ่งกลุ่มผู้สอบตามตัวแปรเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา ซึ่งเป็นข้อมูลทุติยภูมิ การเก็บรวบรวมข้อมูลนั้น ผู้วิจัยได้

ดำเนินการโดยทำหนังสือขอความ ร่วมมือในการทำวิจัย จากภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย เพื่อติดต่อขอข้อมูลผลการตอบข้อสอบรายข้อ จำแนกตามกลุ่มผู้สอบ จากสำนักทดสอบกลาง ทบวงมหาวิทยาลัย ซึ่งทางเจ้าหน้าที่ของสำนักทดสอบกลางเป็นผู้ทำการสุ่มข้อมูลให้ โดยการเขียนโปรแกรมเพื่อใช้สุ่มผู้สอบและชุดของผลการตอบตามจำนวนที่ผู้วิจัยกำหนด

การวิเคราะห์ข้อมูล

ผู้วิจัยดำเนินการวิเคราะห์ข้อมูลตามลำดับต่อไปนี้

1. คำนวณค่าสถิติพื้นฐาน เพื่อบรรยายลักษณะการแจกแจงของคะแนนของแบบสอบวิชาภาษาอังกฤษและวิชาคณิตศาสตร์ ด้วยโปรแกรมคอมพิวเตอร์ SPSS 9.05 for Windows

2. ทำการทดสอบความเป็นเอกมิติ (unidimensionality) ของแบบสอบแต่ละฉบับ โดยการวิเคราะห์ตัวประกอบ (factor analysis) ด้วยวิธีวิเคราะห์ตัวประกอบสำคัญ (principle component analysis) และหมุนแกนด้วยวิธี Varimax ตามวิธีการของ Lord (1980) โดยใช้โปรแกรมคอมพิวเตอร์ SPSS 9.05 for Windows

3. วิเคราะห์คุณภาพของแบบสอบวิชาอังกฤษและวิชาคณิตศาสตร์ ด้วยโปรแกรมคอมพิวเตอร์ SPSS 9.05 for Windows โดยคำนวณค่าความเที่ยงเชิงความสอดคล้องภายใน (internal consistency) ของแบบสอบแต่ละฉบับ ในวิชาภาษาอังกฤษคำนวณด้วยสูตร KR-20 ของคูเดอร์ - ริชาร์ดสัน (Kuder-Richardson) ดังนี้

$$r_{tt} = \left[\frac{n}{n-1}\right]\left[1-\frac{\sum pq}{s_x^2}\right]$$

เมื่อ r_{tt} แทน ค่าความเที่ยง

n แทน จำนวนข้อสอบ

$\sum pq$ แทน ผลรวมของความแปรปรวนรายข้อ

s_x^2 แทน ความแปรปรวนของคะแนนรวม

p แทน สัดส่วนของผู้ที่ตอบถูก

q แทน สัดส่วนของผู้ที่ตอบผิด (1-p)

ส่วนแบบสอบวิชาคณิตศาสตร์ ในส่วนของข้อสอบแบบเลือกตอบ มี 2 ตอน คือตอนที่ 1 มี 24 ข้อ ข้อละ 3 คะแนนและตอนที่ 2 มี 4 ข้อ ข้อละ 4 คะแนน จึงทำการคำนวณค่าความเที่ยงด้วยสูตรแอลฟาของครอนบาค (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

หลังจากที่คำนวณค่าความเที่ยงของแบบสอบวิชาคณิตศาสตร์ แล้วพบว่าค่าความเที่ยงที่ได้มีค่าต่ำมาก คือมีค่าเท่ากับ 0.38 อาจทำให้ผลการวิเคราะห์ในขั้นต่อไปมีความคลาดเคลื่อนเกินไป ผู้วิจัยจึงทำการสุ่มตัดผู้สอบที่มีคะแนนต่ำมาก ๆ บางส่วนออกจากกลุ่มตัวอย่าง โดยตัดออกจากกลุ่มย่อย คือ กลุ่มเพศชาย หญิง เขตกรุงเทพและปริมณฑล และนอกเขตกรุงเทพและปริมณฑล กลุ่มละ 100 คน รวมเป็น 400 คน จึงเหลือกลุ่มตัวอย่างทั้งหมด 3,600 คน หลังจากนั้นจึงนำกลุ่มตัวอย่างที่เหลือไปคำนวณค่าสถิติพื้นฐานและหาค่าความเที่ยงใหม่อีกครั้ง ซึ่งค่าความเที่ยงค่าใหม่ที่ได้มีค่าดีขึ้น จึงทำการวิเคราะห์ในขั้นต่อไป

4. เนื่องจากผู้วิจัยไม่มีรายละเอียดเกี่ยวกับโครงสร้างเนื้อหาที่ใช้ในการพัฒนาแบบสอบอย่างเป็นทางการจึงไม่สามารถตรวจสอบความตรงเชิงเนื้อหา (content validity) ได้ ผู้วิจัยจึงวิเคราะห์ คุณภาพของแบบสอบในด้านความตรงโดยทำการตรวจสอบความตรงเชิงโครงสร้าง (construct validity) ของแบบสอบวิชาอังกฤษและวิชาคณิตศาสตร์ ด้วยวิธีวิเคราะห์องค์ประกอบเชิงยืนยัน (confirmatory factor analysis : CFA) เนื่องจากเป็นกระบวนการที่มีเกณฑ์ตัดสินความตรงที่มีความเป็นปรนัย โดยใช้โปรแกรม LISREL 8.10 แต่เนื่องด้วยข้อจำกัดของโปรแกรมไม่สามารถวิเคราะห์ข้อสอบในแบบสอบวิชาภาษาอังกฤษทั้งหมดได้ในครั้งเดียว ผู้วิจัยจึงได้แบ่งข้อสอบในแบบสอบวิชาภาษาอังกฤษตามลักษณะของเนื้อหา ให้มีลักษณะใกล้เคียงกันมากที่สุด ออกเป็นกลุ่มได้ 5 กลุ่มและแยกการวิเคราะห์ออกเป็นโมเดลย่อย ๆ ตามกลุ่มที่แบ่งไว้เพื่อสร้างสเกลองค์ประกอบทั้ง 5 โมเดลจากข้อสอบทั้งหมด 100 ข้อ เพื่อนำไปวิเคราะห์ความตรงเชิงโครงสร้างของโมเดลองค์ประกอบความรู้วิชาภาษาอังกฤษ ซึ่งเป็นโมเดลการวิเคราะห์องค์ประกอบยืนยันอันดับที่สอง

(second order confirmatory factor analysis) ที่ประกอบด้วยโมเดลย่อยดังกล่าว โดยทำการวิเคราะห์ด้วยโปรแกรมคอมพิวเตอร์ LISREL 8.10 การแบ่งข้อสอบในแบบสอบวิชาภาษาอังกฤษออกเป็นกลุ่มย่อยมีรายละเอียดดังนี้

กลุ่มที่ 1 ความเข้าใจในการอ่าน ระดับข้อเท็จจริง จำนวน 25 ข้อ
กลุ่มที่ 2 ความเข้าใจในการอ่าน ระดับตีความ จำนวน 20 ข้อ
กลุ่มที่ 3 ความเข้าใจในการอ่าน ระดับประเมินค่า จำนวน 9 ข้อ
กลุ่มที่ 4 ความสามารถในการใช้ภาษา จำนวน 16 ข้อ
กลุ่มที่ 5 ความรู้เกี่ยวกับโครงสร้างไวยากรณ์ จำนวน 30 ข้อ

สำหรับแบบสอบวิชาคณิตศาสตร์ซึ่งมีข้อสอบทั้งหมดจำนวน 28 ข้อสามารถวิเคราะห์ความตรงเชิงโครงสร้างของแบบสอบด้วยวิธีวิเคราะห์องค์ประกอบเชิงยืนยัน โดยใช้โปรแกรม LISREL 8.10 ได้ในครั้งเดียว หลังจากนั้นจึงดำเนินการวิเคราะห์ดังนี้

4.1 คำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันของข้อสอบแต่ละข้อ ในแต่ละกลุ่มเนื้อหา โดยใช้โปรแกรมคอมพิวเตอร์ SPSS 9.05 for Windows

4.2 วิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) ของแบบสอบวิชาภาษาอังกฤษและวิชาคณิตศาสตร์ โดยใช้โปรแกรมคอมพิวเตอร์ LISREL 8.10 สำหรับแบบสอบวิชาภาษาอังกฤษจะทำการแยกวิเคราะห์ออกเป็นโมเดลย่อย ๆ แต่ละโมเดลตามเนื้อหาที่แบ่งไว้ เพื่อสร้างสเกลขององค์ประกอบทั้ง 5 องค์ประกอบ จากข้อสอบทั้งหมด 100 ข้อ สำหรับการสร้างสเกลองค์ประกอบคำนวณได้จากผลคูณระหว่างสัมประสิทธ์คะแนนองค์ประกอบกับคะแนนมาตรฐาน (standard score) ของข้อสอบแต่ละข้อ ซึ่งสามารถเขียนให้อยู่ในรูปสมการได้ดังนี้

$$F = \sum_{i=1}^{n} a_i Z_i$$

เมื่อ F แทน คะแนนองค์ประกอบตัวที่ i
a แทน ค่าสัมประสิทธิ์คะแนนองค์ประหอบของข้อสอบข้อที่ i
n แทน จำนวนตัวบ่งชี้
Z แทน ค่าคะแนนมาตรฐาน (standard score) ของข้อสอบข้อที่ i

$$Z = \frac{X - \overline{X}}{S.D.}$$

เมื่อ X แทน คะแนนดิบ

$\overline{X}$ แทน ค่าเฉลี่ยของคะแนนดิบ

S.D. แทน ส่วนเบี่ยงเบนมาตรฐานของคะแนนดิบ

การวิเคราะห์ข้อมูลด้วยโปรแกรม LISREL 8.1 ในการวิจัยนี้ ค่าสถิติที่ใช้ตรวจสอบความสอดคล้องกลมกลืนของโมเดลกับข้อมูลผลการตอบของผู้สอบ คือ ค่าสถิติไค-สแควร์ (chi-square) ค่าดัชนีวัดระดับความกลมกลืน (goodness of fit index = GFI) ค่าดัชนีวัดระดับความกลมกลืนที่ปรับแก้แล้ว (adjustes goodness of fit index = AGFI) ค่าดัชนีวัดความสอดคล้องปกติ (normed fit index = NFI) และค่าดัชนีรากกำลังสองเฉลี่ยของเศษ (root mean squared = RMR)

5. ทำการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ จำแนกตามตัวแปรเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา ซึ่งดำเนินการดังนี้

5.1 นำข้อมูลผลการตอบข้อสอบรายข้อที่เตรียมไว้ ซึ่งเป็นข้อมูลแบบมีการให้คะแนนแบบ 2 ค่าคือ ถูกและผิด มาประมาณค่าพารามิเตอร์ของข้อสอบรายข้อ ค่าความสามารถของผู้สอบแต่ละคนและค่าความแปรปรวนร่วมของข้อมูล ด้วยโปรแกรม BILOG version 3.04 (Mislevy and Bock, 1990). ซึ่งมีวิธีประมาณค่าอยู่ 2 แบบ คือแบบ marginal maximum likelihood estimation (MMLE) และแบบ marginal bayesian estimation (MBE) ในการวิจัยครั้งนี้ใช้การประมาณค่าแบบ MBE ทั้งนี้เพราะ Cohen และ Kim (1993) พบว่าการประมาณค่าด้วยวิธีของ Bayes จะทำให้อัตราความคลาดเคลื่อนชนิดที่ 2 น้อยกว่าแบบ MMLE

ผลการวิเคราะห์ข้อสอบด้วยโปรแกรม BILOG 3.04 โดยใช้โมเดล 3 พารามิเตอร์ พบว่ามีข้อสอบบางข้อที่ไม่สามารถประมาณค่าพารามิเตอร์ได้ ดังนั้นผู้วิจัยจึงได้วิเคราะห์ข้อมูลอีกครั้งโดยใช้โมเดล 2 พารามิเตอร์ พบว่าโปรแกรมสามารถประมาณค่าพารามิเตอร์ได้ทุกข้อ ผู้วิจัยจึงเลือกใช้โมเดล 2 พารามิเตอร์ ในการวิเคราะห์ เนื่องจาก Kim และคณะ (1994) กล่าวว่าการใช้โมเดล 2 พารามิเตอร์และขนาดกลุ่มตัวอย่าง 1,000 คน เป็นเงื่อนไขที่จะทำให้ได้ผลการประมาณค่าดีที่สุด

การวิเคราะห์ด้วยโปรแกรม BILOG 3.04 นี้จะต้องมี batch file ซึ่งเป็นแฟ้มให้โปรแกรมดำเนินการวิเคราะห์ตามผู้ใช้กำหนด ในการวิเคราะห์จะแยกกันระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ และแฟ้มข้อมูลผลการตอบข้อสอบรายข้อของผู้สอบแต่ละคนที่ใช้ในการวิเคราะห์จะต้องมีรูปแบบดังนี้คือ บรรทัดแรกเป็นคำตอบที่ถูกต้อง (key) บรรทัดที่สองเป็น omit key ซึ่งมักจะใช้เลข 9 และบรรทัดสามเป็นต้นไปจะเป็นเลขประจำตัวของผู้สอบและผลการตอบรายข้อ ผลที่ได้จากการวิเคราะห์จะประกอบด้วยแฟ้มข้อมูล 3 แฟ้ม ได้แก่ 1) แฟ้มข้อมูลที่ใช้นามสกุล ph1 ซึ่งผลการวิเคราะห์จะประกอบด้วยค่าสถิติรายข้อ ได้แก่ ร้อยละของการตอบข้อสอบถูกในแต่ละข้อ

สัมประสิทธิ์สหสัมพันธ์ระหว่างข้อกับข้อสอบที่เหลือและจำนวนผู้สอบในแต่ละข้อ 2) แฟ้มข้อมูลที่ใช้นามสกุล ph2 ซึ่งผลการวิเคราะห์ประกอบด้วย ค่าพารามิเตอร์ของข้อสอบและผลการทดสอบความเหมาะสมของข้อมูลต่อโมเดล โดยใช้ดัชนี χ^2 และ 3) แฟ้มข้อมูลที่ใช้นามสกุล cov เป็นแฟ้มข้อมูลที่เก็บค่าความแปรปรวนร่วมซึ่งเป็นแฟ้มข้อมูลที่จะนำไปวิเคราะห์ต่อในโปรแกรม EQUATE version 2.0

5.2 หลังจากประมาณค่าพารามิเตอร์ของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบแล้ว ต่อจากนั้นนำมาเทียบมาตรของค่าพารามิเตอร์ สำหรับกลุ่มผู้สอบที่ต้องการตรวจสอบแต่ละกลุ่มให้อยู่บนมาตรวัดเดียวกัน โดยใช้การแปลงสมการเชิงเส้น ด้วยโปรแกรมคอมพิวเตอร์ EQUATE version 2.0 (Baker, 1993) ในวิชาภาษาอังกฤษ จะใช้ค่าพารามิเตอร์ของผู้สอบเพศหญิงเป็นหลักและเปรียบเทียบค่าพารามิเตอร์ของเพศชาย และใช้ค่าพารามิเตอร์ของผู้สอบจากเขตกรุงเทพมหานครและปริมณฑลเป็นหลัก และปรับเทียบค่าพารามิเตอร์ของผู้สอบนอกเขตกรุงเทพมหานครและปริมณฑล ส่วนในวิชาคณิตศาสตร์ จะใช้ค่าพารามิเตอร์ของผู้สอบเพศชายเป็นหลักและเปรียบเทียบค่าพารามิเตอร์ของเพศหญิง และใช้ค่าพารามิเตอร์ของผู้สอบจากเขตกรุงเทพมหานครและปริมณฑลเป็นหลัก และปรับเทียบค่าพารามิเตอร์ของผู้สอบนอกเขตกรุงเทพมหานครและปริมณฑล ซึ่งสมการเชิงเส้นอย่างง่ายที่ใช้แปลงค่าพารามิเตอร์จากมาตรหนึ่งไปหนึ่งยังอีกมาตรหนึ่ง ดังนี้

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

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

เมื่อ a^*_{iF} แทน ค่าอำนาจจำแนกของข้อสอบข้อที่ i กลุ่มผู้สอบกลุ่มเปรียบเทียบที่เทียบมาตรแล้ว

b^*_{iF} แทน ค่าความยากของข้อสอบข้อที่ i กลุ่มผู้สอบกลุ่มเปรียบเทียบที่เทียบมาตรแล้ว

a_{iF} แทน ค่าอำนาจจำแนกของข้อสอบข้อที่ i กลุ่มผู้สอบกลุ่มเปรียบเทียบก่อนเทียบมาตร

b_{iF} แทน ค่าความยากของข้อสอบข้อที่ i กลุ่มผู้สอบกลุ่มเปรียบเทียบก่อนเทียบมาตร

A , K แทน ค่าคงที่ซึ่งจะทำให้ค่าเฉลี่ยและความแปรปรวนของค่า b ที่แปลงแล้วของกลุ่มที่จะนำไปเทียบมาตร มีค่าเท่ากับค่าเฉลี่ยและความแปรปรวนของกลุ่มที่เป็นหลัก โดยที่ A เป็นค่าความชัน (Slope) และ K เป็นค่า intercept

ในการวิเคราะห์ด้วยโปรแกรม EQUATE 2.0 นั้น จะต้องมี batch file ซึ่งเป็นแฟ้มคำสั่งหรือจะเลือกใช้แบบ interaction mode ซึ่งจะมีลักษณะเป็นคำถามถามผู้ใช้ทีละคำถามให้ผู้ใช้ตอบเกี่ยวกับลักษณะข้อมูลและความต้องการของผู้ใช้ ซึ่งในการวิจัยครั้งนี้ผู้วิจัยใช้แบบ interactive mode ข้อมูลที่นำมาใช้เป็นเมตริกซ์ที่ดัดแปลงมาจากเมตริกซ์ความแปรปรวนร่วมที่ได้จากโปรแกรม BILOG 3.04 โดยลบค่าที่ไม่ต้องการใช้ออกไป โปรแกรม EQUATE 2.0 นี้จะใช้เฉพาะค่าพารามิเตอร์ของข้อสอบเท่านั้น ในการวิเคราะห์จะใช้เมตริกซ์ 2 เมตริกซ์ คือเมตริกซ์ของกลุ่มอ้างอิงและเมตริกซ์ของกลุ่มเปรียบเทียบ ซึ่งจะแยกกันอยู่คนละแฟ้ม ผลที่ได้จากโปรแกรม EQUATE 2.0 จะได้ค่าความชัน (A) ค่า intercept (K) และแฟ้มข้อมูลค่าพารามิเตอร์ของข้อสอบที่เทียบมาตรแล้ว

5.3 วิเคราะห์การทำหน้าที่ต่างกันของแบบสอบ (differential test functioning : DTF)

5.3.1 การคำนวณค่าดัชนีการทำหน้าที่ต่างกันของแบบสอบ (DTF index) จากดัชนีการทำหน้าที่ต่างกันของข้อสอบแบบชดเชย (compensatory DIF index : CDIF) มีสูตรการคำนวณ ดังนี้

$$DTF = \sum_{i=1}^{n} \left[Cov(d_i, D) + \mu_{di}\mu_D \right] \quad \text{..... 18)}$$

ดัชนีการทำหน้าที่ต่างกันที่ระดับแบบสอบ $(CDIF)$ มีค่าเท่ากับ

$$CDIF_i = Cov(d_i, D) + \mu_{di}\mu_D \quad \text{..... 19)}$$

ดังนั้นจากสมการที่ 18) และ 19) จะได้ว่า

$$DTF = \sum_{i=1}^{n} CDIF_i$$

เมื่อ

$CDIF_i$ แทน ดัชนีการทำหน้าที่ต่างกันของข้อสอบแบบชดเชย (compensatory DIF index : CDIF) ของข้อสอบข้อที่ i

d_i แทน ความแตกต่างของความน่าจะเป็นของการตอบข้อสอบข้อที่ i ถูกต้องของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ โดยที่ $d_{is} = P_{iF}(\theta_s) - P_{iR}(\theta_s)$ เป็นผลต่างของความน่าจะเป็นของการตอบข้อสอบข้อ i ถูกของผู้สอบคนที่ s

$P_{iF}(\theta_s)$ และ $P_{iR}(\theta_s)$ แทน ความน่าจะเป็นของการตอบข้อสอบถูกของผู้สอบ s ซึ่งเป็นสมาชิกของกลุ่มเปรียบเทียบและกลุ่มอ้างอิง ที่

ข้อสอบข้อที่ i ตามลำดับ

D แทน ความแตกต่างระหว่างคะแนนจริงของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ โดยที่ $D_s = T_{sR} - T_{sF}$ เป็นผลต่างของคะแนน

จริงหรือสัดส่วนของการตอบถูกเมื่ออยู่กลุ่มย่อยต่างกัน เมื่อสัดส่วนของการตอบข้อสอบถูกของผู้สอบ (an examinee's expected proportion : EPC) หรือคะแนนจริง (T) มีค่าเท่ากับ $T_s = \sum_{i=1}^{n} P_i(\theta_s)$

$Cov(d_i,D)$ แทน ความแปรปรวนร่วมระหว่างความแตกต่างของความน่าจะเป็นของการตอบข้อสอบข้อที่ i ถูก กับความแตกต่างระหว่างคะแนนจริงของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ

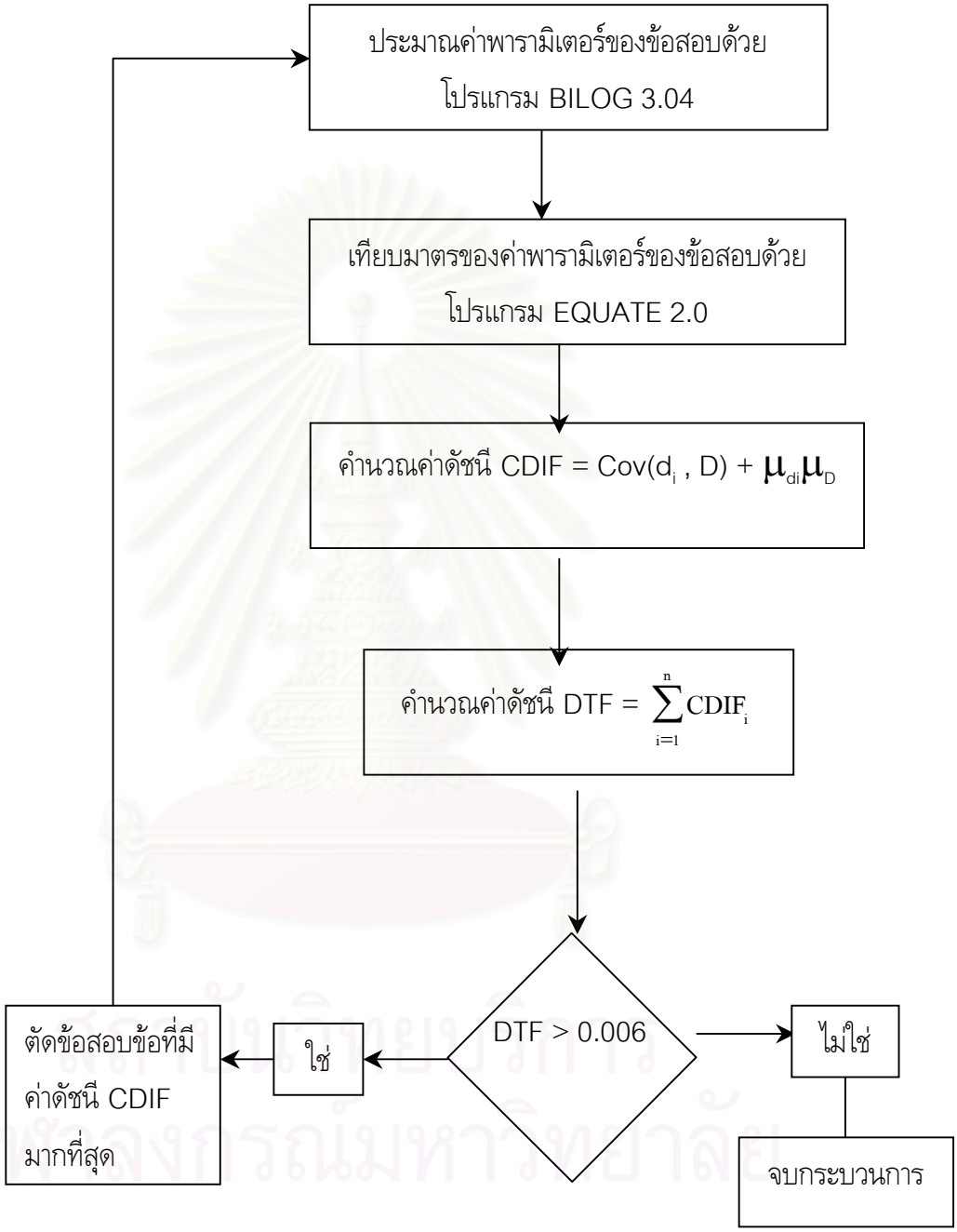
μ_{di} แทน ค่าเฉลี่ยของผลต่างของความน่าจะเป็นในการตอบข้อสอบข้อ i

ถูกต้อง (d_{is})

μ_D แทน ค่าเฉลี่ยของผลต่างของคะแนนจริง (D_s)

5.3.2 ถ้าค่าดัชนี DTF ที่คำนวณได้มีค่ามากกว่า 0.006 แสดงว่าแบบสอบดังกล่าวทำหน้าที่ต่างกัน (Fleer , 1993 อ้างถึงใน Raju และคณะ ,1995) จะตัดข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด คือข้อสอบที่มีค่าดัชนี CDIF/DTF เป็นบวกและมีค่ามากที่สุดออกจากแบบสอบ แล้วทำการประมาณค่าพารามิเตอร์ของข้อสอบที่เหลือด้วยโปรแกรม BILOG 3.04 หลังจากนั้นเทียบมาตรของค่าพารามิเตอร์ของข้อสอบที่เหลือ ให้อยู่บนมาตรเดียวกันสำหรับกลุ่มผู้สอบแต่ละกลุ่ม ด้วยโปรแกรม EQUATE 2.0 แล้วจึงทำการคำนวณค่าดัชนี DTF ทำซ้ำเช่นนี้จนกว่าค่าดัชนี DTF ที่ได้มีค่าน้อยกว่าหรือเท่ากับ 0.006 ดังนั้นสามารถสรุปกระบวนการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบได้ดัง แผนภาพที่ 4

แผนภาพที่ 4 ขั้นตอนการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบ (DTF)



เมื่อ

$$d_{is} = P_{iF}(\theta_s) - P_{iR}(\theta_s)$$

$$D_s = \sum_{i=1}^{n} P_{Fi}(\theta_s) - \sum_{i=1}^{n} P_{Ri}(\theta_s)$$

5.4 วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (differential item functioning : DIF) โดยการคำนวณดัชนีการทำหน้าที่ต่างกันของข้อสอบแบบไม่ชดเชย (non-compensatory DIF index : NCDIF) มีสูตรการคำนวณดังนี้

$$NCDIF_i = \sigma^2_{d_i} + \mu^2_{d_i}$$

โดยที่

$NCDIF_i$ แทน ดัชนีการทำหน้าที่ต่างกันของข้อสอบแบบไม่ชดเชย (non-compensatory DIF index) ของข้อสอบข้อที่ i

$\sigma^2_{d_i}$ แทน ความแปรปรวนของความแตกต่างระหว่างความน่าจะเป็นในการตอบข้อสอบข้อที่ i ถูกของกลุ่มเปรียบเทียบและกลุ่มอ้างอิง

μ_{d_i} แทน ค่าเฉลี่ยของความแตกต่างของความน่าจะเป็นในการตอบข้อสอบข้อที่ i ถูกระหว่างกลุ่มเปรียบเทียบและกลุ่มอ้างอิง

6. วิเคราะห์คุณภาพของแบบสอบหลังจากการตัดข้อสอบที่ทำหน้าที่ต่างกันออกแล้ว จำแนกตามตัวแปรต่าง ๆ โดยดำเนินการดังนี้

6.1 คำนวณค่าความเที่ยงเชิงความสอดคล้องภายในของแบบสอบแต่ละฉบับหลังการตัดข้อสอบที่ทำหน้าที่ต่างกันออก โดยใช้สูตร KR-20 สำหรับแบบสอบวิชาภาษาอังกฤษ และใช้สูตร แอลฟาของครอนบาค (Cronbach's Alpha) สำหรับแบบสอบวิชาคณิตศาสตร์

6.2 วิเคราะห์ความตรงเชิงโครงสร้างของแบบสอบหลังการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ ด้วยวิธีวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) โดยตัดข้อสอบทั้งหมดและบางข้อ ที่ทำหน้าที่ต่างกันออกไป เมื่อจำแนกผู้สอบตามตัวแปรต่าง ๆ ด้วยโปรแกรมคอมพิวเตอร์ LISREL 8.10

7. เปรียบเทียบคุณภาพของแบบสอบด้านความเที่ยง ก่อนและหลังการตัดข้อสอบทั้งหมดหรือบางข้อที่ทำหน้าที่ต่างกันออกจากแบบสอบ โดยดำเนินการดังนี้

7.1 ทำการปรับขยายจำนวนข้อสอบให้เท่ากับจำนวนข้อสอบก่อนตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบสอบในทุกกรณีเพื่อให้จำนวนข้อสอบเท่ากัน และสามารถนำมาเปรียบเทียบกันได้ จากนั้นคำนวณค่าความเที่ยงของแบบสอบฉบับใหม่ โดยใช้สูตรของสเปียร์แมน (Spearman C.) ดังนี้

$$r_{tt}' = \frac{N(r_{tt})}{1+(N-1)r_{tt}}$$

เมื่อ

r_{tt}' แทน ค่าความเที่ยงของแบบสอบฉบับใหม่

r_{tt} แทน ค่าความเที่ยงของแบบสอบฉบับเดิม

N แทน จำนวนเท่าของข้อสอบที่เพิ่มขึ้น

7.2 ทดสอบความมีนัยสำคัญความแตกต่างของค่าความเที่ยงฉบับใหม่ที่ได้จากการปรับขยายจำนวนข้อสอบหลังจากตัดข้อสอบที่ทำหน้าที่ต่างกันในทุกกรณี ออกจากแบบสอบฉบับเดิมกับค่าความเที่ยงของแบบสอบฉบับเดิม โดยใช้สูตรดังนี้ (Fledt ,1980 อ้างถึงใน พิศิษฐ ตัณฑวณิช ,2532)

$$t = \frac{(r_{tt}-r_{tt}')\sqrt{(N-2)}}{\sqrt{4(1-r_{tt})(1-r_{tt}')(1-r_{t1t2})}}$$

เมื่อ

t แทน ค่าสถิติทดสอบที (t-test)

r_{tt} แทน ค่าความเที่ยงของแบบสอบฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกันออก

r_{tt}' แทน ค่าความเที่ยงของแบบสอบฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันออก

r_{t1t2} แทน ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนที่ได้จากแบบสอบฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบและแบบสอบ

N แทน จำนวนกลุ่มตัวอย่าง

8. เปรียบเทียบคุณภาพด้านความตรงเชิงโครงสร้างของโมเดลแบบสอบวิชาภาษาอังกฤษและวิชาคณิตศาสตร์ โดยพิจารณาจากค่าสถิติวัดระดับความกลมกลืน (goodness of fit measures) ของแบบสอบทั้งกรณีก่อนและหลังการตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบสอบ

9. คำนวณค่าฟังก์ชันสารสนเทศของแบบสอบ (test information function : TIF) ของ แบบสอบแต่ละฉบับ ทั้งในกรณีก่อนตัดข้อสอบที่พบ DIF และหลังตัดข้อสอบที่พบ DIF ออกจากแบบสอบทุกข้อและตัดเฉพาะบางข้อจนแบบสอบไม่ทำหน้าที่ต่างกัน ณ ระดับความสามารถ θ ต่าง ๆ โดยนำค่าพารามิเตอร์ของข้อสอบในแบบสอบแต่ละฉบับมาคำนวณค่าฟังก์ชันสารสนเทศของข้อสอบ หลังจากนั้นนำค่าฟังก์ชันสารสนเทศของข้อสอบมาคำนวณค่าฟังก์ชันสารสนเทศของแบบ

สอบ ณ ระดับความสามารถต่าง ๆ ซึ่งแบ่งระดับความสามารถออกเป็น 13 ระดับ คือ –3.0, -2.5, -2.0, -1.5, -1.0, -0.5, 0.0, 0.5, 1.5, 2.0, 2.5 และ 3.0 สูตรที่ใช้ในการคำนวณมีดังนี้ (Hambleton and Swaminathan, 1985)

$$I(\theta) = \sum \frac{[P_i'(\theta)]^2}{P_i(\theta)Q_i(\theta)}$$

เมื่อ $I(\theta)$ แทน ค่าฟังก์ชันสารสนเทศของแบบสอบ ณ ระดับความสามารถ θ

$P_i(\theta)$ แทน โอกาสในการตอบข้อสอบข้อที่ i ได้ถูกของผู้สอบที่มีระดับความสามารถระดับ θ โดยที่ $P_i(\theta)$ สำหรับโมเดลโลจิสติกแบบสองพารามิเตอร์ คำนวณได้จากสูตร

$$P_i(\theta) = \frac{1}{1 + e^{-Da_i(\theta - b_i)}}$$

โดยที่ a_i แทนค่าอำนาจจำแนกของข้อสอบข้อที่ i , b_i แทนค่าความยากของข้อสอบที่ i , e เป็นค่าคงที่ของลอกาลิทึมธรรมชาติ ซึ่งมีค่าประมาณ 2.71828 และ D เป็นค่าองค์ประกอบของการปรับสเกลซึ่งมีค่าเท่ากับ 1.70

$P_i'(\theta)$ แทน ค่าอนุพันธ์ของโค้งคุณลักษณะข้อสอบ (ICC) ข้อที่ i ณ ระดับความสามารถ θ โดยที่ $P_i'(\theta)$ สำหรับโมเดลโลจิสติกแบบสองพารามิเตอร์ มีค่าเท่ากับ $Da_iP_iQ_i$

$Q_i(\theta)$ แทน $1-P_i(\theta)$

ในการคำนวณค่าฟังก์ชันสารสนเทศของแบบสอบฉบับก่อนการตัดข้อสอบที่พบ DIF จะนำเฉพาะค่าฟังก์ชันสารสนเทศของข้อสอบที่ไม่ถูกตัดออกหลังจากการวิเคราะห์ DIF มาคำนวณค่าฟังก์ชันสารสนเทศของแบบสอบ เพื่อให้แบบสอบทั้งสองฉบับสามารถนำมาเปรียบเทียบกันได้

10. คำนวณค่าดัชนีประสิทธิภาพสัมพัทธ์ (relative efficiency : RE) ของแบบสอบแต่ละฉบับ ณ ระดับความสามารถต่าง ๆ ดังนี้

$$RE(\theta) = \frac{I_{หลัง}(\theta)}{I_{ก่อน}(\theta)}$$

เมื่อ $I_{หลัง}(\theta)$ แทน ค่าฟังก์ชันสารสนเทศของแบบสอบฉบับหลังการตัดข้อสอบที่พบ DIF

$I_{ก่อน}(\theta)$ แทน ค่าฟังก์ชันสารสนเทศของแบบสอบฉบับก่อนการตัดข้อสอบที่พบ DIF

การแปลความหมายของค่าดัชนีประสิทธิภาพสัมพัทธ์

RE = 1 แสดงว่าแบบทั้งสองฉบับมีประสิทธิภาพเท่าเทียมกัน

RE > 1 แสดงว่าแบบสอบฉบับหลังการตัดข้อสอบที่พบ DIF มีประสิทธิภาพมากกว่าแบบสอบฉบับก่อนการตัดข้อสอบที่พบ DIF

RE < 1 แสดงว่าแบบสอบฉบับหลังการตัดข้อสอบที่พบ DIF มีประสิทธิภาพต่ำกว่าแบบสอบฉบับก่อนการตัดข้อสอบที่พบ DIF

11. รวมคะแนนของผู้สอบและจัดอันดับคะแนนของผู้สอบจากแบบสอบแต่ละฉบับทั้งในกรณีก่อนตัดข้อสอบที่พบ DIF หลังจากตัดข้อสอบที่พบ DIF ทุกข้อ โดยใช้ดัชนี NCDIF และหลังจากตัดข้อสอบที่พบ DIF บางข้อ โดยใช้ดัชนี CDIF/DTF เพื่อจัดอันดับผู้สอบ จำแนกตามตัวแปรที่ศึกษา

12. คำนวณค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตำแหน่งของคะแนนรวมของผู้สอบก่อนและหลังการตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบสอบ ทั้งกรณีที่ตัดข้อสอบที่ตรวจพบว่าทำหน้าที่ต่างกันออกจากแบบสอบทุกข้อ และกรณีที่ตัดเฉพาะบางข้อจนแบบสอบทั้งฉบับไม่ทำหน้าที่ต่างกัน โดยใช้สูตรการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนแรงค์ (Spearman rank correlation coefficient) ดังนี้

$$r_s = 1 - \frac{6\sum D^2}{n(n^2 - 1)}$$

เมื่อ

r_s แทน ค่าสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนแรงค์

D แทน ความแตกต่างระหว่างลำดับที่ของคะแนนชุดก่อนและหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน

n แทน จำนวนคู่

และทดสอบความมีนัยสำคัญของสัมประสิทธิ์สหสัมพันธ์ โดยใช้สูตร

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

เมื่อ

r แทน ค่าสัมประสิทธิ์สหสัมพันธ์

n แทน จำนวนคู่

บทที่ 4

ผลการวิเคราะห์ข้อมูล

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ ด้วยกระบวนการ DFIT สำหรับแบบสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษา วิชาภาษาอังกฤษและวิชาคณิตศาสตร์ โดยแบ่งกลุ่มผู้สอบตามตัวแปรเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา ผลการวิเคราะห์ข้อมูลนั้น ผู้วิจัยได้นำเสนอเป็น 5 ตอน ดังนี้

ตอนที่ 1 ผลการวิเคราะห์ข้อมูลเบื้องต้น

ตอนที่ 2 ผลการตรวจสอบคุณภาพของแบบสอบก่อนการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ

ตอนที่ 3 ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบด้วยกระบวนการ DFIT

ตอนที่ 4 ผลการเปรียบเทียบคุณภาพของแบบสอบก่อนและหลังการตัดข้อสอบที่ทำหน้าที่ต่างกัน

ตอนที่ 5 ผลการวิเคราะห์ความสัมพันธ์ระหว่างตำแหน่งของคะแนนรวมของผู้สอบก่อนและหลังการตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบสอบ

ตอนที่ 1 ผลการวิเคราะห์ข้อมูลเบื้องต้น

ผู้วิจัยเสนอผลการวิเคราะห์ข้อมูลเบื้องต้น ของคะแนนสอบวิชาภาษาอังกฤษและวิชาคณิตศาสตร์ สำหรับการสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษา โดยทำการสุ่มกลุ่มตัวอย่างผู้สมัครสอบ จำนวน 4,000 คน ซึ่งมีรายละเอียด ดังแสดงในตารางที่ 4

ตารางที่ 4 ลักษณะการแจกแจงของคะแนนสอบวิชาภาษาอังกฤษและวิชาคณิตศาสตร์

ค่าสถิติ	ภาษาอังกฤษ	คณิตศาสตร์
คะแนนเต็ม	100	88
ค่ามัชฌิมเลขคณิต ($\overline{X}$)	32.66	29.90
ส่วนเบี่ยงเบนมาตรฐาน (S.D.)	11.12	11.93
ค่าต่ำสุด (Min)	13	0
ค่าสูงสุด (Max)	92	73
พิสัย (Range)	79	73
ฐานนิยม (Mode)	24	24
มัธยมฐาน (Med)	30	28
ความเบ้ (Sk)	1.50	0.53
ความโด่ง (Ku)	2.77	0.03
จำนวนข้อสอบแบบหลายตัวเลือก	100	28
จำนวนกลุ่มตัวอย่าง (n)	4,000	3,600

จากตารางที่ 4 พบว่าแบบสอบวิชาภาษาอังกฤษมีคะแนนเต็มในส่วนที่เป็นแบบสอบแบบเลือกตอบ 100 คะแนน ผู้สอบส่วนใหญ่ได้คะแนนค่อนข้างต่ำ มีพิสัยของคะแนนเท่ากับ 79 โดยมีคะแนนต่ำสุดเท่ากับ 13 และมีคะแนนสูงสุดเท่ากับ 92 มีค่าเฉลี่ยของคะแนนเท่ากับ 32.66 และค่าส่วนเบี่ยงเบนมาตรฐาน เท่ากับ 11.12 สำหรับลักษณะการกระจายของคะแนนมีลักษณะเบ้ขวา และมีค่าความโด่ง เท่ากับ 2.77

สำหรับแบบสอบวิชาคณิตศาสตร์ มีคะแนนเต็มในส่วนที่เป็นแบบสอบแบบเลือกตอบ 88 คะแนน ผู้สอบส่วนใหญ่ได้คะแนนค่อนข้างต่ำ มีพิสัยของคะแนนเท่ากับ 73 โดยมีคะแนนต่ำสุดเท่ากับ 0 และมีคะแนนสูงสุดเท่ากับ 73 มีค่าเฉลี่ยของคะแนนเท่ากับ 29.90 และค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 11.93 สำหรับลักษณะการกระจายของคะแนนมีลักษณะเบ้ขวาเล็กน้อย และมีค่าความโด่ง เท่ากับ 0.03

ตอนที่ 2 ผลการวิเคราะห์คุณภาพของแบบสอบก่อนวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ

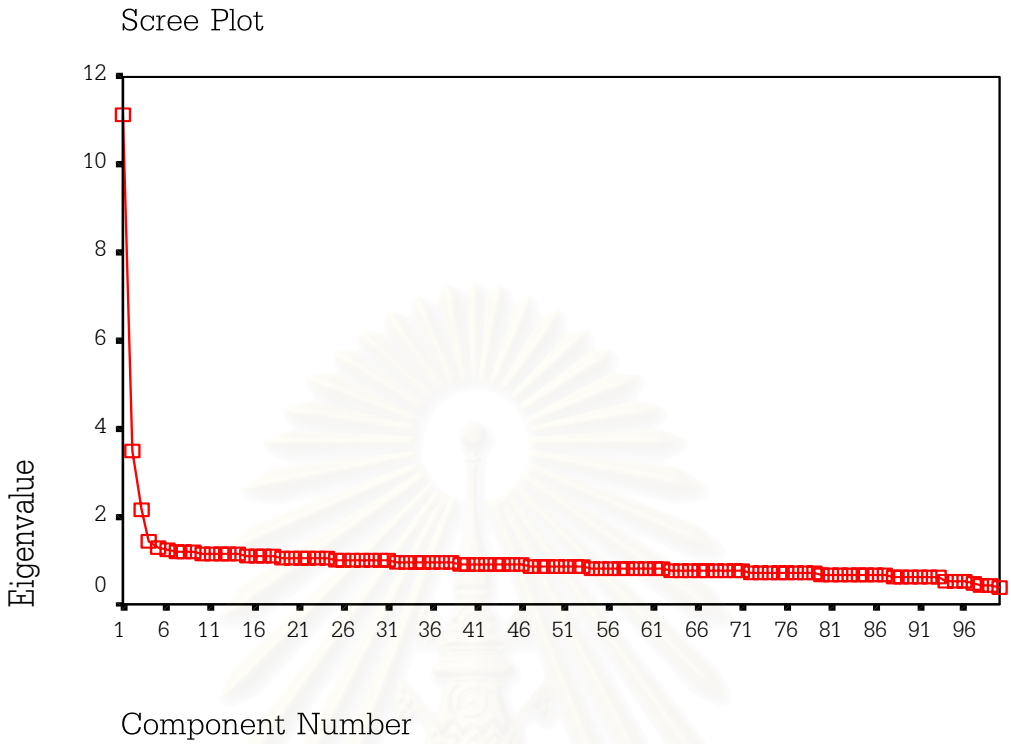
2.1 ผลการทดสอบความเป็นเอกมิติ (unidimensional) ของแบบสอบวิชาภาษาอังกฤษและแบบสอบวิชาคณิตศาสตร์

การตรวจสอบความเป็นเอกมิติของแบบสอบ โดยวิธีการวิเคราะห์ตัวประกอบ (factor analysis) ด้วยวิธีการวิเคราะห์ตัวประกอบสำคัญ (principle component analysis) และหมุนแกนด้วยวิธีแวริแมกซ์ (varimax) โดยใช้โปรแกรมคอมพิวเตอร์ SPSS 9.05 for Windows ในการวิเคราะห์ ผลการวิเคราะห์แบบสอบวิชาภาษาอังกฤษ แสดงดังตารางที่ 5 และผลการวิเคราะห์แบบสอบวิชาคณิตศาสตร์ แสดงดังตารางที่ 6

ตารางที่ 5 ค่าไอเกนและร้อยละของความแปรปรวนของตัวประกอบวิชาภาษาอังกฤษ

ตัวประกอบที่	ผลการสกัดตัวประกอบ (Extraction)		ผลการหมุนแกนตัวประกอบ (Rotation)	
	ค่าไอเกน (Eigenvalue)	ร้อยละของความแปรปรวน (% Variance)	ค่าไอเกน (Eigenvalue)	ร้อยละของความแปรปรวน (% Variance)
1	11.14	11.14	8.33	8.33
2	3.47	3.47	2.53	2.53
3	2.16	2.16	2.49	2.49
4	1.42	1.42	1.63	1.63
5	1.29	1.29	1.63	1.63
6	1.23	1.23	1.42	1.42
7	1.20	1.20	1.39	1.39
8	1.19	1.19	1.34	1.34
9	1.18	1.18	1.31	1.31
10	1.17	1.17	1.29	1.29

จากตารางที่ 5 ผลการสกัดตัวประกอบ (extraction) และการหมุนแกนตัวประกอบ (rotation) พบว่าค่าไอเกนและความแปรปรวนของตัวประกอบตัวแรกมีค่าสูงกว่าตัวประกอบตัวที่ 2 สัดส่วนค่า ไอเกนของตัวประกอบที่ 1 ต่อ ตัวประกอบที่ 2 เป็น 3.3 เท่าและตัวประกอบที่สองก็มีค่าไม่ต่างจากตัวประกอบอื่น ๆ มากนัก จึงทำให้สามารถสรุปได้ว่าแบบสอบวิชาภาษาอังกฤษวัดมิติเด่นเพียงมิติเดียว และค่าไอเกนที่ได้จากการสกัดตัวประกอบแสดงได้ดังแผนภาพที่ 5

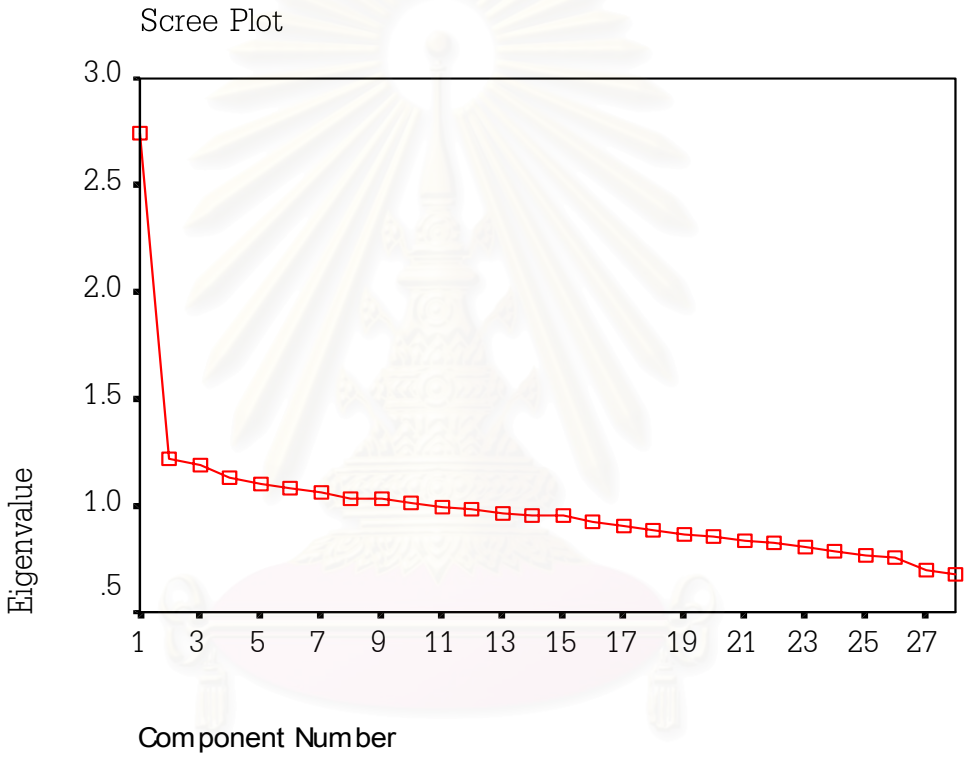


แผนภาพที่ 5 ค่าไอเกนของตัวประกอบที่ได้จากการสกัดตัวประกอบในแบบสอบวิชาภาษาอังกฤษ

ตารางที่ 6 ค่าไอเกนและร้อยละของความแปรปรวนของตัวประกอบวิชาคณิตศาสตร์

ตัวประกอบที่	ผลการสกัดตัวประกอบ (Extraction)		ผลการหมุนแกนตัวประกอบ (Rotation)	
	ค่าไอเกน (Eigenvalue)	ร้อยละของความแปรปรวน (% Variance)	ค่าไอเกน (Eigenvalue)	ร้อยละของความแปรปรวน (% Variance)
1	2.77	9.81	2.00	7.03
2	1.21	4.33	1.39	4.97
3	1.18	4.23	1.35	4.84
4	1.12	4.02	1.25	4.45
5	1.10	3.93	1.16	4.16
6	1.08	3.85	1.13	4.04
7	1.06	3.80	1.11	3.97
8	1.03	3.69	1.09	3.88
9	1.03	3.67	1.07	3.82
10	1.01	3.63	1.06	3.81

จากตารางที่ 6 ผลการสกัดตัวประกอบ (extraction) และการหมุนแกนตัวประกอบ (rotation) พบว่าค่าไอเกนและความแปรปรวนของตัวประกอบตัวแรกมีค่าสูงกว่าตัวประกอบตัวที่สอง และ ตัวประกอบที่สองก็มีค่าไม่ต่างจากตัวประกอบอื่น ๆ มากนัก จึงทำให้พอสรุปได้ว่าแบบสอบวิชาภาษาอังกฤษวัดมิติเด่นเพียงมิติเดียวและค่าไอเกนที่ได้จากการสกัดตัวประกอบแสดงได้ดัง แผนภาพที่ 6



แผนภาพที่ 6 ค่าไอเกนของตัวประกอบที่ได้จากการสกัดตัวประกอบในแบบสอบวิชาคณิตศาสตร์

2.2 ผลการวิเคราะห์คุณภาพของแบบสอบก่อนการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ สำหรับแบบสอบวิชาภาษาอังกฤษและวิชาคณิตศาสตร์ โดยทำการวิเคราะห์ตามทฤษฎีแบบดั้งเดิม (classical test theory : CTT) ซึ่งผู้วิจัยจะนำเสนอ ผลการวิเคราะห์ในด้านค่าความยาก ค่าอำนาจจำแนก และค่าความเที่ยงแบบความสอดคล้องภายใน (internal consistency) ของแบบสอบซึ่งมีรายละเอียดดังตารางที่ 7 และตารางที่ 8

ตารางที่ 7 ค่าความยาก อำนาจจำแนก และค่าความเที่ยงของแบบสอบวิชาภาษาอังกฤษ

ข้อ	ค่าความยาก (p)	ค่าอำนาจจำแนก (r)	ข้อ	ค่าความยาก (p)	ค่าอำนาจจำแนก (r)
1	0.71	0.44	26	0.38	0.22
2	0.49	0.32	27	0.68	0.33
3	0.40	0.65	28	0.34	0.53
4	0.30	0.26	29	0.52	0.34
5	0.77	0.24	30	0.55	0.23
6	0.40	0.58	31	0.39	0.20
7	0.22	-0.01	32	0.37	0.22
8	0.28	0.30	33	0.42	0.24
9	0.67	0.42	34	0.39	0.50
10	0.33	0.29	35	0.35	0.43
11	0.37	0.59	36	0.31	0.22
12	0.39	0.49	37	0.41	0.22
13	0.29	0.19	38	0.50	0.39
14	0.19	0.06	39	0.30	0.27
15	0.18	0.18	40	0.34	0.40
16	0.33	0.43	41	0.38	0.29
17	0.44	0.37	42	0.57	0.41
18	0.34	0.18	43	0.39	0.30
19	0.39	0.15	44	0.49	0.35
20	0.32	0.23	45	0.29	0.26
21	0.33	0.25	46	0.34	0.21
22	0.30	0.08	47	0.20	0.19
23	0.53	0.37	48	0.30	0.15
24	0.35	0.55	49	0.32	0.41
25	0.19	0.16	50	0.34	0.29

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 7 (ต่อ)

ข้อ	ค่าความยาก (p)	ค่าอำนาจจำแนก (r)	ข้อ	ค่าความยาก (p)	ค่าอำนาจจำแนก (r)
51	0.19	0.19	76	0.27	0.18
52	0.17	0.13	77	0.44	0.33
53	0.34	0.17	78	0.23	0.17
54	0.26	-0.03	79	0.42	0.13
55	0.27	0.12	80	0.54	0.27
56	0.25	0.05	81	0.23	0.16
57	0.26	0.28	82	0.31	0.26
58	0.31	0.08	83	0.22	0.13
59	0.25	0.16	84	0.21	0.23
60	0.38	0.17	85	0.35	0.10
61	0.20	0.06	86	0.19	0.10
62	0.32	0.20	87	0.35	0.20
63	0.30	0.20	88	0.32	0.10
64	0.25	0.03	89	0.18	0.12
65	0.35	0.32	90	0.40	0.17
66	0.48	0.39	91	0.26	0.28
67	0.31	0.23	92	0.27	0.14
68	0.58	0.35	93	0.60	0.41
69	0.29	0.04	94	0.22	0.02
70	0.30	0.52	95	0.17	0.17
71	0.40	0.46	96	0.36	0.31
72	0.48	0.18	97	0.25	0.12
73	0.55	0.44	98	0.27	0.12
74	0.17	0.05	99	0.24	0.20
75	0.30	0.12	100	0.26	0.27
ค่าความเที่ยง (KR-20)				0.85	

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

จากตารางที่ 7 เมื่อพิจารณาผลการวิเคราะห์คุณภาพของแบบสอบวิชาภาษาอังกฤษ พบว่ามีความเที่ยงเท่ากับ 0.85 เมื่อพิจารณาเกี่ยวกับค่าความยาก (p) จะเห็นว่าข้อสอบที่มีค่าความยากสูงสุด ได้แก่ข้อสอบข้อที่ 5 มีค่าความยากเท่ากับ 0.77 และข้อสอบที่มีค่าความยากต่ำที่สุด ได้แก่ข้อสอบข้อที่ 52, 74 และ 95 มีค่าความยากเท่ากับ 0.17

เมื่อแบ่งช่วงของค่าความยากข้อสอบตามเกณฑ์การแปลความหมาย (ศิริชัย กาญจนวาสี ,2541) พบว่า ข้อสอบที่มีค่าความยากอยู่ในช่วง 0.00-0.19 มีจำนวน 9 ข้อ คิดเป็นร้อยละ 9.00 ได้แก่ข้อสอบข้อที่ 14, 15, 25, 51, 74, 86, 89 และ 95 ซึ่งเป็นข้อสอบที่มีค่าความยากอยู่ในระดับยากมาก

ข้อสอบที่มีค่าความยากอยู่ในช่วง 0.20-0.39 มีจำนวน 65 ข้อ คิดเป็นร้อยละ 65.00 ได้แก่ ข้อสอบข้อที่ 4, 7, 8, 10, 11, 12, 13, 16, 18, 19, 20, 21, 22, 24, 26, 28, 31, 32, 34, 35, 36, 39, 40, 41, 43, 45, 46, 47, 48, 49, 50, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 67, 69, 70, 75, 76, 78, 81, 82, 83, 84, 85, 87, 88, 91, 92, 94, 96, 97, 98, 99 และ 100 ซึ่งเป็นข้อสอบที่มีค่าความยากอยู่ในระดับค่อนข้างยาก

ข้อสอบที่มีค่าความยากอยู่ในช่วง 0.40-0.59 มีจำนวน 21 ข้อ คิดเป็นร้อยละ 21.00 ได้แก่ ข้อสอบข้อที่ 2, 3, 6, 17, 23, 29, 30, 33, 37, 38, 42, 44, 66, 68, 71, 72, 73, 77, 79, 80 และ 90 ซึ่งเป็นข้อสอบที่มีค่าความยากอยู่ในระดับปานกลาง

ข้อสอบที่มีค่าความยากอยู่ในช่วง 0.60-0.79 มีจำนวน 5 ข้อ คิดเป็นร้อยละ 5.00 ได้แก่ ข้อสอบข้อที่ 1, 5, 9, 27 และ 93 ซึ่งเป็นข้อสอบที่มีค่าความยากอยู่ในระดับค่อนข้างง่าย

เมื่อพิจารณาค่าอำนาจจำแนก (r) พบว่าข้อสอบที่มีค่าอำนาจจำแนกสูงสุดได้แก่ข้อสอบข้อที่ 3 โดยมีค่าอำนาจจำแนกเท่ากับ 0.65 และข้อสอบที่มีค่าอำนาจจำแนกต่ำสุด คือข้อสอบข้อที่ 54 มีค่าอำนาจจำแนกเท่ากับ -0.03 และเมื่อแบ่งค่าอำนาจจำแนกออกเป็นช่วง ๆ ตามเกณฑ์การแปลความหมาย (ศิริชัย กาญจนวาสี ,2541) พบว่า

ข้อสอบที่มีค่าอำนาจจำแนกอยู่ในช่วง ต่ำกว่า 0-0.09 มีจำนวน 11 ข้อ คิดเป็นร้อยละ 11.00 ได้แก่ข้อสอบข้อที่ 7, 14, 22, 54, 56, 58, 61, 64, 69, 74 และ 94 ซึ่งเป็นข้อสอบที่มีค่าอำนาจจำแนกอยู่ในเกณฑ์ต่ำ ควรตัดออก

ข้อสอบที่มีค่าอำนาจจำแนกอยู่ในช่วง 0.10-0.19 มีจำนวน 29 ข้อ คิดเป็นร้อยละ 29.00 ได้แก่ข้อสอบข้อที่ 13, 15, 18, 19, 25, 47, 48, 51, 52, 53, 55, 59, 60, 72, 75, 76, 78, 79, 81, 83, 85, 86, 88, 89, 90, 92, 95, 97 และ 98 ซึ่งเป็นข้อสอบที่มีค่าอำนาจจำแนกอยู่ในเกณฑ์ค่อนข้างต่ำ

ข้อสอบที่มีค่าอำนาจจำแนกอยู่ในช่วง 0.20-0.39 มีจำนวน 42 ข้อ คิดเป็นร้อยละ 42.00 ได้แก่ข้อสอบข้อที่ 2, 4, 5, 8, 10, 17, 20, 21, 23, 26, 27, 29, 30, 31, 32, 33, 36, 37, 38, 39,

41, 43, 44, 45, 46, 50, 57, 62, 63, 65, 66, 67, 68, 77, 80, 82, 84, 87, 91, 96, 99 และ 100 ซึ่งเป็น ข้อสอบที่มีค่าอำนาจจำแนกอยู่ในเกณฑ์พอใช้

ข้อสอบที่มีค่าอำนาจจำแนกอยู่ในช่วง 0.40-0.59 มีจำนวน 17 ข้อ คิดเป็นร้อยละ 17.00 ได้แก่ข้อสอบข้อที่ 1, 6, 9, 11, 12, 16, 24, 28, 34, 35, 40, 42, 49, 70, 71, 73 และ 93 ซึ่งเป็นข้อสอบที่มีค่าอำนาจจำแนกอยู่ในเกณฑ์ดี

ข้อสอบที่มีค่าอำนาจจำแนกอยู่ในช่วง 0.60-1.00 มีจำนวน 1 ข้อ คิดเป็นร้อยละ 1 ได้แก่ ข้อสอบข้อที่ 3 ซึ่งเป็นข้อสอบที่มีค่าอำนาจจำแนกอยู่ในเกณฑ์ดีมาก

ตารางที่ 8 ค่าความยาก อำนาจจำแนก และค่าความเที่ยงของแบบสอบวิชาคณิตศาสตร์

ข้อ	ค่าความยาก (p)	ค่าอำนาจจำแนก (r)	ข้อ	ค่าความยาก (p)	ค่าอำนาจจำแนก (r)
1	0.35	0.30	16	0.33	0.27
2	0.35	0.27	17	0.20	0.14
3	0.58	0.50	18	0.33	0.33
4	0.42	0.41	19	0.24	0.17
5	0.32	0.25	20	0.30	0.21
6	0.45	0.58	21	0.37	0.26
7	0.29	0.23	22	0.44	0.47
8	0.47	0.38	23	0.29	0.31
9	0.19	0.16	24	0.34	0.29
10	0.42	0.35	25	0.29	0.33
11	0.31	0.25	26	0.16	0.14
12	0.30	0.28	27	0.33	0.29
13	0.47	0.46	28	0.20	0.22
14	0.35	0.35			
15	0.44	0.31			
ค่าความเที่ยง (Alpha)			0.60		

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

จากตารางที่ 8 เมื่อพิจารณาผลการวิเคราะห์คุณภาพของแบบสอบวิชาคณิตศาสตร์ 1 พบว่ามีความเที่ยงเท่ากับ 0.60 เมื่อพิจารณาเกี่ยวกับค่าความยาก (p) จะเห็นว่าข้อสอบที่มีค่าความยากสูงสุดได้แก่ข้อสอบข้อที่ 3 มีค่าความยากเท่ากับ 0.58 และข้อสอบที่มีค่าความยากต่ำที่สุด ได้แก่ข้อสอบข้อที่ 26 มีค่าความยากเท่ากับ 0.16

เมื่อแบ่งช่วงของค่าความยากของข้อสอบตามเกณฑ์การแปลความหมาย (ศิริชัย กาญจนวาสี ,2541) พบว่า

ข้อสอบที่มีค่าความยากอยู่ในช่วง 0.00-0.19 มีจำนวน 3 ข้อ คิดเป็นร้อยละ 10.71 ได้แก่ ข้อสอบข้อที่ 9 ,17 และ 26 ซึ่งเป็นข้อสอบที่มีค่าความยากอยู่ในระดับยากมาก

ข้อสอบที่มีค่าความยากอยู่ในช่วง 0.20-0.39 มีจำนวน 17 ข้อ คิดเป็นร้อยละ 60.71 ได้แก่ ข้อสอบข้อที่ 1, 2, 5, 7, 11, 12, 14, 16, 18, 19, 20, 21, 23, 24, 25, 27 และ 28 ซึ่งเป็นข้อสอบที่มีค่าความยากอยู่ในระดับค่อนข้างยาก

ข้อสอบที่มีค่าความยากอยู่ในช่วง 0.40-0.59 มีจำนวน 8 ข้อ คิดเป็นร้อยละ 28.57 ได้แก่ ข้อสอบข้อที่ 3, 4, 6, 8, 10, 13, 15 และ 22 ซึ่งเป็นข้อสอบที่มีค่าความยากอยู่ในระดับปานกลาง

เมื่อพิจารณาค่าอำนาจจำแนก (r) พบว่าข้อสอบที่มีค่าอำนาจจำแนกสูงสุดได้แก่ข้อสอบข้อที่ 6 โดยมีค่าอำนาจจำแนกเท่ากับ 0.55 และข้อสอบที่มีค่าอำนาจจำแนกต่ำสุด คือข้อสอบข้อที่ 17 และ 26 มีค่าอำนาจจำแนกเท่ากับ 0.14 และเมื่อแบ่งค่าอำนาจจำแนกออกเป็นช่วง ๆ ตามเกณฑ์การแปลความหมาย (ศิริชัย กาญจนวาสี ,2541) พบว่า

ข้อสอบที่มีค่าอำนาจจำแนกอยู่ในช่วง 0.10-0.19 มีจำนวน 4 ข้อ คิดเป็นร้อยละ 14.29 ได้แก่ข้อสอบข้อที่ 9, 17, 19 และ 26 ซึ่งเป็นข้อสอบที่มีค่าอำนาจจำแนกอยู่ในเกณฑ์ค่อนข้างต่ำ

ข้อสอบที่มีค่าอำนาจจำแนกอยู่ในช่วง 0.20-0.39 มีจำนวน 19 ข้อ คิดเป็นร้อยละ 67.86 ได้แก่ข้อสอบข้อที่ 1, 2, 5, 7, 8, 10, 11, 12, 14, 15, 16, 18, 20, 21, 23, 24, 25, 27 และ 28 ซึ่งเป็นข้อสอบที่มีค่าอำนาจจำแนกอยู่ในเกณฑ์พอใช้

ข้อสอบที่มีค่าอำนาจจำแนกอยู่ในช่วง 0.40-0.59 มีจำนวน 5 ข้อ คิดเป็นร้อยละ 17.56 ได้แก่ข้อสอบข้อที่ 3, 4, 6, 13 และ 22 ซึ่งเป็นข้อสอบที่มีค่าอำนาจจำแนกอยู่ในเกณฑ์ดี

2.2 การตรวจสอบคุณภาพของแบบสอบในด้านความตรงเชิงโครงสร้างวิชาภาษาอังกฤษและวิชาคณิตศาสตร์ โดยทำการวิเคราะห์ด้วยวิธีวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) ได้ผลการวิเคราะห์ดังนี้

2.2.1 วิชาภาษาอังกฤษ

เนื่องจากข้อจำกัดของโปรแกรม LISREL 8.10 ที่ไม่สามารถวิเคราะห์ข้อสอบทั้งหมดได้ในครั้งเดียว ผู้วิจัยจึงได้แบ่งข้อสอบในแบบสอบวิชาภาษาอังกฤษตามลักษณะของเนื้อหาออกเป็นกลุ่มได้ 5 กลุ่มและแยกการวิเคราะห์ออกเป็นโมเดลย่อย ๆ ตามกลุ่มที่แบ่งไว้เพื่อสร้างสเกลองค์ประกอบทั้ง 5 โมเดลจากข้อสอบทั้งหมด 100 ข้อ เพื่อนำไปวิเคราะห์ความตรงเชิงโครงสร้างของโมเดลองค์ประกอบความรู้วิชาภาษาอังกฤษ ซึ่งเป็นโมเดลการวิเคราะห์องค์ประกอบยืนยันอันดับที่สอง (second order confirmatory factor analysis) ที่ประกอบด้วยโมเดลย่อยดังกล่าว โดยทำการวิเคราะห์ด้วยโปรแกรมคอมพิวเตอร์ LISREL 8.10 การแบ่งข้อสอบออกเป็นกลุ่มย่อยมีรายละเอียดดังตารางที่ 9

ตารางที่ 9 ลักษณะของข้อสอบวิชาภาษาอังกฤษจำแนกตามโครงสร้างเนื้อหา

กลุ่มที่	ลักษณะเนื้อหา	จำนวนข้อ (%)	ข้อที่
1	ความเข้าใจในการอ่าน ระดับข้อเท็จจริง	25 (25%)	41,44, 46, 49, 65, 67, 69, 71, 72, 74, 75, 76, 80, 84, 85, 89, 90, 91, 92, 93, 94, 96, 97, 98 และ 100
2	ความเข้าใจในการอ่าน ระดับตีความ	20 (20%)	19, 21, 22, 38, 43, 45, 47, 48, 63, 66, 68, 70, 73, 77, 79, 81, 82, 83, 88 และ 95
3	ความเข้าใจในการอ่าน ระดับประเมินค่า	9 (9%)	36, 39,40, 42, 64, 78, 86, 87 และ 99
4	ความสามารถในการใช้ภาษา	16 (6%)	1-16
5	ความรู้เกี่ยวกับโครงสร้างไวยากรณ์	30 (30%)	17, 18, 20, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 37, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61 และ 62
	รวม	100 (100%)	

ผลการวิเคราะห์ความตรงเชิงโครงสร้างของแบบสอบวิชาภาษาอังกฤษ โดยการวิเคราะห์องค์ประกอบเชิงยืนยัน ด้วยโปรแกรม LISREL 8.10 เพื่อสร้างสเกลองค์ประกอบจากข้อสอบจำนวน 100 ข้อ ตามโมเดลย่อยต่าง ๆ 5 โมเดล มีรายละเอียดดังนี้

1. **โมเดลด้านความเข้าใจในการอ่าน ระดับข้อเท็จจริง** (FACT)

ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับข้อเท็จจริง พบว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ดีมาก พิจารณาได้จากค่าไค-สแควร์ (Chi-Square = 224.294) ซึ่งมีค่าความน่าจะเป็นเข้าใกล้ 1 (P = 0.989) ที่องศาอิสระเท่ากับ 275 (df = 275) นั่นคือค่าไค-สแควร์ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่ายอมรับสมมติฐานหลักที่ว่าโมเดลการวิจัยสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ส่วนค่าดัชนีวัดความกลมกลืน (GFI) มีค่าเท่ากับ 0.981 ค่าดัชนีวัดความกลมกลืนที่ปรับแก้แล้ว (AGFI) มีค่าเท่ากับ 0.978 และค่าดัชนีวัดความสอดคล้อง (NFI) มีค่าเท่ากับ 0.850 รวมทั้งค่าดัชนีรากกำลังสองเฉลี่ยของเศษ (RMR) มีค่าเท่ากับ 0.005 ซึ่งมีค่าน้อยมากจนเข้าใกล้ศูนย์ ดังรายละเอียดในตารางที่ 10 และรายละเอียดของโมเดลด้านความเข้าใจในการอ่านระดับข้อเท็จจริง สามารถแสดงได้ดังแผนภาพที่ 7

เมื่อพิจาราณาผลการวิเคราะห์โมเดลด้านความเข้าใจในการอ่านระดับข้อเท็จจริง จากตารางที่ 10 และจากแผนภาพที่ 7 พบว่าน้ำหนักองค์ประกอบส่วนใหญ่ที่กำหนดข้อมูลจำเพาะของโมเดล CFA ไว้แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ยกเว้นตัวแปร X72 ที่น้ำหนักองค์ประกอบไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญโดยน้ำหนักองค์ประกอบที่มีนัยสำคัญมีค่าอยู่ระหว่าง 0.08-0.24 ค่าสัมประสิทธิ์การพยากรณ์ (R-square) ของตัวแปร X100 มีค่าสูงสุดเท่ากับ 0.27 และค่าสัมประสิทธิ์ของการพยากรณ์ของตัวแปร X75 มีค่าต่ำสุดคือเท่ากับ 0.09

สำหรับการสร้างสเกลองค์ประกอบ ผู้วิจัยได้นำค่าสัมประสิทธ์คะแนนองค์ประกอบที่ได้จากการวิเคราะห์ใช้ในการสร้างสเกลองค์ประกอบด้านความเข้าใจในการอ่านระดับข้อเท็จจริง (FACT) เพื่อให้ได้ตัวแปรใหม่สำหรับนำไปวิเคราะห์เพื่อตรวจสอบความตรงเชิงโครงสร้างของแบบสอบวิชาภาษาอังกฤษต่อไป ซึ่งได้สเกลองค์ประกอบของความเข้าใจในการอ่านระดับข้อเท็จจริง ดังสมการต่อไปนี้

$$FACT = 0.16Z_{X41} + 0.15Z_{X44} + 0.22Z_{X46} + 0.31Z_{X49} + 0.21Z_{X65} + 0.15Z_{X67} + 0.16Z_{X69} + 0.23Z_{X71} + 0.01Z_{X72} + 0.18Z_{X74} + 0.09Z_{X75} + 0.14Z_{X76} + 0.11Z_{X80} + 0.17Z_{X84} + 0.17Z_{X85} + 0.20Z_{X89} + 0.13Z_{X90} + 0.32Z_{X91} + 0.28Z_{X92} + 0.18Z_{X93} + 0.18Z_{X94} + 0.21Z_{X96} + 0.11Z_{X97} + 0.24Z_{X98} + 0.40Z_{X100}$$

ตารางที่ 10 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่านระดับข้อเท็จจริง (FACT) ของแบบสอบวิชาภาษาอังกฤษ

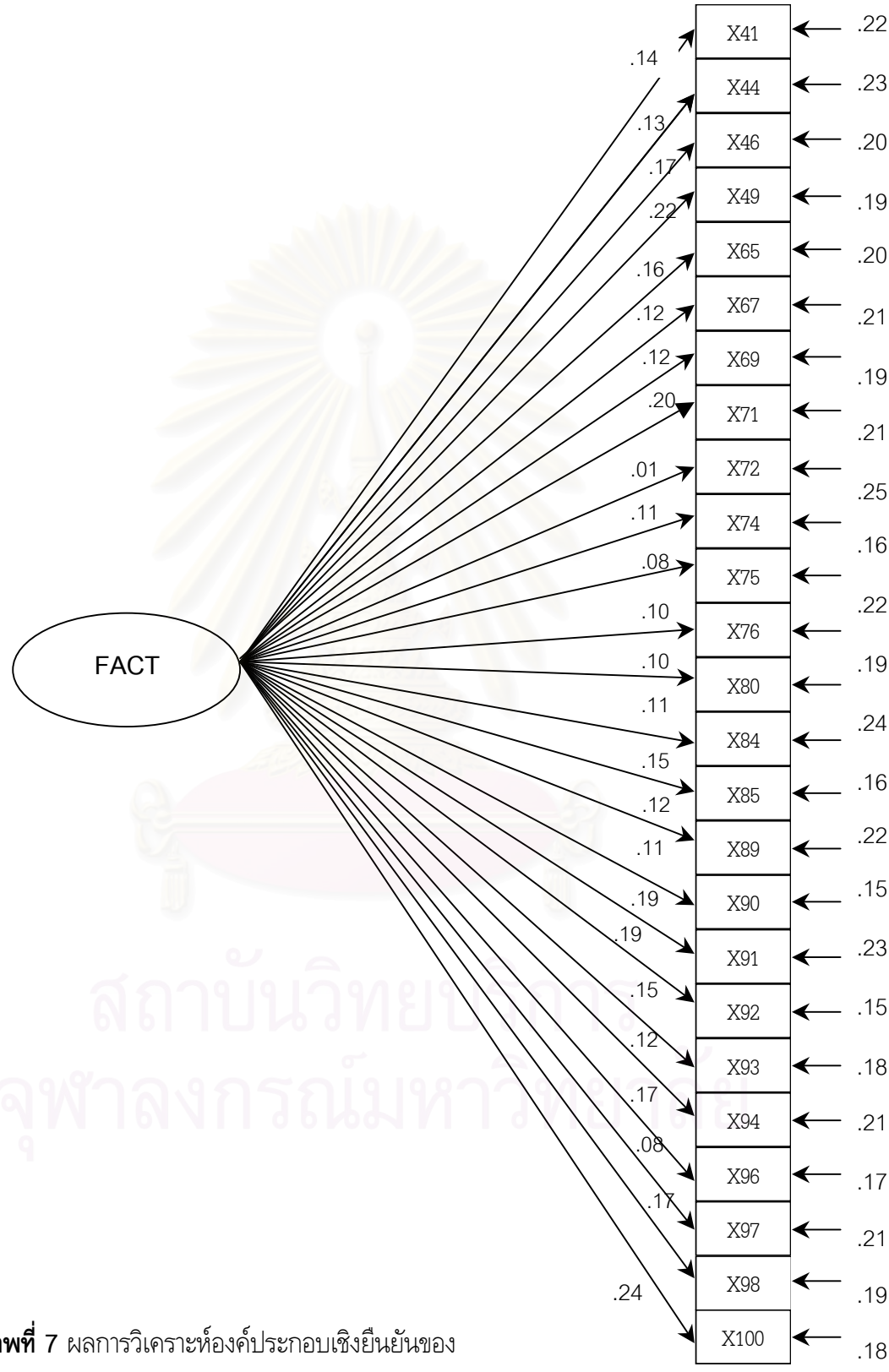
ตัวแปร	เมตริกซ์น้ำหนักองค์ประกอบ			R Square	สปส. คะแนนองค์ประกอบ
	สปส.	SE.	t		
X41	0.14	0.02	7.60**	0.08	0.16
X44	0.13	0.02	7.33**	0.07	0.15
X46	0.17	0.02	9.91**	0.13	0.22
X49	0.22	0.02	13.09**	0.21	0.31
X65	0.16	0.02	9.34**	0.11	0.21
X67	0.12	0.02	7.00**	0.06	0.15
X69	0.12	0.02	7.13**	0.07	0.16
X71	0.20	0.02	10.63**	0.14	0.23
X72	0.01	0.02	0.67	0.00	0.01
X74	0.11	0.02	7.48**	0.07	0.18
X75	0.08	0.02	4.41**	0.03	0.09
X76	0.10	0.02	6.27**	0.05	0.14
X80	0.10	0.02	5.69**	0.04	0.11
X84	0.11	0.02	7.19**	0.07	0.17
X85	0.15	0.02	8.28**	0.09	0.17
X89	0.12	0.02	7.95**	0.08	0.20
X90	0.11	0.02	6.34**	0.05	0.13
X91	0.19	0.02	12.22**	0.19	0.32
X92	0.19	0.02	11.75**	0.17	0.28
X93	0.15	0.02	8.49**	0.09	0.18
X94	0.12	0.02	7.45**	0.07	0.18
X96	0.17	0.02	9.78**	0.12	0.21
X97	0.08	0.02	5.07**	0.03	0.11
X98	0.17	0.02	10.22**	0.13	0.24
X100	0.24	0.02	15.03**	0.27	0.40

Chi-Square = 224.294 df = 275 P = 0.989 GFI = 0.981

AGFI = 0.978 NFI = 0.850 RMR = 0.005

** $p < 0.01$

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



แผนภาพที่ 7 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของ
โมเดลด้านความเข้าใจในการอ่านระดับข้อเท็จจริง

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

2. **โมเดลด้านความเข้าใจในการอ่าน ระดับตีความ** (INTERP)

ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับตีความพบว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ดีมาก พิจารณาได้จากค่าไค-สแควร์ (Chi-0Square = 120.493) ซึ่งมีค่าความน่าจะเป็นเข้าใกล้ 1 (P = 0.998) ที่องศาอิสระเท่ากับ 170 (df = 170) นั่นคือค่าไค-สแควร์ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่ายอมรับสมมติฐานหลักที่ว่าโมเดลการวิจัยสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ส่วนค่าดัชนีวัดความกลมกลืน (GFI) มีค่าเท่ากับ 0.987 ค่าดัชนีวัดความกลมกลืนที่ปรับแก้แล้ว (AGFI) มีค่าเท่ากับ 0.985 และค่าดัชนีวัดความสอดคล้อง (NFI) มีค่าเท่ากับ 0.881 รวมทั้งค่าดัชนีรากกำลังสองเฉลี่ยของเศษ (RMR) มีค่าเท่ากับ 0.005 ซึ่งมีค่าน้อยมากจนเข้าใกล้ศูนย์ ดังรายละเอียดในตารางที่ 11 และรายละเอียดของโมเดลด้านความเข้าใจในการอ่านระดับตีความ สามารถแสดงได้ดังแผนภาพที่ 8

เมื่อพิจาราณาผลการวิเคราะห์โมเดลด้านความเข้าใจในการอ่านระดับตีความ จากตารางที่ 11 และจากแผนภาพที่ 8 พบว่าน้ำหนักองค์ประกอบส่วนใหญ่ที่กำหนดข้อมูลจำเพาะของโมเดล CFA ไว้แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ยกเว้นตัวแปร X22 ที่น้ำหนักองค์ประกอบไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญโดยน้ำหนักองค์ประกอบที่มีนัยสำคัญมีค่าอยู่ระหว่าง 0.06-0.30 และค่าสัมประสิทธิ์การพยากรณ์ (R-square) ของตัวแปร X70 มีค่าสูงสุดเท่ากับ 0.37

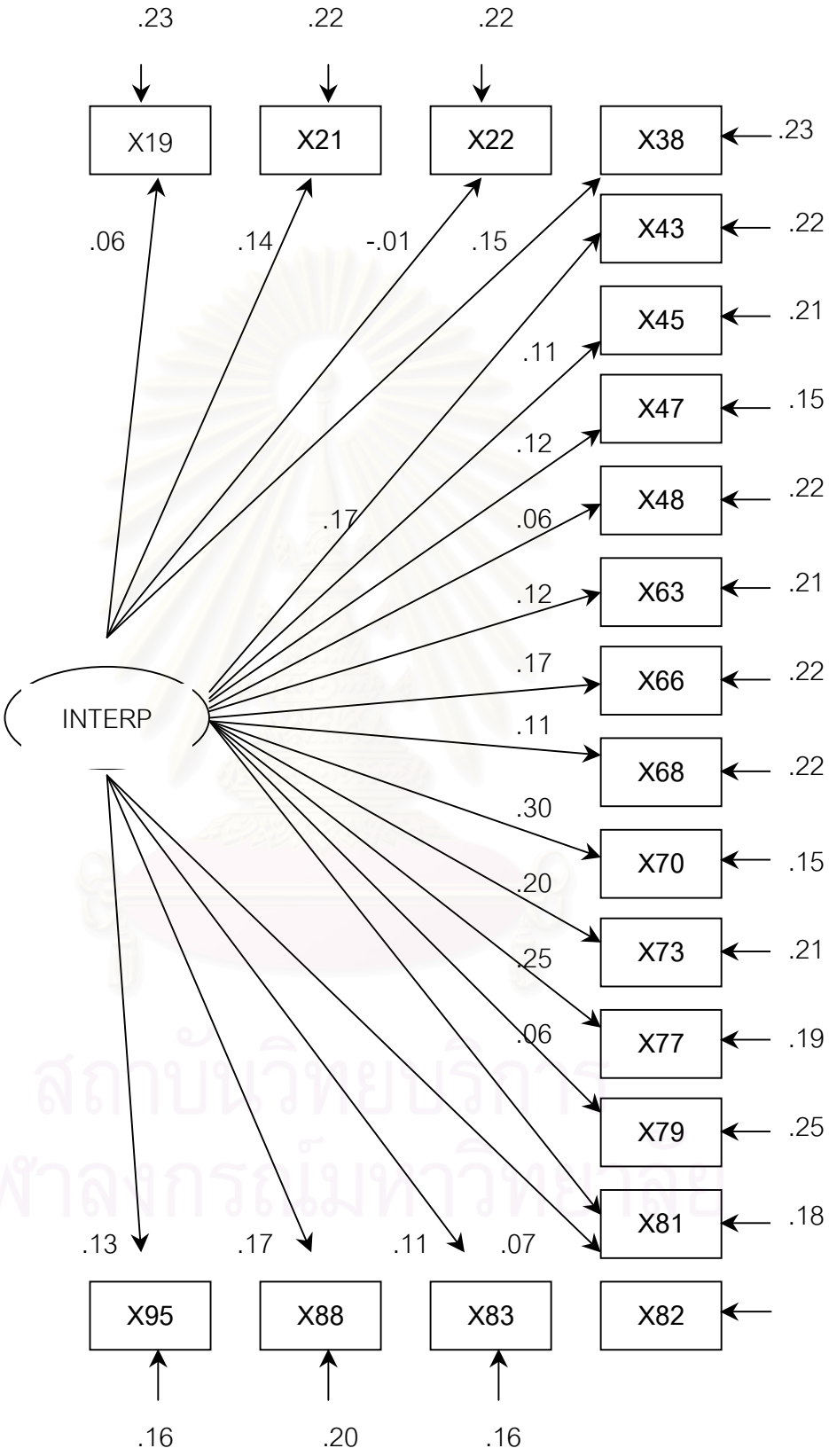
สำหรับการสร้างสเกลองค์ประกอบ ผู้วิจัยได้นำค่าสัมประสิทธ์คะแนนองค์ประกอบที่ได้จากการวิเคราะห์ใช้ในการสร้างสเกลองค์ประกอบด้านความเข้าใจในการอ่านระดับตีความ (INTERP) เพื่อให้ได้ตัวแปรใหม่สำหรับนำไปวิเคราะห์เพื่อตรวจสอบความตรงเชิงโครงสร้างของแบบสอบวิชาภาษาอังกฤษต่อไป ซึ่งได้สเกลองค์ประกอบของความเข้าใจในการอ่านระดับตีความ ดังสมการต่อไปนี้

$$\begin{aligned} INTERP = {} & 0.08Z_{X19} + 0.19Z_{X21} - 0.02Z_{X22} + 0.20Z_{X38} + 0.24Z_{X43} + 0.15Z_{X45} + \\ & 0.25Z_{X47} + 0.08Z_{X48} + 0.18Z_{X63} + 0.22Z_{X66} + 0.16Z_{X68} + 0.59Z_{X70} + \\ & 0.28Z_{X73} + 0.40Z_{X77} + 0.08Z_{X79} + 0.21Z_{X81} + 0.10Z_{X82} + 0.21Z_{X83} + \\ & 0.28Z_{X88} + 0.24Z_{X95} \end{aligned}$$

ตารางที่ 11 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับตีความ (INTERP) ของแบบสอบวิชาภาษาอังกฤษ

ตัวแปร	เมตริกซ์น้ำหนักองค์ประกอบ			R	สปส. คะแนน
	สปส.	SE.	t	**Square**	องค์ประกอบ
X19	0.06	0.02	3.50**	0.02	0.08
X21	0.14	0.02	7.69**	0.08	0.19
X22	-0.01	0.02	-0.62	0.00	-0.02
X38	0.15	0.02	8.35**	0.10	0.20
X43	0.17	0.02	9.44**	0.12	0.24
X45	0.11	0.02	6.12**	0.05	0.15
X47	0.12	0.02	8.10**	0.09	0.25
X48	0.06	0.02	3.37**	0.02	0.08
X63	0.12	0.02	7.04**	0.07	0.18
X66	0.17	0.02	8.93**	0.11	0.22
X68	0.11	0.02	6.34**	0.06	0.16
X70	0.30	0.02	17.33**	0.37	0.59
X73	0.20	0.02	10.89**	0.16	0.28
X77	0.25	0.02	14.01**	0.25	0.40
X79	0.06	0.02	3.38**	0.02	0.08
X81	0.13	0.02	7.76**	0.08	0.21
X82	0.07	0.02	4.19**	0.03	0.10
X83	0.11	0.02	7.24**	0.07	0.21
X88	0.17	0.02	9.48**	0.12	0.25
X95	0.13	0.02	8.37**	0.10	0.24
Chi-Square = 120.493		df = 170	P = 0.998	GFI = 0.987	
AGFI = 0.985		NFI = 0.881	RMR = 0.005		

** $p < 0.01$



แผนภาพที่ 8 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่านระดับตีความ

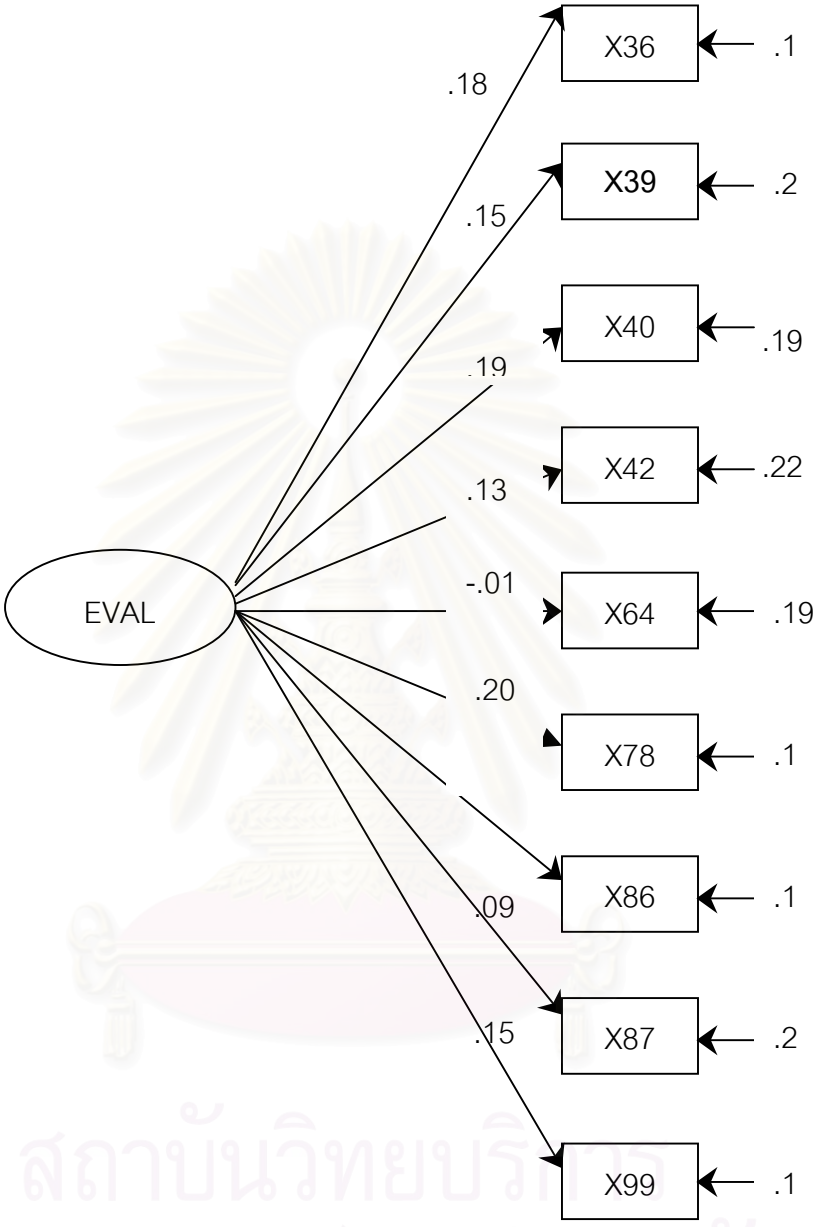
3. **โมเดลด้านความเข้าใจในการอ่าน ระดับประเมินค่า** (EVAL)

ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับประเมินค่า พบว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ดีมาก พิจารณาได้จากค่าไค-สแควร์ (Chi-Square = 26.062) ซึ่งมีค่าความน่าจะเป็นมีค่าเท่ากับ 0.515 (P = 0.515) ที่องศาอิสระเท่ากับ 27 (df = 27) นั่นคือค่าไค-สแควร์ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่ายอมรับสมมติฐานหลักที่ว่าโมเดลการวิจัยสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ส่วนค่าดัชนีวัดความกลมกลืน (GFI) มีค่าเท่ากับ 0.994 ค่าดัชนีวัดความกลมกลืนที่ปรับแก้แล้ว (AGFI) มีค่าเท่ากับ 0.990 และค่าดัชนีวัดความสอดคล้อง (NFI) มีค่าเท่ากับ 0.902 รวมทั้งค่าดัชนีรากกำลังสองเฉลี่ยของเศษ (RMR) มีค่าเท่ากับ 0.005 ซึ่งมีค่าน้อยมากจนเข้าใกล้ศูนย์ ดังรายละเอียดในตารางที่ 12 และรายละเอียดของโมเดลด้านความ เข้าใจในการอ่านระดับประเมินค่า สามารถแสดงได้ดังแผนภาพที่ 9

ตารางที่ 12 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับประเมินค่า (EVAL) ของแบบสอบวิชาภาษาอังกฤษ

ตัวแปร	เมตริกซ์น้ำหนักองค์ประกอบ			R Square	สปส. คะแนนองค์ประกอบ
	สปส.	SE.	t		
X36	0.18	0.02	8.59**	0.15	0.49
X39	0.15	0.02	6.89**	0.10	0.36
X40	0.19	0.02	8.74**	0.16	0.49
X42	0.13	0.02	5.94**	0.07	0.29
X64	-0.01	0.02	-0.70	0.00	-0.04
X78	0.20	0.02	9.51	0.20	0.61
X86	0.08	0.02	4.70	0.04	0.28
X87	0.09	0.02	4.28	0.04	0.21
X99	0.15	0.02	7.54	0.12	0.45
Chi-Square = 26.062		df = 27	P = 0.515	GFI = 0.994	
AGFI = 0.990		NFI = 0.902	RMR = 0.005		

** $p < 0.01$



แผนภาพที่ 9 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่านระดับประเมินค่า

เมื่อพิจาราณาผลการวิเคราะห์โมเดลด้านความเข้าใจในการอ่านระดับประเมินค่า จากตารางที่ 12 และจากแผนภาพที่ 9 พบว่าน้ำหนักองค์ประกอบส่วนใหญ่ที่กำหนดข้อมูลจำเพาะของโมเดล CFA ไว้แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ยกเว้นตัวแปร X64 ที่น้ำหนักองค์ประกอบไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ โดยน้ำหนักองค์ประกอบที่มีนัยสำคัญมีค่าอยู่ระหว่าง 0.08-0.20 และค่าสัมประสิทธิ์การพยากรณ์ (R-square) ของ ตัวแปร X78 มีค่าสูงสุดเท่ากับ 0.20

สำหรับการสร้างสเกลองค์ประกอบ ผู้วิจัยได้นำค่าสัมประสิทธ์คะแนนองค์ประกอบที่ได้จากการวิเคราะห์ใช้ในการสร้างสเกลองค์ประกอบด้านความเข้าใจในการอ่านระดับประเมินค่า (EVAL) เพื่อให้ได้ตัวแปรใหม่สำหรับนำไปวิเคราะห์เพื่อตรวจสอบความตรงเชิงโครงสร้างของแบบสอบวิชาภาษาอังกฤษต่อไป ซึ่งได้สเกลองค์ประกอบของความเข้าใจในการอ่านระดับประเมินค่า ดังสมการต่อไปนี้

$$EVAL = 0.49Z_{X36} + 0.36Z_{X39} + 0.49Z_{X40} + 0.29Z_{X42} - 0.04Z_{X64} + 0.61Z_{X78} + 0.28Z_{X86} + 0.21Z_{X87} + 0.45Z_{X99}$$

4. โมเดลด้านความสามารถในการใช้ภาษา (LANG)

ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความสามารถในการใช้ภาษา พบว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ พิจารณาได้จากค่าไค-สแควร์ (Chi-Square = 122.084) ซึ่งมีค่าความน่าจะเป็นเท่ากับ 0.058 ที่องศาอิสระเท่ากับ 101 (df = 101) นั่นคือค่าไค-สแควร์ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่ายอมรับสมมติฐานหลักที่ว่าโมเดลการวิจัยสอดคล้อง กลมกลืนกับข้อมูลเชิงประจักษ์ ส่วนค่าดัชนีวัดความกลมกลืน (GFI) มีค่าเท่ากับ 0.984 ค่าดัชนีวัดความกลมกลืนที่ปรับแก้แล้ว (AGFI) มีค่าเท่ากับ 0.978 และค่าดัชนีวัดความสอดคล้อง (NFI) มีค่าเท่ากับ 0.949 รวมทั้งค่าดัชนีรากกำลังสองเฉลี่ยของเศษ (RMR) มีค่าเท่ากับ 0.006 ซึ่งมีค่าน้อยมากจนเข้าใกล้ศูนย์ ดังรายละเอียดในตารางที่ 13 และรายละเอียดของโมเดลด้านความเข้าใจในการอ่านระดับประเมินค่า สามารถแสดงได้ดังแผนภาพที่ 10

เมื่อพิจาราณาผลการวิเคราะห์โมเดลด้านความสามารถในการใช้ภาษา จากตารางที่ 13 และจากแผนภาพที่ 10 พบว่าน้ำหนักองค์ประกอบทุกค่าที่กำหนดข้อมูลจำเพาะของโมเดล CFA ไว้แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 และ 0.05 โดยน้ำหนักองค์ประกอบที่มีนัยสำคัญมีค่าอยู่ระหว่าง –0.04-0.38 และค่าสัมประสิทธิ์การพยากรณ์ (R-square) ของตัวแปร X3 มีค่าสูงสุดเท่ากับ 0.59

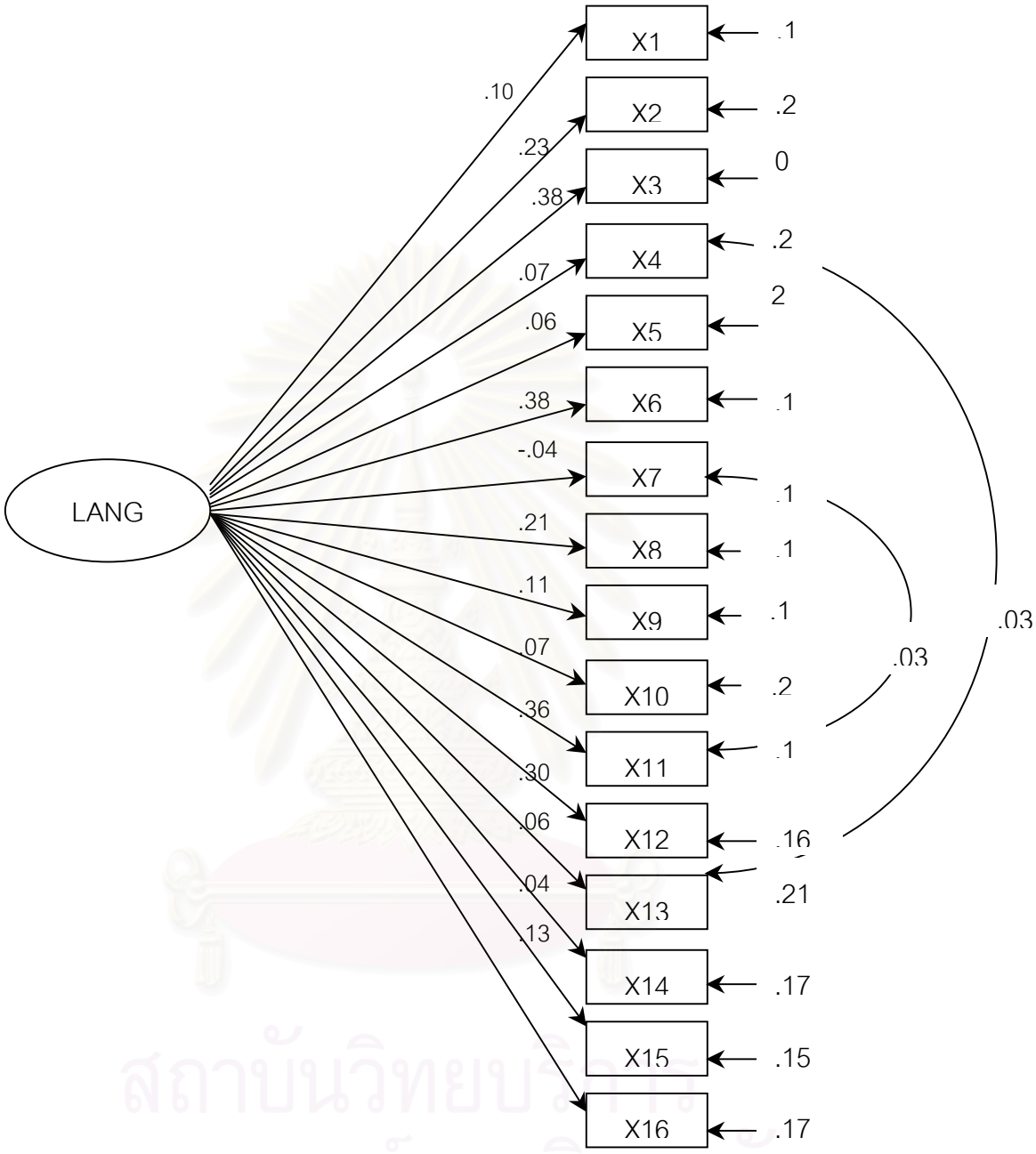
ตารางที่ 13 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความสามารถในการใช้ภาษา (LANG) ของแบบสอบวิชาภาษาอังกฤษ

ตัวแปร	เมตริกซ์น้ำหนักองค์ประกอบ			R Square	สปส. คะแนนองค์ประกอบ
	สปส.	SE.	t		
X1	0.10	0.01	7.25**	0.06	0.07
X2	0.23	0.02	13.97**	0.21	0.17
X3	0.38	0.01	26.60**	0.59	0.56
X4	0.07	0.02	4.57**	0.02	0.05
X5	0.06	0.01	4.88**	0.03	0.05
X6	0.38	0.01	26.38**	0.58	0.54
X7	-0.04	0.01	-2.49*	0.01	-0.11
X8	0.21	0.02	14.42**	0.22	0.20
X9	0.11	0.02	7.13**	0.06	0.07
X10	0.07	0.02	4.27**	0.02	0.05
X11	0.36	0.02	23.98**	0.50	0.46
X12	0.30	0.02	19.16**	0.35	0.28
X13	0.06	0.02	3.62**	0.02	0.03
X14	0.04	0.01	2.54*	0.01	0.03
X15	0.13	0.01	9.28**	0.10	0.13
X16	0.26	0.02	16.73**	0.28	0.23
Chi-Square = 122.084		df = 101	P = 0.058	GFI = 0.984	
AGFI = 0.978		NFI = 0.949		RMR = 0.006	

* $p < 0.05$, ** $p < 0.01$

สำหรับการสร้างสเกลองค์ประกอบ ผู้วิจัยได้นำค่าสัมประสิทธ์คะแนนองค์ประกอบที่ได้จากการวิเคราะห์ไปใช้ในการสร้างสเกลองค์ประกอบด้านความสามารถในการใช้ภาษา (LANG) เพื่อให้ได้ตัวแปรใหม่สำหรับนำไปวิเคราะห์เพื่อตรวจสอบความตรงเชิงโครงสร้างของแบบสอบวิชาภาษาอังกฤษต่อไป ซึ่งได้สเกลองค์ประกอบของความสามารถในการใช้ภาษา ดังสมการต่อไปนี้

$$LANG = 0.07Z_{X1} + 0.17Z_{X2} + 0.56Z_{X3} + 0.05Z_{X4} + 0.05Z_{X5} + 0.54Z_{X6} - 0.11Z_{X7} + 0.20Z_{X8} + 0.07Z_{X9} + 0.05Z_{X10} + 0.46Z_{X11} + 0.28Z_{X12} + 0.03Z_{X13} + 0.03Z_{X14} + 0.13Z_{X15} + 0.23Z_{X16}$$



แผนภาพที่ 10 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความความสามารถในการใช้ภาษา

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

5. โมเดลด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ (GRAM)

ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ พบว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ดี พิจารณาได้จากค่าไค-สแควร์ (Chi-Square = 431.465) ซึ่งมีค่าความน่าจะเป็นเท่ากับ 0.166 ที่องศาอิสระเท่ากับ 405 (df = 405) นั่นคือค่าไค-สแควร์ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่ายอมรับสมมติฐานหลักที่ว่าโมเดลการวิจัยสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ส่วนค่าดัชนีวัดความกลมกลืน (GFI) มีค่าเท่ากับ 0.967 ค่าดัชนีวัดความกลมกลืนที่ปรับแก้แล้ว (AGFI) มีค่าเท่ากับ 0.962 และค่าดัชนีวัดความสอดคล้อง (NFI) มีค่าเท่ากับ 0.797 รวมทั้งค่าดัชนีรากกำลังสองเฉลี่ยของเศษ (RMR) มีค่าเท่ากับ 0.007 ซึ่งมีค่าน้อยมากจนเข้าใกล้ศูนย์ ดังรายละเอียดในตารางที่ 14 และรายละเอียดของโมเดลด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ สามารถแสดงได้ดังแผนภาพที่ 11

เมื่อพิจารณาผลการวิเคราะห์โมเดลด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ จากตารางที่ 14 และจากแผนภาพที่ 11 พบว่าน้ำหนักองค์ประกอบส่วนใหญ่ที่กำหนดข้อมูลจำเพาะของโมเดล CFA ไว้แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 และ 0.05 ยกเว้นตัวแปร X54 และ X61 ที่น้ำหนักองค์ประกอบไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ โดยน้ำหนักองค์ประกอบที่มีนัยสำคัญมีค่าอยู่ระหว่าง 0.04-0.32 และค่าสัมประสิทธิ์การพยากรณ์ (R-square) ของตัวแปร X28 มีค่าสูงสุดเท่ากับ 0.41

สำหรับการสร้างสเกลองค์ประกอบ ผู้วิจัยได้นำค่าสัมประสิทธ์คะแนนองค์ประกอบที่ได้จากการวิเคราะห์ใช้ในการสร้างสเกลองค์ประกอบด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ (GRAM) เพื่อให้ได้ตัวแปรใหม่สำหรับนำไปวิเคราะห์เพื่อตรวจสอบความตรงเชิงโครงสร้างของแบบสอบวิชาภาษาอังกฤษต่อไป ซึ่งได้สเกลองค์ประกอบของความรู้เกี่ยวกับโครงสร้างไวยากรณ์ ดังสมการต่อไปนี้

$$\begin{aligned}GRAM = {} & 0.13Z_{X17} + 0.05Z_{X18} + 0.22Z_{X20} + 0.14Z_{X23} + 0.32Z_{X24} + 0.12Z_{X25} + \\ & 0.08Z_{X26} + 0.12Z_{X27} + 0.48Z_{X28} + 0.14Z_{X29} + 0.09Z_{X30} + 0.06Z_{X31} + \\ & 0.11Z_{X32} + 0.09Z_{X33} + 0.31Z_{X34} + 0.35Z_{X25} + 0.08Z_{X37} + 0.25Z_{X50} + \\ & 0.10Z_{X51} + 0.13Z_{X52} + 0.05Z_{X53} - 0.02Z_{X54} + 0.10Z_{X55} + 0.05Z_{X56} + \\ & 0.25Z_{X57} + 0.06Z_{X58} + 0.16Z_{X59} + 0.06Z_{X60} + 0.03Z_{X61} + 0.10Z_{X62}\end{aligned}$$

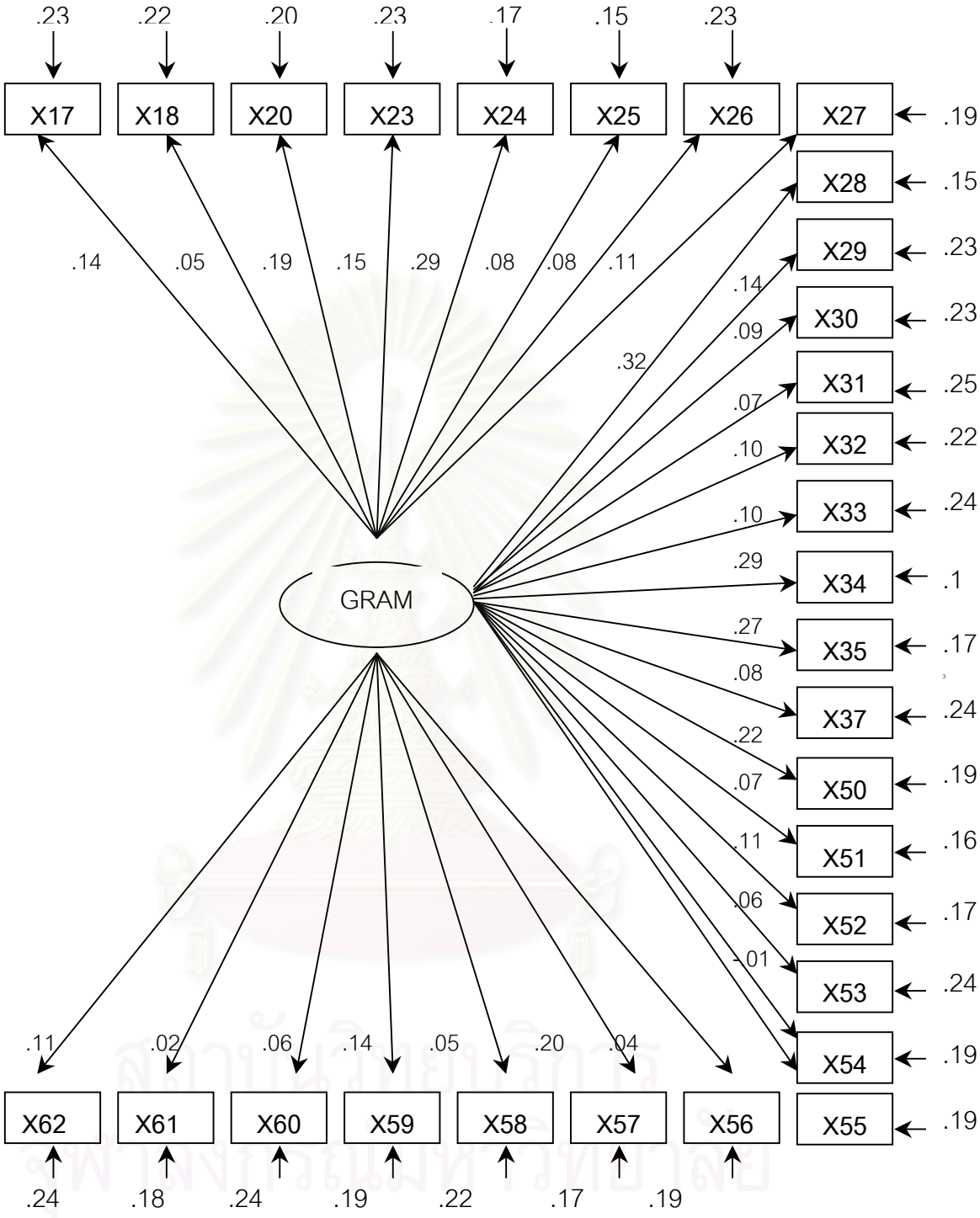
ตารางที่ 14 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ (GRAM) ของแบบสอบวิชาภาษาอังกฤษ

ตัวแปร	เมตริกซ์น้ำหนักองค์ประกอบ			R Square	สปส. คะแนนองค์ประกอบ
	สปส.	SE.	t		
X17	0.14	0.02	7.85**	0.08	0.13
X18	0.05	0.02	2.77**	0.01	0.05
X20	0.19	0.02	11.45**	0.16	0.22
X23	0.15	0.02	8.23**	0.08	0.14
X24	0.29	0.02	17.40**	0.34	0.32
X25	0.08	0.01	5.79**	0.04	0.12
X26	0.08	0.02	4.68**	0.03	0.08
X27	0.11	0.02	6.67**	0.06	0.12
X28	0.32	0.02	19.82**	0.41	0.48
X29	0.14	0.02	8.11**	0.08	0.14
X30	0.09	0.02	5.23**	0.04	0.09
X31	0.07	0.02	3.78**	0.02	0.06
X32	0.10	0.02	6.13**	0.05	0.11
X33	0.10	0.02	5.61**	0.04	0.09
X34	0.29	0.02	17.24**	0.33	0.31
X35	0.27	0.02	16.26**	0.29	0.35
X37	0.08	0.02	4.67**	0.03	0.08
X50	0.22	0.02	13.10**	0.20	0.25
X51	0.07	0.02	4.75**	0.03	0.10
X52	0.11	0.02	6.88**	0.06	0.13
X53	0.06	0.02	3.11**	0.01	0.05
X54	-0.01	0.02	-0.81	0.00	-0.02
X55	0.08	0.02	5.15**	0.03	0.10
X56	0.04	0.02	2.49*	0.01	0.05
X57	0.20	0.02	12.51**	0.18	0.25
X58	0.05	0.02	3.23**	0.01	0.06
X59	0.14	0.02	8.39**	0.09	0.16
X60	0.06	0.02	3.41**	0.02	0.06
X61	0.02	0.02	1.47	0.00	0.03
X62	0.11	0.02	5.96**	0.05	0.10

Chi-Square = 431.465 df = 405 P = 0.166 GFI = 0.967

AGFI = 0.962 NFI = 0.797 RMR = 0.007

* $p < 0.05$, ** $p < 0.01$



แผนภาพที่ 11 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์

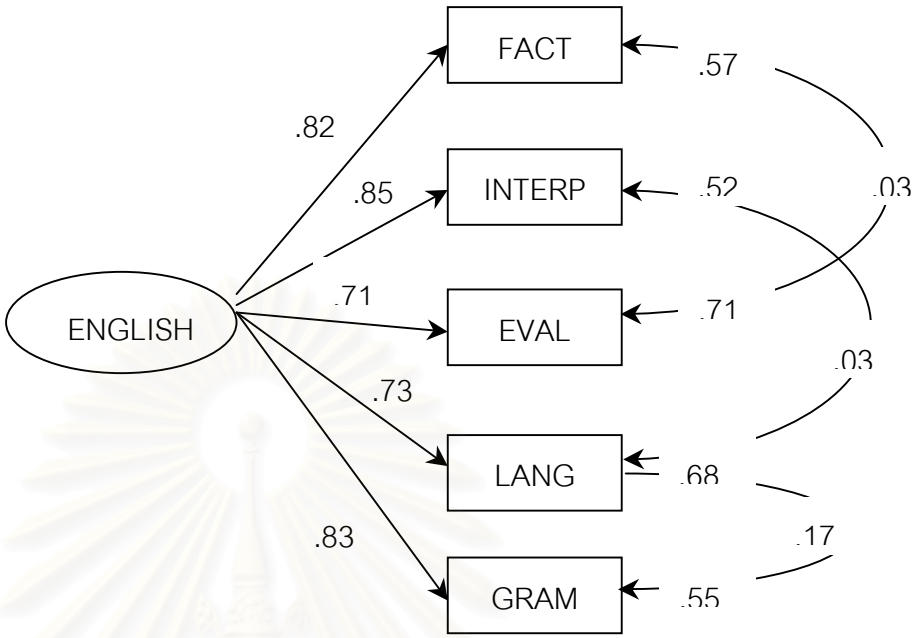
โมเดลองค์ประกอบความรู้วิชาภาษาอังกฤษ (ENGLISH)

ผลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับที่สอง (second order confirmatory factor analysis) ของโมเดลองค์ประกอบความรู้วิชาภาษาอังกฤษ (ENGLISH) พบว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ดี โดยพิจารณาจากค่าไค-สแควร์ (Chi-Square = 1.618) ซึ่งมีค่าความน่าจะเป็นเท่ากับ 0.445 ที่องศาอิสระเท่ากับ 2 (df = 2) นั่นคือค่าไค-สแควร์ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่ายอมรับสมมติฐานหลักที่ว่าโมเดลการวิจัยสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ส่วนค่าดัชนีวัดความกลมกลืน (GFI) มีค่าเท่ากับ 0.999 ค่าดัชนีวัดความกลมกลืนที่ปรับแก้แล้ว (AGFI) มีค่าเท่ากับ 0.995 และค่าดัชนีวัดความสอดคล้อง (NFI) มีค่าเท่ากับ 0.99 รวมทั้งค่าดัชนีรากกำลังสองเฉลี่ยของเศษ (RMR) มีค่าเท่ากับ 0.012 ซึ่งมีค่าน้อยมากจนเข้าใกล้ศูนย์ ดังรายละเอียดใน ตารางที่ 15 และรายละเอียดของโมเดลด้านความเข้าใจในการอ่านระดับประเมินค่า สามารถแสดงได้ดังแผนภาพที่ 12

ตารางที่ 15 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลองค์ประกอบความรู้วิชาภาษาอังกฤษ (ENGLISH)

องค์ประกอบ	เมตริกซ์น้ำหนักองค์ประกอบ			R	สปส. คะแนนองค์ประกอบ
	สปส.	SE.	t	Square	
ความเข้าใจในการอ่านระดับข้อเท็จจริง (FACT)	0.82	0.05	29.66**	0.67	0.14
ความเข้าใจในการอ่านระดับตีความ (INTERP)	0.85	0.05	31.81**	0.73	0.20
ความเข้าใจในการอ่านระดับประเมินค่า (EVAL)	0.71	0.04	23.95**	0.50	0.09
ความสามารถในการใช้ภาษา (LANG)	0.73	0.05	24.35**	0.53	0.03
ความรู้เกี่ยวกับโครงสร้างไวยากรณ์ (GRAM)	0.83	0.05	30.73**	0.69	0.15
Chi-Square = 1.618	df = 2		P = 0.445	GFI = 0.999	
AGFI = 0.995	NFI = 0.999		RMR = 0.012		

** $p < 0.01$



แผนภาพที่ 12 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลองค์ประกอบความรู้วิชาภาษาอังกฤษ

เมื่อพิจารณาผลการวิเคราะห์โมเดลความรู้วิชาภาษาอังกฤษ จากตารางที่ 15 และจากแผนภาพที่ 12 พบว่าน้ำหนักองค์ประกอบทุกค่าที่กำหนดข้อมูลจำเพาะของโมเดล CFA ไว้แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 โดยองค์ประกอบด้านความเข้าใจในการอ่านระดับข้อเท็จจริง (FACT) มีน้ำหนักองค์ประกอบมากที่สุด คือ 0.82 และองค์ประกอบด้านความเข้าใจในการอ่านระดับประเมินค่า (EVAL) มีค่าน้ำหนักองค์ประกอบน้อยที่สุด คือ 0.71 ค่าสัมประสิทธิ์การพยากรณ์ (R-square) ขององค์ประกอบด้านความเข้าใจในการอ่านระดับข้อเท็จจริง (FACT) มีค่าสูงสุดเท่ากับ 0.73 และค่าสัมประสิทธิ์ของการพยากรณ์ขององค์ประกอบด้านความเข้าใจในการอ่านระดับประเมินค่า (EVAL) มีค่าต่ำสุดคือเท่ากับ 0.50

จากผลการวิเคราะห์องค์ประกอบเชิงยืนยันสามารถสรุปได้ว่า แบบสอบวิชาภาษาอังกฤษมีความตรงเชิงโครงสร้างตามโมเดลความรู้วิชาภาษาอังกฤษ

2.2.2 วิชาคณิตศาสตร์

ผลการตรวจสอบความตรงเชิงโครงสร้างของแบบสอบวิชาคณิตศาสตร์ ด้วยวิธีวิเคราะห์องค์ประกอบเชิงยืนยัน โดยทำการวิเคราะห์ด้วยโปรแกรม LISREL 8.10 ได้ผลการวิเคราะห์ซึ่งมีรายละเอียดดังนี้

โมเดลความรู้วิชาคณิตศาสตร์ (MATH)

ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความรู้วิชาคณิตศาสตร์ พบว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ดี พิจารณาได้จากค่าไค-สแควร์ (Chi-Square = 380.184) ซึ่งมีค่าความน่าจะเป็นเท่ากับ 0.076 ที่องศาอิสระเท่ากับ 345 (df = 345) นั่นคือค่าไค-สแควร์ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่ายอมรับสมมติฐานหลักที่ว่าโมเดลการวิจัยสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ส่วนค่าดัชนีวัดความกลมกลืน (GFI) มีค่าเท่ากับ 0.974 ค่าดัชนีวัดความกลมกลืนที่ปรับแก้แล้ว (AGFI) มีค่าเท่ากับ 0.969 และค่าดัชนีวัดความสอดคล้อง (NFI) มีค่าเท่ากับ 0.736 รวมทั้งค่าดัชนีรากกำลังสองเฉลี่ยของเศษ (RMR) มีค่าเท่ากับ 0.06 ซึ่งมีค่าน้อยมากจนเข้าใกล้ศูนย์ ดังรายละเอียดในตารางที่ 16 และรายละเอียดของโมเดลด้านความรู้ความรู้วิชาคณิตศาสตร์ สามารถแสดงได้ดังแผนภาพที่ 13

เมื่อพิจารณาผลการวิเคราะห์โมเดลความรู้วิชาคณิตศาสตร์ จากตารางที่ 16 และจากแผนภาพที่ 13 พบว่าน้ำหนักองค์ประกอบส่วนใหญ่ที่กำหนดข้อมูลจำเพาะของโมเดล CFA ไว้แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 และ 0.05 ยกเว้นตัวแปร X9, X17, X20 และ X26 ที่น้ำหนักองค์ประกอบไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ โดยน้ำหนักองค์ประกอบที่มีนัยสำคัญมีค่าอยู่ระหว่าง 0.11-0.82 และค่าสัมประสิทธิ์การพยากรณ์ (R-square) ของตัวแปร X6 มีค่าสูงสุดเท่ากับ 0.30

จากผลการวิเคราะห์องค์ประกอบเชิงยืนยันสามารถสรุปได้ว่าแบบสอบวิชาคณิตศาสตร์มี ความตรงเชิงโครงสร้างตามโมเดลความรู้วิชาคณิตศาสตร์

ตารางที่ 16 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความรู้วิชาคณิตศาสตร์ (MATH)

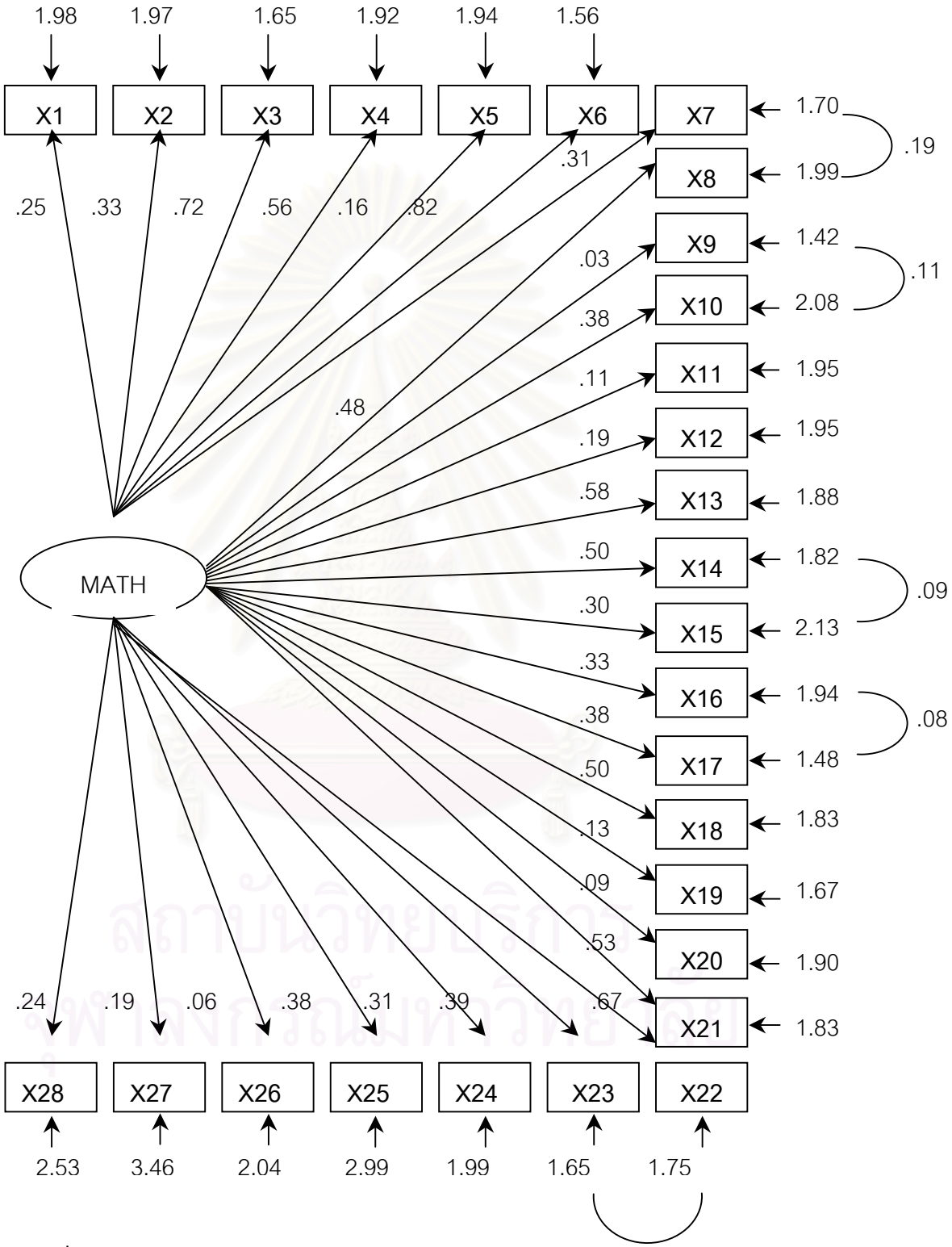
ตัวแปร	เมตริกซ์น้ำหนักองค์ประกอบ			R Square	สปส. คะแนนองค์ประกอบ
	สปส.	SE.	t		
X1	0.25	0.05	4.81**	0.03	0.04
X2	0.33	0.05	6.18**	0.05	0.05
X3	0.72	0.05	13.27**	0.24	0.15
X4	0.56	0.05	10.53**	0.14	0.08
X5	0.16	0.05	3.01**	0.01	0.02
X6	0.82	0.05	15.72**	0.30	0.14
X7	0.31	0.05	6.36**	0.06	0.04
X8	0.48	0.04	8.93**	0.11	0.06
X9	0.03	0.05	0.62	0.00	0.00
X10	0.38	0.05	6.95**	0.07	0.05
X11	0.11	0.05	2.09*	0.01	0.02
X12	0.19	0.05	3.65**	0.02	0.03
X13	0.58	0.05	10.85**	0.15	0.09
X14	0.50	0.05	9.63**	0.12	0.08
X15	0.30	0.05	5.54**	0.04	0.04
X16	0.33	0.05	6.24**	0.05	0.05
X17	0.06	0.05	1.34	0.00	0.01
X18	0.50	0.05	9.27**	0.12	0.09
X19	0.13	0.05	2.60**	0.01	0.02
X20	0.09	0.05	1.69	0.00	0.01
X21	0.53	0.05	10.04**	0.13	0.08
X22	0.67	0.05	12.26**	0.20	0.12
X23	0.39	0.05	7.84**	0.08	0.08
X24	0.29	0.05	5.52**	0.04	0.04
X25	0.38	0.07	5.89**	0.05	0.04
X26	0.06	0.05	1.05	0.00	0.01
X27	0.19	0.07	2.69**	0.01	0.02
X28	0.24	0.06	4.10**	0.02	0.03

Chi-Square = 380.184 df = 345 P = 0.076 GFI = 0.974

AGFI = 0.969 NFI = 0.736 RMR = 0.06

* $p < 0.05$, ** $p < 0.01$

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



แผนภาพที่ 13 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลความรู้วิชาคณิตศาสตร์

ตอนที่ 3 การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบด้วยกระบวนการ DFIT

ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ สำหรับแบบสอบวิชาภาษาอังกฤษและวิชาคณิตศาสตร์ จำแนกตามตัวแปรเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา ซึ่งมีการวิเคราะห์การทำหน้าที่ต่างกันทั้งในรายข้อและรายฉบับ ในการวิเคราะห์นั้นใช้กลุ่มผู้สอบเพศชายและกลุ่มเขตกรุงเทพมหานครและปริมณฑล เป็นกลุ่มอ้างอิง ผลการวิเคราะห์สรุปได้ดังนี้

3.1 วิชาภาษาอังกฤษ

สำหรับแบบสอบวิชาภาษาอังกฤษนั้น ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ เมื่อจำแนกกลุ่มผู้สอบตามตัวแปรเพศ ได้ผลการวิเคราะห์ดังตารางที่ 17 - 19 และตัวแปรสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา ได้ผลการวิเคราะห์ดังตารางที่ 20 - 22

3.1.1 เมื่อจำแนกกลุ่มสอบตามตัวแปรเพศ

ตารางที่ 17 ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามเพศ

ข้อที่	ดัชนี NCDIF	ดัชนี CDIF	**ข้อที่**	ดัชนี NCDIF	ดัชนี CDIF
1	0.096*	0.242	21	0.009*	-0.045
2	0.007*	-0.047	22	0.002	0.031
3	0.014*	0.014	23	0.008*	0.055
4	0.001	0.023	24	0.034*	0.026
5	0.011*	0.094	25	0.005	-0.022
6	0.025*	0.017	26	0.004	0.016
7	0.000	0.008	27	0.012*	0.090
8	0.019*	-0.008	28	0.016*	0.010
9	0.034*	0.142	29	0.008*	0.059
10	0.004	0.022	30	0.009*	0.050
11	0.013*	0.014	31	0.001	0.008
12	0.039*	-0.017	32	0.001	0.018
13	0.000	-0.003	33	0.006	0.047
14	0.001	-0.019	34	0.008*	0.001
15	0.003	-0.010	35	0.002	-0.004
16	0.028*	0.004	36	0.001	-0.005
17	0.009*	0.057	37	0.000	-0.008
18	0.001	-0.005	38	0.021*	0.034
19	0.004	0.003	39	0.006	-0.014
20	0.003	-0.021	40	0.011*	0.028

ตารางที่ 17 (ต่อ)

ข้อที่	ดัชนี NCDIF	ดัชนี CDIF	**ข้อที่**	ดัชนี NCDIF	ดัชนี CDIF
41	0.007*	0.046	71	0.001	-0.006
42	0.008*	0.062	72	0.001	0.028
43	0.000	-0.011	73	0.019*	0.061
44	0.005	0.031	74	0.000	0.002
45	0.004	0.035	75	0.000	-0.001
46	0.001	0.015	76	0.001	-0.008
47	0.004	0.025	77	0.007*	-0.028
48	0.001	-0.025	78	0.001	-0.009
49	0.001	-0.007	79	0.001	0.009
50	0.005	-0.018	80	0.000	0.017
51	0.003	-0.004	81	0.005	-0.006
52	0.000	-0.010	82	0.004	-0.039
53	0.008*	0.048	83	0.001	0.024
54	0.002	0.023	84	0.001	-0.019
55	0.002	-0.024	85	0.004	0.014
56	0.000	0.003	86	0.001	-0.021
57	0.001	-0.002	87	0.001	0.022
58	0.000	0.002	88	0.002	-0.027
59	0.004	-0.009	89	0.002	-0.019
60	0.001	-0.017	90	0.000	0.002
61	0.000	-0.010	91	0.019*	0.026
62	0.002	-0.029	92	0.002	-0.019
63	0.003	0.004	93	0.012*	0.070
64	0.001	0.004	94	0.001	-0.020
65	0.000	0.011	95	0.001	-0.010
66	0.009*	0.040	96	0.011*	0.028
67	0.006	0.008	97	0.001	0.005
68	0.012*	0.085	98	0.000	0.007
69	0.000	-0.006	99	0.005	-0.006
70	0.002	0.008	100	0.005	-0.014
จำนวนข้อสอบที่พบ DIF				30 ข้อ	
ดัชนี DTF				1.225*	

*หมายเหตุ พิจารณาตามเกณฑ์จุดตัดของ Fleer (1993 อ้างถึงใน Raju และคณะ (1995)) ข้อสอบและแบบสอบทำหน้าที่ต่างกันเมื่อดัชนี NCDIF และดัชนี DTF มีค่ามากกว่า 0.006

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

จากตารางที่ 17 ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบวิชาภาษาอังกฤษสำหรับกลุ่มผู้สอบจำแนกตามเพศ พบว่าในการวิเคราะห์การทำหน้าที่ต่างกันที่ระดับข้อสอบ มีข้อสอบที่ทำหน้าที่ต่างกัน 30 ข้อ ได้แก่ข้อที่ 1, 2, 3, 5, 6, 8, 9, 11, 12, 16, 17, 21, 23, 24, 27, 28, 29, 30, 34, 38, 40, 41, 42, 53, 66, 68, 73, 91, 93 และ 96 และในการวิเคราะห์การทำหน้าที่ต่างกันที่ระดับแบบสอบ พบว่าค่าดัชนี DTF มีค่าเท่ากับ 1.225 ซึ่งมีค่ามากกว่า 0.006 แสดงว่าแบบสอบวิชาภาษาอังกฤษทำหน้าที่ต่างกันเมื่อแบ่งกลุ่มผู้สอบตามตัวแปรเพศ หลังจากดำเนินการคัดเลือกข้อสอบออกจากแบบสอบทีละข้อตามกระบวนการ DFIT จากตารางที่ 18 พบว่าข้อสอบข้อที่ 1 มีค่าดัชนี CDIF สูงที่สุด คือมีค่าเท่ากับ 0.242 เมื่อตัดข้อสอบข้อที่ 1 ออกจากแบบสอบและทำการวิเคราะห์ค่าดัชนี CDIF และดัชนี DTF ใหม่อีกครั้ง ได้ผลดังตารางที่ 18

ตารางที่ 18 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามเพศ เมื่อตัดข้อสอบข้อที่ 1 ออกจากแบบสอบ

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
1*	ตัดออกจากการวิเคราะห์	21*	-0.068
2*	-0.024	22	0.020
3*	0.073	23*	0.045
4	0.018	24*	-0.036
5*	0.040	25	0.000
6*	-0.032	26	0.030
7	0.003	27*	0.048
8*	-0.045	28*	-0.039
9*	0.097	29*	0.050
10	0.027	30*	0.047
11*	-0.016	31	0.014
12*	-0.084	32	0.011
13	-0.001	33	0.122
14	-0.012	34*	-0.016
15	-0.031	35	0.000
16*	0.011	36	-0.006
17*	0.058	37	-0.003
18	-0.010	38*	0.043
19	0.032	39	0.024
20	-0.025	40*	0.038

ตารางที่ 18 (ต่อ)

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
41*	0.024	71	0.005
42*	0.038	72	0.019
43	0.074	73*	0.055
44	0.028	74	0.005
45	0.023	75	0.001
46	0.027	76	0.003
47	0.044	77	-0.017
48	-0.029	78	-0.013
49	0.001	79	0.029
50	-0.033	80	-0.009
51	-0.024	81	-0.002
52	0.006	82	0.035
53*	0.049	83	-0.016
54	0.011	84	0.016
55	-0.030	85	-0.038
56	0.014	86	-0.013
57	-0.002	87	0.022
58	0.004	88	-0.007
59	-0.030	89	-0.007
60	-0.031	90	0.010
61	-0.006	91*	0.063
62	-0.021	92	-0.026
63	-0.007	93*	0.058
64	-0.001	94	-0.019
65	-0.004	95	0.007
66*	0.052	96*	0.035
67	0.025	97	0.012
68*	0.041	98	0.017
69	0.003	99	0.024
70	-0.006	100	-0.030
ดัชนี DTF	0.789*		

*หมายเหตุ พิจารณาตามเกณฑ์จุดตัดของ Fleer (1993 อ้างถึงใน Raju และคณะ (1995)) ข้อสอบและแบบสอบทำหน้าที่ต่างกันเมื่อดัชนี NCDIF และดัชนี DTF มีค่ามากกว่า 0.006

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

จากตารางที่ 18 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามตัวแปรเพศ เมื่อตัดข้อสอบข้อที่ 1 ออกจากแบบสอบ พบว่าดัชนี DTF มีค่าเท่ากับ 0.789 ซึ่งมีค่ามากกว่า 0.006 จากตารางพบว่าข้อที่ 33 มีค่าดัชนี CDIF สูงที่สุด คือมีค่าเท่ากับ 0.122 จึงทำการตัดข้อสอบข้อที่ 33 ออกจากแบบสอบและทำการวิเคราะห์ค่าดัชนี CDIF และค่าดัชนี DTF ใหม่อีกครั้งตามกระบวนการ DFIT จากผลการวิเคราะห์ตามกระบวนการ DFIT พบว่าจะต้องดำเนินการตัดข้อสอบข้อที่ 1, 33, 9, 91, 17, 68, 79, 41, 47, 90, 46, 66 และ 20 ออกจากแบบสอบตามลำดับได้ซึ่งรายละเอียดของผลการวิเคราะห์มีดังตารางที่ 65 - 75 ในภาคผนวก ผลการวิเคราะห์ตามกระบวนการ DFIT สำหรับแบบสอบวิชาภาษาอังกฤษ เมื่อแบ่งกลุ่มผู้สอบตามเพศ ในขั้นตอนสุดท้ายได้ผลดังตาราง ที่ 19

ตารางที่ 19 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามเพศ เมื่อตัดข้อสอบข้อที่ 1, 33, 9, 91, 17, 68, 79, 41, 47, 90, 46, 66 และ 20 ออกจากแบบสอบ

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
1*	ตัดออกจากการวิเคราะห์	21*	-0.009
2*	-0.008	22	0.001
3*	0.008	23*	0.006
4	0.006	24*	0.009
5*	0.009	25	-0.016
6*	0.005	26	-0.003
7	0.004	27*	0.008
8*	-0.002	28*	0.007
9*	ตัดออกจากการวิเคราะห์	29*	0.010
10	0.001	30*	0.001
11*	-0.003	31	-0.001
12*	0.006	32	0.008
13	0.000	33	ตัดออกจากการวิเคราะห์
14	-0.008	34*	0.004
15	-0.002	35	-0.003
16*	0.007	36	0.003
17*	ตัดออกจากการวิเคราะห์	37	-0.002
18	0.002	38*	-0.015
19	-0.006	39	-0.002
20	ตัดออกจากการวิเคราะห์	40*	-0.001

ตารางที่ 19 (ต่อ)

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
41*	ตัดออกจากการวิเคราะห์	71	-0.003
42*	0.007	72	0.011
43	0.001	73*	-0.004
44	0.001	74	0.004
45	0.009	75	0.001
46	ตัดออกจากการวิเคราะห์	76	-0.006
47	ตัดออกจากการวิเคราะห์	77	0.001
48	-0.004	78	0.002
49	-0.003	79	ตัดออกจากการวิเคราะห์
50	0.002	80	0.000
51	0.005	81	-0.018
52	-0.004	82	0.002
53*	0.007	83	-0.004
54	0.007	84	0.005
55	-0.003	85	0.000
56	-0.001	86	-0.008
57	0.002	87	0.007
58	0.004	88	-0.005
59	0.006	89	-0.011
60	-0.003	90	ตัดออกจากการวิเคราะห์
61	-0.002	91*	ตัดออกจากการวิเคราะห์
62	-0.007	92	-0.002
63	0.011	93*	0.007
64	0.006	94	-0.006
65	0.007	95	-0.006
66*	ตัดออกจากการวิเคราะห์	96*	-0.004
67	-0.007	97	0.000
68*	ตัดออกจากการวิเคราะห์	98	0.003
69	-0.002	99	-0.009
70	-0.009	100	-0.003
ดัชนี DTF	0.005		

*หมายเหตุ พิจารณาตามเกณฑ์จุดตัดของ Fleer (1993 อ้างถึงใน Raju และคณะ (1995)) ข้อสอบและแบบสอบทำหน้าที่ต่างกันเมื่อดัชนี NCDIF และดัชนี DTF มีค่ามากกว่า 0.006

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

จากตารางที่ 19 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามตัวแปรเพศ เมื่อตัดข้อสอบข้อที่ 1, 33, 9, 91, 17, 68, 79, 41, 47, 90, 46, 66 และ 20 ออกจากแบบสอบ พบว่าดัชนี DTF มีค่าเท่ากับ 0.005 ซึ่งมีค่าน้อยกว่า 0.006 แสดงว่าแบบสอบดังกล่าวไม่ทำหน้าที่ต่างกัน หลังจากที่ทำการตัดข้อสอบที่มีค่าดัชนี CDIF สูงที่สุดออกทีละข้อรวมทั้งสิ้น 13 ข้อ

3.1.2 เมื่อจำแนกกลุ่มผู้สอบตามตัวแปรสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

ตารางที่ 20 ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

ข้อที่	ดัชนี NCDIF	ดัชนี CDIF	**ข้อที่**	ดัชนี NCDIF	ดัชนี CDIF
1	0.015*	0.035	21	0.001	0.013
2	0.000	-0.004	22	0.001	0.012
3	0.004	0.006	23	0.004	0.004
4	0.001	0.009	24	0.027*	0.019
5	0.002	0.018	25	0.001	0.011
6	0.004	0.000	26	0.002	-0.004
7	0.000	-0.007	27	0.003	0.003
8	0.000	0.000	28	0.003	0.007
9	0.013*	-0.006	29	0.000	0.006
10	0.006	0.025	30	0.000	-0.001
11	0.006	0.012	31	0.002	0.020
12	0.000	0.007	32	0.001	0.005
13	0.000	0.004	33	0.125*	0.045
14	0.003	-0.014	34	0.004	0.006
15	0.000	-0.004	35	0.001	-0.011
16	0.013*	-0.002	36	0.001	0.004
17	0.006	-0.009	37	0.004	0.027
18	0.000	-0.008	38	0.002	-0.001
19	0.000	0.002	39	0.011*	0.009
20	0.000	0.006	40	0.004	-0.004

ตารางที่ 20 (ต่อ)

ข้อที่	ดัชนี NCDIF	ดัชนี CDIF	**ข้อที่**	ดัชนี NCDIF	ดัชนี CDIF
41	0.001	-0.011	71	0.003	0.019
42	0.009*	0.014	72	0.002	0.020
43	0.002	0.001	73	0.011*	0.024
44	0.008*	0.000	74	0.004	-0.008
45	0.002	0.020	75	0.001	0.005
46	0.003	0.008	76	0.001	-0.014
47	0.002	-0.020	77	0.009*	-0.007
48	0.002	-0.012	78	0.001	-0.012
49	0.001	-0.011	79	0.004	-0.016
50	0.000	-0.006	80	0.003	0.016
51	0.001	-0.005	81	0.001	-0.011
52	0.009*	-0.010	82	0.005	0.030
53	0.001	0.014	83	0.001	-0.003
54	0.002	-0.013	84	0.000	0.005
55	0.003	-0.015	85	0.000	-0.003
56	0.009*	-0.028	86	0.001	0.010
57	0.009*	-0.006	87	0.002	-0.018
58	0.003	-0.017	88	0.002	-0.019
59	0.001	0.003	89	0.002	-0.009
60	0.007*	0.038	90	0.000	-0.001
61	0.004	-0.008	91	0.002	0.006
62	0.001	-0.014	92	0.000	-0.009
63	0.002	-0.019	93	0.012*	0.027
64	0.003	-0.006	94	0.001	-0.011
65	0.002	-0.017	95	0.003	0.015
66	0.007*	0.004	96	0.002	0.010
67	0.001	-0.005	97	0.000	0.000
68	0.006	0.029	98	0.000	-0.002
69	0.000	-0.003	99	0.001	-0.003
70	0.005	0.008	100	0.001	-0.004

จำนวนข้อที่พบ DIF	16 ข้อ
ค่าดัชนี DTF	0.302*

*หมายเหตุ พิจารณาตามเกณฑ์จุดตัดของ Fleer (1993 อ้างถึงใน Raju และคณะ (1995)) ข้อสอบและแบบสอบทำหน้าที่ต่างกันเมื่อดัชนี NCDIF และดัชนี DTF มีค่ามากกว่า 0.006

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

จากตารางที่ 20 ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบในวิชาภาษาอังกฤษ เมื่อจำแนกกลุ่มผู้สอบตามตัวแปรสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษาพบว่าในการวิเคราะห์การทำหน้าที่ต่างกันที่ระดับข้อสอบ มีข้อสอบที่ทำหน้าที่ต่างกันทั้งหมดจำนวน 16 ข้อ ได้แก่ข้อที่ 1, 9, 16, 24, 33, 39, 42, 44, 52, 56, 57, 60, 66, 73, 77 และ 93 และในการวิเคราะห์การทำหน้าที่ต่างกันที่ระดับแบบสอบ พบว่ามีค่าดัชนี DTF เท่ากับ 0.302 ซึ่งมีค่ามากกว่า 0.006 แสดงว่า แบบสอบดังกล่าวทำหน้าที่ต่างกัน หลังจากดำเนินการคัดเลือกข้อสอบออกจากแบบสอบทีละข้อตามกระบวนการ DFIT จากตารางที่ 20 พบว่าข้อสอบข้อที่ 33 มีค่าดัชนี CDIF สูงที่สุด คือมีค่าเท่ากับ 0.045 เมื่อตัดข้อสอบข้อที่ 33 ออกจากแบบสอบและทำการวิเคราะห์ค่าดัชนี CDIF และดัชนี DTF ใหม่อีกครั้ง ได้ผลดังตารางที่ 21

ตารางที่ 21 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา เมื่อตัดข้อสอบข้อที่ 33 ออกจากแบบสอบ

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
1*	0.017	21	0.008
2	-0.002	22	0.003
3	0.007	23	0.003
4	0.004	24	0.020
5	0.007	25	-0.004
6	0.004	26	0.001
7	-0.001	27	0.003
8	-0.003	28	0.005
9*	0.000	29	0.005
10	0.011	30	0.003
11	0.006	31	-0.003
12	0.004	32	0.004
13	0.002	33*	ตัดออกจากการวิเคราะห์
14	-0.006	34	0.006
15	-0.002	35	-0.005
16*	-0.002	36	0.002
17	0.002	37	0.011
18	0.005	38	0.000
19	0.003	39*	0.005
20	-0.001	40	-0.001

ตารางที่ 22 (ต่อ)

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
41	-0.005	71	0.010
42*	0.007	72	0.008
43	0.001	73*	0.011
44*	0.002	74	-0.004
45	0.008	75	0.002
46	0.004	76	-0.006
47	-0.008	77*	-0.004
48	-0.006	78	-0.005
49	-0.005	79	-0.007
50	-0.003	80	0.007
51	-0.002	81	-0.004
52*	-0.001	82	0.013
53	0.006	83	-0.001
54	-0.006	84	0.002
55	-0.007	85	-0.001
56*	-0.009	86	0.004
57*	0.006	87	-0.007
58	-0.007	88	-0.007
59	0.002	89	-0.004
60*	0.012	90	-0.001
61	-0.004	91	0.003
62	-0.006	92	-0.003
63	-0.008	93*	0.013
64	-0.003	94	-0.005
65	-0.007	95	0.006
66*	0.002	96	0.004
67	-0.002	97	0.000
68	0.012	98	-0.001
69	0.000	99	-0.001
70	0.007	100	-0.001
ดัชนี DTF	0.128*		

*หมายเหตุ พิจารณาตามเกณฑ์จุดตัดของ Fleer (1993 อ้างถึงใน Raju และคณะ (1995)) ข้อสอบและแบบสอบทำหน้าที่ต่างกันเมื่อดัชนี NCDIF และดัชนี DTF มีค่ามากกว่า 0.006

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

จากตารางที่ 21 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามตัวแปรสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา เมื่อตัดข้อสอบข้อที่ 33 ออกจากแบบสอบ พบว่าดัชนี DTF มีค่าเท่ากับ 0.128 ซึ่งมีค่ามากกว่า 0.006 จากตารางพบว่าข้อที่ 24 มีค่าดัชนี CDIF สูงที่สุด คือมีค่าเท่ากับ 0.020 จึงทำการตัดข้อสอบข้อที่ 24 ออกจากแบบสอบและทำการวิเคราะห์ค่าดัชนี CDIF และดัชนี DTF ใหม่อีกครั้งตามกระบวนการ DFIT จากผลการวิเคราะห์ตามกระบวนการ DFIT พบว่าจะต้องดำเนินการตัดข้อสอบข้อที่ 33, 24, 37, 68, 1, 73 และ 66 ออกจากแบบสอบตามลำดับได้ซึ่งรายละเอียดของผลการวิเคราะห์มีดังตารางที่ 76 - 80 ในภาคผนวก ผลการวิเคราะห์ตามกระบวนการ DFIT สำหรับแบบสอบวิชาภาษาอังกฤษ เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษาในขั้นตอนสุดท้ายได้ผลดังตารางที่ 22

ตารางที่ 22 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา เมื่อตัดข้อสอบข้อที่ 33, 24, 37, 68, 1, 73 และ 66 ออกจากแบบสอบ

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
1*	ตัดออกจากการวิเคราะห์	21	0.008
2	0.001	22	0.003
3	0.000	23	0.012
4	0.004	24*	ตัดออกจากการวิเคราะห์
5	0.006	25	0.001
6	-0.004	26	-0.007
7	-0.003	27	0.011
8	-0.008	28	-0.001
9*	0.012	29	0.004
10	0.017	30	-0.002
11	-0.003	31	0.010
12	0.003	32	0.006
13	0.001	33*	ตัดออกจากการวิเคราะห์
14	-0.009	34	-0.002
15	-0.002	35	-0.004
16*	-0.012	36	0.006
17	0.008	37	ตัดออกจากการวิเคราะห์
18	-0.002	38	0.008
19	0.001	39*	-0.008
20	0.003	40	0.009

ตารางที่ 22 (ต่อ)

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
41	0.000	71	0.004
42*	0.017	72	0.006
43	0.007	73*	ตัดออกจากการวิเคราะห์
44*	0.010	74	-0.021
45	0.006	75	-0.004
46	0.008	76	-0.003
47	-0.005	77*	0.007
48	-0.008	78	-0.002
49	-0.002	79	-0.012
50	-0.003	80	0.001
51	-0.005	81	0.000
52*	-0.009	82	0.004
53	0.003	83	0.002
54	-0.009	84	0.004
55	-0.010	85	0.001
56*	-0.014	86	0.016
57*	-0.006	87	-0.004
58	-0.021	88	-0.003
59	-0.003	89	-0.008
60*	0.006	90	-0.002
61	-0.010	91	0.008
62	-0.005	92	-0.004
63	-0.007	93*	0.020
64	-0.011	94	-0.005
65	-0.002	95	0.000
66*	ตัดออกจากการวิเคราะห์	96	0.010
67	-0.005	97	-0.001
68	ตัดออกจากการวิเคราะห์	98	0.000
69	0.002	99	0.004
70	-0.007	100	0.003
ดัชนี DTF	0.004		

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

จากตารางที่ 22 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามตัวแปรสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา เมื่อตัดข้อสอบข้อที่ 33, 24, 37, 68, 1, 73 และ 66 ออก พบว่าดัชนี DTF มีค่าเท่ากับ 0.004 ซึ่งมีค่าน้อยกว่า 0.006 แสดงว่าแบบสอบดังกล่าวไม่ทำหน้าที่ต่างกัน หลังจากที่ทำการตัดข้อสอบที่มีค่าดัชนี CDIF สูงที่สุดออกทีละข้อ รวมทั้งสิ้น 7 ข้อ

สำหรับการเปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบวิชาภาษาอังกฤษเมื่อจำแนกผู้สอบตามตัวแปรเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา ได้ผลดังตารางที่ 23

ตารางที่ 23 เปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบวิชาภาษาอังกฤษ เมื่อจำแนกผู้สอบตามตัวแปรเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

ตัวแปร / ดัชนี	เพศ	สถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา
NCDIF	1, 2, 3, 5, 6, 8, 9, 11, 12, 16, 17, 21, 23, 24, 27, 28, 29, 30, 34, 38, 40, 41, 42, 53, 66, 68, 73, 91, 93 และ 96	1, 9, 16, 24, 33, 39, 42, 44, 52, 56, 57, 60, 66, 73, 77 และ 93
รวม	30 ข้อ	16 ข้อ
CDIF/DTF	1, 9, 17, 20, 33, 41, 46, 47, 66, 68, 79, 90 และ 91	1, 24, 33, 37, 66 , 68 และ 73
รวม	13 ข้อ	7 ข้อ

หมายเหตุ อักษรตัวเข้ม หมายถึง ข้อสอบที่ตรวจพบ DIF ทั้งในกรณีจำแนกผู้สอบตามตัวแปรเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

อักษรขีดเส้นใต้ หมายถึงข้อสอบที่พบ DIF ทั้งโดยดัชนี NCDIF และ CDIF/DTF

จากตารางที่ 23 การเปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบวิชาภาษาอังกฤษเมื่อจำแนกผู้สอบตามตัวแปรเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา พบว่าเมื่อจำแนกกลุ่มผู้สอบตามตัวแปรเพศจะพบข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด คือ 30 ข้อ

และหลังจากตัดข้อสอบออกทีละข้อจนแบบสอบไม่ทำหน้าที่ต่างกัน ต้องตัดข้อสอบออกทั้งสิ้น 13 ข้อ มีข้อสอบที่ถูกตัดออกจำนวน 7 ข้อที่ตรงกับข้อสอบที่พบ DIF นอกจากนี้เมื่อจำแนกกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา พบข้อสอบที่ทำหน้าที่ต่างกันทั้งหมด 16 ข้อและหลังจากตัดข้อสอบออกทีละข้อจนแบบสอบไม่ทำหน้าที่ต่างกัน ต้องตัดข้อสอบออกทั้งสิ้น 7 ข้อ มีข้อสอบที่ถูกตัดออกจำนวน 5 ข้อที่ตรงกับข้อสอบที่พบ DIF และมีข้อสอบ 8 ข้อ ได้แก่ ข้อที่ 1, 9, 16, 24, 42, 66, 73 และ 93 ที่พบ DIF เมื่อแบ่งกลุ่มผู้สอบตามตัวแปรที่ศึกษาทั้ง 2 ตัวแปร

3.2 วิชาคณิตศาสตร์

สำหรับแบบสอบวิชาคณิตศาสตร์นั้น ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ เมื่อจำแนกกลุ่มผู้สอบตามตัวแปรเพศ ได้ผลการวิเคราะห์ดังตารางที่ 24 - 26 และตัวแปรสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา ได้ผลการวิเคราะห์ดังตารางที่ 27

3.2.1 เมื่อจำแนกกลุ่มสอบตามตัวแปรเพศ

ตารางที่ 24 ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบวิชาคณิตศาสตร์ สำหรับกลุ่มผู้สอบจำแนกตามเพศ

ข้อที่	ดัชนี NCDIF	ดัชนี CDIF	ข้อที่	ดัชนี NCDIF	ดัชนี CDIF
1	0.001	-0.006	16	0.009*	0.033
2	0.001	-0.004	17	0.005	-0.016
3	0.012*	0.024	18	0.001	-0.008
4	0.001	0.009	19	0.018*	0.045
5	0.001	-0.002	20	0.001	0.003
6	0.011*	-0.002	21	0.004	0.003
7	0.002	0.017	22	0.013*	-0.023
8	0.010*	0.024	23	0.007*	-0.021
9	0.001	0.006	24	0.007*	0.022
10	0.001	-0.001	25	0.011*	0.033
11	0.003	0.020	26	0.000	0.001
12	0.003	-0.003	27	0.013*	0.027
13	0.003	-0.011	28	0.002	-0.013
14	0.000	0.002			
15	0.003	-0.001			
รวมจำนวนข้อสอบที่พบ DIF		10 ข้อ			
ดัชนี DTF		0.374			

*หมายเหตุ พิจารณาตามเกณฑ์จุดตัดของ Fleer (1993 อ้างถึงใน Raju และคณะ (1995)) ข้อสอบและแบบสอบทำหน้าที่ต่างกันเมื่อดัชนี NCDIF และดัชนี DTF มีค่ามากกว่า 0.006

จากตารางที่ 24 ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบในวิชาคณิตศาสตร์ เมื่อจำแนกกลุ่มผู้สอบตามตัวแปรเพศ พบว่าในการวิเคราะห์การทำหน้าที่ต่างกันที่ระดับ ข้อสอบ มีข้อสอบที่ทำหน้าที่ต่างกันทั้งหมดจำนวน 10 ข้อ ได้แก่ข้อที่ 3, 6, 16, 19, 22, 24, 25 และ 27 และในการวิเคราะห์การทำหน้าที่ต่างกันที่ระดับแบบสอบ พบว่ามีค่าดัชนี DTF เท่ากับ 0.374 ซึ่งมีค่ามากกว่า 0.006 แสดงว่าแบบสอบดังกล่าวทำหน้าที่ต่างกัน หลังจากดำเนินการคัดเลือกข้อสอบออกจากแบบสอบทีละข้อตามกระบวนการ DFIT จากตารางที่ 24 พบว่าข้อสอบข้อที่ 19 มีค่าดัชนี CDIF สูงที่สุด คือมีค่าเท่ากับ 0.045 เมื่อตัดข้อสอบข้อที่ 19 ออกจากแบบสอบและทำการวิเคราะห์ค่าดัชนี CDIF และดัชนี DTF ใหม่อีกครั้ง ได้ผลดังตารางที่ 25

ตารางที่ 25 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาคณิตศาสตร์ สำหรับกลุ่มผู้สอบจำแนกตามเพศ เมื่อตัดข้อสอบข้อที่ 19 ออกจากแบบสอบ

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
1	-0.001	16*	0.006
2	0.000	17	-0.004
3*	0.000	18	-0.001
4	0.000	19*	ตัดออกจากการวิเคราะห์
5	-0.001	20	0.001
6*	-0.001	21	-0.001
7	0.002	22*	-0.004
8.	0.004	23	-0.004
9	0.000	24*	0.005
10	-0.001	25*	0.007
11	0.003	26	0.001
12	-0.001	27*	0.003
13	-0.003	28	-0.003
14	0.000		
15	-0.001		
ดัชนี DTF	0.007		

*หมายเหตุ พิจารณาตามเกณฑ์จุดตัดของ Fleer (1993 อ้างถึงใน Raju และคณะ (1995)) ข้อสอบและแบบสอบทำหน้าที่ต่างกันเมื่อดัชนี NCDIF และดัชนี DTF มีค่ามากกว่า 0.006

จากตารางที่ 25 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาคณิตศาสตร์ สำหรับกลุ่มผู้สอบจำแนกตามตัวแปรเพศ เมื่อตัดข้อสอบข้อที่ 19 ออกจากแบบสอบ พบว่าดัชนี DTF มีค่าเท่ากับ

0.007 ซึ่งมีค่ามากกว่า 0.006 จากตารางพบว่าข้อที่ 25 มีค่าดัชนี CDIF สูงที่สุด คือมีค่าเท่ากับ 0.007 เมื่อตัดข้อสอบข้อที่ 25 ออกจากแบบสอบและทำการวิเคราะห์ค่าดัชนี CDIF และดัชนี DTF ได้ผลดัง ตารางที่ 26

ตารางที่ 26 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาคณิตศาสตร์ สำหรับกลุ่มผู้สอบจำแนกตามเพศ เมื่อตัดข้อสอบข้อที่ 19 และข้อที่ 25 ออกจากแบบสอบ

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
1	-0.001	16*	0.005
2	0.000	17	-0.003
3*	0.001	18	-0.001
4	0.000	19*	ตัดออกจากการวิเคราะห์
5	-0.001	20	0.001
6*	-0.001	21	0.000
7	0.002	22*	-0.003
8	0.004	23	-0.003
9	0.001	24*	0.004
10	0.000	25*	ตัดออกจากการวิเคราะห์
11	0.003	26	0.001
12	-0.001	27*	0.002
13	-0.002	28	-0.002
14	0.001		
15	-0.001		
ดัชนี DTF	0.005		

*หมายเหตุ พิจารณาตามเกณฑ์จุดตัดของ Fleer (1993 อ้างถึงใน Raju และคณะ (1995)) ข้อสอบและแบบสอบทำหน้าที่ต่างกันเมื่อดัชนี NCDIF และดัชนี DTF มีค่ามากกว่า 0.006

จากตารางที่ 26 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาคณิตศาสตร์ สำหรับกลุ่มผู้สอบจำแนกตามตัวแปรเพศ เมื่อตัดข้อสอบข้อที่ 19 และข้อที่ 25 ออก พบว่าดัชนี DTF มีค่าเท่ากับ 0.005 ซึ่งมีค่าน้อยกว่า 0.006 แสดงว่าแบบสอบวิชาคณิตศาสตร์ที่ตัดข้อสอบข้อที่ 19 และ 25 ออกไม่ทำหน้าที่ต่างกัน หลังจากที่ทำการตัดข้อสอบที่มีค่าดัชนี CDIF สูงที่สุดออกทีละข้อ รวมทั้งสิ้น 2 ข้อ และผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบวิชาคณิตศาสตร์ เมื่อจำแนกตามตัวแปรสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา ซึ่งมีการวิเคราะห์การทำหน้าที่ต่างกันทั้งในรายข้อและรายฉบับ ได้ผลการวิเคราะห์ดังตารางที่ 27 ซึ่งมีรายละเอียดดังนี้

3.2.2 เมื่อจำแนกกลุ่มผู้สอบตามตัวแปรสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

ตารางที่ 27 ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบวิชาคณิตศาสตร์ สำหรับกลุ่มผู้สอบจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

ข้อที่	ดัชนี NCDIF	ดัชนี CDIF	ข้อที่	ดัชนี NCDIF	ดัชนี CDIF
1	0.004	0.000	16	0.015*	-0.003
2	0.004	-0.002	17	0.000	0.000
3	0.009*	0.002	18	0.000	0.000
4	0.000	0.000	19	0.001	0.001
5	0.000	0.000	20	0.009*	-0.001
6	0.004	0.002	21	0.001	-0.001
7	0.003	-0.001	22	0.005	0.000
8	0.006	0.000	23	0.000	0.000
9	0.000	0.000	24	0.032*	0.002
10	0.000	0.000	25	0.002	0.001
11	0.001	0.000	26	0.001	0.000
12	0.002	-0.001	27	0.008*	0.000
13.	0.007*	0.001	28	0.003	0.001
14	0.002	-0.001			
15	0.000	0.000			
รวมข้อสอบที่ DIF		6 ข้อ			
ดัชนี DTF		0.004			

*หมายเหตุ พิจารณาตามเกณฑ์จุดตัดของ Fleer (1993 อ้างถึงใน Raju และคณะ (1995)) ข้อสอบและแบบสอบทำหน้าที่ต่างกันเมื่อดัชนี NCDIF และดัชนี DTF มีค่ามากกว่า 0.006

จากตารางที่ 27 ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบวิชาคณิตศาสตร์ สำหรับกลุ่มผู้สอบ จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา พบว่าในการวิเคราะห์การทำหน้าที่ต่างกันที่ระดับข้อสอบ มีข้อสอบที่ทำหน้าที่ต่างกัน 6 ข้อ ได้แก่ข้อที่ 3, 13, 16, 20, 24 และ 27 และในการวิเคราะห์การทำหน้าที่ต่างกันที่ระดับแบบสอบ พบว่าค่าดัชนี DTF มีค่าเท่ากับ 0.004 ซึ่งมีค่าน้อยกว่า 0.006 แสดงว่าแบบสอบวิชาคณิตศาสตร์ ไม่ทำหน้าที่ต่างกัน เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

สำหรับการเปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบวิชาคณิตศาสตร์เมื่อจำแนกผู้สอบตามตัวแปรเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา ได้ผลดังตารางที่ 28

ตารางที่ 28 เปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบวิชาคณิตศาสตร์เมื่อจำแนกผู้สอบตามตัวแปรเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

ตัวแปร / ดัชนี	เพศ	สถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา
NCDIF	**3**, 6, 8, **16**, <u>19</u>, 22, 23, **24**, <u>25</u> และ **27**	**3**, 13, **16**, 20, **24** และ **27**
รวม	10 ข้อ	6 ข้อ
CDIF/DTF	<u>19</u> และ <u>25</u>	—
รวม	2 ข้อ	0 ข้อ

หมายเหตุ อักษรตัวเข้ม หมายถึง ข้อสอบที่ตรวจพบ DIF เมื่อจำแนกผู้สอบตามตัวแปรเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

อักษรขีดเส้นใต้ หมายถึงข้อสอบที่พบ DIF ทั้งโดยดัชนี NCDIF และ CDIF/DTF

จากตารางที่ 28 การเปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบวิชาคณิตศาสตร์เมื่อจำแนกผู้สอบตามตัวแปรเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา จะเห็นว่า เมื่อจำแนกกลุ่มผู้สอบตามตัวแปรเพศจะพบข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด คือ 10 ข้อ และหลังจากการตัดข้อสอบออกทีละข้อจนแบบสอบไม่ทำหน้าที่ต่างกัน ต้องตัดข้อสอบออกทั้งสิ้น 2 ข้อ เมื่อจำแนกกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา พบข้อสอบที่ DIF ทั้งหมด 6 ข้อ และมีข้อสอบ 4 ข้อ ได้แก่ ข้อที่ 3, 16, 24 และ 27 ที่ทำหน้าที่ต่างกัน เมื่อแบ่งกลุ่มผู้สอบตามตัวแปรที่ศึกษาทั้ง 2 ตัวแปร

ตอนที่ 4 ผลการเปรียบเทียบคุณภาพของแบบสอบก่อนและหลังการตัดข้อสอบที่ทำหน้าที่ต่างกัน

จากการวิเคราะห์คุณภาพของแบบสอบวิชาภาษาอังกฤษและวิชาคณิตศาสตร์ ทั้งก่อนและหลังการตัดข้อสอบที่ทำหน้าที่ต่างกัน ด้วยกระบวนการ DFIT เมื่อจำแนกกลุ่มผู้สอบตามตัวแปรเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา ผลการวิเคราะห์และเปรียบเทียบคุณภาพของแบบสอบในด้านความเที่ยง ความตรงเชิงโครงสร้าง และค่าฟังก์ชันสารสนเทศของแบบสอบ ซึ่งมีรายละเอียดดังนี้

4.1 ความเที่ยงของแบบสอบ

4.1.1 วิชาภาษาอังกฤษ

ผลการวิเคราะห์คุณภาพของแบบสอบในด้านความเที่ยงทั้งก่อนและหลังการตัดข้อสอบที่ทำหน้าที่ต่างกัน มีรายละเอียดดังตารางที่ 29

จากตารางที่ 29 ผลการเปรียบเทียบค่าความเที่ยงของแบบสอบวิชาภาษาอังกฤษก่อนและหลังการตัดข้อสอบที่พบ DIF จากตัวแปรเพศ และสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา พบว่าก่อนการตัดข้อสอบที่พบ DIF แบบสอบมีค่าความเที่ยงเท่ากับ 0.85 เมื่อตัดข้อสอบที่พบ DIF ในทุกกรณี ทำให้ค่าความเที่ยงของแบบสอบลดลง โดยกรณีที่ตัดข้อสอบทุกข้อที่พบ DIF เมื่อแบ่งกลุ่มผู้สอบตามตัวแปรเพศ ทำให้ค่าความเที่ยงของแบบสอบลดลงมากที่สุด คือมีค่าเท่ากับ 0.70 และเมื่อพิจารณาค่าความเที่ยงของแบบสอบฉบับใหม่คือหลังจากปรับขยายจำนวนข้อสอบให้เท่ากับ 100 ข้อในทุกกรณี แล้วคำนวณค่าความเที่ยงใหม่ โดยใช้สูตรของสเปียร์แมน พบว่าค่าความเที่ยงใหม่ที่ได้ในเกือบทุกกรณีแตกต่างจากค่าความเที่ยงของแบบสอบก่อนการตัดข้อสอบที่พบ DIF ออกจากแบบสอบอย่างมีนัยสำคัญทางสถิติ ที่ระดับ 0.01 ยกเว้นกรณีเมื่อแบ่งกลุ่มผู้สอบตามเพศและตัดข้อสอบที่พบ DIF ออกจากแบบสอบบางข้อ โดยใช้ดัชนี CDIF/DTF ที่ความเที่ยงของแบบสอบฉบับหลังตัดข้อสอบออกจากแบบสอบไม่แตกต่างจากก่อนตัดข้อสอบ

ตารางที่ 29 ผลการเปรียบเทียบค่าความเที่ยงของแบบสอบวิชาภาษาอังกฤษก่อนและหลังการตัดข้อสอบที่พบ DIF จากตัวแปรเพศ และสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

ลักษณะของแบบสอบ	จำนวนข้อสอบ (ข้อ)	ค่าความเที่ยงของแบบสอบ KR-20	จำนวนข้อสอบที่ขยาย (ข้อ)	ค่าความเที่ยงของแบบสอบฉบับใหม่	ค่าสถิติ t ก่อนและหลังตัดข้อสอบที่ DIF
ก่อนการตัดข้อสอบที่พบ DIF	100	0.85	100	0.85	-
หลังจากการตัดข้อสอบที่พบ DIF เมื่อแบ่งกลุ่มผู้สอบตามตัวแปรเพศ โดยใช้					
1. ดัชนี NCDIF	70	0.70	100	0.77	25.79**
2. ดัชนี CDIF/DTF	87	0.83	100	0.85	0.00
เมื่อแบ่งกลุ่มผู้สอบตามตัวแปรสถานที่ตั้งโรงเรียน โดยใช้					
1. ดัชนี NCDIF	84	0.81	100	0.84	10.19**
2. ดัชนี CDIF/DTF	93	0.83	100	0.84	10.19**

**p < 0.01

4.1.2 วิชาคณิตศาสตร์

ผลการวิเคราะห์คุณภาพของแบบสอบในด้านความเที่ยงทั้งก่อนและหลังการตัดข้อสอบที่ทำหน้าที่ต่างกัน มีรายละเอียดดังตารางที่ 30

จากตารางที่ 30 ผลการเปรียบเทียบค่าความเที่ยงของแบบสอบวิชาคณิตศาสตร์ก่อนและหลังการตัดข้อสอบที่พบ DIF จากตัวแปรเพศ และสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา พบว่าก่อนการตัดข้อสอบที่พบ DIF แบบสอบมีค่าความเที่ยงเท่ากับ 0.60 เมื่อตัดข้อสอบที่พบ DIF ใน เกือบทุกกรณี ทำให้ค่าความเที่ยงของแบบสอบลดลง โดยกรณีที่ตัดข้อสอบทุกข้อที่พบ DIF เมื่อแบ่งกลุ่มผู้สอบตามตัวแปรเพศ ทำให้ค่าความเที่ยงของแบบสอบลดลงมากที่สุด คือมีค่าเท่ากับ 0.42 และเมื่อพิจารณาค่าความเที่ยงของแบบสอบฉบับใหม่คือหลังจากปรับขยายจำนวนข้อสอบให้เท่ากับ 28 ข้อในทุกกรณี แล้วคำนวณค่าความเที่ยงใหม่ โดยใช้สูตรของสเปียร์แมน พบว่าเมื่อแบ่งกลุ่มตัวอย่างตามเพศ ทั้งในกรณีที่ตัดข้อสอบที่พบ DIF บางข้อและทุกข้อ ค่าความเที่ยงใหม่ที่ได้แตกต่างจาก

ค่าความเที่ยงของแบบสอบก่อนการตัดข้อสอบที่พบ DIF ออกจากแบบสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 แต่ในกรณีเมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งของโรงเรียนที่จบการศึกษา ค่าความเที่ยงใหม่ที่ได้ไม่แตกต่างจากค่าความเที่ยงเดิม

ตารางที่ 30 ผลการเปรียบเทียบค่าความเที่ยงของแบบสอบวิชาคณิตศาสตร์ก่อนและหลังการตัดข้อสอบที่พบ DIF จากตัวแปรเพศ และสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

ลักษณะของแบบสอบ	จำนวนข้อสอบ (ข้อ)	ค่าความเที่ยงของแบบสอบ (Alpha)	จำนวนข้อสอบที่ขยาย (ข้อ)	ค่าความเที่ยงของแบบสอบฉบับใหม่	ค่าสถิติ t ก่อนและหลังตัดข้อสอบที่ DIF
ก่อนการตัดข้อสอบที่พบ DIF	28	0.60	28	0.60	-
หลังจากการตัดข้อสอบที่พบ DIF เมื่อแบ่งกลุ่มผู้สอบตามตัวแปรเพศ โดยใช้					
1. ดัชนี NCDIF	18	0.42	28	0.53	7.68**
2. ดัชนี CDIF/DTF	26	0.59	28	0.61	-2.82**
เมื่อแบ่งกลุ่มผู้สอบตามตัวแปรสถานที่ตั้งโรงเรียน โดยใช้					
1. ดัชนี NCDIF	22	0.54	28	0.60	0.00
2. ดัชนี CDIF/DTF	28	0.60	28	0.60	0.00

**p < 0.01

4.2 การตรวจสอบคุณภาพของแบบสอบด้านความตรงของแบบสอบ

4.2.1 วิชาภาษาอังกฤษ

จากการวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) หลังจากตัดข้อสอบที่พบ DIF ทั้งในกรณีที่ตัดข้อสอบที่พบ DIF ทุกข้อและตัดเฉพาะบางข้อ โดยทำการวิเคราะห์แต่ละโมเดลที่มีการตัดข้อสอบออกผลการตัดข้อสอบที่พบ DIF ออกจากโมเดล มีรายละเอียดดังตารางที่ 31 และผลการวิเคราะห์ความตรงของแบบสอบวิชาภาษาอังกฤษ ในแต่ละโมเดลเมื่อแบ่งกลุ่มผู้สอบตามตัวแปรต่าง ๆ มีรายละเอียดดังตารางที่ 32 – 49

ตารางที่ 31 จำนวนข้อสอบที่ตัดออกจากแบบสอบวิชาภาษาอังกฤษ จำแนกตามโครงสร้างเนื้อหา ตัวแปรที่ใช้แบ่งกลุ่มผู้สอบและดัชนีที่ใช้ในการตัดข้อสอบออกจากแบบสอบ

โมเดลย่อยที่	จำนวนข้อก่อนตัดข้อสอบที่พบ DIF	ตัวแปร							
		เพศ				สถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน			
		NCDIF		CDIF/DTF		NCDIF		CDIF/DTF	
		จำนวนข้อที่ตัด	ข้อที่	จำนวนข้อที่ตัด	ข้อที่	จำนวนข้อที่ตัด	ข้อที่	จำนวนข้อที่ตัด	ข้อที่
1	25	4	41,91,93 และ96	4	41,46,90 และ91	2	44และ93	-	-
2	20	5	21,38,66, 68และ73	4	47,66,68 และ79	3	66,73 และ77	3	66,68 และ73
3	9	2	40และ42	-	-	2	39และ42	-	-
4	16	10	1-3,5,6, 8,9,11,12 และ16	2	1และ9	3	1,9 และ16	1	1
5	30	9	17,23,24, 27-30,34 และ53	3	17,20 และ33	6	24,33,52,56, 57 และ60	3	24,33 และ37
รวม (ข้อ)	100	30		13		16		7	

เมื่อโมเดลที่ 1 ประกอบด้วยเนื้อหาเรื่อง ความเข้าใจในการอ่าน ระดับข้อเท็จจริง

โมเดลที่ 2 ประกอบด้วยเนื้อหาเรื่อง ความเข้าใจในการอ่าน ระดับตีความ

โมเดลที่ 3 ประกอบด้วยเนื้อหาเรื่อง ความเข้าใจในการอ่าน ระดับประเมินค่า

โมเดลที่ 4 ประกอบด้วยเนื้อหาเรื่อง ความสามารถในการใช้ภาษา

โมเดลที่ 5 ประกอบด้วยเนื้อหาเรื่อง ความรู้เกี่ยวกับโครงสร้างไวยากรณ์

จากตารางที่ 31 แสดงจำนวนข้อสอบที่ตัดออกจากแบบสอบวิชาภาษาอังกฤษ จำแนกตามโครงสร้างเนื้อหา ตัวแปรที่ใช้แบ่งกลุ่มผู้สอบและดัชนีที่ใช้ในการตัดข้อสอบออกจากแบบสอบ พบว่าเมื่อจำแนกกลุ่มผู้สอบตามเพศ ในการวิเคราะห์ DIF โดยใช้ดัชนี NCDIF พบข้อสอบที่ DIF ในทุกลักษณะเนื้อหา โดยลักษณะเนื้อหาเรื่องความสามารถในการใช้ภาษาเป็นกลุ่มเนื้อหาที่พบข้อสอบที่ DIF มากที่สุด จำนวน 10 ข้อ รองลงมาเป็นเนื้อหาเรื่องความรู้เกี่ยวกับโครงสร้างไวยากรณ์ พบข้อสอบที่ DIF จำนวน 9 ข้อ ในการวิเคราะห์ DTF โดยใช้ดัชนี CDIF/DTF พบว่าลักษณะเนื้อหาเรื่องความเข้าใจในการอ่านระดับข้อเท็จจริง และระดับตีความ เป็นเนื้อหาที่พบ DIF มากที่สุด จำนวน 4 ข้อเท่ากัน

เมื่อจำแนกกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา ในการวิเคราะห์ DIF โดยใช้ดัชนี NCDIF พบข้อสอบที่ DIF ในทุกลักษณะเนื้อหา เนื้อหาเรื่องความรู้เกี่ยวกับโครงสร้างไวยากรณ์ พบข้อสอบที่ DIF มากที่สุด จำนวน 6 ข้อ ในการวิเคราะห์ DTF โดยใช้ดัชนี CDIF/DTF พบว่าลักษณะเนื้อหาเรื่องความเข้าใจในการอ่านระดับตีความและความรู้เกี่ยวกับโครงสร้างไวยากรณ์เป็นเนื้อหาที่พบ DIF มากที่สุด จำนวนเนื้อหาละ 3 ข้อเท่ากัน

หลังจากนั้นทำการตรวจสอบคุณภาพของแบบสอบในด้านความตรงเชิงโครงสร้างวิชาภาษาอังกฤษ หลังจากการวิเคราะห์ DIF และ DTF เมื่อจำแนกกลุ่มผู้สอบตามตัวแปรที่ศึกษาแล้ว โดยทำการวิเคราะห์ด้วยวิธีวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) เพื่อสร้างสเกลองค์ประกอบตามโมเดลย่อยต่าง ๆ 5 โมเดล ได้ผลการวิเคราะห์ มีรายละเอียดดังนี้

4.2.1.1 เมื่อจำแนกกลุ่มผู้สอบตามตัวแปรเพศ

ก) หลังจากการวิเคราะห์ DIF

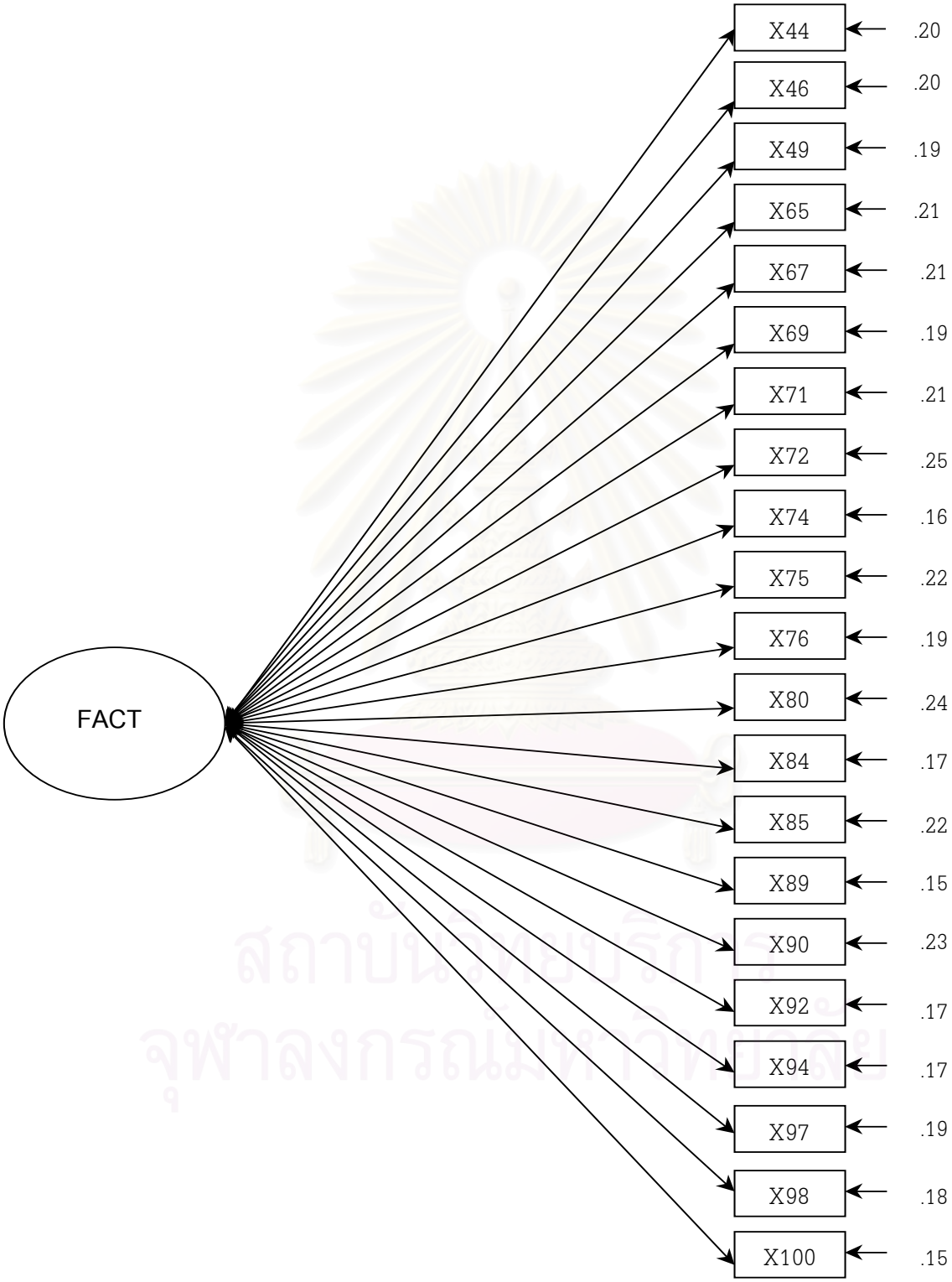
1. โมเดลด้านความเข้าใจในการอ่านระดับข้อเท็จจริง หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับข้อเท็จจริง พบว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ดีมาก พิจารณาได้จากค่าไค-สแควร์ (Chi-Square = 142.664) ซึ่งมีค่าความน่าจะเป็นเข้าใกล้ 1 (P = 0.995) ที่องศาอิสระเท่ากับ 189 (df = 189) นั่นคือค่าไค-สแควร์ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่ายอมรับสมมติฐานหลักที่ว่าโมเดลการวิจัยสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ส่วนค่าดัชนีวัดความกลมกลืน (GFI) มีค่าเท่ากับ 0.986 ค่าดัชนีวัดความกลมกลืนที่ปรับแก้แล้ว (AGFI) มีค่าเท่ากับ 0.982 และค่าดัชนีวัดความสอดคล้อง (NFI) มีค่าเท่ากับ 0.868 รวมทั้งค่าดัชนีรากกำลังสองเฉลี่ยของเศษ (RMR) มีค่าเท่ากับ 0.005 ซึ่งมีค่าน้อยมากจนเข้าใกล้ศูนย์ ดังรายละเอียดในตารางที่ 32

ตารางที่ 32 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่านระดับข้อเท็จจริง (FACT) ของแบบสอบวิชาภาษาอังกฤษ หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

ตัวแปร	เมตริกซ์น้ำหนักองค์ประกอบ			R Square	สปส. คะแนนองค์ประกอบ
	สปส.	SE.	t		
X44	0.12	0.02	6.32**	0.06	0.15
X46	0.17	0.02	9.89**	0.13	0.26
X49	0.23	0.02	13.03**	0.22	0.37
X65	0.15	0.02	8.56**	0.10	0.22
X67	0.11	0.02	6.43**	0.06	0.16
X69	0.12	0.02	7.36**	0.08	0.20
X71	0.19	0.02	10.58**	0.15	0.27
X72	0.01	0.02	0.69	0.00	0.02
X74	0.12	0.02	7.89**	0.09	0.23
X75	0.07	0.02	3.98**	0.02	0.10
X76	0.10	0.02	5.86**	0.05	0.15
X80	0.09	0.02	4.94**	0.03	0.12
X84	0.11	0.02	6.68**	0.06	0.19
X85	0.15	0.02	8.39**	0.10	0.21
X89	0.11	0.02	8.00**	0.07	0.21
X90	0.11	0.02	5.99**	0.05	0.15
X92	0.20	0.02	11.83**	0.19	0.35
X94	0.12	0.02	7.41**	0.08	0.21
X97	0.08	0.02	4.76**	0.03	0.13
X98	0.17	0.02	10.12**	0.14	0.28
X100	0.24	0.02	14.88**	0.28	0.48
Chi-Square = 142.664		df = 189	P = 0.995	GFI = 0.986	
AGFI = 0.982		NFI = 0.868	RMR = 0.005		

** $p < 0.01$



แผนภาพที่ 14 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันโมเดลด้านความเข้าใจในการอ่านระดับข้อเท็จจริง หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มตามเพศ

เมื่อพิจาราณาผลการวิเคราะห์โมเดลด้านความเข้าใจในการอ่านระดับข้อเท็จจริง จากตารางที่ 32 และจากแผนภาพที่ 14 พบว่าน้ำหนักองค์ประกอบส่วนใหญ่ที่กำหนดข้อมูลจำเพาะของโมเดล CFA ไว้แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ยกเว้นตัวแปร X72 ที่น้ำหนักองค์ประกอบไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญโดยน้ำหนักองค์ประกอบที่มีนัยสำคัญมีค่าอยู่ระหว่าง 0.07-0.24 และค่าสัมประสิทธิ์การพยากรณ์ (R-square) ของตัวแปร X100 มีค่าสูงสุดเท่ากับ 0.27

สำหรับการสร้างสเกลองค์ประกอบ ผู้วิจัยได้นำค่าสัมประสิทธ์คะแนนองค์ประกอบที่ได้จากการวิเคราะห์ใช้ในการสร้างสเกลองค์ประกอบด้านความเข้าใจในการอ่านระดับข้อเท็จจริง (FACT) เพื่อให้ได้ตัวแปรใหม่สำหรับนำไปวิเคราะห์เพื่อตรวจสอบความตรงเชิงโครงสร้างของแบบสอบวิชาภาษาอังกฤษต่อไป ซึ่งได้สเกลองค์ประกอบของความเข้าใจในการอ่านระดับข้อเท็จจริง ดังสมการต่อไปนี้

$$\begin{aligned} FACT = {} & 0.15Z_{X44} + 0.26Z_{X46} + 0.37Z_{X49} + 0.22Z_{X65} + 0.16Z_{X67} + 0.20Z_{X69} + \\ & 0.27Z_{X71} + 0.02Z_{X72} + 0.23Z_{X74} + 0.10Z_{X75} + 0.15Z_{X76} + 0.12Z_{X80} + \\ & 0.19Z_{X84} + 0.21Z_{X85} + 0.21Z_{X89} + 0.15Z_{X90} + 0.35Z_{X92} + 0.21Z_{X94} + \\ & 0.13Z_{X97} + 0.28Z_{X98} + 0.48Z_{X100} \end{aligned}$$

2. โมเดลด้านความเข้าใจในการอ่าน ระดับตีความ หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

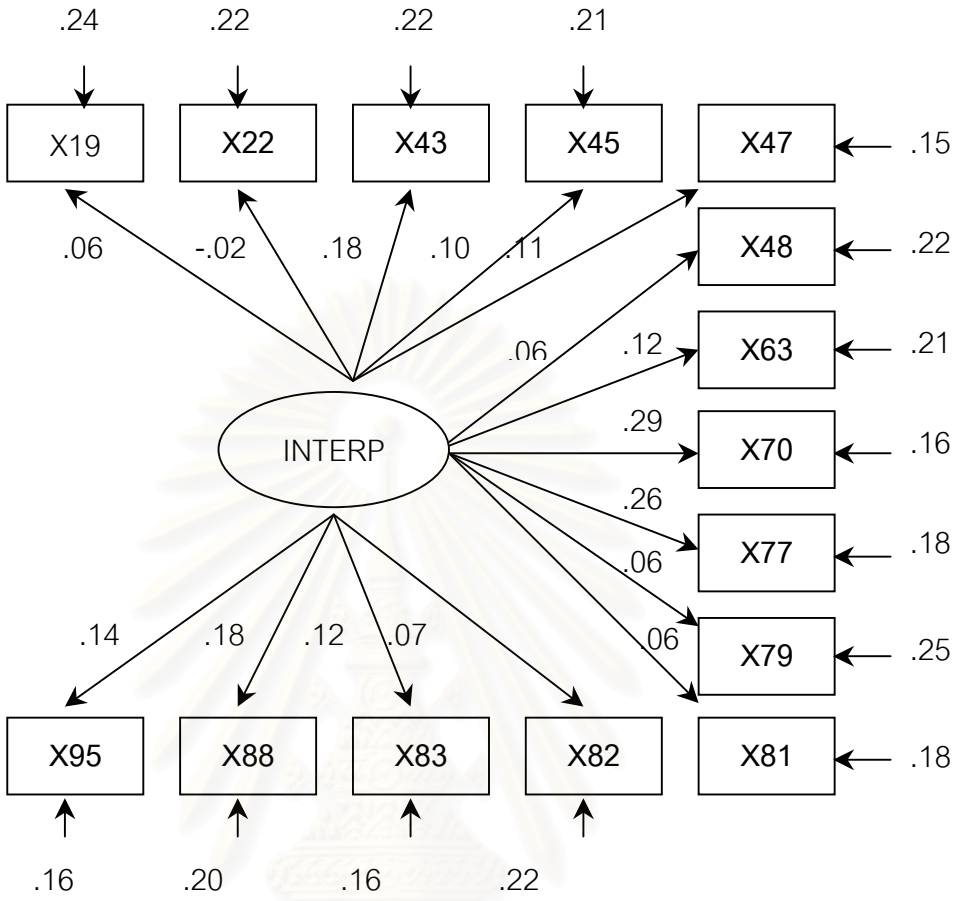
ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับตีความพบว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ดีมาก พิจารณาได้จากค่าไค-สแควร์ (Chi-Square = 62.715) ซึ่งมีค่าความน่าจะเป็นเข้าใกล้ 1 (P = 0.987) ที่องศาอิสระเท่ากับ 90 (df = 90) นั่นคือค่าไค-สแควไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่ายอมรับสมมติฐานหลักที่ว่าโมเดลการวิจัยสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ส่วนค่าดัชนีวัดความกลมกลืน (GFI) มีค่าเท่ากับ 0.991 ค่าดัชนีวัดความกลมกลืนที่ปรับแก้แล้ว (AGFI) มีค่าเท่ากับ 0.988 และค่าดัชนีวัดความสอดคล้อง (NFI) มีค่าเท่ากับ 0.899 รวมทั้งค่าดัชนีรากกำลังสองเฉลี่ยของเศษ (RMR) มีค่าเท่ากับ 0.004 ซึ่งมีค่าน้อยมากจนเข้าใกล้ศูนย์ ดังรายละเอียดในตารางที่ 33 และรายละเอียดของโมเดลด้านความเข้าใจในการอ่านระดับตีความ สามารถแสดงได้ดังแผนภาพที่ 15

ตารางที่ 33 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับตีความ (INTERP) ของแบบสอบวิชาภาษาอังกฤษ หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

ตัวแปร	เมตริกซ์น้ำหนักองค์ประกอบ			R	สปส. คะแนนองค์ประกอบ
	สปส.	SE.	t	**Square**	
X19	0.06	0.02	3.13**	0.02	0.09
X22	-0.02	0.02	-1.04	0.00	-0.03
X43	0.18	0.02	9.26**	0.13	0.29
X45	0.10	0.02	5.19**	0.04	0.16
X47	0.11	0.02	7.30**	0.08	0.28
X48	0.06	0.02	3.40**	0.02	0.10
X63	0.12	0.02	6.40**	0.06	0.20
X70	0.29	0.02	15.55**	0.36	0.67
X77	0.26	0.02	13.83**	0.28	0.52
X79	0.06	0.02	2.88**	0.01	0.08
X81	0.13	0.02	7.37**	0.08	0.25
X82	0.07	0.02	3.69**	0.02	0.11
X83	0.12	0.02	7.38**	0.08	0.27
X88	0.18	0.02	9.50**	0.13	0.31
X95	0.14	0.02	8.49**	0.11	0.31
Chi-Square = 62.715		df = 90	P = 0.987	GFI = 0.991	
AGFI = 0.988		NFI = 0.899	RMR = 0.004		

** $p < 0.01$

เมื่อพิจาราณาผลการวิเคราะห์โมเดลด้านความเข้าใจในการอ่านระดับตีความ จากตารางที่ 33 และจากแผนภาพที่ 15 พบว่าน้ำหนักองค์ประกอบเกือบทุกค่าที่กำหนดข้อมูลจำเพาะของโมเดล CFA ไว้แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ยกเว้นตัวแปร X22 ที่น้ำหนักองค์ประกอบไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ โดยน้ำหนักองค์ประกอบที่มีนัยสำคัญมีค่าอยู่ระหว่าง 0.06-0.29 และค่าสัมประสิทธิ์การพยากรณ์ (R-square) ของตัวแปร X70 มีค่าสูงสุดเท่ากับ 0.36



แผนภาพที่ 15 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่านระดับตีความ หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

สำหรับการสร้างสเกลองค์ประกอบ ผู้วิจัยได้นำค่าสัมประสิทธิ์คะแนนองค์ประกอบที่ได้จากการวิเคราะห์ใช้ในการสร้างสเกลองค์ประกอบด้านความเข้าใจในการอ่านระดับตีความ (INTERP) เพื่อให้ได้ตัวแปรใหม่สำหรับนำไปวิเคราะห์เพื่อตรวจสอบความตรงเชิงโครงสร้างของแบบสอบวิชาภาษาอังกฤษต่อไป ซึ่งได้สเกลองค์ประกอบของความเข้าใจในการอ่านระดับตีความ ดังสมการต่อไปนี้

$$\begin{aligned}INTERP = {} & 0.09Z_{X19} - 0.03Z_{X22} + 0.29Z_{X43} + 0.16Z_{X45} + 0.28Z_{X47} + 0.10Z_{X48} + \\ & 0.20Z_{X63} + 0.67Z_{X70} + 0.52Z_{X77} + 0.08Z_{X79} + 0.25Z_{X81} + 0.11Z_{X82} + \\ & 0.26Z_{X83} + 0.31Z_{X88} + 0.33Z_{X95}\end{aligned}$$

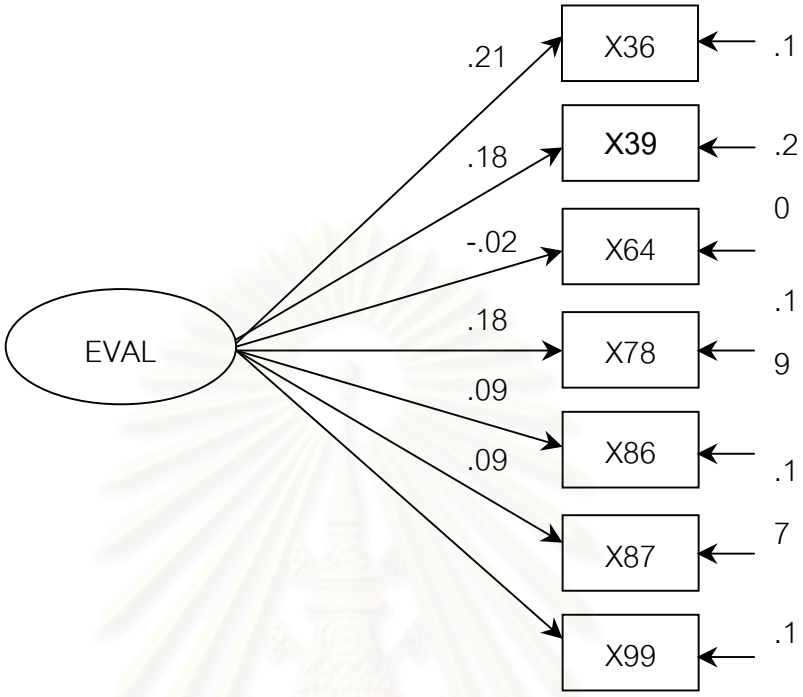
3. โมเดลด้านความเข้าใจในการอ่าน ระดับประเมินค่า หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับประเมินค่า พบว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ดีมาก พิจารณาได้จากค่าไค-สแควร์ (Chi-Square = 9.269) ซึ่งมีค่าความน่าจะเป็นเข้าใกล้ 1 (P = 0.813) ที่องศาอิสระเท่ากับ 14 (df = 14) นั่นคือค่าไค-สแควร์ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่ายอมรับสมมติฐานหลักที่ว่าโมเดลการวิจัยสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ส่วนค่าดัชนีวัดความกลมกลืน (GFI) มีค่าเท่ากับ 0.997 ค่าดัชนีวัดความกลมกลืนที่ปรับแก้แล้ว (AGFI) มีค่าเท่ากับ 0.995 และค่าดัชนีวัดความสอดคล้อง (NFI) มีค่าเท่ากับ 0.940 รวมทั้งค่าดัชนีรากกำลังสองเฉลี่ยของเศษ (RMR) มีค่าเท่ากับ 0.003 ซึ่งมีค่าน้อยมากจนเข้าใกล้ศูนย์ ดังรายละเอียดในตารางที่ 34 และรายละเอียดของโมเดลด้านความเข้าใจในการอ่านระดับประเมินค่า สามารถแสดงได้ดังแผนภาพที่ 16

ตารางที่ 34 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับประเมินค่า (EVAL) ของแบบสอบวิชาภาษาอังกฤษ หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

ตัวแปร	เมตริกซ์น้ำหนักองค์ประกอบ			R Square	สปส. คะแนนองค์ประกอบ
	สปส.	SE.	t		
X36	0.21	0.03	8.38**	0.20	0.67
X39	0.18	0.02	7.26**	0.13	0.49
X64	-0.02	0.02	-0.83	0.00	-0.05
X78	0.18	0.02	7.88**	0.17	0.61
X86	0.09	0.02	4.71**	0.05	0.34
X87	0.09	0.02	3.75**	0.03	0.22
X99	0.13	0.02	6.07**	0.09	0.42
Chi-Square = 9.269		df = 14	P = 0.813	GFI = 0.997	
AGFI = 0.995		NFI = 0.940	RMR = 0.003		

** $p < 0.01$



แผนภาพที่ 16 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่านระดับประเมินค่า หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

เมื่อพิจาราณาผลการวิเคราะห์โมเดลด้านความเข้าใจในการอ่านระดับประเมินค่า จากตารางที่ 34 และจากแผนภาพที่ 16 พบว่าน้ำหนักองค์ประกอบส่วนใหญ่ที่กำหนดข้อมูลจำเพาะของโมเดล CFA ไว้แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ยกเว้นตัวแปร X64 ที่น้ำหนักองค์ประกอบไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ โดยน้ำหนักองค์ประกอบที่มีนัยสำคัญมีค่าอยู่ระหว่าง 0.09-0.18 และค่าสัมประสิทธิ์การพยากรณ์ (R-square) ของตัวแปร X36 มีค่าสูงสุดเท่ากับ 0.20

สำหรับการสร้างสเกลองค์ประกอบ ผู้วิจัยได้นำค่าสัมประสิทธ์คะแนนองค์ประกอบที่ได้จากการวิเคราะห์ใช้ในการสร้างสเกลองค์ประกอบด้านความเข้าใจในการอ่านระดับประเมินค่า (EVAL) เพื่อให้ได้ตัวแปรใหม่สำหรับนำไปวิเคราะห์เพื่อตรวจสอบความตรงเชิงโครงสร้างของแบบสอบวิชาภาษาอังกฤษต่อไป ซึ่งได้สเกลองค์ประกอบของความเข้าใจในการอ่านระดับประเมินค่า ดังสมการต่อไปนี้

$$EVAL = 0.67Z_{X36} + 0.49Z_{X39} - 0.05Z_{X64} + 0.61Z_{X78} + 0.34Z_{X86} + 0.22Z_{X87} + 0.42Z_{X99}$$

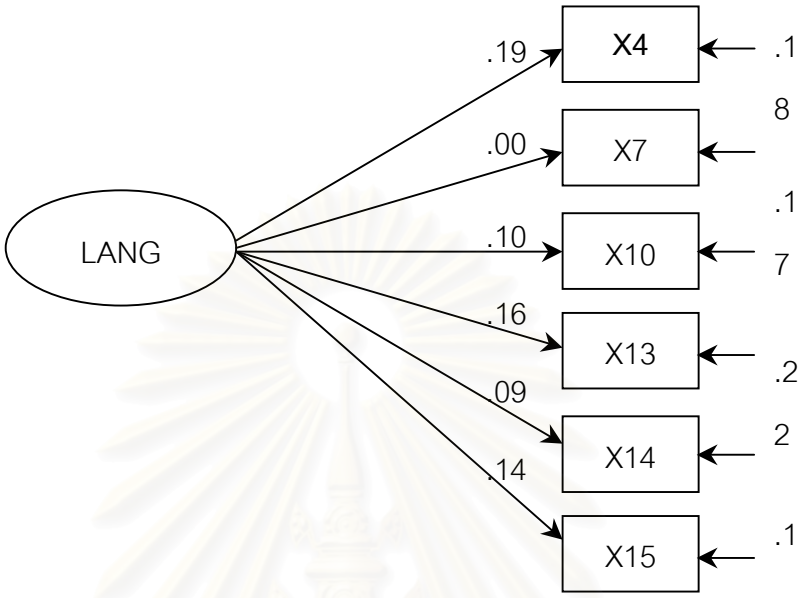
4. โมเดลด้านความสามารถในการใช้ภาษา หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความสามารถในการใช้ภาษา พบว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ พิจารณาได้จากค่าไค-สแควร์ (Chi-Square = 5.300) ซึ่งมีค่าความน่าจะเป็นเท่ากับ 0.807 ที่องศาอิสระเท่ากับ 9 (df = 9) นั่นคือค่าไค-สแควร์ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่ายอมรับสมมติฐานหลักที่ว่าโมเดลการวิจัยสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ส่วนค่าดัชนีวัดความกลมกลืน (GFI) มีค่าเท่ากับ 0.998 ค่าดัชนีวัดความกลมกลืนที่ปรับแก้แล้ว (AGFI) มีค่าเท่ากับ 0.996 และค่าดัชนีวัดความสอดคล้อง (NFI) มีค่าเท่ากับ 0.937 รวมทั้งค่าดัชนีรากกำลังสองเฉลี่ยของเศษ (RMR) มีค่าเท่ากับ 0.003 ซึ่งมีค่าน้อยมากจนเข้าใกล้ศูนย์ ดังรายละเอียดในตารางที่ 35 และรายละเอียดของโมเดลด้านความสามารถในการใช้ภาษา สามารถแสดงได้ดังแผนภาพที่ 17

ตารางที่ 35 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความสามารถในการใช้ภาษา (LANG) ของแบบสอบวิชาภาษาอังกฤษ หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

ตัวแปร	เมตริกซ์น้ำหนักองค์ประกอบ			R Square	สปส. คะแนนองค์ประกอบ
	สปส.	SE.	t		
X4	0.19	0.03	6.31**	0.17	0.68
X7	0.00	0.02	0.08	0.00	0.01
X10	0.10	0.03	3.69**	0.04	0.28
X13	0.16	0.03	5.67**	0.11	0.53
X14	0.09	0.02	3.88**	0.05	0.35
X15	0.14	0.2	5.78**	0.12	0.62
Chi-Square = 5.300		df = 9	P = 0.807	GFI = 0.998	
AGFI = 0.996		NFI = 0.937	RMR = 0.003		

** $p < 0.01$



แผนภาพที่ 17 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความสามารถในการใช้ภาษา หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

เมื่อพิจารณาผลการวิเคราะห์โมเดลด้านความสามารถในการใช้ภาษา จากตารางที่ 35 และจากแผนภาพที่ 17 พบว่าน้ำหนักองค์ประกอบส่วนใหญ่ที่กำหนดข้อมูลจำเพาะของโมเดล CFA ไว้แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ยกเว้นตัวแปร X7 ที่น้ำหนักองค์ประกอบไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ โดยน้ำหนักองค์ประกอบที่มีนัยสำคัญมีค่าอยู่ระหว่าง 0.09-0.19 และค่าสัมประสิทธิ์การพยากรณ์ (R-square) ของตัวแปร X3 มีค่าสูงสุดเท่ากับ 0.59

สำหรับการสร้างสเกลองค์ประกอบ ผู้วิจัยได้นำค่าสัมประสิทธ์คะแนนองค์ประกอบที่ได้จากการวิเคราะห์ใช้ในการสร้างสเกลองค์ประกอบด้านความสามารถในการใช้ภาษา (LANG) เพื่อให้ได้ตัวแปรใหม่สำหรับนำไปวิเคราะห์เพื่อตรวจสอบความตรงเชิงโครงสร้างของแบบสอบวิชาภาษาอังกฤษต่อไป ซึ่งได้สเกลองค์ประกอบของความสามารถในการใช้ภาษา ดังสมการต่อไปนี้

$$LANG = 0.68Z_{X4} + 0.01Z_{X7} + 0.28Z_{X10} + 0.53Z_{X13} + 0.35Z_{X14} + 0.62Z_{X15}$$

5. โมเดลด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ พบว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ดี พิจารณาได้จากค่าไค-สแควร์ (Chi-Square = 132.394) ซึ่งมีค่าความน่าจะเป็นเท่ากับ 0.999 ที่องศาอิสระเท่ากับ 189 (df = 189) นั่นคือค่าไค-สแควร์ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่ายอมรับสมมติฐานหลักที่ว่าโมเดลการวิจัยสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ส่วนค่าดัชนีวัดความกลมกลืน (GFI) มีค่าเท่ากับ 0.987 ค่าดัชนีวัดความกลมกลืนที่ปรับแก้แล้ว (AGFI) มีค่าเท่ากับ 0.984 และค่าดัชนีวัดความสอดคล้อง (NFI) มีค่าเท่ากับ 0.807 รวมทั้งค่าดัชนีรากกำลังสองเฉลี่ยของเศษ (RMR) มีค่าเท่ากับ 0.005 ซึ่งมีค่า น้อยมากจนเข้าใกล้ศูนย์ ดังรายละเอียดในตารางที่ 36 และรายละเอียดของโมเดลด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ สามารถแสดงได้ดังแผนภาพที่ 18

เมื่อพิจารณาผลการวิเคราะห์โมเดลด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ จากตารางที่ 36 และจากแผนภาพที่ 18 พบว่าน้ำหนักองค์ประกอบส่วนใหญ่ที่กำหนดข้อมูลจำเพาะของโมเดล CFA ไว้แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ยกเว้นตัวแปร X54 และ X61 ที่น้ำหนักองค์ประกอบไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ โดยน้ำหนักองค์ประกอบที่มีนัยสำคัญมีค่าอยู่ระหว่าง 0.06-0.22 และค่าสัมประสิทธิ์การพยากรณ์ (R-square) ของตัวแปร X35 มีค่าสูงสุดเท่ากับ 0.21

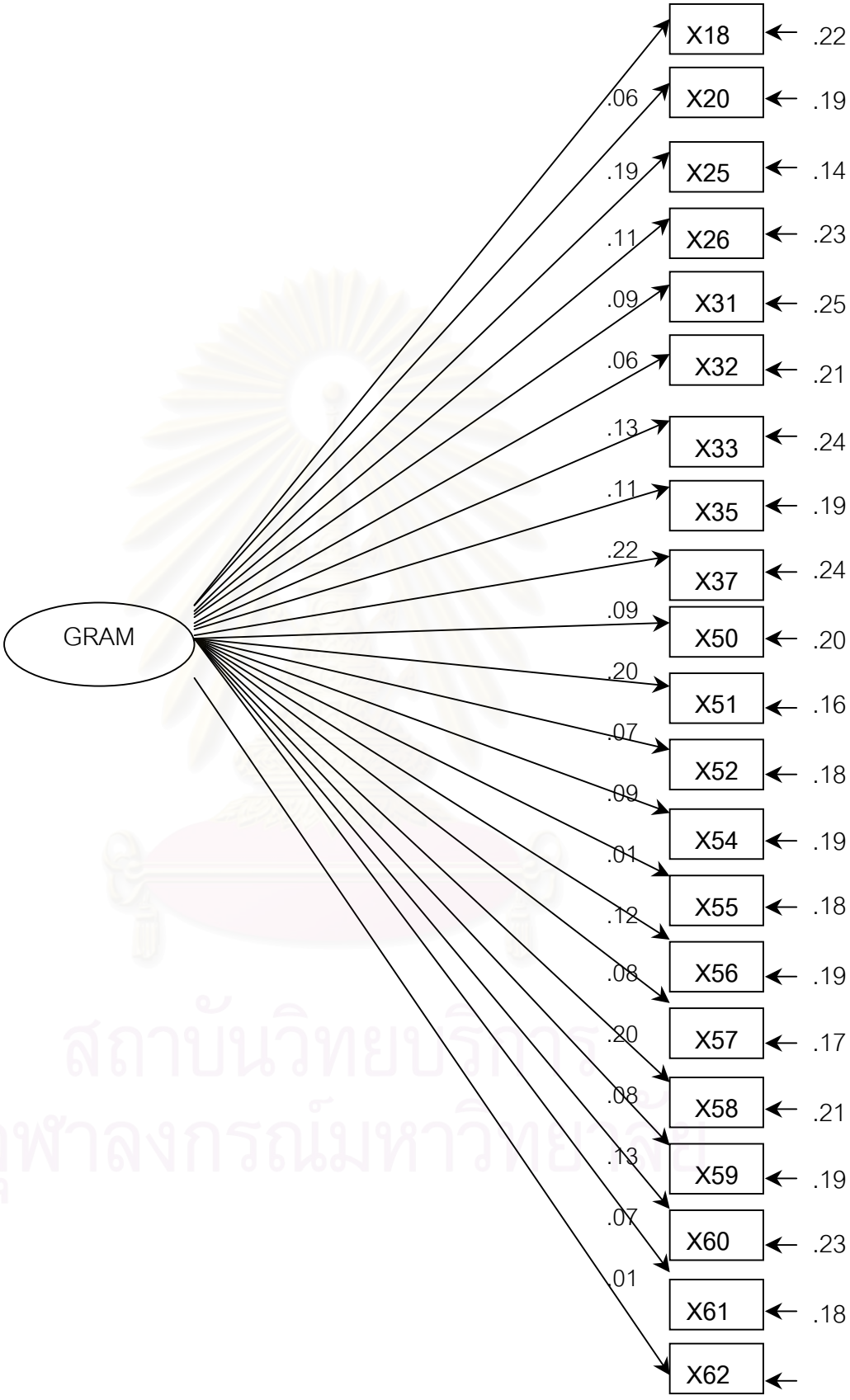
สำหรับการสร้างสเกลองค์ประกอบ ผู้วิจัยได้นำค่าสัมประสิทธ์คะแนนองค์ประกอบที่ได้จากการวิเคราะห์ใช้ในการสร้างสเกลองค์ประกอบด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ (GRAM) เพื่อให้ได้ตัวแปรใหม่สำหรับนำไปวิเคราะห์เพื่อตรวจสอบความตรงเชิงโครงสร้างของแบบสอบวิชาภาษาอังกฤษต่อไป ซึ่งได้สเกลองค์ประกอบของความรู้เกี่ยวกับโครงสร้างไวยากรณ์ ดังสมการต่อไปนี้

$$\begin{aligned} GRAM = {} & 0.11Z_{X18} + 0.37Z_{X20} + 0.30Z_{X25} + 0.15Z_{X26} + 0.09Z_{X31} + 0.23Z_{X32} + \\ & 0.18Z_{X33} + 0.45Z_{X35} + 0.15Z_{X37} + 0.39Z_{X50} + 0.17Z_{X51} + 0.20Z_{X52} + \\ & 0.02Z_{X54} + 0.26Z_{X55} + 0.15Z_{X56} + 0.44Z_{X57} + 0.14Z_{X58} + 0.26Z_{X59} + \\ & 0.12Z_{X50} + 0.03Z_{X61} + 0.22Z_{X62} \end{aligned}$$

ตารางที่ 36 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ (GRAM) ของแบบสอบวิชาภาษาอังกฤษหลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

ตัวแปร	เมตริกซ์น้ำหนักองค์ประกอบ			R Square	สปส. คะแนนองค์ประกอบ
	สปส.	SE.	t		
X18	0.06	0.02	3.28**	0.02	0.11
X20	0.19	0.02	10.00**	0.16	0.37
X25	0.11	0.01	7.22**	0.08	0.30
X16	0.09	0.02	4.27**	0.04	0.15
X31	0.06	0.02	3.01**	0.02	0.09
X32	0.13	0.02	6.73**	0.07	0.23
X33	0.11	0.02	5.57**	0.05	0.18
X35	0.22	0.02	11.64**	0.21	0.45
X37	0.09	0.02	4.72**	0.04	0.15
X50	0.20	0.02	10.58**	0.17	0.39
X51	0.07	0.02	4.44**	0.03	0.17
X52	0.09	0.02	5.44**	0.05	0.20
X54	0.01	0.02	0.69	0.00	0.02
X55	0.12	0.02	7.01**	0.08	0.26
X56	0.08	0.02	4.31**	0.03	0.15
X57	0.20	0.02	10.99**	0.19	0.44
X58	0.08	0.02	4.04**	0.03	0.14
X59	0.13	0.02	7.27**	0.08	0.26
X60	0.07	0.02	3.74**	0.02	0.12
X61	0.01	0.02	0.74	0.00	0.03
X62	0.14	0.02	6.82**	0.07	0.22
Chi-Square = 132.394		df = 189	P = 0.999	GFI = 0.987	
AGFI = 0.984		NFI = 0.807	RMR = 0.005		

** $p < 0.01$



แผนภาพที่ 18 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

โมเดลองค์ประกอบความรู้วิชาภาษาอังกฤษ หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

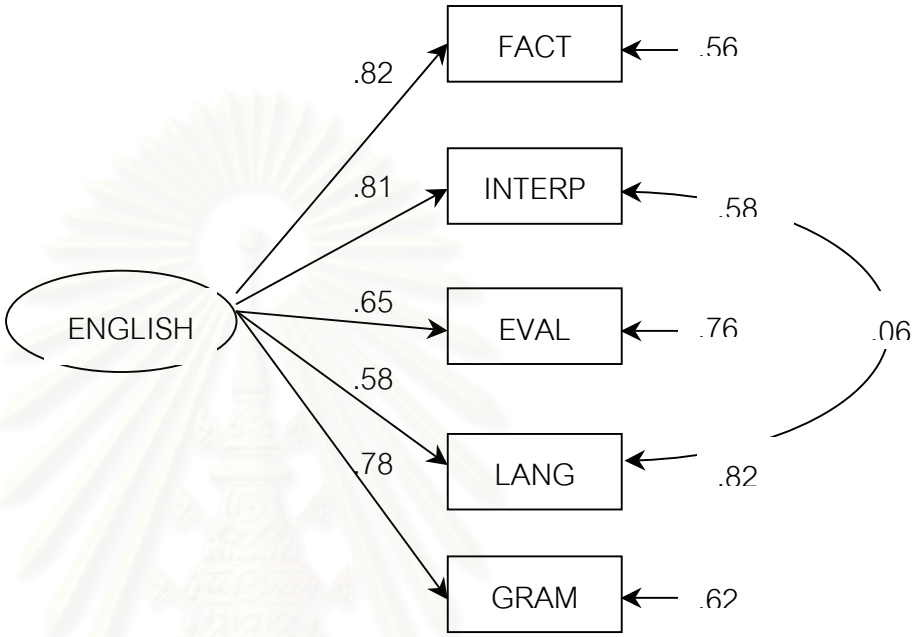
ผลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับที่สอง (second order confirmatory factor analysis) ของโมเดลองค์ประกอบความรู้วิชาภาษาอังกฤษ (ENGLISH) พบว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ดีมาก โดยพิจารณาจากค่าไค-สแควร์ (Chi-Square = 0.932) ซึ่งมีค่าความน่าจะเป็นเท่ากับ 0.920 ที่องศาอิสระเท่ากับ 4 (df = 4) นั่นคือค่าไค-สแควร์ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่ายอมรับสมมติฐานหลักที่ว่าโมเดลการวิจัยสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ส่วนค่าดัชนีวัดความกลมกลืน (GFI) มีค่าเท่ากับ 1.00 ค่าดัชนีวัดความกลมกลืนที่ปรับแก้แล้ว (AGFI) มีค่าเท่ากับ 0.999 และค่าดัชนีวัดความสอดคล้อง (NFI) มีค่าเท่ากับ 1.00 รวมทั้งค่าดัชนีรากกำลังสองเฉลี่ยของเศษ (RMR) มีค่าเท่ากับ 0.009 ซึ่งมีค่าน้อยมากจนเข้าใกล้ศูนย์ ดังรายละเอียดในตารางที่ 37

ตารางที่ 37 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลองค์ประกอบความรู้วิชาภาษาอังกฤษ (ENGLISH) หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

องค์ประกอบ	เมตริกซ์น้ำหนักองค์ประกอบ			R Square	สปส. คะแนนองค์ประกอบ
	สปส.	SE.	t		
ความเข้าใจในการอ่าน ระดับข้อเท็จจริง (FACT)	0.82	0.05	30.03**	0.68	0.18
ความเข้าใจในการอ่าน ระดับตีความ (INTERP)	0.81	0.05	29.15**	0.66	0.19
ความเข้าใจในการอ่าน ระดับประเมินค่า (EVAL)	0.65	0.04	21.80**	0.42	0.10
ความสามารถในการใช้ภาษา (LANG)	0.58	0.04	18.26**	0.33	0.10
ความรู้เกี่ยวกับโครงสร้างไวยากรณ์ (GRAM)	0.78	0.05	27.72**	0.61	0.14
Chi-Square = 0.932	df = 4	P = 0.920		GFI = 1.00	
AGFI = 0.999	NFI = 1.00	RMR = 0.009			

** $p < 0.01$

รายละเอียดของโมเดลด้านความเข้าใจในการอ่านระดับประเมินค่า สามารถแสดงได้ดังแผนภาพที่ 19



แผนภาพที่ 19 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลองค์ประกอบความรู้วิชาภาษาอังกฤษหลังจากการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

เมื่อพิจารณาผลการวิเคราะห์โมเดลความรู้วิชาภาษาอังกฤษ จากตารางที่ 37 และจากแผนภาพที่ 19 พบว่าน้ำหนักองค์ประกอบทุกค่าที่กำหนดข้อมูลจำเพาะของโมเดล CFA ไว้แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 โดยองค์ประกอบด้านความเข้าใจในการอ่านระดับข้อเท็จจริง (FACT) มีน้ำหนักองค์ประกอบมากที่สุด คือ 0.82 และองค์ประกอบด้านความสามารถในการใช้ภาษา (LANG) มีค่าน้ำหนักองค์ประกอบน้อยที่สุด คือ 0.58 ค่าสัมประสิทธิ์การพยากรณ์ (R-square) ขององค์ประกอบด้านความเข้าใจในการอ่านระดับข้อเท็จจริง (FACT) มีค่าสูงสุดเท่ากับ 0.68 และค่าสัมประสิทธิ์ของการพยากรณ์ขององค์ประกอบด้านความสามารถในการใช้ภาษา (LANG) มีค่าต่ำสุดคือเท่ากับ 0.32

จากผลการวิเคราะห์องค์ประกอบเชิงยืนยันสามารถสรุปได้ว่า เมื่อจำแนกกลุ่มผู้สอบตามเพศแบบสอบวิชาภาษาอังกฤษหลังการตัดข้อสอบที่พบ DIF ออกจากแบบสอบทุกข้อแล้ว มีความตรงเชิงโครงสร้างตามโมเดลความรู้วิชาภาษาอังกฤษ

ข) หลังจากการวิเคราะห์ DTF

1. โมเดลด้านความเข้าใจในการอ่าน ระดับข้อเท็จจริง หลังการวิเคราะห์ DTF

เมื่อแบ่งกลุ่มผู้สอบตามเพศ

ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับข้อเท็จจริง พบว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ดีมาก พิจารณาได้จากค่าไค-สแควร์ (Chi-Square = 154.865) ซึ่งมีค่าความน่าจะเป็นเข้าใกล้ 1 (P = 0.967) ที่องศาอิสระเท่ากับ 189 (df = 189) นั่นคือค่าไค-สแควร์ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่ายอมรับสมมติฐานหลักที่ว่าโมเดลการวิจัยสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ส่วนค่าดัชนีวัดความกลมกลืน (GFI) มีค่าเท่ากับ 0.984 ค่าดัชนีวัดความกลมกลืนที่ปรับแก้แล้ว (AGFI) มีค่าเท่ากับ 0.981 และค่าดัชนีวัดความสอดคล้อง (NFI) มีค่าเท่ากับ 0.859 รวมทั้งค่าดัชนีรากกำลังสองเฉลี่ยของเศษ (RMR) มีค่าเท่ากับ 0.005 ซึ่งมีค่าน้อยมากจนเข้าใกล้ศูนย์ ดังรายละเอียดในตารางที่ 38 และรายละเอียดของโมเดลด้านความเข้าใจในการอ่านระดับข้อเท็จจริง สามารถแสดงได้ดังแผนภาพที่ 20

เมื่อพิจาราณาผลการวิเคราะห์โมเดลด้านความเข้าใจในการอ่านระดับข้อเท็จจริง จากตารางที่ 38 และจากแผนภาพที่ 20 พบว่าน้ำหนักองค์ประกอบส่วนใหญ่ที่กำหนดข้อมูลจำเพาะของโมเดล CFA ไว้แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ยกเว้นตัวแปร X72 ที่น้ำหนักองค์ประกอบไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ โดยน้ำหนักองค์ประกอบที่มีนัยสำคัญมีค่าอยู่ระหว่าง 0.07-0.24 และค่าสัมประสิทธิ์การพยากรณ์ (R-square) ของตัวแปร X100 มีค่าสูงสุดเท่ากับ 0.27

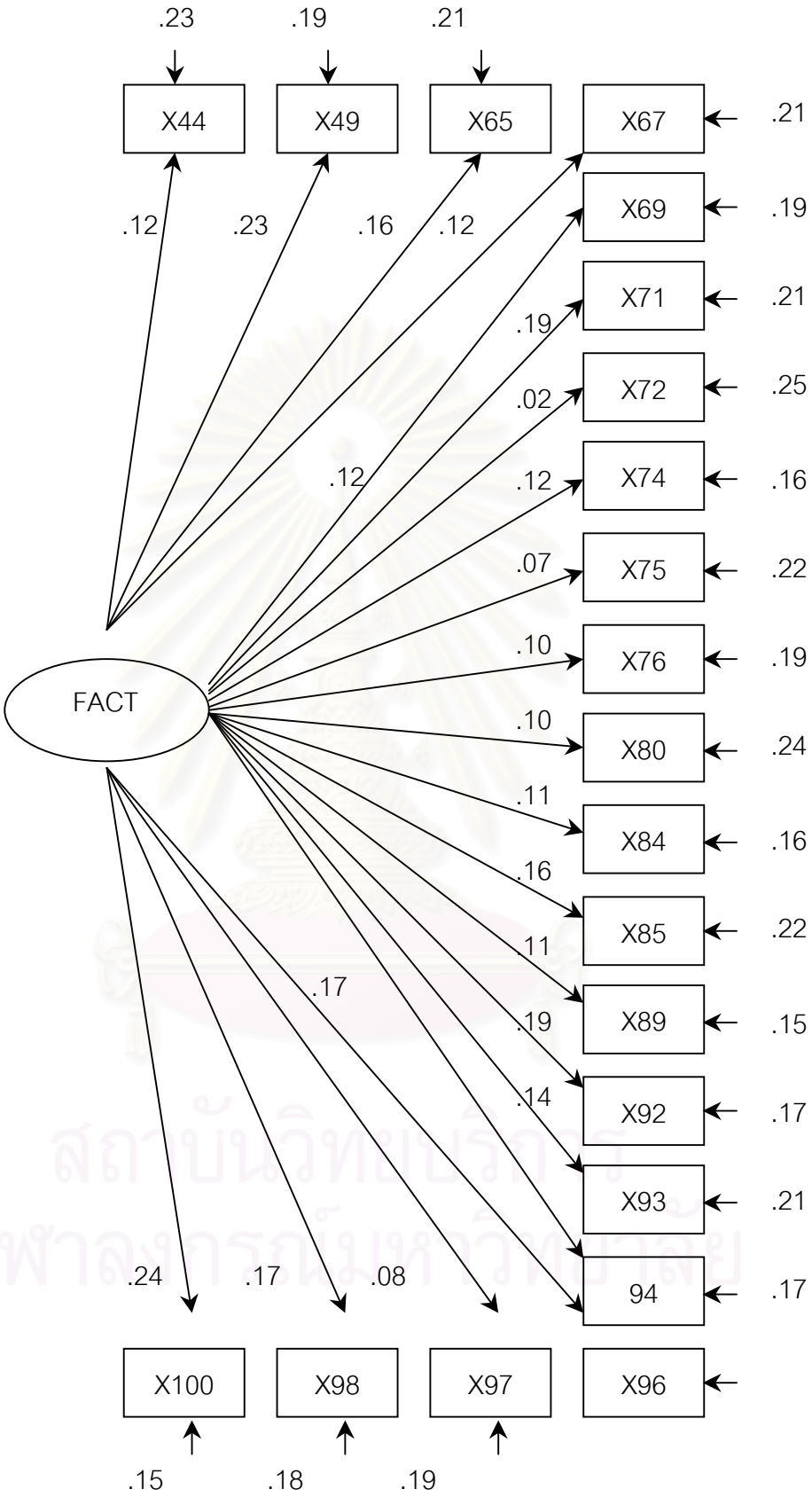
สำหรับการสร้างสเกลองค์ประกอบ ผู้วิจัยได้นำค่าสัมประสิทธ์คะแนนองค์ประกอบที่ได้จากการวิเคราะห์ใช้ในการสร้างสเกลองค์ประกอบด้านความเข้าใจในการอ่านระดับข้อเท็จจริง (FACT) เพื่อให้ได้ตัวแปรใหม่สำหรับนำไปวิเคราะห์เพื่อตรวจสอบความตรงเชิงโครงสร้างของแบบสอบวิชาภาษาอังกฤษต่อไป ซึ่งได้สเกลองค์ประกอบของความเข้าใจในการอ่านระดับข้อเท็จจริง ดังสมการต่อไปนี้

$$FACT = 0.15Z_{X44} + 0.36Z_{X49} + 0.23Z_{X65} + 0.17Z_{X67} + 0.19Z_{X69} + 0.27Z_{X71} + 0.03Z_{X72} + 0.22Z_{X74} + 0.10Z_{X75} + 0.16Z_{X76} + 0.12Z_{X80} + 0.20Z_{X84} + 0.22Z_{X85} + 0.22Z_{X89} + 0.33Z_{X92} + 0.20Z_{X93} + 0.21Z_{X94} + 0.24Z_{X96} + 0.13Z_{X97} + 0.28Z_{X98} + 0.48Z_{X100}$$

ตารางที่ 38 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่านระดับข้อเท็จจริง (FACT) ของแบบสอบวิชาภาษาอังกฤษ หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

ตัวแปร	เมตริกซ์น้ำหนักองค์ประกอบ			R Square	สปส. คะแนนองค์ประกอบ
	สปส.	SE.	t		
X44	0.12	0.02	6.46**	0.06	0.15
X49	0.23	0.02	12.75**	0.21	0.36
X65	0.16	0.02	8.81**	0.11	0.23
X67	0.12	0.02	6.68**	0.06	0.17
X69	0.12	0.02	7.11**	0.07	0.19
X71	0.19	0.02	10.52**	0.15	0.27
X72	0.02	0.02	1.10	0.00	0.03
X74	0.12	0.02	7.64**	0.08	0.22
X75	0.07	0.02	4.10**	0.02	0.10
X76	0.10	0.02	6.02**	0.05	0.16
X80	0.10	0.02	5.22**	0.04	0.12
X84	0.11	0.02	6.88**	0.07	0.20
X85	0.16	0.02	8.59**	0.10	0.20
X89	0.11	0.02	7.38**	0.08	0.22
X92	0.19	0.02	11.40**	0.17	0.33
X93	0.14	0.02	7.73**	0.08	0.20
X94	0.12	0.02	7.44**	0.08	0.21
X96	0.17	0.02	9.24**	0.12	0.24
X97	0.08	0.02	4.87**	0.03	0.13
X98	0.17	0.02	9.96**	0.13	0.28
X100	0.24	0.02	14.73**	0.28	0.48
Chi-Square = 154.865		df = 189	P = 0.967	GFI = 0.984	
AGFI = 0.981		NFI = 0.859	RMR = 0.005		

** $p < 0.01$



แผนภาพที่ 20 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่านระดับข้อเท็จจริง หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

2. โมเดลด้านความเข้าใจในการอ่าน ระดับตีความ หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับตีความ พบว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ดีมาก พิจารณาได้จากค่าไค-สแควร์ (Chi-Square = 70.563) ซึ่งมีค่าความน่าจะเป็นเข้าใกล้ 1 (P = 0.995) ที่องศาอิสระเท่ากับ 104 (df = 104) นั่นคือค่าไค-สแควร์ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่ายอมรับสมมติฐานหลักที่ว่าโมเดลการวิจัยสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ส่วนค่าดัชนีวัดความกลมกลืน (GFI) มีค่าเท่ากับ 0.991 ค่าดัชนีวัดความกลมกลืนที่ปรับแก้แล้ว (AGFI) มีค่าเท่ากับ 0.988 และค่าดัชนีวัดความสอดคล้อง (NFI) มีค่าเท่ากับ 0.910 รวมทั้งค่าดัชนีรากกำลังสองเฉลี่ยของเศษ (RMR) มีค่าเท่ากับ 0.005 ซึ่งมีค่าน้อยมากจนเข้าใกล้ศูนย์ ดังรายละเอียดในตารางที่ 39 และรายละเอียดของโมเดลด้านความเข้าใจในการอ่านระดับตีความ สามารถแสดงได้ดังแผนภาพที่ 21

เมื่อพิจาราณาผลการวิเคราะห์โมเดลด้านความเข้าใจในการอ่านระดับตีความ จากตารางที่ 39 และจากแผนภาพที่ 21 พบว่าน้ำหนักองค์ประกอบส่วนใหญ่ที่กำหนดข้อมูลจำเพาะของโมเดล CFA ไว้แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ยกเว้นตัวแปร X22 ที่น้ำหนักองค์ประกอบไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ โดยน้ำหนักองค์ประกอบที่มีนัยสำคัญมีค่าอยู่ระหว่าง 0.06-0.31 และค่าสัมประสิทธิ์การพยากรณ์ (R-square) ของตัวแปร X70 มีค่าสูงสุดเท่ากับ 0.40

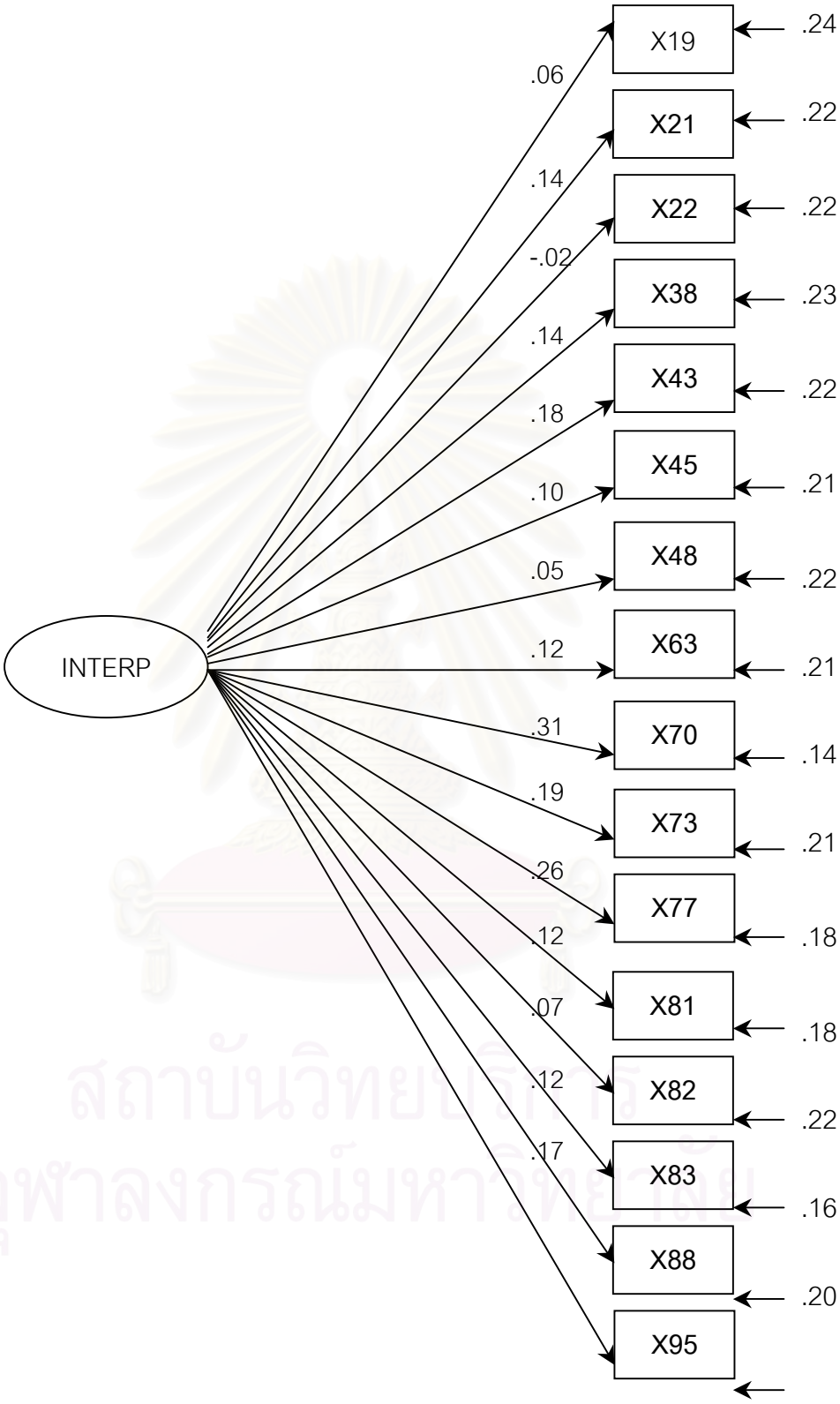
สำหรับการสร้างสเกลองค์ประกอบ ผู้วิจัยได้นำค่าสัมประสิทธ์คะแนนองค์ประกอบที่ได้จากการวิเคราะห์ใช้ในการสร้างสเกลองค์ประกอบด้านความเข้าใจในการอ่านระดับตีความ (INTERP) เพื่อให้ได้ตัวแปรใหม่สำหรับนำไปวิเคราะห์เพื่อตรวจสอบความตรงเชิงโครงสร้างของแบบสอบวิชาภาษาอังกฤษต่อไป ซึ่งได้สเกลองค์ประกอบของความเข้าใจในการอ่านระดับตีความ ดังสมการต่อไปนี้

$$INTERP = 0.08Z_{X19} + 0.20Z_{X21} - 0.03Z_{X22} + 0.20Z_{X38} + 0.26Z_{X43} + 0.15Z_{X45} + 0.08Z_{X48} + 0.18Z_{X63} + 0.70Z_{X70} + 0.29Z_{X73} + 0.46Z_{X77} + 0.22Z_{X81} + 0.11Z_{X82} + 0.23Z_{X83} + 0.26Z_{X88} + 0.27Z_{X95}$$

ตารางที่ 39 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับตีความ (INTERP) ของแบบสอบวิชาภาษาอังกฤษ หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

ตัวแปร	เมตริกซ์น้ำหนักองค์ประกอบ			R	สปส. คะแนนองค์ประกอบ
	สปส.	SE.	t	**Square**	
X19	0.06	0.02	3.30**	0.02	0.08
X21	0.14	0.02	7.54**	0.08	0.20
X22	-0.02	0.02	-0.96	0.00	-0.03
X38	0.14	0.02	7.65**	0.08	0.20
X43	0.18	0.02	9.34**	0.12	0.26
X45	0.10	0.02	5.64**	0.05	0.15
X48	0.05	0.02	2.99**	0.01	0.08
X63	0.12	0.02	6.45**	0.06	0.18
X70	0.31	0.02	17.40**	0.40	0.70
X73	0.19	0.02	10.27**	0.15	0.29
X77	0.26	0.02	14.15**	0.27	0.46
X81	0.12	0.02	7.41**	0.08	0.22
X82	0.07	0.02	4.00**	0.02	0.11
X83	0.12	0.02	7.24**	0.08	0.23
X88	0.17	0.02	9.26**	0.12	0.26
X95	0.13	0.02	8.42**	0.10	0.27
Chi-Square = 70.563		df = 90	P = 0.995	GFI = 0.991	
AGFI = 0.988		NFI = 0.910	RMR = 0.005		

** $p < 0.01$



แผนภาพที่ 21 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่านระดับตีความ หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

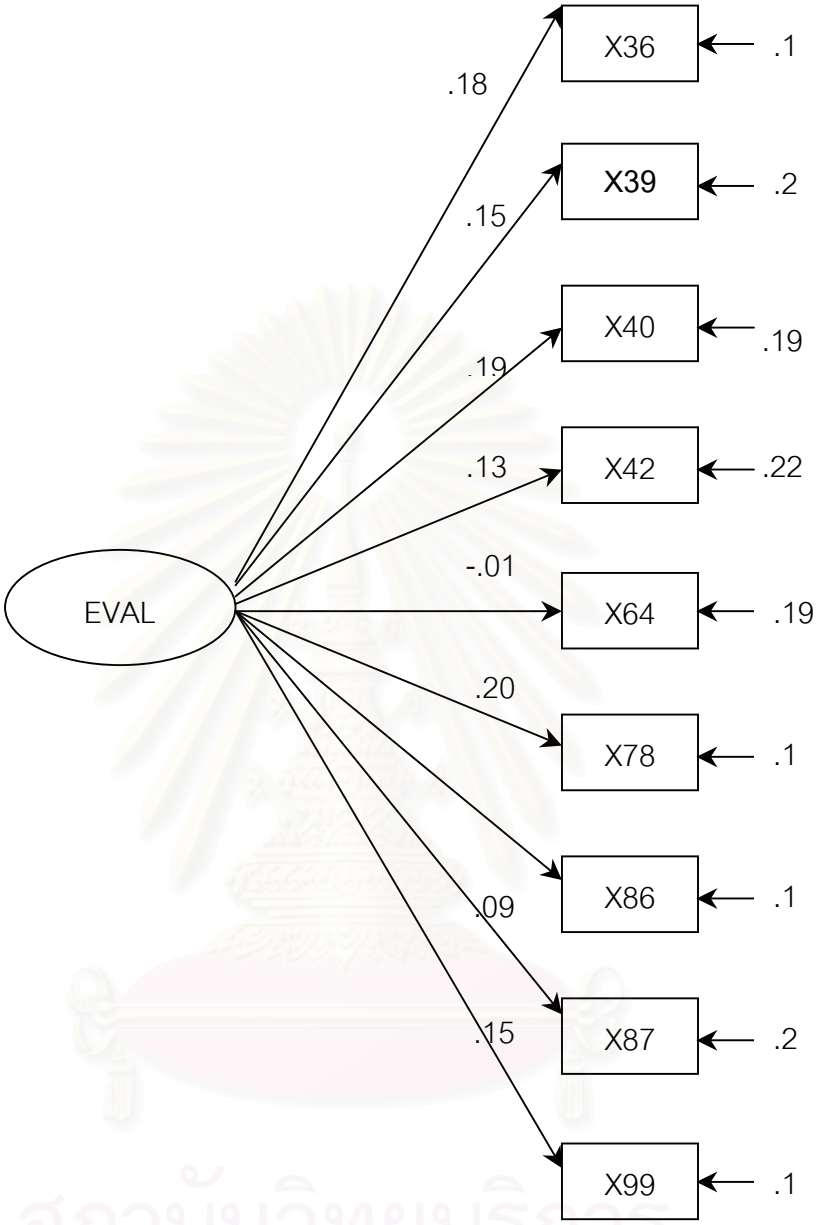
2. โมเดลด้านความเข้าใจในการอ่าน ระดับประเมินค่า หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับประเมินค่า พบว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ดีมาก พิจารณาได้จากค่าไค-สแควร์ (Chi-Square = 26.062) ซึ่งมีค่าความน่าจะเป็นเท่ากับ 0.515 (P = 0.515) ที่องศาอิสระเท่ากับ 27 (df = 27) นั่นคือค่าไค-สแควร์ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่ายอมรับสมมติฐานหลักที่ว่าโมเดลการวิจัยสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ส่วนค่าดัชนีวัดความกลมกลืน (GFI) มีค่าเท่ากับ 0.994 ค่าดัชนีวัดความกลมกลืนที่ปรับแก้แล้ว (AGFI) มีค่าเท่ากับ 0.990 และค่าดัชนีวัดความสอดคล้อง (NFI) มีค่าเท่ากับ 0.902 รวมทั้งค่าดัชนีรากกำลังสองเฉลี่ยของเศษ (RMR) มีค่าเท่ากับ 0.005 ซึ่งมีค่าน้อยมากจนเข้าใกล้ศูนย์ ดังรายละเอียดในตารางที่ 40 และรายละเอียดของโมเดลด้านความเข้าใจในการอ่านระดับประเมินค่า สามารถแสดงได้ดังแผนภาพที่ 22

ตารางที่ 40 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับประเมินค่า (EVAL) ของแบบสอบวิชาภาษาอังกฤษ หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

ตัวแปร	เมตริกซ์น้ำหนักองค์ประกอบ			R Square	สปส. คะแนนองค์ประกอบ
	สปส.	SE.	t		
X36	0.18	0.02	8.59**	0.15	0.49
X39	0.15	0.02	6.89**	0.10	0.36
X40	0.19	0.02	8.74**	0.16	0.49
X42	0.13	0.02	5.94**	0.07	0.29
X64	-0.01	0.02	-0.70	0.00	-0.04
X78	0.20	0.02	9.51**	0.20	0.61
X86	0.08	0.02	4.70**	0.04	0.28
X87	0.09	0.02	4.28**	0.04	0.21
X99	0.15	0.02	7.54**	0.12	0.45
Chi-Square = 26.062		df = 27	P = 0.515	GFI = 0.994	
AGFI = 0.990		NFI = 0.902	RMR = 0.005		

** $p < 0.01$



แผนภาพที่ 22 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่านระดับประเมินค่า หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

เมื่อพิจารณาผลการวิเคราะห์โมเดลด้านความเข้าใจในการอ่านระดับประเมินค่า จากตารางที่ 40 และจากแผนภาพที่ 22 พบว่าน้ำหนักองค์ประกอบส่วนใหญ่ที่กำหนดข้อมูลจำเพาะของโมเดล CFA ไว้แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ยกเว้นตัวแปร X64 ที่น้ำหนักองค์ประกอบไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ โดยน้ำหนักองค์ประกอบที่มีนัยสำคัญมีค่าอยู่ระหว่าง 0.08-0.20 และค่าสัมประสิทธิ์การพยากรณ์ (R-square) ของตัวแปร X78 มีค่าสูงสุดเท่ากับ 0.20

สำหรับการสร้างสเกลองค์ประกอบ ผู้วิจัยได้นำค่าสัมประสิทธ์คะแนนองค์ประกอบที่ได้จากการวิเคราะห์ใช้ในการสร้างสเกลองค์ประกอบด้านความเข้าใจในการอ่านระดับประเมินค่า (EVAL) เพื่อให้ได้ตัวแปรใหม่สำหรับนำไปวิเคราะห์เพื่อตรวจสอบความตรงเชิงโครงสร้างของแบบสอบวิชาภาษาอังกฤษต่อไป ซึ่งได้สเกลองค์ประกอบของความเข้าใจในการอ่านระดับประเมินค่า ดังสมการต่อไปนี้

$$EVAL = 0.49Z_{X36} + 0.36Z_{X39} + 0.49Z_{X40} + 0.29Z_{X42} - 0.04Z_{X64} + 0.61Z_{X78} + 0.28Z_{X86} + 0.21Z_{X87} + 0.45Z_{X99}$$

4. โมเดลด้านความสามารถในการใช้ภาษา หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความสามารถในการใช้ภาษา พบว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ พิจารณาได้จากค่าไค-สแควร์ (Chi-Square = 91.634) ซึ่งมีค่าความน่าจะเป็นเท่ากับ 0.093 ที่องศาอิสระเท่ากับ 75 (df = 75) นั่นคือค่าไค-สแควร์ไม่ แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่ายอมรับสมมติฐานหลักที่ว่าโมเดลการวิจัยสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ส่วนค่าดัชนีวัดความกลมกลืน (GFI) มีค่าเท่ากับ 0.986 ค่าดัชนีวัดความกลมกลืนที่ปรับแก้แล้ว (AGFI) มีค่าเท่ากับ 0.980 และค่าดัชนีวัดความสอดคล้อง (NFI) มีค่าเท่ากับ 0.958 รวมทั้งค่าดัชนีรากกำลังสองเฉลี่ยของเศษ (RMR) มีค่าเท่ากับ 0.005 ซึ่งมีค่าน้อยมากจนเข้าใกล้ศูนย์ ดังรายละเอียดในตารางที่ 41 และรายละเอียดของโมเดลด้านความสามารถในการใช้ภาษาแสดงได้ดังแผนภาพที่ 23

เมื่อพิจารณาผลการวิเคราะห์โมเดลด้านความสามารถในการใช้ภาษา จากตารางที่ 41 และจากแผนภาพที่ 23 พบว่าน้ำหนักองค์ประกอบทุกค่าที่กำหนดข้อมูลจำเพาะของโมเดล CFA ไว้แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 โดยน้ำหนักองค์ประกอบมีค่าอยู่ระหว่าง –0.04-0.38 ค่าสัมประสิทธิ์การพยากรณ์ (R-square) ของตัวแปร X3 มีค่าสูงสุดเท่ากับ 0.59

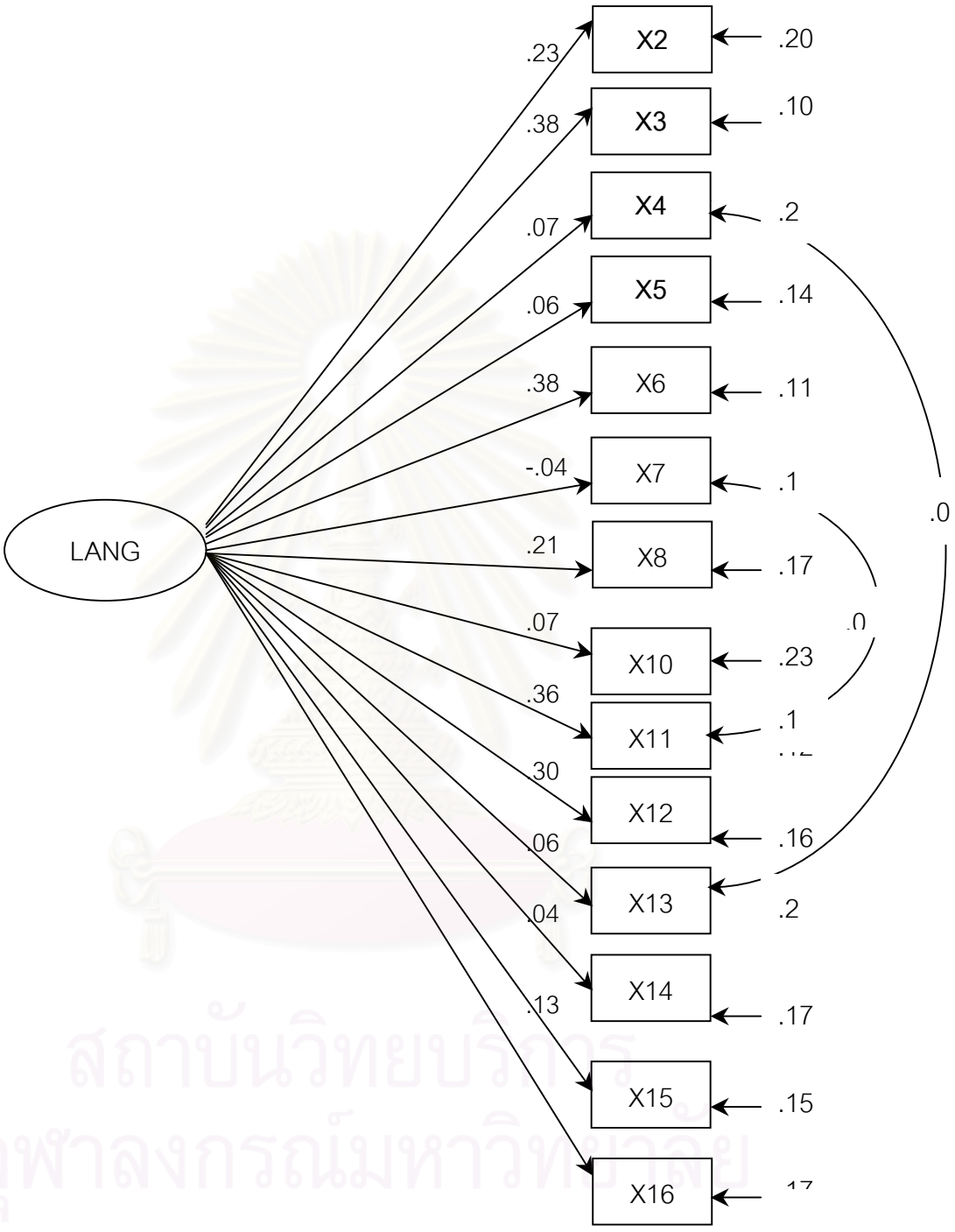
ตารางที่ 41 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความสามารถในการใช้ภาษา (LANG) ของแบบสอบวิชาภาษาอังกฤษ หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

ตัวแปร	เมตริกซ์น้ำหนักองค์ประกอบ			R Square	สปส. คะแนนองค์ประกอบ
	สปส.	SE.	t		
X2	0.23	0.02	13.95**	0.21	0.17
X3	0.38	0.01	26.58**	0.59	0.56
X4	0.07	0.02	4.37**	0.02	0.05
X5	0.06	0.01	4.88**	0.03	0.07
X6	0.38	0.01	26.37**	0.58	0.55
X7	-0.04	0.01	-2.60**	0.01	-0.12
X8	0.21	0.01	14.38**	0.22	0.20
X10	0.07	0.02	4.11**	0.02	0.05
X11	0.36	0.01	23.93**	0.50	0.46
X12	0.30	0.02	19.12**	0.35	0.28
X13	0.06	0.02	3.58**	0.02	0.04
X14	0.04	0.01	2.55**	0.01	0.03
X15	0.13	0.01	9.24**	0.10	0.13
X16	0.26	0.02	16.73**	0.28	0.23
Chi-Square = 91.634		df = 75	P = 0.093	GFI = 0.986	
AGFI = 0.980		NFI = 0.958	RMR = 0.005		

** $p < 0.01$

สำหรับการสร้างสเกลองค์ประกอบ ผู้วิจัยได้นำค่าสัมประสิทธ์คะแนนองค์ประกอบที่ได้จากการวิเคราะห์ไปใช้ในการสร้างสเกลองค์ประกอบด้านความสามารถในการใช้ภาษา (LANG) เพื่อให้ได้ตัวแปรใหม่สำหรับนำไปวิเคราะห์เพื่อตรวจสอบความตรงเชิงโครงสร้างของแบบสอบวิชาภาษาอังกฤษต่อไป ซึ่งได้สเกลองค์ประกอบของความสามารถในการใช้ภาษา ดังสมการต่อไปนี้

$$LANG = 0.17Z_{X2} + 0.56Z_{X3} + 0.05Z_{X4} + 0.07Z_{X5} + 0.55Z_{X6} - 0.12Z_{X7} + 0.20Z_{X8} + 0.05Z_{X10} + 0.46Z_{X11} + 0.28Z_{X12} + 0.04Z_{X13} + 0.03Z_{X14} + 0.13Z_{X15} + 0.23Z_{X16}$$



แผนภาพที่ 23 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความสามารถในการใช้ภาษาหลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

5. โมเดลด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ พบว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ดี พิจารณาได้จากค่าไค-สแควร์ (Chi-Square = 350.164) ซึ่งมีค่าความน่าจะเป็นเท่ากับ 0.143 ที่องศาอิสระเท่ากับ 324 (df = 324) นั่นคือค่าไค-สแควร์ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่ายอมรับสมมติฐานหลักที่ว่าโมเดลการวิจัยสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ส่วนค่าดัชนีวัดความกลมกลืน (GFI) มีค่าเท่ากับ 0.971 ค่าดัชนีวัดความกลมกลืนที่ปรับแก้แล้ว (AGFI) มีค่าเท่ากับ 0.966 และค่าดัชนีวัดความสอดคล้อง (NFI) มีค่าเท่ากับ 0.809 รวมทั้งค่าดัชนีรากกำลังสองเฉลี่ยของเศษ (RMR) มีค่าเท่ากับ 0.006 ซึ่งมีค่า น้อยมากจนเข้าใกล้ศูนย์ ดังรายละเอียดในตารางที่ 42 และรายละเอียดของโมเดลด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ สามารถแสดงได้ดังแผนภาพที่ 24

เมื่อพิจารณาผลการวิเคราะห์โมเดลด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ จากตารางที่ 42 และจากแผนภาพที่ 24 พบว่าน้ำหนักองค์ประกอบส่วนใหญ่ที่กำหนดข้อมูลจำเพาะของโมเดล CFA ไว้แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 และ 0.05 ยกเว้นตัวแปร X54 และ X61 ที่น้ำหนักองค์ประกอบไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ โดยน้ำหนักองค์ประกอบที่มีนัยสำคัญมีค่าอยู่ระหว่าง 0.04-0.33 และค่าสัมประสิทธิ์การพยากรณ์ (R-square) ของตัวแปร X28 มีค่าสูงสุดเท่ากับ 0.43

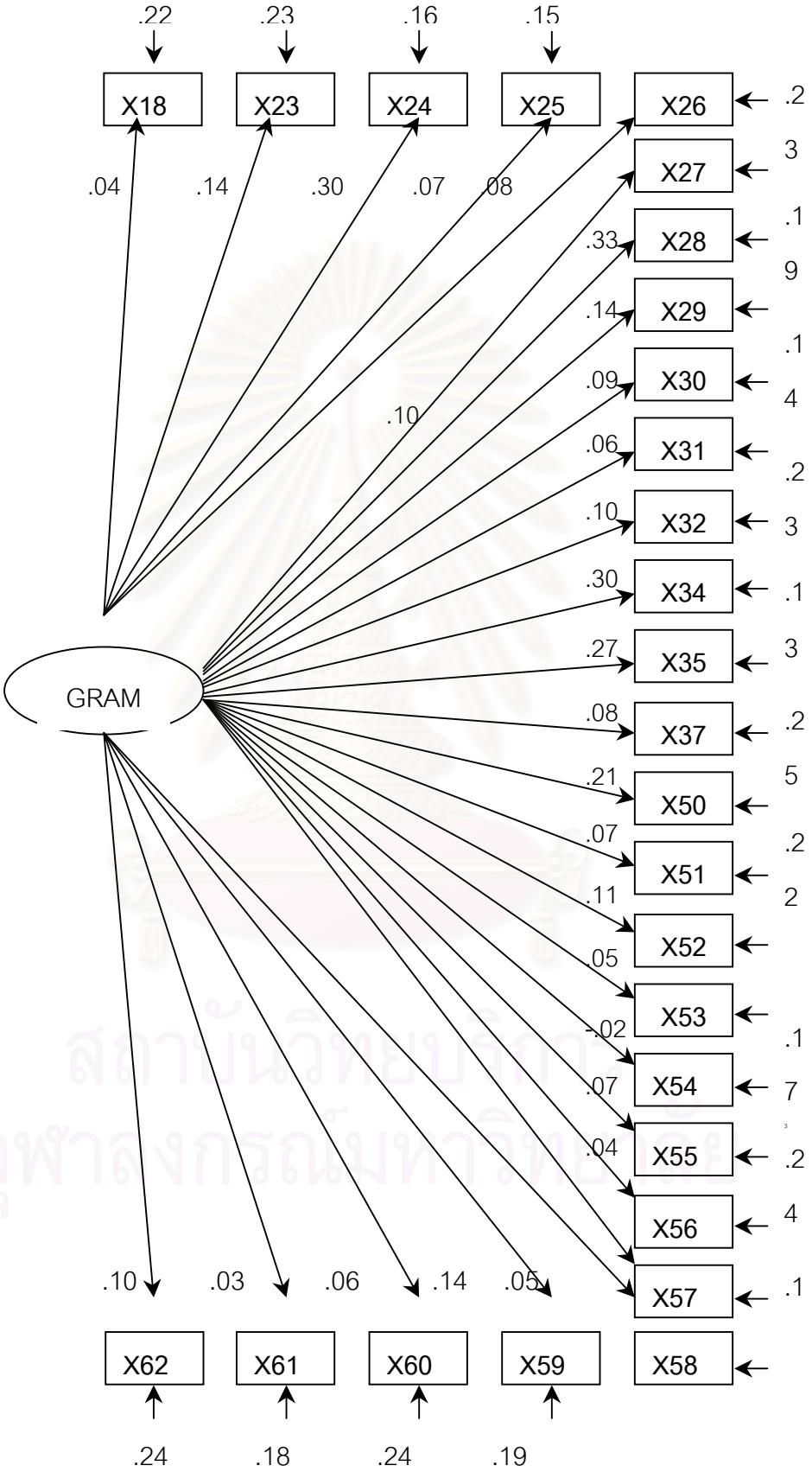
สำหรับการสร้างสเกลองค์ประกอบ ผู้วิจัยได้นำค่าสัมประสิทธ์คะแนนองค์ประกอบที่ได้จากการวิเคราะห์ใช้ในการสร้างสเกลองค์ประกอบด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ (GRAM) เพื่อให้ได้ตัวแปรใหม่สำหรับนำไปวิเคราะห์เพื่อตรวจสอบความตรงเชิงโครงสร้างของแบบสอบวิชาภาษาอังกฤษต่อไป ซึ่งได้สเกลองค์ประกอบของความรู้เกี่ยวกับโครงสร้างไวยากรณ์ ดังสมการต่อไปนี้

$$\begin{aligned} GRAM = {} & 0.04Z_{X18} + 0.14Z_{X23} + 0.38Z_{X24} + 0.12Z_{X25} + 0.08Z_{X26} + 0.12Z_{X27} + \\ & 0.53Z_{X28} + 0.14Z_{X29} + 0.09Z_{X30} + 0.06Z_{X31} + 0.10Z_{X32} + 0.36Z_{X34} + \\ & 0.37Z_{X35} + 0.08Z_{X27} + 0.26Z_{X50} + 0.10Z_{X51} + 0.14Z_{X52} + 0.05Z_{X53} - \\ & 0.02Z_{X54} + 0.09Z_{X55} + 0.04Z_{X56} + 0.26Z_{X57} + 0.06Z_{X58} + 0.16Z_{X59} + \\ & 0.06Z_{X60} + 0.03Z_{X61} + 0.10Z_{X62} \end{aligned}$$

ตารางที่ 42 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ (GRAM) ของแบบสอบวิชาภาษาอังกฤษหลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

ตัวแปร	เมตริกซ์น้ำหนักองค์ประกอบ			R Square	สปส. คะแนนองค์ประกอบ
	สปส.	SE.	t		
X18	0.04	0.02	2.33*	0.01	0.04
X23	0.14	0.02	7.74**	0.08	0.14
X24	0.30	0.02	18.02**	0.37	0.38
X25	0.07	0.02	5.29**	0.04	0.12
X26	0.08	0.02	4.63**	0.03	0.08
X27	0.10	0.02	6.34**	0.05	0.12
X28	0.33	0.02	20.07**	0.43	0.53
X29	0.14	0.02	7.65**	0.07	0.14
X30	0.09	0.02	4.89**	0.03	0.09
X31	0.06	0.02	3.53**	0.02	0.06
X32	0.10	0.02	5.56**	0.04	0.10
X34	0.30	0.02	17.59**	0.35	0.36
X35	0.27	0.02	16.34**	0.30	0.37
X37	0.08	0.02	4.54**	0.03	0.08
X50	0.21	0.02	12.70**	0.19	0.26
X51	0.07	0.02	4.77**	0.03	0.10
X52	0.11	0.02	7.00**	0.06	0.14
X53	0.05	0.02	2.98**	0.01	0.05
X54	-0.02	0.02	-1.24	0.00	-0.02
X55	0.07	0.02	4.68**	0.03	0.09
X56	0.04	0.02	2.18*	0.01	0.04
X57	0.19	0.02	12.15**	0.18	0.26
X58	0.05	0.02	3.04**	0.01	0.06
X59	0.14	0.02	8.31**	0.09	0.16
X60	0.06	0.02	3.35**	0.02	0.06
X61	0.03	0.02	1.70	0.00	0.03
X62	0.10	0.02	5.50**	0.04	0.10
Chi-Square = 350.164		df = 324	P = 0.143	GFI = 0.971	
AGFI = 0.966		NFI = 0.809	RMR = 0.006		

*p < 0.05 , ** p < 0.01


.22
.23
.16
.15
X18
X23
X24
X25
X26
X27
X28
X29
X30
X31
X32
X34
X35
X37
X50
X51
X52
X53
X54
X55
X56
X57
X58
X62
X61
X60
X59
GRAM
.04
.14
.30
.07
.08
.33
.14
.09
.10
.06
.10
.30
.27
.08
.21
.07
.11
.05
-.02
.07
.04
.10
.03
.06
.14
.05
.24
.18
.24
.19
.2
3
.1
9
.1
4
.2
3
.1
3
.2
5
.2
2
.1
7
.2
4
.1

แผนภาพที่ 24 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลความรู้เกี่ยวกับโครงสร้างไวยากรณ์หลังจากการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

โมเดลองค์ประกอบความรู้วิชาภาษาอังกฤษ หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

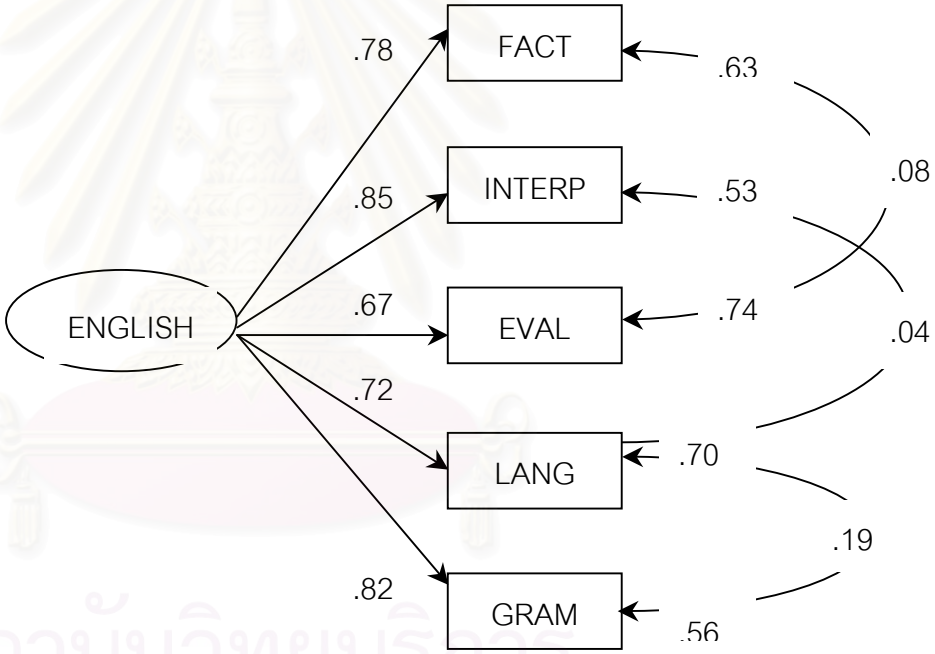
ผลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับที่สอง (second order confirmatory factor analysis) ของโมเดลองค์ประกอบความรู้วิชาภาษาอังกฤษ (ENGLISH) พบว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ดีมาก โดยพิจารณาจากค่าไค-สแควร์ (Chi-Square = 1.086) ซึ่งมีค่าความน่าจะเป็นเท่ากับ 0.581 ที่องศาอิสระเท่ากับ 2 (df = 2) นั่นคือค่าไค-สแควร์ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่ายอมรับสมมติฐานหลักที่ว่าโมเดลการวิจัยสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ส่วนค่าดัชนีวัดความกลมกลืน (GFI) มีค่าเท่ากับ 1.00 ค่าดัชนีวัดความกลมกลืนที่ปรับแก้แล้ว (AGFI) มีค่าเท่ากับ 0.997 และค่าดัชนีวัดความสอดคล้อง (NFI) มีค่าเท่ากับ 1.00 รวมทั้งค่าดัชนีรากกำลังสองเฉลี่ยของเศษ (RMR) มีค่าเท่ากับ 0.009 ซึ่งมีค่าน้อยมากจนเข้าใกล้ศูนย์ ดังรายละเอียดใน ตารางที่ 43 และรายละเอียดของโมเดลด้านความเข้าใจในการอ่านระดับประเมินค่า สามารถแสดงได้ดังแผนภาพที่ 25

ตารางที่ 43 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลองค์ประกอบความรู้วิชาภาษาอังกฤษ (ENGLISH) หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

องค์ประกอบ	เมตริกซ์น้ำหนักองค์ประกอบ			R Square	สปส. คะแนนองค์ประกอบ
	สปส.	SE.	t		
ความเข้าใจในการอ่านระดับข้อเท็จจริง (FACT)	0.78	0.05	27.25**	0.60	0.13
ความเข้าใจในการอ่านระดับตีความ (INTERP)	0.85	0.05	31.06**	0.72	0.23
ความเข้าใจในการอ่านระดับประเมินค่า (EVAL)	0.67	0.04	22.16**	0.45	0.08
ความสามารถในการใช้ภาษา (LANG)	0.72	0.06	23.03**	0.51	0.02
ความรู้เกี่ยวกับโครงสร้างไวยากรณ์ (GRAM)	0.82	0.05	29.76**	0.68	0.17
Chi-Square = 1.086	df = 2	P = 0.581		GFI = 1.00	
AGFI = 0.997	NFI = 1.00	RMR = 0.009			

** $p < 0.01$

เมื่อพิจารณาผลการวิเคราะห์โมเดลความรู้วิชาภาษาอังกฤษ จากตารางที่ 43 และจากแผนภาพที่ 25 พบว่าน้ำหนักองค์ประกอบทุกค่าที่กำหนดข้อมูลจำเพาะของโมเดล CFA ไว้แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 โดยองค์ประกอบด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ (GRAM) มีน้ำหนักองค์ประกอบมากที่สุด คือ 1.49 และองค์ประกอบด้านความเข้าใจในการอ่านระดับประเมินค่า (EVAL) มีค่าน้ำหนักองค์ประกอบน้อยที่สุด คือ 1.03 ค่าสัมประสิทธิ์การพยากรณ์ (R-square) ขององค์ประกอบด้านความเข้าใจในการอ่านระดับตีความ (INTERP) มีค่าสูงสุดเท่ากับ 0.72 และค่าสัมประสิทธิ์ของการพยากรณ์ขององค์ประกอบด้านความเข้าใจในการอ่านระดับประเมินค่า (EVAL) มีค่าต่ำสุดคือเท่ากับ 0.45



แผนภาพที่ 25 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลองค์ประกอบความรู้วิชาภาษาอังกฤษหลังจากการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามตัวแปรเพศ

จากผลการวิเคราะห์องค์ประกอบเชิงยืนยันสามารถสรุปได้ว่า เมื่อจำแนกกลุ่มผู้สอบตามเพศแบบสอบวิชาภาษาอังกฤษหลังการตัดข้อสอบที่พบ DTF ออกจากแบบสอบทุกข้อแล้ว มีความตรงเชิงโครงสร้างตามโมเดลองค์ประกอบความรู้วิชาภาษาอังกฤษ

4.2.1.2 เมื่อจำแนกกลุ่มผู้สอบตามตัวแปรสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

ข) หลังจากการวิเคราะห์ DIF

1. โมเดลด้านความเข้าใจในการอ่าน ระดับข้อเท็จจริง (FACT) หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งของโรงเรียนที่จบการศึกษา

ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับข้อเท็จจริงพบว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ดีมาก พิจารณาได้จากค่าไค-สแควร์ (Chi-Square = 183.765) ซึ่งมีค่าความน่าจะเป็นเข้าใกล้ 1 (P = 0.989) ที่องศาอิสระเท่ากับ 230 (df = 230) นั่นคือค่าไค-สแควร์ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่ายอมรับสมมติฐานหลักที่ว่าโมเดลการวิจัยสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ส่วนค่าดัชนีวัดความกลมกลืน (GFI) มีค่าเท่ากับ 0.983 ค่าดัชนีวัดความกลมกลืนที่ปรับแก้แล้ว (AGFI) มีค่าเท่ากับ 0.980 และค่าดัชนีวัดความสอดคล้อง (NFI) มีค่าเท่ากับ 0.863 รวมทั้งค่าดัชนีรากกำลังสองเฉลี่ยของเศษ (RMR) มีค่าเท่ากับ 0.005 ซึ่งมีค่าน้อยมากจนเข้าใกล้ศูนย์ ดังรายละเอียดในตารางที่ 44 และรายละเอียดของโมเดลด้านความเข้าใจในการอ่านระดับข้อเท็จจริง สามารถแสดงได้ดังแผนภาพที่ 26

เมื่อพิจารณาผลการวิเคราะห์โมเดลด้านความเข้าใจในการอ่านระดับข้อเท็จจริง จากตารางที่ 44 และจากแผนภาพที่ 26 พบว่าน้ำหนักองค์ประกอบส่วนใหญ่ที่กำหนดข้อมูลจำเพาะของโมเดล CFA ไว้แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ยกเว้นตัวแปร X72 ที่น้ำหนักองค์ประกอบไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ โดยน้ำหนักองค์ประกอบที่มีนัยสำคัญมีค่าอยู่ระหว่าง 0.08-0.24 และค่าสัมประสิทธิ์การพยากรณ์ (R-square) ของตัวแปร X100 มีค่าสูงสุดเท่ากับ 0.27

สำหรับการสร้างสเกลองค์ประกอบ ผู้วิจัยได้นำค่าสัมประสิทธิ์คะแนนองค์ประกอบที่ได้จากการวิเคราะห์ใช้ในการสร้างสเกลองค์ประกอบด้านความเข้าใจในการอ่านระดับข้อเท็จจริง (FACT) เพื่อให้ได้ตัวแปรใหม่สำหรับนำไปวิเคราะห์เพื่อตรวจสอบความตรงเชิงโครงสร้างของแบบสอบวิชาภาษาอังกฤษต่อไป ซึ่งได้สเกลองค์ประกอบของความเข้าใจในการอ่านระดับข้อเท็จจริง ดังสมการต่อไปนี้

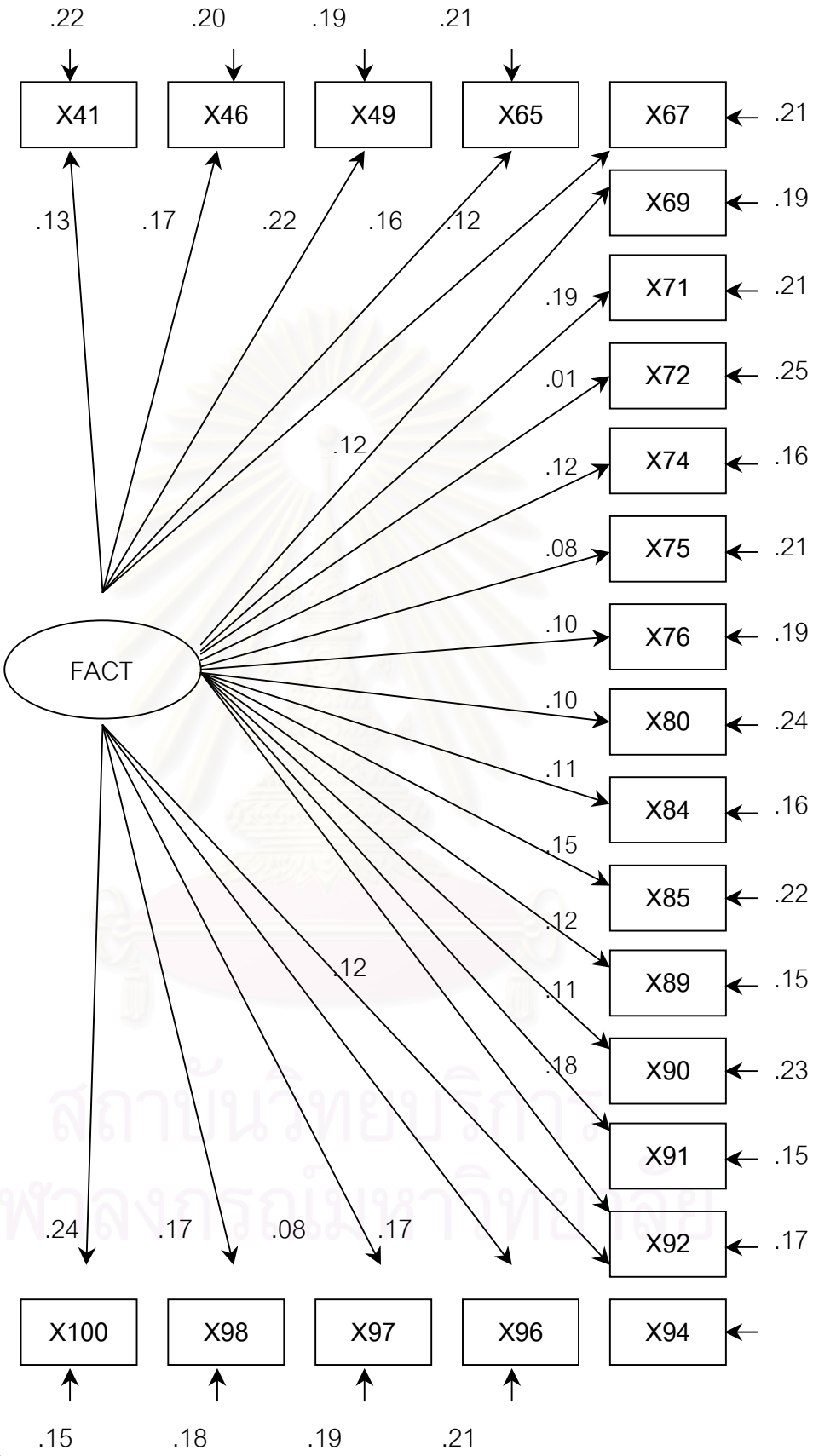
$$\begin{aligned} FACT = {} & 0.16Z_{X41} + 0.23Z_{X46} + 0.32Z_{X49} + 0.21Z_{X65} + 0.16Z_{X67} + 0.17Z_{X69} + \\ & 0.24Z_{X71} + 0.01Z_{X72} + 0.20Z_{X74} + 0.10Z_{X75} + 0.15Z_{X76} + 0.11Z_{X80} + \\ & 0.18Z_{X84} + 0.19Z_{X85} + 0.22Z_{X89} + 0.13Z_{X90} + 0.32Z_{X91} + 0.31Z_{X92} + \\ & 0.19Z_{X94} + 0.22Z_{X96} + 0.12Z_{X97} + 0.26Z_{X98} + 0.43Z_{X100} \end{aligned}$$

ตารางที่ 44 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่านระดับข้อเท็จจริง (FACT) ของแบบสอบวิชาภาษาอังกฤษ หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

ตัวแปร	เมตริกซ์น้ำหนักองค์ประกอบ			R Square	สปส. คะแนนองค์ประกอบ
	สปส.	SE.	t		
X41	0.13	0.02	7.26**	0.07	0.16
X46	0.17	0.02	9.75**	0.12	0.23
X49	0.23	0.02	12.99**	0.21	0.32
X65	0.16	0.02	9.23**	0.11	0.21
X67	0.12	0.02	6.89**	0.06	0.16
X69	0.12	0.02	7.30**	0.07	0.17
X71	0.19	0.02	10.41**	0.14	0.24
X72	0.01	0.02	0.55	0.00	0.01
X74	0.12	0.02	7.68**	0.08	0.20
X75	0.08	0.02	4.34**	0.03	0.10
X76	0.10	0.02	6.21**	0.05	0.15
X80	0.10	0.02	5.35**	0.04	0.11
X84	0.11	0.02	7.00**	0.07	0.18
X85	0.15	0.02	8.28**	0.09	0.19
X89	0.12	0.02	8.03**	0.09	0.22
X90	0.11	0.02	6.11**	0.05	0.13
X91	0.18	0.02	11.71**	0.17	0.32
X92	0.20	0.02	12.06**	0.18	0.31
X94	0.12	0.02	7.58**	0.08	0.19
X96	0.17	0.02	9.49**	0.12	0.22
X97	0.08	0.02	5.11**	0.04	0.12
X98	0.17	0.02	10.31**	0.14	0.26
X100	0.24	0.02	14.99**	0.27	0.43
Chi-Square = 183.765		df = 230	P = 0.989	GFI = 0.983	
AGFI = 0.980		NFI = 0.863	RMR = 0.005		

** $p < 0.01$

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



แผนภาพที่ 26 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับข้อเท็จจริง หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

2. โมเดลด้านความเข้าใจในการอ่าน ระดับตีความ (INTERP) หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งของโรงเรียนที่จบการศึกษา

ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับตีความพบว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ดีมาก พิจารณาได้จากค่าไค-สแควร์ (Chi-Square = 64.791) ซึ่งมีค่าความน่าจะเป็นเท่ากับ 1 (P = 1.00) ที่องศาอิสระเท่ากับ 119 (df = 119) นั่นคือค่าไค-สแควร์ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่ายอมรับสมมติฐานหลักที่ว่าโมเดลการวิจัยสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ดังรายละเอียดในตารางที่ 45 และรายละเอียดของโมเดลด้านความเข้าใจในการอ่านระดับตีความ สามารถแสดงได้ดังแผนภาพที่ 27

เมื่อพิจาราณาผลการวิเคราะห์โมเดลด้านความเข้าใจในการอ่านระดับตีความ จากตารางที่ 45 และจากแผนภาพที่ 27 พบว่าน้ำหนักองค์ประกอบส่วนใหญ่ที่กำหนดข้อมูลจำเพาะของโมเดล CFA ไว้แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ยกเว้นตัวแปร X22 ที่น้ำหนักองค์ประกอบไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ โดยน้ำหนักองค์ประกอบที่มีนัยสำคัญมีค่าอยู่ระหว่าง 0.07-0.27 และค่าสัมประสิทธิ์การพยากรณ์ (R-square) ของตัวแปร X70 มีค่าสูงสุดเท่ากับ 0.30

สำหรับการสร้างสเกลองค์ประกอบ ผู้วิจัยได้นำค่าสัมประสิทธ์คะแนนองค์ประกอบที่ได้จากการวิเคราะห์ไปใช้ในการสร้างสเกลองค์ประกอบด้านความเข้าใจในการอ่านระดับตีความ (INTERP) เพื่อให้ได้ตัวแปรใหม่สำหรับนำไปวิเคราะห์เพื่อตรวจสอบความตรงเชิงโครงสร้างของแบบสอบวิชาภาษาอังกฤษต่อไป ซึ่งได้สเกลองค์ประกอบของความเข้าใจในการอ่านระดับตีความ ดังสมการต่อไปนี้

$$\begin{aligned} INTERP = {} & 0.12Z_{X19} + 0.25Z_{X21} - 0.01Z_{X22} + 0.26Z_{X38} + 0.32Z_{X43} + 0.20Z_{X45} + \\ & 0.34Z_{X47} + 0.12Z_{X48} + 0.24Z_{X63} + 0.19Z_{X68} + 0.61Z_{X70} + 0.12Z_{X79} + \\ & 0.27Z_{X81} + 0.13Z_{X82} + 0.27Z_{X83} + 0.33Z_{X88} + 0.33Z_{X95} \end{aligned}$$

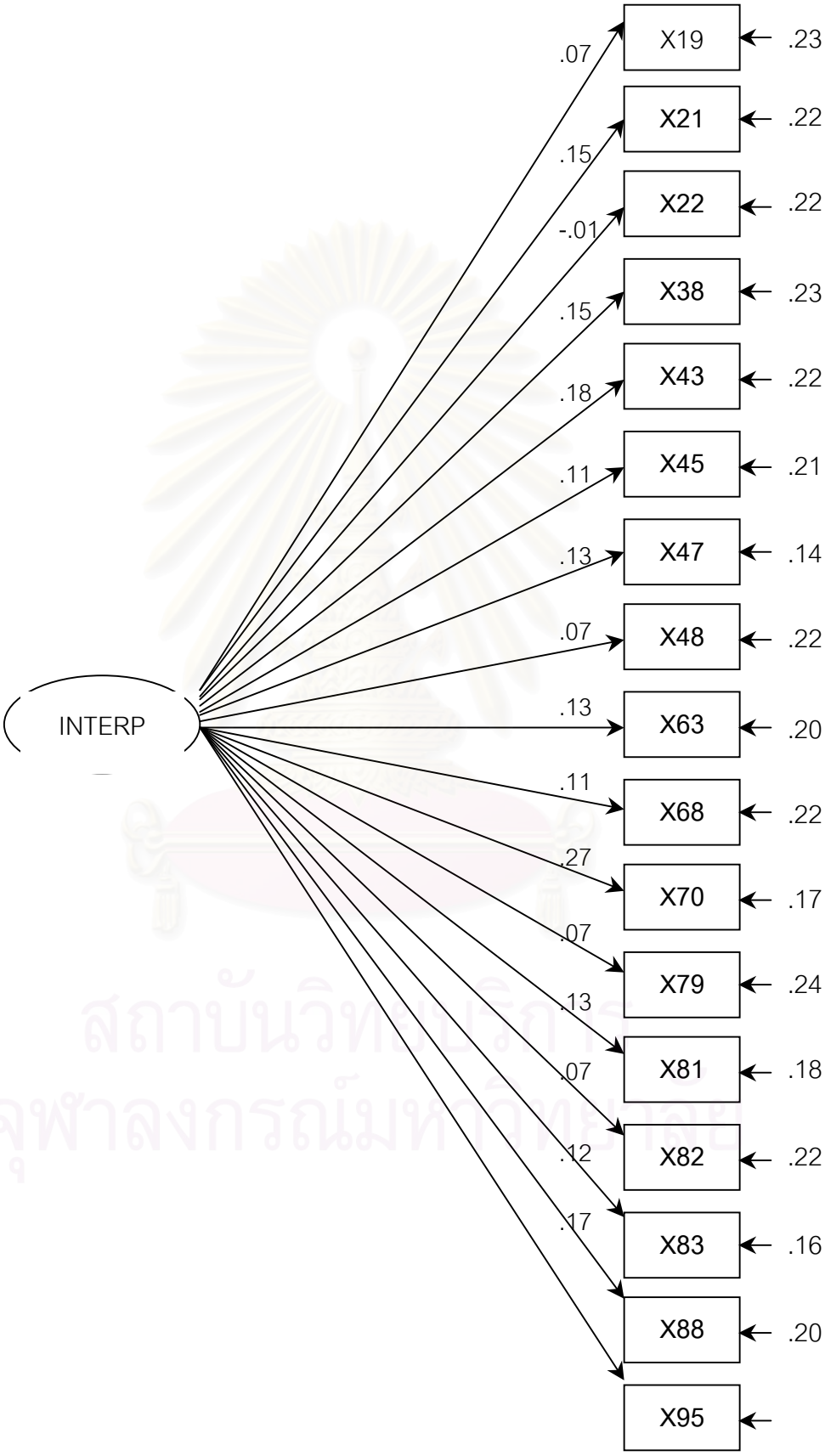
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 45 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับตีความ (INTERP) ของแบบสอบวิชาภาษาอังกฤษ หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตาม **สถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา**

ตัวแปร	เมตริกซ์น้ำหนักองค์ประกอบ			R	สปส. คะแนน
	สปส.	SE.	t	**Square**	องค์ประกอบ
X19	0.07	0.02	3.61**	0.02	0.12
X21	0.15	0.02	7.41**	0.09	0.25
X22	-0.01	0.02	-0.32	0.00	-0.01
X38	0.15	0.02	7.63**	0.09	0.26
X43	0.18	0.02	9.15**	0.13	0.32
X45	0.11	0.02	5.92**	0.06	0.20
X47	0.13	0.02	7.99**	0.10	0.34
X48	0.07	0.02	3.71**	0.02	0.12
X63	0.13	0.02	6.74**	0.07	0.24
X68	0.11	0.02	5.57**	0.05	0.19
X70	0.27	0.02	16.86**	0.30	0.61
X79	0.07	0.02	3.64**	0.02	0.11
X81	0.13	0.02	7.25**	0.08	0.27
X82	0.07	0.02	3.76**	0.02	0.13
X83	0.12	0.02	6.94**	0.08	0.27
X88	0.17	0.02	9.13**	0.13	0.33
X95	0.14	0.02	8.19**	0.11	0.33
Chi-Square = 64.791		df = 119	P = 1.00		

** $p < 0.01$

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

แผนภาพที่ 27 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่านระดับตีความ หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

3. โมเดลด้านความเข้าใจในการอ่าน ระดับประเมินค่า (EVAL) หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งของโรงเรียนที่จบการศึกษา

ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับประเมินค่า พบว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ดีมาก พิจารณาได้จากค่าไค-สแควร์ (Chi-Square = 6.499) ซึ่งมีค่าความน่าจะเป็นเข้าใกล้ 1 (P = 0.952) ที่องศาอิสระเท่ากับ 14 (df = 14) นั่นคือค่าไค-สแควร์ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่ายอมรับสมมติฐานหลักที่ว่าโมเดลการวิจัยสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ส่วนค่าดัชนีวัดความกลมกลืน (GFI) มีค่าเท่ากับ 0.998 ค่าดัชนีวัดความกลมกลืนที่ปรับแก้แล้ว (AGFI) มีค่าเท่ากับ 0.996 และค่าดัชนีวัดความสอดคล้อง (NFI) มีค่าเท่ากับ 0.961 รวมทั้งค่าดัชนีรากกำลังสองเฉลี่ยของเศษ (RMR) มีค่าเท่ากับ 0.003 ซึ่งมีค่าน้อยมากจนเข้าใกล้ศูนย์ ดังรายละเอียดในตารางที่ 46 และรายละเอียดของโมเดลด้านความเข้าใจในการอ่านระดับประเมินค่า สามารถแสดงได้ดังแผนภาพที่ 28

เมื่อพิจาราณาผลการวิเคราะห์โมเดลด้านความเข้าใจในการอ่านระดับประเมินค่า จากตารางที่ 46 และจากแผนภาพที่ 28 พบว่าน้ำหนักองค์ประกอบส่วนใหญ่ที่กำหนดข้อมูลจำเพาะของโมเดล CFA ไว้แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ยกเว้นตัวแปร X64 โดยน้ำหนักองค์ประกอบที่มีนัยสำคัญมีค่าอยู่ระหว่าง 0.08-0.20 และค่าสัมประสิทธิ์การพยากรณ์ (R-square) ของตัวแปร X78 มีค่าสูงสุดเท่ากับ 0.20

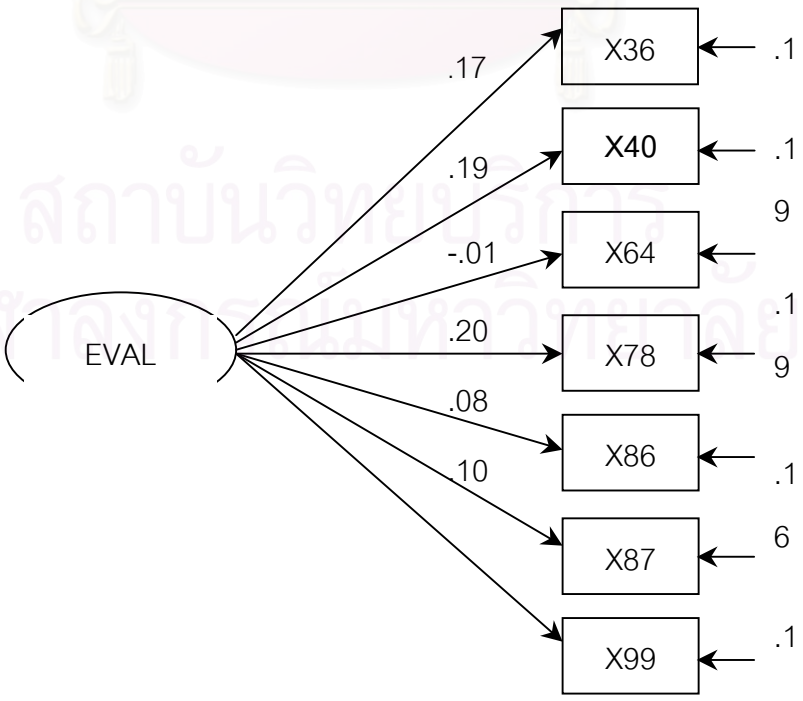
สำหรับการสร้างสเกลองค์ประกอบ ผู้วิจัยได้นำค่าสัมประสิทธ์คะแนนองค์ประกอบที่ได้จากการวิเคราะห์ใช้ในการสร้างสเกลองค์ประกอบด้านความเข้าใจในการอ่านระดับประเมินค่า (EVAL) เพื่อให้ได้ตัวแปรใหม่สำหรับนำไปวิเคราะห์เพื่อตรวจสอบความตรงเชิงโครงสร้างของแบบสอบวิชาภาษาอังกฤษต่อไป ซึ่งได้สเกลองค์ประกอบของความเข้าใจในการอ่านระดับประเมินค่า ดังสมการต่อไปนี้

$$EVAL = 0.47Z_{X36} + 0.54Z_{X40} - 0.04Z_{X64} + 0.69Z_{X78} + 0.30Z_{X86} + 0.24Z_{X87} + 0.53Z_{X99}$$

ตารางที่ 46 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับประเมินค่า (EVAL) ของแบบสอบวิชาภาษาอังกฤษ หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

ตัวแปร	เมตริกซ์น้ำหนักองค์ประกอบ			R Square	สปส. คะแนนองค์ประกอบ
	สปส.	SE.	t		
X36	0.17	0.02	7.21**	0.13	0.47
X40	0.19	0.02	7.95**	0.16	0.54
X64	-0.01	0.02	-0.62	0.00	-0.04
X78	0.20	0.02	8.66**	0.20	0.69
X86	0.08	0.02	4.35**	0.04	0.30
X87	0.10	0.02	4.19**	0.04	0.24
X99	0.16	0.02	7.32**	0.13	0.53
Chi-Square = 6.499		df = 14	P = 0.952	GFI = 0.998	
AGFI = 0.996		NFI = 0.961	RMR = 0.003		

** p < 0.01



แผนภาพที่ 28 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่านระดับประเมินค่าหลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

4. โมเดลด้านความสามารถในการใช้ภาษา (LANG) หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งของโรงเรียนที่จบการศึกษา

ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความสามารถในการใช้ภาษา พบว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ พิจารณาได้จากค่าไค-สแควร์ (Chi-Square = 77.105) ซึ่งมีค่าความน่าจะเป็นเท่ากับ 0.094 ที่องศาอิสระเท่ากับ 63 (df = 63) นั่นคือค่าไค-สแควร์ไม่ แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่ายอมรับสมมติฐานหลักที่ว่าโมเดลการวิจัยสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ส่วนค่าดัชนีวัดความกลมกลืน (GFI) มีค่าเท่ากับ 0.987 ค่าดัชนีวัดความกลมกลืนที่ปรับแก้แล้ว (AGFI) มีค่าเท่ากับ 0.981 และค่าดัชนีวัดความสอดคล้อง (NFI) มีค่าเท่ากับ 0.959 รวมทั้งค่าดัชนีรากกำลังสองเฉลี่ยของเศษ (RMR) มีค่าเท่ากับ 0.005 ซึ่งมีค่าน้อยมากจนเข้าใกล้ศูนย์ ดังรายละเอียดในตารางที่ 47 และรายละเอียดของโมเดลด้านความสามารถในการใช้ภาษาสามารถแสดงได้ดังแผนภาพที่ 29

เมื่อพิจารณาผลการวิเคราะห์โมเดลด้านความสามารถในการใช้ภาษา จากตารางที่ 47 และจากแผนภาพที่ 29 พบว่าน้ำหนักองค์ประกอบทุกค่าที่กำหนดข้อมูลจำเพาะของโมเดล CFA ไว้แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 และ 0.05 โดยน้ำหนักองค์ประกอบที่มีนัยสำคัญมีค่าอยู่ระหว่าง 0.03-0.38 และค่าสัมประสิทธิ์การพยากรณ์ (R-square) ของตัวแปร X3 มีค่าสูงสุดเท่ากับ 0.59

สำหรับการสร้างสเกลองค์ประกอบ ผู้วิจัยได้นำค่าสัมประสิทธ์คะแนนองค์ประกอบที่ได้จากการวิเคราะห์ไปใช้ในการสร้างสเกลองค์ประกอบด้านความสามารถในการใช้ภาษา (LANG) เพื่อให้ได้ตัวแปรใหม่สำหรับนำไปวิเคราะห์เพื่อตรวจสอบความตรงเชิงโครงสร้างของแบบสอบวิชาภาษาอังกฤษต่อไป ซึ่งได้สเกลองค์ประกอบของความสามารถในการใช้ภาษา ดังสมการต่อไปนี้

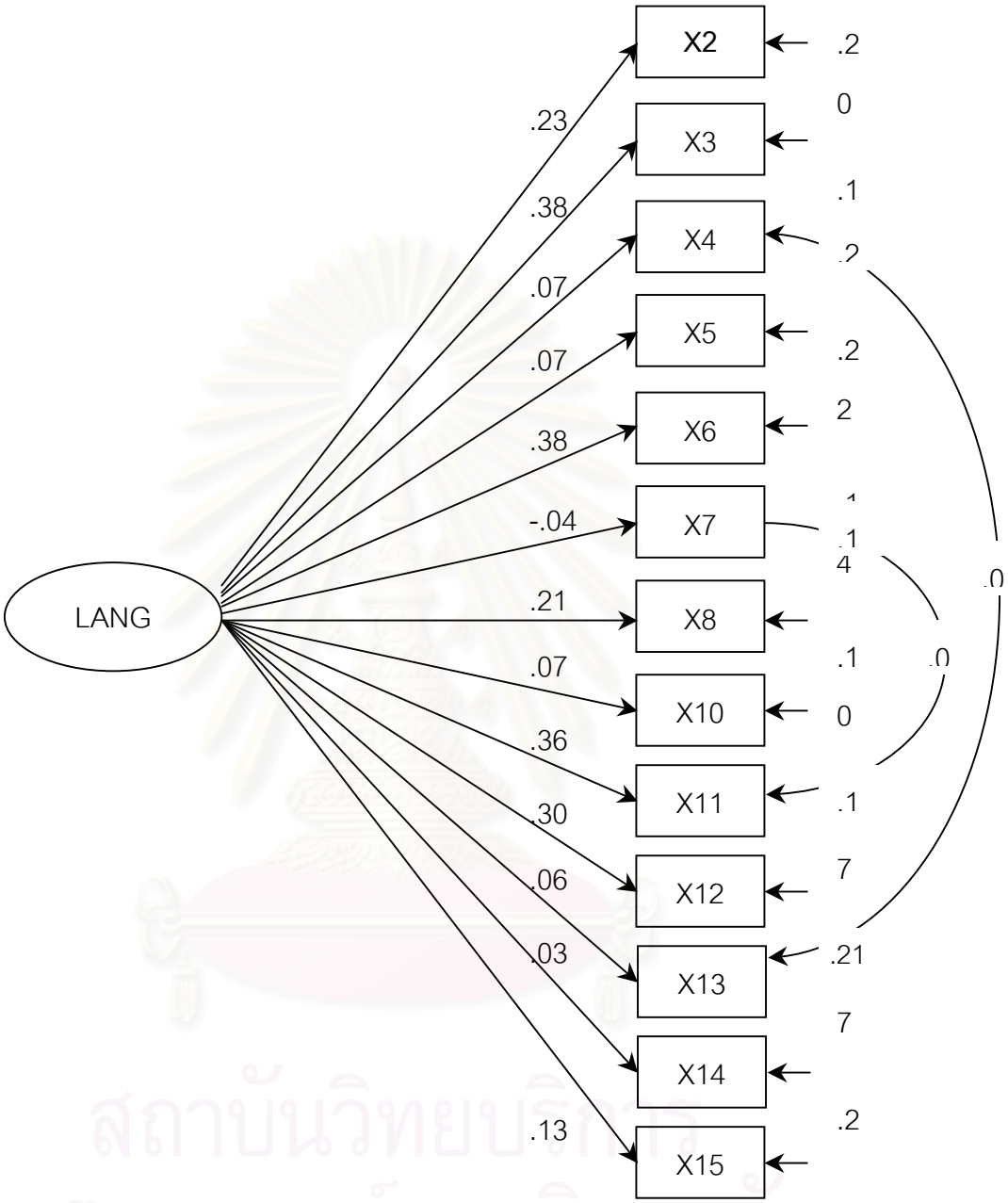
$$LANG = 0.18Z_{X2} + 0.61Z_{X3} + 0.04Z_{X4} + 0.08Z_{X5} + 0.59Z_{X6} - 0.12Z_{X7} + 0.20Z_{X8} + 0.05Z_{X10} + 0.49Z_{X11} + 0.30Z_{X12} + 0.04Z_{X13} + 0.03Z_{X14} + 0.13Z_{X15}$$

ตารางที่ 47 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความสามารถในการใช้ภาษา (LANG) ของแบบสอบวิชาภาษาอังกฤษ หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

ตัวแปร	เมตริกซ์น้ำหนักองค์ประกอบ			R Square	สปส. คะแนนองค์ประกอบ
	สปส.	SE.	t		
X2	0.23	0.02	13.81**	0.20	0.18
X3	0.38	0.02	26.44**	0.59	0.61
X4	0.07	0.02	4.21**	0.02	0.04
X5	0.07	0.01	5.01**	0.03	0.08
X6	0.38	0.02	26.19**	0.58	0.59
X7	-0.04	0.01	-2.52*	0.01	-0.12
X8	0.21	0.01	14.07**	0.21	0.20
X10	0.07	0.02	4.05**	0.02	0.05
X11	0.36	0.02	23.77**	0.50	0.49
X12	0.30	0.02	18.96**	0.35	0.30
X13	0.06	0.02	3.49**	0.01	0.04
X14	0.03	0.01	2.41*	0.01	0.03
X15	0.13	0.01	9.11**	0.09	0.13
Chi-Square = 77.105		df = 63	P = 0.094	GFI = 0.987	
AGFI = 0.981		NFI = 0.959	RMR = 0.005		

*p < 0.05 , ** p < 0.01

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



แผนภาพที่ 29 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความสามารถในการใช้ภาษาหลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

5. โมเดลด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ (GRAM) หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งของโรงเรียนที่จบการศึกษา

ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ พบว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ดี พิจารณาได้จากค่าไค-สแควร์ (Chi-Square = 277.284) ซึ่งมีค่าความน่าจะเป็นเท่ากับ 0.131 ที่องศาอิสระเท่ากับ 252 (df = 252) นั่นคือค่าไค-สแควร์ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่ายอมรับสมมติฐานหลักที่ว่าโมเดลการวิจัยสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ส่วนค่าดัชนีวัดความกลมกลืน (GFI) มีค่าเท่ากับ 0.974 ค่าดัชนีวัดความกลมกลืนที่ปรับแก้แล้ว (AGFI) มีค่าเท่ากับ 0.969 และค่าดัชนีวัดความสอดคล้อง (NFI) มีค่าเท่ากับ 0.796 รวมทั้งค่าดัชนีรากกำลังสองเฉลี่ยของเศษ (RMR) มีค่าเท่ากับ 0.006 ซึ่งมีค่า น้อยมากจนเข้าใกล้ศูนย์ ดังรายละเอียดในตารางที่ 48 และรายละเอียดของโมเดลด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ สามารถแสดงได้ดังแผนภาพที่ 30

เมื่อพิจารณาผลการวิเคราะห์โมเดลด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ จากตารางที่ 48 และจากแผนภาพที่ 30 พบว่าน้ำหนักองค์ประกอบส่วนใหญ่ที่กำหนดข้อมูลจำเพาะของโมเดล CFA ไว้แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ยกเว้นตัวแปร X54 และ X61 ที่น้ำหนักองค์ประกอบไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ โดยน้ำหนักองค์ประกอบที่มีนัยสำคัญมีค่าอยู่ระหว่าง 0.05-0.31 และค่าสัมประสิทธิ์การพยากรณ์ (R-square) ของตัวแปร X28 มีค่าสูงสุดเท่ากับ 0.38

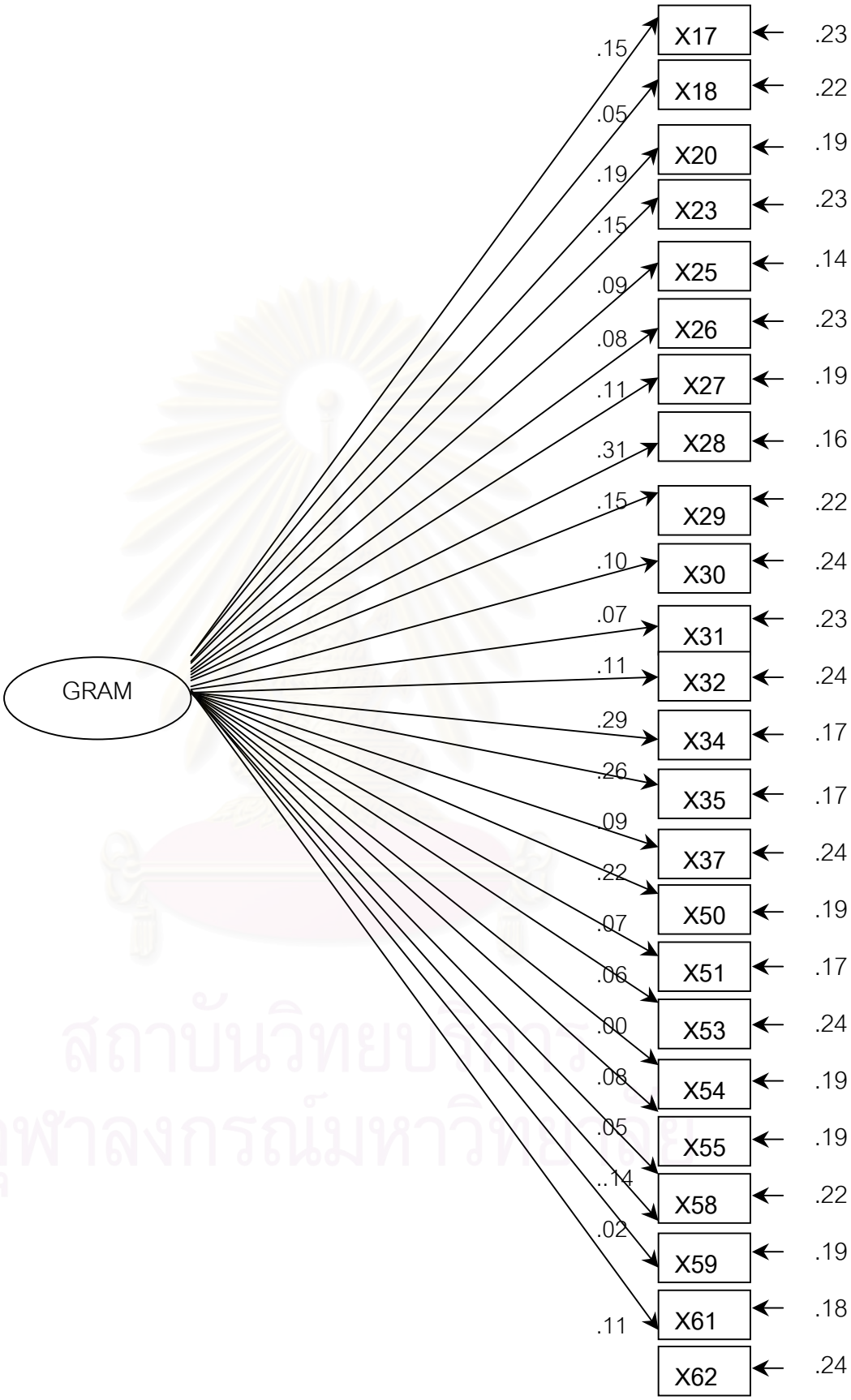
สำหรับการสร้างสเกลองค์ประกอบ ผู้วิจัยได้นำค่าสัมประสิทธ์คะแนนองค์ประกอบที่ได้จากการวิเคราะห์ใช้ในการสร้างสเกลองค์ประกอบด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ (GRAM) เพื่อให้ได้ตัวแปรใหม่สำหรับนำไปวิเคราะห์เพื่อตรวจสอบความตรงเชิงโครงสร้างของแบบสอบวิชาภาษาอังกฤษต่อไป ซึ่งได้สเกลองค์ประกอบของความรู้เกี่ยวกับโครงสร้างไวยากรณ์ ดังสมการต่อไปนี้

$$\begin{aligned} GRAM = {} & 0.17Z_{X17} + 0.06Z_{X18} + 0.26Z_{X20} + 0.17Z_{X23} + 0.16Z_{X25} + 0.10Z_{X26} + \\ & 0.16Z_{X27} + 0.52Z_{X28} + 0.18Z_{X28} + 0.11Z_{X30} + 0.08Z_{X31} + 0.13Z_{X32} + \\ & 0.45Z_{X34} + 0.40Z_{X35} + 0.10Z_{X37} + 0.30Z_{X50} + 0.11Z_{X51} + 0.07Z_{X53} - \\ & 0.01Z_{X54} + 0.12Z_{X55} + 0.07Z_{X58} + 0.19Z_{X59} + 0.03Z_{X61} + 0.13Z_{X62} \end{aligned}$$

ตารางที่ 48 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ (GRAM) ของแบบสอบวิชาภาษาอังกฤษหลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

ตัวแปร	เมตริกซ์น้ำหนักองค์ประกอบ			R Square	สปส. คะแนนองค์ประกอบ
	สปส.	SE.	t		
X17	0.15	0.02	8.14**	0.09	0.17
X18	0.05	0.02	2.78**	0.01	0.06
X20	0.19	0.02	11.10**	0.16	0.26
X23	0.15	0.02	8.26**	0.09	0.17
X25	0.09	0.02	6.02**	0.05	0.16
X26	0.08	0.02	4.66**	0.03	0.10
X27	0.11	0.02	6.82**	0.06	0.16
X28	0.31	0.02	18.07**	0.38	0.52
X29	0.15	0.02	8.45**	0.09	0.18
X30	0.10	0.02	5.31**	0.04	0.11
X31	0.07	0.02	3.83**	0.02	0.08
X32	0.11	0.02	6.16**	0.05	0.13
X34	0.29	0.02	16.76**	0.33	0.45
X35	0.26	0.02	15.33**	0.28	0.40
X37	0.09	0.02	4.88**	0.03	0.10
X50	0.22	0.02	12.74**	0.20	0.30
X51	0.07	0.02	4.43**	0.03	0.11
X53	0.06	0.02	3.27**	0.01	0.07
X54	0.00	0.02	-0.22	0.00	-0.01
X55	0.08	0.02	5.15**	0.04	0.12
X58	0.06	0.02	3.17**	0.01	0.07
X59	0.14	0.02	8.28**	0.09	0.19
X61	0.02	0.02	1.13	0.00	0.03
X62	0.11	0.02	6.10**	0.05	0.13
Chi-Square = 277.284		df = 252	P = 0.131	GFI = 0.974	
AGFI = 0.969		NFI = 0.796	RMR = 0.006		

** $p < 0.01$



แผนภาพที่ 30 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

โมเดลองค์ประกอบความรู้วิชาภาษาอังกฤษ หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

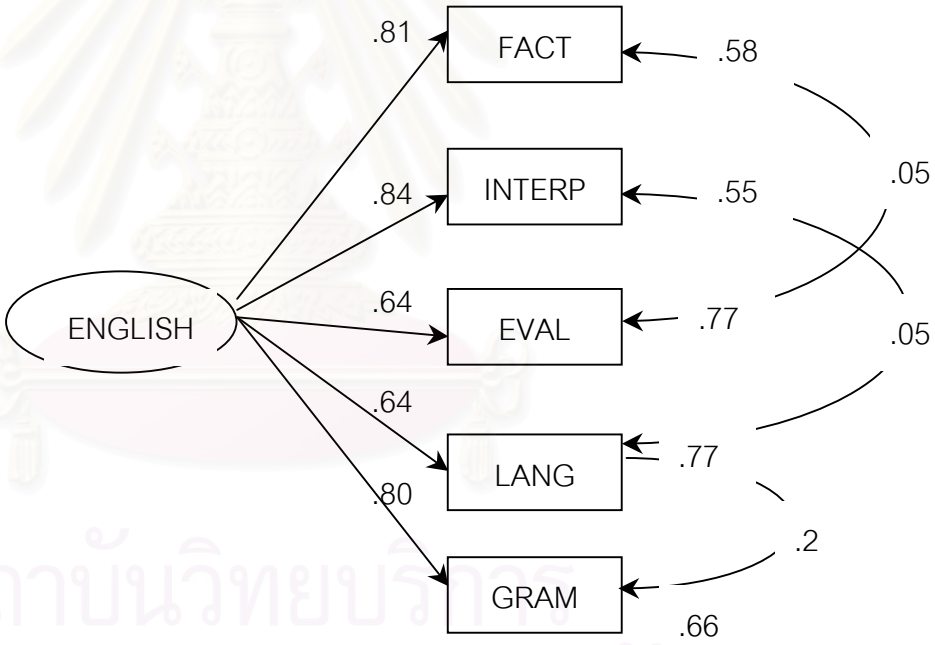
ผลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับที่สอง (second order confirmatory factor analysis) ของโมเดลองค์ประกอบความรู้วิชาภาษาอังกฤษ (ENGLISH) พบว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ดี โดยพิจารณาจากค่าไค-สแควร์ (Chi-Square = 2.301) ซึ่งมีค่าความน่าจะเป็นเท่ากับ 0.316 ที่องศาอิสระเท่ากับ 2 (df = 2) นั่นคือค่าไค-สแควร์ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่ายอมรับสมมติฐานหลักที่ว่าโมเดลการวิจัยสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ส่วนค่าดัชนีวัดความกลมกลืน (GFI) มีค่าเท่ากับ 0.999 ค่าดัชนีวัดความกลมกลืนที่ปรับแก้แล้ว (AGFI) มีค่าเท่ากับ 0.993 และค่าดัชนีวัดความสอดคล้อง (NFI) มีค่าเท่ากับ 0.999 รวมทั้งค่าดัชนีรากกำลังสองเฉลี่ยของเศษ (RMR) มีค่าเท่ากับ 0.017 ซึ่งมีค่าน้อยมากจนเข้าใกล้ศูนย์ ดังรายละเอียดในตารางที่ 49 และรายละเอียดของโมเดลองค์ประกอบความรู้วิชาภาษาอังกฤษ สามารถแสดงได้ดังแผนภาพที่ 31

ตารางที่ 49 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลองค์ประกอบความรู้วิชาภาษาอังกฤษ (ENGLISH) หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

องค์ประกอบ	เมตริกซ์น้ำหนักองค์ประกอบ			R Square	สปส. คะแนนองค์ประกอบ
	สปส.	SE.	t		
ความเข้าใจในการอ่าน ระดับข้อเท็จจริง (FACT)	0.81	0.05	28.82**	0.66	0.18
ความเข้าใจในการอ่าน ระดับตีความ (INTERP)	0.84	0.05	30.35**	0.70	0.22
ความเข้าใจในการอ่าน ระดับประเมินค่า (EVAL)	0.64	0.05	20.66**	0.41	0.08
ความสามารถในการใช้ภาษา (LANG)	0.64	0.06	20.10**	0.41	0.01
ความรู้เกี่ยวกับโครงสร้างไวยากรณ์ (GRAM)	0.80	0.05	28.52**	0.64	0.16
Chi-Square = 2.301	df = 2	P = 0.316	GFI = 0.999		
AGFI = 0.993	NFI = 0.999	RMR = 0.017			

** $p < 0.01$

เมื่อพิจารณาผลการวิเคราะห์โมเดลความรู้วิชาภาษาอังกฤษ จากตารางที่ 49 และจากแผนภาพที่ 31 พบว่าน้ำหนักองค์ประกอบทุกค่าที่กำหนดข้อมูลจำเพาะของโมเดล CFA ไว้แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 โดยองค์ประกอบด้านความเข้าใจในการอ่านระดับตีความ (INTERP) มีน้ำหนักองค์ประกอบมากที่สุด คือ 0.84 และองค์ประกอบด้านความเข้าใจในการอ่านระดับประเมินค่า (EVAL) และองค์ประกอบด้านความสามารถในการใช้ภาษา มีค่าน้ำหนักองค์ประกอบน้อยที่สุด คือ 0.64 ค่าสัมประสิทธิ์การพยากรณ์ (R-square) ขององค์ประกอบด้านความเข้าใจในการอ่านระดับตีความ (INTERP) มีค่าสูงสุดเท่ากับ 0.70 และค่าสัมประสิทธิ์ของการพยากรณ์ขององค์ประกอบด้านความเข้าใจในการอ่านระดับประเมินค่า (EVAL) มีค่าต่ำสุดคือเท่ากับ 0.41



แผนภาพที่ 31 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลองค์ประกอบความรู้วิชาภาษาอังกฤษหลังจากการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามตัวแปรสถานที่ตั้งของโรงเรียนที่จบการศึกษา

จากผลการวิเคราะห์องค์ประกอบเชิงยืนยันสามารถสรุปได้ว่า แบบสอบวิชาภาษาอังกฤษหลังการตัดข้อสอบที่พบ DIF ออกจากแบบสอบทุกข้อแล้ว เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา มีความตรงเชิงโครงสร้างตามโมเดลความรู้วิชาภาษาอังกฤษ

ค) หลังจากการวิเคราะห์ DTF

1. โมเดลด้านความเข้าใจในการอ่าน ระดับข้อเท็จจริง (FACT) หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งของโรงเรียนที่จบการศึกษา

ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่านระดับข้อเท็จจริง หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา ได้ผลเช่นเดียวกับผลการวิเคราะห์องค์ประกอบเชิงยืนยันก่อนการวิเคราะห์ DTF เนื่องจากไม่มีข้อสอบที่ถูกตัดออกจากโมเดลนี้

2. โมเดลด้านความเข้าใจในการอ่าน ระดับตีความ (INTERP) หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับตีความพบว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์พอดี พิจารณาได้จากค่าไค-สแควร์ (Chi-Square = 72.583) ซึ่งมีค่าความน่าจะเป็นเท่ากับ 1 (P = 1.00) ที่องศาอิสระเท่ากับ 119 (df = 119) นั่นคือค่าไค-สแควร์ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่ายอมรับสมมติฐานหลักที่ว่าโมเดลการวิจัยสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ดังรายละเอียดในตารางที่ 50

เมื่อพิจาราณาผลการวิเคราะห์โมเดลด้านความเข้าใจในการอ่านระดับตีความ จากตารางที่ 50 และจากแผนภาพที่ 32 พบว่าน้ำหนักองค์ประกอบส่วนใหญ่ที่กำหนดข้อมูลจำเพาะของโมเดล CFA ไว้แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ยกเว้นตัวแปร X22 ที่น้ำหนักองค์ประกอบไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ โดยน้ำหนักองค์ประกอบที่มีนัยสำคัญมีค่าอยู่ระหว่าง 0.06-0.29 และค่าสัมประสิทธิ์การพยากรณ์ (R-square) ของตัวแปร X70 มีค่าสูงสุดเท่ากับ 0.36

สำหรับการสร้างสเกลองค์ประกอบ ผู้วิจัยได้นำค่าสัมประสิทธ์คะแนนองค์ประกอบที่ได้จากการวิเคราะห์ใช้ในการสร้างสเกลองค์ประกอบด้านความเข้าใจในการอ่านระดับตีความ (INTERP) เพื่อให้ได้ตัวแปรใหม่สำหรับนำไปวิเคราะห์เพื่อตรวจสอบความตรงเชิงโครงสร้างของแบบสอบวิชาภาษาอังกฤษต่อไป ซึ่งได้สเกลองค์ประกอบของความเข้าใจในการอ่านระดับตีความ ดังสมการต่อไปนี้

$$\begin{aligned}INTERP = {} & 0.09Z_{X19} + 0.22Z_{X21} - 0.03Z_{X22} + 0.21Z_{X38} + 0.28Z_{X43} + 0.16Z_{X45} + \\ & 0.26Z_{X47} + 0.10Z_{X48} + 0.19Z_{X63} + 0.64Z_{X70} + 0.48Z_{X77} + 0.08Z_{X79} + \\ & 0.24Z_{X81} + 0.11Z_{X82} + 0.25Z_{X83} + 0.29Z_{X88} + 0.29Z_{X95}\end{aligned}$$

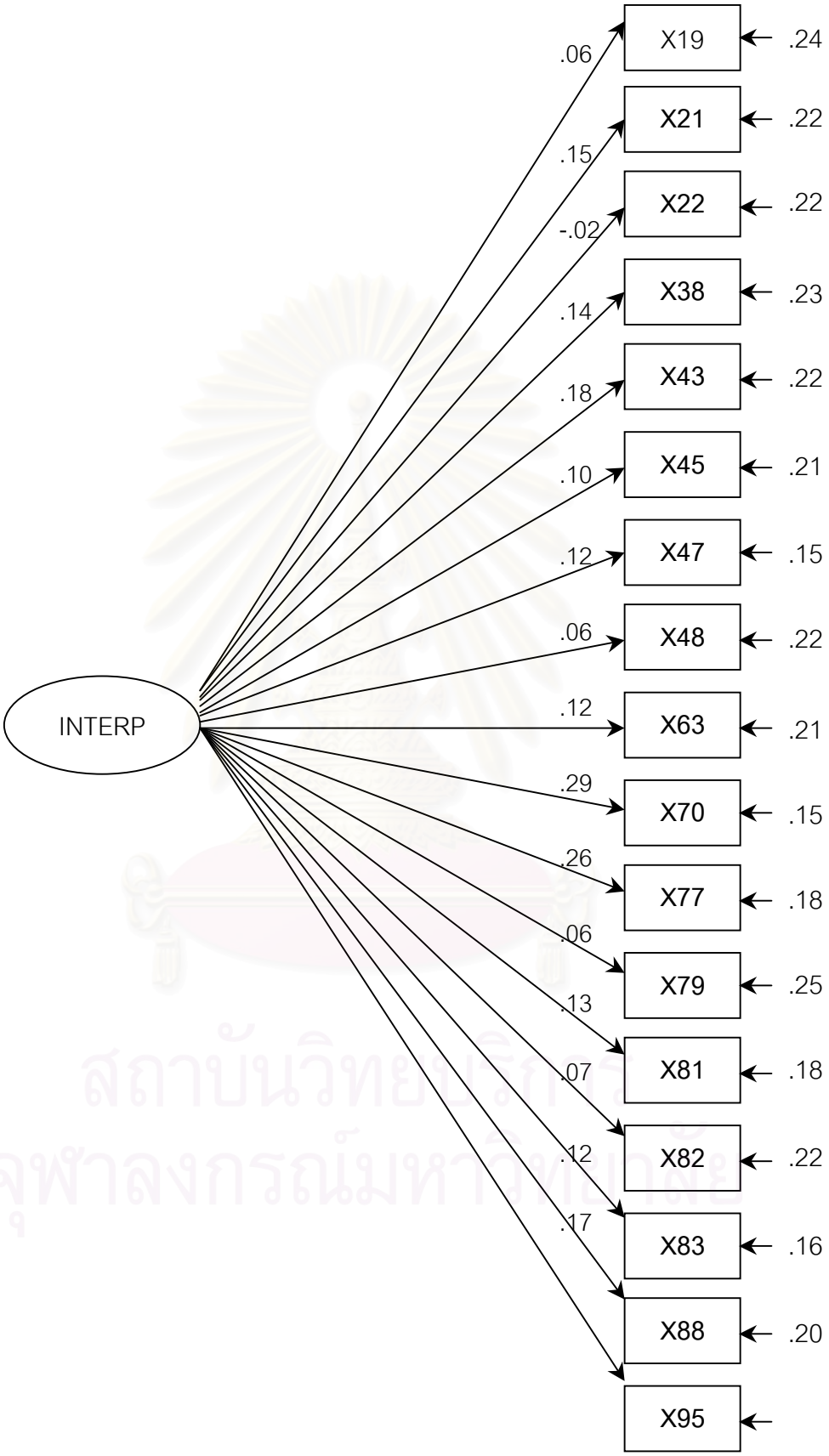
ตารางที่ 50 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่าน ระดับตีความ

(INTERP) ของแบบสอบวิชาภาษาอังกฤษ หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

ตัวแปร	เมตริกซ์น้ำหนักองค์ประกอบ			R	สปส. คะแนนองค์ประกอบ
	สปส.	SE.	t	**Square**	
X19	0.06	0.02	3.36**	0.02	0.09
X21	0.15	0.02	7.75**	0.09	0.22
X22	-0.02	0.02	-0.90	0.00	-0.03
X38	0.14	0.02	7.40**	0.08	0.21
X43	0.18	0.02	9.50**	0.13	0.28
X45	0.10	0.2	5.54**	0.05	0.16
X47	0.12	0.02	7.52**	0.08	0.26
X48	0.06	0.02	3.48**	0.02	0.10
X63	0.12	0.02	6.56**	0.06	0.19
X70	0.29	0.02	16.14**	0.36	0.64
X77	0.26	0.02	14.11**	0.28	0.48
X79	0.06	0.02	3.17**	0.02	0.08
X81	0.13	0.02	7.57**	0.08	0.24
X82	0.07	0.02	3.78**	0.02	0.11
X83	0.12	0.02	7.43**	0.08	0.25
X88	0.17	0.02	9.53**	0.13	0.29
X95	0.14	0.02	8.56**	0.11	0.29
Chi-Square = 72.583		df = 119		P = 1.00	

** p < 0.01

รายละเอียดของโมเดลด้านความเข้าใจในการอ่านระดับตีความ สามารถแสดงได้ดังแผนภาพที่ 32



แผนภาพที่ 32 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่านระดับ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตีความ หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

3. โมเดลด้านความเข้าใจในการอ่าน ระดับประเมินค่า (EVAL) หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความเข้าใจในการอ่านระดับประเมินค่าหลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา ได้ผลเช่นเดียวกับผลการวิเคราะห์องค์ประกอบเชิงยืนยันก่อนการวิเคราะห์ DTF เนื่องจากไม่มีข้อสอบที่ถูกตัดออกจากโมเดลนี้

4. โมเดลด้านความสามารถในการใช้ภาษา (LANG) หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์โรงเรียนที่จบการศึกษา

ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความสามารถในการใช้ภาษา พบว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ พิจารณาได้จากค่าไค-สแควร์ (Chi-Square = 100.648) ซึ่งมีค่าความน่าจะเป็นเท่ากับ 0.134 ที่องศาอิสระเท่ากับ 88 (df = 88) นั่นคือค่าไค-สแควร์ไม่ แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่ายอมรับสมมติฐานหลักที่ว่าโมเดลการวิจัยสอดคล้อง กลมกลืนกับข้อมูลเชิงประจักษ์ ส่วนค่าดัชนีวัดความกลมกลืน (GFI) มีค่าเท่ากับ 0.986 ค่าดัชนีวัดความกลมกลืนที่ปรับแก้แล้ว (AGFI) มีค่าเท่ากับ 0.980 และค่าดัชนีวัดความสอดคล้อง (NFI) มีค่าเท่ากับ 0.955 รวมทั้งค่าดัชนีรากกำลังสองเฉลี่ยของเศษ (RMR) มีค่าเท่ากับ 0.005 ซึ่งมีค่าน้อยมากจนเข้าใกล้ศูนย์ ดังรายละเอียดในตารางที่ 51

เมื่อพิจารณาผลการวิเคราะห์โมเดลด้านความสามารถในการใช้ภาษา จากตารางที่ 51 และจากแผนภาพที่ 33 พบว่าน้ำหนักองค์ประกอบทุกค่าที่กำหนดข้อมูลจำเพาะของโมเดล CFA ไว้แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 โดยน้ำหนักองค์ประกอบที่มีนัยสำคัญมีค่าอยู่ระหว่าง -0.04-0.38 และค่าสัมประสิทธิ์การพยากรณ์ (R-square) ของตัวแปร X3 มีค่าสูงสุดเท่ากับ 0.59

สำหรับการสร้างสเกลองค์ประกอบ ผู้วิจัยได้นำค่าสัมประสิทธิ์คะแนนองค์ประกอบที่ได้จากการวิเคราะห์ไปใช้ในการสร้างสเกลองค์ประกอบด้านความสามารถในการใช้ภาษา (LANG) เพื่อให้ได้ตัวแปรใหม่สำหรับนำไปวิเคราะห์เพื่อตรวจสอบความตรงเชิงโครงสร้างของแบบสอบวิชาภาษาอังกฤษต่อไป ซึ่งได้สเกลองค์ประกอบของความสามารถในการใช้ภาษา ดังสมการต่อไปนี้

$$LANG = 0.17Z_{X2} + 0.56Z_{X3} + 0.05Z_{X4} + 0.07Z_{X5} + 0.55Z_{X6} - 0.12Z_{X7} + 0.20Z_{X8} + 0.05Z_{X10} + 0.46Z_{X11} + 0.28Z_{X12} + 0.04Z_{X13} + 0.03Z_{X14} +$$

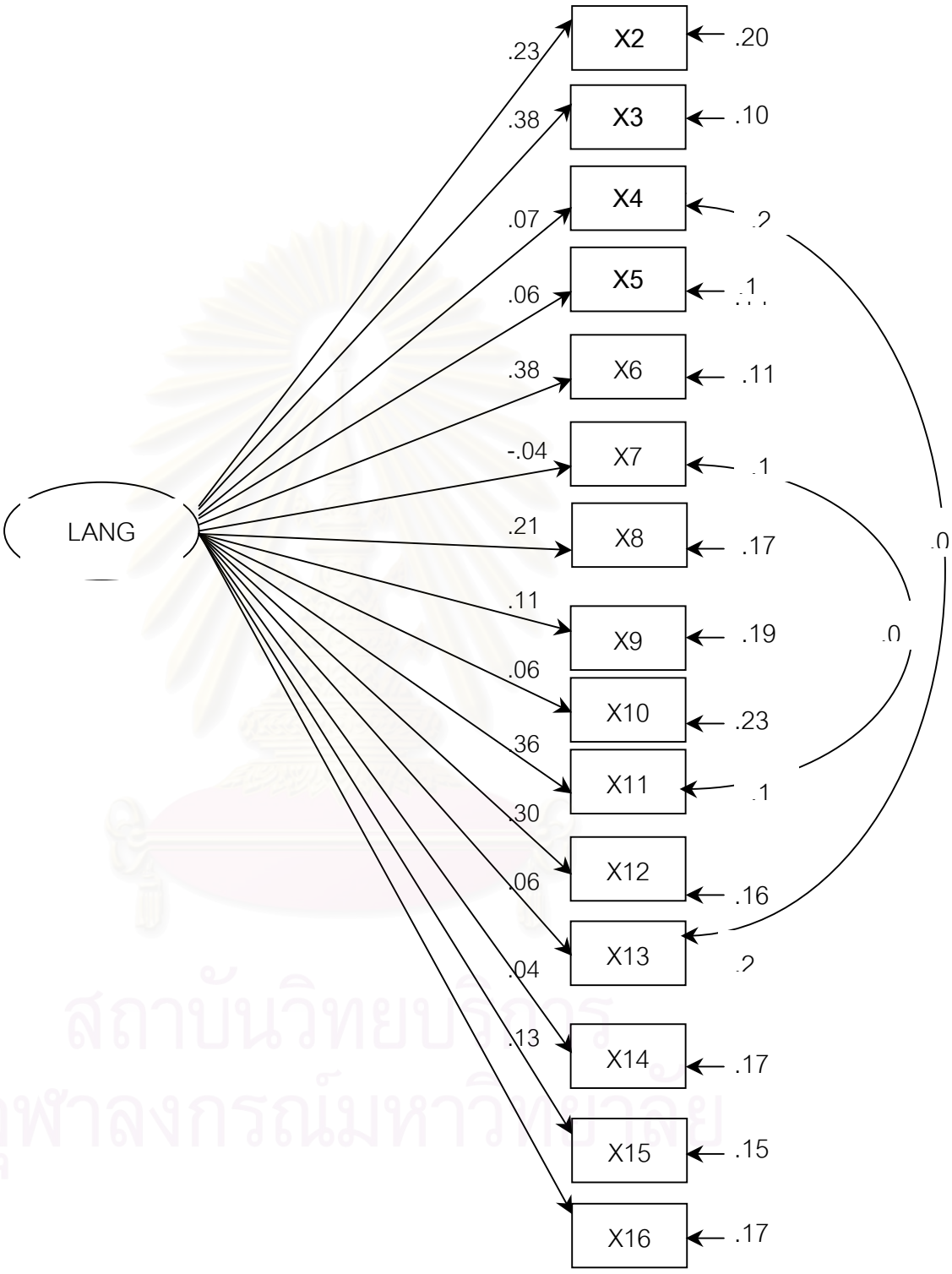
$0.13Z_{X15} + 0.23Z_{X16}$

ตารางที่ 51 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความสามารถในการใช้ภาษา (LANG) ของแบบสอบวิชาภาษาอังกฤษ หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

ตัวแปร	เมตริกซ์น้ำหนักองค์ประกอบ			R Square	สปส. คะแนนองค์ประกอบ
	สปส.	SE.	t		
X2	0.23	0.02	13.93**	0.20	0.17
X3	0.38	0.01	26.59**	0.59	0.56
X4	0.07	0.02	4.38**	0.02	0.04
X5	0.06	0.01	4.88**	0.03	0.06
X6	0.38	0.01	26.36**	0.58	0.55
X7	-0.04	0.01	-2.53*	0.01	-0.11
X8	0.21	0.01	14.39**	0.22	0.20
X9	0.11	0.01	7.06**	0.06	0.08
X10	0.07	0.02	4.18**	0.02	0.05
X11	0.36	0.01	24.00**	0.50	0.46
X12	0.30	0.02	19.17**	0.35	0.28
X13	0.06	0.02	3.60**	0.02	0.04
X14	0.04	0.01	2.52*	0.01	0.03
X15	0.13	0.01	9.24**	0.10	0.13
X16	0.26	0.02	16.75**	0.28	0.23
Chi-Square = 100.648		df = 88	P = 0.134	GFI = 0.986	
AGFI = 0.980		NFI = 0.955	RMR = 0.005		

*p < 0.05 , **p < 0.01

รายละเอียดของโมเดลด้านความเข้าใจในการอ่านระดับประเมินค่า สามารถแสดงได้ดังแผนภาพที่ 33



แผนภาพที่ 33 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความสามารถในการใช้ภาษาหลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

5. โมเดลด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ (GRAM) หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบ

ตามสถานที่ตั้งของโรงเรียนที่จบการศึกษา

ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ พบว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ดี พิจารณาได้จากค่าไค-สแควร์ (Chi-Square = 328.364) ซึ่งมีค่าความน่าจะเป็นเท่ากับ 0.422 ที่องศาอิสระเท่ากับ 324 (df = 324) นั่นคือค่าไค-สแควร์ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่ายอมรับสมมติฐานหลักที่ว่าโมเดลการวิจัยสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ส่วนค่าดัชนีวัดความกลมกลืน (GFI) มีค่าเท่ากับ 0.973 ค่าดัชนีวัดความกลมกลืนที่ปรับแก้แล้ว (AGFI) มีค่าเท่ากับ 0.968 และค่าดัชนีวัดความสอดคล้อง (NFI) มีค่าเท่ากับ 0.809 รวมทั้งค่าดัชนีรากกำลังสองเฉลี่ยของเศษ (RMR) มีค่าเท่ากับ 0.006 ซึ่งมีค่า น้อยมากจนเข้าใกล้ศูนย์ ดังรายละเอียดในตารางที่ 52 และรายละเอียดของโมเดลด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ สามารถแสดงได้ดังแผนภาพที่ 34

เมื่อพิจารณาผลการวิเคราะห์โมเดลด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ จากตารางที่ 52 และจากแผนภาพที่ 34 พบว่าน้ำหนักองค์ประกอบส่วนใหญ่ที่กำหนดข้อมูลจำเพาะของโมเดล CFA ไว้แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ยกเว้นตัวแปร X54 และ X61 ที่น้ำหนักองค์ประกอบไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ โดยน้ำหนักองค์ประกอบที่มีนัยสำคัญมีค่าอยู่ระหว่าง 0.04-0.31 และค่าสัมประสิทธิ์การพยากรณ์ (R-square) ของตัวแปร X28 มีค่าสูงสุดเท่ากับ 0.38

สำหรับการสร้างสเกลองค์ประกอบ ผู้วิจัยได้นำค่าสัมประสิทธ์คะแนนองค์ประกอบที่ได้จากการวิเคราะห์ใช้ในการสร้างสเกลองค์ประกอบด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ (GRAM) เพื่อให้ได้ตัวแปรใหม่สำหรับนำไปวิเคราะห์เพื่อตรวจสอบความตรงเชิงโครงสร้างของแบบสอบวิชาภาษาอังกฤษต่อไป ซึ่งได้สเกลองค์ประกอบของความรู้เกี่ยวกับโครงสร้างไวยากรณ์ ดังสมการต่อไปนี้

$$\begin{aligned}GRAM = {} & 0.16Z_{X17} + 0.05Z_{X18} + 0.24Z_{X20} + 0.16Z_{X23} + 0.15Z_{X25} + 0.09Z_{X26} + \\ & 0.14Z_{X27} + 0.49Z_{X28} + 0.16Z_{X29} + 0.10Z_{X30} + 0.07Z_{X31} + 0.12Z_{X32} + \\ & 0.42Z_{X34} + 0.37Z_{X35} + 0.29Z_{X50} + 0.10Z_{X51} + 0.14Z_{X52} + 0.06Z_{X53} - \\ & 0.01Z_{X54} + 0.12Z_{X55} + 0.06Z_{X56} + 0.28Z_{X57} + 0.07Z_{X58} + 0.17Z_{X59} + \\ & 0.06Z_{X60} + 0.03Z_{X61} + 0.12Z_{X62}\end{aligned}$$

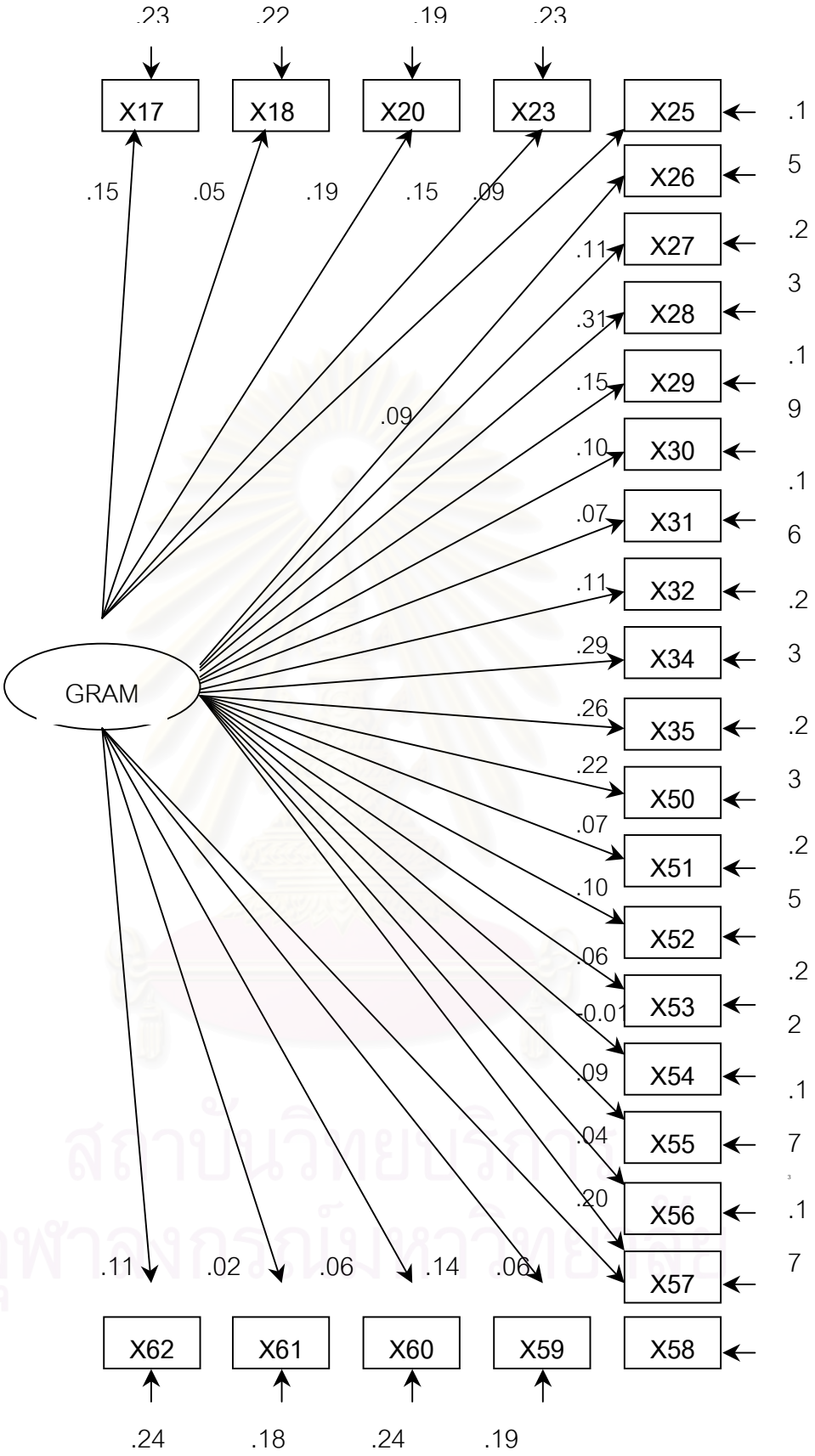
ตารางที่ 52 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความรู้เกี่ยวกับโครงสร้างไวยากรณ์ (GRAM) ของแบบสอบวิชาภาษาอังกฤษหลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

ตัวแปร	เมตริกซ์น้ำหนักองค์ประกอบ			R	สปส. คะแนน
	สปส.	SE.	t	Square	องค์ประกอบ
X17	0.15	0.02	8.25**	0.09	0.16
X18	0.05	0.02	2.79**	0.01	0.05
X20	0.19	0.02	11.29**	0.16	0.24
X23	0.15	0.02	8.21**	0.09	0.16
X25	0.09	0.01	6.18**	0.05	0.15
X26	0.09	0.02	4.81**	0.03	0.09
X27	0.11	0.02	6.70**	0.06	0.14
X28	0.31	0.02	18.53**	0.38	0.49
X29	0.15	0.02	8.29**	0.09	0.16
X30	0.10	0.02	5.37**	0.04	0.10
X31	0.07	0.02	3.70**	0.02	0.07
X32	0.11	0.02	6.33**	0.05	0.12
X34	0.29	0.02	17.06**	0.33	0.42
X35	0.26	0.02	15.53**	0.28	0.37
X50	0.22	0.02	12.96**	0.20	0.29
X51	0.07	0.02	4.50**	0.03	0.10
X52	0.10	0.02	6.42**	0.05	0.14
X53	0.06	0.02	3.13**	0.01	0.06
X54	-0.01	0.02	-0.33	0.00	-0.01
X55	0.09	0.02	5.46**	0.04	0.12
X56	0.04	0.02	2.77**	0.01	0.06
X57	0.20	0.02	12.32**	0.18	0.28
X58	0.06	0.02	3.32**	0.02	0.07
X59	0.14	0.02	8.24**	0.09	0.17
X60	0.06	0.02	3.44**	0.02	0.06
X61	0.02	0.02	1.21	0.00	0.03
X62	0.11	0.02	6.21**	0.05	0.12

Chi-Square = 328.364 df = 324 P = 0.422 GFI = 0.973

AGFI = 0.968 NFI = 0.793 RMR = 0.006

** $p < 0.01$



แผนภาพที่ 34 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลความรู้เกี่ยวกับโครงสร้างไวยากรณ์หลังจากการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

โมเดลองค์ประกอบความรู้วิชาภาษาอังกฤษ หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

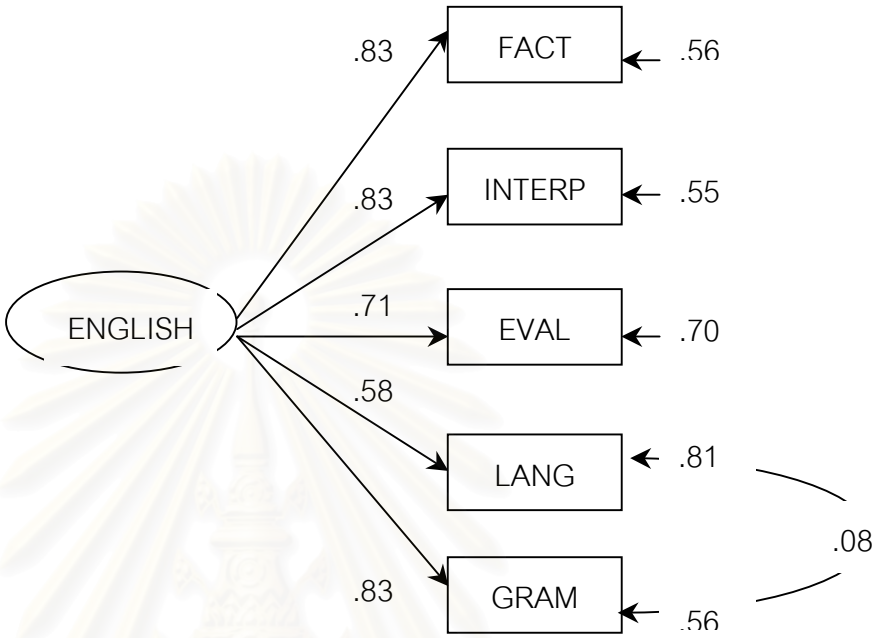
ผลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับที่สอง (second order confirmatory factor analysis) ของโมเดลองค์ประกอบความรู้วิชาภาษาอังกฤษ (ENGLISH) พบว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ดี โดยพิจารณาจากค่าไค-สแควร์ (Chi-Square = 6.413) ซึ่งมีค่าความน่าจะเป็นเท่ากับ 0.170 ที่องศาอิสระเท่ากับ 4 (df = 4) นั่นคือค่าไค-สแควร์ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่ายอมรับสมมติฐานหลักที่ว่าโมเดลการวิจัยสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ส่วนค่าดัชนีวัดความกลมกลืน (GFI) มีค่าเท่ากับ 0.997 ค่าดัชนีวัดความกลมกลืนที่ปรับแก้แล้ว (AGFI) มีค่าเท่ากับ 0.990 และค่าดัชนีวัดความสอดคล้อง (NFI) มีค่าเท่ากับ 0.997 รวมทั้งค่าดัชนีรากกำลังสองเฉลี่ยของเศษ (RMR) มีค่าเท่ากับ 0.023 ซึ่งมีค่าน้อยมากจนเข้าใกล้ศูนย์ ดังรายละเอียดในตารางที่ 53

ตารางที่ 53 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลองค์ประกอบความรู้วิชาภาษาอังกฤษ (ENGLISH) หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

องค์ประกอบ	เมตริกซ์น้ำหนักองค์ประกอบ			R Square	สปส. คะแนนองค์ประกอบ
	สปส.	SE.	t		
ความเข้าใจในการอ่านระดับข้อเท็จจริง (FACT)	0.83	0.05	30.53**	0.68	0.16
ความเข้าใจในการอ่านระดับตีความ (INTERP)	0.83	0.05	30.74**	0.70	0.18
ความเข้าใจในการอ่านระดับประเมินค่า (EVAL)	0.71	0.04	24.72**	0.51	0.11
ความสามารถในการใช้ภาษา (LANG)	0.58	0.04	18.82**	0.34	0.05
ความรู้เกี่ยวกับโครงสร้างไวยากรณ์ (GRAM)	0.83	0.05	30.64**	0.69	0.16
Chi-Square = 6.413	df = 4		P = 0.170	GFI = 0.997	
AGFI = 0.990	NFI = 0.997		RMR = 0.023		

** $p < 0.01$

รายละเอียดของโมเดลด้านความเข้าใจในการอ่านระดับประเมินค่า สามารถแสดงได้ดัง แผนภาพที่ 35



แผนภาพที่ 35 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลองค์ประกอบความรู้วิชาภาษาอังกฤษหลังจากการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

เมื่อพิจารณาผลการวิเคราะห์โมเดลความรู้วิชาภาษาอังกฤษ จากตารางที่ 53 และจากแผนภาพที่ 35 พบว่าน้ำหนักองค์ประกอบทุกค่าที่กำหนดข้อมูลจำเพาะของโมเดล CFA ไว้แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 โดยองค์ประกอบด้านความเข้าใจในการอ่าน ระดับข้อเท็จจริง (FACT) ระดับตีความ (INTERP) และความรู้เกี่ยวกับโครงสร้างไวยากรณ์ (GRAM) มีน้ำหนักองค์ประกอบมากที่สุด คือ 0.83 องค์ประกอบด้านความเข้าใจในการอ่านระดับประเมินค่า มีค่าน้ำหนักองค์ประกอบเท่ากับ 0.71 และด้านความสามารถในการใช้ภาษา (LANG) มีค่าน้ำหนักองค์ประกอบเท่ากับ 0.58 และค่าสัมประสิทธิ์การพยากรณ์ (R-square) ขององค์ประกอบด้านความเข้าใจในการอ่านระดับตีความ (INTERP) และความรู้เกี่ยวกับโครงสร้างไวยากรณ์ (GRAM) มีค่าสูงสุดเท่ากับ 0.70

จากผลการวิเคราะห์องค์ประกอบเชิงยืนยันสามารถสรุปได้ว่า เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา แบบสอบวิชาภาษาอังกฤษหลังการตัดข้อสอบที่พบ DIF ออกจากแบบสอบบางข้อจนแบบสอบไม่ทำหน้าที่ต่างกันแล้ว มีความตรงเชิงโครงสร้างตามโมเดลความรู้วิชาภาษาอังกฤษ

ผลการเปรียบเทียบดัชนีความตรงจากการวิเคราะห์องค์ประกอบของแบบวิชาภาษาอังกฤษก่อนและหลังการตัดข้อสอบที่พบ DIF จากการจำแนกกลุ่มด้วยตัวแปรเพศ และสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา มีรายละเอียดดังตารางที่ 54

ตารางที่ 54 ผลการเปรียบเทียบดัชนีความตรงจากการวิเคราะห์องค์ประกอบของแบบสอบวิชาภาษาอังกฤษ ก่อนและหลังการตัดข้อสอบที่พบ DIF จากตัวแปรเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

ดัชนีของความตรง	χ^2	df	RMR	P	GFI	AGFI	NFI
ก่อนการตัดข้อสอบที่พบ DIF	1.618	2	0.012	0.445	0.999	0.995	0.999
หลังการตัดข้อสอบที่พบ DIF							
จำแนกกลุ่มด้วยตัวแปรเพศ โดยใช้	0.932	4	0.009	0.920	1.000	0.999	1.000
	1.086	2	0.009	0.581	1.000	0.997	1.000
1. ดัชนี NCDIF							
2. ดัชนี CDIF/DTF							
จำแนกกลุ่มด้วยตัวแปรสถานที่ตั้งโรงเรียน โดยใช้	2.301	2	0.017	0.316	0.999	0.993	0.999
	6.413	4	0.022	0.170	0.997	0.990	0.997
1. ดัชนี NCDIF							
2. ดัชนี CDIF/DTF							

จากตารางที่ 54 ผลการเปรียบเทียบดัชนีความตรงจากการวิเคราะห์องค์ประกอบของแบบสอบวิชาภาษาอังกฤษก่อนและหลังการตัดข้อสอบที่พบ DIF ในกรณีต่าง ๆ พบว่า ทุกกรณีเป็นโมเดลที่มีความสอดคล้องกับข้อมูลเชิงประจักษ์ ดัชนีวัดระดับความกลมกลืน (GFI) มีค่าเข้าใกล้ 1 ดัชนีวัดระดับความกลมกลืนที่ปรับแก้แล้ว (AGFI) มีค่าเข้าใกล้ 1 ดัชนีวัดระดับความสอดคล้องปกติ (NFI) มีค่าเข้าใกล้ 1 และดัชนีรากกำลังสองเฉลี่ยของเศษ (RMR) มีค่าเข้าใกล้ศูนย์

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

4.2.2 วิชาคณิตศาสตร์

จากการวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) หลังจากตัดข้อสอบที่พบ DIF ทั้งในกรณีที่ตัดข้อสอบที่พบ DIF ทุกข้อและตัดเฉพาะบางข้อโดยแบ่งกลุ่มผู้สอบตามตัวแปรที่ศึกษา ได้ผลการวิเคราะห์ ซึ่งมีรายละเอียดดังตารางที่ 55 - 57

4.2.2.1 เมื่อแบ่งกลุ่มผู้สอบตามเพศ

โมเดลความรู้วิชาคณิตศาสตร์ หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลความรู้วิชาคณิตศาสตร์ พบว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ดี พิจารณาได้จากค่าไค-สแควร์ (Chi-Square = 146.907) ซึ่งมีค่าความน่าจะเป็นเท่ากับ 0.228 ที่องศาอิสระเท่ากับ 135 (df = 135) นั่นคือค่าไค-สแควร์ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่ายอมรับสมมติฐานหลักที่ว่าโมเดลการวิจัยสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ส่วนค่าดัชนีวัดความกลมกลืน (GFI) มีค่าเท่ากับ 0.984 ค่าดัชนีวัดความกลมกลืนที่ปรับแก้แล้ว (AGFI) มีค่าเท่ากับ 0.980 และค่าดัชนีวัดความสอดคล้อง (NFI) มีค่าเท่ากับ 0.666 รวมทั้งค่าดัชนีรากกำลังสองเฉลี่ยของเศษ (RMR) มีค่าเท่ากับ 0.056 ซึ่งมีค่าน้อยมากจนเข้าใกล้ศูนย์ ดังรายละเอียดในตารางที่ 55 และรายละเอียดของโมเดลด้านความรู้วิชาคณิตศาสตร์ (MATH) สามารถแสดงได้ดังแผนภาพที่ 36

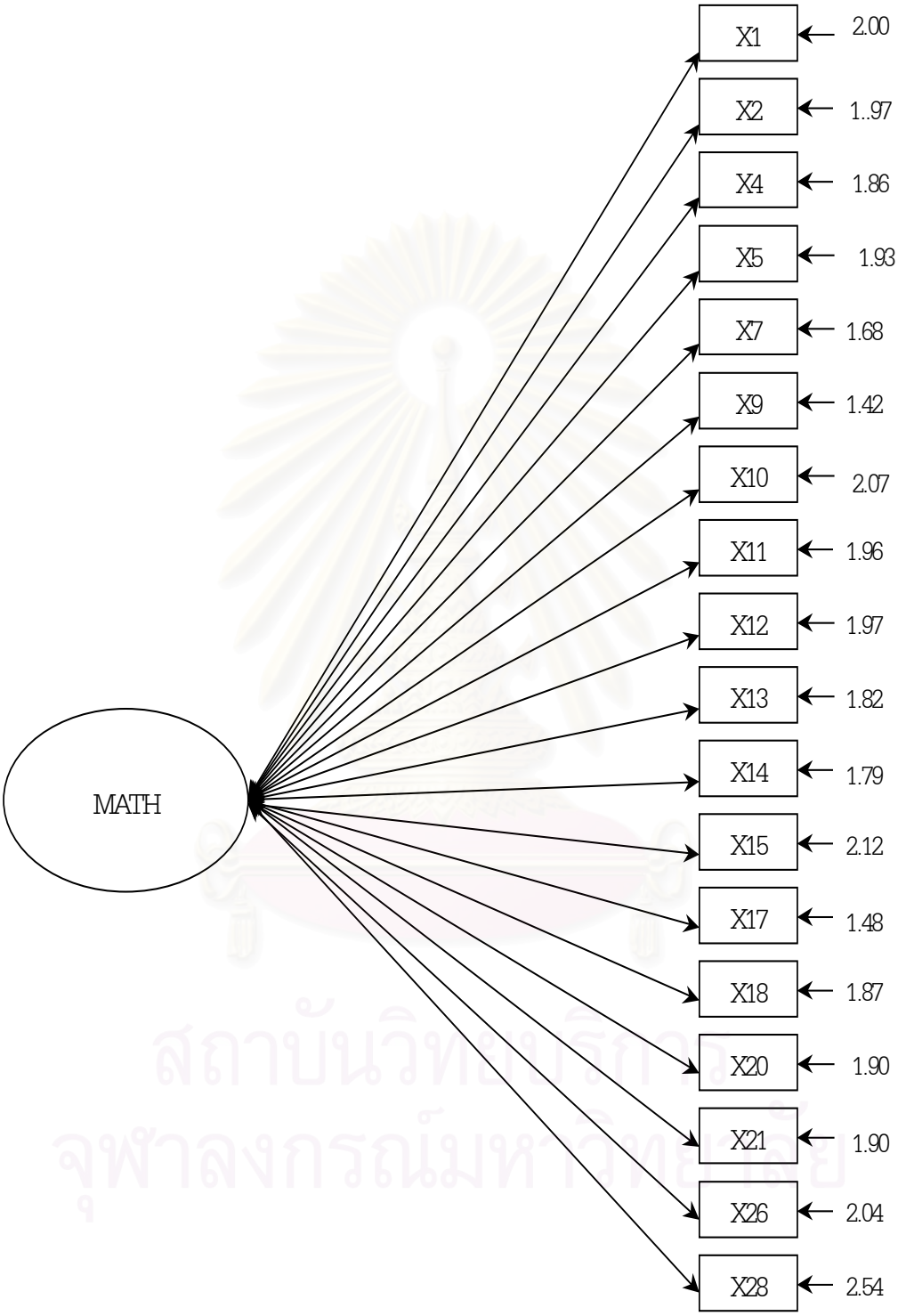
เมื่อพิจารณาผลการวิเคราะห์โมเดลความรู้วิชาคณิตศาสตร์ จากตารางที่ 55 และจากแผนภาพที่ 36 พบว่าน้ำหนักองค์ประกอบส่วนใหญ่ที่กำหนดข้อมูลจำเพาะของโมเดล CFA ไว้แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 และ 0.05 ยกเว้นตัวแปร X9, X11, X20 และ X26 ที่น้ำหนักองค์ประกอบไม่แตกต่างจากศูนย์อย่างนัยสำคัญ โดยน้ำหนักองค์ประกอบที่มีนัยสำคัญมีค่าอยู่ระหว่าง 0.11-0.63 และค่าสัมประสิทธิ์การพยากรณ์ (R-square) ของตัวแปร X13 มีค่าสูงสุดเท่ากับ 0.18

จากผลการวิเคราะห์องค์ประกอบเชิงยืนยันสามารถสรุปได้ว่า เมื่อแบ่งกลุ่มผู้สอบตามเพศ ของแบบสอบวิชาคณิตศาสตร์หลังการตัดข้อสอบที่พบ DIF ออกจากแบบสอบทุกข้อแล้ว มีความตรงเชิงโครงสร้างตามโมเดลความรู้วิชาคณิตศาสตร์

ตารางที่ 55 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความรู้วิชาคณิตศาสตร์ (MATH) หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

ตัวแปร	เมตริกซ์น้ำหนักองค์ประกอบ			R	สปส. คะแนนองค์ประกอบ
	สปส.	SE.	t	Square	
X1	0.22	0.06	3.52**	0.02	0.05
X2	0.32	0.06	5.02**	0.05	0.08
X4	0.60	0.07	9.12**	0.16	0.15
X5	0.17	0.06	2.83**	0.02	0.04
X7	0.34	0.06	5.71**	0.06	0.09
X9	0.02	0.05	0.46	0.00	0.01
X10	0.39	0.06	5.96**	0.07	0.09
X11	0.04	0.06	0.67	0.00	0.01
X12	0.15	0.06	2.42*	0.01	0.04
X13	0.63	0.07	9.16**	0.18	0.16
X14	0.53	0.06	8.39**	0.14	0.14
X15	0.31	0.06	4.76**	0.04	0.07
X17	0.11	0.05	2.10*	0.01	0.04
X18	0.45	0.06	7.09**	0.10	0.11
X20	0.07	0.06	1.09	0.00	0.02
X21	0.45	0.06	7.11**	0.10	0.11
X26	0.08	0.06	1.25	0.00	0.02
X28	0.22	0.07	3.12**	0.02	0.04
Chi-Square = 146.907		df = 135	P = 0.228	GFI = 0.984	
AGFI = 0.980		NFI = 0.666	RMR = 0.056		

*p < 0.05 , ** p < 0.01



แผนภาพที่ 36 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันโมเดลความรู้วิชาคณิตศาสตร์ หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

โมเดลความรู้วิชาคณิตศาสตร์ หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลความรู้วิชาคณิตศาสตร์ พบว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ดี พิจารณาได้จากค่าไค-สแควร์ (Chi-Square = 321.840) ซึ่งมีค่าความน่าจะเป็นเท่ากับ 0.127 ที่องศาอิสระเท่ากับ 295 (df = 295) นั่นคือค่าไค-สแควร์ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่ายอมรับสมมติฐานหลักที่ว่าโมเดลการวิจัยสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ส่วนค่าดัชนีวัดความกลมกลืน (GFI) มีค่าเท่ากับ 0.976 ค่าดัชนีวัดความกลมกลืนที่ปรับแก้แล้ว (AGFI) มีค่าเท่ากับ 0.972 และค่าดัชนีวัดความสอดคล้อง (NFI) มีค่าเท่ากับ 0.755 รวมทั้งค่าดัชนีรากกำลังสองเฉลี่ยของเศษ (RMR) มีค่าเท่ากับ 0.058 ซึ่งมีค่าน้อยจนเข้าใกล้ศูนย์ ดังรายละเอียดในตารางที่ 56 และรายละเอียดของโมเดลความรู้วิชาคณิตศาสตร์ สามารถแสดงได้ ดังแผนภาพที่ 37

เมื่อพิจารณาผลการวิเคราะห์โมเดลความรู้วิชาคณิตศาสตร์ จากตารางที่ 56 และจากแผนภาพที่ 37 พบว่าน้ำหนักองค์ประกอบส่วนใหญ่ที่กำหนดข้อมูลจำเพาะของโมเดล CFA ไว้แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 และ 0.05 ยกเว้นตัวแปร X9, X17, X20 และ X26 ที่น้ำหนักองค์ประกอบไม่แตกต่างจากศูนย์อย่างนัยสำคัญ โดยน้ำหนักองค์ประกอบที่มีนัยสำคัญมีค่าอยู่ระหว่าง 0.11-0.63 และค่าสัมประสิทธิ์การพยากรณ์ (R-square) ของตัวแปร X6 มีค่าสูงสุดเท่ากับ 0.30

จากผลการวิเคราะห์องค์ประกอบเชิงยืนยันสามารถสรุปได้ว่า เมื่อแบ่งกลุ่มผู้สอบตามเพศ ของแบบสอบวิชาคณิตศาสตร์หลังการตัดข้อสอบที่พบ DIF ออกจากแบบสอบบางข้อจนแบบสอบไม่ทำหน้าที่ต่างกันแล้ว แบบสอบมีความตรงเชิงโครงสร้างตามโมเดลความรู้วิชาคณิตศาสตร์

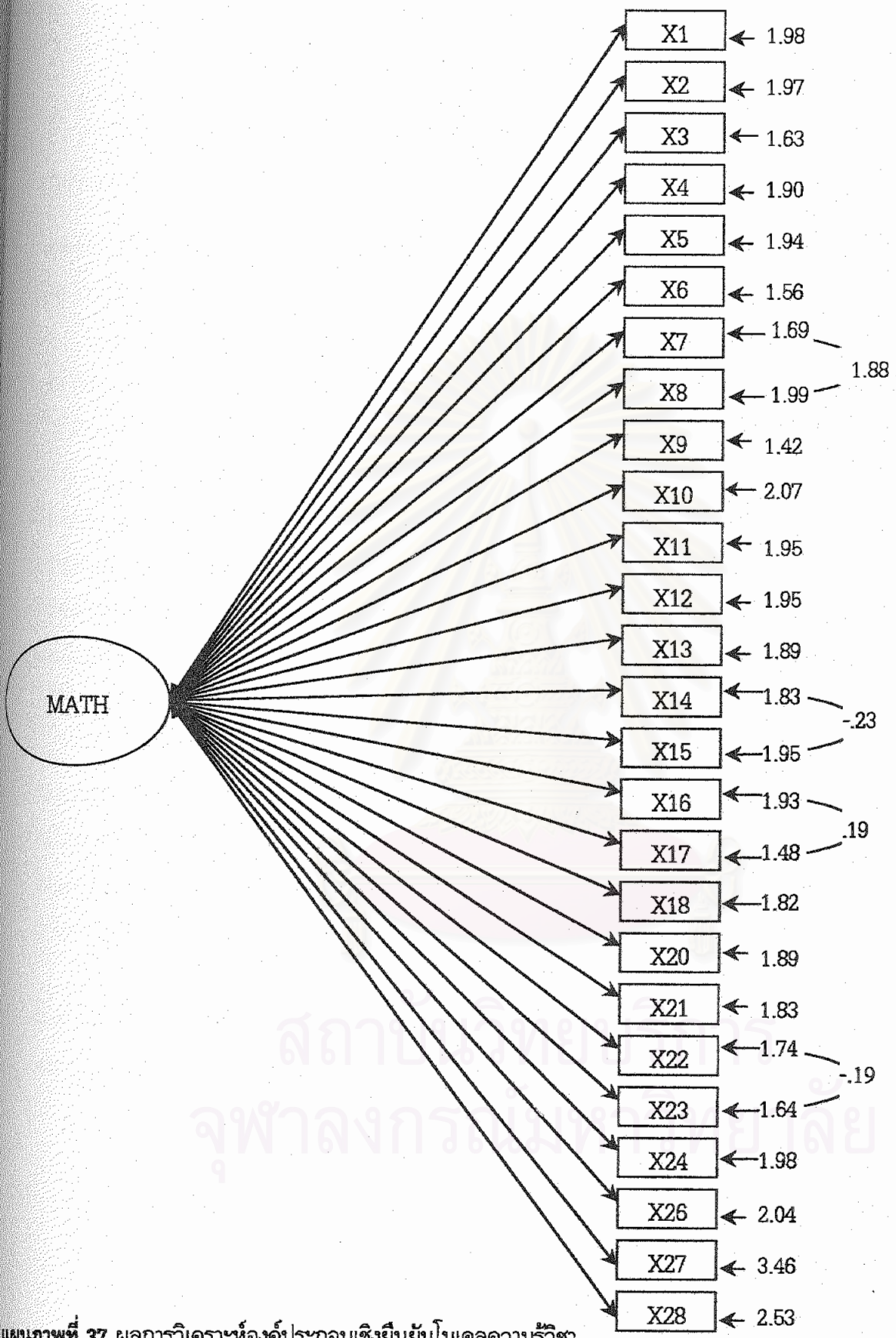
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 56 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความรู้วิชาคณิตศาสตร์ (MATH) หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

ตัวแปร	เมตริกซ์น้ำหนักองค์ประกอบ			R Square	สปส. คะแนนองค์ประกอบ
	สปส.	SE.	t		
X1	0.24	0.05	4.61**	0.03	0.03
X2	0.32	0.05	5.96**	0.05	0.05
X3	0.73	0.05	13.37**	0.25	0.16
X4	0.57	0.05	10.60**	0.15	0.08
X5	0.16	0.05	2.98**	0.01	0.02
X6	0.82	0.05	15.67**	0.30	0.15
X7	0.32	0.05	6.43**	0.06	0.05
X8	0.48	0.05	8.75**	0.10	0.06
X9	0.03	0.04	0.63	0.00	0.01
X10	0.38	0.05	6.96**	0.07	0.05
X11	0.11	0.05	2.07*	0.01	0.02
X12	0.19	0.05	3.55**	0.02	0.03
X13	0.57	0.05	10.68**	0.15	0.08
X14	0.50	0.05	9.43**	0.12	0.08
X15	0.29	0.05	5.35**	0.04	0.04
X16	0.33	0.05	6.30**	0.05	0.05
X17	0.07	0.04	1.51	0.00	0.01
X18	0.51	0.05	9.37**	0.12	0.10
X20	0.10	0.05	1.93	0.01	0.02
X21	0.53	0.05	10.04**	0.13	0.08
X22	0.67	0.05	12.30**	0.21	0.13
X23	0.39	0.05	7.86**	0.08	0.08
X24	0.30	0.05	5.63**	0.04	0.04
X26	0.07	0.05	1.21	0.00	0.01
X27	0.18	0.07	2.58**	0.01	0.01
X28	0.24	0.06	4.08**	0.02	0.03

Chi-Square = 321.840	df = 296	P = 0.127	GFI = 0.976
AGFI = 0.972	NFI = 0.755	RMR = 0.058	

*p < 0.05 , ** p < 0.01



แผนภาพที่ 37 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันโมเดลความรู้วิชาคณิตศาสตร์ หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

4.2.2.2 เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

โมเดลความรู้วิชาคณิตศาสตร์ หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งของโรงเรียนที่จบการศึกษา

ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลความรู้วิชาคณิตศาสตร์ พบว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ดี พิจารณาได้จากค่าไค-สแควร์ (Chi-Square = 231.138) ซึ่งมีค่าความน่าจะเป็นเท่ากับ 0.102 ที่องศาอิสระเท่ากับ 206 (df = 206) นั่นคือค่าไค-สแควร์ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่ายอมรับสมมติฐานหลักที่ว่าโมเดลการวิจัยสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ส่วนค่าดัชนีวัดความกลมกลืน (GFI) มีค่าเท่ากับ 0.980 ค่าดัชนีวัดความกลมกลืนที่ปรับแก้แล้ว (AGFI) มีค่าเท่ากับ 0.975 และค่าดัชนีวัดความสอดคล้อง (NFI) มีค่าเท่ากับ 0.751 รวมทั้งค่าดัชนีรากกำลังสองเฉลี่ยของเศษ (RMR) มีค่าเท่ากับ 0.059 ซึ่งมีค่าน้อยจนเข้าใกล้ศูนย์ดังรายละเอียดในตารางที่ 57 และรายละเอียดของโมเดลด้านความรู้วิชาคณิตศาสตร์ (MATH) สามารถแสดงได้ดังแผนภาพที่ 38

เมื่อพิจารณาผลการวิเคราะห์โมเดลความรู้วิชาคณิตศาสตร์ จากตารางที่ 57 และจากแผนภาพที่ 38 พบว่าน้ำหนักองค์ประกอบส่วนใหญ่ที่กำหนดข้อมูลจำเพาะของโมเดล CFA ไว้แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ยกเว้นตัวแปร X9, X11, X17 และ X26 ที่น้ำหนักองค์ประกอบไม่แตกต่างจากศูนย์อย่างนัยสำคัญ โดยน้ำหนักองค์ประกอบที่มีนัยสำคัญมีค่าอยู่ระหว่าง 0.15-0.82 และค่าสัมประสิทธิ์การพยากรณ์ (R-square) ของตัวแปร X6 มีค่าสูงสุดเท่ากับ 0.30

จากผลการวิเคราะห์องค์ประกอบเชิงยืนยันสามารถสรุปได้ว่า เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา แบบสอบวิชาคณิตศาสตร์หลังการตัดข้อสอบที่พบ DIF ออกจากแบบสอบทุกข้อแล้ว มีความตรงเชิงโครงสร้างตามโมเดลความรู้วิชาคณิตศาสตร์

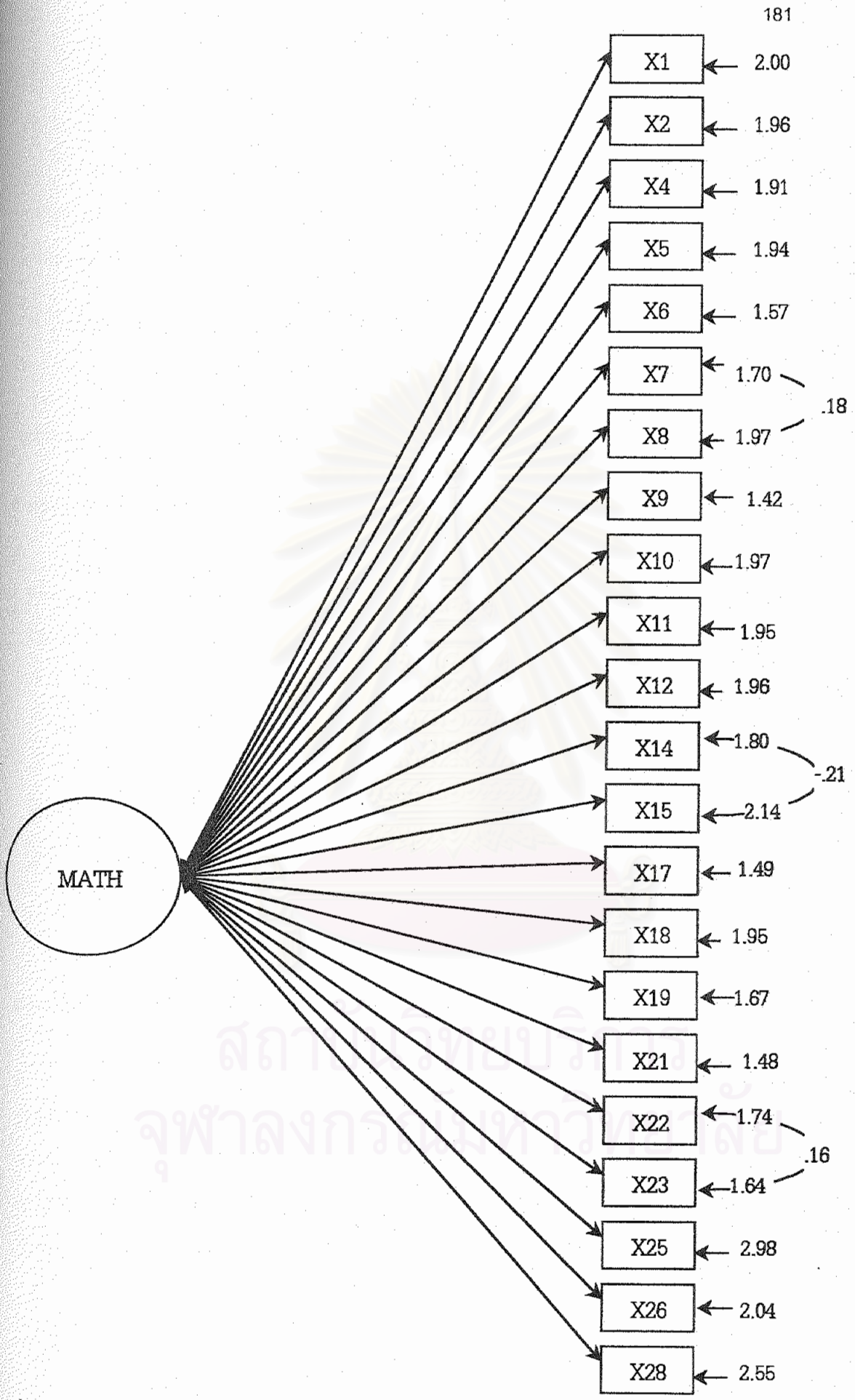
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 57 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลด้านความรู้วิชาคณิตศาสตร์ (MATH) หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

ตัวแปร	เมตริกซ์น้ำหนักองค์ประกอบ			R Square	สปส. คะแนนองค์ประกอบ
	สปส.	SE.	t		
X1	0.21	0.06	3.83**	0.02	0.04
X2	0.34	0.06	6.13**	0.06	0.06
X4	0.56	0.06	9.76**	0.14	0.10
X5	0.15	0.05	2.66**	0.01	0.03
X6	0.82	0.06	14.61**	0.30	0.18
X7	0.31	0.05	5.88**	0.05	0.05
X8	0.50	0.06	8.78**	0.11	0.08
X9	0.01	0.05	0.26	0.00	0.00
X10	0.38	0.06	6.51**	0.06	0.06
X11	0.10	0.05	1.91	0.01	0.02
X12	0.16	0.05	2.86**	0.01	0.03
X14	0.52	0.06	9.35**	0.13	0.11
X15	0.28	0.06	4.84**	0.04	0.05
X17	0.05	0.05	1.12	0.00	0.01
X18	0.49	0.05	8.80**	0.11	0.09
X19	0.15	0.05	2.86**	0.01	0.03
X21	0.55	0.05	9.94**	0.14	0.11
X22	0.67	0.06	12.01**	0.21	0.13
X23	0.40	0.05	7.59**	0.09	0.10
X25	0.39	0.07	5.72**	0.05	0.05
X26	0.03	0.06	0.56	0.00	0.00
X28	0.22	0.06	3.74**	0.02	0.03
Chi-Square = 231.138		df = 206	P = 0.102	GFI = 0.980	
AGFI = 0.975		NFI = 0.751	RMR = 0.059		

**p < 0.01

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



แผนภาพที่ 38 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันโมเดลความรู้วิชาคณิตศาสตร์ หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

ผลการเปรียบเทียบดัชนีความตรงจากการวิเคราะห์องค์ประกอบของแบบวิชาคณิตศาสตร์ก่อนและหลังการตัดข้อสอบที่พบ DIF จากตัวแปรเพศ และสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา มีรายละเอียดดังตารางที่ 58

ตารางที่ 58 ผลการเปรียบเทียบดัชนีความตรงจากการวิเคราะห์องค์ประกอบของแบบสอบวิชาคณิตศาสตร์ ก่อนและหลังการตัดข้อสอบที่พบ DIF จากตัวแปรเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

ดัชนีของความตรง	χ^2	df	RMR	P	GFI	AGFI	NFI
ก่อนการตัดข้อสอบที่พบ DIF	380.184	345	0.060	0.076	0.974	0.969	0.736
หลังการตัดข้อสอบที่พบ DIF							
จำแนกกลุ่มด้วยตัวแปรเพศ โดยใช้							
1. ดัชนี NCDIF	146.907	135	0.056	0.228	0.984	0.980	0.666
2. ดัชนี CDIF/DTF		295		0.127	0.976	0.972	0.755
	321.840		0.058				
จำแนกกลุ่มด้วยตัวแปรสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน โดยใช้							
1. ดัชนี NCDIF		206		0.102	0.980	0.975	0.751
2. ดัชนี CDIF/DTF		--		--	--	--	--
	231.138		0.059				
	--		--				

จากตารางที่ 58 ผลการเปรียบเทียบดัชนีความตรงจากการวิเคราะห์องค์ประกอบของแบบสอบวิชาคณิตศาสตร์ก่อนและหลังการตัดข้อสอบที่พบ DIF ในกรณีต่าง ๆ พบว่า ทุกกรณีเป็นโมเดลที่มีความสอดคล้องกับข้อมูลเชิงประจักษ์ ดัชนีวัดระดับความกลมกลืน (GFI) มีค่าเข้าใกล้ 1 ดัชนีวัดระดับความกลมกลืนที่ปรับแก้แล้ว (AGFI) มีค่าเข้าใกล้ 1 ดัชนีวัดระดับความสอดคล้องปกติ (NFI) มีค่าเข้าใกล้ 1 และดัชนีรากกำลังสองเฉลี่ยของเศษ (RMR) มีค่าเข้าใกล้ศูนย์

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

4.3 ค่าฟังก์ชันสารสนเทศของแบบสอบ

ผลการคำนวณค่าฟังก์ชันสารสนเทศของแบบสอบวิชาภาษาอังกฤษและวิชาคณิตศาสตร์

ทั้งในกรณีก่อนการตัดข้อสอบที่พบ DIF และหลังตัดข้อสอบที่พบ DIF โดยที่ค่าฟังก์ชันสารเทศของแบบสอบนี้คำนวณมาจากผลรวมของค่าฟังก์ชันสารสนเทศของข้อสอบทุกข้อในแบบสอบฉบับนั้น ๆ ที่ระดับความสามารถต่าง ๆ 13 ระดับ ผลการคำนวณมีรายละเอียดดังนี้

4.3.1 วิชาภาษาอังกฤษ

ผลการคำนวณค่าฟังก์ชันสารสนเทศของแบบสอบวิชาภาษาอังกฤษเมื่อแบ่งกลุ่มผู้สอบตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา มีรายละเอียดดังตารางที่ 59 - 60 และแผนภาพที่ 39 - 42

ตารางที่ 59 ค่าฟังก์ชันสารสนเทศของแบบสอบ (TIF) และค่าดัชนีประสิทธิภาพสัมพัทธ์ (RE) ของแบบสอบวิชาภาษาอังกฤษ ก่อนและหลังการตัดข้อสอบที่พบ DIF โดยใช้ดัชนี NCDIF และดัชนี CDIF/DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ ที่ระดับความสามารถต่าง ๆ

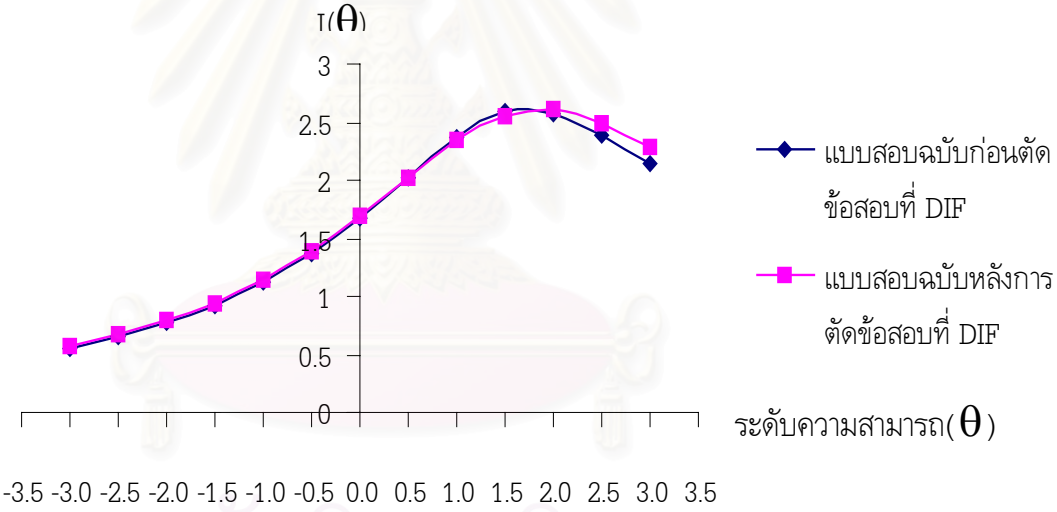
ระดับความสามารถ	ดัชนี NCDIF			ดัชนี CDIF/DTF		
	TIF		RE	TIF		RE
	ก่อนตัด	หลังตัด		ก่อนตัด	หลังตัด	
-3.0	0.556	0.565	1.016	0.754	0.640	0.849
-2.5	0.653	0.665	1.018	0.907	0.784	0.864
-2.0	0.769	0.795	1.034	1.114	0.987	0.886
-1.5	0.926	0.944	1.019	1.436	1.305	0.909
-1.0	1.120	1.149	1.026	1.950	1.832	0.939
-0.5	1.363	1.388	1.018	2.814	2.753	0.978
0.0	1.673	1.688	1.009	4.252	4.404	1.036
0.5	2.029	2.018	0.995	6.178	6.905	1.118
1.0	2.377	2.344	0.986	7.612*	8.871*	1.165
1.5	2.540	2.561	0.991	7.331	8.205	1.119
2.0	2.567*	2.614*	1.018	5.865	6.157	1.050
2.5	2.385	2.490	1.044	4.394	4.425	1.007
3.0	2.141	2.282	1.066	3.335	3.294	0.988

* หมายถึง ระดับความสามารถที่ TIF มีค่าสูงที่สุด

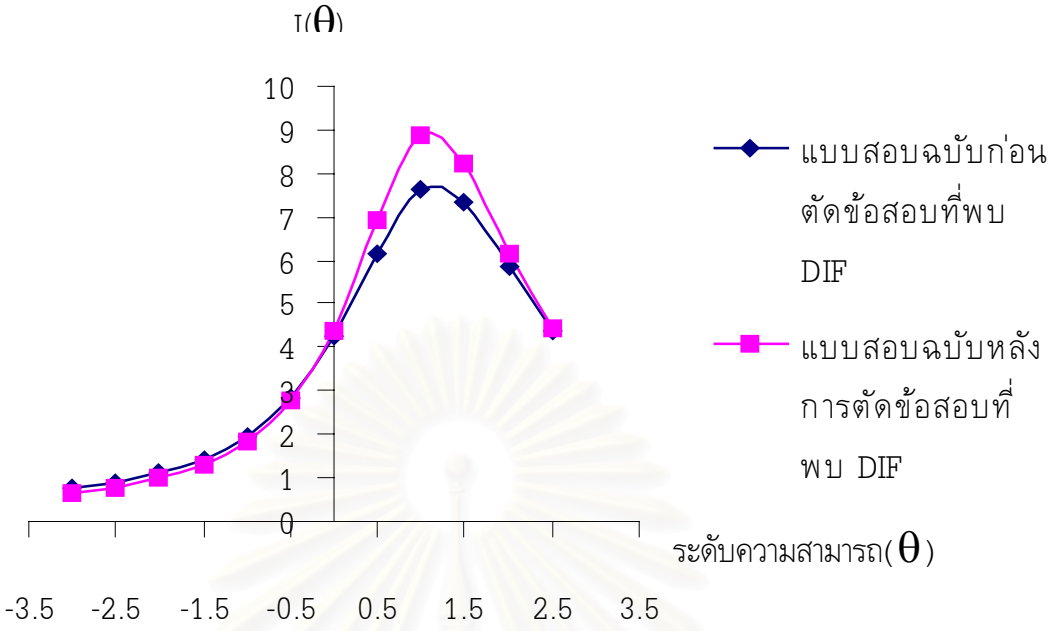
จากตารางที่ 59 พบว่าค่าสารสนเทศของแบบสอบวิชาภาษาอังกฤษ ฉบับก่อนและหลังการตัดข้อสอบที่พบ DIF ทุกข้อโดยใช้ดัชนี NCDIF มีค่าสูงสุดที่ระดับความสามารถ 2.0 เมื่อพิจารณาค่าดัชนีประสิทธิภาพสัมพัทธ์ของแบบสอบแล้วพบว่า แบบสอบฉบับหลังการตัดข้อสอบที่พบ DIF ทุกข้อมีประสิทธิภาพสูงกว่าแบบสอบฉบับก่อนการตัดข้อสอบที่พบ DIF ในช่วงระดับความสามารถ -3.0 ถึง 0.0 และ 2.0 ถึง 3.0

สำหรับค่าสารสนเทศของแบบสอบ ฉบับก่อนและหลังการตัดข้อสอบที่พบ DIF บางข้อโดยใช้ดัชนี CDIF/DTF มีค่าสูงสุดที่ระดับความสามารถ 1.0 เมื่อพิจารณาค่าดัชนีประสิทธิภาพสัมพัทธ์ของแบบสอบแล้วพบว่า แบบสอบฉบับหลังการตัดข้อสอบที่พบ DIF บางข้อ มีประสิทธิภาพสูงกว่าแบบสอบฉบับก่อนการตัดข้อสอบที่พบ DIF ในช่วงระดับความสามารถ 0.0 ถึง 2.5

แผนภาพที่ 39 โค้งฟังก์ชันสารสนเทศของแบบสอบวิชาภาษาอังกฤษ ฉบับก่อนและหลังการตัดข้อสอบที่พบ DIF ทุกข้อ โดยใช้ดัชนี NCDIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ



จากแผนภาพที่ 39 พบว่าแบบสอบวิชาภาษาอังกฤษฉบับก่อนและหลังการตัดข้อสอบที่พบ DIF ทุกข้อโดยใช้ดัชนี NCDIF มีโค้งของค่าฟังก์ชันสารสนเทศของแบบสอบคล้ายคลึงกันมาก แบบสอบทั้งสองฉบับนี้เหมาะสมกับผู้สอบที่มีความสามารถระดับสูง (θ = 1.5 ถึง 3.0) และพบว่าที่ระดับความสามารถ 2.0 ถึง 3.0 แบบสอบฉบับหลังการตัดข้อสอบที่พบ DIF ออกจากแบบสอบจะมีค่าฟังก์ชันสารสนเทศสูงกว่าฉบับก่อนการตัดข้อสอบที่พบ DIF



แผนภาพที่ 40 โค้งฟังก์ชันสารสนเทศของแบบสอบวิชาภาษาอังกฤษ ฉบับก่อนและหลังการตัดข้อสอบ
ที่พบ DIF บางข้อ โดยใช้ดัชนี CDIF/DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

จากแผนภาพที่ 40 พบว่าแบบสอบวิชาภาษาอังกฤษฉบับก่อนและหลังการตัดข้อสอบที่พบ DIF บางข้อโดยใช้ดัชนี CDIF/DTF มีโค้งของค่าฟังก์ชันสารสนเทศของแบบสอบคล้ายคลึงกันมาก แบบสอบทั้งสองฉบับนี้เหมาะสมกับผู้สอบที่มีความสามารถระดับค่อนข้างสูง (θ = 0.5 ถึง 2.5) และพบว่าที่ระดับความสามารถ 0.5 ถึง 2.5 แบบสอบฉบับหลังการตัดข้อสอบที่พบ DIF ออกจากแบบสอบจะมีค่าฟังก์ชันสารสนเทศสูงกว่าฉบับก่อนการตัดข้อสอบที่พบ DIF

ส่วนผลการวิเคราะห์ค่าฟังก์ชันสารสนเทศของแบบสอบวิชาภาษาอังกฤษ ฉบับก่อนและหลังการตัดข้อสอบที่พบ DIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา มีรายละเอียดดังตารางที่ 60 และแผนภาพที่ 41 และ 42

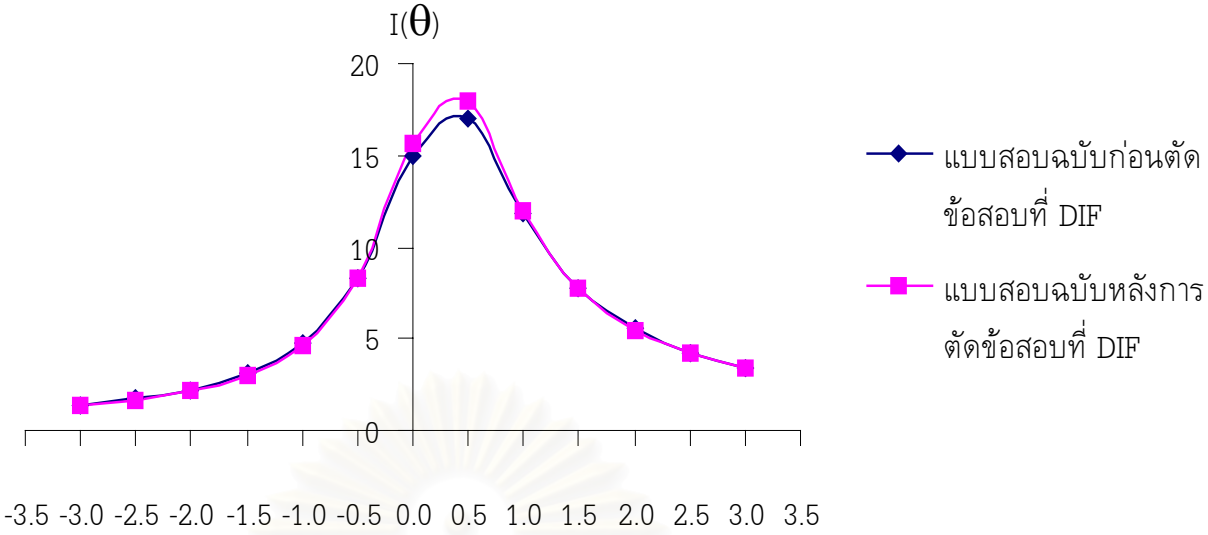
จากตารางที่ 60 พบว่าค่าสารสนเทศของแบบสอบวิชาภาษาอังกฤษ ฉบับก่อนและหลังการตัดข้อสอบที่พบ DIF ทุกข้อโดยใช้ดัชนี NCDIF มีค่าสูงสุดที่ระดับความสามารถ 0.5 เมื่อพิจารณาค่าดัชนีประสิทธิภาพสัมพัทธ์ของแบบสอบแล้วพบว่า แบบสอบฉบับหลังการตัดข้อสอบที่พบ DIF ทุกข้อมีประสิทธิภาพสูงกว่าแบบสอบฉบับก่อนการตัดข้อสอบที่พบ DIF ในช่วงระดับความสามารถ –0.5 ถึง 1.0

ตารางที่ 60 ค่าฟังก์ชันสารสนเทศของแบบสอบ (TIF) และค่าดัชนีประสิทธิภาพสัมพัทธ์ (RE) ของแบบสอบวิชาภาษาอังกฤษ ก่อนและหลังการตัดข้อสอบที่พบ DIF โดยใช้ดัชนี NCDIF และดัชนี CDIF/DTF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา ที่ระดับความสามารถต่าง ๆ

ระดับความสามารถ	ดัชนี NCDIF			ดัชนี CDIF/DTF		
	TIF		RE	TIF		RE
	ก่อนตัด	หลังตัด		ก่อนตัด	หลังตัด	
-3.0	1.354	1.300	0.960	1.591	1.516	0.953
-2.5	1.706	1.640	0.961	2.017	1.922	0.953
-2.0	2.237	2.157	0.964	2.656	2.547	0.959
-1.5	3.114	3.023	0.971	3.697	3.57	0.966
-1.0	4.762	4.664	0.979	5.604	5.491	0.980
-0.5	8.326	8.347	1.003	9.554	9.530	0.997
0.0	14.909	15.650	1.050	16.591	16.963	1.022
0.5	17.052*	18.008*	1.056	19.002*	19.473*	1.025
1.0	11.857	12.022	1.014	13.651	13.743	1.007
1.5	7.778	7.738	0.995	9.153	9.129	0.997
2.0	5.565	5.497	0.988	6.527	6.465	0.991
2.5	4.275	4.216	0.986	4.927	4.861	0.987
3.0	3.431	3.376	0.984	3.878	3.804	0.981

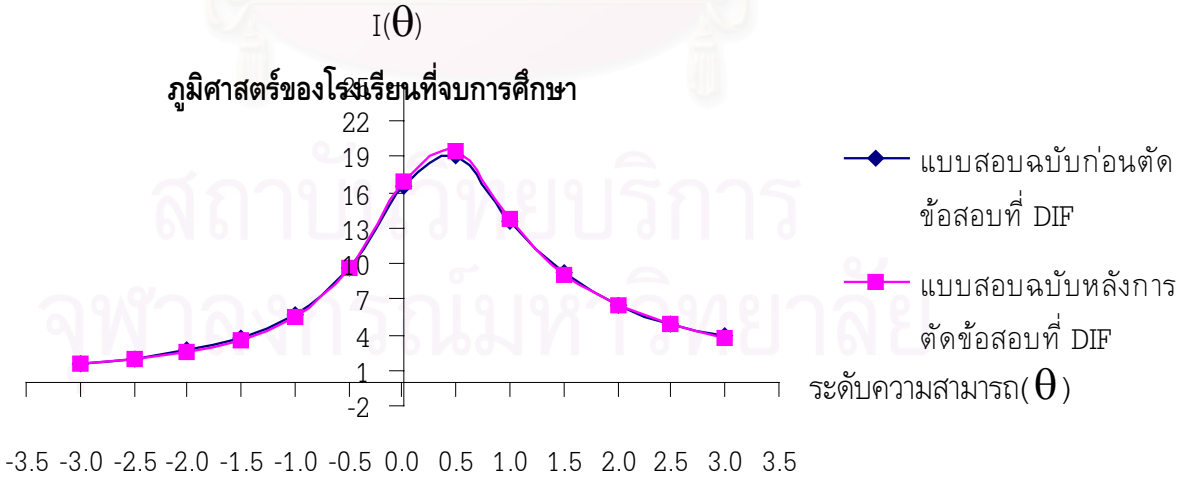
*หมายถึง ระดับความสามารถที่ TIF มีค่าสูงที่สุด

สำหรับค่าสารสนเทศของแบบสอบ ฉบับก่อนและหลังการตัดข้อสอบที่พบ DIF บางข้อโดยใช้ดัชนี CDIF/DTF มีค่าสูงสุดที่ระดับความสามารถ 0.5 เมื่อพิจารณาค่าดัชนีประสิทธิภาพสัมพัทธ์ของแบบสอบแล้วพบว่า แบบสอบฉบับหลังการตัดข้อสอบที่พบ DIF บางข้อ มีประสิทธิภาพสูงกว่าแบบสอบฉบับก่อนการตัดข้อสอบที่พบ DIF ในช่วงระดับความสามารถ 0.0 ถึง 1.0



แผนภาพที่ 41 โค้งฟังก์ชันสารสนเทศของแบบสอบวิชาภาษาอังกฤษ ฉบับก่อนและหลังการตัดข้อสอบที่พบ DIF ทุกข้อ โดยใช้ดัชนี NCDIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

จากแผนภาพที่ 41 พบว่าแบบสอบวิชาภาษาอังกฤษฉบับก่อนและหลังการตัดข้อสอบที่พบ DIF ทุกข้อโดยใช้ดัชนี NCDIF มีโค้งของค่าฟังก์ชันสารสนเทศของแบบสอบคล้ายคลึงกัน แบบสอบทั้งสองฉบับนี้เหมาะสมกับผู้สอบที่มีความสามารถระดับปานกลางค่อนข้างสูง (θ = 0.0 ถึง 0.5) และพบว่าที่ระดับความสามารถ -0.5 ถึง 1.0 แบบสอบฉบับหลังการตัดข้อสอบที่พบ DIF ออกจากแบบสอบจะมีค่าฟังก์ชันสารสนเทศสูงกว่าฉบับก่อนการตัดข้อสอบที่พบ DIF



แผนภาพที่ 42 โค้งฟังก์ชันสารสนเทศของแบบสอบวิชาภาษาอังกฤษ ฉบับก่อนและหลังการตัดข้อสอบที่พบ DIF บางข้อ โดยใช้ดัชนี CDIF/DTF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทาง

จากแผนภาพที่ 42 พบว่าแบบสอบวิชาภาษาอังกฤษฉบับก่อนและหลังการตัดข้อสอบที่พบ DIF บางข้อโดยใช้ดัชนี CDIF/DTF มีโค้งของค่าฟังก์ชันสารสนเทศของแบบสอบคล้ายคลึงกันมาก แบบสอบทั้งสองฉบับนี้เหมาะสมกับผู้สอบที่มีความสามารถระดับปานกลางค่อนข้างสูง (θ = 0.0 ถึง 0.5) และพบว่าที่ระดับความสามารถ 0.0 ถึง 1.0 แบบสอบฉบับหลังการตัดข้อสอบที่พบ DIF ออกจากแบบสอบจะมีค่าฟังก์ชันสารสนเทศสูงกว่าฉบับก่อนการตัดข้อสอบที่พบ DIF

4.3.2 วิชาคณิตศาสตร์

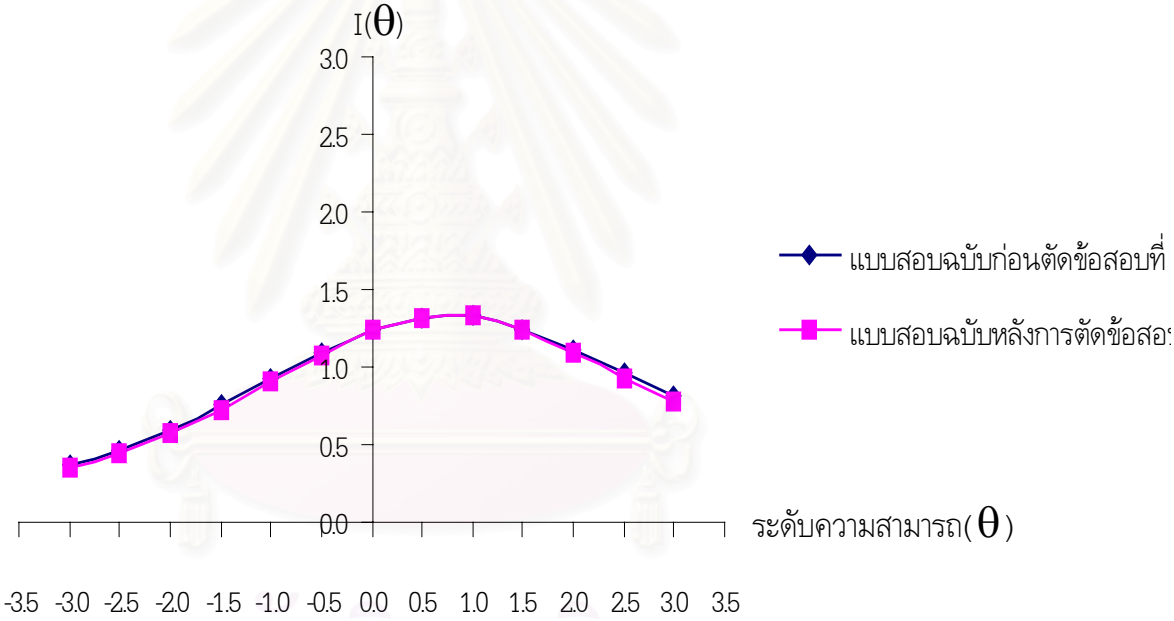
ผลการคำนวณค่าฟังก์ชันสารสนเทศของแบบสอบวิชาคณิตศาสตร์ เมื่อแบ่งกลุ่มผู้สอบตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา มีรายละเอียดดังตารางที่ 61 - 62 และแผนภาพที่ 43 – 45

ตารางที่ 61 ค่าฟังก์ชันสารสนเทศของแบบสอบ (TIF) และค่าดัชนีประสิทธิภาพสัมพัทธ์ (RE) ของแบบสอบวิชาคณิตศาสตร์ ก่อนและหลังการตัดข้อสอบที่พบ DIF โดยใช้ดัชนี NCDIF และดัชนี CDIF/DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ ที่ระดับความสามารถต่าง ๆ

ระดับความสามารถ	ดัชนี NCDIF			ดัชนี CDIF/DTF		
	TIF		RE	TIF		RE
	ก่อนตัด	หลังตัด		ก่อนตัด	หลังตัด	
-3.0	0.364	0.344	0.945	0.594	0.593	0.998
-2.5	0.468	0.443	0.947	0.810	0.810	1.000
-2.0	0.597	0.570	0.955	1.101	1.101	1.000
-1.5	0.752	0.724	0.963	1.474	1.473	0.999
-1.0	0.925	0.901	0.974	1.908	1.908	1.000
-0.5	1.101	1.082	0.983	2.330	2.329	1.000
0.0	1.242	1.238	0.997	2.611	2.617	1.002
0.5	1.324	1.324	1.000	2.663*	2.666*	1.001
1.0	1.326*	1.325*	0.999	2.480	2.477	0.999
1.5	1.247	1.240	0.994	2.157	2.158	1.000
2.0	1.115	1.095	0.982	1.797	1.800	1.002
2.5	0.963	0.931	0.967	1.468	1.473	1.003
3.0	0.809	0.776	0.959	1.186	1.192	1.005

* หมายถึง ระดับความสามารถที่ TIF มีค่าสูงที่สุด

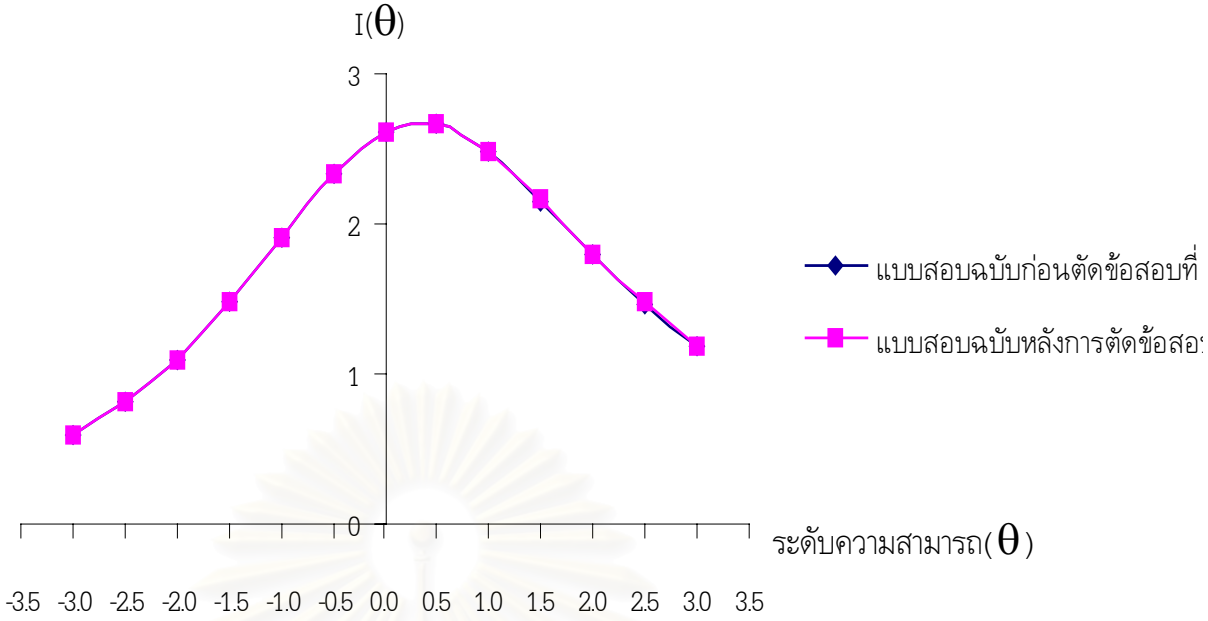
จากตารางที่ 61 พบว่าค่าสารสนเทศของแบบสอบวิชาคณิตศาสตร์ ฉบับก่อนและหลังการตัดข้อสอบที่พบ DIF ทุกข้อโดยใช้ดัชนี NCDIF มีค่าสูงสุดที่ระดับความสามารถ 1.0 เมื่อพิจารณาค่าดัชนีประสิทธิภาพสัมพัทธ์ของแบบสอบแล้วพบว่า แบบสอบฉบับหลังการตัดข้อสอบที่พบ DIF ทุกข้อมีประสิทธิภาพต่ำกว่าแบบสอบฉบับก่อนการตัดข้อสอบที่พบ DIF ในทุกระดับความสามารถ ยกเว้นที่ระดับ 0.5 ที่แบบสอบทั้งสองฉบับมีประสิทธิภาพเท่าเทียมกันสำหรับค่าสารสนเทศของแบบสอบ ฉบับก่อนและหลังการตัดข้อสอบที่พบ DIF บางข้อโดยใช้ดัชนี CDIF/DTF มีค่าสูงสุดที่ระดับความสามารถ 0.5 เมื่อพิจารณาค่าดัชนีประสิทธิภาพสัมพัทธ์ของแบบสอบแล้วพบว่า แบบสอบฉบับหลังการตัดข้อสอบที่พบ DIF บางข้อ มีประสิทธิภาพเกือบเท่าเทียมแบบสอบฉบับก่อนการตัดข้อสอบที่พบ DIF ในทุกระดับความสามารถ



แผนภาพที่ 43 โค้งฟังก์ชันสารสนเทศของแบบสอบวิชาคณิตศาสตร์ ฉบับก่อนและหลังการตัดข้อสอบ

ที่พบ DIF ทุกข้อ โดยใช้ดัชนี NCDIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

จากแผนภาพที่ 43 พบว่าแบบสอบวิชาคณิตศาสตร์ ฉบับก่อนและหลังการตัดข้อสอบที่พบ DIF ทุกข้อโดยใช้ดัชนี NCDIF มีโค้งของค่าฟังก์ชันสารสนเทศของแบบสอบคล้ายคลึงกันมาก แบบสอบทั้งสองฉบับนี้เหมาะสมกับผู้สอบที่มีความสามารถระดับค่อนข้างสูง (θ = 0.0 ถึง 1.5) และพบว่าที่ทุกระดับความสามารถ แบบสอบฉบับหลังการตัดข้อสอบที่พบ DIF ออกจากแบบสอบจะมีค่าฟังก์ชันสารสนเทศทัดเทียมกับแบบสอบฉบับก่อนการตัดข้อสอบที่พบ DIF



แผนภาพที่ 44 โค้งฟังก์ชันสารสนเทศของแบบสอบวิชาคณิตศาสตร์ ฉบับก่อนและหลังการตัดข้อสอบ
ที่พบ DIF บางข้อ โดยใช้ดัชนี CDIF/DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

จากแผนภาพที่ 44 พบว่าแบบสอบวิชาภาษาอังกฤษฉบับก่อนและหลังการตัดข้อสอบที่พบ DIF บางข้อโดยใช้ดัชนี CDIF/DTF มีโค้งของค่าฟังก์ชันสารสนเทศของแบบสอบคล้ายคลึงกันมาก แบบสอบทั้งสองฉบับนี้เหมาะสมกับผู้สอบที่มีความสามารถระดับค่อนข้างสูง (θ = 0.0 ถึง 1.5) และพบว่าที่เกือบทุกระดับความสามารถ แบบสอบฉบับหลังการตัดข้อสอบที่พบ DIF ออกจากแบบสอบจะมีค่าฟังก์ชันสารสนเทศสูงทัดเทียมกับแบบสอบฉบับก่อนการตัดข้อสอบที่พบ DIF

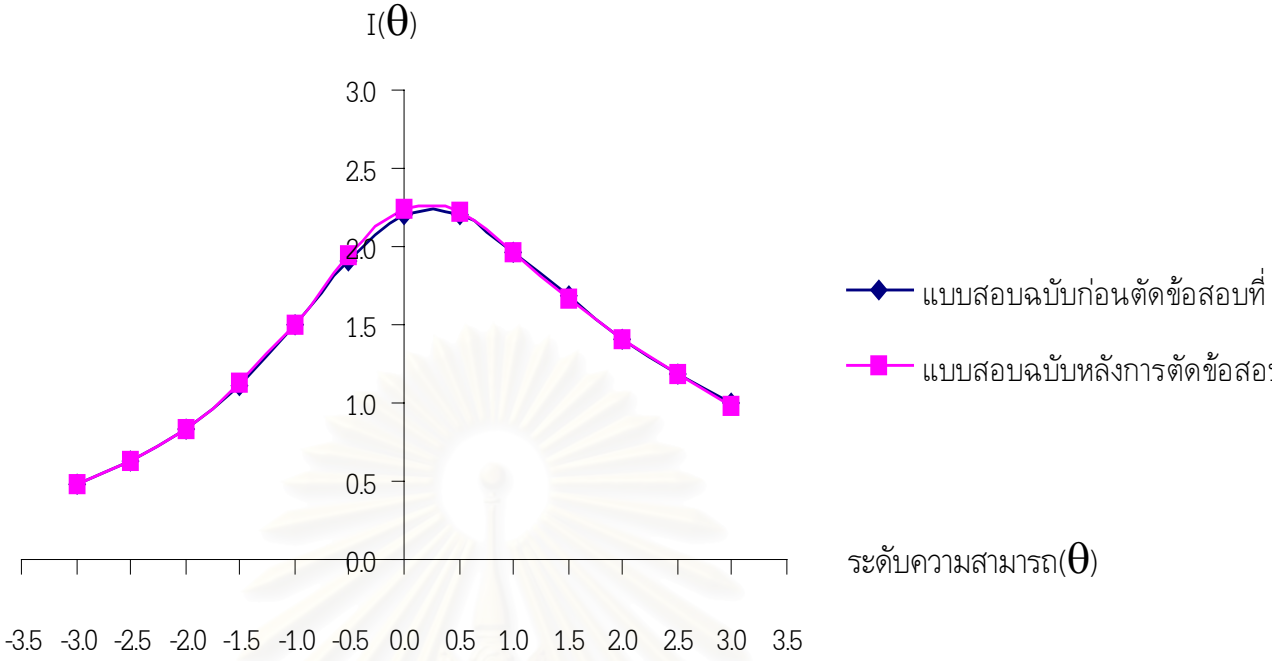
ผลการวิเคราะห์ค่าฟังก์ชันสารสนเทศของแบบสอบและค่าดัชนีประสิทธิภาพสัมพัทธ์ของแบบสอบวิชาคณิตศาสตร์ ก่อนและหลังการตัดข้อสอบที่พบ DIF โดยใช้ดัชนี NCDIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา ที่ระดับความสามารถต่าง ๆ มีรายละเอียดดังตารางที่ 62

ตารางที่ 62 ค่าฟังก์ชันสารสนเทศของแบบสอบ (TIF) และค่าดัชนีประสิทธิภาพสัมพัทธ์ (RE) ของแบบสอบวิชาคณิตศาสตร์ ก่อนและหลังการตัดข้อสอบที่พบ DIF โดยใช้ดัชนี NCDIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา ที่ระดับความสามารถต่าง ๆ

ระดับความสามารถ	ดัชนี NCDIF		
	TIF		RE
	ก่อนตัด	หลังตัด	
-3.0	0.473	0.473	1.000
-2.5	0.625	0.624	0.998
-2.0	0.833	0.835	1.002
-1.5	1.119	1.121	1.002
-1.0	1.492	1.505	1.009
-0.5	1.907	1.944	1.019
0.0	2.197*	2.246*	1.023
0.5	2.196	2.230	1.015
1.0	1.967	1.971	1.002
1.5	1.68	1.672	0.995
2.0	1.416	1.406	0.993
2.5	1.189	1.184	0.996
3.0	0.996	0.989	0.993

* หมายถึง ระดับความสามารถที่ TIF มีค่าสูงที่สุด

จากตารางที่ 62 พบว่าค่าสารสนเทศของแบบสอบวิชาคณิตศาสตร์ ฉบับก่อนและหลังการตัดข้อสอบที่พบ DIF ทุกข้อโดยใช้ดัชนี NCDIF มีค่าสูงสุดที่ระดับความสามารถ 0.0 เมื่อพิจารณาค่าดัชนีประสิทธิภาพสัมพัทธ์ของแบบสอบแล้วพบว่า ในช่วงระดับความสามารถ –2.0 ถึง 1.0 แบบสอบฉบับหลังการตัดข้อสอบที่พบ DIF ทุกข้อ มีประสิทธิภาพสูงกว่าแบบสอบฉบับก่อนการตัดข้อสอบที่พบ DIF ในเกือบทุกระดับความสามารถ



แผนภาพที่ 45 โค้งฟังก์ชันสารสนเทศของแบบสอบวิชาคณิตศาสตร์ ฉบับก่อนและหลังการตัดข้อสอบ

ที่พบ DIF ทุกข้อ โดยใช้ดัชนี NCDIF เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

จากแผนภาพที่ 45 พบว่าแบบสอบวิชาคณิตศาสตร์ ฉบับก่อนและหลังการตัดข้อสอบที่พบ DIF ทุกข้อโดยใช้ดัชนี NCDIF มีโค้งของค่าฟังก์ชันสารสนเทศของแบบสอบคล้ายคลึงกันมาก แบบสอบทั้งสองฉบับนี้เหมาะสมกับผู้สอบที่มีความสามารถระดับค่อนข้างสูง (θ = -0.5 ถึง 0.5) และพบว่าที่ระดับความสามารถ –0.5 ถึง 0.5 แบบสอบฉบับหลังการตัดข้อสอบที่พบ DIF ออกจากแบบสอบจะมีค่าฟังก์ชันสารสนเทศสูงกว่าแบบสอบฉบับก่อนการตัดข้อสอบที่พบ DIF

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตอนที่ 5 ผลการวิเคราะห์ความสัมพันธ์ระหว่างตำแหน่งของคะแนนรวมของผู้สอบก่อนและหลังการตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบสอบ

จากการวิเคราะห์ความสัมพันธ์ระหว่างตำแหน่งของคะแนนรวมของผู้สอบก่อนและหลังการตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบสอบวิชาภาษาอังกฤษและวิชาคณิตศาสตร์ ทั้งในกรณีที่ตัดข้อสอบทุกข้อและบางข้อที่ทำหน้าที่ต่างกัน ได้ผลการวิเคราะห์ดังตารางที่ 63 และ ตารางที่ 64

ตารางที่ 63 ผลการวิเคราะห์ความสัมพันธ์ระหว่างตำแหน่งของคะแนนรวมของผู้สอบก่อนและหลังการตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบสอบวิชาภาษาอังกฤษ

ตัวแปร	ค่าสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนแรงค์ (r_s)
1. เพศ กรณีตัดข้อสอบที่พบ DIF โดยใช้	
1.1 ดัชนี NCDIF	0.831**
1.2 ดัชนี CDIF/DTF	0.948**
2. พื้นที่ กรณีตัดข้อสอบที่พบ DIF โดยใช้	
2.1 ดัชนี NCDIF	0.976**
2.2 ดัชนี CDIF/DTF	0.988**

**p < 0.01

จากตารางที่ 63 ผลการวิเคราะห์ความสัมพันธ์ระหว่างตำแหน่งของคะแนนรวมของผู้สอบก่อนและหลังการตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบสอบวิชาภาษาอังกฤษ ทั้งกรณีที่ตัดข้อสอบที่ตรวจพบว่าทำหน้าที่ต่างกันออกจากแบบสอบทุกข้อ และกรณีที่ตัดเฉพาะบางข้อจนแบบสอบทั้งฉบับไม่ทำหน้าที่ต่างกัน พบว่าตำแหน่งของคะแนนรวมของผู้สอบก่อนและหลังการตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบสอบ มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ในทุกกรณี ซึ่งกรณีการตัด ข้อสอบที่พบการทำหน้าที่ต่างกันออกเป็นบางข้อ โดยการแบ่งกลุ่มผู้สอบตามตัวแปรสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน มีค่าสัมประสิทธิ์สหสัมพันธ์สูงที่สุดคือ 0.988 และกรณีการตัดข้อสอบที่พบการทำหน้าที่ต่างกันออกทุกข้อ โดยการแบ่งกลุ่มผู้สอบตามตัวแปรเพศ มีค่าสัมประสิทธิ์สหสัมพันธ์ต่ำที่สุดคือ 0.831

ตารางที่ 64 ผลการวิเคราะห์ความสัมพันธ์ระหว่างตำแหน่งของคะแนนรวมของผู้สอบก่อนและหลังการตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบสอบวิชาคณิตศาสตร์

ตัวแปร	ค่าสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนแรงค์ (r_s)
1. เพศ กรณีตัดข้อสอบที่พบ DIF โดยใช้	
1.1 ดัชนี NCDIF	0.872**
1.3 ดัชนี CDIF/DTF	0.976**
2. พื้นที่ กรณีตัดข้อสอบที่พบ DIF โดยใช้	
2.1 ดัชนี NCDIF	0.936**
2.2 ดัชนี CDIF/DTF	—

**p < 0.01

จากตารางที่ 64 ผลการวิเคราะห์ความสัมพันธ์ระหว่างตำแหน่งของคะแนนรวมของผู้สอบก่อนและหลังการตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบสอบวิชาคณิตศาสตร์ ทั้งกรณีที่ตัดข้อสอบที่ตรวจพบว่าทำหน้าที่ต่างกันออกจากแบบสอบทุกข้อ และกรณีที่ตัดเฉพาะบางข้อจนแบบสอบทั้งฉบับไม่ทำหน้าที่ต่างกัน พบว่าตำแหน่งของคะแนนรวมของผู้สอบก่อนและหลังการตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบสอบ มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ในทุกกรณี ซึ่งกรณีการตัดข้อสอบที่พบการทำหน้าที่ต่างกันออกเป็นบางข้อ โดยการแบ่งกลุ่มผู้สอบตามตัวแปรเพศ มีค่าสัมประสิทธิ์สหสัมพันธ์สูงที่สุดคือ 0.976 และกรณีการตัดข้อสอบที่พบการทำหน้าที่ต่างกันออกทุกข้อโดยการแบ่งกลุ่มผู้สอบตามตัวแปรเพศ มีค่าสัมประสิทธิ์สหสัมพันธ์ต่ำที่สุดคือ 0.872

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 5

สรุปผลการวิจัย อภิปรายผลและข้อเสนอแนะ

การวิจัยครั้งนี้มีวัตถุประสงค์ ดังต่อไปนี้

1. เพื่อวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบสำหรับแบบสอบคัดเลือกบุคคลเข้าศึกษาในระดับอุดมศึกษาวิชาภาษาอังกฤษและวิชาคณิตศาสตร์ สำหรับกลุ่มผู้สอบจำแนกตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

2. เพื่อเปรียบเทียบค่าความเที่ยง ความตรง และค่าฟังก์ชันสารสนเทศของแบบสอบ สำหรับแบบสอบก่อนและหลังจากตัดข้อสอบที่ทำหน้าที่ต่างกันในแต่ละกลุ่มผู้สอบออกจากแบบสอบ โดยเปรียบเทียบระหว่าง

2.1 แบบสอบก่อนและหลังจากตัดข้อสอบทุกข้อที่ทำหน้าที่ต่างกันออกจากแบบสอบ

2.2 แบบสอบก่อนและหลังจากตัดข้อสอบบางข้อที่ทำหน้าที่ต่างกันออกจากแบบสอบจนแบบสอบทั้งฉบับไม่ทำหน้าที่ต่างกัน

3. เพื่อวิเคราะห์ความสัมพันธ์ระหว่างตำแหน่งของคะแนนรวมของผู้สอบก่อนและหลังการตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบสอบ

กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้คือผู้สมัครสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษา ปีการศึกษา 2543 ครั้งที่ 1/มีนาคม วิชาภาษาอังกฤษ รหัสวิชา 03 จำนวน 4,000 คน และวิชาคณิตศาสตร์ 1 รหัสวิชา 04 จำนวน 3,600 คน

แบบสอบที่ใช้ในการศึกษาประกอบด้วยแบบสอบวิชาภาษาอังกฤษ สร้างโดยคณะกรรมการพัฒนาแบบสอบวิชาภาษาอังกฤษของสำนักทดสอบกลาง ทบวงมหาวิทยาลัย เป็นแบบสอบแบบปรนัย 4 ตัวเลือก จำนวน 100 ข้อ และแบบสอบวิชาคณิตศาสตร์ 1 สร้างโดยคณะกรรมการพัฒนาแบบสอบวิชาคณิตศาสตร์ 1 ของสำนักทดสอบกลาง ทบวงมหาวิทยาลัย ซึ่งในการวิจัยครั้งนี้ใช้เฉพาะส่วนที่เป็นแบบสอบแบบปรนัย 4 ตัวเลือก จำนวน 28 ข้อ

การวิเคราะห์ข้อมูล ทำการวิเคราะห์ข้อมูลเบื้องต้น โดยการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนสอบวิชาภาษาอังกฤษและวิชาคณิตศาสตร์ หลังจากนั้นทดสอบความเป็นเอกมิติของแบบสอบ ด้วยวิธีวิเคราะห์องค์ประกอบ (factor analysis) โดยใช้โปรแกรม SPSS 9.05 for Windows วิเคราะห์คุณภาพของแบบสอบด้านค่าความเที่ยงแบบความสอดคล้องภายใน (internal consistency) โดยใช้

โปรแกรม SPSS 9.05 for Windows วิเคราะห์ความตรงเชิงโครงสร้างของแบบสอบ (construct validity) ด้วยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) โดยใช้โปรแกรม LISREL 8.10 วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ (DFIT) และคำนวณค่าฟังก์ชันสารสนเทศของแบบสอบ (TIF) โดยใช้โปรแกรม BILOG 3.04 , EQUATE 2.0 และ Microsoft Excel

สรุปผลการวิจัย

ผลการวิจัยสรุปได้ดังนี้

1. ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ สำหรับแบบสอบวิชาภาษาอังกฤษและวิชาคณิตศาสตร์ โดยใช้กระบวนการ DFIT ได้ผลการวิเคราะห์ดังนี้

1.1 ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ โดยใช้ดัชนี NCDIF พบว่าแบบสอบวิชาภาษาอังกฤษ เมื่อจำแนกกลุ่มผู้สอบตามเพศ มีข้อสอบที่พบ DIF จำนวน 30 ข้อ คิดเป็นร้อยละ 30.00 ของจำนวนข้อสอบทั้งหมด และเมื่อจำแนกกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา มีข้อสอบที่พบ DIF จำนวน 16 ข้อ คิดเป็นร้อยละ 16.00 ของจำนวนข้อสอบทั้งหมด นอกจากนี้ในแบบสอบวิชาคณิตศาสตร์ เมื่อจำแนกกลุ่มผู้สอบตามเพศ มีข้อสอบที่พบ DIF จำนวน 10 ข้อ คิดเป็นร้อยละ 35.71 ของจำนวนข้อสอบทั้งหมด และเมื่อจำแนกกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา มีข้อสอบที่พบ DIF จำนวน 6 ข้อ คิดเป็นร้อยละ 21.43 ของจำนวนข้อสอบทั้งหมด

1.2 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบ โดยใช้ดัชนี CDIF/DTF พบว่าเมื่อจำแนกกลุ่มผู้สอบตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษาแล้วแบบสอบวิชาภาษาอังกฤษทั้งฉบับทำหน้าที่ต่างกัน หลังจากที่ตัดข้อสอบออกทีละข้อจนแบบสอบไม่ทำหน้าที่ต่างกัน พบว่าเมื่อจำแนกกลุ่มผู้สอบตามเพศ ต้องตัดข้อสอบออกจำนวน 13 ข้อ คิดเป็นร้อยละ 13.00 ของจำนวนข้อสอบทั้งหมด และมีข้อสอบ 7 ข้อ คิดเป็นร้อยละ 53.85 ของข้อสอบที่ถูกตัดออก เป็นข้อสอบที่ตรวจพบ DIF จากการวิเคราะห์การทำหน้าที่ต่างกันที่ระดับข้อสอบโดยใช้ดัชนี NCDIF ด้วย เมื่อจำแนกกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา ต้องตัดข้อสอบออกจำนวน 7 ข้อ คิดเป็นร้อยละ 7.00 ของจำนวนข้อสอบทั้งหมด และมีข้อสอบ 5 ข้อ คิดเป็นร้อยละ 71.43 ของจำนวนข้อสอบที่ถูกตัดออก เป็นข้อสอบที่ตรวจพบ DIF จากการวิเคราะห์การทำหน้าที่ต่างกันที่ระดับ

ข้อสอบโดยใช้ดัชนี NCDIF ด้วย นอกจากนี้สำหรับแบบสอบวิชาคณิตศาสตร์ เมื่อจำแนกกลุ่มผู้สอบตามเพศ พบว่าแบบสอบดังกล่าวทำหน้าที่ต่างกัน หลังจากที่ตัด ข้อสอบออกทีละข้อ จนแบบสอบไม่ทำหน้าที่ต่างกัน ต้องตัดข้อสอบออกจำนวน 2 ข้อ คิดเป็นร้อยละ 7.14 ของจำนวนข้อสอบทั้งหมดและข้อสอบทั้ง 2 ข้อนั้นเป็นข้อสอบที่พบ DIF จากการวิเคราะห์การทำหน้าที่ต่างกันที่ระดับข้อสอบโดยใช้ดัชนี NCDIF ด้วย แต่เมื่อจำแนกกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา พบว่าแบบสอบวิชาคณิตศาสตร์ ไม่ทำหน้าที่ต่างกัน

2. ผลการวิเคราะห์คุณภาพของแบบสอบวิชาภาษาอังกฤษและวิชาคณิตศาสตร์ ในด้านความเที่ยง ความตรง และค่าฟังก์ชันสารสนเทศของแบบสอบ ฉบับก่อนและหลังจากการตัดข้อสอบที่พบ DIF ออกจากแบบสอบในทุกเงื่อนไข สำหรับกลุ่มผู้สอบในแต่ละกลุ่ม สรุปได้ดังนี้

2.1 วิชาภาษาอังกฤษ

ความเที่ยงแบบความสอดคล้องภายในของแบบสอบหลังการตัดข้อสอบออกจากแบบสอบทั้งในกรณีการวิเคราะห์ DIF ซึ่งตัดข้อสอบที่พบ DIF ทุกข้อออกจากแบบสอบ และกรณีการวิเคราะห์ DTF ซึ่งตัดข้อสอบที่พบ DIF เฉพาะบางข้อจนแบบสอบไม่ทำหน้าที่ต่างกัน พบว่าค่าความเที่ยงของแบบสอบจะแตกต่างจากฉบับก่อนทำการวิเคราะห์อย่างมีนัยสำคัญทางสถิติ โดยมีค่าลดลงเล็กน้อย เกือบทุกกรณี ยกเว้นกรณีการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ พบว่าความเที่ยงของแบบสอบมีค่าเท่ากับแบบสอบฉบับก่อนการวิเคราะห์ และในกรณีการวิเคราะห์ DIF โดยตัดข้อสอบที่พบ DIF ทุกข้อ เมื่อแบ่งกลุ่มผู้สอบตามเพศ ทำให้ค่าความเที่ยงของแบบสอบมีค่าลดลงมากที่สุด

ในด้านความตรงเชิงโครงสร้างของแบบสอบวิชาภาษาอังกฤษ สำหรับแบบสอบฉบับก่อนตัดข้อสอบที่พบ DIF และหลังการตัดข้อสอบที่พบ DIF ในกรณีต่าง ๆ ทุกกรณี เมื่อพิจารณาค่าดัชนีความตรง พบว่าแบบสอบทุกฉบับมีความตรงเชิงโครงสร้างไม่แตกต่างกัน

ส่วนผลการเปรียบเทียบค่าฟังก์ชันสารสนเทศของแบบสอบ พบว่าเมื่อแบ่งกลุ่มผู้สอบตามเพศ แบบสอบวิชาภาษาอังกฤษฉบับหลังการตัดข้อสอบที่พบ DIF ทุกข้อ ให้ค่าฟังก์ชันสารสนเทศของแบบสอบสูงกว่าแบบสอบฉบับก่อนตัดข้อสอบ ที่ระดับความสามารถปานกลางลงมา ($\theta \leq 0.0$) และที่ระดับความสามารถสูงขึ้นไป ($\theta \geq 2.0$) และแบบสอบวิชาภาษาอังกฤษฉบับหลังการตัดข้อสอบที่พบ DIF บางข้อ จะให้ค่าฟังก์ชันสารสนเทศของแบบสอบสูงกว่าแบบสอบฉบับก่อนตัดข้อสอบ ที่ระดับความสามารถปานกลางขึ้นไป ($\theta \geq 0.0$) และเมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา แบบสอบวิชาภาษาอังกฤษฉบับหลังการตัดข้อสอบที่พบ DIF ทั้ง

ในกรณีที่ตัดทุกข้อและตัดเฉพาะบางข้อ ให้ค่าฟังก์ชันสารสนเทศของแบบสอบสูงกว่าแบบสอบฉบับก่อนการตัดข้อสอบ ที่ระดับความสามารถปานกลางค่อนข้างสูง ($0.0 \leq \theta \leq 1.0$)

2.2 วิชาคณิตศาสตร์

ค่าความเที่ยงแบบความสอดคล้องภายในของแบบสอบ เมื่อแบ่งกลุ่มผู้สอบตามเพศ พบว่า ความเที่ยงของแบบสอบฉบับหลังตัดข้อสอบที่พบ DIF ทุกข้อและตัดเฉพาะบางข้อจนแบบสอบไม่ทำหน้าที่ต่างกัน มีค่าแตกต่างจากค่าความเที่ยงของแบบสอบฉบับก่อนทำการวิเคราะห์ อย่างมีนัยสำคัญทางสถิติ โดยแบบสอบฉบับหลังตัดข้อสอบที่พบ DIF ทุกข้อ มีค่าความเที่ยงลดลง แต่แบบสอบฉบับหลังตัดข้อสอบที่พบ DIF เฉพาะบางข้อ มีค่าความเที่ยงของแบบสอบเพิ่มขึ้น ส่วนในกรณีเมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา พบว่าค่าความเที่ยงของแบบสอบฉบับหลังตัดข้อสอบที่พบ DIF ทุกข้อและตัดเฉพาะบางข้อ ไม่แตกต่างจากค่าความเที่ยงของแบบสอบฉบับก่อนทำการวิเคราะห์ อย่างมีนัยสำคัญทางสถิติ

ในด้านความตรงเชิงโครงสร้างของแบบสอบวิชาคณิตศาสตร์ สำหรับแบบสอบฉบับก่อนตัดข้อสอบที่พบ DIF และหลังการตัดข้อสอบที่พบ DIF ในกรณีต่าง ๆ ทุกกรณี เมื่อพิจารณาค่าดัชนีความตรง พบว่าแบบสอบทุกฉบับมีความตรงเชิงโครงสร้างไม่แตกต่างกัน

ส่วนผลการเปรียบเทียบค่าฟังก์ชันสารสนเทศของแบบสอบ พบว่าเมื่อแบ่งกลุ่มผู้สอบตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา แบบสอบวิชาคณิตศาสตร์ฉบับหลังตัดข้อสอบที่พบ DIF ทั้งในกรณีที่ตัดทุกข้อและตัดเฉพาะบางข้อ ให้ค่าฟังก์ชันสารสนเทศของแบบสอบใกล้เคียงกันมากกับแบบสอบฉบับก่อนตัดข้อสอบ ในทุกระดับความสามารถ พิจารณาได้จากแผนภาพที่ 43-45 ที่เส้นกราฟของค่าฟังก์ชันสารสนเทศของแบบสอบทั้งสองฉบับทับกันเกือบพอดี ที่ทุกระดับความสามารถ

3. ในการวิเคราะห์ความสัมพันธ์ระหว่างตำแหน่งของคะแนนรวมของผู้สอบก่อนและหลังการตัดข้อสอบที่พบ DIF ของแบบสอบวิชาภาษาอังกฤษและวิชาคณิตศาสตร์ โดยการหาค่าสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนแรงค์ พบว่าตำแหน่งของคะแนนรวมของผู้สอบก่อนและหลังการตัดข้อสอบ ในทุกกรณีมีความสัมพันธ์ในทางบวกซึ่งกันและกันอย่างมีนัยสำคัญทางสถิติ สำหรับแบบสอบวิชาภาษาอังกฤษ ค่าสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนแรงค์ มีค่าอยู่ระหว่าง 0.83-0.99 โดยเมื่อแบ่งกลุ่มผู้สอบตามเพศ พบว่าตำแหน่งของคะแนนรวมของผู้สอบก่อนและหลังการตัดข้อสอบที่พบ DIF ทุกข้อจำนวน 30 ข้อ มีค่าสัมประสิทธิ์สหสัมพันธ์ต่ำสุด คือมีค่าเท่ากับ 0.83 และเมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา พบว่าว่าตำแหน่งของคะแนนรวมของผู้สอบก่อน

และหลังการตัดข้อสอบที่พบ DIF บางข้อ จำนวน 7 ข้อ มีค่าสัมประสิทธิ์สหสัมพันธ์สูงสุด คือมีค่าเท่ากับ 0.99 ส่วนในแบบสอบวิชาคณิตศาสตร์ ค่าสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนแรงค์ มีค่าอยู่ระหว่าง 0.87-0.98 โดยเมื่อแบ่งกลุ่มผู้สอบตามเพศ พบว่าตำแหน่งของคะแนนรวมของผู้สอบก่อนและหลังการตัดข้อสอบที่พบ DIF ทุกข้อ จำนวน 10 ข้อ มีค่าสัมประสิทธิ์สหสัมพันธ์ต่ำสุด คือมีค่าเท่ากับ 0.87 และเมื่อแบ่งกลุ่มผู้สอบตามเพศ พบว่าว่าตำแหน่งของคะแนนรวมของผู้สอบก่อนและหลังการตัดข้อสอบที่พบ DIF บางข้อ จำนวน 2 ข้อ มีค่าสัมประสิทธิ์สหสัมพันธ์สูงสุด คือมีค่าเท่ากับ 0.98

อภิปรายผลการวิจัย

จากผลการวิจัยครั้งนี้ มีประเด็นสำคัญที่ควรนำมาอภิปรายดังต่อไปนี้

1. ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ โดยใช้กระบวนการ DFIT พบว่าในทุกเงื่อนไขข้อสอบที่ถูกระบุว่าทำหน้าที่ต่างกันและถูกตัดออกจากแบบสอบ โดยใช้ดัชนี CDIF/DTF มีจำนวนน้อยกว่าข้อสอบที่ถูกระบุว่าทำหน้าที่ต่างกันโดยใช้ดัชนี NCDIF ในวิชาภาษาอังกฤษ เมื่อแบ่งกลุ่มผู้สอบตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา ดัชนี CDIF/DTF ระบุข้อที่ DIF และถูกตัดออกเพียง ร้อยละ 13.00 และ 7.00 ตามลำดับ ขณะที่ดัชนี NCDIF ระบุข้อที่ DIF ถึงร้อยละ 30.00 และ 16.00 ตามลำดับ ในวิชาคณิตศาสตร์ เมื่อแบ่งกลุ่มผู้สอบตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา ดัชนี CDIF/DTF ระบุข้อที่ DIF และถูกตัดออกเพียง ร้อยละ 7.14 และ 0.00 ตามลำดับ ขณะที่ดัชนี NCDIF ระบุข้อที่ DIF ถึงร้อยละ 35.71 และ 21.43 ตามลำดับ ซึ่งสอดคล้องกับผลการวิจัยของ Ellis และ Mead (2000) ที่ได้ทำการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ สำหรับแบบสอบถาม 16PF (The Sixteen Personality Factor) ที่แปลเป็นภาษาสเปน ผลการวิจัยพบว่า ดัชนี CDIF/DTF ระบุข้อสอบที่พบ DIF และถูกตัดออก มีจำนวนน้อยกว่าดัชนี NCDIF โดยดัชนี CDIF/DTF ระบุข้อสอบที่ถูกตัดออกเพียงประมาณร้อยละ 17.00 ในขณะที่ดัชนี NCDIF ระบุข้อสอบที่ถูกตัดออกถึงร้อยละ 50.00 การตัดข้อสอบที่พบ DIF ทุกข้อออกจากแบบสอบ ในทางปฏิบัติแล้วไม่สามารถทำได้ เนื่องจากอาจทำให้แบบสอบขาดความตรงตามเนื้อหาได้ ดังนั้นดัชนี CDIF/DTF จึงมีประโยชน์ในการลดจำนวนข้อสอบที่ต้องถูกตัดออก หรือข้อสอบที่ต้องนำกลับมาปรับปรุงแก้ไข ข้อสอบที่พบ DIF เป็นการให้ข้อมูลเกี่ยวกับโอกาสในการตอบข้อสอบได้ถูกต้องเปรียบเทียบระหว่างกลุ่มย่อย เพื่อนำไปประกอบการพัฒนาแบบสอบให้มีคุณภาพดีขึ้น การเลือกใช้ดัชนี CDIF/DTF หรือ NCDIF ในการตัดข้อสอบที่ DIF ขึ้นอยู่

กับวัตถุประสงค์ ดัชนี CDIF/DTF มีประโยชน์ในการตรวจสอบคุณภาพของแบบสอบทั้งฉบับในด้านความยุติธรรม ในขณะที่ดัชนี NCDIF มีประโยชน์ในการตรวจสอบคุณภาพของข้อสอบรายข้อ ถ้าต้องการคัดเลือกข้อสอบที่มีคุณภาพเพื่อเก็บในคลังข้อสอบ ดัชนี NCDIF มีความเหมาะสมมากกว่า

2. ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ ระหว่างกลุ่มผู้สอบเมื่อจำแนกตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา พบว่าเมื่อจำแนกกลุ่มผู้สอบตามเพศจะมีข้อสอบที่ถูกระบุว่าเกิด DIF และถูกตัดออกมากกว่าการจำแนกกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษาของผู้สอบ ทั้งในแบบสอบวิชาภาษาอังกฤษและวิชาคณิตศาสตร์ ซึ่งสอดคล้องกับผลการวิจัยของ เกษร หว่างจิตร (2539) ที่ได้วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบในแบบสอบคัดเลือกบุคคลเข้าศึกษาในระดับบัณฑิตศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย วิชาภาษาไทยและวิชาภาษาอังกฤษ ที่พบว่าเมื่อจำแนกกลุ่มผู้สอบตามเพศพบว่ามีข้อสอบที่ระบุว่าเกิด DIF มากที่สุด รองลงมาคือ ภูมิลำเนา สังกัดสถานศึกษา ตามลำดับ ทั้งนี้อาจเป็นเพราะว่าเพศเป็นตัวแปรที่มีความสัมพันธ์ค่อนข้างสูงกับผลสัมฤทธิ์ทางการเรียนด้านภาษาและทักษะการคำนวณ เช่นผลการวิจัยของ นิฏฐา กัลยาประสิทธิ์ (2539) ที่ได้ศึกษาเปรียบเทียบความรู้ความสามารถ ในการเขียนภาษาอังกฤษเชิงอธิบายของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่มีเพศต่างกัน พบว่านักเรียนเพศหญิง มีทั้งความรู้และความสามารถในการเขียนภาษาอังกฤษเชิงอธิบายสูงกว่าเพศชาย และจากข้อค้นพบของ Giray (1995) และ Katherine (1996) ที่พบว่าเพศชายจะได้เปรียบในการทำข้อสอบเกี่ยวกับการคำนวณและพีชคณิต ในขณะที่ผู้หญิงจะได้เปรียบในการทำข้อสอบเกี่ยวกับคำศัพท์และเรขาคณิต ซึ่งสอดคล้องกับคำกล่าวของ Maier และ Casselman (1970 อ้างถึงใน Peter และ Dawin, 1994) ที่กล่าวว่าเพศชายจะมีความสามารถในการแก้ปัญหาทางคณิตศาสตร์ ส่วนเพศหญิงจะมีความสามารถในทางภาษา

3. จากผลการเปรียบเทียบค่าความเที่ยง ความตรง และค่าฟังก์ชันสารสนเทศของแบบสอบสำหรับแบบสอบก่อนและหลังจากตัดข้อสอบที่ทำหน้าที่ต่างกันในแต่ละกลุ่มผู้สอบออกจากแบบสอบ มีประเด็นที่ควรนำมาอภิปรายดังนี้

3.1 ผลการวิเคราะห์คุณภาพของแบบสอบในด้านความเที่ยงของแบบสอบวิชาภาษาอังกฤษและวิชาคณิตศาสตร์ ฉบับก่อนและหลังการตัดข้อสอบที่พบ DIF ในแต่ละกรณี พบว่าส่วนใหญ่ค่าความเที่ยงของแบบสอบแตกต่างกันอย่างมีนัยสำคัญ โดยภาพรวมจากผลการวิเคราะห์ค่าความเที่ยงของแบบสอบทั้งสองวิชา พบว่าถ้าตัดข้อสอบออกจากแบบสอบมากขึ้นเท่าใดจะทำให้ค่าความเที่ยงของ

แบบสอบฉบับนั้นยิ่งมีค่าลดลง ถึงแม้ว่าจะทำการปรับขยายให้มีจำนวนข้อสอบเท่าเดิมแล้วก็ตาม อาจเป็นเพราะว่าการตัดข้อสอบออกจากแบบสอบนั้น อาจตัดข้อที่มีผลทำให้การกระจายของคะแนนสูงออกไป ทำให้ความแปรปรวนของคะแนนที่สังเกตได้ลดลง จึงส่งผลให้ความเที่ยงของแบบสอบมีค่าลดลง

3.2 ส่วนค่าความตรงเชิงโครงสร้างของแบบสอบ ฉบับก่อนและหลังการตัดข้อสอบที่พบ DIF ในกลุ่มของผู้สอบที่จำแนกตามตัวแปรแต่ละตัวแปรออกจากแบบสอบ พบว่าดัชนีค่าความตรงเชิงโครงสร้างของแบบสอบฉบับก่อนการวิเคราะห์และฉบับหลังการตัดข้อสอบที่พบ DIF มีค่าไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ผลสอดคล้องกันทั้งในวิชาภาษาอังกฤษและวิชาคณิตศาสตร์ การที่แบบสอบฉบับหลังตัดข้อสอบที่พบ DIF ออกจากแบบสอบแล้วแบบสอบยังมีความตรงเชิงโครงสร้าง แต่ความเที่ยงของแบบสอบมีค่าลดลง อาจอธิบายได้ว่าจากนิยามความตรงคือคุณสมบัติด้านการวัดได้ตรงตามคุณลักษณะที่มุ่งวัด สามารถประมาณค่าได้จากสัดส่วนความแปรปรวนของคะแนนจริงที่ตรงประเด็นกับลักษณะที่มุ่งวัด จากนิยามความตรง ดังนั้นความแปรปรวนของคะแนนที่สังเกตได้ $\left(\sigma_X^2\right)$ จึงประกอบด้วยความแปรปรวนของคะแนนที่เที่ยงตรงประเด็นตามคุณลักษณะที่มุ่งวัด $\left(\sigma_V^2\right)$ ความแปรปรวนของคะแนนที่เที่ยงแต่ไม่ตรงประเด็นตามลักษณะที่มุ่งวัด $\left(\sigma_I^2\right)$ และความแปรปรวนของคะแนนความคลาดเคลื่อน $\left(\sigma_E^2\right)$ (ศิริชัย กาญจนวาสี, 2544) การตัดข้อสอบที่ DIF ออก อาจทำให้ความแปรปรวนของคะแนนที่สังเกตได้มีค่าลดลงหรือความแปรปรวนของคะแนนความคลาดเคลื่อนมีค่าสูงขึ้น ซึ่งจะทำให้สัดส่วนความแปรปรวนที่เที่ยงต่ำลง $\left(\sigma_T^2 / \sigma_X^2\right)$ จึงทำให้ค่าความเที่ยงของแบบสอบต่ำลง เนื่องจากความแปรปรวนที่เที่ยงมีทั้งความแปรปรวนของส่วนที่ตรงประเด็นตามลักษณะที่มุ่งวัดกับส่วนที่ไม่ตรงประเด็น ดังนั้นสัดส่วนความแปรปรวนที่เที่ยงที่ลดลงนั้น อาจมาจากการลดลงของความแปรปรวนในส่วนที่ไม่ตรงประเด็นตามลักษณะที่มุ่งวัด จึงทำให้แบบสอบมีค่าความตรงสูงขึ้น ในขณะที่ค่าความเที่ยงมีค่าลดลง

3.3 สำหรับผลการเปรียบเทียบค่าฟังก์ชันสารสนเทศของแบบสอบ เมื่อแบ่งกลุ่มผู้สอบตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษาของผู้สอบ หลังการตัดข้อสอบที่พบ DIF ทุกข้อและเฉพาะบางข้อในแบบสอบวิชาภาษาอังกฤษ พบว่าแบบสอบจะให้ค่าฟังก์ชันสารสนเทศของแบบสอบสูงกว่าแบบสอบฉบับก่อนตัดข้อสอบค่อนข้างชัด ที่ระดับความสามารถปานกลางขึ้นไป สาเหตุที่ทำให้ค่าฟังก์ชันสารสนเทศของแบบสอบฉบับก่อนตัดข้อสอบต่ำกว่า อาจเนื่องมาจากยังมีความคลาดเคลื่อนที่เกิดจากการทำหน้าที่ต่างกันของข้อสอบรวมอยู่ ซึ่งข้อสอบกลุ่มนั้น ๆ อาจเข้าข้างกลุ่มใดกลุ่มหนึ่งทำให้แบบสอบขาดความแม่นยำในการประมาณค่าความสามารถของผู้สอบและก่อให้เกิดความคลาดเคลื่อนในการวัด เนื่องจากค่าสารสนเทศของแบบสอบมีความสัมพันธ์ผกผันกับ

ความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถ ความคลาดเคลื่อนในการวัดยิ่งมีมากขึ้น สารสนเทศในการประมาณค่าความสามารถของผู้สอบก็ยิ่งลดลง (Hambleton and Swaminathan, 1985) ดังนั้นการตัดข้อสอบที่ DIF ออกจากแบบสอบ ซึ่งข้อสอบเหล่านี้อาจเป็นสาเหตุที่ทำให้เกิดความคลาดเคลื่อนในการวัด จึงทำให้ค่าฟังก์ชันสารสนเทศของแบบสอบฉบับหลังตัดข้อสอบที่ DIF มีค่าสูงขึ้น

จากผลการเปรียบเทียบค่าความเที่ยง ความตรง และค่าฟังก์ชันสารสนเทศของแบบสอบฉบับก่อนและหลังตัดข้อสอบที่พบ DIF จะเห็นได้ว่าผลการเปรียบเทียบค่าฟังก์ชันสารสนเทศของแบบสอบทั้งสองฉบับ มีความแตกต่างกันค่อนข้างชัดเจนกว่าผลการเปรียบเทียบค่าความเที่ยงและความตรง ซึ่งค่าฟังก์ชันสารสนเทศของแบบสอบ น่าจะเป็นดัชนีวัดคุณภาพที่มีความถูกต้องและเหมาะสมกว่า เนื่องจากว่าอยู่บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ (IRT) เช่นเดียวกันกับกระบวนการ ดี เอฟ ไอ ที ซึ่งเป็นกระบวนการที่ใช้ในการวิเคราะห์ DIF และ DTF ในการวิจัยครั้งนี้ นอกจากนี้จากผลการเปรียบเทียบคุณภาพของแบบสอบทั้งสามดัชนียังเห็นว่าการตัดข้อสอบทุกข้อที่พบ DIF อาจจะไม่เป็นผลดีเสมอไป สอดคล้องกับผลการวิจัยของ Roznowski และ Reith (1999) ที่ได้ทำการตรวจสอบคุณภาพในการทดสอบและความตรงเชิงทำนายของข้อสอบที่ลำเอียง 3 ระดับ คือไม่ลำเอียง ลำเอียงปานกลาง และลำเอียงมาก โดยใช้วิธี Mantel-Haenzel และแบ่งกลุ่มตัวอย่างตามเพศและเชื้อชาติ พบว่าคุณภาพในการทดสอบของข้อสอบที่ DIF ส่วนใหญ่มีประสิทธิภาพเท่าเทียมกับข้อสอบที่ไม่ DIF ผลการวิจัยแสดงให้เห็นว่าข้อสอบที่ DIF บางข้ออาจเป็นข้อสอบที่มีคุณภาพดีได้ และ Rudner และคณะ (1984 , อ้างถึงใน สุพัฒน์ สุกมลสันต์ , 2531) กล่าวว่าข้อสอบที่ลำเอียงไม่จำเป็นต้องเป็นข้อสอบที่ไม่ดีเสมอไป เพราะในบางสถานการณ์ข้อสอบที่ลำเอียงอาจเป็นข้อสอบที่ให้สารสนเทศที่ดีที่สุดหรือเป็นตัวทำนายที่ดีที่สุดได้ ดังนั้นดัชนี CDIF/DTF จึงมีประโยชน์ในการลดจำนวนข้อสอบที่ถูกตัดออกหรือข้อสอบที่ต้องปรับปรุงลงได้ ซึ่งสอดคล้องกับข้อเสนอแนะของ สุรศักดิ์ อมรรัตนศักดิ์ (2531) และ O'Neil และ McPeek (1993) ที่ว่าเมื่อพบข้อสอบที่ DIF สิ่งสำคัญคือการปรับปรุงข้อสอบนั้น ๆ ให้มีคุณภาพดีขึ้นหรือพยายามปรับให้มีจำนวนข้อสอบที่เข้าข้างเฉพาะกลุ่มใดกลุ่มหนึ่งในจำนวนเท่า ๆ กันมากกว่าการที่จะมุ่งตัดข้อสอบที่พบ DIF นั้นออกจากแบบสอบ

4. จากการวิเคราะห์ความสัมพันธ์ระหว่างตำแหน่งของคะแนนรวมของผู้สอบก่อนและหลังการตัดข้อสอบที่พบ DIF ของแบบสอบวิชาภาษาอังกฤษและวิชาคณิตศาสตร์ โดยการหาค่าสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนแรงค์ พบว่าตำแหน่งของคะแนนรวมของผู้สอบก่อนและหลังตัดข้อสอบมีความสัมพันธ์ในทางบวกซึ่งกันและกันอย่างมีนัยสำคัญ แบบสอบวิชาภาษาอังกฤษมีค่าสัมประสิทธิ์สห

สัมพันธ์สเปียร์แมนแรงค์อยู่ระหว่าง 0.83 – 0.99 ส่วนในแบบสอบวิชาคณิตศาสตร์มีค่าสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนแรงค์อยู่ระหว่าง 0.87 – 0.98 ค่าสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนแรงค์ที่ได้มีค่าลดลงขึ้นอยู่กับจำนวนข้อสอบที่พบ DIF และถูกตัดออกไป กล่าวคือถ้าตัดข้อสอบที่พบ DIF ออกไปมากเท่าใด ตำแหน่งของคะแนนรวมของผู้สอบหลังตัดข้อสอบที่ DIF ก็จะแตกต่างจากก่อนการตัดข้อสอบเท่านั้น ถึงแม้ค่าสัมประสิทธิ์สหสัมพันธ์ที่คำนวณได้จะมีค่าสูง นั่นก็หมายถึง ตำแหน่งของคะแนนรวมของผู้สอบก่อนและหลังตัดข้อสอบที่ DIF ไม่แตกต่างกันมากนัก แต่ผู้วิจัยเห็นว่าการตัดข้อสอบที่ทำหน้าที่ต่างกันออก แล้วจึงทำการรวมคะแนนเพื่อนำผลคะแนนนั้นไปใช้ เป็นวิธีที่ให้ความยุติธรรมกับผู้สอบ โดยไม่คำนึงถึงความแตกต่างในเรื่องเพศหรือสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษาของผู้สอบ ดังนั้นถึงแม้ตำแหน่งของคะแนนรวมของผู้สอบจะเปลี่ยนแปลงเพียงเล็กน้อย ก็เป็นสิ่งที่ควรได้รับการพิจารณา

ข้อเสนอแนะ

ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1. จากผลการเปรียบเทียบคุณภาพของแบบสอบหลังการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบและตัดข้อสอบที่ DIF ออกจากแบบสอบ พบว่าแบบสอบฉบับหลังตัดข้อสอบยังมีความตรงเชิงโครงสร้าง แต่ความเที่ยงของแบบสอบฉบับหลังตัดข้อสอบมีค่าลดลงจากแบบสอบฉบับก่อนตัดข้อสอบเล็กน้อยหรือในบางกรณียังมีค่าเท่าเดิม ซึ่งความเที่ยงและความตรงนี้เป็นการวิเคราะห์คุณภาพของแบบสอบตามทฤษฎีการทดสอบแบบดั้งเดิม (classical test theory : CTT) ส่วนค่าสารสนเทศของแบบสอบ ซึ่งเป็นวิธีการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (item response theory : IRT) พบว่าแบบสอบฉบับหลังตัดข้อสอบที่ DIF มีค่าสารสนเทศเพิ่มขึ้น จึงมีข้อเสนอแนะ ดังนี้

1.1 ถ้าใช้ผลคะแนนดิบในการตัดสินผลการคัดเลือกตามทฤษฎีการทดสอบแบบดั้งเดิมแล้ว ไม่ควรตัดข้อสอบที่ DIF ทุกข้อออกจากแบบสอบ ควรตัดเฉพาะบางข้อ เพราะจะมีผลกระทบต่อค่าความเที่ยงแบบสอบ

1.2 ถ้าใช้คะแนนความสามารถ (θ) ในการตัดสินผลการคัดเลือกตามทฤษฎีการตอบสนองข้อสอบแล้ว ควรตัดข้อสอบที่ DIF ออกจากแบบสอบก่อน เพราะว่าทำให้ค่าสารสนเทศของแบบสอบมีค่าเพิ่มขึ้น ความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ

จะลดลง แต่ในการตัดข้อสอบออกนั้น ต้องคำนึงถึงความตรงตามเนื้อหา (content validity) ของแบบสอบเป็นสำคัญด้วย โดยอาจมีการสร้างข้อสอบทดแทนข้อสอบที่เกิด DIF

2. จากผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ สำหรับแบบสอบวิชาภาษาอังกฤษและวิชาคณิตศาสตร์ 1 พบว่าเมื่อจำแนกกลุ่มผู้สอบตามเพศ พบข้อสอบที่ DIF และข้อสอบที่ถูกตัดออกมากกว่าสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา ในวิชาคณิตศาสตร์ ข้อสอบที่พบ DIF มากที่สุดเป็นข้อสอบที่มีเนื้อหาเรื่อง วิธีจัดหมู่ วิธีสับเปลี่ยน สถิติและความน่าจะเป็น ซึ่งส่วนใหญ่ข้อคำถามเป็นโจทย์ปัญหา ส่วนในวิชาภาษาอังกฤษข้อสอบที่พบ DIF มากที่สุด เป็นข้อสอบที่มีเนื้อหาเรื่องการใช้ภาษา ซึ่งข้อคำถามมีลักษณะเป็นสถานการณ์ต่าง ๆ ดังนั้นคณะกรรมการพัฒนาแบบสอบหรือผู้ที่เกี่ยวข้อง น่าจะได้มีการศึกษาวิเคราะห์ถึงสาเหตุของการเกิด DIF เช่น ลักษณะของภาษา หรือสถานการณ์ที่ใช้ เพื่อที่จะได้หลีกเลี่ยงสาเหตุที่จะนำไปสู่การเกิด DIF และทำให้ได้แบบสอบที่มีความยุติธรรมกับผู้สอบทุกกลุ่ม

ข้อเสนอแนะในการวิจัยครั้งต่อไป

1. จากการศึกษาครั้งนี้ผู้วิจัยตัดสินข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบจากค่าสถิติเท่านั้น ดังนั้นควรมีการศึกษาวิเคราะห์ข้อสอบที่ผลการวิจัยครั้งนี้พบว่าทำหน้าที่ต่างกัน เพื่อค้นหาสาเหตุที่ทำให้ข้อสอบเหล่านั้นทำหน้าที่ต่างกัน ว่ามาจากสาเหตุใด สำหรับการพัฒนาแบบสอบคัดเลือกในอนาคต ควรมีการศึกษาวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบในปีที่ผ่านมา แล้วนำข้อสอบที่พบ DIF เหล่านั้นไปวิเคราะห์ถึงสาเหตุของการเกิด DIF เช่นลักษณะของภาษา เนื้อหาหรือสถานการณ์ที่ใช้ในข้อคำถามเปรียบเทียบระหว่างปี เพื่อที่จะได้ทราบสาเหตุ อันจะเป็นประโยชน์ต่อผู้พัฒนาแบบสอบในการปรับปรุงแก้ไขการเขียนข้อสอบเพื่อหลีกเลี่ยงสิ่งที่นำไปสู่การเกิด DIF ในการพัฒนาแบบสอบในครั้งต่อไป

2. ควรมีการศึกษาวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ สำหรับแบบสอบคัดเลือกอื่น ๆ เช่น แบบสอบแข่งขันเพื่อบรรจุเข้ารับราชการครู เป็นต้น หรือแบบสอบอื่น ๆ เช่น แบบสอบมาตรฐานต่าง ๆ ที่ได้รับการแปลเป็นภาษาไทย เพื่อดูลักษณะของภาษาที่ใช้ว่าทำให้เกิด DIF ได้หรือไม่ เมื่อจำแนกกลุ่มผู้สอบตามตัวแปรต่าง ๆ

3. ในการศึกษาครั้งนี้ใช้อัตราส่วนระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่ศึกษา คือ 1:1 ซึ่งจากผลการศึกษาของ จิติมา วรรณศรี (2539) พบว่าอัตราส่วน 1:1 เป็นอัตราส่วนที่ให้ประสิทธิภาพในการตรวจสอบดีที่สุด แต่ในทางปฏิบัติการสอบคัดเลือกเข้าศึกษาต่อในสถาบันการศึกษาต่าง ๆ มักมี

อัตราส่วนของประชากรระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่แตกต่างกัน ดังนั้นควรมีการศึกษาว่าการใช้อัตราส่วนกลุ่มตัวอย่างระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบในอัตราส่วนใด จึงจะยังให้ผลการตรวจสอบที่น่าเชื่อถือ

4. ควรมีการเปรียบเทียบผลของการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบด้วยกระบวนการ DFIT กับวิธีการอื่น ๆ เช่น วิธีซิบเทสส์ (sibtest) วิธีซิบเทสส์ปรับใหม่ (modified sibtest) เป็นต้น โดยใช้ข้อมูลจำลอง (simulate data) ศึกษาตัวแปรต่าง ๆ ที่มีผลต่อประสิทธิภาพของกระบวนการตรวจสอบ เช่น อัตราส่วนระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ความยาวของแบบสอบ หรือจำนวนของข้อสอบที่ DIF ในแบบสอบเป็นต้น เพื่อหาวิธีการในการตรวจสอบที่ดีที่สุด

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รายการอ้างอิง

ภาษาไทย

กาญจนา วัธนสุนทร. (2537). **การพัฒนาเกณฑ์การตัดสินข้อสอบลำเอียงทางเพศ**. วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย .

เกษร หว่างจิตร. (2539). **การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบสำหรับแบบสอบคัดเลือกระดับบัณฑิตศึกษาวิชาภาษาไทยและภาษาอังกฤษด้วยวิธีแมนเทล-เฮนส์เซล**. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย .

คมศักดิ์ ชื่นชม. (2539). **การศึกษาผลการวิเคราะห์ความลำเอียงที่ใช้วิธีต่างกันของแบบทดสอบวัดจริยธรรมด้านความซื่อสัตย์**. ปริญญานิพนธ์การศึกษามหาบัณฑิต สาขาวิชาการวัดผลการศึกษา มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร.

จิติมา วรรณศรี. (2539). **การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-เฮนส์เซลกับวิธีซิบเทสท์ เมื่อความยาวแบบสอบ ขนาดกลุ่มตัวอย่างและอัตราส่วนของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบต่างกัน**. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

ทดสอบกลาง , สำนัก , ทบวงมหาวิทยาลัย. (2540). **คุณสมบัติและเกณฑ์การคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาระบบใหม่**. กรุงเทพฯ : ศรีเมืองการพิมพ์.

ทดสอบกลาง , สำนัก , ทบวงมหาวิทยาลัย. (2541). **ข้อมูลการสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษา ทบวงมหาวิทยาลัย ปีการศึกษา 2541**. มปท.

ทัศนีย์ พีรมนตรี. (2530). **การวิเคราะห์ความลำเอียงของแบบสอบวิชาคณิตศาสตร์โครงการตรวจสอบคุณภาพการศึกษา ชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526**. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิยาลัย .

นงลักษณ์ วิรัชชัย. (2542). โมเดลลิสเรล : สถิติวิเคราะห์สำหรับการวิจัย. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย.

นพมาศ พิพัฒนสุข. (2541). **การเปรียบเทียบประสิทธิภาพระหว่างวิธีแมนเทล-แฮนส์เซลกับวิธีถดถอยโลจิสติกในการตรวจสอบทำหน้าที่ต่างกันของข้อสอบเมื่อใช้เกณฑ์จับคู่แตกต่างกันใน**

แบบสอบชนิดพหุมิติ. วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

นิคม กีรติวรางกูร. (2542). **การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีการวิเคราะห์องค์ประกอบจำกัด วิธีแมนเทล-เฮนส์เซลกับการตอบสนองข้อสอบ**. วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

นิฏฐา กัลยาประเสริฐ. (2539). **การเปรียบเทียบความรู้ความสามารถในการเขียนภาษาอังกฤษเชิงอธิบาย ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่มีเพศและขนาดชั้นเรียนต่างกัน ในโรงเรียนมัธยมศึกษากรุงเทพมหานคร**. วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชามัธยมศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

ปรมินทร์ อริเดช. (2539). **การใช้จีอาร์เอ็ม จีพีซีเอ็มและโมเดลโลจิสติกในการเปรียบเทียบฟังก์ชันสารสนเทศของมาตรประมาณค่าที่มีวิธีการให้คะแนนแบบทวิวิภาคและแบบพหุวิภาค**. วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

พิศิษฐ ตัณฑวณิช. (2532) การทดสอบความมีนัยสำคัญของค่าครอนบาช แอลฟ่า. **วิธีวิทยาการวิจัย** ปีที่ 4 3(กย. – ธค.) : 43-45.

เยาวดี วิบูลย์ศรี. (2539). **การวัดผลและการสร้างแบบสอบผลสัมฤทธิ์ (Measurement and Achievement Test Construction)**. กรุงเทพฯ : สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.

วิไลลักษณ์ บุญเคลือบ, จริยา เกตุเผือก และ นงลักษณ์ เขียนงาม. (2544). **ความคิดเห็นของบุคลากรเขต**

การศึกษา 8 เกี่ยวกับการคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาระบบใหม่ ปีการศึกษา 2543. รายงานการวิจัย คณะศึกษาศาสตร์ มหาวิทยาลัยเชียงใหม่.

ศิริชัย กาญจนวาสี. (2532). โมเดลเชิงสาเหตุ : การสร้างและการวิเคราะห์. **วิธีวิทยาการวิจัย** ปีที่ 4 3(กย. – ธค.) : 1-24.

ศิริชัย กาญจนวาสี. (2541). **ทฤษฎีการวัดและประเมิน (Theories of Measurement and Evaluation)**. เอกสารประกอบการบรรยาย รายวิชา 2702639 ทฤษฎีการวัดและประเมิน.

ศิริชัย กาญจนวาสี. (2544). **ทฤษฎีการทดสอบแบบดั้งเดิม (Claasical Test Theory)**, พิมพ์ครั้งที่ 4. กรุงเทพฯ : โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

สุพัฒน์ สุกมลสันต์. (2534) . **การวิเคราะห์ความลำเอียงของข้อทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ปี พ.ศ. 2531-2533.** สถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย.

สุรศักดิ์ อมรรัตนศักดิ์. (2531). **การศึกษาเปรียบเทียบผลของวิธีวิเคราะห์ความลำเอียงของข้อสอบที่แตกต่างกัน 4 วิธี.** วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย .

ภาษาอังกฤษ

Baker, F.B. (1993). **EQUATE 2.0 : Computer Program for Equating Two Metrics in Item Response Theory** [Computer Program]. Madison : University of Wisconsin, Laboratory of Experimental Design.

Camilli, G. and Shepard, L.A. (1994). **Methods for Identifying Biased Test Items.** (n.p.) :

SAGE Publications.

Chang, H-H., Mazzeo, J. and Roussos, L. (1996). Detecting DIF for Polytomously Scored

Items : An Adaptation of the SIBTEST Procedure. **Journal of Educational Measurement** 33(3) : 333-353.

Doolittle, A.E. and Cleary, T.A. (1987). Gender Based Differential Item Performance in Mathematics Achievement Items. **Journal of Educational Measurement** 24(2) : 157-166.

Ellis, B.B. and Mead, A.D. (2000). Assessment of the Measurement Equivalence of A Spanish Translation of the 16PF Questionnaire. **Educational and Psychological Measurement** 60(5) : 787-807.

Flowers, C.P. and Oshima, T.C. (1994). **The Consistency of DIF/DTF across Different**

Test Administrations: A Multidimensional Perspective. ERIC Digest. Available from : http://ericae.net/ericdb/ED377223.htm. 22/7/43.

Flowers, C.P. and Others. (1996). **A Description and Demonstration of the Polytomous-DFIT Framework.** ERIC Digest. Available from : http://ericae.net/ericdb/ED401319.htm. 27/7/43.

Flowers, C.P. and Others. (1997). **The Relationship between Polytomous-DFIT and Other Polytomous DIF Procedures.** ERIC Digest. Available from : http://ericae.net/ericdb/ED401319.htm. 26/7/43.

Giray, B. (1995). Diffenrential Item Functioning (DIF) Analysis of Computation, Word Problem and Geometry Questions Across Gender and SES Groups. **Studies in Educational Evaluation** 21(4) : 439-456.

Hambleton, R.K. and Swaminathan, H. (1985). **Item Response Theory : Principles and Applications**. Bostion : Kluwer-Nijhoff Publishing .

Hambleton, R.K., Swaminathan, H. and Rogers, H.J. (1991). **Fundamentals of Item Response Theory**. California : SAGE Publications.

Holland, P.W. and Wainer, H. (1993). **Differential Item Functioning**. Hillsdale, NJ: Lawrence Erlbaum Associates.

Katherine, E.R. and Meichu, F. (1996). Examining Gender DIF on a Multiple-Choice Test of Mathematics : a Confirmatory Approach. **Education Measurement** 15 : 15-20.

Larry, R.P. (1999). **Differential Functioning of Items and Tests versus the Mantel-Haenszel Technique for Detecting Differential Item Functioning in Translated Test.** ERIC Digest. Available from : http://ericae.net/ericdb/ED430995.htm. 22/7/43.

Lord, F.M. (1980). **Applications of Item Response Theory to Practical Testing**

Problems. Hillsdale, NJ : Lawrence Erlbaum Associates.

Mislevy, R.J. and Bock, R.D. (1990). **BILOG 3 : Item Analysis and Test Scoring with Binary Logistic Models [Computer Program].** Chicago : Scientific Software, Inc.

Narayan, P. and Swaminathan, H. (1994). Performance of the Mantel-Haenszel and Simultaneous Item Bias Procedures for Detecting Differential Item Functioning. **Applied Psychological Measurement** 18(4) : 315-328.

Ning, W. and Suzanne, L. (1996). Detection of Gender-Related Differential Item Functioning in a Mathematics Performance Assessment. **Applied Measurement in Education** 9(2) : 175-199.

O'Neil, A.K. and McPeek, M.W. (1993) Item and Test Characteristics that are Associated with Differential Item Functioning. **Differential Item Functioning.** Hillsdale, NJ : Erlbaum Associates. : 255-276.

Oshima, T.C. and Others. (1993**). Evaluation of DTF and DIF in Two-Dimensional IRT**. ERIC Digest. Available from : http://ericae.net/ericdb/ED365707.htm. 22/7/43.

Oshima, T.C. , Raju, N.S., and Flowers, C.P. (1997). Development and Demonstration of

Multidimensional IRT-Based Internal Measures of Differential Functioning of Items and Tests. **Journal of Educational Measurement** 34(3) : 253-272.

Oshima, T.C. , Raju, N.S., Flowers, C.P. and Slinde, J.A. (1998). Differential Bundle Functioning the DFIT Framework: Procedures for Identifying Possible Sources of Differential Functioning. **Applied Measurement in Education** 11(4) : 353-369.

Peter, H. and Darwin, P.H. (1994) Human Self Assessment in Multiple Choices Testing. **Journal of Educational Measurement** 31(2) : 149-160.

Potenza, M.T. and Dorans, N.J. (1995). DIF Assessment for Polytomously Scored Items :

A Framework for Classification and Evaluation. **Applied Psychological Measurement** 19(1) : 23-37.

Raju, N.S., Drasgow, F. and Slinde, J.A. (1993). An Empirical Comparison of the Area Methods Lord's Chisquare test and the Mantel-Hanszel technique for Assessing Differential Item Functioning. **Educational and Psychological Measurement.** 53(2) :301-314.

Raju, N.S. , van der Linden, W.J., and Fleer, P.F. (1995). IRT-Based Internal Measures of

Differential Functioning of Items and Tests. **Applied Psychological Measurement** 19(4): 353-368.

Rogers, H.J. and Swaminathan, H. (1993). A Comparison of Logistic Regression and Mantel-Haenszel Procedures for Detecting Difference Item Functioning . **Applied Psychological Measurement** 17(2) : 105-116.

Roznowski, M. And Reith, J. (1999). Examining the Measurement Quality of Tests Containing Diffentially Functioning Items : Do Biased Items Result in Poor Measurement?. **Educational and Psychological Measurement** 59(2) : 248-269.

Ryan, K.E. (1991). The Performance of the Mantel-Haenszel Procedure Across Samples

and Matching Criteria. **Journal of Evaluational Measurement** 28(4) : 325-337.

Shealy, R.T. and Stout, W.F. (1993). A Model-based Standardization Approach that Separates True Bias/DIF from Ability Group Differences and Detects Test Bias/DTF as well as Item Bias/DIF. **Psychometrika** 58(2) : 159-194.

Suzanne, L. (1996). Gender-Related Diffential Item Functioning on a Middle-School Mathematics Performance Assessment. **Educational Measurement** 15 : 21-27.

Welch, C.J. and Miller, T.R. (1995). Assessing Differential Item Functioning in Direct Writing Assessment : Problem and Example. **Journal of Evaluational Measurement** 32(2) : 163-178.

Zwick, R., Donoghue, J., and Grima, A. (1993). Assessment of differential item functioning for performance tasks. **Journal of Educational Measurement** 30(3) : 233-251.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตัวอย่างผลการวิเคราะห์ความตรงเชิงโครงสร้างด้วยวิธีวิเคราะห์องค์ประกอบเชิงยืนยัน
สำหรับแบบสอบวิชาภาษาอังกฤษ หลังการวิเคราะห์ DTF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

DATE: 7/ 2/ 1

TIME: 16:42

DOS L I S R E L 8.10

BY

KARL G JORESKOG AND DAG SORBOM

This program is published exclusively by

Scientific Software International, Inc.

1525 East 53rd Street - Suite 530

Chicago, Illinois 60615, U.S.A.

Voice: (800)247-6113, (312)684-4920, Fax: (312)684-4979

Copyright by Scientific Software International, Inc., 1981-93.

Partial copyright by Microsoft Corp., 1993 and Media Cybernetics Inc., 1993.

Use of this program is subject to the terms specified in the

Universal Copyright Convention.

The following lines were read from file enewst.pro:

CONFIRMATORY FACTOR ANALYSIS ENGLIS FOR SECOND ORDER

DA NI=5 NO=1000 MA=CM

LA

FACT INTERP EVAL LANG GRAM

KM

1

0.66 1

0.60 0.57 1

0.55 0.65 0.49 1

0.64 0.70 0.55 0.78 1

SD

1.80 1.71 1.54 1.87 1.81

MO NX=5 NK=1 LX=FU,FR PH=SY TD=SY

FR TD(4,5) TD(1,3) TD(2,4)

LK

ENGLISH

OU SE TV RS MR FS MI ND=3 AD=OFF

CONFIRMATORY FACTOR ANALYSIS ENGLIS FOR SECOND ORDER

NUMBER OF INPUT VARIABLES 5

NUMBER OF Y - VARIABLES 0

NUMBER OF X - VARIABLES 5

NUMBER OF ETA - VARIABLES 0

NUMBER OF KSI - VARIABLES 1

NUMBER OF OBSERVATIONS 1000

CONFIRMATORY FACTOR ANALYSIS ENGLIS FOR SECOND ORDER

COVARIANCE MATRIX TO BE ANALYZED

	FACT	INTERP	EVAL	LANG	GRAM
FACT	3.240				
INTERP	2.031	2.924			
EVAL	1.663	1.501	2.372		
LANG	1.851	2.079	1.411	3.497	
GRAM	2.085	2.167	1.533	2.640	3.276

CONFIRMATORY FACTOR ANALYSIS ENGLIS FOR SECOND ORDER

Number of Iterations = 6

LISREL ESTIMATES (MAXIMUM LIKELIHOOD)

LAMBDA-X

	ENGLISH

FACT 1.397

(.051)

27.249

INTERP 1.454

(.047)

31.062

EVAL 1.032

(.047)

22.162

LANG 1.338

(.058)

23.035

GRAM 1.490

(.050)

29.762

PHI

ENGLISH

1.000

THETA-DELTA

	FACT	INTERP	EVAL	LANG	GRAM
	---	---	---	---	---
FACT	1.287				
	(.079)				
	16.394				
INTERP	- -	.810			
		(.064)			

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

12.745

EVAL	.221	- -	1.307	
	(.056)		(.069)	
	3.980		18.845	
LANG	- -	.132	- -	1.705
		(.051)		(.105)
		2.586		16.205
GRAM	- -	- -	- -	.646	1.056
				(.071)	(.072)
				9.076	14.589

SQUARED MULTIPLE CORRELATIONS FOR X - VARIABLES

FACT	INTERP	EVAL	LANG	GRAM
---	---	---	---	---
.603	.723	.449	.512	.678

GOODNESS OF FIT STATISTICS

CHI-SQUARE WITH 2 DEGREES OF FREEDOM = 1.086 (P = 0.581)

ESTIMATED NON-CENTRALITY PARAMETER (NCP) = 0.0

90 PERCENT CONFIDENCE INTERVAL FOR NCP = (0.0 ; 5.493)

MINIMUM FIT FUNCTION VALUE = 0.00109

POPULATION DISCREPANCY FUNCTION VALUE (F0) = 0.0

90 PERCENT CONFIDENCE INTERVAL FOR F0 = (0.0 ; 0.00550)

ROOT MEAN SQUARE ERROR OF APPROXIMATION (RMSEA) = 0.0

90 PERCENT CONFIDENCE INTERVAL FOR RMSEA = (0.0 ; 0.0524)

P-VALUE FOR TEST OF CLOSE FIT (RMSEA < 0.05) = 0.939

EXPECTED CROSS-VALIDATION INDEX (ECVI) = 0.0271

90 PERCENT CONFIDENCE INTERVAL FOR ECVI = (0.0280 ; 0.0335)

ECVI FOR SATURATED MODEL = 0.0300

ECVI FOR INDEPENDENCE MODEL = 2.926

CHI-SQUARE FOR INDEPENDENCE MODEL WITH 10 DEGREES OF FREEDOM = 2912.989

INDEPENDENCE AIC = 2922.989

MODEL AIC = 27.086

SATURATED AIC = 30.000

INDEPENDENCE CAIC = 2952.528

MODEL CAIC = 103.887

SATURATED CAIC = 118.616

ROOT MEAN SQUARE RESIDUAL (RMR) = 0.00931

STANDARDIZED RMR = 0.00311

GOODNESS OF FIT INDEX (GFI) = 1.00

ADJUSTED GOODNESS OF FIT INDEX (AGFI) = 0.997

PARSIMONY GOODNESS OF FIT INDEX (PGFI) = 0.133

NORMED FIT INDEX (NFI) = 1.00

NON-NORMED FIT INDEX (NNFI) = 1.002

PARSIMONY NORMED FIT INDEX (PNFI) = 0.200

COMPARATIVE FIT INDEX (CFI) = 1.000

INCREMENTAL FIT INDEX (IFI) = 1.000

RELATIVE FIT INDEX (RFI) = 0.998

CRITICAL N (CN) = 8475.504

CONFIRMATORY FACTOR ANALYSIS ENGLIS FOR SECOND ORDER

MODIFICATION INDICES AND EXPECTED CHANGE

NO NON-ZERO MODIFICATION INDICES FOR LAMBDA-X

NO NON-ZERO MODIFICATION INDICES FOR PHI

MODIFICATION INDICES FOR THETA-DELTA

FACT	INTERP	EVAL	LANG	GRAM
--------	--------	--------	--------	--------

FACT	- -				
INTERP	.022	- -			
EVAL	- -	.022	- -		
LANG	1.067	- -	1.067	- -	
GRAM	.570	- -	.570	- -	- -

EXPECTED CHANGE FOR THETA-DELTA

	FACT	INTERP	EVAL	LANG	GRAM
FACT	- -				
INTERP	.011	- -			
EVAL	- -	-.008	- -		
LANG	-.065	- -	.048	- -	
GRAM	.045	- -	-.033	- -	- -

COMPLETELY STANDARDIZED EXPECTED CHANGE FOR THETA-DELTA

	FACT	INTERP	EVAL	LANG	GRAM
FACT	- -				
INTERP	.003	- -			
EVAL	- -	-.003	- -		
LANG	-.019	- -	.017	- -	
GRAM	.014	- -	-.012	- -	- -

MAXIMUM MODIFICATION INDEX IS 1.07 FOR ELEMENT (4, 3) OF THETA-DELTA

CONFIRMATORY FACTOR ANALYSIS ENGLIS FOR SECOND ORDER

COVARIANCES

X - KSI

	FACT	INTERP	EVAL	LANG	GRAM

	--------	--------	--------	--------	--------
ENGLISH	1.397	1.454	1.032	1.338	1.490

CONFIRMATORY FACTOR ANALYSIS ENGLIS FOR SECOND ORDER

FACTOR SCORES REGRESSIONS

KSI

	FACT	INTERP	EVAL	LANG	GRAM
	--------	--------	--------	--------	--------
ENGLISH	.126	.228	.080	.019	.170

CONFIRMATORY FACTOR ANALYSIS ENGLIS FOR SECOND ORDER

STANDARDIZED SOLUTION

LAMBDA-X

	ENGLISH

FACT	1.397
INTERP	1.454
EVAL	1.032
LANG	1.338
GRAM	1.490

PHI

ENGLISH

1.000

CONFIRMATORY FACTOR ANALYSIS ENGLIS FOR SECOND ORDER

COMPLETELY STANDARDIZED SOLUTION

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

LAMBDA-X

	ENGLISH
FACT	.776
INTERP	.850
EVAL	.670
LANG	.716
GRAM	.823

PHI

	ENGLISH
	1.000

THETA-DELTA

	FACT	INTERP	EVAL	LANG	GRAM
FACT	.397				
INTERP	- -	.277			
EVAL	.080	- -	.551		
LANG	- -	.041	- -	.488	
GRAM	- -	- -	- -	.191	.322

THE PROBLEM USED 4400 BYTES (= 1.5% OF AVAILABLE WORKSPACE)

TIME USED: .3 SECONDS

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตัวอย่างผลการวิเคราะห์ความตรงเชิงโครงสร้างด้วยวิธีวิเคราะห์องค์ประกอบ
สำหรับแบบสอบวิชาคณิตศาสตร์ 1 หลังการวิเคราะห์ DIF เมื่อแบ่งกลุ่มผู้สอบตามเพศ

DATE: 6/23/ 1

TIME: 3:35

DOS L I S R E L 8.10

BY

KARL G JORESKOG AND DAG SORBOM

This program is published exclusively by

Scientific Software International, Inc.

1525 East 53rd Street - Suite 530

Chicago, Illinois 60615, U.S.A.

Voice: (800)247-6113, (312)684-4920, Fax: (312)684-4979

Copyright by Scientific Software International, Inc., 1981-93.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Partial copyright by Microsoft Corp., 1993 and Media Cybernetics Inc., 1993.
Use of this program is subject to the terms specified in the
Universal Copyright Convention.

The following lines were read from file mathnews.pro:

CONFIRMATORY FACTOR ANALYSIS
DA NI=18 NO=1000 MA=CM
LA
X1 X2 X4 X5 X7 X9 X10 X11 X12 X13 X14 X15 X17 X18 X20 X21 X26 X28
KM
1
.07 1
.04 .07 1
.01 .07 .05 1
.01 .01 .05 .03 1
.04 .02 -.01 -.01 .01 1
.01 .10 .14 .03 .13 .05 1
.00 .05 .01 -.01 .01 .03 .05 1
.01 .02 .04 .01 .01 .04 .04 .01 1
.14 .11 .16 .04 .10 .03 .08 .04 .08 1
.05 .06 .17 .04 .13 -.02 .05 -.03 .03 .15 1
.01 .07 .06 .07 .10 .02 .04 .02 .04 .09 .09 1
-.01 .04 .03 .04 .03 .00 .05 .03 .01 .03 .05 .00 1
.05 .05 .11 .05 .04 -.03 .08 -.03 .01 .18 .09 .09 -.01 1
.05 .01 .02 .03 .00 .03 .10 .03 .01 -.02 -.02 .00 -.01 -.01 1
.03 .06 .17 .00 .11 -.02 .07 -.03 .01 .09 .14 .02 .01 .12 .02 1
-.02 .03 .01 .06 .01 .03 -.02 .03 .00 .03 .04 .01 .09 .01 .04 -.02 1
-.01 .01 .09 -.02 .04 -.01 .00 .00 .02 .04 .06 -.01 .06 .05 .09 .06 .03 1
SD
1.43 1.44 1.49 1.40 1.34 1.19 1.49 1.40 1.41 1.49 1.44 1.49 1.22 1.44 1.38

1.45 1.43 1.61

MO NX=18 NK=1 LX=FU,FR TD=SY

LK

K1

OU SE TV RS MR FS MI ND=3 AD=OFF

CONFIRMATORY FACTOR ANALYSIS

NUMBER OF INPUT VARIABLES 18

NUMBER OF Y - VARIABLES 0

NUMBER OF X - VARIABLES 18

NUMBER OF ETA - VARIABLES 0

NUMBER OF KSI - VARIABLES 1

NUMBER OF OBSERVATIONS 1000

CONFIRMATORY FACTOR ANALYSIS

COVARIANCE MATRIX TO BE ANALYZED

	X1	X2	X4	X5	X7	X9
X1	2.045					
X2	.144	2.074				
X4	.085	.150	2.220			
X5	.020	.141	.104	1.960		
X7	.019	.019	.100	.056	1.796	
X9	.068	.034	-.018	-.017	.016	1.416
X10	.021	.215	.311	.063	.260	.089
X11	- -	.101	.021	-.020	.019	.050
X12	.020	.041	.084	.020	.019	.067
X13	.298	.236	.355	.083	.200	.053
X14	.103	.124	.365	.081	.251	-.034
X15	.021	.150	.133	.146	.200	.035
X17	-.017	.070	.055	.068	.049	- -

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

X18	.103	.104	.236	.101	.077	-.051
X20	.099	.020	.041	.058	- -	.049
X21	.062	.125	.367	- -	.214	-.035
X26	-.041	.062	.021	.120	.019	.051
X28	-.023	.023	.216	-.045	.086	-.019

COVARIANCE MATRIX TO BE ANALYZED

	X10	X11	X12	X13	X14	X15
X10	2.220					
X11	.104	1.960				
X12	.084	.020	1.988			
X13	.178	.083	.168	2.220		
X14	.107	-.060	.061	.322	2.074	
X15	.089	.042	.084	.200	.193	2.220
X17	.091	.051	.017	.055	.088	- -
X18	.172	-.060	.020	.386	.187	.193
X20	.206	.058	.019	-.041	-.040	- -
X21	.151	-.061	.020	.194	.292	.043
X26	-.043	.060	- -	.064	.082	.021
X28	- -	- -	.045	.096	.139	-.024

COVARIANCE MATRIX TO BE ANALYZED

	X17	X18	X20	X21	X26	X28
X17	1.488					
X18	-.018	2.074				
X20	-.017	-.020	1.904			
X21	.018	.251	.040	2.103		
X26	.157	.021	.079	-.041	2.045	
X28	.118	.116	.200	.140	.069	2.592

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

CONFIRMATORY FACTOR ANALYSIS

Number of Iterations = 7

LISREL ESTIMATES (MAXIMUM LIKELIHOOD)

LAMBDA-X

	K1
X1	.220
	(.062)
	3.524
X2	.315
	(.063)
	5.015
X4	.601
	(.066)
	9.118
X5	.173
	(.061)
	2.831
X7	.335
	(.059)
	5.715
X9	.024
	(.052)
	.462
X10	.388
	(.065)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

5.955

X11 .039

(.061)

.637

X12 .149

(.062)

2.419

X13 .630

(.066)

9.515

X14 .532

(.063)

8.395

X15 .310

(.065)

4.765

X17 .112

(.053)

2.104

X18 .447

(.063)

7.088

X20 .066

(.060)

1.089

X21 .451

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

(.063)

7.106

X26 .078

(.062)

1.254

X28 .219

(.070)

3.117

PHI

K1

1.000

THETA-DELTA

X1	X2	X4	X5	X7	X9
1.996	1.974	1.859	1.930	1.684	1.416
(.091)	(.093)	(.101)	(.088)	(.080)	(.063)
21.857	21.325	18.364	22.035	20.996	22.341

THETA-DELTA

X10	X11	X12	X13	X14	X15
2.070	1.958	1.966	1.824	1.791	2.124
(.099)	(.088)	(.089)	(.102)	(.094)	(.099)
20.870	22.334	22.121	17.892	19.112	21.430

THETA-DELTA

X17	X18	X20	X21	X26	X28

1.476	1.874	1.900	1.899	2.039	2.544
(.067)	(.093)	(.085)	(.094)	(.091)	(.116)
22.177	20.173	22.304	20.161	22.288	21.967

SQUARED MULTIPLE CORRELATIONS FOR X - VARIABLES

X1	X2	X4	X5	X7	X9
.024	.048	.163	.015	.062	.000

SQUARED MULTIPLE CORRELATIONS FOR X - VARIABLES

X10	X11	X12	X13	X14	X15
.068	.001	.011	.179	.137	.043

SQUARED MULTIPLE CORRELATIONS FOR X - VARIABLES

X17	X18	X20	X21	X26	X28
.008	.096	.002	.097	.003	.019

GOODNESS OF FIT STATISTICS

CHI-SQUARE WITH 135 DEGREES OF FREEDOM = 146.907 (P = 0.228)

ESTIMATED NON-CENTRALITY PARAMETER (NCP) = 11.907

90 PERCENT CONFIDENCE INTERVAL FOR NCP = (0.0 ; 45.451)

MINIMUM FIT FUNCTION VALUE = 0.147

POPULATION DISCREPANCY FUNCTION VALUE (F0) = 0.0119

90 PERCENT CONFIDENCE INTERVAL FOR F0 = (0.0 ; 0.0455)

ROOT MEAN SQUARE ERROR OF APPROXIMATION (RMSEA) = 0.00940

90 PERCENT CONFIDENCE INTERVAL FOR RMSEA = (0.0 ; 0.0184)

P-VALUE FOR TEST OF CLOSE FIT (RMSEA < 0.05) = 1.000

EXPECTED CROSS-VALIDATION INDEX (ECVI) = 0.219

90 PERCENT CONFIDENCE INTERVAL FOR ECVI = (0.207 ; 0.253)

ECVI FOR SATURATED MODEL = 0.342

ECVI FOR INDEPENDENCE MODEL = 0.476

CHI-SQUARE FOR INDEPENDENCE MODEL WITH 153 DEGREES OF FREEDOM = 439.434

INDEPENDENCE AIC = 475.434

MODEL AIC = 218.907

SATURATED AIC = 342.000

INDEPENDENCE CAIC = 581.774

MODEL CAIC = 431.586

SATURATED CAIC = 1352.226

ROOT MEAN SQUARE RESIDUAL (RMR) = 0.0561

STANDARDIZED RMR = 0.0276

GOODNESS OF FIT INDEX (GFI) = 0.984

ADJUSTED GOODNESS OF FIT INDEX (AGFI) = 0.980

PARSIMONY GOODNESS OF FIT INDEX (PGFI) = 0.777

NORMED FIT INDEX (NFI) = 0.666

NON-NORMED FIT INDEX (NNFI) = 0.953

PARSIMONY NORMED FIT INDEX (PNFI) = 0.587

COMPARATIVE FIT INDEX (CFI) = 0.958

INCREMENTAL FIT INDEX (IFI) = 0.961

RELATIVE FIT INDEX (RFI) = 0.621

CRITICAL N (CN) = 1198.781

CONFIRMATORY FACTOR ANALYSIS

COVARIANCES

X - KSI

X1	X2	X4	X5	X7	X9
--------	--------	--------	--------	--------	--------

K1	.220	.315	.601	.173	.335	.024

X - KSI

	X10	X11	X12	X13	X14	X15
K1	.388	.039	.149	.630	.532	.310

X - KSI

	X17	X18	X20	X21	X26	X28
K1	.112	.447	.066	.451	.078	.219

CONFIRMATORY FACTOR ANALYSIS

FACTOR SCORES REGRESSIONS

KSI

	X1	X2	X4	X5	X7	X9
K1	.052	.076	.154	.043	.094	.008

KSI

	X10	X11	X12	X13	X14	X15
K1	.089	.009	.036	.164	.141	.069

KSI

	X17	X18	X20	X21	X26	X28
K1	.036	.113	.016	.113	.018	.041

THE PROBLEM USED 30984 BYTES (= 10.3% OF AVAILABLE WORKSPACE)

TIME USED: 2.8 SECONDS

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

การวิเคราะห์ด้วยโปรแกรม BILOG 3.04

ตัวอย่าง batch file ซึ่งเป็นแฟ้มคำสั่งในโปรแกรม BILOG 3.04

```
>TITLE DETECTION OF DIF AND DTF;
>COMMENTS 2 PLM ENGLISH 100 ITEMS 1000 PERSON MALE FOCAL GROUP;
>GLOBAL DFNAME='C:\BILOG\EFMA1.DAT',NPARM=2,OMITS,SAVE;
>SAVE COV='C:\BILOG\EFMA1.COV';
>LENGTH NITEMS=100;
>INPUT NTOT=100,NALT=2,NIDC=5,KFNAME='C:\BILOG\EFMA1.DAT',
      OFNAME='C:\BILOG\EFMA1.DAT';
       (5A1,100A1)
>TEST TNAME=FOCAL;
>CALIB FLOAT;
```

เมื่อ

TITLE = ชื่อเรื่องหรือหัวเรื่องของแฟ้มข้อมูล

COMMENT = คำอธิบายรายละเอียดของแฟ้มข้อมูล

DFNAME = ชื่อแฟ้มข้อมูลกาตอบข้อสอบรายข้อของผู้สอบ

NPARM = จำนวนพารามิเตอร์ที่เลือก

COV = การกำหนดชื่อแฟ้มข้อมูลที่เก็บค่าความแปรปรวนร่วมของข้อมูลที่ได้จากการวิเคราะห์

NITEMS = จำนวนข้อสอบที่ใช้ในการวิเคราะห์

NTOT = จำนวนข้อสอบที่ใช้ในการวิเคราะห์

NALT = จำนวนตัวเลือก

NIDC = จำนวนสดมภ์ของรหัสประจำตัวผู้สอบ (ID)

(5A1,100A1) = (จำนวนสดมภ์ของรหัสประจำตัวผู้สอบ (ID) , จำนวนข้อสอบที่ใช้ในการวิเคราะห์)

TNAME = ชื่อของแบบสอบ

FLOAT = เป็นการยอมให้โปรแกรมประมาณค่าเฉลี่ยของการกระจายของข้อมูลพร้อม ๆ กับการประมาณค่าพารามิเตอร์

ตัวอย่าง printout แฟ้มข้อมูลนามสกุล cov ที่นำไปดัดแปลงเพื่อใช้ในการวิเคราะห์ด้วยโปรแกรม EQUATE version 2.0 ต่อไป

```
>TITLE DETECTION OF DIF AND DTF;
>COMMENTS 2 PLM ENGLISH 100 ITEMS 1000 PERSON MALE FOCAL GROUP;
0001FOCAL     .580035   -.628593    .000000    .005476    .003969
              .008201    .000000    .000000    .000000
0002FOCAL     .249884   1.156074    .000000    .002396   -.011540
              .080316    .000000    .000000    .000000
0003FOCAL     .829025   1.444951    .000000    .005772   -.007951
              .016470    .000000    .000000    .000000
0004FOCAL     .172791   4.020860    .000000    .001716   -.038824
              .943697    .000000    .000000    .000000
0005FOCAL     .266849  -2.380734    .000000    .002860    .023430
              .218080    .000000    .000000    .000000
.
.
.
0096FOCAL     .263762   1.659089    .000000    .002227   -.013269
              .102869    .000000    .000000    .000000
0097FOCAL     .126817   5.465491    .000000    .001297   -.055246
             2.472246    .000000    .000000    .000000
0098FOCAL     .197798   3.445921    .000000    .001948   -.032897
              .605026    .000000    .000000    .000000
0099FOCAL     .224370   3.387754    .000000    .002121   -.030594
              .482667    .000000    .000000    .000000
0100FOCAL     .245915   4.143776    .000000    .002159   -.034908
```

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

.609501 .000000 .000000 .000000

การวิเคราะห์ด้วยโปรแกรม EQUATE 2.0

ตัวอย่างแฟ้มคำสั่ง ซึ่งเป็นแบบ interaction mode ของโปรแกรม EQUATE 2.0

TYPE IN A TITLE FOR THE COMPUTER RUN

ENGLISH FOR SEX VARIABLE DELETE ITEM 1

ENTER NUMBER OF ABILITY SCALE POINTS N=

21

RESPONSE MODE DICHOTOMOUS, GRADED OR NOMINAL? D/G/N

D

ENTER THE NUMBER OF PARAMETERS IN THE ICC MODEL

2

ENTER THE NAME OF THE “FROM METRIC” ITEM PARAMETER FILE

c:\equate\english\efma11.cov

ENTER FORMAT OF “FROM METRIC” FILE

(2f12.6)

ENTER NUMBER OF ITEMS IN THE “FROM” TEST

99

IS FROM METRIC LOGISTIC OR NORMAL? L/N

L

ENTER THE NAME OF THE “TO METRIC” ITEM PARAMETER FILE

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

c:\equate\english\erfe11.cov

ENTER FORMAT OF “TO METRIC” FILE

(2f12.6)

ENTER NUMBER OF ITEMS IN THE “TO” TEST

99

IS TO METRIC LOGISTIC OR NORMAL? L/N

L

ENTER NAME OF FILE TO STORE TRANFORMED ITEM PARAMETERS

c:\equate\english\esex11.dat

TRANSFORM THETAS? Y/N

N

ARE THESE SPECIFICATIONS OK? Y/N

Y

SPECIFY THE ANCHOR ITEM IDs IN THE “FROM” INSTRUMENT

ENTER LIST OF ANCHO ITEMS SEPERATE WITH COMMAS TERMINATE WITH COLON

1-99:

SPECIFY THE ANCHOR ITEM IDs IN THE “TO” INSTRUMENT

ENTER LIST OF ANCHO ITEMS SEPERATE WITH COMMAS TERMINATE WITH COLON

1-99:

ตัวอย่าง printout ของโปรแกรม EQUATE 2.0

EQUATE VERSION 2.0

DATE: DAY = 18 MONTH = 3 YEAR = 21

START TIME = 0:34:18

ENGLISH FOR SEX VARIABLE DELETE ITEM 1

NUMBER OF ABILITY SCALE POINTS = 21

DICHOTOMOUS RESPONSE MODEL

ICC MODEL HAS 2 PARAMETERS

FROM METRIC ITEM PARAMETER FILE NAME IS

c:\equate\english\efma11.cov

FILE FORMAT IS (2f12.6)

NUMBER OF ITEMS IN “FROM” TEST IS 99

FROM TEST IS IN LOGISTIC OGIVE METRIC

TO METRIC ITEM PARAMETER FILE NAME IS

c:\equate\english\erfe11.cov

FILE FORMAT IS (2f12.6)

NUMBER OF ITEMS IN “TO” TEST IS 99

TO TEST IS IN LOGISTIC OGIVE METRIC

TRANSFORMED ITEM PARAMETER FILE NAME IS

c:\equate\english\esex11.dat

TRANSFORMED PARAMETERS WILL BE IN

LIGISTIC OGIVE METRIC

ANCHOR ITEM IDs FOR “FROM” INDTRUMENT ARE

1-99:

ANCHOR ITEM IDs FOR “TO” INSTRUMENT ARE

1-99:

INITIAL VALUE FOR A = 0.9120 INITIAL VALUE FOR K = 0.2533

FUNCTION AT INITIAL VALUE = 3.850785

NUMBER OF ITERATIONS PERFORMED = 4

METRIC TRANFORMATION COEFFICIENTS ARE

A = 1.0298 K = -0.1714

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

FUNCTION VALUE = 0.191536

SUMMARY STATISTICS FOR TRANSFORMED ITEMS

MEAN B = 3.184 VARAINACE B = 7.381 STD DEV B = 2.717

MEAN A = 0.229 VARAINACE A = 0.034 STD DEV A = 0.185

LABORATORY OF EXPERIMENTAL DESIGN

DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

UNIVERSITY OF WISCONSIN

PROGRAM DATED 15 FEBRUARY 1993

END TIME = 0:35:47

ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบ สำหรับแบบสอบวิชาภาษาอังกฤษ เมื่อแบ่งกลุ่มผู้สอบตามเพศ

ตารางที่ 65 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามเพศ เมื่อตัดข้อสอบข้อที่ 1 และ 33 ออกจากแบบสอบ

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
1*	ตัดออกจากการวิเคราะห์	36	-0.001
2*	-0.007	37	-0.002
3*	0.008	38*	0.010
4	0.010	39	-0.002
5*	0.021	40*	0.011
6*	0.014	41*	0.017
7	0.003	42*	0.017
8*	-0.003	43	0.002
9*	0.044	44	0.010
10	0.008	45	0.013
11*	0.006	46	0.010
12*	-0.006	47	0.014
13	0.000	48	-0.011
14	-0.008	49	0.000
15	-0.006	50	-0.007
16*	0.002	51	-0.004
17*	0.023	52	0.000

18	-0.003	53*	0.019
19	0.009	54	0.009
20	-0.007	55	-0.010
21*	-0.022	56	0.004
22	0.011	57	0.000
23*	0.018	58	0.003
24*	0.016	59	-0.007
25	-0.006	60	-0.012
26	0.008	61	-0.003
27*	0.023	62	-0.010
28*	0.005	63	0.001
29*	0.021	64	0.002
30*	0.019	65	0.000
31	0.004	66*	0.018
32	0.007	67	0.005
33	ตัดออกจากการวิเคราะห์	68*	0.025
34*	0.000	69	-0.001
35	0.000	70	0.001

ตารางที่ 65 (ต่อ)

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
71	0.001	86	-0.007
72	0.011	87	0.010
73*	0.019	88	-0.006
74	0.003	89	-0.005
75	0.000	90	0.004
76	-0.002	91*	0.016
77	-0.003	92	-0.008
78	-0.003	93*	0.025
79	0.012	94	-0.009
80	-0.004	95	0.000
81	-0.013	96*	0.009
82	0.012	97.	0.003
83	-0.006	98	0.005
84	0.006	99	0.003
85	-0.012	100	-0.005
ดัชนี DTF	0.381*		

*หมายเหตุ พิจารณาตามเกณฑ์จุดตัดของ Fleer (1993 อ้างถึงใน Raju และคณะ (1995)) ข้อสอบและแบบสอบทำหน้าที่ต่างกันเมื่อดัชนี NCDIF และดัชนี DTF มีค่ามากกว่า 0.006

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

จากตารางที่ 65 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามตัวแปรเพศ เมื่อตัดข้อสอบข้อที่ 1 และ 33 ออกจากแบบสอบ พบว่าดัชนี DTF มีค่าเท่ากับ 0.381 ซึ่งมีค่ามากกว่า 0.006 จากตารางพบว่าข้อที่ 9 มีค่าดัชนี CDIF สูงที่สุด คือมีค่าเท่ากับ 0.044 เมื่อตัดข้อสอบข้อที่ 9 ออกจากแบบสอบและทำการวิเคราะห์ค่าดัชนี CDIF และดัชนี DTF ใหม่ ได้ผลดังตารางที่ 66

ตารางที่ 66 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามเพศ เมื่อตัดข้อสอบข้อที่ 1 33 และ 9 ออกจากแบบสอบ

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
1*	ตัดออกจากการวิเคราะห์	11*	0.001
2*	0.001	12*	-0.017
3*	0.000	13	-0.001
4	0.006	14	-0.010
5*	0.010	15	-0.009
6*	0.002	16*	-0.007
7	0.002	17*	0.017
8*	-0.006	18	-0.003
9*	ตัดออกจากการวิเคราะห์	19	0.012
ตารางที่ 66(ต่อ)	0.006	20	-0.004

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
21*	-0.025	61	-0.002
22	0.008	62	-0.009
23*	0.009	63	-0.001
24*	0.002	64	0.001
25	-0.003	65	-0.004
26	0.007	66*	0.014
27*	0.012	67	0.006
28*	-0.004	68*	0.013
29*	0.013	69	0.001
30*	0.012	70	0.001
31	0.004	71	0.004
32	0.003	72	0.007
33	ตัดออกจากการวิเคราะห์	73*	0.008
34*	-0.001	74	0.005
35	0.004	75	0.000
36	0.000	76	0.000
37	-0.001	77	0.004
38*	0.004	78	-0.002
39	0.002	79	0.014

40*	0.007	80	-0.006
41*	0.011	81	-0.008
42*	0.008	82	0.011
43	0.006	83	-0.003
44	0.006	84	0.004
45	0.008	85	-0.012
46	0.012	86	-0.004
47	0.015	87	0.007
48	-0.011	88	-0.001
49	0.003	89	-0.002
50	-0.008	90	0.005
51	-0.007	91*	0.017
52	0.004	92	-0.007
53*	0.014	93*	0.014
54	0.006	94	-0.007
55	-0.010	95	0.006
56	0.007	96*	0.005
57	0.002	97	0.003
58	0.003	98	0.006
59	-0.009	99	0.007
60	-0.014	100	-0.005
ดัชนี DTF	0.195*		

*หมายเหตุ พิจารณาตามเกณฑ์จุดตัดของ Fleer (1993 อ้างถึงใน Raju และคณะ (1995)) ข้อสอบและแบบสอบทำหน้าที่ต่างกันเมื่อดัชนี NCDIF และดัชนี DTF มีค่ามากกว่า 0.006

จากตารางที่ 66 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามตัวแปรเพศ เมื่อตัดข้อสอบข้อที่ 1 33 และ 9 ออกจากแบบสอบ พบว่าดัชนี DTF มีค่าเท่ากับ 0.195 ซึ่งมีค่ามากกว่า 0.006 จากตารางพบว่าข้อที่ 91 มีค่าดัชนี CDIF สูงที่สุด คือมีค่าเท่ากับ 0.017 เมื่อตัดข้อสอบข้อที่ 91 ออกจากแบบสอบและทำการวิเคราะห์ค่าดัชนี CDIF และดัชนี DTF ใหม่ได้ผลดังตารางที่ 67

ตารางที่ 67 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามเพศ เมื่อตัดข้อสอบข้อที่ 1, 33, 9 และ 91 ออกจากแบบสอบ

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
1*	ตัดออกจากการวิเคราะห์	31	0.003
2*	0.001	32	0.003
3*	0.000	33	ตัดออกจากการวิเคราะห์

4	0.007	34*	0.000
5*	0.011	35	0.004
6*	0.001	36	0.001
7	0.002	37	-0.001
8*	-0.004	38*	-0.001
9*	ตัดออกจากการวิเคราะห์	39	0.003
10	0.004	40*	0.004
11*	0.000	41*	0.013
12*	-0.013	42*	0.007
13	-0.002	43	0.006
14	-0.006	44	0.003
15	-0.008	45	0.008
16*	-0.005	46	0.012
17*	0.016	47	0.013
18	-0.002	48	-0.010
19	0.009	49	0.003
20	-0.003	50	-0.008
21*	-0.023	51	-0.005
22	0.008	52	0.004
23*	0.007	53*	0.011
24*	0.002	54	0.007
25	-0.005	55	-0.009
26	0.004	56	0.005
27*	0.012	57	0.002
28*	-0.003	58	0.004
29*	0.011	59	-0.006
30*	0.008	60	-0.013

ตารางที่ 67 (ต่อ)

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
61	-0.002	81	-0.010
62	-0.009	82	0.009
63	0.001	83	-0.002
64	0.002	84	0.004
65	-0.003	85	-0.010
66*	0.011	86	-0.004
67	0.003	87	0.007
68*	0.014	88	0.000
69.	0.001	89	-0.004
70	0.005	90	0.006

71	0.004	91*	ตัดออกจากการวิเคราะห์
72	0.008	92	-0.006
73*	0.003	93*	0.011
74	0.005	94	-0.007
75	0.000	95	0.005
76	-0.001	96*	0.003
77	0.006	97	0.002
78	0.000	98	0.005
79	0.013	99	0.004
80.	-0.005	100	-0.004
ดัชนี DTF	0.160*		

*หมายเหตุ พิจารณาตามเกณฑ์จุดตัดของ Fleer (1993 อ้างถึงใน Raju และคณะ (1995)) ข้อสอบและแบบสอบทำหน้าที่ต่างกันเมื่อดัชนี NCDIF และดัชนี DTF มีค่ามากกว่า 0.006

จากตารางที่ 67 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามตัวแปรเพศ เมื่อตัดข้อสอบข้อที่ 1, 33, 9 และ 91 ออกจากแบบสอบ พบว่าดัชนี DTF มีค่าเท่ากับ 0.160 ซึ่งมีค่ามากกว่า 0.006 จากตารางพบว่าข้อที่ 17 มีค่าดัชนี CDIF สูงที่สุด คือมีค่าเท่ากับ 0.016 เมื่อตัดข้อสอบข้อที่ 17 ออกจากแบบสอบและทำการวิเคราะห์ค่าดัชนี CDIF และดัชนี DTF ใหม่ ได้ผลดังตารางที่ 68

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 68 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามเพศ เมื่อตัดข้อสอบข้อที่ 1, 33, 9, 91 และ 17 ออกจากแบบสอบ

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
1*	ตัดออกจากการวิเคราะห์	41*	0.018
2*	-0.003	42*	0.007
3*	0.005	43	0.002
4	0.008	44	0.002
5*	0.013	45	0.012
6*	0.004	46	0.008

7	0.003	47	0.008
8*	-0.002	48	-0.007
9*	ตัดออกจากการวิเคราะห์	49	-0.001
10	0.002	50	-0.004
11*	0.000	51	0.000
12*	0.001	52	-0.001
13	-0.001	53*	0.008
14	-0.007	54	0.009
15	-0.003	55	-0.006
16*	0.003	56	0.002
17*	ตัดออกจากการวิเคราะห์	57	0.001
18	0.000	58	0.003
19	0.002	59	0.000
20	-0.002	60	-0.008
21*	-0.014	61	-0.002
22	0.009	62	-0.009
23*	0.006	63	0.006
24*	0.009	64	0.004
25	-0.010	65	0.002
26	0.000	66*	0.006
27*	0.012	67	-0.002
28*	0.004	68*	0.020
29*	0.010	69	-0.001
30*	0.004	70	-0.002
31	0.001	71	-0.001
32	0.005	72	0.008
33	ตัดออกจากการวิเคราะห์	73*	-0.001
34*	0.001	74	0.004
35	0.000	75	0.001
36	0.002	76	-0.004
37	-0.001	77	0.003
38*	0.008	78	0.000
39	0.000	79	0.009
40*	0.001	80	-0.003

ตารางที่ 68 (ต่อ)

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
81	-0.013	91*	ตัดออกจากการวิเคราะห์
82	0.005	92	-0.003
83	-0.003	93*	0.009
84	0.003	94	-0.006
85	-0.005	95	0.001

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

86	0.005	96*	-0.001
87	0.006	97	0.000
88	-0.002	98	0.004
89	-0.007	99	-0.003
90	0.003	100	-0.002
ดัชนี DTF		0.137*	

*หมายเหตุ พิจารณาตามเกณฑ์จุดตัดของ Fleer (1993 อ้างถึงใน Raju และคณะ (1995)) ข้อสอบและแบบสอบทำหน้าที่ต่างกันเมื่อดัชนี NCDIF และดัชนี DTF มีค่ามากกว่า 0.006

จากตารางที่ 68 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามตัวแปรเพศ เมื่อตัดข้อสอบข้อที่ 1, 33, 9, 91 และ 17 ออกจากแบบสอบ พบว่าดัชนี DTF มีค่าเท่ากับ 0.137 ซึ่งมีค่ามากกว่า 0.006 จากตารางพบว่าข้อที่ 68 มีค่าดัชนี CDIF สูงที่สุด คือมีค่าเท่ากับ 0.020 เมื่อตัดข้อสอบข้อที่ 68 ออกจากแบบสอบและทำการวิเคราะห์ค่าดัชนี CDIF และดัชนี DTF ใหม่ ได้ผลดังตารางที่ 69

ตารางที่ 69 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามเพศ เมื่อตัดข้อสอบข้อที่ 1, 33, 9, 91,17 และ 68 ออกจากแบบสอบ

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
1*	ตัดออกจากการวิเคราะห์	16*	0.000
2*	0.002	17*	ตัดออกจากการวิเคราะห์
3*	0.002	18	-0.001
4	0.003	19	0.006
5*	0.005	20	-0.001
6*	0.005	21*	-0.012
7	0.001	22	0.004
8*	-0.001	23*	0.004
9*	ตัดออกจากการวิเคราะห์	24*	0.006
10	0.002	25	-0.002
11*	0.001	26	0.002
12*	-0.003	27*	0.006
13	-0.001	28*	0.001
14	-0.003	29*	0.005
15	-0.004	30*	0.004

ตารางที่ 69 (ต่อ)

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
31	0.002	66*	0.006
32	0.001	67	0.003

33	ตัดออกจากการวิเคราะห์	68*	ตัดออกจากการวิเคราะห์
34*	0.000	69	-0.003
35	0.001	70	0.000
36	0.000	71	0.001
37	0.000	72	0.004
38*	-0.001	73*	0.001
39	0.002	74	0.002
40*	0.002	75	0.000
41*	0.006	76	0.000
42*	0.003	77	0.003
43	0.004	78	0.000
44	0.002	79	0.008*
45	0.004	80	-0.003
46	0.006	81	-0.004
47	0.006	82	0.005
48	-0.005	83	-0.001
49	0.001	84	0.002
50	-0.003	85	-0.005
51	-0.003	86	-0.002
52	0.002	87	0.003
53*	0.006	88	0.001
54	0.004	89	-0.002
55	-0.004	90	0.003
56	0.003	91*	ตัดออกจากการวิเคราะห์
57	0.001	92	-0.003
58	0.002	93*	0.006
59	-0.003	94	-0.004
60	-0.007	95	0.003
61	-0.001	96*	0.001
62	-0.005	97	0.001
63	0.001	98	0.002
64	0.001	99	0.002
65	-0.001	100	-0.001
ดัชนี DTF	0.082*		

*หมายเหตุ พิจารณาตามเกณฑ์จุดตัดของ Fleer (1993 อ้างถึงใน Raju และคณะ (1995)) ข้อสอบและแบบสอบทำหน้าที่ต่างกันเมื่อดัชนี NCDIF และดัชนี DTF มีค่ามากกว่า 0.006

จากตารางที่ 69 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามตัวแปรเพศ เมื่อตัดข้อสอบข้อที่ 1, 33, 9, 91, 17 และ 68 ออกจากแบบสอบ พบว่าดัชนี

DTF มีค่าเท่ากับ 0.082 ซึ่งมีค่ามากกว่า 0.006 และมีนัยสำคัญทางสถิติที่ระดับ 0.01จากตารางพบว่าข้อที่ 79 มีค่าดัชนี CDIF สูงที่สุด คือมีค่าเท่ากับ 0.008 เมื่อตัดข้อสอบข้อที่ 79 ออกจากแบบสอบและทำการวิเคราะห์ค่าดัชนี CDIF และดัชนี DTF ได้ผลดังตารางที่ 70

ตารางที่ 70 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามเพศ เมื่อตัดข้อสอบข้อที่ 1, 33, 9, 91, 17, 68 และ 79 ออกจากแบบสอบ

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
1*	ตัดออกจากการวิเคราะห์	31	0.001
2*	0.000	32	0.003
3*	0.003	33	ตัดออกจากการวิเคราะห์
4	0.006	34*	-0.006
5*	0.009	35	0.001
6*	0.004	36	0.001
7	0.002	37	-0.001
8*	-0.001	38*	-0.005
9*	ตัดออกจากการวิเคราะห์	39	0.001
10	0.002	40*	0.001
11*	0.000	41*	0.012
12*	-0.001	42*	0.005
13	-0.001	43	0.003
14	-0.005	44	0.002
15	-0.003	45	0.008
16*	0.001	46	0.007
17*	ตัดออกจากการวิเคราะห์	47	0.007
18	0.000	48	-0.006
19	0.003	49	-0.007
20	-0.006	50	-0.003
21*	-0.023	51	-0.001
22	0.006	52	0.001
23*	0.005	53*	0.007
24*	0.007	54	0.007
25	-0.006	55	-0.005
26	0.001	56	0.003
27*	0.004	57	0.001
28*	0.003	58	0.003
29*	0.007	59	-0.001
30*	0.004	60	-0.007

ตารางที่ 70 (ต่อ)

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
61	-0.002	81	-0.009
62	-0.007	82	0.004
63	0.004	83	-0.001
64	0.002	84	0.003
65	0.001	85	-0.005
66*	0.006	86	-0.003
67	0.000	87	0.005
68*	ตัดออกจากการวิเคราะห์	88	-0.001
69	0.000	89	-0.004
70	-0.006	90	0.003
71	0.000	91*	ตัดออกจากการวิเคราะห์
72	0.006	92	-0.003
73*	0.000	93*	0.007
74	0.003	94	-0.005
75	0.000	95	0.001
76	-0.001	96*	0.000
77	0.003	97	0.001
78	0.000	98	0.003
79	ตัดออกจากการวิเคราะห์	99	-0.001
80	-0.002	100	-0.002
ดัชนี DTF	0.054		

*หมายเหตุ พิจารณาตามเกณฑ์จุดตัดของ Raju และคณะ (1995) ข้อสอบและแบบสอบทำหน้าที่ต่างกันเมื่อดัชนี NCDIF และดัชนี DTF มีค่ามากกว่า 0.006 และมีนัยสำคัญทางสถิติที่ระดับ 0.01

จากตารางที่ 70 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามตัวแปรเพศ เมื่อตัดข้อสอบข้อที่ 1, 33, 9, 91, 17, 68 และ 79 ออกจากแบบสอบ พบว่าดัชนี DTF มีค่าเท่ากับ 0.054 ซึ่งมีค่ามากกว่า 0.006 จากตารางพบว่าข้อที่ 41 มีค่าดัชนี CDIF สูงที่สุด คือมีค่าเท่ากับ 0.012 เมื่อตัดข้อสอบข้อที่ 41 ออกจากแบบสอบและทำการวิเคราะห์ค่าดัชนี CDIF และดัชนี DTF ใหม่ ได้ผลดังตารางที่ 71

ตารางที่ 71 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามเพศ เมื่อตัดข้อสอบข้อที่ 1, 33, 9, 91, 17, 68, 79 และ 41 ออกจากแบบสอบ

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
1*	ตัดออกจากการวิเคราะห์	6*	0.004
2*	0.002	7	0.001
3*	0.002	8*	-0.002

4	0.004	9*	ตัดออกจากการวิเคราะห์
5*	0.006	10	0.003

ตารางที่ 71 (ต่อ)

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
11*	0.000	56	0.004
12*	-0.005	57	0.002
13	-0.006	58	-0.004
14	-0.004	59	-0.003
15	-0.004	60	-0.008
16*	-0.001	61	-0.001
17*	ตัดออกจากการวิเคราะห์	62	-0.006
18	-0.005	63	0.001
19	0.005	64	0.001
20	-0.001	65	-0.001
21*	-0.023	66*	0.005
22	0.004	67	0.001
23*	0.004	68*	ตัดออกจากการวิเคราะห์
24*	0.005	69	0.001
25	-0.003	70	-0.008
26	0.002	71	0.002
27*	0.006	72	0.004
28*	0.000	73*	0.001
29*	0.006	74	0.003
30*	0.004	75	-0.006
31	0.002	76	0.000
32	0.001	77	0.004

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

33	ตัดออกจากการวิเคราะห์	78	-0.010
34*	-0.001	79	ตัดออกจากการวิเคราะห์
35	0.002	80	-0.003
36	-0.004	81	-0.005
37	-0.001	82	0.005
38*	-0.001	83	-0.001
39	0.002	84	0.002
40*	0.003	85	-0.006
41*	ตัดออกจากการวิเคราะห์	86	-0.002
42*	0.004	87	0.003
43	-0.004	88	0.001
44	0.002	89	-0.002
45	0.004	90	0.003
46	0.006	91*	ตัดออกจากการวิเคราะห์
47	0.008	92	-0.003
48	-0.006	93*	0.006
49	0.002	94	-0.004
50	-0.004	95	0.003
51	-0.003	96*	0.001
52	0.003	97	0.001
53*	0.006	98	0.003
54	0.004	99	-0.002
55	-0.005	100	0.001
ดัชนี DTF	0.018*		

จากตารางที่ 71 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามตัวแปรเพศ เมื่อตัดข้อสอบข้อที่ 1, 33, 9, 91, 17, 68, 79 และ 41 ออกจากแบบสอบ พบว่าดัชนี DTF มีค่าเท่ากับ 0.018 ซึ่งมีค่ามากกว่า 0.006 จากตารางพบว่าข้อที่ 47 มีค่าดัชนี CDIF สูงที่สุด คือมีค่าเท่ากับ 0.008 เมื่อตัดข้อสอบข้อที่ 47 ออกจากแบบสอบและทำการวิเคราะห์ค่าดัชนี CDIF และดัชนี DTF ใหม่ ได้ผลดังตารางที่ 72

ตารางที่ 72 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบ จำแนกตามเพศ เมื่อตัดข้อสอบข้อที่ 1, 33, 9, 91, 17, 68, 79, 41 และ 47 ออกจากแบบสอบ

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
1*	ตัดออกจากการวิเคราะห์	31	0.005
2*	-0.002	32	0.012
3*	0.001	33	ตัดออกจากการวิเคราะห์
4	0.020	34*	0.004
5*	0.018	35	0.007
6*	-0.008	36	0.004
7	0.006	37	-0.001
8*	-0.008	38*	-0.017

9*	ตัดออกจากการวิเคราะห์	39	0.006
10	0.009	40*	0.005
11*	-0.002	41*	ตัดออกจากการวิเคราะห์
12*	-0.017	42*	0.017
13	-0.002	43	0.011
14	-0.016	44	0.007
15	-0.012	45	0.019
16*	-0.005	46	0.013
17*	ตัดออกจากการวิเคราะห์	47	ตัดออกจากการวิเคราะห์
18	0.001	48	0.017
19	0.007	49	0.003
20	-0.005	50	-0.013
21*	-0.043	51	-0.003
22	0.022	52	0.004
23*	-0.200	53*	0.010
24*	0.000	54	0.011
25	-0.022	55	-0.014
26	0.002	56	0.008
27*	0.019	57	0.007
28*	-0.005	58	0.010
29*	0.014	59	-0.002
30*	0.011	60	-0.021

ตารางที่ 72 (ต่อ)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
61	-0.005	81	-0.028
62	-0.021	82	0.015
63	0.014	83	-0.006
64	0.009	84	0.010
65	0.005	85	-0.013
66*	0.008	86	-0.010
67	-0.004	87	0.015
68*	ตัดออกจากการวิเคราะห์	88	-0.001
69	0.000	89	-0.015
70	-0.001	90	0.150
71	0.006	91*	ตัดออกจากการวิเคราะห์
72	0.010	92	-0.009
73*	0.049	93*	0.021
74	0.012	94	-0.015
75	0.000	95	0.003
76	-0.005	96*	0.001
77	-0.005	97	0.002
78	0.001	98	0.010
79	ตัดออกจากการวิเคราะห์	99	-0.002
80	-0.006	100	-0.007
ดัชนี DTF	0.011*		

*หมายเหตุ พิจารณาตามเกณฑ์จุดตัดของ Fleer (1993 อ้างถึงใน Raju และคณะ (1995)) ข้อสอบและแบบสอบทำหน้าที่ต่างกันเมื่อดัชนี NCDIF และดัชนี DTF มีค่ามากกว่า 0.006

จากตารางที่ 72 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามตัวแปรเพศ เมื่อตัดข้อสอบข้อที่ 1, 33, 9, 91, 17, 68, 79, 41 และ 47 ออกจากแบบสอบ พบว่าดัชนี DTF มีค่าเท่ากับ 0.011 ซึ่งมีค่ามากกว่า 0.006 จากตารางพบว่าข้อที่ 90 มีค่าดัชนี CDIF สูงที่สุด คือมีค่าเท่ากับ 0.150 เมื่อตัดข้อสอบข้อที่ 90 ออกจากแบบสอบและทำการวิเคราะห์ค่าดัชนี CDIF และดัชนี DTF ใหม่ ได้ผลดังตารางที่ 38

ตารางที่ 73 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามเพศ เมื่อตัดข้อสอบข้อที่ 1, 33, 9, 91, 17, 68, 79, 41, 47 และ 90 ออกจากแบบสอบ

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
1*	ตัดออกจากการวิเคราะห์	41*	ตัดออกจากการวิเคราะห์
2*	0.003	42*	0.005
3*	0.000	43	-0.007
4	0.005	44	0.003
5*	0.007	45	0.005
6*	0.000	46	0.011
7	0.002	47	ตัดออกจากการวิเคราะห์
8*	-0.003	48	-0.008
9*	ตัดออกจากการวิเคราะห์	49	0.003
10	0.004	50	-0.006
11*	0.000	51	-0.005
12*	-0.019	52	0.005
13	-0.001	53*	0.007
14	-0.005	54	0.006
15	-0.007	55	-0.007
16*	-0.004	56	0.005
17*	ตัดออกจากการวิเคราะห์	57	0.004
18	-0.002	58	0.003
19	0.008	59	-0.005
20	-0.002	60	-0.011
21*	-0.023	61	-0.001
22	0.006	62	-0.007
23*	0.005	63	0.002
24*	0.001	64	0.001
25	-0.004	65	-0.001
26	0.003	66*	0.010
27*	0.008	67	0.002
28*	-0.004	68*	ตัดออกจากการวิเคราะห์
29*	0.005	69	-0.002
30*	0.006	70	0.001
31	0.003	71	0.004
32	0.002	72	0.005
33	ตัดออกจากการวิเคราะห์	73*	0.001
34*	-0.001	74	-0.005
35	0.005	75	0.000
36	0.001	76	0.001
37	0.000	77	0.008
38*	-0.001	78	-0.011

39	0.004	79	ตัดออกจากการวิเคราะห์
40*	0.003	80	-0.003

ตารางที่ 73 (ต่อ)

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
81	-0.006	91*	ตัดออกจากการวิเคราะห์
82	0.008	92	-0.005
83	-0.001	93*	0.008
84	0.003	94	-0.006
85	-0.008	95	0.006
86	-0.012	96*	0.002
87	0.005	97	-0.002
88	0.003	98	0.004
89	-0.003	99	0.004
90	ตัดออกจากการวิเคราะห์	100	-0.002
ดัชนี DTF	0.011*		

*หมายเหตุ พิจารณาตามเกณฑ์จุดตัดของ Fleer (1993 อ้างถึงใน Raju และคณะ (1995)) ข้อสอบและแบบสอบทำหน้าที่ต่างกันเมื่อดัชนี NCDIF และดัชนี DTF มีค่ามากกว่า 0.006

จากตารางที่ 73 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามตัวแปรเพศ เมื่อตัดข้อสอบข้อที่ 1, 33, 9, 91, 17, 68, 79, 41, 47และ 90 ออกจากแบบสอบ พบว่าดัชนี DTF มีค่าเท่ากับ 0.011 ซึ่งมีค่ามากกว่า 0.006 จากตารางพบว่าข้อที่ 46 มีค่าดัชนี CDIF สูงที่สุด คือมีค่าเท่ากับ 0.011 เมื่อตัดข้อสอบข้อที่ 46 ออกจากแบบสอบและทำการวิเคราะห์ค่าดัชนี CDIF และดัชนี DTF ใหม่ ได้ผลดังตารางที่ 74

ตารางที่ 74 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามเพศ เมื่อตัดข้อสอบข้อที่ 1, 33, 9, 91, 17, 68, 79, 41, 47, 90 และ 46 ออกจากแบบสอบ

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
1*	ตัดออกจากการวิเคราะห์	21*	-0.018
2*	0.003	22	0.005
3*	0.000	23*	0.004
4	0.005	24*	0.001
5*	0.006	25	-0.003
6*	0.000	26	-0.003
7	0.001	27*	0.007
8*	-0.002	28*	-0.004
9*	ตัดออกจากการวิเคราะห์	29*	0.007
10	0.003	30*	0.005
11*	0.000	31	0.003
12*	-0.011	32	0.001
13	0.000	33	ตัดออกจากการวิเคราะห์

14	-0.004	34*	0.000
15	-0.006	35	0.005
16*	-0.004	36	0.001
17*	ตัดออกจากการวิเคราะห์	37	0.000
18	-0.001	38*	-0.001
19	0.007	39	0.004
20	-0.001	40*	0.004

ตารางที่ 74 (ต่อ)

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
41*	ตัดออกจากการวิเคราะห์	71	0.004
42*	0.004	72	0.005
43	0.007	73*	0.001
44	0.003	74	0.004
45	0.004	75	0.000
46	ตัดออกจากการวิเคราะห์	76	0.001
47	ตัดออกจากการวิเคราะห์	77	0.008
48	-0.007	78	-0.001
49	0.003	79	ตัดออกจากการวิเคราะห์
50	-0.006	80	-0.003
51	-0.005	81	-0.006
52	0.005	82	0.007
53*	0.007	83	0.000
54	0.005	84	0.003
55	-0.006	85	-0.007
56	0.005	86	-0.002
57	0.003	87	0.004
58	0.003	88	0.003
59	-0.005	89	-0.002
60	-0.010	90	ตัดออกจากการวิเคราะห์
61	-0.001	91*	ตัดออกจากการวิเคราะห์
62	-0.006	92	-0.005
63	0.002	93*	0.007
64	0.002	94	-0.005
65	-0.001	95	0.006
66*	0.009	96*	0.001
67	0.001	97	0.001
68*	ตัดออกจากการวิเคราะห์	98	0.004
69	0.001	99	0.004
70	0.001	100	-0.002
ดัชนี DTF	0.009*		

*หมายเหตุ พิจารณาตามเกณฑ์จุดตัดของ Fleer (1993 อ้างถึงใน Raju และคณะ (1995)) ข้อสอบและแบบสอบ

ทำหน้าที่ต่างกันเมื่อดัชนี NCDIF และดัชนี DTF มีค่ามากกว่า 0.006

จากตารางที่ 74 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามตัวแปรเพศ เมื่อตัดข้อสอบข้อที่ 1, 33, 9, 91, 17, 68, 79, 41, 47, 90 และ 46 ออกจากแบบสอบ พบว่าดัชนี DTF มีค่าเท่ากับ 0.009 ซึ่งมีค่ามากกว่า 0.006 จากตารางพบว่าข้อที่ 66 มีค่าดัชนี CDIF สูงที่สุด คือมีค่าเท่ากับ 0.009 เมื่อตัดข้อสอบข้อที่ 66 ออกจากแบบสอบและทำการวิเคราะห์ค่าดัชนี CDIF และดัชนี DTF ใหม่ ได้ผลดังตารางที่ 75

ตารางที่ 75 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามเพศ เมื่อตัดข้อสอบข้อที่ 1, 33, 9, 91, 17, 68, 79, 41, 47, 90, 46 และ 66 ออกจากแบบสอบ

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
1*	ตัดออกจากการวิเคราะห์	41*	ตัดออกจากการวิเคราะห์
2*	0.009	42*	0.000
3*	-0.003	43	0.010
4	-0.001	44	0.003
5*	-0.002	45	-0.007
6*	0.000	46	ตัดออกจากการวิเคราะห์
7	0.000	47	ตัดออกจากการวิเคราะห์
8*	-0.002	48	-0.007
9*	ตัดออกจากการวิเคราะห์	49	0.007
10	0.003	50	-0.007
11*	0.001	51	-0.011
12*	-0.020	52	0.010
13	-0.001	53*	0.004
14	-0.001	54	-0.002
15	-0.008	55	-0.006
16*	-0.008	56	0.006
17*	ตัดออกจากการวิเคราะห์	57	0.004
18	-0.003	58	0.001
19	0.014	59	-0.010
20	0.037	60	-0.011
21*	-0.019	61	0.001
22	0.000	62	-0.002
23*	0.002	63	-0.005
24*	-0.003	64	-0.002
25	0.005	65	-0.006
26	0.006	66*	ตัดออกจากการวิเคราะห์
27*	-0.001	67	0.007
28*	-0.009	68*	ตัดออกจากการวิเคราะห์

29*	0.002	69	0.004
30*	0.006	70	0.003
31	0.004	71	0.007
32	-0.003	72	-0.001
33	ตัดออกจากการวิเคราะห์	73*	0.004
34*	-0.001	74	0.003
35	0.009	75	-0.001
36	0.000	76	0.005
37	0.000	77	0.010
38*	-0.008	78	-0.002
39	-0.008	79	ตัดออกจากการวิเคราะห์
40*	0.005	80	-0.003

ตารางที่ 75 (ต่อ)

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
81	0.005	91*	ตัดออกจากการวิเคราะห์
82	0.008	92	-0.006
83	0.002	93*	0.005
84	0.001	94	-0.004
85	-0.010	95	0.012
86	0.002	96*	0.005
87	0.001	97	0.003
88	0.008	98	0.004
89	0.004	99	0.011
90	ตัดออกจากการวิเคราะห์	100	-0.001
ดัชนี DTF	0.009*		

*หมายเหตุ พิจารณาตามเกณฑ์จุดตัดของ Fleer (1993 อ้างถึงใน Raju และคณะ (1995)) ข้อสอบและแบบสอบทำหน้าที่ต่างกันเมื่อดัชนี NCDIF และดัชนี DTF มีค่ามากกว่า 0.006

จากตารางที่ 75 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามตัวแปรเพศ เมื่อตัดข้อสอบข้อที่ 1, 33, 9, 91, 17, 68, 79, 41,47, 90, 46 และ 66 ออกจากแบบสอบ พบว่าดัชนี DTF มีค่าเท่ากับ 0.009 ซึ่งมีค่ามากกว่า 0.006 จากตารางพบว่าข้อที่ 20 มีค่าดัชนี CDIF สูงที่สุด คือมีค่าเท่ากับ 0.037 เมื่อตัดข้อสอบข้อที่ 20 ออกจากแบบสอบและทำการวิเคราะห์ค่าดัชนี CDIF และดัชนี DTF ใหม่ ได้ผลดังตารางที่ 19 ในบทที่ 4

ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบ สำหรับแบบสอบวิชาภาษาอังกฤษ เมื่อแบ่งกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา

ตารางที่ 76 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา เมื่อตัดข้อสอบข้อที่ 33 และข้อที่ 24 ออกจากแบบสอบ

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
1*	-0.001	11	0.007
2	-0.004	12	0.001
3	0.008	13	0.002
4	0.002	14	0.001
5	0.007	15	-0.001
6	0.006	16*	0.010
7	0.002	17	-0.007
8	-0.002	18	-0.001
9*	-0.012	19	0.000
10	0.000	20	-0.003

ตารางที่ 76 (ต่อ)

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
21	-0.001	61	0.005
22	0.004	62	-0.001
23	-0.003	63	-0.002
24*	ตัดออกจากการวิเคราะห์	64	0.004
25	0.004	65	-0.003
26	0.006	66*	-0.005
27	-0.003	67	0.004
28	0.005	68	0.005
29	0.003	69	-0.003
30	0.005	70	0.006
31	0.003	71	0.006
32	-0.001	72	0.006
33*	ตัดออกจากการวิเคราะห์	73*	-0.002
34	0.004	74	0.004
35	0.000	75	0.003
36	-0.003	76	-0.002
37	0.009	77*	-0.015
38	-0.001	78	-0.004
39*	0.007	79	0.003
40	-0.003	80	0.000
41	-0.001	81	-0.004
42*	-0.004	82	0.005
43	-0.004	83	-0.003
44*	-0.009	84	0.000
45	0.005	85	-0.002

46	-0.004	86	0.000
47	-0.003	87	-0.004
48	0.001	88	-0.006
49	-0.003	89	0.003
50	-0.001	90	0.002
51	0.002	91	0.000
52*	0.005	92	-0.003
53	0.006	93*	-0.002
54	0.002	94	-0.001
55	0.002	95	0.004
56*	0.002	96	0.000
57*	0.005	97	0.001
58	0.001	98	-0.002
59	0.003	99	-0.003
60*	0.007	100	-0.005
ดัชนี DTF	0.105*		

*หมายเหตุ พิจารณาตามเกณฑ์จุดตัดของ Fleer (1993 อ้างถึงใน Raju และคณะ (1995)) ข้อสอบและแบบสอบทำหน้าที่ต่างกันเมื่อดัชนี NCDIF และดัชนี DTF มีค่ามากกว่า 0.006

จากตารางที่ 76 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามตัวแปรสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา เมื่อตัดข้อสอบข้อที่ 33 และข้อที่ 24 ออก พบว่าดัชนี DTF มีค่าเท่ากับ 0.105 ซึ่งมีค่ามากกว่า 0.006 แสดงว่าแบบสอบดังกล่าวยังทำหน้าที่ต่างกัน จากตารางพบว่าข้อที่ 37 มีค่าดัชนี CDIF สูงที่สุด คือมีค่าเท่ากับ 0.009 เมื่อตัดข้อสอบข้อที่ 37 ออกจากแบบสอบและทำการวิเคราะห์ค่าดัชนี CDIF และดัชนี DTF ได้ผลดังตารางที่ 77

ตารางที่ 77 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา เมื่อตัดข้อสอบข้อที่ 33 , 24 และ 37 ออกจากแบบสอบ

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
1*	0.017	31	0.011
2	-0.005	32	0.002
3	0.006	33*	ตัดออกจากการวิเคราะห์
4	0.006	34	0.003
5	0.015	35	-0.005
6	0.003	36	0.000
7	0.000	37	ตัดออกจากการวิเคราะห์
8	-0.004	38	0.002
9*	-0.008	39*	0.004
10	0.011	40	-0.001
11	0.004	41	-0.004

12	0.001	42*	0.006
13	0.003	43	-0.002
14	-0.005	44*	-0.005
15	-0.002	45	0.011
16*	0.003	46	0.001
17	-0.006	47	-0.009
18	-0.003	48	-0.005
19	0.001	49	-0.006
20	-0.001	50	-0.004
21	0.004	51	-0.002
22	0.007	52*	0.004
23	0.002	53	0.011
24*	ตัดออกจากการวิเคราะห์	54	-0.005
25	0.007	55	-0.005
26	0.003	56*	-0.008
27	0.002	57*	0.006
28	0.003	58	-0.007
29	0.006	59	0.002
30	0.005	60*	0.016

ตารางที่ 77 (ต่อ)

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
61	-0.001	81	-0.006
62	-0.007	82	0.017
63	-0.009	83	-0.003
64	0.000	84	0.003
65	-0.008	85	-0.003
66*	0.001	86	0.004
67	0.002	87	-0.010
68	0.018	88	-0.011
69	-0.003	89	-0.002
70	0.004	90	0.001
71	0.008	91	0.004
72	0.013	92	-0.005
73*	0.011	93*	0.012
74	-0.003	94	-0.005
75	0.004	95	0.006
76	-0.007	96	0.006
77*	-0.012	97	0.001

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

78	-0.006	98	-0.002
79	-0.005	99	-0.002
80	0.008	100	-0.004
ดัชนี DTF		0.078*	

*หมายเหตุ พิจารณาตามเกณฑ์จุดตัดของ Fleer (1993 อ้างถึงใน Raju และคณะ (1995)) ข้อสอบและแบบสอบ

จากตารางที่ 77 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามตัวแปรสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา เมื่อตัดข้อสอบข้อที่ 33 , 24 และ 37 ออก พบว่าดัชนี DTF มีค่าเท่ากับ 0.078 ซึ่งมีค่ามากกว่า 0.006 แสดงว่าแบบสอบดังกล่าวยังทำหน้าที่ต่างกัน จากตารางพบว่าข้อที่ 68 มีค่าดัชนี CDIF สูงที่สุด คือมีค่าเท่ากับ 0.018 เมื่อตัดข้อสอบข้อที่ 68 ออกจากแบบสอบและทำการวิเคราะห์ค่าดัชนี CDIF และดัชนี DTF ได้ผลดังตารางที่ 78

ตารางที่ 78 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา เมื่อตัดข้อสอบข้อที่ 33 , 24 ,37 และ 68 ออกจากแบบสอบ

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
1*	0.025	41	-0.003
2	-0.003	42*	0.014
3	0.000	43	0.004
4	0.006	44*	0.007
5	0.011	45	0.009
6	-0.003	46	0.007
7	-0.002	47	-0.008
8	-0.004	48	-0.008
9*	0.005	49	-0.006

10	0.016	50	-0.004
11	-0.002	51	-0.004
12	0.002	52*	-0.006
13	0.002	53	0.007
14	-0.009	54	-0.009
15	-0.003	55	-0.010
16*	-0.008	56*	-0.013
17	0.001	57*	0.001
18	-0.003	58	-0.011
19	0.001	59	-0.001
20	-0.001	60*	0.014
21	0.004	61	-0.008
22	0.005	62	-0.007
23	0.007	63	-0.010
24*	ตัดออกจากการวิเคราะห์	64	-0.006
25	0.004	65	-0.007
26	-0.005	66*	0.008
27	0.006	67	-0.004
28	0.000	68	ตัดออกจากการวิเคราะห์
29	0.004	69	0.000
30	-0.002	70	0.000
31	0.011	71	0.005
32	0.005	72	0.010
33*	ตัดออกจากการวิเคราะห์	73*	0.018
34	-0.001	74	-0.009
35	-0.005	75	0.001
36	0.004	76	-0.007
37	ตัดออกจากการวิเคราะห์	77*	-0.012
38	0.003	78	-0.004
39*	-0.003	79	-0.011
40	0.003	80	0.010

ตารางที่ 78 (ต่อ)

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
81	-0.003	91	0.005
82	0.017	92	-0.003
83	0.000	93*	0.020
84	0.003	94	-0.006
85	-0.001	95	0.004
86	0.006	96	0.008
87	-0.008	97	-0.001
88	-0.007	98	-0.001
89	-0.007	99	0.001

90	-0.002	100	0.002
ดัชนี DTF	0.072*		

*หมายเหตุ พิจารณาตามเกณฑ์จุดตัดของ Fleer (1993 อ้างถึงใน Raju และคณะ (1995)) ข้อสอบและแบบสอบทำหน้าที่ต่างกันเมื่อดัชนี NCDIF และดัชนี DTF มีค่ามากกว่า 0.006

จากตารางที่ 78 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามตัวแปรสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา เมื่อตัดข้อสอบข้อที่ 33 , 24 , 37 และ 68 ออก พบว่าดัชนี DTF มีค่าเท่ากับ 0.072 ซึ่งมีค่ามากกว่า 0.006 แสดงว่าแบบสอบดังกล่าวยังทำหน้าที่ต่างกัน จากตารางพบว่าข้อที่ 1 มีค่าดัชนี CDIF สูงที่สุด คือมีค่าเท่ากับ 0.025 เมื่อตัดข้อสอบข้อที่ 1 ออกจากแบบสอบและทำการวิเคราะห์ค่าดัชนี CDIF และดัชนี DTF ได้ผลดังตารางที่ 79

ตารางที่ 79 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา เมื่อตัดข้อสอบข้อที่ 33 , 24 ,37 ,68 และ 1 ออกจากแบบสอบ

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
1*	ตัดออกจากการวิเคราะห์	16*	0.001
2	0.000	17	0.001
3	0.005	18	-0.004
4	0.002	19	0.000
5	0.003	20	0.001
6	0.002	21	0.003
7	-0.001	22	0.001
8	-0.001	23	0.003
9*	0.004	24*	ตัดออกจากการวิเคราะห์
10	0.005	25	0.001
11	0.004	26	-0.002
12	0.002	27	0.003
13	0.000	28	0.003
14	-0.003	29	0.001
15	-0.001	30	-0.004

ตารางที่ 79 (ต่อ)

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
31	0.003	66*	0.004
32	0.002	67	-0.005
33*	ตัดออกจากการวิเคราะห์	68	ตัดออกจากการวิเคราะห์
34	0.003	69	0.000
35	-0.002	70	0.003
36	0.002	71	0.005
37	ตัดออกจากการวิเคราะห์	72	0.002
38	0.002	73*	0.007
39*	0.001	74	-0.003
40	0.001	75	0.000
41	-0.003	76	-0.002
42*	0.006	77*	0.000
43	0.002	78	-0.007
44*	0.004	79	-0.006
45	0.003	80	0.003
46	0.003	81	-0.001
47	-0.004	82	0.004
48	-0.003	83	0.000
49	-0.002	84	0.001
50	-0.003	85	0.000
51	-0.001	86	0.002
52*	-0.005	87	-0.002
53	0.001	88	-0.004
54	-0.003	89	-0.003
55	-0.003	90	-0.001
56*	-0.005	91	0.002
57*	0.002	92	-0.004
58	-0.004	93*	0.007
59	0.000	94	-0.003
60*	0.004	95	0.002
61	-0.003	96	0.003
62	-0.002	97	0.000
63	-0.003	98	0.000
64	-0.002	99	0.001
65	-0.004	100	-0.002
ดัชนี DTF	0.019*		

*หมายเหตุ พิจารณาตามเกณฑ์จุดตัดของ Fleer (1993 อ้างถึงใน Raju และคณะ (1995)) ข้อสอบและแบบสอบทำหน้าที่ต่างกันเมื่อดัชนี NCDIF และดัชนี DTF มีค่ามากกว่า 0.006

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

จากตารางที่ 79 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามตัวแปรสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา เมื่อตัดข้อสอบข้อที่ 33, 24, 37, 68 และ 1 ออก พบว่าดัชนี DTF มีค่าเท่ากับ 0.019 ซึ่งมีค่ามากกว่า 0.006 แสดงว่าแบบสอบดังกล่าวยังทำหน้าที่ต่างกัน จากตารางพบว่าข้อที่ 73 มีค่าดัชนี CDIF สูงที่สุด คือมีค่าเท่ากับ 0.007 เมื่อตัดข้อสอบข้อที่ 73 ออกจากแบบสอบและทำการวิเคราะห์ค่าดัชนี CDIF และดัชนี DTF ได้ผลดังตารางที่ 80

ตารางที่ 80 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา เมื่อตัดข้อสอบข้อที่ 33, 24, 37, 68, 1 และ 73 ออกจากแบบสอบ

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
1*	ตัดออกจากการวิเคราะห์	31	0.010
2	0.004	32	0.000
3	0.009	33*	ตัดออกจากการวิเคราะห์
4	0.003	34	0.012
5	0.009	35	-0.006
6	0.006	36	0.004
7	0.000	37	ตัดออกจากการวิเคราะห์
8	0.000	38	-0.007
9*	-0.025	39*	0.025
10	0.008	40	-0.013
11	0.011	41	-0.010
12	0.009	42*	0.000
13	0.003	43	-0.003
14	-0.004	44*	-0.008
15	-0.001	45	0.012
16*	0.015	46	-0.001
17	-0.013	47	-0.015
18	-0.004	48	-0.005
19	0.000	49	-0.010
20	0.007	50	-0.002
21	0.006	51	0.000
22	0.008	52*	0.010
23	-0.005	53	0.011
24*	ตัดออกจากการวิเคราะห์	54	-0.004
25	0.007	55	-0.006
26	0.002	56*	-0.007
27	-0.002	57*	0.020

28	0.010	58	-0.006
29	0.002	59	0.007
30	0.001	60*	0.019

ตารางที่ 80 (ต่อ)

ข้อที่	ดัชนี CDIF	ข้อที่	ดัชนี CDIF
61	0.002	81	-0.010
62	-0.007	82	0.016
63	-0.012	83	-0.005
64	0.003	84	0.001
65	-0.015	85	-0.002
66*	0.081	86	0.005
67	0.000	87	-0.012
68	ตัดออกจากการวิเคราะห์	88	-0.012
69	-0.004	89	-0.002
70	0.015	90	0.000
71	0.022	91	-0.001
72	0.012	92	-0.005
73*	ตัดออกจากการวิเคราะห์	93*	0.007
74	0.003	94	-0.006
75	0.006	95	0.019
76	-0.009	96	0.001
77*	-0.011	97	0.001
78	-0.008	98	-0.003
79	-0.004	99	-0.005
80	0.005	100	-0.004
ดัชนี DTF	0.009*		

*หมายเหตุ พิจารณาตามเกณฑ์จุดตัดของ Fleer (1993 อ้างถึงใน Raju และคณะ (1995)) ข้อสอบและแบบสอบทำหน้าที่ต่างกันเมื่อดัชนี NCDIF และดัชนี DTF มีค่ามากกว่า 0.006

จากตารางที่ 80 ผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบวิชาภาษาอังกฤษ สำหรับกลุ่มผู้สอบจำแนกตามตัวแปรสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา เมื่อตัดข้อสอบข้อที่ 33, 24, 37, 68, 1 และ 73 ออก พบว่าดัชนี DTF มีค่าเท่ากับ 0.009 ซึ่งมีค่ามากกว่า 0.006 แสดงว่าแบบสอบดังกล่าวยังทำหน้าที่ต่างกัน จากตารางพบว่าข้อที่ 66 มีค่าดัชนี CDIF สูงที่สุด คือมีค่าเท่ากับ 0.081 เมื่อตัดข้อสอบข้อที่ 66 ออกจากแบบสอบและทำการวิเคราะห์ค่าดัชนี CDIF และดัชนี DTF ได้ผลดังตารางที่ 22

ในบทที่ 4 ผลการวิเคราะห์ข้อมูล ดังที่กล่าวมาแล้ว

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวรักชนก ยี่สุ่นศรี เกิดเมื่อวันที่ 29 มีนาคม พ.ศ. 2521 สำเร็จการศึกษาหลักสูตรปริญญาวิทยาศาสตรบัณฑิต สาขาวิชาคณิตศาสตร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2541 และเข้าศึกษาต่อในหลักสูตรครุศาสตรมหาบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2542

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย