การวิเคราะห์เปรียบเทียบประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี

นางสาวสุภะ อภิญญาภิบาล

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ ที่ส่งผ่านทางบัณฑิตวิทยาลัย
The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR)
are the thesis authors' files submitted through the University Graduate School.

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาครุศาสตรดุษฎีบัณฑิต
สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา
คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2558
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

A COMPARATIVE ANALYSIS OF THE EFFICIENCY OF DIFFERENTIAL ITEM FUNCTIONING DETECTION IN DICHOTOMOUSLY SCORED ITEMS AMONG LOGISTIC REGRESSION, SIBTEST AND RASCHTREE METHODS

Miss Supa Apinyapibal

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Educational Measurement and Evaluation
Department of Educational Research and Psychology
Faculty of Education
Chulalongkorn University
Academic Year 2015

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การวิเคราะห์เปรียบเทียบประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี
โดย	นางสาวสุภะ อภิญญาภิบาล
สาขาวิชา	การวัดและประเมินผลการศึกษา
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ ดร.ณัฏฐภรณ์ หลาวทอง
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม	ศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาดุษฎีบัณฑิต

.....คณบดีคณะครุศาสตร์

(รองศาสตราจารย์ ดร. บัญชา ชลาภิรมย์)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ

(รองศาสตราจารย์ ดร.ศิริเดช สุชีวะ)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(ผู้ช่วยศาสตราจารย์ ดร.ณัฏฐภรณ์ หลาวทอง)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม

(ศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี)

.....กรรมการ

(รองศาสตราจารย์ ดร. โชติกา ภาษีผล)

.....กรรมการ

(รองศาสตราจารย์ ดร. กมลวรรณ ตังธนกานนท์)

.....กรรมการภายนอกมหาวิทยาลัย

(รองศาสตราจารย์ ดร. สมถวิล วิจิตรวรรณา)

สุภะ อภิญญาภิบาล : การวิเคราะห์เปรียบเทียบประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี (A COMPARATIVE ANALYSIS OF THE EFFICIENCY OF DIFFERENTIAL ITEM FUNCTIONING DETECTION IN DICHOTOMOUSLY SCORED ITEMS AMONG LOGISTIC REGRESSION, SIBTEST AND RASCHTREE METHODS) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: ผศ. ดร. ณัฏฐภรณ์ หลาวทอง, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม: ศ. ดร.ศิริชัย กาญจนวาสี, 305 หน้า.

การศึกษาครั้งนี้ มีวัตถุประสงค์เพื่อเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี โดยการทดสอบระดับนัยสำคัญและการวัดขนาดอิทธิพล ข้อมูลที่ใช้ศึกษาจำลองภายใต้ทฤษฎีการตอบสนองข้อสอบ โมเดล 1 พารามิเตอร์ แล้วจัดกระทำข้อมูลตามปัจจัย 4 ปัจจัย คือ (1) ค่าความยากของข้อสอบ 3 ระดับ (2) ความยาวของแบบสอบ 2 ขนาด (3) สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด และ (4) ขนาดกลุ่มตัวอย่าง 3 ขนาด รวมข้อมูลที่ศึกษาทั้งหมด 36 เงื่อนไข ในแต่ละเงื่อนไขจำลองซ้ำ 50 ครั้ง และข้อมูลเชิงประจักษ์ศึกษาการทำหน้าที่ต่างกันของข้อสอบตามตัวแปร 3 ตัว คือ เพศ สังกัดของโรงเรียน และภูมิภาค แล้ววิเคราะห์ข้อมูลด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบทั้งหมด ใช้ระดับนัยสำคัญ .05

ผลการวิจัยสรุปได้ดังนี้

1. อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญของวิธีซิปเทสท์มีค่าสูงกว่าวิธีการถดถอยโลจิสติกและวิธีราสช์ทรีภายใต้เกือบทุกเงื่อนไข และอำนาจการทดสอบของวิธีราสช์ทรีมีค่าสูงกว่าวิธีการถดถอยโลจิสติกภายใต้เกือบทุกเงื่อนไข ส่วนอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพลของวิธีซิปเทสท์มีค่าสูงกว่าวิธีการถดถอยโลจิสติกภายใต้เกือบทุกเงื่อนไข

2. อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญของวิธีราสช์ทรีมีค่าต่ำกว่าวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ภายใต้เกือบทุกเงื่อนไข และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์มีค่าต่ำกว่าวิธีการถดถอยโลจิสติกภายใต้เกือบทุกเงื่อนไข ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพลของวิธีการถดถอยโลจิสติกมีค่าต่ำกว่าวิธีซิปเทสท์ภายใต้เกือบทุกเงื่อนไข

3. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับข้อมูลเชิงประจักษ์ด้วยวิธีราสช์ทรีมีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ แต่มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าเกณฑ์ที่กำหนด ที่ระดับ 10%

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาควิชา	วิจัยและจิตวิทยาการศึกษา	ลายมือชื่อนิสิต .....
สาขาวิชา	การวัดและประเมินผลการศึกษา	ลายมือชื่อ อ.ที่ปรึกษาหลัก .....
ปีการศึกษา	2558	ลายมือชื่อ อ.ที่ปรึกษาร่วม .....

5384474327 : MAJOR EDUCATIONAL MEASUREMENT AND EVALUATION
KEYWORDS: DIFFERENTIAL ITEM FUNCTIONING / LOGISTIC REGRESSION / SIBTEST / RASCHTREE / DICHOTOMOUSLY

SUPA APINYAPIBAL: A COMPARATIVE ANALYSIS OF THE EFFICIENCY OF DIFFERENTIAL ITEM FUNCTIONING DETECTION IN DICHOTOMOUSLY SCORED ITEMS AMONG LOGISTIC REGRESSION, SIBTEST AND RASCHTREE METHODS. ADVISOR: ASST. PROF. NUTTAPORN LAWTHONG, Ph.D., CO-ADVISOR: PROF. SIRICHAI KANJANAWASEE, Ph.D., 305 pp.

The objectives of this study were to compare power rate and Type I error rate of differential item functioning (DIF) with dichotomously scored items among Logistic Regression, SIBTEST and Raschtree methods by significance tests and effect size measures. In this study, the data were simulated under the IRT theory of one-parameter item response, simulating dichotomous response under the condition of 4 factors which were (1) difficulty of 3-level items (2) 2-size test length (3) proportion of 2-size DIF and (4) 3-size sample size. The total studied data were 36 conditions each of which was replicated 50 times. The empirical data were used to study DIF in 3 factors: sex, stratum and region. The data were analyzed by Logistic Regression, SIBTEST and Raschtree methods. The significance level of .05 was applied in all DIF analyses.

Results of the research were the following:

1. The power rate of SIBTEST method in detecting differential item functioning by significance tests was higher than Logistic Regression and Raschtree methods under almost all conditions and the power rate of Raschtree method was higher than Logistic Regression method under almost all conditions. The power rate of SIBTEST method in detecting differential item functioning by effect size measures was higher than Logistic Regression method under almost all conditions.

2. Type I error rate of Raschtree method in detecting differential item functioning by significance tests was lower than Logistic Regression and SIBTEST methods under almost all conditions and type I error rate of SIBTEST method was lower than Logistic Regression method under almost all conditions. Type I error rate of Logistic Regression method in detecting differential item functioning by effect size measures was lower than SIBTEST method under almost all conditions.

3. The detection result of DIF in empirical data of the power rate of Raschtree method was higher than Logistic Regression and SIBTEST methods, but Type I error rate was higher than the criterion which was set at 10% level.

Department:	Educational Research and Psychology	Student's Signature .....
		Advisor's Signature .....
Field of Study:	Educational Measurement and Evaluation	Co-Advisor's Signature .....
Academic Year:	2015	

กิตติกรรมประกาศ

วิทยานิพนธ์เล่มนี้สำเร็จลุล่วงไปได้ด้วยดี ด้วยได้รับความเมตตากรุณาอย่างยิ่งจาก ผู้ช่วยศาสตราจารย์ ดร.ณัฏฐภรณ์ หลาวทอง และ ศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี อาจารย์ที่ปรึกษาวิทยานิพนธ์ ผู้ถ่ายทอดองค์ความรู้ด้านการวัดและประเมิน และวิธีวิทยาการวิจัย เสียสละเวลาให้คำปรึกษา แนะนำ และแก้ไขข้อบกพร่องต่าง ๆ แก่ผู้วิจัยตลอดจนให้กำลังใจ ดูแล ติดตามความก้าวหน้าในการทำวิทยานิพนธ์ของผู้วิจัย จนทำให้วิทยานิพนธ์เล่มนี้สำเร็จลุล่วงได้ ผู้วิจัยรู้สึกซาบซึ้งเป็นอย่างยิ่ง จึงขอกราบขอบพระคุณอาจารย์ทั้งสองท่านเป็นอย่างสูง

ขอกราบขอบพระคุณคณะกรรมการสอบวิทยานิพนธ์ รองศาสตราจารย์ ดร.ศิริเดช สุชีวะ รองศาสตราจารย์ ดร.โชติกา ภาษีผล รองศาสตราจารย์ ดร.กมลวรรณ ตังธนกานนท์ และ รองศาสตราจารย์ ดร.สมถวิล วิจิตรวรรณา ท่านทั้งหลายได้ให้ข้อคิดและข้อเสนอแนะอันเป็น ประโยชน์เพื่อปรับแก้วิทยานิพนธ์ให้มีความสมบูรณ์ยิ่งขึ้น

ขอกราบขอบพระคุณคณาจารย์ภาควิชาวิจัยและจิตวิทยาการศึกษา ที่ได้ให้ความรู้ ประสบการณ์ และกำลังใจในการศึกษาและการทำวิจัยแก่ผู้วิจัย และขอขอบคุณบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย ที่ให้การสนับสนุนทุนวิจัย “ทุน 90 ปี จุฬาลงกรณ์มหาวิทยาลัย” จาก กองทุนรัชดาภิเษกสมโภช ประจำปี 2557

ผู้วิจัยขอกราบขอบพระคุณ อาจารย์สมกิจ กิจพูนวงศ์ อาจารย์ ดร.อรินทร์ น่วมถนอม อาจารย์ ดร.อุไร จักษ์ตรีมงคล อาจารย์ ดร.อัญชลี สุขในสิทธิ์ อาจารย์ ดร.เรืองเดช ศิริกิจ และ อาจารย์ ดร.อนุสรณ์ เกิดศรี ท่านเหล่านี้ ได้ให้คำปรึกษา แนะนำ ในการวิเคราะห์ข้อมูลแก่ผู้วิจัย รวมทั้งขอกราบขอบพระคุณผู้อำนวยการสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (สสวท.) ที่ให้ความอนุเคราะห์ข้อมูลโครงการประเมินผลนักเรียนนานาชาติ (PISA 2009)

ขอขอบคุณเพื่อน ๆ รุ่นพี่ และรุ่นน้อง สาขาวิชาการวัดและประเมินผลการศึกษา ทั้ง ภาคในเวลาและภาคนอกเวลา ที่ให้คำแนะนำ ความช่วยเหลือและเป็นกำลังใจ ในการทำ วิทยานิพนธ์เล่มนี้ ครอบครัวของผู้วิจัยที่เป็นกำลังใจที่สำคัญยิ่ง และขอขอบคุณอีกหลาย ๆ ท่าน ที่ไม่ได้กล่าวถึง ที่ได้มีส่วนช่วยให้วิทยานิพนธ์เล่มนี้ประสบความสำเร็จ

ท้ายนี้ ประโยชน์ที่พึงจะเกิดจากวิทยานิพนธ์เล่มนี้ ผู้วิจัยขอมอบตอบแทนพระคุณพ่อ แม่ และครูอาจารย์ทุกท่านที่ให้การอบรมสั่งสอนจนผู้วิจัยสำเร็จการศึกษา

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญ ..... ช

สารบัญตาราง ..... ฎ

สารบัญภาพ ..... ป

บทที่ 1 บทนำ 1

ความเป็นมาและความสำคัญของปัญหา 1

คำถามการวิจัย 10

วัตถุประสงค์การวิจัย 10

สมมติฐานการวิจัย 11

ขอบเขตการวิจัย 12

ข้อจำกัดในการวิจัย 14

คำจำกัดความที่ใช้ในการวิจัย 14

ประโยชน์ที่จะได้รับ 17

บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง 18

ตอนที่ 1 มโนทัศน์ของการทำหน้าที่ต่างกันของข้อสอบ 19

1.1 ความเป็นมาของการทำหน้าที่ต่างกันของข้อสอบ 19

1.2 ความหมายของการทำหน้าที่ต่างกันของข้อสอบ 20

1.3 ประเภทของการทำหน้าที่ต่างกันของข้อสอบ 21

ตอนที่ 2 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 24

2.1 วิธีการถดถอยโลจิสติก (logistic regression; LR) 26

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

หน้า

2.2 วิธีซิปเทสท์ (Simultaneous Item Bias Test; SIBTEST) 30
2.3 วิธีราสช์ทรี (Raschtree)..... 35
2.4 วิธีการวัดพื้นที่ของราชู (Raju)..... 44
2.5 เกณฑ์สำหรับเปรียบเทียบคุณภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ..... 46
2.6 ขนาดอิทธิพลในการทำหน้าที่ต่างกันของข้อสอบ 48
ตอนที่ 3 โครงการประเมินผลนักเรียนนานาชาติ (Programme for International Student Assessment; PISA) 49
ตอนที่ 4 งานวิจัยที่เกี่ยวกับการศึกษาการทำหน้าที่ต่างกันของข้อสอบในแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค 58
4.1 งานวิจัยที่เกี่ยวข้องกับการศึกษาปัจจัยที่มีผลต่อการทำหน้าที่ต่างกันของข้อสอบในแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค..... 58
4.2 งานวิจัยที่เกี่ยวข้องกับการศึกษาเปรียบเทียบประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค 64
4.3 สรุปประเด็นการศึกษาการทำหน้าที่ต่างกันของข้อสอบในแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค 70
4.4 กรอบแนวคิดในการวิจัย 71
บทที่ 3 วิธีดำเนินการวิจัย..... 74
ตอนที่ 1 การกำหนดแนวทางในการวิเคราะห์ข้อมูล 74
ตอนที่ 2 การจัดกระทำข้อมูลตามปัจจัยที่ศึกษาและการจำลองข้อมูล 76
ตอนที่ 3 ลักษณะข้อมูลเชิงประจักษ์..... 80
ตอนที่ 4 การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ 83
ตอนที่ 5 การวิเคราะห์อำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 87
ตอนที่ 6 การเปรียบเทียบประสิทธิภาพของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ... 88

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

หน้า

บทที่ 4 ผลการวิเคราะห์ข้อมูล 90
ตอนที่ 1 ผลการวิเคราะห์ค่าเฉลี่ยของอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 91
1.1 การตรวจสอบความถูกต้องของข้อมูลจำลอง 92
1.2 ประสิทธิภาพด้านอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 โดยการทดสอบระดับนัยสำคัญ ของวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี .. 96
1.3 ผลการวิเคราะห์อำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 โดยการทดสอบระดับนัยสำคัญ ของวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี 101
1.4 ประสิทธิภาพด้านอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 โดยการวัดขนาดอิทธิพล ของวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ 106
1.5 ผลการวิเคราะห์อำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 โดยการวัดขนาดอิทธิพล ของวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ 108
ตอนที่ 2 ผลการเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 112
2.1 การเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี 112
2.2 การเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ระหว่างวิธีการถดถอย โลจิสติกและวิธีซิปเทสท์ 162
ตอนที่ 3 ผลการเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ในข้อมูลเชิงประจักษ์ 195
3.1. การวิเคราะห์ข้อมูลด้านสถิติเชิงบรรยายของคะแนนจากแบบสอบ 196
3.2 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 209
3.3 ประสิทธิภาพด้านอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 241
บทที่ 5 สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ 242

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

หน้า

สรุปผลการวิจัย 243
อภิปรายผลการวิจัย 262
ข้อเสนอแนะ 276
รายการอ้างอิง 279
ภาคผนวก 285
ภาคผนวก ก ตัวอย่างผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีการถดถอยโลจิสติก (Logistic Regression) 286
ภาคผนวก ข ตัวอย่างผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีซิปเทสท์ (SIBTEST) 291
ภาคผนวก ค ตัวอย่างผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีราสช์ทรี (Raschtree) 296
ภาคผนวก ง ตัวอย่างผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีการวัดพื้นที่ของราชู (Raju) 301
ประวัติผู้เขียนวิทยานิพนธ์ 305

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญตาราง

หน้า

ตารางที่ 1 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่มีการตรวจให้คะแนนแบบทวิวิภาค และพหุวิภาค 26

ตารางที่ 2 คุณภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 47

ตารางที่ 3 ค่า R^2 ของการวัดขนาดอิทธิพลในการทำหน้าที่ต่างกันของข้อสอบ 48

ตารางที่ 4 รอบของการประเมินโครงการประเมินผลนักเรียนนานาชาติ PISA 50

ตารางที่ 5 ข้อสอบ PISA 2009 ด้านการรู้เรื่องการอ่าน ที่ตรวจให้คะแนนแบบทวิวิภาค 51

ตารางที่ 6 การศึกษาปัจจัยที่มีผลต่อการทำหน้าที่ต่างกันของข้อสอบในแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค 62

ตารางที่ 7 การศึกษาเปรียบเทียบประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค 68

ตารางที่ 8 แผนผังการจำลองข้อมูล 79

ตารางที่ 9 จำนวนผู้สอบโครงการ PISA 2009 การรู้เรื่องการอ่านที่ใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ จำแนกตามตัวแปร 3 ตัว 81

ตารางที่ 10 ภาพรวมของการจำลองข้อมูลจำแนกตามปัจจัยและเงื่อนไขของปัจจัยที่แปรเปลี่ยน 91

ตารางที่ 11 ผลการตรวจสอบความถูกต้องของข้อมูลจำลองจำแนกตามปัจจัยที่ศึกษา 36 เงื่อนไข . 92

ตารางที่ 12 ผลการตรวจสอบค่าพารามิเตอร์ความยาก (b) และส่วนเบี่ยงเบนมาตรฐาน (*SD*) ของข้อมูลจำลอง 94

ตารางที่ 13 ค่าความเที่ยง (Reliability) ของแบบสอบจากการจำลองข้อมูล 96

ตารางที่ 14 ค่าเฉลี่ยร้อยละและส่วนเบี่ยงเบนมาตรฐานของอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 โดยการทดสอบระดับนัยสำคัญในภาพรวม 97

ตารางที่ 15 ค่าเฉลี่ยร้อยละ (*M*) และส่วนเบี่ยงเบนมาตรฐาน (*SD*) ของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี จำแนกตามปัจจัยที่ศึกษา..... 102

ตารางที่ 16 ค่าเฉลี่ยร้อยละ (*M*) และส่วนเบี่ยงเบนมาตรฐาน (*SD*) ของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี จำแนกตามปัจจัยที่ศึกษา 103

ตารางที่ 17 ค่าเฉลี่ยร้อยละและส่วนเบี่ยงเบนมาตรฐานของอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 โดยการวัดขนาดอิทธิพล ในภาพรวม 106

ตารางที่ 18 ค่าเฉลี่ยร้อยละ (*M*) และส่วนเบี่ยงเบนมาตรฐาน (*SD*) ของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ด้วยวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ จำแนกตามปัจจัยที่ศึกษา 109

ตารางที่ 19 ค่าเฉลี่ยร้อยละ (*M*) และส่วนเบี่ยงเบนมาตรฐาน (*SD*) ของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพลด้วยวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ จำแนกตามปัจจัยที่ศึกษา 109

ตารางที่ 20 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอย โลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบต่างกัน 3 ระดับ 113

ตารางที่ 21 ผลการวิเคราะห์อิทธิพลอย่างง่ายของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบต่างกัน 3 ระดับ 114

ตารางที่ 22 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญ ด้วยวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบต่างกัน 3 ระดับ โดยใช้วิธีการทดสอบของ Games-Howell ... 115

ตารางที่ 23 ผลการวิเคราะห์อิทธิพลอย่างง่ายของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบเดียวกัน 3 ระดับ 116

ตารางที่ 24 ผลการเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบ b_L b_M และ b_H โดยใช้วิธีการทดสอบของ Bonferroni และวิธีการทดสอบของ Games-Howell..... 117

ตารางที่ 25 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทปส์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบต่างกัน 3 ระดับ 119

ตารางที่ 26 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ของอำนาจการทดสอบในการตรวจสอบ การทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยความยาวของแบบสอบต่างกัน 2 ขนาด 121

ตารางที่ 27 ผลการเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยความยาวของแบบสอบขนาดเดียวกัน 2 ขนาด โดยใช้วิธีการทดสอบของ Games-Howell..... 122

ตารางที่ 28 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด 123

ตารางที่ 29 ผลการเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันขนาดเดียวกัน 2 ขนาด โดยใช้วิธีการทดสอบของ Games-Howell..... 124

ตารางที่ 30 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด 126

ตารางที่ 31 ผลการวิเคราะห์อิทธิพลอย่างง่ายของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด 127

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 32 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ด้วยวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้วิธีการทดสอบของ Games-Howell 129

ตารางที่ 33 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญ ด้วยวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้วิธีการทดสอบของ Games-Howell..... 130

ตารางที่ 34 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ด้วยวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้วิธีการทดสอบของ Bonferroni 131

ตารางที่ 35 ผลการวิเคราะห์อิทธิพลอย่างง่ายของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยกลุ่มตัวอย่างขนาดเดียวกัน 3 ขนาด 132

ตารางที่ 36 ผลการเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของขนาดกลุ่มตัวอย่างขนาดเดียวกัน 3 ขนาด โดยใช้วิธีการทดสอบของ Games-Howell 133

ตารางที่ 37 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด 135

ตารางที่ 38 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบต่างกัน 3 ระดับ 138

ตารางที่ 39 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบต่างกัน 3 ระดับ โดยใช้วิธีการทดสอบของ Bonferroni และวิธีการทดสอบของ Games-Howell 139

ตารางที่ 40 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบ b_L b_M และ b_H โดยใช้วิธีการทดสอบของ Bonferroni และวิธีการทดสอบของ Games-Howell 141

ตารางที่ 41 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยความยาวของแบบสอบต่างกัน 2 ขนาด 142

ตารางที่ 42 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยความยาวของแบบสอบขนาดเดียวกัน 2 ขนาด โดยใช้วิธีการทดสอบของ Games-Howell 143

ตารางที่ 43 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด 144

ตารางที่ 44 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันขนาดเดียวกัน 2 ขนาด โดยใช้วิธีการทดสอบของ Games-Howell 145

ตารางที่ 45 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด 146

ตารางที่ 46 ผลการวิเคราะห์อิทธิพลอย่างง่ายของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยขนาด กลุ่มตัวอย่างต่างกัน 3 ขนาด 148

Chulalongkorn University

ตารางที่ 47 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ด้วยวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้วิธีการทดสอบของ Bonferroni 149

ตารางที่ 48 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ด้วยวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้วิธีการทดสอบของ Games-Howell 150

ตารางที่ 49 ผลการวิเคราะห์อิทธิพลอย่างง่ายของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างขนาดเดียวกัน 3 ขนาด 151

ตารางที่ 50 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของขนาดกลุ่มตัวอย่างขนาดเดียวกัน 3 ขนาด โดยใช้วิธีการทดสอบของ Bonferroni และวิธีการทดสอบของ Games-Howell 152

ตารางที่ 51 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด 154

ตารางที่ 52 สรุปผลการวิเคราะห์ค่าเฉลี่ยของอำนาจการทดสอบ (Power) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ด้วยวิธีการถดถอยโลจิสติก (Logistic Regression) วิธีซิปเทสท์ (SIBTEST) และวิธีราสช์ทรี (Raschtree) ภายใต้เงื่อนไขที่แตกต่างกันของปัจจัยที่ศึกษา 157

ตารางที่ 53 สรุปผลการเปรียบเทียบอำนาจการทดสอบ (Power) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ด้วยวิธีการถดถอยโลจิสติก (Logistic Regression) วิธีซิปเทสท์ (SIBTEST) และวิธีราสช์ทรี (Raschtree) ภายใต้เงื่อนไขของปัจจัยที่ศึกษาต่างระดับกัน 158

ตารางที่ 54 สรุปผลการเปรียบเทียบอำนาจการทดสอบ (Power) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก (Logistic Regression) วิธีซิปเทสท์ (SIBTEST) และวิธีราสช์ทรี (Raschtree) ภายใต้เงื่อนไขของปัจจัยที่ศึกษาระดับเดียวกัน 160

ตารางที่ 55 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ระหว่างวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบต่างกัน 3 ระดับ 162

ตารางที่ 56 ผลการวิเคราะห์อิทธิพลอย่างง่ายของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ด้วยวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบต่างกัน 3 ระดับ..... 164

ตารางที่ 57 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ด้วยวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบต่างกัน 3 ระดับ โดยใช้วิธีการทดสอบของ Games-Howell... 165

ตารางที่ 58 ผลการวิเคราะห์อิทธิพลอย่างง่ายของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ด้วยวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบเดียวกัน 3 ระดับ..... 166

ตารางที่ 59 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ด้วยวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบต่างกัน 3 ระดับ..... 167

ตารางที่ 60 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ระหว่างวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยความยาวของแบบสอบต่างกัน 2 ขนาด 168

ตารางที่ 61 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ระหว่างวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด 169

ตารางที่ 62 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ระหว่างวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด 170

ตารางที่ 63 ผลการวิเคราะห์อิทธิพลอย่างง่ายของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพล ด้วยวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด 172

ตารางที่ 64 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ด้วยวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้วิธีการทดสอบของ Games-Howell 173

ตารางที่ 65 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ด้วยวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้วิธีการทดสอบของ Games-Howell 174

ตารางที่ 66 ผลการวิเคราะห์อิทธิพลอย่างง่ายของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ระหว่างวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างขนาดเดียวกัน 3 ขนาด 175

ตารางที่ 67 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ด้วยวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด 176

ตารางที่ 68 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพลระหว่างวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบต่างกัน 3 ระดับ 178

ตารางที่ 69 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบ b_L b_M และ b_H โดยใช้การทดสอบ t-test 179

ตารางที่ 70 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพลระหว่างวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยความยาวของแบบสอบต่างกัน 2 ขนาด 180

ตารางที่ 71 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยความยาวของแบบสอบเดียวกัน 2 ขนาด โดยใช้การทดสอบ t-test 181

ตารางที่ 72 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพลระหว่างวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยสัดส่วนของข้อสอบที่ ทำหน้าที่ต่างกัน 2 ขนาด 182

ตารางที่ 73 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันขนาดเดียวกัน 2 ขนาด โดยใช้การทดสอบ t-test 183

ตารางที่ 74 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลระหว่างวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด 184

ตารางที่ 75 ผลการวิเคราะห์อิทธิพลอย่างง่ายของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพล ด้วยวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด 185

ตารางที่ 76 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ด้วยวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้วิธีการทดสอบของ Games-Howell 186

ตารางที่ 77 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ด้วยวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้วิธีการทดสอบของ Bonferroni 187

ตารางที่ 78 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างขนาดเดียวกัน 3 ขนาด โดยใช้การทดสอบ t-test 188

ตารางที่ 79 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ด้วยวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด 189

ตารางที่ 80 สรุปผลการวิเคราะห์ค่าเฉลี่ยของอำนาจการทดสอบ (Power) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ด้วยวิธีการถดถอยโลจิสติก (Logistic regression) และวิธีซิปเทสท์ (SIBTEST) ภายใต้เงื่อนไขที่แตกต่างกันของปัจจัยที่ศึกษา 191

ตารางที่ 81 สรุปผลการเปรียบเทียบอำนาจการทดสอบ (Power) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ด้วยวิธีการถดถอยโลจิสติก (Logistic regression) และวิธีซิปเทสท์ (SIBTEST) ภายใต้เงื่อนไขของปัจจัยที่ศึกษาต่างระดับกัน 192

ตารางที่ 82 สรุปผลการเปรียบเทียบอำนาจการทดสอบ (Power) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ด้วยวิธีการถดถอยโลจิสติก (Logistic regression) และวิธีซิปเทสท์ (SIBTEST) ภายใต้เงื่อนไขของปัจจัยที่ศึกษาระดับเดียวกัน 194

ตารางที่ 83 สถิติเชิงบรรยายของคะแนนจากแบบสอบ 196

ตารางที่ 84 ค่าความเที่ยงแบบความสอดคล้องภายในโดยสูตรสัมประสิทธิ์แอลฟาของครอนบาคของแบบสอบด้านการรู้เรื่องการอ่าน 198

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 85 ค่าความยาก (p) และค่าอำนาจจำแนก (r) ของแบบสอบฉบับที่ 1 200
ตารางที่ 86 ค่าความยาก (p) และค่าอำนาจจำแนก (r) ของแบบสอบฉบับที่ 2 201
ตารางที่ 87 ค่าความยาก (p) และค่าอำนาจจำแนก (r) ของแบบสอบฉบับที่ 3 201
ตารางที่ 88 ค่าความยาก (p) และค่าอำนาจจำแนก (r) ของแบบสอบฉบับที่ 4 202
ตารางที่ 89 ค่าความยาก (p) และค่าอำนาจจำแนก (r) ของแบบสอบฉบับที่ 5 202
ตารางที่ 90 ค่าความยาก (p) และค่าอำนาจจำแนก (r) ของแบบสอบฉบับที่ 6 203
ตารางที่ 91 ค่าความยาก (p) และค่าอำนาจจำแนก (r) ของแบบสอบฉบับที่ 7 203
ตารางที่ 92 ค่าความยาก (p) และค่าอำนาจจำแนก (r) ของแบบสอบฉบับที่ 8 204
ตารางที่ 93 ค่าความยาก (p) และค่าอำนาจจำแนก (r) ของแบบสอบฉบับที่ 9 204
ตารางที่ 94 ค่าความยาก (p) และค่าอำนาจจำแนก (r) ของแบบสอบฉบับที่ 10 205
ตารางที่ 95 ค่าความยาก (p) และค่าอำนาจจำแนก (r) ของแบบสอบฉบับที่ 11 205
ตารางที่ 96 ค่าความยาก (p) และค่าอำนาจจำแนก (r) ของแบบสอบฉบับที่ 12 205
ตารางที่ 97 ค่าความยาก (p) และค่าอำนาจจำแนก (r) ของแบบสอบฉบับที่ 13 206
ตารางที่ 98 ผลการวิเคราะห์ความแปรปรวนแบบทางเดียวของคะแนน ระหว่างแบบสอบ 13 ฉบับ 206
ตารางที่ 99 ผลการเปรียบเทียบคะแนนระหว่างแบบสอบ 13 ฉบับ โดยใช้วิธีการทดสอบของ Games-Howell 207
ตารางที่ 100 รายละเอียดของตัวแปรที่ใช้ในการแบ่งกลุ่มเพื่อวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ 210
ตารางที่ 101 ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ 211
ตารางที่ 102 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการวัดพื้นที่ของราชู วิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ของแบบสอบฉบับที่ 1 212
ตารางที่ 103 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการวัดพื้นที่ของราชู วิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ของแบบสอบฉบับที่ 2 214

ตารางที่ 104 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการวัดพื้นที่ของราชู วิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ของแบบสอบฉบับที่ 4..... 216

ตารางที่ 105 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการวัดพื้นที่ของราชู วิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ของแบบสอบฉบับที่ 5..... 218

ตารางที่ 106 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการวัดพื้นที่ของราชู วิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ของแบบสอบฉบับที่ 6 จำแนกตามตัวแปรเพศ 220

ตารางที่ 107 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการวัดพื้นที่ของราชู วิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ของแบบสอบฉบับที่ 6 จำแนกตามตัวแปรสังกัดของโรงเรียน 222

ตารางที่ 108 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการวัดพื้นที่ของราชู วิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ของแบบสอบฉบับที่ 7 224

ตารางที่ 109 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการวัดพื้นที่ของราชู วิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ของแบบสอบฉบับที่ 8 จำแนกตามตัวแปรเพศ 225

ตารางที่ 110 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการวัดพื้นที่ของราชู วิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ของแบบสอบฉบับที่ 8 จำแนกตามตัวแปรสังกัดของโรงเรียน 227

ตารางที่ 111 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการวัดพื้นที่ของราชู วิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ของแบบสอบฉบับที่ 8 จำแนกตามตัวแปรภูมิภาค..... 228

ตารางที่ 112 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการวัดพื้นที่ของราชู วิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ของแบบสอบฉบับที่ 9 230

ตารางที่ 113 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการวัดพื้นที่ของราชู วิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ของแบบสอบฉบับที่ 11 231

ตารางที่ 114 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการวัดพื้นที่ของราชู วิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ของแบบสอบฉบับที่ 13 จำแนกตามตัวแปรเพศ..... 233

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 115 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการวัดพื้นที่ของราชู วิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ของแบบสอบฉบับที่ 13 จำแนกตามตัวแปรสังกัดของโรงเรียน 235

ตารางที่ 116 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการวัดพื้นที่ของราชู วิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ของแบบสอบฉบับที่ 13 จำแนกตามตัวแปรภูมิภาค 237

ตารางที่ 117 สรุปผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 238

ตารางที่ 118 จำนวนข้อของการเกิดและไม่เกิดการทำหน้าที่ต่างกันของข้อสอบ 239

ตารางที่ 119 ร้อยละของอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 จำแนกตามวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 241

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญภาพ

ภาพที่ 1 ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป (uniform DIF) 22
ภาพที่ 2 ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป (nonuniform DIF) 22
ภาพที่ 3 ข้อสอบทำหน้าที่ต่างกันแบบมีทิศทางเดียวกัน (unidirectional DIF) 23
ภาพที่ 4 ข้อสอบทำหน้าที่ต่างกันแบบไม่มีทิศทาง (non-unidirectional DIF) 24
ภาพที่ 5 แนวทางการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบใน Rasch model 37
ภาพที่ 6 ตัวอย่างผลการตรวจสอบการทำหน้าที่ต่างกันด้วยวิธีราสช์ทรี 43
ภาพที่ 7 ตัวอย่างข้อสอบ PISA 2009 การรู้เรื่องการอ่าน ด้านการบูรณาการและตีความ 54
ภาพที่ 8 ตัวอย่างข้อสอบ PISA 2009 การรู้เรื่องการอ่าน ด้านการสะท้อนและประเมิน 55
ภาพที่ 9 ตัวอย่างข้อสอบ PISA 2009 การรู้เรื่องการอ่าน ด้านการเข้าถึงและค้นสาระ 56
ภาพที่ 10 กรอบแนวคิดในการวิจัย กรณีศึกษาการจำลองข้อมูล 72
ภาพที่ 11 กรอบแนวคิดในการวิจัย กรณีศึกษาข้อมูลเชิงประจักษ์ 73
ภาพที่ 12 อำนาจการทดสอบของวิธีการตรวจสอบ 3 วิธี ภายใต้ค่าความยากของข้อสอบต่างกัน 3 ระดับ 113
ภาพที่ 13 อำนาจการทดสอบของวิธีการตรวจสอบ 3 วิธี ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด 127
ภาพที่ 14 อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบ 3 วิธี ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด 147
ภาพที่ 15 อำนาจการทดสอบของวิธีการตรวจสอบ 2 วิธี ภายใต้ค่าความยากของข้อสอบต่างกัน 3 ระดับ 163
ภาพที่ 16 อำนาจการทดสอบของวิธีการตรวจสอบ 2 วิธี ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด 171
ภาพที่ 17 อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบ 2 วิธี ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด 185
ภาพที่ 18 ผลการวิเคราะห์ DIF ด้วยวิธีราสช์ทรี ในแบบสอบฉบับที่ 1 212

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ภาพที่ 19 ผลการวิเคราะห์ DIF ด้วยวิธีราสช์ทรี ในแบบสอบฉบับที่ 2 213
ภาพที่ 20 ผลการวิเคราะห์ DIF ด้วยวิธีราสช์ทรี ในแบบสอบฉบับที่ 4 215
ภาพที่ 21 ผลการวิเคราะห์ DIF ด้วยวิธีราสช์ทรี ในแบบสอบฉบับที่ 5 217
ภาพที่ 22 ผลการวิเคราะห์ DIF ด้วยวิธีราสช์ทรี ในแบบสอบฉบับที่ 6 จำแนกตามตัวแปรเพศ 219
ภาพที่ 23 ผลการวิเคราะห์ DIF ด้วยวิธีราสช์ทรี ในแบบสอบฉบับที่ 6 จำแนกตามตัวแปรสังกัดของโรงเรียน 221
ภาพที่ 24 ผลการวิเคราะห์ DIF ด้วยวิธีราสช์ทรี ในแบบสอบฉบับที่ 7 223
ภาพที่ 25 ผลการวิเคราะห์ DIF ด้วยวิธีราสช์ทรี ในแบบสอบฉบับที่ 8 จำแนกตามตัวแปรเพศ 225
ภาพที่ 26 ผลการวิเคราะห์ DIF ด้วยวิธีราสช์ทรี ในแบบสอบฉบับที่ 8 จำแนกตามตัวแปรสังกัดของโรงเรียน 226
ภาพที่ 27 ผลการวิเคราะห์ DIF ด้วยวิธีราสช์ทรี ในแบบสอบฉบับที่ 8 จำแนกตามตัวแปรภูมิภาค 228
ภาพที่ 28 ผลการวิเคราะห์ DIF ด้วยวิธีราสช์ทรี ในแบบสอบฉบับที่ 9 229
ภาพที่ 29 ผลการวิเคราะห์ DIF ด้วยวิธีราสช์ทรี ในแบบสอบฉบับที่ 11 231
ภาพที่ 30 ผลการวิเคราะห์ DIF ด้วยวิธีราสช์ทรี ในแบบสอบฉบับที่ 13 จำแนกตามตัวแปรเพศ ... 232
ภาพที่ 31 ผลการวิเคราะห์ DIF ด้วยวิธีราสช์ทรี ในแบบสอบฉบับที่ 13 จำแนกตามตัวแปรสังกัดของโรงเรียน 234
ภาพที่ 32 ผลการวิเคราะห์ DIF ด้วยวิธีราสช์ทรี ในแบบสอบฉบับที่ 13 จำแนกตามตัวแปรภูมิภาค 236

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 1
บทนำ

ความเป็นมาและความสำคัญของปัญหา

“ความตรง” เป็นหัวใจสำคัญของคุณภาพแบบสอบ ในการสร้างและการตรวจสอบคุณภาพของแบบสอบจะต้องคำนึงถึงคุณภาพด้านความตรงเป็นสำคัญ เนื่องด้วยความตรงเป็นคุณสมบัติของแบบสอบที่สะท้อนถึงความสามารถในการวัดได้ถูกต้องแม่นยำ ถ้าผลการวัดมีค่าที่ใกล้เคียงกับค่าคุณลักษณะที่แท้จริงเท่าใด ก็ถือว่าการวัดมีความตรงมากขึ้นเท่านั้น ส่วนการทำหน้าที่ต่างกันของข้อสอบและแบบสอบเป็นวิธีหนึ่งของการตรวจสอบคุณภาพด้านความตรง โดยเป็นการตรวจสอบในประเด็นของความอยุติธรรมของข้อสอบและแบบสอบ (item and test unfairness) (ศิริชัย กาญจนวาสี, 2555)

ความอยุติธรรมของข้อสอบเกิดขึ้นในกรณีที่ทำให้ผู้สอบระหว่างกลุ่มย่อยมีความได้เปรียบหรือเสียเปรียบกัน เดิมใช้คำว่า “ความลำเอียงของข้อสอบ” (item bias) ซึ่งเป็นภาษาที่ใช้กันโดยทั่วไปและมีความหมายในเชิงลบ สำหรับการตัดสินว่าข้อสอบมีความลำเอียงหรือไม่นั้น มักจะพิจารณาอิทธิพลที่สังเกตได้ของผู้สอบกลุ่มย่อยที่นำมาศึกษา โดยไม่คำนึงถึงวิธีการทางสถิติ จึงทำให้เกิดความคลุมเครือเกี่ยวกับเกณฑ์ที่ใช้ในการตัดสินความลำเอียง ต่อมานักวิธีวิทยาการวิจัยได้พัฒนาวิธีเพื่อใช้วิเคราะห์ดัชนีความลำเอียงของข้อสอบ โดยมุ่งเน้นความแตกต่างระหว่างกลุ่มผู้สอบที่ตอบข้อสอบข้อเดียวกัน และเลือกเกณฑ์ที่ใช้ในการจับคู่กลุ่มผู้สอบ (matching criterion) แล้วนำสารสนเทศทางสถิติมาใช้เป็นเกณฑ์ในการตัดสินความลำเอียงของข้อสอบ ดังนั้น จึงเปลี่ยนมาใช้คำว่า “การทำหน้าที่ต่างกันของข้อสอบ” (differential item functioning; DIF) แทน ซึ่งเป็นคำที่มีความหมายเป็นกลางและมีความเหมาะสมมากกว่า (Holland & Wainer, 1993)

การทำหน้าที่ต่างกันของข้อสอบจะเปรียบเทียบผลการตอบข้อสอบระหว่างผู้สอบกลุ่มย่อยอย่างน้อยสองกลุ่ม ที่มีความสามารถระดับเดียวกัน โดยผู้สอบกลุ่มที่ได้เปรียบ เรียกว่า กลุ่มอ้างอิง (reference group; R) และผู้สอบกลุ่มที่เสียเปรียบ เรียกว่า กลุ่มเปรียบเทียบ (focal group; F) ซึ่งเป็นกลุ่มที่สนใจศึกษาการทำหน้าที่ต่างกันของข้อสอบ ข้อสอบทำหน้าที่ต่างกัน เมื่อผู้สอบที่มีความสามารถระดับเดียวกันแต่อยู่ต่างกลุ่มกันมีโอกาสของการตอบข้อสอบได้ถูกต้องไม่เท่ากัน (Li and Stout, 1996 อ้างถึงใน วลีมาศ แซ่อึ้ง, 2543) ซึ่งขนาดและทิศทางของข้อสอบที่ทำหน้าที่ต่างกันจะแปรเปลี่ยนไปตามระดับความสามารถที่แตกต่างกัน โดยข้อสอบที่ทำหน้าที่ต่างกันแบ่งเป็นสองประเภท (Mellenbergh, 1982 อ้างถึงใน วลีมาศ แซ่อึ้ง, 2543) คือ ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป (uniform DIF) และข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป (nonuniform DIF)

ข้อสอบทำหน้าที่ต่างกันประเภทแรกเกิดขึ้นเมื่อไม่มีปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับการเป็นสมาชิกของกลุ่ม ส่วนข้อสอบทำหน้าที่ต่างกันประเภทหลังเกิดขึ้นเมื่อมีปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับการเป็นสมาชิกของกลุ่ม ซึ่งตามทฤษฎีการตอบสนองข้อสอบ (item response theory; IRT) สามารถพิจารณาปฏิสัมพันธ์ได้จากความแตกต่างของค่าพารามิเตอร์อำนาจจำแนกของข้อสอบระหว่างผู้สอบกลุ่มย่อยสองกลุ่ม กล่าวคือ ถ้าข้อสอบระหว่างผู้สอบกลุ่มย่อยสอง กลุ่ม มีค่าอำนาจจำแนกเท่ากันแล้ว โค้งลักษณะข้อสอบ (item characteristic curves; ICCs) ระหว่างกลุ่มผู้สอบจะขนานกัน แสดงว่า ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป แต่ถ้าข้อสอบระหว่างผู้สอบกลุ่มย่อยสองกลุ่ม มีค่าอำนาจจำแนกไม่เท่ากันแล้ว โค้งลักษณะข้อสอบระหว่างกลุ่มผู้สอบจะไม่ขนานกัน แสดงว่า ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป (Camilli & Shepard, 1994)

สำหรับวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) แบ่งตามมิติลักษณะของตัวแปรเกณฑ์ ได้เป็นสองกลุ่ม ดังนี้ (Potenza and Dorans, 1995; Feinstein, 1995 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555)

1. กลุ่มวิธีที่ใช้คะแนนสังเกตได้ (observed score) ซึ่งวิเคราะห์ตามทฤษฎีแบบดั้งเดิม (classical test theory; CTT) ใช้คะแนนรวมของผู้สอบเป็นเกณฑ์การจับคู่ของกลุ่มผู้สอบ วิธีการตรวจสอบที่สำคัญ ได้แก่ การวิเคราะห์ความแปรปรวน (ANOVA) วิธีการถดถอยโลจิสติก (logistic regression; LR) วิธีแปลงค่าความยากของข้อสอบ (transformed item difficulty; TID) วิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel; MH) และวิธีดัชนีมาตรฐาน (standardization; STND)

2. กลุ่มวิธีที่ใช้คุณลักษณะแฝง (latent variable) ซึ่งวิเคราะห์บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ (item response theory; IRT) ใช้ค่าประมาณความสามารถของผู้สอบเป็นเกณฑ์การจับคู่กลุ่มผู้สอบ วิธีการตรวจสอบที่สำคัญ ได้แก่ วิธีวัดพื้นที่ความแตกต่างระหว่างโค้งการตอบสนองข้อสอบ (IRT-D^2) วิธีไค-สแควร์ของลอร์ด (Lord's χ^2) วิธีอัตราส่วนไลค์ลิฮูดทั่วไป (general IRT likelihood ratio) วิธีอัตราส่วนไลค์ลิฮูด ลอกลิเนียร์ (loglinear IRT likelihood ratio) และวิธีซิปเทสท์ (Simultaneous Item Bias Test; SIBTEST)

วิธีที่ใช้หลักการของตารางการณ์จร (contingency table) ดังเช่น วิธีแมนเทล-แฮนส์เซล และวิธีในกลุ่ม IRT ทั้งสองวิธีมีแนวคิดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบคล้ายกัน โดยใช้หลักการวิเคราะห์ความแตกต่างของผลการตอบข้อสอบจากผู้สอบสองกลุ่มที่มีความสามารถระดับเดียวกัน ในวิธีตารางการณ์จรจะใช้คะแนนรวมที่สังเกตได้แทนระดับความสามารถ ส่วนวิธีในกลุ่ม IRT จะประมาณค่าความสามารถโดยใช้ฟังก์ชันทางคณิตศาสตร์ตามลักษณะโมเดล IRT จุดเด่นของวิธีตารางการณ์จร คือ สามารถนำไปประยุกต์ใช้กับกลุ่มตัวอย่างที่มีขนาดเล็ก มีวิธีดำเนินการตรวจสอบง่าย และไม่จำเป็นต้องใช้โปรแกรมคอมพิวเตอร์ที่ซับซ้อนเหมือนกับวิธีในกลุ่ม IRT (Camilli &

Shepard, 1994) ในการวิเคราะห์ด้วยวิธีนี้จะนำข้อมูลมาสร้างตารางการณ์จรแบบ 3 ทิศทาง คือ กลุ่มผู้สอบ (กลุ่มอ้างอิง/กลุ่มเปรียบเทียบ) ผลการตอบข้อสอบ (ถูก/ผิด) และคะแนนรวม (k ระดับ) ซึ่งวิธีที่ใช้หลักการของตารางการณ์จรที่นิยมใช้กันโดยทั่วไป คือ วิธีแมนเทล-แฮนส์เซล (MH) พัฒนาโดย Holland และ Thayer (1988 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555) เป็นที่ยอมรับจากนักวิธีวิทยาการวิจัยอย่างกว้างขวางว่าเป็นวิธีที่ใช้ง่าย สะดวก และประหยัด มีขั้นตอนการคำนวณที่ไม่สลับซับซ้อน มีการทดสอบทางสถิติแบบนันพาราเมตริก ซึ่งไม่จำเป็นต้องใช้โมเดลประมาณค่า

วิธีที่ใช้คะแนนสังเกตได้ (observed score) ซึ่งวิเคราะห์ตามทฤษฎีแบบดั้งเดิม (CTT) ดังเช่น วิธีการถดถอยโลจิสติก (LR) เป็นวิธีที่พัฒนามาจากวิธีลอกลิเนียร์ ของ Mellenberg ซึ่งเสนอโดย Swaminathan และ Rogers (1990 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555)วิธีนี้อยู่บนพื้นฐานของโมเดลการวิเคราะห์สมการถดถอยโลจิสติก ซึ่งเป็นโมเดลที่มีพื้นฐานเป็นแบบจำลองที่สามารถเพิ่มตัวแปรความสามารถและปฏิสัมพันธ์เข้าไปในสมการได้ และสอดคล้องกับธรรมชาติของการทำหน้าที่ต่างกันของข้อสอบได้ดี จึงสามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ทั้งแบบเอกรูปและอเนกรูป โมเดลถดถอยโลจิสติก เป็นโมเดลที่มีความยืดหยุ่นและใช้ได้ง่าย ในการวิเคราะห์สมการถดถอยโลจิสติกหรือเรียกว่าการวิเคราะห์โลจิท (logit analysis) เป็นการวิเคราะห์สมการทำนาย เมื่อต้องการศึกษาผลของตัวแปรทำนายที่มีต่อตัวแปรเกณฑ์ซึ่งมีลักษณะเป็นทวิวิภาค (dichotomous variable) โดยใช้ฟังก์ชันโลจิสติก (logistic function) ในการแสดงความสัมพันธ์ระหว่างค่าของตัวแปรทำนายกับค่าความน่าจะเป็นของการเกิดเหตุการณ์ตามตัวแปรเกณฑ์

สำหรับวิธีที่ใช้คุณลักษณะแฝง (latent variable) ซึ่งวิเคราะห์บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ (IRT) ดังเช่น วิธีซิปเทสท์ (SIBTEST) พัฒนาโดย Shealy และ Stout (1993 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555) เพื่อใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) การทำหน้าที่ต่างกันของแบบสอบ (DTF) และการทำหน้าที่ต่างกันของกลุ่มข้อสอบ (DBF) วิธีนี้สามารถวิเคราะห์ได้ทั้งในแบบสอบเอกมิติ (unidimensional test) และแบบสอบพหุมิติ (multidimensional test) (Stoult, Li, and Nandakumar, 1997 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555) วิธีซิปเทสท์ใช้สถิติทดสอบแบบนันพาราเมตริก ซึ่งพัฒนาบนพื้นฐานของทฤษฎี IRT ชนิดพหุมิติ แต่ไม่ต้องใช้ฟังก์ชันการตอบสนองข้อสอบหรือการประมาณค่าความสามารถแฝง ได้รับการออกแบบมาสำหรับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบมีทิศทางเดียว (unidirectional DIF) (Li and Stout, 1996 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555) จุดเด่น คือ สามารถคำนวณได้ง่ายไม่ซับซ้อน ประหยัดค่าใช้จ่าย และไม่จำเป็นต้องใช้กลุ่มตัวอย่างที่มีขนาดใหญ่ ทั้งยังใช้สถิติทดสอบนัยสำคัญ (Narayanan and Swaminathan, 1996 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555) นอกจากนี้ ยังสามารถนำไปประยุกต์ใช้กับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบพหุวิภาค (Chang, Mazzeo, and Roussos, 1995 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555)

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

วิธีการวัดพื้นที่ของราชูจะคำนวณพื้นที่ในช่วงเปิด (open-interval or exact area) ส่วนวิธีการวัดพื้นที่ของคิมและโคเฮนจะคำนวณพื้นที่ในช่วงปิด (closed-interval area) จุดเด่นของวิธีการวัดพื้นที่ทั้งสองวิธี คือ สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและแบบอเนกรูปได้ถูกต้องแม่นยำ โดยจะใช้วิธีการคำนวณทางคณิตศาสตร์ที่เรียกว่า “การอินทิเกรทแบบต่อเนื่อง” (continuous integration) คำนวณพื้นที่ระหว่างโค้งลักษณะข้อสอบจากผู้สอบสองกลุ่ม ภายใต้โมเดลโลจิสติกแบบ 1, 2 และ 3 พารามิเตอร์ ทั้งยังสามารถคำนวณพื้นที่ชนิดคิดเครื่องหมาย (signed) และชนิดไม่คิดเครื่องหมาย (unsigned) ในการตัดสินข้อสอบที่ทำหน้าที่ต่างกันของวิธีการวัดพื้นที่ของราชูจะใช้สถิติทดสอบนัยสำคัญ ส่วนวิธีการวัดพื้นที่ของคิมและโคเฮนจะไม่ใช้การทดสอบทางสถิติ แต่จะนำขนาดของพื้นที่ไปเปรียบเทียบกับเกณฑ์ที่กำหนดไว้ ดังนั้น จึงเป็นที่ยอมรับกันโดยทั่วไปว่าวิธีการตรวจสอบการทำหน้าที่ต่างกันดังกล่าวเป็นวิธีที่มีประสิทธิภาพสูง (วลีมาศ แซ่อึ้ง, 2543)

วิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบใน Rasch model โดยส่วนใหญ่อยู่บนพื้นฐานการเปรียบเทียบค่าพารามิเตอร์ของข้อสอบที่ประมาณค่าระหว่างกลุ่มผู้สอบสองกลุ่มหรือมากกว่าที่ได้กำหนดไว้ล่วงหน้า เช่น กลุ่มผู้ชายเป็นกลุ่มเปรียบเทียบ และกลุ่มผู้หญิงเป็นกลุ่มอ้างอิง (Strobl, Kopf, & Zeileis, 2015) ข้อดี คือ ถ้าตรวจสอบพบ DIF ผลลัพธ์ดังกล่าว สามารถอธิบายได้อย่างชัดเจนว่าข้อสอบง่ายหรือยากสำหรับผู้สอบกลุ่มใด เพื่อแก้ไขข้อสอบสำหรับแต่ละกลุ่ม และสามารถขจัดหรือหลีกเลี่ยงรูปแบบของแบบสอบในอนาคต ส่วนข้อจำกัดของโมเดลการทดสอบนี้ คือ นักวิจัยเป็นผู้กำหนดกลุ่มที่ชัดเจนในการตรวจสอบ DIF ตัวแปรในการทดสอบ เช่น อายุ เพศ ศาสนา

และภาษา ดังนั้น ถ้าการวิเคราะห์ภายหลังพบความแตกต่างระหว่างกลุ่มในตัวแปรที่ไม่ได้กำหนดไว้ก่อน จะทำให้ผลนั้นไม่สามารถอธิบายได้

นอกจากวิธีการตรวจสอบที่ได้กล่าวไปแล้ว ยังมีอีกแนวทางหนึ่ง เรียกว่า แนวทางกลุ่มแฝง (the latent class approach) ของ Rost (1990 cited in Strobl, Kopf and Zeileis, 2015) นับเป็นการทดสอบใน Rasch model ที่เข้มงวด เพราะมุ่งทดสอบความแตกต่างของพารามิเตอร์ของข้อสอบระหว่างกลุ่มของผู้สอบทั้งหมด โดยไม่พิจารณาเรื่องตัวแปรร่วมของบุคคล ทำให้มีข้อจำกัดในการอธิบายผลอย่างไม่ชัดเจน

แนวทางใหม่ที่ผสมผสานระหว่างสองแนวทางที่กล่าวข้างต้น คือ การทดสอบทุกกลุ่มแบบทำซ้ำ ภายในขอบเขตของตัวแปรร่วมที่เป็นไปได้ ทำให้สามารถอธิบายผลลัพธ์ได้ชัดเจน และเป็นการตรวจสอบ DIF ได้อย่างกว้างขวาง วิธีนี้อยู่บนพื้นฐานเทคนิคโมเดลการแบ่งส่วนแบบทำซ้ำ (model-based recursive partitioning) ใช้สถิติทดสอบโครงสร้างที่เปลี่ยนแปลงจากเศรษฐมิติ (econometrics) โมเดลการแบ่งส่วนแบบทำซ้ำ เป็นแนวทางแบบเซมิพาราเมตริก มีเป้าหมายเพื่อตรวจสอบความแตกต่างในค่าพารามิเตอร์ของโมเดลทางสถิติระหว่างกลุ่มผู้สอบที่กำหนดจากการรวมกันของตัวแปรร่วม วิธีนี้ เรียกว่า ราสช์ทรี (Raschtree) นำเสนอโดย Strobl et al. (2015) ข้อดีของวิธีราสช์ทรี คือ สามารถตรวจสอบได้ทั้งกลุ่มที่กำหนดไว้ล่วงหน้าและกลุ่มแฝง โดยการทำหน้าที่ต่างกันของข้อสอบจะแสดงผลได้โดยอัตโนมัติจากการตรวจสอบในกลุ่มผู้สอบ อีกทั้งในขั้นตอนของกำหนดจุดตัดเพื่อตัดสินกลุ่มที่ทำหน้าที่ต่างกัน จะเกิดจากความแตกต่างของค่าพารามิเตอร์ที่เด่นชัด และผลวิเคราะห์จะแสดงเป็นแผนภาพที่สามารถอธิบายกลุ่มข้อสอบและกลุ่มผู้สอบให้เข้าใจได้ง่าย โมเดลการแบ่งส่วนแบบทำซ้ำยังสัมพันธ์กับวิธี classification trees ซึ่งงานวิจัยของ Vaughn and Wang (2010) พบว่า วิธี classification trees ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบมีอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 เท่าเทียมกับวิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก แต่ยังไม่มีการวิจัยศึกษาประสิทธิภาพของวิธีราสช์ทรีในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

จากการศึกษาของ Rogers and Swaminathan (1993) พบว่า วิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติกมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปเท่าเทียมกัน ส่วน Narayanan and Swaminathan (1994) ได้ศึกษาการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป พบว่า เมื่อการแจกแจงค่าความสามารถระหว่างกลุ่มผู้สอบมีค่าเท่ากัน วิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีประสิทธิภาพเท่าเทียมกัน แต่เมื่อการแจกแจงค่าความสามารถระหว่างกลุ่มผู้สอบมีค่าไม่เท่ากัน วิธีซิปเทสท์มีประสิทธิภาพสูงกว่าวิธีแมนเทล์-แฮนส์เซล ต่อมาวิธีแมนเทล-แฮนส์เซลได้ถูกปรับปรุงขั้นตอนการวิเคราะห์ข้อมูล โดยแบ่งกลุ่มผู้สอบออกเป็นสองกลุ่มตามระดับคะแนนสอบ คือ กลุ่มผู้สอบที่มีความสามารถสูงและกลุ่มผู้สอบที่มีความสามารถต่ำ แล้ว

วิเคราะห์แยกกลุ่มผู้สอบตามขั้นตอนการวิเคราะห์เดิม พบว่า วิธีแมนเทล-แฮนส์เซลที่ปรับปรุงขั้นตอนการวิเคราะห์มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูงขึ้น และสามารถตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปได้ดีขึ้น (Mazor, Clauser, & Hambletor, 1994; เสรี ชัดแช้ม, 2539)

ต่อมาจากการศึกษาของ Narayanan and Swaminathan (1996) พบว่า วิธีการถดถอยโลจิสติกมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปสูงกว่าวิธีแมนเทล-แฮนส์เซล ภายใต้เกือบทุกเงื่อนไขของการตรวจสอบ ทำให้ทราบว่าวิธีแมนเทล-แฮนส์เซลไม่สามารถระบุข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปได้ ส่วนวิธีซิปเทสท์และวิธีการถดถอยโลจิสติกมีประสิทธิภาพในการตรวจสอบเท่าเทียมกัน ภายใต้เกือบทุกเงื่อนไข นอกจากนี้ วลีมาศ แซ่อึ้ง (2543) ได้ศึกษาการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป พบว่า วิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ที่ปรับปรุงขั้นตอนการวิเคราะห์มีอำนาจการทดสอบเท่าเทียมกันภายใต้เกือบทุกเงื่อนไข และทั้งสองวิธีมีอำนาจการทดสอบสูงกว่าวิธีซิปเทสท์แบบเดิมและวิธีแมนเทล-แฮนส์เซลที่ปรับปรุงขั้นตอนการวิเคราะห์ ภายใต้เกือบทุกเงื่อนไข ดังนั้น อาจสรุปได้ว่าวิธีแมนเทล-แฮนส์เซลที่ปรับปรุงขั้นตอนการวิเคราะห์ ทำให้มีอัตราการตรวจสอบสูงขึ้น แต่เมื่อเปรียบเทียบกับวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์แล้ว ก็ยังมีข้อจำกัดในเรื่องการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป ส่วนวิธีซิปเทสท์ที่ปรับปรุงขั้นตอนการวิเคราะห์มีอำนาจการทดสอบสูงขึ้นเท่าเทียมกับวิธีการถดถอยโลจิสติก

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อศึกษากับกลุ่มตัวอย่างจำนวนมาก ส่วนใหญ่ค่าสถิติของการทดสอบจะมีนัยสำคัญ ทำให้สรุปว่าข้อสอบทำหน้าที่ต่างกัน ซึ่งเกิดจากความคลาดเคลื่อนของข้อมูล เช่นการศึกษาของ Kim และคนอื่น ๆ (2007 อ้างถึงใน ธเกียรติกมล ทองงอก, 2554) ที่พบว่า กลุ่มตัวอย่างขนาดใหญ่จะมีความไวในการตรวจพบการทำหน้าที่ต่างกันค่อนข้างสูง ดังนั้น เพื่อแก้ไขความคลาดเคลื่อนที่อาจเกิดขึ้นได้ จึงควรใช้การวัดขนาดอิทธิพล (effect size measure) มาช่วยในการตัดสินประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เพื่อให้ผลลัพธ์มีความถูกต้องยิ่งขึ้น ดังเช่นการศึกษาของ Jodoin และ Gierl (2001 อ้างถึงใน ธเกียรติกมล ทองงอก, 2554) พบว่า การใช้ขนาดอิทธิพลโดยสถิติ R^2 ในวิธีการถดถอยโลจิสติก ร่วมกับการทดสอบนัยสำคัญในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ จะได้ค่าของการสรุปผิดที่ว่าข้อสอบทำหน้าที่ต่างกันทั้งที่ความเป็นจริงทำหน้าที่ไม่ต่างกัน (false positive) ลดลงเกือบเป็นศูนย์

ในการศึกษาวิจัยเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบที่ผ่านมา จะเห็นว่า มีการเปรียบเทียบประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในด้านความถูกต้องแม่นยำของการตรวจสอบ โดยพิจารณาได้จากค่าอำนาจการทดสอบหรืออัตราความถูกต้องของการตรวจพบข้อที่ทำหน้าที่ต่างกัน (power rate) และความคลาดเคลื่อนของการตรวจสอบหรืออัตราความคลาดเคลื่อนประเภทที่ 1 (type I error rate) โดยวิธีที่นำมาศึกษาเพื่อเปรียบเทียบ

ประสิทธิภาพการตรวจสอบที่ใช้กันจำนวนมาก ได้แก่ วิธีแมนเทล-แฮนส์เซล วิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ซึ่งจะเห็นได้จากงานวิจัยที่นำทั้งสามวิธีนี้ มาศึกษาเปรียบเทียบกันเอง หรือศึกษาเปรียบเทียบกับวิธีอื่น ๆ ด้วย (DeMars, 2009; Escorial & Navas, 2007; Hidalgo & LÓPEZ-PINA, 2004; Kalaycioglu & Berberoglu, 2010; Moses, Miao, & Dorans, 2010; Vaughn & Wang, 2010) โดยนำปัจจัยที่มีผลต่ออำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันมาศึกษา เช่น ความยากของข้อสอบ ความยาวของแบบสอบ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน และขนาดกลุ่มตัวอย่าง เป็นต้น

ความยากของข้อสอบ (Difficulty of item) จากการศึกษาของ Rogers and Swaminathan (1993) พบว่า ในการตรวจสอบการทำหน้าที่ต่างกันแบบเอกรูปเมื่อข้อสอบมีค่าความยากปานกลาง และค่าอำนาจจำแนกสูง จะทำให้อำนาจการทดสอบของวิธีการถดถอยโลจิสติกและวิธีแมนเทล-แฮนส์เซล มีค่าสูงสุด สำหรับการศึกษาของ Narayanan and Swaminathan (1994) พบว่า การตรวจสอบการทำหน้าที่ต่างกันแบบเอกรูปภายใต้เกือบทุกเงื่อนไขลักษณะของข้อสอบที่มีค่าความยากปานกลางและค่าอำนาจจำแนกสูง จะทำให้อำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลและวิธีซิปเทสท์มีค่าสูงใกล้เคียงกัน ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ลักษณะของข้อสอบไม่มีผลต่อวิธีการตรวจสอบทั้งสองวิธี การวิจัยครั้งนี้ศึกษาความยากของข้อสอบที่ระดับต่ำ ปานกลาง และสูง เพื่อให้ครอบคลุมการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ความยากของข้อสอบทุกระดับ

ความยาวของแบบสอบ (Test length) เป็นปัจจัยที่ส่งผลต่ออำนาจการทดสอบ กล่าวคือ แบบสอบที่มีความยาวมากกว่าย่อมส่งผลให้มีความน่าเชื่อถือมากกว่า ทำให้การจับคู่เปรียบเทียบระหว่างผู้สอบมีความถูกต้องมากขึ้น จากการศึกษาของ Swaminathan and Rogers (1990) โดยใช้ข้อมูลจำลองและตรวจสอบด้วยวิธีแมนเทล-แฮนส์เซล กับวิธีการถดถอยโลจิสติก ใช้แบบสอบที่มีความยาว 40 ข้อ 60 ข้อ และ 80 ข้อ พบว่า เมื่อใช้แบบสอบที่ยาว อำนาจการทดสอบของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจะดีกว่าการใช้แบบสอบที่สั้น ต่อมาในปี 1993 ทั้งสองคนได้ศึกษาอีกครั้งหนึ่ง โดยใช้แบบสอบที่มีความยาว 40 ข้อ และ 80 ข้อ ตรวจสอบด้วยวิธีแมนเทล-แฮนส์เซล กับวิธีการถดถอยโลจิสติก พบว่า ความยาวของแบบสอบไม่มีผลต่ออำนาจการทดสอบ ส่วนการศึกษาของ กาญจนา วัธนสุนทร (2537) พบว่า ความยาวของแบบสอบไม่มีผลต่ออัตราการตรวจสอบของวิธีแมนเทล-แฮนส์เซล และวิธีซิปเทสท์ ซึ่งขัดแย้งกับการศึกษาของจิติมา วรรณศรี (2539) ที่พบว่า เมื่อใช้แบบสอบความยาว 60 ข้อ จะมีผลทำให้วิธีแมนเทล-แฮนส์เซลและวิธีซิปเทสท์มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันดีที่สุด การวิจัยครั้งนี้ศึกษากับแบบสอบที่มีความยาว 40 ข้อ และ 60 ข้อ ซึ่งเป็นขนาดความยาวของแบบสอบที่เป็นตัวแทนของการทดสอบ

ผลสัมฤทธิ์ทางการเรียนในสถานการณ์ที่เป็นจริง และยังพบว่าข้อสอบที่มีความยาวปานกลางขึ้นไปจะส่งผลต่อประสิทธิภาพในการตรวจสอบมากที่สุด (จิติมา วรรณศรี, 2539)

สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน (Proportion of DIF) จากการศึกษาของ Rogers and Swaminathan (1993) พบว่า ในการตรวจสอบการทำหน้าที่ต่างกันแบบเอกรูปของวิธีการถดถอยโลจิสติก เมื่อสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันใบแบบสอบมีจำนวนลดลง จะมีผลทำให้อำนาจการทดสอบของวิธีวิธีการถดถอยโลจิสติกมีค่าเพิ่มมากขึ้น สำหรับการศึกษาของ Narayanan and Swaminathan (1994) พบว่า การตรวจสอบการทำหน้าที่ต่างกันแบบเอกรูป เมื่อสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบมีจำนวนลดลง มีผลทำให้อำนาจการทดสอบของวิธีซิปเทสท์และวิธีการถดถอยโลจิสติกมีค่าเพิ่มขึ้น และยังส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งสองวิธีมีค่าเพิ่มขึ้นด้วย การศึกษาที่ผ่านมา พบว่า หากแบบสอบมาตรฐานวัดผลสัมฤทธิ์ทางการเรียนมีข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10-15 ถือว่าไม่ผิดปกติ แต่หากมีข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 20 ขึ้นไป ถือว่าอยู่ในภาวะไม่ปกติ (Narayanan and Swaminathan, 1994; cited in Clauser and Mazor, 1998) ดังนั้น การวิจัยครั้งนี้จึงศึกษาสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน ร้อยละ 10 และร้อยละ 20

ผลการเปรียบเทียบวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ผ่านมา พบว่า ปัจจัยสำคัญที่ส่งผลต่ออำนาจการทดสอบ คือ ขนาดกลุ่มตัวอย่าง (Sample size) เช่น การศึกษาของ Swaminathan and Rogers (1990) พบว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและแบบอเนกรูป เมื่อเพิ่มขนาดกลุ่มตัวอย่างจะมีผลทำให้อำนาจการทดสอบของวิธีการถดถอยโลจิสติกมีค่าเพิ่มมากขึ้นเกือบทุกเงื่อนไข การศึกษาของ Rogers and Swaminathan (1993) พบว่า เมื่อขนาดกลุ่มตัวอย่างเพิ่มขึ้น จะทำให้อำนาจการทดสอบของวิธีการถดถอยโลจิสติกและวิธีแมนเทล-แฮนส์เซล มีค่าเพิ่มมากขึ้น นอกจากนี้ Narayanan and Swaminathan (1994) ได้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป พบว่า ขนาดกลุ่มตัวอย่างและอัตราส่วนของกลุ่มอ้างอิงต่อกลุ่มเปรียบเทียบมีผลต่ออำนาจการทดสอบของวิธีการตรวจสอบ กล่าวคือ เมื่อเพิ่มขนาดกลุ่มตัวอย่างจะทำให้อำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลและวิธีซิปเทสท์มีค่าเพิ่มมากขึ้น สำหรับการศึกษาของ จิติมา วรรณศรี (2539) พบว่า เมื่อขนาดกลุ่มตัวอย่าง 200 คน และ 600 คน วิธีแมนเทล-แฮนส์เซล และวิธีซิปเทสท์สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ถูกต้อง 50% แต่ถ้าขนาดกลุ่มตัวอย่าง 1000 คน สามารถตรวจสอบได้ถูกต้อง 100% นอกจากนี้ ยังมีการศึกษาของ Mazor และคนอื่น ๆ (1992 อ้างถึงใน ทองอยู่ สาระ, 2543) Narayanan and Swaminathan (1996) ได้ข้อค้นพบที่สอดคล้องกัน คือ เมื่อขนาดกลุ่มตัวอย่างใหญ่ขึ้น อำนาจการทดสอบจะเพิ่มขึ้น การวิจัยครั้งนี้ศึกษาขนาดกลุ่มตัวอย่างสามขนาด คือ 250 คน 500 คน และ

1000 คน เพื่อให้ครอบคลุมการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ขนาดกลุ่มตัวอย่างขนาดเล็ก ขนาดปานกลาง และขนาดใหญ่

จากที่กล่าวมาจะเห็นว่า งานวิจัยที่ผ่านมาเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค จะใช้วิธีการตรวจสอบอยู่สองแนวทาง ได้แก่ แนวทางที่กำหนดกลุ่มในการตรวจสอบไว้ล่วงหน้า (pre-specified groups) ซึ่งมีข้อดีในการอธิบายผลการตรวจสอบได้ชัดเจน แต่มีข้อจำกัดในเรื่องผลการตรวจสอบ DIF ในภายหลังที่เกิดจากตัวแปรที่ไม่ได้กำหนดไว้ตั้งแต่ต้น กับแนวทางที่ตรวจสอบกับกลุ่มผู้สอบทั้งหมด โดยไม่พิจารณาถึงตัวแปรที่เกี่ยวข้องกับตัวผู้สอบ ซึ่งมีข้อจำกัดในการอธิบายผลการตรวจสอบได้ไม่ชัดเจน ในปัจจุบันจากการวิจัยของ Strobl et al. (2015) ได้พัฒนาแนวทางใหม่ที่ผสมผสานสองแนวทางแรกไว้ด้วยกัน ที่เรียกว่า ราสช์ทรี คือ กำหนดกลุ่มในการตรวจสอบไว้ล่วงหน้าและพิจารณาถึงปฏิสัมพันธ์ของตัวแปรที่ใช้ในการตรวจสอบทั้งหมด ทำให้ผลการตรวจสอบสามารถอธิบายได้ชัดเจน และยังเพิ่มมิติรายละเอียดของผลการตรวจสอบได้มากกว่าวิธีการที่เคยใช้กันมา แต่ยังไม่มีการวิจัยศึกษาประสิทธิภาพของวิธีราสช์ทรีในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบว่ามีความเท่าเทียมหรือแตกต่างกับวิธีการตรวจสอบอื่น ๆ อย่างไร ดังนั้น ควรมีการศึกษาเปรียบเทียบประสิทธิภาพของผลการตรวจสอบของวิธีใหม่ คือ ราสช์ทรี กับวิธีการตรวจสอบที่วิเคราะห์ตามแนวทฤษฎีแบบดั้งเดิม คือ วิธีการถดถอยโลจิสติก และวิธีที่วิเคราะห์ตามแนวทฤษฎีการตอบสนองข้อสอบ คือ วิธีซิปเทสท์ โดยแบ่งการศึกษาเป็น 1) การทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปในข้อมูลจำลอง ทำให้ศึกษาการทำหน้าที่ต่างกันของข้อสอบได้หลากหลายเงื่อนไข และ 2) การทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและแบบอเนกรูปในข้อมูลเชิงประจักษ์ ทั้งนี้เนื่องจากผู้วิจัยไม่ได้จัดกระทำตามเงื่อนไขใด ๆ ในข้อมูลเชิงประจักษ์ เพื่อต้องการใช้ผลการสอบจากสถานการณ์ที่เป็นจริง จึงใช้วิธีการวัดพื้นที่ของราชูซึ่งเป็นวิธีที่มีประสิทธิภาพสูงดังได้กล่าวไปแล้วเป็นวิธีเกณฑ์ในเปรียบเทียบประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ จากผลการศึกษาดังกล่าว จะทำให้ทราบว่าวิธีการตรวจสอบที่พัฒนาขึ้นใหม่ โดย Strobl et al. (2015) มีประสิทธิภาพแตกต่างจากวิธีตรวจสอบตามแนวทฤษฎีแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบอย่างไร ซึ่งจะเป็นประโยชน์ต่อการเลือกใช้วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่เหมาะสม เมื่อข้อสอบมีค่าความยากของข้อสอบ ความยาวของแบบสอบ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน และขนาดกลุ่มตัวอย่างที่แตกต่างกัน อีกทั้งเป็นประโยชน์ต่อการเลือกใช้วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่เหมาะสมกับสถานการณ์สอบระดับประเทศ ภายใต้บริบทที่มีผู้สอบจำนวนมาก

คำถามการวิจัย

1. อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี โดยการทดสอบระดับนัยสำคัญและการวัดขนาดอิทธิพล ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ ความยากของข้อสอบ ความยาวของแบบสอบ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน และขนาดกลุ่มตัวอย่าง ให้ผลแตกต่างกันอย่างไร

2. อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี โดยการทดสอบระดับนัยสำคัญและการวัดขนาดอิทธิพล ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ ความยากของข้อสอบ ความยาวของแบบสอบ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน และขนาดกลุ่มตัวอย่าง ให้ผลแตกต่างกันอย่างไร

3. อำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค สำหรับข้อมูลเชิงประจักษ์ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ให้ผลแตกต่างกันอย่างไร

วัตถุประสงค์การวิจัย

1. เพื่อเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี โดยการทดสอบระดับนัยสำคัญและการวัดขนาดอิทธิพล ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ ความยากของข้อสอบ ความยาวของแบบสอบ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน และขนาดกลุ่มตัวอย่าง

2. เพื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี โดยการทดสอบระดับนัยสำคัญและการวัดขนาดอิทธิพล ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ ความยากของข้อสอบ ความยาวของแบบสอบ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน และขนาดกลุ่มตัวอย่าง

3. เพื่อเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค สำหรับข้อมูลเชิงประจักษ์ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี

สมมติฐานการวิจัย

Swaminathan and Rogers (1990 อ้างถึงใน ธเกียรติกมล ทองงอก, 2554) ได้ศึกษาเปรียบเทียบการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติกและวิธีแมนเทล-แฮนส์เซล พบว่า เมื่อใช้แบบสอบที่มีความยาวเพิ่มขึ้น ทำให้ความถูกต้องในการตรวจสอบของวิธีการถดถอยโลจิสติกเพิ่มมากขึ้นทั้งสองรูปแบบ กรณีกลุ่มตัวอย่างขนาดเล็กและแบบสอบสั้น วิธีการถดถอยโลจิสติกสามารถตรวจสอบได้ถูกต้องร้อยละ 50 ส่วนกรณีแบบสอบยาวและกลุ่มตัวอย่างขนาดใหญ่ ตรวจสอบได้ถูกต้องร้อยละ 75 ต่อมาจากการศึกษาของ Rogers and Swaminathan (1993) ได้ศึกษาซ้ำกรณีของความยากและอำนาจจำแนกของข้อสอบ พบว่า วิธีแมนเทล-แฮนส์เซลและ วิธีการถดถอยโลจิสติกมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปเท่าเทียมกัน และวิธีการถดถอยโลจิสติกตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปได้ดีในกรณีที่ข้อสอบมีความยากปานกลางและอำนาจจำแนกสูง ยกเว้นการกระจายค่าสถิติของวิธีการถดถอยโลจิสติกไม่เป็นไปตามคาดไว้ในกรณีข้อสอบยากมากและอำนาจจำแนกสูง ส่วน Narayanan and Swaminathan (1994) ได้ศึกษาการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป พบว่า เมื่อการแจกแจงค่าความสามารถระหว่างกลุ่มผู้สอบมีค่าเท่ากัน วิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีประสิทธิภาพเท่าเทียมกัน แต่เมื่อการแจกแจงค่าความสามารถระหว่างกลุ่มผู้สอบมีค่าไม่เท่ากัน วิธีซิปเทสท์มีประสิทธิภาพสูงกว่าวิธีแมนเทล์-แฮนส์เซล สำหรับ วลีมาศ แซ่อึ้ง (2543) ได้ศึกษาการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป พบว่า วิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ที่ปรับปรุงขั้นตอนการวิเคราะห์มีอำนาจการทดสอบเท่าเทียมกันภายใต้เกือบทุกเงื่อนไข และทั้งสองวิธีมีอำนาจการทดสอบสูงกว่าวิธีซิปเทสท์แบบเดิมและวิธีแมนเทล-แฮนส์เซลที่ปรับปรุงขั้นตอนการวิเคราะห์ ภายใต้เกือบทุกเงื่อนไข ส่วนวิธีราสช์ทรี ที่มีพื้นฐานจากโมเดลการแบ่งส่วนแบบทำซ้ำ (model-based recursive partitioning) เหมือนกับวิธี classification trees ซึ่งจากการศึกษาของ Vaughn and Wang (2010) พบว่า วิธี classification trees ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบมีอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 เท่าเทียมกับวิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติก จากที่กล่าวมายังไม่มีหลักฐานสนับสนุนว่าวิธีการตรวจสอบทั้งสามวิธีมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเท่าเทียมกันหรือไม่ ดังนั้นผู้วิจัยจึงตั้งสมมติฐานการวิจัย ดังนี้

1. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาคระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี โดยการทดสอบระดับนัยสำคัญและการวัดขนาดอิทธิพล ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ ความยากของข้อสอบ ความยาวของแบบสอบ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน และขนาดกลุ่มตัวอย่าง มีอำนาจการทดสอบแตกต่างกัน

2. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี โดยการทดสอบระดับนัยสำคัญและการวัดขนาดอิทธิพล ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ ความยากของข้อสอบ ความยาวของแบบสอบ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน และขนาดกลุ่มตัวอย่าง มีอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกัน

3. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค สำหรับข้อมูลเชิงประจักษ์ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี มีอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกัน

ขอบเขตการวิจัย

1. ข้อมูลจำลอง มีขอบเขตการดำเนินการ ดังนี้

1.1 การจำลองข้อมูลภายใต้ทฤษฎีการตอบสนองข้อสอบ โมเดล 1 พารามิเตอร์ ซึ่งในการจำลองข้อมูล Swaminathan และ Gifford (1980 cited in Lautenschlager and Park, 1988) เสนอให้ใช้ค่าอำนาจจำแนก (ค่า a) ในช่วงตั้งแต่ .60 ถึง 2 ค่าความยาก (ค่า b) ในช่วงตั้งแต่ -2 ถึง 2 ในการศึกษาครั้งนี้ผู้วิจัยใช้ค่าพารามิเตอร์ b ในช่วงดังกล่าวและแบ่งค่า b ออกเป็นสามช่วง คือ ต่ำ ปานกลาง และสูง โดยใช้สเกลในแต่ละช่วงเท่ากัน ผลการแบ่งค่า b ปรากฏว่า ช่วงต่ำมีค่าตั้งแต่ -2 ถึงน้อยกว่า -0.67 $(\bar{b} = -1.34)$ ช่วงปานกลางมีค่าตั้งแต่ -0.67 ถึง 0.67 $(\bar{b} = 0.00)$ และช่วงสูงมีค่าตั้งแต่มากกว่า 0.67 ถึง 2 $(\bar{b} = 1.34)$ สำหรับค่าพารามิเตอร์ a กำหนดให้มีค่าคงที่เท่ากับ 1 และค่าพารามิเตอร์ c เท่ากับ 0 ส่วนพารามิเตอร์ของผู้สอบ (θ) กำหนดให้มี Mean เท่ากับ 0 และ SD เท่ากับ 1

1.2 การจำลองการตอบข้อสอบที่มีโครงสร้างวัดความสามารถเอกมิติที่ให้คะแนนแบบทวิวิภาค โดยกำหนดการทำหน้าที่ต่างกันแบบเอกรูป (uniform DIF) กำหนดค่าพารามิเตอร์อำนาจจำแนกมีค่าคงที่ ผลการตอบภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ ความยากของข้อสอบ 3 ระดับ คือ ค่า b ต่ำ ค่า b ปานกลาง และค่า b สูง ความยาวของแบบสอบ 2 ขนาด คือ 40 ข้อ และ 60 ข้อ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด คือ ร้อยละ 10 และร้อยละ 20 และขนาดกลุ่มตัวอย่าง 3 ขนาด คือ 250, 500 และ 1000 คน มีข้อมูลที่ศึกษาทั้งหมด 36 เงื่อนไข (3 ระดับ x 2 ขนาด x 2 ขนาด x 3 ขนาด)

2. ข้อมูลเชิงประจักษ์จากโครงการประเมินผลนักเรียนนานาชาติ หรือ PISA 2009 ด้านการรู้เรื่องการอ่าน (reading literacy) คัดเลือกเฉพาะข้อสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค กลุ่มตัวอย่างเป็นนักเรียนอายุ 15 ปี ที่กำลังศึกษาอยู่ในชั้นมัธยมศึกษาปีที่ 1 ขึ้นไป จากกลุ่มโรงเรียนทุก

สังกัด ครอบคลุมทุกพื้นที่ทั่วประเทศ ศึกษาการทำหน้าที่ต่างกันของข้อสอบตามตัวแปร 3 ตัว ได้แก่ เพศ สังกัดของโรงเรียน และภูมิภาค

3 . เงื่อนไขที่ศึกษาการทำหน้าที่ต่างกันของข้อสอบ ประกอบด้วย

3.1 ความยากของข้อสอบ 3 ระดับ

1) ค่า b ต่ำ $(\bar{b} = -1.34)$

2) ค่า b ปานกลาง $(\bar{b} = 0.00)$

3) ค่า b สูง $(\bar{b} = 1.34)$

3.2 ความยาวของแบบสอบ 2 ขนาด

1) แบบสอบที่มีความยาว 40 ข้อ

2) แบบสอบที่มีความยาว 60 ข้อ

3.3 สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด

1) ร้อยละ 10

2) ร้อยละ 20

3.4 ขนาดกลุ่มตัวอย่างประกอบด้วยกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีจำนวนกลุ่มตัวอย่างในแต่ละกลุ่มเท่ากันมี 3 ขนาด

1) จำนวน 250 คน

2) จำนวน 500 คน

3) จำนวน 1000 คน

4. ตัวแปรที่ใช้ในการวิจัย ประกอบด้วย

4.1 ตัวแปรอิสระ ได้แก่ วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 3 วิธี คือ

4.1.1 วิธีการถดถอยโลจิสติก

4.1.2 วิธีซิปเทสท์

4.1.3 วิธีราสช์ทรี

4.2 ตัวแปรตาม ได้แก่ ประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี โดยการทดสอบระดับนัยสำคัญ และระหว่างวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ โดยการวัดขนาดอิทธิพล ซึ่งพิจารณาจาก

4.2.1 อำนาจการทดสอบ (power rate)

4.2.2 อัตราความคลาดเคลื่อนประเภทที่ 1 (Type I error rate)

5. ในการศึกษาครั้งนี้ ผู้วิจัยใช้วิธีการวัดพื้นที่ชนิดไม่คิดเครื่องหมายของ Raju (1990) ภายใต้โมเดล 1 พารามิเตอร์ เป็นวิธีเกณฑ์ เพื่อใช้เป็นเกณฑ์เปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบในข้อมูลเชิงประจักษ์ ซึ่งตรวจสอบด้วยวิธีที่ศึกษา 3 วิธี ได้แก่ วิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธี

ราสช์ทรี กล่าวคือ ถ้าวิธีที่ศึกษาระบุข้อสอบที่ทำหน้าที่ต่างกันได้ตรงกับข้อสอบที่ถูกระบุด้วยวิธีการวัดพื้นที่ของราชู แสดงว่า วิธีที่ศึกษาตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ถูกต้อง แต่ถ้าวิธีที่ศึกษาระบุข้อสอบที่ทำหน้าที่ต่างกันได้ไม่ตรงกับข้อสอบที่ถูกระบุด้วยวิธีการวัดพื้นที่ของราชู แสดงว่า วิธีที่ศึกษาตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ไม่ถูกต้อง

6. ข้อตกลงเบื้องต้น เนื่องจากการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีราสช์ทรี ผู้นำเสนอวิธีนี้ยังไม่ได้ระบุดัชนีการทำหน้าที่ต่างกันเป็นรายข้อที่ชัดเจน แต่จากงานวิจัยได้แสดงให้เห็นผลของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่เกิดขึ้น ดังนั้น ผู้วิจัยจึงได้ทดลองวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบกับข้อมูลจำลองในเงื่อนไขที่ 1 ด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี พบว่า ข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบด้วยวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ส่วนใหญ่จะตรงกับข้อสอบที่มีความแตกต่างของค่าพารามิเตอร์ความยากของข้อสอบ ที่ค่า $\pm$ 4.000 ขึ้นไป ในวิธีราสช์ทรี จึงอาจกล่าวได้ว่า ข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบด้วยวิธีราส์ทรี คือ ข้อสอบที่มีความแตกต่างของค่าพารามิเตอร์ความยากของข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ตั้งแต่ $\pm$ 4.000 ขึ้นไป และเพื่อให้สามารถเปรียบเทียบประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้งสามวิธีได้ ดังนั้น ผู้วิจัยจึงกำหนดเกณฑ์การตัดสินการทำหน้าที่ต่างกันของข้อสอบเป็นรายข้อสำหรับวิธีราสช์ทรี โดยใช้ผลต่างของค่าพารามิเตอร์ความยากของข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ เป็นตัวชี้วัดการเกิดการทำหน้าที่ต่างกันของข้อสอบ ที่เกณฑ์ $|b| \geq 4.000$

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ข้อจำกัดในการวิจัย

เนื่องจากการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีราสช์ทรี เป็นการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้งฉบับของแบบสอบ จึงไม่มีการกำหนดดัชนีในการระบุขนาดอิทธิพลของการทำหน้าที่ต่างกันของข้อสอบเป็นรายข้อ ดังนั้น การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพล จึงใช้วิธีการตรวจสอบ 2 วิธี คือ วิธีการถดถอยโลจิสติกและวิธีซิปเทสท์

คำจำกัดความที่ใช้ในการวิจัย

การทำหน้าที่ต่างกันของข้อสอบ หมายถึง โอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกันสำหรับผู้สอบที่มีความสามารถหรือคุณลักษณะในระดับเท่ากัน แต่มาจากกลุ่มประชากรที่แตกต่างกัน

การตรวจให้คะแนนแบบทวิวิภาค หมายถึง การให้คะแนนการตอบข้อสอบแบบมี 2 ค่า คือ ตอบถูก ได้ 1 คะแนน และตอบผิด ได้ 0 คะแนน

กลุ่มอ้างอิง หมายถึง กลุ่มผู้สอบที่คาดว่าจะได้เปรียบในการตอบข้อสอบ โดยมีโอกาสตอบข้อสอบได้ถูกต้องมากกว่าอีกกลุ่มหนึ่ง ที่มีระดับความสามารถเท่ากัน สำหรับข้อมูลจำลองจะกำหนดกลุ่มอ้างอิงจากการจำลองข้อมูล สำหรับข้อมูลเชิงประจักษ์ ได้แก่ กลุ่มผู้สอบเพศหญิง เนื่องจากมีความสามารถด้านภาษามากกว่าเพศชาย ส่วนตัวแปรสังกัดของโรงเรียน และภูมิภาค จะกำหนดกลุ่มอ้างอิงจากการตรวจสอบโดยวิธีราสช์ทรี

กลุ่มเปรียบเทียบ หมายถึง กลุ่มผู้สอบที่คาดว่าจะเสียเปรียบในการตอบข้อสอบ โดยมีโอกาสข้อสอบได้ถูกต้องน้อยกว่าอีกกลุ่มหนึ่ง ที่มีระดับความสามารถเท่ากัน สำหรับข้อมูลจำลองจะกำหนดกลุ่มเปรียบเทียบจากการจำลองข้อมูล สำหรับข้อมูลเชิงประจักษ์ ได้แก่ กลุ่มผู้สอบเพศชาย เนื่องจากมีความสามารถด้านภาษาน้อยกว่าเพศหญิง ส่วนตัวแปรสังกัดของโรงเรียน และภูมิภาค จะกำหนดกลุ่มเปรียบเทียบจากการตรวจสอบโดยวิธีราสช์ทรี

วิธีการถดถอยโลจิสติก หมายถึง วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่วิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิม โดยการคำนวณดัชนีการทำหน้าที่ต่างกันของข้อสอบจากผลการตอบข้อสอบถูกระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ใช้โมเดลการถดถอยโลจิสติกและทดสอบระดับนัยสำคัญด้วยสถิติ χ^2

วิธีซิปเทสท์ หมายถึง วิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่วิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ ซึ่งจะคำนวณดัชนีการทำหน้าที่ต่างกันของข้อสอบจากค่าเฉลี่ยสัดส่วนการตอบข้อสอบถูกระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบในชุดแบบสอบที่ใช้ศึกษา (studied subtest) โดยใช้คะแนนการจับคู่เปรียบเทียบระหว่างกลุ่มผู้สอบจากชุดแบบสอบที่มีความตรง (valid subtest) และทดสอบระดับนัยสำคัญด้วยสถิติ Z

วิธีราสช์ทรี หมายถึง วิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยโมเดล 1 พารามิเตอร์ ที่มีพื้นฐานการทำงานจากโมเดลการแบ่งส่วนแบบทำซ้ำ (model-based recursive partitioning) โดยตรวจสอบความแตกต่างของค่าพารามิเตอร์ข้อสอบระหว่างกลุ่มผู้สอบ จากผลของปฏิสัมพันธ์ระหว่างตัวแปรทุกตัว

วิธีการวัดพื้นที่ของราชู หมายถึง วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่วิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ โดยการคำนวณดัชนีการทำหน้าที่ต่างกันของข้อสอบจากพื้นที่ระหว่างโค้งลักษณะข้อสอบโดยใช้การอินทิเกรทพื้นที่ในช่วงเปิด จากผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ และทดสอบระดับนัยสำคัญด้วยสถิติ Z ในการวิจัยครั้งนี้ ผู้วิจัยใช้วิธีการวัดพื้นที่ชนิดไม่คิดเครื่องหมาย ภายใต้โมเดล 1 พารามิเตอร์ ชนิดกำหนดค่า a เท่ากับ 1 และค่า c เท่ากับ 0

กลุ่มข้อสอบที่มีค่าความยากต่ำ หมายถึง กลุ่มข้อสอบที่มีค่าความยากเฉลี่ย -1.34 โดยมีค่าความยากตั้งแต่ -2.00 ถึงน้อยกว่า -0.67 ที่ใช้หลักการวิเคราะห์แบบ IRT

กลุ่มข้อสอบที่มีค่าความยากปานกลาง หมายถึง กลุ่มข้อสอบที่มีค่าความยากเฉลี่ย 0.00 โดยมีค่าความยากตั้งแต่ -0.67 ถึง 0.67 ใช้หลักการวิเคราะห์แบบ IRT

กลุ่มข้อสอบที่มีค่าความยากสูง หมายถึง กลุ่มข้อสอบที่มีค่าความยากเฉลี่ย 1.34 โดยมีค่าความยากตั้งแต่มากกว่า 0.67 ถึง 2.00 ใช้หลักการวิเคราะห์แบบ IRT

ความยาวของแบบสอบ หมายถึง จำนวนข้อสอบในแบบสอบแต่ละฉบับ มี 2 ขนาด คือ จำนวน 40 ข้อ และจำนวน 60 ข้อ

สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน หมายถึง ข้อสอบที่ทำหน้าที่ต่างกันจากข้อสอบทั้งหมดในแต่ละฉบับ โดยคิดจำนวนเป็นร้อยละ มี 2 ขนาด คือ ร้อยละ 10 และร้อยละ 20

ขนาดกลุ่มตัวอย่าง หมายถึง จำนวนผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่ใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ มี 3 ขนาด คือ 250 คน 500 คน และ 1000 คน

ประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ หมายถึง ความถูกต้องของการระบุการทำหน้าที่ต่างกันของข้อสอบ จากการตรวจสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ซึ่งพิจารณาได้จากอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1

อำนาจการทดสอบ หมายถึง ร้อยละของจำนวนข้อสอบที่ระบุได้ถูกต้องว่าทำหน้าที่ต่างกัน ซึ่งตรวจสอบด้วยวิธีที่ศึกษา 3 วิธี ต่อจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งหมดในแบบสอบกรณีข้อมูลจำลอง และต่อจำนวนข้อสอบที่ระบุว่าทำหน้าที่ต่างกันทั้งหมดในแบบสอบซึ่งตรวจสอบด้วยวิธีที่ใช้เป็นเกณฑ์กรณีข้อมูลเชิงประจักษ์

อัตราความคลาดเคลื่อนประเภทที่ 1 หมายถึง ร้อยละของจำนวนข้อสอบที่ระบุผิดพลาดว่าทำหน้าที่ต่างกัน ทั้งที่ความเป็นจริงทำหน้าที่ไม่ต่างกัน ซึ่งตรวจสอบด้วยวิธีที่ศึกษา 3 วิธี ต่อจำนวนข้อสอบที่ทำหน้าที่ไม่ต่างกันทั้งหมดในแบบสอบกรณีข้อมูลจำลอง และต่อจำนวนข้อสอบที่ระบุว่าทำหน้าที่ไม่ต่างกันทั้งหมดในแบบสอบซึ่งตรวจสอบด้วยวิธีที่ใช้เป็นเกณฑ์กรณีข้อมูลเชิงประจักษ์

เกณฑ์ของอัตราความคลาดเคลื่อนประเภทที่ 1 หมายถึง เกณฑ์ที่ใช้เปรียบเทียบกับอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบที่ศึกษา 3 วิธี ได้แก่ วิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ในการศึกษาครั้งนี้ผู้วิจัยใช้เกณฑ์ที่ระดับ 10%

ขนาดอิทธิพล หมายถึง ขนาดหรือค่าความเข้มของการทำหน้าที่ต่างกันของข้อสอบ ซึ่งเป็นระดับความสัมพันธ์ของตัวแปรอิสระกับตัวแปรตาม โดยจะตัดสินข้อสอบที่ทำหน้าที่ต่างกันเมื่อขนาดอิทธิพลมีขนาดปานกลางและขนาดใหญ่ จากค่าสถิติ R^2 ตามเกณฑ์ของ Jodoin and Gierl (2001) และเมื่อขนาดอิทธิพลมีขนาดเล็ก ขนาดปานกลางและขนาดใหญ่ จากค่า $\hat{\beta}$ ตามเกณฑ์ของ Roussos และ Stout (1996 อ้างถึงใน Magis et al., 2010) ตามเกณฑ์ดังนี้

1. เกณฑ์ของ Jodoin and Gierl (2001) ซึ่งกำหนดเกณฑ์การตัดสินขนาดอิทธิพลของวิธีการถดถอยโลจิสติกใน 3 ระดับ ดังนี้

ค่าความแตกต่างของ R^2 มีค่าน้อยกว่า 0.035 แสดงว่า ขนาดอิทธิพลมีขนาดเล็ก ยอมรับสมมติฐานว่าง (Null hypothesis) ไม่เกิด DIF

ค่าความแตกต่างของ R^2 มีค่าระหว่าง 0.035 ถึง 0.069 แสดงว่า ขนาดอิทธิพลมีขนาดปานกลาง

ค่าความแตกต่างของ R^2 มีค่ามากกว่าหรือเท่ากับ 0.07 แสดงว่า ขนาดอิทธิพลมีขนาดใหญ่

2. เกณฑ์ของ Roussos และ Stout (1996 อ้างถึงใน Magis et al., 2010) ซึ่งกำหนดเกณฑ์การตัดสินขนาดอิทธิพลของวิธีซิปเทสท์ใน 3 ระดับ ดังนี้

ค่าสัมบูรณ์ของ $\hat{\beta}$ มีค่าน้อยกว่า 0.059 แสดงว่า ขนาดอิทธิพลมีขนาดเล็ก ปฏิเสธสมมติฐานว่าง (Null hypothesis) เกิด DIF

ค่าสัมบูรณ์ของ $\hat{\beta}$ มีค่าระหว่าง 0.059 ถึง 0.087 แสดงว่า ขนาดอิทธิพลมีขนาดปานกลาง

ค่าสัมบูรณ์ของ $\hat{\beta}$ มีค่ามากกว่าหรือเท่ากับ 0.088 แสดงว่า ขนาดอิทธิพลมีขนาดใหญ่

ประโยชน์ที่จะได้รับ

การวิจัยครั้งนี้เป็นการวิเคราะห์เปรียบเทียบประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ตามเกณฑ์การทดสอบระดับนัยสำคัญและการวัดขนาดอิทธิพล โดยศึกษาข้อมูลจำลองตามปัจจัยที่แปรเปลี่ยน ร่วมกับศึกษาข้อมูลเชิงประจักษ์ ซึ่งเป็นการขยายขอบเขตการศึกษาเรื่องการทำหน้าที่ต่างกันของข้อสอบ ผลการศึกษามีประโยชน์ดังนี้

1. ประโยชน์ด้านวิชาการ เป็นการขยายขอบเขตการศึกษาการทำหน้าที่ต่างกันของวิธีราสช์ทรีที่สามารถนำเสนอผลการตรวจสอบที่เข้าใจได้ง่าย มีแง่มุมที่แตกต่างจากวิธีตรวจสอบอื่น ๆ นอกจากนี้ ยังนำมาเปรียบเทียบประสิทธิภาพในการตรวจสอบกับวิธีการตรวจสอบที่มีจุดเด่นและเป็นที่นิยม เพื่อจะได้เลือกวิธีการตรวจสอบที่เหมาะสมกับลักษณะของข้อสอบที่ต้องการนำไปใช้ ภายใต้ปัจจัยที่แปรเปลี่ยน ได้แก่ ความยากของข้อสอบ ความยาวของแบบสอบ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน และขนาดกลุ่มตัวอย่าง

2. ประโยชน์ด้านการนำไปใช้ กรณีการศึกษากับข้อมูลเชิงประจักษ์ ทำให้ได้สารสนเทศในการเลือกใช้วิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่เหมาะสมกับข้อสอบที่ใช้ในบริบทของการศึกษาที่ต้องการใช้ผลการสอบเพื่อประเมินความสามารถของผู้สอบ

บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

การศึกษาครั้งนี้ ผู้วิจัยขอนำเสนอแนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง แบ่งเป็น 4 ตอน ได้แก่

ตอนที่ 1 มโนทัศน์ของการทำหน้าที่ต่างกันของข้อสอบ

1.1 ความเป็นมาของการทำหน้าที่ต่างกันของข้อสอบ

1.2 ความหมายของการทำหน้าที่ต่างกันของข้อสอบ

1.3 ประเภทของการทำหน้าที่ต่างกันของข้อสอบ

ตอนที่ 2 วิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

2.1 วิธีการถดถอยโลจิสติก

2.2 วิธีซิปเทสท์

2.3 วิธีราสช์ทรี

2.4 วิธีการวัดพื้นที่ของราชู

2.5 เกณฑ์สำหรับเปรียบเทียบคุณภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

2.6 ขนาดอิทธิพลกับการศึกษาการทำหน้าที่ต่างกันของข้อสอบ

ตอนที่ 3 โครงการประเมินผลนักเรียนนานาชาติ (Programme for International Student Assessment; PISA)

ตอนที่ 4 งานวิจัยที่เกี่ยวกับการศึกษาการทำหน้าที่ต่างกันของข้อสอบในแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค

4.1 งานวิจัยที่เกี่ยวข้องกับการศึกษาปัจจัยที่มีผลต่อการทำหน้าที่ต่างกันของข้อสอบในแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค

4.2 งานวิจัยที่เกี่ยวข้องกับการศึกษาเปรียบเทียบประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค

4.3 สรุปประเด็นการศึกษาการทำหน้าที่ต่างกันของข้อสอบในแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค

4.4 กรอบแนวคิดในการวิจัย

โดยมีรายละเอียดดังนี้

ตอนที่ 1 มโนทัศน์ของการทำหน้าที่ต่างกันของข้อสอบ

1.1 ความเป็นมาของการทำหน้าที่ต่างกันของข้อสอบ

การศึกษาความลำเอียงของข้อสอบมีการศึกษาอย่างจริงจังในช่วงปี ค.ศ.1960 เพื่อพัฒนาวิธีการสำหรับศึกษาความแตกต่างทางวัฒนธรรมและเพื่อรวบรวมสำหรับการยืนยัน เนื่องจากความไม่เท่าเทียมกันระหว่างนักเรียนผิวดำและลาตินอเมริกา กับนักเรียนผิวขาวในการทดสอบความสามารถทางสมอง ซึ่งข้อสอบเกี่ยวข้องกับเนื้อหาที่นักเรียนชนกลุ่มน้อย เช่น นักเรียนที่เป็นคนผิวดำหรือเชื้อชาติที่มาจากประเทศอื่น ๆ มีโอกาสที่จะเรียนรู้ได้น้อยกว่า และเป้าหมายที่เฉพาะคือการระบุข้อสอบที่มีความลำเอียงต่อนักเรียนชนกลุ่มน้อยและตัดข้อสอบเหล่านั้นออกจากแบบสอบ ในระยะเวลาเดียวกันก็ได้มีการศึกษาเพื่อรวบรวมความสมบูรณ์ของแบบสอบ โดยเฉพาะแบบสอบที่ใช้เพื่อการคัดเลือก เนื่องมาจากความลำเอียงต่อชนกลุ่มน้อยทำให้คะแนนของพวกเขาต่ำกว่าตำแหน่งบนเกณฑ์การวัดที่ควรจะเป็น

สิ่งที่แบ่งแยกได้สองอย่าง คือ 1) เป้าหมายที่ต้องการ และ 2) โมเดลและวิธีการที่แตกต่าง การใช้เกณฑ์ภายนอกยังคงมีอยู่ในการตรวจสอบและประเมินความลำเอียงในแบบสอบที่ใช้เพื่อการคัดเลือก ในขณะที่ไม่ค่อยมีการใช้เกณฑ์ภายนอกเพื่อพิจารณาให้เกิดเป็นประโยชน์อย่างฉับพลันสำหรับประเมินความลำเอียงในข้อสอบ ผลที่ตามมา ผู้รวบรวมได้ศึกษาความลำเอียงของข้อสอบใช้การออกแบบการใช้แบบสอบที่สมบูรณ์ ทั้งทางตรงและทางอ้อม แทนที่เกณฑ์เพื่อจับคู่กลุ่มโดยคำนึงถึงความสามารถ การศึกษาทั้งสองประเภทรู้จักในระยะแรกว่าเป็นการศึกษาความลำเอียงซึ่งเกณฑ์ยังไม่เพียงพอ โดยเฉพาะถ้าตัวเกณฑ์มีความลำเอียงเอง (Angoff, 1993)

การศึกษาความลำเอียงของข้อสอบและแบบสอบ โดยเฉพาะการใช้คำว่า “ความลำเอียง” ผู้เขียนหลายคน (Ironson, 1982; Linn and Drasgow, 1987; Linn et al., 1981; and Shepard, Camilli and Averill, 1981 cited in Angoff, 1993) ได้นิยาม ความลำเอียงของข้อสอบ ไว้มากมายในแนวทางเดียวกัน โดยเฉพาะข้อสอบลำเอียง เมื่อผู้สอบที่มีความสามารถเท่ากัน (หรือประสิทธิภาพ) จากต่างกลุ่มกัน แต่มีความน่าจะเป็นในการตอบข้อสอบถูกไม่เท่ากัน (ความหมายเหมือนกันในหลักการ การนำไปใช้กับความลำเอียงในแบบสอบ) Shepard et al. (cited in Angoff, 1993) กล่าวว่า ลักษณะของความลำเอียงเป็นการขาดความตรงที่เป็นอันตรายต่อกลุ่มหนึ่งมากกว่าอีกกลุ่มหนึ่ง สิ่งแรกของคำนิยาม เป็นข้อสังเกตที่อ้างถึงการสังเกตง่าย ๆ ในผลการปฏิบัติที่แตกต่างกัน สิ่งที่สองเป็นการรับรู้จากการประเมิน และอ้างถึงผลภายในกลุ่ม ดังนั้น ในความหมายแรก ความลำเอียงของข้อสอบควรจะเป็นการระบุ แต่มีเพียงข้อค้นพบทางสถิติ ต่อมาจึงเป็นการตีความและพิจารณา บางข้อสอบที่ผิดแปลกไปอาจมีความลำเอียง ในการรับรู้ที่เป็นความไม่ยุติธรรมซึ่งเป็นการเสียเปรียบต่อชนกลุ่มน้อย อย่างไรก็ตาม อีกด้านหนึ่งถูกพิจารณาเป็นความยุติธรรม ในการรับรู้ที่เรียกว่าผลลัพธ์

ทางการศึกษาที่สำคัญ ความเหมาะสมสำหรับนักเรียนทั้งหมด แต่รู้และเข้าใจถึงความไม่เท่าเทียมกันของนักเรียนทั้งหมด

คำว่า “ความลำเอียง” ได้ถูกใช้พร้อม ๆ กัน แต่ไม่สามารถเข้าใจได้ทั้งหมด ความหมายที่แตกต่างกันสองทาง ในทางสังคมและทางสถิติ ผลที่ตามมาเป็นข้อสงสัยที่ไม่ควรเกิดขึ้น นำไปสู่บรรยากาศที่สับสนทางการเมือง ในสิ่งที่ความลำเอียงถือเป็นสาเหตุใหญ่ในความไม่เท่าเทียมกันของคะแนน บางคำแนะนำได้ใช้คำอื่นแทนความลำเอียงสำหรับการสังเกตทางสถิติ แยกต่างหากเพื่อการตัดสินหรือตีความหมายและการนำไปใช้ และในอีกด้านหนึ่ง คำที่ใช้เพื่ออธิบายในการตัดสินและประเมินความลำเอียงในทางการรับรู้ทางสังคม ในท้ายที่สุด คำว่า “การทำหน้าที่ต่างกันของข้อสอบ” (DIF) ถูกนำมาใช้ (Angoff, 1993; Masters & Keeves, 1999) โดยอ้างถึงการสังเกตอย่างง่าย ๆ ที่ข้อสอบได้แสดงความแตกต่างกันในทางสถิติในกลุ่มที่ต่างกัน (หลังจากควบคุมความแตกต่างกันในเรื่องความสามารถของกลุ่ม) อย่างไรก็ตาม DIF เป็นคำที่ดีกว่าสำหรับการสังเกตที่เป็นปรนัย คือ ปราศจากการตัดสินล่วงหน้าของความไม่เท่าเทียมกันของกลุ่ม อิสระจากการประเมินคุณค่า ยอมให้ความลำเอียงเป็นประเด็นทางสังคมเพื่อตรวจสอบโดยแยกจากกัน

1.2 ความหมายของการทำหน้าที่ต่างกันของข้อสอบ

เมื่อกล่าวถึงการทำหน้าที่ต่างกันของข้อสอบ ซึ่งมีความเชื่อมโยงกับความลำเอียงของข้อสอบ ดังนั้นในการให้ความหมายของการทำหน้าที่ต่างกันของข้อสอบ จึงต้องนำเสนอในส่วนของความลำเอียงของข้อสอบร่วมด้วย ดังนี้

Scheuneman (1979 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555) กล่าวว่า ความลำเอียงของข้อสอบ หมายถึง สัดส่วนของผู้สอบที่ตอบข้อสอบได้ถูกต้องไม่เท่ากัน ในแต่ละกลุ่มประชากรที่ใช้ในการศึกษา เมื่อกลุ่มผู้สอบมีคะแนนเท่ากันและข้อสอบมีความเป็นเอกพันธ์

Dorans and Kulick (1986 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555) กล่าวว่า ความลำเอียงของข้อสอบ หมายถึง โอกาสในการตอบข้อสอบได้ถูกต้องของผู้สอบกลุ่มหนึ่ง มีค่าต่ำกว่าหรือสูงกว่าผู้สอบอีกกลุ่มหนึ่งที่มีระดับความสามารถเดียวกัน

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

Wiersma and Jurs (1990) กล่าวว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง โอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกันของผู้สอบที่มีปริมาณความรู้เท่ากัน ที่มาจากตั้งแต่สองกลุ่มขึ้นไป

Embretson and Reise (2000) กล่าวว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ข้อสอบที่มีความสัมพันธ์กับคุณลักษณะแฝงต่างกันในกลุ่มผู้สอบตั้งแต่สองกลุ่มขึ้นไป

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

Kline (2005) กล่าวว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง โอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกัน ของผู้สอบที่มีระดับคุณลักษณะพื้นฐานเท่ากัน แต่มาจากต่างกลุ่มกัน

Brennan (2006) กล่าวว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ระดับความสำเร็จในการตอบข้อสอบที่ต่างกันของผู้สอบต่างกลุ่มที่มีความสามารถเท่ากันในเป้าหมายที่ต้องการวัด

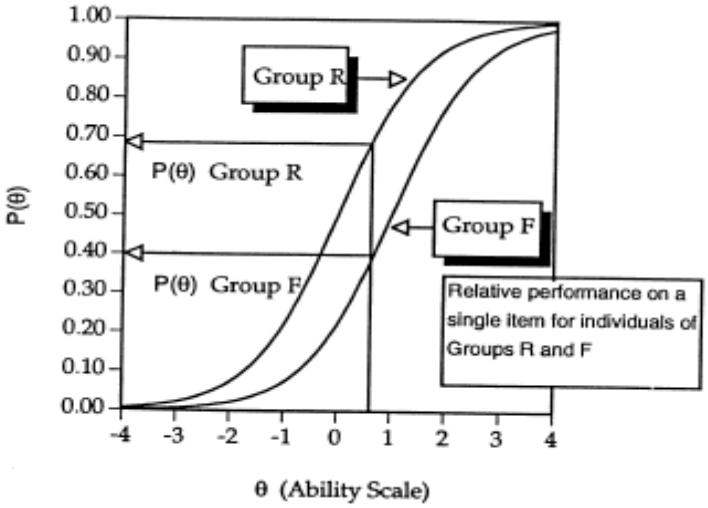
จากความหมายของการทำหน้าที่ต่างกันของข้อสอบ สรุปได้ว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง โอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกัน สำหรับผู้สอบที่มีความสามารถหรือคุณลักษณะในระดับเท่ากันที่มาจากกลุ่มประชากรที่แตกต่างกัน

1.3 ประเภทของการทำหน้าที่ต่างกันของข้อสอบ

การทำหน้าที่ต่างกันของข้อสอบ เป็นการเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มผู้สอบอย่างน้อยสองกลุ่มขึ้นไป โดยปกตินิยมทำการเปรียบเทียบสองกลุ่ม ประกอบด้วย กลุ่มที่สนใจศึกษาและคาดว่าจะเป็นกลุ่มที่เสียเปรียบในการตอบข้อสอบได้ถูกต้อง เรียกว่า กลุ่มเปรียบเทียบ (focal group หรือ กลุ่ม F) และกลุ่มที่คาดว่าจะได้เปรียบในการตอบข้อสอบได้ถูกต้อง เรียกว่า กลุ่มอ้างอิง (reference group หรือ กลุ่ม R) (ศิริชัย กาญจนวาสี, 2555)

ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ จะพบลักษณะของข้อสอบที่ทำหน้าที่ต่างกันได้สองประเภท ได้แก่ การทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป (uniform DIF) และการทำหน้าที่ต่างกันแบบอเนกรูป (nonuniform DIF) (Mellenbergh, 1982 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555)

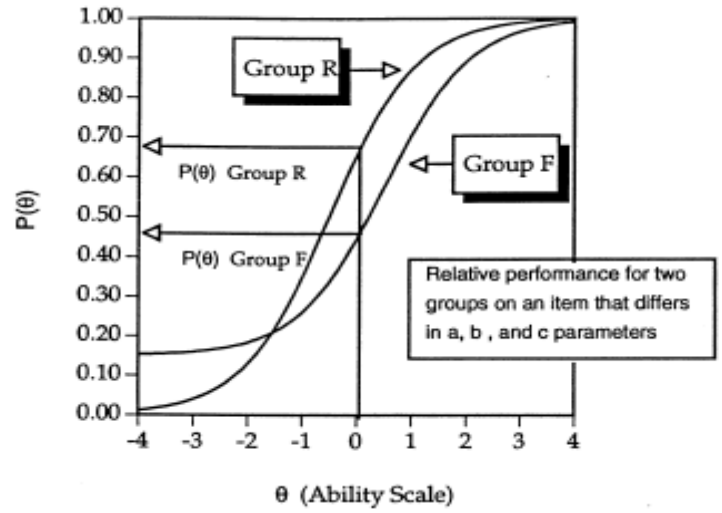
1. การทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป (uniform DIF) หมายถึง ข้อสอบที่ทำให้ผู้สอบกลุ่มหนึ่งมีโอกาสในการตอบข้อสอบได้ถูกต้องมากกว่าผู้สอบอีกกลุ่มหนึ่งอย่างสม่ำเสมอกัน ในทุกระดับความสามารถของผู้สอบ เมื่อพิจารณาจากโค้งลักษณะข้อสอบ จะไม่พบปฏิสัมพันธ์ (interaction) ระหว่างระดับความสามารถของผู้สอบกับการเป็นสมาชิกของกลุ่มผู้สอบ (group membership) ดังภาพที่ 1



ภาพที่ 1 ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป (uniform DIF)

ที่มา: Camilli and Shepard (1994)

2. การทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป (nonuniform DIF) หมายถึง ข้อสอบที่ทำให้โอกาสในการตอบข้อสอบได้ถูกต้องของผู้สอบระหว่างกลุ่มเกิดขึ้นอย่างไม่สม่ำเสมอกัน ในทุกระดับความสามารถของผู้สอบ เมื่อพิจารณาจากโค้งลักษณะข้อสอบ จะพบปฏิสัมพันธ์ (interaction) ระหว่างระดับความสามารถของผู้สอบกับการเป็นสมาชิกของกลุ่มผู้สอบ (group membership) ดังภาพที่ 2



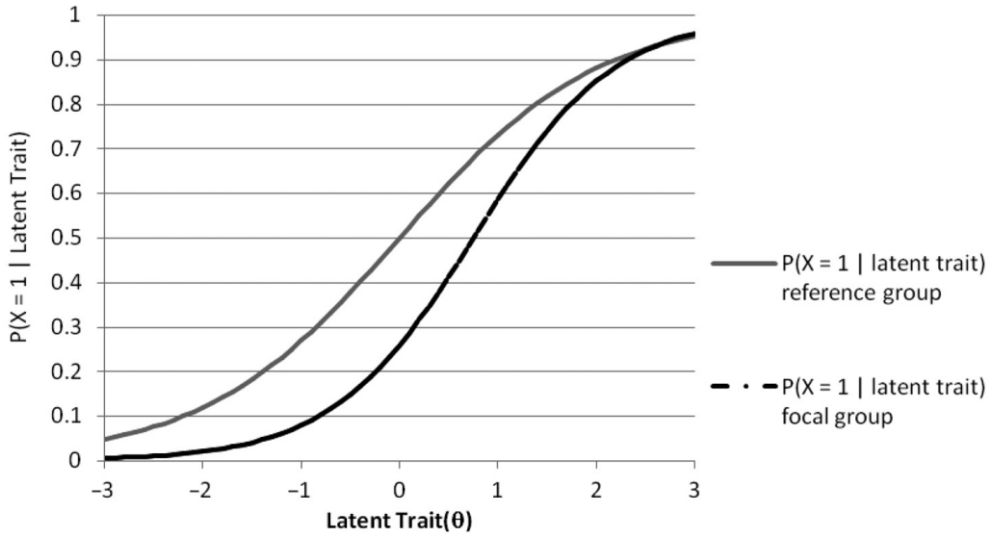
ภาพที่ 2 ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป (nonuniform DIF)

ที่มา: Camilli and Shepard (1994)

ตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory; IRT) สามารถพิจารณาปฏิสัมพันธ์ ได้จากความแตกต่างของค่าพารามิเตอร์อำนาจจำแนกข้อสอบ ระหว่างผู้สอบกลุ่มย่อยสองกลุ่ม กล่าวคือ ถ้าข้อสอบทำหน้าที่ต่างกันแบบเอกรูปแล้ว โค้งลักษณะข้อสอบ (Item Characteristic Curves; ICCs) ระหว่างผู้สอบกลุ่มย่อยสองกลุ่มจะขนานกันหรือมีฟังก์ชันนการตอบสนองข้อสอบ (Item Response Functions; IRFs) เหมือนกัน แต่ถ้าข้อสอบทำหน้าที่ต่างกันแบบอเนกรูปแล้ว โค้งลักษณะข้อสอบระหว่างผู้สอบกลุ่มย่อยสองกลุ่มจะไม่ขนานกันหรือมีฟังก์ชันการตอบสนองข้อสอบต่างกัน ดังนั้นความแตกต่างระหว่างโค้งลักษณะข้อสอบทั้งสองแบบจะบ่งบอกถึงขนาดและทิศทางของข้อสอบที่ทำหน้าที่ต่างกัน ซึ่งสามารถคำนวณได้โดยใช้สูตรการคำนวณพื้นที่ของราชู (1990 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555)

ข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป สามารถจำแนกได้เป็นสองลักษณะ (Swaminathan & Rogers, 1990) ดังนี้

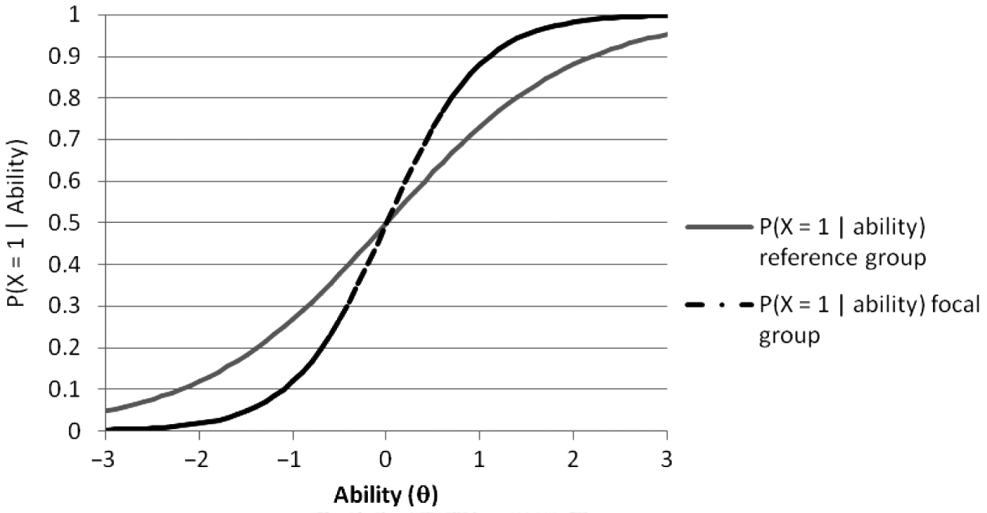
1. ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป โดยมีปฏิสัมพันธ์เป็นลำดับ (ordinal interaction) เป็นการทำหน้าที่ต่างกันสำหรับกลุ่มผู้สอบ ซึ่งเกิดขึ้นเมื่อโค้งลักษณะข้อสอบต่างกันอย่างไม่สม่ำเสมอ แต่ไม่ตัดกันหรืออาจตัดกันนอกช่วงความสามารถของผู้สอบตรงปลายสุดของช่วงความสามารถต่ำหรือสูง อาจเรียกข้อสอบลักษณะนี้ว่า ข้อสอบทำหน้าที่ต่างกันแบบมีทิศทางเดียว (unidirectional DIF) ดังภาพที่ 3



ภาพที่ 3 ข้อสอบทำหน้าที่ต่างกันแบบมีทิศทางเดียวกัน (unidirectional DIF)

ที่มา: Walker (2011)

2. ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป โดยมีปฏิสัมพันธ์ไม่เป็นลำดับ (disordinal interaction) เป็นการทำหน้าที่ต่างกันสำหรับกลุ่มผู้สอบ ซึ่งเกิดขึ้นเมื่อโค้งลักษณะข้อสอบตัดกันระหว่างช่วงความสามารถของผู้สอบ หรือเรียกว่า ข้อสอบทำหน้าที่ต่างกันแบบไม่มีทิศทาง (non-unidirectional DIF) ดังภาพที่ 4



ภาพที่ 4 ข้อสอบทำหน้าที่ต่างกันแบบไม่มีทิศทาง (non-unidirectional DIF)

ที่มา: Walker (2011)

ตอนที่ 2 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

นักวัดผลทางการศึกษา ได้จำแนกวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ไว้ดังนี้

Haladyna (1999) ได้จำแนกวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ไว้ 5 วิธี ได้แก่

1. วิธีใช้โมเดล IRT (IRT methods)
2. วิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel (MH) statistic)
3. วิธีดัชนีมาตรฐาน (standardization)
4. วิธีซิปเทสท์ (SIBTEST)
5. วิธีการถดถอยโลจิสติก (logistic regression)

Masters and Keeves (1999) ได้จำแนกวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบตามทฤษฎีการวัด เป็น 2 ประเภท ได้แก่

1. ทฤษฎีการวัดแบบดั้งเดิม (classical approaches)
   1.1 วิธีแปลงค่าความยากของข้อสอบ (transformed item difficulty methods: TID)
   1.2 วิธีดัชนีค่าอำนาจจำแนกของข้อสอบ (item discrimination procedures)

1.3 วิธีตารางการณ์จร (contingency table approaches)

1.3.1 วิธีไคสแควร์ (chi-square methods)

1.3.2 วิธีลอก-ลีเนียร์ (log-linear methods)

1.3.3 วิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel procedure)

1.4 วิธีดัชนีมาตรฐาน (standardization procedure)

1.5 วิธีการถดถอยโลจิสติก (logistic regression method)

1.6 วิธีวิเคราะห์ตัวลวง (distractor analysis)

2. ทฤษฎีการตอบสนองข้อสอบ (item response theory approaches)

2.1 วิธีสามพารามิเตอร์ (three-parameter methods)

2.2 วิธีราส์ชโมเดล (the Rasch model)

Brennan (2006) ได้จำแนกวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ไว้ 6 วิธี ได้แก่

1. วิธีความแตกต่างในค่าพารามิเตอร์ (difference in IRT parameters)

2. วิธีความแตกต่างในค่าฟังก์ชันการตอบสนองข้อสอบ (difference in item response functions)

2.1 วิธีความแตกต่างของตัวแปรพหุในค่าพารามิเตอร์ (multivariate difference in parameters)

2.2 วิธีวัดพื้นที่ความแตกต่าง (area between IRFs)

2.3 วิธีอัตราส่วนไลค์ลิฮูด (likelihood ratio test)

3. วิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel statistic)

4. วิธีการถดถอยโลจิสติก (logistic regression)

5. วิธีดัชนีมาตรฐาน (standardized difference)

6. วิธีซิปเทสท์ (SIBTEST)

วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ดังกล่าวข้างต้น สามารถจำแนกตามลักษณะการตรวจให้คะแนนได้เป็นสองประเภท คือ ข้อสอบที่มีการให้คะแนนแบบทวิวิภาค หรือสองค่า (dichotomous) และข้อสอบที่มีการให้คะแนนแบบพหุวิภาค หรือหลายค่า (polytomous) นอกจากนี้ยังจำแนกได้อีกสองมิติ ได้แก่ มิติลักษณะของตัวแปรเกณฑ์ แบ่งเป็นกลุ่มวิธีที่ใช้คะแนนสังเกตได้ (observed score) และกลุ่มวิธีที่ใช้คะแนนของตัวแปรแฝง (latent variable) และมิติลักษณะของสถิติวิเคราะห์ แบ่งเป็นกลุ่มวิธีที่ใช้สถิติพาราเมตริก (parametric approach) และกลุ่มวิธีที่ใช้สถิตินันพาราเมตริก (nonparametric approach) สรุปได้ดังตารางที่ 1 (Potenza and Dorans, 1995; Feinstein, 1995 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555)

ตารางที่ 1 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่มีการตรวจให้คะแนนแบบทวิวิภาคและพหุวิภาค

ประเภทและตัวแปรเกณฑ์	พาราเมตริก	นันพาราเมตริก
1. DIF แบบทวิวิภาค		
1.1 คะแนนสังเกตได้	ANOVA	TID
	Logistic Regression	MH
		STND
1.2 ตัวแปรแฝง	IRT-D^2	SIBTEST
	Lord's χ^2	
	General IRTLR	
	Loglinear IRTLR	
2. DIF แบบพหุวิภาค		
2.1 คะแนนสังเกตได้	ANOVA	Polytomous STND
	Polytomous Logistic Regression	GMH
2.2 ตัวแปรแฝง	General IRTLR	Polytomous
		SIBTEST
	PCM	GPCM

วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบทวิวิภาค ในกลุ่มวิธีที่ใช้คะแนนสังเกตได้หรือวิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิม และกลุ่มวิธีที่ใช้คะแนนของตัวแปรแฝงหรือวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบที่สำคัญมีดังนี้

2.1 วิธีการถดถอยโลจิสติก (logistic regression; LR)

วิธีการถดถอยโลจิสติก เป็นวิธีที่ Swaminathan and Rogers (1990) ได้พัฒนามาจากวิธีลอก-ลิเนียร์ของ Mellenberg (1982 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555) และวิธีแมนเทล-แฮนส์เซลของ Holland และ Thayer (1988 อ้างถึงใน วลีมาศ แซ่อึ้ง, 2543) วิธีนี้อยู่บนพื้นฐานของโมเดลการวิเคราะห์สมการถดถอยโลจิสติกซึ่งเป็นโมเดลที่มีพื้นฐานเป็นแบบจำลองที่สามารถเพิ่มตัวแปรความสามารถและปฏิสัมพันธ์เข้าไปในสมการได้ และสอดคล้องกับธรรมชาติของการทำหน้าที่ต่างกันของแบบสอบได้ดี จึงสามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและแบบอเนกรูปได้

โมเดลการถดถอยโลจิสติกเป็นโมเดลที่มีความยืดหยุ่นและใช้ได้ง่ายในการวิเคราะห์สมการถดถอยโลจิสติกหรือเรียกว่าการวิเคราะห์โลจิท (logit analysis) เป็นการวิเคราะห์สมการทำนาย เมื่อ

ต้องการศึกษาผลของตัวแปรทำนายที่มีต่อตัวแปรเกณฑ์ ซึ่งมีลักษณะเป็นทวิวิภาค (dichotomous variable) โดยใช้ฟังก์ชันโลจิสติก (logistic function) ในการแสดงความสัมพันธ์ระหว่างค่าของตัวแปรทำนายกับค่าความน่าจะเป็นของการเกิดเหตุการณ์ตามตัวแปรเกณฑ์ (ศิริชัย กาญจนวาสี, 2555)

ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ โดยแบ่งสมการสำหรับผู้สอบออกเป็นสองกลุ่ม สมการมาตรฐานของโมเดลการถดถอยโลจิสติก สำหรับคำนวณโอกาสของการตอบข้อสอบถูกเป็นดังนี้

$$P(U_{ij}=1|\theta_{ij}) = \frac{e^{(\beta_{0j}+\beta_{1j}\theta_{ij})}}{1+e^{(\beta_{0j}+\beta_{1j}\theta_{ij})}} \; ; \; i = 1,2,\ldots,n : j = 1,2$$

เมื่อ U_{ij} = โอกาสของการตอบข้อสอบถูกของคนที่ i ในกลุ่ม j

θ_{ij} = ความสามารถของผู้สอบที่ i ในกลุ่ม j

β_{0j} = พารามิเตอร์จุดตัด (intercept parameter)

β_{1j} = พารามิเตอร์ความชัน (slope parameter)

จากโมเดลดังกล่าว

ถ้า $\beta_{01} = \beta_{02}$ และ $\beta_{11} = \beta_{12}$ เป็นลักษณะโค้งถดถอยโลจิสติก สำหรับผู้สอบสองกลุ่มเหมือนกัน แสดงว่า ข้อสอบทำหน้าที่ไม่ต่างกัน (no DIF)

ถ้า $\beta_{01} \neq \beta_{02}$ แต่ $\beta_{11} = \beta_{12}$ เป็นลักษณะโค้งถดถอยโลจิสติก สำหรับผู้สอบสองกลุ่มเท่าเทียมกันหรือขนานกัน แสดงว่า ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป (uniform DIF)

ถ้า $\beta_{01} = \beta_{02}$ แต่ $\beta_{11} \neq \beta_{12}$ เป็นลักษณะโค้งถดถอยโลจิสติก สำหรับผู้สอบสองกลุ่มไม่เท่าเทียมกันหรือไม่ขนานกัน แสดงว่า ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป (nonuniform DIF)

จากสมการถดถอยโลจิสติกดังกล่าว สามารถเขียนเป็นสมการใหม่ในรูปการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบได้ดังนี้

$$P(U_{ij}=1|\theta_{ij}) = \frac{e^{Z_{ij}}}{1+e^{Z_{ij}}}$$

โดยที่ $log\left[\frac{P}{1-P}\right] = Z_{ij} = \tau_0 + \tau_1\theta_{ij} + \tau_2 G_j + \tau_3\theta_{ij}G_j$

เมื่อ $P(U_{ij}=1|\theta_{ij})$ = โอกาสในการตอบข้อสอบถูกของผู้สอบที่ i ในกลุ่ม j ซึ่งมีความสามารถ θ

θ_{ij} = ความสามารถของผู้สอบคนที่ i ในกลุ่ม j

G_j = กลุ่มผู้สอบ (G=1 สมาชิกกลุ่ม 1 หรือกลุ่มเปรียบเทียบ, G=2 สมาชิกกลุ่ม 2 หรือกลุ่มอ้างอิง)

$\theta_{ij}G_j$ = ปฏิสัมพันธ์ของตัวแปรอิสระ *2* ตัว คือ θ_{ij} กับ G_j

τ_0 = พารามิเตอร์จุดตัด

τ_1 = ผลของความสามารถของผู้สอบ

τ_2 = ความแตกต่างระหว่างกลุ่มในการตอบข้อสอบถูก

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

τ_3 = ปฏิสัมพันธ์ระหว่างกลุ่มกับความสามารถผู้สอบ

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

ในการตัดสินว่าข้อสอบทำหน้าที่ต่างกันแบบเอกรูปหรืออเนกรูป พิจารณาจากพารามิเตอร์ τ_2 และ τ_3 ดังนี้

ถ้า $\tau_2 \neq 0$ และ $\tau_3 = 0$ แสดงว่า ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป

ถ้า $\tau_3 \neq 0$ และ $\tau_2 = 0$ หรือไม่ก็ได้ แสดงว่า ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป

สำหรับการประมาณค่าพารามิเตอร์ตามโมเดลโลจิสติก ใช้วิธีประมาณค่าด้วยวิธีไลค์ลิฮูดสูงสุด (maximum likelihood estimation; MLE) ซึ่งเขียนในรูปฟังก์ชันได้ดังนี้

$$L(U_{ij}|\theta) = \prod_{i=1}^{n}\prod_{j=1}^{k} P\left(U_{ij}\right)^{u_{ij}}\left[1-P(U_{ij})\right]^{1-U_{ij}}$$

โดยที่ k และ n แทนขนาดกลุ่มตัวอย่างและความยาวของแบบสอบตามลำดับ สำหรับค่าประมาณของพารามิเตอร์โดยใช้ไลค์ลิฮูดสูงสุดมีการแจกแจงแบบปกติของตัวแปรพหุในรูปเชิงเส้นกำกับ (asymptotically multivariate normal) ซึ่งมีค่าเฉลี่ยของเวคเตอร์ τ และ เมทริกซ์ความแปรปรวน ความแปรปรวนร่วมในรูป Σ ในขณะที่ Σ^{-1} เป็นเมทริกซ์สารสนเทศกำหนด ดังนี้

$$\Sigma^{-1} = -\mathrm{E}\left[\frac{\delta^2}{\delta_{\tau_r}\delta_{\tau_s}}\ln \mathrm{L}\right] \;;\; r,s = 0,1,2,3$$

เมื่อ E และ ln L แทนค่าความคาดหมายของเมทริกซ์และลอการิทึมของฟังก์ชันไลค์ลิฮูด ตามลำดับ ดังนั้นการแจกแจงของการประมาณค่าพารามิเตอร์ด้วยวิธี MLE จะอยู่ในรูป ดังนี้

$$\tau \sim N(\tau, \Sigma)$$

โดยที่ $\tau = [\tau_0, \ \tau_1, \ \tau_2, \ \tau_3]$ ส่วนความคลาดเคลื่อนมาตรฐานเชิงเส้นกำกับของค่าประมาณของ τ_s (S = 0, 1, 2, 3) เมื่อ S เป็นสมาชิกแนวเส้นทแยงมุมของ Σ สามารถคำนวณได้จากสูตรดังนี้

$$SE(\hat{\tau}_s) = \sqrt{\Sigma^{ss}}$$

ในการทดสอบสมมติฐานของการทำหน้าที่ต่างกันของข้อสอบจะทดสอบสมาชิกของ τ_s ซึ่งสมมติฐานที่สนใจ คือ $H_0 : \tau_2 = 0$ และ $H_0 : \tau_3 = 0$ สมมติฐานทั้งสองสามารถทดสอบพร้อม ๆ กัน ดังนี้

$$H_0 : C\tau = 0$$

$$H_1 : C\tau \neq 0$$

โดยที่ C เป็นเมทริกซ์ขนาด 2 x 4 ดังนี้

$$C = \begin{vmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix}$$

ส่วนการทดสอบนัยสำคัญของสมมติฐานจะใช้สถิติไค-สแควร์ ที่ระดับชั้นความเป็นอิสระเท่ากับ 2 (df = 2) ดังนี้

$$\chi^2 = \hat{\tau}' C' (C \Sigma C')^{-1} C \hat{\tau}'$$

ถ้า χ^2 มีค่ามากกว่า $\chi^2_{(\alpha, \ 2)}$ แสดงว่า ปฏิเสธสมมติฐานศูนย์ นั่นคือ ข้อสอบทำหน้าที่ต่างกัน (DIF) นั่นเอง

2.2 วิธีซิปเทสท์ (Simultaneous Item Bias Test; SIBTEST)

วิธีซิปเทสท์ พัฒนาโดย Shealy และ Stout (1993 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555) เพื่อใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) การทำหน้าที่ต่างกันของแบบสอบ (DTF) และการทำหน้าที่ต่างกันของกลุ่มข้อสอบ (Differential Bundle Functioning; DBF) วิธีนี้สามารถวิเคราะห์ได้ทั้งในแบบสอบเอกมิติและแบบสอบพหุมิติ (Stout, Li and Nandakumar, 1997 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555) วิธีซิปเทสท์ใช้สถิติทดสอบแบบนันพาราเมตริก ซึ่งพัฒนามาบนพื้นฐานของทฤษฎี IRT ชนิดพหุมิติ แต่ไม่ต้องใช้ฟังก์ชันการตอบสนองข้อสอบหรือการประมาณค่าความสามารถแฝง วิธีซิปเทสท์ถูกออกแบบมาสำหรับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบมีทิศทางเดียว (unidirectional DIF) ดังนั้นจึงไม่มีความไวในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบไม่มีทิศทาง (nondirectional DIF) (Li and Stout, 1996 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555) จุดเด่นของวิธีซิปเทสท์ คือ สามารถคำนวณได้ง่ายไม่ซับซ้อน ประหยัดค่าใช้จ่ายและไม่จำเป็นต้องใช้กลุ่มตัวอย่างที่มีขนาดใหญ่ ทั้งยังใช้สถิติทดสอบนัยสำคัญ (Narayanan and Swaminathan, 1996) นอกจากนี้ ยังสามารถนำไปประยุกต์ใช้กับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบพหุวิภาค (polytomous DIF) (Chang, Mazzeo and Roussos, 1995 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555)

ในการศึกษาการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีซิปเทสท์ของแบบสอบเอกมิติ ถือว่าข้อสอบในแบบสอบจะต้องมุ่งวัดคุณลักษณะหรือความสามารถแฝงเพียงลักษณะเดียว ความสามารถแฝงสามารถจำแนกเป็น ความสามารถเป้าหมายที่ต้องการวัด (taget ability; θ) กับ ความสามารถแทรกซ้อนที่ไม่ใช้เป้าหมายของการวัด (nuisance ability; η) ตัวอย่างเช่น แบบสอบคำศัพท์ในวิชาภาษาอังกฤษ ข้อสอบบางข้ออาจถามความรู้สำหรับผู้ชายเป็นพิเศษ เช่น ความรู้เกี่ยวกับกีฬา ในขณะที่ข้อสอบบางข้ออาจถามความรู้เกี่ยวกับผู้หญิงโดยเฉพาะ เช่น ความรู้เกี่ยวกับงานในบ้าน จากสถานการณ์ดังกล่าว ทักษะความรู้เกี่ยวกับคำศัพท์ในวิชาภาษาอังกฤษ เป็นความสามารถเป้าหมายที่ต้องการวัด (θ) ส่วนความรู้ทางด้านกีฬาและงานในบ้านเป็นความสามารถแทรกซ้อน ที่ไม่ใช่เป้าหมายของการวัด (η_1 และ η_2) ข้อสอบทุกข้อในแบบสอบจะวัดความสามารถเป้าหมาย ส่วนข้อสอบบางข้อทำหน้าที่ต่างกันจะวัดทั้งความสามารถเป้าหมายและความสามารถแทรกซ้อน (Nandakumar, 1993 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555)

ความสามารถ θ และ η ได้มาจากผลการตอบข้อสอบระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ส่วนผลการตอบข้อสอบได้มาจากการสุ่มจากกลุ่มผู้ตอบในแต่ละกลุ่ม ซึ่งแทนด้วย $U = (U_1, U_2, ..., U_N)$ โดยที่ U_i เท่ากับ 1 ถ้าตอบข้อสอบข้อที่ i ถูก และ U_i เท่ากับ 0 ถ้าตอบข้อสอบผิด ส่วน N เป็นจำนวนข้อสอบ ในการสร้าง U ตามโมเดล โดยทั่ว ๆ ไป จะต้องประกอบด้วย

2 องค์ประกอบ คือ (1) พารามิเตอร์ความสามารถของผู้สอบจำนวน d มิติ และ (2) ฟังก์ชันการตอบสนองข้อสอบ (IRFs) ซึ่งฟังก์ชันการตอบสนองข้อสอบในแต่ละข้อจะเป็นตัวกำหนดความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง สำหรับในที่นี้ d = 2 เพราะมีพารามิเตอร์ความสามารถสองชนิด คือ ความสามารถเป้าหมายที่ต้องการวัด (θ) และความสามารถแทรกซ้อนที่ไม่ต้องการวัด (η) ส่วนเวคเตอร์ความสามารถที่กำหนดจากผู้สอบในแต่ละกลุ่มแทนด้วย (θ,η) โดยที่การแจกแจงของ (θ,η) ถูกเลือกขึ้นมาอย่างสุ่มจากกลุ่มผู้ตอบ และความสามารถที่สุ่มมาจากกลุ่มผู้สอบแทนด้วย (Θ,η) โดยสมมติว่าข้อสอบทุกข้อของแบบสอบจะวัดความสามารถ θ ส่วนข้อสอบบางข้อที่ทำหน้าที่ต่างกันจะวัดทั้งความสามารถ θ และ η ซึ่งความสามารถ η อาจมีเพียงความสามารถเดียวหรือมากกว่าหนึ่งความสามารถก็ได้ สำหรับ IRF ข้อที่ i ขึ้นอยู่กับความสามารถ θ เพียงอย่างเดียว แทนด้วย $P_i(\theta)$ ส่วน IRF ข้อที่ i ซึ่งขึ้นอยู่กับความสามารถ θ และ η แทนด้วย $P_i(\theta,\eta)$ ดังนี้ (Shealy and Stout, 1993 อ้างถึงใน วลีมาศ แซ่อึ้ง, 2543)

$$P_i(\theta) = c_i + \frac{(1-c_i)}{1+exp[-1.7a_{i\theta}(\theta-b_{i\theta})]} \quad , i = 1, \dots, N$$

$$P_i(\theta,\eta) = c_i + \frac{(1-c_i)}{1+exp\{-1.7[a_{i\theta}(\theta-b_{i\theta})+a_{i\eta}(\eta-b_{i\eta})]\}} \quad , i = n+1, \dots, N$$

ค่า IRF ทั้ง 2 ลักษณะดังกล่าวเป็นโมเดลการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ ซึ่งมีคุณสมบัติไม่แปรเปลี่ยน (Invariance) ไปตามกลุ่มผู้สอบ ทั้งยังมีข้อตกลงเกี่ยวกับความเป็นอิสระต่อกันในการตอบข้อสอบ (local independence) ของ U เมื่อให้ IRF แทนด้วย $P_i(\theta,\eta)$ สามารถกำหนดในรูปของความน่าจะเป็น ดังนี้ (Li and Stout, 1996 อ้างถึงใน วลีมาศ แซ่อึ้ง, 2543)

$$P[U|(\Theta=\theta,\eta=\eta)] = \prod_{i=1}^{N} P_i(\theta,\eta)^{u_i}[1-P_i(\theta,\eta)]^{1-u_i}$$

Shealy และ Stout (1993 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555) ได้ใช้ marginal IRFs อธิบายการทำหน้าที่ต่างกันของข้อสอบ ดังนี้

$$M_{ig}(\theta) = \int_{\eta} P_i(\theta,\eta) f_g(\eta|\theta)^{d\eta}$$

เมื่อ $M_{ig}(\theta)$ = marginal IRF สำหรับความสามารถเป้าหมายที่ต้องการวัด θ ของผู้สอบกลุ่มอ้างอิงหรือกลุ่มเปรียบเทียบ

$P_i(\theta, \eta)$ = IRF ของข้อสอบข้อที่ i

$f_g(\eta|\theta)$ = การแจกแจงแบบมีเงื่อนไขของกลุ่มผู้สอบ

การเปรียบเทียบ marginal IRF ระหว่างกลุ่มอ้างอิง (R) กับกลุ่มเปรียบเทียบ (F) จะทำให้ทราบถึงทิศทางของการได้เปรียบหรือเสียเปรียบ กล่าวคือ ถ้า $M_{iF}(\theta) < M_{iR}(\theta)$ ทุกค่าของ θ แสดงว่าข้อสอบทำหน้าที่ต่างกันแบบมีทิศทางเดียว โดยข้อสอบจะเข้าข้างผู้สอบกลุ่มอ้างอิง และถ้า $M_{iF}(\theta) > M_{iR}(\theta)$ ทุกค่าของ θ แสดงว่าข้อสอบทำหน้าที่ต่างกันแบบมีทิศทางเดียว โดยข้อสอบจะเข้าข้างผู้สอบกลุ่มเปรียบเทียบ การทำหน้าที่ต่างกันของข้อสอบแบบมีทิศทางเดียวอาจเรียกอีกอย่างหนึ่งว่า “การทำหน้าที่ต่างกันแบบไม่ตัดกัน” (Noncrossing DIF)

ในการเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบตามวิธีซิปเทสท์ จะแบ่งข้อสอบออกเป็น 2 ชุดย่อย (subtests) คือ (1) ชุดแบบสอบที่มีความตรง (Valid subtests) หรือ ชุดแบบสอบที่ใช้ในการจับคู่เปรียบเทียบ (Matching subtests) แบบสอบชุดนี้ประกอบด้วยข้อสอบที่ไม่ทำหน้าที่ต่างกัน และ (2) ชุดแบบสอบที่ต้องการศึกษา (Studied subtests) ประกอบด้วยข้อสอบที่สงสัยว่าทำหน้าที่ต่างกัน ถ้าแบบสอบชุดแรกมีจำนวน n ข้อ (ข้อที่ 1 ถึง n) แล้ว แบบสอบชุดที่สองจะมีจำนวน N-n ข้อ (ข้อที่ n + 1 ถึง N) เมื่อ N เป็นจำนวนข้อสอบทั้งหมด

ฟังก์ชันการตอบสนองข้อสอบของแบบสอบที่ต้องการศึกษา จากผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ กำหนดในรูปฟังก์ชัน marginal ดังนี้

$$M_{SR}(\theta) = \sum_{i=n+1}^{N} M_{iR}(\theta)$$

$$M_{SF}(\theta) = \sum_{i=n+1}^{N} M_{iF}(\theta)$$

เมื่อ $M_{SR}(\theta)$ = ผลรวมของ marginal IRFs ของข้อสอบที่ต้องการศึกษาจากผู้สอบกลุ่มอ้างอิง ณ ระดับสามารถ θ

$M_{SF}(\theta)$ = ผลรวมของ marginal IRFs ของข้อสอบที่ต้องการศึกษาจากผู้สอบกลุ่มเปรียบเทียบ ณ ระดับสามารถ θ

n = จำนวนข้อสอบในชุดแบบสอบที่มีความตรง

N = จำนวนข้อสอบทั้งหมด

ขนาดของความแตกต่างระหว่าง $M_{SR}(\theta)$ และ $M_{SF}(\theta)$ แสดงถึงปริมาณความเข้มของการทดสอบ การทำหน้าที่ต่างกันของข้อสอบแบบมีทิศทางเดียว หรือการทำหน้าที่ต่างกันแบบไม่ตัดกันจากชุดแบบสอบที่ต้องการศึกษา ณ ระดับความสามารถ θ ซึ่งสามารถคำนวณตลอดช่วงความสามารถ θ ของผู้สอบ ด้วยการอินทิเกรท ดังนี้

$$\beta_{uni} = \int_{\theta}[M_{SR}(\theta) - M_{SF}(\theta)]f_p(\theta)d\theta$$

เมื่อ β_{uni} = ดัชนีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบมีทิศทางเดียว

$f_p(\theta)$ = ฟังก์ชันความหนาแน่นของโอกาสการแจกแจงความสามารถ θ ของผู้สอบทั้ง 2 กลุ่ม

ดัชนี β_{uni} ที่คำนวณได้จากสูตรดังกล่าว นำมาทดสอบสมมติฐานของการทำหน้าที่ต่างกันของข้อสอบแบบมีทิศทางเดียว ดังนี้

H_0 : $\beta_{uni} = 0$

H_1 : $\beta_{uni} > 0$

สมมติฐานอื่น (H_1) มีลักษณะทิศทางเดียว ซึ่งใช้ทดสอบการทำหน้าที่ต่างกันของข้อสอบที่เข้าข้างผู้สอบกลุ่มอ้างอิง สำหรับค่าประมาณของ β_{uni} คำนวณได้จากคะแนนรวมของชุดแบบสอบที่มีความตรงและชุดแบบสอบที่ต้องการศึกษา ซึ่งกำหนดด้วยสัญลักษณ์ ดังนี้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

$$X = \sum_{i=1}^{n} U_i$$

$$Y = \sum_{i=n+1}^{N} U_i$$

เมื่อ X = คะแนนรวมของชุดแบบสอบที่มีความตรง

Y = คะแนนรวมของชุดแบบสอบที่ต้องการศึกษา

U_i = ผลการตอบข้อสอบข้อที่ i (ตอบถูกได้ 1 คะแนน และตอบผิดได้ 0 คะแนน)

นำคะแนนรวมของชุดแบบสอบที่มีความตรง (X) เป็นเกณฑ์ในการจับคู่ผู้ที่มีความสามารถระดับเดียวกัน แล้วคำนวณคะแนนเฉลี่ยรายข้อจากผลการตอบข้อสอบชุดแบบสอบที่ต้องการศึกษาระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีความสามารถระดับเดียวกันมาจับคู่เปรียบเทียบกัน ณ X = k โดยเขียนในรูปสัญลักษณ์ได้ ดังนี้

$$\bar{Y}_{Rk} - \bar{Y}_{Fk} \qquad ; k = 0, 1, 2, \ldots, n$$

เมื่อ $\bar{Y}_{Rk}$ = ค่าเฉลี่ยของคะแนนรายข้อ จากชุดแบบสอบที่ต้องการศึกษาของผู้สอบกลุ่มอ้างอิง ซึ่งได้คะแนน X=k

$\bar{Y}_{Fk}$ = ค่าเฉลี่ยของคะแนนรายข้อ จากชุดแบบสอบที่ต้องการศึกษาของผู้สอบกลุ่มเปรียบเทียบ ซึ่งได้คะแนน X=k

k = คะแนนรวมจากชุดแบบสอบที่มีความตรง

ค่า $\bar{Y}_{Rk} - \bar{Y}_{Fk}$ ดังกล่าวเป็นความแตกต่างของผลการตอบข้อสอบในชุดแบบสอบที่ต้องการศึกษาระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีความสามารถระดับเดียวกัน ถ้า $\bar{Y}_{Rk} - \bar{Y}_{Fk}$ = 0 ทุกคะแนน k แสดงว่า ข้อสอบที่ต้องการศึกษาทำหน้าที่ไม่ต่างกัน แต่ถ้า $\bar{Y}_{Rk} - \bar{Y}_{Fk}$ > 0 ทุกคะแนน k แสดงว่า ข้อสอบทำหน้าที่ต้างกันแบบมีทิศทางเดียง โดยจะเข้าข้างผู้สอบกลุ่มอ้างอิง ค่าความแตกต่างของผลการตอบข้อสอบสามารถประมาณค่าในรูป β_{uni} ได้ดังนี้

$$\hat{\beta}_{uni} = \sum_{k=0}^{n} \hat{P}_k(\bar{Y}_{Rk} - \bar{Y}_{Fk})$$

โดยที่ $$\hat{P}_k = \frac{(J_{Rk}+J_{Fk})}{\sum_{k=0}^{n}(J_{Rk}+J_{Fk})}$$

เมื่อ P_k = สัดส่วนของจำนวนผู้สอบกลุ่มเปรียบเทียบและกลุ่มอ้างอิงที่ได้คะแนนรวม X = k จากจำนวนผู้สอบทั้งหมด

J_{Fk} = จำนวนผู้สอบกลุ่มเปรียบเทียบที่ได้คะแนนรวม X = k

J_{Rk} = จำนวนผู้สอบกลุ่มอ้างอิงที่ได้คะแนนรวม X = k

สำหรับการทดสอบสมมติฐานศูนย์ของ no DIF ใช้สถิติ B_{uni} ดังนี้

$$B_{uni} = \frac{\hat{\beta}_{uni}}{\hat{\sigma}(\hat{\beta}_{uni})}$$

โดยที่

$$\hat{\sigma}(\hat{\beta}_{uni}) = \sqrt{\sum_{k=0}^{n} \hat{P}_k^2 \left[\frac{1}{J_{Rk}}\hat{\sigma}^2(Y|k,R) + \frac{1}{J_{Fk}}\hat{\sigma}^2(Y|k,F)\right]}$$

เมื่อ $\hat{\sigma}(\hat{\beta}_{uni})$ = ค่าประมาณความคลาดเคลื่อนมาตรฐานของ β_{uni}

$\hat{\sigma}^2(Y|k,g)$ = ค่าประมาณความแปรปรวนของคะแนนจากชุดแบบสอบที่ต้องการศึกษา สำหรับผู้สอบกลุ่ม $\mathbf{g}$ (R หรือ F) ซึ่งมีคะแนนรวมเท่ากับ k

J_{gk} = จำนวนผู้สอบกลุ่ม $\mathbf{g}$ (R หรือ F) ซึ่งตอบชุดแบบสอบที่มีความตรงแล้วได้คะแนนรวม X = k

B_{uni} มีการแจกแจงในลักษณะปกติมาตรฐาน [N(0, 1)] จึงสามารถทดสอบนัยสำคัญด้วยสถิติทดสอบ Z ถ้าผลการทดสอบปรากฏว่า $B_{uni} > Z_{\alpha}$ แสดงว่า การทดสอบมีนัยสำคัญ จึงปฏิเสธ H0 นั่นคือ ข้อสอบที่นำมาตรวจสอบทำหน้าที่ต่างกัน โดยจะเข้าข้างผู้สอบกลุ่มอ้างอิง เมื่อ B_{uni} มีค่าเป็นบวก และจะเข้าข้างผู้สอบกลุ่มเปรียบเทียบเมื่อ B_{uni} มีค่าเป็นลบ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2.3 วิธีราสช์ทรี (Raschtree)

วิธีการตรวจสอบ DIF ใน Rasch model มีหลากหลายวิธี เช่น วิธีทดสอบ Wald (the item-specific Wald test) ใช้ตรวจสอบ DIF ในข้อสอบเป็นรายข้อ (Fischer & Molenaar, 1995 cited in Strobl et al., 2015) หรือวิธีทดสอบอัตราส่วนไลค์ลิฮูด (the likelihood ratio test) ให้ค่าความพอเหมาะของแบบสอบที่สะท้อนถึง DIF (Andersen, 1972; Gustafsson, 1980 cited in Strobl et al., 2015) ซึ่งวิธีเหล่านี้อยู่บนพื้นฐานการเปรียบเทียบค่าพารามิเตอร์ของข้อสอบที่ประมาณค่าระหว่างกลุ่มผู้สอบสองกลุ่มหรือมากกว่าที่ได้กำหนดไว้ล่วงหน้า เช่น กลุ่มผู้ชายเป็นกลุ่มเปรียบเทียบ และกลุ่มผู้หญิงเป็นกลุ่มอ้างอิง (Strobl et al., 2015) วิธีเหล่านี้เรียกว่า แนวทางที่กำหนดกลุ่มไว้ล่วงหน้า

ข้อดีของโมเดลการทดสอบที่แสดงผลเป็นกลุ่ม คือ ถ้าตรวจสอบพบ DIF ผลลัพธ์ดังกล่าวสามารถอธิบายได้อย่างชัดเจนว่าข้อสอบง่ายหรือยากสำหรับผู้สอบกลุ่มใด เพื่อแก้ไขข้อสอบสำหรับ

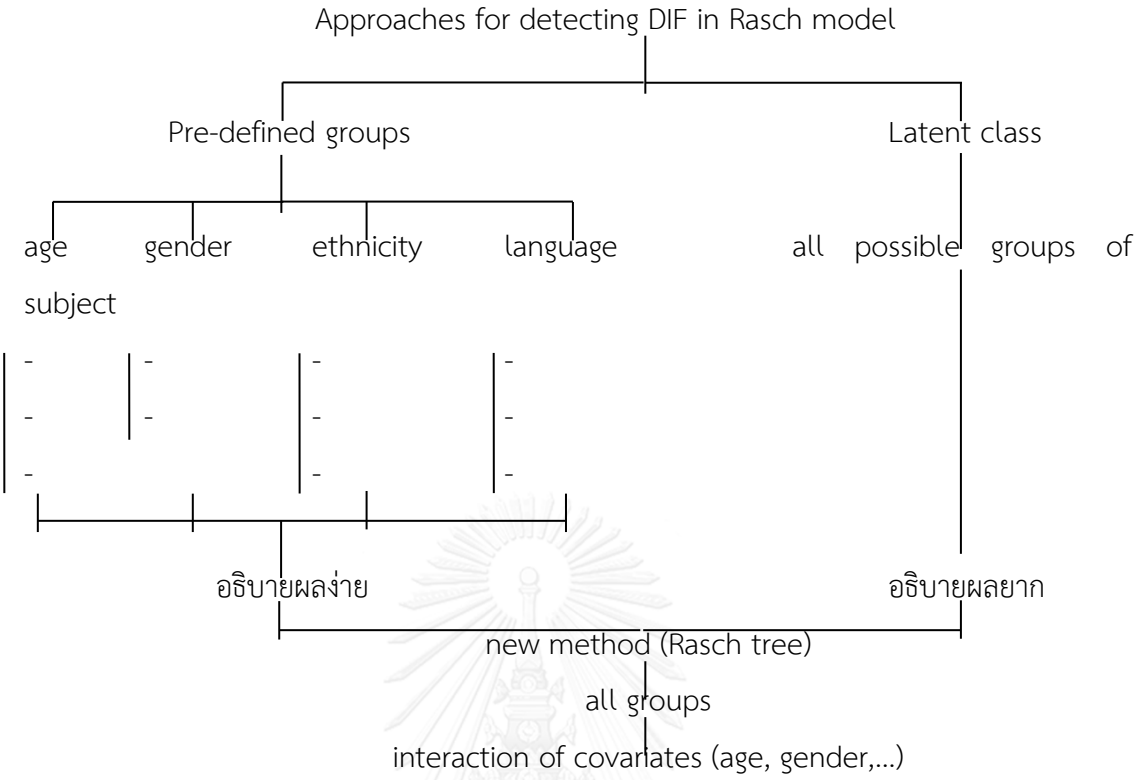
แต่ละกลุ่ม และสามารถขจัดหรือหลีกเลี่ยงรูปแบบของแบบสอบในอนาคต ส่วนข้อจำกัดของโมเดลการทดสอบนี้ คือ นักวิจัยเป็นผู้กำหนดกลุ่มที่ชัดเจนในการตรวจสอบ DIF ตัวแปรในการทดสอบ เช่น อายุ เพศ ศาสนา และภาษา ดังนั้น ถ้าการวิเคราะห์ภายหลังพบความแตกต่างระหว่างกลุ่มในตัวแปรที่ไม่ได้กำหนดไว้ก่อน จะทำให้ผลนั้นไม่สามารถอธิบายได้

ในการตรวจสอบ DIF ยังมีอีกแนวทางหนึ่ง เรียกว่า แนวทางกลุ่มแฝง (the latent class approach) ของ Rost (1990 cited in Strobl et al., 2015) นับเป็นการทดสอบใน Rasch model ที่เข้มงวด เพราะมุ่งทดสอบความแตกต่างของพารามิเตอร์ของข้อสอบระหว่างกลุ่มของผู้สอบทั้งหมด โดยไม่พิจารณาเรื่องตัวแปรร่วมของบุคคล ทำให้มีข้อจำกัดในการอธิบายผลอย่างไม่ชัดเจน

แนวทางใหม่ที่ผสมผสานระหว่างสองแนวทางที่กล่าวข้างต้น คือ การทดสอบทุกกลุ่มแบบทำซ้ำ ภายในขอบเขตของตัวแปรร่วมที่เป็นไปได้ ทำให้สามารถอธิบายผลลัพธ์ได้ชัดเจน และเป็นการตรวจสอบ DIF ได้อย่างกว้างขวาง วิธีนี้อยู่บนพื้นฐานเทคนิคโมเดลการแบ่งส่วนแบบทำซ้ำ (model-based recursive partitioning) ใช้สถิติทดสอบโครงสร้างที่เปลี่ยนแปลงจากเศรษฐมิติ (econometrics) โมเดลการแบ่งส่วนแบบทำซ้ำ เป็นแนวทางแบบเซมิพาราเมตริก มีเป้าหมายเพื่อตรวจสอบความแตกต่างในค่าพารามิเตอร์ของโมเดลทางสถิติระหว่างกลุ่มผู้สอบที่กำหนดจากการรวมกันของตัวแปรร่วม วิธีนี้ เรียกว่า ราสช์ทรี นำเสนอโดย Strobl et al. (2015)

แนวทางการตรวจสอบ DIF สามารถสรุปได้ ดังนี้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



ภาพที่ 5 แนวทางการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบใน Rasch model

ที่มา: Strobl et al. (2015)

โมเดลนี้ พื้นที่ของตัวแปรร่วมถูกแบ่งส่วนแบบทำซ้ำ เพื่อระบุกลุ่มของผู้สอบด้วยค่าของตัวแปรจัดประเภทหรือตัวแปรต่อเนื่องที่แตกต่างกัน ข้อดี คือ ค่าพารามิเตอร์เป็นโมเดลแบบพาราเมตริกมากกว่าเป็นค่าตัวแปรเดี่ยว ๆ ที่ผันแปรระหว่างกลุ่ม เช่น พารามิเตอร์ส่วนตัด (intercept) หรือพารามิเตอร์ความชัน (slope) ในโมเดลการถดถอยเชิงเส้นตรง ดั้งนั้น ค่าพารามิเตอร์ข้อสอบของ Rasch model จะผันแปรระหว่างกลุ่มของผู้สอบ

ขั้นตอนตามลำดับที่ใช้ในการสร้างราสช์ทรี มีดังนี้

1. ประมาณค่าพารามิเตอร์ของข้อสอบจากผู้สอบทั้งหมดในกลุ่มตัวอย่าง เริ่มจากกลุ่มตัวอย่างทั้งหมด

2. ประเมินความคงที่ของค่าพารามิเตอร์ของข้อสอบโดยพิจารณาจากตัวแปรร่วมแต่ละตัว

3. ถ้าพบความไม่คงที่อย่างมีนัยสำคัญ ให้ทำการแบ่งส่วนตัวอย่างตามตัวแปรร่วมที่มีค่าความไม่คงที่ที่เข้มแข็ง และจุดตัดดูที่ค่าสูงที่สุดที่ทำให้โมเดลเหมาะสม

4. ทำข้อที่ 1-3 ซ้ำ ๆ ในกลุ่มตัวอย่างย่อย จนกระทั่งค่าความไม่คงที่ไม่มีนัยสำคัญทางสถิติ

ขั้นตอนที่ 1 การประมาณค่าพารามิเตอร์ของข้อสอบ

ใช้ common conditional maximum likelihood ในการประมาณค่าพารามิเตอร์ของข้อสอบ ให้ $\theta_i, i = 1, \dots, n$ แทน ค่าพารามิเตอร์ของผู้สอบ $\beta_j, j = 1, \dots, m$ แทนค่าพารามิเตอร์ของข้อสอบ และ u_{ij} แทน ผลการตอบของผู้สอบคนที่ i ข้อสอบที่ j ภายใต้ Rasch model ดังนี้

$$P(U_{ij} = u_{ij} | \theta_i, \beta_j) = \frac{e^{u_{ij} \cdot (\theta_i - \beta_j)}}{1 + e^{\theta_i - \beta_j}}$$

คะแนนดิบของผู้สอบ r_i สร้างค่าสถิติที่เพียงพอสำหรับค่าพารามิเตอร์ของผู้สอบ ค่าพารามิเตอร์ของข้อสอบประมาณจากค่าเฉลี่ยที่ทำซ้ำ ๆ จากเงื่อนไขแบบ likelihood

สมการที่ 1

$$L_c(\beta | r_1, \dots, r_n) = \prod_{i=1}^{n} L_c(\beta | r_i) = \prod_{i=1}^{n} \frac{e^{-\sum_{j=1}^{m} u_{ji} \cdot \beta_j}}{\gamma_{r_i}(\beta)}$$

เมื่อ γ_{r_i} เป็นฟังก์ชันสมมาตร (symmetric function) ของ order r_i

ขั้นตอนที่ 2 การทดสอบความไม่คงที่ของค่าพารามิเตอร์

ในขั้นตอนการทดสอบค่าพารามิเตอร์ของข้อสอบที่ผันแปรระหว่างกลุ่มผู้สอบจากตัวแปรร่วม ใช้การทดสอบการเปลี่ยนแปลงโครงสร้างจากเศรษฐมิติ (econometrics) หลักการเดียวกันนำมาประยุกต์ใช้ตรวจสอบการเปลี่ยนแปลงในค่าพารามิเตอร์ของข้อสอบใน Rasch model บนพิสัยของตัวแปรร่วมของบุคคล ค่าพารามิเตอร์ของข้อสอบถูกประมาณค่าครั้งแรกจากกลุ่มตัวอย่างทั้งหมด หลังจากนั้นส่วนเบี่ยงเบนแต่ละค่าจากโมเดลร่วมเป็นลำดับที่คำนึงถึงตัวแปรร่วม เช่น อายุ ถ้ามี DIF ที่เป็นระบบ ที่เกี่ยวข้องกับกลุ่มเกิดจากตัวแปรร่วม ขั้นตอนจะแสดงการเปลี่ยนแปลงที่เป็นระบบของค่าพารามิเตอร์ของข้อสอบ ถ้าไม่มี DIF ค่าจะขึ้นลงเพียงเล็กน้อย

สำหรับการทดสอบการเปลี่ยนแปลงโครงสร้างทางสถิติในค่าพารามิเตอร์ ใช้การทดสอบ generalized M-fluctuation test ที่สร้างขึ้นจากกรอบโมเดลการแบ่งส่วนแบบทำซ้ำ ของ Zeileis, Hothorn, and Hornik (2008) แนวคิดของการทดสอบจะคำนวณโมเดลส่วนเบี่ยงเบน “subject-wise” และได้สถิติทดสอบจากการรู้ค่าการแจกแจง (distribution) ของส่วนเบี่ยงเบน

การวัดส่วนเบี่ยงเบนของโมเดล likelihood เมื่อ $i = 1, \dots, n$ ค่าสังเกต คือ ฟังก์ชันคะแนนรายคน $\varphi(u_i, \hat{\beta})$ ที่เกี่ยวเนื่องจากค่าลักษณะเฉพาะตัว (individual contributions) เป็น log-likelihood $\Psi(u_i, \hat{\beta})$ โดยคำนึงถึงเวคเตอร์ของพารามิเตอร์ ค่าลักษณะเฉพาะตัว คำนวณอย่างง่ายจากเงื่อนไข likelihood สำหรับ Rasch model

กระบวนการผลรวมที่สะสมนิยามได้ดังสมการที่ 2

สมการที่ 2

$$W_l(t) = \hat{V}^{-1/2} n^{-1/2} \sum_{i=1}^{\lfloor n.t \rfloor} \varphi(u_{(i|l)}, \hat{\beta}) \qquad (0 \le t \le 1)$$

เมื่อ ดัชนี $(i|l)$ แทน ค่าสังเกตลำดับที่ i โดยคำนึงถึงตัวแปรร่วมที่ l $\lfloor . \rfloor$ แทน ส่วนจำนวนเต็ม และ $\hat{V} = \sum_{i=1}^{n} \varphi(u_i, \hat{\beta}) \varphi(u_i, \hat{\beta})^T$ เป็น “outer-product-of-gradients” ประมาณค่าจาก covariance matrix ภายใต้สมมติฐานศูนย์ค่าความคงที่ของค่าพารามิเตอร์ คือ กระบวนการผลรวมที่สะสม $W_l(.)$ แสดงไปสู่การ converge ไป (m-1) –dimensional Brownian bridge ซึ่งสามารถใช้เป็นพื้นฐานสำหรับสถิติอ้างอิง

Cumulative aggregation ดำเนินการบนลำดับที่มีอิทธิพลจากตัวแปรร่วมที่ l: $i = 1, \dots, n$ ส่วนเบี่ยงเบนแต่ละค่าเป็นลำดับโดยคำนึงถึงตัวแปรร่วมและ aggregated ขึ้นอยู่กับ $\lfloor n.t \rfloor$ ส่วนประกอบในแต่ละขั้นตอน เมื่อ $W_l(t)$ พิจารณาว่าเป็นฟังก์ชันของ fraction t ของขนาดกลุ่มตัวอย่าง โมเดลศูนย์ไม่เปลี่ยนแปลงโครงสร้างต่อวิถีของกระบวนการสุ่มด้วยค่าเฉลี่ยเป็นศูนย์คงที่

ข้อดีของวิธีนี้ คือ โมเดลไม่มีการประมาณค่าซ้ำในการแบ่งทั้งหมดในตัวแปรร่วมทั้งหมด เพราะส่วนเบี่ยงเบนแต่ละค่ายังคงเหมือนกันและมีเพียงลำดับที่ต้องการปรับใหม่เพื่อประเมินตัวแปรร่วมที่แตกต่างกัน

ค่าส่วนเบี่ยงเบนที่เป็นระบบใน $W_l(.)$ อาจใช้สถิติทดสอบที่แตกต่างกัน ขึ้นอยู่กับตัวแปรร่วมที่ l เป็นตัวแปรต่อเนื่องหรือตัวแปรจัดประเภท ถ้าเป็นตัวแปรต่อเนื่อง สถิติทดสอบเป็นดังสมการที่ 3

สมการที่ 3

$$S_l = \max_{i=\underline{i},\dots,\bar{i}} \left(\frac{i}{n}.\frac{n-i}{n}\right)^{-1} \left\| W_l\left(\frac{i}{n}\right) \right\| 2.$$

การตีความหมายเช่นเดียวกับ maximum Lagrange-multiplier statistic (รู้จักในชื่อสถิติคะแนน) สำหรับตัวเลือกที่เบี่ยงเบนเดี่ยวๆ เหนือจุดตัดทั้งหมดใน $[\underline{i}, \bar{i}]$ การแจกแจงจำกัดเป็น the supremum of a tied-down Bessel process คำนวณได้จากค่า p value

ถ้าตัวแปรร่วม l เป็นตัวแปรจัดประเภท (ค่า x_{il} จากประเภท $q = 1, \dots, Q$) ใช้สถิติทดสอบดังสมการที่ 4

สมการที่ 4

$$S_l = \sum_{q=1}^{Q} n\left(\sum_{i=1}^{n} I(x_{il} = q)\right)^{-1} \left\|\Delta_q W_l\left(\frac{i}{n}\right)\right\|_{2'}^{2}$$

เมื่อ Δ_q เป็นระยะที่เพิ่มขึ้นเป็นช่วง ๆ ภายในกลุ่มที่ q สถิติทดสอบนี้ไม่แปรเปลี่ยนไปตามลำดับของกลุ่ม Q และผู้สอบภายในแต่ละกลุ่ม สถิติทดสอบค้นพบความไม่คงที่ในกลุ่มย่อย Q การแจกแจงที่จำกัดคือ χ^2 ด้วยชั้นความเป็นอิสระ (degree of freedom) ที่ (Q-1).(m-1) คำนวณจากค่า p value การทดสอบใช้ทั้งตัวแปรนามบัญญัติ (nominal) และตัวแปรเรียงอันดับ (ordinal) การตรวจสอบลำดับที่มีประสิทธิภาพจะอยู่ในขั้นตอนเมื่อได้เลือกจุดตัดแล้ว

ใน Rasch model ฟังก์ชันเป้าหมายตามเงื่อนไข log-likelihood ใช้เพื่อประมาณค่าพารามิเตอร์ ค่าลักษณะเฉพาะตามเงื่อนไข log-likelihood คำนวณอย่างง่าย เป็น log $L_c(\beta|r_i)$ ดังสมการที่ 5

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สมการที่ 5

$$\Psi(u_i, \beta) = -\sum_{j=1}^{m} u_{ij}.\beta_j - \log\left(\gamma_{ri}(\beta)\right)$$

สำหรับการคำนวณทดสอบการเปลี่ยนแปลงโครงสร้าง ค่าลักษณะเฉพาะตัวไปสู่ฟังก์ชันคะแนนได้มาจากสมการที่ 5 ลักษณะเฉพาะตัวของค่าพารามิเตอร์ของข้อสอบที่ j สำหรับผู้สอบคนที่ i เป็นดังสมการที่ 6

สมการที่ 6

$$\varphi(u_i, \beta)_j = \frac{\partial \Psi(u_i, \beta)}{\partial \beta_j} = -u_{ij} - \frac{1}{\gamma_{r_i}(\beta)}.\frac{\partial_{\gamma_{r_i}}(\beta)}{\partial \beta_j}$$

อนุพันธ์ของฟังก์ชันสมมาตร $\gamma_{r_i}(\beta)$ เป็นฟังก์ชันสมมาตรที่ทำซ้ำด้วยความแน่นอนในความหมายที่ละไว้ การนำไปใช้กับ Rasch model ขั้นตอนวิธีทั้งหมดใช้การคำนวณอนุพันธ์ของ Liou (1994 cited in Strobl, Kopf and Zeileis, 2010)

เมื่อค่าลักษณะเฉพาะตัวไปสู่ฟังก์ชันคะแนนใน Rasch model จากสมการที่ 6 เป็นลำดับที่พิจารณาจากตัวแปรร่วม l และใส่ในสมการที่ 2 ความไม่คงที่ของค่าพารามิเตอร์ในค่าพารามิเตอร์ของข้อสอบสามารถทดสอบทางสถิติโดยใช้วิธีการแบ่งส่วนแบบทำซ้ำ

ผลลัพธ์จากกระบวนการดังกล่าว อธิบายได้อย่างง่าย ๆ คือ สถิติทดสอบความไม่คงที่ของค่าพารามิเตอร์ S_l เกี่ยวข้องกับค่า p value ที่คำนวณได้ในแต่ละตัวแปร สถิติทดสอบสำหรับตัวแปรต่อเนื่อง เช่น อายุ ใช้สมการที่ 3 และตัวแปรจัดประเภท เช่น เพศ และตัวแปรจัดประเภทที่เป็นลำดับ เช่น คะแนนแรงจูงใจ ใช้สมการที่ 4 ค่า p values ได้มาจากการแจกแจงที่จำกัด

การแบ่งทำอย่างต่อเนื่องจนกระทั่งค่า p value ทั้งหมด เกินจากระดับนัยสำคัญ (ตามปกติ คือ 5%) ชี้ให้เห็นว่า ค่าความไม่คงที่ของพารามิเตอร์ไม่มีนัยสำคัญทางสถิติ หรือกระทั่งจำนวนของค่าสังเกตในกลุ่มย่อยต่ำกว่าจุดที่รับรู้

ขั้นตอนที่ 3 การเลือกจุดตัด (Selecting the cutpoints)

หลังจากเลือกตัวแปรร่วมในการแบ่งแล้ว จุดตัดจะถูกกำหนดจากค่า partitioned likelihood ที่เพิ่มให้สูงที่สุด อยู่เหนือจุดตัดอื่น ๆ ภายในพิสัยของตัวแปรนั้น ๆ คือ ผลรวมของ likelihood ของค่าสังเกตก่อนและหลังจากจุดตัด

จุดตัดนี้ได้มาโดยตรงจากข้อมูล ในขณะที่วิธีมาตรฐาน เช่น แผนภาพ หรือ การทดสอบอัตราส่วน likelihood ต้องมีการกำหนดกลุ่มเปรียบเทียบและกลุ่มอ้างอิงไว้ล่วงหน้า ใช้ค่ามัธยฐานหรือค่าเฉลี่ยเป็นจุดตัดในการแบ่งตัวอย่างไปสู่กลุ่มเปรียบเทียบหรือกลุ่มอ้างอิง ซึ่งวิธีนี้ไม่ค่อยมีหลักการและปกปิดความแตกต่างของพารามิเตอร์ที่เป็นจริงซึ่งสัมพันธ์กับจุดตัดอื่น การใช้ค่ามัธยฐานเป็นการกำหนดจุดตัดที่ไม่ค่อยมีหลักการอาจทำให้ผลลัพธ์การทดสอบไม่มีนัยสำคัญ แสดง DIF ที่เด่นชัดเกิดจากตัวแปรนั้น ๆ

วิธีราสช์ทรี ขับเคลื่อนด้วยข้อมูลสามารถตรวจสอบทั้งความมีอยู่ของค่าความไม่คงที่ของค่าพารามิเตอร์โดยคำนึงถึงตัวแปรนั้น ๆ และการเปลี่ยนแปลงค่าพารามิเตอร์ที่เกิดขึ้น

ตัวแปรในการแบ่งส่วนที่เป็นค่าต่อเนื่อง กำหนดกลุ่มย่อยเป็น $L(\xi) = \{i|x_{il} \leq \xi\}$ และ $R(\xi) = \{i|x_{il} > \xi\}$ ทางด้านซ้ายและด้านขวาของจุดตัด ξ ตามลำดับ ทั้งสองกลุ่มย่อย ค่าพารามิเตอร์ $\hat{\beta}^{(L)}$ และ $\hat{\beta}^{(R)}$ ประมาณค่าโดยแบ่งส่วนตามที่อธิบายข้างต้น

การกำหนดจุดตัดที่ดี ξ partitioned log-likelihood ที่ทำให้ค่าสูงที่สุด อยู่เหนือจุดตัดอื่น ๆ ξ (จำนวนกลุ่มตัวอย่างย่อยน้อยที่สุด) เป็นดังนี้

$$\sum_{i \in L(\xi)} \Psi\left(u_i \hat{\beta}^{(L)}\right) + \sum_{i \in R(\xi)} \Psi\left(u_i \hat{\beta}^{(R)}\right)$$

วิธีนี้ประยุกต์ใช้ได้กับตัวแปรร่วมที่เป็นค่าต่อเนื่องและตัวแปรลำดับ สำหรับตัวแปรร่วมจัดประเภทที่ไม่เป็นลำดับ ประเภท Q (Q categories) สามารถแบ่งย่อยไปสู่สองกลุ่ม การแบ่งเป็นสองส่วนจะถูกเลือกมาหนึ่งส่วนที่มีค่า partitioned likelihood สูงที่สุด

ข้อสังเกตที่สำคัญ คือ จุดตัดที่ดีจะถูกกำหนดเพียงถ้าตัวแปรสร้างค่าความไม่คงที่ของพารามิเตอร์อย่างมีนัยสำคัญ ซึ่งป้องกันความลำเอียงในการเลือกตัวแปร โดยเฉพาะการใช้การทดสอบอัตราส่วน likelihood มาตรฐาน (ใช้การแจกแจง χ^2) เพื่อประเมินนัยสำคัญของจุดตัดที่ดีเป็นสิ่งที่ไม่ถูกต้องทางสถิติ ด้วยเหตุนี้การเลือกจุดตัดที่ดีควรใช้การแจกแจง asymptotic ของสถิติอัตราส่วน likelihood ที่มีค่าสูงที่สุด มากกว่าใช้ χ^2 ในความเป็นจริงสถิติ Lagrange-multiplier จากสมการที่ 3 นำมาใช้ในวิธีราสช์ทรี เป็น asymptotic เทียบเท่ากับสถิติอัตราส่วน likelihood ที่มีค่าสูงที่สุด แต่ไม่มีการประมาณค่าซ้ำในโมเดล ดังนั้น วิธีราสช์ทรี นำเสนอกรอบแนวคิดทางสถิติที่มีเหตุมีผล (เหมาะสม) ในการตรวจสอบตัวแปรและกำหนดจุดตัดโดยอัตโนมัติ ทำให้ DIF มีความเข้มแข็ง

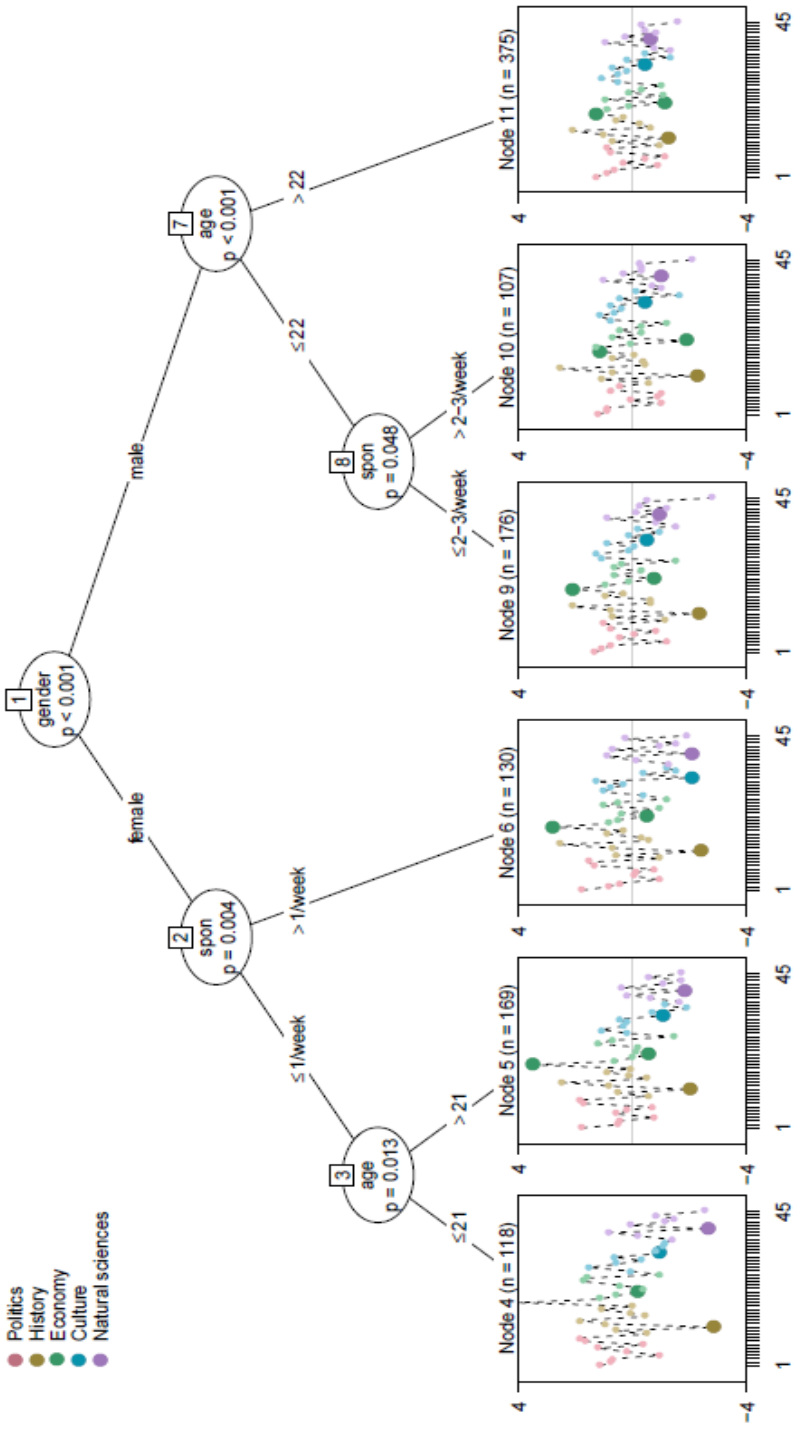
ขั้นตอนที่ 4 เกณฑ์ในการหยุดทำงาน (Stopping criteria)

เกณฑ์ในการหยุดทำงานมี 2 ชนิด ได้แก่ 1) แบ่งอย่างต่อเนื่อง กรณีที่ค่าความไม่คงที่ของพารามิเตอร์มีนัยสำคัญ ถ้าค่าความไม่คงที่ไม่มีนัยสำคัญโดยพิจารณาจากตัวแปรร่วมใด ๆ การแบ่งจะหยุดทำงาน ดังนั้น ตามปกติระดับนัยสำคัญจะกำหนดที่ 5% เป็นเกณฑ์ในการหยุดทำงานที่สำคัญ 2) เกณฑ์ในการหยุดทำงานเมื่อขนาดตัวอย่างมีค่าน้อยที่สุดในแต่ละโหนด ขนาดโหนดที่น้อยที่สุดจะถูกเลือกเป็นพื้นฐานที่เพียงพอในการประมาณค่าพารามิเตอร์ในแต่ละกลุ่มย่อย และจะเพิ่มขึ้นเมื่อจำนวนพารามิเตอร์ของข้อสอบที่จะประมาณค่าใหญ่ขึ้น ตัวอย่างเช่น ที่ระดับนัยสำคัญ 5% ขนาดโหนดที่น้อยที่สุดที่ใช้ คือ 20 คน

ข้อควรพิจารณา คือ เมื่อจำนวนตัวแปรร่วมมีขนาดใหญ่ในชุดของข้อมูล และตัวแปรร่วมทั้งหมดนั้นถูกทดสอบ DIF การทดสอบที่หลากหลายกลายเป็นประเด็นขึ้นมา เช่นเดียวกับสถิติที่ใช้ทดสอบ DIF การตรวจสอบความเป็นจริงว่า การทดสอบที่หลากหลายจะนำไปสู่อัตราการเกิดความคลาดเคลื่อนประเภทที่ 1 (false-positive) ที่เพิ่มขึ้น เมื่อจำนวนตัวแปรร่วมใหญ่ขึ้น ใช้ค่าปรับแก้

Bonferroni สำหรับค่า p value ที่เป็นเกณฑ์ในการแบ่ง ยิ่งไปกว่านั้น วิธีการแบ่งส่วนแบบทำซ้ำสร้างกระบวนการทดสอบแบบปิด ดังนั้นระดับนัยสำคัญมีไว้เพื่อต้นไม้ทั้งหมด ไม่เพียงแค่แต่ละส่วนที่แบ่ง สิ่งนี้ยืนยันว่า ตรวจสอบ DIF อย่างไม่คลาดเคลื่อน

ภาพที่ 6 ตัวอย่างผลการตรวจสอบการทำหน้าที่ต่างกันด้วยวิธีราสช์ทรี



ที่มา: Strobl, Kopf, and Zeileis (2010)

2.4 วิธีการวัดพื้นที่ของราชู (Raju)

Raju (1990) ได้เสนอสูตรการคำนวณพื้นที่ชนิดคิดเครื่องหมายในช่วงเปิด (open-interval signed area or exact signed area; ESA) และพื้นที่ชนิดไม่คิดเครื่องหมายในช่วงเปิด (open-interval unsigned area or exact unsigned area; EUA) เพื่อใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบทวิวิภาค ภายใต้ทฤษฎี IRT โมเดลแบบ 1, 2 และ 3 พารามิเตอร์ สูตรทั่วไปที่ใช้ในการคำนวณพื้นที่ทั้งสองชนิดมีลักษณะดังนี้

$$ESA_{\mathrm{kl}} = \int_{-\infty}^{+\infty} [\mathrm{P_R}(\theta) - \mathrm{P_F}(\theta)]\, \mathrm{d}\theta$$

$$EUA_{kl} = \int_{-\infty}^{+\infty} |\mathrm{P_R}(\theta) - \mathrm{P_F}(\theta)|\, \mathrm{d}\theta$$

โดยที่ k แทน โมเดลการตอบสนองข้อสอบ

ถ้า k = 1 แสดงว่า เป็นโมเดลแบบ 1 พารามิเตอร์

ถ้า k = 2 แสดงว่า เป็นโมเดลแบบ 2 พารามิเตอร์

ถ้า k = 3 แสดงว่า เป็นโมเดลแบบ 3 พารามิเตอร์

และ l แทน พารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่ i

ถ้า l = 0 แสดงว่า เป็นดัชนีชนิดคิดเครื่องหมาย โดยมี a_i แบบเท่ากันหรือไม่เท่ากัน

ถ้า l = 1 แสดงว่า เป็นดัชนีชนิดไม่คิดเครื่องหมาย โดยมี a_i แบบเท่ากัน

ถ้า l = 2 แสดงว่า เป็นดัชนีชนิดไม่คิดเครื่องหมาย โดยมี a_i แบบไม่เท่ากัน

ค่า IRFs ระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบภายใต้โมเดลแบบ 3 พารามิเตอร์ สามารถคำนวณได้จากสูตร

$$P(\theta) = c + (1 - c)P^*(\theta)$$

โดยที่ $P^*(\theta) = \{1 + epx[-Da(\theta - b)]\}^{-1}$

a, b และ c เป็นพารามิเตอร์ที่แสดงคุณลักษณะของข้อสอบ และ D เป็นค่าคงที่ โดยปกติกำหนดให้เท่ากับ 1.7 ถ้าพารามิเตอร์ c เท่ากับ 0 แล้ว IRF จะมีเพียง 2 พารามิเตอร์ คือ a กับ b และถ้าพารามิเตอร์ a เท่ากับ 1 แล้ว จะเหลือเพียงพารามิเตอร์ b เท่านั้น

การคำนวณพื้นที่ชนิด ESA และ EUA ระหว่าง IRFs ภายใต้โมเดลโลจิสติกแบบ 1 พารามิเตอร์ หรือโมเดล Rasch สามารถคำนวณได้จากสูตร (วลีมาศ แซ่อึ้ง, 2543)

$$ESA_{10} = \hat{b}_F - \hat{b}_R$$

$$EUA_{11} = \left|\hat{b}_F - \hat{b}_R\right|$$

เนื่องจากค่าประมาณความยากของข้อสอบมีการแจกแจงลักษณะเชิงเส้นกำกับ (asymptotically) ดังนั้น ค่าเฉลี่ยและค่าความแปรปรวนของ ESA_{10} สามารถคำนวณได้จากสูตร

$$\mu(ESA_{10}) = b_F - b_R$$

$$\sigma^2(ESA_{10}) = Var(\hat{b}_F) + Var(\hat{b}_R)$$

โดยที่

$$Var(\hat{b}_i) = \left[\sum_{j=1}^{n} P_i(\theta_j)Q_i(\theta_j)\right]$$

และ $$Q_i(\theta_j) = 1 - P_i(\theta_j)$$

เมื่อ $Var(\hat{b})$ = ความแปรปรวนของค่าประมาณความยากของข้อสอบข้อที่ i

N = จำนวนผู้สอบในแต่ละกลุ่ม

$P_i(\theta_j)$ = โอกาสของผู้สอบคนที่ j ที่มีระดับความสามารถ θ จะตอบข้อสอบข้อที่ i ถูก

$Q_i(\theta_j)$ = โอกาสของผู้สอบคนที่ j ที่มีระดับความสามารถ θ จะตอบข้อสอบข้อที่ i ผิด

ถ้ากำหนดให้ $X = \hat{b}_F - \hat{b}_R$ และ X มีข้อตกลงการแจกแจงแบบปกติ ดังนั้น ค่าเฉลี่ยและความแปรปรวนของ EUA_{11} สามารถคำนวณได้จากสูตร

$$\mu(EUA_{11}) = \mu(ESA_{10})[1 - 2\Phi(z_0)] + (2/\pi)^{1/2}\sigma(ESA_{10})exp(-z_0^2/2)$$

$$\sigma^2(EUA_{11}) = \sigma^2(ESA_{10}) + \mu^2(ESA_{10}) - \mu^2(EUA_{11})$$

โดยที่ $\Phi(z_0) = \int_{-\infty}^{z_0} g(z)dz$

$z_0 = [0 - \mu(ESA_{10})]/\sigma(ESA_{10})$

เมื่อ g(z) = ฟังก์ชันความหนาแน่นของความน่าจะเป็นแบบปกติ

ในการตัดสินว่าข้อสอบที่นำมาตรวจสอบทำหน้าที่ต่างกันหรือไม่ Raju (1990) ได้เสนอให้นำดัชนีการวัดพื้นที่มาทดสอบนัยสำคัญโดยใช้สถิติ Z ภายใต้สมมติฐานการแจกแจงแบบปกติ ในการทดสอบดังกล่าวแบ่งออกเป็น 2 ชนิด ดังนี้

ชนิดที่ 1 การทดสอบนัยสำคัญของ ESA

นำดัชนี ESA ของข้อสอบข้อที่ i ไปทดสอบความแตกต่างกับ 0 โดยใช้สถิติทดสอบ Z ดังนี้

$$Z_i(ESA) = \frac{\hat{b}_{iF} - \hat{b}_{iR}}{\sqrt{Var(\hat{b}_{iF}) + Var(\hat{b}_{iR})}}$$

ชนิดที่ 2 การทดสอบนัยสำคัญของ EUA

ในการทดสอบนัยสำคัญของดัชนี EUA สามารถแบ่งออกเป็น 2 กรณี คือ กรณีที่ 1 เมื่อ $\hat{a}_{iR} = \hat{a}_{iF}$ จะทดสอบเหมือนกับดัชนี ESA สำหรับ กรณีที่ 2 $\hat{a}_{iR} \neq \hat{a}_{iF}$ จะนำดัชนี EUA ของข้อสอบข้อที่ i ไปทดสอบความแตกต่างกับ 0 โดยใช้สถิติทดสอบ Z ดังนี้

$$Z_i(H) = \frac{H_i}{\sqrt{Var(H_i)}}$$

2.5 เกณฑ์สำหรับเปรียบเทียบคุณภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

คุณภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ พิจารณาจากดัชนีบ่งชี้สำคัญสองตัว ได้แก่ อำนาจการทดสอบหรืออัตราความถูกต้องของการตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน (power rate) และความคลาดเคลื่อนของการตรวจสอบหรืออัตราความคลาดเคลื่อนประเภทที่ 1

(type I error rate) ซึ่งเป็นโอกาสของการเกิดความคลาดเคลื่อนในลักษณะที่ตรวจพบว่า ข้อสอบทำหน้าที่ต่างกัน ทั้งที่ความเป็นจริงข้อสอบไม่ได้ทำหน้าที่ต่างกัน (ศิริชัย กาญจนวาสี, 2555)

ในการคำนวณค่าสถิติตามวิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ มีจุดมุ่งหมายเพื่อทดสอบนัยสำคัญของผลการตรวจสอบ โดยมีสมมติฐานศูนย์ของการทดสอบ คือ ข้อสอบไม่ได้ทำหน้าที่ต่างกัน หรือ H_0: no DIF ผลการทดสอบสมมติฐานของวิธีการตรวจสอบ DIF วิธีต่าง ๆ นำสู่การตัดสินใจว่า ยอมรับสมมติฐานศูนย์ (accept H_0) หรือปฏิเสธสมมติฐานศูนย์ (reject H_0) ผลของการตัดสินใจมีโอกาสเกิดเหตุการณ์ 4 ลักษณะ ดังตารางที่ 2

ตารางที่ 2 คุณภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

การตัดสินใจตามผลการตรวจสอบ	H_0: no DIF	
	ความเป็นจริง	
	H_0 ถูก	H_0 ผิด
accept H_0	-ตัดสินถูก (true negative) -ระดับความเชื่อมั่น (1-α)	-ตัดสินผิด (false negative) - ความคลาดเคลื่อนประเภทที่ 2 (type II error, β)
reject H_0	-ตัดสินผิด (false positive) -ความคลาดเคลื่อนประเภทที่ 1 (type I error, α)	-ตัดสินถูก (true positive) -อำนาจการทดสอบ (1-β)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากค่าสถิติที่คำนวณตามวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ นำไปสู่การตัดสินใจสรุปผลการตรวจสอบ ดังนี้

1) ตัดสินใจถูก มีโอกาสเกิดขึ้น 2 ลักษณะ คือ การสรุปถูกว่า

-ข้อสอบ no DIF ตามความเป็นจริง (true negative)

-ข้อสอบ DIF ตามความเป็นจริง (true positive)

2) ตัดสินผิด มีโอกาสเกิดขึ้น 2 ลักษณะ คือ การสรุปผิดว่า

-ข้อสอบ DIF ทั้งที่ความจริงข้อสอบ no DIF (false positive)

-ข้อสอบ no DIF ทั้งที่ความจริงข้อสอบ DIF (false negative)

เนื่องจากอำนาจการทดสอบ (1-β) กับ β เป็นค่าดัชนีมีสเกลที่ผกผันกันและความคลาดเคลื่อนประเภทที่ 1(α) กับ 1-α ต่างก็เป็นดัชนีมีสเกลที่ผกผันเช่นเดียวกัน ดังนั้น การพิจารณา

ดัชนีบ่งชี้คุณภาพ 2 ตัว คือ อัตราความถูกต้องของการตรวจพบ DIF (power rate) และอัตราความคลาดเคลื่อนประเภทที่ 1 (type I error rate) ก็เพียงพอที่จะให้สารสนเทศครบทั้ง 4 เหตุการณ์

2.6 ขนาดอิทธิพลในการทำหน้าที่ต่างกันของข้อสอบ

Zumbo (1999 อ้างถึงใน ธเกียรติกมล ทองงอก, 2554) ได้จำแนกการวัดขนาดอิทธิพลในการทำหน้าที่ต่างกันของข้อสอบ พิจารณาได้ดังตารางที่ 3

ตารางที่ 3 ค่า R^2 ของการวัดขนาดอิทธิพลในการทำหน้าที่ต่างกันของข้อสอบ

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

จากตารางที่ 3 พบว่า การคำนวณ R^2 สำหรับการวัดขนาดอิทธิพลในการทำหน้าที่ต่างกันของข้อสอบในข้อสอบที่มีรูปแบบการให้คะแนนแบบพหุวิภาคมีเพียงวิธีเดียว คือ R^2 ที่เป็นลำดับขั้น (R-squared for ordinal) ส่วนข้อสอบที่มีรูปแบบการให้คะแนนแบบทวิวิภาคมีการคำนวณ R^2 สามวิธี ประกอบด้วย

1. วิธีการคำนวณ R^2 ของ Nagelkerke (Nagelkerke R-squared) เป็นกลุ่มนามบัญญัติ (nominal) ที่มีวิธีการคำนวณง่ายที่สุด สามารถใช้โปรแกรมสำเร็จรูป SPSS ในการคำนวณได้

2. วิธีการคำนวณ R^2 ของ WLS (Weighted-least-squares Squared) เป็นกลุ่มนามบัญญัติ (nominal) วิธีนี้เหมาะสำหรับใช้ประโยชน์กับกลุ่มตัวอย่างที่มีตัวแปรเป็นลำดับขั้นภายในกลุ่ม

3. วิธีการคำนวณ R^2 สำหรับการจัดลำดับข้อมูลหรือ R^2 ที่เป็นลำดับขั้น (R-squared for ordinal) เป็นกลุ่มลำดับ วิธีนี้เหมาะกับกลุ่มตัวแปรที่อยู่ในงานวิจัยประเภทพฤติกรรมศาสตร์และสังคมศาสตร์ เป็นตัวแปรที่มีความเป็นคุณลักษณะแฝงที่มีความต่อเนื่อง

Jodoin and Gierl (2001) ได้เสนอดัชนี $R^2\Delta$ เพื่อวัดขนาดอิทธิพลของการทำหน้าที่ต่างกันของข้อสอบของวิธีการถดถอยโลจิสติก เป็น 3 ระดับ ดังนี้

DIF ขนาดเล็ก หรือ ระดับ A : $R^2\Delta < 0.035$

DIF ขนาดปานกลาง หรือ ระดับ B : ปฏิเสธสมมติฐานว่าง และ $0.035 \leq R^2\Delta < 0.07$

DIF ขนาดใหญ่ หรือ ระดับ A : ปฏิเสธสมมติฐานว่าง และ $R^2\Delta \geq 0.07$

Roussos and Stout (1996 อ้างถึงใน Magis et al., 2010) กำหนดเกณฑ์การตัดสินขนาดอิทธิพลของวิธีซิปเทสท์เป็น 3 ระดับ ดังนี้

DIF ขนาดเล็ก หรือ ระดับ A: ปฏิเสธสมมติฐานว่าง $|\hat{\beta}_u| < 0.059$

DIF ขนาดปานกลาง หรือ ระดับ B: ปฏิเสธสมมติฐานว่าง $0.059 \leq |\hat{\beta}_u| < 0.088$

DIF ขนาดใหญ่ หรือ ระดับ C: ปฏิเสธสมมติฐานว่าง $|\hat{\beta}_u| \geq 0.088$

ตอนที่ 3 โครงการประเมินผลนักเรียนนานาชาติ (Programme for International Student Assessment; PISA)

โครงการประเมินผลนักเรียนนานาชาติ หรือที่เรียกว่า PISA เป็นโครงการที่ดำเนินการโดยองค์การเพื่อความร่วมมือและพัฒนาทางเศรษฐกิจ (Organisation for Economic Co-operation and Development; OECD) เพื่อหาตัวชี้วัดคุณภาพการศึกษา สำหรับประเทศสมาชิกและประเทศร่วมโครงการ ทั้งหมดประมาณ 90% ของเขตเศรษฐกิจโลก PISA ต้องการตรวจสอบว่า นักเรียนที่จบการศึกษาภาคบังคับหรือกลุ่มอายุ 15 ปี ได้รับความรู้และทักษะที่จำเป็นสำหรับการอยู่ในสังคมอย่างมีคุณภาพหรือไม่เพียงใด โดยเน้นให้ความสำคัญกับตัวชี้วัดศักยภาพการแข่งขันทางเศรษฐกิจ ได้แก่ การรู้เรื่อง (literacy) สามด้าน คือ การรู้เรื่องการอ่าน (reading literacy) การรู้เรื่องคณิตศาสตร์ (mathematics literacy) และการรู้เรื่องวิทยาศาสตร์ (scientific literacy) (สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี, 2553, 2554ก)

วัตถุประสงค์ของ PISA คือ ติดตามแนวโน้มการเปลี่ยนแปลงคุณภาพการเรียนรู้ของนักเรียนเพื่อให้ข้อมูลแก่ระดับนโยบาย ดังนั้น จึงมีการประเมินผลทุก ๆ สามปี โดยใช้ชื่อโครงการตามปีที่เก็บข้อมูล การประเมินผลแต่ละครั้งจะครอบคลุมทั้งสามด้าน แต่จะให้น้ำหนักความสำคัญแต่ละวิชาสลับกันไป โครงการแรก หรือการประเมินรอบแรก คือ PISA 2000 ให้น้ำหนักการอ่านเป็นหัวใจหลัก PISA 2003 ให้น้ำหนักในด้านคณิตศาสตร์ และ PISA 2006 ให้น้ำหนักในด้านวิทยาศาสตร์ ทั้งนี้ วิชาที่เป็นหัวใจหลักจะมีน้ำหนัก 60% ของภารกิจการประเมิน ส่วนวิชารองจะมีน้ำหนักวิชาละ 20% สำหรับการประเมินรอบสอง คือ PISA 2009 ให้น้ำหนักแก่การประเมินการอ่านเป็นสำคัญ PISA 2012 ให้น้ำหนักในด้านคณิตศาสตร์ และ PISA 2015 ให้น้ำหนักในด้านวิทยาศาสตร์ เพื่อติดตามดูว่าในเวลาที่เปลี่ยนไป นักเรียนมีการเปลี่ยนแปลงมากน้อยเพียงใด (สถาบันส่งเสริมการสอนวิทยาศาสตร์

และเทคโนโลยี, 2553) สำหรับรอบการประเมินผลนักเรียนนานาชาติ แสดงได้ดังตารางที่ 4 (สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี, 2554ข)

ตารางที่ 4 รอบของการประเมินโครงการประเมินผลนักเรียนนานาชาติ PISA

โครงการ	รอบที่หนึ่ง (first cycle)			รอบที่สอง (second cycle)		
ปีที่ประเมิน	PISA 2000	PISA 2003	PISA 2006	PISA 2009	PISA 2012	PISA 2015
วิชาหลัก	**การอ่าน**	**คณิตศาสตร์**	**วิทยาศาสตร์**	**การอ่าน**	**คณิตศาสตร์**	**วิทยาศาสตร์**
วิชารอง	คณิตศาสตร์ วิทยาศาสตร์	วิทยาศาสตร์ การอ่าน	การอ่าน คณิตศาสตร์	คณิตศาสตร์ วิทยาศาสตร์	วิทยาศาสตร์ การอ่าน	การอ่าน คณิตศาสตร์
		+การแก้ปัญหา	+ตัวเลือกอื่น	+ตัวเลือกอื่น	+การแก้ปัญหา	+ตัวเลือกอื่น

เนื่องจากทักษะการอ่านเป็นพื้นฐานสำคัญของการเรียนรู้ทุกประเภท การรู้เรื่องการอ่านจึงไม่เพียงแต่สำคัญเฉพาะบุคคล แต่ยังสำคัญต่อเศรษฐกิจโดยรวม เพราะต้นทุนกำลังคน (human capital) เป็นต้นทุนที่สำคัญที่สุด นักเศรษฐศาสตร์ได้ใช้เวลาหลายปีศึกษาและพัฒนาตัวแบบทางเศรษฐกิจที่แสดงว่าระดับคุณภาพการศึกษาของประเทศเป็นตัวทำนายศักยภาพความเจริญเติบโตทางเศรษฐกิจของประเทศนั้น ๆ (OECD, 2009 อ้างถึงใน สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี, 2553)

PISA 2009 ได้นิยามการรู้เรื่องการอ่าน คือ ความเข้าใจ (understanding) การใช้ (using) การสะท้อน (reflecting) และความรักและผูกพันกับการอ่าน (engaging) ในถ้อยความที่เป็นข้อเขียน (written texts) ที่ได้อ่าน เพื่อไปบรรลุเป้าหมายของแต่ละคน เพื่อพัฒนาความรู้และศักยภาพของตน และเพื่อมีส่วนร่วมในกระบวนการของสังคม (สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี, 2554ข) โดยข้อสอบจะวัด 3 สมรรถนะ ได้แก่ การบูรณาการและตีความ (integrate and interpret) การสะท้อนและประเมิน (reflect and evaluate) และการเข้าถึงและค้นสาระ (access and retrieve)

จากที่กล่าวมา การวิจัยครั้งนี้จึงศึกษาการทำหน้าที่ต่างกันของข้อสอบกับข้อมูลเชิงประจักษ์จากโครงการประเมินผลนักเรียนนานาชาติ PISA 2009 ด้านการรู้เรื่องการอ่าน คัดเลือกเฉพาะข้อสอบที่มีรูปแบบการตรวจให้คะแนนแบบทวิวิภาค จำนวน 39 ข้อ แต่ละข้อจะมีอยู่ในแบบสอบ 4 ฉบับ ฉบับของแบบสอบมีจำนวน 13 ฉบับ รายละเอียดดังตารางที่ 5

ตารางที่ 5 ข้อสอบ PISA 2009 ด้านการรู้เรื่องการอ่าน ที่ตรวจให้คะแนนแบบทวิวิภาค

ข้อที่	Unit item code	Unit name	สมรรถนะ	ฉบับ (booklet)													รวมข้อ
				1	2	3	4	5	6	7	8	9	10	11	12	13	
1	R055Q01	Drugged Spiders	การบูรณาการและตีความ				✓				✓			✓		✓	4
2	R067Q01	Aesop	การบูรณาการและตีความ	✓	✓							✓				✓	4
3	R083Q01	Household Work	การบูรณาการและตีความ		✓		✓	✓		✓							4
4	R083Q04	Household Work	การบูรณาการและตีความ		✓		✓	✓		✓							4
5	R101Q01	Rhino	การบูรณาการและตีความ		✓		✓	✓		✓							4
6	R101Q02	Rhino	การบูรณาการและตีความ		✓		✓	✓		✓							4
7	R101Q03	Rhino	การสะท้อนและประเมิน		✓		✓	✓		✓							4
8	R101Q04	Rhino	การบูรณาการและตีความ		✓		✓	✓		✓							4
9	R101Q05	Rhino	การบูรณาการและตีความ		✓		✓	✓		✓							4
10	R102Q07	Shirts	การบูรณาการและตีความ	✓	✓							✓				✓	4
11	R111Q01	Exchange	การบูรณาการและตีความ				✓				✓			✓		✓	4
12	R220Q02B	South Pole	การบูรณาการและตีความ	✓	✓							✓				✓	4
13	R220Q04	South Pole	การบูรณาการและตีความ	✓	✓							✓				✓	4
14	R220Q05	South Pole	การบูรณาการและตีความ	✓	✓							✓				✓	4

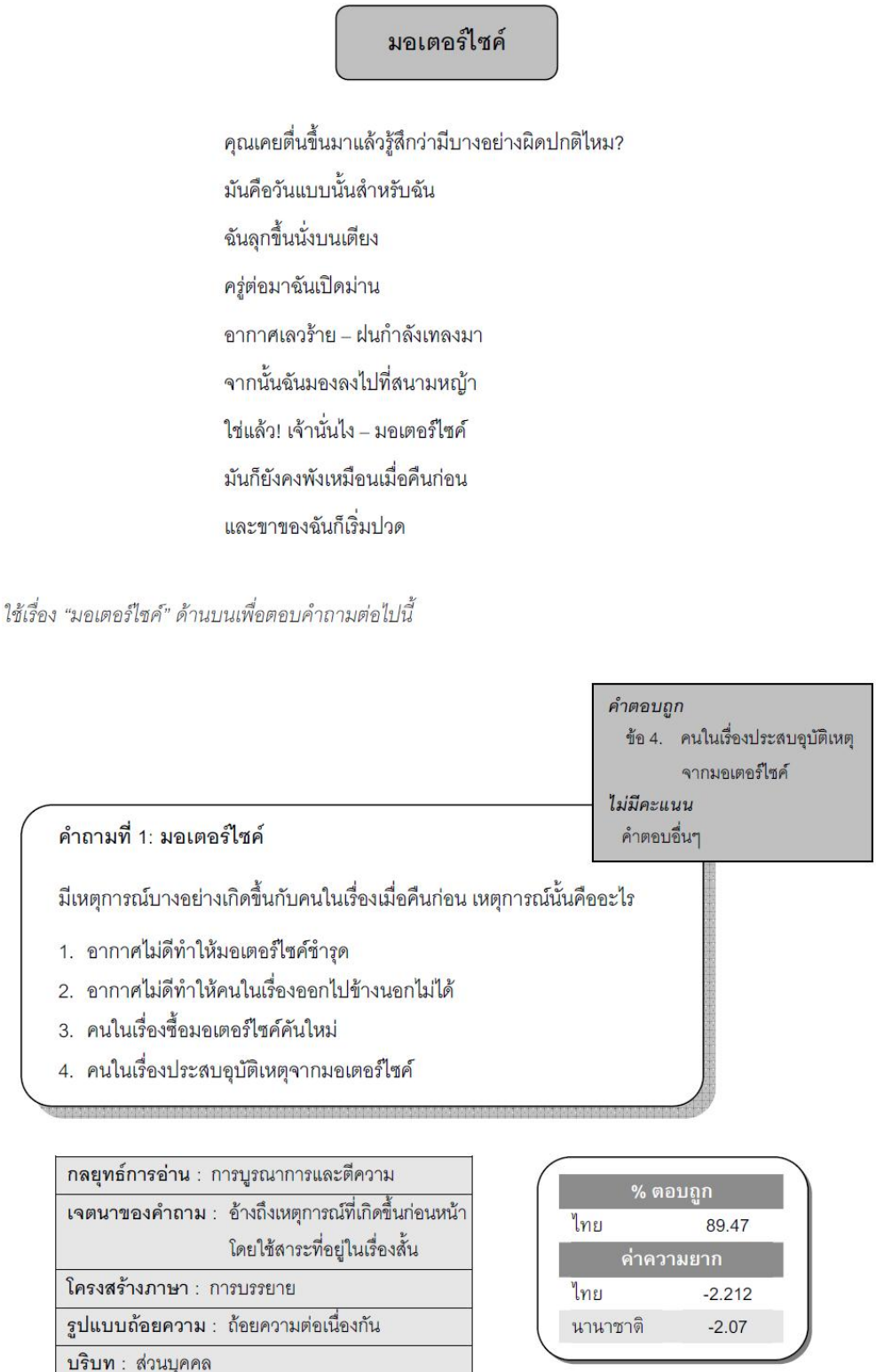
ตารางที่ 5 (ต่อ)

ข้อที่	Unit item code	Unit name	สมรรถนะ	ฉบับ (booklet)													รวมข้อ
				1	2	3	4	5	6	7	8	9	10	11	12	13	
15	R220Q06	South Pole	การบูรณาการและตีความ	✓	✓							✓				✓	4
16	R227Q01	Optician	การบูรณาการและตีความ				✓				✓			✓		✓	4
17	R404Q03	Sleep	การบูรณาการและตีความ					✓	✓				✓			✓	4
18	R404Q06	Sleep	การบูรณาการและตีความ					✓	✓				✓			✓	4
19	R412Q01	World Languages	การเข้าถึงและค้นสาระ						✓	✓	✓	✓					4
20	R412Q05	World Languages	การบูรณาการและตีความ						✓	✓	✓	✓					4
21	R414Q02	Mobile Phone Safety	การบูรณาการและตีความ	✓		✓	✓		✓								4
22	R414Q09	Mobile Phone Safety	การบูรณาการและตีความ	✓		✓	✓		✓								4
23	R414Q11	Mobile Phone Safety	การสะท้อนและประเมิน	✓		✓	✓		✓								4
24	R424Q03	Fair Trade	การสะท้อนและประเมิน					✓	✓				✓			✓	4
25	R424Q07	Fair Trade	การสะท้อนและประเมิน					✓	✓				✓			✓	4
26	R437Q01	Narcissus	การบูรณาการและตีความ						✓	✓	✓	✓					4
27	R437Q06	Narcissus	การบูรณาการและตีความ						✓	✓	✓	✓					4
28	R442Q07	If Galileo raised his head	การบูรณาการและตีความ		✓		✓	✓		✓							4

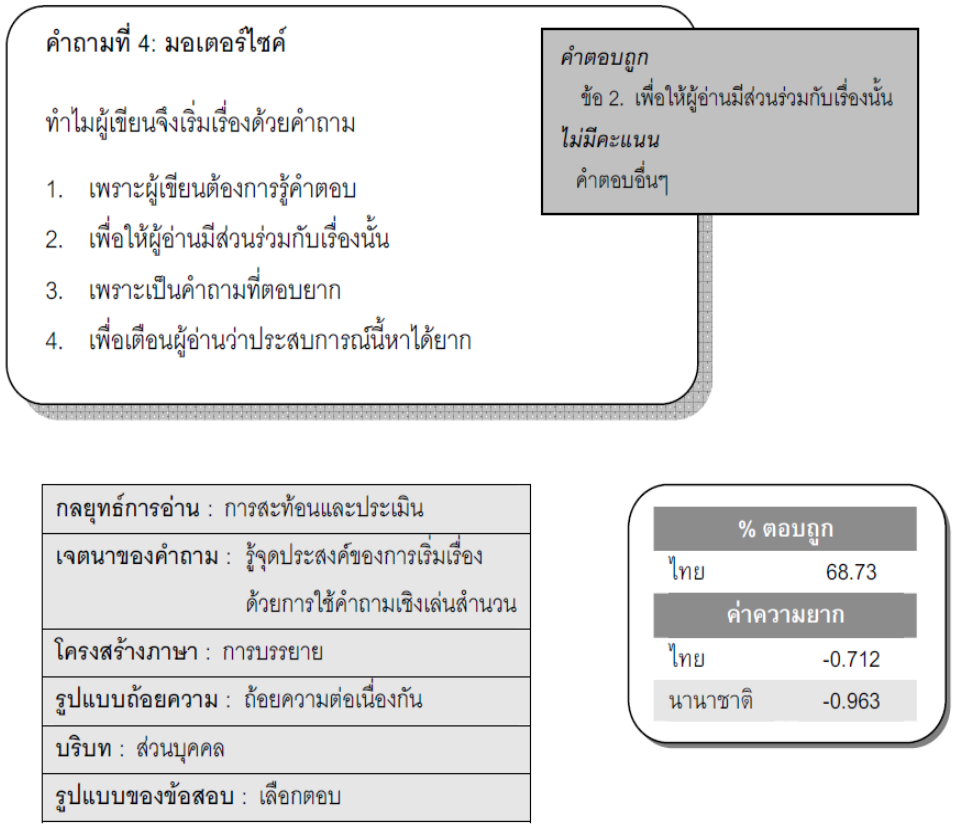
ตารางที่ 5 (ต่อ)

ข้อที่	Unit item code	Unit name	สมรรถนะ	ฉบับ (booklet)													รวมข้อ
				1	2	3	4	5	6	7	8	9	10	11	12	13	
29	R447Q04	Acne Vulgaris	การสะท้อนและประเมิน	✓		✓	✓		✓								4
30	R447Q05	Acne Vulgaris	การเข้าถึงและค้นสาระ	✓		✓	✓		✓								4
31	R452Q04	The Play's the Thing	การบูรณาการและตีความ	✓		✓	✓		✓								4
32	R452Q07	The Play's the Thing	การบูรณาการและตีความ	✓		✓	✓		✓								4
33	R453Q01	Summer Job	การบูรณาการและตีความ						✓	✓	✓	✓					4
34	R455Q04	Chocolate and Health	การบูรณาการและตีความ					✓	✓				✓			✓	4
35	R456Q01	Biscuits	การเข้าถึงและค้นสาระ		✓				✓					✓	✓		4
36	R458Q01	Telecommuting	การบูรณาการและตีความ	✓		✓	✓		✓								4
37	R458Q04	Telecommuting	การบูรณาการและตีความ	✓		✓	✓		✓								4
38	R460Q05	Gulf of Mexico	การเข้าถึงและค้นสาระ		✓				✓					✓	✓		4
39	R460Q06	Gulf of Mexico	การบูรณาการและตีความ		✓				✓					✓	✓		4
			รวมจำนวนข้อ	15	17	9	20	13	22	13	8	11	5	6	3	14	156

ตัวอย่างข้อสอบ PISA 2009 การรู้เรื่องการอ่าน ทั้ง 3 ด้าน แสดงดังภาพที่ 7-9



ภาพที่ 7 ตัวอย่างข้อสอบ PISA 2009 การรู้เรื่องการอ่าน ด้านการบูรณาการและตีความ



ภาพที่ 8 ตัวอย่างข้อสอบ PISA 2009 การรู้เรื่องการอ่าน ด้านการสะท้อนและประเมิน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ประชาธิปไตยในเอเธนส์

ตอนที่ 1

ธูซิดิดีส เป็นนักประวัติศาสตร์และทหารที่มีชีวิตอยู่เมื่อศตวรรษที่ 5 ก่อนคริสตกาลในยุคกรีกโบราณ เขาเกิดในกรุงเอเธนส์ ในช่วงสงครามเพโลพอนเนสเซียน (431 – 404 ปีก่อนคริสตกาล) ระหว่างเอเธนส์กับสปาร์ต้า เขามีตำแหน่งควบคุมกองเรือที่มีภารกิจคือป้องกันเมืองแอมฟิโพลิสในนครเทรซ เขาไปถึงเมืองช้าไปและเมืองตกอยู่ในมือของบราซิดัสนายพลชาวสปาร์ต้า ซึ่งเนรเทศธูซิดิดีสออกไปเป็นเวลา 20 ปี สิ่งนี้ช่วยให้เขามีโอกาสเก็บรวบรวมข้อมูลจากทั้งสองฝ่ายที่ทำสงครามกันและเปิดโอกาสให้เขามีผลงานค้นคว้าเรื่อง *ประวัติศาสตร์ของสงครามเพโลพอนเนสเซียน*

ธูซิดิดีสนับว่าเป็นหนึ่งในยอดนักประวัติศาสตร์ยุคโบราณ เขามุ่งความสนใจในเรื่องสาเหตุจากธรรมชาติและพฤติกรรมของแต่ละคนมากกว่าเรื่องโชคชะตาหรือการกระทำของพระเจ้า เพื่ออธิบายวิวัฒนาการของประวัติศาสตร์ ในงานของเขาจะไม่เพียงบรรยายข้อเท็จจริงเท่านั้นแต่จะพยายามหาเหตุผลที่ชักจูงให้ตัวละครหลักเกิดการกระทำในลักษณะนั้น ธูซิดิดีส เน้นในเรื่องพฤติกรรมของแต่ละบุคคลทำให้บางครั้งเขาเริ่มแต่งสุนทรพจน์ขึ้นมา สิ่งเหล่านี้ช่วยให้เขาอธิบายแรงจูงใจของตัวละครในประวัติศาสตร์

ตอนที่ 2

ธูซิดิดีสเขียนในฐานะของเพริคลิส (ศตวรรษที่ 5 ก่อนคริสตกาล) ผู้ปกครองชาวเอเธนส์ สุนทรพจน์ต่อไปนี้เป็นเกียรติแก่ทหารที่พลีชีพในปีแรกของสงครามเพโลพอนเนสเซียน

ระบบการปกครองของพวกเราไม่ได้ลอกเลียนมาจากรัฐข้างเคียง เราเป็นแบบอย่างให้ผู้อื่นมากกว่าที่เราจะเป็นผู้ลอกเลียนแบบ ระบบของเราเรียกว่าประชาธิปไตย เนื่องจากเป็นการบริหารที่ขึ้นอยู่กับคนส่วนใหญ่มากกว่าคนเพียงไม่กี่คน กฎหมายของเราพยายามให้สิทธิเท่าเทียมกันในเรื่องของแต่ละบุคคล แต่ทว่าชื่อเสียงเกียรติยศต่อสาธารณะขึ้นอยู่กับคุณความดีมากกว่าระดับชนชั้นในสังคม

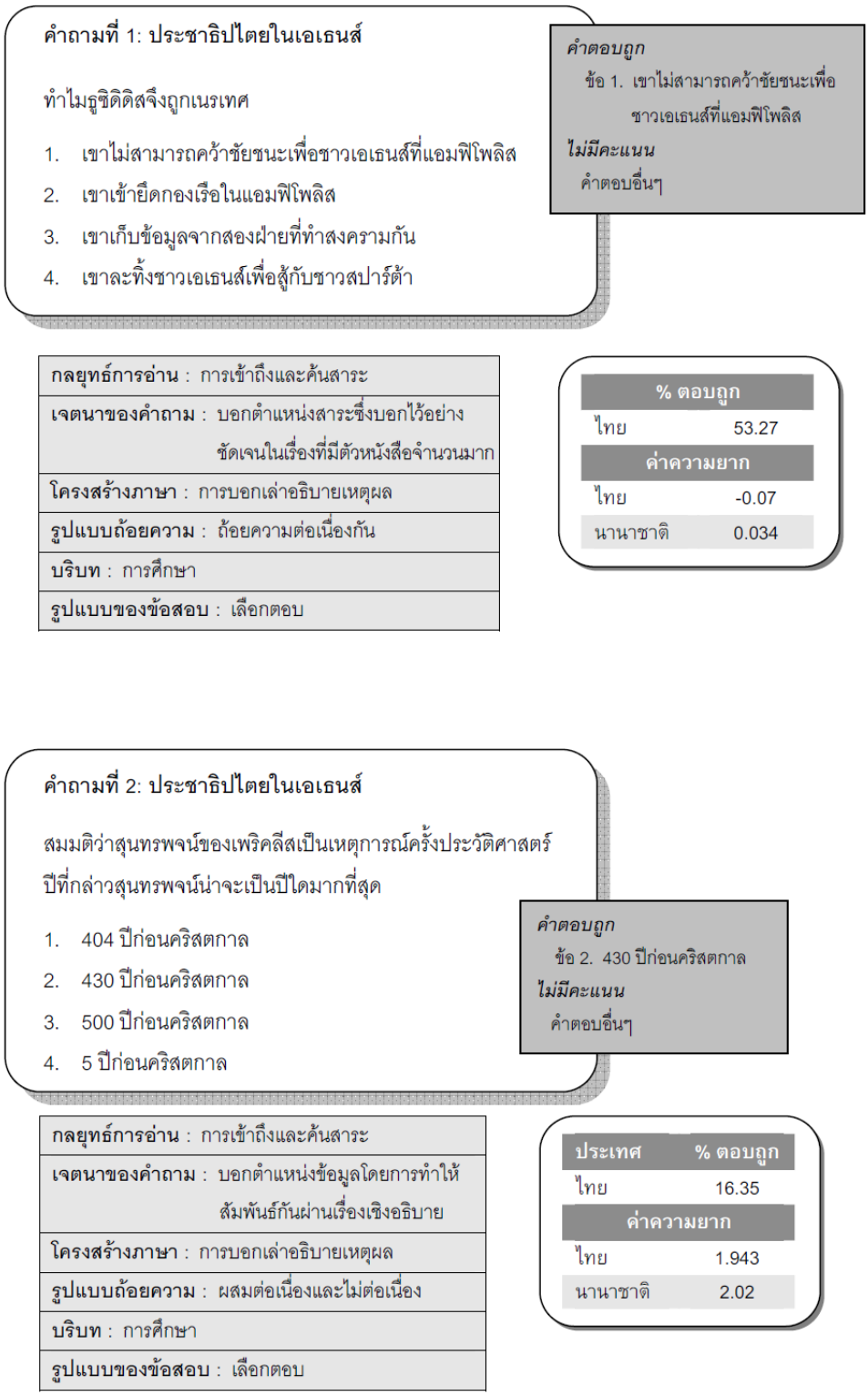
ระดับชนชั้นในสังคมไม่ได้กีดกันบุคคลจากการรับตำแหน่งทางสังคมเช่นกัน (...) และในเวลาเดียวกันเราจะไม่เข้าไปยุ่งเกี่ยวกับเรื่องส่วนตัว เราจะไม่ฝ่าฝืนกฎหมายในเรื่องสาธารณะ เราจะทำตามผู้ที่อยู่ในตำแหน่งที่มีอำนาจสั่งการ และเราจะทำตามกฎหมายโดยเฉพาะส่วนที่ป้องกันการกดขี่ และกฎหมายที่ไม่ได้เขียนเป็นลายลักษณ์อักษรซึ่งถือเป็นความน่าละอายที่จะละเมิด

ยิ่งไปกว่านั้น เรายังมีสิ่งที่น่าเพลิดเพลินใจอีกมากมาย เกมการแข่งขันและการบวงสรวงที่เราเฉลิมฉลองกันตลอดปี จากความสง่างามของสถานที่พักอาศัยส่วนตัวของพวกเรา ซึ่งสร้างความพึงพอใจในแต่ละวันและช่วยกำจัดความวิตกกังวลต่างๆ ออกไป ในขณะที่ผู้อาศัยในเมืองจำนวนมากสามารถดึงผลผลิตจากทั่วโลกมาสู่เอเธนส์ ดังนั้นชาวเอเธนส์จึงคุ้นเคยกับผลไม้ของประเทศอื่น เช่นเดียวกับคนในประเทศนั้นๆ

ธูซิดิดีส ประวัติศาสตร์ของสงครามเพโลพอนเนสเซียน (ฉบับปรับปรุง)

ภาพที่ 9 ตัวอย่างข้อสอบ PISA 2009 การรู้เรื่องการอ่าน ด้านการเข้าถึงและค้นสาระ

ใช้ "ประชาธิปไตยในเอเธนส์" ในหน้าที่ผ่านมาเพื่อตอบคำถามต่อไปนี้



ภาพที่ 9 (ต่อ)

ตอนที่ 4 งานวิจัยที่เกี่ยวกับการศึกษาการทำหน้าที่ต่างกันของข้อสอบในแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค

เอกสารและงานวิจัยที่เกี่ยวข้องกับการศึกษาการทำหน้าที่ต่างกันของข้อสอบที่มีการตรวจให้คะแนนแบบทวิวิภาคนำเสนอได้ดังนี้

4.1 งานวิจัยที่เกี่ยวข้องกับการศึกษาปัจจัยที่มีผลต่อการทำหน้าที่ต่างกันของข้อสอบในแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค

เกษร หว่างจิตร (2539) ทำการศึกษาวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทยและวิชาภาษาอังกฤษ ของศูนย์ทดสอบทางการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ด้วยวิธีแมนเทล-เฮนส์เซล เมื่อจำแนกกลุ่มผู้สอบตามเพศ ภูมิลำเนา ประสบการณ์ในการสอบ และสังกัดของสถานศึกษา โดยมีกลุ่มตัวอย่างเป็นผู้สอบวิชาภาษาไทย จำนวน 506 คน และผู้สอบวิชาภาษาอังกฤษ จำนวน 501 คน ผลการวิจัย พบว่า ข้อสอบที่พบ DIF ส่วนใหญ่เป็นแบบอเนกรูป ข้อสอบที่ DIF ทั้งแบบเอกรูปและอเนกรูปเกิดขึ้นเมื่อการวิเคราะห์จำแนกกลุ่มผู้สอบตามเพศมากที่สุด ลักษณะของข้อสอบที่พบ DIF ส่วนใหญ่เป็นข้อสอบที่มีค่าอำนาจจำแนกค่อนข้างต่ำ ค่าความเที่ยง และค่าความตรงตามทฤษฎีของแบบสอบก่อนและหลังตัดข้อสอบที่พบ DIF ออกจากแบบสอบส่วนใหญ่ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ

รักชนก ยี่สุ่นศรี (2544) ทำการศึกษาวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบด้วยกระบวนการ ดี เอฟ ไอ ที กับแบบสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาของทบวงมหาวิทยาลัย ปีการศึกษา 2543 ครั้งที่ 1 วิชาภาษาอังกฤษและวิชาคณิตศาสตร์ จากกลุ่มตัวอย่าง จำนวน 4,000 คน และ 3,600 คน ตามลำดับ ผลการวิจัย พบว่า แบบสอบวิชาภาษาอังกฤษทำหน้าที่ต่างกันตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษาของผู้สอบ ส่วนแบบสอบวิชาคณิตศาสตร์ทำหน้าที่ต่างกันตามเพศของผู้สอบ แบบสอบฉบับก่อนและหลังตัดข้อสอบที่พบ DIF มีค่าความตรงเชิงโครงสร้างไม่แตกต่างกัน แบบสอบฉบับหลังตัดข้อสอบที่พบ DIF ส่วนใหญ่มีค่าความเที่ยงลดลง แต่ค่าฟังก์ชันสารสนเทศของแบบสอบมากขึ้น และความสัมพันธ์ระหว่างตำแหน่งของคะแนนรวมของผู้สอบก่อนและหลังตัดข้อสอบที่พบ DIF มีความสัมพันธ์กันในทางบวกอย่างมีนัยสำคัญ

สุมาลี แกล้วทนงค์ (2547) ทำการศึกษาสาเหตุของการทำหน้าที่ต่างกันของข้อสอบสาระการเรียนรู้ภาษาไทย และสาระการเรียนรู้สังคมศึกษา ศาสนาและวัฒนธรรม กลุ่มตัวอย่างเป็นนักเรียนระดับชั้นมัธยมศึกษาปีที่ 1 สำนักงานเขตพื้นที่การศึกษาสงขลา พัทลุง ตรัง และสตูล จำนวน 1,320 คน ผลการวิจัย พบว่า ข้อสอบที่ทำหน้าที่ต่างกันด้านเพศ ด้านภาษาพูด และด้านเชื้อชาติของแบบสอบสาระการเรียนรู้ภาษาไทย มีสาเหตุมาจากเนื้อเรื่องที่สนใจและภาษาที่ใช้ในแบบสอบที่เป็นสำนวน

และคำศัพท์ คำศัพท์เฉพาะ และคำราชาศัพท์หรือคำศัพท์ยาก ตามลำดับ ส่วนแบบสอบสาระการเรียนรู้สังคมศึกษา ศาสนาและวัฒนธรรม มีสาเหตุมาจากเรื่องที่สนใจและเนื้อเรื่องเกี่ยวกับวัฒนธรรม

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

French and Maller (2007) ทำการศึกษาการขจัดความลำเอียงออกจากข้อสอบและการใช้ขนาดของอิทธิพลกับวิธีการวิเคราะห์การถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ใช้แบบสอบที่มีช่วงความยาว 20 ถึง 80 ข้อ ผ่านการจำลองข้อมูลและใช้กลุ่มตัวอย่างจริง 50 คน ในแต่ละเงื่อนไขการทดสอบ เพื่อยืนยันความคงที่ของผลการทดสอบ เงื่อนไขการทดสอบประกอบด้วยการขจัดความลำเอียงออกจากข้อสอบและไม่มีการขจัดความลำเอียงออกจากข้อสอบไขว้กับการใช้และไม่ใช้เกณฑ์ขนาดของอิทธิพล ผลการวิจัย พบว่า การขจัดความลำเอียงออกจากข้อสอบทำให้อำนาจการทดสอบโดยรวมและความคลาดเคลื่อนประเภทที่ 1 ดีขึ้นไม่มาก แต่ก็มีประโยชน์ภายใต้เงื่อนไขที่แน่นอน การทดสอบการวิเคราะห์การถดถอยโลจิสติกที่ไม่มีการขจัดความลำเอียงออกจากข้อสอบใช้ได้ดีเทียบเท่ากับเกณฑ์ในการจำแนกอื่น ๆ และเป็นทางเลือกที่จะนำไปปฏิบัติในสถานการณ์ต่าง ๆ

Scherbaum and Goldstein (2008) ทำการศึกษาตรวจสอบความสัมพันธ์ระหว่างการทำหน้าที่ต่างกันของข้อสอบที่อยู่บนฐานของเชื้อชาติกับความยากของข้อสอบ ใช้ข้อมูลจาก the Civic Education Study of the International Association for the Evaluation of Educational Achievement (CES) ในปี 1999 กลุ่มตัวอย่างเป็นนักเรียนผิวขาว จำนวน 1,701 คน และนักเรียนผิวดำ จำนวน 427 คน การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบใช้วิธีดัชนีมาตรฐาน (STND) และ differential functioning of items and tests (DFIT) procedure บนพื้นฐานของ IRT ผลการวิจัย พบว่า มีความสัมพันธ์ที่เป็นสาระสำคัญระหว่างความยากของข้อสอบกับการทำหน้าที่ต่างกันของข้อสอบ เมื่อใช้เทคนิคการวิเคราะห์ DIF และแหล่งข้อมูลที่แตกต่างกัน ในการจำลองข้อมูลที่ข้อสอบ ไม่มี DIF ก็ไม่พบความสัมพันธ์ระหว่างความยากของข้อสอบกับการทำหน้าที่ต่างกันของข้อสอบ วิธีดัชนีมาตรฐาน พบความสัมพันธ์ระหว่างความยากของข้อสอบกับการทำหน้าที่ต่างกันของข้อสอบเพียงเล็กน้อย

Chulalongkorn University

Woods (2008) ทำการศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีอัตราส่วนไลค์ลิฮูด (Likelihood-Ratio) ในกรณีที่มีผลของการแจกแจงไม่เป็นปกติ ผ่านการจำลองข้อมูล โดยมีตัวแปรที่จัดกระทำ 3 ตัว ได้แก่ การเกิดของ DIF ความแตกต่างของค่าเฉลี่ยในค่าความสามารถของผู้สอบระหว่างกลุ่ม และการแจกแจงค่าความสามารถของผู้สอบ ใช้ข้อสอบที่เป็นแบบ binary จำนวน 24 ข้อ ผลการวิจัย พบว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี IRT-LR ทำงานได้ดีในกรณีการแจกแจงแฝงเป็นปกติทั้งกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ แต่เมื่อการแจกแจงเบ้ในกลุ่มหนึ่งหรือทั้งสองกลุ่ม ผลการตรวจสอบจะผิดเพี้ยน โดยเฉพาะเมื่อความคลาดเคลื่อนประเภทที่ 1 เพิ่มสูงขึ้น การประมาณค่าพารามิเตอร์ของข้อสอบของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบจะไม่ถูกต้อง ความแตกต่างของกลุ่มในค่าเฉลี่ยและความแปรปรวนจะเกินกว่าการประมาณค่าจริง

DeMars (2009) ทำการศึกษาปรับปรุงกระบวนการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธีแมนเทล-แฮนส์เซลและวิธีการวิเคราะห์การถดถอยโลจิสติกที่รวมการแก้ไขการถดถอย โดยใช้การจำลองข้อมูล ปัจจัยที่นำมาวิเคราะห์ ได้แก่ ความยาวของแบบสอบ ระดับของผลกระทบต่อกลุ่ม และขนาดของกลุ่มตัวอย่าง ผลการวิจัย พบว่า การใช้การแก้ไขแบบการถดถอย เชิงเส้นและไม่เป็นเชิงเส้น ช่วยลดการสูงขึ้นของความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซลและวิธีการวิเคราะห์การถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันแบบเอกรูปอย่างมีประสิทธิผล แต่ไม่ได้มีการทดสอบกับวิธีการวิเคราะห์การถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันแบบอเนกรูป สำหรับกรณีกลุ่มตัวอย่างขนาดใหญ่และความแตกต่างค่าเฉลี่ยของกลุ่มมีขนาดใหญ่ Δ effect size ได้รับการประมาณที่ถูกต้องมากขึ้น

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

Woods and Grimm (2011) ทำการศึกษาตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปโดยโมเดล Multiple Indicator Multiple Cause (MIMIC) ซึ่งปฏิสัมพันธ์ระหว่างตัวแปรแฝงเป็นสิ่งที่เพิ่มเข้ามาในโมเดล MIMIC เพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป แนวทางนี้ใช้สถานการณ์จำลองด้วยกลุ่มเปรียบเทียบขนาดเล็ก และตัวอย่างเชิงประจักษ์ โดยใช้สเกล agoraphobic cognitions เงื่อนไขที่ใช้ศึกษา ได้แก่ ขนาดของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ จำนวนของข้อสอบ และการเกิดและไม่เกิด DIF โมเดล MIMIC แบบมีปฏิสัมพันธ์นำมาเปรียบเทียบกับโมเดล MIMIC แบบไม่มีปฏิสัมพันธ์ ที่ดีเท่ากับวิธีตรวจสอบอัตราส่วนไลค์ลิฮูด ที่ใช้ item response theory (IRT-LR-DIF) ผลการวิจัย พบว่า เมื่อสมการโครงสร้าง latent moderated นำมาประมาณค่าปฏิสัมพันธ์ ความคลาดเคลื่อนประเภทที่ 1 ในโมเดล MIMIC แบบมีปฏิสัมพันธ์เพิ่มขึ้นอย่างรุนแรง

DeMars and Lau (2011) ทำการศึกษาการทำหน้าที่ต่างกันของข้อสอบกับ latent classes ดูความถูกต้องในการตรวจสอบคนที่ตอบสนองแตกต่าง โดยใช้โมเดลผสมของทฤษฎีการตอบสนองข้อสอบ กับข้อมูลจำลอง เงื่อนไขในการศึกษา ได้แก่ ปัจจัยอิทธิพล จำนวนข้อสอบที่พบ DIF และจำนวนข้อสอบที่ไม่แปรเปลี่ยน ผลการวิจัย พบว่า การใช้โมเดลผสมของทฤษฎีการตอบสนองข้อสอบ สำหรับตรวจสอบ DIF แฝง ยังมีปัญหา การค้นพบ class membership ยังไม่ดีในทุกเงื่อนไขการทดสอบ การประมาณค่าพารามิเตอร์อำนาจจำแนกให้ผลดีสำหรับข้อที่ไม่แปรเปลี่ยนเช่นเดียวกันกับข้อที่พบ DIF ในกรณีที่ไม่มีอิทธิพลของกลุ่ม แต่เมื่อมีอิทธิพลของกลุ่ม การจำแนกข้อที่พบ DIF จะลำเอียงทางบวก เมื่อไม่มีอิทธิพลของกลุ่ม การประมาณค่าผลของ DIF มีแนวโน้มที่จะลำเอียงทางบวก โดยทั่วไปมีข้อสอบจำนวนน้อยที่มีการประมาณค่าที่ลำเอียงและมีความคลาดเคลื่อนมาตรฐานขนาดใหญ่

Chulalongkorn University

ตารางที่ 6 การศึกษาปัจจัยที่มีผลต่อการทำหน้าที่ต่างกันของข้อสอบในแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค

ผู้วิจัย	ปี	เครื่องมือ/ข้อมูลที่ใช้วิเคราะห์	วิธีตรวจสอบ	ปัจจัยที่นำมาศึกษา	ข้อค้นพบ
เกษร หว่างจิตร	2539	ข้อสอบวิชาภาษาไทยและภาษาอังกฤษ ของศูนย์ทดสอบทางการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย	MH	-เพศ -ภูมิลำเนา -ประสบการณ์ในการสอบ -สังกัดของสถานศึกษา	ข้อสอบที่พบ DIF ทั้งแบบเอกรูปและอเนกรูปเกิดขึ้นเมื่อวิเคราะห์จำแนกผู้สอบตามกลุ่มเพศมากที่สุด
รักชนก ยี่สุ่นศรี	2544	แบบสอบวิชาภาษาอังกฤษและวิชาคณิตศาสตร์ คัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาของทบวงมหาวิทยาลัย	DFIT procedure	-เพศ -สถานที่ตั้งของโรงเรียนที่จบการศึกษา -ความตรงเชิงโครงสร้าง -ความเที่ยง -ฟังก์ชันสารสนเทศของแบบสอบ -ความสัมพันธ์ระหว่างตำแหน่งของคะแนนรวม	แบบสอบวิชาภาษาอังกฤษทำหน้าที่ต่างกันตามเพศและสถานที่ตั้งของโรงเรียนที่จบการศึกษา ส่วนแบบสอบวิชาคณิตศาสตร์ ทำหน้าที่ต่างกันตามเพศ -ค่าความตรงเชิงโครงสร้าง ไม่แตกต่างกัน ค่าความเที่ยงลดลง และค่าฟังก์ชันสารสนเทศมากขึ้น -ความสัมพันธ์ในทางบวกอย่างมีนัยสำคัญระหว่างตำแหน่งของคะแนนรวมของผู้สอบก่อนและหลังตัดข้อสอบที่พบ DIF
สุมาลี แกล้วทนงค์	2547	แบบสอบสาระการเรียนรู้ภาษาไทย และสาระการเรียนรู้สังคมศึกษา ศาสนา และวัฒนธรรม ชั้น ม.1	MH	-เพศ -ภาษาพูด -เชื้อชาติ	-แบบสอบสาระการเรียนรู้ภาษาไทย ด้านเพศ ภาษาพูด และเชื้อชาติ มีสาเหตุมาจากเนื้อเรื่องที่สนใจ และภาษาที่ใช้ในแบบสอบ -แบบสอบสาระการเรียนรู้สังคมศึกษา ศาสนา และวัฒนธรรม ด้านเพศ ภาษาพูด และเชื้อชาติ มีสาเหตุมาจากเนื้อเรื่องที่สนใจ และเนื้อเรื่องเกี่ยวกับวัฒนธรรม
Sheppard, et al.	2006	Hogan personality inventory	MH	-เพศ -เชื้อชาติ	1 ใน 3 ของข้อคำถามเป็นข้อที่พบ DIF โดยส่วนใหญ่ข้อที่พบ DIF จะเกี่ยวกับเพศมากกว่าเชื้อชาติ
French และ Maller	2007	ข้อมูลจำลอง	LR	-การขจัดความลำเอียงออกจากข้อสอบ -การใช้เกณฑ์ขนาดของอิทธิพล	การขจัดความลำเอียงออกจากข้อสอบทำให้อำนาจการทดสอบโดยรวมและความคลาดเคลื่อนประเภทที่ 1 ดีขึ้นเล็กน้อย

ตารางที่ 6 (ต่อ)

ผู้วิจัย	ปี	เครื่องมือ/ข้อมูลที่ใช้วิเคราะห์	วิธีตรวจสอบ	ปัจจัยที่นำมาศึกษา	ข้อค้นพบ
Scherbaum และ Goldstein	2008	the Civic Education Study of the International Association for the Evaluation of Educational Achievement (CES)	STND และ IRT procedure	-เชื้อชาติ -ความยากของข้อสอบ	มีความสัมพันธ์ที่เป็นสาระสำคัญระหว่างความยากของข้อสอบกับการทำหน้าที่ต่างกันของข้อสอบเมื่อใช้เทคนิคการวิเคราะห์ DIF และแหล่งข้อมูลที่แตกต่างกัน
Woods	2008	ข้อมูลจำลอง	IRT LR	-การเกิดของ DIF -ความแตกต่างของค่าเฉลี่ยในค่าความสามารถของผู้สอบระหว่างกลุ่ม -การแจกแจงค่าความสามารถของผู้สอบ	-วิธี IRT LR ทำงานได้ดีกรณีการแจกแจงแผ่งเป็นปกติทั้งกลุ่ม R และกลุ่ม F -เมื่อการแจกแจงเบ้ในกลุ่มหนึ่งหรือทั้งสองกลุ่ม ผลการตรวจสอบจะผิดเพี้ยน -เมื่อความคลาดเคลื่อนประเภทที่ 1 เพิ่มสูงขึ้น การประมาณค่าพารามิเตอร์ของข้อสอบของกลุ่ม R และกลุ่ม F จะไม่ถูกต้อง -ความแตกต่างของกลุ่มในค่าเฉลี่ยและความแปรปรวนจะเกินกว่าการประมาณค่าจริง
Demars	2009	ข้อมูลจำลอง	MH และ LR	-ความยาวของแบบสอบ -ระดับของผลกระทบต่อกลุ่ม -ขนาดของกลุ่มตัวอย่าง	-การแก้ไขแบบ linear regression และ nonlinear ช่วยลดการสูงขึ้นของความคลาดเคลื่อนประเภทที่ 1 ของวิธี MH และวิธี LR ในการตรวจสอบ uniform DIF อย่างมีประสิทธิผล -กรณีกลุ่มตัวอย่างขนาดใหญ่และความแตกต่างค่าเฉลี่ยของกลุ่มมีขนาดใหญ่ Δ effect size ได้รับการประมาณที่ถูกต้อง

ตารางที่ 6 (ต่อ)

ผู้วิจัย	ปี	เครื่องมือ/ข้อมูลที่ใช้วิเคราะห์	วิธีตรวจสอบ	ปัจจัยที่นำมาศึกษา	ข้อค้นพบ
Shin และ Wang	2009	ข้อมูลจำลอง	MIMIC	-โมเดลการตอบสนองข้อสอบ -ขนาดกลุ่มตัวอย่าง -ความยาวของแบบสอบ -ร้อยละของข้อสอบที่พบ DIF ในแบบสอบ -รูปแบบของ DIF -ความยาวของข้อสอบ anchor	MIMIC ที่ความยาวของข้อสอบ anchor 1, 2, 4 และ 10 ข้อ เมื่อข้อสอบที่พบ DIF มีร้อยละ 40 ในแบบสอบ สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้เป็นอย่างดี
Woods และ Grimm	2011	ข้อมูลจำลอง	MIMIC	-ขนาดของกลุ่ม -จำนวนข้อสอบ -การเกิดและไม่เกิด DIF	เมื่อ approach สมการโครงสร้าง latent moderated นำมาประมาณค่าปฏิสัมพันธ์ ความคลาดเคลื่อนประเภทที่ 1 ในโมเดล MIMIC แบบมีปฏิสัมพันธ์ เพิ่มขึ้นอย่างรุนแรง
DeMars และ Lau	2011	ข้อมูลจำลอง	IRT mixture models	-ปัจจัยอิทธิพล -จำนวนข้อสอบที่พบ DIF -จำนวนข้อสอบที่ไม่แปรเปลี่ยน	-การใช้ IRT mixture models สำหรับตรวจสอบ DIF แฝง ยังมีปัญหา -การค้นพบ class membership ยังไม่ดีในทุกเงื่อนไขการทดสอบ -การประมาณค่าพารามิเตอร์อำนาจจำแนกให้ผลดีสำหรับข้อที่ไม่แปรเปลี่ยน เช่นเดียวกันกับข้อที่พบ DIF ในกรณีที่ไม่มีอิทธิพลของกลุ่ม -เมื่อมีอิทธิพลของกลุ่ม การจำแนกข้อที่พบ DIF จะลำเอียงทางบวก

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

4.2 งานวิจัยที่เกี่ยวข้องกับการศึกษาเปรียบเทียบประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค

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

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

นพมาศ พิพัฒนสุข (2541) ทำการศึกษาเปรียบเทียบประสิทธิภาพระหว่างวิธีแมนเทล-แฮนส์เซลกับวิธีถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อใช้เกณฑ์จับคู่เปรียบเทียบแตกต่างกัน โดยใช้แบบสอบความสามารถทางคณิตศาสตร์ซึ่งเป็นแบบสอบชนิดพหุมิติ กลุ่มตัวอย่างเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 จำนวน 1,076 คน ผลการวิจัย พบว่า วิธีแมนเทล-แฮนส์เซลมีประสิทธิภาพมากกว่าวิธีถดถอยโลจิสติก เมื่อใช้เกณฑ์จับคู่คะแนนรวม และมีประสิทธิภาพไม่แตกต่างกันเมื่อใช้เกณฑ์จับคู่คะแนนแบบสอบย่อย ส่วนวิธีถดถอยโลจิสติกเมื่อใช้เกณฑ์จับคู่คะแนนหลายแบบสอบย่อยมีความเหมาะสมในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบชนิดพหุมิติ

นิคม กีรติวรางกูร (2542) ทำการศึกษาเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีการวิเคราะห์องค์ประกอบจำกัด (RFA) วิธีแมนเทล-แฮนส์เซลและการตอบสนองข้อสอบ เมื่อขนาดกลุ่มตัวอย่าง ความยาวของแบบสอบ ค่าความยากของข้อสอบ ค่าอำนาจจำแนกของข้อสอบ และขนาดความลำเอียงของข้อสอบแตกต่างกัน โดยใช้ข้อมูลจำลอง ผลการวิจัย พบว่า โดยภาพรวมวิธี RFA มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูงที่สุด วิธีแมนเทล-แฮนส์เซลมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูง ภายใต้เงื่อนไขแบบสอบที่มีค่าความยากต่ำ ค่าอำนาจจำแนกสูง ที่ขนาดความยาวแบบสอบ 75 ข้อ เมื่อใช้กลุ่มตัวอย่างขนาด 1,000 คน และวิธี IRT แบบ 2 พารามิเตอร์ มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูง ภายใต้เงื่อนไขแบบสอบที่มีค่าความยากต่ำ ที่ขนาดความยาว แบบสอบ 75 ข้อ เมื่อใช้กลุ่มตัวอย่างขนาด 1,000 คน

วลีมาศ แซ่อึ้ง (2543) ทำการศึกษาเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก โดยศึกษาข้อมูลจำลองและจัดกระทำข้อมูลตามปัจจัย 4 ตัว ได้แก่ ลักษณะของข้อสอบ คือ ค่าความยากและค่าอำนาจจำแนก ความยาวของแบบสอบ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบ และขนาดกลุ่มตัวอย่าง ผลการศึกษา พบว่า วิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีอำนาจการ

ทดสอบเท่าเทียมกันภายใต้เกือบทุกเงื่อนไข นอกจากนี้ ทั้งสองวิธีมีอำนาจการทดสอบสูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลภายใต้เงื่อนไขเกือบทุกเงื่อนไข และทั้งสี่วิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ภายในเกณฑ์ ที่ระดับ 10% ภายใต้เกือบทุกเงื่อนไข

อิทธิฤทธิ์ พงษ์ปิยะรัตน์ (2551) ทำการศึกษาวิเคราะห์ข้อสอบและตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวิเคราะห์พหุระดับ กับแบบสอบความรู้วิชาคณิตศาสตร์ กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 1,588 คน ผู้บริหาร จำนวน 32 คน จากโรงเรียนในสังกัดสำนักงานเขตพื้นที่การศึกษาลพบุรีเขต 1 และ 2 ผลการวิจัย พบว่า ผลการประมาณค่าพารามิเตอร์ข้อสอบด้วยโมเดล HGLM-2L และ HGLM-3L มีความสัมพันธ์อย่างสัมบูรณ์กับผลการประมาณค่าด้วยโปรแกรม BILOG-MG ผลการประมาณค่าพารามิเตอร์ผู้สอบด้วยโมเดล HGLM-2L มีความสัมพันธ์อย่างสัมบูรณ์กับผลการประมาณค่าด้วยโปรแกรม BILOG-MG ส่วนโมเดล HGLM-3L มีระดับค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.793 ผลการวิเคราะห์ระดับนักเรียน และระดับโรงเรียน พบว่า ตัวแปรผลการเรียนวิชาคณิตศาสตร์ในภาคเรียนที่ผ่านมา ตัวแปรขนาดโรงเรียนและตัวแปรความเป็นผู้นำทางวิชาการของผู้บริหาร ส่งผลต่อค่าเฉลี่ยของโอกาสในการตอบข้อสอบได้ถูกต้องในแต่ละโรงเรียน ตามลำดับ และการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยโมเดล HGLM สามารถตรวจสอบข้อสอบที่พบ DIF ได้สอดคล้องกับผลการวิเคราะห์จากโปรแกรม BILOG-MG

ธเกียรติกมล ทองงอก (2554) ศึกษาเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 โดยการจำลองข้อมูลและข้อมูลเชิงประจักษ์ในวิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin และ Gierl กับเกณฑ์ Zumbo และ Thomas ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน ขนาดของการทำหน้าที่ต่างกัน จำนวนข้อสอบที่ทำหน้าที่ต่างกัน และความยาวของแบบสอบทั้งฉบับ ผลการวิจัย พบว่า วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin และ Gierl มีอัตราความถูกต้องสูงกว่าเกณฑ์ Zumbo และ Thomas ภายใต้เกือบทุกเงื่อนไข และการศึกษาในข้อมูลเชิงประจักษ์ พบว่า ขนาดอิทธิพลตามเกณฑ์ Jodoin และ Gierl มีอัตราความถูกต้องสูงกว่า และมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าเกณฑ์ Zumbo และ Thomas

Hidalgo and LÓPEZ-PINA (2004) ทำการศึกษาการทำหน้าที่ต่างกันของข้อสอบและขนาดของอิทธิพล โดยเปรียบเทียบระหว่างวิธีการวิเคราะห์การถดถอยโลจิสติก วิธีแมนเทล-แฮนส์เซล และวิธีแมนเทล-แฮนส์เซลแบบปรับปรุง เงื่อนไขของการศึกษา ได้แก่ ค่าความยากและค่าอำนาจจำแนกของข้อสอบ ขนาดของการทำหน้าที่ต่างกันของข้อสอบ และข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูปและแบบอเนกรูป โดยใช้การจำลองข้อมูล ผลการวิจัย พบว่า วิธีการวิเคราะห์การถดถอยโลจิสติกตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปที่สมมาตร ได้ดีกว่าวิธีแมนเทล-แฮนส์เซล และวิธีแมนเทล-แฮนส์เซลแบบปรับปรุง แต่เมื่อพิจารณาการวัดขนาดของอิทธิพลของการทำหน้าที่

ต่างกันของข้อสอบบนพื้นฐานการวิเคราะห์การถดถอยโลจิสติก ยังขาดความไวต่อการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีเงื่อนไขเฉพาะ

Kalaycioglu and Berberoglu (2010) ทำการศึกษาการทำหน้าที่ต่างกันของข้อสอบวิชาวิทยาศาสตร์และคณิตศาสตร์ สำหรับการสอบเข้าเรียนในมหาวิทยาลัยในประเทศตุรกี (University Entrance Examination: UEE) กลุ่มตัวอย่างเป็นนักเรียน 35,372 คน เป็นนักเรียนหญิง 11,368 คน และนักเรียนชาย 24,004 คน ใช้วิธีการตรวจสอบ DIF 4 วิธี ได้แก่ วิธีแมนเทล - แฮนส์เซล (MH) วิธีการถดถอยโลจิสติก (LR) วิธี restricted factor analysis (RFA) และวิธี IRT log likelihood ratio (IRT - LR) ผลการวิจัย พบว่า มีข้อสอบที่มี DIF ใน UEE แต่ข้อสอบที่มี DIF ไม่ปัญหามากนักต่อความตรงในเชิงปริมาณของ UEE แหล่งที่มาของ DIF มาจากปัจจัยที่เกี่ยวข้องกับเนื้อหาวิชาข้อสอบที่วัดทักษะการคิด และลักษณะรูปแบบของข้อสอบ ในระหว่างปัจจัยเหล่านี้ข้อสอบที่วัดทักษะการคิดมีประสิทธิผลในการเกิด DIF ระหว่างกลุ่มเพศ

Moses et al. (2010) ทำการศึกษาเปรียบเทียบกลยุทธ์สำหรับการประมาณค่าการทำหน้าที่ต่างกันของข้อสอบอย่างมีเงื่อนไขกับแบบสอบผลสัมฤทธิ์วิชาวิทยาศาสตร์ และวิชาประวัติศาสตร์ โดยใช้การจำลองข้อมูลจริง จำนวน 52,896 คน และ 325,250 คน ตามลำดับ ใช้วิธีการตรวจสอบการทำหน้าที่ต่างกัน 4 วิธี ได้แก่ raw data, การวิเคราะห์การถดถอยโลจิสติก, log-linear models และ kernel smoothing ผลการวิจัย พบว่า วิธีการวิเคราะห์การถดถอยโลจิสติกเป็นกลยุทธ์ที่ใช้ได้ดีมากที่สุดในแง่ของความลำเอียงและความผันแปรของการประมาณค่า วิธี log-linear models มีข้อได้เปรียบเรื่องความยืดหยุ่น ซึ่งเป็นการทดแทนความผันแปรของการประมาณค่าที่มากขึ้นเมื่อกลุ่มตัวอย่างมีขนาดใหญ่ขึ้น วิธี Kernel smoothing มีความถูกต้องน้อยที่สุด เนื่องมาจากปัญหาการประมาณค่าเมื่อกลุ่มอ้างอิงและกลุ่มเปรียบเทียบมีความแตกต่างในค่าความสามารถโดยรวม

Vaughn and Wang (2010) ทำการศึกษาการใช้ classification trees เพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยเปรียบเทียบความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบกับวิธีแมนเทล-แฮนส์เซล และวิธีการวิเคราะห์การถดถอยโลจิสติก เครื่องมือเป็นข้อสอบที่ให้คะแนนแบบ 2 ค่า จำนวน 40 ข้อ โดยใช้การจำลองข้อมูล ปัจจัยที่นำมาพิจารณาในการศึกษามี 2 องค์ประกอบ ได้แก่ ขนาดกลุ่มตัวอย่างในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ และเงื่อนไขของการแจกแจงค่าความสามารถ ผลการวิจัย พบว่า classification trees เป็นวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่เป็นทางเลือกใหม่ในวิธีการแบบเดิม ค่าความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบเทียบเท่ากับวิธีแมนเทล-แฮนส์เซล และวิธีการวิเคราะห์การถดถอยโลจิสติก

ตารางที่ 7 การศึกษาเปรียบเทียบประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค

ผู้วิจัย	ปี	เครื่องมือ/ข้อมูลที่ใช้วิเคราะห์	วิธีตรวจสอบ	ปัจจัยที่นำมาศึกษา	ข้อค้นพบ
จิติมา วรรณศรี	2539	ข้อมูลจำลองจากโปรแกรม IRTDATA	MH กับ SIBTEST	-ความยาวของแบบสอบ -ขนาดกลุ่มตัวอย่าง	วิธี MH กับวิธี SIBTEST มีประสิทธิภาพเท่าเทียมกันทุกเงื่อนไขที่ศึกษา
รัชรินทร์ มุคดา	2540	ข้อมูลจำลองจากโปรแกรม IRTDATA	MH กับ LR	-DIF แบบอเนกรูป -กลุ่มความสามารถ -ค่าความยาก -ค่าอำนาจจำแนก	วิธี MH กับวิธี LR มีประสิทธิภาพเท่ากัน ทุกกลุ่มความสามารถของผู้สอบ
นพมาศ พิพัฒนสุข	2541	แบบสอบความสามารถทางคณิตศาสตร์ชั้นประถมศึกษาปีที่ 6	MH กับ LR	เกณฑ์จับคู่เปรียบเทียบ	-วิธี MH มีประสิทธิภาพมากกว่า วิธี LR เมื่อใช้เกณฑ์จับคู่คะแนนรวม -วิธี LR มีความเหมาะสมในการตรวจสอบ DIF เมื่อใช้เกณฑ์จับคู่คะแนนหลายแบบสอบย่อย
นิคม กีรติวรางกูร	2542	ข้อมูลจำลอง	RFA, MH และ IRT	-ขนาดกลุ่มตัวอย่าง -ความยาวของแบบสอบ -ค่าความยาก -ค่าอำนาจจำแนก -ขนาดความลำเอียงของข้อสอบ	โดยภาพรวมวิธี RFA มีประสิทธิภาพสูงที่สุด
วลีมาศ แซ่อึ้ง	2543	ข้อมูลจำลอง	SIBTEST ปรับใหม่, SIBTEST, MH และ LR	-ค่าความยากและค่าอำนาจจำแนก -ความยาวของแบบสอบ -สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน -ขนาดกลุ่มตัวอย่าง	-วิธี SIBTEST ปรับใหม่ และวิธี LR มีอำนาจการทดสอบเท่าเทียมกัน และสูงกว่าอีกสองวิธี -ทั้งสี่วิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ภายในเกณฑ์
อิทธิฤทธิ์ พงษ์ปิยะรัตน์	2551	แบบสอบความรู้วิชาคณิตศาสตร์ ชั้น ม.3	โมเดล HGLM และโปรแกรม BILOG-MG	-โมเดล HGLM -โปรแกรม BILOG-MG	โมเดล HGLM ตรวจสอบข้อสอบที่พบ DIF สอดคล้องกับผลการวิเคราะห์จากโปรแกรม BILOG-MG

ตารางที่ 7 (ต่อ)

ผู้วิจัย	ปี	เครื่องมือ/ข้อมูลที่ใช้วิเคราะห์	วิธีตรวจสอบ	ปัจจัยที่นำมาศึกษา	ข้อค้นพบ
ธเกียรติกมล ทองงอก	2554	ข้อมูลจำลอง, โครงการสอบระดับชาติ ชั้น ป.6	LR	-เกณฑ์การวัดขนาดอิทธิพล -รูปแบบของ DIF -ขนาดของ DIF -จำนวนของข้อสอบที่ DIF -ความยาวของแบบสอบ	-การวัดขนาดอิทธิพลตามเกณฑ์ Jodoin และ Gierl มีอัตราความถูกต้องสูงกว่าเกณฑ์ Zumbo และ Thomas -ข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปมีอัตราความถูกต้องสูงกว่าแบบเอกรูป ทั้ง 2 เกณฑ์ -ขนาดของ DIF ที่เพิ่มขึ้น มีผลต่ออัตราความถูกต้องที่เพิ่มขึ้น -จำนวนของข้อสอบที่ DIF ร้อยละ 20 มีอัตราความถูกต้องสูงกว่าร้อยละ 10 ทั้ง 2 เกณฑ์ -การศึกษาในข้อมูลเชิงประจักษ์ให้ผลสอดคล้องกับข้อมูลจำลอง
Hidalgo และ LÓPEZ-PINA	2004	ข้อมูลจำลอง	LR, MH และ Modified MH	-ค่าความยาก, อำนาจจำแนก -ขนาดของ DIF -ประเภทของ DIF	วิธี LR ตรวจสอบ uniform DIF ได้ดีกว่าอีก 2 วิธี แต่ยังขาดความไวในการตรวจสอบ DIF ที่มีเงื่อนไขเฉพาะ
Kalaycıoglu และ Berberoglu	2010	University Entrance Examination: UEE	MH, LR, RFA และ IRT - LR	-เพศ -แหล่งที่มาของ DIF	-มีข้อสอบที่พบ DIF ใน UEE -แหล่งที่มาของ DIF มาจากปัจจัยที่เกี่ยวข้องกับเนื้อหาวิชา ข้อสอบที่วัดทักษะการคิด และลักษณะรูปแบบของข้อสอบ -ข้อสอบที่วัดทักษะการคิดมีประสิทธิผลในการเกิด DIF ระหว่างกลุ่มเพศ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 7 (ต่อ)

ผู้วิจัย	ปี	เครื่องมือ/ข้อมูลที่ใช้วิเคราะห์	วิธีตรวจสอบ	ปัจจัยที่นำมาศึกษา	ข้อค้นพบ
Moses, Miao, และ Dorans	2010	แบบสอบผลสัมฤทธิ์วิชาวิทยาศาสตร์ และวิชาประวัติศาสตร์	raw data, LR, log-linear models และ kernel smoothing	ขนาดกลุ่มตัวอย่าง	-วิธี LR เป็นกลยุทธ์ที่ใช้ได้ดีมากที่สุดในแง่ของความลำเอียงและความผันแปรของการประมาณค่า -วิธี log-linear models มีข้อได้เปรียบเรื่องความยืดหยุ่น ซึ่งเป็นการทดแทนความผันแปรของการประมาณค่าที่มากขึ้นเมื่อกลุ่มตัวอย่างมีขนาดใหญ่ขึ้น -วิธี Kernel smoothing มีความถูกต้องน้อยที่สุดเนื่องมาจากปัญหาการประมาณค่าเมื่อกลุ่ม R และกลุ่ม F มีความแตกต่างในค่าความสามารถโดยรวม
Vaughn และ Wang	2010	ข้อมูลจำลอง	MH, LR และ Classification trees	-ขนาดกลุ่มตัวอย่างในกลุ่ม R และกลุ่ม F -เงื่อนไขของการแจกแจงค่าความสามารถ	Classification trees เป็นวิธีการตรวจสอบDIF ที่มีค่าความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบเทียบเท่ากับวิธี MH และวิธี LR

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

4.3 สรุปประเด็นการศึกษาการทำหน้าที่ต่างกันของข้อสอบในแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค

4.3.1 การศึกษาปัจจัยที่มีผลต่อการทำหน้าที่ต่างกันของข้อสอบ

จากการสังเคราะห์งานวิจัยที่เกี่ยวข้องกับการกับการศึกษาการทำหน้าที่ต่างกันของข้อสอบในแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค จะพบการศึกษาผลการเกิดการทำหน้าที่ต่างกันของข้อสอบที่จำแนกตามปัจจัยต่าง ๆ ได้แก่ เพศ เชื้อชาติ ความยากของข้อสอบ การแจกแจงค่าความสามารถของผู้สอบ ขนาดของกลุ่มตัวอย่าง ความยาวของแบบสอบ จำนวนข้อสอบที่พบ DIF โดยมีงานวิจัยที่เกี่ยวข้อง เช่น เกษร หว่างจิตร (2539) รักชนก ยี่สุ่นศรี (2544) สุมาลี แกล้วทนงค์ (2547) Sheppard et al. (2006) French and Maller (2007) Scherbaum and Goldstein (2008) Woods (2008) DeMars (2009) Shin and Wang (2009) Woods and Grimm (2011) และ DeMars and Lau (2011) เป็นต้น

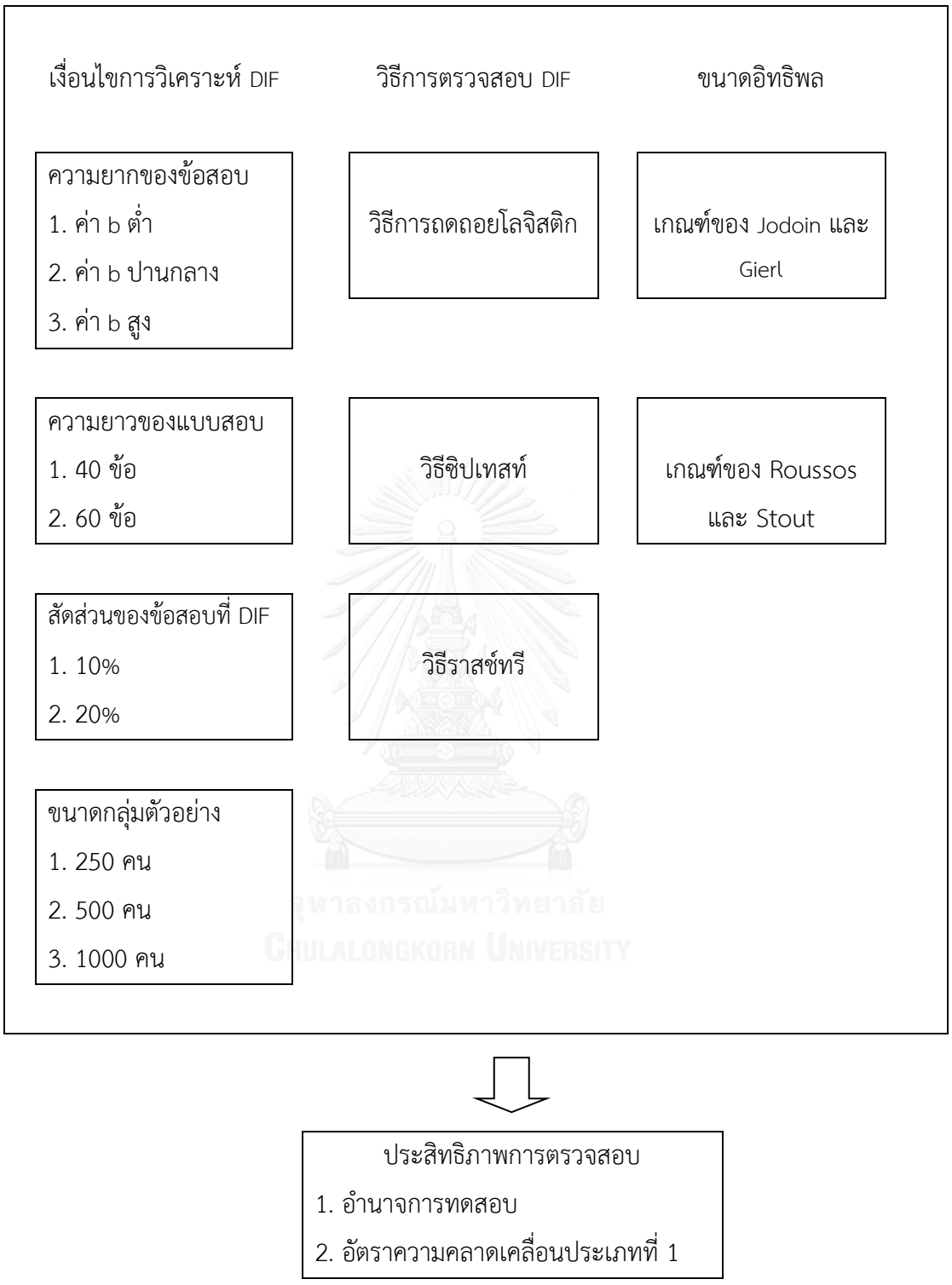
4.3.2 การศึกษาเปรียบเทียบประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

จากการสังเคราะห์งานวิจัยที่เกี่ยวข้องกับการกับการศึกษาการทำหน้าที่ต่างกันของข้อสอบในแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค จะพบการศึกษาเปรียบเทียบประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีที่วิเคราะห์ตามทฤษฎีแบบเดิม (CTT) และวิธีที่วิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (IRT) ซึ่งวิธีการตรวจสอบที่นิยมนำมาเปรียบเทียบกัน ได้แก่ วิธี Mantel-Haenzel และวิธี logistic regression โดยเปรียบเทียบประสิทธิภาพในด้านอำนาจการทดสอบ (power rate) และความคลาดเคลื่อนประเภทที่ 1 (type I error rate) ซึ่งมีงานวิจัยที่เกี่ยวข้อง เช่น จิติมา วรรณศรี (2539) รัชรินทร์ มุคดา (2540) นพมาศ พิพัฒนสุข (2541) นิคม กีรติวรางกูร (2542) วลีมาศ แซ่อึ้ง (2543) อิทธิฤทธิ์ พงษ์ปิยะรัตน์ (2551) ธเกียรติกมล ทองงอก (2554) Hidalgo and LÓPEZ-PINA (2004) Kalaycioglu and Berberoglu (2010) Moses et al. (2010) และ Vaughn and Wang (2010) เป็นต้น

4.4 กรอบแนวคิดในการวิจัย

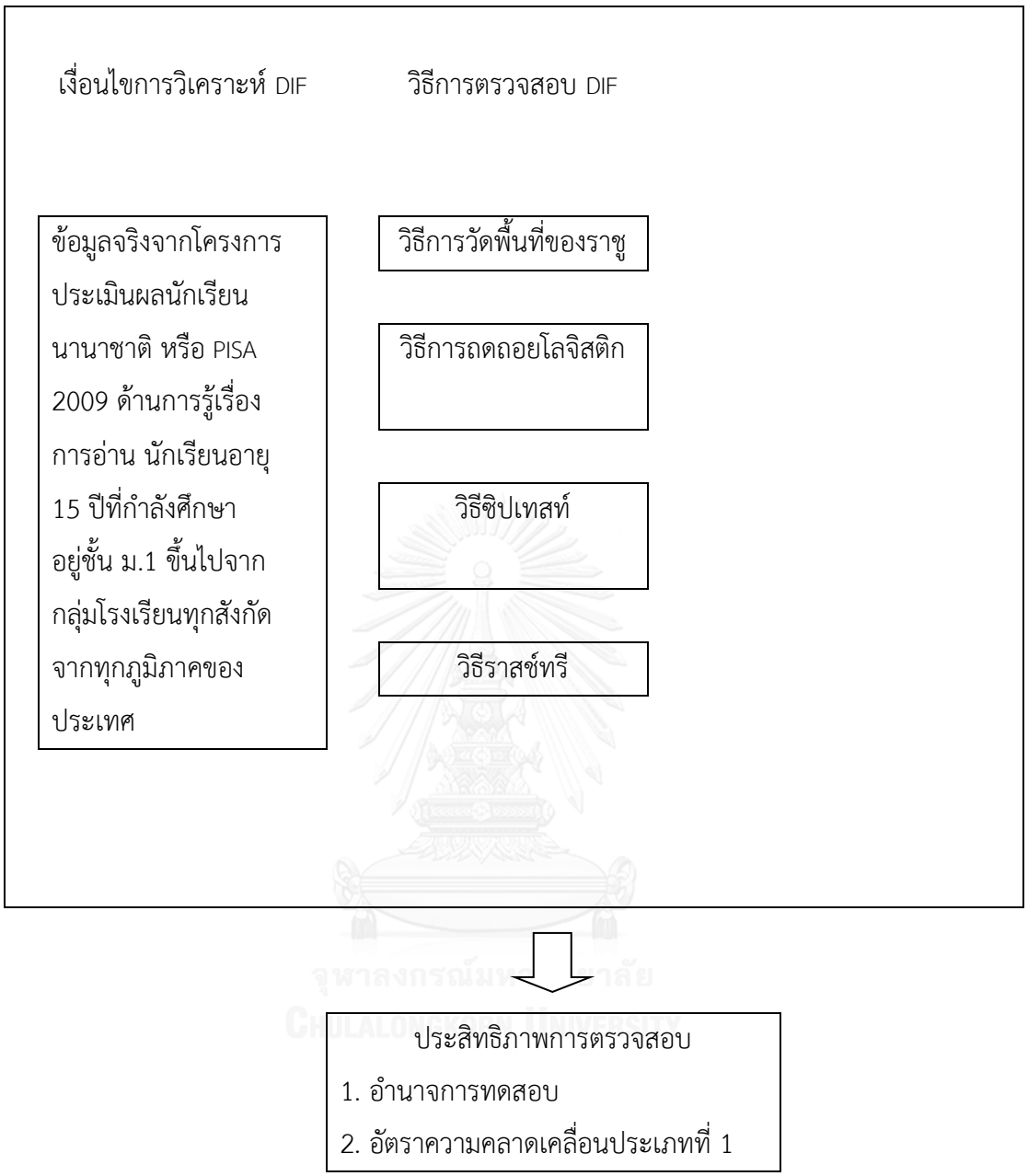
การศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบทำให้สามารถสรุปกรอบแนวคิดในการวิจัยได้ดังภาพที่ 10 และภาพที่ 11

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



ภาพที่ 10 กรอบแนวคิดในการวิจัย กรณีศึกษาการจำลองข้อมูล

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



ภาพที่ 11 กรอบแนวคิดในการวิจัย กรณีศึกษาข้อมูลเชิงประจักษ์

บทที่ 3
วิธีดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการวิจัยเพื่อเปรียบเทียบประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี โดยใช้การจำลองข้อมูลภายใต้ทฤษฎีการตอบสนองข้อสอบ โมเดล 1 พารามิเตอร์ กำหนดการทำหน้าที่ต่างกันแบบเอกรูป แล้วจัดกระทำข้อมูลตามปัจจัยที่ศึกษา 4 ปัจจัย คือ ความยากของข้อสอบ 3 ระดับ ความยาวของแบบสอบ 2 ขนาด สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด และขนาดกลุ่มตัวอย่าง 3 ขนาด มีข้อมูลที่ศึกษาทั้งหมด 36 เงื่อนไข (3 ระดับ x 2 ขนาดx 2 ขนาด x 3 ขนาด) ต่อจากนั้นจึงนำข้อมูลดังกล่าวมาวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีที่ศึกษา 3 วิธี โดยการทดสอบระดับนัยสำคัญ และการวัดขนาดอิทธิพลของวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ แล้วจึงเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 แยกตามการทดสอบระดับนัยสำคัญและการวัดขนาดอิทธิพล ต่อจากนั้นจึงศึกษากับข้อมูลเชิงประจักษ์จากโครงการประเมินผลนักเรียนนานาชาติ หรือ PISA 2009 ด้านการรู้เรื่องการอ่าน (reading literacy) ศึกษาการทำหน้าที่ต่างกันแบบเอกรูปและอเนกรูป คัดเลือกเฉพาะข้อสอบที่มีรูปแบบการตรวจให้คะแนนแบบทวิวิภาค กลุ่มตัวอย่างเป็นนักเรียนอายุ 15 ปี ที่กำลังศึกษาอยู่ในชั้นมัธยมศึกษาปีที่ 1 ขึ้นไป จากกลุ่มโรงเรียนทุกสังกัด ครอบคลุมทุกพื้นที่ ทั่วประเทศ สำหรับวิธีดำเนินการวิจัยแบ่งออกเป็น 6 ขั้นตอน ดังนี้

ตอนที่ 1 การกำหนดแนวทางในการวิเคราะห์ข้อมูล

ตอนที่ 2 การจัดกระทำข้อมูลตามปัจจัยที่ศึกษาและการจำลองข้อมูล

ตอนที่ 3 ลักษณะข้อมูลเชิงประจักษ์

ตอนที่ 4 การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ

ตอนที่ 5 การวิเคราะห์อำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1

ตอนที่ 6 การเปรียบเทียบประสิทธิภาพของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

ตอนที่ 1 การกำหนดแนวทางในการวิเคราะห์ข้อมูล

การจำลองข้อมูล (simulation data) ใช้วิธีการทางคณิตศาสตร์ เพื่อกำหนดสถานการณ์ตามเงื่อนไขที่ซับซ้อน ทำให้ศึกษาการทำหน้าที่ต่างกันของข้อสอบได้หลายเงื่อนไข ในขณะที่ข้อมูลโดยทั่วไปไม่สามารถดำเนินการได้อย่างครบถ้วนและสมบูรณ์ (Harwell, Stone, Hsu, & Kirisci,

1996) ข้อมูลเชิงประจักษ์ในการศึกษาครั้งนี้ เป็นผลการประเมินนักเรียนนานาชาติหรือ PISA 2009 ด้านการรู้เรื่องการอ่าน (reading literacy) คัดเลือกมาเฉพาะข้อสอบที่มีรูปแบบการตรวจให้คะแนนแบบทวิวิภาค กลุ่มตัวอย่างเป็นนักเรียนอายุ 15 ปี ที่กำลังศึกษาอยู่ในชั้นมัธยมศึกษาปีที่ 1 ขึ้นไป จากกลุ่มโรงเรียนทุกสังกัด ครอบคลุมทุกพื้นที่ทั่วประเทศ วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบตามสภาพที่เกิดขึ้นจริง ใช้วิธีการวัดพื้นที่ของราชู เป็นวิธีเกณฑ์ในการตรวจสอบ ถ้าวิธีตรวจสอบวิธีใดตรวจพบข้อสอบที่ทำหน้าที่ต่างกันสอดคล้องกับวิธีเกณฑ์ ถือว่า วิธีนั้นตรวจสอบการทำหน้าที่ต่างกันได้ถูกต้อง ข้อจำกัดของการศึกษาข้อมูลเชิงประจักษ์ คือ ผู้วิจัยไม่สามารถจัดกระทำกับข้อมูลให้เกิดเงื่อนไขต่าง ๆ เหมือนกับเงื่อนไขที่ศึกษาโดยการจำลองข้อมูล แนวทางในการวิเคราะห์ข้อมูลมีการดำเนินการ ดังนี้

1. ศึกษาค้นคว้าเกี่ยวกับมโนทัศน์ของการทำหน้าที่ต่างกันของข้อสอบ ขนาดอิทธิพลในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ และงานวิจัยที่เกี่ยวข้องจากเอกสาร วารสารวิชาการและงานวิจัย

2. ศึกษาโปรแกรมในการวิเคราะห์ข้อมูล ได้แก่ โปรแกรม WinGen เพื่อใช้จำลองข้อมูล โปรแกรม BILOG-MG เพื่อใช้ตรวจสอบค่าพารามิเตอร์ของข้อสอบ โปรแกรม SPSS โปรแกรม SIBTEST และโปรแกรม R package psychotree และ package difR เพื่อใช้วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบทั้ง 4 วิธี

3. กำหนดแนวทางในการวิเคราะห์ข้อมูล ดังนี้

3.1 ข้อมูลจำลอง

1) การจำลองข้อมูล ภายใต้ทฤษฎีการตอบสนองข้อสอบ โมเดล 1 พารามิเตอร์ จำลองการตอบข้อสอบที่มีโครงสร้างวัดความสามารถเอกมิติที่ให้คะแนนแบบทวิวิภาค กำหนดการทำหน้าที่ต่างกันแบบเอกรูป ผลการตอบข้อสอบภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ ความยากของข้อสอบ 3 ระดับ คือ ค่า b ต่ำ ค่า b ปานกลาง และค่า b สูง ความยาวของแบบสอบ 2 ขนาด คือ 40 ข้อ และ 60 ข้อ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด คือ ร้อยละ 10 และร้อยละ 20 และขนาดกลุ่มตัวอย่าง 3 ขนาด คือ 250, 500 และ 1,000 คน มีข้อมูลที่ศึกษาทั้งหมด 36 เงื่อนไข (3 ระดับ x 2 ขนาดx 2 ขนาด x 3 ขนาด) ทุกเงื่อนไขจำลองข้อมูลซ้ำ 50 ครั้ง

2) ตรวจสอบข้อมูลที่จำลองขึ้นว่าเป็นไปตามเงื่อนไขของปัจจัยที่แปรเปลี่ยนหรือไม่ พร้อมกับวิเคราะห์คุณภาพเบื้องต้นของข้อมูล

3) วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญ ด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี

4) วิเคราะห์ขนาดอิทธิพล (measure of effect size) จากข้อสอบที่ทำหน้าที่ต่างกันของวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์

5) คำนวณอำนาจการทดสอบ (power rate) และอัตราความคลาดเคลื่อนประเภทที่ 1 (type I error rate) จากการทดสอบระดับนัยสำคัญและการวัดขนาดอิทธิพล

6) เปรียบเทียบประสิทธิภาพของการระบุการทำหน้าที่ต่างกันของข้อสอบจากการตรวจสอบระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี พิจารณาจากอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1

7) สรุปผลการวิเคราะห์

3.2 ข้อมูลเชิงประจักษ์

1) เตรียมข้อมูลเชิงประจักษ์ พิจารณาความสมบูรณ์ครบถ้วนของข้อมูล จากผลการประเมินนักเรียนนานาชาติหรือ PISA 2009 ด้านการรู้เรื่องการอ่าน (reading literacy) คัดเลือกมาเฉพาะข้อสอบที่มีรูปแบบการตรวจให้คะแนนแบบทวิวิภาค กลุ่มตัวอย่างเป็นนักเรียนอายุ 15 ปี ที่กำลังศึกษาอยู่ในชั้นมัธยมศึกษาปีที่ 1 ขึ้นไป จากกลุ่มโรงเรียนทุกสังกัด ครอบคลุมทุกพื้นที่ทั่วประเทศ

2) วิเคราะห์สถิติพื้นฐาน และค่าพารามิเตอร์ข้อสอบ

3) วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีราสช์ทรี เพื่อให้ทราบว่ากลุ่มอ้างอิงและกลุ่มเปรียบเทียบจะแบ่งที่ประเภทใดของตัวแปรแต่ละตัว ต่อจากนั้นจึงวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวัดพื้นที่ของราชู เพื่อให้ทราบข้อสอบที่ทำหน้าที่ต่างกัน

4) วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี

5) คำนวณอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1

6) เปรียบเทียบประสิทธิภาพของการระบุการทำหน้าที่ต่างกันของข้อสอบจากการตรวจสอบระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี พิจารณาจากอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1

7) สรุปผลการวิเคราะห์

4. สรุปผลการวิเคราะห์ และอภิปรายผล

5. เขียนรายงานการวิจัย

ตอนที่ 2 การจัดกระทำข้อมูลตามปัจจัยที่ศึกษาและการจำลองข้อมูล

ผู้วิจัยจัดกระทำข้อมูลตามปัจจัยที่ศึกษา 4 ปัจจัย ดังนี้

1. ความยากของข้อสอบ 3 ระดับ

1) ค่า b ต่ำ $(\bar{b} = -1.34)$

2) ค่า b ปานกลาง $(\bar{b} = 0.00)$

3) ค่า b สูง $(\bar{b} = 1.34)$

2. ความยาวของแบบสอบ 2 ขนาด

1) แบบสอบที่มีความยาว 40 ข้อ

2) แบบสอบที่มีความยาว 60 ข้อ

3. สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด

1) ร้อยละ 10

2) ร้อยละ 20

4. ขนาดกลุ่มตัวอย่างประกอบด้วยกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีจำนวนกลุ่มตัวอย่างในแต่ละกลุ่มเท่ากันมี 3 ขนาด

1) จำนวน 250 คน

2) จำนวน 500 คน

3) จำนวน 1000 คน

ผู้วิจัยศึกษาในสถานการณ์จำลองโดยใช้โปรแกรม WinGen (Windows software that generates IRT parameters and item responses) โดย Han (2007) จำลองข้อมูลภายใต้ทฤษฎีการตอบสนองข้อสอบ โมเดล 1 พารามิเตอร์ จำลองการตอบแบบสอบที่มีโครงสร้างวัดความสามารถเอกมิติ ข้อสอบทุกข้อมีรายการตอบ คือ 0 และ 1

โปรแกรม WinGen มีลักษณะที่สำคัญ ดังนี้

1. สามารถสร้างเซตของค่าพารามิเตอร์ข้อสอบ (item parameters) และเซตของค่าพารามิเตอร์ผู้สอบ (examinee ability parameters) เพื่อสร้างข้อมูลการตอบสนองข้อสอบตามการแจกแจงได้หลายชนิด

2. ง่ายต่อการใช้งานและการเข้าถึงรายละเอียดของโปรแกรม การใช้งานบนโปรแกรม Windows มีแนวทางสำหรับการจำลองข้อมูล คือ การสร้างหรืออ่านในค่าพารามิเตอร์ความสามารถของผู้สอบ การสร้างหรืออ่านค่าพารามิเตอร์ข้อสอบ และการจำลองข้อมูลการตอบสนองข้อสอบ

3. ได้รับการพัฒนาบน Microsoft.NET frameworks 2.0 คอมพิวเตอร์ที่ได้รับการพัฒนาซอฟต์แวร์ล่าสุด โปรแกรมสามารถทำงานบน 32 bit ชุด Windows เช่น Windows XP หรือ 64 bit ชุด Windows Vista และมีการเพิ่มประสิทธิภาพการใช้งานระบบบแหล่งข้อมูลได้อย่างมีประสิทธิภาพ เพื่อให้โปรแกรมสามารถจัดการกับข้อมูลที่มีขนาดใหญ่โดยใช้เวลาน้อย สามารถใช้งานซอฟต์แวร์นี้ได้อย่างง่ายและได้รับการแปลงเป็นซอฟต์แวร์บนเว็บไซต์

4. ส่งเสริมให้เครื่องมือวิจัยมีประสิทธิภาพ เพื่อการวิจัยต่าง ๆ ในโปรแกรม WinGen ข้อมูลการจำลองแบบสามารถจำลองได้ถึง 1,000,000 ชุด

การจำลองข้อมูลด้วยโปรแกรม WinGen ประกอบด้วย 3 ส่วน ได้แก่

1. การจำลองพารามิเตอร์ผู้สอบ แฟ้มข้อมูลมีนามสกุลเป็น *.wge (WinGen data file for examinees) ได้แฟ้มข้อมูลจำนวน 36 แฟ้ม

2. การจำลองพารามิเตอร์ข้อสอบ แฟ้มข้อมูลมีนามสกุลเป็น *.wgi (WinGen data file for item parameters) ได้แฟ้มข้อมูลจำนวน 36 แฟ้ม

3. การจำลองข้อมูลการตอบ แฟ้มข้อมูลมีนามสกุลเป็น *.wgr (WinGen data file for generated responses) ได้แฟ้มข้อมูลจำนวน 3,600 แฟ้ม

ในการกำหนดเงื่อนไขของการจำลองข้อมูลโดยใช้โปรแกรม WinGen เป็นไปตามเงื่อนไขของการจัดกระทำข้อมูลที่ได้กำหนดไว้ข้างต้น ตามขั้นตอนดังนี้

1. กำหนดจำนวนผู้สอบเท่ากับ 250 คน 500 คน และ 1,000 คน ความสามารถของผู้สอบมีการแจกแจงแบบปกติ (Mean เท่ากับ 0 และ SD เท่ากับ 1)

2. กำหนดความยาวของแบบสอบทั้งฉบับ คือ 40 ข้อ และ 60 ข้อ รายการคำตอบเป็น 2 ค่า คือ 0 และ 1 โมเดล 1 พารามิเตอร์ และความยากของข้อสอบ คือ ค่า b ต่ำ $(\bar{b} = -1.34)$ ค่า b ปานกลาง $(\bar{b} = 0.00)$ และค่า b สูง $(\bar{b} = 1.34)$ ส่วน SD เท่ากับ 0.222 ซึ่งจะมีผลทำให้ค่าพารามิเตอร์ของข้อสอบที่ได้จากการจำลองข้อมูลมีค่าอยู่ในช่วงที่ต้องการศึกษาโดยใช้สูตรดังนี้

$$SD = \frac{x_{max} - x_{min}}{6}$$

โดยที่ $x_{\max}$ แทน ค่าพารามิเตอร์ของข้อสอบที่มีค่าสูงสุด

$x_{\min}$ แทน ค่าพารามิเตอร์ของข้อสอบที่มีค่าต่ำสุด

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

3. กำหนดจำนวนข้อสอบที่ทำหน้าที่ต่างกัน คือ ร้อยละ 10 และร้อยละ 20 ซึ่งในกรณีข้อสอบ 40 ข้อ จะมีจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 4 ข้อ (กำหนดข้อที่ 1 ถึงข้อที่ 4) และ 8 ข้อ (กำหนดข้อที่ 1 ถึงข้อที่ 8) ส่วนในกรณีข้อสอบ 60 ข้อ จะมีจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 6 ข้อ (กำหนดข้อที่ 1 ถึงข้อที่ 6) และ 12 ข้อ (กำหนดข้อที่ 1 ถึงข้อที่ 12)

4. กำหนดค่าความยากของข้อสอบที่ทำหน้าที่ต่างกันในข้อ 3 ให้มีค่าแปรเปลี่ยนไป คือ ข้อสอบครึ่งแรกในแต่ละเงื่อนไขมีการทำหน้าที่ต่างกันขนาด 0.2 ส่วนข้อสอบครึ่งหลังในแต่ละเงื่อนไขมีการทำหน้าที่ต่างกันขนาด -0.2 เพื่อให้ค่าความยากของข้อสอบทั้งฉบับมีค่าเท่าเดิมในแต่ละลักษณะ

5. ทำซ้ำ 50 รอบ ในแต่ละเงื่อนไข

ตารางที่ 8 แผนผังการจำลองข้อมูล

ความยากของข้อสอบ	ความยาวของแบบสอบ	สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน	ขนาดกลุ่มตัวอย่าง	เงื่อนไข
ค่า b ต่ำ	40 ข้อ	10%	250	เงื่อนไขที่ 1
			500	เงื่อนไขที่ 2
			1000	เงื่อนไขที่ 3
		20%	250	เงื่อนไขที่ 4
			500	เงื่อนไขที่ 5
			1000	เงื่อนไขที่ 6
	60 ข้อ	10%	250	เงื่อนไขที่ 7
			500	เงื่อนไขที่ 8
			1000	เงื่อนไขที่ 9
		20%	250	เงื่อนไขที่ 10
			500	เงื่อนไขที่ 11
			1000	เงื่อนไขที่ 12
ค่า b ปานกลาง	40 ข้อ	10%	250	เงื่อนไขที่ 13
			500	เงื่อนไขที่ 14
			1000	เงื่อนไขที่ 15
		20%	250	เงื่อนไขที่ 16
			500	เงื่อนไขที่ 17
			1000	เงื่อนไขที่ 18
	60 ข้อ	10%	250	เงื่อนไขที่ 19
			500	เงื่อนไขที่ 20
			1000	เงื่อนไขที่ 21
		20%	250	เงื่อนไขที่ 22
			500	เงื่อนไขที่ 23
			1000	เงื่อนไขที่ 24
ค่า b สูง	40 ข้อ	10%	250	เงื่อนไขที่ 25
			500	เงื่อนไขที่ 26
			1000	เงื่อนไขที่ 27
		20%	250	เงื่อนไขที่ 28
			500	เงื่อนไขที่ 29
			1000	เงื่อนไขที่ 30
	60 ข้อ	10%	250	เงื่อนไขที่ 31
			500	เงื่อนไขที่ 32
			1000	เงื่อนไขที่ 33
		20%	250	เงื่อนไขที่ 34
			500	เงื่อนไขที่ 35
			1000	เงื่อนไขที่ 36

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตอนที่ 3 ลักษณะข้อมูลเชิงประจักษ์

รายงานผลการประเมิน PISA 2009 การอ่าน คณิตศาสตร์ และวิทยาศาสตร์ บทสรุปเพื่อการบริหาร ได้วิเคราะห์และรายงานผลแยกตามกลุ่มโรงเรียน และภาคพื้นที่ ซึ่งกลุ่มโรงเรียนหรือสังกัดของโรงเรียนมี 8 กลุ่ม และภาคพื้นที่หรือภูมิภาคมี 9 กลุ่ม นอกจากตัวแปรสังกัดของโรงเรียนและภูมิภาคแล้ว ตัวแปรเพศก็ถือว่าเป็นตัวแปรที่ใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบอีกตัวหนึ่ง จากงานวิจัยที่ผ่านมา พบว่า ตัวแปรเพศเป็นตัวแปรที่ส่งผลต่อการทำหน้าที่ต่างกันของข้อสอบประกอบกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีราสช์ทรีสามารถตรวจสอบทั้งกลุ่มผู้สอบที่กำหนดไว้ล่วงหน้าและไม่ได้กำหนดไว้ ดังนั้น ในการวิจัยครั้งนี้ จึงศึกษาการทำหน้าที่ต่างกันของข้อสอบตามตัวแปร 3 ตัว ได้แก่ เพศ สังกัดของโรงเรียน และภูมิภาค โดย PISA 2009 มีจำนวนผู้สอบทั้งหมด 6,225 คน แต่เมื่อได้พิจารณาความสมบูรณ์ครบถ้วนของผลการตอบแล้ว ผู้วิจัยได้คัดเลือกผู้สอบเพื่อนำมาใช้ในการศึกษาการทำหน้าที่ต่างกันของข้อสอบ จำนวน 5,875 คน ซึ่งพิจารณาจากการตอบข้อสอบจำนวนร้อยละ 75 ของแบบสอบทั้งฉบับ รายละเอียดจำนวนผู้สอบแยกตามตัวแปรทั้ง 3 ตัว แสดงดังตารางที่ 9

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 9 จำนวนผู้สอบโครงการ PISA 2009 การรู้เรื่องการอ่านที่ใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ จำแนกตามตัวแปร 3 ตัว

ตัวแปร	กลุ่มตัวแปร		ฉบับ (booklet)													รวม
			1	2	3	4	5	6	7	8	9	10	11	12	13	จำนวน
เพศ	หญิง		272	237	255	282	243	241	244	209	254	279	287	265	248	3,316
	ชาย		196	221	209	191	223	204	192	184	184	184	162	205	204	2,559
	รวมจำนวน (คน)		**468**	**458**	**464**	**473**	**466**	**445**	**436**	**393**	**438**	**463**	**449**	**470**	**452**	**5,875**
สังกัด	สพฐ.1 (สปช.เดิม)		29	30	31	31	33	29	26	29	23	24	27	29	30	371
	สพฐ.2 (สศ.เดิม)		220	216	222	228	222	207	207	190	218	236	221	230	216	2,833
	สช.		29	29	27	27	31	25	27	25	27	31	31	32	29	370
	กทม.		36	35	38	37	37	34	29	25	34	33	34	36	32	440
	กศท.		37	34	39	40	36	38	35	28	35	33	31	36	35	457
	สาธิต		37	38	34	36	33	36	36	35	37	38	36	36	36	468
	อศ.1 (สช.อาชีว)		33	34	31	30	33	24	29	20	24	28	27	32	30	375
	อศ.2 (อาชีวรัฐ)		47	42	42	44	41	52	47	41	40	40	42	39	44	561
	รวมจำนวน (คน)		**468**	**458**	**464**	**473**	**466**	**445**	**436**	**393**	**438**	**463**	**449**	**470**	**452**	**5,875**

ตารางที่ 9 (ต่อ)

ตัวแปร	กลุ่มตัวแปร	ฉบับ (booklet)													รวม
		1	2	3	4	5	6	7	8	9	10	11	12	13	จำนวน
ภูมิภาค	กทม. และปริมณฑล	114	109	115	113	116	107	102	88	104	109	110	113	106	1,406
	กลาง	34	28	31	33	28	27	29	26	29	29	27	31	30	382
	เหนือบน	33	33	33	37	34	32	33	27	35	37	34	35	32	435
	เหนือล่าง	35	35	32	36	34	29	26	25	27	33	30	32	30	404
	อีสานบน	72	71	75	74	75	70	71	67	72	74	72	74	70	937
	อีสานล่าง	67	64	71	69	71	71	67	60	64	63	62	69	71	869
	ใต้	63	64	61	59	56	55	54	52	56	63	60	62	65	770
	ตะวันออก	18	21	16	21	19	19	19	15	17	20	20	20	22	247
	ตะวันตก	32	33	30	31	33	35	35	33	34	35	34	34	26	425
	รวมจำนวน (คน)	**468**	**458**	**464**	**473**	**466**	**445**	**436**	**393**	**438**	**463**	**449**	**470**	**452**	**5,875**

ตอนที่ 4 การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ

ผู้วิจัยวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 3 วิธี คือ วิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี

1. การวิเคราะห์ DIF ด้วยวิธีการถดถอยโลจิสติก

ในการวิเคราะห์ DIF ด้วยวิธีการถดถอยโลจิสติก ผู้วิจัยใช้โปรแกรม SPSS โมเดลสมการถดถอย (Swaminathan and Rogers, 1990; อ้างถึงใน ธเกียรติกมล ทองงอก, 2554) คือ

$$z = \beta_0 + \beta_1 x + \beta_2 G + \beta_3 XG$$

เมื่อ x แทน ระดับความสามารถของผู้สอบเป็นคะแนนสอบ

G แทน สมาชิกของกลุ่มผู้สอบ คือ กลุ่มอ้างอิงหรือกลุ่มเปรียบเทียบ

XG แทน ปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับกลุ่มผู้สอบ

β_0 แทน ค่า Intercept (จุดตัดแกน)

β_1 แทน พารามิเตอร์ที่เกี่ยวข้องกับความสามารถที่แตกต่าง

β_2 แทน พารามิเตอร์ที่เกี่ยวข้องกับความสามารถที่แตกต่างระหว่างกลุ่ม

β_3 แทน พารามิเตอร์ที่เกี่ยวข้องกับปฏิสัมพันธ์ระหว่างกลุ่มกับความสามารถ

ผู้วิจัยใช้คะแนนรวมของแบบสอบ (total test score) แทนระดับความสามารถของผู้สอบ ส่วนสมาชิกของกลุ่มผู้สอบจะกำหนดให้ G = 1 เมื่อผู้สอบอยู่ในกลุ่มอ้างอิง และ G = 2 เมื่อผู้สอบอยู่ในกลุ่มเปรียบเทียบ การตัดสินข้อสอบที่ทำหน้าที่ต่างกันจะพิจารณาค่าพารามิเตอร์ของ β_2 กับ β_3 จากโมเดลข้างต้น กล่าวคือ

ถ้า $\beta_2 = \beta_3 = 0$ แสดงว่า ข้อสอบทำหน้าที่ไม่ต่างกัน

ถ้า $\beta_2 \neq 0$ และ $\beta_3 = 0$ แสดงว่า ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป

ถ้า $\beta_3 \neq 0$ (ส่วน $\beta_2 = 0$ หรือไม่ก็ได้) แสดงว่า ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป

สำหรับสถิติทดสอบสมมติฐาน ใช้สถิติ χ^2 ที่ระดับชั้นความเป็นอิสระเท่ากับ 2 (df = 2) และผู้วิจัยใช้เกณฑ์ตัดสินข้อสอบที่ทำหน้าที่ต่างกันที่ระดับนัยสำคัญ .05

2. การวิเคราะห์ DIF ด้วยวิธีซิปเทสท์

ในการวิเคราะห์ DIF ใช้โปรแกรม SIBTEST ซึ่งมีสูตรในการคำนวณ (Stout and Roussos, 1992; อ้างถึงใน วลีมาศ แซ่อึ้ง, 2543) ดังนี้

$$\hat{\beta}_{uni} = \sum_{k=0}^{n} \hat{P}_k (\bar{Y}_{Rk} - \bar{Y}_{Fk})$$

เมื่อ $\hat{\beta}_{uni}$ แทน ดัชนีซิปเทสท์

$\bar{Y}_{Rk}$ แทน ค่าเฉลี่ยของคะแนนรวมจากชุดแบบสอบที่ต้องการศึกษาของผู้สอบกลุ่มอ้างอิง ซึ่งได้คะแนน X = k

$\bar{Y}_{Fk}$ แทน ค่าเฉลี่ยของคะแนนรวมจากชุดแบบสอบที่ต้องการศึกษาของผู้สอบกลุ่มเปรียบเทียบ ซึ่งได้คะแนน X = k

$\hat{P}_k$ แทน สัดส่วนของผู้สอบทั้งหมด (กลุ่มอ้างอิงและกลุ่มเปรียบเทียบ) ซึ่งตอบชุดแบบสอบที่มีความตรง โดยใช้คะแนนรวม X = k

k แทน คะแนนรวมจากชุดแบบสอบที่มีความตรง

นำค่า $\hat{\beta}_{uni}$ ดังกล่าวมาคำนวณสถิติ B_{uni} เพื่อนำไปทดสอบสมมติฐาน H_0 โดยนำไปเปรียบเทียบกับสถิติ Z ที่ระดับนัยสำคัญ .05 ดังนี้

$$B_{uni} = \frac{\hat{\beta}_{uni}}{\hat{\sigma}(\hat{\beta}_{uni})}$$

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

เกณฑ์การตัดสินข้อสอบทำหน้าที่ต่างกันที่ระดับนัยสำคัญ .05 ถ้าผลการทดสอบ พบว่า $B_{uni} > Z_{\alpha}$ แสดงว่า ปฏิเสธ H_0 นั่นคือ ข้อสอบทำหน้าที่ต่างกัน

3. การวิเคราะห์ DIF ด้วยวิธีราสช์ทรี

ผู้วิจัยใช้โปรแกรม R package psychotree version 0.12-2 ของ Zeileis et al. (2012 cited in Strobl, Kopf and Zeileis, 2015) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทำงาน 4 ขั้นตอน ดังนี้

1. การประมาณค่าพารามิเตอร์ของข้อสอบ ซึ่งมีสูตรในการคำนวณ (Strobl et al., 2015) ดังนี้

$$P(U_{ij} = u_{ij}|\theta_i, \beta_j) = \frac{e^{u_{ij}.(\theta_i - \beta_j)}}{1 + e^{\theta_i - \beta_j}}$$

$$L_c(\beta|r_1, \ldots, r_n) = \prod_{i=1}^{n} L_c(\beta|r_i) = \prod_{i=1}^{n} \frac{e^{-\sum_{j=1}^{m} u_{ji}.\beta_j}}{\gamma_{r_i}(\beta)}$$

2. การทดสอบความไม่คงที่ของค่าพารามิเตอร์ ซึ่งมีสูตรในการคำนวณ (Strobl et al., 2015) ดังนี้

$$\varphi(u_i, \beta)_j = \frac{\partial \Psi(u_i, \beta)}{\partial \beta_j} = -u_{ij} - \frac{1}{\gamma_{r_i}(\beta)}.\frac{\partial_{\gamma_{r_i}}(\beta)}{\partial \beta_j}$$

3. การเลือกจุดตัด (Selecting the cutpoints) ซึ่งมีสูตรในการคำนวณ (Strobl et al., 2015) ดังนี้

$$\sum_{i \in L(\xi)} \Psi\left(u_i \hat{\beta}^{(L)}\right) + \sum_{i \in R(\xi)} \Psi\left(u_i \hat{\beta}^{(R)}\right)$$

4. เกณฑ์ในการหยุดทำงาน (Stopping criteria)

เมื่อค่าความไม่คงที่ของพารามิเตอร์ข้อสอบไม่มีนัยสำคัญ โดยพิจารณาจากตัวแปรร่วมทุกตัว กำหนดที่ระดับนัยสำคัญ .05

การพิจารณาการเกิดการทำหน้าที่ต่างกันของข้อสอบของวิธีราสช์ทรี พิจารณาจากค่าพารามิเตอร์ความยากของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ กล่าวคือ ถ้าผลต่างของค่าพารามิเตอร์ความยากของทั้ง 2 กลุ่ม มีค่ามากกว่าหรือเท่ากับ 0.4000 หรือมีค่าน้อยกว่าหรือเท่ากับ -0.4000 แสดงว่า ข้อสอบข้อนั้นทำหน้าที่ต่างกัน

4. การวิเคราะห์ DIF ด้วยวิธีการวัดพื้นที่ของราชู

ผู้วิจัยใช้โปรแกรม R package difR version 4.6 ของ Magis, Beland, and Raiche (2015) ซึ่งมีสูตรในการคำนวณ ดังนี้

$$EUA_{11} = \left|\hat{b}_F - \hat{b}_R\right|$$

กำหนดให้ $X = \hat{b}_F - \hat{b}_R$ และ X มีข้อตกลงการแจกแจงแบบปกติ ดังนั้น ค่าเฉลี่ยและความแปรปรวนของ EUA_{11} สามารถคำนวณได้จากสูตร

$$\mu(EUA_{11}) = \mu(ESA_{10})[1 - 2\Phi(z_0)] + (2/\pi)^{1/2}\sigma(ESA_{10})exp(-z_0^2/2)$$

$$\sigma^2(EUA_{11}) = \sigma^2(ESA_{10}) + \mu^2(ESA_{10}) - \mu^2(EUA_{11})$$

โดยที่ $\Phi(z_0) = \int_{-\infty}^{z_0} g(z)dz$

$$z_0 = [0 - \mu(ESA_{10})]/\sigma(ESA_{10})$$

เมื่อ g(z) = ฟังก์ชันความหนาแน่นของความน่าจะเป็นแบบปกติ

ในการตัดสินข้อสอบที่ทำหน้าที่ต่างกัน ให้นำดัชนีการวัดพื้นที่มาทดสอบนัยสำคัญโดยใช้สถิติ Z ภายใต้สมมติฐานการแจกแจงแบบปกติ ดังนี้

$$Z_i(EUA) = \frac{\hat{b}_{iF} - \hat{b}_{iR}}{\sqrt{Var(\hat{b}_{iF}) + Var(\hat{b}_{iR})}}$$

ตอนที่ 5 การวิเคราะห์อำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1

ในการวิเคราะห์อำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ แบ่งเป็น 2 กรณี ดังนี้

1. กรณีข้อมูลจำลอง ผู้วิจัยนำผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี มาเปรียบเทียบกับจำนวนข้อสอบที่ทำหน้าที่ต่างกันในแต่ละเงื่อนไขการตรวจสอบ

2. กรณีข้อมูลเชิงประจักษ์ ผู้วิจัยนำผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี มาเปรียบเทียบกับวิธีการวัดพื้นที่ของราชู โดยถือว่าผลจากการวิเคราะห์ DIF ด้วยวิธีการวัดพื้นที่ของราชูเป็นเกณฑ์ที่ถูกต้อง กล่าวคือ ถ้าวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ระบุข้อสอบที่ทำหน้าที่ต่างกันได้ตรงกับข้อสอบที่ถูกระบุด้วยวิธีการวัดพื้นที่ของราชู แสดงว่าวิธีที่ศึกษา 3 วิธี สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ถูกต้อง แต่ถ้าวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ระบุข้อสอบที่ทำหน้าที่ต่างกันได้ไม่ตรงกับข้อสอบที่ถูกระบุด้วยวิธีการวัดพื้นที่ของราชู แสดงว่าวิธีที่ศึกษา 3 วิธี ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบไม่ถูกต้อง สำหรับสูตรในการคำนวณมีดังนี้

$$Power = \frac{n_1}{N_1} \times 100$$

$$E_1 = \frac{n_2}{N_2} \times 100$$

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

เมื่อ $Power$ แทน อำนาจการทดสอบ

E_1 แทน อัตราความคลาดเคลื่อนประเภทที่ 1

n_1 แทน จำนวนข้อสอบที่ระบุถูกต้องว่าทำหน้าที่ต่างกัน

n_2 แทน จำนวนข้อสอบที่ระบุผิดพลาดว่าทำหน้าที่ต่างกัน

N_1 แทน จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งหมดในแบบสอบ หรือจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งหมดในแบบสอบที่ตรวจสอบด้วยวิธีเกณฑ์

N_2 แทน จำนวนข้อสอบที่ทำหน้าที่ไม่ต่างกันทั้งหมดในแบบสอบหรือจำนวนข้อสอบที่ทำหน้าที่ไม่ต่างกันทั้งหมดในแบสอบที่ตรวจสอบด้วยวิธีเกณฑ์

ตอนที่ 6 การเปรียบเทียบประสิทธิภาพของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

1. การเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1

การเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้ปัจจัยที่ศึกษา 4 ปัจจัย คือ ความยากของข้อสอบ 3 ระดับ ความยาวของแบบสอบ 2 ขนาด สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด และขนาดกลุ่มตัวอย่าง 3 ขนาด ผู้วิจัยใช้สถิติวิเคราะห์ความแปรปรวนแบบสองทาง (two-way ANOVA) เพื่อทดสอบผลของวิธีการตรวจสอบ ผลของปัจจัยที่ศึกษา และผลของปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบกับปัจจัยที่ศึกษา ถ้าผลการทดสอบดังกล่าวมีนัยสำคัญทางสถิติ จะใช้การวิเคราะห์อิทธิพลอย่างง่าย (simple effect) ทดสอบในแต่ละระดับของตัวแปร ถ้าผลการทดสอบดังกล่าวมีนัยสำคัญทางสถิติจะเปรียบเทียบความแตกต่างเป็นรายคู่ โดยใช้วิธีการทดสอบของ Bonferroni กรณีค่าความแปรปรวนของทุกกลุ่มเท่ากัน และวิธีการทดสอบของ Games-Howell กรณีค่าความแปรปรวนของทุกกลุ่มไม่เท่ากัน โดยจะทดสอบใน 3 กรณี ดังนี้

กรณีที่ 1 ทดสอบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธีการตรวจสอบ ภายใต้ปัจจัยที่ศึกษา 4 ปัจจัย คือ ความยากของข้อสอบ 3 ระดับ ความยาวของแบบสอบ 2 ขนาด สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด และขนาดกลุ่มตัวอย่าง 3 ขนาด ที่ต่างระดับกัน

กรณีที่ 2 ทดสอบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 3 วิธี ภายใต้ปัจจัยที่ศึกษา 4 ปัจจัย คือ ความยากของข้อสอบ 3 ระดับ ความยาวของแบบสอบ 2 ขนาด สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาดและขนาดกลุ่มตัวอย่าง 3 ขนาด ที่ระดับเดียวกัน

กรณีที่ 3 ทดสอบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 3 วิธี ภายใต้ปัจจัยที่ศึกษา 4 ปัจจัย คือ ความยากของข้อสอบ 3 ระดับ ความยาวของแบบสอบ 2 ขนาด สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาดและขนาดกลุ่มตัวอย่าง 3 ขนาด ที่ต่างระดับกัน

2. กำหนดเกณฑ์ที่ใช้เปรียบเทียบประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ โดยการวัดขนาดอิทธิพล ดังนี้

2.1 เกณฑ์ของ Jodoin และ Gierl (2001) ซึ่งกำหนดเกณฑ์การตัดสินขนาดอิทธิพลของวิธีการถดถอยโลจิสติกใน 3 ระดับ ดังนี้

ค่าความแตกต่างของ R^2 มีค่าน้อยกว่า 0.035 แสดงว่า ขนาดอิทธิพลมีขนาดเล็ก ยอมรับสมมติฐานว่าง (Null hypothesis) ไม่เกิด DIF

ค่าความแตกต่างของ R^2 มีค่าระหว่าง 0.035 ถึง 0.069 แสดงว่า ขนาดอิทธิพลมีขนาดปานกลาง

ค่าความแตกต่างของ R^2 มีค่ามากกว่าหรือเท่ากับ 0.07 แสดงว่า ขนาดอิทธิพลมีขนาดใหญ่

2.2 เกณฑ์ของ Roussos and Stout (1996) ซึ่งกำหนดเกณฑ์การตัดสินขนาดอิทธิพลของวิธีซิปเทสท์ใน 3 ระดับ ดังนี้

ค่าสัมบูรณ์ของ $\hat{\beta}$ มีค่าน้อยกว่า 0.059 แสดงว่า ขนาดอิทธิพลมีขนาดเล็ก ปฏิเสธสมมติฐานว่าง (Null hypothesis) เกิด DIF

ค่าสัมบูรณ์ของ $\hat{\beta}$ มีค่าระหว่าง 0.059 ถึง 0.087 แสดงว่า ขนาดอิทธิพลมีขนาดปานกลาง

ค่าสัมบูรณ์ของ $\hat{\beta}$ มีค่ามากกว่าหรือเท่ากับ 0.088 แสดงว่า ขนาดอิทธิพลมีขนาดใหญ่

วิธีใดเกิดอำนาจการทดสอบในการตรวจสอบสูงและเกิดอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำ แสดงว่ามีประสิทธิภาพสูงสุดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 4
ผลการวิเคราะห์ข้อมูล

การศึกษาครั้งนี้ มีวัตถุประสงค์เพื่อเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี โดยการทดสอบระดับนัยสำคัญและการวัดขนาดอิทธิพล ในข้อมูลจำลอง ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ ความยากของข้อสอบ 3 ระดับ ความยาวของแบบสอบ 2 ขนาด สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด และขนาดกลุ่มตัวอย่าง 3 ขนาด สำหรับข้อมูลเชิงประจักษ์ ศึกษาการทำหน้าที่ต่างกันของข้อสอบตามตัวแปร 3 ตัว คือ เพศ สังกัดของโรงเรียน และภูมิภาค

ข้อมูลที่ใช้ในการศึกษาครั้งนี้เป็นข้อมูลจำลอง กำหนดการทำหน้าที่ต่างกันแบบเอกรูป ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย จำนวน 36 เงื่อนไข (3 ระดับ x 2 ขนาดx 2 ขนาด x 3 ขนาด) และในแต่ละเงื่อนไขจำลองซ้ำ 50 รอบ และข้อมูลเชิงประจักษ์จากโครงการประเมินผลนักเรียนนานาชาติ หรือ PISA 2009 ด้านการรู้เรื่องการอ่าน (reading literacy) ศึกษาการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและอเนกรูป นำข้อมูลดังกล่าวมาตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 3 วิธี คือ วิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี สำหรับผลการวิเคราะห์ข้อมูลที่จะกล่าวต่อไป จำแนกออกเป็น 3 ตอน ดังนี้

ตอนที่ 1 ผลการวิเคราะห์ค่าเฉลี่ยของอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1

ตอนที่ 2 ผลการเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1

ตอนที่ 3 ผลการเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ในข้อมูลเชิงประจักษ์

ในการนำเสนอผลการวิเคราะห์ข้อมูล เพื่อความสะดวก ผู้วิจัยจึงกำหนดสัญลักษณ์ ดังนี้

DIF	หมายถึง	การทำหน้าที่ต่างกันของข้อสอบ
DI	หมายถึง	ความยากของข้อสอบ
b_L	หมายถึง	ค่าความยากต่ำ
b_M	หมายถึง	ค่าความยากปานกลาง
b_H	หมายถึง	ค่าความยากสูง
TL	หมายถึง	ความยาวของแบบสอบ
%DIF	หมายถึง	สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน

Size	หมายถึง	ขนาดกลุ่มตัวอย่าง
LR	หมายถึง	วิธีการถดถอยโลจิสติก
LRj	หมายถึง	ผลการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl
SIBTEST	หมายถึง	วิธีซิปเทสท์
SIBTESTs	หมายถึง	ผลการวัดขนาดอิทธิพลตามเกณฑ์ Roussos and Stout
Raschtree	หมายถึง	วิธีราสช์ทรี
Raju	หมายถึง	วิธีการวัดพื้นที่ของราชู
θ	หมายถึง	ค่าพารามิเตอร์ความสามารถของผู้สอบ
M	หมายถึง	ค่าเฉลี่ย
SD	หมายถึง	ส่วนเบี่ยงเบนมาตรฐาน
Power	หมายถึง	อำนาจการทดสอบ
E_1	หมายถึง	อัตราความคลาดเคลื่อนประเภทที่ 1

ตอนที่ 1 ผลการวิเคราะห์ค่าเฉลี่ยของอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1

การศึกษาข้อมูลจำลองภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน จำนวน 36 เงื่อนไข แสดงรายละเอียดดังตารางที่ 10

ตารางที่ 10 ภาพรวมของการจำลองข้อมูลจำแนกตามปัจจัยและเงื่อนไขของปัจจัยที่แปรเปลี่ยน

ปัจจัยที่ศึกษา	เงื่อนไข
ความยากของข้อสอบ (Difficulty of item)	1. ค่าความยากต่ำ
	2. ค่าความยากปานกลาง
	3. ค่าความยากสูง
ความยาวของแบบสอบ (Test length)	1. 40 ข้อ
	2. 60 ข้อ
สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน (%DIF)	1. 10%
	2. 20%
ขนาดกลุ่มตัวอย่าง (Sample size)	1. 250 คน
	2. 500 คน
	3. 1000 คน

ผู้วิจัยนำเสนอค่าพารามิเตอร์ของข้อมูลจากการจำลองตามเงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ ปัจจัยความยากของข้อสอบ มี 3 ระดับ คือ ค่าความยากต่ำ ค่าความยากปานกลางและ

ค่าความยากสูง ปัจจัยความยาวของแบบสอบ มี 2 ขนาด คือ ความยาว 40 ข้อ และความยาว 60 ข้อ ปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน มี 2 ขนาด คือ 10% และ 20% และปัจจัยขนาดกลุ่มตัวอย่าง มี 3 ขนาด คือ 250 คน 500 คน และ 1000 คน มีข้อมูลทั้งหมดที่ต้องจัดกระทำเพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ จำนวน 36 เงื่อนไข (3x2x2x3) ทุกเงื่อนไขจำลองข้อมูลซ้ำ 50 ครั้ง

1.1 การตรวจสอบความถูกต้องของข้อมูลจำลอง

ผลการจำลองข้อมูล 36 เงื่อนไข ได้ค่าพารามิเตอร์ของข้อสอบและค่าพารามิเตอร์ของผู้สอบสอดคล้องกับความยากของข้อสอบและผู้สอบที่ต้องการศึกษา กล่าวคือ ค่าความยาก (b) มีค่าอยู่ในช่วงต่ำ $(\bar{b} = -1.34)$ ช่วงปานกลาง $(\bar{b} = 0.00)$ และช่วงสูง $(\bar{b} = 1.34)$ ส่วน SD อยู่ในช่วง 0.222 ค่าพารามิเตอร์ของผู้สอบ (θ) มี Mean อยู่ในช่วง 0 และ SD อยู่ในช่วง 1 รายละเอียดดังตารางที่ 11

ตารางที่ 11 ผลการตรวจสอบความถูกต้องของข้อมูลจำลองจำแนกตามปัจจัยที่ศึกษา 36 เงื่อนไข

เงื่อนไข	ปัจจัยที่แปรเปลี่ยน				ค่าพารามิเตอร์ของข้อสอบและผู้สอบ			
	DI	TL	% DIF	Size	Parameter b		θ	
					M	*SD*	*M*	*SD*
1	b_L	40	10	250	-1.330	0.205	-0.020	1.018
2	b_L	40	10	500	-1.339	0.224	-0.033	0.978
3	b_L	40	10	1000	-1.347	0.222	0.021	0.959
4	b_L	40	20	250	-1.349	0.206	0.015	1.002
5	b_L	40	20	500	-1.348	0.222	0.006	1.021
6	b_L	40	20	1000	-1.347	.0209	0.030	0.994
7	b_L	60	10	250	-1.332	0.216	-0.018	1.065
8	b_L	60	10	500	-1.344	0.197	-0.025	1.027
9	b_L	60	10	1000	-1.350	0.197	-0.024	1.006
10	b_L	60	20	250	-1.336	0.194	0.015	1.040
11	b_L	60	20	500	-1.334	0.223	-0.014	0.934
12	b_L	60	20	1000	-1.343	0.220	0.025	0.988
13	b_M	40	10	250	0.001	0.240	0.015	1.007
14	b_M	40	10	500	0.006	0.220	0.002	0.976
15	b_M	40	10	1000	-0.014	0.208	0.023	1.005

ตารางที่ 11 (ต่อ)

เงื่อนไข	ปัจจัยที่แปรเปลี่ยน				ค่าพารามิเตอร์ของข้อสอบและผู้สอบ			
	DI	TL	% DIF	Size	Parameter b		θ	
					M	*SD*	*M*	*SD*
16	b_M	40	20	250	0.028	0.205	0.014	1.016
17	b_M	40	20	500	0.008	0.205	0.012	1.022
18	b_M	40	20	1000	-0.001	0.187	0.010	1.003
19	b_M	60	10	250	0.009	0.229	-0.019	0.990
20	b_M	60	10	500	-0.018	0.190	0.027	0.956
21	b_M	60	10	1000	-0.012	0.197	0.017	0.995
22	b_M	60	20	250	-0.004	0.227	-0.020	0.933
23	b_M	60	20	500	0.019	0.204	0.009	1.030
24	b_M	60	20	1000	0.003	0.210	0.030	1.004
25	b_H	40	10	250	1.338	0.178	0.000	0.988
26	b_H	40	10	500	1.331	0.209	0.003	0.993
27	b_H	40	10	1000	1.349	0.199	-0.017	0.993
28	b_H	40	20	250	1.335	0.208	-0.007	1.014
29	b_H	40	20	500	1.348	0.220	-0.001	0.997
30	b_H	40	20	1000	1.347	0.214	0.001	0.989
31	b_H	60	10	250	1.345	0.210	-0.014	0.969
32	b_H	60	10	500	1.334	0.222	0.007	0.939
33	b_H	60	10	1000	1.327	0.197	-0.012	1.050
34	b_H	60	20	250	1.351	0.177	0.006	0.976
35	b_H	60	20	500	1.344	0.207	-0.010	0.994
36	b_H	60	20	1000	1.335	0.221	0.000	0.982

จากการจำลองข้อมูล 36 เงื่อนไข ๆ ละ 50 รอบ ผู้วิจัยเลือกข้อมูลจำลองในแต่ละเงื่อนไขมาเงื่อนไขละ 1 รอบ เพื่อนำมาตรวจสอบความถูกต้องของค่าความยากและส่วนเบี่ยงเบนมาตรฐานโดยโปรแกรม BILOGMG 3.0 ผลจากการตรวจสอบได้ค่าพารามิเตอร์ของข้อสอบและส่วนเบี่ยงเบนมาตรฐานใกล้เคียงกับความยากของข้อสอบที่กำหนดไว้ รายละเอียดดังตารางที่ 12

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 12 ผลการตรวจสอบค่าพารามิเตอร์ความยาก (b) และส่วนเบี่ยงเบนมาตรฐาน (*SD*) ของข้อมูลจำลอง

เงื่อนไข	ความยากของข้อสอบ	ค่าความยาก			ส่วนเบี่ยงเบนมาตรฐาน		
		จำลองข้อมูล	ตรวจสอบโดย BILOG		จำลองข้อมูล	ตรวจสอบโดย BILOG	
			กลุ่ม F	กลุ่ม R		กลุ่ม F	กลุ่ม R
1	($\bar{b}$ = -1.34)	-1.330	-1.324	-1.310	0.205	0.238	0.260
2		-1.339	-1.363	-1.346	0.224	0.280	0.238
3		-1.347	-1.442	-1.416	0.222	0.253	0.227
4		-1.349	-1.417	-1.322	0.206	0.254	0.239
5		-1.348	-1.377	-1.346	0.222	0.231	0.239
6		-1.347	-1.391	-1.347	0.209	0.261	0.209
7		-1.332	-1.212	-1.233	0.216	0.219	0.218
8		-1.344	-1.305	-1.307	0.197	0.228	0.224
9		-1.350	-1.386	-1.360	0.197	0.222	0.216
10		-1.336	-1.312	-1.298	0.194	0.260	0.193
11		-1.334	-1.448	-1.461	0.223	0.276	0.262
12		-1.343	-1.417	-1.421	0.220	0.248	0.238
13	($\bar{b}$ = 0.00)	0.001	-0.014	-0.023	0.240	0.254	0.270
14		0.006	0.004	0.003	0.220	0.249	0.252
15		0.014	-0.041	-0.051	0.208	0.211	0.214
16		0.028	-0.002	0.032	0.205	0.236	0.232
17		0.008	-0.003	-0.001	0.205	0.219	0.212
18		0.001	-0.007	-0.009	0.187	0.192	0.180
19		0.009	0.011	0.027	0.229	0.231	0.242
20		0.018	-0.049	-0.036	0.190	0.229	0.198
21		0.012	-0.041	-0.028	0.197	0.211	0.209
22		0.004	0.034	0.008	0.227	0.279	0.282
23		0.019	0.020	0.007	0.204	0.229	0.209
24		0.003	-0.031	-0.025	0.210	0.245	0.226

ตารางที่ 12 (ต่อ)

เงื่อนไข	ความยากของข้อสอบ	ค่าความยาก			ส่วนเบี่ยงเบนมาตรฐาน		
		จำลองข้อมูล	ตรวจสอบโดย BILOG		จำลองข้อมูล	ตรวจสอบโดย BILOG	
			กลุ่ม F	กลุ่ม R		กลุ่ม F	กลุ่ม R
25	($\bar{b}$ = 1.34)	1.338	1.409	1.381	0.178	0.232	0.196
26		1.331	1.354	1.389	0.209	0.231	0.230
27		1.349	1.362	1.379	0.199	0.231	0.205
28		1.335	1.353	1.257	0.208	0.235	0.270
29		1.348	1.357	1.361	0.220	0.220	0.245
30		1.347	1.372	1.357	0.214	0.237	0.207
31		1.345	1.455	1.467	0.210	0.267	0.274
32		1.334	1.411	1.428	0.222	0.268	0.236
33		1.327	1.313	1.284	0.197	0.214	0.184
34		1.351	1.382	1.449	0.177	0.242	0.241
35		1.344	1.432	1.416	0.207	0.265	0.230
36		1.335	1.402	1.400	0.221	0.250	0.221

ต่อจากนั้น ผู้วิจัยได้ตรวจสอบค่าความเที่ยงของแบบสอบจากการจำลองข้อมูล ในแต่ละเงื่อนไข ๆ ละ 1 รอบ ที่ได้ตรวจสอบความถูกต้องของค่าความยากและส่วนเบี่ยงเบนมาตรฐานไปแล้ว พบว่า ค่าความเที่ยงมีค่าอยู่ระหว่าง 0.8438 ถึง 0.9595 รายละเอียดดังตารางที่ 13

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 13 ค่าความเที่ยง (Reliability) ของแบบสอบจากการจำลองข้อมูล

เงื่อนไข	ค่าความเที่ยง		เงื่อนไข	ค่าความเที่ยง	
	กลุ่ม F	กลุ่ม R		กลุ่ม F	กลุ่ม R
1	0.8591	0.8566	19	0.9595	0.9538
2	0.8563	0.8518	20	0.9531	0.9519
3	0.8470	0.8438	21	0.9544	0.9550
4	0.8520	0.8505	22	0.9541	0.9513
5	0.8520	0.8533	23	0.9592	0.9542
6	0.8502	0.8497	24	0.9565	0.9566
7	0.8923	0.8971	25	0.8444	0.8468
8	0.8963	0.8941	26	0.8527	0.8539
9	0.8896	0.8900	27	0.8547	0.8508
10	0.8975	0.8840	28	0.8563	0.8682
11	0.8817	0.8855	29	0.8539	0.8539
12	0.8850	0.8821	30	0.8513	0.8510
13	0.9424	0.9334	31	0.8772	0.8817
14	0.9295	0.9348	32	0.8845	0.8897
15	0.9337	0.9350	33	0.8906	0.8897
16	0.9370	0.9321	34	0.8869	0.8878
17	0.9325	0.9367	35	0.8829	0.8887
18	0.9376	0.9348	36	0.8872	0.8865

1.2 ประสิทธิภาพด้านอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 โดยการทดสอบระดับนัยสำคัญ ของวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี

ค่าเฉลี่ยร้อยละของอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ทุกเงื่อนไขการตรวจสอบประสิทธิภาพการทำหน้าที่ต่างกันของข้อสอบภายใต้วิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ยึดเงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ผลวิเคราะห์อำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 โดยการทดสอบระดับนัยสำคัญในภาพรวม รายละเอียดดังตารางที่ 14

ตารางที่ 14 ค่าเฉลี่ยร้อยละและส่วนเบี่ยงเบนมาตรฐานของอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 โดยการทดสอบระดับนัยสำคัญในภาพรวม

DI	TL	%DIF	Size	อำนาจการทดสอบ						อัตราความคลาดเคลื่อนประเภทที่ 1					
				LR		SIBTEST		Raschtree		LR		SIBTEST		Raschtree	
				M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
b_L	40	10	250	7.50	2.52	22.50	18.79	5.00	2.00	6.61	4.43	5.56	12.33	1.39	1.71
			500	14.00	1.63	57.00	23.47	19.50	7.72	7.11	3.78	5.44	9.18	2.94	2.46
			1000	18.50	1.00	67.00	11.60	30.00	1.63	6.33	3.43	4.72	8.66	0.44	0.97
		20	250	8.50	4.87	15.50	16.72	8.00	4.14	6.88	3.93	6.06	11.31	4.06	2.76
			500	14.75	7.01	32.00	15.96	33.00	5.66	7.13	4.16	4.44	8.53	4.56	3.18
			1000	19.75	5.90	67.50	18.60	31.00	3.55	6.44	3.69	5.94	9.76	0.50	1.02
	60	10	250	10.67	6.41	23.67	18.39	11.33	3.27	5.93	2.55	5.81	10.53	3.81	2.61
			500	9.00	7.24	37.67	19.03	19.33	5.75	6.44	3.74	7.67	12.36	2.48	2.20
			1000	17.33	5.01	48.33	9.16	22.67	2.42	5.93	3.41	5.37	10.97	0.26	.078
		20	250	10.83	4.22	17.83	3.86	17.00	3.36	6.50	3.63	4.75	2.94	6.38	3.61
			500	9.33	5.28	38.67	7.10	34.00	4.09	6.46	3.48	4.67	3.15	4.17	3.14
			1000	16.50	6.72	67.00	8.07	29.50	6.22	6.54	3.57	5.38	2.80	0.58	1.37

ตารางที่ 14 (ต่อ)

DI	TL	%DIF	Size	อำนาจการทดสอบ						อัตราความคลาดเคลื่อนประเภทที่ 1					
				LR		SIBTEST		Raschtree		LR		SIBTEST		Raschtree	
				M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
b_M	40	10	250	9.00	1.15	27.50	10.50	9.50	2.52	6.06	3.12	5.28	2.67	2.00	1.66
			500	14.50	1.91	58.50	7.55	27.00	6.63	4.83	2.55	4.50	3.46	0.72	1.28
			1000	27.50	3.42	84.00	6.73	29.50	3.42	5.50	3.00	4.50	2.59	0.00	0.00
		20	250	10.75	3.01	28.50	6.99	21.00	6.23	5.63	3.64	5.13	3.48	3.38	3.07
			500	15.25	6.04	49.50	5.32	31.00	5.95	7.06	3.77	5.75	3.24	1.38	1.64
			1000	28.25	6.71	84.00	4.41	26.25	7.05	5.69	3.09	5.31	3.79	0.00	0.00
	60	10	250	11.67	4.27	29.67	8.52	12.67	4.50	5.15	3.02	5.41	3.51	2.67	2.27
			500	19.33	5.89	50.67	4.84	25.00	4.52	7.07	3.58	4.26	3.25	0.63	1.34
			1000	25.67	4.80	81.67	5.85	30.33	4.46	6.26	3.54	4.70	3.49	0.07	0.38
		20	250	11.33	3.85	31.50	7.49	23.50	4.27	4.88	3.03	7.92	3.69	4.67	2.68
			500	14.00	6.88	46.33	6.97	29.83	8.42	5.58	3.30	5.04	2.95	1.25	1.83
			1000	29.17	4.04	80.00	7.39	32.33	5.10	6.17	4.14	4.67	3.10	0.04	0.29

ตารางที่ 14 (ต่อ)

DI	TL	%DIF	Size	อำนาจการทดสอบ						อัตราความคลาดเคลื่อนประเภทที่ 1					
				LR		SIBTEST		Raschtree		LR		SIBTEST		Raschtree	
				M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
b_H	40	10	250	10.00	7.12	19.00	7.39	11.50	5.26	5.50	2.96	4.56	2.93	3.56	2.62
			500	14.50	5.00	27.50	8.70	22.00	5.16	5.56	3.38	5.00	2.85	2.89	2.76
			1000	21.50	7.19	36.00	10.71	27.50	12.15	7.44	3.26	5.89	2.70	0.28	0.70
		20	250	14.25	8.71	15.75	5.60	22.75	4.78	5.19	2.68	6.69	3.86	8.13	3.80
			500	11.50	3.16	21.25	4.40	25.50	4.24	5.63	2.89	3.56	2.82	3.25	2.20
			1000	22.75	6.50	41.75	7.59	36.25	8.97	7.00	4.40	5.63	3.82	0.31	0.74
	60	10	250	13.67	5.28	9.67	4.80	6.00	4.20	7.22	3.71	5.81	3.64	2.15	2.01
			500	16.67	6.15	24.00	10.58	25.67	7.09	6.70	4.29	4.70	3.58	2.96	2.85
			1000	23.00	7.56	40.67	4.13	26.00	7.16	6.70	3.48	4.70	3.05	1.96	3.72
		20	250	10.50	3.92	12.33	5.31	15.83	3.46	6.71	3.97	6.75	3.70	6.21	2.71
			500	14.67	4.92	26.50	3.83	35.17	4.13	5.46	3.64	4.42	3.03	3.25	3.08
			1000	27.67	9.18	43.17	6.35	33.83	4.86	7.00	4.25	4.96	2.92	0.42	1.01

จากตารางที่ 14 ค่าเฉลี่ยร้อยละของอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 โดยการทดสอบระดับนัยสำคัญ ของวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี สรุปได้ดังนี้

1.2.1 อำนาจการทดสอบ

การตรวจสอบการทำหน้าที่ต่างกันโดยการทดสอบระดับนัยสำคัญด้วยวิธีการถดถอยโลจิสติก มีร้อยละเฉลี่ยสูงสุดของอำนาจการทดสอบ คิดเป็นร้อยละ 29.17 ภายใต้เงื่อนไขความยากของข้อสอบแบบความยากปานกลาง ความยาวของแบบสอบ 60 ข้อ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 20 และขนาดกลุ่มตัวอย่าง 1000 คน ส่วนร้อยละเฉลี่ยต่ำสุด คิดเป็นร้อยละ 7.50 ภายใต้เงื่อนไขความยากของข้อสอบแบบความยากต่ำ ความยาวของแบบสอบ 40 ข้อ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และขนาดกลุ่มตัวอย่าง 250 คน

การตรวจสอบการทำหน้าที่ต่างกันโดยการทดสอบระดับนัยสำคัญด้วยวิธีซิปเทสท์ มีร้อยละเฉลี่ยสูงสุดของอำนาจการทดสอบ คิดเป็นร้อยละ 84.00 ภายใต้เงื่อนไขความยากของข้อสอบแบบความยากปานกลาง ความยาวของแบบสอบ 40 ข้อ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และขนาดกลุ่มตัวอย่าง 1000 คน และภายใต้เงื่อนไขความยากของข้อสอบแบบความยากปานกลาง ความยาวของแบบสอบ 40 ข้อ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 20 และขนาดกลุ่มตัวอย่าง 1000 คน ส่วนร้อยละเฉลี่ยต่ำสุด คิดเป็นร้อยละ 9.67 ภายใต้เงื่อนไขความยากของข้อสอบแบบความยากสูง ความยาวของแบบสอบ 60 ข้อ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และขนาดกลุ่มตัวอย่าง 250 คน

การตรวจสอบการทำหน้าที่ต่างกันโดยการทดสอบระดับนัยสำคัญด้วยวิธีราสช์ทรี มีร้อยละเฉลี่ยสูงสุดของอำนาจการทดสอบ คิดเป็นร้อยละ 36.25 ภายใต้เงื่อนไขความยากของข้อสอบแบบความยากสูง ความยาวของแบบสอบ 40 ข้อ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 20 และขนาดกลุ่มตัวอย่าง 1000 คน ส่วนร้อยละเฉลี่ยต่ำสุด คิดเป็นร้อยละ 5.00 ภายใต้เงื่อนไขความยากของข้อสอบแบบความยากต่ำ ความยาวของแบบสอบ 40 ข้อ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และขนาดกลุ่มตัวอย่าง 250 คน

1.2.2 อัตราความคลาดเคลื่อนประเภทที่ 1

การตรวจสอบการทำหน้าที่ต่างกันโดยการทดสอบระดับนัยสำคัญด้วยวิธีการถดถอยโลจิสติก มีร้อยละเฉลี่ยสูงสุดของอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 7.44 ภายใต้เงื่อนไขความยากของข้อสอบแบบความยากสูง ความยาวของแบบสอบ 40 ข้อ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และขนาดกลุ่มตัวอย่าง 1000 คน ส่วนร้อยละเฉลี่ยต่ำสุด คิดเป็นร้อยละ

4.83 ภายใต้เงื่อนไขความยากของข้อสอบแบบความยากปานกลาง ความยาวของแบบสอบ 40 ข้อ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และขนาดกลุ่มตัวอย่าง 500 คน

การตรวจสอบการทำหน้าที่ต่างกันโดยการทดสอบระดับนัยสำคัญด้วยวิธีซิปเทสท์ มีร้อยละเฉลี่ยสูงสุดของอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 7.92 ภายใต้เงื่อนไขความยากของข้อสอบแบบความยากปานกลาง ความยาวของแบบสอบ 60 ข้อ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 20 และขนาดกลุ่มตัวอย่าง 250 คน ส่วนร้อยละเฉลี่ยต่ำสุด คิดเป็นร้อยละ 3.56 ภายใต้เงื่อนไขความยากของข้อสอบแบบความยากสูง ความยาวของแบบสอบ 40 ข้อ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 20 และขนาดกลุ่มตัวอย่าง 500 คน

การตรวจสอบการทำหน้าที่ต่างกันโดยการทดสอบระดับนัยสำคัญด้วยวิธีราสช์ทรี มีร้อยละเฉลี่ยสูงสุดของอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 8.13 ภายใต้เงื่อนไขความยากของข้อสอบแบบความยากสูง ความยาวของแบบสอบ 40 ข้อ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 20 และขนาดกลุ่มตัวอย่าง 250 คน ส่วนร้อยละเฉลี่ยต่ำสุด คิดเป็นร้อยละ 0.00 ภายใต้เงื่อนไขความยากของข้อสอบแบบความยากปานกลาง ความยาวของแบบสอบ 40 ข้อ ขนาดกลุ่มตัวอย่าง 1000 คน ทั้งในสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และร้อยละ 20

1.3 ผลการวิเคราะห์อำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 โดยการทดสอบระดับนัยสำคัญ ของวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี

ค่าเฉลี่ยร้อยละของอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี จำแนกตามปัจจัยที่ศึกษา แสดงดังตารางที่ 15 และ 16

ตารางที่ 15 ค่าเฉลี่ยร้อยละ (*M*) และส่วนเบี่ยงเบนมาตรฐาน (*SD*) ของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี จำแนกตามปัจจัยที่ศึกษา

ปัจจัย	LR		SIBTEST		Raschtree	
	M	*SD*	*M*	*SD*	*M*	*SD*
Difficulty of Item						
b_L	13.06	4.27	41.22	19.75	21.69	10.01
b_M	18.03	7.59	54.32	22.97	24.83	7.25
b_H	16.72	5.67	26.47	11.65	24.00	9.27
Test length						
40 ข้อ	15.71	6.19	41.93	22.79	23.13	9.17
60 ข้อ	16.17	6.40	39.41	20.81	23.89	8.59
%DIF						
10%	15.78	5.89	41.39	21.44	20.03	8.54
20%	16.10	6.67	39.95	22.25	26.99	7.72
Sample size						
250 คน	10.72	1.94	21.12	7.19	13.67	6.35
500 คน	13.96	2.89	39.13	13.09	27.25	5.42
1000 คน	23.13	4.49	61.76	18.69	29.60	3.68

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 16 ค่าเฉลี่ยร้อยละ (*M*) และส่วนเบี่ยงเบนมาตรฐาน (*SD*) ของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี จำแนกตามปัจจัยที่ศึกษา

ปัจจัย	LR		SIBTEST		Raschtree	
	M	*SD*	*M*	*SD*	*M*	*SD*
Difficulty of Item						
b_L	6.52	0.38	5.48	0.87	2.63	2.01
b_M	5.82	0.74	5.21	0.96	1.40	1.51
b_H	6.34	0.81	5.22	0.95	2.95	2.34
Test length						
40 ข้อ	6.20	0.79	5.22	0.75	2.21	2.11
60 ข้อ	6.26	0.66	5.39	1.06	2.44	2.02
%DIF						
10%	6.24	0.75	5.22	0.80	1.73	1.28
20%	6.22	0.71	5.39	1.03	2.92	2.49
Sample size						
250 คน	6.02	0.76	5.81	0.95	4.03	2.01
500 คน	6.25	0.80	4.95	1.03	2.54	1.29
1000 คน	6.42	0.57	5.15	0.50	0.41	0.53

จากตารางที่ 15-16 ค่าเฉลี่ยร้อยละของอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี จำแนกตามปัจจัยที่ศึกษา สรุปผลได้ดังนี้

ปัจจัยความยากของข้อสอบ (DI) ในการตรวจสอบการทำหน้าที่ต่างกันด้วยวิธีการถดถอยโลจิสติก พบว่า ภายใต้ความยากของข้อสอบ b_M มีค่าเฉลี่ยของอำนาจการทดสอบสูงสุด ประมาณร้อยละ 18.03 และมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ประมาณร้อยละ 5.82 แต่ภายใต้ความยากของข้อสอบ b_L มีค่าเฉลี่ยของอำนาจการทดสอบต่ำสุด ประมาณร้อยละ 13.06 และมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ประมาณร้อยละ 6.52 สำหรับวิธีซิปเทสท์ พบว่า ภายใต้ความยากของข้อสอบ b_M มีค่าเฉลี่ยของอำนาจการทดสอบสูงสุด ประมาณ

ร้อยละ 54.32 และมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ประมาณร้อยละ 5.21 แต่ภายใต้ความยากของข้อสอบ b_H มีค่าเฉลี่ยของอำนาจการทดสอบต่ำสุด ประมาณร้อยละ 26.47 และภายใต้ความยากของข้อสอบ b_L มีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ประมาณร้อยละ 5.48 ส่วนวิธีราสช์ทรี พบว่า ภายใต้ความยากของข้อสอบ b_M มีค่าเฉลี่ยของอำนาจการทดสอบสูงสุด ประมาณร้อยละ 24.83 และมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ประมาณร้อยละ 1.40 แต่ภายใต้ความยากของข้อสอบ b_L มีค่าเฉลี่ยของอำนาจการทดสอบต่ำสุด ประมาณร้อยละ 21.69 และภายใต้ความยากของข้อสอบ b_H มีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ประมาณร้อยละ 2.95

จากผลดังกล่าวเมื่อพิจารณาวิธีการตรวจสอบทั้ง 3 วิธี ภายใต้ความยากของข้อสอบต่างกัน 3 ระดับ พบว่า ภายใต้ความยากของข้อสอบที่มีค่าความยากปานกลาง (b_M) วิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี มีค่าเฉลี่ยของอำนาจการทดสอบสูงกว่าภายใต้ความยากของข้อสอบที่มีค่าความยากต่ำและสูง และมีค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าภายใต้ความยากของข้อสอบที่มีค่าความยากต่ำและสูงด้วย

ปัจจัยความยาวของแบบสอบ (TL) ในการตรวจสอบการทำหน้าที่ต่างกันด้วยวิธีการถดถอยโลจิสติก พบว่า ภายใต้ความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ มีค่าเฉลี่ยของอำนาจการทดสอบ ประมาณร้อยละ 15.71 และ 16.17 ตามลำดับ และมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณร้อยละ 6.20 และ 6.22 ตามลำดับ สำหรับวิธีซิปเทสท์ พบว่า ภายใต้ความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ มีค่าเฉลี่ยของอำนาจการทดสอบ ประมาณร้อยละ 41.93 และ 39.41 ตามลำดับ และมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณร้อยละ 5.22 และ 5.39 ตามลำดับ ส่วนวิธีราสช์ทรี พบว่า ภายใต้ความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ มีค่าเฉลี่ยของอำนาจการทดสอบ ประมาณร้อยละ 23.13 และ 23.89 ตามลำดับ และมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณร้อยละ 2.21 และ 2.44 ตามลำดับ

จากผลดังกล่าวเมื่อพิจารณาวิธีการตรวจสอบทั้ง 3 วิธี ภายใต้ความยาวของแบบสอบต่างกัน 2 ขนาด พบว่า ภายใต้ความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ ค่าเฉลี่ยของอำนาจการทดสอบของวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี มีค่าใกล้เคียงกัน และมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ใกล้เคียงกันด้วย

ปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน (%DIF) ในการตรวจสอบการทำหน้าที่ต่างกันด้วยวิธีการถดถอยโลจิสติก พบว่า ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และร้อยละ 20 มีค่าเฉลี่ยของอำนาจการทดสอบ ประมาณร้อยละ 15.78 และ 16.10 ตามลำดับ และมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณร้อยละ 6.24 และ 6.22 ตามลำดับ สำหรับวิธี ซิปเทสท์ พบว่า ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และร้อยละ 20

มีค่าเฉลี่ยของอำนาจการทดสอบ ประมาณร้อยละ 41.39 และ 39.95 ตามลำดับ และมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณร้อยละ 5.22 และ 5.39 ตามลำดับ ส่วนวิธีราสช์ทรี พบว่า ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และร้อยละ 20 มีค่าเฉลี่ยของอำนาจการทดสอบ ประมาณร้อยละ 20.03 และ 26.99 ตามลำดับ และมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณร้อยละ 1.73 และ 2.92 ตามลำดับ

จากผลดังกล่าวเมื่อพิจารณาวิธีการตรวจสอบทั้ง 3 วิธี ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด พบว่า ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และร้อยละ 20 ค่าเฉลี่ยของอำนาจการทดสอบของวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี มีค่าใกล้เคียงกัน และมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ใกล้เคียงกันด้วย

ปัจจัยขนาดกลุ่มตัวอย่าง (Size) ในการตรวจสอบการทำหน้าที่ต่างกันด้วยวิธีการถดถอยโลจิสติก พบว่า ภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน มีค่าเฉลี่ยของอำนาจการทดสอบสูงสุด ประมาณร้อยละ 23.13 และมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ประมาณร้อยละ 6.42 แต่ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน มีค่าเฉลี่ยของอำนาจการทดสอบต่ำสุด ประมาณร้อยละ 10.72 และมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ประมาณร้อยละ 6.02 สำหรับวิธีซิปเทสท์ พบว่า ภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน มีค่าเฉลี่ยของอำนาจการทดสอบสูงสุด ประมาณร้อยละ 61.76 และมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณร้อยละ 5.15 แต่ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน มีค่าเฉลี่ยของอำนาจการทดสอบต่ำสุด ประมาณร้อยละ 21.12 และมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ประมาณร้อยละ 5.81 ส่วนวิธีราสช์ทรี พบว่า ภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน มีค่าเฉลี่ยของอำนาจการทดสอบสูงสุด ประมาณร้อยละ 29.60 และมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ประมาณร้อยละ 0.41 แต่ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน มีค่าเฉลี่ยของอำนาจการทดสอบต่ำสุด ประมาณร้อยละ 13.67 และมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ประมาณร้อยละ 4.03

จากผลดังกล่าวเมื่อพิจารณาวิธีการตรวจสอบทั้ง 3 วิธี ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด พบว่า ภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน วิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี มีค่าเฉลี่ยของอำนาจการทดสอบสูงกว่าภายใต้ขนาดกลุ่มตัวอย่าง 250 คน และ 500 คน ส่วนค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าใกล้เคียงกันทั้ง 3 ขนาด

1.4 ประสิทธิภาพด้านอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 โดยการวัดขนาดอิทธิพล ของวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์

ค่าเฉลี่ยร้อยละของอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ทุกเงื่อนไขการตรวจสอบประสิทธิภาพการทำหน้าที่ต่างกันของข้อสอบภายใต้วิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ โดยการวัดขนาดอิทธิพล ยึดเงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ผลวิเคราะห์อำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในภาพรวม รายละเอียดดังตารางที่ 17

ตารางที่ 17 ค่าเฉลี่ยร้อยละและส่วนเบี่ยงเบนมาตรฐานของอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 โดยการวัดขนาดอิทธิพล ในภาพรวม

DI	TL	%DIF	Size	อำนาจการทดสอบ				อัตราความคลาดเคลื่อนประเภทที่ 1			
				LRj		SIBTESTs		LRj		SIBTESTs	
				M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
b_L	40	10	250	2.50	2.52	22.50	18.79	2.83	2.50	5.56	12.33
			500	9.50	1.00	57.00	23.47	4.94	3.33	5.44	9.18
			1000	16.50	1.00	67.00	11.60	5.78	2.98	4.72	8.66
		20	250	2.25	3.28	15.50	16.72	2.06	2.00	6.06	11.31
			500	9.25	6.04	32.00	15.96	4.81	3.73	4.44	8.53
			1000	18.75	6.50	67.50	18.60	6.00	3.59	5.94	9.76
	60	10	250	1.67	2.34	23.67	18.39	0.78	1.13	5.81	10.53
			500	4.00	3.10	37.67	19.03	3.26	3.14	7.67	12.36
			1000	15.67	6.25	48.33	9.16	4.85	3.10	5.37	10.97
	60	20	250	1.67	1.87	17.83	3.86	1.21	1.53	4.75	2.94
			500	3.83	3.66	38.67	7.10	2.46	2.31	4.67	3.15
			1000	14.33	7.13	67.00	8.07	5.38	3.38	5.38	2.80
b_M	40	10	250	5.50	1.00	27.50	10.50	3.28	2.45	5.28	2.67
			500	13.50	2.52	58.50	7.55	4.67	2.44	4.50	3.46
			1000	27.50	4.12	84.00	6.73	5.50	3.04	4.50	2.59
		20	250	4.50	2.07	28.50	6.99	3.38	2.51	5.13	3.48
			500	14.25	6.09	49.50	5.32	6.63	3.78	5.75	3.24
			1000	28.00	6.68	84.00	4.41	5.69	3.09	5.31	3.79

ตารางที่ 17 (ต่อ)

DI	TL	%DIF	Size	อำนาจการทดสอบ				อัตราความคลาดเคลื่อนประเภทที่ 1			
				LRj		SIBTESTs		LRj		SIBTESTs	
				M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
b_M	60	10	250	1.00	1.10	29.67	8.52	0.78	1.06	5.41	3.51
			500	15.00	5.62	50.67	4.84	5.59	3.42	4.26	3.25
			1000	25.67	4.80	81.67	5.85	6.26	3.56	4.70	3.49
	60	20	250	2.00	2.83	31.50	7.49	0.63	1.10	7.92	3.69
			500	12.83	6.58	46.33	6.97	4.58	3.03	5.04	2.95
			1000	29.17	4.04	80.00	7.39	6.17	4.14	4.67	3.10
b_H	40	10	250	4.00	3.65	19.00	7.39	1.78	2.07	4.56	2.93
			500	9.00	3.83	27.50	8.70	4.22	3.02	5.00	2.85
			1000	20.50	6.61	36.00	10.71	6.78	3.03	5.89	2.70
		20	250	4.50	3.51	15.75	5.60	1.94	1.79	6.69	3.86
			500	7.50	2.33	21.25	4.40	4.44	2.95	3.56	2.82
			1000	20.25	7.89	41.75	7.59	6.69	4.36	5.63	3.82
b_H	60	10	250	0.33	0.82	9.67	4.80	0.41	0.81	5.81	3.64
			500	7.33	2.42	24.00	10.58	2.74	2.89	4.70	3.58
			1000	18.33	8.71	40.67	4.13	6.07	3.39	4.70	3.05
		20	250	0.50	0.90	12.33	5.31	0.79	1.29	6.75	3.70
			500	6.00	4.18	26.50	3.83	1.92	2.22	4.42	3.03
			1000	23.83	9.63	43.17	6.35	6.00	3.57	4.96	2.92

จากตารางที่ 17 ค่าเฉลี่ยร้อยละของอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 โดยการวัดขนาดอิทธิพล ของวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ สรุปได้ดังนี้

1.4.1 อำนาจการทดสอบ

การตรวจสอบการทำหน้าที่ต่างกันโดยการวัดขนาดอิทธิพลด้วยวิธีการถดถอยโลจิสติก มีร้อยละเฉลี่ยสูงสุดของอำนาจการทดสอบ คิดเป็นร้อยละ 29.17 ภายใต้เงื่อนไขความยากของข้อสอบแบบความยากปานกลาง ความยาวของแบบสอบ 60 ข้อ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 20 และขนาดกลุ่มตัวอย่าง 1000 คน ส่วนร้อยละเฉลี่ยต่ำสุด คิดเป็นร้อยละ 0.33 ภายใต้เงื่อนไขความ

ยากของข้อสอบแบบความยากสูง ความยาวของแบบสอบ 60 ข้อ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และขนาดกลุ่มตัวอย่าง 250 คน

การตรวจสอบการทำหน้าที่ต่างกันโดยการวัดขนาดอิทธิพลด้วยวิธีซิปเทสท์ มีร้อยละเฉลี่ยสูงสุดของอำนาจการทดสอบ คิดเป็นร้อยละ 84.00 ภายใต้เงื่อนไขความยากของข้อสอบแบบความยากปานกลาง ความยาวของแบบสอบ 40 ข้อ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และขนาดกลุ่มตัวอย่าง 1000 คน และภายใต้เงื่อนไขความยากของข้อสอบแบบความยากปานกลาง ความยาวของแบบสอบ 40 ข้อ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 20 และขนาดกลุ่มตัวอย่าง 1000 คน ส่วนร้อยละเฉลี่ยต่ำสุด คิดเป็นร้อยละ 9.67 ภายใต้เงื่อนไขความยากของข้อสอบแบบความยากสูง ความยาวของแบบสอบ 60 ข้อ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และขนาดกลุ่มตัวอย่าง 250 คน

1.4.2 อัตราความคลาดเคลื่อนประเภทที่ 1

การตรวจสอบการทำหน้าที่ต่างกันโดยการวัดขนาดอิทธิพลด้วยวิธีการถดถอยโลจิสติก มีร้อยละเฉลี่ยสูงสุดของอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 6.78 ภายใต้เงื่อนไขความยากของข้อสอบแบบความยากสูง ความยาวของแบบสอบ 40 ข้อ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และขนาดกลุ่มตัวอย่าง 1000 คน ส่วนร้อยละเฉลี่ยต่ำสุด คิดเป็นร้อยละ 0.41 ภายใต้เงื่อนไขความยากของข้อสอบแบบความยากสูง ความยาวของแบบสอบ 60 ข้อ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และขนาดกลุ่มตัวอย่าง 250 คน

การตรวจสอบการทำหน้าที่ต่างกันโดยการวัดขนาดอิทธิพลด้วยวิธีซิปเทสท์ มีร้อยละเฉลี่ยสูงสุดของอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 7.92 ภายใต้เงื่อนไขความยากของข้อสอบแบบความยากปานกลาง ความยาวของแบบสอบ 60 ข้อ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 20 และขนาดกลุ่มตัวอย่าง 250 คน ส่วนร้อยละเฉลี่ยต่ำสุด คิดเป็นร้อยละ 3.56 ภายใต้เงื่อนไขความยากของข้อสอบแบบความยากสูง ความยาวของแบบสอบ 40 ข้อ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 20 และขนาดกลุ่มตัวอย่าง 500 คน

1.5 ผลการวิเคราะห์อำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 โดยการวัดขนาดอิทธิพล ของวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์

ค่าเฉลี่ยร้อยละของอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ด้วยวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ จำแนกตามปัจจัยที่ศึกษา แสดงดังตารางที่ 18 และ 19

ตารางที่ 18 ค่าเฉลี่ยร้อยละ (*M*) และส่วนเบี่ยงเบนมาตรฐาน (*SD*) ของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ด้วยวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ จำแนกตามปัจจัยที่ศึกษา

ปัจจัย	LRj		SIBTESTs	
	M	*SD*	*M*	*SD*
Difficulty of Item				
b_L	8.33	6.51	41.22	19.75
b_M	14.91	10.50	54.32	22.97
b_H	10.17	8.29	26.47	11.65
Test length				
40 ข้อ	12.10	8.25	41.93	22.79
60 ข้อ	10.18	9.46	39.41	20.81
%DIF				
10%	10.97	8.49	41.39	21.44
20%	11.30	9.34	39.95	22.25
Sample size				
250 คน	2.53	1.70	21.12	7.19
500 คน	9.33	3.85	39.13	13.09
1000 คน	21.54	5.14	61.76	18.69

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 19 ค่าเฉลี่ยร้อยละ (*M*) และส่วนเบี่ยงเบนมาตรฐาน (*SD*) ของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ด้วยวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ จำแนกตามปัจจัยที่ศึกษา

ปัจจัย	LRj		SIBTESTs	
	M	*SD*	*M*	*SD*
Difficulty of Item				
b_L	3.70	1.82	5.48	0.87
b_M	4.43	2.04	5.21	0.96
b_H	3.65	2.34	5.22	0.95

ตารางที่ 19 (ต่อ)

ปัจจัย	LRj		SIBTESTs	
	M	*SD*	*M*	*SD*
Test length				
40 ข้อ	4.52	1.66	5.22	0.75
60 ข้อ	3.33	2.27	5.39	1.06
%DIF				
10%	3.92	2.02	5.22	0.80
20%	3.93	2.13	5.39	1.03
Sample size				
250 คน	1.65	1.06	5.81	0.95
500 คน	4.19	1.36	4.95	1.03
1000 คน	5.93	0.54	5.15	0.50

จากตารางที่ 18-19 ค่าเฉลี่ยร้อยละของอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ด้วยวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ จำแนกตามปัจจัยที่ศึกษา สรุปผลได้ดังนี้

ปัจจัยความยากของข้อสอบ (DI) ในการตรวจสอบการทำหน้าที่ต่างกันด้วยวิธีการถดถอยโลจิสติก พบว่า ภายใต้ความยากของข้อสอบ b_M มีค่าเฉลี่ยของอำนาจการทดสอบสูงสุด ประมาณร้อยละ 14.91 และมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ประมาณร้อยละ 4.43 แต่ภายใต้ความยากของข้อสอบ b_L มีค่าเฉลี่ยของอำนาจการทดสอบต่ำสุด ประมาณร้อยละ 8.33 และภายใต้ความยากของข้อสอบ b_H มีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ประมาณร้อยละ 3.65 สำหรับวิธีซิปเทสท์ พบว่า ภายใต้ความยากของข้อสอบ b_M มีค่าเฉลี่ยของอำนาจการทดสอบสูงสุด ประมาณร้อยละ 54.32 และมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ประมาณร้อยละ 5.21 แต่ภายใต้ความยากของข้อสอบ b_H มีค่าเฉลี่ยของอำนาจการทดสอบต่ำสุด ประมาณร้อยละ 26.47 และภายใต้ความยากของข้อสอบ b_L มีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ประมาณร้อยละ 5.48

จากผลดังกล่าวเมื่อพิจารณาวิธีการตรวจสอบทั้ง 2 วิธี ภายใต้ความยากของข้อสอบต่างกัน 3 ระดับ พบว่า ภายใต้ความยากของข้อสอบที่มีค่าความยากปานกลาง (b_M) วิธีการถดถอยโลจิสติก

และวิธีซิปเทสท์ มีค่าเฉลี่ยของอำนาจการทดสอบสูงกว่าภายใต้ความยากของข้อสอบที่มีค่าความยากต่ำและสูง และมีค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 ใกล้เคียงกันด้วย

ปัจจัยความยาวของแบบสอบ (TL) ในการตรวจสอบการทำหน้าที่ต่างกันด้วยวิธีการถดถอยโลจิสติก พบว่า ภายใต้ความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ มีค่าเฉลี่ยของอำนาจการทดสอบ ประมาณร้อยละ 12.10 และ 10.18 ตามลำดับ และมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณร้อยละ 4.52 และ 3.33 ตามลำดับ สำหรับวิธีซิปเทสท์ พบว่า ภายใต้ความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ มีค่าเฉลี่ยของอำนาจการทดสอบ ประมาณร้อยละ 41.93 และ 39.41 ตามลำดับ และมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณร้อยละ 5.22 และ 5.39 ตามลำดับ

จากผลดังกล่าวเมื่อพิจารณาวิธีการตรวจสอบทั้ง 2 วิธี ภายใต้ความยาวของแบบสอบต่างกัน 2 ขนาด พบว่า ภายใต้ความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ ค่าเฉลี่ยของอำนาจการทดสอบของวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ มีค่าใกล้เคียงกัน และมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ใกล้เคียงกันด้วย

ปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน (%DIF) ในการตรวจสอบการทำหน้าที่ต่างกันด้วยวิธีการถดถอยโลจิสติก พบว่า ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และร้อยละ 20 มีค่าเฉลี่ยของอำนาจการทดสอบ ประมาณร้อยละ 10.97 และ 11.30 ตามลำดับ และมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณร้อยละ 3.92 และ 3.93 ตามลำดับ สำหรับวิธี ซิปเทสท์ พบว่า ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และร้อยละ 20 มีค่าเฉลี่ยของอำนาจการทดสอบ ประมาณร้อยละ 41.39 และ 39.95 ตามลำดับ และมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณร้อยละ 5.22 และ 5.39 ตามลำดับ

จากผลดังกล่าวเมื่อพิจารณาวิธีการตรวจสอบทั้ง 2 วิธี ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด พบว่า ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และร้อยละ 20 ค่าเฉลี่ยของอำนาจการทดสอบของวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี มีค่าใกล้เคียงกัน และมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ใกล้เคียงกันด้วย

ปัจจัยขนาดกลุ่มตัวอย่าง (Size) ในการตรวจสอบการทำหน้าที่ต่างกันด้วยวิธีการถดถอยโลจิสติก พบว่า ภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน มีค่าเฉลี่ยของอำนาจการทดสอบสูงสุด ประมาณร้อยละ 21.54 และมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ประมาณร้อยละ 5.93 แต่ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน มีค่าเฉลี่ยของอำนาจการทดสอบต่ำสุด ประมาณร้อยละ 2.53 และมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ประมาณร้อยละ 1.65 สำหรับวิธีซิปเทสท์ พบว่า ภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน มีค่าเฉลี่ยของอำนาจการทดสอบสูงสุด ประมาณร้อยละ 61.76 และมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณร้อยละ

5.15 แต่ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน มีค่าเฉลี่ยของอำนาจการทดสอบต่ำสุด ประมาณร้อยละ 21.12 และมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ประมาณร้อยละ 5.81

จากผลดังกล่าวเมื่อพิจารณาวิธีการตรวจสอบทั้ง 2 วิธี ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด พบว่า ภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน วิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ มีค่าเฉลี่ยของอำนาจการทดสอบสูงกว่าภายใต้ขนาดกลุ่มตัวอย่าง 250 คน และ 500 คน ส่วนค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์มีค่าใกล้เคียงกันทั้ง 3 ขนาด

ตอนที่ 2 ผลการเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1

2.1 การเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี

2.1.1 การเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี

ผู้วิจัยนำอำนาจการทดสอบของวิธีการตรวจสอบ 3 วิธี ภายใต้ปัจจัยที่ศึกษา 4 ปัจจัย จากตารางที่ 14 ไปเปรียบเทียบโดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง (two-way ANOVA) เพื่อทดสอบผลของวิธีการตรวจสอบ ผลของปัจจัยที่ศึกษา และผลของปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบและปัจจัยที่ศึกษา ถ้าผลการทดสอบดังกล่าวมีนัยสำคัญทางสถิติ จะใช้การวิเคราะห์อิทธิพลอย่างง่าย (simple effect) ทดสอบในแต่ละระดับของตัวแปร แล้วเปรียบเทียบเป็นรายคู่โดยใช้วิธีการทดสอบของ Bonferroni กรณีค่าความแปรปรวนของทุกกลุ่มเท่ากัน และวิธีการทดสอบของ Games-Howell กรณีค่าความแปรปรวนของทุกกลุ่มไม่เท่ากัน โดยจะทดสอบใน 3 กรณีต่อไปนี้ 1) ทดสอบอำนาจการทดสอบของแต่ละวิธีการตรวจสอบภายใต้ปัจจัยที่ศึกษาต่างระดับกัน 2) ทดสอบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 3 วิธี ภายใต้ปัจจัยที่ศึกษาระดับเดียวกัน และ 3) ทดสอบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 3 วิธี ภายใต้ปัจจัยที่ศึกษาต่างระดับกัน ผลการวิเคราะห์แสดงดังตารางต่อไปนี้

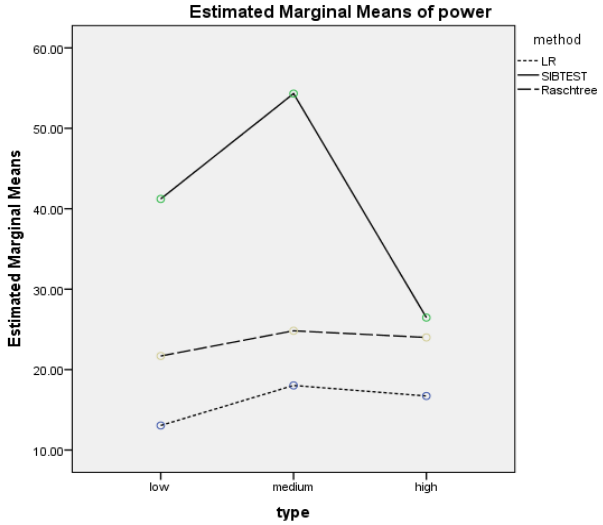
ตารางที่ 20 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบต่างกัน 3 ระดับ

แหล่งของความแปรปรวน	df	SS	MS	F	p-value
ผลหลัก					
วิธีการตรวจสอบ DIF	2	11562.052	5781.026	37.165**	.000
ความยากของข้อสอบ (DI)	2	1901.948	950.974	6.114**	.003
ปฏิสัมพันธ์ 2 ทาง					
วิธีการตรวจสอบ DIF x DI	4	2981.820	745.455	4.792**	.001
ความคลาดเคลื่อน	99	15399.519	155.551		
รวม	107	31845.339			

test of homogeneity of variances (F = 7.802, p = .000)

**p < .01

จากตารางที่ 20 เมื่อพิจารณาผลการทดสอบปฏิสัมพันธ์ 2 ทาง (two way interaction) พบว่า ปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบ DIF และความยากของข้อสอบ (DI) มีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญด้วยวิธีการตรวจสอบต่างกันภายใต้ความยากของข้อสอบต่างกัน จะส่งผลให้อำนาจการทดสอบมีค่าแตกต่างกัน แสดงเป็นกราฟได้ดังนี้



ภาพที่ 12 อำนาจการทดสอบของวิธีการตรวจสอบ 3 วิธี ภายใต้ค่าความยากของข้อสอบต่างกัน 3 ระดับ

ดังนั้น จึงนำอำนาจการทดสอบดังกล่าว ไปเปรียบเทียบความแตกต่างโดยเปรียบเทียบใน 3 กรณี ดังนี้

กรณีที่ 1 เปรียบเทียบอำนาจการทดสอบของแต่ละวิธีการตรวจสอบ 3 วิธี ภายใต้ความยากของข้อสอบต่างกัน 3 ระดับ โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย (simple effect) ผลการวิเคราะห์แสดงดังตารางที่ 21

กรณีที่ 2 เปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 3 วิธี ภายใต้ความยากของข้อสอบเดียวกัน 3 ระดับ โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย (simple effect) ผลการวิเคราะห์แสดงดังตารางที่ 23

กรณีที่ 3 เปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 3 วิธี ภายใต้ความยากของข้อสอบต่างกัน 3 ระดับ โดยใช้วิธีการทดสอบของ Bonferroni และวิธีการทดสอบของ Games-Howell ผลการเปรียบเทียบแสดงดังตารางที่ 25

ตารางที่ 21 ผลการวิเคราะห์อิทธิพลอย่างง่ายของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบต่างกัน 3 ระดับ

แหล่งของความแปรปรวน	df	SS	MS	F	p-value
วิธี LR					
ความยากของข้อสอบ	2	159.910	79.955	2.218	.125
ความคลาดเคลื่อน	33	1189.724	36.052		
รวม	35	1349.634			
วิธี SIBTEST					
ความยากของข้อสอบ	2	4660.633	2330.317	6.638**	.004
ความคลาดเคลื่อน	33	11584.236	351.037		
รวม	35	16244.870			
วิธี Raschtree					
ความยากของข้อสอบ	2	63.225	31.612	.397	.675
ความคลาดเคลื่อน	33	2625.559	79.562		
รวม	35	2688.784			

test of homogeneity of variances (F = 4.060, p = .027)

**p < .01

จากตารางที่ 21 พบว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญ ด้วยวิธีการถดถอยโลจิสติกและวิธีราสช์ทรี ภายใต้ความยากของข้อสอบต่างกัน 3 ระดับ มีค่าไม่แตกต่างกัน ส่วนวิธีซิปเทสท์มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อำนาจการทดสอบโดยการทดสอบระดับนัยสำคัญของวิธีซิปเทสท์ภายใต้ความยากของข้อสอบอย่างน้อย 1 ระดับที่ให้ผลต่างไปจากอำนาจการทดสอบภายใต้ความยากของข้อสอบอื่นๆ ดังนั้น จึงนำอำนาจการทดสอบไปเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Games-Howell ผลการเปรียบเทียบแสดงดังตารางที่ 22

ตารางที่ 22 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญ ด้วยวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบต่างกัน 3 ระดับ โดยใช้วิธีการทดสอบของ Games-Howell

	b_L (41.22)		b_M (54.32)		b_H (26.47)	
DI (M)	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value
b_L (41.22)	-	-	13.10	.312	14.75	.093
b_M (54.32)			-	-	27.85**	.005
b_H (26.47)					-	-

**p < .01

จากตารางที่ 22 พบว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีซิปเทสท์ภายใต้ความยากของข้อสอบ b_M มีค่าสูงสุด รองลงมาได้แก่ ภายใต้ความยากของข้อสอบ b_L และ b_H ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบของวิธีซิปเทสท์ภายใต้ความยากของข้อสอบต่างกัน 3 ระดับ พบว่า อำนาจการทดสอบภายใต้ความยากของข้อสอบ b_L มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ความยากของข้อสอบ b_M และ b_H ส่วนอำนาจการทดสอบภายใต้ความยากของข้อสอบ b_M มีค่าสูงกว่าอำนาจการทดสอบภายใต้ความยากของข้อสอบ b_H อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

จากผลดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีซิปเทสท์ เมื่อความยากของข้อสอบมีค่าความยากปานกลางจะส่งผลให้อำนาจการทดสอบมีค่าสูงสุด ประมาณร้อยละ 54.32 แตกต่างจากภายใต้เงื่อนไขความยากของข้อสอบมีค่าความยากสูง ประมาณร้อยละ 27.85

ตารางที่ 23 ผลการวิเคราะห์อิทธิพลอย่างง่ายของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบเดียวกัน 3 ระดับ

แหล่งของความแปรปรวน	df	SS	MS	F	p-value
b_L					
วิธีการตรวจสอบ	2	4997.596	2498.798	14.740**	.000
ความคลาดเคลื่อน	33	5594.329	169.525		
รวม	35	10591.925			
b_M					
วิธีการตรวจสอบ	2	8930.490	4465.245	21.010**	.000
ความคลาดเคลื่อน	33	7013.558	212.532		
รวม	35	15944.048			
b_H					
วิธีการตรวจสอบ	2	615.786	307.893	3.640*	.037
ความคลาดเคลื่อน	33	2791.632	84.595		
รวม	35	3407.418			

test of homogeneity of variances b_L (F = 14.304, p = .000), b_M (F = 14.870, p = .000), b_H (F = 2.621, p = .088)

**p < .01, *p<.05

จากตารางที่ 23 พบว่า ภายใต้ความยากของข้อสอบ b_L และ b_M อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และภายใต้ความยากของข้อสอบ b_H อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ดังนั้น จึงนำผลการทดสอบที่มีนัยสำคัญทางสถิติไปเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Bonferroni และวิธีการทดสอบของ Games-Howell ผลการเปรียบเทียบแสดงดังตารางที่ 24

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 24 ผลการเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบ b_L b_M และ b_H โดยใช้วิธีการทดสอบของ Bonferroni และวิธีการทดสอบของ Games-Howell

วิธีการตรวจสอบ DIF (M)		LR (13.06)		SIBTEST (41.22)		Raschtree (21.69)	
		Mean difference	p-value	Mean difference	p-value	Mean difference	p-value
b_L	LR (13.06)	-	-	28.16**	.001	8.63*	.038
	SIBTEST (41.22)			-	-	19.53*	.019
	Raschtree (21.69)					-	--
วิธีการตรวจสอบ DIF (M)		LR (18.03)		SIBTEST (54.32)		Raschtree (24.83)	
b_M	LR (18.03)	-	-	36.29**	.000	6.80	.086
	SIBTEST (54.32)			-	-	29.49**	.003
	Raschtree (24.83)					-	-
วิธีการตรวจสอบ DIF (M)		LR (16.72)		SIBTEST (26.47)		Raschtree (24.00)	
b_H	LR (16.72)	-	-	9.75*	.042	7.28	.184
	SIBTEST (26.47)			-	-	2.47	1.000
	Raschtree (24.00)					-	

**p < .01, *p<.05

จากตารางที่ 24 การทดสอบความแตกต่างของอำนาจการทดสอบระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบ b_L b_M และ b_H ปรากฏผลดังนี้

ภายใต้ความยากของข้อสอบ b_L อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีซิปเทสท์มีค่าสูงสุด รองลงมาได้แก่วิธีราสช์ทรีและวิธีการถดถอยโลจิสติก ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อำนาจการทดสอบของวิธีซิปเทสท์มีค่าสูงกว่าอำนาจการทดสอบของวิธีการถดถอยโลจิสติกและวิธีราสช์ทรีอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และ .05 ตามลำดับ และอำนาจการทดสอบของวิธีราสช์ทรีมีค่าสูงกว่าอำนาจการทดสอบของวิธีการถดถอยโลจิสติกอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

ภายใต้ความยากของข้อสอบ b_M อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีซิปเทสท์มีค่าสูงสุด รองลงมาได้แก่วิธีราสช์ทรีและ

วิธีการถดถอยโลจิสติก ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อำนาจการทดสอบของวิธีซิปเทสท์มีค่าสูงกว่าอำนาจการทดสอบของวิธีการถดถอยโลจิสติกและวิธีราสช์ทรีอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนอำนาจการทดสอบระหว่างวิธีการถดถอยโลจิสติกกับวิธีราสช์ทรีมีค่าไม่แตกต่างกัน

ภายใต้ความยากของข้อสอบ b_H อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีซิปเทสท์มีค่าสูงสุด รองลงมาได้แก่วิธีราสช์ทรีและวิธีการถดถอยโลจิสติก ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อำนาจการทดสอบของวิธีซิปเทสท์มีค่าสูงกว่าอำนาจการทดสอบของวิธีการถดถอยโลจิสติกอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบระหว่างวิธีการตรวจสอบคู่อื่น ๆ ที่เหลือมีค่าไม่แตกต่างกัน

จากผลดังกล่าว แสดงว่า ภายใต้ความยากของข้อสอบ b_L b_M และ b_H อำนาจการทดสอบของวิธีซิปเทสท์มีค่าสูงกว่าอำนาจการทดสอบของวิธีการถดถอยโลจิสติกและวิธีราสช์ทรีภายใต้ทุกเงื่อนไขของความยากของข้อสอบ สำหรับอำนาจการทดสอบของวิธีราสช์ทรีมีค่าสูงกว่าอำนาจการทดสอบของวิธีการถดถอยโลจิสติกภายใต้เงื่อนไขความยากของข้อสอบ b_L ส่วนอำนาจการทดสอบภายใต้เงื่อนไขความยากของข้อสอบ b_M และ b_H ทั้ง 2 วิธีมีค่าใกล้เคียงกัน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 25 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทปส์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบต่างกัน 3 ระดับ

วิธีการ	LR		LR		LR		SIBTEST		SIBTEST		SIBTEST		Raschtree		Raschtree		Raschtree	
ตรวจสอบ DIF	(b_L)	13.06	(b_M)	18.03	(b_H)	16.72	(b_L)	41.22	(b_M)	54.32	(b_H)	26.47	(b_L)	21.69	(b_M)	24.83	(b_H)	24.00
	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value
LR (b_L = 13.06)	-	-	4.97	.147	3.66	.198	28.16**	.001	41.26**	.000	13.41**	.002	8.63*	.038	11.77**	.000	10.94**	.001
LR (b_M = 18.03)			-	-	1.31	.882	23.19**	.002	36.29**	.000	8.44*	.047	3.66	.324	6.80	.086	5.97	.099
LR (b_H = 16.72)					-	-	24.50**	.001	37.60**	.000	9.75*	.042	4.97	.153	8.11**	.006	7.28	.184
SIBTEST (b_L = 41.22)							-	-	13.10	.312	14.75	.093	19.53*	.019	16.39*	.017	17.22*	.015
SIBTEST (b_M = 54.32)									-	-	27.85**	.005	32.63**	.000	29.49**	.003	30.32**	.001
SIBTEST (b_H = 26.47)											-	-	4.78	.293	1.64	.683	2.47	1.000
Raschtree (b_L=21.69)													-	-	3.14	1.000	2.31	1.000
Raschtree (b_M=24.83)															-	-	0.83	1.000
Raschtree (b_H=24.00)																	-	-

**p < .01, *p < .05

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตารางที่ 25 การทดสอบความแตกต่างของอำนาจการทดสอบระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ภายใต้ปัจจัยความยากของข้อสอบต่างกัน 3 ระดับ โดยเปรียบเทียบค่าเฉลี่ยของอำนาจการทดสอบซึ่งมีทั้งหมด 9 ค่า เมื่อพิจารณาค่าเฉลี่ยของอำนาจการทดสอบที่มีค่าสูงสุด 3 อันดับแรก พบว่า ค่าเฉลี่ยของอำนาจการทดสอบของวิธีซิปเทสท์ (b_M) มีค่าสูงสุด รองลงมาได้แก่ วิธีซิปเทสท์ (b_L) และวิธีซิปเทสท์ (b_H) ตามลำดับ เมื่อพิจารณาค่าเฉลี่ยของอำนาจการทดสอบที่มีค่าต่ำสุด 3 อันดับแรก พบว่า ค่าเฉลี่ยของอำนาจการทดสอบของวิธีการถดถอยโลจิสติก (b_L) มีค่าต่ำสุด รองลงมาได้แก่ วิธีการถดถอยโลจิสติก (b_H) และวิธีการถดถอยโลจิสติก (b_M) ตามลำดับ ผลการเปรียบเทียบอำนาจการทดสอบของวิธีการตรวจสอบภายใต้ความยากของข้อสอบดังกล่าวกับอำนาจการทดสอบของวิธีการตรวจสอบภายใต้ความยากของข้อสอบอื่น ๆ ปรากฏผลดังนี้

อำนาจการทดสอบของวิธีซิปเทสท์ (b_M) มีค่าไม่แตกต่างกับอำนาจการทดสอบของวิธีซิปเทสท์ (b_L) แต่มีค่าสูงกว่าอำนาจการทดสอบของวิธีการตรวจสอบภายใต้ความยากของข้อสอบอื่น ๆ ที่เหลืออย่างมีนัยสำคัญทางสถิติที่ระดับ .01

อำนาจการทดสอบของวิธีซิปเทสท์ (b_L) มีค่าไม่แตกต่างกับอำนาจการทดสอบของวิธีซิปเทสท์ (b_M) และวิธีซิปเทสท์ (b_H) แต่มีค่าสูงกว่าอำนาจการทดสอบของวิธีการตรวจสอบภายใต้ความยากของข้อสอบอื่น ๆ ที่เหลืออย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และที่ระดับ .05

อำนาจการทดสอบของวิธีซิปเทสท์ (b_H) มีค่าไม่แตกต่างกับอำนาจการทดสอบของวิธีซิปเทสท์ (b_L) วิธีราสช์ทรี (b_L) วิธีราสช์ทรี (b_M) และวิธีราสช์ทรี (b_H) แต่มีค่าสูงกว่าอำนาจการทดสอบของวิธีการถดถอยโลจิสติกและวิธีราสช์ทรีภายใต้ความยากของข้อสอบอื่น ๆ ที่เหลืออย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และที่ระดับ .05

อำนาจการทดสอบของวิธีการถดถอยโลจิสติก (b_L) มีค่าต่ำกว่าอำนาจการทดสอบของวิธีซิปเทสท์ (b_L) วิธีซิปเทสท์ (b_M) วิธีซิปเทสท์ (b_H) วิธีราสช์ทรี (b_M) และวิธีราสช์ทรี (b_H) อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และมีค่าต่ำกว่าวิธีราสช์ทรี (b_L) อย่างมีนัยสำคัญที่ระดับ .05 แต่มีค่าไม่แตกต่างกับวิธีการตรวจสอบภายใต้ความยากของข้อสอบอื่น ๆ ที่เหลือ

อำนาจการทดสอบของวิธีการถดถอยโลจิสติก (b_H) มีค่าต่ำกว่าอำนาจการทดสอบของวิธีซิปเทสท์ (b_L) วิธีซิปเทสท์ (b_M) และวิธีราสช์ทรี (b_M) อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และมีค่าต่ำกว่าวิธีซิปเทสท์ (b_H) อย่างมีนัยสำคัญที่ระดับ .05 แต่มีค่าไม่แตกต่างกับวิธีการตรวจสอบภายใต้ความยากของข้อสอบอื่น ๆ ที่เหลือ

อำนาจการทดสอบของวิธีการถดถอยโลจิสติก (b_M) มีค่าต่ำกว่าอำนาจการทดสอบของวิธีซิปเทสท์ (b_L) และวิธีซิปเทสท์ (b_M) อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และมีค่าต่ำกว่าวิธีซิปเทสท์

(b_H) อย่างมีนัยสำคัญที่ระดับ .05 แต่มีค่าไม่แตกต่างกับวิธีการตรวจสอบภายใต้ความยากของข้อสอบอื่น ๆ ที่เหลือ

จากผลการเปรียบเทียบดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันข้อสอบโดยการทดสอบระดับนัยสำคัญ เมื่อใช้วิธีซิปเทสท์ภายใต้ความยากของข้อสอบที่มีค่าความยากปานกลาง (b_M) จะส่งผลให้อำนาจการทดสอบมีค่าสูงสุด โดยมีค่าสูงกว่าวิธีการถดถอยโลจิสติกและวิธีราสช์ทรีในทุกเงื่อนไขของการตรวจสอบ ส่วนการตรวจสอบด้วยวิธีการถดถอยโลจิสติกภายใต้ความยากของข้อสอบที่มีค่าความยากต่ำ (b_L) จะส่งผลให้อำนาจการทดสอบมีค่าต่ำสุด โดยมีค่าต่ำกว่าวิธีซิปเทสท์และวิธีราสช์ทรีเกือบทุกเงื่อนไขของการตรวจสอบ

ตารางที่ 26 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยความยาวของแบบสอบต่างกัน 2 ขนาด

แหล่งของความแปรปรวน	df	SS	MS	F	p-value
ผลหลัก					
วิธีการตรวจสอบ DIF	2	11562.052	5781.026	29.164**	.000
ความยาวของแบบสอบ (TL)	1	5.070	5.070	.026	.873
ปฏิสัมพันธ์ 2 ทาง					
วิธีการตรวจสอบ DIF x TL	2	59.324	29.662	.150	.861
ความคลาดเคลื่อน	102	20218.893	198.224		
รวม	107	31845.339			

test of homogeneity of variances (F = 11.043, p = .000)

**p < .01

จากตารางที่ 26 เมื่อพิจารณาผลการทดสอบปฏิสัมพันธ์ 2 ทาง (two way interaction) พบว่า ปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบ DIF และความยาวของแบบสอบ (TL) ไม่มีนัยสำคัญทางสถิติ แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญด้วยวิธีการตรวจสอบต่างกันภายใต้ความยาวของแบบสอบต่างกันแล้ว อำนาจการทดสอบมีค่าไม่แตกต่างกัน นั่นคือ วิธีการตรวจสอบและความยาวของแบบสอบไม่มีผลร่วมกันต่ออำนาจการทดสอบ

เมื่อพิจารณาผลการทดสอบของปัจจัยความยาวของแบบสอบ (TL) พบว่า ความยาวของแบบสอบไม่มีนัยสำคัญทางสถิติ แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ภายใต้ความยาวของแบบสอบต่างกันแล้ว อำนาจการทดสอบมีค่าไม่แตกต่างกัน

เมื่อพิจารณาผลการทดสอบของวิธีการตรวจสอบ DIF พบว่า วิธีการตรวจสอบมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญด้วยวิธีการตรวจสอบต่างกัน จะส่งผลให้อำนาจการทดสอบมีค่าแตกต่างกัน ดังนั้น จึงนำอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 3 วิธี ภายใต้ความยาวของแบบสอบทั้ง 2 ขนาด ไปเปรียบเทียบความแตกต่างเป็นรายคู่ โดยใช้วิธีการทดสอบของ Games-Howell ผลการเปรียบเทียบแสดงดังตารางที่ 27

ตารางที่ 27 ผลการเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยความยาวของแบบสอบขนาดเดียวกัน 2 ขนาด โดยใช้วิธีการทดสอบของ Games-Howell

วิธีการตรวจสอบ DIF (M)		LR (15.71)		SIBTEST (41.93)		Raschtree (23.13)	
		Mean difference	p-value	Mean difference	p-value	Mean difference	p-value
40 ข้อ	LR (15.71)	-	-	26.22**	.000	7.42*	.021
	SIBTEST (41.93)			-	-	18.80**	.010
	Raschtree (23.13)					-	--
วิธีการตรวจสอบ DIF (M)		LR (16.17)		SIBTEST (39.41)		Raschtree (23.89)	
60 ข้อ	LR (16.17)	-	-	23.24**	.001	7.72*	.012
	SIBTEST (39.41)			-	-	15.52*	.020
	Raschtree (23.89)					-	-

**p < .01

จากตารางที่ 27 การทดสอบความแตกต่างของอำนาจการทดสอบระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้ปัจจัยของความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ ปรากฏผลดังนี้

ภายใต้ความยาวของแบบสอบ 40 ข้อ อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีซิปเทสท์มีค่าสูงสุด รองลงมาได้แก่ วิธีราสช์ทรีและวิธีการถดถอยโลจิสติก ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อำนาจการทดสอบของวิธีซิปเทสท์มีค่าสูงกว่าอำนาจการทดสอบของวิธีการถดถอยโลจิสติกและวิธีราสช์ทรีอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนอำนาจการทดสอบของวิธีราสช์ทรีมีค่าสูงกว่าอำนาจการทดสอบของวิธีการถดถอยโลจิสติก อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภายใต้ความยาวของแบบสอบ 60 ข้อ อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีซิปเทสท์มีค่าสูงสุด รองลงมาได้แก่ วิธีราสช์ทรีและวิธีการถดถอยโลจิสติก ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อำนาจการทดสอบของวิธีซิปเทสท์มีค่าสูงกว่าอำนาจการทดสอบของวิธีการถดถอยโลจิสติกและวิธีราสช์ทรีอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และที่ระดับ .05 ตามลำดับ ส่วนอำนาจการทดสอบของวิธีราสช์ทรีมีค่าสูงกว่าอำนาจการทดสอบของวิธีการถดถอยโลจิสติก อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

จากผลดังกล่าว แสดงว่า ภายใต้ความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ อำนาจการทดสอบของวิธีซิปเทสท์มีค่าสูงกว่าอำนาจการทดสอบของวิธีการถดถอยโลจิสติกและวิธีราสช์ทรีภายใต้ทุกเงื่อนไขของความยาวของแบบสอบ ส่วนอำนาจการทดสอบของวิธีราสช์ทรีก็มีค่าสูงกว่าอำนาจการทดสอบของวิธีการถดถอยโลจิสติกภายใต้ทุกเงื่อนไขของความยาวของแบบสอบ

ตารางที่ 28 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด

แหล่งของความแปรปรวน	df	SS	MS	F	p-value
ผลหลัก					
วิธีการตรวจสอบ DIF	2	11562.052	5781.026	29.739**	.000
สัดส่วนของ DIF (%DIF)	1	102.161	102.161	.526	.470
ปฏิสัมพันธ์ 2 ทาง					
วิธีการตรวจสอบ DIF x %DIF	2	353.141	176.571	.908	.406
ความคลาดเคลื่อน	102	19827.985	194.392		
รวม	107	31845.339			

test of homogeneity of variances (F = 11.264, p = .000)

*p < .01

จากตารางที่ 28 เมื่อพิจารณาผลการทดสอบปฏิสัมพันธ์ 2 ทาง (two way interaction) พบว่า ปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบ DIF และสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน (%DIF) ไม่มีนัยสำคัญทางสถิติ แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญด้วยวิธีการตรวจสอบต่างกันภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกัน

แล้ว อำนาจการทดสอบมีค่าไม่แตกต่างกัน นั่นคือ วิธีการตรวจสอบและสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน ไม่มีผลร่วมกันต่ออำนาจการทดสอบ

เมื่อพิจารณาผลการทดสอบของปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน (%DIF) พบว่า สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันไม่มีนัยสำคัญทางสถิติ แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ภายใต้ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกันแล้ว อำนาจการทดสอบมีค่าไม่แตกต่างกัน

เมื่อพิจารณาผลการทดสอบของวิธีการตรวจสอบ DIF พบว่า วิธีการตรวจสอบมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญด้วยวิธีการตรวจสอบต่างกัน จะส่งผลให้อำนาจการทดสอบมีค่าแตกต่างกัน ดังนั้น จึงนำอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 3 วิธี ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน ทั้ง 2 ขนาด ไปเปรียบเทียบความแตกต่างเป็นรายคู่ โดยใช้วิธีการทดสอบของ Games-Howell ผลการเปรียบเทียบแสดงดังตารางที่ 29

ตารางที่ 29 ผลการเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันขนาดเดียวกัน 2 ขนาด โดยใช้วิธีการทดสอบของ Games-Howell

วิธีการตรวจสอบ DIF (M)		LR (15.78)		SIBTEST (41.39)		Raschtree (20.03)	
		Mean difference	p-value	Mean difference	p-value	Mean difference	p-value
10%	LR (15.78)	-	-	25.61**	.000	4.25	.208
	SIBTEST (41.39)			-	-	21.36**	.002
	Raschtree (20.03)					-	--
วิธีการตรวจสอบ DIF (M)		LR (16.10)		SIBTEST (39.95)		Raschtree (26.99)	
20%	LR (16.10)	-	-	23.85**	.001	10.89**	.000
	SIBTEST (39.95)			-	-	12.96	.073
	Raschtree (26.99)					-	-

**p < .01

จากตารางที่ 29 การทดสอบความแตกต่างของอำนาจการทดสอบระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และร้อยละ 20 ปรากฏผลดังนี้

ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีซิปเทสท์มีค่าสูงสุด รองลงมาได้แก่ วิธีราสช์ทรีและวิธีการถดถอยโลจิสติก ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อำนาจการทดสอบของวิธีซิปเทสท์มีค่าสูงกว่าอำนาจการทดสอบของวิธีการถดถอยโลจิสติกและวิธีราสช์ทรีอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนอำนาจการทดสอบระหว่างวิธีการถดถอยโลจิสติกกับวิธีราสช์ทรีมีค่าไม่แตกต่างกัน

ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 20 อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีซิปเทสท์มีค่าสูงสุด รองลงมาได้แก่ วิธีราสช์ทรีและวิธีการถดถอยโลจิสติก ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อำนาจการทดสอบของวิธีซิปเทสท์และวิธีราสช์ทรีมีค่าสูงกว่าอำนาจการทดสอบของวิธีการถดถอยโลจิสติกอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนอำนาจการทดสอบระหว่างวิธีซิปเทสท์กับวิธีราสช์ทรีมีค่าไม่แตกต่างกัน

จากผลดังกล่าว แสดงว่า ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 อำนาจการทดสอบของวิธีซิปเทสท์มีค่าสูงกว่าอำนาจการทดสอบของวิธีการถดถอยโลจิสติกและวิธีราสช์ทรี ส่วนภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 20 อำนาจการทดสอบของวิธีซิปเทสท์และวิธีราสช์ทรีมีค่าใกล้เคียงกันและทั้ง 2 วิธี มีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติก

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 30 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด

แหล่งของความแปรปรวน	df	SS	MS	F	p-value
ผลหลัก					
วิธีการตรวจสอบ DIF	2	11562.052	5781.026	75.670**	.000
ขนาดกลุ่มตัวอย่าง (Size)	2	9514.721	4757.360	62.271**	.000
ปฏิสัมพันธ์ 2 ทาง					
วิธีการตรวจสอบ DIF x Size	4	3205.168	801.292	10.488**	.000
ความคลาดเคลื่อน	99	7563.399	76.398		
รวม	107	31845.339			

test of homogeneity of variances (F = 19.100, p = .000)

*p < .01

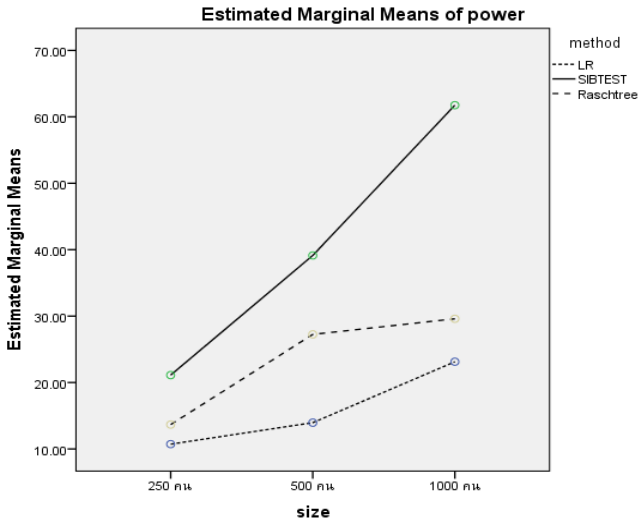
จากตารางที่ 30 เมื่อพิจารณาผลการทดสอบปฏิสัมพันธ์ 2 ทาง (two way interaction) พบว่า ปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบ DIF และขนาดกลุ่มตัวอย่าง (Size) มีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญด้วยวิธีการตรวจสอบต่างกันภายใต้ขนาดกลุ่มตัวอย่างต่างกัน จะส่งผลให้อำนาจการทดสอบมีค่าแตกต่างกัน ดังนั้น จึงนำอำนาจการทดสอบดังกล่าว ไปเปรียบเทียบความแตกต่างโดยเปรียบเทียบใน 3 กรณี ดังนี้

กรณีที่ 1 เปรียบเทียบอำนาจการทดสอบของแต่ละวิธีการตรวจสอบ 3 วิธี ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย (simple effect) ผลการวิเคราะห์แสดงดังตารางที่ 31

กรณีที่ 2 เปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 3 วิธี ภายใต้ขนาดกลุ่มตัวอย่างเดียวกัน 3 ขนาด โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย (simple effect) ผลการวิเคราะห์แสดงดังตารางที่ 35

กรณีที่ 3 เปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 3 วิธี ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้วิธีการทดสอบของ Games-Howell ผลการเปรียบเทียบแสดงดังตารางที่ 37

Chulalongkorn University



ภาพที่ 13 อำนาจการทดสอบของวิธีการตรวจสอบ 3 วิธี ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด

ตารางที่ 31 ผลการวิเคราะห์อิทธิพลอย่างง่ายของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด

แหล่งของความแปรปรวน	df	SS	MS	F	p-value
วิธี LR					
ขนาดกลุ่มตัวอย่าง	2	994.576	497.288	46.219**	.000
ความคลาดเคลื่อน	33	355.058	10.759		
รวม	35	1349.634			
วิธี SIBTEST					
ขนาดกลุ่มตัวอย่าง	2	9951.771	4975.885	26.093**	.000
ความคลาดเคลื่อน	33	6293.099	190.700		
รวม	35	16244.870			
วิธี Raschtree					
ขนาดกลุ่มตัวอย่าง	2	1773.541	886.771	31.973**	.000
ความคลาดเคลื่อน	33	915.243	27.735		
รวม	35	2688.784			

test of homogeneity of variances LR (F = 5.371, p = .010), SIBTEST (F = 9.511, p = .001), Raschtree (F = 2.729, p = .080)

**p < .01

จากตารางที่ 31 การเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญของแต่ละวิธีการตรวจสอบ 3 วิธี ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด ปรากฏผลดังนี้

อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อำนาจการทดสอบของวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่างอย่างน้อย 1 ขนาด ที่ให้ผลต่างไปจากอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่างขนาดอื่นๆ ดังนั้น จึงนำอำนาจการทดสอบไปเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Games-Howell ผลการเปรียบเทียบแสดงดังตารางที่ 32

อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีซิปเทสท์ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อำนาจการทดสอบของวิธีซิปเทสท์ภายใต้ขนาดกลุ่มตัวอย่างอย่างน้อย 1 ขนาด ที่ให้ผลต่างไปจากอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่างขนาดอื่นๆ ดังนั้น จึงนำอำนาจการทดสอบไปเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Games-Howell ผลการเปรียบเทียบแสดงดังตารางที่ 33

อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีราสช์ทรีภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อำนาจการทดสอบของวิธีราสช์ทรีภายใต้ขนาดกลุ่มตัวอย่างอย่างน้อย 1 ขนาด ที่ให้ผลต่างไปจากอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่างขนาดอื่นๆ ดังนั้น จึงนำอำนาจการทดสอบไปเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Bonferroni ผลการเปรียบเทียบแสดงดังตารางที่ 34

ตารางที่ 32 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญ ด้วยวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้วิธีการทดสอบของ Games-Howell

Size (M)	250 คน (10.72)		500 คน (13.96)		1000 คน (23.13)	
	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value
250 คน (10.72)	-	-	3.24*	.012	12.41**	.000
500 คน (13.96)			-	-	9.17**	.000
1000 คน (23.13)					-	-

**p < .01

จากตารางที่ 32 พบว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน มีค่าสูงสุด รองลงมาได้แก่ภายใต้ขนาดกลุ่มตัวอย่าง 500 คน และ 250 คน ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบด้วยวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด พบว่า อำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่าง 500 คน มีค่าสูงกว่าอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่าง 250 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน มีค่าสูงกว่าอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่าง 250 คน และ 500 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

จากผลดังกล่าว แสดงว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีการถดถอยโลจิสติก ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน 500 คน และ 1000 คน มีค่าแตกต่างกัน ถ้าเพิ่มขนาดกลุ่มตัวอย่างจาก 250 คน เป็น 1000 คน จะส่งผลให้อำนาจการทดสอบมีค่าเพิ่มขึ้น ประมาณร้อยละ 12.41

ตารางที่ 33 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญ ด้วยวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้วิธีการทดสอบของ Games-Howell

Size (M)	250 คน (21.12)		500 คน (39.13)		1000 คน (61.76)	
	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value
250 คน (21.12)	-	-	18.01**	.002	40.64**	.000
500 คน (39.13)			-	-	22.63**	.007
1000 คน (61.76)					-	-

**p < .01

จากตารางที่ 33 พบว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีซิปเทสท์ภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน มีค่าสูงสุด รองลงมาได้แก่ภายใต้ขนาดกลุ่มตัวอย่าง 500 คน และ 250 คน ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบด้วยวิธีซิปเทสท์ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด พบว่า อำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน มีค่าสูงกว่าอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่าง 250 คน และ 500 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่าง 500 คน มีค่าสูงกว่าอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่าง 250 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

จากผลดังกล่าว แสดงว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีซิปเทสท์ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน 500 คน และ 1000 คน มีค่าแตกต่างกัน ถ้าเพิ่มขนาดกลุ่มตัวอย่างจาก 250 คน เป็น 500 คน และจาก 250 คน เป็น 1000 คน จะส่งผลให้อำนาจการทดสอบมีค่าเพิ่มขึ้น ประมาณร้อยละ 18.01 และร้อยละ 40.64 ตามลำดับ

ตารางที่ 34 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญ ด้วยวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้วิธีการทดสอบของ Bonferroni

Size (M)	250 คน (13.67)		500 คน (27.25)		1000 คน (29.60)	
	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value
250 คน (13.67)	-	-	13.58**	.000	15.93**	.000
500 คน (27.25)			-	-	2.35	.849
1000 คน (29.60)					-	-

**p < .01

จากตารางที่ 34 พบว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีราสช์ทรีภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน มีค่าสูงสุด รองลงมาได้แก่ภายใต้ขนาดกลุ่มตัวอย่าง 500 คน และ 250 คน ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบด้วยวิธีราสช์ทรีภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด พบว่า อำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่าง 500 คน มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน ส่วนอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่าง 500 คน และ 1000 คน มีค่าสูงกว่าอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่าง 250 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

จากผลดังกล่าว แสดงว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีราสช์ทรีภายใต้ขนาดกลุ่มตัวอย่าง 500 คน และ 1000 คน มีค่าใกล้เคียงกัน แต่ถ้าเพิ่มขนาดกลุ่มตัวอย่างจาก 250 คน เป็น 500 คน และจาก 250 คน เป็น 1000 คน จะส่งผลให้อำนาจการทดสอบมีค่าเพิ่มขึ้น ประมาณร้อยละ 13.58 และร้อยละ 15.93 ตามลำดับ

ตารางที่ 35 ผลการวิเคราะห์อิทธิพลอย่างง่ายของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยกลุ่มตัวอย่างขนาดเดียวกัน 3 ขนาด

แหล่งของความแปรปรวน	df	SS	MS	F	p-value
250 คน					
วิธีการตรวจสอบ	2	688.835	344.418	10.793**	.000
ความคลาดเคลื่อน	33	1053.083	31.912		
รวม	35	1741.918			
500 คน					
วิธีการตรวจสอบ	2	3806.404	1903.202	27.326**	.000
ความคลาดเคลื่อน	33	2298.403	69.649		
รวม	35	6104.806			
1000 คน					
วิธีการตรวจสอบ	2	10271.981	5135.991	40.240**	.000
ความคลาดเคลื่อน	33	4211.913	127.634		
รวม	35	14483.894			

test of homogeneity of variances 250 คน (F = 10.253, p = .000), 500 คน (F = 15.926, p = .000), 1000 คน (F = 33.432, p = .000)

**p < .01

จากตารางที่ 35 พบว่า ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน 500 คน และ 1000 คน อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรีมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ดังนั้น จึงนำผลการทดสอบภายใต้ขนาดกลุ่มตัวอย่างทั้ง 3 ขนาด ไปเปรียบเทียบความแตกต่างเป็นรายคู่ โดยใช้วิธีการทดสอบของ Games-Howell ผลการเปรียบเทียบแสดงดังตารางที่ 36

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 36 ผลการเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของขนาดกลุ่มตัวอย่างขนาดเดียวกัน 3 ขนาด โดยใช้วิธีการทดสอบของ Games-Howell

วิธีการตรวจสอบ DIF (M)		LR (10.72)		SIBTEST (21.12)		Raschtree (13.67)	
		Mean difference	p-value	Mean difference	p-value	Mean difference	p-value
250 คน	LR (10.72)	-	-	10.40**	.001	2.95	.305
	SIBTEST (21.12)			-	-	7.45*	.035
	Raschtree (13.67)					-	--
วิธีการตรวจสอบ DIF (M)		LR (13.96)		SIBTEST (39.13)		Raschtree (27.25)	
500 คน	LR (13.96)	-	-	25.17**	.000	13.29**	.000
	SIBTEST (39.13)			-	-	11.88*	.028
	Raschtree (27.25)					-	-
วิธีการตรวจสอบ DIF (M)		LR (23.13)		SIBTEST (61.76)		Raschtree (29.60)	
1000 คน	LR (23.13)	-	-	38.63**	.000	6.47**	.003
	SIBTEST (61.76)			-	-	32.16**	.000
	Raschtree (29.60)					-	

**p < .01

จากตารางที่ 36 การทดสอบความแตกต่างของอำนาจการทดสอบระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ภายใต้ปัจจัยขนาดกลุ่มตัวอย่างขนาดเดียวกัน 3 ขนาด ปรากฏผลดังนี้

ภายใต้ขนาดกลุ่มตัวอย่าง 500 คน และ 1000 คน ได้ผลสอดคล้องกัน กล่าวคือ อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีซิปเทสท์มีค่าสูงสุด รองลงมาได้แก่ วิธีราสช์ทรีและวิธีการถดถอยโลจิสติกตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อำนาจการทดสอบของวิธีซิปเทสท์มีค่าสูงกว่าอำนาจการทดสอบของวิธีการถดถอยโลจิสติกและวิธีราสช์ทรีอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และ .05 ส่วนอำนาจการทดสอบของวิธีราสช์ทรีมีค่าสูงกว่าอำนาจการทดสอบของวิธีการถดถอยโลจิสติกอย่างมีนัยสำคัญที่ระดับ .01

ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญของวิธีซิปเทสท์มีค่าสูงสุด รองลงมาได้แก่ วิธีราสช์ทรีและวิธีการถดถอยโลจิสติกตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ

ดังกล่าว พบว่า อำนาจการทดสอบของวิธีซิปเทสท์มีค่าสูงกว่าอำนาจการทดสอบของวิธีการถดถอยโลจิสติกและวิธีราสช์ทรีอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และ .05 ตามลำดับ ส่วนอำนาจการทดสอบของวิธีการถดถอยโลจิสติกกับวิธีราสช์ทรีมีค่าไม่แตกต่างกัน

จากผลดังกล่าว แสดงว่า ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน 500 คน และ 1000 คน อำนาจการทดสอบของซิปเทสท์มีค่าสูงกว่าวิธีการถดถอยโลจิสติกและวิธีราสช์ทรี ส่วนอำนาจการทดสอบของวิธีราสช์ทรีมีค่าสูงกว่าวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่าง 500 คน และ 1000 คน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 37 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด

วิธีการ	LR		LR		LR		SIBTEST		SIBTEST		SIBTEST		Raschtree		Raschtree		Raschtree	
ตรวจสอบ DIF	250 คน	10.72	500 คน	13.96	1000 คน	23.13	250 คน	21.12	500 คน	39.13	1000	61.76	250 คน	13.67	500 คน	27.25	1000	29.60
	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value
LR (250 = 10.72)	-	-	3.24*	.012	12.41**	.000	10.40**	.001	28.41**	.000	51.04**	.000	2.95	.305	16.53**	.000	18.88**	.000
LR (500 = 13.96)			-	-	9.17**	.000	7.16**	.006	25.17**	.000	47.80**	.000	0.29	.889	13.29**	.000	15.64**	.000
LR (1000 = 23.13)					-	-	2.01	.419	16.00**	.001	38.63**	.000	9.46**	.000	4.12	.055	6.47**	.003
SIBTEST (250 = 21.12)							-	-	18.01**	.002	40.64**	.000	7.45*	.035	6.13*	.028	8.48**	.002
SIBTEST (500 = 39.13)									-	-	22.63**	.007	25.45**	.000	11.88*	.028	9.53*	.031
SIBTEST (1000=61.76)											-	-	48.09**	.000	34.51**	.000	32.16**	.000
Raschtree (250=13.67)													-	-	13.58**	.000	15.93**	.000
Raschtree (500=27.25)															-	-	2.35	.849
Raschtree (1000=29.60)																	-	-

**p < .01, *p < .05

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตารางที่ 37 การทดสอบความแตกต่างของอำนาจการทดสอบระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ภายใต้ปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยเปรียบเทียบค่าเฉลี่ยของอำนาจการทดสอบซึ่งมีทั้งหมด 9 ค่า เมื่อพิจารณาค่าเฉลี่ยของอำนาจการทดสอบที่มีค่าสูงสุด 3 อันดับแรก พบว่า ค่าเฉลี่ยของอำนาจการทดสอบของวิธีซิปเทสท์ (1000 คน) มีค่าสูงสุด รองลงมาได้แก่ วิธีซิปเทสท์ (500 คน) และวิธีราสช์ทรี (1000 คน) ตามลำดับ เมื่อพิจารณาค่าเฉลี่ยของอำนาจการทดสอบที่มีค่าต่ำสุด 3 อันดับแรก พบว่า ค่าเฉลี่ยของอำนาจการทดสอบของวิธีการถดถอยโลจิสติก (250 คน) มีค่าต่ำสุด รองลงมาได้แก่ วิธีราสช์ทรี (250 คน) และวิธีการถดถอยโลจิสติก (500 คน) ตามลำดับ ผลการเปรียบเทียบอำนาจการทดสอบของวิธีการตรวจสอบภายใต้ขนาดกลุ่มตัวอย่างดังกล่าวกับอำนาจการทดสอบของวิธีการตรวจสอบภายใต้ขนาดกลุ่มตัวอย่างอื่น ๆ ปรากฏผลดังนี้

อำนาจการทดสอบของวิธีซิปเทสท์ (1000 คน) มีค่าสูงกว่าวิธีการตรวจสอบภายใต้ขนาดกลุ่มตัวอย่างขนาดอื่น ๆ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

อำนาจการทดสอบของวิธีซิปเทสท์ (500 คน) มีค่าสูงกว่าวิธีการตรวจสอบภายใต้ขนาดกลุ่มตัวอย่างขนาดอื่น ๆ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และที่ระดับ .05

อำนาจการทดสอบของวิธีราสช์ทรี (1000 คน) มีค่าไม่แตกต่างกับอำนาจการทดสอบของวิธีราสช์ทรี (500 คน) แต่มีค่าสูงกว่าวิธีการตรวจสอบภายใต้ขนาดกลุ่มตัวอย่างขนาดอื่น ๆ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และที่ระดับ .05

อำนาจการทดสอบของวิธีการถดถอยโลจิสติก (250 คน) มีค่าต่ำกว่าอำนาจการทดสอบของวิธีการถดถอยโลจิสติก (1000 คน) วิธีซิปเทสท์ (250 คน) วิธีซิปเทสท์ (500 คน) วิธีซิปเทสท์ (1000 คน) วิธีราสช์ทรี (500 คน) และวิธีราสช์ทรี (1000 คน) อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แต่มีค่าไม่แตกต่างกับวิธีการตรวจสอบภายใต้ขนาดกลุ่มตัวอย่างขนาดอื่น ๆ ที่เหลือ

อำนาจการทดสอบของวิธีราสช์ทรี (250 คน) มีค่าต่ำกว่าอำนาจการทดสอบของวิธีการถดถอยโลจิสติก (1000 คน) วิธีซิปเทสท์ (500 คน) วิธีซิปเทสท์ (1000 คน) วิธีราสช์ทรี (500 คน) และวิธีราสช์ทรี (1000 คน) อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และมีค่าต่ำกว่าอำนาจการทดสอบของวิธีซิปเทสท์ (250 คน) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับวิธีการตรวจสอบภายใต้ขนาดกลุ่มตัวอย่างขนาดอื่น ๆ ที่เหลือ

อำนาจการทดสอบของวิธีการถดถอยโลจิสติก (500 คน) มีค่าต่ำกว่าอำนาจการทดสอบของวิธีการถดถอยโลจิสติก (1000 คน) วิธีซิปเทสท์ (250 คน) วิธีซิปเทสท์ (500 คน) วิธีซิปเทสท์ (1000 คน) วิธีราสช์ทรี (500 คน) และวิธีราสช์ทรี (1000 คน) อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แต่มีค่าไม่แตกต่างกับวิธีการตรวจสอบภายใต้ขนาดกลุ่มตัวอย่างขนาดอื่น ๆ ที่เหลือ

จากผลการเปรียบเทียบดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญ เมื่อใช้วิธีซิปเทสท์ภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน จะส่งผลให้อำนาจการทดสอบมีค่าสูงสุด โดยมีค่าสูงกว่าวิธีการถดถอยโลจิสติกและวิธีราสช์ทรีในทุกเงื่อนไขของการตรวจสอบ ส่วนการตรวจสอบด้วยวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่าง 250 คน จะส่งผลให้อำนาจการทดสอบมีค่าต่ำสุด โดยมีค่าต่ำกว่าวิธีซิปเทสท์และวิธีราสช์ทรีเกือบทุกเงื่อนไขของการตรวจสอบ

2.1.2 การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี

ผู้วิจัยนำอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบ 3 วิธี ภายใต้ปัจจัยที่ศึกษา 4 ปัจจัย จากตารางที่ 14 ไปเปรียบเทียบโดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง (two-way ANOVA) เพื่อทดสอบผลของวิธีการตรวจสอบ ผลของปัจจัยที่ศึกษา และผลของปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบและปัจจัยที่ศึกษา ถ้าผลการทดสอบดังกล่าวมีนัยสำคัญทางสถิติจะใช้การวิเคราะห์อิทธิพลอย่างง่าย (simple effect) ทดสอบในแต่ละระดับของตัวแปร แล้วเปรียบเทียบเป็นรายคู่โดยใช้วิธีการทดสอบของ Bonferroni กรณีค่าความแปรปรวนของทุกกลุ่มเท่ากัน และวิธีการทดสอบของ Games-Howell กรณีค่าความแปรปรวนของทุกกลุ่มไม่เท่ากัน โดยจะทดสอบใน 3 กรณีต่อไปนี้ 1) ทดสอบอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธีการตรวจสอบภายใต้ปัจจัยที่ศึกษาต่างระดับกัน 2) ทดสอบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 3 วิธี ภายใต้ปัจจัยที่ศึกษาระดับเดียวกัน และ 3) ทดสอบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 3 วิธี ภายใต้ปัจจัยที่ศึกษาต่างระดับกัน ผลการวิเคราะห์แสดงดังตารางต่อไปนี้

ตารางที่ 38 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบต่างกัน 3 ระดับ

แหล่งของความแปรปรวน	df	SS	MS	F	p-value
ผลหลัก					
วิธีการตรวจสอบ DIF	2	299.582	149.791	85.724**	.000
ความยากของข้อสอบ (DI)	2	12.316	6.158	3.524*	.033
ปฏิสัมพันธ์ 2 ทาง					
วิธีการตรวจสอบ DIF x DI	4	7.471	1.868	1.069	.376
ความคลาดเคลื่อน	99	172.990	1.747		
รวม	107	492.359			

test of homogeneity of variances (F = 4.511, p = .000)

**p < .01, *p < .05

จากตารางที่ 38 เมื่อพิจารณาผลการทดสอบปฏิสัมพันธ์ 2 ทาง (two way interaction) พบว่า ปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบ DIF และความยากของข้อสอบ (DI) ไม่มีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญด้วยวิธีการตรวจสอบต่างกันภายใต้ความยากของข้อสอบต่างกัน อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าไม่แตกต่างกัน นั่นคือ วิธีการตรวจสอบและความยากของข้อสอบไม่มีผลร่วมกันต่ออัตราความคลาดเคลื่อนประเภทที่ 1

เมื่อพิจารณาผลการทดสอบของปัจจัยความยากของข้อสอบ (DI) พบว่า ความยากของข้อสอบมีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญภายใต้ความยากของข้อสอบต่างกัน จะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าแตกต่างกัน ดังนั้น จึงนำอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธีการตรวจสอบ 3 วิธี ภายใต้ความยากของข้อสอบต่างกัน 3 ระดับ ไปเปรียบเทียบความแตกต่างเป็นรายคู่ โดยใช้วิธีการทดสอบของ Bonferroni และวิธีการทดสอบของ Games-Howell ผลการเปรียบเทียบแสดงดังตารางที่ 39

เมื่อพิจารณาผลการทดสอบของวิธีการตรวจสอบ DIF พบว่า วิธีการตรวจสอบ DIF มีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีการตรวจสอบต่างกัน จะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1

มีค่าแตกต่างกัน ดังนั้น จึงนำอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธีการตรวจสอบ 3 วิธี ภายใต้ความยากของข้อสอบเดียวกัน 3 ระดับ ไปเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Bonferroni และวิธีการทดสอบของ Games-Howell ผลการเปรียบเทียบแสดงดังตารางที่ 40

ตารางที่ 39 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบต่างกัน 3 ระดับ โดยใช้วิธีการทดสอบของ Bonferroni และวิธีการทดสอบของ Games-Howell

DI (M)		b_L (6.52)		b_M (5.82)		b_H (6.34)	
		Mean difference	p-value	Mean difference	p-value	Mean difference	p-value
LR	b_L (6.52)	-	-	0.70*	.025	0.18	.762
	b_M (5.82)			-	-	0.52	.249
	b_H (6.34)					-	-
DI (M)		b_L (5.48)		b_M (5.21)		b_H (5.22)	
SIBTEST	b_L (5.48)	-	-	0.27	1.000	0.26	1.000
	b_M (5.21)			-	-	0.01	1.000
	b_H (5.22)					-	-
DI (M)		b_L (2.63)		b_M (1.40)		b_H (2.95)	
Raschtree	b_L (2.63)	-	-	1.23	.414	0.32	1.000
	b_M (1.40)			-	-	1.55	.194
	b_H (2.95)					-	-

test of homogeneity of variances LR (F = 6.268, p = .005), SIBTEST (F = .311, p = .735), Raschtree (F = .613, p = .548)

*p < .05

จากตารางที่ 39 การทดสอบความแตกต่างของอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบ b_L b_M และ b_H ปรากฏผลดังนี้

อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีการถดถอยโลจิสติก ภายใต้ความยากของข้อสอบ b_L มีค่าสูงสุด

รองลงมาได้แก่ ภายใต้ความยากของข้อสอบ b_H และ b_M ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกภายใต้ความยากของข้อสอบต่างกัน 3 ระดับ พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ความยากของข้อสอบ b_L มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ความยากของข้อสอบ b_M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ความยากของข้อสอบ b_L กับ b_H และภายใต้ความยากของข้อสอบ b_M กับ b_H มีค่าไม่แตกต่างกัน

อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีซิปเทสท์ และวิธีราสช์ทรี ได้ผลสอดคล้องกัน กล่าวคือ อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ความยากของข้อสอบ b_L กับ b_M ภายใต้ความยากของข้อสอบ b_L กับ b_H และภายใต้ความยากของข้อสอบ b_M กับ b_H มีค่าไม่แตกต่างกัน

จากผลดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีการถดถอยโลจิสติก เมื่อความยากของข้อสอบมีค่าความยากต่ำจะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าสูงสุด ประมาณร้อยละ 6.52 แตกต่างจากภายใต้เงื่อนไขความยากของข้อสอบมีค่าความยากปานกลาง ประมาณร้อยละ 0.70 ส่วนวิธีซิปเทสท์และวิธีราสช์ทรีภายใต้ความยากข้อสอบที่มีค่าความยากต่ำ ความยากปานกลาง และความยากสูง ของแต่ละวิธี มีอัตราความคลาดเคลื่อนประเภทที่ 1 ใกล้เคียงกัน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 40 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบ b_L b_M และ b_H โดยใช้วิธีการทดสอบของ Bonferroni และวิธีการทดสอบของ Games-Howell

วิธีการตรวจสอบ DIF (M)		LR (6.52)		SIBTEST (5.48)		Raschtree (2.63)	
		Mean difference	p-value	Mean difference	p-value	Mean difference	p-value
b_L	LR (6.52)	-	-	1.04**	.005	3.89**	.000
	SIBTEST (5.48)			-	-	2.85**	.001
	Raschtree (2.63)					-	--
วิธีการตรวจสอบ DIF (M)		LR (5.82)		SIBTEST (5.21)		Raschtree (1.40)	
b_M	LR (5.82)	-	-	0.61	.558	4.42**	.000
	SIBTEST (5.21)			-	-	3.81**	.000
	Raschtree (1.40)					-	-
วิธีการตรวจสอบ DIF (M)		LR (6.34)		SIBTEST (5.22)		Raschtree (2.95)	
b_H	LR (6.34)	-	-	1.12	.246	3.39**	.000
	SIBTEST (5.22)			-	-	2.27**	.003
	Raschtree (2.95)					-	

test of homogeneity of variances b_L (F = 14.636, p = .000), b_M (F = 2.883, p = .070), b_H (F = 3.015, p = .063)

**p < .01

จากตารางที่ 40 การทดสอบความแตกต่างของอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบ b_L b_M และ b_H ปรากฏผลดังนี้

ภายใต้ความยากของข้อสอบ b_L อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีการถดถอยโลจิสติกมีค่าสูงสุด รองลงมาได้แก่วิธีซิปเทสท์และวิธีราสช์ทรี ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์และวิธีราสช์ทรี อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีราสช์ทรี อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

ภายใต้ความยากของข้อสอบ b_M และ b_H ได้ผลสอดคล้องกัน กล่าวคือ อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีการถดถอยโลจิสติกมีค่าสูงสุด รองลงมาได้แก่วิธีซิปเทสท์และวิธีราสช์ทรี ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีราสช์ทรี อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

จากผลดังกล่าว แสดงว่า ภายใต้ความยากของข้อสอบ b_L อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าวิธีซิปเทสท์และวิธีราสช์ทรี ส่วนภายใต้ความยากของข้อสอบ b_M และ b_H อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์มีค่าสูงใกล้เคียงกัน ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีราสช์ทรีมีค่าต่ำภายใต้ทุกเงื่อนไขของความยากของข้อสอบ

ตารางที่ 41 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยความยาวของแบบสอบต่างกัน 2 ขนาด

แหล่งของความแปรปรวน	df	SS	MS	F	p-value
ผลหลัก					
วิธีการตรวจสอบ DIF	2	299.582	149.791	79.575**	.000
ความยาวของแบบสอบ (TL)	1	.639	.639	.340	.561
ปฏิสัมพันธ์ 2 ทาง					
วิธีการตรวจสอบ DIF x TL	2	.133	.066	.035	.965
ความคลาดเคลื่อน	102	192.005	1.882		
รวม	107	492.359			

test of homogeneity of variances (F = 8.032, p = .000)

**p < .01

จากตารางที่ 41 เมื่อพิจารณาผลการทดสอบปฏิสัมพันธ์ 2 ทาง (two way interaction) พบว่า ปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบ DIF และความยาวของแบบสอบ (TL) ไม่มีนัยสำคัญทางสถิติ แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญด้วย

วิธีการตรวจสอบต่างกันภายใต้ความยาวของแบบสอบต่างกันแล้ว อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าไม่แตกต่างกัน นั่นคือ วิธีการตรวจสอบและความยาวของแบบสอบไม่มีผลร่วมกันต่ออัตราความคลาดเคลื่อนประเภทที่ 1

เมื่อพิจารณาผลการทดสอบของปัจจัยความยาวของแบบสอบ (TL) พบว่า ความยาวของแบบสอบไม่มีนัยสำคัญทางสถิติ แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ภายใต้ความยาวของแบบสอบต่างกันแล้ว อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าไม่แตกต่างกัน

เมื่อพิจารณาผลการทดสอบของวิธีการตรวจสอบ DIF พบว่า วิธีการตรวจสอบมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญด้วยวิธีการตรวจสอบต่างกัน จะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าแตกต่างกัน ดังนั้น จึงนำอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 3 วิธี ภายใต้ความยาวของแบบสอบทั้ง 2 ขนาด ไปเปรียบเทียบความแตกต่างเป็นรายคู่ โดยใช้วิธีการทดสอบของ Games-Howell ผลการเปรียบเทียบแสดงดังตารางที่ 42

ตารางที่ 42 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยความยาวของแบบสอบขนาดเดียวกัน 2 ขนาด โดยใช้วิธีการทดสอบของ Games-Howell

วิธีการตรวจสอบ DIF (M)		LR (6.20)		SIBTEST (5.22)		Raschtree (2.21)	
		Mean difference	p-value	Mean difference	p-value	Mean difference	p-value
40 ข้อ	LR (6.20)	-	-	0.98**	.002	3.99**	.000
	SIBTEST (5.22)			-	-	3.01**	.000
	Raschtree (2.21)					-	--
วิธีการตรวจสอบ DIF (M)		LR (6.26)		SIBTEST (5.39)		Raschtree (2.44)	
60 ข้อ	LR (6.26)	-	-	0.87*	.016	3.82**	.000
	SIBTEST (5.39)			-	-	2.95**	.000
	Raschtree (2.44)					-	-

test of homogeneity of variances 40 ข้อ (F = 10.733, p = .000), 60 ข้อ (F = 9.381, p = .000)

**p < .01

จากตารางที่ 42 การทดสอบความแตกต่างของอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้ปัจจัยของความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ ปรากฏผลดังนี้

ภายใต้ความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ ได้ผลสอดคล้องกัน กล่าวคือ อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีการถดถอยโลจิสติกมีค่าสูงสุด รองลงมาได้แก่ วิธีซิปเทสท์และวิธีราสช์ทรี ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีราสช์ทรีอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าวิธีซิปเทสท์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และ .05

จากผลดังกล่าว แสดงว่า ภายใต้ความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าวิธีซิปเทสท์และวิธีราสช์ทรี ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีราสช์ทรีมีค่าต่ำภายใต้ทุกเงื่อนไขของความยาวของแบบสอบ

ตารางที่ 43 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด

แหล่งของความแปรปรวน	df	SS	MS	F	p-value
ผลหลัก					
วิธีการตรวจสอบ DIF	2	299.582	149.791	84.951**	.000
สัดส่วนของ DIF (%DIF)	1	5.391	5.391	3.058	.083
ปฏิสัมพันธ์ 2 ทาง					
วิธีการตรวจสอบ DIF x %DIF	2	7.534	3.767	2.136	.123
ความคลาดเคลื่อน	102	179.852	1.763		
รวม	107	492.359			

test of homogeneity of variances (F = 13.368, p = .000)

*p < .01

จากตารางที่ 43 เมื่อพิจารณาผลการทดสอบปฏิสัมพันธ์ 2 ทาง (two way interaction) พบว่า ปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบ DIF และสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน (%DIF) ไม่มีนัยสำคัญทางสถิติ แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญด้วยวิธีการตรวจสอบต่างกันภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีขนาดต่างกันแล้ว อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าไม่แตกต่างกัน นั่นคือ วิธีการตรวจสอบและสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน ไม่มีผลร่วมกันต่ออัตราความคลาดเคลื่อนประเภทที่ 1

เมื่อพิจารณาผลการทดสอบของสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน (%DIF) พบว่า สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันไม่มีนัยสำคัญทางสถิติ แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีขนาดต่างกันแล้ว อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าไม่แตกต่างกัน

เมื่อพิจารณาผลการทดสอบของวิธีการตรวจสอบ DIF พบว่า วิธีการตรวจสอบมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญด้วยวิธีการตรวจสอบต่างกัน จะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าแตกต่างกัน ดังนั้น จึงนำอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 3 วิธี ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันทั้ง 2 ขนาด ไปเปรียบเทียบความแตกต่างเป็นรายคู่ โดยใช้วิธีการทดสอบของ Games-Howell ผลการเปรียบเทียบแสดงดังตารางที่ 44

ตารางที่ 44 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันขนาดเดียวกัน 2 ขนาด โดยใช้วิธีการทดสอบของ Games-Howell

วิธีการตรวจสอบ DIF (M)		LR (6.24)		SIBTEST (5.22)		Raschtree (1.73)	
		Mean difference	p-value	Mean difference	p-value	Mean difference	p-value
10%	LR (6.24)	-	-	1.02**	.001	4.51**	.000
	SIBTEST (5.22)			-	-	3.49**	.000
	Raschtree (1.73)					-	--
วิธีการตรวจสอบ DIF (M)		LR (6.22)		SIBTEST (5.39)		Raschtree (2.92)	
20%	LR (6.22)	-	-	0.83*	.023	3.30**	.000
	SIBTEST (5.39)			-	-	2.47**	.002
	Raschtree (2.92)					-	-

test of homogeneity of variances 10% (F = 6.628, p = .003), 20% (F = 17.339, p = .000)

**p < .01

จากตารางที่ 44 การทดสอบความแตกต่างของอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 10% และ 20% ปรากฏผลดังนี้

ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และร้อยละ 20 ได้ผลสอดคล้องกัน กล่าวคือ อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีการถดถอย โลจิสติกมีค่าสูงสุด รองลงมาได้แก่ วิธีซิปเทสท์และวิธีราสช์ทรี ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีราสช์ทรีอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และ .05

จากผลดังกล่าว แสดงว่า ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และร้อยละ 20 อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์และวิธีราสช์ทรี ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีราสช์ทรีมีค่าต่ำภายใต้ทุกเงื่อนไขของสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน

ตารางที่ 45 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด

แหล่งของความแปรปรวน	df	SS	MS	F	p-value
ผลหลัก					
วิธีการตรวจสอบ DIF	2	299.582	149.791	138.410**	.000
ขนาดกลุ่มตัวอย่าง (Size)	2	30.459	15.230	14.072**	.000
ปฏิสัมพันธ์ 2 ทาง					
วิธีการตรวจสอบ DIF x Size	4	55.177	13.794	12.746**	.000
ความคลาดเคลื่อน	99	107.140	1.082		
รวม	107	492.359			

test of homogeneity of variances (F = 4.447, p = .000)

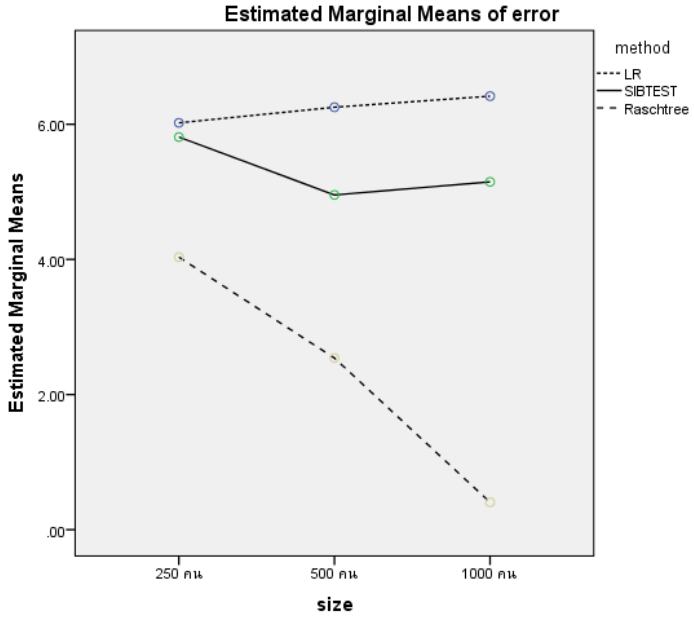
*p < .01

จากตารางที่ 45 เมื่อพิจารณาผลการทดสอบปฏิสัมพันธ์ 2 ทาง (two way interaction) พบว่า ปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบ DIF และขนาดกลุ่มตัวอย่าง (Size) มีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญด้วยวิธีการตรวจสอบต่างกันภายใต้ขนาดกลุ่มตัวอย่างต่างกัน จะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าแตกต่างกัน ดังนั้น จึงนำอัตราความคลาดเคลื่อนประเภทที่ 1 ดังกล่าว ไปเปรียบเทียบความแตกต่างโดยเปรียบเทียบใน 3 กรณี ดังนี้

กรณีที่ 1 เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธีการตรวจสอบ 3 วิธี ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย (simple effect) ผลการวิเคราะห์แสดงดังตารางที่ 46

กรณีที่ 2 เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 3 วิธี ภายใต้ขนาดกลุ่มตัวอย่างเดียวกัน 3 ขนาด โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย (simple effect) ผลการวิเคราะห์แสดงดังตารางที่ 49

กรณีที่ 3 เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 3 วิธี ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้วิธีการทดสอบเป็นรายคู่ของ Games-Howell ผลการเปรียบเทียบแสดงดังตารางที่ 51



ภาพที่ 14 อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบ 3 วิธี ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด

ตารางที่ 46 ผลการวิเคราะห์อิทธิพลอย่างง่ายของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด

แหล่งของความแปรปรวน	df	SS	MS	F	p-value
วิธี LR					
ขนาดกลุ่มตัวอย่าง	2	.945	.473	.917	.410
ความคลาดเคลื่อน	33	17.009	.515		
รวม	35	17.954			
วิธี SIBTEST					
ขนาดกลุ่มตัวอย่าง	2	4.845	2.423	3.302*	.049
ความคลาดเคลื่อน	33	24.214	.734		
รวม	35	29.059			
วิธี Raschtree					
ขนาดกลุ่มตัวอย่าง	2	79.846	39.923	19.986**	.000
ความคลาดเคลื่อน	33	65.918	1.998		
รวม	35	145.764			

test of homogeneity of variances SIBTEST (F = .813, p = .452), Raschtree (F = 6.606, p = .004)

**p < .01, *p < .05

จากตารางที่ 46 การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญของแต่ละวิธีการตรวจสอบ 3 วิธี ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด ปรากฏผลดังนี้

อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด มีค่าไม่แตกต่างกัน แสดงว่า ขนาดกลุ่มตัวอย่างไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก

อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีซิปเทสท์ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่า อำนาจการทดสอบของวิธีซิปเทสท์ภายใต้ขนาดกลุ่มตัวอย่างอย่างน้อย 1 ขนาด ที่ให้ผลต่างไปจากอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ขนาดกลุ่ม

ตัวอย่างขนาดอื่นๆ ดังนั้น จึงนำอัตราความคลาดเคลื่อนประเภทที่ 1 ไปเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Bonferroni ผลการเปรียบเทียบแสดงดังตารางที่ 47

อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีราสช์ทรีภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีราสช์ทรีภายใต้ขนาดกลุ่มตัวอย่างอย่างน้อย 1 ขนาด ที่ให้ผลต่างไปจากอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ขนาดกลุ่มตัวอย่างขนาดอื่นๆ ดังนั้น จึงนำอัตราความคลาดเคลื่อนประเภทที่ 1 ไปเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Games-Howell ผลการเปรียบเทียบแสดงดังตารางที่ 48

ตารางที่ 47 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ด้วยวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้วิธีการทดสอบของ Bonferroni

Size (M)	250 คน (5.81)		500 คน (4.95)		1000 คน (5.15)	
	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value
250 คน (5.81)	-	-	0.86 -	.059	0.66	.200
500 คน (4.95)			-	-	0.20	1.000
1000 คน (5.15)					-	-

*p < .05

จากตารางที่ 47 พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีซิปเทสท์ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน มีค่าสูงสุด รองลงมาได้แก่ภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน และ 500 คน ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ด้วยวิธีซิปเทสท์ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน 500 คน และ 1000 คน มีค่าไม่แตกต่างกัน

จากผลดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีซิปเทสท์ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน ไม่ส่งผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าเพิ่มขึ้นหรือลดลง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 48 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ด้วยวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้วิธีการทดสอบของ Games-Howell

Size (M)	250 คน (4.03)		500 คน (2.54)		1000 คน (0.41)	
	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value
250 คน (4.03)	-	-	1.49	.104	3.62**	.000
500 คน (2.54)			-	-	2.13**	.000
1000 คน (0.41)					-	-

**p < .01, *p < .05

จากตารางที่ 48 พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีราสช์ทรีภายใต้ขนาดกลุ่มตัวอย่าง 250 คน มีค่าสูงสุด รองลงมาได้แก่ภายใต้ขนาดกลุ่มตัวอย่าง 500 คน และ 1000 คน ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ด้วยวิธีราสช์ทรีภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน และ 500 คน มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

จากผลดังกล่าว แสดงว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีราสช์ทรีภายใต้ขนาดกลุ่มตัวอย่าง 250 คน และ 500 คน มีค่าแตกต่างกับขนาดกลุ่มตัวอย่าง 1000 คน ถ้าเพิ่มขนาดกลุ่มตัวอย่างจาก 250 คน เป็น 1000 คน จะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าลดลง ประมาณร้อยละ 3.62

ตารางที่ 49 ผลการวิเคราะห์อิทธิพลอย่างง่ายของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างขนาดเดียวกัน 3 ขนาด

แหล่งของความแปรปรวน	df	SS	MS	F	p-value
250 คน					
วิธีการตรวจสอบ	2	28.605	14.302	7.750**	.002
ความคลาดเคลื่อน	33	60.898	1.845		
รวม	35	89.502			
500 คน					
วิธีการตรวจสอบ	2	85.186	42.593	38.159**	.000
ความคลาดเคลื่อน	33	36.835	1.116		
รวม	35	122.021			
1000 คน					
วิธีการตรวจสอบ	2	240.969	120.484	422.618**	.000
ความคลาดเคลื่อน	33	9.408	.285		
รวม	35	250.377			

test of homogeneity of variances 250 คน ($F = 4.756$, $p = .015$), 500 คน ($F = 1.285$, $p = .290$), 1000 คน ($F = .572$, $p = .570$)

**$p < .01$

จากตารางที่ 49 พบว่า ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน 500 คน และ 1000 คน อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรีมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ดังนั้น จึงนำผลการทดสอบภายใต้ขนาดกลุ่มตัวอย่างทั้ง 3 ขนาด ไปเปรียบเทียบความแตกต่างเป็นรายคู่ โดยใช้วิธีการทดสอบของ Bonferroni และวิธีการทดสอบของ Games-Howell ผลการเปรียบเทียบแสดงดังตารางที่ 50

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 50 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ภายใต้เงื่อนไขของขนาดกลุ่มตัวอย่างขนาดเดียวกัน 3 ขนาด โดยใช้วิธีการทดสอบของ Bonferroni และวิธีการทดสอบของ Games-Howell

วิธีการตรวจสอบ DIF (M)		LR (6.02)		SIBTEST (5.81)		Raschtree (4.03)	
		Mean difference	p-value	Mean difference	p-value	Mean difference	p-value
250 คน	LR (6.02)	-	-	0.21	.821	1.99*	.017
	SIBTEST (5.81)			-	-	1.78*	.035
	Raschtree (4.03)					-	--
วิธีการตรวจสอบ DIF (M)		LR (6.25)		SIBTEST (4.95)		Raschtree (2.54)	
500 คน	LR (6.25)	-	-	1.30*	.015	3.71**	.000
	SIBTEST (4.95)			-	-	2.41**	.000
	Raschtree (2.54)					-	-
วิธีการตรวจสอบ DIF (M)		LR (6.42)		SIBTEST (5.15)		Raschtree (0.41)	
1000 คน	LR (6.42)	-	-	1.27**	.000	6.01**	.000
	SIBTEST (5.15)			-	-	4.74**	.000
	Raschtree (0.41)					-	

**p < .01, *p<.05

จากตารางที่ 50 การทดสอบความแตกต่างของอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ภายใต้ปัจจัยขนาดกลุ่มตัวอย่างขนาดเดียวกัน 3 ขนาด ปรากฏผลดังนี้

ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีการถดถอยโลจิสติกมีค่าสูงสุด รองลงมาได้แก่ วิธีซิปเทสท์และวิธีราสช์ทรีตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีราสช์ทรีอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกกับวิธีซิปเทสท์มีค่าไม่แตกต่างกัน

ภายใต้ขนาดกลุ่มตัวอย่าง 500 คน อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีการถดถอยโลจิสติกมีค่าสูงสุด รองลงมาได้แก่ วิธีซิปเทสท์และวิธีราสช์ทรีตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อน

ประเภทที่ 1 ระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีราสช์ทรีอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

ภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญด้วยวิธีการถดถอยโลจิสติกมีค่าสูงสุด รองลงมาได้แก่ วิธีซิปเทสท์และวิธีราสช์ทรีตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบดังกล่าว พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีราสช์ทรีอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

จากผลดังกล่าว แสดงว่า ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน 500 คน และ 1000 คน อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าวิธีซิปเทสท์และวิธีราสช์ทรีภายใต้เกือบทุกเงื่อนไขของขนาดกลุ่มตัวอย่าง ในขณะที่อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีราสช์ทรีมีค่าต่ำกว่าวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ภายใต้ทุกเงื่อนไขของขนาดกลุ่มตัวอย่าง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 51 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด

วิธีการ	LR		LR		LR		SIBTEST		SIBTEST		SIBTEST		Raschtree		Raschtree		Raschtree	
ตรวจสอบ DIF	250 คน	6.02	500 คน	6.25	1000	6.42	250 คน	5.81	500 คน	4.95	1000 คน	5.15	250 คน	4.03	500 คน	2.54	1000	0.41
	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value
LR (250 = 6.02)	-	-	0.23	1.000	0.40	.561	0.21	.821	1.07**	.008	0.87**	.003	1.99*	.017	3.48**	.000	5.61**	.000
LR (500 = 6.25)			-	-	0.17	1.000	0.44	.230	1.30*	.015	1.10**	.001	2.22**	.003	3.71**	.000	5.84**	.000
LR (1000 = 6.42)					-	-	0.61	.070	1.47**	.000	1.27**	.000	2.39**	.002	3.88**	.000	6.01**	.000
SIBTEST (250 = 5.81)							-	-	0.86	.059	0.66	.200	1.78*	.035	3.27**	.000	5.40**	.000
SIBTEST (500 = 4.95)									-	-	0.20	1.000	0.92	.177	2.41**	.000	4.54**	.000
SIBTEST (1000=5.15)											-	-	1.12	.087	2.61**	.000	4.74**	.000
Raschtree (250=4.03)													-	-	1.49	.104	3.62**	.000
Raschtree (500=2.54)															-	-	2.13**	.000
Raschtree (1000=0.41)																	-	-

**p < .01, *p < .05

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตารางที่ 51 การทดสอบความแตกต่างของอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ภายใต้ปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ซึ่งมีทั้งหมด 9 ค่า เมื่อพิจารณาค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ที่มีค่าสูงสุด 3 อันดับแรก พบว่า ค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก (1000 คน) มีค่าสูงสุด รองลงมาได้แก่ วิธีการถดถอยโลจิสติก (500 คน) และวิธีการถดถอยโลจิสติก (250 คน) ตามลำดับ เมื่อพิจารณาค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ที่มีค่าต่ำสุด 3 อันดับแรก พบว่า ค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีราสช์ทรี (1000 คน) มีค่าต่ำสุด รองลงมาได้แก่ วิธีราสช์ทรี (500 คน) และวิธีราสช์ทรี (250 คน) ตามลำดับ ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบภายใต้ขนาดกลุ่มตัวอย่างดังกล่าวกับอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบภายใต้ขนาดกลุ่มตัวอย่างอื่นๆ ปรากฏผลดังนี้

อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก (1000 คน) มีค่าไม่แตกต่างกับอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก (250 คน) การถดถอยโลจิสติก (500 คน) และวิธีซิปเทสท์ (250 คน) แต่มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบภายใต้ขนาดกลุ่มตัวอย่างขนาดอื่น ๆ ที่เหลืออย่างมีนัยสำคัญทางสถิติที่ระดับ .01

อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก (500 คน) มีค่าไม่แตกต่างกับอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก (250 คน) การถดถอยโลจิสติก (1000 คน) และวิธีซิปเทสท์ (250 คน) แต่มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบภายใต้ขนาดกลุ่มตัวอย่างขนาดอื่น ๆ ที่เหลืออย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และที่ระดับ .05

อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก (250 คน) มีค่าไม่แตกต่างกับอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก (500 คน) การถดถอยโลจิสติก (1000 คน) และวิธีซิปเทสท์ (250 คน) แต่มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบภายใต้ขนาดกลุ่มตัวอย่างขนาดอื่น ๆ ที่เหลืออย่างมีนัยสำคัญทางสถิติที่ระดับ .01

อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีราสช์ทรี (1000 คน) มีค่าต่ำกว่าวิธีการตรวจสอบภายใต้ขนาดกลุ่มตัวอย่างขนาดอื่น ๆ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีราสช์ทรี (500 คน) มีค่าไม่แตกต่างกับอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีราสช์ทรี (250 คน) แต่มีค่าต่ำกว่าวิธีการตรวจสอบภายใต้ขนาดกลุ่มตัวอย่างขนาดอื่น ๆ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และที่ระดับ .05

อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีราสช์ทรี (250 คน) มีค่าต่ำกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก (500 คน) และวิธีการถดถอยโลจิสติก (1000 คน) อย่างมีนัยสำคัญที่สถิติที่ระดับ .01 และมีค่าต่ำกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก (250 คน) และวิธีซิปเทสท์ (250 คน) อย่างมีนัยสำคัญที่สถิติที่ระดับ .05 แต่มีค่าไม่แตกต่างกับอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบภายใต้ขนาดกลุ่มตัวอย่างอื่น ๆ ที่เหลือ

จากผลการเปรียบเทียบดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญ เมื่อใช้วิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่าง 250 คน 500 คน และ 1000 คน จะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าสูงกว่าวิธีซิปเทสท์และวิธีราสช์ทรีเกือบทุกเงื่อนไขของการตรวจสอบ ส่วนการตรวจสอบด้วยวิธีราสช์ทรีภายใต้ขนาดกลุ่มตัวอย่าง 250 คน 500 คน และ 1000 คน จะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าต่ำกว่าวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์เกือบทุกเงื่อนไขของการตรวจสอบ

2.1.3 สรุปผลการเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี

สรุปผลการเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้การจัดกระทำปัจจัย 4 ปัจจัย คือ ความยากของข้อสอบ 3 ระดับ ความยาวของแบบสอบ 2 ขนาด สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด และขนาดกลุ่มตัวอย่าง 3 ขนาด สรุปได้ดังตารางที่ 52-54

ตารางที่ 52 สรุปผลการวิเคราะห์ค่าเฉลี่ยของอำนาจการทดสอบ (Power) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ด้วยวิธีการถดถอยโลจิสติก (Logistic Regression) วิธีซิปเทสท์ (SIBTEST) และวิธีราสช์ทรี (Raschtree) ภายใต้เงื่อนไขที่แตกต่างกันของปัจจัยที่ศึกษา

ปัจจัย	วิธีการตรวจสอบ DIF					
	LR		SIBTEST		Raschtree	
	Power	E_1	Power	E_1	Power	E_1
Difficulty of Item						
b_L	13.06	6.52	41.22	5.48	21.69	2.63
b_M	18.03	5.82	54.32	5.21	24.83	1.40
b_H	16.72	6.34	26.47	5.22	24.00	2.95
Test length						
40 ข้อ	15.71	6.20	41.93	5.22	23.13	2.21
60 ข้อ	16.17	6.26	39.41	5.39	23.89	2.44
%DIF						
10%	15.78	6.24	41.39	5.22	20.03	1.73
20%	16.10	6.22	39.95	5.39	26.99	2.92
Sample size						
250 คน	10.72	6.02	21.12	5.81	13.67	4.03
500 คน	13.96	6.25	39.13	4.95	27.25	2.54
1000 คน	23.13	6.42	61.76	5.15	29.60	0.41
ค่าเฉลี่ยรวม	15.94	6.23	40.67	5.30	23.51	2.33

จากตารางที่ 52 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขที่แตกต่างกันของปัจจัยที่ศึกษา พบว่า วิธีซิปเทสท์มีค่าเฉลี่ยรวมของอำนาจการทดสอบสูงสุด และเมื่อพิจารณาในแต่ละปัจจัยแต่ละระดับ พบว่า มีค่าเฉลี่ยสูงกว่าอำนาจการทดสอบของวิธีการถดถอยโลจิสติกและวิธีราสช์ทรีทุกเงื่อนไขของปัจจัยที่ศึกษา สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีราสช์ทรีมีค่าเฉลี่ยรวมต่ำสุด และเมื่อพิจารณาในแต่ละปัจจัยแต่ละระดับ พบว่า มีค่าเฉลี่ยต่ำกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ทุกเงื่อนไขของปัจจัยที่ศึกษา

ตารางที่ 53 สรุปผลการเปรียบเทียบอำนาจการทดสอบ (Power) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ด้วยวิธีการถดถอยโลจิสติก (Logistic Regression) วิธีซิปเทสท์ (SIBTEST) และวิธีราสช์ทรี (Raschtree) ภายใต้เงื่อนไขของปัจจัยที่ศึกษาต่างระดับกัน

ปัจจัย		วิธีการตรวจสอบ DIF					
		LR		SIBTEST		Raschtree	
		Power	E_1	Power	E_1	Power	E_1
★b_L	● b_M	0	+	0	0	0	0
	● b_H	0	0	0	0	0	0
★b_M	● b_L	0	-	0	0	0	0
	● b_H	0	0	+	0	0	0
★b_H	● b_L	0	0	0	0	0	0
	● b_M	0	0	-	0	0	0
★40 ข้อ	● 60 ข้อ	0	0	0	0	0	0
★10%	● 20%	0	0	0	0	0	0
★250 คน	● 500 คน	-	0	-	0	-	0
	● 1000 คน	-	0	-	0	-	+
★500 คน	● 250 คน	+	0	+	0	+	0
	● 1000 คน	-	0	-	0	0	+
★1000 คน	● 250 คน	+	0	+	0	+	-
	● 500 คน	+	0	+	0	0	-

0 หมายถึง Power หรือ E_1 ของวิธีการตรวจสอบ DIF ภายใต้ปัจจัย ★ มีค่าเท่ากับ Power หรือ E_1 ของวิธีการตรวจสอบ DIF ภายใต้ปัจจัย ●

+ หมายถึง Power หรือ E_1 ของวิธีการตรวจสอบ DIF ภายใต้ปัจจัย ★ มีค่าสูงกว่า Power หรือ E_1 ของวิธีการตรวจสอบ DIF ภายใต้ปัจจัย ●

- หมายถึง Power หรือ E_1 ของวิธีการตรวจสอบ DIF ภายใต้ปัจจัย ★ มีค่าต่ำกว่า Power หรือ E_1 ของวิธีการตรวจสอบ DIF ภายใต้ปัจจัย ●

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

จากตารางที่ 53 สรุปผลการเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยที่ศึกษาต่างระดับกัน พบประเด็นที่สำคัญ ดังนี้

1. ปัจจัยค่าความยากของข้อสอบ (b_L b_M และ b_H) พบว่า มีผลต่ออำนาจการทดสอบของวิธีซิปเทสท์ กล่าวคือ เมื่อความยากของข้อสอบ b_M จะส่งผลให้อำนาจการทดสอบสูงกว่าภายใต้เงื่อนไขความยากของข้อสอบ b_H สำหรับวิธีการถดถอยโลจิสติก พบว่า ปัจจัยค่าความยากของข้อสอบมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 กล่าวคือ เมื่อความยากของข้อสอบ b_L จะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าภายใต้เงื่อนไขความยากของข้อสอบ b_M ส่วนวิธีราสช์ทรี พบว่า ปัจจัยค่าความยากของข้อสอบไม่มีผลต่ออำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีนี้

2. ปัจจัยความยาวของแบบสอบ (40 ข้อ และ 60 ข้อ) พบว่า ไม่มีผลต่ออำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบทั้ง 3 วิธี

3. ปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน (10% และ 20%) พบว่า ไม่มีผลต่ออำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบทั้ง 3 วิธี

4. ปัจจัยขนาดกลุ่มตัวอย่าง (250 คน 500 คน และ 1000 คน) พบว่า มีผลต่ออำนาจการทดสอบของวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี กล่าวคือ สำหรับวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ ได้ผลสอดคล้องกัน เมื่อขนาดกลุ่มตัวอย่าง 1000 คน จะส่งผลให้อำนาจการทดสอบสูงกว่าภายใต้ขนาดกลุ่มตัวอย่าง 250 คน และ 500 คน อีกทั้ง เมื่อขนาดกลุ่มตัวอย่าง 500 คน จะส่งผลให้อำนาจการทดสอบสูงกว่าภายใต้ขนาดกลุ่มตัวอย่าง 250 คน ส่วนวิธีราสช์ทรี พบว่า เมื่อใช้กลุ่มตัวอย่างขนาด 500 คน และ 1000 คน จะส่งผลให้อำนาจการทดสอบสูงกว่าภายใต้ขนาดกลุ่มตัวอย่าง 250 คน นอกจากนี้ ปัจจัยขนาดกลุ่มตัวอย่างมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีราสช์ทรี กล่าวคือ เมื่อขนาดกลุ่มตัวอย่าง 250 คน และ 500 คน จะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน

ตารางที่ 54 สรุปผลการเปรียบเทียบอำนาจการทดสอบ (Power) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก (Logistic Regression) วิธีซิปเทสท์ (SIBTEST) และวิธีราสช์ทรี (Raschtree) ภายใต้เงื่อนไขของปัจจัยที่ศึกษาระดับเดียวกัน

ปัจจัย	อำนาจการทดสอบ (Power)						อัตราความคลาดเคลื่อนประเภทที่ 1 (E_1)					
	◆LR		◆SIBTEST		◆Raschtree		◆LR		◆SIBTEST		◆Raschtree	
	■ SIBTEST	■ Raschtree	■ LR	■ Raschtree	■ LR	■ SIBTEST	■ SIBTEST	■ Raschtree	■ LR	■ Raschtree	■ LR	■ SIBTEST
b_L	-	-	+	+	+	-	+	+	-	+	-	-
b_M	-	0	+	+	0	-	0	+	0	+	-	-
b_H	-	0	+	0	0	0	0	+	0	+	-	-
40 ข้อ	-	-	+	+	+	-	+	+	-	+	-	-
60 ข้อ	-	-	+	+	+	-	+	+	-	+	-	-
10%	-	0	+	+	0	-	+	+	-	+	-	-
20%	-	-	+	0	+	0	+	+	-	+	-	-
250 คน	-	0	+	+	0	-	0	+	0	+	-	-
500 คน	-	-	+	+	+	-	+	+	-	+	-	-
1000 คน	-	-	+	+	+	-	+	+	-	+	-	-

0 หมายถึง Power หรือ E_1 ของวิธีการตรวจสอบ DIF ภายใต้ปัจจัย ◆ มีค่าเท่ากับ Power หรือ E_1 ของวิธีการตรวจสอบ DIF ภายใต้ปัจจัย ■

\+ หมายถึง Power หรือ E_1 ของวิธีการตรวจสอบ DIF ภายใต้ปัจจัย ◆ มีค่าสูงกว่า Power หรือ E_1 ของวิธีการตรวจสอบ DIF ภายใต้ปัจจัย ■

\- หมายถึง Power หรือ E_1 ของวิธีการตรวจสอบ DIF ภายใต้ปัจจัย ◆ มีค่าต่ำกว่า Power หรือ E_1 ของวิธีการตรวจสอบ DIF ภายใต้ปัจจัย ■

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตารางที่ 54 สรุปผลการเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยที่ศึกษาระดับเดียวกัน พบประเด็นที่สำคัญ ดังนี้

1. ปัจจัยค่าความยากของข้อสอบ (b_L b_M และ b_H) พบว่า วิธีซิปเทสท์มีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติกและวิธีราสช์ทรีเกือบทุกระดับความยากของข้อสอบ ยกเว้นความยากของข้อสอบ b_H อำนาจการทดสอบของวิธีซิปเทสท์และวิธีราสช์ทรีมีค่าใกล้เคียงกัน สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีราสช์ทรีทุกระดับความยากของข้อสอบ

2. ปัจจัยความยาวของแบบสอบ (40 ข้อ และ 60 ข้อ) พบว่า วิธีซิปเทสท์มีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติกและวิธีราสช์ทรีทั้งความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีราสช์ทรีทั้งความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ

3. ปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน (10% และ 20%) พบว่า วิธีซิปเทสท์มีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติกและวิธีราสช์ทรี ยกเว้น สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 20% อำนาจการทดสอบของวิธีซิปเทสท์และวิธีราสช์ทรีมีค่าใกล้เคียงกัน สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีราสช์ทรีทั้งสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 10% และ 20%

4. ปัจจัยขนาดกลุ่มตัวอย่าง (250 คน 500 คน และ 1000 คน) พบว่า วิธีซิปเทสท์มีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติกและวิธีราสช์ทรีทุกขนาดกลุ่มตัวอย่าง 250 คน 500 คน และ 1000 คน สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีราสช์ทรีทุกขนาดกลุ่มตัวอย่าง 250 คน 500 คน และ 1000 คน

2.2 การเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ระหว่างวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์

2.2.1 การเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ระหว่างวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์

ผู้วิจัยนำอำนาจการทดสอบของวิธีการตรวจสอบ 2 วิธี ภายใต้ปัจจัยที่ศึกษา 4 ปัจจัย จากตารางที่ 17 ไปเปรียบเทียบโดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง (two-way ANOVA) เพื่อทดสอบผลของวิธีการตรวจสอบ ผลของปัจจัยที่ศึกษา และผลของปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบและปัจจัยที่ศึกษา ถ้าผลการทดสอบดังกล่าวมีนัยสำคัญทางสถิติ จะใช้การวิเคราะห์อิทธิพลอย่างง่าย (simple effect) ทดสอบในแต่ละระดับของตัวแปร แล้วเปรียบเทียบเป็นรายคู่โดยใช้วิธีการทดสอบของ Bonferroni กรณีค่าความแปรปรวนของทุกกลุ่มเท่ากัน และวิธีการทดสอบของ Games-Howell กรณีค่าความแปรปรวนของทุกกลุ่มไม่เท่ากัน โดยจะทดสอบใน 3 กรณีต่อไปนี้ 1) ทดสอบอำนาจการทดสอบของแต่ละวิธีการตรวจสอบภายใต้ปัจจัยที่ศึกษาต่างระดับกัน 2) ทดสอบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 2 วิธี ภายใต้ปัจจัยที่ศึกษาระดับเดียวกัน และ 3) ทดสอบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 2 วิธี ภายใต้ปัจจัยที่ศึกษาต่างระดับกัน ผลการวิเคราะห์แสดงดังตารางต่อไปนี้

ตารางที่ 55 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ระหว่างวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบต่างกัน 3 ระดับ

แหล่งของความแปรปรวน	df	SS	MS	F	p-value
ผลหลัก					
วิธีการตรวจสอบ DIF	1	15699.625	15699.625	73.912**	.000
ความยากของข้อสอบ (DI)	2	3232.483	1616.242	7.609**	.001
ปฏิสัมพันธ์ 2 ทาง					
วิธีการตรวจสอบ DIF x DI	2	1704.915	852.458	4.013*	.023
ความคลาดเคลื่อน	66	14019.131	212.411		
รวม	71	34656.154			

test of homogeneity of variances (F = 7.802, p = .000)

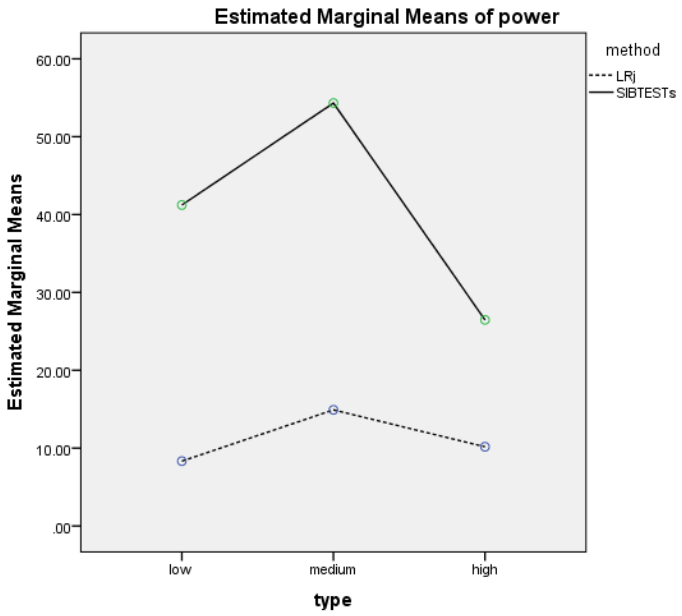
**p < .01, *p < .05

จากตารางที่ 55 เมื่อพิจารณาผลการทดสอบปฏิสัมพันธ์ 2 ทาง (two way interaction) พบว่า ปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบ DIF และความยากของข้อสอบ (DI) มีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพลด้วยวิธีการตรวจสอบต่างกันภายใต้ความยากของข้อสอบต่างกัน จะส่งผลให้อำนาจการทดสอบมีค่าแตกต่างกัน ดังนั้น จึงนำอำนาจการทดสอบดังกล่าว ไปเปรียบเทียบความแตกต่างโดยเปรียบเทียบใน 3 กรณี ดังนี้

กรณีที่ 1 เปรียบเทียบอำนาจการทดสอบของแต่ละวิธีการตรวจสอบ 2 วิธี ภายใต้ความยากของข้อสอบต่างกัน 3 ระดับ โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย (simple effect) ผลการวิเคราะห์แสดงดังตารางที่ 56

กรณีที่ 2 เปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 2 วิธี ภายใต้ความยากของข้อสอบเดียวกัน 3 ระดับ โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย (simple effect) ผลการวิเคราะห์แสดงดังตารางที่ 58

กรณีที่ 3 เปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 2 วิธี ภายใต้ความยากของข้อสอบต่างกัน 3 ระดับ ผลการเปรียบเทียบแสดงดังตารางที่ 59



ภาพที่ 15 อำนาจการทดสอบของวิธีการตรวจสอบ 2 วิธี ภายใต้ค่าความยากของข้อสอบต่างกัน 3 ระดับ

ตารางที่ 56 ผลการวิเคราะห์อิทธิพลอย่างง่ายของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ด้วยวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบต่างกัน 3 ระดับ

แหล่งของความแปรปรวน	df	SS	MS	F	p-value
วิธี LRj					
ความยากของข้อสอบ	2	276.765	138.383	1.875	.169
ความคลาดเคลื่อน	33	2434.894	73.785		
รวม	35	2711.660			
วิธี SIBTESTs					
ความยากของข้อสอบ	2	4660.633	2330.317	6.638**	.004
ความคลาดเคลื่อน	33	11584.236	351.037		
รวม	35	16244.870			

test of homogeneity of variances (F = 4.060, p = .027)

**p < .01

จากตารางที่ 56 พบว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพล ด้วยวิธีการถดถอยโลจิสติกภายใต้ความยากของข้อสอบต่างกัน 3 ระดับ มีค่าไม่แตกต่างกัน ส่วนวิธีซิปเทสท์มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่าอำนาจการทดสอบโดยการวัดขนาดอิทธิพลของวิธีซิปเทสท์ภายใต้ความยากของข้อสอบอย่างน้อย 1 ลักษณะที่ให้ผลต่างไปจากอำนาจการทดสอบภายใต้ความยากของข้อสอบอื่น ๆ ดังนั้น จึงนำอำนาจการทดสอบไปเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Games-Howell ผลการเปรียบเทียบแสดงดังตารางที่ 57

ตารางที่ 57 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพล ด้วยวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบต่างกัน 3 ระดับ โดยใช้วิธีการทดสอบของ Games-Howell

DI (M)	b_L (41.22)		b_M (54.32)		b_H (26.47)	
	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value
b_L (41.22)	-	-	13.10	.312	14.75	.093
b_M (54.32)			-	-	27.85**	.005
b_H (26.47)					-	-

**p < .01

จากตารางที่ 57 พบว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลด้วยวิธีซิปเทสท์ภายใต้ความยากของข้อสอบ b_M มีค่าสูงสุด รองลงมาได้แก่ ภายใต้ความยากของข้อสอบ b_L และ b_H ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบของวิธีซิปเทสท์ภายใต้ความยากของข้อสอบต่างกัน 3 ระดับ พบว่า อำนาจการทดสอบภายใต้ความยากของข้อสอบ b_L มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ความยากของข้อสอบ b_M และ b_H ส่วนอำนาจการทดสอบภายใต้ความยากของข้อสอบ b_M มีค่าสูงกว่าอำนาจการทดสอบภายใต้ความยากของข้อสอบ b_H อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

จากผลดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลด้วยวิธีซิปเทสท์ เมื่อความยากของข้อสอบมีค่าความยากปานกลางจะส่งผลให้อำนาจการทดสอบมีค่าสูงสุด ประมาณร้อยละ 54.32 แตกต่างจากภายใต้เงื่อนไขความยากของข้อสอบมีค่าความยากสูง ประมาณร้อยละ 27.85

ตารางที่ 58 ผลการวิเคราะห์อิทธิพลอย่างง่ายของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ด้วยวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบเดียวกัน 3 ระดับ

แหล่งของความแปรปรวน	df	SS	MS	F	p-value
b_L					
วิธีการตรวจสอบ	1	6492.815	6492.815	30.030**	.000
ความคลาดเคลื่อน	22	4756.662	216.212		
รวม	23	11249.477			
b_M					
วิธีการตรวจสอบ	1	9318.889	9318.889	29.228**	.000
ความคลาดเคลื่อน	22	7014.329	318.833		
รวม	23	16333.217			
b_H					
วิธีการตรวจสอบ	1	1592.836	1592.836	15.587**	.001
ความคลาดเคลื่อน	22	2248.140	102.188		
รวม	23	3840.976			

**$p < .01$

จากตารางที่ 58 พบว่า ภายใต้ความยากของข้อสอบ b_L b_M และ b_H อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลของวิธีซิปเทสท์มีค่าสูงกว่าอำนาจการทดสอบของวิธีการถดถอยโลจิสติกอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 จากผลดังกล่าวแสดงว่า ภายใต้ความยากของข้อสอบ 3 ระดับ อำนาจการทดสอบของวิธีซิปเทสท์มีค่าสูงภายใต้ทุกเงื่อนไขของความยากของข้อสอบ

ตารางที่ 59 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ด้วยวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบต่างกัน 3 ระดับ

วิธีการตรวจสอบ DIF	LRj		LRj		LRj		SIBTESTs		SIBTESTs		SIBTESTs	
DI (M)	b_L (8.33)		b_M (14.91)		b_H (10.17)		b_L (41.22)		b_M (54.32)		b_H (26.47)	
	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value
LRj (b_L=8.33)	-	-	6.58	.208	1.84	1.000	32.89**	.000	45.99**	.000	18.14**	.000
LRj (b_M=14.91)			-	-	4.74	.558	26.31**	.001	39.41**	.000	11.56*	.018
LRj (b_H=10.17)					-	-	31.05**	.000	44.15**	.000	16.30**	.001
SIBTESTs (b_L=41.22)							-	-	13.10	.312	14.75	.093
SIBTESTs (b_M=54.32)									-	-	27.85**	.005
SIBTESTs (b_H=26.47)											-	-

**p < .01, *p < .05

จากตารางที่ 59 การทดสอบความแตกต่างของอำนาจการทดสอบระหว่างวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ ภายใต้ปัจจัยความยากของข้อสอบต่างกัน 3 ระดับ โดยเปรียบเทียบค่าเฉลี่ยของอำนาจการทดสอบซึ่งมีทั้งหมด 6 ค่า เมื่อพิจารณาค่าเฉลี่ยของอำนาจการทดสอบที่มีค่าสูงสุด พบว่า ค่าเฉลี่ยของอำนาจการทดสอบของวิธีซิปเทสท์ (b_M) มีค่าสูงสุด รองลงมาได้แก่ วิธีซิปเทสท์ (b_L) เมื่อพิจารณาค่าเฉลี่ยของอำนาจการทดสอบที่มีค่าต่ำสุด พบว่า ค่าเฉลี่ยของอำนาจการทดสอบของวิธีการถดถอยโลจิสติก (b_L) มีค่าต่ำสุด รองลงมาได้แก่ วิธีการถดถอยโลจิสติก (b_H) ผลการเปรียบเทียบอำนาจการทดสอบของวิธีการตรวจสอบภายใต้ความยากของข้อสอบดังกล่าวกับอำนาจการทดสอบของวิธีการตรวจสอบภายใต้ความยากของข้อสอบอื่นๆ ปรากฏผลดังนี้

อำนาจการทดสอบของวิธีซิปเทสท์ (b_M) มีค่าไม่แตกต่างกับอำนาจการทดสอบของวิธีซิปเทสท์ (b_L) แต่มีค่าสูงกว่าอำนาจการทดสอบของวิธีการตรวจสอบภายใต้ความยากของข้อสอบอื่น ๆ ที่เหลืออย่างมีนัยสำคัญทางสถิติที่ระดับ .01

อำนาจการทดสอบของวิธีซิปเทสท์ (b_L) มีค่าไม่แตกต่างกับอำนาจการทดสอบของวิธีซิปเทสท์ (b_M) และวิธีซิปเทสท์ (b_H) แต่มีค่าสูงกว่าอำนาจการทดสอบของวิธีการตรวจสอบภายใต้ความยากของข้อสอบอื่น ๆ ที่เหลืออย่างมีนัยสำคัญทางสถิติที่ระดับ .01

อำนาจการทดสอบของวิธีการถดถอยโลจิสติก (b_L) มีค่าต่ำกว่าอำนาจการทดสอบของวิธีซิปเทสท์ (b_L) วิธีซิปเทสท์ (b_M) และวิธีซิปเทสท์ (b_H) อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แต่มีค่าไม่แตกต่างกับวิธีการตรวจสอบภายใต้ความยากของข้อสอบอื่น ๆ ที่เหลือ

อำนาจการทดสอบของวิธีการถดถอยโลจิสติก (b_H) มีค่าต่ำกว่าอำนาจการทดสอบของวิธีซิปเทสท์ (b_L) วิธีซิปเทสท์ (b_M) และวิธีซิปเทสท์ (b_H) อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แต่มีค่าไม่แตกต่างกับวิธีการตรวจสอบภายใต้ความยากของข้อสอบอื่น ๆ ที่เหลือ

จากผลการเปรียบเทียบดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพล เมื่อใช้วิธีซิปเทสท์ภายใต้ความยากของข้อสอบที่มีค่าความยากปานกลาง จะส่งผลให้อำนาจการทดสอบมีค่าสูงสุด โดยมีค่าสูงกว่าวิธีการถดถอยโลจิสติกในทุกเงื่อนไขการตรวจสอบ ส่วนการตรวจสอบด้วยวิธีการถดถอยโลจิสติกภายใต้ความยากของข้อสอบที่มีค่าความยากต่ำ จะส่งผลให้อำนาจการทดสอบมีค่าต่ำสุด โดยมีค่าต่ำกว่าวิธีซิปเทสท์ในทุกเงื่อนไขการตรวจสอบ

ตารางที่ 60 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ระหว่างวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยความยาวของแบบสอบต่างกัน 2 ขนาด

แหล่งของความแปรปรวน	df	SS	MS	F	p-value
ผลหลัก					
วิธีการตรวจสอบ DIF	1	15699.625	15699.625	56.587**	.000
ความยาวของแบบสอบ (TL)	1	88.867	88.867	.320	.573
ปฏิสัมพันธ์ 2 ทาง					
วิธีการตรวจสอบ DIF x TL	1	1.623	1.623	.006	.939
ความคลาดเคลื่อน	68	18866.040	277.442		
รวม	71	34656.154			

test of homogeneity of variances (F = 7.898, p = .000)

**p < .01

จากตารางที่ 60 เมื่อพิจารณาผลการทดสอบปฏิสัมพันธ์ 2 ทาง (two way interaction) พบว่า ปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบ DIF และความยาวของแบบสอบ (TL) ไม่มีนัยสำคัญทางสถิติ แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลด้วยวิธีการตรวจสอบต่างกันภายใต้ความยาวของแบบสอบต่างกันแล้ว อำนาจการทดสอบมีค่าไม่แตกต่างกัน นั่นคือ วิธีการตรวจสอบและความยาวของแบบสอบไม่มีผลร่วมกันต่ออำนาจการทดสอบ

เมื่อพิจารณาผลการทดสอบของปัจจัยความยาวของแบบสอบ (TL) พบว่า ความยาวของแบบสอบไม่มีนัยสำคัญทางสถิติ แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพล ภายใต้ความยาวของแบบสอบต่างกันแล้ว อำนาจการทดสอบมีค่าไม่แตกต่างกัน

เมื่อพิจารณาผลการทดสอบของวิธีการตรวจสอบ DIF พบว่า วิธีการตรวจสอบมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลด้วยวิธีการตรวจสอบต่างกัน จะส่งผลให้อำนาจการทดสอบมีค่าแตกต่างกัน จากผลดังกล่าวแสดงว่า ภายใต้ความยาวของแบบสอบ 2 ขนาด อำนาจการทดสอบของวิธีซิปเทสท์มีค่าสูงภายใต้ทุกเงื่อนไขของความยาวของแบบสอบ

ตารางที่ 61 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ระหว่างวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด

แหล่งของความแปรปรวน	df	SS	MS	F	p-value
ผลหลัก					
วิธีการตรวจสอบ DIF	1	15699.625	15699.625	56.375**	.000
สัดส่วนของ DIF (%DIF)	1	5.572	5.572	.020	.888
ปฏิสัมพันธ์ 2 ทาง					
วิธีการตรวจสอบ DIF x %DIF	1	14.089	14.089	.051	.823
ความคลาดเคลื่อน	68	18936.868	278.483		
รวม	71	34656.154			

test of homogeneity of variances (F = 7.263, p = .000)

**p < .01

จากตารางที่ 61 เมื่อพิจารณาผลการทดสอบปฏิสัมพันธ์ 2 ทาง (two way interaction) พบว่า ปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบ DIF และสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน (%DIF) ไม่มีนัยสำคัญทางสถิติ แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลด้วยวิธีการตรวจสอบต่างกันภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกันแล้ว อำนาจการทดสอบมีค่าไม่แตกต่างกัน นั่นคือ วิธีการตรวจสอบและสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน ไม่มีผลร่วมกันต่ออำนาจการทดสอบ

เมื่อพิจารณาผลการทดสอบของปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน (%DIF) พบว่า สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันไม่มีนัยสำคัญทางสถิติ แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพล ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกันแล้ว อำนาจการทดสอบมีค่าไม่แตกต่างกัน

Chulalongkorn University

เมื่อพิจารณาผลการทดสอบของวิธีการตรวจสอบ DIF พบว่า วิธีการตรวจสอบมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลด้วยวิธีการตรวจสอบต่างกัน จะส่งผลให้อำนาจการทดสอบมีค่าแตกต่างกัน จากผลดังกล่าวแสดงว่า ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด อำนาจการทดสอบของวิธีซิปเทสท์มีค่าสูงภายใต้ทุกเงื่อนไขของสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน

ตารางที่ 62 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพล ระหว่างวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด

แหล่งของความแปรปรวน	df	SS	MS	F	p-value
ผลหลัก					
วิธีการตรวจสอบ DIF	1	15699.625	15699.625	152.858**	.000
ขนาดกลุ่มตัวอย่าง (Size)	2	10773.327	5386.663	52.447**	.000
ปฏิสัมพันธ์ 2 ทาง					
วิธีการตรวจสอบ DIF x Size	2	1404.537	702.268	6.838**	.002
ความคลาดเคลื่อน	66	6778.666	102.707		
รวม	71	34656.154			

test of homogeneity of variances (F = 20.577, p = .000)

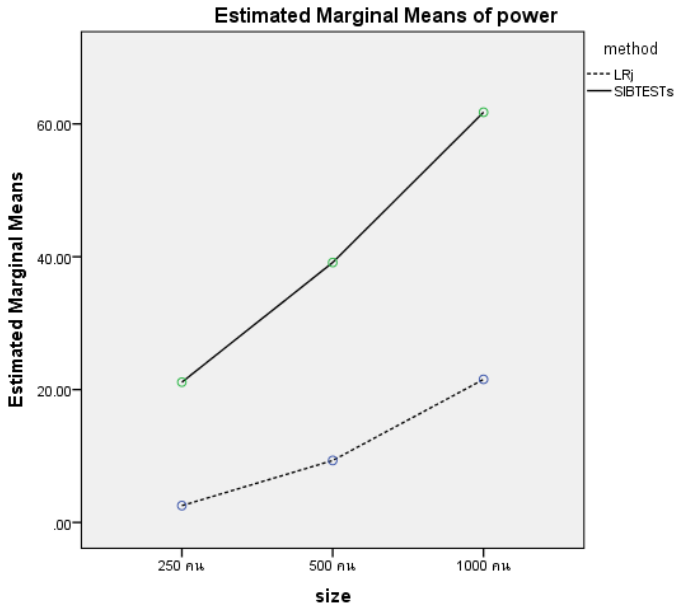
**p < .01

จากตารางที่ 62 เมื่อพิจารณาผลการทดสอบปฏิสัมพันธ์ 2 ทาง (two way interaction) พบว่า ปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบ DIF และขนาดกลุ่มตัวอย่าง (Size) มีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลด้วยวิธีการตรวจสอบต่างกันภายใต้ขนาดกลุ่มตัวอย่างต่างกัน จะส่งผลให้อำนาจการทดสอบมีค่าแตกต่างกัน ดังนั้น จึงนำอำนาจการทดสอบดังกล่าว ไปเปรียบเทียบความแตกต่างโดยเปรียบเทียบใน 3 กรณีดังนี้

กรณีที่ 1 เปรียบเทียบอำนาจการทดสอบของแต่ละวิธีการตรวจสอบ 2 วิธี ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย (simple effect) ผลการวิเคราะห์แสดงดังตารางที่ 63

กรณีที่ 2 เปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 2 วิธี ภายใต้ขนาดกลุ่มตัวอย่างเดียวกัน 3 ขนาด โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย (simple effect) ผลการวิเคราะห์แสดงดังตารางที่ 66

กรณีที่ 3 เปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 2 วิธี ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด ผลการเปรียบเทียบแสดงดังตารางที่ 67



ภาพที่ 16 อำนาจการทดสอบของวิธีการตรวจสอบ 2 วิธี ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด

Chulalongkorn University

ตารางที่ 63 ผลการวิเคราะห์อิทธิพลอย่างง่ายของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพล ด้วยวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด

แหล่งของความแปรปรวน	df	SS	MS	F	p-value
วิธี LRj					
ขนาดกลุ่มตัวอย่าง	2	2226.093	1113.046	75.645**	.000
ความคลาดเคลื่อน	33	485.567	14.714		
รวม	35	2711.660			
วิธี SIBTESTs					
ขนาดกลุ่มตัวอย่าง	2	9951.771	4975.885	26.093**	.000
ความคลาดเคลื่อน	33	6293.099	190.700		
รวม	35	16244.870			

test of homogeneity of variances LRj (F = 7.814, p = .002), SIBTESTs (F = 9.511, p = .001)

**p < .01

จากตารางที่ 63 การเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลของแต่ละวิธีการตรวจสอบ 2 วิธี ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด ปรากฏผลดังนี้

อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลด้วยวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อำนาจการทดสอบของวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่างอย่างน้อย 1 ขนาด ที่ให้ผลต่างไปจากอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่างขนาดอื่น ๆ ดังนั้นจึงนำอำนาจการทดสอบไปเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Games-Howell ผลการเปรียบเทียบแสดงดังตารางที่ 64

อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลด้วยวิธีซิปเทสท์ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อำนาจการทดสอบของวิธีซิปเทสท์ภายใต้ขนาดกลุ่มตัวอย่างอย่างน้อย 1 ขนาด ที่ให้ผลต่างไปจากอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่างขนาดอื่นๆ ดังนั้น จึงนำอำนาจการทดสอบไปเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Games-Howell ผลการเปรียบเทียบแสดงดังตารางที่ 65

ตารางที่ 64 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ด้วยวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้วิธีการทดสอบของ Games-Howell

Size (M)	250 คน (2.53)		500 คน (9.33)		1000 คน (21.54)	
	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value
250 คน (2.53)	-	-	6.80**	.000	19.01**	.000
500 คน (9.33)			-	-	12.21**	.000
1000 คน (21.54)					-	-

**p < .01

จากตารางที่ 64 พบว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลด้วยวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน มีค่าสูงสุด รองลงมาได้แก่ภายใต้ขนาดกลุ่มตัวอย่าง 500 คน และ 250 คน ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบด้วยวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด พบว่า อำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน มีค่าสูงกว่าอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่าง 250 คน และ 500 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่าง 500 คน มีค่าสูงกว่าอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่าง 250 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

จากผลดังกล่าว แสดงว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลด้วยวิธีการถดถอยโลจิสติก ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน 500 คน และ 1000 คน มีค่าแตกต่างกัน ถ้าเพิ่มขนาดกลุ่มตัวอย่างจาก 250 คน เป็น 500 คน และจาก 250 คน เป็น 1000 คน จะส่งผลให้อำนาจการทดสอบมีค่าเพิ่มขึ้น ประมาณร้อยละ 6.80 และร้อยละ 19.01 ตามลำดับ

ตารางที่ 65 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพล ด้วยวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้วิธีการทดสอบของ Games-Howell

Size (M)	250 คน (21.12)		500 คน (39.13)		1000 คน (61.76)	
	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value
250 คน (21.12)	-	-	18.01**	.002	40.64**	.000
500 คน (39.13)			-	-	22.63**	.007
1000 คน (61.76)					-	-

**p < .01

จากตารางที่ 65 พบว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลด้วยวิธีซิปเทสท์ภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน มีค่าสูงสุด รองลงมาได้แก่ภายใต้ขนาดกลุ่มตัวอย่าง 500 คน และ 250 คน ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบด้วยวิธีซิปเทสท์ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด พบว่า อำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน มีค่าสูงกว่าอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่าง 250 คน และ 500 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่าง 500 คน มีค่าสูงกว่าอำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่าง 250 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

จากผลดังกล่าว แสดงว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลด้วยวิธีซิปเทสท์ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน 500 คน และ 1000 คน มีค่าแตกต่างกัน ถ้าเพิ่มขนาดกลุ่มตัวอย่างจาก 250 คน เป็น 500 คน และจาก 250 คน เป็น 1000 คน จะส่งผลให้อำนาจการทดสอบมีค่าเพิ่มขึ้น ประมาณร้อยละ 18.01 และร้อยละ 40.64 ตามลำดับ

ตารางที่ 66 ผลการวิเคราะห์อิทธิพลอย่างง่ายของอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ระหว่างวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างขนาดเดียวกัน 3 ขนาด

แหล่งของความแปรปรวน	df	SS	MS	F	p-value
250 คน					
วิธีการตรวจสอบ	1	2072.042	2072.042	75.923**	.000
ความคลาดเคลื่อน	22	600.413	27.292		
รวม	23	2672.455			
500 คน					
วิธีการตรวจสอบ	1	5328.240	5328.240	57.268**	.000
ความคลาดเคลื่อน	22	2046.882	93.040		
รวม	23	7375.122			
1000 คน					
วิธีการตรวจสอบ	1	9703.880	9703.880	51.674**	.000
ความคลาดเคลื่อน	22	4131.370	187.790		
รวม	23	13835.249			

**p < .01

จากตารางที่ 66 พบว่า ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน 500 คน และ 1000 คน อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลของวิธีซิปเทสท์มีค่าสูงกว่าอำนาจการทดสอบของวิธีการถดถอยโลจิสติกอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 จากผลดังกล่าว แสดงว่า ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด อำนาจการทดสอบของวิธีซิปเทสท์มีค่าสูงภายใต้ทุกเงื่อนไขของขนาดกลุ่มตัวอย่าง

ตารางที่ 67 ผลการเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ด้วยวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด

วิธีตรวจ DIF	LRj		LRj		LRj		SIBTESTs		SIBTESTs		SIBTESTs	
Size (M)	250 คน (2.53)		500 คน (9.33)		1000 คน (21.54)		250 คน (21.12)		500 คน (39.13)		1000 คน (61.76)	
	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value
LRj (250=2.53)	-	-	6.80**	.000	19.01**	.000	18.59**	.000	36.60**	.000	59.23**	.000
LRj (500=9.33)			-	-	12.21**	.000	11.79**	.000	29.80**	.000	52.43**	.000
LRj (1000=21.54)					-	-	0.42	.870	17.59**	.001	40.22**	.000
SIBTESTs (250=21.12)							-	-	18.01**	.002	40.64**	.000
SIBTESTs (500=39.13)									-	-	22.63**	.007
SIBTESTs (1000=61.76)											-	-

**p < .01

จากตารางที่ 67 การทดสอบความแตกต่างของอำนาจการทดสอบระหว่างวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ ภายใต้ปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยเปรียบเทียบค่าเฉลี่ยของอำนาจการทดสอบซึ่งมีทั้งหมด 6 ค่า เมื่อพิจารณาค่าเฉลี่ยของอำนาจการทดสอบที่มีค่าสูงสุด พบว่า ค่าเฉลี่ยของอำนาจการทดสอบของวิธีซิปเทสท์ (1000 คน) มีค่าสูงสุด รองลงมาได้แก่ วิธีซิปเทสท์ (500 คน) เมื่อพิจารณาค่าเฉลี่ยของอำนาจการทดสอบที่มีค่าต่ำสุด พบว่า ค่าเฉลี่ยของอำนาจการทดสอบของวิธีการถดถอยโลจิสติก (250 คน) มีค่าต่ำสุด รองลงมาได้แก่ วิธีการถดถอยโลจิสติก (500 คน) ผลการเปรียบเทียบอำนาจการทดสอบของวิธีการตรวจสอบภายใต้ขนาดกลุ่มตัวอย่างดังกล่าวกับอำนาจการทดสอบของวิธีการตรวจสอบภายใต้ขนาดกลุ่มตัวอย่างอื่น ๆ ปรากฏผลดังนี้

อำนาจการทดสอบของวิธีซิปเทสท์ (1000 คน) มีค่าสูงกว่าอำนาจการทดสอบของวิธีการตรวจสอบภายใต้ความยากของข้อสอบอื่น ๆ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

อำนาจการทดสอบของวิธีซิปเทสท์ (500 คน) มีค่าสูงกว่าอำนาจการทดสอบของวิธีการตรวจสอบภายใต้ความยากของข้อสอบอื่น ๆ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

อำนาจการทดสอบของวิธีการถดถอยโลจิสติก (250 คน) มีค่าต่ำกว่าอำนาจการทดสอบของวิธีการตรวจสอบภายใต้ความยากของข้อสอบอื่น ๆ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

อำนาจการทดสอบของวิธีการถดถอยโลจิสติก (500 คน) มีค่าต่ำกว่าอำนาจการทดสอบของวิธีการตรวจสอบภายใต้ความยากของข้อสอบอื่น ๆ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

จากผลการเปรียบเทียบดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพล เมื่อใช้วิธีซิปเทสท์ภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน จะส่งผลให้อำนาจการทดสอบมีค่าสูงสุด โดยมีค่าสูงกว่าวิธีการถดถอยโลจิสติกในทุกเงื่อนไขการตรวจสอบ ส่วนการตรวจสอบด้วยวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่าง 250 คน จะส่งผลให้อำนาจการทดสอบมีค่าต่ำสุด โดยมีค่าต่ำกว่าวิธีซิปเทสท์ในทุกเงื่อนไขการตรวจสอบ

2.2.2 การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพล ระหว่างวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์

ผู้วิจัยนำอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบ 2 วิธี ภายใต้ปัจจัยที่ศึกษา 4 ปัจจัย จากตารางที่ 17 ไปเปรียบเทียบโดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง (two-way ANOVA) เพื่อทดสอบผลของวิธีการตรวจสอบ ผลของปัจจัยที่ศึกษา และผลของปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบและปัจจัยที่ศึกษา ถ้าผลการทดสอบดังกล่าวมีนัยสำคัญทางสถิติจะใช้การวิเคราะห์อิทธิพลอย่างง่าย (simple effect) ทดสอบในแต่ละระดับของตัวแปร แล้วเปรียบเทียบเป็นรายคู่โดยใช้วิธีการทดสอบของ Bonferroni กรณีค่าความแปรปรวนของทุกกลุ่มเท่ากัน และวิธีการทดสอบของ Games-Howell กรณีค่าความแปรปรวนของทุกกลุ่มไม่เท่ากัน ยกเว้นการเปรียบเทียบภายใต้ความยาวของแบบสอบ 2 ขนาด และสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด และการเปรียบเทียบระหว่างวิธีการตรวจสอบ 2 วิธี ภายใต้ปัจจัยที่ศึกษาระดับเดียวกัน จะใช้การทดสอบ t-test กรณีกลุ่มตัวอย่างเป็นอิสระต่อกัน (Independent sample) โดยจะทดสอบใน 3 กรณีต่อไปนี้ 1) ทดสอบอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธีการตรวจสอบภายใต้ปัจจัยที่ศึกษาต่างระดับกัน 2) ทดสอบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 2 วิธี ภายใต้ปัจจัยที่ศึกษาระดับเดียวกัน และ 3) ทดสอบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 2 วิธี ภายใต้ปัจจัยที่ศึกษาต่างระดับกัน ผลการวิเคราะห์แสดงดังตารางต่อไปนี้

ตารางที่ 68 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ระหว่างวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบต่างกัน 3 ระดับ

แหล่งของความแปรปรวน	df	SS	MS	F	p-value
ผลหลัก					
วิธีการตรวจสอบ DIF	1	34.238	34.238	13.233**	.000
ความยากของข้อสอบ (DI)	2	1.777	.888	.343	.711
ปฏิสัมพันธ์ 2 ทาง					
วิธีการตรวจสอบ DIF x DI	2	3.413	1.706	.660	.520
ความคลาดเคลื่อน	66	170.757	2.587		
รวม	71	210.184			

test of homogeneity of variances (F = 7.202, p = .000)

**p < .01

จากตารางที่ 68 เมื่อพิจารณาผลการทดสอบปฏิสัมพันธ์ 2 ทาง (two way interaction) พบว่า ปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบ DIF และความยากของข้อสอบ (DI) ไม่มีนัยสำคัญทางสถิติ แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพลด้วยวิธีการตรวจสอบต่างกันภายใต้ความยากของข้อสอบต่างกันแล้ว อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าไม่แตกต่างกัน นั่นคือ วิธีการตรวจสอบและความยากของข้อสอบไม่มีผลร่วมกันต่ออัตราความคลาดเคลื่อนประเภทที่ 1

เมื่อพิจารณาผลการทดสอบของปัจจัยความยากของข้อสอบ (DI) พบว่า ความยากของข้อสอบไม่มีนัยสำคัญทางสถิติ แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพลภายใต้ความยากของข้อสอบต่างกันแล้ว อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าไม่แตกต่างกัน

เมื่อพิจารณาผลการทดสอบของวิธีการตรวจสอบ DIF พบว่า วิธีการตรวจสอบมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพลด้วยวิธีการตรวจสอบต่างกัน จะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าแตกต่างกัน ดังนั้น จึงนำอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 2 วิธี ภายใต้ความยากของข้อสอบขนาดเดียวกัน 3 ระดับ ไปเปรียบเทียบเป็นรายคู่ โดยใช้การทดสอบ t-test กรณีกลุ่มตัวอย่างเป็นอิสระต่อกัน (Independent sample) ผลการเปรียบเทียบแสดงดังตารางที่ 69

ตารางที่ 69 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยความยากของข้อสอบ b_L b_M และ b_H โดยใช้การทดสอบ t-test

วิธีการตรวจสอบ DIF (M)		LRj (3.70)		SIBTESTs (5.48)	
		Mean difference	p-value	Mean difference	p-value
b_L	LRj (3.70)	-	-	1.78**	.007
	SIBTESTs (5.48)			-	-
วิธีการตรวจสอบ DIF (M)		LRj (4.43)		SIBTESTs (5.21)	
b_M	LRj (4.43)	-	-	0.78	.250
	SIBTESTs (5.21)			-	-
วิธีการตรวจสอบ DIF (M)		LRj (3.65)		SIBTESTs (5.22)	
b_H	LRj (3.65)	-	-	1.57*	.048
	SIBTESTs (5.22)			-	-

test of homogeneity of variances b_L (F = 13.283, p = .001), b_M (F = 6.525, p = .018), b_H (F = 17.181, p = .000)

**p < .01, *p < .05

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตารางที่ 69 พบว่า ภายใต้ความยากของข้อสอบ b_L และ b_H อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลของวิธีซิปเทสท์มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และที่ระดับ .05 ตามลำดับ ส่วนภายใต้ความยากของข้อสอบ b_M อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์มีค่าไม่แตกต่างกัน

จากผลดังกล่าว แสดงว่า ภายใต้ความยากของข้อสอบ 3 ระดับ อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์มีค่าสูงภายใต้เกือบทุกเงื่อนไขของความยากของข้อสอบ

ตารางที่ 70 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ระหว่างวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยความยาวของแบบสอบต่างกัน 2 ขนาด

แหล่งของความแปรปรวน	df	SS	MS	F	p-value
ผลหลัก					
วิธีการตรวจสอบ DIF	1	34.238	34.238	14.301**	.000
ความยาวของแบบสอบ (TL)	1	4.759	4.759	1.988	.163
ปฏิสัมพันธ์ 2 ทาง					
วิธีการตรวจสอบ DIF x TL	1	8.385	8.385	3.502	.066
ความคลาดเคลื่อน	68	162.803	2.394		
รวม	71	210.184			

test of homogeneity of variances (F = 13.615, p = .000)

**p < .01

จากตารางที่ 70 เมื่อพิจารณาผลการทดสอบปฏิสัมพันธ์ 2 ทาง (two way interaction) พบว่า ปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบ DIF และความยาวของแบบสอบ (TL) ไม่มีนัยสำคัญทางสถิติ แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลด้วยวิธีการตรวจสอบต่างกันภายใต้ความยาวของแบบสอบต่างกันแล้ว อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าไม่แตกต่างกัน นั่นคือ วิธีการตรวจสอบและความยาวของแบบสอบไม่มีผลร่วมกันต่ออัตราความคลาดเคลื่อนประเภทที่ 1

เมื่อพิจารณาผลการทดสอบของปัจจัยความยาวของแบบสอบ (TL) พบว่า ความยาวของแบบสอบไม่มีนัยสำคัญทางสถิติ แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพล ภายใต้ความยาวของแบบสอบต่างกันแล้ว อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าไม่แตกต่างกัน

เมื่อพิจารณาผลการทดสอบของวิธีการตรวจสอบ DIF พบว่า วิธีการตรวจสอบมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลด้วยวิธีการตรวจสอบต่างกัน จะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าแตกต่างกัน ดังนั้น จึงนำอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 2 วิธี ภายใต้ความยาวของ แบบสอบขนาดเดียวกัน 2 ขนาด ไปเปรียบเทียบเป็นรายคู่ โดยใช้การทดสอบ t-test กรณีกลุ่มตัวอย่างเป็นอิสระต่อกัน (Independent sample) ผลการเปรียบเทียบแสดงดังตารางที่ 71

Chulalongkorn University

ตารางที่ 71 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยความยาวของแบบสอบเดียวกัน 2 ขนาด โดยใช้การทดสอบ t-test

วิธีการตรวจสอบ DIF (M)		LRj (4.52)		SIBTESTs (5.22)	
		Mean difference	p-value	Mean difference	p-value
40 ข้อ	LRj (4.52)	-	-	0.70	.117
	SIBTESTs (5.22)			-	-
วิธีการตรวจสอบ DIF (M)		LRj (3.33)		SIBTESTs (5.39)	
60 ข้อ	LRj (3.33)	-	-	2.06**	.002
	SIBTESTs (5.39)			-	-

test of homogeneity of variances 40 ข้อ (F = 10.908, p = .002), 60 ข้อ (F = 22.527, p = .000)

**p < .01

จากตารางที่ 71 พบว่า ภายใต้ความยาวของแบบสอบ 60 ข้อ อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลของวิธีซิปเทสท์มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนภายใต้ความยาวของแบบสอบ 40 ข้อ อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลของวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์มีค่าไม่แตกต่างกัน

จากผลดังกล่าว แสดงว่า เมื่อเพิ่มความยาวของแบบสอบจาก 40 ข้อ เป็น 60 ข้อ จะมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์มีค่าเพิ่มขึ้นประมาณร้อยละ 2.06

ตารางที่ 72 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ระหว่างวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด

แหล่งของความแปรปรวน	df	SS	MS	F	p-value
ผลหลัก					
วิธีการตรวจสอบ DIF	1	34.238	34.238	13.254**	.001
สัดส่วนของ DIF (%DIF)	1	.165	.165	.064	.801
ปฏิสัมพันธ์ 2 ทาง					
วิธีการตรวจสอบ DIF x %DIF	1	.119	.119	.046	.831
ความคลาดเคลื่อน	68	175.662	2.583		
รวม	71	210.184			

test of homogeneity of variances (F = 13.077, p = .000)

**p < .01

จากตารางที่ 72 เมื่อพิจารณาผลการทดสอบปฏิสัมพันธ์ 2 ทาง (two way interaction) พบว่า ปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบ DIF และสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน (%DIF) ไม่มีนัยสำคัญทางสถิติ แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลด้วยวิธีการตรวจสอบต่างกันภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกันแล้ว อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าไม่แตกต่างกัน นั่นคือ วิธีการตรวจสอบและสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน ไม่มีผลร่วมกันต่ออัตราความคลาดเคลื่อนประเภทที่ 1

เมื่อพิจารณาผลการทดสอบของปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน (%DIF) พบว่า สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันไม่มีนัยสำคัญทางสถิติ แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพล ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนต่างกันแล้ว อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าไม่แตกต่างกัน

เมื่อพิจารณาผลการทดสอบของวิธีการตรวจสอบ DIF พบว่า วิธีการตรวจสอบมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลด้วยวิธีการตรวจสอบต่างกัน จะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าแตกต่างกัน ดังนั้น จึงนำอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 2 วิธี ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันขนาดเดียวกัน 2 ขนาด ไปเปรียบเทียบเป็นรายคู่ โดยใช้การทดสอบ t-test กรณีกลุ่มตัวอย่างเป็นอิสระต่อกัน (Independent sample) ผลการเปรียบเทียบแสดงดังตารางที่ 73

ตารางที่ 73 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน ขนาดเดียวกัน 2 ขนาด โดยใช้การทดสอบ t-test

วิธีการตรวจสอบ DIF (M)		LRj (3.92)		SIBTESTs (5.22)	
		Mean difference	p-value	Mean difference	p-value
10%	LRj (3.92)	-	-	1.30*	.0.19
	SIBTESTs (5.22)			-	-
วิธีการตรวจสอบ DIF (M)		LRj (3.93)		SIBTESTs (5.39)	
20%	LRj (3.93)	-	-	1.46*	.015
	SIBTESTs (5.39)			-	-

test of homogeneity of variances 10% (F = 18.657, p = .000), 20% (F = 19.608, p = .000)

**p < .05

จากตารางที่ 73 พบว่า ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และร้อยละ 20 อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลของวิธีซิปเทสท์มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 จากผลดังกล่าว แสดงว่า ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์มีค่าสูงภายใต้ทุกเงื่อนไขของสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน

ตารางที่ 74 ผลการวิเคราะห์ความแปรปรวนแบบ 2 ทาง ของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพล ระหว่างวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด

แหล่งของความแปรปรวน	df	SS	MS	F	p-value
ผลหลัก					
วิธีการตรวจสอบ DIF	1	34.238	34.238	37.537**	.000
ขนาดกลุ่มตัวอย่าง (Size)	2	39.200	19.600	21.489**	.000
ปฏิสัมพันธ์ 2 ทาง					
วิธีการตรวจสอบ DIF x Size	2	76.547	38.274	41.962**	.000
ความคลาดเคลื่อน	66	60.199	.912		
รวม	71	210.184			

test of homogeneity of variances (F = 2.372, p = .048)

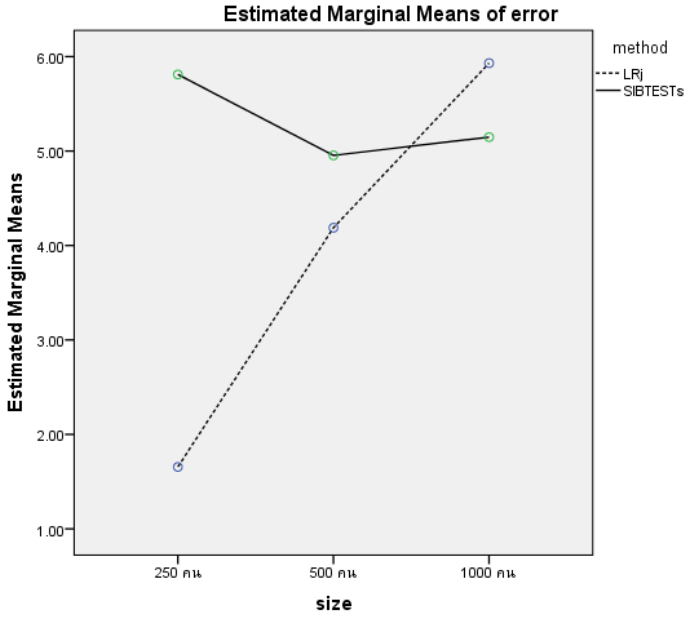
**p < .01

จากตารางที่ 74 เมื่อพิจารณาผลการทดสอบปฏิสัมพันธ์ 2 ทาง (two way interaction) พบว่า ปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบ DIF และขนาดกลุ่มตัวอย่าง (Size) มีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลด้วยวิธีการตรวจสอบต่างกันภายใต้ขนาดกลุ่มตัวอย่างต่างกัน จะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าแตกต่างกัน ดังนั้น จึงนำอัตราความคลาดเคลื่อนประเภทที่ 1 ดังกล่าว ไปเปรียบเทียบความแตกต่างโดยเปรียบเทียบใน 3 กรณี ดังนี้

กรณีที่ 1 เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธีการตรวจสอบ 2 วิธี ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย (simple effect) ผลการวิเคราะห์แสดงดังตารางที่ 75

กรณีที่ 2 เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 2 วิธี ภายใต้ขนาดกลุ่มตัวอย่างเดียวกัน 3 ขนาด โดยใช้การทดสอบ t-test กรณีกลุ่มตัวอย่างเป็นอิสระต่อกัน (Independent sample) ผลการวิเคราะห์แสดงดังตารางที่ 78

กรณีที่ 3 เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 2 วิธี ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด ผลการเปรียบเทียบแสดงดังตารางที่ 79



ภาพที่ 17 อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบ 2 วิธี ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด

ตารางที่ 75 ผลการวิเคราะห์อิทธิพลอย่างง่ายของอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพล ด้วยวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด

แหล่งของความแปรปรวน	df	SS	MS	F	p-value
วิธี LRj					
ขนาดกลุ่มตัวอย่าง	2	110.902	55.451	50.851**	.000
ความคลาดเคลื่อน	33	35.985	1.090		
รวม	35	146.887			
วิธี SIBTESTs					
ขนาดกลุ่มตัวอย่าง	2	4.845	2.423	3.302*	.049
ความคลาดเคลื่อน	33	24.214	.734		
รวม	35	29.059			

test of homogeneity of variances LRj (F = 4.142, p = .025), SIBTESTs (F = .813, p = .452)

**p < .01, *p < .05

จากตารางที่ 75 การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลของแต่ละวิธีการตรวจสอบ 2 วิธี ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด ปรากฏผลดังนี้

อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลด้วยวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่างอย่างน้อย 1 ขนาด ที่ให้ผลต่างไปจากอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ขนาดกลุ่มตัวอย่างขนาดอื่นๆ ดังนั้น จึงนำอัตราความคลาดเคลื่อนประเภทที่ 1 ไปเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Games-Howell ผลการเปรียบเทียบแสดงดังตารางที่ 76

อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลด้วยวิธีซิปเทสท์ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ภายใต้ขนาดกลุ่มตัวอย่างอย่างน้อย 1 ขนาด ที่ให้ผลต่างไปจากอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ขนาดกลุ่มตัวอย่างขนาดอื่นๆ ดังนั้น จึงนำอัตราความคลาดเคลื่อนประเภทที่ 1 ไปเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Bonferroni ผลการเปรียบเทียบแสดงดังตารางที่ 77

ตารางที่ 76 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ด้วยวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้วิธีการทดสอบของ Games-Howell

Size (M)	250 คน (1.65)		500 คน (4.19)		1000 คน (5.93)	
	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value
250 คน (1.65)	-	-	2.54**	.000	4.28**	.000
500 คน (4.19)			-	-	1.74**	.003
1000 คน (5.93)					-	-

**p < .01

จากตารางที่ 76 พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลด้วยวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่าง

1000 คน มีค่าสูงสุด รองลงมาได้แก่ภายใต้ขนาดกลุ่มตัวอย่าง 500 คน และ 250 คน ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ด้วยวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน และ 500 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ขนาดกลุ่มตัวอย่าง 500 คน มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

จากผลดังกล่าว แสดงว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลด้วยวิธีการถดถอยโลจิสติก ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน 500 คน และ 1000 คน มีค่าแตกต่างกัน ถ้าเพิ่มขนาดกลุ่มตัวอย่างจาก 250 คน เป็น 500 คน และจาก 250 คน เป็น 1000 คน จะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าเพิ่มขึ้นประมาณร้อยละ 2.54 และร้อยละ 4.28 ตามลำดับ

ตารางที่ 77 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ด้วยวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้วิธีการทดสอบของ Bonferroni

Size (M)	250 คน (5.81)		500 คน (4.95)		1000 คน (5.15)	
	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value
250 คน (5.81)	-	-	0.86 -	.059	0.66	.200
500 คน (4.95)			-	-	0.20	1.000
1000 คน (5.15)					-	-

*p < .05

จากตารางที่ 77 พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลด้วยวิธีซิปเทสท์ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน มีค่าสูงสุด รองลงมาได้แก่ภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน และ 500 คน ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ด้วยวิธีซิปเทสท์ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน 500 คน และ 1000 คน มีค่าไม่แตกต่างกัน

จากผลดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลด้วยวิธีซิปเทสท์ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน 500 คน และ 1000 คน มีอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่แตกต่างกัน

ตารางที่ 78 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างขนาดเดียวกัน 3 ขนาด โดยใช้การทดสอบ t-test

วิธีการตรวจสอบ DIF (M)		LRj (1.65)		SIBTESTs (5.81)	
		Mean difference	p-value	Mean difference	p-value
250 คน	LRj (1.65)	-	-	4.16**	.000
	SIBTESTs (5.81)			-	-
วิธีการตรวจสอบ DIF (M)		LRj (4.19)		SIBTESTs (4.95)	
500 คน	LRj (4.19)	-	-	0.76	.134
	SIBTESTs (4.95)			-	-
วิธีการตรวจสอบ DIF (M)		LRj (5.93)		SIBTESTs (5.15)	
1000 คน	LRj (5.93)	-	-	0.78**	.001
	SIBTESTs (5.15)			-	-

test of homogeneity of variances 250 คน (F = .717, p = .406), 500 คน (F = 1.445, p = .242), 1000 คน (F = .070, p = .794)

**p < .01

จากตารางที่ 78 พบว่า ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน และ 1000 คน อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลของวิธีซิปเทสท์มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนภายใต้ขนาดกลุ่มตัวอย่าง 500 คน อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลของวิธีการถดถอยโลจิสติกกับวิธีซิปเทสท์มีค่าไม่แตกต่างกัน จากผลดังกล่าว แสดงว่า ภายใต้ขนาดกลุ่มตัวอย่าง 3 ขนาด อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์มีค่าสูงภายใต้เกือบทุกเงื่อนไขของขนาดกลุ่มตัวอย่าง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 79 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ด้วยวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด

วิธีตรวจ DIF	LRj		LRj		LRj		SIBTESTs		SIBTESTs		SIBTESTs	
Size (M)	250 คน (1.65)		500 คน (4.19)		1000 คน (5.93)		250 คน (5.81)		500 คน (4.95)		1000 คน (5.15)	
	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value
LRj (250=1.65)	-	-	2.54**	.000	4.28**	.000	4.16**	.000	3.30**	.000	3.50**	.000
LRj (500=4.19)			-	-	1.74**	.003	1.62**	.003	0.76	.134	0.96*	.038
LRj (1000=5.93)					-	-	0.12	.706	0.98**	.008	0.78**	.001
SIBTESTs (250=5.81)							-	-	0.86	.059	0.66	.200
SIBTESTs (500=4.95)									-	-	0.20	1.000
SIBTESTs (1000=5.15)											-	-

**p < .01, *p < .05

จากตารางที่ 79 การทดสอบความแตกต่างของอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ ภายใต้ปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ซึ่งมีทั้งหมด 6 ค่า เมื่อพิจารณาค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ที่มีค่าสูงสุด พบว่า ค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก (1000 คน) มีค่าสูงสุด รองลงมาได้แก่ วิธีซิปเทสท์ (250 คน) เมื่อพิจารณาค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ที่มีค่าต่ำสุด พบว่า ค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก (250 คน) มีค่าต่ำสุด รองลงมาได้แก่ วิธีการถดถอยโลจิสติก (500 คน) ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบภายใต้ขนาดกลุ่มตัวอย่างดังกล่าวกับอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบภายใต้ขนาดกลุ่มตัวอย่างอื่นๆ ปรากฏผลดังนี้

อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก (1000 คน) มีค่าไม่แตกต่างกับอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการซิปเทสท์ (250 คน) แต่มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบภายใต้ขนาดกลุ่มตัวอย่างขนาดอื่น ๆ ที่เหลืออย่างมีนัยสำคัญทางสถิติที่ระดับ .01

อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ (250 คน) มีค่าไม่แตกต่างกับอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก (1000 คน) วิธีซิปเทสท์ (500 คน) และวิธีซิปเทสท์ (1000 คน) แต่มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบภายใต้ขนาดกลุ่มตัวอย่างขนาดอื่น ๆ ที่เหลืออย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และที่ระดับ .05

อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก (250 คน) มีค่าต่ำกว่าวิธีการตรวจสอบภายใต้ขนาดกลุ่มตัวอย่างขนาดอื่น ๆ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก (500 คน) มีค่าต่ำกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก (1000 คน) วิธีซิปเทสท์ (250 คน) และวิธีซิปเทสท์ (1000 คน) อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และที่ระดับ .05 แต่มีค่าไม่แตกต่างกับวิธีซิปเทสท์ (500 คน)

จากผลการเปรียบเทียบดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญ เมื่อใช้วิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน จะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าสูงสุด โดยมีค่าสูงกว่าวิธีซิปเทสท์เกือบทุกเงื่อนไขของการตรวจสอบ ส่วนการตรวจสอบด้วยวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่าง 250 คน จะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าต่ำสุด โดยมีค่าต่ำกว่าวิธีซิปเทสท์ในทุกเงื่อนไขของการตรวจสอบ

2.2.3 สรุปผลการเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพล ระหว่างวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์

สรุปผลการเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพล ระหว่างวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ ภายใต้การจัดกระทำปัจจัย 4 ปัจจัย คือ ความยากของข้อสอบ 3 ระดับ ความยาวของแบบสอบ 2 ขนาด สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด และขนาดกลุ่มตัวอย่าง 3 ขนาด สรุปได้ดังตารางที่ 80-82

ตารางที่ 80 สรุปผลการวิเคราะห์ค่าเฉลี่ยของอำนาจการทดสอบ (Power) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ด้วยวิธีการถดถอยโลจิสติก (Logistic regression) และวิธีซิปเทสท์ (SIBTEST) ภายใต้เงื่อนไขที่แตกต่างกันของปัจจัยที่ศึกษา

ปัจจัย	วิธีการตรวจสอบ DIF			
	LRj		SIBTESTs	
	Power	E_1	Power	E_1
Difficulty of Item				
b_L	8.33	3.70	41.22	5.48
b_M	14.91	4.43	54.32	5.21
b_H	10.17	3.65	26.47	5.22
Test length				
40 ข้อ	12.10	4.52	41.93	5.22
60 ข้อ	10.18	3.33	39.41	5.39
%DIF				
10%	10.97	3.92	41.39	5.22
20%	11.30	3.93	39.95	5.39
Sample size				
250 คน	2.53	1.65	21.12	5.81
500 คน	9.33	4.19	39.13	4.95
1000 คน	21.54	5.93	61.76	5.15
ค่าเฉลี่ยรวม	11.14	3.93	40.67	5.30

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตารางที่ 80 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพลด้วยวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขที่แตกต่างกันของปัจจัยที่ศึกษา พบว่า วิธีซิปเทสท์มีค่าเฉลี่ยรวมของอำนาจการทดสอบสูงสุด และเมื่อพิจารณาในแต่ละปัจจัยแต่ละระดับ ยังพบว่า มีค่าเฉลี่ยสูงกว่าอำนาจการทดสอบของวิธีการถดถอยโลจิสติกทุกเงื่อนไขของปัจจัยที่ศึกษาอีกด้วย สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกมีค่าเฉลี่ยรวมต่ำสุด และเมื่อพิจารณาในแต่ละปัจจัยแต่ละระดับ ยังพบว่า มีค่าเฉลี่ยต่ำกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ ยกเว้นปัจจัยขนาดกลุ่มตัวอย่าง 1000 คน

ตารางที่ 81 สรุปผลการเปรียบเทียบอำนาจการทดสอบ (Power) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ด้วยวิธีการถดถอยโลจิสติก (Logistic regression) และวิธีซิปเทสท์ (SIBTEST) ภายใต้เงื่อนไขของปัจจัยที่ศึกษาต่างระดับกัน

ปัจจัย		วิธีการตรวจสอบ DIF			
		LRj		SIBTESTs	
		Power	E_1	Power	E_1
★b_L	● b_M	0	0	0	0
	● b_H	0	0	0	0
★b_M	● b_L	0	0	0	0
	● b_H	0	0	+	0
★b_H	● b_L	0	0	0	0
	● b_M	0	0	-	0
★40 ข้อ	● 60 ข้อ	0	0	0	0
★10%	● 20%	0	0	0	0
★250 คน	● 500 คน	-	-	-	0
	● 1000 คน	-	-	-	0
★500 คน	● 250 คน	+	+	+	0
	● 1000 คน	-	-	-	0
★1000 คน	● 250 คน	+	+	+	0
	● 500 คน	+	+	+	0

0 หมายถึง Power หรือ E_1 ของวิธีการตรวจสอบ DIF ภายใต้ปัจจัย ★ มีค่าเท่ากับ Power หรือ E_1 ของวิธีการตรวจสอบ DIF ภายใต้ปัจจัย ●

\+ หมายถึง Power หรือ E_1 ของวิธีการตรวจสอบ DIF ภายใต้ปัจจัย ★ มีค่าสูงกว่า Power หรือ E_1 ของวิธีการตรวจสอบ DIF ภายใต้ปัจจัย ●

\- หมายถึง Power หรือ E_1 ของวิธีการตรวจสอบ DIF ภายใต้ปัจจัย ★ มีค่าต่ำกว่า Power หรือ E_1 ของวิธีการตรวจสอบ DIF ภายใต้ปัจจัย ●

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตารางที่ 81 สรุปผลการเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ด้วยวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยที่ศึกษาต่างระดับกัน พบประเด็นที่สำคัญดังนี้

1. ปัจจัยค่าความยากของข้อสอบ (b_L b_M และ b_H) พบว่า มีผลต่ออำนาจการทดสอบของวิธีซิปเทสท์ กล่าวคือ เมื่อความยากของข้อสอบ b_M จะส่งผลให้อำนาจการทดสอบสูงกว่าภายใต้เงื่อนไขความยากของข้อสอบ b_H สำหรับวิธีการถดถอยโลจิสติก พบว่า ปัจจัยค่าความยากของข้อสอบไม่มีผลต่ออำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีนี้

2. ปัจจัยความยาวของแบบสอบ (40 ข้อ และ 60 ข้อ) พบว่า ไม่มีผลต่ออำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบทั้ง 2 วิธี

3. ปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน (10% และ 20%) พบว่า ไม่มีผลต่ออำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบทั้ง 2 วิธี

4. ปัจจัยขนาดกลุ่มตัวอย่าง (250 คน 500 คน และ 1000 คน) พบว่า มีผลต่ออำนาจการทดสอบของวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ สอดคล้องกัน กล่าวคือ เมื่อขนาดกลุ่มตัวอย่าง 1000 คน จะส่งผลให้อำนาจการทดสอบสูงกว่าภายใต้ขนาดกลุ่มตัวอย่าง 250 คน และ 500 คน อีกทั้ง เมื่อขนาดกลุ่มตัวอย่าง 500 คน จะส่งผลให้อำนาจการทดสอบสูงกว่าภายใต้ขนาดกลุ่มตัวอย่าง 250 คน

นอกจากนี้ ปัจจัยขนาดกลุ่มตัวอย่างมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก กล่าวคือ เมื่อขนาดกลุ่มตัวอย่าง 1000 คน จะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าภายใต้ขนาดกลุ่มตัวอย่าง 250 คน และ 500 คน อีกทั้ง เมื่อขนาดกลุ่มตัวอย่าง 500 คน จะส่งผลให้อำนาจการทดสอบสูงกว่าภายใต้ขนาดกลุ่มตัวอย่าง 250 คน

ตารางที่ 82 สรุปผลการเปรียบเทียบอำนาจการทดสอบ (Power) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ด้วยวิธีการถดถอยโลจิสติก (Logistic regression) และวิธีซิปเทสท์ (SIBTEST) ภายใต้เงื่อนไขของปัจจัยที่ศึกษาระดับเดียวกัน

ปัจจัย	อำนาจการทดสอบ (Power)		อัตราความคลาดเคลื่อนประเภทที่ 1 (E_1)	
	◆ LRj ■ SIBTESTs	◆ SIBTESTs ■ LRj	◆ LRj ■ SIBTESTs	◆ SIBTESTs ■ LRj
b_L	-	+	-	+
b_M	-	+	0	0
b_H	-	+	-	+
40 ข้อ	-	+	0	0
60 ข้อ	-	+	-	+
10%	-	+	-	+
20%	-	+	-	+
250 คน	-	+	-	+
500 คน	-	+	0	0
1000 คน	-	+	+	-

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

0 หมายถึง Power หรือ E_1 ของวิธีการตรวจสอบ DIF ภายใต้ปัจจัย ◆ มีค่าเท่ากับ Power หรือ E_1 ของวิธีการตรวจสอบ DIF ภายใต้ปัจจัย ■

\+ หมายถึง Power หรือ E_1 ของวิธีการตรวจสอบ DIF ภายใต้ปัจจัย ◆ มีค่าสูงกว่า Power หรือ E_1 ของวิธีการตรวจสอบ DIF ภายใต้ปัจจัย ■

\- หมายถึง Power หรือ E_1 ของวิธีการตรวจสอบ DIF ภายใต้ปัจจัย ◆ มีค่าต่ำกว่า Power หรือ E_1 ของวิธีการตรวจสอบ DIF ภายใต้ปัจจัย ■

จากตารางที่ 82 สรุปผลการเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ด้วยวิธีการ

ถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยที่ศึกษาระดับเดียวกัน พบประเด็นที่สำคัญดังนี้

1. ปัจจัยค่าความยากของข้อสอบ (b_L b_M และ b_H) พบว่า วิธีซิปเทสท์มีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติกทุกระดับความยากของข้อสอบ สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีซิปเทสท์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีการถดถอยโลจิสติกที่ระดับความยากของข้อสอบ b_Lและ b_H

2. ปัจจัยความยาวของแบบสอบ (40 ข้อ และ 60 ข้อ) พบว่า วิธีซิปเทสท์มีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติก ทั้งความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีซิปเทสท์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีการถดถอยโลจิสติก ที่ความยาวของแบบสอบ 60 ข้อ

3. ปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน (10% และ 20%) พบว่า วิธีซิปเทสท์มีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติก ทั้งสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 10% และ 20% สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีซิปเทสท์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีการถดถอยโลจิสติก ทั้งสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 10% และ 20%

4. ปัจจัยขนาดกลุ่มตัวอย่าง (250 คน 500 คน และ 1000 คน) พบว่า วิธีซิปเทสท์มีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติก ทุกขนาดกลุ่มตัวอย่าง 250 คน 500 คน และ 1000 คน สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีซิปเทสท์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีการถดถอยโลจิสติก ที่ขนาดกลุ่มตัวอย่าง 250 คน ในขณะที่วิธีการถดถอยโลจิสติก มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีซิปเทสท์ ที่ขนาดกลุ่มตัวอย่าง 1000 คน

Chulalongkorn University

ตอนที่ 3 ผลการเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ในข้อมูลเชิงประจักษ์

ข้อมูลเชิงประจักษ์ จากผลการประเมินนักเรียนนานาชาติหรือ PISA 2009 ด้านการรู้เรื่องการอ่าน (reading literacy) ข้อสอบมีรูปแบบการตรวจให้คะแนนแบบทวิวิภาค จำนวน 39 ข้อ แต่ละข้อจะมีอยู่ในแบบสอบ 4 ฉบับ ฉบับของแบบสอบมีจำนวน 13 ฉบับ

กลุ่มตัวอย่างเป็นนักเรียนอายุ 15 ปี ที่กำลังศึกษาอยู่ในชั้นมัธยมศึกษาปีที่ 1 ขึ้นไป จากกลุ่มโรงเรียนทุกสังกัด ครอบคลุมทุกพื้นที่ทั่วประเทศ ผู้สอบจำนวน 6,255 คน พิจารณาคัดเลือกข้อมูลที่มีความสมบูรณ์ครบถ้วนมาใช้ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ คงเหลือ 5,875 คน จำแนกตามตัวแปรที่ใช้ในการศึกษา ได้แก่ เพศ สังกัดโรงเรียน และภูมิภาค

3.1. การวิเคราะห์ข้อมูลด้านสถิติเชิงบรรยายของคะแนนจากแบบสอบ

การวิเคราะห์คุณภาพเบื้องต้น ค่าความเที่ยงแบบความสอดคล้องภายในโดยสูตรสัมประสิทธิ์แอลฟาของครอนบาค (Alpha's Cronbach) ในแบบสอบด้านการรู้เรื่องการอ่าน จำนวน 13 ฉบับ รายละเอียดดังตารางต่อไปนี้

ตารางที่ 83 สถิติเชิงบรรยายของคะแนนจากแบบสอบ

ฉบับที่	กลุ่มนักเรียน	จำนวนคน	จำนวนข้อ	Max	Min	Range	M	SD
1	หญิง	272	15	15	2	13	7.437	2.882
	ชาย	196	15	14	1	13	6.811	2.558
	รวม	468	15	15	1	14	7.175	2.765
2	หญิง	237	17	17	2	15	9.422	3.137
	ชาย	221	17	16	0	16	8.643	3.445
	รวม	458	17	17	0	17	9.046	3.308
3	หญิง	255	9	9	0	9	3.988	1.833
	ชาย	209	9	9	1	8	3.593	1.744
	รวม	464	9	9	0	9	3.810	1.802
4	หญิง	282	20	20	1	19	9.252	3.718
	ชาย	191	20	19	2	17	7.853	3.286
	รวม	473	20	20	1	19	8.687	3.613
5	หญิง	243	13	13	2	11	6.893	2.502
	ชาย	223	13	13	0	13	5.852	2.801
	รวม	466	13	13	0	13	6.395	2.697
6	หญิง	241	22	21	3	18	12.062	3.269
	ชาย	204	22	21	3	18	10.628	3.710
	รวม	445	22	21	3	18	11.405	3.547

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 83 (ต่อ)

ฉบับที่	กลุ่มนักเรียน	จำนวนคน	จำนวนข้อ	Max	Min	Range	M	SD
7	หญิง	244	13	13	1	12	6.291	2.294
	ชาย	192	13	12	0	12	5.729	2.365
	รวม	436	13	13	0	13	6.044	2.339
8	หญิง	209	8	8	1	7	4.541	1.572
	ชาย	184	8	8	0	8	4.033	1.780
	รวม	393	8	8	0	8	4.303	1.690
9	หญิง	254	11	11	0	11	6.339	2.193
	ชาย	184	11	11	0	11	5.386	2.491
	รวม	438	11	11	0	11	5.938	2.367
10	หญิง	279	5	5	0	5	3.007	1.336
	ชาย	184	5	5	0	5	2.609	1.486
	รวม	463	5	5	0	5	2.849	1.409
11	หญิง	287	6	6	0	6	3.829	1.395
	ชาย	162	6	6	0	6	3.420	1.519
	รวม	449	6	6	0	6	3.682	1.453
12	หญิง	265	3	3	0	3	2.049	0.799
	ชาย	205	3	3	0	3	1.995	0.819
	รวม	470	3	3	0	3	2.026	0.807
13	หญิง	248	14	14	1	13	7.835	3.102
	ชาย	204	14	14	1	13	7.020	3.250
	รวม	452	14	14	1	13	7.467	3.192
รวม นร.หญิง		3316						
รวม นร.ชาย		2559						
รวมทั้งหมด		5875						

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตารางที่ 83 ผลการสอบของนักเรียนที่เข้าสอบทั้งหมด 5,875 คน เป็นนักเรียนหญิง 3,316 คน นักเรียนชาย 2,559 คน แบบสอบที่มีจำนวนข้อสูงสุด ได้แก่ ฉบับที่ 6 จำนวน 22 ข้อ ค่าคะแนนสูงสุด 21 คะแนน ค่าคะแนนต่ำสุด 3 คะแนน ค่าเฉลี่ยของคะแนนเท่ากับ 11.405 คะแนน ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 3.547 ส่วนแบบสอบที่มีจำนวนข้อต่ำสุด ได้แก่ ฉบับที่ 12 จำนวน 3 ข้อ ค่าคะแนนสูงสุด 3 คะแนน ค่าคะแนนต่ำสุด 0 คะแนน ค่าเฉลี่ยของคะแนนเท่ากับ 2.026 คะแนน ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.807

ค่าความเที่ยงแบบความสอดคล้องภายในโดยสูตรสัมประสิทธิ์แอลฟาของครอนบาคของแบบสอบด้านการรู้เรื่องการอ่าน จำนวน 13 ฉบับ รายละเอียดดังตารางที่ 84

ตารางที่ 84 ค่าความเที่ยงแบบความสอดคล้องภายในโดยสูตรสัมประสิทธิ์แอลฟาของครอนบาคของแบบสอบด้านการรู้เรื่องการอ่าน

ฉบับที่	กลุ่มนักเรียน	จำนวนผู้สอบ	จำนวนข้อสอบ	ค่าความเที่ยง
1	หญิง	272	15	.661
	ชาย	196	15	.539
	รวม	468	15	.620
2	หญิง	237	17	.677
	ชาย	221	17	.734
	รวม	458	17	.711
3	หญิง	255	9	.488
	ชาย	209	9	.420
	รวม	464	9	.465
4	หญิง	282	20	.718
	ชาย	191	20	.641
	รวม	473	20	.701
5	หญิง	243	13	.591
	ชาย	223	13	.670
	รวม	466	13	.643

ตารางที่ 84 (ต่อ)

ฉบับที่	กลุ่มนักเรียน	จำนวนผู้สอบ	จำนวนข้อสอบ	ค่าความเที่ยง
6	หญิง	241	22	.618
	ชาย	204	22	.694
	รวม	445	22	.670
7	หญิง	244	13	.525
	ชาย	192	13	.555
	รวม	436	13	.544
8	หญิง	209	8	.389
	ชาย	184	8	.523
	รวม	393	8	.469
9	หญิง	254	11	.576
	ชาย	184	11	.656
	รวม	438	11	.629
10	หญิง	279	5	.469
	ชาย	184	5	.562
	รวม	463	5	.518
11	หญิง	287	6	.475
	ชาย	162	6	.531
	รวม	449	6	.504
12	หญิง	265	3	.215
	ชาย	205	3	.207
	รวม	470	3	.212
13	หญิง	248	14	.730
	ชาย	204	14	.747
	รวม	452	14	.740

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตารางที่ 84 แบบสอบด้านการรู้เรื่องการอ่าน จำนวน 13 ฉบับ มีค่าความเที่ยงอยู่ระหว่าง .212 ถึง .740 โดยแบบสอบที่มีค่าความเที่ยงสูงสุด 3 อันดับแรก ได้แก่ แบบสอบฉบับที่ 13 จำนวน 14 ข้อ มีค่าความเที่ยงสูงสุดเท่ากับ .740 รองลงมาได้แก่ แบบสอบฉบับที่ 2 จำนวน 17 ข้อ มีค่าความเที่ยงเท่ากับ .711 และแบบสอบฉบับที่ 4 จำนวน 20 ข้อ มีค่าความเที่ยงเท่ากับ .701 จากนั้นวิเคราะห์คุณภาพรายข้อตามทฤษฎีการทดสอบแบบดั้งเดิม ได้แก่ ค่าความยาก (p) และค่าอำนาจจำแนก (r) ดังตารางที่ 85-97

ตารางที่ 85 ค่าความยาก (p) และค่าอำนาจจำแนก (r) ของแบบสอบฉบับที่ 1

ข้อที่	ความหมาย		การแปลความหมาย		ข้อที่	ความหมาย		การแปลความหมาย	
	p	r	p	r		p	r	p	r
1	0.70	0.32	ค่อนข้างง่าย	จำแนกพอใช้ได้	9	0.27	0.30	ค่อนข้างยาก	จำแนกพอใช้ได้
2	0.78	0.39	ค่อนข้างง่าย	จำแนกพอใช้ได้	10	0.31	0.36	ค่อนข้างยาก	จำแนกพอใช้ได้
3	0.45	0.50	ยากง่ายปานกลาง	จำแนกได้ดี	11	0.79	0.34	ค่อนข้างง่าย	จำแนกพอใช้ได้
4	0.44	0.41	ยากง่ายปานกลาง	จำแนกได้ดี	12	0.41	0.36	ยากง่ายปานกลาง	จำแนกพอใช้ได้
5	0.76	0.48	ค่อนข้างง่าย	จำแนกได้ดี	13	0.32	0.46	ค่อนข้างยาก	จำแนกได้ดี
6	0.50	0.48	ยากง่ายปานกลาง	จำแนกได้ดี	14	0.33	0.46	ค่อนข้างยาก	จำแนกได้ดี
7	0.25	0.31	ค่อนข้างยาก	จำแนกพอใช้ได้	15	0.31	0.46	ค่อนข้างยาก	จำแนกได้ดี
8	0.49	0.27	ยากง่ายปานกลาง	จำแนกพอใช้ได้					

จากตารางที่ 85 เมื่อวิเคราะห์ข้อสอบรายข้อในแบบสอบด้านการรู้เรื่องการอ่านฉบับที่ 1 จำนวน 15 ข้อ พบว่า มีค่าความยากระหว่าง 0.25 ถึง 0.79 และมีค่าอำนาจจำแนกระหว่าง 0.27 ถึง 0.50

ตารางที่ 86 ค่าความยาก (p) และค่าอำนาจจำแนก (r) ของแบบสอบฉบับที่ 2

ข้อที่	ความหมาย		การแปลความหมาย		ข้อที่	ความหมาย		การแปลความหมาย	
	p	r	p	r		p	r	p	r
1	0.69	0.31	ค่อนข้างง่าย	จำแนกพอใช้ได้	10	0.54	0.52	ยากง่ายปานกลาง	จำแนกได้ดี
2	0.36	0.45	ค่อนข้างยาก	จำแนกได้ดี	11	0.48	0.28	ยากง่ายปานกลาง	จำแนกพอใช้ได้
3	0.70	0.32	ค่อนข้างง่าย	จำแนกพอใช้ได้	12	0.80	0.45	ค่อนข้างง่าย	จำแนกได้ดี
4	0.49	0.35	ยากง่ายปานกลาง	จำแนกพอใช้ได้	13	0.56	0.48	ยากง่ายปานกลาง	จำแนกได้ดี
5	0.71	0.52	ค่อนข้างง่าย	จำแนกได้ดี	14	0.19	0.47	ยากมาก	จำแนกได้ดี
6	0.30	0.49	ค่อนข้างยาก	จำแนกได้ดี	15	0.83	0.37	ง่ายมาก	จำแนกพอใช้ได้
7	0.47	0.46	ยากง่ายปานกลาง	จำแนกได้ดี	16	0.44	0.51	ยากง่ายปานกลาง	จำแนกได้ดี
8	0.33	0.33	ค่อนข้างยาก	จำแนกพอใช้ได้	17	0.37	0.44	ค่อนข้างยาก	จำแนกได้ดี
9	0.81	0.47	ง่ายมาก	จำแนกได้ดี					

จากตารางที่ 86 เมื่อวิเคราะห์ข้อสอบรายข้อในแบบสอบด้านการรู้เรื่องการอ่านฉบับที่ 2 จำนวน 17 ข้อ พบว่า มีค่าความยากระหว่าง 0.19 ถึง 0.83 และมีค่าอำนาจจำแนกระหว่าง 0.28 ถึง 0.52

ตารางที่ 87 ค่าความยาก (p) และค่าอำนาจจำแนก (r) ของแบบสอบฉบับที่ 3

ข้อที่	ความหมาย		การแปลความหมาย		ข้อที่	ความหมาย		การแปลความหมาย	
	p	r	p	r		p	r	p	r
1	0.22	0.40	ค่อนข้างยาก	จำแนกได้ดี	6	0.53	0.50	ยากง่ายปานกลาง	จำแนกได้ดี
2	0.53	0.42	ยากง่ายปานกลาง	จำแนกได้ดี	7	0.43	0.44	ยากง่ายปานกลาง	จำแนกได้ดี
3	0.27	0.40	ค่อนข้างยาก	จำแนกได้ดี	8	0.36	0.44	ค่อนข้างยาก	จำแนกได้ดี
4	0.29	0.44	ค่อนข้างยาก	จำแนกได้ดี	9	0.33	0.52	ค่อนข้างยาก	จำแนกได้ดี
5	0.85	0.36	ง่ายมาก	จำแนกพอใช้ได้					

จากตารางที่ 87 เมื่อวิเคราะห์ข้อสอบรายข้อในแบบสอบด้านการรู้เรื่องการอ่านฉบับที่ 3 จำนวน 9 ข้อ พบว่า มีค่าความยากระหว่าง 0.22 ถึง 0.85 และมีค่าอำนาจจำแนกระหว่าง 0.36 ถึง 0.50

ตารางที่ 88 ค่าความยาก (p) และค่าอำนาจจำแนก (r) ของแบบสอบฉบับที่ 4

ข้อที่	ความหมาย		การแปลความหมาย		ข้อที่	ความหมาย		การแปลความหมาย	
	p	r	p	r		p	r	p	r
1	0.52	0.43	ยากง่ายปานกลาง	จำแนกได้ดี	11	0.25	0.40	ค่อนข้างยาก	จำแนกได้ดี
2	0.36	0.46	ค่อนข้างยาก	จำแนกได้ดี	12	0.51	0.25	ยากง่ายปานกลาง	จำแนกพอใช้ได้
3	0.68	0.24	ค่อนข้างง่าย	จำแนกพอใช้ได้	13	0.26	0.23	ค่อนข้างยาก	จำแนกพอใช้ได้
4	0.46	0.45	ยากง่ายปานกลาง	จำแนกได้ดี	14	0.19	0.50	ยากมาก	จำแนกได้ดี
5	0.73	0.41	ค่อนข้างง่าย	จำแนกได้ดี	15	0.28	0.33	ค่อนข้างยาก	จำแนกพอใช้ได้
6	0.28	0.47	ค่อนข้างยาก	จำแนกได้ดี	16	0.87	0.30	ง่ายมาก	จำแนกพอใช้ได้
7	0.44	0.41	ยากง่ายปานกลาง	จำแนกได้ดี	17	0.52	0.50	ยากง่ายปานกลาง	จำแนกได้ดี
8	0.33	0.29	ค่อนข้างยาก	จำแนกพอใช้ได้	18	0.48	0.38	ยากง่ายปานกลาง	จำแนกพอใช้ได้
9	0.43	0.53	ยากง่ายปานกลาง	จำแนกได้ดี	19	0.35	0.33	ค่อนข้างยาก	จำแนกพอใช้ได้
10	0.36	0.39	ค่อนข้างยาก	จำแนกพอใช้ได้	20	0.40	0.47	ยากง่ายปานกลาง	จำแนกได้ดี

จากตารางที่ 88 เมื่อวิเคราะห์ข้อสอบรายข้อในแบบสอบด้านการรู้เรื่องการอ่านฉบับที่ 4 จำนวน 20 ข้อ พบว่า มีค่าความยากระหว่าง 0.19 ถึง 0.87 และมีค่าอำนาจจำแนกระหว่าง 0.23 ถึง 0.53

ตารางที่ 89 ค่าความยาก (p) และค่าอำนาจจำแนก (r) ของแบบสอบฉบับที่ 5

ข้อที่	ความหมาย		การแปลความหมาย		ข้อที่	ความหมาย		การแปลความหมาย	
	p	r	p	r		p	r	p	r
1	0.42	0.48	ยากง่ายปานกลาง	จำแนกได้ดี	8	0.55	0.54	ยากง่ายปานกลาง	จำแนกได้ดี
2	0.65	0.35	ค่อนข้างง่าย	จำแนกพอใช้ได้	9	0.48	0.41	ยากง่ายปานกลาง	จำแนกได้ดี
3	0.47	0.39	ยากง่ายปานกลาง	จำแนกพอใช้ได้	10	0.45	0.47	ยากง่ายปานกลาง	จำแนกได้ดี
4	0.79	0.48	ค่อนข้างง่าย	จำแนกได้ดี	11	0.67	0.44	ค่อนข้างง่าย	จำแนกได้ดี
5	0.27	0.48	ค่อนข้างยาก	จำแนกได้ดี	12	0.24	0.48	ค่อนข้างยาก	จำแนกได้ดี
6	0.51	0.47	ยากง่ายปานกลาง	จำแนกได้ดี	13	0.57	0.52	ยากง่ายปานกลาง	จำแนกได้ดี
7	0.34	0.17	ค่อนข้างยาก	จำแนกไม่ดี					

จากตารางที่ 89 เมื่อวิเคราะห์ข้อสอบรายข้อในแบบสอบด้านการรู้เรื่องการอ่านฉบับที่ 5 จำนวน 13 ข้อ พบว่า มีค่าความยากระหว่าง 0.24 ถึง 0.79 และมีค่าอำนาจจำแนกระหว่าง 0.17 ถึง 0.54

ตารางที่ 90 ค่าความยาก (p) และค่าอำนาจจำแนก (r) ของแบบสอบฉบับที่ 6

ข้อที่	ความหมาย		การแปลความหมาย		ข้อที่	ความหมาย		การแปลความหมาย	
	p	r	p	r		p	r	p	r
1	0.76	0.40	ค่อนข้างง่าย	จำแนกได้ดี	12	0.30	0.36	ค่อนข้างยาก	จำแนกพอใช้ได้
2	0.52	0.42	ยากง่ายปานกลาง	จำแนกได้ดี	13	0.80	0.41	ค่อนข้างง่าย	จำแนกได้ดี
3	0.92	0.28	ง่ายมาก	จำแนกพอใช้ได้	14	0.42	0.36	ยากง่ายปานกลาง	จำแนกพอใช้ได้
4	0.45	0.41	ยากง่ายปานกลาง	จำแนกได้ดี	15	0.34	0.39	ค่อนข้างยาก	จำแนกพอใช้ได้
5	0.24	0.21	ค่อนข้างยาก	จำแนกพอใช้ได้	16	0.71	0.50	ค่อนข้างง่าย	จำแนกได้ดี
6	0.41	0.31	ยากง่ายปานกลาง	จำแนกพอใช้ได้	17	0.71	0.36	ค่อนข้างง่าย	จำแนกพอใช้ได้
7	0.28	0.24	ค่อนข้างยาก	จำแนกพอใช้ได้	18	0.89	0.28	ง่ายมาก	จำแนกพอใช้ได้
8	0.57	0.42	ยากง่ายปานกลาง	จำแนกได้ดี	19	0.27	0.40	ค่อนข้างยาก	จำแนกได้ดี
9	0.76	0.40	ค่อนข้างง่าย	จำแนกได้ดี	20	0.33	0.32	ค่อนข้างยาก	จำแนกพอใช้ได้
10	0.42	0.17	ยากง่ายปานกลาง	จำแนกไม่ดี	21	0.53	0.48	ยากง่ายปานกลาง	จำแนกได้ดี
11	0.31	0.32	ค่อนข้างยาก	จำแนกพอใช้ได้	22	0.47	0.41	ยากง่ายปานกลาง	จำแนกได้ดี

จากตารางที่ 90 เมื่อวิเคราะห์ข้อสอบรายข้อในแบบสอบด้านการรู้เรื่องการอ่านฉบับที่ 6 จำนวน 22 ข้อ พบว่า มีค่าความยากระหว่าง 0.24 ถึง 0.92 และมีค่าอำนาจจำแนกระหว่าง 0.17 ถึง 0.50

ตารางที่ 91 ค่าความยาก (p) และค่าอำนาจจำแนก (r) ของแบบสอบฉบับที่ 7

ข้อที่	ความหมาย		การแปลความหมาย		ข้อที่	ความหมาย		การแปลความหมาย	
	p	r	p	r		p	r	p	r
1	0.34	0.46	ค่อนข้างยาก	จำแนกได้ดี	8	0.94	0.20	ง่ายมาก	จำแนกพอใช้ได้
2	0.68	0.31	ค่อนข้างง่าย	จำแนกพอใช้ได้	9	0.45	0.43	ยากง่ายปานกลาง	จำแนกได้ดี
3	0.41	0.39	ยากง่ายปานกลาง	จำแนกพอใช้ได้	10	0.43	0.28	ยากง่ายปานกลาง	จำแนกพอใช้ได้
4	0.64	0.45	ค่อนข้างง่าย	จำแนกได้ดี	11	0.30	0.41	ค่อนข้างยาก	จำแนกได้ดี
5	0.29	0.37	ค่อนข้างยาก	จำแนกพอใช้ได้	12	0.19	0.49	ยากมาก	จำแนกได้ดี
6	0.39	0.45	ค่อนข้างยาก	จำแนกได้ดี	13	0.64	0.49	ค่อนข้างง่าย	จำแนกได้ดี
7	0.32	0.36	ค่อนข้างยาก	จำแนกพอใช้ได้					

จากตารางที่ 91 เมื่อวิเคราะห์ข้อสอบรายข้อในแบบสอบด้านการรู้เรื่องการอ่านฉบับที่ 7 จำนวน 13 ข้อ พบว่า มีค่าความยากระหว่าง 0.19 ถึง 0.94 และมีค่าอำนาจจำแนกระหว่าง 0.20 ถึง 0.49

ตารางที่ 92 ค่าความยาก (p) และค่าอำนาจจำแนก (r) ของแบบสอบฉบับที่ 8

ข้อที่	ความหมาย		การแปลความหมาย		ข้อที่	ความหมาย		การแปลความหมาย	
	p	r	p	r		p	r	p	r
1	0.78	0.44	ค่อนข้างง่าย	จำแนกได้ดี	5	0.38	0.39	ค่อนข้างยาก	จำแนกพอใช้ได้
2	0.65	0.51	ค่อนข้างง่าย	จำแนกได้ดี	6	0.34	0.39	ค่อนข้างยาก	จำแนกพอใช้ได้
3	0.41	0.57	ยากง่ายปานกลาง	จำแนกได้ดี	7	0.30	0.40	ค่อนข้างยาก	จำแนกได้ดี
4	0.86	0.43	ง่ายมาก	จำแนกได้ดี	8	0.58	0.55	ยากง่ายปานกลาง	จำแนกได้ดี

จากตารางที่ 92 เมื่อวิเคราะห์ข้อสอบรายข้อในแบบสอบด้านการรู้เรื่องการอ่านฉบับที่ 8 จำนวน 8 ข้อ พบว่า มีค่าความยากระหว่าง 0.30 ถึง 0.86 และมีค่าอำนาจจำแนกระหว่าง 0.39 ถึง 0.57

ตารางที่ 93 ค่าความยาก (p) และค่าอำนาจจำแนก (r) ของแบบสอบฉบับที่ 9

ข้อที่	ความหมาย		การแปลความหมาย		ข้อที่	ความหมาย		การแปลความหมาย	
	p	r	p	r		p	r	p	r
1	0.72	0.34	ค่อนข้างง่าย	จำแนกพอใช้ได้	7	0.88	0.39	ง่ายมาก	จำแนกพอใช้ได้
2	0.71	0.53	ค่อนข้างง่าย	จำแนกได้ดี	8	0.40	0.45	ยากง่ายปานกลาง	จำแนกได้ดี
3	0.40	0.52	ยากง่ายปานกลาง	จำแนกได้ดี	9	0.39	0.34	ค่อนข้างยาก	จำแนกพอใช้ได้
4	0.40	0.48	ยากง่ายปานกลาง	จำแนกได้ดี	10	0.33	0.39	ค่อนข้างยาก	จำแนกพอใช้ได้
5	0.63	0.61	ค่อนข้างง่าย	จำแนกได้ดีมาก	11	0.65	0.54	ค่อนข้างง่าย	จำแนกได้ดี
6	0.44	0.47	ยากง่ายปานกลาง	จำแนกได้ดี					

จากตารางที่ 93 เมื่อวิเคราะห์ข้อสอบรายข้อในแบบสอบด้านการรู้เรื่องการอ่านฉบับที่ 9 จำนวน 11 ข้อ พบว่า มีค่าความยากระหว่าง 0.33 ถึง 0.88 และมีค่าอำนาจจำแนกระหว่าง 0.34 ถึง 0.61

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 94 ค่าความยาก (p) และค่าอำนาจจำแนก (r) ของแบบสอบฉบับที่ 10

ข้อที่	ความหมาย		การแปลความหมาย		ข้อที่	ความหมาย		การแปลความหมาย	
	p	r	p	r		p	r	p	r
1	0.62	0.62	ค่อนข้างง่าย	จำแนกได้ดีมาก	4	0.69	0.62	ค่อนข้างง่าย	จำแนกได้ดีมาก
2	0.46	0.60	ยากง่ายปานกลาง	จำแนกได้ดีมาก	5	0.67	0.54	ค่อนข้างง่าย	จำแนกได้ดี
3	0.41	0.55	ยากง่ายปานกลาง	จำแนกได้ดี					

จากตารางที่ 94 เมื่อวิเคราะห์ข้อสอบรายข้อในแบบสอบด้านการรู้เรื่องการอ่านฉบับที่ 10 จำนวน 5 ข้อ พบว่า มีค่าความยากระหว่าง 0.41 ถึง 0.69 และมีค่าอำนาจจำแนกระหว่าง 0.54 ถึง 0.62

ตารางที่ 95 ค่าความยาก (p) และค่าอำนาจจำแนก (r) ของแบบสอบฉบับที่ 11

ข้อที่	ความหมาย		การแปลความหมาย		ข้อที่	ความหมาย		การแปลความหมาย	
	p	r	p	r		p	r	p	r
1	0.74	0.55	ค่อนข้างง่าย	จำแนกได้ดี	4	0.92	0.31	ง่ายมาก	จำแนกพอใช้ได้
2	0.53	0.64	ยากง่ายปานกลาง	จำแนกได้ดีมาก	5	0.67	0.56	ค่อนข้างง่าย	จำแนกได้ดี
3	0.36	0.54	ค่อนข้างยาก	จำแนกได้ดี	6	0.47	0.58	ยากง่ายปานกลาง	จำแนกได้ดี

จากตารางที่ 95 เมื่อวิเคราะห์ข้อสอบรายข้อในแบบสอบด้านการรู้เรื่องการอ่านฉบับที่ 11 จำนวน 6 ข้อ พบว่า มีค่าความยากระหว่าง 0.36 ถึง 0.92 และมีค่าอำนาจจำแนกระหว่าง 0.31 ถึง 0.64

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 96 ค่าความยาก (p) และค่าอำนาจจำแนก (r) ของแบบสอบฉบับที่ 12

ข้อที่	ความหมาย		การแปลความหมาย		ข้อที่	ความหมาย		การแปลความหมาย	
	p	r	p	r		p	r	p	r
1	0.89	0.47	ง่ายมาก	จำแนกได้ดี					
2	0.70	0.66	ค่อนข้างง่าย	จำแนกได้ดีมาก					
3	0.44	0.71	ยากง่ายปานกลาง	จำแนกได้ดีมาก					

จากตารางที่ 96 เมื่อวิเคราะห์ข้อสอบรายข้อในแบบสอบด้านการรู้เรื่องการอ่านฉบับที่ 12 จำนวน 3 ข้อ พบว่า มีค่าความยากระหว่าง 0.44 ถึง 0.89 และมีค่าอำนาจจำแนกระหว่าง 0.47 ถึง 0.71

ตารางที่ 97 ค่าความยาก (p) และค่าอำนาจจำแนก (r) ของแบบสอบฉบับที่ 13

ข้อที่	ความหมาย		การแปลความหมาย		ข้อที่	ความหมาย		การแปลความหมาย	
	p	r	p	r		p	r	p	r
1	0.72	0.50	ค่อนข้างง่าย	จำแนกได้ดี	8	0.49	0.46	ยากง่ายปานกลาง	จำแนกได้ดี
2	0.78	0.39	ค่อนข้างง่าย	จำแนกพอใช้ได้	9	0.36	0.44	ค่อนข้างยาก	จำแนกได้ดี
3	0.77	0.46	ค่อนข้างง่าย	จำแนกได้ดี	10	0.49	0.53	ยากง่ายปานกลาง	จำแนกได้ดี
4	0.56	0.56	ยากง่ายปานกลาง	จำแนกได้ดี	11	0.34	0.40	ค่อนข้างยาก	จำแนกได้ดี
5	0.37	0.51	ค่อนข้างยาก	จำแนกได้ดี	12	0.49	0.38	ยากง่ายปานกลาง	จำแนกพอใช้ได้
6	0.37	0.46	ค่อนข้างยาก	จำแนกได้ดี	13	0.58	0.50	ยากง่ายปานกลาง	จำแนกได้ดี
7	0.69	0.61	ค่อนข้างง่าย	จำแนกได้ดีมาก	14	0.48	0.50	ยากง่ายปานกลาง	จำแนกได้ดี

จากตารางที่ 97 เมื่อวิเคราะห์ข้อสอบรายข้อในแบบสอบด้านการรู้เรื่องการอ่านฉบับที่ 13 จำนวน 14 ข้อ พบว่า มีค่าความยากระหว่าง 0.34 ถึง 0.78 และมีค่าอำนาจจำแนกระหว่าง 0.38 ถึง 0.61

ผู้วิจัยนำคะแนนของผู้สอบทั้งหมดมาเปลี่ยนเป็นร้อยละ แล้วจึงวิเคราะห์ความแปรปรวนแบบทางเดียว (one way ANOVA) เพื่อตรวจสอบความแตกต่างของค่าเฉลี่ยระหว่างแบบสอบ 13 ฉบับ ผลแสดงดังตารางที่ 98

ตารางที่ 98 ผลการวิเคราะห์ความแปรปรวนแบบทางเดียวของคะแนน ระหว่างแบบสอบ 13 ฉบับ

แหล่งของความแปรปรวน	df	SS	MS	F	p-value
ระหว่างกลุ่ม	12	270618.265	22551.522	48.803**	.000
ภายในกลุ่ม	5862	2708768.947	462.090		
รวม	5874	2979387.212			

test of homogeneity of variances (F = 24.483, p = .000)

**p < .01

จากตารางที่ 98 พบว่า คะแนนของแบบสอบ 13 ฉบับ มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า คะแนนของแบบสอบอย่างน้อย 1 ฉบับ มีค่าแตกต่างไปจากคะแนนของแบบสอบฉบับอื่น ๆ ดังนั้น จึงนำคะแนนไปเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Games-Howell ผลการเปรียบเทียบแสดงดังตารางที่ 99

ตารางที่ 99 ผลการเปรียบเทียบคะแนนระหว่างแบบสอบ 13 ฉบับ โดยใช้วิธีการทดสอบของ Games-Howell

ฉบับ	1		2		3		4		5		6		7	
	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value
1	-	-	5.37**	.001	5.49**	.001	4.39*	.014	1.35	.998	4.00*	.028	1.34	.997
2			-	-	10.87**	.000	9.77**	.000	4.01	.113	1.37	.995	6.72**	.000
3					-	-	1.09	1.000	6.85**	.000	9.50**	.000	4.15	.058
4							-	-	5.75**	.000	8.40**	.000	3.05	.342
5									-	-	2.64	.624	2.70	.669
6											-	-	5.34**	.000
7													-	-

**p < .01, *p< .05

ตารางที่ 99 (ต่อ)

ฉบับ	8		9		10		11		12		13	
	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value
1	5.95**	.001	6.15**	.000	9.14**	.000	13.52**	.000	19.68**	.000	5.49**	.004
2	0.54	1.000	0.77	1.000	3.76	.473	8.14**	.000	14.30**	.000	0.12	1.000
3	11.44**	.000	11.64**	.000	14.63**	.000	19.02**	.000	25.18**	.000	10.99**	.000
4	10.34**	.000	10.54**	.000	13.54**	.000	17.92**	.000	24.08**	.000	9.89**	.000
5	4.59	.072	4.79*	.038	7.78**	.000	12.16**	.000	18.32**	.000	4.14	.170
6	1.94	.959	2.14	.904	5.13*	.041	9.51**	.000	15.67**	.000	1.49	.996
7	7.29**	.000	7.49**	.000	10.48**	.000	14.86**	.000	21.02**	.000	6.84**	.000
8	-	-	0.20	1.000	3.19	.801	7.57**	.000	13.73**	.000	0.45	1.000
9			-	-	2.99	.851	7.37**	.000	13.53**	.000	0.65	1.000
10					-	-	4.38	.362	10.54**	.000	3.64	.627
11							-	-	6.15*	.017	8.02**	.000
12									-	-	14.18**	.000
13											-	-

**p < .01, *p< .05

จากตารางที่ 99 พบว่า แบบสอบฉบับที่ 1 มีคะแนนต่างจากแบบทดสอบฉบับอื่น ๆ ยกเว้น ฉบับที่ 5 และ 7

แบบสอบฉบับที่ 2 มีคะแนนต่างจากแบบทดสอบฉบับอื่น ๆ ยกเว้น ฉบับที่ 5, 6, 8, 9, 10 และ 13

แบบสอบฉบับที่ 3 มีคะแนนต่างจากแบบทดสอบฉบับอื่น ๆ ยกเว้น ฉบับที่ 4 และ 7

แบบสอบฉบับที่ 4 มีคะแนนต่างจากแบบทดสอบฉบับอื่น ๆ ยกเว้น ฉบับที่ 3 และ 7

แบบสอบฉบับที่ 5 มีคะแนนต่างจากแบบทดสอบฉบับอื่น ๆ ยกเว้น ฉบับที่ 1, 2, 6, 7, 8 และ 13

แบบสอบฉบับที่ 6 มีคะแนนต่างจากแบบทดสอบฉบับอื่น ๆ ยกเว้น ฉบับที่ 2, 5, 8, 9 และ 13

แบบสอบฉบับที่ 7 มีคะแนนต่างจากแบบทดสอบฉบับอื่น ๆ ยกเว้น ฉบับที่ 1, 3, 4, และ 5

แบบสอบฉบับที่ 8 มีคะแนนต่างจากแบบทดสอบฉบับอื่น ๆ ยกเว้น ฉบับที่ 2, 5, 6, 9, 10 และ 13

แบบสอบฉบับที่ 9 มีคะแนนต่างจากแบบทดสอบฉบับอื่น ๆ ยกเว้น ฉบับที่ 2, 6, 8, 10 และ 13

แบบสอบฉบับที่ 10 มีคะแนนต่างจากแบบทดสอบฉบับอื่น ๆ ยกเว้น ฉบับที่ 2, 8, 9, 11 และ 13

แบบสอบฉบับที่ 11 มีคะแนนต่างจากแบบทดสอบฉบับอื่น ๆ ยกเว้น ฉบับที่ 10

แบบสอบฉบับที่ 12 มีคะแนนต่างจากแบบทดสอบฉบับอื่น ๆ ทั้งหมด

แบบสอบฉบับที่ 13 มีคะแนนต่างจากแบบทดสอบฉบับอื่น ๆ ยกเว้น ฉบับที่ 2, 5, 6, 8, 9 และ 10

3.2 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

ก่อนที่จะนำแบบสอบด้านการรู้เรื่องการอ่าน จำนวน 13 ฉบับ ไปวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ต้องนำมาวิเคราะห์ด้วยวิธีราสช์ทรีเพื่อกำหนดกลุ่มอ้างอิงและกลุ่มเปรียบเทียบตามแต่ละตัวแปร ซึ่งตัวแปรเพศ แบ่งเป็น 2 กลุ่มตามลักษณะอยู่แล้ว ส่วนตัวแปรสังกัดของโรงเรียนและภูมิภาค มี 8 กลุ่ม และ 9 กลุ่ม ตามลำดับ ดังนั้น จึงต้องวิเคราะห์ด้วยวิธีราสช์ทรีเพื่อให้ทราบว่า ประเภทใดบ้างในตัวแปรนั้น ๆ จะเป็นกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ และจะวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบกี่ตัวแปรในแต่ละฉบับของแบบสอบ หลังจากนั้น จึงนำไปวิเคราะห์ด้วยวิธีการวัดพื้นที่ของราชูซึ่งเป็นวิธีเกณฑ์ที่ใช้ใน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

การเปรียบเทียบประสิทธิภาพด้านอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี รายละเอียดของตัวแปรที่ใช้ในการแบ่งกลุ่ม ดังตารางที่ 100

ตารางที่ 100 รายละเอียดของตัวแปรที่ใช้ในการแบ่งกลุ่มเพื่อวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ

ตัวแปร	ประเภท	
เพศ	1 = หญิง	2 = ชาย
สังกัดโรงเรียน	1 = สพฐ.1 (สปช. เดิม)	5 = กศท.
	2 = สพฐ.2 (สศ. เดิม)	6 = สาธิต
	3 = สช.	7 = อศ.1 (สช. อาชีว)
	4 = กทม.	8 = อศ.2 (อาชีวรัฐ)
ภูมิภาค	1 = กทม และปริมณฑล	6 = อีสานล่าง
	2 = กลาง	7 = ใต้
	3 = เหนือบน	8 = ตะวันออก
	4 = เหนือล่าง	9 = ตะวันตก
	5 = อีสานบน	

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ผู้วิจัยวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีราสช์ทรีเพื่อให้ทราบว่าจะใช้ตัวแปรใดและแบบสอบฉบับใด และเพื่อแบ่งกลุ่มอ้างอิงกับกลุ่มเปรียบเทียบ จากนั้นจึงวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวัดพื้นที่ของราชู ในตัวแปรเพศที่มีกลุ่มอ้างอิงและกลุ่มเปรียบเทียบแล้วเพื่อยืนยันการเกิดการทำหน้าที่ต่างกันของข้อสอบในแบบสอบแต่ละฉบับ ดังตารางที่ 101

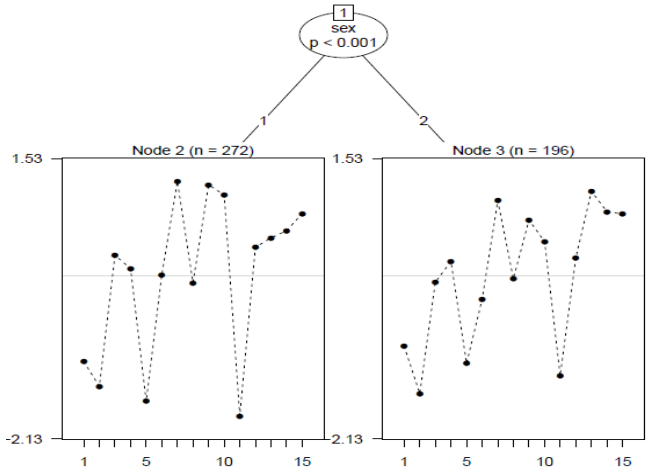
ตารางที่ 101 ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ

แบบสอบ	ตัวแปร			
	เพศ		สังกัด	ภูมิภาค
	Raschtree	Raju		
1	✓	✓	-	-
2	-	✓	-	-
3	-	-	-	-
4	✓	✓	-	-
5	✓	✓	-	-
6	✓	✓	✓(2,6/1,3,4,5,7,8)	-
7	✓	✓	-	-
8	✓	✓	✓(1,2,3,5/4,6,7,8)	✓(1,3,7/2,4,5,6,8,9)
9	-	✓	-	-
10	-	-	-	-
11	-	✓	-	-
12	-	-	-	-
13	✓	✓	✓(4,6/1,2,3,5,7,8)	✓(3,5,7/1,2,4,6,8,9)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตารางที่ 101 ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีราสช์ทรีและวิธีการวัดพื้นที่ของราชู พบว่า แบบสอบฉบับที่ 3 10 และ 12 ไม่พบการทำหน้าที่ต่างกันของข้อสอบ ดังนั้น ตัวแปรเพศจะใช้วิเคราะห์กับแบบสอบฉบับที่ 1, 2, 4, 5, 6, 7, 8, 9, 11 และ 13 สำหรับตัวแปรสังกัดของโรงเรียนจะใช้วิเคราะห์กับแบบสอบฉบับที่ 6, 8 และ 13 ส่วนตัวแปรภูมิภาคจะใช้วิเคราะห์กับแบบสอบฉบับที่ 8 และ 13

ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีราสช์ทรี จำแนกตามตัวแปรเพศ ในแบบสอบฉบับที่ 1 จำนวน 15 ข้อ แสดงดังภาพ



ภาพที่ 18 ผลการวิเคราะห์ DIF ด้วยวิธีราสช์ทรี ในแบบสอบฉบับที่ 1

ตารางที่ 102 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการวัดพื้นที่ของราชู วิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ของแบบสอบฉบับที่ 1

ข้อ	Raju			LR					SIBTEST			Raschtree	
	ค่าสถิติ	p-value		(g)		(g by x)			Beta	p-value		ผลต่าง	
				B	Sig.	B	Sig.					ค่า b	
1	0.7876	0.4310	-	-.437	.499	.033	.735	-	0.077	0.103	-	-0.2020	-
2	-0.4978	0.6186	-	-.437	.585	.102	.457	-	-0.022	0.580	-	0.0931	-
3	-1.6591	0.0971	-	-1.217	.142	.238	.038*	✓	-0.111	0.016*	✓	0.3505	-
4	0.4065	0.6844	-	-.412	.540	.041	.652	-	0.039	0.415	-	-0.0918	-
5	1.9622	0.0497*	✓	-2.085	.041*	.266	.140	✓	0.060	0.122	-	-0.4951	✓
6	-1.5433	0.1228	-	.485	.503	-.015	.884	-	-0.081	0.105	-	0.3156	-
7	-0.9080	0.3639	-	1.446	.041*	-.166	.053	✓	0.002	0.955	-	0.2508	-
8	0.1932	0.8468	-	-.111	.848	-.002	.981	-	0.037	0.444	-	-0.0515	-
9	-1.8326	0.0669	-	1.036	.129	-.088	.294	-	-0.059	0.178	-	0.4592	✓
10	-2.5755	0.0100**	✓	1.550	.026*	-.127	.138	✓	-0.122	0.006**	✓	0.6099	✓
11	2.0040	0.0451*	✓	-.652	.386	.014	.908	-	0.075	0.067	-	-0.5245	✓
12	-0.6320	0.5274	-	.770	.223	-.094	.255	-	-0.004	0.932	-	0.1382	-
13	2.6109	0.0090**	✓	.181	.817	-.111	.249	-	0.138	0.001**	✓	-0.6071	✓
14	1.1382	0.2551	-	-.918	.256	.082	.414	-	0.022	0.617	-	-0.2501	-
15	0.0565	0.9550	-	1.287	.099	-.165	.085	-	0.000	0.996	-	0.0046	-

**p < .01, *p < .05

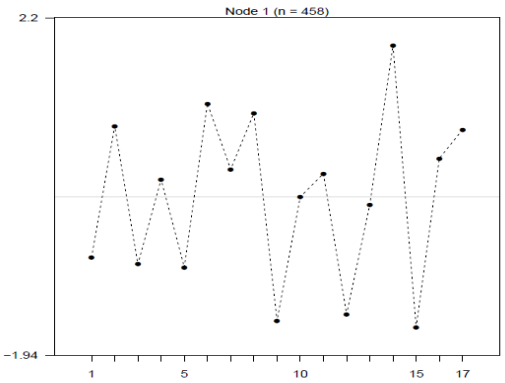
จากตารางที่ 102 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแบบสอบด้านการรู้เรื่องการอ่าน ฉบับที่ 1 จำนวน 15 ข้อ ด้วยวิธีการวัดพื้นที่ของราชูซึ่งเป็นวิธีเกณฑ์ พบข้อสอบที่ทำหน้าที่ต่างกัน 4 ข้อ ได้แก่ ข้อที่ 5, 10, 11 และ 13 ส่วนข้อสอบที่ทำหน้าที่ไม่ต่างกัน 11 ข้อ ได้แก่ ข้อที่ 1, 2, 3, 4, 6, 7, 8, 9, 12, 14 และ 15

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี โดยทั้ง 3 วิธี พบข้อสอบที่ทำหน้าที่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 1 ข้อ ได้แก่ ข้อที่ 10 สำหรับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก และวิธีราสช์ทรี พบข้อสอบที่ทำหน้าที่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 1 ข้อ ได้แก่ ข้อที่ 5 ส่วนการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีซิปเทสท์ และวิธีราสช์ทรี พบข้อสอบที่ทำหน้าที่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 1 ข้อ ได้แก่ ข้อที่ 13 ในขณะที่การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีราสช์ทรี เพียงวิธีเดียว พบข้อสอบที่ทำหน้าที่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 1 ข้อ ได้แก่ ข้อที่ 11

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี โดยทั้ง 3 วิธี พบข้อสอบที่ทำหน้าที่ไม่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 8 ข้อ ได้แก่ ข้อที่ 1, 2, 4, 6, 8, 12, 14 และ 15

จากผลดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแบบสอบด้านการรู้เรื่องการอ่าน ฉบับที่ 1 จำนวน 15 ข้อ วิธีราสช์ทรีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ตรงกับวิธีเกณฑ์มากที่สุด จำนวน 4 ข้อ ส่วนวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ตรงกับวิธีเกณฑ์ จำนวน 2 ข้อ เท่ากัน

ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีราสช์ทรี จำแนกตามตัวแปรเพศ ในแบบสอบฉบับที่ 2 จำนวน 17 ข้อ แสดงดังภาพ



ภาพที่ 19 ผลการวิเคราะห์ DIF ด้วยวิธีราสช์ทรี ในแบบสอบฉบับที่ 2

ตารางที่ 103 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการวัดพื้นที่ของราชู วิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ของแบบสอบฉบับที่ 2

ข้อ	Raju			LR					SIBTEST			Raschtree	
	ค่าสถิติ	p-value		(g)		(g by x)			Beta	p-value		ผลต่าง	
				B	Sig.	B	Sig.					ค่า b	
1	-0.1763	0.8601	-	-.772	.215	.089	.209	-	0.008	0.860	-	0.0000	-
2	-0.3671	0.7135	-	1.663	.045*	-.163	.044*	✓	0.022	0.622	-	0.0000	-
3	-2.1005	0.0357 *	✓	-.871	.173	.163	.032*	✓	-0.068	0.109	-	0.0000	-
4	-0.1281	0.8981	-	-.236	.709	.021	.747	-	0.026	0.588	-	0.0000	-
5	1.0465	0.2953	-	1.306	.125	-.200	.065	-	0.044	0.273	-	0.0000	-
6	-0.4081	0.6832	-	2.070	.035*	-.192	.037*	✓	0.009	0.819	-	0.0000	-
7	-0.6333	0.5265	-	.800	.282	-.072	.349	-	-0.024	0.602	-	0.0000	-
8	0.1205	0.9041	-	-.838	.250	.075	.289	-	0.036	0.434	-	0.0000	-
9	-0.1236	0.9017	-	-.232	.788	.067	.583	-	-0.007	0.837	-	0.0000	-
10	1.5588	0.1190	-	-.911	.247	.059	.490	-	0.090	0.041*	✓	0.0000	-
11	-0.7471	0.4550	-	.515	.396	-.051	.421	-	0.030	0.541	-	0.0000	-
12	0.6814	0.4956	-	-.391	.633	.044	.694	-	0.001	0.985	-	0.0000	-
13	-0.7983	0.4247	-	.849	.245	-.074	.352	-	-0.014	0.756	-	0.0000	-
14	-0.1725	0.8631	-	2.537	.059	-.222	.056	-	-0.008	0.831	-	0.0000	-
15	1.7300	0.0836	-	-2.515	.001**	.280	.006**	✓	0.039	0.266	-	0.0000	-
16	-0.1000	0.9204	-	.495	.543	-.050	.545	-	0.010	0.819	-	0.0000	-
17	0.1887	0.8503	-	-.371	.638	.029	.704	-	0.025	0.573	-	0.0000	-

**p < .01, *p < .05

จุฬาลงกรณ์มหาวิทยาลัย

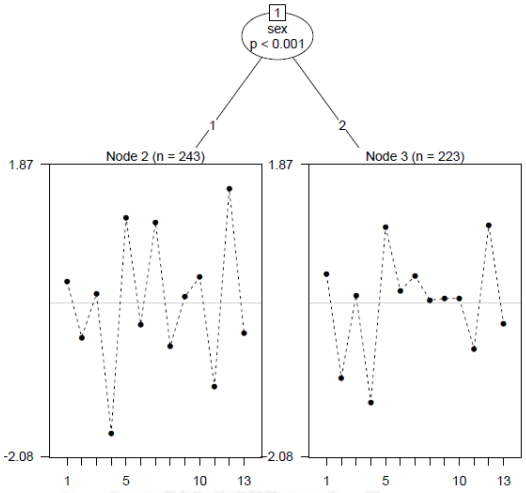
จากตารางที่ 103 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแบบสอบด้านการรู้เรื่องการอ่าน ฉบับที่ 2 จำนวน 17 ข้อ ด้วยวิธีการวัดพื้นที่ของราชูซึ่งเป็นวิธีเกณฑ์ พบข้อสอบที่ทำหน้าที่ต่างกัน 1 ข้อ ได้แก่ ข้อที่ 3 ส่วนข้อสอบที่ทำหน้าที่ไม่ต่างกัน 16 ข้อ ได้แก่ ข้อที่ 1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 และ 17

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี พบว่า วิธีการถดถอยโลจิสติกเพียงวิธีเดียวที่ตรวจพบข้อสอบที่ทำหน้าที่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 1 ข้อ ได้แก่ ข้อที่ 3 สำหรับวิธีซิปเทสท์ตรวจพบข้อสอบที่ทำหน้าที่ต่างกันไม่ตรงกับวิธีเกณฑ์ ส่วนวิธีราสช์ทรีตรวจไม่พบข้อสอบที่ทำหน้าที่ต่างกัน

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี โดยทั้ง 3 วิธี พบข้อสอบที่ทำหน้าที่ไม่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 12 ข้อ ได้แก่ ข้อที่ 1, 4, 5, 7, 8, 9, 11, 12, 13, 14, 16 และ 17

จากผลดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแบบสอบด้านการรู้เรื่องการอ่าน ฉบับที่ 2 จำนวน 17 ข้อ วิธีการถดถอยโลจิสติกตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ตรงกับวิธีเกณฑ์มากที่สุด จำนวน 1 ข้อ ส่วนวิธีซิปเทสท์และวิธีราสช์ทรีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบไม่ตรงกับวิธีเกณฑ์

ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีราสช์ทรี จำแนกตามตัวแปรเพศ ในแบบสอบฉบับที่ 4 จำนวน 20 ข้อ แสดงดังภาพ



ภาพที่ 20 ผลการวิเคราะห์ DIF ด้วยวิธีราสช์ทรี ในแบบสอบฉบับที่ 4

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 104 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการวัดพื้นที่ของราชู วิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ของแบบสอบฉบับที่ 4

ข้อ	Raju			LR					SIBTEST			Raschtree	
	ค่าสถิติ	p-value		(g)		(g by x)			Beta	p-value		ผลต่าง	
				B	Sig.	B	Sig.					ค่า b	
1	0.5850	0.5585	-	-1.334	.040*	.156	.045*	✓	0.041	0.405	-	-0.1281	-
2	1.1104	0.2668	-	.435	.528	-.070	.330	-	0.078	0.102	-	-0.2439	-
3	-2.3975	0.0165*	✓	.483	.394	-.007	.922	-	-0.042	0.414	-	0.5097	✓
4	-0.7397	0.4595	-	-.076	.907	.040	.593	-	-0.020	0.701	-	0.1631	-
5	1.7596	0.0785	-	-1.004	.140	.089	.338	-	0.049	0.267	-	-0.4378	✓
6	-0.2296	0.8184	-	.519	.514	-.033	.675	-	0.001	0.987	-	0.0671	-
7	0.1077	0.9142	-	-.551	.380	.064	.357	-	0.036	0.505	-	-0.0181	-
8	0.8126	0.4165	-	-.116	.845	-.019	.759	-	0.056	0.263	-	-0.1783	-
9	2.1289	0.0333*	✓	-1.195	.117	.092	.274	-	0.084	0.099	-	-0.4672	✓
10	-0.6013	0.5476	-	1.660	.010**	-.168	.012*	✓	0.057	0.212	-	0.1459	-
11	-0.1704	0.8647	-	.868	.240	-.081	.263	-	-0.010	0.836	-	0.0539	-
12	-1.4044	0.1602	-	.226	.671	-.007	.907	-	0.019	0.716	-	0.2989	-
13	-1.1701	0.2420	-	.171	.783	-.004	.951	-	0.032	0.473	-	0.2978	-
14	-0.0908	0.9277	-	.499	.642	-.020	.836	-	-0.011	0.776	-	0.0351	-
15	-0.9594	0.3374	-	.207	.751	-.002	.974	-	0.029	0.541	-	0.2397	-
16	0.7242	0.4689	-	-1.010	.228	.131	.310	-	0.002	0.948	-	-0.2770	-
17	1.3524	0.1762	-	1.011	.128	-.151	.051	-	0.054	0.288	-	-0.2932	-
18	0.4473	0.6546	-	-.010	.986	-.012	.851	-	0.042	0.423	-	-0.0946	-
19	-1.6492	0.0991	-	.098	.872	.025	.693	-	-0.037	0.499	-	0.3804	-
20	0.2832	0.7770	-	.506	.453	-.055	.452	-	0.056	0.275	-	-0.0534	-

**p < .01, *p < .05

จากตารางที่ 104 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแบบสอบด้านการรู้เรื่องการอ่าน ฉบับที่ 4 จำนวน 20 ข้อ ด้วยวิธีการวัดพื้นที่ของราชูซึ่งเป็นวิธีเกณฑ์ พบข้อสอบที่ทำหน้าที่ต่างกัน 2 ข้อ ได้แก่ ข้อที่ 3 และ 9 ส่วนข้อสอบที่ทำหน้าที่ไม่ต่างกัน 18 ข้อ ได้แก่ ข้อที่ 1, 2, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 และ 20

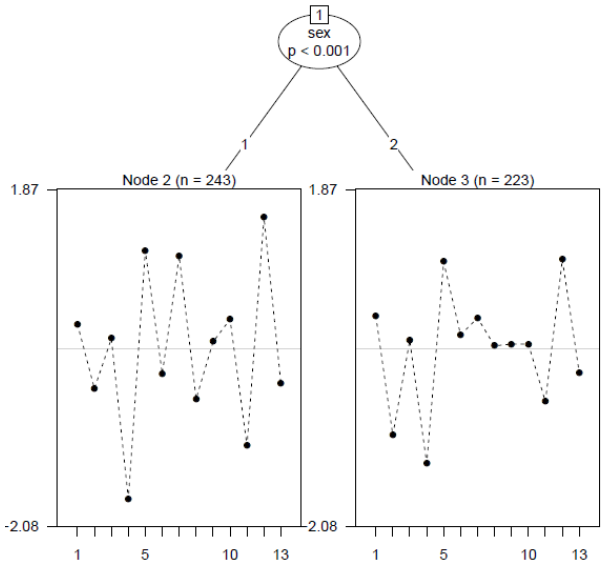
ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี พบว่า วิธีราสช์ทรีเพียงวิธีเดียวที่ตรวจพบข้อสอบที่ทำหน้าที่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 2 ข้อ ได้แก่ ข้อที่ 3 และ 9 สำหรับวิธีการถดถอยโลจิสติกตรวจพบข้อสอบที่ทำหน้าที่ต่างกันไม่ตรงกับวิธีเกณฑ์ ส่วนวิธีซิปเทสท์ตรวจไม่พบข้อสอบที่ทำหน้าที่ต่างกัน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี โดยทั้ง 3 วิธี พบข้อสอบที่ทำหน้าที่ไม่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 15 ข้อ ได้แก่ ข้อที่ 2, 4, 6, 7, 8, 11, 12, 13, 14, 15, 16, 17, 18, 19 และ 20

จากผลดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแบบสอบด้านการรู้เรื่องการอ่าน ฉบับที่ 4 จำนวน 20 ข้อ วิธีราสช์ทรีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ตรงกับวิธีเกณฑ์มากที่สุด จำนวน 2 ข้อ สำหรับวิธีการถดถอยโลจิสติกตรวจสอบการทำหน้าที่ต่างกันของข้อสอบไม่ตรงกับวิธีเกณฑ์ ส่วนวิธีซิปเทสท์ไม่พบการทำหน้าที่ต่างกันของข้อสอบ

ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีราสช์ทรี จำแนกตามตัวแปรเพศ ในแบบสอบฉบับที่ 5 จำนวน 13 ข้อ แสดงดังภาพ



ภาพที่ 21 ผลการวิเคราะห์ DIF ด้วยวิธีราสช์ทรี ในแบบสอบฉบับที่ 5

ตารางที่ 105 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการวัดพื้นที่ของราชู วิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ของแบบสอบฉบับที่ 5

ข้อ	Raju			LR					SIBTEST			Raschtree	
	ค่าสถิติ	p-value		(g)		(g by x)			Beta	p-value		ผลต่าง ค่า b	
				B	Sig.	B	Sig.						
1	0.3849	0.7003	-	-.649	.334	.082	.386	-	0.056	0.236	-	-0.0958	-
2	-2.4089	0.0160*	✓	.354	.526	.026	.773	-	-0.077	0.104	-	0.5453	✓
3	-0.1257	0.9000	-	-.185	.750	.022	.790	-	0.009	0.850	-	0.0216	-
4	1.7408	0.0817	-	-.386	.607	.023	.875	-	0.037	0.313	-	-0.4159	✓
5	-0.6464	0.5180	-	.752	.385	-.077	.484	-	-0.025	0.547	-	0.1333	-
6	2.0853	0.0370*	✓	-.403	.519	-.015	.871	-	0.103	0.031*	✓	-0.4564	✓
7	-3.2525	0.0011**	✓	.900	.111	-.064	.410	-	-0.064	0.182	-	0.7242	✓
8	2.8443	0.0045**	✓	.665	.351	-.220	.050*	✓	0.164	0.000**	✓	-0.6217	✓
9	-0.1482	0.8822	-	-.425	.471	.063	.461	-	-0.014	0.777	-	0.0272	-
10	-1.3581	0.1744	-	1.027	.122	-.105	.270	-	-0.067	0.165	-	0.2889	-
11	2.3431	0.0191*	✓	-.241	.695	-.055	.591	-	0.131	0.003**	✓	-0.5175	✓
12	-2.0786	0.0377*	✓	.578	.554	.005	.965	-	-0.096	0.020*	✓	0.4936	✓
13	0.6011	0.5477	-	.236	.722	-.051	.628	-	-0.020	0.670	-	-0.1266	-

**p < .01, *p < .05

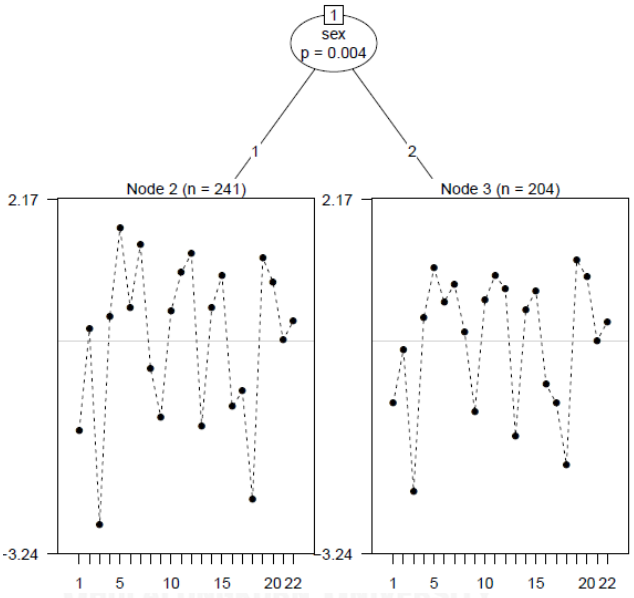
จากตารางที่ 105 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแบบสอบด้านการรู้เรื่องการอ่าน ฉบับที่ 5 จำนวน 13 ข้อ ด้วยวิธีการวัดพื้นที่ของราชูซึ่งเป็นวิธีเกณฑ์ พบข้อสอบที่ทำหน้าที่ต่างกัน 6 ข้อ ได้แก่ ข้อที่ 2, 6, 7, 8, 11 และ 12 ส่วนข้อสอบที่ทำหน้าที่ไม่ต่างกัน 7 ข้อ ได้แก่ ข้อที่ 1, 3, 4, 5, 9, 10 และ 13

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี โดยทั้ง 3 วิธี พบข้อสอบที่ทำหน้าที่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 1 ข้อ ได้แก่ ข้อที่ 8 สำหรับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีซิปเทสท์ และวิธีราสช์ทรี พบข้อสอบที่ทำหน้าที่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 3 ข้อ ได้แก่ ข้อที่ 6, 11 และ 12 ส่วนการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีราสช์ทรีเพียงวิธีเดียว พบข้อสอบที่ทำหน้าที่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 2 ข้อ ได้แก่ ข้อที่ 2 และ 7

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี โดยทั้ง 3 วิธี พบข้อสอบที่ทำหน้าที่ไม่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 6 ข้อ ได้แก่ ข้อที่ 1, 3, 5, 9, 10 และ 13

จากผลดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแบบสอบด้านการรู้เรื่องการอ่าน ฉบับที่ 5 จำนวน 13 ข้อ วิธีราสช์ทรีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ตรงกับวิธีเกณฑ์มากที่สุด จำนวน 6 ข้อ ส่วนวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ตรงกับวิธีเกณฑ์ จำนวน 1 ข้อ และ 4 ข้อ ตามลำดับ

ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีราสช์ทรี จำแนกตามตัวแปรเพศ ในแบบสอบฉบับที่ 6 จำนวน 22 ข้อ แสดงดังภาพ



ภาพที่ 22 ผลการวิเคราะห์ DIF ด้วยวิธีราสช์ทรี ในแบบสอบฉบับที่ 6 จำแนกตามตัวแปรเพศ

ตารางที่ 106 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการวัดพื้นที่ของราชู วิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ของแบบสอบฉบับที่ 6 จำแนกตามตัวแปรเพศ

ข้อ	Raju			LR					SIBTEST			Raschtree	
	ค่าสถิติ	p-value		(g)		(g by x)			Beta	p-value		ผลต่าง	
				B	Sig.	B	Sig.					ค่า b	
1	1.6915	0.0907	-	-1.224	.179	.088	.324		0.061	0.131	-	-0.4091	✓
2	-1.3885	0.1650	-	1.943	.020*	-.141	.050*	✓	-0.036	0.465	-	0.3066	-
3	1.3477	0.1778	-	-3.598	.007**	.361	.012*	✓	0.002	0.928	-	-0.5105	✓
4	-0.0752	0.9401	-	-.013	.987	.003	.969	-	0.015	0.759	-	0.0182	-
5	-2.5038	0.0123*	✓	1.976	.029*	-.122	.081	✓	-0.038	0.349	-	0.6043	✓
6	0.3888	0.6974	-	-.457	.546	.023	.708	-	0.067	0.183	-	-0.0877	-
7	-2.5838	0.0098**	✓	.676	.418	-.017	.802	-	-0.066	0.149	-	0.6013	✓
8	2.5753	0.0100**	✓	.012	.988	-.056	.427	-	0.157	0.001**	✓	-0.5633	✓
9	0.3732	0.7090	-	-.115	.897	.012	.891	-	0.021	0.600	-	-0.0806	-
10	0.7447	0.4564	-	-.899	.194	.046	.420		0.124	0.012*	✓	-0.1666	-
11	-0.2057	0.8370	-	.226	.789	-.021	.755	-	0.044	0.341	-	0.0370	-
12	-2.3681	0.0179*	✓	-.263	.776	.067	.356	-	-0.078	0.098	-	0.5419	✓
13	-0.5208	0.6025	-	-.535	.589	.097	.351	-	-0.053	0.155	-	0.1442	-
14	-0.1392	0.8893	-	.963	.226	-.084	.199	-	0.027	0.583	-	0.0305	-
15	-1.0527	0.2925	-	.188	.833	.005	.949	-	-0.044	0.348	-	0.2326	-
16	1.4971	0.1344	-	-.662	.498	.046	.633	-	0.047	0.258	-	-0.3406	-
17	-0.7550	0.4503	-	-.801	.331	.101	.201	-	-0.025	0.570	-	0.1855	-
18	1.6226	0.1047	-	-.658	.555	.017	.883	-	0.042	0.175	-	-0.5312	✓
19	-0.2180	0.8274	-	.236	.813	-.014	.851	-	0.005	0.908	-	0.0375	-
20	0.2815	0.7783	-	1.160	.164	-.108	.104	-	0.092	0.040*	✓	-0.0739	-
21	-0.0227	0.9819	-	.478	.588	-.033	.665	-	-0.011	0.806	-	0.0101	-
22	-0.0540	0.9569	-	.778	.342	-.067	.332	-	0.014	0.770	-	0.0137	-

**p < .01, *p < .05

จากตารางที่ 106 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแบบสอบด้านการรู้เรื่องการอ่าน ฉบับที่ 6 จำแนกตามตัวแปรเพศ จำนวน 22 ข้อ ด้วยวิธีการวัดพื้นที่ของราชูซึ่งเป็นวิธีเกณฑ์ พบข้อสอบที่ทำหน้าที่ต่างกัน 4 ข้อ ได้แก่ ข้อที่ 5, 7, 8 และ 12 ส่วนข้อสอบที่ทำหน้าที่ไม่ต่างกัน 18 ข้อ ได้แก่ ข้อที่ 1, 2, 3, 4, 6, 9, 10, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21 และ 22

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี พบว่า วิธีการถดถอยโลจิสติก และวิธีราสช์ทรีตรวจพบข้อสอบที่ทำหน้าที่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 1 ข้อ ได้แก่ ข้อที่ 5 สำหรับวิธีซิปเทสท์ และวิธีราสช์ทรีตรวจพบข้อสอบที่ทำ

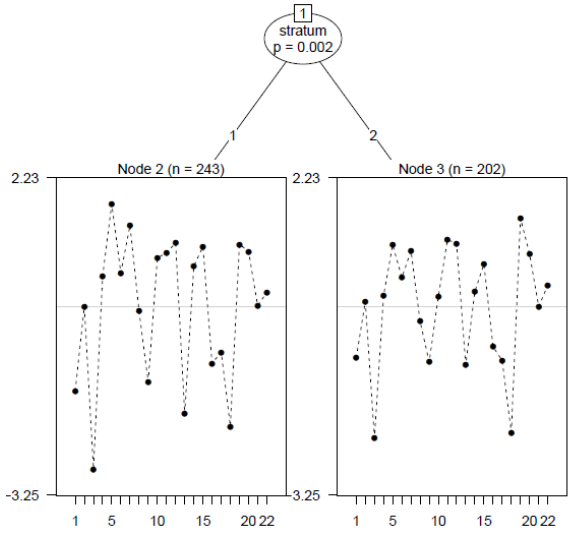
จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

หน้าที่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 1 ข้อ ได้แก่ ข้อที่ 8 ส่วนวิธีราสช์ทรีเพียงวิธีเดียวที่ตรวจพบข้อสอบที่ทำหน้าที่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 2 ข้อ ได้แก่ ข้อที่ 7 และ 12

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี โดยทั้ง 3 วิธี พบข้อสอบที่ทำหน้าที่ไม่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 12 ข้อ ได้แก่ ข้อที่ 4, 6, 9, 11, 13, 14, 15, 16, 17, 19, 21 และ 22

จากผลดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแบบสอบด้านการรู้เรื่องการอ่าน ฉบับที่ 6 จำแนกตามตัวแปรเพศ จำนวน 22 ข้อ วิธีราสช์ทรีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ตรงกับวิธีเกณฑ์มากที่สุด จำนวน 4 ข้อ สำหรับวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ตรงกับวิธีเกณฑ์ จำนวน 1 ข้อ เท่ากัน

ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีราสช์ทรี จำแนกตามตัวแปรสังกัดของโรงเรียน ในแบบสอบฉบับที่ 6 จำนวน 22 ข้อ แสดงดังภาพ



ภาพที่ 23 ผลการวิเคราะห์ DIF ด้วยวิธีราสช์ทรี ในแบบสอบฉบับที่ 6 จำแนกตามตัวแปรสังกัดของโรงเรียน

ตารางที่ 107 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการวัดพื้นที่ของราชู วิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ของแบบสอบฉบับที่ 6 จำแนกตามตัวแปรสังกัดของโรงเรียน

ข้อ	Raju			LR					SIBTEST			Raschtree	
	ค่าสถิติ	p-value		(g)		(g by x)			Beta	p-value		ผลต่าง	
				B	Sig.	B	Sig.					ค่า b	
1	2.1732	0.0298*	✓	.065	.942	-.068	.449	-	0.124	0.002**	✓	-0.5721	✓
2	0.4105	0.6814	-	-.328	.687	.020	.780	-	0.009	0.857	-	-0.0875	-
3	1.2730	0.2030	-	-1.644	.177	.134	.321	-	0.031	0.249	-	-0.5332	✓
4	-1.4908	0.1360	-	1.373	.102	-.089	.206	-	-0.052	0.299	-	0.3343	-
5	-2.7182	0.0066**	✓	1.374	.123	-.073	.303	-	-0.083	0.069	-	0.7029	✓
6	-0.3109	0.7559	-	-.878	.265	.073	.270	-	0.039	0.439	-	0.0829	-
7	-1.7493	0.0802	-	.210	.802	.003	.963	-	-0.017	0.729	-	0.4412	✓
8	-0.8222	0.4109	-	.548	.499	-.032	.652	-	-0.012	0.819	-	0.1760	-
9	1.2938	0.1957	-	1.678	.065	-.208	.022*	✓	0.061	0.156	-	-0.3458	-
10	-2.9687	0.0030**	✓	.087	.902	.028	.641	-	-0.043	0.401	-	0.6637	✓
11	1.0279	0.3040	-	-1.785	.053	.121	.104	-	0.089	0.058	-	-0.2182	-
12	0.1030	0.9179	-	.332	.719	-.033	.659	-	0.049	0.235	-	0.0014	-
13	3.0060	0.0026**	✓	-.171	.861	-.077	.450	-	0.124	0.002**	✓	-0.8417	✓
14	-1.9574	0.0503	-	.551	.497	-.013	.848	-	-0.080	0.116	-	0.4412	✓
15	-1.2281	0.2194	-	1.008	.267	-.059	.418	-	-0.052	0.315	-	0.3008	-
16	1.1276	0.2595	-	.829	.396	-.108	.260	-	0.039	0.368	-	-0.2844	-
17	-0.7119	0.4765	-	-.247	.759	.039	.608	-	-0.010	0.819	-	0.1458	-
18	-0.4787	0.6322	-	-.756	.473	.103	.365	-	-0.007	0.825	-	0.1057	-
19	1.9205	0.0548	-	.234	.819	-.060	.458	-	0.087	0.050*	✓	-0.4524	✓
20	-0.1001	0.9203	-	2.969	.001**	-.256	.000**	✓	0.009	0.844	-	0.0452	-
21	-0.0702	0.9440	-	-.087	.921	.015	.851	-	-0.018	0.705	-	0.0168	-
22	0.5998	0.5487	-	.614	.453	-.068	.327	-	0.080	0.108	-	-0.1225	-

**p < .01, *p < .05

จากตารางที่ 107 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแบบสอบด้านการรู้เรื่องการอ่าน ฉบับที่ 6 จำแนกตามตัวแปรสังกัดของโรงเรียน จำนวน 22 ข้อ ด้วยวิธีการวัดพื้นที่ของราชูซึ่งเป็นวิธีเกณฑ์ พบข้อสอบที่ทำหน้าที่ต่างกัน 4 ข้อ ได้แก่ ข้อที่ 1, 5, 10 และ 13 ส่วนข้อสอบที่ทำหน้าที่ไม่ต่างกัน 18 ข้อ ได้แก่ ข้อที่ 2, 3, 4, 6, 7, 8, 9, 11, 12, 14, 15, 16, 17, 18, 19, 20, 21 และ 22

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี พบว่า วิธีซิปเทสท์ และวิธีราสช์ทรีตรวจพบข้อสอบที่ทำหน้าที่ต่างกันตรงกับวิธีเกณฑ์

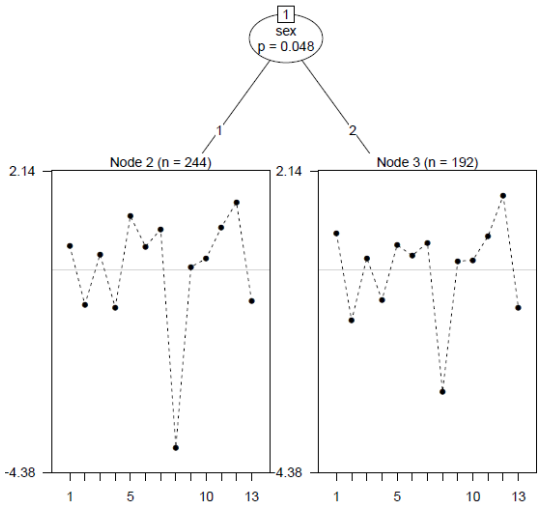
จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จำนวน 2 ข้อ ได้แก่ ข้อที่ 1 และ 13 สำหรับวิธีราสช์ทรีเพียงวิธีเดียวตรวจพบข้อสอบที่ทำหน้าที่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 2 ข้อ ได้แก่ ข้อที่ 5 และ 10 ส่วนวิธีการถดถอยโลจิสติกตรวจพบข้อสอบที่ทำหน้าที่ต่างกันไม่ตรงกับวิธีเกณฑ์

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี โดยทั้ง 3 วิธี พบข้อสอบที่ทำหน้าที่ไม่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 12 ข้อ ได้แก่ ข้อที่ 2, 4, 6, 8, 11, 12, 15, 16, 17, 18, 21 และ 22

จากผลดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแบบสอบด้านการรู้เรื่องการอ่าน ฉบับที่ 6 จำแนกตามตัวแปรสังกัดของโรงเรียน จำนวน 22 ข้อ วิธีราสช์ทรีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ตรงกับวิธีเกณฑ์มากที่สุด จำนวน 4 ข้อ สำหรับวิธีซิปเทสท์ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ตรงกับวิธีเกณฑ์ จำนวน 2 ข้อ ส่วนวิธีการถดถอยโลจิสติกตรวจสอบการทำหน้าที่ต่างกันของข้อสอบไม่ตรงกับวิธีเกณฑ์

ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีราสช์ทรี จำแนกตามตัวแปรเพศ ในแบบสอบฉบับที่ 7 จำนวน 13 ข้อ แสดงดังภาพ



ภาพที่ 24 ผลการวิเคราะห์ DIF ด้วยวิธีราสช์ทรี ในแบบสอบฉบับที่ 7

ตารางที่ 108 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการวัดพื้นที่ของราชู วิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ของแบบสอบฉบับที่ 7

ข้อ	Raju			LR					SIBTEST			Raschtree	
	ค่าสถิติ	p-value		(ɡ)		(ɡ by x)			Beta	p-value		ผลต่าง ค่า b	
				B	Sig.	B	Sig.						
1	1.1236	0.2612	-	-1.435	.081	.163	.178	-	0.088	0.050*	✓	-0.2634	-
2	-1.4523	0.1464		.338	.595	-.020	.859	-	0.002	0.973	-	0.3421	-
3	-0.3297	0.7416	-	-.299	.661	.045	.670	-	0.046	0.362	-	0.0759	-
4	0.7751	0.4383	-	-1.571	.043*	.253	.071	✓	0.052	0.263	-	-0.1616	-
5	-2.6391	0.0083**	✓	2.487	.001**	-.289	.009**	✓	-0.079	0.077	-	0.6249	✓
6	-0.8607	0.3894	-	.356	.634	-.034	.766	-	-0.022	0.645	-	0.1938	-
7	-1.2332	0.2175	-	.466	.510	-.043	.677	-	-0.028	0.552	-	0.2863	-
8	2.4463	0.0144*	✓	-2.849	.028*	.336	.170	✓	0.061	0.021*	✓	-1.2032	✓
9	0.5798	0.5620	-	-.337	.635	.019	.862	-	0.046	0.356	-	-0.1222	-
10	-0.1350	0.8926	-	.031	.959	-.025	.789	-	0.070	0.164	-	0.0336	-
11	-0.8218	0.4112	-	.925	.228	-.123	.268	-	-0.004	0.932	-	0.1915	-
12	0.5084	0.6112	-	1.567	.165	-.244	.097	-	0.051	0.148	✓	-0.1447	-
13	-0.6041	0.5458	-	-1.051	.210	.237	.125	-	-0.033	0.473	-	0.1469	-

**p < .01, *p < .05

จากตารางที่ 108 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแบบสอบด้านการรู้เรื่องการอ่าน ฉบับที่ 7 จำนวน 13 ข้อ ด้วยวิธีการวัดพื้นที่ของราชูซึ่งเป็นวิธีเกณฑ์ พบข้อสอบที่ทำหน้าที่ต่างกัน 2 ข้อ ได้แก่ ข้อที่ 5 และ 8 ส่วนข้อสอบที่ทำหน้าที่ไม่ต่างกัน 11 ข้อ ได้แก่ ข้อที่ 1, 2, 3, 4, 6, 7, 9, 10, 11, 12 และ 13

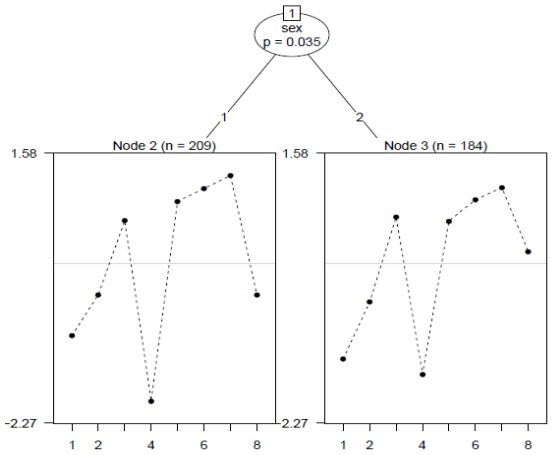
ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี โดยทั้ง 3 วิธี พบข้อสอบที่ทำหน้าที่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 1 ข้อ ได้แก่ ข้อที่ 8 สำหรับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก และวิธีราสช์ทรี พบข้อสอบที่ทำหน้าที่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 1 ข้อ ได้แก่ ข้อที่ 5

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี โดยทั้ง 3 วิธี พบข้อสอบที่ทำหน้าที่ไม่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 8 ข้อ ได้แก่ ข้อที่ 2, 3, 6, 7, 9, 10, 11 และ 13

จากผลดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแบบสอบด้านการรู้เรื่องการอ่าน ฉบับที่ 7 จำนวน 13 ข้อ วิธีการถดถอยโลจิสติกและวิธีราสช์ทรีตรวจสอบการทำ

หน้าที่ต่างกันของข้อสอบได้ตรงกับวิธีเกณฑ์มากที่สุด จำนวน 2 ข้อ ส่วนวิธีซิปเทสท์ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ตรงกับวิธีเกณฑ์ จำนวน 1 ข้อ

ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีราสช์ทรี จำแนกตามตัวแปรเพศ ในแบบสอบฉบับที่ 8 จำนวน 8 ข้อ แสดงดังภาพ



ภาพที่ 25 ผลการวิเคราะห์ DIF ด้วยวิธีราสช์ทรี ในแบบสอบฉบับที่ 8 จำแนกตามตัวแปรเพศ

ตารางที่ 109 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการวัดพื้นที่ของราชู วิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ของแบบสอบฉบับที่ 8 จำแนกตามตัวแปรเพศ

ข้อ	Raju			LR					SIBTEST			Raschtree	
	ค่าสถิติ	p-value		(g)		(g by x)			Beta	p-value		ผลต่าง ค่า b	
				B	Sig.	B	Sig.						
1	-1.0715	0.2839	-	-.032	.965	.129	.531	-	-0.082	0.054	-	0.3306	-
2	-0.3732	0.7090	-	1.693	.028*	-.405	.032*	✓	-0.015	0.769	-	0.1056	-
3	0.1313	0.8955	-	-.539	.602	.114	.596	-	-0.011	0.831	-	-0.0510	-
4	1.4348	0.1513	-	-.261	.760	-.047	.858	-	0.022	0.550	-	-0.3806	-
5	-1.3067	0.1913	-	-1.503	.055	.380	.021*	✓	-0.054	0.305	-	0.2822	-
6	-0.8091	0.4185	-	-.140	.862	.054	.738	-	-0.012	0.816	-	0.1594	-
7	-0.8708	0.3839	-	.807	.358	-.135	.435	-	0.002	0.960	-	0.1703	-
8	2.5931	0.0095**	✓	.005	.995	-.185	.335	-	0.124	0.015*	✓	-0.6165	✓

**p < .01, *p < .05

จากตารางที่ 109 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแบบสอบด้านการรู้เรื่องการอ่าน ฉบับที่ 8 จำแนกตามตัวแปรเพศ จำนวน 8 ข้อ ด้วยวิธีการวัดพื้นที่ของราชูซึ่งเป็นวิธี

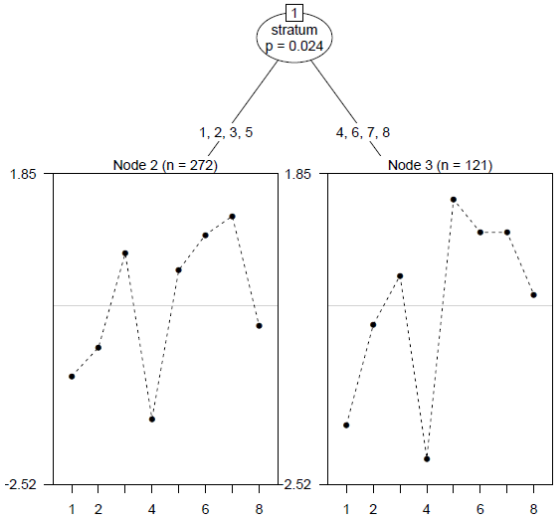
เกณฑ์ พบข้อสอบที่ทำหน้าที่ต่างกัน 1 ข้อ ได้แก่ ข้อที่ 8 ส่วนข้อสอบที่ทำหน้าที่ไม่ต่างกัน 7 ข้อ ได้แก่ ข้อที่ 1, 2, 3, 4, 5, 6 และ 7

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี พบว่า วิธีซิปเทสท์ และวิธีราสช์ทรีตรวจพบข้อสอบที่ทำหน้าที่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 1 ข้อ ได้แก่ ข้อที่ 8 ส่วนวิธีการถดถอยโลจิสติกตรวจพบข้อสอบที่ทำหน้าที่ต่างกันไม่ตรงกับวิธีเกณฑ์

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี โดยทั้ง 3 วิธี พบข้อสอบที่ทำหน้าที่ไม่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 5 ข้อ ได้แก่ ข้อที่ 1, 3, 4, 6 และ 7

จากผลดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแบบสอบด้านการรู้เรื่องการอ่าน ฉบับที่ 8 จำแนกตามตัวแปรเพศ จำนวน 8 ข้อ วิธีซิปเทสท์และวิธีราสช์ทรีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ตรงกับวิธีเกณฑ์มากที่สุด จำนวน 1 ข้อ สำหรับวิธีการถดถอยโลจิสติกตรวจสอบการทำหน้าที่ต่างกันของข้อสอบไม่ตรงกับวิธีเกณฑ์

ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีราสช์ทรี จำแนกตามตัวแปรสังกัดของโรงเรียน ในแบบสอบฉบับที่ 8 จำนวน 8 ข้อ แสดงดังภาพ



ภาพที่ 26 ผลการวิเคราะห์ DIF ด้วยวิธีราสช์ทรี ในแบบสอบฉบับที่ 8 จำแนกตามตัวแปรสังกัดของโรงเรียน

ตารางที่ 110 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการวัดพื้นที่ของราชู วิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ของแบบสอบฉบับที่ 8 จำแนกตามตัวแปรสังกัดของโรงเรียน

ข้อ	Raju			LR					SIBTEST			Raschtree	
	ค่าสถิติ	p-value		(g)		(g by x)			Beta	p-value		ผลต่างค่า b	
				B	Sig.	B	Sig.						
1	-2.1536	0.0313*	✓	.846	.309	-.006	.981	-	-0.099	0.025*	✓	0.6816	✓
2	1.2700	0.2041	-	-1.008	.237	.184	.385	-	0.040	0.469	-	-0.3194	-
3	-1.2221	0.2217	-	-.215	.849	.183	.456	-	-0.115	0.024*	✓	0.3119	-
4	-1.5114	0.1307	-	-1.262	.269	.723	.084	-	-0.091	0.008**	✓	0.5582	✓
5	3.5830	0.0003**	✓	-.325	.701	-.135	.431	-	0.211	0.000**	✓	-0.9947	✓
6	0.0695	0.9446	-	-.435	.619	.102	.564	-	0.003	0.959	-	-0.0332	-
7	-0.9381	0.3482	-	.050	.957	.062	.738	-	-0.076	0.167	-	0.2289	-
8	1.7474	0.0806	-	-.678	.450	.059	.779	-	0.105	0.062	-	-0.4333	✓

**p < .01, *p < .05

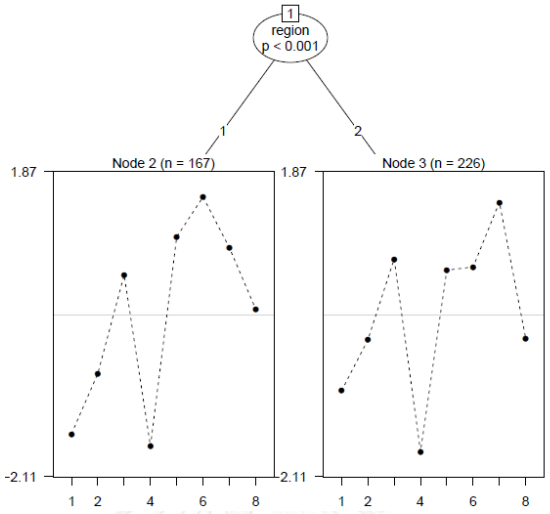
จากตารางที่ 110 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแบบสอบด้านการรู้เรื่องการอ่าน ฉบับที่ 8 จำแนกตามตัวแปรสังกัดของโรงเรียน จำนวน 8 ข้อ ด้วยวิธีการวัดพื้นที่ของราชูซึ่งเป็นวิธีเกณฑ์ พบข้อสอบที่ทำหน้าที่ต่างกัน 2 ข้อ ได้แก่ ข้อที่ 1 และ 5 ส่วนข้อสอบที่ทำหน้าที่ไม่ต่างกัน 6 ข้อ ได้แก่ ข้อที่ 2, 3, 4, 6, 7 และ 8

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี พบว่า วิธีซิปเทสท์ และวิธีราสช์ทรีตรวจพบข้อสอบที่ทำหน้าที่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 2 ข้อ ได้แก่ ข้อที่ 1 และ 5 ส่วนวิธีการถดถอยโลจิสติกตรวจพบข้อสอบที่ทำหน้าที่ต่างกันไม่ตรงกับวิธีเกณฑ์

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี โดยทั้ง 3 วิธี พบข้อสอบที่ทำหน้าที่ไม่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 3 ข้อ ได้แก่ ข้อที่ 2, 6 และ 7

จากผลดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแบบสอบด้านการรู้เรื่องการอ่าน ฉบับที่ 8 จำแนกตามตัวแปรสังกัดของโรงเรียน จำนวน 8 ข้อ วิธีซิปเทสท์และวิธีราสช์ทรีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ตรงกับวิธีเกณฑ์มากที่สุด จำนวน 2 ข้อ สำหรับวิธีการถดถอยโลจิสติกตรวจสอบไม่พบการทำหน้าที่ต่างกันของข้อสอบ

ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีราสช์ทรี จำแนกตามตัวแปรภูมิภาค ในแบบสอบฉบับที่ 8 จำนวน 8 ข้อ แสดงดังภาพ



ภาพที่ 27 ผลการวิเคราะห์ DIF ด้วยวิธีราสช์ทรี ในแบบสอบฉบับที่ 8 จำแนกตามตัวแปรภูมิภาค

ตารางที่ 111 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการวัดพื้นที่ของราชู วิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ของแบบสอบฉบับที่ 8 จำแนกตามตัวแปรภูมิภาค

ข้อ	Raju			LR					SIBTEST			Raschtree	
	ค่าสถิติ	p-value		(g)		(g by x)			Beta	p-value		ผลต่าง ค่า b	
				B	Sig.	B	Sig.						
1	1.8557	0.0635	-	-2.150	.005**	.444	.030*	✓	0.083	0.053	-	-0.5624	✓
2	1.7020	0.0888	-	-.077	.920	-.114	.549	-	0.132	0.006**	✓	-0.4396	✓
3	0.9676	0.3332	-	1.299	.240	-.329	.148	-	0.066	0.192	-	-0.2104	-
4	-0.3142	0.7533	-	.299	.723	-.052	.840	-	-0.035	0.364	-	0.0720	-
5	-1.7590	0.0786	-	1.525	.071	-.243	.151	-	-0.065	0.209	-	0.4357	✓
6	-3.6376	0.0003**	✓	.571	.522	.071	.686	-	-0.161	0.001**	✓	0.9151	✓
7	2.4403	0.0147*	✓	-.816	.352	.018	.915	-	0.157	0.001**	✓	-0.5930	✓
8	-1.5768	0.1148	-	1.290	.148	-.182	.368	-	-0.109	0.031*	✓	0.3827	-

**p < .01, *p < .05

จากตารางที่ 111 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแบบสอบด้านการรู้เรื่องการอ่าน ฉบับที่ 8 จำแนกตามตัวแปรภูมิภาค จำนวน 8 ข้อ ด้วยวิธีการวัดพื้นที่ของราชูซึ่งเป็นวิธีเกณฑ์

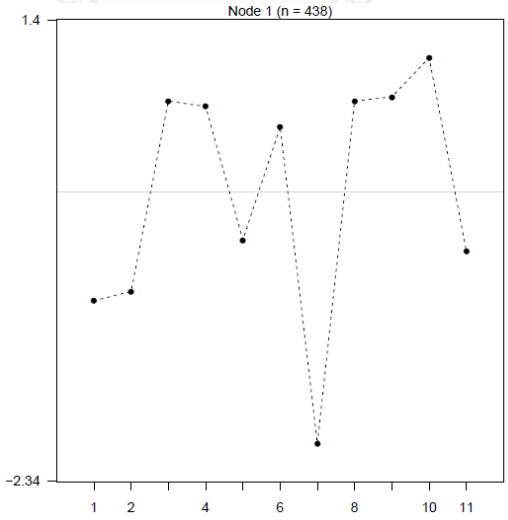
พบข้อสอบที่ทำหน้าที่ต่างกัน 2 ข้อ ได้แก่ ข้อที่ 6 และ 7 ส่วนข้อสอบที่ทำหน้าที่ไม่ต่างกัน 6 ข้อ ได้แก่ ข้อที่ 1, 2, 3, 4, 5 และ 8

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี พบว่า วิธีซิปเทสท์ และวิธีราสช์ทรีตรวจพบข้อสอบที่ทำหน้าที่ต่างกันตรงกับวิธีเกณฑ์จำนวน 2 ข้อ ได้แก่ ข้อที่ 6 และ 7 ส่วนวิธีการถดถอยโลจิสติกตรวจพบข้อสอบที่ทำหน้าที่ต่างกันไม่ตรงกับวิธีเกณฑ์

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี โดยทั้ง 3 วิธี พบข้อสอบที่ทำหน้าที่ไม่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 2 ข้อ ได้แก่ ข้อที่ 3 และ 4

จากผลดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแบบสอบด้านการรู้เรื่องการอ่าน ฉบับที่ 8 จำแนกตามตัวแปรภูมิภาค จำนวน 8 ข้อ วิธีซิปเทสท์และวิธีราสช์ทรีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ตรงกับวิธีเกณฑ์มากที่สุด จำนวน 2 ข้อ สำหรับวิธีการถดถอยโลจิสติกตรวจสอบการทำหน้าที่ต่างกันของข้อสอบไม่ตรงกับวิธีเกณฑ์

ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีราสช์ทรี จำแนกตามตัวแปรเพศ ในแบบสอบฉบับที่ 9 จำนวน 11 ข้อ แสดงดังภาพ



ภาพที่ 28 ผลการวิเคราะห์ DIF ด้วยวิธีราสช์ทรี ในแบบสอบฉบับที่ 9

ตารางที่ 112 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการวัดพื้นที่ของราชู วิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ของแบบสอบฉบับที่ 9

ข้อ	Raju			LR					SIBTEST			Raschtree	
	ค่าสถิติ	p-value		(g)		(g by x)			Beta	p-value		ผลต่างค่า b	
				B	Sig.	B	Sig.						
1	1.5966	0.1104	-	-1.239	.047*	.135	.226	✓	0.094	0.042*	✓	0.0000	-
2	0.9944	0.3200	-	.046	.953	-.049	.749	-	0.053	0.233	-	0.0000	-
3	-1.4079	0.1592	-	.356	.664	-.001	.995	-	-0.073	0.130	-	0.0000	-
4	-1.6819	0.0926	-	.971	.205	-.095	.417	-	-0.053	0.285	-	0.0000	-
5	0.0017	0.9986	-	.467	.609	-.055	.747	-	-0.056	0.223	-	0.0000	-
6	-1.4588	0.1446	-	.051	.944	.041	.721	-	-0.035	0.486	-	0.0000	-
7	2.0690	0.0385*	✓	-.987	.254	.084	.660	-	0.045	0.163	-	0.0000	-
8	-0.4109	0.6812	-	-.176	.810	.028	.801	-	0.036	0.461	-	0.0000	-
9	-0.6506	0.5153	-	1.530	.019*	-.252	.011*	✓	0.031	0.553	-	0.0000	-
10	-0.9984	0.3181	-	-.057	.939	.025	.817	-	0.010	0.828	-	0.0000	-
11	0.9128	0.3613	-	.454	.554	-.126	.373	-	0.005	0.914	-	0.0000	-

*p < .05

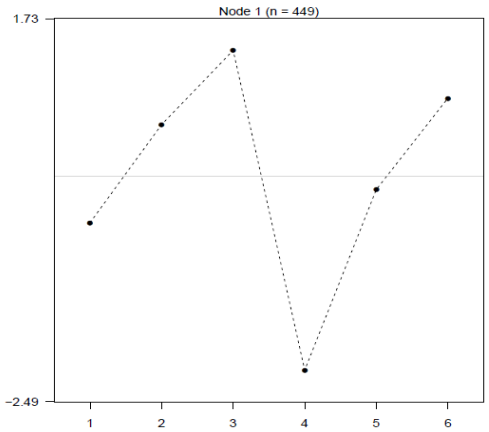
จากตารางที่ 112 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแบบสอบด้านการรู้เรื่องการอ่าน ฉบับที่ 9 จำนวน 11 ข้อ ด้วยวิธีการวัดพื้นที่ของราชูซึ่งเป็นวิธีเกณฑ์ พบข้อสอบที่ทำหน้าที่ต่างกัน 1 ข้อ ได้แก่ ข้อที่ 7 ส่วนข้อสอบที่ทำหน้าที่ไม่ต่างกัน 10 ข้อ ได้แก่ ข้อที่ 1, 2, 3, 4, 5, 6, 8, 9, 10, และ 11

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี พบว่า วิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ตรวจพบข้อสอบที่ทำหน้าที่ต่างกันไม่ตรงกับวิธีเกณฑ์ ส่วนวิธีราสช์ทรีตรวจไม่พบข้อสอบที่ทำหน้าที่ต่างกัน

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี โดยทั้ง 3 วิธี พบข้อสอบที่ทำหน้าที่ไม่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 8 ข้อ ได้แก่ ข้อที่ 2, 3, 4, 5, 6, 8, 10 และ 11

จากผลดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแบบสอบด้านการรู้เรื่องการอ่าน ฉบับที่ 9 จำนวน 11 ข้อ วิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบไม่ตรงกับวิธีเกณฑ์ ส่วนวิธีราสช์ทรีตรวจสอบไม่พบการทำหน้าที่ต่างกันของข้อสอบ

ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีราสช์ทรี จำแนกตามตัวแปรเพศ ในแบบสอบฉบับที่ 11 จำนวน 6 ข้อ แสดงดังภาพ



ภาพที่ 29 ผลการวิเคราะห์ DIF ด้วยวิธีราสช์ทรี ในแบบสอบฉบับที่ 11

ตารางที่ 113 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการวัดพื้นที่ของราชู วิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ของแบบสอบฉบับที่ 11

ข้อ	Raju			LR					SIBTEST			Raschtree	
	ค่าสถิติ	p-value		(g)		(g by x)			Beta	p-value		ผลต่าง ค่า b	
				B	Sig.	B	Sig.						
1	0.1072	0.9146	-	-.867	.295	.295	.275	-	-0.002	0.970	-	0.0000	-
2	1.8430	0.0653	-	.986	.325	-.440	.096	-	0.127	0.022*	✓	0.0000	-
3	-2.1942	0.0282*	✓	2.651	.006**	-.512	.025*	✓	-0.087	0.120	-	0.0000	-
4	0.8260	0.4088	-	-.214	.782	-.068	.808	-	0.066	0.061	-	0.0000	-
5	-0.1663	0.8679	-	-2.015	.028*	.661	.020*	✓	-0.007	0.898	-	0.0000	-
6	-0.8576	0.3911	-	.975	.275	-.204	.371	-	-0.012	0.828	-	0.0000	-

**p < .01, *p < .05

จากตารางที่ 113 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแบบสอบด้านการรู้เรื่องการอ่าน ฉบับที่ 11 จำนวน 6 ข้อ ด้วยวิธีการวัดพื้นที่ของราชูซึ่งเป็นวิธีเกณฑ์ พบข้อสอบที่ทำหน้าที่ต่างกัน 1 ข้อ ได้แก่ ข้อที่ 3 ส่วนข้อสอบที่ทำหน้าที่ไม่ต่างกัน 5 ข้อ ได้แก่ ข้อที่ 1, 2, 4, 5 และ 6

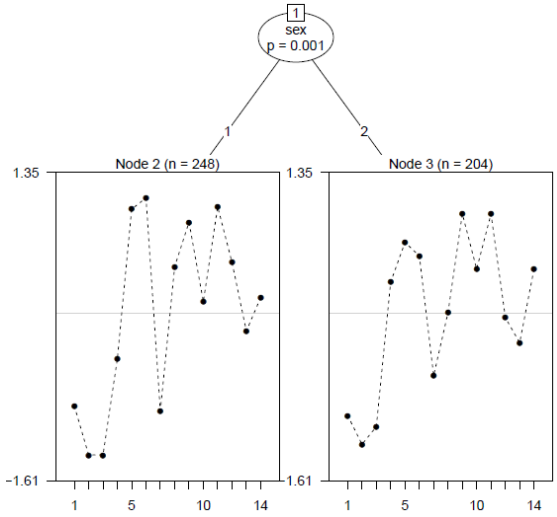
ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี พบว่า วิธีการถดถอยโลจิสติกเพียงวิธีเดียวตรวจพบข้อสอบที่ทำหน้าที่ต่างกันตรงกับ

วิธีเกณฑ์ จำนวน 1 ข้อ ได้แก่ ข้อที่ 3 สำหรับวิธีซิปเทสท์ตรวจพบข้อสอบที่ทำหน้าที่ต่างกันไม่ตรงกับวิธีเกณฑ์ ส่วนวิธีราสช์ทรีตรวจไม่พบข้อสอบที่ทำหน้าที่ต่างกัน

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี โดยทั้ง 3 วิธี พบข้อสอบที่ทำหน้าที่ไม่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 3 ข้อ ได้แก่ ข้อที่ 1, 4 และ 6

จากผลดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแบบสอบด้านการรู้เรื่องการอ่าน ฉบับที่ 11 จำนวน 6 ข้อ วิธีการถดถอยโลจิสติกตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ตรงกับวิธีเกณฑ์มากที่สุด จำนวน 1 ข้อ สำหรับวิธีซิปเทสท์ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบไม่ตรงกับวิธีเกณฑ์ ส่วนวิธีราสช์ทรีตรวจสอบไม่พบการทำหน้าที่ต่างกันของข้อสอบ

ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีราสช์ทรี จำแนกตามตัวแปรเพศ ในแบบสอบฉบับที่ 13 จำนวน 14 ข้อ แสดงดังภาพ



ภาพที่ 30 ผลการวิเคราะห์ DIF ด้วยวิธีราสช์ทรี ในแบบสอบฉบับที่ 13 จำแนกตามตัวแปรเพศ

ตารางที่ 114 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการวัดพื้นที่ของราชู วิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ของแบบสอบฉบับที่ 13 จำแนกตามตัวแปรเพศ

ข้อ	Raju			LR					SIBTEST			Raschtree	
	ค่าสถิติ	p-value		(g)		(g by x)			Beta	p-value		ผลต่าง ค่า b	
				B	Sig.	B	Sig.						
1	-0.3889	0.6974	-	1.089	.109	-.152	.160	-	-0.026	0.537	-	0.0922	-
2	0.3694	0.7118	-	-1.704	.010**	.287	.009**	✓	-0.004	0.919	-	-0.1023	-
3	1.0481	0.2946	-	-.686	.300	.080	.464	-	0.032	0.419	-	-0.2714	-
4	3.1522	0.0016**	✓	1.025	.152	-.268	.007**	✓	0.151	0.001**	✓	-0.7395	✓
5	-1.3455	0.1785	-	-.942	.217	.162	.072	-	-0.085	0.067	-	0.3184	-
6	-2.3317	0.0197*	✓	-.051	.941	.076	.349	-	-0.147	0.002**	✓	0.5563	✓
7	1.4188	0.1559	-	.171	.840	-.075	.590	-	0.067	0.096	-	-0.3391	-
8	-1.8938	0.0583	-	.153	.803	.039	.619	-	-0.053	0.278	-	0.4428	✓
9	0.2991	0.7649	-	.967	.150	-.133	.086	-	0.061	0.218	-	-0.0808	-
10	1.3402	0.1802	-	-.970	.153	.086	.316	-	0.090	0.064	-	-0.3130	-
11	-0.3087	0.7575	-	.931	.159	-.112	.140	-	0.028	0.549	-	0.0665	-
12	-2.2755	0.0229*	✓	2.226	.000**	-.239	.001**	✓	-0.016	0.756	-	0.5309	✓
13	-0.4681	0.6397	-	.364	.558	-.032	.708	-	-0.017	0.720	-	0.1124	-
14	1.1703	0.2419	-	-1.472	.025*	.157	.058	✓	0.068	0.164	-	-0.2732	-

**p < .01, *p < .05

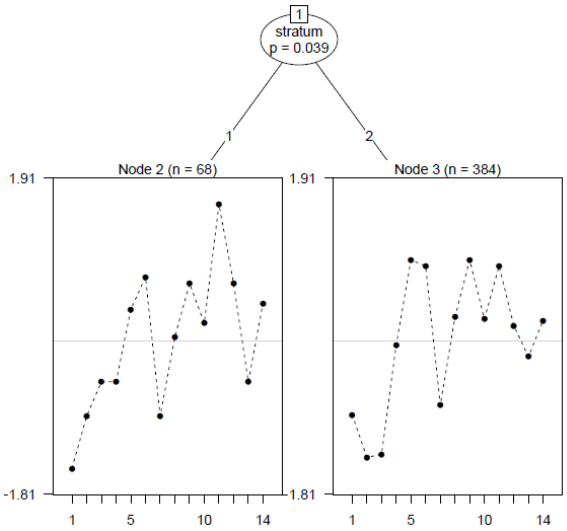
จากตารางที่ 114 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแบบสอบด้านการรู้เรื่องการอ่าน ฉบับที่ 13 จำแนกตามตัวแปรเพศ จำนวน 14 ข้อ ด้วยวิธีการวัดพื้นที่ของราชูซึ่งเป็นวิธีเกณฑ์ พบข้อสอบที่ทำหน้าที่ต่างกัน 3 ข้อ ได้แก่ ข้อที่ 4, 6 และ 12 ส่วนข้อสอบที่ทำหน้าที่ไม่ต่างกัน 11 ข้อ ได้แก่ ข้อที่ 1, 2, 3, 5, 7, 8, 9, 10, 11, 13 และ 14

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี โดยทั้ง 3 วิธี พบข้อสอบที่ทำหน้าที่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 1 ข้อ ได้แก่ ข้อที่ 4 สำหรับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติกและวิธีราสช์ทรีพบข้อสอบที่ทำหน้าที่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 1 ข้อ ได้แก่ ข้อที่ 12 ส่วนการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีซิปเทสท์และวิธีราสช์ทรีพบข้อสอบที่ทำหน้าที่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 1 ข้อ ได้แก่ ข้อที่ 6

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี โดยทั้ง 3 วิธี พบข้อสอบที่ทำหน้าที่ไม่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 8 ข้อ ได้แก่ ข้อที่ 1, 3, 5, 7, 9, 10, 11 และ 13

จากผลดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแบบสอบด้านการรู้เรื่องการอ่าน ฉบับที่ 13 จำแนกตามตัวแปรเพศ จำนวน 14 ข้อ วิธีราสช์ทรีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ตรงกับวิธีเกณฑ์มากที่สุด จำนวน 3 ข้อ ส่วนวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ตรงกับวิธีเกณฑ์ จำนวน 2 ข้อ เท่ากัน

ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีราสช์ทรี จำแนกตามตัวแปรสังกัดของโรงเรียน ในแบบสอบฉบับที่ 13 จำนวน 14 ข้อ แสดงดังภาพ



ภาพที่ 31 ผลการวิเคราะห์ DIF ด้วยวิธีราสช์ทรี ในแบบสอบฉบับที่ 13 จำแนกตามตัวแปรสังกัดของโรงเรียน

ตารางที่ 115 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการวัดพื้นที่ของราชู วิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ของแบบสอบฉบับที่ 13 จำแนกตามตัวแปรสังกัดของโรงเรียน

ข้อ	Raju			LR					SIBTEST			Raschtree	
	ค่าสถิติ	p-value		(g)		(g by x)			Beta	p-value		ผลต่าง ค่า b	
				B	Sig.	B	Sig.						
1	1.3519	0.1764	-	-1.023	.279	.067	.637	-	0.020	0.780	-	-0.6297	✓
2	-1.4683	0.1420	-	.330	.709	.034	.776	-	-0.007	0.907	-	0.4833	✓
3	-2.5360	0.0112*	✓	1.807	.110	-.091	.545	-	-0.194	0.018*	✓	0.8541	✓
4	1.2149	0.2244	-	-1.179	.204	.106	.359	-	0.079	0.349	-	-0.4345	✓
5	1.8474	0.0647	-	-.566	.592	-.006	.958	-	0.131	0.105	-	-0.5838	✓
6	0.5060	0.6129	-	-1.155	.207	.107	.275	-	0.092	0.251	-	-0.1293	-
7	0.2271	0.8203	-	2.275	.227	-.335	.246	-	-0.014	0.867	-	-0.1281	-
8	0.7522	0.4519	-	1.468	.196	-.215	.101	-	0.001	0.989	-	-0.2352	-
9	0.9472	0.3435	-	-.143	.887	-.026	.810	-	0.108	0.186	-	-0.2761	-
10	0.2080	0.8352	-	-.718	.451	.086	.424	-	0.069	0.511	-	-0.0471	-
11	-1.9908	0.0465*	✓	.582	.615	.001	.995	-	-0.104	0.216	-	0.7256	✓
12	-1.4158	0.1568	-	1.811	.098	-.170	.146	-	-0.052	0.542	-	0.4950	✓
13	0.8129	0.4163	-	-.414	.650	.017	.881	-	0.016	0.809	-	-0.2960	-
14	-0.5282	0.5974	-	1.178	.309	-.111	.382	-	-0.053	0.483	-	0.2017	-

*p < .05

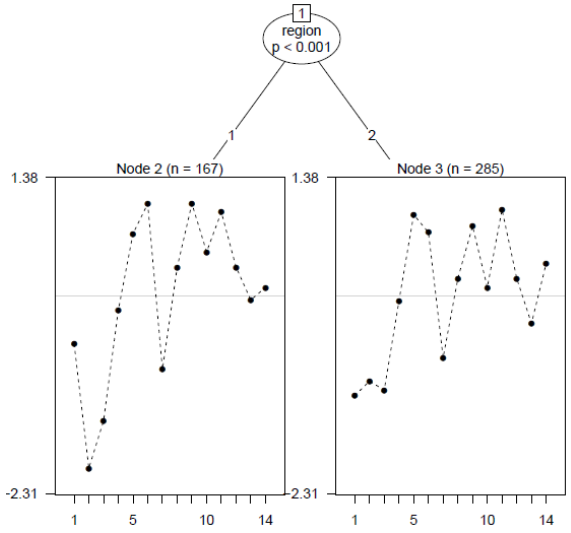
จากตารางที่ 115 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแบบสอบด้านการรู้เรื่องการอ่าน ฉบับที่ 13 จำแนกตามตัวแปรสังกัดของโรงเรียน จำนวน 14 ข้อ ด้วยวิธีการวัดพื้นที่ของราชูซึ่งเป็นวิธีเกณฑ์ พบข้อสอบที่ทำหน้าที่ต่างกัน 2 ข้อ ได้แก่ ข้อที่ 3 และ 11 ส่วนข้อสอบที่ทำหน้าที่ไม่ต่างกัน 12 ข้อ ได้แก่ ข้อที่ 1, 2, 4, 5, 6, 7, 8, 9, 10, 12, 13 และ 14

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี พบว่า วิธีซิปเทสท์และวิธีราสช์ทรีตรวจพบข้อสอบที่ทำหน้าที่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 1 ข้อ ได้แก่ ข้อที่ 3 สำหรับวิธีราสช์ทรีเพียงวิธีเดียวตรวจพบข้อสอบที่ทำหน้าที่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 1 ข้อ ได้แก่ ข้อที่ 11 ส่วนวิธีการถดถอยโลจิสติกตรวจไม่พบข้อสอบที่ทำหน้าที่ต่างกัน

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี โดยทั้ง 3 วิธี พบข้อสอบที่ทำหน้าที่ไม่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 7 ข้อ ได้แก่ ข้อที่ 6, 7, 8, 9, 10, 13 และ 14

จากผลดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแบบสอบด้านการรู้เรื่องการอ่าน ฉบับที่ 13 จำแนกตามตัวแปรสังกัดของโรงเรียน จำนวน 14 ข้อ วิธีราสช์ทรีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ตรงกับวิธีเกณฑ์มากที่สุด จำนวน 2 ข้อ สำหรับวิธีซิปเทสท์ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ตรงกับวิธีเกณฑ์ จำนวน 1 ข้อ ส่วนวิธีการถดถอยโลจิสติกตรวจสอบไม่พบการทำหน้าที่ต่างกันของข้อสอบ

ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีราสช์ทรี จำแนกตามตัวแปรภูมิภาคในแบบสอบฉบับที่ 13 จำนวน 14 ข้อ แสดงดังภาพ



ภาพที่ 32 ผลการวิเคราะห์ DIF ด้วยวิธีราสช์ทรี ในแบบสอบฉบับที่ 13 จำแนกตามตัวแปรภูมิภาค

ตารางที่ 116 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการวัดพื้นที่ของราชู วิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ของแบบสอบฉบับที่ 13 จำแนกตามตัวแปรภูมิภาค

ข้อ	Raju			LR					SIBTEST			Raschtree	
	ค่าสถิติ	p-value		(g)		(g by x)			Beta	p-value		ผลต่าง ค่า b	
				B	Sig.	B	Sig.						
1	-2.4007	0.0164*	✓	.819	.247	-.017	.880	-	-0.108	0.013*	✓	0.5987	✓
2	3.4848	0.0005**	✓	-1.011	.173	-.018	.882	-	0.136	0.000**	✓	-1.0135	✓
3	1.3897	0.1646	-	.146	.844	-.103	.414	-	0.053	0.196	-	-0.3597	-
4	0.4302	0.6670	-	.540	.460	-.100	.320	-	0.040	0.381	-	-0.1086	-
5	0.8488	0.3960	-	-.271	.721	-.001	.987	-	0.013	0.768	-	-0.2215	-
6	-1.3840	0.1664	-	1.367	.078	-.127	.146	-	-0.057	0.248	-	0.3350	-
7	0.5528	0.5804	-	-.132	.880	-.009	.948	-	0.024	0.563	-	-0.1338	-
8	-0.5849	0.5586	-	-.078	.902	.023	.767	-	-0.033	0.503	-	0.1317	-
9	-1.0821	0.2792	-	.317	.655	-.012	.880	-	-0.074	0.133	-	0.2605	-
10	-1.7568	0.0790	-	-1.137	.096	.209	.015*	✓	-0.118	0.018*	✓	0.4146	✓
11	0.0658	0.9476	-	.638	.367	-.086	.287	-	0.013	0.804	-	-0.0284	-
12	-0.5828	0.5600	-	-.098	.866	.023	.746	-	-0.051	0.314	-	0.1317	-
13	-1.1610	0.2457	-	.705	.289	-.062	.491	-	-0.082	0.082	-	0.2729	-
14	1.1271	0.2597	-	-.665	.306	.045	.585	-	0.069	0.173	-	-0.2794	-

**p < .01, *p < .05

จากตารางที่ 116 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแบบสอบด้านการรู้เรื่องการอ่าน ฉบับที่ 13 จำแนกตามตัวแปรภูมิภาค จำนวน 14 ข้อ ด้วยวิธีการวัดพื้นที่ของราชูซึ่งเป็นวิธีเกณฑ์ พบข้อสอบที่ทำหน้าที่ต่างกัน 2 ข้อ ได้แก่ ข้อที่ 1 และ 2 ส่วนข้อสอบที่ทำหน้าที่ไม่ต่างกัน 12 ข้อ ได้แก่ ข้อที่ 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 และ 14

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี พบว่า วิธีซิปเทสท์และวิธีราสช์ทรีตรวจพบข้อสอบที่ทำหน้าที่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 2 ข้อ ได้แก่ ข้อที่ 1 และ 2 ส่วนวิธีการถดถอยโลจิสติกตรวจพบข้อสอบที่ทำหน้าที่ต่างกันไม่ตรงกับวิธีเกณฑ์

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี โดยทั้ง 3 วิธี พบข้อสอบที่ทำหน้าที่ไม่ต่างกันตรงกับวิธีเกณฑ์ จำนวน 11 ข้อ ได้แก่ ข้อที่ 3, 4, 5, 6, 7, 8, 9, 11, 12, 13 และ 14

จากผลดังกล่าว แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแบบสอบด้านการรู้เรื่องการอ่าน ฉบับที่ 13 จำแนกตามตัวแปรภูมิภาค จำนวน 14 ข้อ วิธีซิปเทสท์และวิธีราสช์ทรี

ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ตรงกับวิธีเกณฑ์มากที่สุด จำนวน 2 ข้อ ส่วนวิธีการถดถอยโลจิสติกตรวจสอบการทำหน้าที่ต่างกันของข้อสอบไม่ตรงกับวิธีเกณฑ์

จากการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี สรุปผลได้ดังตารางที่ 117

ตารางที่ 117 สรุปผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

ฉบับ	จำนวน	จำนวนข้อสอบ DIF (ข้อ)				จำนวนข้อสอบ No DIF (ข้อ)			
	(ข้อ)	Raju	LR	SIBTEST	Raschtree	Raju	LR	SIBTEST	Raschtree
1	15	4	4	3	5	11	11	12	10
2	17	1	4	1	0	16	13	16	17
4	20	2	2	0	3	18	18	20	17
5	13	6	1	4	7	7	12	9	6
6(1)	22	4	3	3	7	18	19	19	15
6(2)	22	4	2	3	8	18	20	19	14
7	13	2	3	2	2	11	10	11	11
8 (1)	8	1	2	1	1	7	6	7	7
8(2)	8	2	0	4	4	6	8	4	4
8(3)	8	2	1	4	5	6	7	4	3
9	11	1	1	1	0	10	10	10	11
11	6	1	2	1	0	5	4	5	6
13(1)	14	3	4	2	4	11	10	12	10
13(2)	14	2	0	1	7	12	14	13	7
13(3)	14	2	1	3	3	12	13	11	11
รวม	205	37	30	33	56	168	175	172	149

หมายเหตุ (1) เพศ (2) สังกัดของโรงเรียน (3) ภูมิภาค

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตารางที่ 117 ผลการตรวจสอบด้วยวิธีการวัดพื้นที่ของราชูซึ่งเป็นวิธีเกณฑ์ พบข้อสอบที่ทำหน้าที่ต่างกัน จำนวน 37 ข้อ เมื่อเปรียบเทียบผลการตรวจสอบระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี พบว่า วิธีราสช์ทรีตรวจพบข้อสอบที่ทำหน้าที่ต่างกันในปริมาณมากที่สุด จำนวน 56 ข้อ และตรงกับผลการตรวจด้วยวิธีเกณฑ์ จำนวน 34 ข้อ ส่วนวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ตรวจพบข้อสอบที่ทำหน้าที่ต่างกันในปริมาณใกล้เคียงกัน จำนวน 30 ข้อ และ 33 ข้อ

ตามลำดับ และตรงกับผลการตรวจด้วยวิธีเกณฑ์ จำนวน 10 และ 20 ข้อ ตามลำดับ ผลสรุปของจำนวนการตรวจพบการทำหน้าที่ต่างกันของข้อสอบ ดังตารางที่ 118

ตารางที่ 118 จำนวนข้อของการเกิดและไม่เกิดการทำหน้าที่ต่างกันของข้อสอบ

ฉบับ	การทำหน้าที่ต่างกันของข้อสอบ		LR		SIB		Rasch	
			DIF	No DIF	DIF	No DIF	DIF	No DIF
1	ข้อสอบทำหน้าที่ต่างกัน	DIF	2	2	2	2	4	0
	ข้อสอบทำหน้าที่ไม่ต่างกัน	No DIF	2	9	1	10	1	10
2	ข้อสอบทำหน้าที่ต่างกัน	DIF	1	0	0	1	0	1
	ข้อสอบทำหน้าที่ไม่ต่างกัน	No DIF	3	13	1	15	0	16
4	ข้อสอบทำหน้าที่ต่างกัน	DIF	0	2	0	2	2	0
	ข้อสอบทำหน้าที่ไม่ต่างกัน	No DIF	2	16	0	18	1	17
5	ข้อสอบทำหน้าที่ต่างกัน	DIF	1	5	4	2	6	0
	ข้อสอบทำหน้าที่ไม่ต่างกัน	No DIF	0	7	0	7	1	6
6(1)	ข้อสอบทำหน้าที่ต่างกัน	DIF	1	3	1	3	4	0
	ข้อสอบทำหน้าที่ไม่ต่างกัน	No DIF	2	16	2	16	3	15
6(2)	ข้อสอบทำหน้าที่ต่างกัน	DIF	0	4	2	2	4	0
	ข้อสอบทำหน้าที่ไม่ต่างกัน	No DIF	2	16	1	17	4	14
7	ข้อสอบทำหน้าที่ต่างกัน	DIF	2	0	1	1	2	0
	ข้อสอบทำหน้าที่ไม่ต่างกัน	No DIF	1	10	1	10	0	11
8(1)	ข้อสอบทำหน้าที่ต่างกัน	DIF	0	1	1	0	1	0
	ข้อสอบทำหน้าที่ไม่ต่างกัน	No DIF	2	5	0	7	0	7
8(2)	ข้อสอบทำหน้าที่ต่างกัน	DIF	0	2	2	0	2	0
	ข้อสอบทำหน้าที่ไม่ต่างกัน	No DIF	0	6	2	4	2	4
8(3)	ข้อสอบทำหน้าที่ต่างกัน	DIF	0	2	2	0	2	0
	ข้อสอบทำหน้าที่ไม่ต่างกัน	No DIF	1	5	2	4	3	3

ตารางที่ 118 (ต่อ)

ฉบับ	การทำหน้าที่ต่างกันของข้อสอบ		LR		SIB		Rasch	
			DIF	No DIF	DIF	No DIF	DIF	No DIF
9	ข้อสอบทำหน้าที่ต่างกัน	DIF	0	1	0	1	0	1
	ข้อสอบทำหน้าที่ไม่ต่างกัน	No DIF	1	9	1	9	0	10
11	ข้อสอบทำหน้าที่ต่างกัน	DIF	1	0	0	1	0	1
	ข้อสอบทำหน้าที่ไม่ต่างกัน	No DIF	1	4	1	4	0	5
13(1)	ข้อสอบทำหน้าที่ต่างกัน	DIF	2	1	2	1	3	0
	ข้อสอบทำหน้าที่ไม่ต่างกัน	No DIF	2	9	0	11	1	10
13(2)	ข้อสอบทำหน้าที่ต่างกัน	DIF	0	2	1	1	2	0
	ข้อสอบทำหน้าที่ไม่ต่างกัน	No DIF	0	12	0	12	5	7
13(3)	ข้อสอบทำหน้าที่ต่างกัน	DIF	0	2	2	0	2	0
	ข้อสอบทำหน้าที่ไม่ต่างกัน	No DIF	1	11	1	11	1	11
รวม	ข้อสอบทำหน้าที่ต่างกัน	DIF	10	27	20	17	34	3
รวม	ข้อสอบทำหน้าที่ไม่ต่างกัน	No DIF	20	148	13	155	22	146

หมายเหตุ (1) เพศ (2) สังกัดของโรงเรียน (3) ภูมิภาค

จากตารางที่ 118 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีการถดถอยโลจิสติกกับวิธีการวัดพื้นที่ของราชูซึ่งเป็นวิธีเกณฑ์ พบข้อร่วมของการทำหน้าที่ต่างกันของข้อสอบระหว่างสองวิธี จำนวน 10 ข้อ จากข้อสอบ 205 ข้อ คิดเป็นร้อยละ 4.88

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีซิปเทสท์กับวิธีการวัดพื้นที่ของราชูซึ่งเป็นวิธีเกณฑ์ พบข้อร่วมของการทำหน้าที่ต่างกันของข้อสอบระหว่างสองวิธี จำนวน 20 ข้อ จากข้อสอบ 205 ข้อ คิดเป็นร้อยละ 9.76

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีราสช์ทรีโดยการทดสอบระดับนัยสำคัญกับวิธีการวัดพื้นที่ของราชูซึ่งเป็นวิธีเกณฑ์ พบข้อร่วมของการทำหน้าที่ต่างกันของข้อสอบระหว่างสองวิธี จำนวน 34 ข้อ จากข้อสอบ 205 ข้อ คิดเป็นร้อยละ 16.59

3.3 ประสิทธิภาพด้านอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1

จากผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและผลสรุปการเกิดและไม่เกิดการทำหน้าที่ต่างกันของข้อสอบ นำมาพิจารณาประสิทธิภาพด้านอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับแบบสอบด้านการรู้เรื่องการอ่าน ดังตารางที่ 119

ตารางที่ 119 ร้อยละของอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 จำแนกตามวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

ร้อยละของอำนาจการทดสอบ			ร้อยละของอัตราความคลาดเคลื่อนประเภทที่ 1		
LR	SIBTEST	Raschtree	LR	SIBTEST	Raschtree
27.03	54.05	91.89	11.90	7.74	13.10

จากตารางที่ 119 ร้อยละของอำนาจการทดสอบของวิธีการถดถอยโลจิสติก ประมาณร้อยละ 27.03 วิธีซิปเทสท์ ประมาณร้อยละ 54.05 และวิธีราสช์ทรี ประมาณร้อยละ 91.89

ร้อยละของอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก ประมาณร้อยละ 11.90 วิธีซิปเทสท์ ประมาณร้อยละ 7.74 และวิธีราสช์ทรี ประมาณร้อยละ 13.10

เกณฑ์การพิจารณาประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบดูจากอำนาจการทดสอบในการตรวจสอบที่สูง และมีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบที่ต่ำ แสดงถึงประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ดี จากผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการตรวจสอบทั้ง 3 วิธี พบว่า วิธีราสช์ทรีให้ค่าร้อยละของอำนาจการทดสอบสูงสุด และมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ซึ่งเกินเกณฑ์ที่กำหนดที่ระดับร้อยละ 10 จึงยังไม่อาจกล่าวได้ว่า วิธีราสช์ทรีมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ส่วนวิธีซิปเทสท์มีอำนาจการทดสอบประมาณร้อยละ 54.05 และมีอัตราความคลาดเคลื่อนประเภทที่ 1 ร้อยละ 7.74 ซึ่งอยู่ในเกณฑ์ที่กำหนดที่ระดับร้อยละ 10 โดยร้อยละของอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 สอดคล้องกับการตรวจสอบในข้อมูลจำลอง สำหรับวิธีการถดถอยโลจิสติกให้ค่าร้อยละของอำนาจการทดสอบต่ำสุด ในขณะที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าเกณฑ์ที่กำหนด ดังนั้น จึงกล่าวได้ว่า วิธีซิปเทสท์มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบดีกว่าวิธีการถดถอยโลจิสติกและวิธีราสช์ทรี

บทที่ 5
สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

การศึกษาครั้งนี้ มีวัตถุประสงค์เพื่อเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี โดยการทดสอบระดับนัยสำคัญและการวัดขนาดอิทธิพล ในข้อมูลจำลอง ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ ความยากของข้อสอบ ความยาวของแบบสอบ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน และขนาดกลุ่มตัวอย่าง ในแต่ปัจจัยแบ่งออกเป็นลักษณะย่อย ๆ ดังนี้ ความยากของข้อสอบ 3 ระดับ คือ ค่าความยากต่ำ (b_L) ค่าความยากปานกลาง (b_M) และค่าความยากสูง (b_H) ความยาวของแบบสอบ 2 ขนาด คือ 40 ข้อ และ 60 ข้อ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด คือ ร้อยละ 10 และร้อยละ 20 และขนาดกลุ่มตัวอย่าง 3 ขนาด คือ 250 คน 500 คน และ 1000 คน และข้อมูลเชิงประจักษ์ศึกษาการทำหน้าที่ต่างกันของข้อสอบตามตัวแปร 3 ตัว คือ เพศ สังกัดของโรงเรียน และภูมิภาค

ข้อมูลที่ใช้ ในการศึกษาครั้งนี้เป็นข้อมูลจำลองภายใต้ทฤษฎีการตอบสนองข้อสอบ โมเดล 1 พารามิเตอร์ กำหนดการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป โดยจัดกระทำข้อมูลตามปัจจัยที่แปรเปลี่ยน 4 ปัจจัย จำนวน 36 เงื่อนไข (3 ระดับ x 2 ขนาดx 2 ขนาด x 3 ขนาด) และในแต่ละเงื่อนไขจำลองซ้ำ 50 รอบ ผู้วิจัยตรวจสอบค่าพารามิเตอร์ของข้อมูลจำลองให้เป็นไปตามเงื่อนไขของปัจจัยที่ศึกษา จากนั้นจึงนำข้อมูลดังกล่าวมาตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 3 วิธี คือ วิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี สำหรับข้อมูลเชิงประจักษ์จากโครงการประเมินผลนักเรียนนานาชาติ หรือ PISA 2009 ด้านการรู้เรื่องการอ่าน (reading literacy) ศึกษาการทำหน้าที่ต่างกันแบบเอกรูปและอเนกรูป คัดเลือกมาเฉพาะข้อสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค ตรวจสอบความสมบูรณ์ครบถ้วนของการตอบข้อสอบ วิเคราะห์ค่าสถิติพื้นฐาน คุณภาพรายข้อและคุณภาพทั้งฉบับ จากนั้นจึงวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีราสช์ทรี เพื่อระบุกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ แล้ววิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวัดพื้นที่ของราชู เพื่อระบุข้อสอบที่ทำหน้าที่ต่างกัน

ผู้วิจัยนำข้อมูลจำลอง 36 เงื่อนไข และข้อมูลเชิงประจักษ์ ไปวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ 3 วิธี คือ วิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ส่วนการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพล วิเคราะห์ 2 วิธี คือ วิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ การคำนวณอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภท

ที่ 1 ของวิธีการตรวจสอบทั้ง 3 วิธี ในข้อมูลจำลองเปรียบเทียบกับข้อสอบที่กำหนดให้ทำหน้าที่ต่างกัน ส่วนการเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบทั้ง 3 วิธี ใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง (two way ANOVA) เพื่อทดสอบผลของวิธีการตรวจสอบ ผลของปัจจัยที่ศึกษา และผลของปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบและปัจจัยที่ศึกษา ถ้าผลการทดสอบดังกล่าวมีนัยสำคัญทางสถิติจะใช้การวิเคราะห์อิทธิพลอย่างง่าย (simple effect) ทดสอบในแต่ละระดับของตัวแปร แล้วเปรียบเทียบความแตกต่างเป็นรายคู่โดยใช้วิธีการทดสอบของ Bonferroni กรณีค่าความแปรปรวนของทุกกลุ่มเท่ากัน และวิธีการทดสอบของ Games-Howell กรณีค่าความแปรปรวนของทุกกลุ่มไม่เท่ากัน โดยทดสอบใน 3 กรณีต่อไปนี้ 1) ทดสอบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธีการตรวจสอบภายใต้ปัจจัยที่ศึกษาต่างระดับกัน 2) ทดสอบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 3 วิธี ภายใต้ปัจจัยที่ศึกษาระดับเดียวกัน และ 3) ทดสอบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 3 วิธี ภายใต้ปัจจัยที่ศึกษาต่างระดับกัน สำหรับข้อมูลเชิงประจักษ์ ผู้วิจัยใช้วิธีการวัดพื้นที่ของราชูเป็นเกณฑ์ในการเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบทั้ง 3 วิธี

สรุปผลการวิจัย

ผลการวิจัยสรุปตามวัตถุประสงค์การวิจัย ได้ดังนี้

1. การเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยที่ศึกษา ปรากฏผลดังนี้

1.1 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญภายใต้ปัจจัยความยากของข้อสอบต่างกัน 3 ระดับ พบว่า อำนาจการทดสอบของวิธีการถดถอยโลจิสติกภายใต้ความยากของข้อสอบ b_M มีค่าสูงสุด ประมาณร้อยละ 18.03 รองลงมาได้แก่ภายใต้ความยากของข้อสอบ b_H และ b_L ตามลำดับ มีค่าประมาณร้อยละ 16.72 และ 13.06 ตามลำดับ สำหรับวิธีซิปเทสท์ พบว่า อำนาจการทดสอบภายใต้ความยากของข้อสอบ bM มีค่าสูงสุด ประมาณร้อยละ 54.32 รองลงมาได้แก่ภายใต้ความยากของข้อสอบ b_L และ b_H ตามลำดับ มีค่าประมาณร้อยละ 41.22 และ 26.47 ตามลำดับ ส่วนวิธีราสช์ทรี พบว่า อำนาจการทดสอบภายใต้ความยากของข้อสอบ b_M มีค่าสูงสุด ประมาณร้อยละ 24.83 รองลงมาได้แก่ภายใต้ความยากของข้อสอบ b_H และ b_L ตามลำดับ มีค่าประมาณร้อยละ 24.00 และ 21.69 ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบ

ระหว่างวิธีการตรวจสอบ 3 วิธีดังกล่าว ภายใต้ความยากของข้อสอบต่างกัน 3 ระดับ โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง พบว่า ผลของปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบและความยากของข้อสอบมีนัยสำคัญทางสถิติที่ระดับ .01 ดังนั้นจึงเปรียบเทียบ 3 กรณี คือ 1) เปรียบเทียบอำนาจการทดสอบของแต่ละวิธีการตรวจสอบ 3 วิธีภายใต้ความยากของข้อสอบต่างกัน 3 ระดับ 2) เปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 3 วิธีภายใต้ความยากของข้อสอบเดียวกัน 3 ระดับ และ 3) เปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 3 วิธีภายใต้ความยากของข้อสอบต่างกัน 3 ระดับ ผลปรากฏดังนี้

กรณีที่ 1 เมื่อเปรียบเทียบอำนาจการทดสอบของแต่ละวิธีการตรวจสอบ 3 วิธีภายใต้ความยากของข้อสอบต่างกัน 3 ระดับ โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย พบว่า ปัจจัยความยากของข้อสอบต่างกัน 3 ระดับ ไม่มีผลต่ออำนาจการทดสอบของวิธีการถดถอยโลจิสติกและวิธีราสช์ทรี ในขณะที่ปัจจัยค่าความยากของข้อสอบ มีผลต่ออำนาจการทดสอบของวิธีซิปเทสท์ กล่าวคือ เมื่อความยากของข้อสอบ b_M จะส่งผลให้อำนาจการทดสอบสูงกว่าภายใต้เงื่อนไขความยากของข้อสอบ b_H ส่วนอำนาจการทดสอบภายใต้ความยากของข้อสอบ b_L มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ความยากของข้อสอบ b_M และ b_H

กรณีที่ 2 เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 3 วิธีภายใต้ความยากของข้อสอบเดียวกัน 3 ระดับ โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย พบว่า วิธีซิปเทสท์มีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติกและวิธีราสช์ทรีเกือบทุกระดับความยากของข้อสอบ ยกเว้นความยากของข้อสอบ b_H อำนาจการทดสอบของวิธีซิปเทสท์และวิธีราสช์ทรีมีค่าใกล้เคียงกัน

กรณีที่ 3 เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 3 วิธีภายใต้ความยากของข้อสอบต่างกัน 3 ระดับ โดยใช้วิธีการทดสอบของ Bonferroni และวิธีการทดสอบของ Games-Howell พบว่า อำนาจการทดสอบของวิธีซิปเทสท์ภายใต้ความยากของข้อสอบที่มีค่าความยากปานกลาง (b_M) มีค่าสูงสุด โดยมีค่าสูงกว่าวิธีการถดถอยโลจิสติกและวิธีราสช์ทรีในทุกเงื่อนไขของการตรวจสอบ ส่วนอำนาจการทดสอบด้วยวิธีการถดถอยโลจิสติกภายใต้ความยากของข้อสอบที่มีค่าความยากต่ำ (b_L) มีค่าต่ำสุด โดยมีค่าต่ำกว่าวิธีซิปเทสท์และวิธีราสช์ทรีเกือบทุกเงื่อนไขของการตรวจสอบ

1.2 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญภายใต้ปัจจัยความยาวของแบบสอบต่างกัน 2 ขนาด พบว่า อำนาจการทดสอบของวิธีการถดถอยโลจิสติกภายใต้ความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ มีค่าประมาณร้อยละ 15.71 และ 16.17 ตามลำดับ สำหรับวิธีซิปเทสท์ พบว่า อำนาจการทดสอบภายใต้ความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ มีค่าประมาณร้อยละ 41.93 และ 39.41 ตามลำดับ ส่วนวิธีราสช์ทรี พบว่า อำนาจการทดสอบภายใต้ความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ มีค่าประมาณร้อยละ 23.13 และ 23.89

ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 3 วิธีดังกล่าว ภายใต้ลักษณะความยาวของแบบสอบต่างกัน 2 ขนาด โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง พบว่า ผลของปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบและความยาวของแบบสอบไม่มีนัยสำคัญทางสถิติ แสดงว่า วิธีการตรวจสอบและความยาวของแบบสอบไม่มีผลร่วมกันต่ออำนาจการทดสอบ เมื่อพิจารณาผลของความยาวของแบบสอบ พบว่า ไม่มีนัยสำคัญทางสถิติ แสดงว่า ความยาวของแบบสอบไม่มีผลต่ออำนาจการทดสอบ และเมื่อพิจารณาผลของวิธีการตรวจสอบ พบว่า มีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า วิธีการตรวจสอบมีผลต่ออำนาจการทดสอบ ดังนั้น จึงเปรียบเทียบเพียงกรณีเดียว คือ เปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 3 วิธีภายใต้ความยาวของแบบสอบขนาดเดียวกัน 2 ขนาด โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย ผลปรากฏดังนี้

เมื่อเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญระหว่างวิธีการตรวจสอบ 3 วิธีภายใต้ความยาวของแบบสอบขนาดเดียวกัน 2 ขนาด โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย พบว่า วิธีซิปเทสท์มีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติกและวิธีราสช์ทรีทั้งความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ และอำนาจทดสอบของวิธีราสช์ทรีสูงกว่าวิธีการถดถอยโลจิสติกทั้งความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ

1.3 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญภายใต้ปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด พบว่า อำนาจการทดสอบของวิธีการถดถอยโลจิสติกภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และร้อยละ 20 มีค่าประมาณร้อยละ 15.78 และ 16.10 ตามลำดับ สำหรับวิธีซิปเทสท์ พบว่า อำนาจการทดสอบภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และร้อยละ 20 มีค่าประมาณร้อยละ 41.39 และ 39.95 ตามลำดับ ส่วนวิธีราสช์ทรี พบว่า อำนาจการทดสอบภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และร้อยละ 20 มีค่าประมาณร้อยละ 20.03 และ 26.99 ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 3 วิธีดังกล่าว ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง พบว่า ผลของปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบและสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันไม่มีนัยสำคัญทางสถิติ แสดงว่า วิธีการตรวจสอบและสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันไม่มีผลร่วมกันต่ออำนาจการทดสอบ เมื่อพิจารณาผลของสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน พบว่า ไม่มีนัยสำคัญทางสถิติ แสดงว่า สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันไม่มีผลต่ออำนาจการทดสอบ และเมื่อพิจารณาผลของวิธีการตรวจสอบ พบว่า มีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า วิธีการตรวจสอบมีผลต่ออำนาจการทดสอบ ดังนั้น จึงเปรียบเทียบเพียงกรณีเดียว คือ เปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 3 วิธีภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย ผลปรากฏดังนี้

เมื่อเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญระหว่างวิธีการตรวจสอบ 3 วิธีภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย พบว่า วิธีซิปเทสท์มีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติกและวิธีราสช์ทรี ยกเว้น สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน ร้อยละ 20 อำนาจการทดสอบของวิธีซิปเทสท์และวิธีราสช์ทรีมีค่าใกล้เคียงกัน และอำนาจการทดสอบของวิธีราสช์ทรีมีค่าสูงกว่าวิธีการถดถอยโลจิสติก ที่สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน ร้อยละ 20

1.4 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญภายใต้ปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด พบว่า อำนาจการทดสอบของวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน มีค่าสูงสุด ประมาณร้อยละ 23.13 รองลงมาได้แก่ภายใต้ขนาดกลุ่มตัวอย่าง 500 คน และ 250 คน ตามลำดับ มีค่าประมาณร้อยละ 13.96 และ 10.72 ตามลำดับ สำหรับวิธีซิปเทสท์ พบว่า อำนาจการทดสอบภายใต้ภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน มีค่าสูงสุด ประมาณร้อยละ 61.76 รองลงมาได้แก่ภายใต้ภายใต้ขนาดกลุ่มตัวอย่าง 500 คน และ 250 คน ตามลำดับ มีค่าประมาณร้อยละ 39.13 และ 21.12 ตามลำดับ ส่วนวิธีราสช์ทรี พบว่า อำนาจการทดสอบภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน มีค่าสูงสุด ประมาณร้อยละ 29.60 รองลงมาได้แก่ภายใต้ขนาดกลุ่มตัวอย่าง 500 คน และ 250 คน ตามลำดับ มีค่าประมาณร้อยละ 27.25 และ 13.67 ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 3 วิธีดังกล่าว ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง พบว่า ผลของปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบและขนาดกลุ่มตัวอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ดังนั้น จึงเปรียบเทียบ 3 กรณี คือ 1) เปรียบเทียบอำนาจการทดสอบของแต่ละวิธีการตรวจสอบ 3 วิธีภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด 2) เปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 3 วิธีภายใต้ขนาดกลุ่มตัวอย่างขนาดเดียวกัน 3 ขนาด และ 3) เปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 3 วิธีภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด ผลปรากฏดังนี้

กรณีที่ 1 เมื่อเปรียบเทียบอำนาจการทดสอบของแต่ละวิธีการตรวจสอบ 3 วิธีภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย พบว่า ปัจจัยขนาดกลุ่มตัวอย่าง มีผลต่ออำนาจการทดสอบของวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี กล่าวคือ สำหรับวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ ได้ผลสอดคล้องกัน เมื่อขนาดกลุ่มตัวอย่าง 1000 คน จะส่งผลให้อำนาจการทดสอบสูงกว่าภายใต้ขนาดกลุ่มตัวอย่าง 250 คน และ 500 คน อีกทั้ง เมื่อขนาดกลุ่มตัวอย่าง 500 คน จะส่งผลให้อำนาจการทดสอบสูงกว่าภายใต้ขนาดกลุ่มตัวอย่าง 250 คน ส่วนวิธีราสช์ทรี พบว่า เมื่อใช้กลุ่มตัวอย่างขนาด 500 คน และ 1000 คน จะส่งผลให้อำนาจการทดสอบสูงกว่าภายใต้ขนาดกลุ่มตัวอย่าง 250 คน

กรณีที่ 2 เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 3 วิธีภายใต้ขนาดกลุ่มตัวอย่างขนาดเดียวกัน 3 ขนาด โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย พบว่า วิธีซิปเทสท์มีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติกและวิธีราสช์ทรีทุกขนาดกลุ่มตัวอย่าง 250 คน 500 คน และ 1000 คน และวิธีราสช์ทรีมีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติก ที่ขนาดกลุ่มตัวอย่าง 500 คน และ 1000 คน

กรณีที่ 3 เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 3 วิธีภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้วิธีการทดสอบของ Bonferroni และวิธีการทดสอบของ Games-Howell พบว่า อำนาจการทดสอบของวิธีซิปเทสท์ภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน มีค่าสูงกว่าอำนาจการทดสอบของวิธีการถดถอยโลจิสติกและวิธีราสช์ทรีในทุกเงื่อนไขของการตรวจสอบ ส่วนการตรวจสอบด้วยวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่าง 250 คน มีค่าต่ำกว่าวิธีซิปเทสท์และวิธีราสช์ทรีเกือบทุกเงื่อนไขของการตรวจสอบ

2. การเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพล ระหว่างวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยที่ศึกษาปรากฏผลดังนี้

2.1 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลภายใต้ปัจจัยความยากของข้อสอบต่างกัน 3 ระดับ พบว่า อำนาจการทดสอบของวิธีการถดถอยโลจิสติกภายใต้ความยากของข้อสอบ b_M มีค่าสูงสุด ประมาณร้อยละ 14.91 รองลงมาได้แก่ภายใต้ความยากของข้อสอบ b_H และ b_L ตามลำดับ มีค่าประมาณร้อยละ 10.17 และ 8.33 ตามลำดับ สำหรับวิธีซิปเทสท์ พบว่า อำนาจการทดสอบภายใต้ความยากของข้อสอบ b_M มีค่าสูงสุด ประมาณร้อยละ 54.32 รองลงมาได้แก่ภายใต้ความยากของข้อสอบ b_L และ b_H ตามลำดับ มีค่าประมาณร้อยละ 41.22 และ 26.47 ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 2 วิธีดังกล่าว ภายใต้ความยากของข้อสอบต่างกัน 3 ระดับ โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง พบว่า ผลของปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบและความยากของข้อสอบมีนัยสำคัญทางสถิติที่ระดับ .05 ดังนั้นจึงเปรียบเทียบ 3 กรณี คือ 1) เปรียบเทียบอำนาจการทดสอบของแต่ละวิธีการตรวจสอบ 2 วิธีภายใต้ความยากของข้อสอบต่างกัน 3 ระดับ 2) เปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 2 วิธีภายใต้ความยากของข้อสอบเดียวกัน 3 ระดับ และ 3) เปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 2 วิธีภายใต้ความยากของข้อสอบต่างกัน 3 ระดับ ผลปรากฏดังนี้

กรณีที่ 1 เมื่อเปรียบเทียบอำนาจการทดสอบของแต่ละวิธีการตรวจสอบ 2 วิธีภายใต้ความยากของข้อสอบต่างกัน 3 ระดับ โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย พบว่า มีผลต่ออำนาจการทดสอบของวิธีซิปเทสท์ กล่าวคือ เมื่อความยากของข้อสอบ b_M จะส่งผลให้อำนาจการทดสอบสูงกว่า

ภายใต้เงื่อนไขความยากของข้อสอบ b_H ส่วนอำนาจการทดสอบภายใต้ความยากของข้อสอบ bL มีค่าไม่แตกต่างกับอำนาจการทดสอบภายใต้ความยากของข้อสอบ b_M และ b_H สำหรับวิธีการถดถอยโลจิสติก พบว่า ปัจจัยค่าความยากของข้อสอบไม่มีผลต่ออำนาจการทดสอบ

กรณีที่ 2 เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 2 วิธีภายใต้ความยากของข้อสอบเดียวกัน 3 ระดับ โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย พบว่า วิธีซิปเทสท์มีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติกทุกระดับความยากของข้อสอบ

กรณีที่ 3 เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 2 วิธีภายใต้ความยากของข้อสอบต่างกัน 3 ระดับ โดยใช้วิธีการทดสอบของ Bonferroni และวิธีการทดสอบของ Games-Howell พบว่า อำนาจการทดสอบของวิธีซิปเทสท์ภายใต้ความยากของข้อสอบที่มีค่าความยากปานกลางมีค่าสูงสุด โดยมีค่าสูงกว่าวิธีการถดถอยโลจิสติกในทุกเงื่อนไขการตรวจสอบ ส่วนการตรวจสอบด้วยวิธีการถดถอยโลจิสติกภายใต้ความยากของข้อสอบที่มีค่าความยากต่ำมีค่าต่ำสุด โดยมีค่าต่ำกว่าวิธีซิปเทสท์ในทุกเงื่อนไขการตรวจสอบ

2.2 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลภายใต้ปัจจัยความยาวของแบบสอบต่างกัน 2 ขนาด พบว่า อำนาจการทดสอบของวิธีการถดถอยโลจิสติกภายใต้ความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ มีค่าประมาณร้อยละ 12.10 และ 10.18 ตามลำดับ สำหรับวิธีซิปเทสท์ พบว่า อำนาจการทดสอบภายใต้ความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ มีค่าประมาณร้อยละ 41.93 และ 39.41 ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 2 วิธีดังกล่าว ภายใต้ลักษณะความยาวของแบบสอบต่างกัน 2 ขนาด โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง พบว่า ผลของปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบและความยาวของแบบสอบไม่มีนัยสำคัญทางสถิติ แสดงว่า วิธีการตรวจสอบและความยาวของแบบสอบไม่มีผลร่วมกันต่ออำนาจการทดสอบ เมื่อพิจารณาผลของความยาวของแบบสอบ พบว่า ไม่มีนัยสำคัญทางสถิติ แสดงว่า ความยาวของแบบสอบไม่มีผลต่ออำนาจการทดสอบ และเมื่อพิจารณาผลของวิธีการตรวจสอบ พบว่า มีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า วิธีการตรวจสอบมีผลต่ออำนาจการทดสอบ ดังนั้น จึงเปรียบเทียบเพียงกรณีเดียว คือ เปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 2 วิธีภายใต้ความยาวของแบบสอบขนาดเดียวกัน 2 ขนาด โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย ผลปรากฏดังนี้

เมื่อเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลระหว่างวิธีการตรวจสอบ 2 วิธีภายใต้ความยาวของแบบสอบขนาดเดียวกัน 2 ขนาด โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย พบว่า วิธีซิปเทสท์มีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติก ทั้งความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ

2.3 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลภายใต้ปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด พบว่า อำนาจการทดสอบของวิธีการถดถอยโลจิสติกภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และร้อยละ 20 มีค่าประมาณร้อยละ 10.97 และ 11.30 ตามลำดับ สำหรับวิธีซิปเทสท์ พบว่า อำนาจการทดสอบภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และร้อยละ 20 มีค่าประมาณร้อยละ 41.39 และ 39.95 ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 2 วิธีดังกล่าว ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง พบว่า ผลของปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบและสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันไม่มีนัยสำคัญทางสถิติ แสดงว่า วิธีการตรวจสอบและสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันไม่มีผลร่วมกันต่ออำนาจการทดสอบ เมื่อพิจารณาผลของสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน พบว่า ไม่มีนัยสำคัญทางสถิติ แสดงว่า สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันไม่มีผลต่ออำนาจการทดสอบ และเมื่อพิจารณาผลของวิธีการตรวจสอบ พบว่า มีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า วิธีการตรวจสอบมีผลต่ออำนาจการทดสอบ ดังนั้น จึงเปรียบเทียบเพียงกรณีเดียว คือ เปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 2 วิธีภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย ผลปรากฏดังนี้

เมื่อเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลระหว่างวิธีการตรวจสอบ 2 วิธีภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย พบว่า วิธีซิปเทสท์มีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติก ทั้งสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน ร้อยละ 10 และร้อยละ 20

2.4 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลภายใต้ปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด พบว่า อำนาจการทดสอบของวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน มีค่าสูงสุด ประมาณร้อยละ 21.54 รองลงมาได้แก่ภายใต้ขนาดกลุ่มตัวอย่าง 500 คน และ 250 คน ตามลำดับ มีค่าประมาณร้อยละ 9.33 และ 2.53 ตามลำดับ สำหรับวิธีซิปเทสท์ พบว่า อำนาจการทดสอบภายใต้ภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน มีค่าสูงสุด ประมาณร้อยละ 61.76 รองลงมาได้แก่ภายใต้ภายใต้ขนาดกลุ่มตัวอย่าง 500 คน และ 250 คนตามลำดับ มีค่าประมาณร้อยละ 39.13 และ 21.12 ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 2 วิธีดังกล่าว ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง พบว่า ผลของปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบและขนาดกลุ่มตัวอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ดังนั้นจึงเปรียบเทียบ 3 กรณี คือ 1) เปรียบเทียบอำนาจการทดสอบของแต่ละวิธีการตรวจสอบ 2 วิธีภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด 2) เปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 2 วิธีภายใต้ขนาดกลุ่มตัวอย่างขนาดเดียวกัน 3 ขนาด และ

3) เปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 2 วิธีภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด ผลปรากฏดังนี้

กรณีที่ 1 เมื่อเปรียบเทียบอำนาจการทดสอบของแต่ละวิธีการตรวจสอบ 2 วิธีภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย พบว่า มีผลต่ออำนาจการทดสอบของวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ สอดคล้องกัน กล่าวคือ เมื่อขนาดกลุ่มตัวอย่าง 1000 คน จะส่งผลให้อำนาจการทดสอบสูงกว่าภายใต้ขนาดกลุ่มตัวอย่าง 250 คน และ 500 คน อีกทั้ง เมื่อขนาดกลุ่มตัวอย่าง 500 คน จะส่งผลให้อำนาจการทดสอบสูงกว่าภายใต้ขนาดกลุ่มตัวอย่าง 250 คน

กรณีที่ 2 เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 2 วิธีภายใต้ขนาดกลุ่มตัวอย่างขนาดเดียวกัน 3 ขนาด โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย พบว่า วิธีซิปเทสท์มีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติก ทุกขนาดกลุ่มตัวอย่าง 250 คน 500 คน และ 1000 คน

กรณีที่ 3 เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 2 วิธีภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้วิธีการทดสอบของ Bonferroni และวิธีการทดสอบของ Games-Howell พบว่า อำนาจการทดสอบของวิธีซิปเทสท์ภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน มีค่าสูงสุด โดยมีค่าสูงกว่าวิธีการถดถอยโลจิสติกในทุกเงื่อนไขการตรวจสอบ ส่วนการตรวจสอบด้วยวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่าง 250 คน มีค่าต่ำสุด โดยมีค่าต่ำกว่าวิธีซิปเทสท์ในทุกเงื่อนไขการตรวจสอบ

3. การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยที่ศึกษา ปรากฏผลดังนี้

3.1 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญภายใต้ปัจจัยความยากของข้อสอบต่างกัน 3 ระดับ พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกภายใต้ความยากของข้อสอบ b_L มีค่าสูงสุด ประมาณร้อยละ 6.52 รองลงมาได้แก่ภายใต้ความยากของข้อสอบ b_H และ b_M ตามลำดับ มีค่าประมาณร้อยละ 6.34 และ 5.82 ตามลำดับ สำหรับวิธีซิปเทสท์ พบว่า อำนาจการทดสอบภายใต้ความยากของข้อสอบ b_L มีค่าสูงสุด ประมาณร้อยละ 5.48 รองลงมาได้แก่ภายใต้ความยากของข้อสอบ b_H และ b_M ตามลำดับ มีค่าประมาณร้อยละ 5.22 และ 5.21 ตามลำดับ ส่วนวิธีราสช์ทรี พบว่า อำนาจการทดสอบภายใต้ความยากของข้อสอบ b_H มีค่าสูงสุด ประมาณร้อยละ 2.95 รองลงมาได้แก่ภายใต้ความยากของข้อสอบ b_L และ b_M ตามลำดับ มีค่าประมาณร้อยละ 2.63 และ 1.40 ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 3 วิธีดังกล่าว ภายใต้ความยากของข้อสอบต่างกัน

3 ระดับ โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง พบว่า ผลของปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบและความยากของข้อสอบไม่มีนัยสำคัญทางสถิติ แสดงว่า วิธีการตรวจสอบและความยากของข้อสอบไม่มีผลร่วมกันต่ออัตราความคลาดเคลื่อนประเภทที่ 1 เมื่อพิจารณาผลของความยากของข้อสอบ พบว่า มีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่า ความยากของข้อสอบมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 และเมื่อพิจารณาผลของวิธีการตรวจสอบ พบว่า มีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า วิธีการตรวจสอบมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ดังนั้น จึงเปรียบเทียบ 2 กรณี คือ 1) เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธีการตรวจสอบ 3 วิธีภายใต้ความยากของข้อสอบต่างกัน 3 ระดับ และ 2) เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 3 วิธีภายใต้ความยากของข้อสอบเดียวกัน 3 ระดับ ผลปรากฏดังนี้

กรณีที่ 1 เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธีการตรวจสอบ 3 วิธีภายใต้ความยากของข้อสอบต่างกัน 3 ระดับ โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย พบว่า ปัจจัยค่าความยากของข้อสอบมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1ของวิธีการถดถอยโลจิสติก กล่าวคือ เมื่อความยากของข้อสอบ b_L จะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าภายใต้เงื่อนไขความยากของข้อสอบ b_M ส่วนวิธีซิปเทสท์และวิธีราสช์ทรี พบว่า ปัจจัยค่าความยากของข้อสอบไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีนี้

กรณีที่ 2 เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 3 วิธีภายใต้ความยากของข้อสอบเดียวกัน 3 ระดับ โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย พบว่า วิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีราสช์ทรีทุกระดับความยากของข้อสอบ และวิธีการถดถอยโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีซิปเทสท์ ภายใต้ความยากของข้อสอบ b_L

3.2 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญภายใต้ปัจจัยปัจจัยความยาวของแบบสอบต่างกัน 2 ขนาด พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกภายใต้ความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ มีค่าประมาณร้อยละ 6.20 และ 6.26 ตามลำดับ สำหรับวิธีซิปเทสท์ พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ มีค่าประมาณร้อยละ 5.22 และ 5.39 ตามลำดับ ส่วนวิธีราสช์ทรี พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ มีค่าประมาณร้อยละ 2.21 และ 2.44 ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 3 วิธีดังกล่าว ภายใต้ลักษณะความยาวของแบบสอบต่างกัน 2 ขนาด โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง พบว่า ผลของปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบและความยาวของแบบสอบไม่มีนัยสำคัญทางสถิติ แสดงว่า วิธีการตรวจสอบและความยาว

ของแบบสอบไม่มีผลร่วมกันต่ออัตราความคลาดเคลื่อนประเภทที่ 1 เมื่อพิจารณาผลของความยาวของแบบสอบ พบว่า ไม่มีนัยสำคัญทางสถิติ แสดงว่า ความยาวของแบบสอบไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 และเมื่อพิจารณาผลของวิธีการตรวจสอบ พบว่า มีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า วิธีการตรวจสอบมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ดังนั้น จึงเปรียบเทียบเพียงกรณีเดียว คือ เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 3 วิธีภายใต้ความยาวของแบบสอบขนาดเดียวกัน 2 ขนาด โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย ผลปรากฏดังนี้

เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญระหว่างวิธีการตรวจสอบ 3 วิธีภายใต้ความยาวของแบบสอบขนาดเดียวกัน 2 ขนาด โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย พบว่า วิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีราสช์ทรีทั้งความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ

3.3 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญภายใต้ปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และร้อยละ 20 มีค่าประมาณร้อยละ 6.24 และ 6.22 ตามลำดับ สำหรับวิธีซิปเทสท์ พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และร้อยละ 20 มีค่าประมาณร้อยละ 5.22 และ 5.39 ตามลำดับ ส่วนวิธีราสช์ทรี พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน ร้อยละ 10 และร้อยละ 20 มีค่าประมาณร้อยละ 1.73 และ 2.92 ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 3 วิธีดังกล่าว ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง พบว่า ผลของปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบและสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันไม่มีนัยสำคัญทางสถิติ แสดงว่า วิธีการตรวจสอบและสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันไม่มีผลร่วมกันต่ออัตราความคลาดเคลื่อนประเภทที่ 1 เมื่อพิจารณาผลของสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน พบว่า ไม่มีนัยสำคัญทางสถิติ แสดงว่า สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 และเมื่อพิจารณาผลของวิธีการตรวจสอบ พบว่า มีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า วิธีการตรวจสอบมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ดังนั้น จึงเปรียบเทียบเพียงกรณีเดียว คือ เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 3 วิธีภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย ผลปรากฏดังนี้

เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญระหว่างวิธีการตรวจสอบ 3 วิธีภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย พบว่า วิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีราสช์ทรีทั้งสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน ร้อยละ 10 และร้อยละ 20

3.4 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญภายใต้ปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน มีค่าสูงสุด ประมาณร้อยละ 6.42 รองลงมาได้แก่ภายใต้ขนาดกลุ่มตัวอย่าง 500 คน และ 250 คน ตามลำดับ มีค่าประมาณร้อยละ 6.25 และ 6.02 ตามลำดับ สำหรับวิธีซิปเทสท์ พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน มีค่าสูงสุด ประมาณร้อยละ 5.81 รองลงมาได้แก่ภายใต้ภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน และ 500 คน ตามลำดับ มีค่าประมาณร้อยละ 5.15 และ 4.95 ตามลำดับ ส่วนวิธีราสช์ทรี พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน มีค่าสูงสุด ประมาณร้อยละ 4.03 รองลงมาได้แก่ภายใต้ขนาดกลุ่มตัวอย่าง 500 คน และ 1000 คน ตามลำดับ มีค่าประมาณร้อยละ 2.54 และ 0.41 ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 3 วิธีดังกล่าว ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง พบว่า ผลของปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบและขนาดกลุ่มตัวอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ดังนั้นจึงเปรียบเทียบ 3 กรณี คือ 1) เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธีการตรวจสอบ 3 วิธีภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด 2) เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 3 วิธีภายใต้ขนาดกลุ่มตัวอย่างขนาดเดียวกัน 3 ขนาด และ 3) เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 3 วิธีภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด ผลปรากฏดังนี้

กรณีที่ 1 เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธีการตรวจสอบ 3 วิธีภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย พบว่า ปัจจัยขนาดกลุ่มตัวอย่างมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีราสช์ทรี กล่าวคือ เมื่อขนาดกลุ่มตัวอย่าง 250 คน และ 500 คน จะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน

กรณีที่ 2 เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 3 วิธีภายใต้ขนาดกลุ่มตัวอย่างขนาดเดียวกัน 3 ขนาด โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีราสช์ทรีทุกขนาดกลุ่มตัวอย่าง 250 คน 500 คน และ 1000 คน

กรณีที่ 3 เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 3 วิธีภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้วิธีการทดสอบของ Bonferroni และวิธีการทดสอบของ Games-Howell พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน มีค่าสูงสุด โดยมีค่าสูงกว่าวิธีซิปเทสท์และวิธีราสช์ทรีเกือบทุกเงื่อนไขของการตรวจสอบ ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีราสช์ทรีภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน มีค่าต่ำสุด โดยมีค่าต่ำกว่าวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ในทุกเงื่อนไขของการตรวจสอบ

4. การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพล ระหว่างวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยที่ศึกษา ปรากฏผลดังนี้

4.1 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลภายใต้ปัจจัยความยากของข้อสอบต่างกัน 3 ระดับ พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกภายใต้ความยากของข้อสอบ b_M มีค่าสูงสุด ประมาณร้อยละ 4.43 รองลงมาได้แก่ภายใต้ความยากของข้อสอบ b_L และ b_H ตามลำดับ มีค่าประมาณร้อยละ 3.70 และ 3.65 ตามลำดับ สำหรับวิธีซิปเทสท์ พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ความยากของข้อสอบ b_L มีค่าสูงสุด ประมาณร้อยละ 5.48 รองลงมาได้แก่ภายใต้ความยากของข้อสอบ b_H และ b_M ตามลำดับ มีค่าประมาณร้อยละ 5.22 และ 5.21 ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 2 วิธีดังกล่าว ภายใต้ความยากของข้อสอบต่างกัน 3 ระดับ โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง พบว่า ผลของปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบและความยากของข้อสอบไม่มีนัยสำคัญทางสถิติ แสดงว่า วิธีการตรวจสอบและความยากของข้อสอบไม่มีผลร่วมกันต่ออัตราความคลาดเคลื่อนประเภทที่ 1 เมื่อพิจารณาผลของความยากของข้อสอบ พบว่าไม่มีนัยสำคัญทางสถิติ แสดงว่า ความยากของข้อสอบไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 และเมื่อพิจารณาผลของวิธีการตรวจสอบ พบว่า มีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า วิธีการตรวจสอบมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ดังนั้นจึงเปรียบเทียบเพียงกรณีเดียว คือ เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 2 วิธีภายใต้ความยากของข้อสอบเดียวกัน 3 ระดับ ผลปรากฏดังนี้

เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 2 วิธีภายใต้ความยากของข้อสอบเดียวกัน 3 ระดับ โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย พบว่า วิธีซิปเทสท์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีการถดถอยโลจิสติกที่ระดับความยากของข้อสอบ b_Lและ b_H

4.2 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลภายใต้ปัจจัยความยาวของแบบสอบต่างกัน 2 ขนาด พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1ของวิธีการถดถอยโลจิสติกภายใต้ความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ มีค่าประมาณร้อยละ 4.52 และ 3.33 ตามลำดับ สำหรับวิธีซิปเทสท์ พบว่า อำนาจการทดสอบภายใต้ความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ มีค่าประมาณร้อยละ 5.22 และ 5.39 ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 2 วิธีดังกล่าว ภายใต้ลักษณะความยาวของแบบสอบต่างกัน 2 ขนาด โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง พบว่า ผลของปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบและความยาวของแบบสอบไม่มีนัยสำคัญทางสถิติ แสดงว่า วิธีการตรวจสอบและความยาวของแบบสอบไม่มีผลร่วมกันต่ออัตราความคลาดเคลื่อนประเภทที่ 1 เมื่อพิจารณาผลของความยาวของแบบสอบ พบว่า ไม่มีนัยสำคัญทางสถิติ แสดงว่า ความยาวของแบบสอบไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 และเมื่อพิจารณาผลของวิธีการตรวจสอบ พบว่า มีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า วิธีการตรวจสอบมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ดังนั้น จึงเปรียบเทียบเพียงกรณีเดียว คือ เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1ระหว่างวิธีการตรวจสอบ 2 วิธีภายใต้ความยาวของแบบสอบขนาดเดียวกัน 2 ขนาด โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย ผลปรากฏดังนี้

เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลระหว่างวิธีการตรวจสอบ 2 วิธีภายใต้ความยาวของแบบสอบขนาดเดียวกัน 2 ขนาด โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย พบว่า วิธีซิปเทสท์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีการถดถอยโลจิสติก ภายใต้ความยาวของแบบสอบ 60 ข้อ

4.3 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลภายใต้ปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และร้อยละ 20 มีค่าประมาณร้อยละ 3.92 และ 3.93 ตามลำดับ สำหรับวิธีซิปเทสท์ พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และร้อยละ 20 มีค่าประมาณร้อยละ 5.22 และ 5.39 ตามลำดับ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 2 วิธีดังกล่าว ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง พบว่า ผลของปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบและสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันไม่มีนัยสำคัญทางสถิติ แสดงว่า วิธีการตรวจสอบและสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันไม่มีผลร่วมกันต่ออัตราความคลาดเคลื่อนประเภทที่ 1 เมื่อพิจารณาผลของสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน พบว่า ไม่มีนัยสำคัญทางสถิติ แสดงว่า สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 และเมื่อพิจารณาผลของวิธีการ

ตรวจสอบ พบว่า มีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า วิธีการตรวจสอบมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ดังนั้น จึงเปรียบเทียบเพียงกรณีเดียว คือ เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 2 วิธีภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย ผลปรากฏดังนี้

เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลระหว่างวิธีการตรวจสอบ 2 วิธีภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย พบว่า วิธีซิปเทสท์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีการถดถอยโลจิสติก ทั้งสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และร้อยละ 20

4.4 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลภายใต้ปัจจัยขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน มีค่าสูงสุด ประมาณร้อยละ 5.93 รองลงมาได้แก่ภายใต้ขนาดกลุ่มตัวอย่าง 500 คน และ 250 คน ตามลำดับ มีค่าประมาณร้อยละ 4.19 และ 1.65 ตามลำดับ สำหรับวิธีซิปเทสท์ พบว่า อำนาจการทดสอบภายใต้ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน มีค่าสูงสุด ประมาณร้อยละ 5.81 รองลงมาได้แก่ภายใต้ภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน และ 500 คน ตามลำดับ มีค่าประมาณร้อยละ 5.15 และ 4.95 ตามลำดับ เมื่อเปรียบเทียบอำนาจการทดสอบระหว่างวิธีการตรวจสอบ 2 วิธีดังกล่าว ภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง พบว่า ผลของปฏิสัมพันธ์ระหว่างวิธีการตรวจสอบและขนาดกลุ่มตัวอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ดังนั้นจึงเปรียบเทียบ 3 กรณี คือ 1) เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธีการตรวจสอบ 2 วิธีภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด 2) เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 2 วิธีภายใต้ขนาดกลุ่มตัวอย่างขนาดเดียวกัน 3 ขนาด และ 3) เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1ระหว่างวิธีการตรวจสอบ 2 วิธีภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด ผลปรากฏดังนี้

กรณีที่ 1 เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธีการตรวจสอบ 2 วิธีภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และ .05 ตามลำดับ แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ขนาดกลุ่มตัวอย่างมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบทั้ง 2 วิธี ดังนั้น จึงนำอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธีการตรวจสอบทั้ง 2 วิธีภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาดมาเปรียบเทียบเป็นรายคู่ โดยใช้วิธีการทดสอบของ Bonferroni และวิธีการทดสอบของ Games-Howell พบว่า

ปัจจัยขนาดกลุ่มตัวอย่างมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก กล่าวคือ ภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน จะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าภายใต้ขนาดกลุ่มตัวอย่าง 250 คน และ 500 คน อีกทั้ง ภายใต้ขนาดกลุ่มตัวอย่าง 500 คน จะส่งผลให้อำนาจการทดสอบสูงกว่าภายใต้ขนาดกลุ่มตัวอย่าง 250 คน ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ไม่มีผลจากปัจจัยขนาดกลุ่มตัวอย่าง

กรณีที่ 2 เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 2 วิธีภายใต้ขนาดกลุ่มตัวอย่างขนาดเดียวกัน 3 ขนาด โดยใช้การวิเคราะห์อิทธิพลอย่างง่าย พบว่า วิธีซิปเทสท์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีการถดถอยโลจิสติก ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน ในขณะที่วิธีการถดถอยโลจิสติก มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีซิปเทสท์ ภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน

กรณีที่ 3 เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 2 วิธีภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด โดยใช้วิธีการทดสอบของ Bonferroni และวิธีการทดสอบของ Games-Howell พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน จะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าสูงสุด โดยมีค่าสูงกว่าวิธีซิปเทสท์เกือบทุกเงื่อนไขของการตรวจสอบ ส่วนการตรวจสอบด้วยวิธีการถดถอยโลจิสติกภายใต้ขนาดกลุ่มตัวอย่าง 250 คน มีค่าต่ำสุด โดยมีค่าต่ำกว่าวิธีซิปเทสท์ในทุกเงื่อนไขการตรวจสอบ

5. การเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการถดถอยโลจิสติกวิธีซิปเทสท์ และวิธีราสช์ทรีในข้อมูลเชิงประจักษ์

5.1 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ จำแนกตามวิธีที่ศึกษา

ผลการตรวจสอบการทำหน้าที่ต่างกันด้วยวิธีการวัดพื้นที่ของราชูจากแบบสอบทั้งหมด จำนวน 205 ข้อ พบข้อสอบที่ทำหน้าที่ต่างกัน จำนวน 37 ข้อ และข้อสอบที่ทำหน้าที่ไม่ต่างกัน จำนวน 168 ข้อ

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก พบข้อสอบที่ทำหน้าที่ต่างกัน จำนวน 30 ข้อ ส่วนข้อสอบที่ทำหน้าที่ไม่ต่างกัน จำนวน 175 ข้อ โดยพบข้อร่วมของการทำหน้าที่ต่างกันระหว่างวิธีการถดถอยโลจิสติกกับวิธีเกณฑ์ จำนวน 10 ข้อ

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีซิปเทสท์ พบข้อสอบที่ทำหน้าที่ต่างกัน จำนวน 33 ข้อ ส่วนข้อสอบที่ทำหน้าที่ไม่ต่างกัน จำนวน 172 ข้อ โดยพบข้อร่วมของการทำหน้าที่ต่างกันระหว่างวิธีการถดถอยโลจิสติกกับวิธีเกณฑ์ จำนวน 20 ข้อ

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีราสช์ทรี พบข้อสอบที่ทำหน้าที่ต่างกัน จำนวน 56 ข้อ ส่วนข้อสอบที่ทำหน้าที่ไม่ต่างกัน จำนวน 149 ข้อ โดยพบข้อร่วมของการทำหน้าที่ต่างกันระหว่างวิธีการถดถอยโลจิสติกกับวิธีเกณฑ์ จำนวน 34 ข้อ

5.2 ประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

อำนาจการทดสอบของวิธีการถดถอยโลจิสติก ประมาณร้อยละ 27.03 วิธีซิปเทสท์ ประมาณร้อยละ 54.05 และวิธีราสช์ทรี ประมาณร้อยละ 91.89 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก ประมาณร้อยละ 11.90 วิธีซิปเทสท์ ประมาณร้อยละ 7.74 และวิธีราสช์ทรี ประมาณร้อยละ 13.10 จากผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการตรวจสอบทั้ง 3 วิธี พบว่า วิธีราสช์ทรีให้ค่าร้อยละของอำนาจการทดสอบสูงสุด แต่มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ซึ่งเกินเกณฑ์ที่กำหนดที่ระดับร้อยละ 10 จึงยังไม่อาจกล่าวได้ว่า วิธีราสช์ทรีมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ส่วนวิธีซิปเทสท์มีอำนาจการทดสอบประมาณร้อยละ 54.05 และมีอัตราความคลาดเคลื่อนประเภทที่ 1 ร้อยละ 7.74 ซึ่งอยู่ในเกณฑ์ที่กำหนดที่ระดับร้อยละ 10 โดยร้อยละของอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 สอดคล้องกับการตรวจสอบในข้อมูลจำลอง สำหรับวิธีการถดถอยโลจิสติกให้ค่าร้อยละของอำนาจการทดสอบต่ำสุด ในขณะที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าเกณฑ์ที่กำหนด

จากสรุปผลการวิจัยที่กล่าวมา สามารถสรุปเป็นประเด็นสำคัญ ได้ดังนี้

1. ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้เงื่อนไขที่แตกต่างกันของปัจจัยที่ศึกษา พบว่า วิธีซิปเทสท์มีค่าเฉลี่ยรวมของอำนาจการทดสอบสูงสุด และเมื่อพิจารณาในแต่ละปัจจัยแต่ละระดับ พบว่า มีค่าเฉลี่ยสูงกว่าอำนาจการทดสอบของวิธีการถดถอยโลจิสติกและวิธีราสช์ทรีทุกเงื่อนไขของปัจจัยที่ศึกษา สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีราสช์ทรีมีค่าเฉลี่ยรวมต่ำสุด และเมื่อพิจารณาในแต่ละปัจจัยแต่ละระดับ พบว่า มีค่าเฉลี่ยต่ำกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ทุกเงื่อนไขของปัจจัยที่ศึกษา

2. การเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยที่ศึกษาต่างระดับกัน พบประเด็นที่สำคัญ ดังนี้

2.1 ปัจจัยค่าความยากของข้อสอบ (b_L b_M และ b_H) พบว่า มีผลต่ออำนาจการทดสอบของวิธีซิปเทสท์ กล่าวคือ เมื่อความยากของข้อสอบ b_M จะส่งผลให้อำนาจการทดสอบสูงกว่าภายใต้เงื่อนไขความยากของข้อสอบ b_H สำหรับวิธีการถดถอยโลจิสติก พบว่า ปัจจัยค่าความยากของข้อสอบมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 กล่าวคือ เมื่อความยากของข้อสอบ b_L จะส่งผลให้อัตรา

ความคลาดเคลื่อนประเภทที่ 1 สูงกว่าภายใต้เงื่อนไขความยากของข้อสอบ b_M ส่วนวิธีราสช์ทรี พบว่า ปัจจัยค่าความยากของข้อสอบไม่มีผลต่ออำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีนี้

2.2 ปัจจัยความยาวของแบบสอบ (40 ข้อ และ 60 ข้อ) พบว่า ไม่มีผลต่ออำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบทั้ง 3 วิธี

2.3 ปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน (10% และ 20%) พบว่า ไม่มีผลต่ออำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบทั้ง 3 วิธี

2.4 ปัจจัยขนาดกลุ่มตัวอย่าง (250 คน 500 คน และ 1000 คน) พบว่า มีผลต่ออำนาจการทดสอบของวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี กล่าวคือ สำหรับวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ ได้ผลสอดคล้องกัน เมื่อขนาดกลุ่มตัวอย่าง 1000 คน จะส่งผลให้อำนาจการทดสอบสูงกว่าภายใต้ขนาดกลุ่มตัวอย่าง 250 คน และ 500 คน อีกทั้ง เมื่อขนาดกลุ่มตัวอย่าง 500 คน จะส่งผลให้อำนาจการทดสอบสูงกว่าภายใต้ขนาดกลุ่มตัวอย่าง 250 คน ส่วนวิธีราสช์ทรี พบว่า เมื่อใช้กลุ่มตัวอย่างขนาด 500 คน และ 1000 คน จะส่งผลให้อำนาจการทดสอบสูงกว่าภายใต้ขนาดกลุ่มตัวอย่าง 250 คน นอกจากนี้ ปัจจัยขนาดกลุ่มตัวอย่างมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีราสช์ทรี กล่าวคือ เมื่อขนาดกลุ่มตัวอย่าง 250 คน และ 500 คน จะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน ในขณะที่ปัจจัยขนาดกลุ่มตัวอย่างไม่มีผลต่ออำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์

3. การเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการทดสอบระดับนัยสำคัญ ด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ภายใต้เงื่อนไขของปัจจัยที่ศึกษาระดับเดียวกัน พบประเด็นที่สำคัญ ดังนี้

3.1 ปัจจัยค่าความยากของข้อสอบ (b_L b_M และ b_H) พบว่า วิธีซิปเทสท์มีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติกและวิธีราสช์ทรีเกือบทุกระดับความยากของข้อสอบ ยกเว้นความยากของข้อสอบ b_H อำนาจการทดสอบของวิธีซิปเทสท์และวิธีราสช์ทรีมีค่าใกล้เคียงกัน ส่วนวิธีราสช์ทรีมีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติก ภายใต้ความยากของข้อสอบ b_L สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีราสช์ทรีทุกระดับความยากของข้อสอบ ส่วนวิธีการถดถอยโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีซิปเทสท์ ภายใต้ความยากของข้อสอบ b_L

3.2 ปัจจัยความยาวของแบบสอบ (40 ข้อ และ 60 ข้อ) พบว่า วิธีซิปเทสท์มีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติกและวิธีราสช์ทรีทั้งความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ

ส่วนวิธีราสช์ทรีมีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติก ภายใต้ความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีราสช์ทรีทั้งความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ ส่วนวิธีการถดถอยโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีซิปเทสท์ ภายใต้ความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ

3.3 ปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน (10% และ 20%) พบว่า วิธีซิปเทสท์มีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติกและวิธีราสช์ทรี ยกเว้น สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 20% อำนาจการทดสอบของวิธีซิปเทสท์และวิธีราสช์ทรีมีค่าใกล้เคียงกัน ส่วนวิธีราสช์ทรีมีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติก ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 20% สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีราสช์ทรีทั้งสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 10% และ 20% ส่วนวิธีการถดถอยโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีซิปเทสท์ ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 10% และ 20%

3.4 ปัจจัยขนาดกลุ่มตัวอย่าง (250 คน 500 คน และ 1000 คน) พบว่า วิธีซิปเทสท์มีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติกและวิธีราสช์ทรีทุกขนาดกลุ่มตัวอย่าง 250 คน 500 คน และ 1000 คน ส่วนวิธีราสช์ทรีมีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติก ภายใต้ขนาดกลุ่มตัวอย่าง 500 คน และ 1000 คนสำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีราสช์ทรีทุกขนาดกลุ่มตัวอย่าง 250 คน 500 คน และ 1000 คน ส่วนวิธีการถดถอยโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีซิปเทสท์ ภายใต้ขนาดกลุ่มตัวอย่าง 500 คน และ 1000 คน

4. ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ด้วยวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขที่แตกต่างกันของปัจจัยที่ศึกษา พบว่า วิธีซิปเทสท์มีค่าเฉลี่ยรวมของอำนาจการทดสอบสูงสุด และเมื่อพิจารณาในแต่ละปัจจัยแต่ละระดับ ยังพบว่า มีค่าเฉลี่ยสูงกว่าอำนาจการทดสอบของวิธีการถดถอยโลจิสติกทุกเงื่อนไขของปัจจัยที่ศึกษาอีกด้วย สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกมีค่าเฉลี่ยรวมต่ำสุด และเมื่อพิจารณาในแต่ละปัจจัยแต่ละระดับ ยังพบว่า มีค่าเฉลี่ยต่ำกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ ยกเว้นปัจจัยขนาดกลุ่มตัวอย่าง 1000 คน อัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้ง 2 วิธี มีค่าใกล้เคียงกัน

5. การเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ด้วยวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยที่ศึกษาต่างระดับกัน พบประเด็นที่สำคัญ ดังนี้

5.1 ปัจจัยค่าความยากของข้อสอบ (b_L b_M และ b_H) พบว่า มีผลต่ออำนาจการทดสอบของวิธีซิปเทสท์ กล่าวคือ เมื่อความยากของข้อสอบ b_M จะส่งผลให้อำนาจการทดสอบสูงกว่าภายใต้เงื่อนไขความยากของข้อสอบ b_H สำหรับวิธีการถดถอยโลจิสติก พบว่า ปัจจัยค่าความยากของข้อสอบไม่มีผลต่ออำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีนี้

5.2 ปัจจัยความยาวของแบบสอบ (40 ข้อ และ 60 ข้อ) พบว่า ไม่มีผลต่ออำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบทั้ง 2 วิธี

5.3 ปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน (10% และ 20%) พบว่า ไม่มีผลต่ออำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบทั้ง 2 วิธี

5.4 ปัจจัยขนาดกลุ่มตัวอย่าง (250 คน 500 คน และ 1000 คน) พบว่า มีผลต่ออำนาจการทดสอบของวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ สอดคล้องกัน กล่าวคือ เมื่อขนาดกลุ่มตัวอย่าง 1000 คน จะส่งผลให้อำนาจการทดสอบสูงกว่าภายใต้ขนาดกลุ่มตัวอย่าง 250 คน และ 500 คน อีกทั้ง เมื่อขนาดกลุ่มตัวอย่าง 500 คน จะส่งผลให้อำนาจการทดสอบสูงกว่าภายใต้ขนาดกลุ่มตัวอย่าง 250 คน

นอกจากนี้ ปัจจัยขนาดกลุ่มตัวอย่างมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก กล่าวคือ เมื่อขนาดกลุ่มตัวอย่าง 1000 คน จะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าภายใต้ขนาดกลุ่มตัวอย่าง 250 คน และ 500 คน อีกทั้ง เมื่อขนาดกลุ่มตัวอย่าง 500 คน จะส่งผลให้อำนาจการทดสอบสูงกว่าภายใต้ขนาดกลุ่มตัวอย่าง 250 คน สำหรับวิธีซิปเทสท์ พบว่า ปัจจัยขนาดกลุ่มตัวอย่างไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีนี้

6. การเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพล ด้วยวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขของปัจจัยที่ศึกษาระดับเดียวกัน พบประเด็นที่สำคัญ ดังนี้

6.1 ปัจจัยค่าความยากของข้อสอบ (b_L b_M และ b_H) พบว่า วิธีซิปเทสท์มีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติกทุกระดับความยากของข้อสอบ สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีซิปเทสท์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีการถดถอยโลจิสติกที่ระดับความยากของข้อสอบ b_Lและ b_H

6.2 ปัจจัยความยาวของแบบสอบ (40 ข้อ และ 60 ข้อ) พบว่า วิธีซิปเทสท์มีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติก ทั้งความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีซิปเทสท์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีการถดถอยโลจิสติก ที่ความยาวของแบบสอบ 60 ข้อ

6.3 ปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน (10% และ 20%) พบว่า วิธีซิปเทสท์มีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติก ทั้งสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 10% และ 20% สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีซิปเทสท์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีการถดถอยโลจิสติก ทั้งสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 10% และ 20%

6.4 ปัจจัยขนาดกลุ่มตัวอย่าง (250 คน 500 คน และ 1000 คน) พบว่า วิธีซิปเทสท์มีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติก ทุกขนาดกลุ่มตัวอย่าง 250 คน 500 คน และ 1000 คน สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีซิปเทสท์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีการถดถอยโลจิสติก ที่ขนาดกลุ่มตัวอย่าง 250 คน ในขณะที่วิธีการถดถอยโลจิสติก มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีซิปเทสท์ ที่ขนาดกลุ่มตัวอย่าง 1000 คน

โดยสรุป ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ได้แก่ ความยากของข้อสอบ ความยาวของแบบสอบ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน และขนาดกลุ่มตัวอย่าง วิธีที่มีอำนาจการทดสอบสูงสุด ได้แก่ วิธีซิปเทสท์ รองลงมา ได้แก่ วิธีราสช์ทรี และวิธีการถดถอยโลจิสติก ตามลำดับ สำหรับวิธีที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ได้แก่ วิธีการถดถอยโลจิสติก รองลงมา ได้แก่ วิธีซิปเทสท์ และวิธีราสช์ทรี ตามลำดับ

ส่วนการการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพล ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ได้แก่ ความยากของข้อสอบ ความยาวของแบบสอบ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน และขนาดกลุ่มตัวอย่าง พบว่า วิธีซิปเทสท์ มีอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีการถดถอยโลจิสติก

สำหรับการการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในข้อมูลเชิงประจักษ์ พบว่า วิธีราสช์ทรี มีอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์

อภิปรายผลการวิจัย

1. การเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ภายใต้เงื่อนไขที่แตกต่างกันของปัจจัยที่ศึกษา ปรากฏผลดังนี้

1.1 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญภายใต้ความยากของข้อสอบต่างกัน 3 ระดับ พบว่า มีปฏิสัมพันธ์ระหว่างความยากของข้อสอบกับวิธีการตรวจสอบ ทำให้ส่งผลต่อการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ กล่าวคือ เมื่อความยากของ

ข้อสอบปานกลาง จะมีผลทำให้อำนาจการทดสอบของวิธีซิปเทสท์มีค่าสูงกว่าความยากของข้อสอบสูง ทั้งนี้อาจเนื่องมาจากข้อสอบที่มีค่าความยากปานกลาง และมีอำนาจจำแนกสูง (การจำลองข้อมูลกำหนดค่า a เท่ากับ 1) เป็นข้อสอบที่มีคุณภาพเหมาะสม ดังนั้น หากข้อสอบเกิดการทำหน้าที่ต่างกันจะทำให้อำนาจการทดสอบของวิธีการตรวจสอบมีค่าสูงสุด สอดคล้องกับการศึกษาของ Narayanan and Swaminathan (1994) ที่พบว่า การตรวจสอบการทำหน้าที่ต่างกันแบบเอกรูปภายใต้เกือบทุกเงื่อนไขลักษณะของข้อสอบที่มีค่าความยากปานกลางและค่าอำนาจจำแนกสูง จะทำให้อำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลและวิธีซิปเทสท์มีค่าสูงใกล้เคียงกัน

เมื่อเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ภายใต้ความยากของข้อสอบ 3 ระดับ พบว่า ภายใต้ความยากของข้อสอบต่ำ และความยากของข้อสอบปานกลาง อำนาจการทดสอบของวิธีซิปเทสท์มีค่าสูงกว่าวิธีการถดถอยโลจิสติกและวิธีราสช์ทรี ส่วนภายใต้ความยากของข้อสอบสูง อำนาจการทดสอบของวิธีซิปเทสท์มีค่าสูงกว่าวิธีการถดถอยโลจิสติก ผลการศึกษาดังกล่าวสอดคล้องกับสมมติฐานที่กำหนดไว้ในข้อ 1 แสดงว่า เมื่อความยากของข้อสอบต่ำ และความยากของข้อสอบปานกลาง เมื่อใช้วิธีซิปเทสท์สามารถตรวจสอบได้ถูกต้องมากกว่าวิธีการถดถอยโลจิสติกและวิธีราสช์ทรี ทั้งนี้อาจเนื่องมาจากในการเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบตามวิธีซิปเทสท์ จะแบ่งข้อสอบออกเป็น 2 ชุดย่อย คือ (1) ชุดแบบสอบที่มีความตรงหรือชุดแบบสอบที่ใช้ในการจับคู่เปรียบเทียบ แบบสอบชุดนี้ประกอบด้วยข้อสอบที่ไม่ทำหน้าที่ต่างกัน และ (2) ชุดแบบสอบที่ต้องการศึกษา ประกอบด้วยข้อสอบที่สงสัยว่าทำหน้าที่ต่างกัน ทำให้การตรวจสอบของวิธีซิปเทสท์มีความถูกต้อง อีกทั้งยังเป็นวิธีที่ถูกออกแบบมาเพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป และข้อสอบทำหน้าที่ต่างกันแบบมีทิศทางเดียว สอดคล้องกับการศึกษาของ Narayanan and Swaminathan (1994) ที่พบว่า การตรวจสอบการทำหน้าที่ต่างกันแบบเอกรูปภายใต้เกือบทุกเงื่อนไขลักษณะของข้อสอบที่มีค่าความยากปานกลางและค่าอำนาจจำแนกสูง จะทำให้อำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลและวิธีซิปเทสท์มีค่าสูงใกล้เคียงกัน

1.2 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญภายใต้ความยาวของแบบสอบต่างกัน 2 ขนาด พบว่า ความยาวของแบบสอบไม่มีผลต่ออำนาจการทดสอบของวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี แสดงว่า ในการตรวจสอบด้วย 3 วิธีดังกล่าว ถ้าใช้ความยาวของแบบสอบ 40 ข้อ จะมีความถูกต้องใกล้เคียงกับการตรวจสอบโดยใช้ความยาว 60 ข้อ กล่าวคือ เมื่อเพิ่มความยาวของแบบสอบ จะไม่มีผลต่ออำนาจการทดสอบของวิธีการตรวจสอบทั้ง 3 วิธี ทั้งนี้อาจเนื่องมาจากความยาวของแบบสอบมีค่าใกล้เคียงกัน จึงทำให้อำนาจการทดสอบมีค่าไม่แตกต่างกันมากนัก ผลการศึกษาดังกล่าวสอดคล้องกับผลการศึกษาของ

กาญจนา วัธนสุนทร (2537) ที่พบว่า ความยาวของแบบสอบไม่มีผลต่ออัตราการตรวจสอบของวิธีแมนเทล-แฮนส์เซล และวิธีซิปเทสท์ แต่ไม่สอดคล้องกับการศึกษาของ Swaminathan and Rogers (1990) โดยใช้ข้อมูลจำลองและตรวจสอบด้วยวิธีแมนเทล-แฮนส์เซล กับวิธีการถดถอยโลจิสติก ใช้แบบสอบที่มีความยาว 40 ข้อ 60 ข้อ และ 80 ข้อ พบว่า เมื่อใช้แบบสอบที่ยาว อำนาจการทดสอบของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจะดีกว่าการใช้แบบสอบที่สั้น

เมื่อเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้ความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ พบว่า ได้ผลการตรวจสอบสอดคล้องกัน กล่าวคือ ภายใต้ทุกเงื่อนไขการตรวจสอบ อำนาจการทดสอบของวิธีซิปเทสท์มีค่าสูงกว่าวิธีการถดถอยโลจิสติกและวิธีราสช์ทรี และอำนาจการทดสอบของวิธีราสช์ทรีมีค่าสูงกว่าวิธีการถดถอยโลจิสติก ภายใต้ความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ ผลการศึกษาดังกล่าวสอดคล้องกับสมมติฐานที่กำหนดไว้ในข้อ 1 แสดงว่า เมื่อความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ เมื่อใช้วิธีซิปเทสท์สามารถตรวจสอบได้ถูกต้องมากกว่าวิธีการถดถอยโลจิสติกและวิธีราสช์ทรี และเมื่อใช้วิธีราสช์ทรีสามารถตรวจสอบได้ถูกต้องมากกว่าวิธีการถดถอยโลจิสติก สอดคล้องกับและการศึกษาของจิติมา วรรณศรี (2539) ที่พบว่า เมื่อใช้แบบสอบความยาว 60 ข้อ จะมีผลทำให้วิธีแมนเทล-แฮนส์เซลและวิธีซิปเทสท์มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันดีที่สุด

1.3 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด พบว่า สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบไม่มีผลต่ออำนาจการทดสอบของวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี แสดงว่า ในการตรวจสอบด้วย 3 วิธีดังกล่าว ถ้าแบบสอบมีสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 จะมีความถูกต้องใกล้เคียงกับการตรวจสอบในแบบสอบที่มีสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 20 กล่าวคือ แม้ว่าสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีขนาดเพิ่มขึ้น จะไม่มีผลต่ออำนาจการทดสอบของวิธีการตรวจสอบทั้ง 3 วิธี ทั้งนี้อาจเนื่องมาจากแบบสอบที่มีความยาว 40 ข้อ และ 60 ข้อ มีสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และร้อยละ 20 คิดเป็นจำนวน 4 ข้อ กับ 8 ข้อ และ 6 ข้อ กับ 12 ข้อ ตามลำดับนั้น ความแตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันมีน้อย ทำให้ไม่ส่งผลต่ออำนาจการทดสอบ ผลการศึกษาดังกล่าวไม่สอดคล้องกับผลการศึกษาของ Rogers and Swaminathan (1993) ที่พบว่า ในการตรวจสอบการทำหน้าที่ต่างกันแบบเอกรูปของวิธีการถดถอยโลจิสติก เมื่อสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันใบแบบสอบมีจำนวนลดลง จะมีผลทำให้อำนาจการทดสอบของวิธีวิธีการถดถอยโลจิสติกมีค่าเพิ่มมากขึ้น และการศึกษาของ Narayanan and Swaminathan (1994) ที่พบว่า การตรวจสอบการทำหน้าที่ต่างกันแบบเอกรูป

เมื่อสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบมีจำนวนลดลง มีผลทำให้อำนาจการทดสอบของวิธีซิปเทสท์และวิธีการถดถอยโลจิสติกมีค่าเพิ่มขึ้น

เมื่อเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และร้อยละ 20 พบว่า วิธีซิปเทสท์มีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติกและวิธีราสช์ทรี ยกเว้น สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 20 อำนาจการทดสอบของวิธีซิปเทสท์และวิธีราสช์ทรีมีค่าใกล้เคียงกัน และอำนาจการทดสอบของวิธีราสช์ทรีมีค่าสูงกว่าวิธีการถดถอยโลจิสติก ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 20 ผลการศึกษาดังกล่าวสอดคล้องกับสมมติฐานที่กำหนดไว้ในข้อ 1 แสดงว่า ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน ร้อยละ 10 เมื่อใช้วิธีซิปเทสท์สามารถตรวจสอบได้ถูกต้องมากกว่าวิธีการถดถอยโลจิสติกและวิธีราสช์ทรี สำหรับภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 20 เมื่อใช้วิธีซิปเทสท์และวิธีราสช์ทรีสามารถตรวจสอบได้ถูกต้องมากกว่าวิธีการถดถอยโลจิสติก

1.4 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด พบว่า ขนาดกลุ่มตัวอย่างมีผลต่ออำนาจการทดสอบของวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี กล่าวคือ เมื่อเพิ่มขนาดกลุ่มตัวอย่างจาก 250 คน เป็น 1000 คน จะมีผลทำให้อำนาจการทดสอบของวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรีมีค่าสูงสุด ผลการศึกษาดังกล่าวสอดคล้องกับผลการศึกษาของ Swaminathan และ Rogers (1990) ที่พบว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและแบบอเนกรูป เมื่อเพิ่มขนาดกลุ่มตัวอย่างจะมีผลทำให้อำนาจการทดสอบของวิธีการถดถอยโลจิสติกมีค่าเพิ่มมากขึ้นเกือบทุกเงื่อนไข และ Narayanan and Swaminathan (1994) ที่ได้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป พบว่า ขนาดกลุ่มตัวอย่างและอัตราส่วนของกลุ่มอ้างอิงต่อกลุ่มเปรียบเทียบมีผลต่ออำนาจการทดสอบของวิธีการตรวจสอบ กล่าวคือ เมื่อเพิ่มขนาดกลุ่มตัวอย่างจะทำให้อำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลและวิธีซิปเทสท์มีค่าเพิ่มมากขึ้น และสอดคล้องกับการศึกษาของ Mazor และคนอื่น ๆ (1992 อ้างถึงใน ทองอยู่ สาระ, 2543) Narayanan and Swaminathan (1996) ได้ข้อค้นพบที่สอดคล้องกัน คือ เมื่อขนาดกลุ่มตัวอย่างใหญ่ขึ้น อำนาจการทดสอบจะเพิ่มขึ้น

เมื่อเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ภายใต้ขนาดกลุ่มตัวอย่าง 3 ขนาด พบว่า ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน 500 คน และ 1000 คน อำนาจการทดสอบของวิธีซิปเทสท์มีค่าสูงกว่าวิธีการถดถอยโลจิสติกและวิธีราสช์ทรี ส่วนภายใต้ขนาดกลุ่มตัวอย่าง 250 คน และ 500 คน อำนาจการทดสอบของวิธีราสช์ทรีมีค่าสูงกว่าวิธีการถดถอยโลจิสติก ผลการศึกษาดังกล่าวสอดคล้องกับสมมติฐานที่กำหนดไว้ในข้อ 1 แสดงว่า เมื่อขนาดกลุ่มตัวอย่าง

250 คน 500 คน และ 1000 คน เมื่อใช้วิธีซิปเทสท์สามารถตรวจสอบได้ถูกต้องมากกว่าวิธีการถดถอยโลจิสติกและวิธีราสช์ทรี ส่วนการตรวจสอบภายใต้ขนาดกลุ่มตัวอย่าง 250 คน และ 500 คน เมื่อใช้วิธีราสช์ทรีสามารถตรวจสอบได้ถูกต้องมากกว่าวิธีการถดถอยโลจิสติก สอดคล้องกับผลการศึกษาของ จิติมา วรรณศรี (2539) พบว่า เมื่อขนาดกลุ่มตัวอย่าง 200 คน และ 600 คน วิธีแมนเทล-แฮนส์เซล และวิธีซิปเทสท์สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ถูกต้อง 50% แต่ถ้าขนาดกลุ่มตัวอย่าง 1000 คน สามารถตรวจสอบได้ถูกต้อง 100%

2. การเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพล ระหว่างวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขที่แตกต่างกันของปัจจัยที่ศึกษา ปรากฏผลดังนี้

2.1 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลภายใต้ความยากของข้อสอบต่างกัน 3 ระดับ พบว่า เมื่อความยากของข้อสอบปานกลาง จะมีผลทำให้อำนาจการทดสอบของวิธีซิปเทสท์มีค่าสูงกว่าความยากของข้อสอบสูง ทั้งนี้อาจเนื่องมาจากข้อสอบที่มีค่าความยากปานกลาง และมีอำนาจจำแนกสูง (การจำลองข้อมูล กำหนดค่า a เท่ากับ 1) เป็นข้อสอบที่มีคุณภาพเหมาะสม ดังนั้น หากข้อสอบเกิดการทำหน้าที่ต่างกัน จะทำให้อำนาจการทดสอบของวิธีการตรวจสอบมีค่าสูงสุด สอดคล้องกับการศึกษาของ Narayanan and Swaminathan (1994) พบว่า การตรวจสอบการทำหน้าที่ต่างกันแบบเอกรูปภายใต้เกือบทุกเงื่อนไขลักษณะของข้อสอบที่มีค่าความยากปานกลางและค่าอำนาจจำแนกสูง จะทำให้อำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลและวิธีซิปเทสท์มีค่าสูงใกล้เคียงกัน

เมื่อเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลระหว่างวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ ภายใต้ความยากของข้อสอบ 3 ระดับ พบว่า ภายใต้ความยากของข้อสอบที่มีค่าความยากต่ำ ค่าความยากปานกลางและค่าความยากสูง อำนาจการทดสอบของวิธีซิปเทสท์มีค่าสูงกว่าวิธีการถดถอยโลจิสติก ผลการศึกษาดังกล่าวสอดคล้องกับสมมติฐานที่กำหนดไว้ในข้อ 1 แสดงว่า เมื่อความยากของข้อสอบที่มีค่าความยากต่ำ ความยากปานกลางและความยากสูง เมื่อใช้วิธีซิปเทสท์สามารถตรวจสอบได้ถูกต้องมากกว่าวิธีการถดถอยโลจิสติก ทั้งนี้อาจเนื่องมาจากในการเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบตามวิธีซิปเทสท์ จะแบ่งข้อสอบออกเป็น 2 ชุดย่อย คือ (1) ชุดแบบสอบที่มีความตรงหรือชุดแบบสอบที่ใช้ในการจับคู่เปรียบเทียบ แบบสอบชุดนี้ประกอบด้วยข้อสอบที่ไม่ทำหน้าที่ต่างกัน และ (2) ชุดแบบสอบที่ต้องการศึกษา ประกอบด้วยข้อสอบที่สงสัยว่าทำหน้าที่ต่างกัน ทำให้การตรวจสอบของวิธีซิปเทสท์มีความถูกต้อง อีกทั้งการพิจารณาข้อสอบที่ทำหน้าที่ต่างกันโดยการวัดขนาดอิทธิพลของวิธีซิปเทสท์จากขนาดอิทธิพลระดับเล็ก ปานกลาง และใหญ่ จึงทำให้อำนาจการ

ทดสอบของวิธีซิปเทสท์ มีค่าสูงกว่าวิธีการถดถอยโลจิสติกที่พิจารณาข้อสอบที่ทำหน้าที่ต่างกันโดยการวัดขนาดอิทธิพลจากขนาดอิทธิพลระดับปานกลาง และใหญ่

2.2 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลภายใต้ความยาวของแบบสอบต่างกัน 2 ขนาด พบว่า ความยาวของแบบสอบไม่มีผลต่ออำนาจการทดสอบของวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ แสดงว่า ในการตรวจสอบด้วย 2 วิธีดังกล่าว ถ้าใช้ความยาวของแบบสอบ 40 ข้อ จะมีความถูกต้องใกล้เคียงกับการตรวจสอบโดยใช้ความยาว 60 ข้อ กล่าวคือ เมื่อเพิ่มความยาวของแบบสอบ จะไม่มีผลต่ออำนาจการทดสอบของวิธีการตรวจสอบทั้ง 2 วิธี ทั้งนี้อาจเนื่องมาจากความยาวของแบบสอบมีค่าใกล้เคียงกัน จึงทำให้อำนาจการทดสอบมีค่าไม่แตกต่างกันมากนัก ผลการศึกษาดังกล่าวสอดคล้องกับผลการศึกษาของ กาญจนา วัธนสุนทร (2537) ที่พบว่าความยาวของแบบสอบไม่มีผลต่ออัตราการตรวจสอบของวิธีแมนเทล-แฮนส์เซล และวิธีซิปเทสท์ แต่ไม่สอดคล้องกับการศึกษาของ Swaminathan and Rogers (1990) ที่ใช้ข้อมูลจำลองและตรวจสอบด้วยวิธีแมนเทล-แฮนส์เซล กับวิธีการถดถอยโลจิสติก กับแบบสอบที่มีความยาว 40 ข้อ 60 ข้อ และ 80 ข้อ พบว่า เมื่อใช้แบบสอบที่ยาว อำนาจการทดสอบของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจะดีกว่าการใช้แบบสอบที่สั้น

เมื่อเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลระหว่างวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ ภายใต้ความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ พบว่า ได้ผลการตรวจสอบสอดคล้องกัน กล่าวคือ ภายใต้ทุกเงื่อนไขการตรวจสอบ อำนาจการทดสอบของวิธีซิปเทสท์มีค่าสูงกว่าวิธีการถดถอยโลจิสติก ผลการศึกษาดังกล่าวสอดคล้องกับสมมติฐานที่กำหนดไว้ในข้อ 1 แสดงว่า เมื่อความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ เมื่อใช้วิธีซิปเทสท์สามารถตรวจสอบได้ถูกต้องมากกว่าวิธีการถดถอยโลจิสติก ทั้งนี้อาจเนื่องมาจากการพิจารณาข้อสอบที่ทำหน้าที่ต่างกันโดยการวัดขนาดอิทธิพลของวิธีซิปเทสท์จากขนาดอิทธิพลระดับเล็ก ปานกลาง และใหญ่ จึงทำให้อำนาจการทดสอบของวิธีซิปเทสท์ มีค่าสูงกว่าวิธีการถดถอยโลจิสติกที่พิจารณาข้อสอบที่ทำหน้าที่ต่างกันโดยการวัดขนาดอิทธิพลจากขนาดอิทธิพลระดับปานกลาง และใหญ่ ผลการศึกษาดังกล่าวสอดคล้องกับและการศึกษาของ จิติมา วรรณศรี (2539) ที่พบว่า เมื่อใช้แบบสอบความยาว 60 ข้อ จะมีผลทำให้วิธีแมนเทล-แฮนส์เซลและวิธีซิปเทสท์มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันดีที่สุด

2.3 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด พบว่า สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบไม่มีผลต่ออำนาจการทดสอบของวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ แสดงว่า ในการตรวจสอบด้วย 2 วิธีดังกล่าว ถ้าแบบสอบมีสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน ร้อยละ 10 จะมีความถูกต้องใกล้เคียงกับการตรวจสอบในแบบสอบที่มีสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 20 กล่าวคือ

แม้ว่าสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีขนาดเพิ่มขึ้น จะไม่มีผลต่ออำนาจการทดสอบของวิธีการตรวจสอบทั้ง 2 วิธี ทั้งนี้อาจเนื่องมาจากแบบสอบที่มีความยาว 40 ข้อ และ 60 ข้อ มีสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และร้อยละ 20 คิดเป็นจำนวน 4 ข้อ กับ 8 ข้อ และ 6 ข้อ กับ 12 ข้อ ตามลำดับนั้น ความแตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันมีน้อย ทำให้ไม่ส่งผลต่ออำนาจการทดสอบ ผลการศึกษาดังกล่าวไม่สอดคล้องกับผลการศึกษาของ Rogers and Swaminathan (1993) ที่พบว่า ในการตรวจสอบการทำหน้าที่ต่างกันแบบเอกรูปของวิธีการถดถอยโลจิสติก เมื่อสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันใบแบบสอบมีจำนวนลดลง จะมีผลทำให้อำนาจการทดสอบของวิธีวิธีการถดถอยโลจิสติกมีค่าเพิ่มมากขึ้น และการศึกษาของ Narayanan and Swaminathan (1994) ที่พบว่า การตรวจสอบการทำหน้าที่ต่างกันแบบเอกรูป เมื่อสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบมีจำนวนลดลง มีผลทำให้อำนาจการทดสอบของวิธีซิปเทสท์และวิธีการถดถอยโลจิสติกมีค่าเพิ่มขึ้น

เมื่อเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลระหว่างวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน พบว่า ได้ผลการตรวจสอบสอดคล้องกัน กล่าวคือ ภายใต้ทุกเงื่อนไขการตรวจสอบอำนาจการทดสอบของวิธีซิปเทสท์มีค่าสูงกว่าวิธีการถดถอยโลจิสติก ผลการศึกษาดังกล่าวสอดคล้องกับสมมติฐานที่กำหนดไว้ในข้อ 1 แสดงว่า เมื่อสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และร้อยละ 20 เมื่อใช้วิธีซิปเทสท์สามารถตรวจสอบได้ถูกต้องมากกว่าวิธีการถดถอยโลจิสติก

2.4 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด พบว่า ขนาดกลุ่มตัวอย่างมีผลต่ออำนาจการทดสอบของวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ กล่าวคือ เมื่อขนาดกลุ่มตัวอย่างจาก 250 คน และ 500 คน เป็น 1000 คน จะมีผลทำให้อำนาจการทดสอบของวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์มีค่าสูงสุด ผลการศึกษาดังกล่าวสอดคล้องกับผลการศึกษาของ Swaminathan and Rogers (1990) ที่พบว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและแบบอเนกรูป เมื่อเพิ่มขนาดกลุ่มตัวอย่างจะมีผลทำให้อำนาจการทดสอบของวิธีการถดถอยโลจิสติกมีค่าเพิ่มมากขึ้นเกือบทุกเงื่อนไข และ Narayanan and Swaminathan (1994) ที่ได้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป พบว่า ขนาดกลุ่มตัวอย่างและอัตราส่วนของกลุ่มอ้างอิงต่อกลุ่มเปรียบเทียบมีผลต่ออำนาจการทดสอบของวิธีการตรวจสอบ กล่าวคือ เมื่อเพิ่มขนาดกลุ่มตัวอย่างจะทำให้อำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลและวิธีซิปเทสท์มีค่าเพิ่มมากขึ้น และสอดคล้องกับการศึกษาของ Mazor และคนอื่น ๆ (1992 อ้างถึงใน ทองอยู่ สาระ, 2543) Narayanan and Swaminathan (1996) ได้ข้อค้นพบที่สอดคล้องกัน คือ เมื่อขนาดกลุ่มตัวอย่างใหญ่ขึ้น อำนาจการทดสอบจะเพิ่มขึ้น

เมื่อเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลระหว่างวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ ภายใต้ขนาดกลุ่มตัวอย่าง 3 ขนาด พบว่า ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน 500 คน และ 1000 คน อำนาจการทดสอบของวิธีซิปเทสท์มีค่าสูงกว่าวิธีการถดถอยโลจิสติก ผลการศึกษาดังกล่าวสอดคล้องกับสมมติฐานที่กำหนดไว้ในข้อ 1 แสดงว่า เมื่อขนาดกลุ่มตัวอย่าง 250 คน 500 คน และ 1000 คน เมื่อใช้วิธีซิปเทสท์สามารถตรวจสอบได้ถูกต้องมากกว่าวิธีการถดถอยโลจิสติก สอดคล้องกับผลการศึกษาของ จิติมา วรรณศรี (2539) พบว่า เมื่อขนาดกลุ่มตัวอย่าง 200 คน และ 600 คน วิธีแมนเทล-แฮนส์เซล และวิธีซิปเทสท์สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ถูกต้อง 50% แต่ถ้าขนาดกลุ่มตัวอย่าง 1000 คน สามารถตรวจสอบได้ถูกต้อง 100%

3. การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ภายใต้เงื่อนไขที่แตกต่างกันของปัจจัยที่ศึกษา ปรากฏผลดังนี้

3.1 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญภายใต้ความยากของข้อสอบต่างกัน 3 ระดับ พบว่า ความยากของข้อสอบมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก กล่าวคือ เมื่อความยากของข้อสอบต่ำจะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าภายใต้เงื่อนไขความยากของข้อสอบปานกลาง ในขณะที่ความยากของข้อสอบไม่มีผลต่ออำนาจการทดสอบของวิธีซิปเทสท์ สอดคล้องกับการศึกษาของ Narayanan and Swaminathan (1994) ที่พบว่า การตรวจสอบการทำหน้าที่ต่างกันแบบเอกรูป ลักษณะของข้อสอบไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซลและวิธีซิปเทสท์

เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ภายใต้ความยากของข้อสอบ 3 ระดับ พบว่า ภายใต้ความยากของข้อสอบต่ำ ค่าความยากปานกลาง และค่าความยากสูง อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์มีค่าสูงกว่าวิธีราสช์ทรี ส่วนภายใต้ความยากของข้อสอบต่ำ อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าวิธีซิปเทสท์ ผลการศึกษาดังกล่าวสอดคล้องกับสมมติฐานที่กำหนดไว้ในข้อ 2 แสดงว่า เมื่อความยากของข้อสอบต่ำ ความยากปานกลาง และความยากสูง เมื่อใช้วิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ตรวจสอบจะมีความคลาดเคลื่อนมากกว่าวิธีราสช์ทรี และเมื่อความยากของข้อสอบต่ำ เมื่อใช้วิธีการถดถอยโลจิสติกตรวจสอบจะมีความคลาดเคลื่อนมากกว่าวิธีซิปเทสท์

3.2 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญภายใต้ความยาวของแบบสอบต่างกัน 2 ขนาด พบว่า ความยาวของแบบสอบไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี แสดงว่า ในการตรวจสอบด้วย 3 วิธีดังกล่าว ถ้าใช้ความยาวของแบบสอบ 40 ข้อ จะมีความคลาดเคลื่อนใกล้เคียงกับการตรวจสอบโดยใช้ความยาว 60 ข้อ กล่าวคือ เมื่อเพิ่มความยาวของแบบสอบ จะไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบทั้ง 3 วิธี

เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้ความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ พบว่า ได้ผลการตรวจสอบสอดคล้องกัน กล่าวคือ ภายใต้ทุกเงื่อนไขการตรวจสอบ อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์มีค่าสูงกว่าวิธีราสช์ทรี และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าวิธีซิปเทสท์ ผลการศึกษาดังกล่าวสอดคล้องกับสมมติฐานที่กำหนดไว้ในข้อ 2 แสดงว่า เมื่อความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ เมื่อใช้วิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ตรวจสอบจะมีความคลาดเคลื่อนมากกว่าวิธีราสช์ทรี และเมื่อใช้วิธีการถดถอยโลจิสติกตรวจสอบจะมีความคลาดเคลื่อนมากกว่าวิธีซิปเทสท์

3.3 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด พบว่า สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี แสดงว่า ในการตรวจสอบด้วย 3 วิธีดังกล่าว ถ้าแบบสอบมีสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน ร้อยละ 10 จะมีความคลาดเคลื่อนใกล้เคียงกับการตรวจสอบในแบบสอบที่มีสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 20 กล่าวคือ แม้ว่าสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีขนาดเพิ่มขึ้น จะไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบทั้ง 3 วิธี ทั้งนี้อาจเนื่องมาจากแบบสอบที่มีความยาว 40 ข้อ และ 60 ข้อ มีสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และร้อยละ 20 คิดเป็นจำนวน 4 ข้อ กับ 8 ข้อ และ 6 ข้อ กับ 12 ข้อ ตามลำดับนั้น ความแตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันมีน้อย ทำให้ไม่ส่งผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ผลการศึกษาดังกล่าวไม่สอดคล้องกับ Narayanan and Swaminathan (1994) ที่พบว่า การตรวจสอบการทำหน้าที่ต่างกันแบบเอกรูป เมื่อสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบมีจำนวนลดลง มีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์และวิธีการถดถอยโลจิสติกมีค่าเพิ่มขึ้น

เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธี

ราสช์ทรี ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน พบว่า ได้ผลการตรวจสอบสอดคล้องกัน กล่าวคือ ภายใต้ทุกเงื่อนไขการตรวจสอบ อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์มีค่าสูงกว่าวิธีราสช์ทรี ผลการศึกษาดังกล่าวสอดคล้องกับสมมติฐานที่กำหนดไว้ในข้อ 2 แสดงว่า เมื่อสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และร้อยละ 20 เมื่อใช้วิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ตรวจสอบจะมีความคลาดเคลื่อนมากกว่าวิธีราสช์ทรี ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และร้อยละ 20 ของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าวิธีซิปเทสท์ ผลการศึกษาดังกล่าวสอดคล้องกับสมมติฐานที่กำหนดไว้ในข้อ 2 แสดงว่า เมื่อแบบสอบมีสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และร้อยละ 20 ใช้การตรวจสอบด้วยวิธีการถดถอยโลจิสติกจะมีความคลาดเคลื่อนมากกว่าวิธีซิปเทสท์

3.4 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด พบว่า ขนาดกลุ่มตัวอย่างมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีราสช์ทรี กล่าวคือ ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน และ 500 คน จะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน สำหรับวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ได้ผลสอดคล้องกัน คือ ขนาดกลุ่มตัวอย่างไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1

เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี ภายใต้ขนาดกลุ่มตัวอย่าง 3 ขนาด พบว่า ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน 500 คน และ 1000 คน อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์มีค่าสูงกว่าวิธีราสช์ทรี ส่วนภายใต้ขนาดกลุ่มตัวอย่าง 500 คน และ 1000 คน อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าวิธีซิปเทสท์ ผลการศึกษาดังกล่าวสอดคล้องกับสมมติฐานที่กำหนดไว้ในข้อ 2 แสดงว่า เมื่อขนาดกลุ่มตัวอย่าง 250 คน 500 คน และ 1000 คน เมื่อใช้วิธีการถดถอยโลจิสติกและวิธีซิปเทสท์มีความคลาดเคลื่อนมากกว่าวิธีราสช์ทรี และภายใต้ขนาดกลุ่มตัวอย่าง 500 คน และ 1000 คน เมื่อใช้วิธีการถดถอยโลจิสติกมีความคลาดเคลื่อนมากกว่าวิธีซิปเทสท์

4. การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพล ระหว่างวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ ภายใต้เงื่อนไขที่แตกต่างกันของปัจจัยที่ศึกษา ปรากฏผลดังนี้

4.1 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลภายใต้ความยากของข้อสอบต่างกัน 3 ระดับ พบว่า ความยากของข้อสอบไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ แสดงว่า ในการตรวจสอบด้วย 2 วิธีดังกล่าว เมื่อความยากของข้อสอบมีค่าความยากเพิ่มขึ้น จะไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบทั้ง 2 วิธี สอดคล้องกับการศึกษาของ Narayanan and Swaminathan (1994) พบว่า การตรวจสอบการทำหน้าที่ต่างกันแบบเอกรูปภายใต้เกือบทุกเงื่อนไขลักษณะของข้อสอบที่มีค่าความยากปานกลางและค่าอำนาจจำแนกสูง ลักษณะของข้อสอบไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซลและวิธีซิปเทสท์

เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลระหว่างวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ ภายใต้ความยากของข้อสอบ 3 ระดับ พบว่า ภายใต้ความยากของข้อสอบที่มีค่าความยากต่ำ และค่าความยากสูง อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์มีค่าสูงกว่าวิธีการถดถอยโลจิสติก ผลการศึกษาดังกล่าวสอดคล้องกับสมมติฐานที่กำหนดไว้ในข้อ 2 แสดงว่า เมื่อความยากของข้อสอบที่มีค่าความยากต่ำและค่าความยากสูง เมื่อใช้วิธีซิปเทสท์จะมีความคลาดเคลื่อนมากกว่าวิธีการถดถอยโลจิสติก ส่วนภายใต้ความยากของข้อสอบที่มีค่าความยากปานกลาง อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์มีค่าใกล้เคียงกัน ผลการศึกษาดังกล่าวไม่สอดคล้องกับสมมติฐานที่กำหนดไว้ในข้อ 2 แสดงว่า เมื่อความยากของข้อสอบที่มีค่าความยากปานกลาง เมื่อใช้วิธีการถดถอยโลจิสติกหรือวิธีซิปเทสท์ก็จะมีความคลาดเคลื่อนใกล้เคียงกัน

4.2 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลภายใต้ความยาวของแบบสอบต่างกัน 2 ขนาด พบว่า ความยาวของแบบสอบไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ แสดงว่า ในการตรวจสอบด้วย 2 วิธีดังกล่าว ถ้าใช้ความยาวของแบบสอบ 40 ข้อ จะมีความคลาดเคลื่อนใกล้เคียงกับการตรวจสอบโดยใช้ความยาว 60 ข้อ กล่าวคือ เมื่อเพิ่มความยาวของแบบสอบ จะไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบทั้ง 2 วิธี

เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลระหว่างวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ ภายใต้ความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ พบว่า ภายใต้ความยาวของแบบสอบ 40 ข้อ อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์มีค่าไม่ต่างกัน ผลการศึกษาดังกล่าวไม่สอดคล้องกับสมมติฐานที่กำหนดไว้ในข้อ 2 แสดงว่า เมื่อความยาวของแบบสอบ 40 ข้อ เมื่อใช้วิธีการถดถอยโลจิสติกหรือวิธีซิปเทสท์ก็จะมีความคลาดเคลื่อนใกล้เคียงกัน ส่วนภายใต้ความยาวของแบบสอบ 60 ข้อ อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์มีค่าสูงกว่า

วิธีการถดถอยโลจิสติก ผลการศึกษาดังกล่าวสอดคล้องกับสมมติฐานที่กำหนดไว้ในข้อ 2 แสดงว่า เมื่อความยาวของแบบสอบ 60 ข้อ เมื่อใช้วิธีซิปเทสท์จะมีความคลาดเคลื่อนมากกว่าวิธีการถดถอยโลจิสติก

4.3 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด พบว่า สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ แสดงว่า ในการตรวจสอบด้วย 2 วิธีดังกล่าว ถ้าแบบสอบมีสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน ร้อยละ 10 จะมีความคลาดเคลื่อนใกล้เคียงกับการตรวจสอบในแบบสอบที่มีสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 20 กล่าวคือ แม้ว่าสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมีขนาดเพิ่มขึ้น จะไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบทั้ง 2 วิธี ผลการศึกษาดังกล่าวไม่สอดคล้องกับ Narayanan and Swaminathan (1994) ที่พบว่า การตรวจสอบการทำหน้าที่ต่างกันแบบเอกรูป เมื่อสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบมีจำนวนลดลง มีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์และวิธีการถดถอยโลจิสติกมีค่าเพิ่มขึ้น

เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลระหว่างวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน พบว่า ได้ผลการตรวจสอบสอดคล้องกัน กล่าวคือ ภายใต้ทุกเงื่อนไขการตรวจสอบ อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์มีค่าสูงกว่าวิธีการถดถอยโลจิสติก ผลการศึกษาดังกล่าวสอดคล้องกับสมมติฐานที่กำหนดไว้ในข้อ 2 แสดงว่า เมื่อสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และร้อยละ 20 เมื่อใช้วิธีซิปเทสท์จะมีความคลาดเคลื่อนมากกว่าวิธีการถดถอยโลจิสติก

4.4 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลภายใต้ขนาดกลุ่มตัวอย่างต่างกัน 3 ขนาด พบว่า ขนาดกลุ่มตัวอย่างมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก กล่าวคือ เมื่อเพิ่มขนาดกลุ่มตัวอย่าง จะมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าเพิ่มขึ้น ในขณะที่ขนาดกลุ่มตัวอย่างไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์

เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวัดขนาดอิทธิพลระหว่างวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ ภายใต้ขนาดกลุ่มตัวอย่าง 3 ขนาด พบว่า ภายใต้ขนาดกลุ่มตัวอย่าง 250 คน อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์มีค่าสูงกว่าวิธีการถดถอยโลจิสติก ในขณะที่ภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าวิธีซิปเทสท์ ผลการศึกษาดังกล่าวสอดคล้องกับสมมติฐานที่กำหนดไว้ในข้อ 2 แสดงว่า เมื่อขนาดกลุ่มตัวอย่าง

250 คน เมื่อใช้วิธีซิปเทสท์จะมีความคลาดเคลื่อนมากกว่าวิธีการถดถอยโลจิสติก แต่เมื่อขนาดกลุ่มตัวอย่าง 1000 คน เมื่อใช้วิธีการถดถอยโลจิสติกจะมีความคลาดเคลื่อนมากกว่าวิธีซิปเทสท์ ส่วนภายใต้ขนาดกลุ่มตัวอย่าง 500 คน อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ มีค่าไม่ต่างกัน ผลการศึกษาดังกล่าวไม่สอดคล้องกับสมมติฐานที่กำหนดไว้ในข้อ 2 แสดงว่า เมื่อขนาดกลุ่มตัวอย่าง 500 คน เมื่อใช้วิธีการถดถอยโลจิสติกหรือวิธีซิปเทสท์ก็จะมีความคลาดเคลื่อนใกล้เคียงกัน

5. การเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรีในข้อมูลเชิงประจักษ์ ปรากฏผลดังนี้

5.1 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ผลการตรวจสอบซึ่งใช้วิธีการวัดพื้นที่ของราชูเป็นวิธีเกณฑ์และเปรียบเทียบผลการตรวจสอบระหว่างวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ในแบบสอบด้านการรู้เรื่องการอ่าน ฉบับที่ 1, 2, 4, 5, 6, 7, 8, 9, 11 และ 13 พบว่า วิธีการถดถอยโลจิสติกและวิธีซิปเทสท์พบข้อสอบที่ทำหน้าที่ต่างกันในปริมาณที่ใกล้เคียงกัน ส่วนวิธีราสช์ทรีพบข้อสอบที่ทำหน้าที่ต่างกันในปริมาณที่สูงกว่า 2 วิธีข้างต้น สำหรับการตรวจพบข้อสอบที่ทำหน้าที่ต่างกันตรงกับผลการตรวจด้วยวิธีเกณฑ์ วิธีราสช์ทรีตรวจพบข้อสอบที่ทำหน้าที่ต่างกันตรงกับผลการตรวจด้วยวิธีเกณฑ์สูงสุด รองลงมา ได้แก่ วิธีซิปเทสท์ และวิธีการถดถอยโลจิสติกตรวจพบข้อสอบที่ทำหน้าที่ต่างกันต่ำสุด ส่วนการตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน ทั้งที่ความเป็นจริงทำหน้าที่ไม่ต่างกัน วิธีราสช์ทรีตรวจพบในปริมาณสูงสุด รองลงมาได้แก่ วิธีการถดถอยโลจิสติก และวิธีซิปเทสท์ตรวจพบในปริมาณต่ำสุด

5.2 การเปรียบเทียบประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

อำนาจการทดสอบของวิธีการถดถอยโลจิสติก ประมาณร้อยละ 27.03 วิธีซิปเทสท์ ประมาณร้อยละ 54.05 และวิธีราสช์ทรี ประมาณร้อยละ 91.89 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก ประมาณร้อยละ 11.90 วิธีซิปเทสท์ ประมาณร้อยละ 7.74 และวิธีราสช์ทรี ประมาณร้อยละ 13.10 จากผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีการตรวจสอบทั้ง 3 วิธี พบว่า อำนาจการทดสอบของวิธีราสช์ทรีมีค่าสูงสุด แต่ในขณะเดียวกันอัตราความคลาดเคลื่อนประเภทที่ 1 ก็มีค่าสูงสุดเช่นเดียวกัน ซึ่งเกินเกณฑ์ที่กำหนดที่ระดับร้อยละ 10 อาจเนื่องมาจากเกณฑ์การตัดสินข้อสอบที่ทำหน้าที่ต่างกันของวิธีราสช์ทรี ทำให้ข้อสอบที่ระบุว่าทำหน้าที่ต่างกันรวมข้อที่มีขนาดอิทธิพลระดับเล็กไว้ด้วย จึงทำให้มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าการศึกษาในข้อมูลจำลอง สำหรับอำนาจการทดสอบของวิธีซิปเทสท์มีค่าอยู่ในระดับปานกลาง และมีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในเกณฑ์ที่กำหนดที่

ระดับร้อยละ 10 โดยร้อยละของอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 สอดคล้องกับการตรวจสอบในข้อมูลจำลอง ส่วนอำนาจการทดสอบวิธีการถดถอยโลจิสติกมีค่าต่ำสุดและมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีซิปเทสท์ จึงกล่าวได้ว่า วิธีซิปเทสท์มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบดีกว่าวิธีการถดถอยโลจิสติกและวิธีราสช์ทรี

6. การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างข้อมูลจำลองกับข้อมูลเชิงประจักษ์

อำนาจการทดสอบของวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในข้อมูลจำลอง พบว่า วิธีซิปเทสท์มีอำนาจการทดสอบสูงสุด รองลงมา ได้แก่ วิธีราสช์ทรี และวิธีการถดถอยโลจิสติก ตามลำดับ ในขณะที่อำนาจการทดสอบของวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในข้อมูลเชิงประจักษ์ พบว่า วิธีราสช์ทรีมีอำนาจการทดสอบสูงสุด รองลงมา ได้แก่ วิธีซิปเทสท์ และวิธีการถดถอยโลจิสติก ตามลำดับ

อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในข้อมูลจำลอง พบว่า วิธีการถดถอยโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด รองลงมา ได้แก่ วิธีซิปเทสท์ และวิธีราสช์ทรี ตามลำดับ ในขณะที่อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในข้อมูลเชิงประจักษ์ พบว่า วิธีราสช์ทรี มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด รองลงมา ได้แก่ วิธีการถดถอยโลจิสติก และวิธีซิปเทสท์

Chulalongkorn University

จากผลดังกล่าว แสดงว่า อำนาจการทดสอบของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในข้อมูลจำลองและข้อมูลเชิงประจักษ์ สอดคล้องกันที่วิธีการถดถอยโลจิสติกมีอำนาจการทดสอบต่ำสุด ในวิธีการตรวจสอบทั้ง 3 วิธี แต่ไม่สอดคล้องกันที่วิธีซิปเทสท์เหมาะกับข้อมูลจำลองที่กำหนดสถานการณ์ตามเงื่อนไขที่ซับซ้อน ในขณะที่วิธีราสช์ทรีเหมาะกับข้อมูลเชิงประจักษ์ที่ใช้ในสถานการณ์การสอบจริง สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในข้อมูลจำลองและข้อมูลเชิงประจักษ์ ไม่สอดคล้องกันที่วิธีราสช์ทรีมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุดในข้อมูลจำลอง แต่กลับมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุดในข้อมูลเชิงประจักษ์ ทั้งนี้อาจเนื่องมาจากเงื่อนไขของปัจจัยที่ศึกษาในข้อมูลจำลองกับข้อมูลเชิงประจักษ์แตกต่างกัน กล่าวคือ ในข้อมูลจำลองจัดกระทำข้อมูลตามปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ได้แก่ ความยากของข้อสอบ 3 ระดับ คือ ค่าความยากต่ำ (b_L) ค่าความยากปานกลาง (b_M) และค่าความยากสูง (b_H) ความยาวของแบบสอบ 2 ขนาด คือ 40 ข้อ และ 60 ข้อ สัดส่วนของข้อสอบที่ทำหน้าที่

ต่างกัน 2 ขนาด คือ ร้อยละ 10 และร้อยละ 20 และขนาดกลุ่มตัวอย่าง 3 ขนาด คือ 250 คน 500 คน และ 1000 คน ส่วนข้อมูลเชิงประจักษ์ศึกษาการทำหน้าที่ต่างกันของข้อสอบตามตัวแปร 3 ตัว คือ เพศ สังกัดของโรงเรียน และภูมิภาค ซึ่งค่าความยากของข้อสอบ ค่าอำนาจจำแนกของข้อสอบ ความยาวของแบบสอบ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน และจำนวนกลุ่มตัวอย่าง อยู่ในสภาพจริง จึงเกิดความแตกต่างของปัจจัยที่ศึกษาขึ้น รวมถึงข้อมูลจำลองกำหนดการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป ในขณะที่ข้อมูลเชิงประจักษ์เป็นการศึกษาข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูปและแบบอเนกรูป

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1.1 จากผลการวิจัย พบว่า ความยากของข้อสอบมีผลต่ออำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญของวิธีซิปเทสท์ กล่าวคือ เมื่อความยากของข้อสอบที่มีค่าความยากปานกลาง จะมีผลทำให้อำนาจการทดสอบของซิปเทสท์มีค่าสูงกว่าข้อสอบที่มีค่าความยากสูง แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญภายใต้ความยากของข้อสอบที่มีค่าความยากปานกลาง วิธีซิปเทสท์สามารถระบุข้อสอบที่ทำหน้าที่ต่างกันได้ถูกต้องแม่นยำ ดังนั้น เมื่อสถานศึกษาหรือหน่วยงานทางการศึกษาได้สร้างข้อสอบขึ้น และได้วิเคราะห์คุณภาพเบื้องต้นของข้อสอบแล้วว่า แบบสอบฉบับนั้นมีค่าความยากปานกลาง สามารถนำวิธีซิปเทสท์ไปจะวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบที่สร้างขึ้นได้ ซึ่งจะให้ผลที่มีความถูกต้องแม่นยำ

1.2 จากผลการวิจัย พบว่า ความยาวของแบบสอบไม่มีผลต่ออำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญของวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี กล่าวคือ อำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบทั้ง 3 วิธี ภายใต้ความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ มีค่าใกล้เคียงกัน แสดงว่า เมื่อตรวจสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้ความยาวของแบบสอบ 40 ข้อ และ 60 ข้อ สามารถระบุข้อสอบที่ทำหน้าที่ต่างกันได้ใกล้เคียงกัน หรืออาจระบุผิดพลาดว่าข้อสอบทำหน้าที่ต่างกัน (ทั้งที่ความเป็นจริงข้อสอบทำหน้าที่ไม่ต่างกัน) ได้ใกล้เคียงกัน ดังนั้น ผู้ทำวิจัย สถานศึกษาหรือหน่วยงานทางการศึกษาที่ต้องการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สามารถสร้างแบบสอบที่ยาว 40 ข้อ ก็มีความเพียงพอที่จะนำมาใช้ศึกษาการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีวิธีการถดถอยโลจิสติก

วิธีซิปเทสท์ หรือวิธีราสช์ทรี ซึ่งจะให้ความถูกต้องในการระบุข้อสอบที่ทำหน้าที่ต่างกันหรือการระบุผิดพลาดว่าข้อสอบทำหน้าที่ไม่ต่างกัน ได้ใกล้เคียงกับแบบสอบที่ยาว 60 ข้อ

1.3 จากผลการวิจัย พบว่า สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันไม่มีผลต่อผลต่ออำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญของวิธีการถดถอยโลจิสติก วิธีซิปเทสท์และวิธีราสช์ทรี กล่าวคือ อำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบทั้ง 3 วิธี ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และร้อยละ 20 มีค่าใกล้เคียงกัน แสดงว่า เมื่อตรวจสอบด้วยวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ และวิธีราสช์ทรี ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 และร้อยละ 20 สามารถระบุข้อสอบที่ทำหน้าที่ต่างกันได้ใกล้เคียงกัน หรืออาจระบุผิดพลาดว่าข้อสอบทำหน้าที่ต่างกัน (ทั้งที่ความเป็นจริงข้อสอบทำหน้าที่ไม่ต่างกัน) ได้ใกล้เคียงกัน ดังนั้น ผู้ทำวิจัย สถานศึกษาหรือหน่วยงานทางการศึกษาที่ต้องการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบกับแบบสอบที่มีสัดส่วนของการทำหน้าที่ต่างกันร้อยละ 10 ก็มีความเพียงพอที่จะนำมาใช้ศึกษาการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีวิธีการถดถอยโลจิสติก วิธีซิปเทสท์ หรือวิธีราสช์ทรี ซึ่งจะให้ความถูกต้องในการระบุข้อสอบที่ทำหน้าที่ต่างกันหรือการระบุผิดพลาดว่าข้อสอบทำหน้าที่ไม่ต่างกัน ได้ใกล้เคียงกับแบบสอบที่มีสัดส่วนของการทำหน้าที่ต่างกันร้อยละ 20

1.4 จากผลการวิจัย พบว่า ขนาดกลุ่มตัวอย่างมีผลต่ออำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญของวิธีราสช์ทรี กล่าวคือ เมื่อขนาดกลุ่มตัวอย่าง 1000 คน จะมีผลทำให้อำนาจการทดสอบของวิธีราสช์ทรีมีค่าสูงขึ้น และมีอัตราความคลาดเคลื่อนประเภทที่ 1 ลดลง แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญภายใต้ขนาดกลุ่มตัวอย่าง 1000 คน วิธีราสช์ทรีสามารถระบุข้อสอบที่ทำหน้าที่ต่างกันได้ถูกต้องแม่นยำ หรืออาจระบุผิดพลาดว่าข้อสอบทำหน้าที่ต่างกัน (ทั้งที่ความเป็นจริงข้อสอบทำหน้าที่ไม่ต่างกัน) ได้น้อยมาก ดังนั้น หน่วยงานทางการศึกษาที่สร้างข้อสอบเพื่อใช้กับกลุ่มผู้สอบจำนวนมาก ตั้งแต่ 1000 คน ขึ้นไป และต้องการนำมาตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยมีปัจจัยที่ศึกษาคล้ายคลึงกับการวิจัยครั้งนี้ สามารถเลือกใช้วิธีราสช์ทรีในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่สร้างขึ้นได้ ซึ่งจะให้ความถูกต้องในการระบุข้อสอบที่ทำหน้าที่ต่างกันหรือการระบุผิดพลาดว่าข้อสอบทำหน้าที่ไม่ต่างกันได้น้อยมาก

2. ข้อเสนอแนะในการวิจัยต่อไป

2.1 จากการวิจัยครั้งนี้ พบว่า อำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีราสช์ทรีในข้อมูลจำลองกับข้อมูลเชิงประจักษ์ได้ผลไม่สอดคล้องกัน ทั้งนี้อาจเนื่องมาจากเกณฑ์ในการตัดสินการทำหน้าที่ต่างกันของข้อสอบที่เป็นข้อตกลงเบื้องต้นในครั้งนี้ ดังนั้น ควรมี

การศึกษาเพิ่มเติมเพื่อพัฒนาเกณฑ์ในการตัดสินข้อสอบที่ทำหน้าที่ต่างกันให้มีความชัดเจน ถูกต้องแม่นยำขึ้น เช่น พัฒนาเกณฑ์โดยเปรียบเทียบผลการตรวจสอบกับวิธีการวัดพื้นที่ของราชู

2.2 ควรมีการศึกษาเพิ่มเติมเกี่ยวกับประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยศึกษากรณีรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป และแบบอเนกรูปทั้งกรณีของข้อสอบที่ทำหน้าที่ต่างกันแบบมีทิศทางเดียว (unidirectional) และข้อสอบที่ทำหน้าที่ต่างกันแบบไม่มีทิศทาง (non-unidirectional) เพื่อขยายผลการศึกษาประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของวิธีราสช์ทรีซึ่งเป็นวิธีการตรวจสอบใหม่ให้กว้างขวางยิ่งขึ้น

2.3 ควรมีการศึกษาเพิ่มเติมเกี่ยวกับประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยศึกษากรณีแบบสอบที่มีการตรวจให้คะแนนแบบพหุวิภาค หรือให้คะแนนแบบหลายค่า (polytomous) นอกจากนี้ ควรศึกษาเพิ่มเติมในกรณีแบบสอบพหุมิติ (multidimension)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

รายการอ้างอิง

Angoff, W. H. (1993). Perspectives on Differential Item Functioning Methodology. In P. W. Holland & H. Wainer (Eds.), *Differential Item Functioning* (pp. 3-5). New Jersey: Lawrence Erlbaum Associates.

Borsboom, D., Mellenbergh, G. J., & Heerden, J. v. (2002). Different Kinds of DIF: A Distinction Between Absolute and Relative Forms of Measurement Invaiance and Bias. *Applied Psychological Measurement,* 26(4), 433-450.

Brennan, R. L. (2006). *Educational Measurement* (4th ed.). Connecticut: Praeger.

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

Clauser, R. E., & Mazor, K. M. (1998). Using statistical procedures to identify Differentially functioning test items. *Educational Measurement: Issues and Practice,* 17(1), 31-44.

DeMars, C. E. (2009). Modification of the Mantel-Haenszel and Logistic Regression DIF Procedures to Incorporate the SIBTEST Regression Correction. *Journal of Educational and Behavioral Statistics,* 34(2), 149-170.

DeMars, C. E., & Lau, A. (2011). Differential Item Functioning Detection With Latent Classes: How Accurately Can We Detect Who Is Responding Differentially?. *Educational and Psychological Measurement,* 71(4), 597-616.

Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory For Psychologists*. New Jersey: Lawrence Erlbaum Associates.

Escorial, S., & Navas, M. J. (2007). Analysis of the Gender Variable in the Eysenck Personality Questionnaire-Revised Scales Using Differential Item Functioning Techniques. *Educational and Psychological Measurement,* 67(6), 990-1001.

French, B. F., & Maller, S. J. (2007). Iterative Purification and Effect Size Use With Logistic Regression for Differential Item Functioning Detection. *Educational and Psychological Measurement,* 67(3), 373-393.

Haladyna, T. M. (1999). *Developing and validating multiple-choice test items* (2nd ed.). New Jersey: Lawrence Erlbaum Associates.

Han, K. T. (2007). User's Manual for WinGen: Windows Software that Generates IRT Model Parameters and Item Responses. (computer program). Retrieved December 30, 2013 from http://www.umass.edu/remp/software/wingen/

Harwell, M., Stone, C. A., Hsu, T.-C., & Kirisci, L. (1996). Monte Carlo studies in Item Response Theory. *Applied Psychological Measurement,* 20(2), 101-125.

Hidalgo, M. D., & LÓPEZ-PINA, J. A. (2004). Differential Item Functioning Detection and Effect Size: A Comparison between Logistic Regression and Mantel-Haenszel Procedures. *Educational and Psychological Measurement,* 64(6), 903-915.

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

Jodoin, M. G., & Gierl, M. J. (2001). Evaluating Type I Error and Power Rates Using an Effect Size Measure With the Logistic Regression Procedure for DIF Detection. *Applied Measurement in Education,* 14(4), 329-349.

Kalaycioglu, D. B., & Berberoglu, G. (2010). Differential Item Functioning Analysis of the Science and Mathematics Items in the University Entrance Examinations in Turkey. *Journal of Psychoeducational Assessment,* 29(5), 467-478.

Kline, T. J. B. (2005). *Psychological Testing: A Practical Approach to Design and Evaluation*. California: Sage Publications.

Lautenschlager, G. J., & Park, D. G. (1988). IRT item bias detection procedures: issues of model Misspecification, robustness, and parameter linking. *Applied Psychological Measurement,* 12(4), 365-376.

Magis, D., Beland, S., & Raiche, G. (2015). Collection of Methods to Detect Dichotomous Differential Item Functioning (DIF). Retrieved November 11, 2015 from https://cran.r-project.org/web/packages/difR/difR.pdf

Magis, D., Beland, S., Tuerlinckx, F., & Boeck, P. D. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods,* 42(3), 847-862.

Masters, G. N., & Keeves, J. P. (1999). *Advances in Measurement in Educational Research and Assessment*. Amsterdam: Pergamon.

Mazor, K. M., Clauser, B. E., & Hambletor, R. K. (1994). Identification of nonuniform differential item functioning using a variation of the Mantel-Haenszel procedure. *Educational and Psychological Measurement,* 54(2), 284-291.

Moses, T., Miao, J., & Dorans, N. J. (2010). A Comparison of Strategies for Estimating Conditional DIF. *Journal of Educational and Behavioral Statistics,* 35(6), 726-743.

Narayanan, P., & 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.

Narayanan, P., & Swaminathan, H. (1996). Identification of items that show Nonuniform DIF. *Applied Psychological Measurement,* 20(3), 257-274.

Raju, N. S. (1990). Determining the significance of estimated signed and unsigned Areas between two item response functions. *Applied Psychological Measurement,* 14(2), 197-207.

Rogers, H. J., & Swaminathan, H. (1993). A Comparison of logistic regression and Mantel-Haenszel procedures for detecting differential item functioning. *Applied Psychological Measurement,* 17(2), 105-116.

Scherbaum, C. A., & Goldstein, H. W. (2008). Examining the Relationship Between Race-Based Differential Item Functioning and Item Difficulty. *Educationaland and Psychological Measurement,* 68(4), 537-553.

Sheppard, R., Han, K. T., Colarelli, S. M., Dai, G., & King, D. W. (2006). Differential Item Functioning by Sex and Race in the Hogan Personality Inventory. *Assessment,* 13(4), 442-453.

Shin, C. L., & Wang, W. C. (2009). Differential Item Functioning Detection Using the Multiple Indicators, Multiple Causes Method With a Pure Short Anchor. *Applied Psychological Measurement,* 33(3), 184-199.

Strobl, C., Kopf, J., & Zeileis, A. (2010). A New Method for Detecting Differential Item Functioning In the Rasch Model. Retrieved May 12, 2012, from https://epub.ub.uni-muenchen.de/11915/1/raschtree_techreport.pdf

Strobl, C., Kopf, J., & Zeileis, A. (2015). Rasch Trees : A New Method for Detecting Differential Item Functioning In the Rasch Model. *Psychometrika,* 80(2), 289-316.

Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using Logistic Regression procedures. *Journal of Educational Measurement,* 27(4), 361-370.

Vaughn, B. K., & Wang, Q. (2010). DIF Trees: Using Classification Trees to Detect Differential Item Functioning. *Educational and Psychological Measurement,* 70(6), 941-952.

Walker, C. M. (2011). What's the DIF? Why Differential Item Functioning Analyses Are an Important Part of Instrument Development and Validation. *Journal of Psychoeducational Assessment,* 29(4), 364-376.

Wiersma, W., & Jurs, S. G. (1990). *Educational Measurement and Testing* (2nd ed.). Massachusetts: Allyn and Bacon.

Woods, C. M. (2008). Likelihood-Ratio DIF Testing: Effects of Nonnormality. *Applied Psychological Measurement,* 32(7), 511-526.

Woods, C. M., & Grimm, K. J. (2011). Testing for Nonuniform Differential Item Functioning With Multiple Indicator Multiple Cause Models. *Applied Psychological Measurement,* 35(5), 339-361.

Zeileis, A., Hothorn, T., & Hornik, K. (2008). Model-based Recursive Partitioning. *Journal of Computational and Graphical Statistics,* 17(2), 492-514.

Zieky, M. (1993). Practical questions in the use of DIF statistics in test development. In P. W. Holland & H. Wainer (Eds.), *Differential Item Functioning* (pp. 337-347). Hillside, New Jersey: Lawrence Erlbaum Associates.

กาญจนา วัธนสุนทร. (2537). การพัฒนาเกณฑ์ตัดสินข้อสอบลำเอียงทางเพศ. วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

เกษร หว่างจิตร. (2539). การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบสำหรับแบบสอบคัดเลือกระดับบัณฑิตศึกษาวิชาภาษาไทยและภาษาอังกฤษด้วยวิธีแมนเทล-แฮนส์เซล. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

จิติมา วรรณศรี. (2539). การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซลกับวิธีซิบเทสท์ เมื่อความยาวของแบบสอบขนาดกลุ่มตัวอย่างและอัตราส่วนของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบต่างกัน. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

ทองอยู่ สาระ. (2543). การเปรียบเทียบอำนาจการตรวจสอบและการจำแนกผิดพลาดในการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบสม่ำเสมอและแบบไม่สม่ำเสมอ ระหว่างวิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติก โดยใช้ความยาวแบบทดสอบและขนาดกลุ่มตัวอย่างต่างกัน. ปริญญานิพนธ์ปริญญามหาบัณฑิต วิชาเอกการวัดผลการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.

ธเกียรติกมล ทองงอก. (2554). ประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในวิธีถดถอยโลจิสติก โดยใช้เกณฑ์ขนาดอิทธิพล 2 วิธี สำหรับข้อสอบที่มีรูปแบบการตรวจให้คะแนนแบบทวิวิภาค. วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

นพมาศ พิพัฒนสุข. (2541). การเปรียบเทียบประสิทธิภาพระหว่างวิธีแมนเทล-แฮนส์เซลกับวิธีถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อใช้เกณฑ์จับคู่เปรียบเทียบแตกต่างกันในแบบสอบชนิดพหุมิติ. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

นิคม กีรติวรางกูร. (2542). การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีการวิเคราห์องค์ประกอบจำกัด แมนเทล-แฮนส์เซลและการตอบสนองข้อสอบ. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

รักชนก ยี่สุ่นศรี. (2544). การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบด้วยกระบวนการ ดี เอฟ ไอ ที สำหรับแบบสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาวิชาภาษาอังกฤษและวิชาคณิตศาสตร์. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

รัชรินทร์ มุคดา. (2540). การเปรียบเทียบประสิทธิภาพระหว่างวิธีแมนเทล-แฮนส์เซลกับวิธีถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันของแบบสอบแบบอเนกรูป ในกรณีที่จัดกลุ่มความสามารถ ค่าความยากของข้อสอบ และค่าอำนาจจำแนกของข้อสอบต่างกัน. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

วลีมาศ แซ่อึ้ง. (2543). การเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป ระหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก. วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนวาสี. (2555). ทฤษฎีการทดสอบแนวใหม่ (***MODERN TEST THEORIES).*** พิมพ์ครั้งที่ 4. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี. (2553). ผลการประเมิน *PISA* 2009 การอ่าน คณิตศาสตร์ และวิทยาศาสตร์ บทสรุปเพื่อการบริหาร. กรุงเทพฯ: อรุณการพิมพ์.

สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี. (2554ก). ปัจจัยที่ทำให้ระบบโรงเรียนประสบความสำเร็จ. กรุงเทพฯ: อรุณการพิมพ์.

สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี. (2554ข). กรอบโครงสร้างการประเมินผลนักเรียนนานาชาติ *PISA* 2009. กรุงเทพฯ: อรุณการพิมพ์.

สุมาลี แกล้วทนงค์. (2547). สาเหตุของการทำหน้าที่ต่างกันของข้อสอบสาระการเรียนรู้ภาษาไทยและสาระการเรียนรู้สังคมศึกษา ศาสนาและวัฒนธรรม. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

เสรี ชัดแช้ม. (2539). การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันแบบไม่สม่ำเสมอของข้อสอบระหว่างวิธีแมนเทล-แฮนส์เซล แบบปกติ กับ วิธีแมนเทล-แฮนส์เซลแบบแบ่งกลุ่มความสามารถของผู้สอบและความยากของข้อสอบ. วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

อิทธิฤทธิ์ พงษ์ปิยะรัตน์. (2551). การวิเคราะห์ข้อสอบและการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ: การวิเคราะห์พหุระดับ. วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

ภาคผนวก

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาคผนวก ก
ตัวอย่างผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีการถดถอยโลจิสติก (Logistic Regression)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Logistic Regression

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	500	100.0
	Missing Cases	0	.0
	Total	500	100.0
Unselected Cases		0	.0
Total		500	100.0

a.If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
0	0
1	1

Categorical Variables Codings

		Frequency	Parameter coding (1)
g	1	250	.000
	2	250	1.000

Block 0: Beginning Block

Iteration History[a,b,c]

Iteration		-2 Log likelihood	Coefficients Constant
Step 0	1	439.867	1.376
	2	433.032	1.660
	3	432.977	1.688
	4	432.977	1.688

a.Constant is included in the model.

b.Initial -2 Log Likelihood: 432.977

c.Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Classification Table[a,b]

Observed			Predicted		
			i1		Percentage Correct
			0	1	
Step 0	i1	0	0	78	.0
		1	0	422	100.0
	Overall Percentage				84.4

a.Constant is included in the model.

b.The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	1.688	.123	187.644	1	.000	5.410

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	g(1)	2.187	1	.139
		x	163.074	1	.000
		g(1) by x	6.319	1	.012
	Overall Statistics		172.392	3	.000

Block 1: Method = Enter

Iteration History[a,b,c,d]

Iteration		-2 Log likelihood	Coefficients			
			Constant	g(1)	x	g(1) by x
Step 1	1	328.030	-.987	-1.528	.075	.042
	2	293.192	-1.594	-2.257	.112	.064
	3	289.660	-1.819	-2.761	.125	.081
	4	289.577	-1.846	-2.899	.127	.085
	5	289.577	-1.847	-2.905	.127	.086
	6	289.577	-1.847	-2.905	.127	.086

a.Method: Enter

b.Constant is included in the model.

c.Initial -2 Log Likelihood: 432.977

d.Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	143.400	3	.000
	Block	143.400	3	.000
	Model	143.400	3	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	289.577[a]	.249	.430

a.Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	9.706	8	.286

Contingency Table for Hosmer and Lemeshow Test

		i1 = 0		i1 = 1		
		Observed	Expected	Observed	Expected	Total
Step 1	1	37	38.416	16	14.584	53
	2	16	14.854	35	36.146	51
	3	7	7.499	46	45.501	53
	4	7	4.276	39	41.724	46
	5	7	3.708	48	51.292	55
	6	1	2.141	39	37.859	40
	7	1	2.352	51	49.648	52
	8	0	1.998	51	49.002	51
	9	2	2.002	64	63.998	66
	10	0	.754	33	32.246	33

Classification Table[a]

Observed			Predicted		
			i1		Percentage Correct
			0	1	
Step 1	i1	0	34	44	43.6
		1	10	412	97.6
	Overall Percentage				89.2

a.The cut value is .500

Variables in the Equation

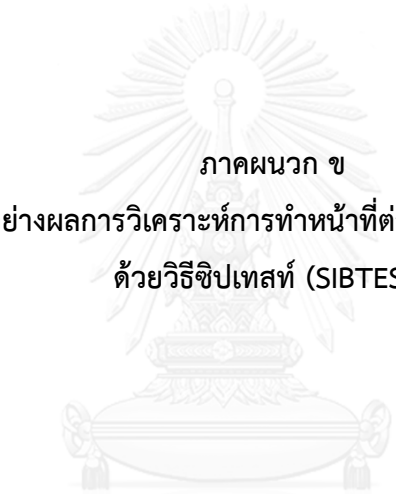
		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	g(1)	-2.905	1.073	7.332	1	.007	.055	.007	.448
	x	.127	.021	37.656	1	.000	1.135	1.090	1.182
	g(1) by x	.086	.037	5.460	1	.019	1.090	1.014	1.171
	Constant	-1.847	.596	9.605	1	.002	.158		

a.Variable(s) entered on step 1: g, x, g * x .

Correlation Matrix

		Constant	g(1)	x	g(1) by x
Step 1	Constant	1.000	-.555	-.933	.526
	g(1)	-.555	1.000	.518	-.957
	x	-.933	.518	1.000	-.564
	g(1) by x	.526	-.957	-.564	1.000

ภาคผนวก ข
ตัวอย่างผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ
ด้วยวิธีซิปเทสท์ (SIBTEST)



จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

SIBTEST OUTPUT

number of items on test = 40

name of file for Ref. grp. scores = C:\con1\con1_r_1.txt

name of file for Focal grp. scores = C:\con1\con1_f_1.txt

minimum no. of examinees per matching score cell = 2

number of runs for this data set = 40

number of examinees in Reference Group = 250

number of examinees in Focal group = 250

Examinee Test Score Summary Statistics

Reference Group: Mean = 32.60

Standard deviation = 8.63

Focal Group: Mean = 32.70

Standard deviation = 8.33

Standardized Score Difference = -.01

Item Statistics

= item number

p = proportion right on item (Classical Test Theory p-value)

r = point biserial (item score-test score correlation)

#: 1 2 3 4 5 6 7 8 9 10

p: 0.842 0.770 0.858 0.878 0.838 0.854 0.848 0.834 0.838 0.850

r: 0.603 0.539 0.540 0.566 0.622 0.568 0.613 0.501 0.532 0.543

#: 11 12 13 14 15 16 17 18 19 20

p: 0.844 0.882 0.832 0.794 0.806 0.710 0.772 0.784 0.852 0.838

r: 0.532 0.510 0.602 0.508 0.570 0.574 0.525 0.610 0.523 0.548

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

#: 21 22 23 24 25 26 27 28 29 30

p: 0.842 0.734 0.850 0.758 0.836 0.780 0.848 0.832 0.832 0.802

r: 0.554 0.592 0.549 0.553 0.559 0.506 0.556 0.537 0.569 0.586

#: 31 32 33 34 35 36 37 38 39 40

p: 0.800 0.790 0.752 0.802 0.862 0.870 0.750 0.784 0.838 0.764

r: 0.550 0.549 0.592 0.513 0.593 0.471 0.523 0.508 0.510 0.598

p-value notation:

R denotes p-value for test of DIF/DBF against Ref. group

F denotes p-value for test of DIF/DBF against Foc. group

E denotes p-value for test of DIF/DBF against either the Ref. or Foc. group.

NOTES:

MS/SSD = Matching Subtest Standardized Score Difference. Standardized difference in mean observed scores between Reference group and Focal group on the matching subtest.

p-elim = proportion of Reference (R) and Focal (F) groups eliminated (not used) in SIBTEST calculations.

Positive Beta estimate indicates DIF/DBF favoring Ref. grp.

Negative Beta estimate indicates DIF/DBF favoring Foc. grp.

FLAG = error flag indicator. FLAG=0 indicates a normal successful completion of a SIBTEST run. All other values of FLAG come with short error messages.

SIBTEST-pooled weighting

Run no.	Suspect Subtest Item Numbers	Beta estimate	standard error	p-value			p-elim R	MS F	SSD	FLAG
1	1	0.001	0.031	0.977	E	.24	.26	-0.01	0	
2	2	0.089	0.039	0.023	E	.24	.24	-0.02	0	
3	3	-0.023	0.031	0.446	E	.24	.25	-0.01	0	
4	4	-0.026	0.032	0.422	E	.26	.25	-0.01	0	
5	5	-0.053	0.032	0.101	E	.26	.24	-0.01	0	
6	6	0.045	0.032	0.162	E	.24	.25	-0.02	0	
7	7	-0.032	0.030	0.284	E	.22	.26	-0.01	0	
8	8	0.026	0.038	0.492	E	.27	.26	-0.01	0	
9	9	-0.006	0.033	0.850	E	.22	.25	-0.01	0	
10	10	-0.004	0.031	0.897	E	.23	.24	-0.01	0	
11	11	0.011	0.034	0.745	E	.25	.26	-0.01	0	
12	12	0.066	0.029	0.024	E	.23	.23	-0.02	0	
13	13	0.010	0.034	0.773	E	.25	.25	-0.01	0	
14	14	-0.018	0.037	0.621	E	.25	.26	-0.01	0	
15	15	0.018	0.038	0.642	E	.27	.27	-0.01	0	
16	16	0.004	0.044	0.931	E	.29	.27	-0.01	0	
17	17	-0.001	0.041	0.971	E	.24	.23	-0.01	0	
18	18	-0.010	0.040	0.797	E	.26	.30	-0.01	0	
19	19	-0.050	0.034	0.136	E	.23	.23	-0.01	0	
20	20	-0.029	0.033	0.386	E	.24	.25	-0.01	0	
21	21	-0.011	0.031	0.734	E	.24	.24	-0.01	0	
22	22	-0.038	0.043	0.367	E	.22	.24	-0.01	0	
23	23	-0.033	0.035	0.347	E	.24	.24	-0.01	0	
24	24	-0.007	0.041	0.866	E	.23	.26	-0.01	0	
25	25	0.037	0.035	0.287	E	.24	.25	-0.01	0	
26	26	-0.107	0.041	0.008	E	.25	.24	0.00	0	

27	27		0.064	0.034	0.060	E	.24	.24	-0.02	0
28	28		0.011	0.035	0.753	E	.21	.22	-0.01	0
29	29		-0.005	0.035	0.883	E	.23	.25	-0.01	0
30	30		-0.018	0.037	0.629	E	.25	.26	-0.01	0
31	31		0.056	0.037	0.133	E	.23	.24	-0.01	0
32	32		-0.055	0.040	0.173	E	.22	.24	-0.01	0
33	33		-0.001	0.042	0.985	E	.26	.25	-0.01	0
34	34		0.065	0.036	0.068	E	.25	.26	-0.02	0
35	35		-0.005	0.028	0.869	E	.26	.26	-0.01	0
36	36		-0.012	0.033	0.725	E	.25	.24	-0.01	0
37	37		-0.002	0.042	0.955	E	.25	.28	-0.01	0
38	38		-0.037	0.038	0.337	E	.23	.25	-0.01	0
39	39		0.065	0.034	0.060	E	.24	.26	-0.02	0
40	40		0.005	0.041	0.910	E	.24	.24	-0.01	0

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาคผนวก ค
ตัวอย่างผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ
ด้วยวิธีราสช์ทรี (Raschtree)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

-- Node 1 ---------------------------------

Number of observations: 500

Parameter instability tests:

	group
statistic	59.19055
p.value	0.02006

Best splitting variable: group

Perform split? yes

Selected split: 1 | 2

-- Node 2 ---------------------------------

Number of observations: 250

Parameter instability tests:

	group
statistic	0
p.value	NA

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Best splitting variable: NA

Perform split? no

-- Node 3 ---------------------------------

Number of observations: 250

Parameter instability tests:

	group
statistic	0
p.value	NA

Best splitting variable: NA

Perform split? no

null device

1

	resp1-C0	resp1-C1	resp2-C0	resp2-C1	resp3-C0	resp3-C1	resp4-C0	resp4-C1
1	0.3739	0.6261	0.5268	0.4732	0.4694	0.5306	0.5448	0.4552
2	0.3739	0.6261	0.5268	0.4732	0.4694	0.5306	0.5448	0.4552
3	0.3739	0.6261	0.5268	0.4732	0.4694	0.5306	0.5448	0.4552
4	0.3739	0.6261	0.5268	0.4732	0.4694	0.5306	0.5448	0.4552
5	0.3739	0.6261	0.5268	0.4732	0.4694	0.5306	0.5448	0.4552
6	0.3739	0.6261	0.5268	0.4732	0.4694	0.5306	0.5448	0.4552
7	0.3739	0.6261	0.5268	0.4732	0.4694	0.5306	0.5448	0.4552
8	0.3739	0.6261	0.5268	0.4732	0.4694	0.5306	0.5448	0.4552
9	0.3739	0.6261	0.5268	0.4732	0.4694	0.5306	0.5448	0.4552

	resp2	resp3	resp4	resp5	resp6	resp7	resp8	resp9	resp10
2	0.6228	0.3930	0.6950	0.2694	7.617e-06	0.3525	0.3525	0.1388	0.04722
3	0.3262	-0.4689	-0.5168	-0.5168	-3.761e-01	0.1133	0.1133	-0.3761	-1.13501

	resp11	resp12	resp13	resp14	resp15	resp16	resp17	resp18	resp19
2	7.617e-06	0.04722	0.4720	0.6591	0.7652	1.153	0.8672	0.9004	4.329e-01
3	2.918e-01	-0.88737	-0.3761	0.1133	0.1862	0.944	0.4268	0.6493	-4.254e-05

	resp20	resp21	resp22	resp23	resp24	resp25	resp26	resp27	resp28
2	0.5106	0.2267	1.526	0.09347	0.8336	0.4329	7.996e-01	0.5860	0.3113
3	-0.2871	-0.1597	0.944	-0.07856	0.2571	0.1500	-4.254e-05	0.0383	0.1862

	resp29	resp30	resp31	resp32	resp33	resp34	resp35	resp36	resp37	resp38
2	0.1831	0.6228	0.3930	0.8672	1.0919	0.7303	-0.09768	0.1831	0.9332	1.0294
3	-0.1597	0.6493	0.5241	0.4596	0.3601	0.1862	0.03830	-0.4689	0.5873	0.2918

	resp39	resp40
2	0.2267	0.96562
3	-0.2438	-0.03898

	resp1	resp2	resp3	resp4	resp5	resp6	resp7	resp8	resp9
2	-0.51542	0.1074	-0.1224	0.1795	-0.2460	-0.5154	-0.16289	-0.16289	-0.3767
3	-0.04365	0.2825	-0.5125	-0.5605	-0.5605	-0.4198	0.06963	0.06963	-0.4198

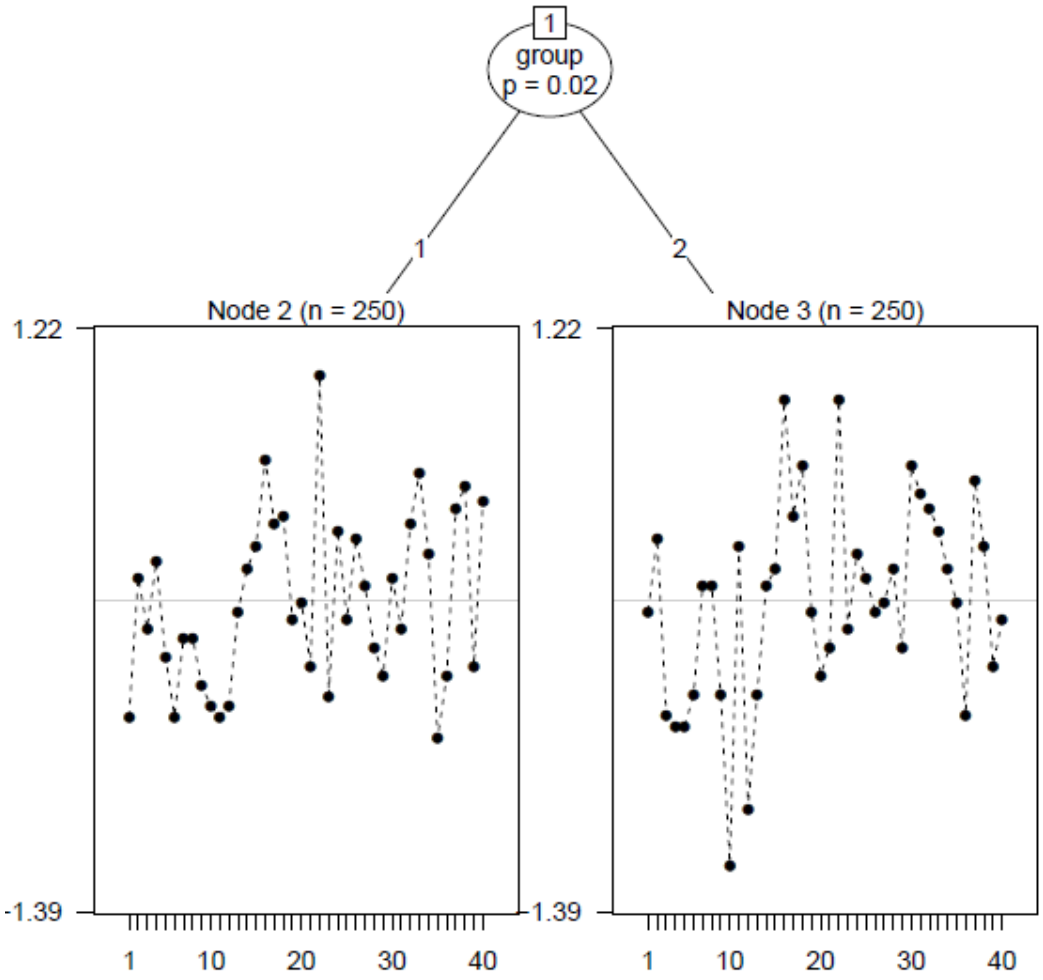
	resp10	resp11	resp12	resp13	resp14	resp15	resp16	resp17	resp18	resp19
2	-0.4682	-0.5154	-0.4682	-0.04339	0.14371	0.2498	0.6377	0.3518	0.3849	-0.08257
3	-1.1787	0.2482	-0.9310	-0.41976	0.06963	0.1425	0.9003	0.3832	0.6057	-0.04370

	resp20	resp21	resp22	resp23	resp24	resp25	resp26	resp27	resp28
2	-0.004834	-0.2888	1.0105	-0.4219	0.3182	-0.08257	0.2842	0.070528	-0.2041
3	-0.330731	-0.2034	0.9003	-0.1222	0.2134	0.10632	-0.0437	-0.005357	0.1425

	resp29	resp30	resp31	resp32	resp33	resp34	resp35	resp36	resp37	resp38
2	-0.3323	0.1074	-0.1224	0.3518	0.5764	0.2149	-0.613107	-0.3323	0.4178	0.5140
3	-0.2034	0.6057	0.4804	0.4159	0.3164	0.1425	-0.005357	-0.5125	0.5437	0.2482

	resp39	resp40
2	-0.2888	0.45020
3	-0.2875	-0.08264

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาคผนวก ง
ตัวอย่างผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ
ด้วยวิธีการวัดพื้นที่ของราชู (Raju)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Detection of Differential Item Functioning using Raju's method
with 1PL model and without item purification

Type of Raju's Z statistic: based on unsigned area

Engine 'ltm' for item parameter estimation

Common discrimination parameter: fixed to 1

No set of anchor items was provided

Raju's statistic:

	Stat.	P-value
i1	0.7876	0.4310
i2	-0.4978	0.6186
i3	-1.6591	0.0971 .
i4	0.4065	0.6844
i5	1.9622	0.0497 *
i6	-1.5433	0.1228
i7	-0.9080	0.3639
i8	0.1932	0.8468
i9	-1.8326	0.0669 .
i10	-2.5755	0.0100 **
i11	2.0040	0.0451 *
i12	-0.6320	0.5274
i13	2.6109	0.0090 **
i14	1.1382	0.2551
i15	0.0565	0.9550

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Detection thresholds: -1.96 and 1.96 (significance level: 0.05)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Items detected as DIF items:

i5
i10
i11
i13

Effect size (ETS Delta scale):

Effect size code:
'A': negligible effect
'B': moderate effect
'C': large effect

	mF-mR	deltaRaju	
i1	0.1866	-0.4385	A
i2	-0.1282	0.3013	A
i3	-0.3714	0.8728	A
i4	0.0914	-0.2148	A
i5	0.4964	-1.1665	B
i6	-0.3443	0.8091	A
i7	-0.2265	0.5323	A
i8	0.0432	-0.1015	A
i9	-0.4449	1.0455	B
i10	-0.6076	1.4279	B
i11	0.5263	-1.2368	B
i12	-0.1431	0.3363	A
i13	0.6379	-1.4991	B
i14	0.2706	-0.6359	A
i15	0.0135	-0.0317	A

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
Effect size codes: 0 'A' 1.0 'B' 1.5 'C'
 (for absolute values of 'deltaRaju')

Output was captured and saved into file
 'D:/difRaju/RAJUresults.txt'
```

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวสุภะ อภิญญาภิบาล เกิดที่กรุงเทพมหานคร สำเร็จการศึกษาระดับปริญญาตรี หลักสูตรบริหารธุรกิจบัณฑิต สาขาวิชาการบัญชี จากมหาวิทยาลัยศรีนครินทรวิโรฒ เมื่อปีการศึกษา 2540 ต่อมาปี พ.ศ.2541 ได้เข้ารับราชการที่สำนักทดสอบทางการศึกษาและจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ ในปีการศึกษา 2549 สำเร็จการศึกษาระดับปริญญาโท หลักสูตรการศึกษามหาบัณฑิต สาขาวิชาการวัดผลการศึกษา มหาวิทยาลัยศรีนครินทรวิโรฒ และในปีการศึกษา 2553 ได้เข้าศึกษาต่อในระดับปริญญาดุษฎีบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย และได้รับทุน “90 ปี จุฬาลงกรณ์มหาวิทยาลัย” จากกองทุนรัชดาภิเษกสมโภช จุฬาลงกรณ์มหาวิทยาลัย ประจำปี 2557

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University