ประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในวิธีถดถอยโลจิสติก โดยใช้เกณฑ์ขนาดอิทธิพล 2 วิธี สำหรับข้อสอบที่มีรูปแบบการตรวจให้คะแนนแบบทวิวิภาค: ข้อมูลจำลองและข้อมูลเชิงประจักษ์

นางสาวธเกียรติกมล ทอองอก

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาครุศาสตรดุษฎีบัณฑิต
สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา
คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2554

EFFICACY OF DETECTION DIF IN LOGISTIC REGRESSION BY USING TWO EFFECT SIZE CRITERIA FOR DICHOTOMOUSLY SCORED ITEMS: SIMULATION AND EMPIRICAL DATA

Miss Tha-kiatkamol Thong-ngok

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 2011

หัวข้อวิทยานิพนธ์ ประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในวิธีถดถอยโลจิสติก โดยใช้เกณฑ์ขนาดอิทธิพล 2 วิธี สำหรับข้อสอบที่มีรูปแบบการตรวจให้คะแนนแบบทวิวิภาค: ข้อมูลจำลองและข้อมูลเชิงประจักษ์

โดย นางสาวธเกียรติกมล ทองงอก
สาขาวิชา การวัดและประเมินผลการศึกษา
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก รองศาสตราจารย์ ดร.โชติกา ภาษีผล
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาดุษฎีบัณฑิต

.....คณบดีคณะครุศาสตร์
(ศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ
(รองศาสตราจารย์ ดร.ศิริเดช สุชีวะ)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(รองศาสตราจารย์ ดร.โชติกา ภาษีผล)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม
(ศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี)

.....กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.ณัฏฐภรณ์ หลาวทอง)

.....กรรมการภายนอกมหาวิทยาลัย
(อาจารย์ ดร.ชูศักดิ์ ขัมภลิขิต)

ธเกียรติกมล ทองงอก : ประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในวิธีถดถอยโลจิสติก โดยใช้เกณฑ์ขนาดอิทธิพล 2 วิธี สำหรับข้อสอบที่มีรูปแบบการตรวจให้คะแนนแบบทวิวิภาค : ข้อมูลจำลองและข้อมูลเชิงประจักษ์ (EFFICACY OF DETECTION DIF IN LOGISTIC REGRESSION BY USING TWO EFFECT SIZE CRITERIA FOR DICHOTOMOUSLY SCORED ITEMS: SIMULATION AND EMPIRICAL DATA) อ.ที่ปรึกษาวิทยานิพนธ์หลัก : รศ.ดร.โชติกา ภาษีผล , อ.ที่ปรึกษาวิทยานิพนธ์ร่วม: ศ.ดร.ศิริชัย กาญจนวาสี, 306 หน้า.

การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อ เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่มีรูปแบบการให้คะแนนแบบทวิวิภาค โดยการจำลอง ข้อมูลและข้อมูลเชิงประจักษ์ ในวิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl กับเกณฑ์ Zumbo and Thomas การศึกษาครั้งนี้จำลองข้อมูลภายใต้ทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์ จำลองผลการตอบภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย รวมข้อมูลที่ศึกษาทั้งหมด 24 เงื่อนไข (2 x 3 x 2 x 2) คือ รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน (อเนกรูป และ เอกรูป) ขนาดของการทำหน้าที่ต่างกัน (0.1, 0.2 และ 0.4) จำนวนข้อสอบที่ทำหน้าที่ต่างกัน (ทั้งฉบับคิดเป็นร้อยละ 10 และ 20) และความยาวของแบบสอบทั้งฉบับ (40 และ 50 ข้อ) ใน ทุกเงื่อนไขจำลองข้อมูลซ้ำ 25 ครั้ง วิเคราะห์ข้อมูลในแต่ละเงื่อนไข ด้วยวิธีถดถอยโลจิสติก ระหว่างการวัด ขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl และเกณฑ์ Zumbo and Thomas การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบทั้งหมดใช้ระดับนัยสำคัญ .05

ผลการวิจัยสรุปได้ดังนี้

1. วิธีถดถอยโลจิสติก โดยการวัด ขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl มีอัตราความถูกต้อง ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูงกว่าเกณฑ์ Zumbo and Thomas ภายใต้เกือบทุกเงื่อนไข

2. ข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปมีอัตราความถูกต้องจากการวัดขนาดอิทธิพลทั้ง 2 เกณฑ์สูงกว่าแบบเอกรูป แบบสอบที่มี จำนวนข้อสอบทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 20 มีอัตราความถูกต้อง จากการวัด ขนาดอิทธิพลทั้ง 2 เกณฑ์สูงกว่า ในแบบสอบที่มี จำนวนข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 10 และเมื่อขนาดอิทธิพลของข้อสอบที่ การทำหน้าที่ต่างกัน เพิ่มขึ้น มีผลทำให้อัตราความถูกต้อง จากการวัด ขนาดอิทธิพล ทั้ง 2 เกณฑ์เพิ่มขึ้นภายใต้เกือบทุกเงื่อนไข

3. ผลการตรวจสอบการทำหน้าที่ต่างกัน ของข้อสอบ ในข้อมูลเชิงประจักษ์ พบว่า ขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl ให้อัตราความถูกต้องสูงกว่า และ มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าเกณฑ์ของ Zumbo and Thomas เมื่อข้อมูลเชิงประจักษ์มีประชากรขนาดใหญ่สามารถตรวจพบ ข้อสอบที่ทำหน้าที่ต่างกัน ด้วยการทดสอบระดับนัยสำคัญอย่างมีนัยสำคัญ ส่งผลให้ความคลาดเคลื่อนประเภทที่ มีแนวโน้มสูงขึ้น

ข้อเสนอแนะ : ภายใต้การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก นักวิจัยควรใช้ ผลการทดสอบระดับนัยสำคัญในการตัดสินข้อสอบที่ทำหน้าที่ต่างกันร่วมกับผลของการวัดขนาดอิทธิพล

ภาควิชา วิจัยและจิตวิทยาการศึกษา ลายมือชื่อนิสิต .....

สาขาวิชา การวัดและประเมินผลการศึกษา ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก .....

ปีการศึกษา 2554 ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์ร่วม .....

##5084241127: MAJOR EDUCATIONAL MEASUREMENT AND EVALUATION
KEYWORD: EFFECT SIZE MEASURES / DIFFERENTIAL ITEM FUNCTIONIG / LOGISTIC REGRESSION PROCEDURE / DICHOTOMOUSLY / SIMULATION / EMPIRICAL DATA / EFFECT SIZE MEASURES

THAKIATKAMOL THONGNGOK: (EFFICACY OF DETECTION DIF IN LOGISTIC REGRESSION BY USING TWO EFFECT SIZE CRITERIA FOR DICHOTOMOUSLY SCORED ITEMS: SIMULATION AND EMPIRICAL DATA) ADVISOR: ASSOC.PROF. SHOTIGA PASIPHOL, Ph.D, CO-ADVISOR: PROF.SIRICHAI KANJANAWASEE, Ph.D, 306 pp.

The objectives of this study were to compare correct identification and Type I error rate of DIF with dichotomously scored items by simulation and empirical data in logistic regression procedure between effect size measures of Jodoin and Gierl's criteria and Zumbo and Thomas's criteria. In this study, the data was simulated under the IRT theory of two-parameter item response, simulating dichotomous response under the condition of 4 varied factors. The total of data studied was 24 conditions (2 x 3 x 2 x 2); 2 forms of DIF Type (Nonuniform and Uniform), 3 amounts of DIF (0.1, 0.2 and 0.4), 2 numbers of items with DIF (10% and 20%), and 2 sizes of Test length (40 and 50 items). The data was replicated 25 times for each condition. In each condition, the data was analyzed with effect size measures of Jodoin and Gierl's criteria and Zumbo and Thomas's criteria. Significance .05 was used in the analysis of all DIF.

The research results were as follows:

1. Logistic regression procedure with effect size measures of Jodoin and Gierl's criteria had higher correct identification of DIF than of Zumbo and Thomas's criteria under almost conditions.

2. Nonuniform DIF had higher correct identification from effect size measures with both criteria than uniform DIF. All items with DIF at 20 percent had higher correct identification from effect size measures with both criteria than all items with DIF at 10 percent. And when the effect size of DIF increased, the correct identification from effect size measured with both criteria increased as well under almost conditions.

3. The detection result of DIF in empirical data revealed that the effect size of Jodoin and Gierl's criteria yielded higher correct identification and lower Type I error rate than of Zumbo and Thomas's criteria. When big size of population was studied in an empirical data, DIF could be detected by significantly testing significance which tended to increase deviation type 1 error rate.

Suggesstion: Under the detection with logistic regression procedure, the result of significance test should be used along with the result of effect size to detect DIF.

Department: Educational Research and Psychology Student's Signature……..
Field of Study: Educational Measurement and Evaluation Advisor's Signature……
Academic Year: 2011 Co-advisor's Signature……

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปด้วยดีด้วยความกรุณาอย่างเป็นที่สุดของอาจารย์ รศ.ดร.โชติกา ภาษีผล อาจารย์ที่ปรึกษาวิทยานิพนธ์ และ ศ.ดร.ศิริชัย กาญจนวาสี อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ซึ่งท่านทั้งสองให้คำแนะนำ แนวคิด แก้ไขข้อบกพร่องในการทำวิทยานิพนธ์ด้วยดีเสมอมา ตลอดจนเป็นผู้ให้ความเมตตามอบโอกาสและประสบการณ์ในการทำงานด้านการศึกษา อันเป็นประสบการณ์ตรงที่ข้าพเจ้าไม่สามารถหาได้ในห้องเรียน ผู้วิจัยขอกราบขอบพระคุณท่านทั้งสองเป็นอย่างสูงมา ณ โอกาสนี้

ขอกราบขอบพระคุณ ดร.ดิเรก ศรีสุโข ศ.ดร.สุวิมล ว่องวาณิช ศ.กิตติคุณ ดร. นงลักษณ์ วิรัชชัย และ รศ.ดร. สิริพันธุ์ สุวรรณมรรคา ที่ท่านเป็นเสมือนปูชนียบุคคล และคณาจารย์ประจำภาควิชาวิจัยและจิตวิทยาทุกท่าน ที่ได้ให้ความเมตตา แนะนำชี้แนะประสิทธิประศาสตร์วิชาการและประสบการณ์ที่มีคุณค่าแก่ลูกศิษย์เสมอมา

ขอกราบขอบพระคุณ รศ.ดร.ศิริเดช สุชีวะ ประธานสอบวิทยานิพนธ์ อาจารย์ ร.ชูศักดิ์ ขัมภลิขิต และ ผศ.ดร.ณัฏฐภรณ์ หลาวทอง กรรมการสอบวิทยานิพนธ์ ที่คอยแนะนำชี้แนะด้านวิชาการที่มีคุณค่าต่อวิทยานิพนธ์ จนทำให้วิทยานิพนธ์นี้เสร็จสมบูรณ์

ขอกราบขอบพระคุณ ดร.สังวรณ์ งัดกระโทก และคุณอาบีด หนุนอนันต์ ที่ให้ความรู้เกี่ยวกับแนวคิดการจำลองข้อมูล ความรู้สถิติและให้ข้อเสนอแนะที่มีคุณค่ายิ่งแก่ผู้วิจัย

ขอขอบคุณ อาจารย์ david magis และคณะแห่ง Katholieke Universiteit Leuven ประเทศเบลเยียม เจ้าของ Package 'difR' และอาจารย์ Seung W. Choi และคณะ แห่ง Northwestern University Feinberg School of Medicine Chicago, Illinois สหรัฐอเมริกา เจ้าของPackage 'lordif' โดยผู้วิจัยนำPackage ของอาจารย์ทั้งสองท่านมาใช้ในการวิเคราะห์ผลการทำหน้าที่ต่างกันของข้อสอบภายใต้โปรแกรมR

ขอขอบคุณอย่างสุดซึ้งแก่ ดร.ชนะศึก นิชานนท์ ที่เป็นดั่งกัลยาณมิตรแนะนำช่วยเหลือ แลกเปลี่ยนประสบการณ์การเรียนรู้ที่สำคัญในชีวิตทั้งในและนอกห้องเรียน ขอขอบคุณ ดร.วราพร เอราวรรณ์ ตลอดจนเพื่อนร่วมรุ่น คุณศักดิ์สิทธิ์ ฤทธิลัน ดร.สาธิดา สกุลรัตนกุลชัย คุณทัศน์ศิรินทร์ สว่างบุญ ดร.ชลี ภัทรพิชยธรรม ที่ให้กำลังใจเอาใจใส่ดูแลกัน และร่วมแลกเปลี่ยนเรียนรู้ด้านวิชาการ และนันทนาการเสมอมาขอบคุณทุกกำลังใจ ดร.ศิริรัตน์ สุคันธพฤษ์ คุณเรืองเดช ศิริกิจ คุณสุภาวดี คำนาดี คุณสุกัญญา จันทวาลย์ คุณสุกัญญา ทองนาค คุณอนันดา สัณฐิติวณิชย์ ขอขอบคุณ คุณณัฐสุรางค์ ยะสูงเนิน คุณชานนท์ จินตะเวช คุณฐกฤต เฉลยวาเรศ คุณอานันต์ชนกวิจิตรนิเทศ และพี่สาวที่น่ารักคุณนิตยา ธนบริบูรณ์พงศ์ รวมถึงพี่น้องในภาควิชาวิจัยและจิตวิทยาการศึกษาทุกคน

ขอขอบคุณจุฬาลงกรณ์มหาวิทยาลัยที่ให้โอกาสผู้วิจัยได้รับการสนับสนุนทุนวิจัยจาก "ทุน 90 ปี จุฬาลงกรณ์มหาวิทยาลัย" กองทุนรัชดาภิเษกสมโภช เพื่ออุดหนุนในการทำวิทยานิพนธ์จนสำเร็จลุล่วงไปด้วยดี

ขอกราบขอบพระคุณ คุณจีระชัย ไกรกังวาร นายกเทศมนตรีเมืองวารินชำราบ คุณบุญยง จินตนะกุล คุณวิเศษ หิรัญเทศ และคุณวีรยาพร จันทรา ที่ให้ความสำคัญต่อการศึกษาและหยิบยื่นโอกาสด้านการศึกษาแก่ผู้วิจัย

ขอกราบขอบพระคุณ คุณปู่พอ.พิเศษ นายแพทย์ธนิต -คุณย่าอุไรวรรณ ทองงอก คุณสันทัด -ผศ.ดร.วรรณวิภา จัตุชัย คุณนิตยา ทองงอก คุณไพบูลย์ -คุณจุฑารัตน์ จงสุวัฒน์ ผศ.ดร.นิพนธ์ -คุณประภาวดี จันทร์โพธิ์ คุณวรรณี หลักคำ คุณจิระศักดิ์ -คุณวาสนา แก้วรักษา และ คุณสุบรรณคุณสิตานันท์ บรรณโก ที่เป็นผู้อบรมเลี้ยงดูและให้กำลังใจแก่ผู้วิจัยเสมอมา

ขอกราบขอบพระคุณบิดามารดา พ.ท.สวัสดิ์ -นางขวัญจิต ทองงอก และน้องชาย คุณธารณ์ ทองงอก ที่มอบความรัก ความห่วงใย คอยเป็นกำลังใจที่สำคัญที่สุด เอาใจใส่ต่อสุขภาพ ให้ความเอื้ออาทรด้วยดี อันเปรียบได้ดั่งขุมพลังมหาศาล และเป็นแรงผลักดันให้ผู้วิจัยมีความเพียรมุ่งมั่นด้านการศึกษาจนทำให้ผู้วิจัยมีวันนี้

ความดีงามทั้งหลายทั้งปวง ข้าพเจ้าขออุทิศแด่การศึกษาของชาติไทย

สารบัญ

หน้า

บทคัดย่อภาษาไทย..... ง
บทคัดย่อภาษาอังกฤษ..... จ
กิตติกรรมประกาศ..... ฉ
สารบัญ..... ช
สารบัญตาราง..... ญ
สารบัญภาพ..... ฒ
บทที่ 1 บทนำ..... 1
ความเป็นมาและความสำคัญของปัญหา..... 1
คำถามวิจัย..... 14
วัตถุประสงค์ของการวิจัย..... 14
สมมติฐานการวิจัย..... 15
ขอบเขตของการวิจัย..... 17
นิยามศัพท์ที่ใช้ในการวิจัย..... 18
ประโยชน์ที่คาดว่าจะได้รับ..... 21
บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง..... 23
ตอนที่ 1 มโนทัศน์ของทฤษฎีทางการสอบแบบดั้งเดิมและการประยุกต์ใช้..... 23
ตอนที่ 2 มโนทัศน์ของทฤษฎีการตอบสนองข้อสอบและการประยุกต์ใช้..... 26
ตอนที่ 3 มโนทัศน์ของการทำหน้าที่ต่างกันของข้อสอบ..... 31
3.1 ความลำเอียงและการทำหน้าที่ต่างกันของข้อสอบ..... 31
3.2 ประเภทของการทำหน้าที่ต่างกันของข้อสอบ..... 34
3.3 หลักการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ..... 37
3.4 ทฤษฎีการตอบสนองข้อสอบและการทำหน้าที่ต่างกันของข้อสอบ..... 38
3.5 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ..... 41
3.6 ประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ..... 45
3.7 ทฤษฎีการตอบสนองข้อสอบสำหรับรูปแบบการให้คะแนนแบบทวิวิภาค.. 46
ตอนที่ 4 มโนทัศน์ของขนาดอิทธิพล..... 47
4.1 การรายงานขนาดอิทธิพลของงานวิจัยในปัจจุบัน..... 47
4.2 ความหมายของขนาดอิทธิพล..... 48

หน้า หน้า

4.3 การตัดสินใจทางการวิจัย..... 48
4.4 ขนาดอิทธิพลกับการศึกษาการทำหน้าที่ต่างกันของข้อสอบ..... 50
ตอนที่ 5 เอกสารและงานวิจัยที่เกี่ยวข้อง..... 53
5.1 งานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบในประเทศ..... 53
5.2 งานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบต่างประเทศ..... 70
5.3 สรุปประเด็นปัญหาที่พบเกี่ยวกับการการทำหน้าที่ต่างกันของข้อสอบ..... 94
5.4 กรอบแนวคิดการวิจัย..... 99
บทที่ 3 วิธีดำเนินการวิจัย..... 102
ตอนที่ 1 ขั้นตอนการวิจัย..... 103
ตอนที่ 2 การจัดกระทำข้อมูลตามปัจจัยที่ศึกษา..... 106
ตอนที่ 3 การจำลองข้อมูล..... 108
ตอนที่ 4 การวิเคราะห์ข้อมูล..... 113
ตอนที่ 5 การเปรียบเทียบประสิทธิภาพของผลการตรวจสอบ..... 115
บทที่ 4 ผลการวิเคราะห์ข้อมูล..... 118
ตอนที่ 1 การคำนวณค่าเฉลี่ยของอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1..... 120
ตอนที่ 2 ผลการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างการวัดขนาดอิทธิพลตามเกณฑ์ที่ศึกษา 129
ตอนที่ 3 สรุปผลการเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1..... 152
ตอนที่ 4 ผลการศึกษาในกรณีข้อมูลเชิงประจักษ์..... 165
บทที่ 5 สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ..... 200
สรุปผลการวิจัย..... 203
อภิปรายผล..... 213
ข้อเสนอแนะ..... 220
รายการอ้างอิง..... 224

หน้า

ภาคผนวก..... 235

ภาคผนวก ก ค่าพารามิเตอร์ข้อสอบรายข้อ ภายใต้ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory model) ชนิด 2 พารามิเตอร์ (two-parameter)..... 236

ภาคผนวก ข ตัวอย่างผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติกโดยการทดสอบระดับนัยสำคัญ และการวัดขนาดอิทธิพลภายใต้ทฤษฎีการตอบสนองข้อสอบ ชนิด 2 พารามิเตอร์..... 251

ภาคผนวก ค Print out ผลการเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ และการวัดขนาดอิทธิพล..... 265

ภาคผนวก ง การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธีแมนเทล-แฮนด์เซลล์ วิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ และการวัดขนาดอิทธิพลในข้อมูลเชิงประจักษ์..... 279

ภาคผนวก จ ผลการวิเคราะห์ประสิทธิผลของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้านอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ และการวัดขนาดอิทธิพล..... 299

ประวัติผู้เขียนวิทยานิพนธ์..... 306

สารบัญตาราง

ตารางที่		หน้า
2.1	ฟังก์ชันทางคณิตศาสตร์ของโมเดลการตอบสนองข้อสอบ.....	39
2.2	วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับรูปแบบการตรวจให้คะแนนแบบทวิวิภาค ที่เป็นเอกมิติ จำแนกตามลักษณะของข้อมูล.....	43
2.3	วิธีการตรวจสอบการทำหน้าที่ต่างกันที่มีการให้คะแนนแบบทวิวิภาค (Dichotomous DIF) และพหุวิภาค (polytomous DIF).....	44
2.4	คุณภาพของการตรวจสอบประสิทธิภาพของการทำหน้าที่ต่างกันของข้อสอบ...	45
2.5	แสดงสัดส่วนของระดับการตอบสนอง k ระดับตามช่วงระดับความสามารถของ b ใช้ค่า b ระหว่าง ±3.00 เมื่อกำหนดค่า a เป็น .50, 1.00 และ 2.00.....	51
2.6	ค่า R^2 ของการวัดขนาดอิทธิพลในการทำหน้าที่ต่างกันของข้อสอบ.....	52
2.7	สรุปงานวิจัยที่ได้ศึกษาเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบในประเทศไทยตั้งแต่อดีตถึงปัจจุบัน.....	62
2.8	สรุปงานวิจัยที่ได้ศึกษาเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบในต่างประเทศถึงปัจจุบัน.....	81
4.1	ภาพรวมของการจำลองข้อมูลจำแนกตามปัจจัยและเงื่อนไขของปัจจัยที่แปรเปลี่ยน.....	120
4.2	การตรวจสอบคุณภาพข้อมูลจำลองตามจำแนกตามปัจจัยที่ศึกษา 24 เงื่อนไข	121
4.3	ร้อยละเฉลี่ย ($\overline{X}$) และส่วนเบี่ยงเบนมาตรฐาน (SD) ร้อยละของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในภาพรวม	122
4.4	ค่าเฉลี่ยร้อยละของอัตราความถูกต้องในทุกวิธีที่ศึกษาภายใต้วิธีถดถอยโลจิสติก.....	125
4.5	ค่าเฉลี่ยร้อยละของอัตราความคลาดเคลื่อนประเภทที่ 1 จำแนกตามวิธีที่ศึกษาภายใต้วิธีถดถอยโลจิสติก.....	126
4.6	การทดสอบ Box's Test และ Bartlett's Test .....…….....…....	130
4.7	การเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่1 ในวิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพลตามเกณฑ์ที่ศึกษาภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย.....	131

หน้า

ตารางที่

4.8 ผลการทดสอบระหว่างกลุ่ม ภายใต้เงื่อนไขปฏิสัมพันธ์สองทางระหว่าง วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบกับปัจจัยที่แปรเปลี่ยน…… 132

4.9 การทดสอบความแปรปรวนผลย่อยระหว่างวิธีการตรวจสอบการทำหน้าที่ ต่างกันของข้อสอบ กับปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน……..... 135

4.10 การทดสอบความแปรปรวนผลย่อยระหว่างวิธีการตรวจสอบการทำหน้าที่ ต่างกันของข้อสอบ กับปัจจัยความยาวของแบบสอบทั้งฉบับ 136

4.11 การทดสอบความแปรปรวนผลย่อยระหว่างวิธีการตรวจสอบการทำหน้าที่ ต่างกันของข้อสอบกับ ปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ..... 138

4.12 การทดสอบความแปรปรวนผลย่อยระหว่าง วิธีการตรวจสอบการทำหน้าที่ ต่างกันของข้อสอบ กับ ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันและ ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน..... 140

4.13 การทดสอบความแปรปรวนผลย่อยระหว่างวิธีการตรวจสอบการทำหน้าที่ ต่างกันของข้อสอบกับ ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกันและปัจจัย ความยาวของแบบสอบทั้งฉบับ……..... 143

4.14 การทดสอบความแปรปรวนผลย่อยระหว่างวิธีการตรวจสอบการทำหน้าที่ ต่างกันของข้อสอบกับ ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันและ ปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ…..……..... 146

4.15 การทดสอบความแปรปรวนผลย่อยระหว่างวิธีการตรวจสอบการทำหน้าที่ ต่างกันของข้อสอบ กับ ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน และปัจจัย ขนาดของการทำหน้าที่ต่างกันของข้อสอบ……..…… 149

4.16 ปัจจัยที่มีผลต่ออัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่1 ในวิธีถดถอยโลจิสติก ของการวัดขนาดอิทธิพล2 เกณฑ์..... 158

4.17 สรุปผลการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อน ประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอย โลจิสติก ระหว่างการวัดขนาดอิทธิพลตามเกณฑ์Jodoin and Gierl และเกณฑ์ Zumbo and Thomas ภายใต้เงื่อนไขเดียวกัน……..….. 160

หน้า

ตารางที่

4.18 สรุปผลการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl และเกณฑ์ Zumbo and Thomas ภายใต้เงื่อนไขต่างกัน..... 162

4.19 สถิติเชิงบรรยายของคะแนนจากแบบสอบรายวิชาคณิตศาสตร์……..…….. 165

4.20 ค่าความเที่ยงแบบความสอดคล้องภายในของแบบสอบโดยสูตรแอลฟาสัมประสิทธิ์ของครอนบาค วิชาคณิตศาสตร์ จำแนกตามกลุ่มผู้สอบ……... 166

4.21 ค่าความยาก (p) และอำนาจจำแนก (r) ของข้อสอบรายข้อ วิชาคณิตศาสตร์ (40 ข้อ)..... 166

4.22 ค่าพารามิเตอร์อำนาจจำแนก (a) และความยาก (b) ของข้อสอบรายข้อวิชาคณิตศาสตร์..... 167

4.23 ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation matrix), KMO, ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานของแบบสอบวิชาคณิตศาสตร์ (n=123,167 คน)…….. 170

4.24 ค่าไอเกนและร้อยละของความแปรปรวนขององค์ประกอบของแบบสอบวิชาคณิตศาสตร์……..…….. 173

4.25 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวัดพื้นที่ของราชูวิชาคณิตศาสตร์..... 175

4.26 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้วิธีถดถอยโลจิสติกโดยการทดสอบระดับนัยสำคัญ ในวิชาคณิตศาสตร์……... 176

4.27 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติกโดยการวัดขนาดอิทธิพลในวิชาคณิตศาสตร์ ตามเกณฑ์ Zumbo and Thomas..... 177

4.28 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติกโดยการวัดขนาดอิทธิพลในวิชาคณิตศาสตร์ ตามเกณฑ์ Jodoin and Gierl…… 178

4.29 สรุปผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ..... .....… 179

4.30 จำนวนข้อของการเกิดและไม่เกิดการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์..... 179

หน้า

ตารางที่

4.31 ร้อยละของอัตราความถูกต้องและร้อยละของอัตราความคลาดเคลื่อนประเภทที่ 1 จำแนกตามการตรวจสอบการทำหน้าที่ต่างกันของแบบสอบตามวิธีที่ศึกษาในวิชาคณิตศาสตร์..... 180

4.32 สถิติเชิงบรรยายของคะแนนจากแบบสอบรายวิชาวิทยาศาสตร์..... 181

4.33 ค่าความเที่ยงแบบความสอดคล้องภายในของแบบสอบโดยสูตรสัมประสิทธิ์แอลฟาของครอนบาค จำแนกตามวิชาวิทยาศาสตร์และกลุ่มผู้สอบ..... 181

4.34 ค่าความยาก (p) และอำนาจจำแนก (r) ของข้อสอบรายข้อ วิชาวิทยาศาสตร์ (50 ข้อ)..... 182

4.35 ค่าพารามิเตอร์อำนาจจำแนก (a) และความยาก (b) ของข้อสอบรายข้อวิชาวิทยาศาสตร์..... 183

4.36 ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation matrix), KMO, ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน ของข้อสอบวิชาวิทยาศาสตร์ (n=110,609 คน)..... 185

4.37 ค่าไอเกนและร้อยละของความแปรปรวนขององค์ประกอบในแบบสอบวิชาวิทยาศาสตร์..... 189

4.38 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธีการวัดพื้นที่ของราชูวิชาวิทยาศาสตร์..... 191

4.39 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้วิธีถดถอยโลจิสติกโดยการทดสอบระดับนัยสำคัญ ในวิชาวิทยาศาสตร์..... 192

4.40 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติกโดยการวัดขนาดอิทธิพลในวิชาคณิตศาสตร์ ตามเกณฑ์ของ Zumbo and Thomas..... 194

4.41 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติกโดยการวัดขนาดอิทธิพลในวิชาคณิตศาสตร์ ตามเกณฑ์ Jodoin and Gierl..... 195

4.42 สรุปผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาวิทยาศาสตร์..... 196

4.43 จำนวนข้อของการเกิดและไม่เกิดการทำหน้าที่ต่างกันของข้อสอบของวิชาวิทยาศาสตร์..... 196

หน้า

ตารางที่

4.44 ร้อยละของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่1 จำแนกตามการตรวจสอบการทำหน้าที่ต่างกันของแบบสอบตามวิธีที่ศึกษาวิชาวิทยาศาสตร์..... 197

4.45 ผลการตรวจสอบ DIF ระหว่างการทดสอบระดับนัยสำคัญ กับการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl..... 198

สารบัญภาพ

หน้า

ภาพที่

2.1	ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป (uniform DIF) ……..…..……	35
2.2	ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป (Nonuniform DIF) โดยมีปฏิสัมพันธ์ไม่เป็นลำดับ (Disordinal interaction)…….....	36
2.3	ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป (Nonuniform DIF) โดยมีปฏิสัมพันธ์เป็นลำดับ (Ordinal interaction) …….....	36
2.4	กรอบแนวคิดในการวิจัย กรณีศึกษาการจำลองข้อมูล……	100
2.5	กรอบแนวคิดในการวิจัย กรณีศึกษาข้อมูลเชิงประจักษ์…….....	101
3.1	ขั้นตอนการดำเนินการศึกษา.....	105
3.2	ขั้นตอนการจำลองข้อมูล…….....	111
3.3	แผนผังของการจำลองข้อมูล.....	112
4.1	เกณฑ์การพิจารณาประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ…….....	129
4.2	อัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน 2 เงื่อนไข.....	152
4.3	อัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ปัจจัยขนาดของการทำหน้าที่ต่างกัน 3 เงื่อนไข.....	153
4.4	อัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 2 เงื่อนไข.....	154
4.5	อัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ปัจจัยความยาวของแบบสอบทั้งฉบับ 2 เงื่อนไข.....	155
4.6	ผลการตรวจสอบความเป็นเอกมิติของแบบสอบวิชาคณิตศาสตร์ 50 ข้อ.....	173
4.7	ผลการตรวจสอบความเป็นเอกมิติของแบบสอบวิชาวิทยาศาสตร์ 50 ข้อ.....	189

สารบัญ

หน้า

บทคัดย่อภาษาไทย..... ง
บทคัดย่อภาษาอังกฤษ..... จ
กิตติกรรมประกาศ..... ฉ
สารบัญ..... ช
สารบัญตาราง..... ญ
สารบัญภาพ..... ฌ
บทที่ 1 บทนำ..... 1
ความเป็นมาและความสำคัญของปัญหา..... 1
คำถามวิจัย..... 14
วัตถุประสงค์ของการวิจัย..... 14
สมมติฐานการวิจัย..... 15
ขอบเขตของการวิจัย..... 17
นิยามศัพท์ที่ใช้ในการวิจัย..... 18
ประโยชน์ที่คาดว่าจะได้รับ..... 21
บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง..... 23
ตอนที่ 1 มโนทัศน์ของทฤษฎีทางการสอบแบบดั้งเดิมและการประยุกต์ใช้..... 23
ตอนที่ 2 มโนทัศน์ของทฤษฎีการตอบสนองข้อสอบและการประยุกต์ใช้..... 26
ตอนที่ 3 มโนทัศน์ของการทำหน้าที่ต่างกันของข้อสอบ..... 31
3.1 ความลำเอียงและการทำหน้าที่ต่างกันของข้อสอบ..... 31
3.2 ประเภทของการทำหน้าที่ต่างกันของข้อสอบ..... 34
3.3 หลักการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ..... 37
3.4 ทฤษฎีการตอบสนองข้อสอบและการทำหน้าที่ต่างกันของข้อสอบ..... 38
3.5 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ..... 41
3.6 ประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ..... 45
3.7 ทฤษฎีการตอบสนองข้อสอบสำหรับรูปแบบการให้คะแนนแบบทวิวิภาค.. 46
ตอนที่ 4 มโนทัศน์ของขนาดอิทธิพล..... 47
4.1 การรายงานขนาดอิทธิพลของงานวิจัยในปัจจุบัน..... 47
4.2 ความหมายของขนาดอิทธิพล..... 48

หน้า หน้า

4.3 การตัดสินใจทางการวิจัย..... 48
4.4 ขนาดอิทธิพลกับการศึกษาการทำหน้าที่ต่างกันของข้อสอบ..... 50
ตอนที่ 5 เอกสารและงานวิจัยที่เกี่ยวข้อง..... 53
5.1 งานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบในประเทศ..... 53
5.2 งานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบต่างประเทศ..... 70
5.3 สรุปประเด็นปัญหาที่พบเกี่ยวกับการการทำหน้าที่ต่างกันของข้อสอบ..... 94
5.4 กรอบแนวคิดการวิจัย..... 99
บทที่ 3 วิธีดำเนินการวิจัย..... 102
ตอนที่ 1 ขั้นตอนการวิจัย..... 103
ตอนที่ 2 การจัดกระทำข้อมูลตามปัจจัยที่ศึกษา..... 106
ตอนที่ 3 การจำลองข้อมูล..... 108
ตอนที่ 4 การวิเคราะห์ข้อมูล..... 113
ตอนที่ 5 การเปรียบเทียบประสิทธิภาพของผลการตรวจสอบ..... 115
บทที่ 4 ผลการวิเคราะห์ข้อมูล..... 118
ตอนที่ 1 การคำนวณค่าเฉลี่ยของอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1..... 120
ตอนที่ 2 ผลการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างการวัดขนาดอิทธิพลตามเกณฑ์ที่ศึกษา 129
ตอนที่ 3 สรุปผลการเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1..... 152
ตอนที่ 4 ผลการศึกษาในกรณีข้อมูลเชิงประจักษ์..... 165
บทที่ 5 สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ..... 200
สรุปผลการวิจัย..... 203
อภิปรายผล..... 213
ข้อเสนอแนะ..... 220
รายการอ้างอิง..... 224

หน้า

ภาคผนวก..... 235

ภาคผนวก ก ค่าพารามิเตอร์ข้อสอบรายข้อ ภายใต้ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory model) ชนิด 2 พารามิเตอร์ (two-parameter)..... 236

ภาคผนวก ข ตัวอย่างผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติกโดยการทดสอบระดับนัยสำคัญ และการวัดขนาดอิทธิพลภายใต้ทฤษฎีการตอบสนองข้อสอบ ชนิด 2 พารามิเตอร์..... 251

ภาคผนวก ค Print out ผลการเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ และการวัดขนาดอิทธิพล..... 265

ภาคผนวก ง การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธีแมนเทล-แฮนด์เซลล์ วิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ และการวัดขนาดอิทธิพลในข้อมูลเชิงประจักษ์..... 279

ภาคผนวก จ ผลการวิเคราะห์ประสิทธิผลของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้านอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ และการวัดขนาดอิทธิพล..... 299

ประวัติผู้เขียนวิทยานิพนธ์..... 306

สารบัญตาราง

ตารางที่		หน้า
2.1	ฟังก์ชันทางคณิตศาสตร์ของโมเดลการตอบสนองข้อสอบ.....	39
2.2	วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับรูปแบบการตรวจให้คะแนนแบบทวิวิภาค ที่เป็นเอกมิติ จำแนกตามลักษณะของข้อมูล.....	43
2.3	วิธีการตรวจสอบการทำหน้าที่ต่างกันที่มีการให้คะแนนแบบทวิวิภาค (Dichotomous DIF) และพหุวิภาค (polytomous DIF).....	44
2.4	คุณภาพของการตรวจสอบประสิทธิภาพของการทำหน้าที่ต่างกันของข้อสอบ...	45
2.5	แสดงสัดส่วนของระดับการตอบสนอง k ระดับตามช่วงระดับความสามารถของ b ใช้ค่า b ระหว่าง ±3.00 เมื่อกำหนดค่า a เป็น .50, 1.00 และ 2.00.....	51
2.6	ค่า R^2 ของการวัดขนาดอิทธิพลในการทำหน้าที่ต่างกันของข้อสอบ.....	52
2.7	สรุปงานวิจัยที่ได้ศึกษาเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบในประเทศไทยตั้งแต่อดีตถึงปัจจุบัน.....	62
2.8	สรุปงานวิจัยที่ได้ศึกษาเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบในต่างประเทศถึงปัจจุบัน.....	81
4.1	ภาพรวมของการจำลองข้อมูลจำแนกตามปัจจัยและเงื่อนไขของปัจจัยที่แปรเปลี่ยน.....	120
4.2	การตรวจสอบคุณภาพข้อมูลจำลองตามจำแนกตามปัจจัยที่ศึกษา 24 เงื่อนไข	121
4.3	ร้อยละเฉลี่ย ($\overline{X}$) และส่วนเบี่ยงเบนมาตรฐาน (SD) ร้อยละของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในภาพรวม	122
4.4	ค่าเฉลี่ยร้อยละของอัตราความถูกต้องในทุกวิธีที่ศึกษาภายใต้วิธีถดถอยโลจิสติก.....	125
4.5	ค่าเฉลี่ยร้อยละของอัตราความคลาดเคลื่อนประเภทที่ 1 จำแนกตามวิธีที่ศึกษาภายใต้วิธีถดถอยโลจิสติก.....	126
4.6	การทดสอบ Box's Test และ Bartlett's Test .....…….....…....	130
4.7	การเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่1 ในวิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพลตามเกณฑ์ที่ศึกษาภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย.....	131

หน้า

ตารางที่

4.8 ผลการทดสอบระหว่างกลุ่ม ภายใต้เงื่อนไขปฏิสัมพันธ์สองทางระหว่าง วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบกับปัจจัยที่แปรเปลี่ยน…… 132

4.9 การทดสอบความแปรปรวนผลย่อยระหว่างวิธีการตรวจสอบการทำหน้าที่ ต่างกันของข้อสอบ กับปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน……..... 135

4.10 การทดสอบความแปรปรวนผลย่อยระหว่างวิธีการตรวจสอบการทำหน้าที่ ต่างกันของข้อสอบ กับปัจจัยความยาวของแบบสอบทั้งฉบับ 136

4.11 การทดสอบความแปรปรวนผลย่อยระหว่างวิธีการตรวจสอบการทำหน้าที่ ต่างกันของข้อสอบกับ ปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ..... 138

4.12 การทดสอบความแปรปรวนผลย่อยระหว่าง วิธีการตรวจสอบการทำหน้าที่ ต่างกันของข้อสอบ กับ ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันและ ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน..... 140

4.13 การทดสอบความแปรปรวนผลย่อยระหว่างวิธีการตรวจสอบการทำหน้าที่ ต่างกันของข้อสอบกับ ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกันและปัจจัย ความยาวของแบบสอบทั้งฉบับ……..... 143

4.14 การทดสอบความแปรปรวนผลย่อยระหว่างวิธีการตรวจสอบการทำหน้าที่ ต่างกันของข้อสอบกับ ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันและ ปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ…..……..... 146

4.15 การทดสอบความแปรปรวนผลย่อยระหว่างวิธีการตรวจสอบการทำหน้าที่ ต่างกันของข้อสอบ กับ ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน และปัจจัย ขนาดของการทำหน้าที่ต่างกันของข้อสอบ……..…… 149

4.16 ปัจจัยที่มีผลต่ออัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่1 ในวิธีถดถอยโลจิสติก ของการวัดขนาดอิทธิพล2 เกณฑ์..... 158

4.17 สรุปผลการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อน ประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอย โลจิสติก ระหว่างการวัดขนาดอิทธิพลตามเกณฑ์Jodoin and Gierl และเกณฑ์ Zumbo and Thomas ภายใต้เงื่อนไขเดียวกัน……..….. 160

หน้า

ตารางที่

4.18 สรุปผลการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl และเกณฑ์ Zumbo and Thomas ภายใต้เงื่อนไขต่างกัน..... 162

4.19 สถิติเชิงบรรยายของคะแนนจากแบบสอบรายวิชาคณิตศาสตร์……..…….. 165

4.20 ค่าความเที่ยงแบบความสอดคล้องภายในของแบบสอบโดยสูตรแอลฟา สัมประสิทธิ์ของครอนบาค วิชาคณิตศาสตร์ จำแนกตามกลุ่มผู้สอบ……... 166

4.21 ค่าความยาก (p) และอำนาจจำแนก (r) ของข้อสอบรายข้อ วิชาคณิตศาสตร์ (40 ข้อ)..... 166

4.22 ค่าพารามิเตอร์อำนาจจำแนก (a) และความยาก (b) ของข้อสอบรายข้อ วิชาคณิตศาสตร์..... 167

4.23 ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation matrix), KMO, ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานของแบบสอบวิชาคณิตศาสตร์ (n=123,167 คน)…….. 170

4.24 ค่าไอเกนและร้อยละของความแปรปรวนขององค์ประกอบของแบบสอบวิชาคณิตศาสตร์……..…….. 173

4.25 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวัดพื้นที่ของราชูวิชาคณิตศาสตร์..... 175

4.26 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้วิธีถดถอยโลจิสติกโดยการทดสอบระดับนัยสำคัญ ในวิชาคณิตศาสตร์……... 176

4.27 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติกโดยการวัดขนาดอิทธิพลในวิชาคณิตศาสตร์ ตามเกณฑ์ Zumbo and Thomas..... 177

4.28 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติกโดยการวัดขนาดอิทธิพลในวิชาคณิตศาสตร์ ตามเกณฑ์ Jodoin and Gierl…… 178

4.29 สรุปผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ..... 179

4.30 จำนวนข้อของการเกิดและไม่เกิดการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์..... 179

หน้า

ตารางที่

4.31 ร้อยละของอัตราความถูกต้องและร้อยละของอัตราความคลาดเคลื่อนประเภทที่ 1 จำแนกตามการตรวจสอบการทำหน้าที่ต่างกันของแบบสอบตามวิธีที่ศึกษาในวิชาคณิตศาสตร์..... 180

4.32 สถิติเชิงบรรยายของคะแนนจากแบบสอบรายวิชาวิทยาศาสตร์..... 181

4.33 ค่าความเที่ยงแบบความสอดคล้องภายในของแบบสอบโดยสูตรสัมประสิทธิ์แอลฟาของครอนบาค จำแนกตามวิชาวิทยาศาสตร์และกลุ่มผู้สอบ..... 181

4.34 ค่าความยาก (p) และอำนาจจำแนก (r) ของข้อสอบรายข้อ วิชาวิทยาศาสตร์ (50 ข้อ)..... 182

4.35 ค่าพารามิเตอร์อำนาจจำแนก (a) และความยาก (b) ของข้อสอบรายข้อวิชาวิทยาศาสตร์..... 183

4.36 ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation matrix), KMO, ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน ของข้อสอบวิชาวิทยาศาสตร์ (n=110,609 คน)..... 185

4.37 ค่าไอเกนและร้อยละของความแปรปรวนขององค์ประกอบในแบบสอบวิชาวิทยาศาสตร์..... 189

4.38 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธีการวัดพื้นที่ของราชูวิชาวิทยาศาสตร์..... 191

4.39 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้วิธีถดถอยโลจิสติกโดยการทดสอบระดับนัยสำคัญ ในวิชาวิทยาศาสตร์..... 192

4.40 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติกโดยการวัดขนาดอิทธิพลในวิชาคณิตศาสตร์ ตามเกณฑ์ของ Zumbo and Thomas..... 194

4.41 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติกโดยการวัดขนาดอิทธิพลในวิชาคณิตศาสตร์ ตามเกณฑ์ Jodoin and Gierl..... 195

4.42 สรุปผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาวิทยาศาสตร์..... 196

4.43 จำนวนข้อของการเกิดและไม่เกิดการทำหน้าที่ต่างกันของข้อสอบของวิชาวิทยาศาสตร์..... 196

หน้า

ตารางที่

4.44 ร้อยละของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่1 จำแนกตามการตรวจสอบการทำหน้าที่ต่างกันของแบบสอบตามวิธีที่ศึกษาวิชาวิทยาศาสตร์..... 197

4.45 ผลการตรวจสอบ DIF ระหว่างการทดสอบระดับนัยสำคัญ กับการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl..... 198

สารบัญภาพ

หน้า

ภาพที่

2.1	ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป (uniform DIF) ……..…..……	35
2.2	ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป (Nonuniform DIF) โดยมีปฏิสัมพันธ์ไม่เป็นลำดับ (Disordinal interaction)…….....	36
2.3	ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป (Nonuniform DIF) โดยมีปฏิสัมพันธ์เป็นลำดับ (Ordinal interaction) …….....	36
2.4	กรอบแนวคิดในการวิจัย กรณีศึกษาการจำลองข้อมูล……	100
2.5	กรอบแนวคิดในการวิจัย กรณีศึกษาข้อมูลเชิงประจักษ์…….....	101
3.1	ขั้นตอนการดำเนินการศึกษา.....	105
3.2	ขั้นตอนการจำลองข้อมูล…….....	111
3.3	แผนผังของการจำลองข้อมูล.....	112
4.1	เกณฑ์การพิจารณาประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ…….....	129
4.2	อัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน 2 เงื่อนไข.....	152
4.3	อัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ปัจจัยขนาดของการทำหน้าที่ต่างกัน 3 เงื่อนไข.....	153
4.4	อัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 2 เงื่อนไข.....	154
4.5	อัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ปัจจัยความยาวของแบบสอบทั้งฉบับ 2 เงื่อนไข.....	155
4.6	ผลการตรวจสอบความเป็นเอกมิติของแบบสอบวิชาคณิตศาสตร์ 50 ข้อ.....	173
4.7	ผลการตรวจสอบความเป็นเอกมิติของแบบสอบวิชาวิทยาศาสตร์ 50 ข้อ.....	189

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

การสอบเพื่อวัดความสามารถพัฒนาการของผู้เรียน ตามจุดมุ่งหมายการศึกษา (education objective) ต้องมีการกำหนดนโยบายตามจุดมุ่งหมายการวัดที่ชัดเจนว่าจุดประสงค์หลักที่ต้องการวัดคือสิ่งใดโดยใช้เครื่องมือวัดที่มีคุณภาพ เครื่องมือวัดทางการศึกษาและจิตวิทยา ส่วนใหญ่ใช้วัด พัฒนาการของผู้เรียน อาทิ แบบสอบ แบบสัมภาษณ์ แบบสังเกตและแบบสอบถาม แบบประเมินการเขียนและแบบประเมินภาคปฏิบัติ ผลการวัดถูกกำหนดค่าให้เป็นคะแนนซึ่ง เป็นเพียงตัวอย่างของพฤติกรรม แต่ไม่ใช่พฤติกรรมทั้งหมดของบุคคล เมื่อแปลผลของคะแนนแล้วต้องมีความหมายที่สื่อถึงคุณลักษณะภายในที่ต้องการจะวัดได้ถูกต้องตามความเป็นจริง คะแนนที่ได้เป็น คะแนนดิบไม่มีความสมบูรณ์มีคุณลักษณะเชิงสัมพัทธ์ (relative) ต้องเปรียบเทียบความหมายกับเกณฑ์มาตรฐานอื่นแล้วจึงแปลความหมายต่อ ได้ (ศิริชัย กาญจนวาสี , 2550) มีรูปแบบการตรวจให้คะแนน 2 แบบ คือ แบบทวิวิภาค (dichotomously scored items) และแบบพหุวิภาค (polytomously scored items) แบบทวิวิภาค เช่น ข้อสอบเลือกตอบ แบบจับคู่ แบบถูก -ผิด ส่วนการตรวจให้คะแนนแบบ พหุวิภาค จะให้คะแนนในรูปลำดับขั้น (ordinal scores) เช่น ข้อสอบภาคปฏิบัติ (performance assessment) การวัดตามสภาพจริง (authentic assessment) การตัดสินคุณภาพแฟ้มสะสมงาน (portfolio assessment) การพิสูจน์ทางคณิตศาสตร์ (mathematical proof) และการทดลองวิทยาศาสตร์ (scientific experiment) (Kim, Chosen, Alagoz and Kim, 2007)

การสร้างเครื่องมือต้องคำนึงถึงคุณภาพเครื่องมือ ในประเด็น วัดได้ตรงตามคุณลักษณะที่ต้องการวัดและเกิดความคลาดเคลื่อนในการวัดน้อยที่สุด การวัดมีความคลาดเคลื่อนเกิดขึ้นเสมอจึงต้องพัฒนาเครื่องมือให้มีคุณภาพสูงสุดเพื่อให้เกิดความคลาดเคลื่อนน้อยที่สุด (ศิริชัย กาญจนวาสี , 2550) ผลของการวัดไม่ได้ขึ้นอยู่กับรูปแบบของข้อคำถามแต่มุ่งเน้นความสำคัญไปที่รูปแบบของการให้คะแนน (Zumbo, 1999) เมื่อมีความเชื่อว่าผู้สอบทุกคนมีความสามารถเท่ากัน การสร้างข้อสอบต้องไม่ให้เกิดการได้ประโยชน์หรือเสียประโยชน์ระหว่างผู้สอบ ดังนั้น การวัดโดยอาศัยเครื่องมือเหล่านี้ต้องอยู่ในสภาพที่ปราศจากความลำเอียง (Allen andYen, 1979; Popham, 1981) คุณภาพที่สำคัญของเครื่องมือวัดทางการศึกษาและจิตวิทยา ประกอบด้วย ความตรง (validity) ความเที่ยง (reliability) ความยากง่าย (difficulty) อำนาจจำแนก (discrimination) มีความยุติธรรม (fairness) ปรนัย (objectivity)

ถามลึก (searching) ยั่วยุให้ตอบ (exemplary) จำเพาะเจาะจง (definition) มีประสิทธิภาพ (efficiency) มีความหมายในการวัด (meaningfulness) เหมาะสมในการนำไปใช้ (usability) คุณภาพที่สำคัญของเครื่องมือวัดประการแรกคือคุณภาพด้านความตรง (บุญธรรม กิจปรีดาบริสุทธิ์, 2543) ปัจจัยที่เกี่ยวข้องกับความตรง อาทิ ความเป็นปรนัย การเก็บรวบรวมข้อมูล การดำเนินการสอบ ความแตกต่างด้านความสามารถของกลุ่มผู้สอบที่ถูกวัดและความสัมพันธ์ระหว่างเครื่องมือกับเกณฑ์ เครื่องมือที่ขาดคุณภาพความตรงย่อมไม่ยุติธรรมเพราะไม่ทำให้ผลของการวัดสามารถสะท้อนถึงความสามารถที่แท้จริงของผู้สอบ กรณีนี้สรุปได้ว่าเครื่องมือที่ไม่สามารถวัดเฉพาะคุณลักษณะแฝงที่ต้องการแต่วัดคุณลักษณะที่ไม่ต้องการของผู้สอบแสดงว่าเครื่องมือขาดความตรง (Shealy and Stout, 1993) เหตุผลอีกประการหนึ่งที่ทำให้คุณภาพด้านความตรงมีความสำคัญที่สุดเกี่ยวกับการสรุปบุคคลจากคะแนนที่ได้จากการวัดโดยเครื่องมือนั้น ถ้าเครื่องมือให้ค่าคะแนนเอนเอียงเข้าข้างผู้สอบกลุ่มใดกลุ่มหนึ่งมากกว่าอีกกลุ่ม โดยผู้สอบทั้งสองกลุ่มมีความสามารถเท่าๆกัน กล่าวได้ว่าเครื่องมือเกิดความลำเอียง (bias) ถือว่าเป็นการละเมิดคุณภาพด้านความยุติธรรมไปโดยปริยาย (Shumacker, 2005)

โชติกา ภาษีผล (2554) กล่าวว่าความตรงของเครื่องมือประกอบด้วยปัจจัยต่างๆ ดังนี้ 1) ปัจจัยจากแบบสอบ คำสั่งไม่ชัดเจน ใช้โครงสร้างภาษาซับซ้อนเกินไป ใช้ภาษากำกวมจนเกิดความสับสนและตีความหมายผิด ใช้คำถามนำ มีคำหรือข้อความในข้อคำถามแนะคำตอบ ข้อสอบมีระดับความยากไม่เหมาะสมยากไปหรือง่ายไป เลือกใช้รูปแบบข้อสอบที่ไม่เหมาะสมกับพฤติกรรมการเรียนรู้ที่ต้องการ แบบสอบสั้นมีจำนวนข้อน้อยเกิน 2) ปัจจัยจากการบริหารการสอบและการตรวจให้คะแนน เวลาที่ใช้สอบไม่เหมาะสม สภาพแวดล้อมการสอบไม่เหมาะสม ขาดมาตรฐานคุมสอบ ไม่ปฏิบัติตามคำแนะนำในการคุมสอบ และการตรวจให้คะแนนขาดความเป็นปรนัย 3) ปัจจัยจากผู้สอบ ความเป็นเอกพันธ์ของกลุ่มผู้สอบ การเดาคำตอบ การจับทางแนวคำตอบได้ นิสัยส่วนตัวของผู้สอบในการทำข้อสอบ และสภาพความไม่พร้อมทางร่างกายและจิตใจ 4) ปัจจัยจากเกณฑ์ที่ใช้อ้างอิง ความชัดเจนของมวลเนื้อเรื่องที่มุ่งวัด ความเหมาะสมของการคัดเลือกเกณฑ์ที่นำมาเทียบ ความเหมาะสมของทฤษฎีที่เกี่ยวข้องกับลักษณะที่มุ่งวัด

คุณภาพด้านความยุติธรรมของเครื่องมือมีความสำคัญและจำเป็นไม่น้อยไปกว่าการตรวจสอบคุณภาพด้านความเที่ยง ความยากง่าย อำนาจจำแนกและคุณภาพด้านอื่น เพราะความ ไม่ยุติธรรมถือเป็นภาวะคุกคาม (threat) ต่อคุณภาพด้านความตรง ทำให้เกิดปัญหาด้านความลำเอียงของเครื่องมือทางการศึกษาและจิตวิทยา (Zumbo, 2005) อดีตเราเรียกข้อสอบที่ขาดความยุติธรรมว่า “ข้อสอบลำเอียง” (Item bias) เพราะมีกลุ่มที่ได้เปรียบและกลุ่มที่เสียเปรียบระหว่างผู้เข้าสอบ (Angoff, 1993) คำว่า “ความลำเอียงของข้อสอบ” เน้นการพิจารณาสังเกตอิทธิพลที่สังเกตได้ของกลุ่มย่อยที่นำมาศึกษา นอกจากนี้คำว่าความลำเอียงของข้อสอบยังสามารถตีความได้ทั้งทางสังคมและทางสถิติจึง เกิดความคลุมเครือในการนำไปใช้ (Hambleton, Swaminathan and Rogers, 1991) เมื่อนักการศึกษาได้พัฒนาวิธีการใหม่เพื่อคำนวณหาดัชนีในการตรวจวัด โดยการมุ่งเน้นความแตกต่างระหว่างผู้สอบต่างกลุ่มกันที่

ตอบข้อสอบข้อเดียวกัน เลือกใช้เกณฑ์ในการจับคู่กลุ่มผู้สอบ นักการศึกษาจึงใช้สารสนเทศทางสถิติมาเป็นเกณฑ์ตัดสิน ความลำเอียงของข้อสอบ คำว่า“การทำหน้าที่ต่างกันของข้อสอบ ” (Differential Item Functioning: DIF) ความหมายจึงเป็นกลางและเหมาะสมในเชิงวิชาการมากกว่าความลำเอียง (Holland and Wainer, 1993; Shealy and Stout, 1993; วลีมาศ แซ่อึ้ง, 2543)

ความยุติธรรมและความลำเอียงในการสอบนับเป็นประเด็นสำคัญของสาขาการทดสอบและประเมินผลโดยเฉพาะในทวีปอเมริกาเหนือมาตั้งแต่ทศวรรษ 1960 เป็นเรื่องจำเป็นที่จะใช้เครื่องมือวัดผลที่ถูกต้องและปราศจากความลำเอียงเพื่อใช้ตัดสิน ในหลายแง่มุมของการศึกษารวมถึงการสอบคัดเลือกต่างๆ วิธีการตรวจหา การทำหน้าที่ต่างกันของข้อสอบและความลำเอียงของข้อสอบ (item bias) มักนำมาใช้ทั่วไปในกระบวนการ พัฒนาเครื่องมือวัดใหม่ๆ โดยดัดแปลงจากของเดิมหรือการอนุมานความตรงของคะแนนสอบ วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจะช่วยตัดสินว่าข้อสอบข้อใด(หรือข้อสอบทั้งชุด) ทำหน้าที่ในลักษณะเดียวกันต่อกลุ่มผู้เข้าสอบที่หลากหลาย กล่าวโดยรวมก็คือเป็นเรื่องของการวัดความไม่แปรปรวนว่าข้อสอบทำหน้าที่แบบเดียวกันกับผู้เข้าสอบแต่ละกลุ่มหรือไม่ ปัจจุบันการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ แตกต่างกัน 5 จุดประสงค์ (Shimizu and Zumbo, 2005)

1. *เพื่อให้มั่นใจได้ว่าการสอบนั้นถูกต้องมีความตรง* แต่ละกลุ่มถูกกำหนดไว้แล้วตามนโยบายและระเบียบข้อบังคับ (เช่น ชนกลุ่มน้อยที่เห็นได้ชัด, เพศ, กลุ่มภาษา) การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเป็นแนวทางหนึ่งของการตรวจสอบความตรงของคะแนนที่ได้จากการสอบ และเป็นการตรวจสอบในประเด็นความยุติธรรมของข้อสอบ เพราะ ส่งผลต่อคุณภาพด้านความตรงในการแสดงความหมายเฉพาะคุณลักษณะที่แฝงอยู่ของคะแนนเหล่านั้น (Camilli and Shepard ,1994; Kim, Chosen, Alagoz and Kim, 2007; ศิริชัย กาญจนวาสี, 2550) Zumbo and Hubley (2003) ตั้งข้อสังเกตไว้ว่าจุดประสงค์นี้เป็นเหตุผลทั้งหมดของการพัฒนาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในยุคแรกๆ การใช้เครื่องมือนี้เป็นเรื่องปกติที่สุดในบริบทของการสอบคนจำนวนมาก ซึ่งบางคนอาจนำคะแนนจากการสอบไปใช้ในการตัดสินใจ เช่น คัดเลือกนักเรียนนักศึกษาเข้าทำงานหรือเข้าเรียนต่อในมหาวิทยาลัย โดยใช้การสอบภาษา ความกังวลเรื่องข้อสอบลำเอียงจะเกิดขึ้นในบริบทของการสอบที่มีการแข่งขันกันสูง การตัดสินใจจะเกี่ยวโยงไปถึงการสอบเพื่อจบการศึกษา สอบวัดทัศนะคติ สอบเพื่อรับประกาศนียบัตร สอบเพื่อขอใบอนุญาต ซึ่งต้องมีความถูกต้อง มีความตรงสูงสุด จากประวัติที่ผ่านมาความกังวลเรื่องข้อสอบลำเอียงจะมุ่งไปที่ความสามารถที่แตกต่างกันของกลุ่มคนตามเพศหรือเชื้อชาติ

2. *เพื่อเป็นหลักฐานหากมีการฟ้องร้อง* การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ สามารถนำไปใช้เป็นหลักฐานได้ในกรณีที่ผู้เข้าสอบที่ได้คะแนนที่ไม่น่าพอใจและสอบไม่ผ่านเกณฑ์อ้างว่ามีการเลือกปฏิบัติ กล่าวสั้นๆ ก็คือผลของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจะช่วยลดความเสี่ยงที่จะมีการฟ้องร้องในเรื่องการเลือกปฏิบัติ

3. *เพื่อใช้ตรวจสอบหากความยากและการเลือกปฏิบัติของข้อสอบเปลี่ยนไปตามกาลเวลา* ในการสอบใหญ่ๆ ที่ใช้ข้อสอบชุดเดิมซ้ำแล้วซ้ำอีกติดต่อกันมายาวนานจะมีข้อกังขาเกิดขึ้นว่าความยากและการเลือกปฏิบัติของข้อสอบจะเปลี่ยนแปลงตามกาลเวลาไปด้วยหรือไม่ กรณีเช่นนี้จะถูกเรียกกันว่า "ข้อสอบเบี่ยงเบน" การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ สามารถตรวจสอบว่าข้อสอบนั้นยังคงมีคุณสมบัติที่จะใช้ประเมินได้อยู่หรือไม่เมื่อใช้ซ้ำหลายครั้งแล้ว

4. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ *เพื่อสามารถเปรียบเทียบแต่ละกลุ่มและขัดขวางการวัดที่จัดกระทำขึ้นเองเพื่อสามารถอธิบายความแตกต่างของแต่ละกลุ่มได้* จุดประสงค์นี้โดยแก่นแล้วเกี่ยวข้องกับความเป็นไปได้ที่จะมี "การคุกคามถึงความตรงภายใน" ในการเปรียบเทียบกลุ่มต่างๆ กลุ่มเหล่านี้ถูกระบุไว้ล่วงหน้าแล้วและมักจะถูกกดดันจากคำถามงานวิจัยว่าที่ผู้วิจัยตั้งคำถามเอาไว้ (เช่น ความแตกต่างด้านเพศส่งผลต่อความสามารถด้านการสอบภาษา)

5. *เพื่อเข้าใจกระบวนการตอบสนองของการสอบ* ระยะหลังนี้มีความสนใจมุ่งไปยังการตรวจสอบกระบวนการรับรู้ของการตอบสนองข้อสอบและความสามารถในการสอบ และยังมีการตรวจสอบว่ามีกระบวนการเหล่านี้เหมือนกันหรือไม่ในแต่ละกลุ่มหรือแต่ละคน ในบริบทนี้กลุ่มทั้งหลายไม่จำเป็นต้องถูกระบุไว้ล่วงหน้าและใช้ชั้นเรียนที่แฝงเร้นแทนที่หรือใช้วิธีอื่นใดเพื่อ "ระบุ" หรือ "สร้าง" กลุ่มต่างๆ ขึ้นมา และ "กลุ่มใหม่" เหล่านี้จะถูกศึกษาเพื่อดูว่าจะสามารถใช้เรียนรู้เรื่องกระบวนการตอบสนองข้อสอบได้หรือไม่ แน่นอนว่าอาจจะตรวจสอบกระบวนการตอบสนองกับกลุ่มที่ไม่ได้รับผลกระทบ เช่น กลุ่มเพศ

การ ศึกษาอย่างกว้างขวางเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบนี้เอง ที่เป็นกระแสสร้างแรงกระตุ้นและทิศทางการพัฒนาการตรวจสอบอีกหลากหลายในอนาคต เช่น การสอบการแปลและการปรับตัวข้ามวัฒนธรรม (Zumbo, 2003) ตัวอย่างของการใช้การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในการสอบภาษาก็น่าจะเป็นการศึกษาผลกระทบของตัวแปรด้านภูมิหลัง เช่น ความมีระเบียบในการศึกษา วัฒนธรรม เศรษฐสถานะ กิจกรรมบันเทิงและงานอดิเรกที่มีต่อความสามารถในการทำข้อสอบ ถึงแม้ทุกวันนี้จะนิยมวิจารณ์กันว่าการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ไม่สามารถให้เหตุผลได้ถึงความสามารถในการสอบที่แตกต่างกัน แต่ก็เห็นได้ชัดเจนจากคำอธิบายเรื่องจุดประสงค์ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบก่อนหน้านี้แล้ว ว่าการวิจารณ์ดังกล่าวนั้นผิด จุดประสงค์เพราะไม่ใช่ว่าการศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทุกชิ้นงานจะมุ่งหา "เหตุผล" ให้วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ตัวอย่างเช่น อาจจะมีคนสนใจหาข้อสอบที่มีการทำหน้าที่ต่างกัน การปฏิบัติการสอบภาษา ดังนั้น "เหตุผล" เพื่อการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ จึงเป็นเรื่องรองที่จะรับประกันว่าเพียงพอที่จะอนุมานคะแนนสอบได้

การศึกษาที่ผ่านมาในเรื่องความลำเอียงของข้อสอบได้มุ่งเน้นความสามารถที่แตกต่างกันของแต่ละกลุ่ม เช่น ด้านเพศ Lumsden and Scott (1987) ทดสอบ t-test และสมการถดถอยพหุ (multiple

regression) เพื่อที่จะสรุปหาความแตกต่างด้านความสามารถระหว่างนักศึกษาเพศชายและเพศหญิงในการสอบเขียนเรียงความและข้อสอบแบบมีตัวเลือกในบริบทการศึกษาทางด้านเศรษฐศาสตร์ อีกตัวอย่างหนึ่งที่ใช้วิธีซับซ้อนมากขึ้นเพื่อสอบหาความลำเอียงในการศึกษาทางด้านเศรษฐศาสตร์ก็คืองานของ Waslstad and Robson (1997) และ Barrett (2001) โดยใช้เทคนิคของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในด้านการสอบภาษาบทสรุปของ Kunann (2000) ชี้ให้เห็นว่าการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในเรื่องความยุติธรรมในการสอบปรากฏขึ้นตอนต้นทศวรรษ 1980 โดยหลักแล้วมุ่งเน้นเรื่องความแตกต่างด้านเพศและภาษาแรกของผู้เข้าสอบที่แตกต่างกัน

การศึกษาเกี่ยวกับความยุติธรรมของข้อสอบมุ่งประเด็นไปที่การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อข้อสอบทำหน้าที่ต่างกันย่อมส่งผลกระทบต่อความตรงของแบบสอบ ถ้าใน แบบสอบมีจำนวนข้อสอบที่ทำหน้าที่ต่างกัน หลายข้อ จะส่งผลให้ ความตรงของแบบสอบลดน้อย ลง (Ackerman, 1992) ข้อสอบทำหน้าที่ต่างกันคือเกิดความสัมพันธ์ทางสถิติระหว่างผล การตอบข้อสอบใน กลุ่มผู้สอบเมื่อกำหนดระดับคะแนนที่ใช้เป็นเกณฑ์ในการจับคู่ทำให้เกิดรูปแบบ ของการทำหน้าที่ต่างกัน โดยพิจารณาจากขนาดและทิศทางซึ่งแปรเปลี่ยนไปตามระดับความสามารถที่ต่างกันของกลุ่มผู้สอบ Mellenbergh (1982) กล่าวว่ารูปแบบการทำหน้าที่ต่างกัน มี 2 ประเภท คือ แบบเอกรูป (uniform) เกิดเมื่อไม่มีปฏิสัมพันธ์ (interaction) ระหว่างระดับความสามารถของผู้สอบกับการเป็นสมาชิกของกลุ่มย่อย นั่นคือโอกาสของการตอบข้อสอบได้ถูกต้องของผู้สอบกลุ่มย่อยกลุ่มหนึ่งสูงกว่าอีกกลุ่มหนึ่งตลอดทุกช่วงความสามารถและแบบอเนกรูป (nonuniform) เกิดเมื่อไม่มีปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับการเป็นสมาชิกของกลุ่มย่อย นั่นคือโอกาสของการตอบข้อสอบได้ถูกต้องของผู้สอบกลุ่มย่อยกลุ่มหนึ่งสูงกว่าผู้สอบกลุ่มย่อยอีกกลุ่มหนึ่งไม่ตลอดทุกช่วงความสามารถ

วิธีการตรวจสอบการทำหน้าที่ต่างกันในปัจจุบันถือว่ามีการพัฒนาอยู่ในระดับที่ดีมาก ทั้งด้านรูปแบบของข้อสอบ วิธีการตอบและวิธีการให้คะแนน โดยมีการปรับแก้และพัฒนาวิธีการตรวจสอบให้มีความเหมาะสมและรัดกุมยิ่งขึ้น (Zwick, Donoghue and Grima, 1993) เมื่อการทำหน้าที่ต่างกันของข้อสอบเกิดจากความน่าจะเป็นในการตอบข้อสอบของผู้สอบที่มีความสามารถระดับเดียวกันแต่อยู่คนละกลุ่มกันทั้งที่ผู้สอบทุกคนมีคุณลักษณะที่ต้องการวัดเท่ากันจึงเกิดการได้เปรียบเสียเปรียบกันระหว่างกลุ่มผู้สอบ การเปรียบเทียบการตอบข้อสอบในกลุ่มผู้สอบ แบ่งเป็นกลุ่มอ้างอิง (reference groups: R) และกลุ่มเปรียบเทียบ (focal groups: F) โดยใช้การจับคู่ตามมโนทัศน์ที่สนใจศึกษา (Zumbo, 1999; Zumbo and Hubley, 2003) กลุ่มอ้างอิง เป็นกลุ่มที่ได้เปรียบในการตอบข้อสอบซึ่งคาดว่าได้ประโยชน์จากการตอบข้อสอบ กลุ่มเปรียบเทียบ เป็นกลุ่มที่นักวิจัยสนใจศึกษา โดยคาดว่าเป็นกลุ่มที่เสียเปรียบจากการตอบข้อสอบ (Holland and Thayer,1988) การเลือกว่ากลุ่มใดเป็นกลุ่มอ้างอิงหรือกลุ่มเปรียบเทียบขึ้นอยู่กับวัตถุประสงค์ของการศึกษา

การจับคู่ความสามารถ ของ ผู้สอบ แบ่งได้ 2 เกณฑ์ คือ ความสามารถที่สังเกตได้กับความสามารถแฝง ซึ่งเป็นความสามารถภาย ในตัวข องบุคคล ที่ไม่สามารถสังเกตได้ (Chang et al., 1996l) ความหมายตามสถิติของการทำหน้าที่ต่างกันของข้อสอบ สามารถเขียนได้ในรูปสมการถดถอย ของ ผล การ ตอบสนอง (Y) กับเกณฑ์ที่ใช้จับคู่ ความสามารถ (ความสามารถที่สังเกตได้ และความสามารถแฝง) การจับคู่ความสามารถของผู้สอบกรณีที่ผลการตอบรายข้อ เป็นแบบทวิวิภาค จะสรุปว่าข้อสอบไม่เกิด การทำหน้าที่ต่างกัน 2 ประเด็น เมื่อนำ คะแนนรวม จากแบบ สอบ 2 ฉบับมาจับคู่ระหว่างกลุ่มผู้สอบสองกลุ่ม (R, F) และเมื่อ นำคะแนนที่ประมาณได้ จากการตอบ ข้อสอบแต่ละข้อ (ความสามารถที่สังเกตได้) มาเป็นเกณฑ์ในการจับคู่ระหว่างกลุ่มผู้สอบสองกลุ่ม ซึ่งเป็น ความสามารถแฝง ข้อสอบไม่เกิด การทำหน้าที่ต่างกัน คือ ค่าความคาดหมายแบบมีเงื่อนไขของผลการตอบสนองข้อสอบสองกลุ่มจะเท่ากันทุกระดับความสามารถ (สุทธิพร สุรธณี, 2550)

การพัฒนาวิธีการตรวจสอบ การทำหน้าที่ต่างกันของข้อสอบ มีอย่างต่อเนื่อง การนำวิธีการตรวจสอบ การทำหน้าที่ต่างกันของข้อสอบ มาใช้ถือเป็นการพัฒนาวิธีการ ที่ก่อให้เกิดประโยชน์และมีความสำคัญต่อการพัฒนาให้เครื่องมือมีความตรงต่อ การวัด (Zumbo, 2005; Zumbo, 1999) สถิติสำหรับตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ แบ่งเป็น สถิติพาราเมตริก (parametric approach) เป็นสถิติที่ มีข้อตกลงเจาะจงเกี่ยวกับโมเดลการตอบสนองรายข้อ และ สถิตินันพาราเมตริก (nonparametric approach) เป็นสถิติที่ ไม่มีข้อตกลงเหมือนสถิติพาราเมตริก (Wangand Su, 2004) หากใช้ฐานแนวคิดตามทฤษฎีทางการสอบ (Test Theory) แบ่งวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบตามกลุ่มทฤษฎีทางการสอบเป็น 2 กลุ่ม ดังนี้

1) ทฤษฎีทางการสอบแบบดั้งเดิม (Classical Test Theory: CTT) เป็นแนวทาง หลักที่ใช้กันกว้างขวาง ค.ศ.1964 Cardalland Coffman (Angoff, 1993) นำวิธีวิเคราะห์ความแปรปรวน (Analysis of variance: ANOVA) มาประยุกต์ใช้ในการตรวจสอบ การทำหน้าที่ต่างกันของข้อสอบ ซึ่งเป็นการทดสอบปฏิสัมพันธ์ระหว่างข้อสอบกับผู้สอบ หากผลการทดสอบทางสถิติมีนัยสำคัญก็แสดงว่าแบบสอบทำหน้าที่ต่างกัน ผลที่เกิดนี้เป็นการตัดสินในภาพรวมของแบบสอบ ถือเป็นจุดเริ่มต้นของการพัฒนาวิธีการตรวจสอบ การทำหน้าที่ต่างกันของข้อสอบ วิธีวิเคราะห์ความแปรปรวน ดังกล่าวนำไปสู่การพัฒนาการตรวจสอบด้วยวิธีอื่น Angoff (1972) เสนอวิธีแปลงค่าความยากของข้อสอบ (Transformed item difficulty; TID) แนวคิดของวิธีนี้ใช้การทดสอบปฏิสัมพันธ์ระหว่างข้อสอบกับกลุ่มผู้สอบ (คล้ายกับเทคนิควิธีวิเคราะห์ความแปรปรวน) Scheuneman (1979) เสนอสถิติไค -สแควร์ (Chi-square: χ^2) พัฒนาเป็นอิสระ จากวิธีทฤษฎีการตอบสนองข้อสอบแต่มีเทคนิคการตรวจสอบคล้ายกัน หลักการตรวจสอบใช้ตารางการณ์จร (Contingency table) แบ่งคะแนนรวมของแบบสอบเป็นช่วงย่อยประมาณ 3-5 ช่วง คำนวณสัดส่วนของผู้สอบที่ตอบข้อสอบถูกในแต่ละช่วงคะแนนแล้วทดสอบด้วย χ^2 ซึ่ง Baker (1981) กล่าวถึงวิธีตรวจสอบของ Scheuneman ว่าผลที่ได้ไม่สอดคล้องกันเนื่องจากอิทธิพลของขนาด

ตัวอย่างระหว่างกลุ่มผู้สอบและการแจกแจงสุ่มไม่เหมือนกับไค -สแควร์ ดังนั้นจึงไม่มีคุณสมบัติเพียงพอที่จะนำไปใช้ทดสอบนัยสำคัญทางสถิติ ต่อมา Scheuneman (1981) ปรับแก้สูตรโดยนำผลการตอบข้อสอบผิดเข้าร่วมวิเคราะห์ด้วย เรียกวิธีนี้ว่า “ไค-สแควร์แบบเต็มรูป” (full chi-square) จุดเด่นของวิธีนี้คือกระบวนการตรวจสอบมีการจับคู่ความสามารถ ในกลุ่มผู้สอบก่อนที่จะเปรียบเทียบผลการตอบข้อสอบทำให้สามารถแก้ปัญหาความแตกต่างด้านความสามารถของผู้สอบได้แต่ยังไม่สมบูรณ์เพราะเป็นการเปรียบเทียบฟังก์ชันการตอบข้อสอบที่ไม่ละเอียด ถ้าแบ่งช่วงคะแนนกว้างระดับความสามารถระหว่างกลุ่มผู้สอบอาจไม่เท่าเทียมกัน การตรวจสอบอาจคลาดเคลื่อนจากความเป็นจริง แนวคิดวิธีนี้เกิดการพัฒนาวิธีตรวจสอบให้มีความเหมาะสมยิ่งขึ้นเป็นลำดับ (อรินทร์ น่วมถนอม, 2549)

ภายใต้ทฤษฎีการวัดแบบดั้งเดิมมีจุดอ่อนของข้อตกลงเบื้องต้น คือฐานความเชื่อเกี่ยวกับคะแนนความคลาดเคลื่อน ค่าพารามิเตอร์ข้อสอบและแบบสอบผันแปรไปตามกลุ่มผู้สอบ จึงประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบเพื่อ มาขยายแนวคิดของโมเดลการวัดแบบดั้งเดิม ช่วย คลายข้อตกลงเบื้องต้นและแก้ไขจุดอ่อนบางประการ (ศิริชัย กาญจนวาสี, 2550) ความก้าวหน้าของเทคโนโลยีทำให้มีเครื่องอำนวยความสะดวกในการคำนวณค่าทางสถิติสำหรับประมวลผลข้อมูลที่มีความซับซ้อนให้ได้ผลลัพธ์ที่ต้องการได้อย่างรวดเร็วและประหยัดเวลา ทำให้ปัจจุบันมีการศึกษาและพัฒนาวิธีการให้สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ภายใต้ทฤษฎีการวัดแบบดั้งเดิมในเงื่อนไขต่างๆ ได้อย่างมีประสิทธิภาพมากยิ่งขึ้น (Su and Wang, 2005)

2) ทฤษฎีการตอบสนองข้อสอบ (Item Respond Theory: IRT) เป็นทฤษฎีการทดสอบแนวใหม่ จำแนกได้ 2 กลุ่ม (ศิริชัย กาญจนวาสี, 2555) (1) ใช้การเปรียบเทียบค่าประมาณพารามิเตอร์ของข้อสอบระหว่างกลุ่มผู้สอบ มีหลักแนวคิดคือใช้การเปรียบเทียบค่าประมาณพารามิเตอร์ของข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ โดยการทดสอบนัยสำคัญทางสถิติ (2) ใช้การเปรียบเทียบค่าประมาณฟังก์ชันการตอบสนองข้อสอบระหว่างกลุ่มผู้สอบ โดยการวัดพื้นที่ระหว่างฟังก์ชันการตอบข้อสอบ มีหลักแนวคิดคือใช้ การวัดพื้นที่ของฟังก์ชันการตอบ สนองข้อสอบ (Item response function; IRFs) ระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบโดยตรง แล้วตัดสินการทำหน้าที่ต่างกันของข้อสอบโดยนำดัชนีที่คำนวณได้ไปเปรียบเทียบกับเกณฑ์ที่กำหนดขึ้นหรือใช้การทดสอบนัยสำคัญทางสถิติ การพัฒนาวิธีในกลุ่มนี้เริ่มต้นจาก Rudner (1977) ได้ประยุกต์ใช้สัดส่วนของการตอบข้อสอบถูกในแต่ละระดับคะแนนรวมแทนความน่าจะเป็นในการตอบข้อสอบถูกและใช้คะแนนรวมแทนความสามารถของผู้สอบแล้วนำไปลงจุดของแต่ละกลุ่มผู้สอบแยกกัน แล้ว เปรียบเทียบจุดเหล่านั้น เพื่อตัดสินว่าข้อสอบข้อใดทำหน้าที่ต่างกัน ต่อมาได้นำวิธีการดังกล่าวมาพัฒนาเป็นดัชนีการวัดพื้นที่แบบไม่คิดเครื่องหมายภายใต้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ จากนั้นนักวิจัยพัฒนาสูตรการวัดพื้นที่อาทิ การวัดพื้นที่ของ Linn และคณะ (1981) การวัดพื้นที่ของ Shepard; Camilli and Williams (1984) การวัดพื้นที่ของ Raju (1990) และการวัดพื้นที่ของ Kim and Cohen (1991) เป็นต้น นักวิจัยนิยมนำวิธีการตรวจสอบ การทำหน้าที่ต่างกันของ

ข้อสอบ โดยทฤษฎีการตอบสนองข้อสอบไปศึกษาอย่างกว้างขวาง แต่มีข้อจำกัด คือ ข้อมูลต้องเป็นไปตามข้อตกลงเบื้องต้นที่เข้มงวด ต้องใช้กลุ่มตัวอย่างขนาดใหญ่และการประมาณค่าพารามิเตอร์ภายใต้โมเดล 2 และ 3 พารามิเตอร์ ตรวจสอบหลายขั้นตอนและแปลผลยาก (Clauser and Mazor, 1998) การวิเคราะห์ต้องใช้โปรแกรมคอมพิวเตอร์ที่มีคำสั่งการใช้งานที่ซับซ้อน การประมาณค่าพารามิเตอร์ต้องเสียเวลาและมีค่าใช้จ่ายมาก จากข้อจำกัดของวิธีการตรวจสอบในกลุ่มทฤษฎีการตอบสนองข้อสอบดังกล่าว จึงพัฒนาวิธีการตรวจสอบโดยใช้ทฤษฎีการทดสอบแบบดั้งเดิมแต่ใช้ฐานคิดและหลักการตรวจสอบของทฤษฎีการตอบสนองข้อสอบ เพื่อสามารถตรวจสอบง่ายใช้ตัวอย่างขนาดเล็กแต่ มีความถูกต้องและแม่นยำสูง การคำนวณไม่ยุ่งยากและไม่ซับซ้อน

Gómez-Benito และคณะ (2009) ว่าการเลือกวิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ว่าในปัจจุบันไม่ได้มุ่งวิธีที่หลากหลายแต่มุ่งศึกษาเชิงลึกในวิธีการตรวจสอบที่มีอยู่เดิมเพื่อให้เกิดสารสนเทศมากยิ่งขึ้น วิธีที่นิยมนำมาใช้ศึกษามากสุดตั้งแต่ช่วงเริ่มต้นศึกษาการทำหน้าที่ต่างกันของข้อสอบ ซึ่งในปัจจุบันก็ยังคงได้รับความนิยมคือ วิธีแมนเทล -แฮนส์เซล (Mantel-Hanszel) (Holland and Thayer, 1988) วิธี SIBTEST (Shealy and Stout,1993) และวิธีถดถอยโลจิสติก (Swaminathan and Rogers, 1990) ทุกวิธีสามารถ ตรวจสอบ การทำหน้าที่ต่างกันของข้อสอบ แบบเอกรูป ได้ มีการคำนวณง่ายไม่ซับซ้อน การแปลผลไม่ยากใช้ได้กับกลุ่มตัวอย่างขนาดเล็กทำให้เสียค่าใช้จ่ายไม่มากแต่มีข้อจำกัด คือไม่อาศัยวิธีการประมาณค่าพารามิเตอร์ของผู้สอบและของข้อสอบแต่ต้องอาศัยคะแนนที่สังเกตได้เป็นหลัก (อ้างถึงใน Oshima and Morris, 2008) ตามบริบทของการสอบทั่วไป ข้อสอบไม่เกิดเฉพาะการทำหน้าที่ต่างกันแบบ เอกรูปเท่านั้น ทำให้ วิธีการตรวจสอบ โดยวิธีแมนเทล -แฮนส์เซล และวิธี SIBTEST ไม่ครอบคลุมบริบทสภาพจริงของข้อสอบ ใน 3 วิธีการตรวจสอบข้างต้นมีเพียงวิธีถดถอยโลจิสติกที่สามารถตรวจสอบได้ทั้งแบบเอกรูปและแบบอเนกรูปซึ่งครอบคลุมบริบทของข้อสอบมากกว่า ปัจจุบันจึงพัฒนาเทคนิคการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยวิธีถดถอยโลจิสติกมากขึ้น

วิธีถดถอยโลจิสติก (Logistic regression: LR) เป็นวิธีตรวจสอบที่ เสนอโดย Swaminathan และ Rogers (1990) โดยดัดแปลงจากวิธีล็อก-ลิเนียร์และวิธีแมนเทล-แฮนส์เซล มีกระบวนการตรวจสอบที่ใช้หลักเดียวกับวิธีที่ใช้ทฤษฎีการตอบสนองข้อสอบ ความสัมพันธ์ของวิธีนี้อยู่ในรูปสมการเส้นถดถอย (Regression Equation) มีจุดเด่นที่เป็นข้อได้เปรียบคือ การใช้โมเดลถดถอยโลจิสติกทำนายความน่าจะเป็นของการตอบข้อสอบถูกภายใต้คะแนนความสามารถ แบบต่อเนื่อง สามารถทดสอบอิทธิพลของปฏิสัมพันธ์ระหว่างระดับความสามารถกับการเป็นสมาชิกของกลุ่มผู้สอบจึงมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและแบบอเนกรูป (Li and Stout, 1996, Gómez-Benito and Hidalgo and Padilla, 2009, Jana Gomez-Benito; Hidalgo and Padillla, 2009) วิธีถดถอยโลจิสติก ยังใช้กับข้อสอบที่มีรูปแบบการให้คะแนนแบบทวิวิภาคและพหุวิภาค ใช้ตัวแปรความสามารถแบบต่อเนื่องทำให้มีความถูกต้องและ มีความแม่นยำในการตรวจสอบ (Zumbo, 1999)

ส่วนข้อจำกัด คือตรวจสอบภายใต้กรอบแนวคิดที่วัดความสามารถมิติเดียว (อรินทร์ น่วมถนอม, 2549) เมื่อตรวจสอบข้อสอบที่วัดความสามารถหลายมิติจะใช้คะแนนรวมแทนความสามารถคำนวณมาจากค่าเฉลี่ยของคะแนนในทุกมิติ คะแนนรวมดังกล่าวเป็นตัวแทนมิติเดียวทำให้ไม่สามารถจำแนกระดับความสามารถของผู้สอบทุกมิติได้อย่างแท้จริง

หลักการของวิธีถดถอยโลจิสติก ทำนายค่าความน่าจะเป็นของการเกิดหรือไม่เกิดเหตุการณ์ที่สนใจ ตัวแปรตามเป็นคะแนนการตอบซึ่งมีสองค่า (Dichotomous Variable) ส่วนตัวแปรอิสระอาจมีตัวเดียวหรือหลายตัว ทำการประมาณค่า Odds แปลงค่า Odds ให้อยู่ในรูป Logit แล้วประมาณค่าสัมประสิทธิ์ของตัวแบบโดยใช้ Maximum Likelihood ค่าสัมประสิทธิ์จะถูกนำไปใช้ในการคำนวณหาค่าความน่าจะเป็นของเหตุการณ์ที่สนใจด้วยการคำนวณหาผลต่างหรือการเปลี่ยนแปลงระหว่าง log odds ของตัวแปรตามเมื่อตัวแปรอิสระแต่ละตัวเปลี่ยนแปลงไป 1 หน่วย การทดสอบความเหมาะสมของตัวแบบหรือที่เรียกว่า Goodness of Fit of the Model นั้น สามารถทำนายค่าความน่าจะเป็นของการเกิดเหตุการณ์ได้ดีที่สุดโดยการคำนวณค่าไค-สแควร์ (Chi-square) ทั้งกรณีที่มีและไม่มีตัวแปรอิสระอยู่ในสมการเรียกว่าค่า -2LL หรือค่า -2log likelihood การวิเคราะห์โดยวิธี ถดถอยโลจิสติก ไม่มีข้อกำหนดเกี่ยวกับตัวแปรตามและค่าความคลาดเคลื่อนที่ต้องมีการแจกแจงแบบปกติ ไม่มีข้อกำหนดของความสัมพันธ์เชิงเส้นตรงระหว่างตัวแปรอิสระและตัวแปรตามเพียงแต่ต้องมีความสัมพันธ์เชิงเส้นตรงระหว่างตัวแปรอิสระที่เป็น Continuous กับค่า Logit ซึ่งสามารถวิเคราะห์ข้อมูลได้ทุกระดับ ของการวัดนับว่าเป็นจุดแข็งของสถิตินี้

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบอาจเกิดความคลาดเคลื่อนจากข้อมูล เมื่อศึกษากับกลุ่มผู้สอบที่มีจำนวนมาก การตรวจสอบมักจะพบค่าสถิติมีนัยสำคัญเสมอ ทำให้ต้องสรุปว่าข้อสอบทำหน้าที่ต่างกัน สอดคล้องกับ Kim, Chosen, Alagoz and Kim (2007) ที่พบว่าเมื่อกลุ่มตัวอย่าง มีขนาดใหญ่จะมีความไวในการตรวจพบการทำหน้าที่ต่างกันของข้อสอบค่อนข้างสูง แม้จะเลือกวิธีใดมาตรวจสอบก็จะได้รับสารสนเทศที่ว่าข้อสอบทำหน้าที่ต่างกันและสอดคล้องกับ Kim (2000) ที่กล่าวว่าการใช้กลุ่มตัวอย่างที่มีขนาดใหญ่เกินไปจะไม่มีประโยชน์ต่อการตรวจสอบ เพื่อแก้ไขความคลาดเคลื่อนต่างๆที่อาจเกิดขึ้น จึงใช้ผลการวัดขนาดอิทธิพล (effect size measure) มาช่วยในการตัดสินประสิทธิภาพของข้อสอบเพื่อ ให้มีความถูกต้องรัดกุมมากยิ่งขึ้น สอดคล้องกับ Gómez-Benito, Hidalgo and Padilla (2009) ที่ศึกษาประสิทธิภาพการตรวจสอบ การทำหน้าที่ต่างกันของข้อสอบ ด้วยการทดสอบระดับนัยสำคัญใน วิธีถดถอยโลจิสติก กับการผลการวัด ขนาดอิทธิพล ที่ใช้สถิติ R^2 ของ Nagelkerke ตามเงื่อนไขที่แตกต่างกันด้านขนาดกลุ่มตัวอย่าง อัตราส่วนระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ขนาดของการทำหน้าที่ต่างกัน ความยาวของข้อสอบและ ด้านร้อยละของจำนวนข้อสอบที่ ทำหน้าที่ต่างกัน พบว่า การวัดขนาดอิทธิพลได้ร้อยละของอัตราความถูกต้องต่ำกว่าที่ได้จากการวัดระดับนัยสำคัญ แต่ก็ให้อัตราความคลาดเคลื่อนที่ต่ำกว่าด้วย (อัตราความคลาดเคลื่อนน้อยจะให้ผลของประสิทธิภาพที่ดี)

แม้ว่าการทดสอบที่มีนัยสำคัญในวิธีถดถอยโลจิสติก สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ดีเมื่อขนาดกลุ่มตัวอย่างเล็กกว่า แต่ก็ควรพิจารณาอัตราความคลาดเคลื่อนของสองวิธีร่วมกันนำผลที่ได้มาสนับสนุนให้พิจารณาขนาดอิทธิพลร่วมกับผลของการทดสอบระดับนัยสำคัญทางสถิติ

การศึกษาขนาดอิทธิพลมีสิ่งที่ควรพิจารณา 2 ประการ (Kim, Chosen, Alagoz and Kim, 2007) คือ 1) สามารถช่วยตอบคำถามเกี่ยวกับความสัมพันธ์ระหว่างตัวแปรที่ศึกษาว่ามีขนาดเล็กมากจนกระทั่งไม่ค่อยมีความหมายหรือใหญ่เกินไปจนน่าจะต้องจัดการอะไรบางอย่าง เนื่องจากขนาดอิทธิพลมีความจำเป็น ในกรณีมีปัญหาเมื่อกลุ่มตัวอย่างมีขนาดใหญ่มากจะมีความไวในการตรวจสอบสูงมาก และ 2) ประเด็นเกี่ยวกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ Zumbo and Hubley (1998) เสนอว่านอกจากใช้ผลการทดสอบระดับนัยสำคัญทางสถิติ (Test of Significance) ควรมีการวัดระดับความเข้มของการทำหน้าที่ต่างกันของข้อสอบหรือการวัดขนาดอิทธิพลร่วมกัน Jodoin and Gierl (2001) ศึกษาประสิทธิภาพของการวัดขนาดอิทธิพลในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ พบว่าการใช้ขนาดอิทธิพลโดยสถิติ R^2 ในวิธีถดถอยโลจิสติก ร่วมกับการทดสอบนัยสำคัญจะได้ค่าของการสรุปผิดลดลงเกือบเป็นศูนย์ในประเด็นที่ว่าข้อสอบทำหน้าที่ต่างกันทั้งที่ความเป็นจริงข้อสอบไม่ได้ทำหน้าที่ต่างกัน (False Positive: FP) Kim, Chosen, Alagoz and Kim (2007) ได้ศึกษาขนาดอิทธิพลของการทำหน้าที่ต่างกันของข้อสอบในข้อมูลเชิงประจักษ์ เพื่อตรวจสอบความสอดคล้องระหว่างวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 5 วิธี คือ ใช้โมเดล IRT ที่ต่างกันระหว่างวิธีถดถอยโลจิสติก วิธีแมนเทล และวิธีแมนเทล-แฮนด์เซล โดยศึกษาจากข้อมูลการประเมินที่มีการให้คะแนนแบบพหุวิภาคในกลุ่มตัวอย่างที่มีขนาดใหญ่ พบว่าการทดสอบขนาดอิทธิพลด้วยสถิติ R^2 ของ Nagelkerke เกิดประโยชน์ต่อการแก้ปัญหาด้านขนาดของกลุ่มตัวอย่างที่ใหญ่เกินไปทำให้ค่าสถิติมีนัยสำคัญ ดังนั้นควรพิจารณาขนาดอิทธิพลร่วมด้วย สอดคล้องกับ Juana, Dolores and Joseluis (2009) ที่ศึกษาประสิทธิภาพของขนาดอิทธิพลในการพัฒนาวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติกโดยการจำลองข้อมูล ผลการวิจัยสนับสนุนให้ศึกษาการวัดขนาดอิทธิพลโดยสถิติ R^2 ร่วมกับการทดสอบนัยสำคัญทางสถิติจะทำให้ได้สารสนเทศมากยิ่งขึ้น

การวัดขนาดอิทธิพลเกี่ยวข้องโดยตรงกับการเลือกใช้เกณฑ์ในการพิจารณาขนาดของการทำหน้าที่ต่างกันของข้อสอบ เนื่องจากเป็นตัวช่วยในการพิจารณาข้อสอบว่าทำหน้าที่ต่างกันหรือไม่ Zumbo and Thomas (1997) กำหนดดัชนีทดสอบสถิติไค -สแควร์ (Chi-square) หรือ χ^2 ที่มี df = 2 โดยใช้อัตราส่วนไลด์ลิฮู้ด (Likelihood Ratio Test) ด้วยวิธีถดถอยโลจิสติก ร่วมกับการวัดขนาดอิทธิพลของการทำหน้าที่ต่างกันของข้อสอบ เน้นประเด็นปัญหาเมื่อค่าสถิติมีนัยสำคัญเมื่อกลุ่มตัวอย่างมีขนาดใหญ่

Zumbo and Thomas (1997) อ้างอิงสูตรการวัดขนาดอิทธิพลโดยใช้ R^2 อย่างเป็นคู่ขนานกับการวัดขนาดอิทธิพลในสถิติตัวอื่น การจัดประเภทของข้อสอบที่ทำหน้าที่ต่างกันโดย ใช้เกณฑ์ของ Zumbo and Thomas ได้กำหนดขนาดอิทธิพล 3 ระดับ ได้แก่ อิทธิพลของการทำหน้าที่ต่างกันขนาด

เล็กน้อย เกิดเมื่อค่าความแตกต่างของ R^2 มีค่าน้อยกว่า 0.13 ($\Delta R^2 < .13$) อิทธิพลของการทำหน้าที่ต่างกันปานกลาง เกิดเมื่อค่าความแตกต่างของ R^2 มีค่าระหว่าง 0 .13 - 0.26 ($.13 \leq \Delta R^2 \leq .26$) อิทธิพลของการทำหน้าที่ต่างกันขนาดใหญ่ เกิดเมื่อค่าความแตกต่างของ R^2 มีค่ามากกว่า .26 ($\Delta R^2 > .26$) ซึ่ง Gierl และ Bruno (Zumbo and Bruno, 1999) ก็ใช้เกณฑ์นี้เช่นกันในการวัดและประเมินทางการศึกษาเกี่ยวกับการตัดสินการทำหน้าที่ต่างกันของข้อสอบหรือแบบสอบ โดยนำเกณฑ์นี้ไปใช้ตัดสินขนาดอิทธิพลของการทำหน้าที่ต่างกันของข้อสอบจากการตรวจสอบด้วยวิธีต่างๆ ได้แก่ วิธีแมนเทล-แฮนส์เซล และวิธีซิบเทสท์ เพื่อใช้ตัดสินข้อสอบ นำไปสู่การสนับสนุนคุณภาพด้านความตรงของการวัด

Jodoin and Gierl (2001) กล่าวว่าเกณฑ์ของ Zumbo and Thomas นี้ยังมีความไวไม่เพียงพอในการวัดขนาดอิทธิพลของการทำหน้าที่ต่างกันของข้อสอบ จึงกำหนดเกณฑ์ตัดสินขนาดอิทธิพลของขึ้นใหม่ ใน 3 ระดับคือ ขนาดเล็กน้อย เกิดเมื่อค่าความแตกต่างของ R^2 มีค่าน้อยกว่า 0.035 ($\Delta R^2 < .035$) ขนาดปานกลาง เกิดเมื่อค่าความแตกต่างมีค่าระหว่าง 0.035 ถึง 0.07 ($.035 \leq \Delta R^2 \leq .07$) และขนาดใหญ่ เกิดเมื่อค่าความแตกต่างของ R^2 มีค่ามากกว่า 0.07 ($\Delta R^2 > .07$)

จากความสำคัญดังกล่าว ทำให้ผู้วิจัยมุ่งศึกษาประสิทธิภาพของ การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้วิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl และขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas ในข้อสอบที่มีรูปแบบการตรวจให้คะแนนแบบทวิวิภาคในข้อมูลจำลอง (simulation data) และข้อมูลเชิงประจักษ์ (empirical data) ดังนี้ (1) กรณีข้อมูลจำลอง ผู้วิจัยจำลองข้อมูลโดยใช้วิธีการทางคณิตศาสตร์และสถิติมาช่วยในการสร้างสถานการณ์ที่ซับซ้อน ผลจากเงื่อนไขของปัจจัยที่กำหนดขึ้นคาดว่าจะให้ผลครอบคลุมในหลายกรณี ดังนั้น การศึกษาข้อมูลจำลองจะช่วยให้สามารถศึกษาการเกิดเหตุการณ์ได้หลายเงื่อนไข การศึกษาครั้งนี้ผู้วิจัยเน้นศึกษาประสิทธิภาพการตรวจสอบด้วย วิธีถดถอยโลจิสติก ภายใต้เงื่อนไขปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน (DIF type) ขนาดของการทำหน้าที่ต่างกัน (amount of DIF) จำนวนข้อสอบที่มีการทำหน้าที่ต่างกัน (item with DIF) และความยาวของแบบสอบทั้งฉบับ (Test length) เพื่อศึกษาว่าปัจจัยที่แปรเปลี่ยนเหล่านี้ส่งผลต่อประสิทธิภาพ (efficacy) ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ หรือไม่ โดย พิจารณา ประสิทธิภาพจากอัตราความถูกต้อง (correct identification) และอัตราความคลาดเคลื่อนประเภทที่ 1 (type I error rate) ควรทำการศึกษา ข้อมูลจำลองควบคู่กับข้อมูลเชิงประจักษ์ ช่วยทำให้เกิดผลลัพธ์อย่างครบถ้วนและสมบูรณ์ยิ่งขึ้น (Harwell. M, 1996) และ (2) กรณีข้อมูลเชิงประจักษ์จากโครงการสอบระดับชาติของสถาบันแห่งหนึ่ง ประจำปี พ.ศ. 2552 มีรูปแบบการตรวจให้คะแนนแบบทวิวิภาค ระดับชั้นประถมศึกษาปีที่ 6 จากโรงเรียนทุกสังกัด วิชาวิทยาศาสตร์และวิชาคณิตศาสตร์

การศึกษาครั้งนี้ ใช้วิธีแมนเทล -แฮนส์เซล (Holland and Thayer, 1988) เป็นวิธีเกณฑ์ในการตรวจสอบ เนื่องจาก เป็นวิธีการตรวจสอบที่จัดในประเภท กลุ่มวิธีที่ใช้คะแนนที่สังเกตได้ (Observed

Score) วิเคราะห์ตามทฤษฎีทางการสอบแบบดั้งเดิม เรียกกลุ่มที่ไม่ใช้ทฤษฎีการตอบสนองข้อสอบ (Non-IRT Approach) เงื่อนไขของวิธีการ คือ ใช้คะแนนรวมของผู้สอบเป็นเกณฑ์การจับคู่ตามความสามารถของกลุ่มผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ เทคนิคทางสถิติเชื่อถือได้ มีการทดสอบทางสถิติแบบนอนพาราเมตริก (nonparametric) โดยการ ทดสอบนัยสำคัญและวัดขนาดอิทธิพล การคำนวณไม่มีความยุ่งยากซับซ้อน ใช้ได้กับกลุ่มตัวอย่างขนาดเล็ก ก็เพียงพอ จึงทำให้ เสียค่าใช้จ่ายน้อย (Hambleton, 1986 อ้างถึงใน กาญจนา วัจนสุนทร, 2538) ซึ่งวิธีแมนเทล -แฮนส์เซล เป็นที่ยอมรับจากนักวิจัยอย่างกว้างขวาง และได้รับความนิยมในการใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่มีรูปแบบการให้คะแนนแบบทวิวิภาค เช่น หน่วยงานบริการทดสอบทางการศึกษาแห่งสหรัฐอเมริกา (Educational Testing Service : ETS) แนะนำให้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซลและเป็นวิธีมาตรฐานที่ใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในโครงการสำคัญ ๆ ของหน่วยงาน

กลุ่มตัวอย่าง (sample size) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบ่งเป็น กลุ่มอ้างอิง (Reference group: R) และกลุ่มเปรียบเทียบ (focal group: F) ซึ่งเป็นกลุ่มเป้าหมายที่ต้องการศึกษาโดยเทียบกับกลุ่มอ้างอิง เมื่อข้อสอบทำหน้าที่ต่างกันคาดว่า กลุ่มอ้างอิง จะได้เปรียบในการตอบข้อสอบและ กลุ่มเปรียบเทียบ คาดว่าจะเสียเปรียบในการตอบข้อสอบ ความสำคัญของขนาดกลุ่มตัวอย่างเกี่ยวข้องกับอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 สำหรับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบตาม วิธีแมนเทล -แฮนส์เซล วิธีซิบเทสท์ และวิธีถดถอยโลจิสติก เมื่อขนาดตัวอย่างเพิ่มขึ้นมีผลทำให้อัตราความถูกต้องของวิธีการตรวจสอบมีค่าเพิ่มขึ้น นั่นคือจำนวนสมาชิกกลุ่มอ้างอิงและกลุ่มเปรียบเทียบเพิ่มขึ้นด้วย วิธีดังกล่าวจำเป็นต้องใช้ขนาดกลุ่มตัวอย่างที่เพียงพอ ดังนั้นขนาดตัวอย่างมีผลต่ออัตราความถูกต้องของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้งสองกลุ่ม การวิจัยครั้งนี้ศึกษาขนาดกลุ่มตัวอย่างโดยพิจารณาจากงานวิจัยที่ผ่านมาจึงกำหนดขนาดกลุ่มตัวอย่างให้มีขนาดเดียวคือจำนวน 2,000 คน แบ่งออกเป็นกลุ่มอ้างอิง จำนวน 1,000 คน และกลุ่มเปรียบเทียบ จำนวน 1,000 คน การศึกษาสถานการณ์จำลอง มีการจัดกระทำกับข้อมูลดังต่อไปนี้

รูปแบบของข้อสอบทำหน้าที่ต่างกัน (DIF type) ขนาดและทิศทางของการทำหน้าที่ต่างกันจะแปรเปลี่ยนไปตามระดับความสามารถที่แตกต่างกันของผู้สอบ การวิจัยครั้งนี้สนใจศึกษารูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน 2 รูปแบบ คือ แบบเอกรูปและแบบอเนกรูป การจำลองข้อมูล ที่มีวิธีการให้คะแนนรายข้อเป็นแบบทวิวิภาค ในเงื่อนไขดังกล่าวใช้โมเดล ของทฤษฎีการตอบสนองข้อสอบชนิด 2 พารามิเตอร์

ขนาดของการทำหน้าที่ต่างกัน (amount of DIF) ขนาดของการทำหน้าที่ต่างกันของข้อสอบหรือพื้นที่รวมแตกต่างระหว่างโค้งคุณลักษณะข้อสอบ 2 กลุ่ม การวิจัยครั้งนี้สนใจศึกษาขนาดอิทธิพลใน 3 ขนาด คือ ขนาด 0.1, 0.2 และ 0.4 (Gómez-Benito, Hidalgo and Padilla, 2009)

ความยาวของแบบสอบทั้งฉบับ (Test length) การศึกษาของ Swaminathan and Rogers (1990) พบว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและแบบอเนกรูป เมื่อใช้แบบสอบที่มีความยาวมากขึ้นจะมีผลทำให้อำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซลและวิธีถดถอยโลจิสติกมีค่ามากขึ้น ยกเว้นในกรณีแบบอเนกรูปของวิธีแมนเทล-แฮนส์เซล และทำการศึกษาใหม่อีกครั้ง (Rogers and Swaminathan, 1993) พบว่าให้ผลขัดแย้งกับครั้งแรก เนื่องจากความยาวของแบบสอบไม่มีผลต่ออำนาจการทดสอบของวิธีแมนเทล-แฮนส์เซล และวิธีถดถอยโลจิสติก ยกเว้นในกรณีแบบอเนกรูปของวิธีถดถอยโลจิสติก สาเหตุที่เลือกแบบสอบที่มีจำนวน 40 และ 50 ข้อ เนื่องจาก Narayanan and Swaminathan (1994) และ Narayanan and Swaminathan (1996) พบว่าการจัดกระทำกับข้อมูลในด้านความยาวของเครื่องมืออาจไม่ต้องกำหนดเงื่อนไขที่หลากหลาย เนื่องจากที่ระดับความยาว 40 ข้อนั้น แม้จะเป็นตัวแทนของการทดสอบผลสัมฤทธิ์ทางการเรียนสั้นๆ แต่มีความน่าเชื่อถือที่ได้มาตรฐาน และสอดคล้องกับการศึกษาที่ผ่านมาพบว่าข้อสอบที่มีความยาวปานกลางขึ้นไปจะส่งผลต่อประสิทธิภาพในการตรวจสอบมากที่สุดและเป็นระดับความยาวที่เหมาะสมกับการนำไปใช้เก็บข้อมูลจริง (จิติมาวรรณศรี, 2539; ญาณภัทร สีหะมงคล, 2540; ปิยะทิพย์ ตินวร, 2549; Kim and Cohen, 1998) การวิจัยครั้งนี้ จึงศึกษาข้อสอบที่มีความยาวทั้งฉบับ 40 และ 50 ข้อ ซึ่งเป็นขนาดความยาวของเครื่องมือการประเมินทางด้านจิตวิทยาทั่วไปที่มีการจัดสอบในสภาพที่สอดคล้องกับความเป็นจริง

จำนวนข้อสอบที่ทำหน้าที่ต่างกัน (the number of item with DIF) การศึกษาที่ผ่านมาพบว่าจำนวนข้อสอบที่ทำหน้าที่ต่างกัน สามารถลดความตรงในการจับคู่ระหว่างตัวแปรและยังส่งผลกระทบไปถึงค่าของอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 หากแบบสอบมาตรฐานวัดผลสัมฤทธิ์ทางการเรียนมีข้อสอบทำหน้าที่ต่างกันทั้งฉบับ คิดเป็นร้อยละ 10 ถึงร้อยละ 15 ยังคงถือว่าไม่ผิดปกติ หากแบบสอบมาตรฐานวัดผลสัมฤทธิ์ทางการเรียนมีข้อสอบทำหน้าที่ต่างกันทั้งฉบับ ร้อยละ 20 ขึ้นไปถือว่า อยู่ในภาวะ ไม่ปกติ (Narayananand Swaminathan, 1994; citing Clauser, 1993) การวิจัยครั้งนี้ศึกษาจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10 และ 20 จากความยาวของแบบสอบทั้งฉบับ 2 ขนาด คือ 40 และ 50 ข้อ สำหรับแบบสอบที่มีความยาว 40 ข้อ มีจำนวนข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 10 และ 20 จะมีข้อสอบที่ทำหน้าที่ต่างกัน 4 และ 8 ข้อ ส่วนแบบสอบที่มีความยาว 50 ข้อ มีจำนวนข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 10 และ 20 จะมีข้อสอบที่ทำหน้าที่ต่างกัน 5 และ 10 ข้อ

การวิจัยครั้งนี้ มีข้อจำกัดที่ผู้วิจัยไม่ได้จัดกระทำตามเงื่อนไขใดๆ ในข้อมูลเชิงประจักษ์เพราะต้องการใช้ผลการสอบจากสถานการณ์ที่เป็นจริงของการจัดสอบระดับประเทศ ภายใต้บริบทที่มีผู้เข้าสอบจำนวนมาก เพื่อมุ่งศึกษาว่าประสิทธิภาพการตรวจสอบด้วยวิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพลตามเกณฑ์ที่ศึกษาให้ผลเช่นใด เพื่อนำสารสนเทศจากผลดังกล่าวไปเป็นแนวทางสำหรับปรับปรุงข้อสอบให้มีความยุติธรรมต่อกลุ่มผู้สอบ นำไปสู่การออกแบบการวัดให้สามารถประเมินและ

ตัดสินผลตามความเหมาะสมกับสภาพการสอบในปัจจุบัน เพื่อช่วยพัฒนาคุณภาพของเครื่องมือที่ใช้ทางการศึกษาและจิตวิทยาต่อไป

คำถามวิจัย

การศึกษาแนวคิด ทฤษฎี หลักการและงานวิจัยเกี่ยวกับการตรวจสอบการทำหน้าที่ ต่างกันของข้อสอบ ผู้วิจัยจึงมุ่ง ศึกษาการวัดขนาดอิทธิพลและผลของประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในวิธีถดถอยโลจิสติก สำหรับข้อสอบที่มีรูปแบบการตรวจให้คะแนนแบบทวิวิภาค กรณีข้อมูลจำลองและข้อมูลเชิงประจักษ์ โดยนำแนวทางจากการศึกษาข้อมูลจำลองไปตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในข้อมูลเชิงประจักษ์ ทั้งนี้เพื่อตอบคำถามการวิจัย 3 ข้อ คือ

1. การศึกษาข้อมูลจำลองตามเงื่อนไขของ ปัจจัยและปฏิสัมพันธ์ ระหว่าง เงื่อนไขของ ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย จะมีผลต่ออัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl และการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas หรือไม่

2. การวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl และตามเกณฑ์ Zumbo and Thomas ภายใต้ การศึกษาข้อมูลจำลองตามเงื่อนไขของ ปัจจัยและปฏิสัมพันธ์ ระหว่าง เงื่อนไขของ ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย จะมีอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแตกต่างกันหรือไม่

3. การวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl และตามเกณฑ์ Zumbo and Thomas เมื่อนำไปตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในข้อมูลเชิงประจักษ์ จะมีอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกันหรือไม่

วัตถุประสงค์ของการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์สำคัญเพื่อศึกษา การวัดขนาดอิทธิพลและผลของประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในวิธีถดถอยโลจิสติก สำหรับข้อสอบที่มีรูปแบบการตรวจให้คะแนนแบบทวิวิภาค กรณีข้อมูลจำลองและข้อมูลเชิงประจักษ์ พิจารณาประสิทธิภาพ จากอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้วิธีถดถอยโลจิสติก ระหว่าง ขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl (2001) และขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas (1997) ว่าใช้เกณฑ์การแบ่งขนาดอิทธิพลใดมี ประสิทธิภาพในการตรวจสอบมากที่สุด ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัยและ ปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ซึ่งประกอบด้วย

รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน ขนาดของการทำหน้าที่ต่างกัน จำนวนข้อสอบที่ทำหน้าที่ต่างกัน และความยาวของแบบสอบทั้งฉบับ สำหรับวัตถุประสงค์ของการวิจัยแยกเป็นวัตถุประสงค์เฉพาะ ดังนี้

1. เพื่อเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีรูปแบบการให้คะแนนแบบทวิวิภาค โดยการจำลอง ข้อมูลในวิธีถดถอยโลจิสติก ระหว่างการวัด ขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl และการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas **ภายใต้เงื่อนไขเดียวกัน** ของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน ขนาดของการทำหน้าที่ต่างกัน จำนวนข้อสอบที่ทำหน้าที่ต่างกัน และความยาวของแบบสอบทั้งฉบับ

2. เพื่อเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีรูปแบบการให้คะแนนแบบทวิวิภาค โดยการจำลอง ข้อมูลในวิธีถดถอยโลจิสติก ด้วยการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl และการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas **ภายใต้เงื่อนไขต่างกัน** ของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน ขนาดของการทำหน้าที่ต่างกัน จำนวนข้อสอบที่ทำหน้าที่ต่างกัน และความยาวของแบบสอบทั้งฉบับ

3. เพื่อเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีรูปแบบการให้คะแนนแบบทวิวิภาค โดยข้อมูลเชิงประจักษ์ในวิธีถดถอยโลจิสติก ระหว่างการวัด ขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl และการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas

สมมติฐานการวิจัย

Swaminathan and Rogers (1990) ศึกษาเปรียบเทียบการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิธีถดถอยโลจิสติก กับวิธีแมนเทล-แฮนส์เซล พบว่า ข้อสอบที่ทำหน้าที่ต่างกัน แบบเอกรูปและแบบอเนกรูป เมื่อใช้แบบสอบที่มีความยาวมากขึ้นทำให้ความถูกต้องในการตรวจสอบของวิธีถดถอยโลจิสติกเพิ่มมากขึ้นทั้ง 2 รูปแบบ กรณีกลุ่มตัวอย่างน้อยและแบบสอบสั้นวิธีนี้สามารถตรวจสอบได้ถูกต้องร้อยละ 50 กรณีแบบสอบยาวและกลุ่มตัวอย่างขนาดใหญ่ ตรวจสอบได้ถูกต้องร้อยละ 75 ต่อมา Rogers and Swaminathan (1993) ศึกษาซ้ำเกี่ยวกับกรณีของความยากและอำนาจจำแนกของข้อสอบ พบว่า วิธีถดถอยโลจิสติก กับวิธีแมนเทล-แฮนส์เซล มีประสิทธิภาพการตรวจสอบ เท่ากันยกเว้นการกระจายค่าสถิติของวิธีถดถอยโลจิสติกไม่เป็นไปตามคาดไว้ในกรณีข้อสอบยากมากและอำนาจจำแนกสูง และวิธีถดถอยโลจิสติกตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปได้ดีในกรณีที่ข้อสอบมีความยากปานกลางและอำนาจจำแนกสูง Narayanan and Swaminathan (1996) เปรียบเทียบประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีแมนเทล-แฮนส์เซล วิธีถดถอยโลจิสติก

และวิธี CRO-SIB พบว่าการตรวจด้วยวิธีถดถอยโลจิสติกและวิธี CRO-SIB เป็นข้อสอบที่มีค่าความยากต่ำ ค่าอำนาจจำแนกสูง Lei et al.(2006) เปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบ แบบปรับเหมาะใช้การจำลองข้อมูลการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีถดถอยโลจิสติก วิธีการทดสอบอัตราส่วนไลด์ลิฮู้ดแบบ IRT (IRT Likelihood Ratio Test) และวิธี CATSIB พบว่าวิธีถดถอยโลจิสติกและวิธีการทดสอบอัตราส่วนไลด์ลิฮู้ดแบบ IRT ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้งแบบมีทิศทางและแบบไม่มีทิศทางได้ดีเท่ากัน และสองวิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ดีกว่าวิธี CATSIB

Uttaro and Millsap (1994) เปรียบเทียบความยาวของแบบสอบ 2 ขนาด คือ 20 และ 40 ข้อ พบว่าเมื่อใช้แบบสอบยาวมากขึ้น ความผิดพลาดประเภทที่ 1 ของวิธีแมนเทล -แฮนส์เซล จะมีค่าลดลง จิติมา วรรณศรี (2539) เปรียบเทียบความยาวแบบสอบ 30, 60 และ 90 ข้อ พบว่าแบบสอบที่มีความยาวปานกลาง (60 ข้อ) ทั้งวิธีแมนเทล-แฮนส์เซลและวิธีชิบเทสท์สามารถตรวจสอบได้มีประสิทธิภาพ ดีที่สุด ดังนั้นข้อสอบที่มีรูปแบบการให้คะแนนแบบทวิวิภาคที่มีความยาวปานกลางขึ้นไปจะส่งผลต่อประสิทธิภาพการตรวจสอบมากที่สุด (จิติมา วรรณศรี, 2539; ญาณภัทร สีหะมงคล, 2540; ปิยะทิพย์ ตินวร, 2549; Kim and Cohen, 1998)

Zumbo and Thomas (1997) แบ่งขนาดอิทธิพลเป็น 3 ระดับ ได้แก่ ขนาดเล็กน้อย (เกิดเมื่อค่าความแตกต่างของ $R^2 < .13$) ขนาดปานกลาง (เกิดเมื่อค่าความแตกต่าง $.13 \leq R^2 \leq .26$) และขนาดใหญ่ (เกิดเมื่อค่าความแตกต่างของ $R^2 > .26$) Jodoin and Gierl (2001) กล่าวว่าตามเกณฑ์นี้ยังมีความไวไม่เพียงพอในการวัดความเข้ม ของการทำหน้าที่ต่างกันของข้อสอบ จึงกำหนดเกณฑ์ตัดสิน ขนาดอิทธิพลของ การทำหน้าที่ต่างกันของข้อสอบ ใน 3 ระดับขึ้นใหม่ คือ ขนาดเล็กน้อย (เกิดเมื่อค่าความแตกต่างของ $R^2 < .035$) ขนาดปานกลาง (เกิดเมื่อค่าความแตกต่าง $.035 \leq R^2 \leq .07$) และขนาดใหญ่ (เกิดเมื่อค่าความแตกต่างของ $R^2 > .07$) สอดคล้องกับที่ Juana Dolores and Joseluis (2009) ที่ศึกษาประสิทธิภาพของขนาดอิทธิพลในการพัฒนาวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก โดยการจำลองข้อมูล จากข้อค้นพบดังกล่าวผู้วิจัยจึงตั้งสมมติฐานดังนี้

1. วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas และการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีรูปแบบการให้คะแนนแบบทวิวิภาค ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน ขนาดของการทำหน้าที่ต่างกัน จำนวนข้อสอบที่ทำหน้าที่ต่างกัน และความยาวของแบบสอบทั้งฉบับ มีอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกัน

2. ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน ขนาดของการทำหน้าที่ต่างกัน จำนวนข้อสอบที่ทำหน้าที่ต่างกัน ความยาวของแบบสอบทั้งฉบับ มีผลทำให้อัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1

ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่มีรูปแบบการให้คะแนนแบบทวิวิภาค ในวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas และในวิธีถดถอยโลจิสติก โดย การวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl แตกต่างกันแตกต่างกัน

3. วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas และการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีรูปแบบการให้คะแนนแบบทวิวิภาค ภายใต้ข้อสอบของ ข้อมูลเชิงประจักษ์ มีผลทำให้ อัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกัน

ขอบเขตของการวิจัย

1. **ข้อมูลจำลอง** (Simulation data)

1.1 จำลองข้อมูลใช้โมเดลภายใต้ทฤษฎีการตอบสนองข้อสอบชนิด 2 พารามิเตอร์ จำลองการตอบข้อสอบที่มีโครงสร้างวัดความสามารถเอกมิติที่ให้คะแนนแบบทวิวิภาค มีจำนวน ผู้สอบ 2,000 คน (แบ่งสองกลุ่มเท่ากัน) ผลการตอบข้อสอบภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน 2 รูปแบบ ขนาดของการทำหน้าที่ต่างกัน 3 ขนาด จำนวนข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด และความยาวของแบบสอบทั้งฉบับ 2 ขนาด มีข้อมูลที่ศึกษาทั้งสิ้น 24 เงื่อนไข (2 รูปแบบ × 3 ขนาด × 2 ขนาด × 2 ขนาด) ทุกเงื่อนไขจำลองข้อมูลซ้ำ 25 ครั้ง

1.2 ประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่มีรูปแบบการตรวจให้คะแนนแบบทวิวิภาค ภายใต้วิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl และการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas

1.3 ตัวแปรที่ศึกษา

1.3.1 ตัวแปรอิสระ

1) การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้วิธีถดถอยโลจิสติก

- ขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl
- ขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas

2) รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน มี 2 เงื่อนไข

- แบบเอกรูป (uniform DIF)
- แบบอเนกรูป (nonuniform DIF)

3) ขนาดของการทำหน้าที่ต่างกัน มี 3 เงื่อนไข คือ ขนาด 0.1, 0.2 และ 0.4

4) จำนวนข้อสอบที่ทำหน้าที่ต่างกัน มี 2 เงื่อนไข

- จำนวนข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 10
- จำนวนข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 20

5) ความยาวของแบบสอบทั้งฉบับ มี 2 เงื่อนไข

- ความยาว 40 ข้อ
- ความยาว 50 ข้อ

1.3.2 ตัวแปรตาม

1) อัตราความถูกต้อง (correct identification)

2) อัตราความคลาดเคลื่อนประเภทที่ 1 (Type I error rate)

2. ข้อมูลเชิงประจักษ์ (Empirical data)

ข้อมูลเชิงประจักษ์จาก "โครงการสอบระดับชาติของสถาบันแห่งหนึ่ง ประจำปี พ.ศ.2552 มีรูปแบบการตรวจให้คะแนนแบบทวิวิภาค ระดับชั้นประถมศึกษาปีที่ 6 จากโรงเรียนทุกสังกัด วิชาวิทยาศาสตร์และวิชาคณิตศาสตร์ " ศึกษาวิธีการตรวจสอบการทำหน้าที่ต่างกันของ ข้อสอบภายใต้ วิธีถดถอยโลจิสติก ระหว่างการวัด ขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl และเกณฑ์ Zumbo and Thomas โดยกลุ่มที่ทำการศึกษาจำแนกตามสภาพภูมิศาสตร์ตามสถานที่ตั้งของสถานศึกษา

กลุ่มที่ทำการศึกษาจำแนกตามสภาพภูมิศาสตร์ของสถานที่ตั้งของสถานศึกษา มีที่มาจากความเชื่อด้านเศรษฐสถานะของครอบครัวนักเรียน ซึ่งแบ่ง สถานะทางเศรษฐสังคมของครัวเรือน โดยจัดแบ่งครัวเรือนเป็นกลุ่มตามฐานะทางเศรษฐกิจและสังคม พิจารณาจากแหล่งรายได้ส่วนใหญ่ของครัวเรือน สถานภาพการทำงาน ประเภทของกิจกรรมในเชิงเศรษฐกิจและอาชีพ ของคนในครอบครัว เป็นหลักการแบ่งประเภทของครัวเรือนตามสถานะทางเศรษฐสังคมนี้ ขึ้นอยู่กับแหล่งที่มาของรายได้ที่ใช้ในการดำรงชีวิตของครัวเรือน และสถานภาพการทำงานของผู้รับเงินรายได้สูงสุดของครัวเรือน ซึ่งปกติได้แก่ หัวหน้าครัวเรือน อย่างไรก็ดี หากรายได้ของสมาชิกหลายคนซึ่งทำงานอาชีพเดียวกันและสถานภาพการทำงานเหมือนกันรวมกันแล้วเป็นรายได้ส่วนใหญ่ซึ่งใช้ในการดำรงชีพของครัวเรือนแล้ว ครัวเรือนนั้นจะถูกจัดเข้ากลุ่มตามอาชีพและสถานภาพการทำงานของบุคคลเหล่านั้น (สำนักงานสถิติแห่งชาติ, 2555)

นิยามศัพท์ที่ใช้ในการวิจัย

ผลการตรวจสอบ การทำหน้าที่ต่างกัน (Differential Item Functioning) หมายถึง ผลที่ปรากฏจากการนำวิธีทางสถิติมาใช้วิเคราะห์ เพื่อบ่งบอกว่ามีการทำหน้าที่ต่างกันของข้อสอบระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ

ข้อสอบที่ทำหน้าต่างกัน หมายถึง ข้อสอบที่ผู้สอบจากกลุ่มที่ต่างกันตาม สภาพภูมิศาสตร์ด้านตำแหน่งที่ตั้งของสถานศึกษา คือ นักเรียนที่สังกัดสถานศึกษาในเขตอำเภอเมือง กับ นอกเขตอำเภอเมืองที่มีความสามารถในสิ่งที่ต้องการวัดเท่ากันแต่มีโอกาสตอบข้อสอบข้อนั้นได้ถูกต้องไม่เท่ากัน

วิธีถดถอยโลจิสติก หมายถึง วิธีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบที่ใช้การคำนวณดัชนีการทำหน้าที่ต่างกันจากผลการตอบข้อสอบถูกระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มี

ความสามารถระดับเดียวกัน แต่มีการตอบข้อสอบได้ถูกต้องต่างกัน ใช้โมเดลการถดถอยโลจิสติก และพิจารณาค่าปฏิสัมพันธ์ระหว่างกลุ่มผู้สอบกับระดับความสามารถของผู้สอบ โดยทดสอบระดับนัยสำคัญด้วยสถิติ χ^2 (Chi-square) ที่ระดับนัยสำคัญ .05 หากค่าสถิติ χ^2 มีนัยสำคัญ ถือว่าข้อสอบข้อนั้นทำหน้าที่ต่างกัน ส่วนการวัดขนาดอิทธิพลในวิธีถดถอยโลจิสติกใช้ค่าสถิติ Nagelkerke: R^2 ซึ่งคล้ายกับค่าสัมประสิทธิ์การอธิบาย (Coefficient of determination: R^2) ใน Multiple Regression (Norusis, 1997)

ประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกัน (Efficacy of detected DIF) หมายถึง ความถูกต้องของการระบุการทำหน้าที่ต่างกันของข้อสอบจากการตรวจสอบด้วยวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl และตามเกณฑ์ Zumbo and Thomas พิจารณาประสิทธิภาพจากอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1

อัตราความถูกต้อง (correct identification) หมายถึง จำนวนข้อที่ตรวจสอบได้ถูกต้องว่าทำหน้าที่ต่างกัน ต่อจำนวนข้อที่ทำหน้าที่ต่างกันทั้งหมดในแบบสอบ คำนวณเป็นค่าร้อยละ

อัตราความคลาดเคลื่อนประเภทที่ 1 (Type I error rate) หมายถึง จำนวนข้อที่ระบุผิดพลาดว่าทำหน้าที่ต่างกันทั้งที่ความจริงทำหน้าที่ไม่ต่างกัน ต่อจำนวนข้อที่ทำหน้าที่ไม่ต่างกันในแบบสอบ คำนวณเป็นค่าร้อยละ

ขนาดอิทธิพล (Effect Size) หมายถึง ขนาดหรือค่าความเข้ม ของการทำหน้าที่ต่างกัน ของข้อสอบซึ่งเป็นระดับความสัมพันธ์ของตัวแปรต้นและตัวแปรตาม การตัดสินขนาดอิทธิพลโดยค่าสถิติ R^2 ซึ่งจะตัดสินข้อสอบทำหน้าที่ต่างกันเมื่อขนาดอิทธิพลมีขนาดปานกลางและขนาดใหญ่ ใช้ 2 เกณฑ์ คือ

1) เกณฑ์ของ Jodoin and Gierl ตัดสินขนาดของ DIF 3 ระดับ ได้แก่

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

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

DIF ขนาดใหญ่ มีค่าความแตกต่าง $\Delta R^2 > .07$

2) เกณฑ์ของ Zumbo and Thomas มีขนาด 3 ระดับได้แก่

DIF ขนาดเล็กน้อย มีค่าความแตกต่าง $\Delta R^2 < .13$

DIF ขนาดปานกลาง มีค่าความแตกต่าง $.13 \leq \Delta R^2 \leq .26$

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

วิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) หมายถึง วิธีการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันตามแนวคิดของ Mantel-Haenszel โดยการวิเคราะห์ความแตกต่างของสัดส่วนการตอบข้อสอบระหว่างผู้สอบกลุ่มย่อย ได้แก่ กลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ที่มีความสามารถในระดับเดียวกัน

ข้อมูลจำลอง หมายถึง ผลการตอบข้อสอบที่สร้างขึ้นตามเงื่อนไขปัจจัยหลัก เป็นตัวแปรที่ต้องการศึกษา โดยใช้ The simulation algorithm สำหรับเมตริกซ์การตอบสนองรายข้อเริ่มต้นที่การ

กำหนดรูปแบบการแจกแจงความสามารถ และกำหนดค่าพารามิเตอร์ของข้อสอบ แล้วดำเนินการจำลองข้อมูลโดยใช้โปรแกรม R ตรวจสอบความถูกต้องของข้อมูลจำลองโดยใช้โปรแกรม MULTILOG-MG

ภายใต้เงื่อนไขเดียวกัน หมายถึง การเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบภายใต้วิธีการที่ศึกษา โดยพิจารณาเฉพาะเงื่อนไขแต่ละระดับของปัจจัยที่แปรเปลี่ยน เช่น พิจารณาเฉพาะเงื่อนไขรูปแบบของการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป เพื่อเปรียบเทียบประสิทธิภาพภายใต้วิธีการที่ศึกษา ระหว่างการวัดขนาดอิทธิพลตามเกณฑ์ของ Jodoin and Gierl และเกณฑ์ Zumbo and Thomas คำตอบที่จะได้รับคือ รูปแบบของการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป ใช้เกณฑ์การวัดขนาดอิทธิพลตามเกณฑ์ของใครให้ผลดีกว่ากัน เป็นต้น

ภายใต้เงื่อนไขต่างกัน หมายถึง การเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบภายใต้ปัจจัยที่ศึกษา โดยพิจารณาเฉพาะวิธีการตรวจสอบแต่ละวิธีการ เช่น พิจารณาเฉพาะวิธีการตรวจสอบการวัดขนาดอิทธิพลตามเกณฑ์ของ Jodoin and Gierl เพื่อเปรียบเทียบประสิทธิภาพของปัจจัยรูปแบบการทำหน้าที่ต่างกันของข้อสอบ คำตอบที่จะได้รับคือ ภายใต้รูปแบบของการทำหน้าที่ต่างกันของข้อสอบทั้ง 2 รูปแบบ เมื่อวัดขนาดอิทธิพลเฉพาะเกณฑ์ของ Jodoin and Gierl รูปแบบไหนให้ประสิทธิภาพที่ดีกว่ากัน เป็นต้น

ข้อมูลเชิงประจักษ์ หมายถึง ข้อมูลจากการสอบแข่งขันตามโครงการสอบระดับชาติของสถาบันแห่งหนึ่ง ประจำปี พ.ศ.2552 มีรูปแบบการตรวจให้คะแนนแบบทวิวิภาค ระดับชั้นประถมศึกษาปีที่ 6 จากโรงเรียนทุกสังกัด วิชาวิทยาศาสตร์และวิชาคณิตศาสตร์

กลุ่มอ้างอิง (Reference group: R) หมายถึง ผู้สอบกลุ่มที่คาดว่าจะได้เปรียบ จากการตอบข้อสอบ โดยมีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง สูงกว่าผู้สอบ อีกกลุ่มทั้งที่มีความสามารถเท่ากัน การศึกษาข้อมูลเชิงประจักษ์ครั้งนี้จำแนกกลุ่มที่ศึกษาตามเศรษฐสถานะของครอบครัวภายใต้ความแตกต่างของรายได้ผู้ปกครองและสิ่งแวดล้อมทางวิชาการเป็นหลัก กลุ่มอ้างอิงคือ นักเรียนที่สังกัดสถานศึกษาในเขตพื้นที่การศึกษาเขตอำเภอเมืองที่มีที่ตั้งสถานศึกษาในเขตกรุงเทพมหานครและในเขตพื้นที่การศึกษาเขต 1 ของทุกจังหวัด

กลุ่มเปรียบเทียบ (Focal group: F) หมายถึง ผู้สอบกลุ่มที่คาดว่าจะเสียเปรียบ จากการตอบข้อสอบ โดยมีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง ต่ำกว่าผู้สอบ อีกกลุ่มทั้งที่มีความสามารถเท่ากัน การศึกษาข้อมูลเชิงประจักษ์ครั้งนี้จำแนกกลุ่มที่ศึกษาตามเศรษฐสถานะของครอบครัวภายใต้ความแตกต่างของรายได้ผู้ปกครองและสิ่งแวดล้อมทางวิชาการเป็นหลัก กลุ่มเปรียบเทียบคือ นักเรียนที่สังกัดสถานศึกษาในเขตพื้นที่การศึกษานอกเขตอำเภอเมืองที่ไม่ได้มีที่ตั้งของสถานศึกษาในเขตกรุงเทพมหานครและสถานศึกษามีที่ตั้งในเขตพื้นที่การศึกษาเขตอื่นๆ ที่ไม่ใช่เขต 1 ของทุกจังหวัด

ขนาดกลุ่มตัวอย่าง (sample size) หมายถึง ขนาดกลุ่มตัวอย่าง เป็นผู้สอบกลุ่มอ้างอิง ต่อผู้สอบกลุ่มเปรียบเทียบ กำหนดให้กลุ่มตัวอย่างที่ศึกษามีขนาดเดียวคืออัตราส่วน 1,000: 1,000

ปัจจัยที่แปรเปลี่ยน หมายถึง ปัจจัยที่ศึกษาซึ่งการจำลองข้อมูลครั้งนี้จำลองผลการตอบข้อสอบใน 4 ปัจจัย คือ รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน ขนาดของการทำหน้าที่ต่างกัน จำนวนข้อสอบที่ทำหน้าที่ต่างกัน และความยาวของแบบสอบทั้งฉบับ

รูปแบบของข้อสอบ ที่ทำหน้าที่ต่างกันแบบเอกรูป (uniform DIF) หมายถึง ข้อสอบที่ทำหน้าที่ต่างกันแบบสม่ำเสมอ เกิดขึ้นเมื่อไม่มีปฏิสัมพันธ์ (interaction) ระหว่างระดับความสามารถของผู้สอบกับการเป็นสมาชิกของกลุ่มย่อยหรือโอกาสของการตอบข้อสอบได้ถูกต้องของผู้สอบกกลุ่มย่อยกลุ่มหนึ่งสูงกว่าผู้สอบกลุ่มย่อยอีกกลุ่มหนึ่งตลอดช่วงความสามารถ

รูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป (Nonuniform DIF) หมายถึง ข้อสอบที่ทำหน้าที่ต่างกันแบบไม่สม่ำเสมอ เกิดขึ้นเมื่อมีปฏิสัมพันธ์ ระหว่างระดับความสามารถของผู้สอบกับการเป็นสมาชิกของกลุ่มย่อยหรือโอกาสของการตอบข้อสอบได้ถูกต้องของผู้สอบกลุ่มย่อยกลุ่มหนึ่งสูงกว่าผู้สอบกลุ่มย่อยอีกกลุ่มหนึ่งตลอดช่วงความสามารถ

ขนาดของการทำหน้าที่ต่างกัน (amount of DIF) หมายถึง ขนาดของการทำหน้าที่ต่าง กันของข้อสอบอย่างมีนัยสำคัญทางสถิติ หรือพื้นที่รวมแตกต่างระหว่างโค้งคุณลักษณะข้อสอบ 2 กลุ่ม 3 ระดับ คือ ขนาด 0.1, 0.2 และ 0.4

จำนวนข้อสอบที่ทำหน้าที่ต่างกัน (the number of items with DIF) หมายถึง จำนวนการทำหน้าที่ต่างกันของข้อสอบในแต่ละฉบับ มี 2 เงื่อนไข คือ ทั้งฉบับคิดเป็นร้อยละ 10 และคิดเป็นร้อยละ 20 ถ้าข้อสอบจำนวน 40 ข้อจะมีข้อสอบที่ทำหน้าที่ต่างกัน 4 และ 8 ข้อ มีสัดส่วนข้อสอบที่ทำหน้าที่ไม่ต่างกันต่อข้อสอบที่ทำหน้าที่ต่างกันเป็น 36 : 4 และ 32 : 8 ถ้าข้อสอบจำนวน 50 ข้อ จะมีข้อสอบที่ทำหน้าที่ต่างกัน 5 และ 10 ข้อ มีสัดส่วนข้อสอบที่ทำหน้าที่ไม่ต่างกันต่อข้อสอบที่ทำหน้าที่ต่างกันเป็น 45 : 5 และ 40 : 10

ความยาวของแบบสอบทั้งฉบับ (Test length) หมายถึง จำนวนข้อสอบในแบบสอบแต่ละฉบับ มี 2 เงื่อนไข คือ จำนวน 40 ข้อ และจำนวน 50 ข้อ

ประโยชน์ที่คาดว่าจะได้รับ

การวิจัยครั้งนี้เป็นการศึกษาประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้วิธีถดถอยโลจิสติก จากข้อสอบที่มีรูปแบบการตรวจให้คะแนนแบบทวิวิภาค โดยการศึกษาข้อมูลจำลองตามเงื่อนไขของปัจจัยที่แปรเปลี่ยนต่างๆ ร่วมกับ การศึกษา ข้อมูลเชิงประจักษ์ ซึ่งเป็นการขยายขอบเขตการศึกษาเกี่ยวกับวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ภายใต้วิธีถดถอยโลจิสติก ผลการศึกษาครั้งนี้มีประโยชน์ดังนี้

1. ประโยชน์ทางวิชาการและทางปฏิบัติ การเลือกใช้วิธีการตรวจสอบให้เหมาะสมกับลักษณะธรรมชาติของข้อสอบทางการศึกษาและจิตวิทยาและผู้สอบ ได้สารสนเทศเกี่ยวกับการตรวจสอบการทำ

หน้าที่ต่างกันของข้อสอบ ทำให้ทราบถึงข้อดีและข้อจำกัดของวิธีการตรวจสอบการทำหน้าที่ ต่างกันของข้อสอบภายใต้ปัจจัยที่แปรเปลี่ยน ได้แก่ รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน ขนาดของการทำหน้าที่ต่างกัน จำนวนข้อสอบที่ทำหน้าที่ต่างกัน ความยาวของแบบสอบทั้งฉบับ ทำให้ทราบถึงปัจจัยที่มีผลต่อวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สามารถตัดสินข้อดีและข้อจำกัดเพื่อนำไปสู่การเลือกใช้วิธีการตรวจสอบที่ได้สารสนเทศสูงสุด

2. ประโยชน์ในการนำผลไปใช้ กรณีศึกษาในข้อมูลเชิงประจักษ์ สารสนเทศที่ได้รับ จะเป็นแนวทางให้นักการศึกษาตัดสินใจเลือกใช้วิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้สถานการณ์การสอบในบริบทจริงได้อย่างเหมาะสม ผู้สร้างแบบสอบสามารถนำสารสนเทศเกี่ยวกับวิธีการ ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้การให้คะแนนแบบทวิวิภาคไปใช้ประโยชน์ในการเลือกข้อสอบที่มีประสิทธิภาพเหมาะสมกับข้อมูลที่ใช้วัดความสามารถ ได้สารสนเทศที่เป็น ประโยชน์ต่อ การวัดและประเมินผลในแง่คุณภาพของเครื่องมือตามเกณฑ์มาตรฐานถือเป็นแนวทางสำหรับการพัฒนาคุณภาพเครื่องมือที่ใช้วัดผล ทำให้การแปลความหมายของคะแนนจากการใช้แบบสอบถูกต้องและชัดเจน นำไปสู่การปรับปรุงข้อสอบ เพื่อให้เกิดความยุติธรรมสำหรับ ผู้เข้าสอบทุกคน ทำให้การพัฒนาแบบทดสอบมีคุณภาพตามเกณฑ์มาตรฐาน การแปลความหมายของคะแนนมีความถูกต้องและชัดเจน

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

การศึกษาครั้งนี้ ผู้วิจัยนำเสนอแนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้องกับมโนทัศน์ของทฤษฎีการสอบ การทำหน้าที่ต่างกันของข้อสอบและการจำลองข้อมูล โดยศึกษาค้นคว้าและรวบรวมข้อมูลจากการสังเคราะห์หนังสือ เอกสาร บทความวิชาการจากในประเทศและต่างประเทศ ผู้วิจัยนำเสนอเอกสารและงานวิจัยที่เกี่ยวข้องเป็น 5 ตอน ดังนี้

ตอนที่ 1 มโนทัศน์ของทฤษฎีทางการสอบแบบดั้งเดิมและการประยุกต์ใช้

ตอนที่ 2 มโนทัศน์ของทฤษฎีการตอบสนองข้อสอบและการประยุกต์ใช้

ตอนที่ 3 มโนทัศน์ของการทำหน้าที่ต่างกันของข้อสอบ

3.1 ความลำเอียงและการทำหน้าที่ต่างกันของข้อสอบ

3.2 ประเภทของการทำหน้าที่ต่างกันของข้อสอบ

3.3 หลักการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

3.4 ทฤษฎีการตอบสนองข้อสอบและการทำหน้าที่ต่างกันของข้อสอบ

3.5 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

3.6 ประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

ตอนที่ 4 มโนทัศน์ของขนาดอิทธิพล

4.1 การรายงานขนาดอิทธิพลของงานวิจัยในปัจจุบัน

4.2 ความหมายของขนาดอิทธิพล

4.3 การตัดสินใจทางการวิจัย

4.4 ขนาดอิทธิพลกับการศึกษาการทำหน้าที่ต่างกันของข้อสอบ

ตอนที่ 5 เอกสารและงานวิจัยที่เกี่ยวข้อง

5.1 งานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบในประเทศ

5.2 งานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบต่างประเทศ

5.3 สรุปประเด็นปัญหาที่พบเกี่ยวกับการศึกษาการทำหน้าที่ต่างกันของข้อสอบ

5.4 กรอบแนวคิดการวิจัย

ตอนที่ 1 มโนทัศน์ของทฤษฎีทางการสอบแบบดั้งเดิม

ทฤษฎีทางการสอบแบบดั้งเดิม (Classical Test Theory: CTT) มีแนวคิดสำคัญว่าคะแนน จากการวัดเป็นคะแนนที่สังเกตได้ (observed score: X) เกิดจากผลรวมเชิงเส้นของคะแนนจริง (True score) และคะแนนความคลาดเคลื่อน (error score) ของการวัด เป็นค่าที่ไม่สามารถสังเกตได้ทั้งสองค่า สามารถเขียนโมเดลการทดสอบแบบดั้งเดิม ดังนี้ $X = T + E$ ความหมาย คือ คะแนนที่สังเกต และวัดได้จากแบบสอบเท่ากับผลรวมของคะแนนจริงกับคะแนนความคลาดเคลื่อน (ศิริชัย กาญจนวาสี , 2550) หากคะแนนความคลาดเคลื่อนมีค่าน้อยหรือใกล้ศูนย์แล้วคะแนนที่สังเกตได้จะมีค่าใกล้เคียงกับคะแนนจริง ทั้งนี้ขึ้นอยู่กับสถานการณ์ในการสอบและคุณภาพของแบบสอบ นักวัดผลจึงพยายามสร้างแบบสอบที่มีคุณภาพสูงและจัดสถานการณ์การสอบให้ลดโอกาสของการเกิดความคลาดเคลื่อนให้มากที่สุดเพื่อใช้คะแนนดิบหรือคะแนนที่สังเกตได้โดยไม่ต้องประมาณค่าคะแนนจริง

1.1 ข้อตกลงเบื้องต้น

ทฤษฎีทางการสอบแบบดั้งเดิม ตั้งอยู่บนพื้นฐานความเชื่อตามข้อตกลงเบื้องต้นของโมเดล (model assumption) และข้อตกลงเบื้องต้นของแบบสอบคู่ขนาน (Assumption of Parallelism) (ศิริชัย กาญจนวาสี, 2550) ดังนี้

1.1.1 ข้อตกลงเบื้องต้นของโมเดล

1) ข้อตกลงเบื้องต้นข้อที่ 1 คะแนนที่ได้จากการวัดมีความสัมพันธ์เชิงเส้นตรง (additive relationship) เป็นเชิงบวกกับคะแนนจริงและคะแนนความคลาดเคลื่อน ซึ่งความสัมพันธ์เชิงเส้นตรงนิยมใช้กันทางสถิติ เช่น โมเดลของการถดถอยพหุ (Multiple Regression) การวิเคราะห์ความแปรปรวน (Analysis of variance) การวิเคราะห์องค์ประกอบ (Factor analysis) เป็นต้น

2) ข้อตกลงเบื้องต้นข้อที่ 2 คะแนนจริงมีสภาวะคงที่ เท่ากับค่าเฉลี่ยของคะแนนที่ได้จากการวัดซ้ำๆ หลายๆ ครั้ง หรือ $T_{บุคคล} = E(X_{บุคคล})$

3) ข้อตกลงเบื้องต้นข้อที่ 3 คะแนนความคลาดเคลื่อนไม่มีความสัมพันธ์กับคะแนนจริง ผู้สอบที่มีคะแนนจริงสูงหรือต่ำย่อมไม่มีความสัมพันธ์กับคะแนนความคลาดเคลื่อนที่เกิดขึ้น และคะแนนความคลาดเคลื่อนจากแบบสอบฉบับหนึ่งจะไม่มีความสัมพันธ์กับคะแนนจริงจากแบบสอบฉบับอื่นๆ

4) ข้อตกลงเบื้องต้นข้อที่ 4 คะแนนความคลาดเคลื่อนของบุคคลไม่สัมพันธ์กัน

1.1.2 ข้อตกลงเบื้องต้นของแบบสอบคู่ขนาน

1) ข้อตกลงเบื้องต้นข้อที่ 5 แบบสอบสองฉบับจะถือว่าเป็นแบบสอบคู่ขนาน (Parallel Test) เมื่อคะแนนจริงของผู้สอบแต่ละคนมีค่าเท่ากันทั้งสองฉบับและความคลาดเคลื่อนมาตรฐานของประชากรที่ทำแบบสอบทั้งสองฉบับมีค่าเท่ากัน

2) ข้อตกลงเบื้องต้นข้อที่ 6 แบบสอบสองฉบับจะถือว่าเป็นแบบสอบทัดเทียมกันก็ต่อเมื่อความคลาดเคลื่อนมาตรฐานของประชากรที่ทำแบบสอบทั้งสองฉบับมีค่าเท่ากัน

1.2 ความคลาดเคลื่อนมาตรฐานของการวัด (Standard Errors of Measurement)

ทฤษฎีทางการสอบแบบดั้งเดิมมีข้อตกลงเบื้องต้นว่า คะแนนที่สังเกตได้ (X) ของผู้สอบใดๆ ที่ทำการสอบซ้ำๆ อย่างเป็นอิสระต่อกันด้วยแบบสอบฉบับเดิมหรือแบบสอบคู่ขนาน การแจกแจงของคะแนนจะกระจุกตัวอยู่รอบๆ คะแนนจริง (T) โดยมีส่วนเบี่ยงเบนมาตรฐานเป็น σ_E หรือ ความคลาดเคลื่อนมาตรฐานของการวัด (Standard Errors of Measurement: SEM) (ศิริชัย กาญจนวาสี, 2550)

ความคลาดเคลื่อนมาตรฐานของการวัด คำนวณได้จากสมการต่อไปนี้

$$SEM = S_X \sqrt{1 - R_{XX'}}$$

เมื่อ SEM หมายถึง ค่าความคลาดเคลื่อนมาตรฐานของการวัด

S_X หมายถึง ค่าส่วนเบี่ยงเบนมาตรฐานของคะแนนที่สังเกตได้ (X)

$R_{XX'}$ หมายถึง สัมประสิทธิ์ความเที่ยงของแบบสอบ

1.3 ทฤษฎีทางการสอบแบบดั้งเดิมและการประยุกต์ใช้

การวิเคราะห์ข้อสอบโดยอาศัยทฤษฎีทางการสอบแบบดั้งเดิม ทำให้นักการศึกษาสามารถนำผลที่วิเคราะห์ไปใช้ประโยชน์ได้หลายประการ ดังต่อไปนี้

1.3.1 การใช้ประโยชน์ในการปรับแก้คะแนนสอบ ผลการวิเคราะห์ข้อสอบจะชี้ให้ผู้ออกข้อสอบเห็นถึงผลบางประการเช่นการเฉลย ข้อสอบข้อใดน่าจะมีปัญหามีความคลุมเครือในคำถามหรือตัวเลือกมีความซ้ำซ้อนกันหรือเนื้อหาของข้อสอบอยู่นอกเหนือไปจากสิ่งที่สอนนักเรียน เป็นต้น ข้อสอบที่มีปัญหาเหล่านี้ต้องได้รับการประเมินโดยคณะกรรมการตรวจข้อสอบซึ่งประกอบด้วยอาจารย์ผู้มีความรู้ความชำนาญในเนื้อหาวิชาที่ทำการสอบว่าจะดำเนินการอย่างไรกับการคิดคะแนน หากปัญหาที่พบมีความรุนแรงไม่มากจนทำให้การตัดสินใจเลือกคำตอบที่ถูกต้องเปลี่ยนไป

1.3.2 การใช้ประโยชน์ในการปรับปรุงคุณภาพข้อสอบ ภายหลังจากการรายงานคะแนนสอบเป็นที่เรียบร้อยแล้วครูผู้สอนสามารถนำผลการวิเคราะห์ข้อสอบแต่ละข้อมาพิจารณาโดยละเอียดว่าข้อสอบข้อใดสมควรได้รับการปรับปรุงแก้ไข ข้อสอบที่ยากเกินไปอาจเกิดจากโจทย์คำถามมีความคลุมเครือต้องทำการปรับแก้ให้โจทย์ชัดเจนขึ้น หรือต้องเพิ่มเติมข้อมูลบางประการเข้าไปเพื่อให้การวินิจฉัยชัดเจนขึ้น ข้อสอบที่ง่ายเกินไปอาจพิจารณาปรับให้ยากขึ้นโดยการปรับแก้ไขโจทย์หรือปรับแก้ไขตัวเลือก ข้อสอบที่มีค่า point-biserial ต่ำมักเกิดจากโจทย์ที่คลุมเครือสร้างความสับสนให้ผู้สอบ สมควรปรับโจทย์คำถามใหม่ นอกจากนี้ครูผู้สอนยังต้องพิจารณาถึงการทำงานของตัวเลือกด้วย ปัญหาที่พบบ่อยในการวิเคราะห์ข้อสอบปรนัย คือมีตัวลวงจำนวนมากที่ไม่ ได้ทำหน้าที่ (มีผู้สอบเลือกน้อยหรือลวง

เฉพาะผู้ที่มีความรู้ดีให้มาเลือก) จากการศึกษาวิจัยข้อสอบปรนัยจำนวนมากพบว่าข้อสอบส่วนใหญ่มักมีตัวเลือกที่ทำงานจริงเพียง 3 ตัวเลือกเท่านั้น ตัวเลือกที่เหลือเป็นตัวเลือกที่ไม่มีประโยชน์ พิมพ์ลงมาในข้อสอบเป็นการเปลืองเนื้อที่หน้ากระดาษและเสียเวลาอ่านโดยใช่เหตุควรพิจารณาตัดตัวลวงที่ไม่ทำงานหรือเปลี่ยนตัวลวงที่น่าจะมีประสิทธิภาพมากขึ้น

1.3.3 การใช้ประโยชน์สำหรับการบริหารคลังข้อสอบ ข้อสอบแต่ละข้อได้มาด้วยความลำบาก ครูผู้สอนต้องใช้เวลาและความคิดอย่างมากเพื่อพัฒนาข้อสอบที่ดีขึ้นมาใช้ ดังนั้นเมื่อนำข้อสอบมาใช้ผลการวิเคราะห์ข้อสอบจะแสดงความเป็นข้อสอบที่ดีมีระดับความยากง่ายเหมาะสม ความสามารถในการจำแนกผู้สอบที่ดีก็ควรจะพิจารณาเลือกเก็บข้อสอบดังกล่าวไว้ในคลังข้อสอบเพื่อที่จะได้นำกลับมาใช้ใหม่ในอนาคต ในการเก็บข้อสอบเข้าในคลังข้อสอบก็ต้องมีการแนบข้อมูลเกี่ยวกับประวัติการใช้งานและผลการวิเคราะห์ข้อสอบในแต่ละครั้งไว้คู่กันด้วย เพื่อที่จะได้เป็นประโยชน์ในการเลือกข้อสอบมาใช้งาน หากอาจารย์ต้องการข้อสอบที่มีระดับความยากง่าย หรือความสามารถในการจำแนกผู้สอบมากน้อยเพียงใดจะได้ดึงเอาข้อสอบที่มีคุณลักษณะตามต้องการออกมาใช้ได้ตามต้องการ

1.3.4 การใช้ประโยชน์สำหรับการพัฒนาคุณภาพการสอน การพิจารณาผลการวิเคราะห์ข้อสอบโดยละเอียดในหัวข้อที่ครูผู้สอนท่านใดท่านหนึ่งรับผิดชอบจะทำให้เกิดการพิจารณาตนเองตามมา ถือว่าเป็นการพัฒนาตนเองของผู้สอนควบคู่ไปกับการพิจารณาคุณภาพของเครื่องมือที่ใช้สำหรับวัดนักเรียน การใช้ทฤษฎีการวัดแบบดั้งเดิมให้ข้อมูลที่เป็นประโยชน์หลายอย่างแต่เนื่องจากวิธีการวิเคราะห์เหล่านี้เป็นเทคนิคที่วางรากฐานอยู่บนทฤษฎีทางการสอบแบบดั้งเดิม (classical test theory) มีข้อจำกัดหลายประการด้วยกันในการนำค่าต่างๆ ที่ได้จากการวิเคราะห์ข้อสอบไปใช้ครูผู้สอนควรคำนึงถึงข้อจำกัดของผลการวิเคราะห์ด้วย

ตอนที่ 2 มโนทัศน์ของทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบเป็นทฤษฎีที่อธิบายความสัมพันธ์ระหว่างลักษณะ (Trait) หรือความสามารถ (Ability) ที่ไม่สามารถสังเกตได้ภายในตัวบุคคลกับพฤติกรรม การตอบข้อคำถามของบุคคลเป็นสิ่งที่สังเกตได้โดยตรง แสดงในรูปแบบฟังก์ชันทางคณิตศาสตร์ที่ แสดงความสัมพันธ์ระหว่างพฤติกรรมการตอบข้อสอบกับระดับความสามรถ เรียกฟังก์ชันนี้ว่า ฟังก์ชันการตอบสนองข้อสอบ (Item Response Function) (Lord and Novick, 1968)

การกำหนดฟังก์ชันการตอบสนองข้อสอบสามารถกำหนดได้หลายรูปแบบ ขึ้นอยู่กับข้อตกลงเบื้องต้นของข้อมูลที่ได้จากการสอบ (Hambleton and Cook, 1977) ทฤษฎีการตอบสนองข้อสอบมีหัวใจหลักอยู่ที่โค้งลักษณะข้อสอบ (Item Characteristic Curve: ICC) (อุทุมพร จามรมาน, 2540) เนื่องจากแสดงความสัมพันธ์ระหว่างระดับความสามารถหรือคุณลักษณะที่วัดด้วยข้อความนั้น กับโอกาสที่เขาจะตอบข้อความนั้นถูกต้อง ทฤษฎีการตอบสนองข้อสอบเป็นทฤษฎีการวัดแนวใหม่ที่อธิบาย

ความสัมพันธ์ระหว่างคุณลักษณะภายใน กับ พฤติกรรมการตอบข้อสอบของบุคคล ว่ามีโอกาสตอบข้อสอบถูกเพียงใด ทฤษฎีนี้มีพื้นฐานความเชื่อที่ว่าพฤติกรรมการตอบข้อสอบของผู้สอบถูกกำหนดโดยคุณลักษณะภายในหรือความสามารถที่มีอยู่ภายในตัวบุคคล (ศิริชัย กาญจนวาสี, 2550)

2.1 หลักการของทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) มีแนวคิดพื้นฐาน คือพฤติกรรมการตอบข้อสอบที่บุคคลแสดงออกเป็นสิ่งที่สังเกตได้โดยตรง สามารถใช้ในการพยากรณ์คุณลักษณะภายในที่เรียกว่าคุณลักษณะ (trait) หรือ คุณลักษณะแฝง (latent traits) หรือ ความสามารถ (abilities) ของบุคคล ความสัมพันธ์ระหว่างการแสดงพฤติกรรมการตอบข้อสอบของผู้สอบและคุณลักษณะภายในอธิบายได้โดยฟังก์ชันคุณลักษณะข้อสอบหรือโค้งลักษณะข้อสอบ (ICC) ซึ่งเป็นฟังก์ชันที่มีลักษณะเพิ่มขึ้นทางเดียว (monotonically increasing function) เมื่อผู้ตอบข้อสอบมีความสามารถสูง ความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องก็จะมากขึ้น (Hambleton, Swaminatan and Roger, 1991)

ทฤษฎีการตอบสนองข้อสอบ อธิบายความสัมพันธ์ระหว่างคุณลักษณะภายในที่อยู่ในตัวบุคคลกับ ผลการตอบข้อสอบของบุคคลนั้น โดยโค้งลักษณะข้อสอบ (ICC) กำหนดลักษณะข้อสอบด้วยค่าพารามิเตอร์ข้อสอบ คือ ความยาก (b) อำนาจจำแนก (a) และการเดา (c) (ศิริชัย กาญจนวาสี, 2550) และพารามิเตอร์ของผู้สอบ (person parameter) อันเป็นคุณลักษณะที่มีอยู่ภายในที่เป็นความสามารถแท้จริงของผู้สอบ ซึ่งไม่ควรแปรเปลี่ยนไปตามชุดข้อสอบ แต่เนื่องจากความสามารถของผู้สอบเป็นคุณลักษณะแฝงที่ไม่สามารถสังเกตหรือวัดได้โดยตรง จำเป็นต้องใช้การทำนาย (predict) คุณลักษณะดังกล่าว โดยอาศัยผลที่ได้จากการสังเกต ที่วัดได้ (Lord and Novick,1968; Hambleton and Cook, 1977; Hambleton and Swaminatan, 1985)

นักวัดผลพยายามหาความสัมพันธ์ระหว่างผลการตอบแบบสอบ ในรูป คะแนน (Test Performance or score) กับระดับความสามารถ (ability) ของผู้ตอบราย บุคคล นำมาเขียนเป็นโมเดลทางคณิตศาสตร์ (Hambleton and Cook, 1977; Hambleton and Swaminatan, 1985) แต่เนื่องจากความสัมพันธ์ดังกล่าวเป็นเพียงฟังก์ชันความสัมพันธ์ในลักษณะทั่วไป นักวัดผลการศึกษาจึงต้องหาโมเดลทางคณิตศาสตร์ที่เหมาะสมเพื่อใช้แทนฟังก์ชันความสัมพันธ์ดังกล่าว โดยอาศัยข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ โมเดลทางคณิตศาสตร์ ที่แสดงความสัมพันธ์ระหว่างผลที่ได้จากการตอบข้อสอบหรือแบบสอบกับระดับความสามารถของผู้สอบสามารถเขียนในรูปของความสัมพันธ์ ดังนี้

$$P = f(U_i / \theta_1, \theta_2, \theta_3, \ldots \theta_k ; \beta_k)$$

เมื่อ P แทน ผลจากการตอบแบบสอบ ((test performance)

f แทน ฟังก์ชัน (function)

U_i แทน ผลการตอบข้อสอบข้อที่ i

(พฤติกรรมที่สนใจศึกษา $U_i = 1$ พฤติกรรมที่ไม่สนใจศึกษา $U_i = 0$)

$\theta_1, \theta_2, \theta_3, \ldots, \theta_k$ แทน ระดับความสามารถ (ability) ที่ 1, 2, 3,…,k

β_k แทน ค่าพารามิเตอร์ของข้อสอบข้อที่ k

2.2 ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบ มีข้อตกลงเบื้องต้นที่สำคัญ 4 ประการ คือ ความเป็นเอกมิติ (unidimensionality) ภายใต้การสอบที่ไม่จำกัดเวลา (nonspeeded test administration) ความเป็นอิสระ (local independence) โมเดลการตอบสนองข้อสอบ (Item Response Models) (Lord and Novick, 1968; Hambleton and Swaminathan, 1985; ศิริชัย กาญจนวาสี, 2550)

2.3.1 ความเป็นเอกมิติ ความเป็น เอกมิติในทฤษฎีการตอบสนองข้อสอบ (IRT) มีข้อตกลงว่า ความสามารถของมนุษย์มีอยู่ทั้งหมด k อย่าง ต่างก็ส่งผลต่อการตอบข้อสอบรวมกันภายใต้แบบสอบ คะแนนของผู้ตอบข้อสอบสามารถอธิบายได้ด้วยความสามารถเพียงอย่างเดียวก็เพียงพอ เรียกข้อตกลงดังกล่าวว่า “ความเป็นเอกมิติ” หากคะแนนของผู้ตอบข้อสอบต้องอธิบายด้วยความสามารถหลายอย่าง เรียกข้อตกลงนั้นว่าการใช้ความสามารถหลายมิติ (multidimensionality) ความหมายของข้อตกลงเบื้องต้นของความเป็นเอกมิติของแบบสอบคือแบบสอบต้องมุ่งวัดคุณลักษณะภายในหรือความสามารถของบุคคลที่เป็นตัวกำหนดพฤติกรรมการตอบข้อสอบแต่ละข้อมีลักษณะเด่นที่สำคัญเพียงลักษณะเดียว (ข้อสอบทุกข้อในแบบสอบมีความเป็นเอกพันธ์) อย่างไรก็ตามข้อตกลงเบื้องต้น ข้อนี้ไม่เข้มงวดนัก ถ้าแบบสอบมีลักษณะเด่น (dominant) ที่จะวัดองค์ประกอบใดองค์ประกอบหนึ่งก็ถือว่าใช้ได้ วิธีการตรวจสอบความเป็นเอกมิติของแบบสอบสามารถทำได้หลายวิธี เช่น การวิเคราะห์องค์ประกอบ (factor analysis) โดยพิจารณาค่าไอเกน (Eigen value) ว่ามีค่าสูงสุดแตกต่างจากค่าอื่นๆ อย่างชัดเจนหรือไม่

2.3.2 ภายใต้การสอบที่ไม่จำกัดเวลา แบบสอบประเภทใช้ความเร็วในการตอบหรือ ที่เรียกว่า Speed Test ผู้ตอบข้อสอบต้องใช้ความสามารถอย่างน้อยสองมิติ คือ มิติด้านความเร็วในการตอบ (response speed) และมิติด้านความสามารถที่แบบสอบต้องการวัด (trait being measured) ดังนั้นเพื่อความเป็นไปตามข้อตกลงเบื้องต้นด้านความเป็นเอกมิติของแบบสอบ ต้องเป็นแบบสอบที่ไม่ใช้ความเร็วในการตอบ (nonspeed test) คือ เป็นแบบสอบที่ถูกกำหนดเวลาไว้อย่างเหมาะสม สำหรับผู้เข้าสอบว่าสามารถทำข้อสอบได้ครบทุกข้อภายในระยะเวลาที่กำหนดและสามารถใช้คะแนนรวมจากการทำ

แบบสอบร่วมกับลักษณะข้อสอบเป็นตัวประมาณค่าความสามารถที่แท้จริงของผู้ตอบข้อสอบโดยไม่มีเงื่อนไขด้านความเร็วเข้ามาเกี่ยวข้อง

2.3.3 ความเป็นอิสระ ความเป็นอิสระในการตอบข้อสอบ ความน่าจะเป็นในการตอบข้อสอบแต่ละข้อถูกต้องเป็นอิสระจากกัน การตอบข้อสอบข้อใดข้อหนึ่งถูกหรือผิดจะไม่มีผลกระทบต่อการตอบข้อสอบข้ออื่นๆ กล่าวในเชิงคณิตศาสตร์ได้ว่าความเป็นอิสระในการตอบข้อสอบ หมายถึง ความน่าจะเป็นในการตอบข้อสอบถูกทั้งหมดมีค่าเท่ากับผลคูณของความน่าจะเป็นในการตอบข้อสอบถูกเป็นรายข้อ นั่นคือผู้สอบที่มีความสามารถ (θ) จะมีความน่าจะเป็นที่จะตอบข้อสอบทั้งข้อ 1 และข้อ 2 ถูกเท่ากันซึ่งได้มาจากความน่าจะเป็นในการตอบข้อสอบข้อที่ 1 ถูก และข้อที่ 2 ถูก ถ้าผู้สอบมีความสามารถ (θ) เท่ากับ 1.5 จะมีความน่าจะเป็นในการตอบข้อสอบข้อที่ 1 ถูกเท่ากับ 0.5 และมีความน่าจะเป็นในการตอบข้อสอบข้อที่ 2 ถูกเท่ากับ 0.6 ดังนั้นผู้สอบที่มีความสามารถ (θ) = 1.5 มีความน่าจะเป็นในการตอบข้อสอบทั้งสองข้อถูกภายใต้เงื่อนไขความเป็นอิสระเป็น (0.5)(0.6) = 0.3 ถ้าแบบสอบมีความเป็นเอกมิติอยู่แล้วความเป็นอิสระในการตอบข้อสอบจะเกิดขึ้นด้วย (Hambleton and Swaminathan, 1985)

กล่าวโดยสรุป ความเป็นอิสระเป็นข้อตกลงที่กำหนดโอกาสที่แต่ละคนจะตอบข้อสอบแต่ละข้อได้ถูกนั้นต้องเป็นอิสระจากกัน โดย 1) ความเป็นอิสระจากกันระหว่างข้อสอบแต่ละข้อ การตอบข้อสอบข้อหนึ่งไม่มีผลกระทบต่อการตอบข้ออื่นๆ ในแบบสอบฉบับนั้น และ 2) ความเป็นอิสระระหว่างผู้ตอบข้อสอบ ผู้ตอบข้อสอบแต่ละคนตอบข้อสอบแต่ละข้อเป็นอิสระจากกัน การทำข้อสอบแต่ละข้ออาจต้องใช้ความสามารถหลายอย่าง ถ้าสามารถกำจัดความสามารถที่ไม่ต้องการวัดออกไปก็จะทำให้การตอบข้อสอบแต่ละข้อของแต่ละคนมีความอิสระ เรียก “ความอิสระอย่างมีเงื่อนไข ” (conditional independence) ถ้าข้อตกลงของความเป็นมิติเดียวเป็นจริง จะมีคุณสมบัติของความเป็นอิสระในการตอบข้อสอบด้วย (Hambleton. 1991; citing Lord. 1980; Lord and Novick 1968) ความเป็นอิสระในการตอบข้อสอบเกิดได้แม้ว่าชุดของข้อมูลไม่ได้มีความเป็นมิติเดียว ความเป็นอิสระในการตอบข้อสอบจะเกิดขึ้นเมื่อกำหนดคุณลักษณะภายในอย่างสมบูรณ์และความสามารถทั้งหลายเหล่านั้นต่างส่งผลต่อการตอบข้อสอบ (Hambleton, 1991)

2.3.4 โค้งคุณลักษณะข้อสอบ เป็นฟังก์ชันแสดงความสัมพันธ์ระหว่างระดับความสามารถของผู้ตอบข้อสอบ (θ_p) กับโอกาสในการตอบข้อสอบได้ถูกต้อง [P (θ_i)] แสดงโดยโค้งคุณลักษณะข้อสอบของแต่ละแบบจำลองที่เลือกมาใช้ในการอธิบาย โค้งคุณลักษณะข้อสอบเป็นฟังก์ชันทางคณิตศาสตร์สามารถใช้อธิบายความสัมพันธ์ระหว่างความน่าจะเป็นหรือโอกาสที่ผู้สอบจะตอบข้อสอบถูกกับระดับความสามารถที่วัดได้โดยใช้ชุดของข้อสอบหรือแบบสอบฉบับนั้น ทั้งนี้ความน่าจะเป็นหรือโอกาสในการตอบข้อสอบถูกจะขึ้นอยู่กับโค้งลักษณะข้อสอบในแต่ละโมเดลที่เลือกใช้โดยที่รูปร่าง (shape) ของโค้งคุณลักษณะข้อสอบในแต่ละข้อมีคุณสมบัติไม่แปรเปลี่ยนไปตามกลุ่มตัวอย่างที่ใช้ ดังนั้น จึงทำให้ความน่าจะเป็นหรือโอกาสในการตอบข้อสอบถูกในแต่ละข้อไม่แปรเปลี่ยนด้วย คุณสมบัตินี้ถือเป็นลักษณะ

เด่นของโมเดลต่างๆ ในทฤษฎีการตอบสนองข้อสอบ โค้งคุณลักษณะข้อสอบมีหลายรูปแบบขึ้นอยู่กับการเลือกใช้จำนวนพารามิเตอร์ของข้อสอบ

2.3 พารามิเตอร์ในทฤษฎีการตอบสนองข้อสอบ

พารามิเตอร์ในทฤษฎีการตอบสนองข้อสอบ แบ่งออกเป็น 2 ชนิด คือ พารามิเตอร์ข้อสอบ (Item parameter) ประกอบด้วย ความยาก (b_i) อำนาจจำแนก (a_i) การเดา (c_i) และความรอบคอบ (γ_i) และพารามิเตอร์ผู้สอบ (person parameter) เป็นระดับความสามารถหรือคุณลักษณะของผู้สอบ (θ) พิสัยของพารามิเตอร์ต่างๆ มีดังนี้ (Hambleton and Swaminathan, 1985; ศิริชัย กาญจนวาสี, 2545)

พารามิเตอร์ความยาก (b_i) ทางทฤษฎีมีค่าตั้งแต่ - ∞ ถึง ∞ แต่ทางปฏิบัติจะมีค่าอยู่ระหว่าง -2.5 ถึง +2.5 ค่าที่เป็นลบแสดงว่าข้อสอบง่าย และค่าที่เป็นบวกแสดงว่าข้อสอบยาก พารามิเตอร์อำนาจจำแนก (a_i) ในทางทฤษฎีมีค่าตั้งแต่ - ∞ ถึง ∞ ควรมีค่าเป็นบวก ตามปกติมีค่าไม่เกิน +2.5 ในทางปฏิบัตินิยมใช้ข้อสอบที่มีค่าพารามิเตอร์อำนาจจำแนกอยู่ระหว่าง +0.5 ถึง +2.5 พารามิเตอร์การเดา (c_i) เป็นค่าแสดงความน่าจะเป็นหรือโอกาสของการตอบข้อสอบได้ถูกต้อง โดยไม่มีความรอบรู้หรือคุณลักษณะในเรื่องนั้นๆ ในทางทฤษฎี พารามิเตอร์การเดามีค่าระหว่าง 0 ถึง 1 โดยทั่วไปนิยมใช้ข้อสอบที่มีค่าพารามิเตอร์การเดาไม่เกิน 0.30

ความรอบคอบ (γ_i) โดย Barton และ Lord (1980) เสนอพารามิเตอร์ที่แสดงถึงความรอบคอบของผู้สอบซึ่งเป็นค่าที่บ่งชี้ว่าผู้สอบที่มีความสามารถสูงอาจตอบข้อสอบได้ไม่ถูกต้องเสมอไป เนื่องจากความไม่รอบคอบในการพิจารณาคำตอบ หรือผู้สอบอาจจะมีสารสนเทศอื่นเกี่ยวกับผู้ออกข้อสอบ จึงเลือกตอบในตัวเลือกที่ไม่ใช่คำตอบที่ถูกต้อง Barton และ Lord กล่าวว่าพารามิเตอร์ตัวนี้จะเหมาะสมในการศึกษาทางทฤษฎีเท่านั้น แต่ในทางปฏิบัติแล้วไม่สามารถพบ ค่าพารามิเตอร์ ตัวนี้ (Hambleton and Swaminatan, 1985)

พารามิเตอร์ผู้สอบเป็นระดับความสามารถของผู้สอบ (θ) ที่ประมาณได้จากโมเดลตามทฤษฎีการตอบสนองข้อสอบ นิยมปรับให้เป็นคะแนนมาตรฐานที่มีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 มีค่าระหว่าง -∞ ถึง ∞ แต่ส่วนใหญ่จะมีค่าอยู่ในช่วง -3.0 ถึง +3.0 ค่าที่เป็นลบแสดงว่าผู้สอบมีความสามารถต่ำและค่าที่เป็นบวกแสดงว่าผู้สอบมีความสามารถสูง

2.4 ทฤษฎีการตอบสนองข้อสอบและการประยุกต์ใช้

ทฤษฎีการตอบสนองข้อสอบ (Item-Response Theory) เกิดขึ้นเมื่อ Lawley ได้เสนอโมเดลของทฤษฎีการตอบสนองข้อสอบไว้ตั้งแต่ ค.ศ.1943 ในระยะนั้นไม่มีการนำโมเดลไปใช้ในทางปฏิบัติ จะมีก็เพียงการเสนอแนวคิดและหลักการ ต่อมาในปี 1970 นับเป็นช่วงที่เริ่มมีการประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบ นำไปสู่ ยุคที่วิธีวิทยาการด้านการวัดมีการพัฒนาขึ้นอย่างรวดเร็ว อาทิ การปรับเทียบ

ข้อสอบ (test equating) การทำหน้าที่ต่างกัน ของข้อสอบ (Differential Item Functioning: DIF) การบริหารการสอบด้วยคอมพิวเตอร์ (computerized test administration) และการสร้างมาตรวัดและการหาปกติวิสัย (scaling and norming) Hambleton, R.K. (1989)

ลักษณะของโมเดลการตอบสนองรายข้อที่ได้รับการพัฒนาขึ้นใหม่ ในช่วงหลายทศวรรษผ่านมา พบว่า นอกจากจะมีโมเดลโลจิสติกแบบเอกมิติ 1, 2 และ 3 พารามิเตอร์ซึ่งใช้กับผลสัมฤทธิ์ทางการเรียน ตลอดจนความถนัดทางการเรียนภายใต้รูปแบบของการตรวจให้คะแนนเป็นแบบทวิวิภาค ก็มีการพัฒนารูปแบบสู่การให้คะแนนแบบพหุวิภาค พัฒนารูปแบบจากการวัดเอกมิติเป็นการวัดแบบพหุมิติและยังมีโมเดลอีกหลายแบบที่นักวัดผลการศึกษาได้พัฒนาจากโมเดลการตอบสนองรายข้อ

การใช้โมเดลตอบสนองรายข้อเป็นประโยชน์ต่อการวัดผลการศึกษา เมื่อข้อมูลสอดคล้องกับข้อตกลงเบื้องต้นของโมเดลที่ใช้ นักวัดผลจะสามารถประมาณค่าความสามารถของผู้สอบได้ โดยพารามิเตอร์นี้เป็นอิสระไม่ขึ้นกับการเปลี่ยนแปลงของข้อสอบ ได้ค่าพารามิเตอร์ของข้อสอบที่ไม่ขึ้นกับกลุ่มผู้สอบ ได้ค่าสถิติที่บ่งชี้ถึงความถูกต้องในการประมาณค่าความสามารถผู้สอบ จำนวนและคุณสมบัติทางสถิติของข้อคำถามและได้มาตรร่วม (common scale) ใช้บรรยายคุณสมบัติผู้สอบและข้อสอบได้ การศึกษาวิจัยเกี่ยวกับทฤษฎีการตอบสนองข้อสอบ นอกจากการพัฒนาโมเดลและ การตรวจสอบโมเดลแล้วยังมีการวิจัยเกี่ยวกับการสร้างมาตรวัดคะแนนความสามารถ (Ability scores) ของผู้สอบ โดยการพัฒนาคะแนนความสามารถในรูปฟังก์ชันของพารามิเตอร์ความสามารถผู้สอบรูปต่างๆ มีการกำหนดน้ำหนักคะแนนแบบต่างๆ และมีการ ศึกษา วิจัยเกี่ยวกับการพัฒนาวิธีประมาณค่าพารามิเตอร์ความสามารถแบบต่างๆ ด้วย

ตอนที่ 3 มโนทัศน์ของการทำหน้าที่ต่างกันของข้อสอบ

3.1 ความลำเอียงและการทำหน้าที่ต่างกันของข้อสอบ

3.1.1 การตรวจสอบความลำเอียงของข้อสอบ โดยนำสารสนเทศการทำหน้าที่ต่างกันของข้อสอบมาวิเคราะห์เชิงตรรกะ (Logical analysis) แล้วให้ผู้เชี่ยวชาญพิจารณาการเขียนข้อสอบ เนื้อหาสาระของข้อสอบและจุดมุ่งหมายของการวัด เพื่อระบุว่าข้อสอบข้อนั้นลำเอียงเข้าข้าง ผู้สอบ กลุ่มใดหรือไม่และเพราะเหตุใดจึงเป็นการตัดสินความลำเอียงของข้อสอบ (ศิริชัย กาญจนวาสี , 2550 อ้างอิงจาก Camilli and Shapard, 1994) นักการศึกษาได้นิยามความหมายของความลำเอียงของข้อสอบ ดังนี้

Scheuneman (1979) และ Rudner, Getson and Knight (1980) กล่าวเกี่ยวกับความลำเอียงของข้อสอบว่ามีความหมายเกี่ยวข้องกับเรื่องสัดส่วนของจำนวนผู้ตอบข้อสอบ โดยความลำเอียงของข้อสอบ เป็นสัดส่วนของผู้สอบที่ตอบข้อสอบได้ถูกต้องไม่เท่ากันในแต่ละกลุ่มประชากรที่ใช้ในการศึกษา เมื่อกลุ่มผู้สอบมีคะแนนเท่ากันและข้อสอบมีความเป็นเอกพันธ์ ส่วน Rudner, Getson and

Knight (1980) กล่าวว่าความลำเอียงของข้อสอบ หมายถึงข้อสอบที่มีค่าความยากสัมพัทธ์สำหรับสมาชิกของผู้สอบกลุ่มหนึ่งมากกว่าสมาชิกของผู้สอบอีกกลุ่มหนึ่ง

Millsap and Everson (1993); Hulin, et al. (1983) กล่าวถึง ความลำเอียงของข้อสอบ (Item bias) และความลำเอียงของการวัด (Measurement bias) ว่าเป็นความไม่ถูกต้องของการวัดอย่างเป็นระบบและถ้าเป็นการวัดความสามารถ ความลำเอียงของข้อสอบจะเกิดขึ้นเมื่อผู้สอบที่มีคุณลักษณะ (trait) ที่ต้องการวัดเท่ากัน มาจากประชากรย่อยต่างกันมีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องต่างกัน ถ้าเป็นการวัดเจตคติการวัดจะเกิดความลำเอียงเมื่อมีความน่าจะเป็นในการตอบข้อสอบในทางบวกแตกต่างกัน

Hulin, Drasgow and Parson (1983) และ Dorans and Kulick (1986) กล่าวเกี่ยวกับความลำเอียงของข้อสอบว่ามีความหมายเกี่ยวข้องกับเรื่องโอกาสในการตอบข้อสอบ Hulin, Drasgow and Parson (1983) ให้ความหมายว่า ความลำเอียงของข้อสอบ หมายถึง โอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกันสำหรับการวัดความสามารถ หรือโอกาสในการตอบข้อสอบในทางบวกแตกต่างกันสำหรับการวัดเจตคติ เมื่อผู้สอบที่มีคุณลักษณะของการวัดในปริมาณเท่ากันแต่มาจากกลุ่มประชากรย่อยที่แตกต่างกัน ส่วนความหมายของความลำเอียงของข้อสอบของ Dorans and Kulick (1986) คือ โอกาสในการตอบข้อสอบได้ถูกต้องของผู้สอบกลุ่มหนึ่งมีค่าต่ำกว่าหรือสูงกว่าผู้สอบอีกกลุ่มหนึ่งที่มีระดับความสามารถเดียวกัน Popham (1981) กล่าวเกี่ยวกับความลำเอียงของข้อสอบว่าเกี่ยวข้องกับความยุติธรรมของการตอบข้อสอบเป็นความโน้มเอียงของข้อสอบเมื่อใช้คะแนนจากข้อสอบนั้น ทำให้การตัดสินผลเป็นไปอย่างไม่ยุติธรรม

3.1.2 การทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning: DIF) เป็นกระบวนการเน้นการใช้วิธีทางสถิติมาตรวจสอบ เพื่อได้สารสนเทศเกี่ยวกับการทำหน้าที่ของข้อสอบ ในกลุ่มผู้สอบกลุ่มย่อยที่มีลักษณะเฉพาะบางอย่างแตกต่างกัน ส่วนความลำเอียงของข้อสอบ (Item bias) มีแนวคิดที่แตกต่างไปเพราะ ความลำเอียงของข้อสอบเป็นกระบวนการตัดสินความยุติธรรม (Fairness) ของข้อสอบ เมื่อความไม่ยุติธรรมของข้อสอบเกิดจากข้อสอบทำให้ผู้เข้าสอบกลุ่มหนึ่งเสียเปรียบผู้เข้าสอบอีกกลุ่มหนึ่ง ความแตกต่างระหว่างกลุ่มผู้เข้าสอบนี้เป็นผลมาจากความแตกต่างของวัฒนธรรมหรือภูมิหลังของผู้สอบมากกว่าความสามารถของผู้สอบ (Angoff, 1993)

การทำหน้าที่ต่างกันของข้อสอบในความหมายโดยทั่วไปบอกถึงความพยายามในการกำหนดเงื่อนไขอย่างมีประสิทธิภาพหรือผลของคะแนนรวมของข้อสอบที่มีความแตกต่างในการทำหน้าที่ในแต่ละกลุ่มของผู้เข้าสอบ ดังนั้น การที่ข้อสอบทำให้ผู้สอบจากต่างกลุ่มที่มีความสามารถเท่ากันเกิดความน่าจะเป็นในการตอบสนองข้อสอบของผู้เข้าสอบได้ถูกต้องแตกต่างกันเพราะถ้าผู้สอบกลุ่มใดมีคุณลักษณะแฝงอื่นสูงกว่าย่อมมีโอกาสที่จะตอบข้อสอบได้ถูกต้องมากกว่าทั้งๆ ที่มีคุณลักษณะแฝงที่ต้องการวัดเท่ากันกับกลุ่มผู้สอบกลุ่มอื่น จึงเกิดการได้เปรียบเสียเปรียบกันระหว่างกลุ่มผู้สอบขึ้น

ลักษณะนี้เดิมใช้คำว่า ความลำเอียงของข้อสอบ (Item Bias) ต่อมาระยะหลังเกิดความคลุมเครือในการที่จะใช้เกณฑ์ในการตัดสินความลำเอียง จึงนำสารสนเทศทางสถิติมาเป็นเกณฑ์ในการตัดสินใจและใช้คำว่า “การทำหน้าที่ต่างกันของข้อสอบ” เนื่องจากเป็นคำที่มีความหมายกลางๆ เหมาะสมในเชิงวิชาการมากกว่าคำว่า“ความลำเอียง”(Holland and Wainer, 1993)

นักการศึกษาได้นิยามความหมายของการทำหน้าที่ต่างกันของข้อสอบไว้ ดังนี้

Dorans and Kulick (1986); Holland and Wainer (1993) และ Mazor, Kanjee and Clauser (1995) กล่าวว่าการทำหน้าที่ต่างกันของข้อสอบมีความหมายเกี่ยวข้องกับเรื่องโอกาสการตอบข้อสอบ โดย Holland and Wainer (1993) กล่าวว่าการทำหน้าที่ต่างกันของข้อสอบ เป็นสารสนเทศทางสถิติของข้อสอบที่ได้จากผลการตอบของผู้สอบต่างกลุ่มกันและมีความสามารถเท่ากันแต่มีโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกัน ส่วน Mazor, Kanjee and Clauser (1995) กล่าวว่าข้อสอบทำหน้าที่ต่างกันเมื่อผู้สอบที่มีความสามารถระดับเดียวกันแต่เป็นสมาชิกกลุ่มย่อยต่างกันมีโอกาสตอบข้อสอบถูกต้องแตกต่างกัน Dorans and Kulick (1986) กล่าวว่า การทำหน้าที่ต่างกันของข้อสอบ เป็นโอกาสในการตอบข้อสอบได้ถูกต้องของผู้สอบกลุ่มหนึ่งมีค่าต่ำกว่าหรือสูงกว่าผู้สอบอีกกลุ่มหนึ่งที่มีระดับความสามารถเดียวกัน

Camilli and Shepard (1994); Millsap and Everson (1993) และศิริชัย กาญจนวาสี (2545) กล่าวเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบว่ามีความเกี่ยวข้องกับความสามารถหลักหรือความสามารถรองของผู้สอบ โดย Camilli and Shepard (1994) กล่าวว่าการทำหน้าที่ต่างกันของข้อสอบ คือ ความเป็นพหุมิติในการวัดของข้อสอบซึ่งแสดงได้จากการแจกแจงความสามารถหลัก (primary ability) ของผู้สอบ 2 กลุ่มขึ้นไป ที่มีความสามารถเท่ากันแต่มีการแจกแจงความสามารถรองแตกต่างกัน (secondary ability) ส่วน Millsap and Everson (1993) กล่าวถึงการทำหน้าที่ต่างกันของข้อสอบว่าเป็นความแตกต่างในการทำหน้าที่ของแบบสอบหรือข้อสอบระหว่างกลุ่มผู้สอบซึ่งถูกจับคู่ตามคุณลักษณะที่วัดโดยแบบสอบหรือข้อสอบนั้น ศิริชัย กาญจนวาสี (2550) กล่าวถึงการทำหน้าที่ต่างกันของข้อสอบว่าเป็นการที่ข้อสอบทำให้ผู้สอบจากต่างกลุ่มกันที่มีความสามารถหรือคุณลักษณะที่มุ่งวัดเท่ากัน มีโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกัน หรือมีฟังก์ชันการตอบสนองข้อสอบแตกต่างกัน การทำหน้าที่ต่างกันของข้อสอบเกิดขึ้นเมื่อนำข้อสอบไปทดสอบกับผู้สอบกลุ่มย่อยต่างกันที่มีความสามารถหลักระดับเดียวกัน หรือมีคุณลักษณะแฝง (Latent Trait) ที่ต้องการวัดเท่ากัน แต่มีความสามารถรองแตกต่างกันทำให้ผู้สอบต่างกลุ่มที่นำมาเปรียบเทียบมีโอกาสตอบถูกแตกต่างกัน

Potenza and Dorans (1995); Scheuneman (1985) (Scheuneman, 1985 cited in Potenza and Dorans, 1995) การทำหน้าที่ต่างกันของข้อสอบว่ามีความหมายเกี่ยวข้องกับสัดส่วนของผู้สอบที่ตอบข้อสอบ คือสัดส่วนของผู้สอบที่ตอบข้อสอบได้ถูกต้องไม่เท่ากันในแต่ละกลุ่มประชากร เมื่อผู้สอบทั้งหมดมีคะแนนเท่ากันทำแบบสอบชุดที่มีความเป็นเอกพันธุ์

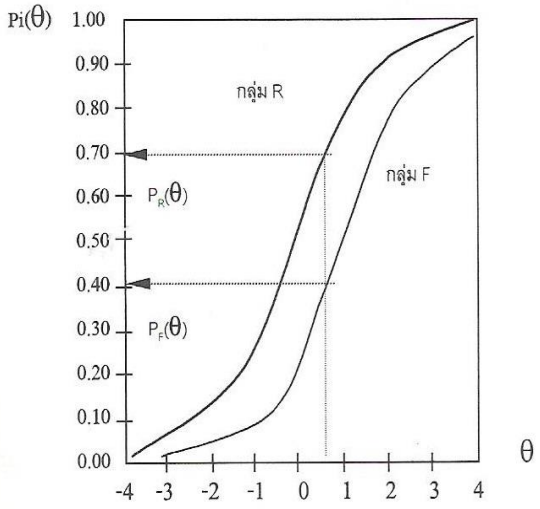
Shealy and Stout (1993) กล่าว ว่าการทำหน้าที่ต่างกันของข้อสอบว่ามีความหมายเกี่ยวข้องกับกับความยุติธรรมของการตอบข้อสอบ โดยการทำหน้าที่ต่างกันของข้อสอบ คือข้อสอบที่เข้าข้าง (favor) ผู้สอบกลุ่มหนึ่งมากกว่าผู้สอบอีกกลุ่มหนึ่งที่นำมาจับคู่เปรียบเทียบกันทำให้ผู้สอบกลุ่มหนึ่งได้ประโยชน์แต่ผู้สอบอีกกลุ่มหนึ่งเสียประโยชน์

สรุปได้ว่า การทำหน้าที่ต่างกันของข้อสอบ เป็นสารสนเทศทางสถิติของข้อสอบ สามารถคำนวณหาดัชนีการทำหน้าที่ต่างกันได้ชัดเจน จากผลการตอบของผู้สอบต่างกลุ่มกันซึ่งถูกจับคู่ตามคุณลักษณะที่วัดโดยแบบสอบหรือข้อสอบนั้น การทำหน้าที่ต่างกันของข้อสอบมีความเกี่ยวข้องกับความสามารถหลักหรือความสามารถรองของผู้สอบ เมื่อผู้สอบทุกคนมีความสามารถเท่ากันแต่มีโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกัน กล่าวคือ ผู้สอบกลุ่มย่อยต่างกันที่มีความสามารถหลักระดับเดียวกัน หรือมีคุณลักษณะแฝง (Latent Trait) ที่ต้องการวัดเท่ากัน แต่มีความสามารถรองแตกต่างกันทำให้ผู้สอบต่างกลุ่มที่นำมาเปรียบเทียบมีโอกาสตอบถูกแตกต่างกัน

3.2 ประเภทของการทำหน้าที่ต่างกันของข้อสอบ

การทำหน้าที่ต่างกันของข้อสอบ (DIF) เป็นการเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มผู้สอบอย่างน้อย 2 กลุ่มขึ้นไป การเปรียบเทียบผลการตอบข้อสอบนิยมทำระหว่างกลุ่มที่มีความสามารถระดับเดียวกัน (ศิริชัย กาญจนวาสี, 2550) กำหนดผู้สอบเป็น 2 กลุ่มคือ กลุ่มอ้างอิง (Reference group) ซึ่งคาดว่าจะได้เปรียบ จากการตอบข้อสอบได้ถูกต้อง จึงน่าจะเป็นผู้ ได้ประโยชน์ จากการตอบ และมีโอกาสในการตอบข้อสอบถูกมากกว่าอีกกลุ่ม กลุ่มเปรียบเทียบ (Focal group) เป็นกลุ่มที่ ผู้วิจัยสนใจศึกษาและคาดว่าจะเป็นกลุ่มที่เสียเปรียบในการตอบข้อสอบหรือเสียประโยชน์ จากการตอบข้อสอบมีโอกาสในการตอบข้อสอบได้ถูกต้องน้อยกว่าผู้สอบอีกกลุ่มหนึ่ง สำหรับเกณฑ์ที่ใช้ในการจำแนกผู้สอบเป็นกลุ่มเปรียบเทียบหรือ กลุ่มอ้างอิง จะใช้เกณฑ์ หลายลักษณะ เช่น เพศ (gender) เชื้อชาติ (race) ภาษา ประสบการณ์ สถาบันการศึกษา ความแตกต่างทางภูมิลำเนา รวมทั้งความแตกต่างทางวัฒนธรรม (cultural difference) เป็นต้น (Holland and Thayer, 1988; Holland and Wainer, 1993; McNamara & Roever, 2004; ศิริชัย กาญจนวาสี, 2550) การทำหน้าที่ต่างกันของข้อสอบ ขนาดและทิศทางของการทำหน้าที่ต่างกันจะแปรเปลี่ยนไปตามระดับความสามารถที่แตกต่างกันของผู้สอบ แบ่งลักษณะข้อสอบที่ทำหน้าที่ต่างกันได้ 2 ประเภท Mellenbergh (1982 อ้างถึงในศิริชัย กาญจนวาสี, 2550) ดังต่อไปนี้

1) ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป (uniform DIF) เกิดเมื่อผู้สอบกลุ่มหนึ่งมีโอกาสในการตอบข้อสอบถูกมากกว่าผู้สอบอีกกลุ่มหนึ่งในทุกระดับความสามารถ พิจารณาได้จากโค้งคุณลักษณะของข้อสอบ (Item Characteristic Curves: ICC) ระหว่างกลุ่มผู้สอบย่อย 2 กลุ่ม โค้งคุณลักษณะของข้อสอบต้องขนานกันหรือไม่เกิดปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบและการเป็นสมาชิกกลุ่ม (group membership) (ศิริชัย กาญจนวาสี, 2550) ดังภาพ2.1

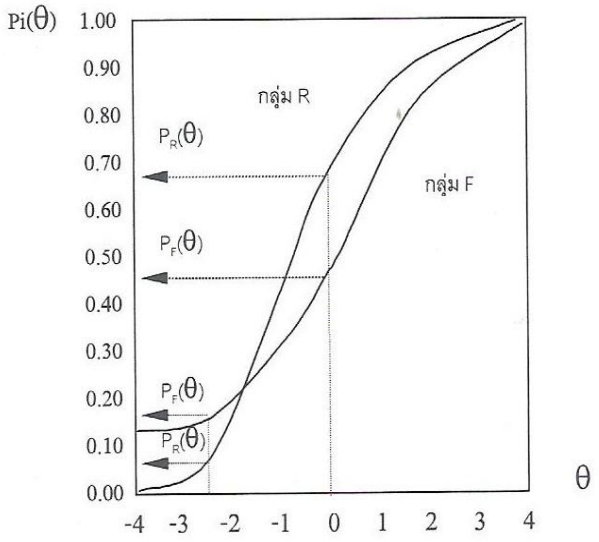


ภาพที่ 2.1 ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป (uniform DIF)

ตามทฤษฎีการตอบสนองข้อสอบ พิจารณา “ปฏิสัมพันธ์” ความแตกต่างของพารามิเตอร์อำนาจจำแนกข้อสอบระหว่างผู้สอบกลุ่มย่อย 2 กลุ่ม (ศิริชัย กาญจนวาสี , 2550) ถ้าข้อสอบทำหน้าที่ต่างกันแบบเอกรูป โค้งคุณลักษณะข้อสอบ (ICC) ระหว่างผู้สอบกลุ่มย่อย 2 กลุ่มจะขนานกันหรือมีฟังก์ชันการตอบสนองข้อสอบเหมือนกัน (Item Response Functions: IRF) แต่ถ้าข้อสอบทำหน้าที่ต่างกันแบบอเนกรูปโค้งคุณลักษณะข้อสอบระหว่างผู้สอบกลุ่มย่อย 2 กลุ่มจะไม่ขนานกันหรือมีฟังก์ชันการตอบสนองข้อสอบไม่เหมือนกัน ดังนั้นความแตกต่างระหว่างโค้งลักษณะข้อสอบทั้งสองแบบจะบ่งบอกถึงขนาดและทิศทางของข้อสอบที่ทำหน้าที่ต่างกัน

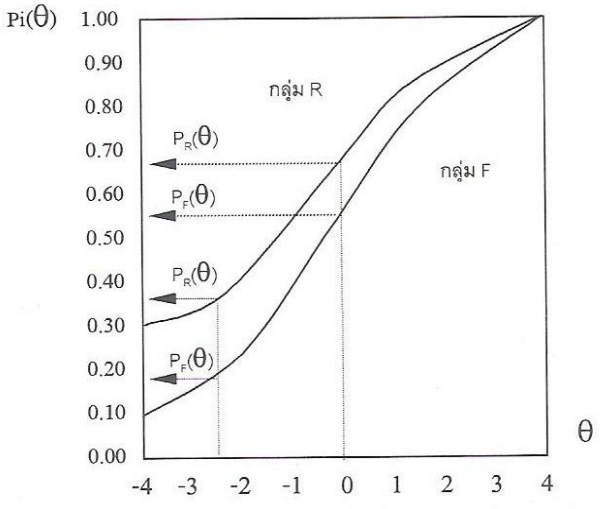
2) ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป (nonuniform DIF) เกิดเมื่อโอกาสในการตอบข้อสอบถูกของผู้สอบระหว่างกลุ่มย่อยสองกลุ่มไม่สม่ำเสมอกัน เมื่อพิจารณาในแต่ละระดับความสามารถเพื่อให้เกิดภาพที่ชัดเจน สามารถพิจารณาได้จากโค้งคุณลักษณะของข้อสอบระหว่างกลุ่มผู้สอบย่อยสองกลุ่ม โค้งคุณลักษณะของข้อสอบจะไม่ขนานกันหรือเกิดปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบและการเป็นสมาชิกของกลุ่ม ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป จำแนกได้ 2 ลักษณะ (ศิริชัย กาญจนวาสี, 2550 อ้างอิงจาก Swaminathan and Roger, 1990)

(1) ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูปโดยมีปฏิสัมพันธ์ไม่เป็นลำดับ (Disordinal interaction) เป็นการทำหน้าที่ต่างกันสำหรับกลุ่มผู้สอบซึ่งเกิดขึ้น เมื่อโค้งลักษณะข้อสอบตัดกันระหว่างช่วงความสามารถของผู้สอบหรือเรียกว่าข้อสอบทำหน้าที่ต่างกันแบบไม่มีทิศทาง (Non-Unidirectional DIF) (ศิริชัย กาญจนวาสี, 2550) ดังภาพที่ 2.2



ภาพที่ 2.2 ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป (Nonuniform DIF)
โดยมีปฏิสัมพันธ์ไม่เป็นลำดับ (Disordinal interaction)

(2) ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูปโดยมีปฏิสัมพันธ์เป็นลำดับ (Ordinal interaction) เป็นการทำหน้าที่ต่างกันสำหรับกลุ่มผู้สอบ เกิดขึ้นเมื่อโค้งลักษณะข้อสอบต่างกันอย่างไม่สม่ำเสมอแต่ไม่ตัดกันหรืออาจตัดกันนอกช่วงความสามารถของผู้สอบตรงปลายสุดของช่วงความสามารถต่ำหรือสูง หรือเรียกว่าข้อสอบทำหน้าที่ต่างกันแบบมีทิศทางเดียว (Unidirectional DIF) (ศิริชัย กาญจนวาสี, 2550) ดังภาพที่ 2.3



ภาพที่ 2.3 ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป (Nonuniform DIF)
โดยมีปฏิสัมพันธ์เป็นลำดับ (Ordinal interaction)

3.3 หลักการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

กระบวนการในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบมี 2 ขั้นตอน คือการตรวจสอบในระหว่างการสร้างข้อสอบโดยการให้ผู้เชี่ยวชาญพิจารณาจากการทำตารางที่วิเคราะห์หลักสูตร (Table of Specification) ว่าข้อสอบสามารถวัดได้ตรงกับจุดมุ่งหมายหรือพฤติกรรมและเนื้อหาที่ต้องการวัดหรือไม่ รวมถึงภาษาที่ใช้ในการสื่อความหมายนั้นมีความชัดเจนหรือคลุมเครือเพื่อป้องกันการเกิดปัญหาในการตีความเมื่อนำข้อสอบนั้นไปใช้ รวมถึงการป้องกันไม่ให้ข้อสอบเกิดประโยชน์ต่อผู้เข้าสอบกลุ่มใดกลุ่มหนึ่ง การตรวจสอบโดยให้ผู้เชี่ยวชาญพิจารณานี้ถือเป็นการตรวจสอบความตรงเชิงเนื้อหา (Content Validity) ขั้นตอนที่ 2 การตรวจสอบโดยใช้สถิติ เมื่อมีการนำแบบสอบไปใช้ ศิริชัย กาญจนวาสี (2550) กล่าวว่าการตรวจการทำหน้าที่ต่างกันของข้อสอบเป็นการเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มผู้สอบอย่างน้อยสองกลุ่มที่มีความสามารถหลัก (Primary Ability) ที่มุ่งวัดเท่ากันแต่คาดว่าจะเกิดการได้เปรียบเสียเปรียบกันระหว่างกลุ่มอ้างอิง (Reference Group) ซึ่งคาดว่าน่าจะได้เปรียบในการตอบข้อสอบข้อนั้นหรือมีโอกาสตอบข้อสอบได้ถูก ต้องมากกว่า ส่วนอีกกลุ่มคือกลุ่มเปรียบเทียบ (Focal Group) ซึ่งเป็นกลุ่มที่สนใจศึกษาและคาดว่าเป็นกลุ่มที่เสียเปรียบ การเปรียบเทียบจำเป็นต้องจับคู่ (Matching) ผู้สอบตามความสามารถ เป็นเงื่อนไขสำคัญของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เกณฑ์การจับคู่ (Matching Criteria) นิยมใช้เกณฑ์ภายนอก (External Criterion) และเกณฑ์ภายใน (Internal Criterion)

3.3.1 การตรวจสอบโดยใช้เกณฑ์ภายนอก หรือการใช้วิธีภายนอก (External Method) โดยนำแบบสอบที่สร้างขึ้นมาหาความสัมพันธ์กับเกณฑ์ที่เป็นมาตรฐาน แต่มีปัญหา ของเกณฑ์ภายนอกที่จะนำมาหาความสัมพันธ์ เพราะถ้าเกณฑ์ไม่มีมาตรฐานแล้วจะทำให้การตรวจสอบขาดความถูกต้อง การวิเคราะห์การทำหน้าที่ต่างกัน โดยใช้เกณฑ์ภายนอกนี้สามารถนำไปใช้ได้ทั้งข้อสอบรายข้อและข้อสอบที่รวมเป็นแบบสอบทั้งฉบับ โดยการใช้คะแนนจากแบบสอบอื่นเป็นเกณฑ์ภายนอก ใช้เทคนิคการวิเคราะห์การถดถอย (Regression Analysis) เพื่อทำการเปรียบเทียบเส้นกราฟความสัมพันธ์ระหว่างตัวแปรเกณฑ์กับตัวแปรทำนายระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ หลักการนี้มีจุดมุ่งหมายเพื่อสร้างสมการทำนายตัวแปรเกณฑ์เป็นคะแนนของแบบสอบอื่นจากตัวแปรทำนาย (คะแนนรายข้อหรือคะแนนแบบสอบระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบในการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบจะใช้คะแนนรวมของแบบสอบทั้งฉบับเป็นตัวแปรทำนาย) สำหรับตัวแปรเกณฑ์ภายนอกอาจใช้คะแนนรวมทั้งฉบับ /เกรดเฉลี่ย /ผลสัมฤทธิ์ในงานที่เกี่ยวข้องของผู้สอบ (Cronbach, 1970 อ้างถึงใน ศิริชัย กาญจนวาสี, 2550)

สมการทำนายสำหรับกลุ่มอ้างอิงและกลุ่มเปรียบเทียบแสดงได้ดังนี้

กลุ่มอ้างอิง (R) $Yi = AR + BRXi$

กลุ่มเปรียบเทียบ (F) $Yi = AF + BFXi$

เมื่อ Yi = คะแนนของตัวแปรเกณฑ์ภายนอก

Xi = คะแนนของตัวแปรทำนาย

A = ค่าคงที่หรือค่าตัดแกน (intercept)

B = ค่าความชัน (slope)

จากฟังก์ชันการทำนาย 2 สมการ สามารถเปรียบเทียบค่าตัดแกน (A) และค่าความชัน (B) ของเส้นกราฟระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบได้ ถ้าเส้นกราฟดังกล่าวมีค่าความชันหรือค่าตัดแกนแตกต่างกันสำหรับข้อสอบใดหรือแบบสอบใดแสดงว่าข้อสอบหรือแบบสอบนั้นมีการทำหน้าที่ต่างกันโดยเข้าข้างกลุ่มผู้สอบที่มีค่าตัดแกนหรือค่าความชันที่สูงกว่า การใช้เกณฑ์ภายนอกมีข้อดี คือเกณฑ์ที่ใช้มีอิสระจากข้อสอบและแบบสอบที่ต้องการตรวจสอบแต่มีจุดอ่อนตรงที่ความเหมาะสมของเกณฑ์ที่จะนำมาใช้ ในทางปฏิบัติเป็นทางยากที่จะหาตัวแปรเกณฑ์ภายนอกจากแบบสอบฉบับอื่นที่มีความตรงเชิงทำนายและมีความยุติธรรมสำหรับกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ถ้าตัวแปรเกณฑ์ภายนอกขาดคุณสมบัติดังกล่าวจะทำให้ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ขาดความแม่นยำและ ขาดความสมบูรณ์

3.3.2 การตรวจสอบโดยใช้เกณฑ์ภายในหรือการใช้วิธีภายใน (Internal Method) ถือเป็นกระบวนการสนับสนุนการตรวจสอบความตรงเชิงโครงสร้าง เป็นวิธีที่ได้รับความนิยมเป็นอย่างมาก เนื่องจากตัดปัญหาเรื่องเกณฑ์ที่ไม่มีมาตรฐาน การตรวจสอบสามารถดำเนินการได้จากการศึกษาโครงสร้างภายในของข้อสอบ โดยพิจารณาคะแนนที่ได้จากผลการตอบข้อสอบของผู้สอบแต่ละกลุ่มว่าวัดในคุณลักษณะที่ต้องการวัดตามโครงสร้างเช่นเดียวกันหรือไม่ด้วยการวิเคราะห์ผลจากการตอบข้อสอบและความสามารถหรือคะแนนจริงของผู้สอบที่ได้จากแบบสอบฉบับนั้น เพื่อนำมาเปรียบเทียบระหว่างผู้สอบจากกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีความสามารถหรือคะแนนจริงเท่ากัน ว่าจะมีผลการตอบหรือโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกันหรือไม่ เพื่อบ่งชี้ถึงการทำหน้าที่ต่างกันของข้อสอบ หลักการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจะเป็นการเปรียบเทียบผลการตอบข้อสอบระหว่างผู้สอบสองกลุ่มที่มีระดับความสามารถเดียวกัน โดยกำหนดให้ผู้สอบกลุ่มหนึ่งเป็นกลุ่มอ้างอิงและผู้สอบอีกกลุ่มหนึ่งเป็นกลุ่มเปรียบเทียบ ถ้าข้อสอบทำหน้าที่ต่างกันแล้วโอกาสในการตอบข้อสอบถูกของผู้สอบแต่ละกลุ่มจะไม่เท่ากัน (ศิริชัย กาญจนวาสี, 2550)

3.4 ทฤษฎีการตอบสนองข้อสอบและการทำหน้าที่ต่างกันของข้อสอบ

3.4.1 รูปแบบการตรวจให้คะแนนของข้อสอบ ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจำแนกได้ 2 ประเภท คือ ทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนแบบทวิวิภาคและแบบพหุวิภาค ทฤษฎีการตอบสนองข้อสอบแต่ละประเภทแบ่งตามข้อตกลงเกี่ยวกับคุณลักษณะภายในที่ใช้ในการตอบข้อสอบ คือโมเดลที่ใช้ข้อตกลงเกี่ยวกับความเป็นเอกมิติและโมเดลที่ใช้ข้อตกลงเกี่ยวกับการใช้

ความสามารถพหุมิติ โมเดลที่ใช้ข้อตกลงเกี่ยวกับความเป็นเอกมิติ (Unidimensionality) และทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนแบบทวิวิภาค (dichotomously IRT Model) โมเดลที่นิยมใช้ประกอบด้วยโมเดลการตอบสนองแบบ 1, 2 และ 3 พารามิเตอร์ โมเดลการตอบสนองข้อสอบทั้งสามแบบมีโค้งลักษณะข้อสอบ (item characteristic curves: ICC) ที่เขียนในรูปฟังก์ชันปกติสะสม (Normal Ogive Function) และฟังก์ชันโลจิส (Logistic Function) (ศิริชัย กาญจนวาสี, 2550) ดังตารางที่ 2.1

ตารางที่ 2.1 ฟังก์ชันทางคณิตศาสตร์ของโมเดลการตอบสนองข้อสอบ

โมเดล (Model)	ฟังก์ชันปกติสะสม (Normal Ogive Function)	ฟังก์ชันโลจิส (Logistic Function)
1พารามิเตอร์	$\mathrm{P}_i(\theta)=\int_{-\infty}^{\theta-b_1}\frac{1}{\sqrt{2\pi}}e^{-Z^{2}/2}dZ$	$\mathrm{P}_i(\theta)=\frac{1}{1+e^{-(\theta-b_i)}}$
2 พารามิเตอร์	$\mathrm{P}_i(\theta)=\int_{-\infty}^{a_1(\theta-b_i)}\frac{1}{\sqrt{2\pi}}e^{-Z^{2}/2}dZ$	$\mathrm{P}_i(\theta)=\frac{1}{1+e^{-Da_1(\theta-b_i)}}$
3 พารามิเตอร์	$\mathrm{P}_i(\theta)=C_i+(1-C_i)\int_{-\infty}^{a_1(\theta-b_i)}\frac{1}{\sqrt{2\pi}}e^{-Z^{2}/2}dZ$	$\mathrm{P}_i(\theta)=c_i+\frac{(1-c_i)}{1+e^{-Da_1(\theta-b_i)}}$

โดยที่ θ คือ ระดับความสามารถของผู้ตอบข้อสอบใดๆ ที่ประมาณได้จากโมเดลตามทฤษฎีการตอบสนองข้อสอบ ปรับให้เป็นคะแนนมาตรฐานที่มีค่าเฉลี่ยเป็น0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 ซึ่ง θ มีพิสัยระหว่าง $\pm\infty$ แต่ในทางปฏิบัติส่วนใหญ่ใช้ค่า θ ระหว่าง ± 3

$P_i(\theta)$ คือ ความน่าจะเป็นที่ผู้ตอบข้อสอบที่ได้มาจากการสุ่มและมีความสามารถ θ ตอบคำถามข้อที่ I ได้ถูกต้อง

b_i คือ ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i ซึ่งเป็นค่าที่แสดงตำแหน่งของโค้งคุณลักษณะ ข้อสอบ (ICC) ตามแกนนอนบนเสกล θ ณ จุดที่โค้งมีความชันมากที่สุด หรือที่เรียกว่าจุดเปลี่ยนโค้งหรือที่ตำแหน่งต่อไปนี้

สำหรับโมเดล 1 และ 2 พารามิเตอร์ $b_i = \theta$ ที่ $P_i(\theta)= 0.5$

สำหรับโมเดล 3 พารามิเตอร์ $b_i = \theta$ ที่ $P_i(\theta)= (1+c_i)/2$

ค่า b_i มีพิสัยระหว่าง $\pm\infty$ แต่ในทางปฏิบัติส่วนใหญ่ใช้ค่า b_i ระหว่าง ±2.5

a_i คือ ค่าอำนาจจำแนกของข้อสอบข้อที่ i ซึ่งเป็นสัดส่วนต่อความชันของโค้งคุณลักษณะข้อสอบ (ICC) ณ จุดเปลี่ยนโค้ง หรือที่จุด $\theta = b_i$ ค่า a_i มีค่าสูงแสดงว่าข้อสอบข้อนั้น

มีความชันที่มีค่าสูงจึงจำแนกผู้ที่มีความสามารถแตกต่างกันได้ดี ค่า a_i มีพิสัยระหว่าง $\pm\infty$ แต่ในทางปฏิบัติส่วนใหญ่ใช้ค่า a_i ระหว่าง +0.5 ถึง +2.5

c_i คือ ความน่าจะเป็นของการเดาได้ถูกต้อง ซึ่งเป็นความน่าจะเป็นที่ผู้สอบมีความสามารถต่ำมากๆ ตอบข้อสอบข้อที่ i ได้ถูก ค่า c_i มีพิสัยระหว่าง 0 ถึง 1 แต่ในทางปฏิบัตินิยมใช้ข้อสอบที่มีค่า c_i ระหว่าง 0 ถึง 0.3

e คือ ค่าคงที่ของลอการิทึมธรรมชาติ มีค่าประมาณ 2.71828

D คือ ค่าองค์ประกอบการปรับเสกลให้โลจิสติกฟังก์ชัน มีค่าใกล้เคียงกับฟังก์ชันปกติสะสม (Normal Ogive Function) มากที่สุดเท่าที่จะเป็นไปได้ มีค่า 1.70

การวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบให้ค่าพารามิเตอร์ที่สามารถอ้างอิงได้กับกลุ่มตัวอย่างทั่วไปจึงมีการนำมาใช้กันมาก ในปัจจุบันจำแนกได้เป็น 2 ประเภท คือ ตามลักษณะการให้คะแนน ทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนแบบทวิวิภาค และแบบพหุวิภาค สำหรับ “ข้อสอบที่มีการตรวจให้คะแนนแบบพหุวิภาค” ถือว่าในปัจจุบันเป็นรูปแบบการตรวจให้คะแนนที่นำมาใช้กันอย่างแพร่หลาย (Kim, Chosen, Alagoz and Kim, 2007)

รูปแบบข้อสอบที่มีการตรวจให้คะแนนแบบพหุวิภาค เช่น การให้คะแนนมาตรประมาณค่าที่วัดคุณลักษณะต่างๆ ข้อคำถามที่มีการกำหนดคะแนนตามลำดับขั้น ถ้าเป็นแบบวัดคุณลักษณะคะแนนแต่ละค่ามักแสดงค่าถึงระดับของคุณลักษณะ แต่ถ้าเป็นแบบสอบผลสัมฤทธิ์ทางการให้คะแนนแต่ละค่าแสดงถึงระดับความสามารถของผู้สอบที่ตอบข้อกระทงนั้น อาจมีวิธีให้คะแนนแตกต่างกันไป เช่น การให้คะแนนตามระดับความมั่นใจในการตอบตามความสามารถในการตัดตัวลวงของข้อกระทงนั้น และการให้คะแนนความรู้บางส่วน (Partial Knowledge) ซึ่งเป็นวิธีที่มีผู้สนใจศึกษากันมาก (Kim, Chosen, Alagoz and Kim, 2007)

การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบตามทฤษฎีการตอบสนองข้อสอบที่มีรูปแบบการตรวจให้คะแนนแบบพหุวิภาคกำลังเป็นที่รู้จักในการนำมาใช้ในวงการวัดผลทางการศึกษา โมเดลในการวิเคราะห์ตามแนวทางดังกล่าวมีด้วยกันหลายโมเดล เช่น Graded Response Model (GRM) Rating Scale Model (RSM) และ Generalized Partial Credit Model (GPCM) โมเดลที่รู้จักกันดีในปัจจุบันได้แก่โมเดล GRM, RSM, GPCM เป็นต้น ขณะที่มีการพัฒนาโมเดลก็มีการพัฒนาโปรแกรมคอมพิวเตอร์เพื่อใช้วิเคราะห์มากขึ้น ตัวอย่างโปรแกรมคอมพิวเตอร์ที่มีการพัฒนาขึ้นและเป็นที่รู้จักกันดี ได้แก่ MULTILOG, BIGSTEPS, PARSCALE ส่วนใหญ่แล้วการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบการพัฒนาในระยะแรกจะใช้กับแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค ต่อมาได้มีผู้พัฒนาโมเดล IRT เพื่อใช้กับแบบสอบและแบบวัดที่มีการตรวจให้คะแนนแบบพหุวิภาคหรือลักษณะมาตรประมาณค่า (rating data) โมเดลในแนวทฤษฎีนี้เรียกว่า polytomous Item Response Models มีโมเดลที่สำคัญคือ

Graded-Response Model (GRM) Modified Graded-Response Model Partial Credit Model (PCM) Generalized Partial Credit Model (GPCM) Rating Scale Model (RSM) และ Nominal Response Model (NRM) (ศิริชัย กาญจนวาสี, 2550) เมื่อพิจารณาทางปฏิบัติแล้ว พบว่าการสอบตามบริบทของประเทศไทยยังนิยมใช้เครื่องมือวัดทางการศึกษาที่มีรูปแบบการตรวจให้คะแนนแบบทวิวิภาคเป็นส่วนใหญ่

3.4.2 การทดสอบการทำหน้าที่ต่างกันของข้อสอบด้วยทฤษฎีการตอบสนองข้อสอบ (Testing DIF with IRT) มีวิธีการหลัก 5 วิธี สำหรับทดสอบสมมติฐานทางสถิติว่าไม่เกิดการทำหน้าที่ต่างกันของข้อสอบ (Marie, 2007) ได้แก่ การทดสอบความแตกต่างของค่าพารามิเตอร์ b (test of b difference) เป็นแนวทางที่ง่ายในการทดสอบเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบ สถิติที่ใช้ในการทดสอบสมมุติฐานศูนย์ไม่ยุ่งยาก ดัชนีการแปลผลพิจารณาจากความแตกต่างของกลุ่มสนใจที่ศึกษาสองกลุ่ม วิธี item drift method วิธี Lord's chi-square วิธี empirical sampling distributions for DIF indices และวิธี measurement of model comparisons

3.4.3 การวัดขนาดของการทำหน้าที่ต่างกันของข้อสอบด้วยทฤษฎีการตอบสนองข้อสอบ (Measure Size of DIF with IRT) มีอย่างน้อย 4 วิธี ได้แก่ วิธี simple area indices วิธี probability difference indices วิธี b parameter difference และวิธี ICC method for small sample

3.5 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

การจำแนกวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สามารถจำแนกตามลักษณะของเกณฑ์ ประกอบด้วย ลักษณะการตรวจให้คะแนน (แบบทวิวิภาคและแบบพหุวิภาค) มิติของตัวแปรเกณฑ์ (กลุ่มวิธีที่ใช้คะแนนสังเกตได้และกลุ่มวิธีที่ใช้คะแนนของตัวแปรแฝง) มิติลักษณะของสถิติวิเคราะห์ (กลุ่มที่ใช้สถิติพาราเมตริกและกลุ่มที่ใช้สถิตินันพาราเมตริก) มีรายละเอียด ดังนี้

3.5.1 จำแนกตามลักษณะการตรวจให้คะแนน

1) กลุ่มวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบทวิวิภาค (Dichotomous DIF Methods) หรือการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบสองค่า (dichotomous DIF procedures) แบบสอบที่มีลักษณะของการตรวจให้คะแนนแบบนี้ได้แก่ แบบสอบชนิดเลือกตอบที่มีการให้คะแนนในการตอบถูกเป็น 1 คะแนน ในขณะที่ตอบผิดได้ 0 คะแนน การนำเสนอการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบถูกกำหนดอย่างจริงจังเพื่อให้เกิดความยุติธรรมในการใช้แบบสอบ และเกิดความตรงต่อการแสดงความหมายที่แฝงอยู่ของคะแนนเหล่านั้น งานวิจัยที่เกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบมีความหลากหลายมากขึ้น เดิมเน้นไปที่ข้อสอบที่มีการตรวจให้คะแนนแบบ ทวิวิภาค ในช่วงระยะเวลาที่ผ่านมามีความพยายามพัฒนาแนวทางเลือกใหม่ของวิธีการวัดซึ่งช่วยจุดประกายให้เกิดประเด็นที่น่าสนใจ และยังมีการทำหน้าที่ต่างกันของ

ข้อสอบชนิดอื่นที่นอกเหนือจากข้อสอบที่มีการตรวจให้คะแนนแบบ ทวิวิภาค (Kim, Chosen, Alagoz and Kim, 2007)

2) กลุ่มวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบพหุวิภาค (Polytomously Methods) เช่น ข้อสอบวัดภาคปฏิบัติ (performance assessment) ข้อสอบความเรียง (essay items) การตัดสินคุณภาพของแฟ้มสะสมงาน (portfolio assessment) ข้อสอบที่วัดการอ่าน (reading item) และข้อสอบที่วัดการเขียน (writing item) รวมไปถึงข้อสอบปลายเปิด (open-ended item) เป็นต้น

3.5.2 จำแนกตามมิติของตัวแปรเกณฑ์

1) กลุ่มวิธีที่ใช้คะแนนที่สังเกตได้ (Observed Score) ค่าพารามิเตอร์แปรเปลี่ยนไปตามกลุ่มผู้สอบ วิธีในกลุ่มนี้มักวิเคราะห์ตามทฤษฎีทางการสอบแบบดั้งเดิม เรียกกลุ่มที่ไม่ใช้ทฤษฎีการตอบสนองข้อสอบ (Non-IRT Approach) ใช้คะแนนรวมของผู้สอบเป็นเกณฑ์การจับคู่ของกลุ่มผู้สอบ วิธีการตรวจสอบที่สำคัญในกลุ่มนี้ ได้แก่ การวิเคราะห์ความแปรปรวน (ANOVA)การวิเคราะห์การถดถอยโลจิสติกพหุวิภาค (Polytomous Logistic Regression) วิธีแมนเทล-แฮนส์เซลทั่วไป (General Mantel-Haenszel) และวิธีดัชนีมาตรฐานพหุวิภาค (Polytomous Standardization)

2) กลุ่มวิธีที่ใช้คะแนนของคุณลักษณะหรือตัวแปรแฝง (Latent Variable) วิเคราะห์บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ ตัวแปรแฝงหรือคุณลักษณะดังกล่าวถูกใช้เป็นเกณฑ์การจับคู่กลุ่มผู้สอบ ค่าพารามิเตอร์คงที่ไม่ว่าจะใช้กลุ่มผู้สอบใด วิธีการตรวจสอบที่สำคัญในกลุ่มนี้ ได้แก่ วิธีการให้คะแนนบางส่วนทั่วไป (Generalized Partial Credit Model: GPCM) วิธีอัตราส่วนไลค์ลิฮูดในรูปทั่วไป (General IRT Likelihood Ratio) วิธีการให้คะแนนบางส่วน (Partial Credit Model: PCM) และวิธีซิปเทสท์พหุวิภาค (Polytomous SIBTEST)

3.5.3 จำแนกตามมิติลักษณะของสถิติวิเคราะห์

1) กลุ่มที่ใช้สถิติพาราเมตริก (Parametric Approach) การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบมีข้อตกลงเบื้องต้นของโมเดลสำหรับอธิบายความสัมพันธ์ระหว่างคะแนนของข้อสอบและการจับคู่ตัวแปร

2) กลุ่มที่ใช้สถิตินันพาราเมตริก (Nonparametric Approach) การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบไม่มีข้อตกลงเบื้องต้นของโมเดล ความหลากหลายของวิธีทางสถิติช่วยพัฒนาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนแบบทวิวิภาคและพหุวิภาค (Penfield, 2005)

วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่สำคัญๆ แสดงในตารางที่ 2.2–2.3 (Feinstein, 1995; Potenza and Dorans, 1995 อ้างถึงใน ศิริชัย กาญจนวาสี, 2550)

ตารางที่ 2.2 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับรูปแบบการตรวจให้คะแนนแบบทวิวิภาค ที่เป็นเอกมิติ จำแนกตามลักษณะของข้อมูล (Marie, 2007)

วิธีการ (Method)	สถิติพาราเมตริกซ์ (Par) /สถิตินันพาราเมตริกซ์ (non-p)	การจับคู่ตัวแปรคะแนนที่สังเกตได้ (Obs)/ คะแนนของตัวแปรแฝง (Lat)	ลักษณะการให้คะแนน ทวิวิภาค (D) /พหุวิภาค (P)	การทดสอบ(T) /การวัด(M)	DIFเอกรูป (U) /DIFอเนกรูป(N)
Mentel-Haenszel	Non-p	Obs	D/P	T/M	U
Standardization	Non-p	Obs	D	M	U
Chi-square methods	Non-p	Obs	D	T	U/N
SIBTEST	Non-p	Lat	D/P	T/M	U/N
Logistic Regression	**Par**	**Obs**	**D/P**	**T/M**	**U/N**
Likelihood ratio test	Par	Obs/Lat	D/P	T/M	U/N
Prob. Diff. indices	Par	Lat	D	M	U/N
B parameter indices	Par	Lat	D	M	U/N
General IRT-LR	Par	Lat	D/P	T/M	U/N
IRT LRT	Par	Lat	D/P	T	U/N
IRT methods	Par	Lat	D/P	T/M	U/N
Lord's chi-squared test	Par	Lat	D	T	U/N
Log linear models	Par	Obs	D/P	T	U/N
Mixed effect models	Par	Lat	D/P	T	U/N

หมายเหตุ: **(Par)**ametric / **(non-p)** arametric, **(Obs)**erved / **(Lat)**ent, **(D)**ichotomously / **(P)**olytomously, **(T)**est DIF / **(M)**easure DIF, **(U)**niform / **(N)**onuniform

จากตารางที่ 2.2 แสดงวิธีตรวจสอบ DIF ใน 14 วิธีการสำหรับข้อมูลที่มีลักษณะเป็นเอกมิติ (uni-dimensional) (Marie, 2007) จำแนกรายละเอียดที่ต้องการศึกษาออกเป็น 5 มิติ ได้แก่ มิติประเภทของสถิติ (พาราเมตริก (Par)ametric และนอนพาราเมตริก (non-p)arametric) มิติประเภทของคะแนน (คะแนนที่สังเกตได้ (Obs)erved และคะแนนแฝง (Lat)ent) มิติรูปแบบการตรวจให้คะแนน (ทวิวิภาค (D)ichotomously และพหุวิภาค(P)olytomously) มิติประเภทของการตรวจสอบ (การตรวจสอบ (T)est DIF และการวัดขนาด (M)easure DIF) มิติรูปแบบ DIF (เอกรูป (U)niform และอเนกรูป (N)onuniform) ข้อมูลในตารางที่นำไปสู่การตัดสินใจเปรียบเทียบการเลือกใช้วิธีการตรวจสอบเพื่อให้สอดคล้องและเหมาะสมตามจุดมุ่งหมายของสารสนเทศที่ต้องการ

วิธีการตรวจสอบการทำหน้าที่ต่างกันที่มีการให้คะแนนแบบทวิวิภาค (Dichotomous DIF) และหุวิภาค (polytomous DIF) แสดงในตารางที่ 2.3

ตารางที่ 2.3 วิธีการตรวจสอบการทำหน้าที่ต่างกันที่มีการให้คะแนนแบบทวิวิภาคและพหุวิภาค

ประเภทและตัวแปรเกณฑ์	พาราเมตริก (Parametric Form)	นันพาราเมตริก (Nonparametric Form)
1. DIF แบบทวิวิภาค		
1.1 คะแนนที่สังเกตได้ (observed score)	1. ANOVA การวิเคราะห์ความแปรปรวน (Cleary and Hilton, 1968) 2. Logistic Regression: LR วิธีถดถอยโลจิสติก (Swaminathanand Rogers, 1990)	1. TID (Transformed Item Difficulty) วิธีแปลงค่าความยากของข้อสอบ (Cleary and Hilton, 1968; Angoff and Ford, 1973) 2. MH (Mantel-Haenszel) วิธีแมนเทล-แฮนส์เซล (Holland and Thayer, 1988, 1989) 3. STND (Standardization) วิธีดัชนีมาตรฐานการปรับให้เป็น มาตรฐานด้วยน้ำหนักตัวประกอบ (Dorans and Kulick, 1986)
1.2 คะแนนของตัวแปรแฝง (Latent variable)	1. IRT-D^2 วิธีวัดพื้นที่ความแตกต่างระหว่างโค้งการตอบสนองข้อสอบ (Linn et al.,1981; Raju,1990; KimandCohen,1991) 2. Lord's χ^2 วิธีไค-สแควร์ของลอร์ด (Lord,1980) 3. General IRT Likelihood Ratio วิธีอัตราส่วนไลค์ลิฮูดทั่วไป(Thissen, Steinberg,andWainer, 1993) 4. Loglinear IRT-LR (Loglinear IRT Likelihood Ratio) วิธีอัตราส่วนไลค์ลิฮูดลอกลิเนียร์ (Thissen,Steinberg,andWainer,1993)	1. SIBTEST วิธีซิปเทสท์ (ShealyandStout, 1993)
2. DIF แบบพหุวิภาค		
2.1 คะแนนที่สังเกตได้ (observed score)	1. ANOVA การวิเคราะห์ความแปรปรวน (Cleary and Hilton, 1968) 2. Polytomous Logistic Regression การวิเคราะห์การถดถอยโลจิสติกพหุวิภาค (Swaminathan and Rogers, 1990)	1. Polytomous STND (Polytomous Standardization) วิธีดัชนีมาตรฐานพหุวิภาค (Dorans and Kulick, 1986) 2. GMH (General Mantel-Haenszel) วิธีแมนเทล-แฮนส์เซลทั่วไป (Holland and Thayer, 1988, 1989)
2.2 คะแนนของตัวแปรแฝง (Latent variable)	1. General IRT-LR (General IRT Likelihood Ratio) วิธีอัตราส่วนไลค์ลิฮูดในรูปทั่วไป (Thissen, Steinberg,andWainer, 1993) 2. PCM (Partial Credit Model) วิธีการให้คะแนนบางส่วน (Master, 1982)	1. Polytomous SIBTEST วิธีซิปเทสท์พหุวิภาค (ShealyandStout, 1993) 2. GPCM (Generalized Partial Credit Model) วิธีการให้คะแนนบางส่วนทั่วไป (Muraki, 1992, 1993)

ที่มา: ศิริชัย กาญจนวาสี (2550)

3.6 ประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

เกณฑ์ในการประเมินหรือตัดสินเกี่ยวกับประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ นิยมตัดสินจากดัชนีบ่งชี้ 2 ตัว คืออัตราความถูกต้องและอัตราความคลาดเคลื่อนของการตรวจสอบ ศิริชัย กาญจนวาสี (2550) กล่าวว่าในการคำนวณค่าสถิติตามวิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบมีจุดมุ่งหมายเพื่อเพื่อทดสอบนัยสำคัญของผลการตรวจสอบ มีสมมติฐานศูนย์คือข้อสอบไม่ได้ทำหน้าที่ต่างกัน (H_0: No DIF) ซึ่งผลการทดสอบสมมติฐานของวิธีที่ใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการต่างๆ นำไปสู่การตัดสินใจว่าจะปฏิเสธสมมุติฐานศูนย์ (Reject H_0) หรือจะยอมรับสมมติฐานศูนย์ (Accept H_0) ผลของการตัดสินใจมีโอกาสเกิดได้ ดังตารางที่ 2.4

ตารางที่ 2.4 คุณภาพของการตรวจสอบประสิทธิภาพของการทำหน้าที่ต่างกันของข้อสอบ

ผลการตัดสินใจ ตามผลการตรวจสอบ	สมมติฐานศูนย์คือข้อสอบไม่ได้ทำหน้าที่ต่างกัน (H_0: No DIF) สถานการณ์จริง	
	H_0 ถูก	H_0 ผิด
ยอมรับสมมติฐานศูนย์ (Accept H_0)	**ตัดสินถูก (True negative)** **ระดับความเชื่อมั่น (1-α)**	ตัดสินผิด (False negative) ความคลาดเคลื่อนประเภทที่ 2 (Type II Error, β)
ปฏิเสธสมมติฐานศูนย์ (Reject H_0)	ตัดสินผิด (False positive) ความคลาดเคลื่อนประเภทที่1 (Type I Error, α)	**ตัดสินถูก (True positive)** **อำนาจการทดสอบ (1-β)**

ที่มา: ศิริชัย กาญจนวาสี (2550)

ตารางที่ 2.4 เมื่อคำนวณค่าสถิติตามวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ นำไปสู่การตัดสินใจสรุปผลการตรวจสอบ คือ กรณีการตัดสินใจถูก กับ กรณีการตัดสินใจผิด รายละเอียดดังนี้

กรณีการตัดสินใจถูก มีโอกาสที่จะเกิดขึ้น 2 ลักษณะ คือ การสรุปถูกว่า

(1) ข้อสอบไม่ได้ทำหน้าที่ต่างกัน (No DIF) ในความเป็นจริง (True negative)

(2) ข้อสอบทำหน้าที่ต่างกัน (DIF) ตามความเป็นจริง (True positive)

กรณีการตัดสินใจผิด มีโอกาสที่จะเกิดขึ้น 2 ลักษณะ คือ

(1) สรุปผิดว่าข้อสอบทำหน้าที่ต่างกัน ทั้งที่ความเป็นจริงข้อสอบไม่ได้ DIF (False positive)

(2) การสรุปผิดว่าข้อสอบไม่ DIF ทั้งที่ความเป็นจริงข้อสอบ DIF (False negative)

เนื่องจากระหว่างคู่ของอำนาจการทดสอบหรืออัตราความถูกต้อง (1-β) กับ ความคลาดเคลื่อนประเภทที่ 2 (β) และคู่ของระดับความเชื่อมั่น (1-α) กับ ความคลาดเคลื่อนประเภทที่ 1 (α) เป็นค่าดัชนีที่มีเสกลผกผันกัน ดังนั้นการพิจารณาบ่งชี้คุณภาพก็สามารถพิจารณาเพียง อัตราความถูกต้อง (correct identification) ของการตรวจพบการทำหน้าที่ต่างกันของข้อสอบ และอัตราความคลาดเคลื่อนประเภทที่ 1 (Type I Error) มีความเพียงพอสำหรับสารสนเทศที่ต้องการ (ศิริชัย กาญจนวาสี, 2550) ดัชนีดังกล่าว คือ อัตราความถูกต้อง (correct identification) หมายถึง จำนวนข้อที่ตรวจสอบได้ถูกต้องว่าทำหน้าที่ต่างกัน ต่อจำนวนข้อที่ทำหน้าที่ต่างกันทั้งหมดในแบบสอบ (คำนวณเป็นค่าร้อยละ) และอัตราความคลาดเคลื่อนประเภทที่ 1 (type I error rate) คือ การระบุผิดพลาดว่าข้อสอบทำหน้าที่ต่างกัน (False Positive: FP) ทั้งที่ความเป็นจริงข้อสอบทำหน้าที่ไม่ต่างกัน คำนวณจาก จำนวนข้อที่ระบุผิดพลาดว่าทำหน้าที่ต่างกันทั้งที่ความจริงทำหน้าที่ไม่ต่างกัน ต่อจำนวนข้อที่ทำหน้าที่ไม่ต่างกันในแบบสอบ (คำนวณเป็นค่าร้อยละ)

3.7 ทฤษฎีการตอบสนองข้อสอบสำหรับรูปแบบการให้คะแนนแบบทวิวิภาค

ทฤษฎีการตอบสนองข้อสอบ ประกอบด้วย ระดับชั้นของโมเดลและวิธีการสำหรับวิเคราะห์รายละเอียดของการตอบสนองข้อสอบรายข้อในทดสอบหรือแบบสอบถาม รูปแบบที่ถูกศึกษามากที่สุดเป็นประเด็นข้อตกลงเบื้องต้นที่เกี่ยวกับการเป็นเอกมิติ (unidimensional IRT) มีข้อสันนิษฐานว่าการตอบสนองของรายการขึ้นอยู่กับตัวแปรเดียวที่ซ่อนเร้นต่อเนื่อง แบบสอบต้องมุ่งวัดคุณลักษณะเด่นเพียงลักษณะเดียว หมายถึง คุณลักษณะภายในหรือความสามารถของบุคคลเป็นตัวกำหนดพฤติกรรมการตอบข้อสอบแต่ละข้อมีลักษณะเด่นที่สำคัญเพียงลักษณะเดียว โดยตัวแปรที่เป็นตัวแปรแฝงถูกวัดในเชิงโครงสร้างของทางจิตวิทยาในรูปแบบข้อคำถาม ศูนย์กลางของการตอบสนองเป็นกลุ่มของข้อคำถามที่ครอบคลุมความสัมพันธ์ของข้อสอบกับโครงสร้างพื้นฐานทางจิตวิทยาหน้าที่ของการตอบสนองรายข้อถูกอธิบายด้วยรูปแบบการตอบสนองรายข้อกับตัวแปรแฝง Thissen and Steinberg (2006) ทฤษฎีการตอบสนองข้อสอบ แบบ 2 พารามิเตอร์ ฟังก์ชันโลจิส (Logistic Function) (ศิริชัย กาญจนวาสี, 2550) มีรูปแบบสมการ ดังนี้

$$P_i(\theta) = \frac{1}{1+e^{-Da_1(\theta-b_i)}}$$

โดยที่ θ คือ ระดับความสามารถของผู้ตอบข้อสอบใดๆ ที่ประมาณได้จากโมเดลตามทฤษฎีการตอบสนองข้อสอบ

$P_i(\theta)$ คือ ความน่าจะเป็นที่ผู้ตอบข้อสอบที่ได้มาจากการสุ่มและมีความสามารถ θ ตอบคำถามข้อที่ I ได้ถูกต้อง

b_i คือ ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i ซึ่งเป็นค่าที่แสดงตำแหน่งของโค้งคุณลักษณะ ข้อสอบ (ICC)

a_i คือ ค่าอำนาจจำแนกของข้อสอบข้อที่ i ซึ่งเป็นสัดส่วนต่อความชันของโค้งคุณลักษณะข้อสอบ (ICC) ณ จุดเปลี่ยนโค้ง

e คือ ค่าคงที่ของลอการิทึมธรรมชาติ มีค่าประมาณ 2.71828

D คือ ค่าองค์ประกอบการปรับเสกลให้โลจิสติกฟังก์ชัน มีค่าใกล้เคียงกับฟังก์ชันปกติสะสม (Normal Ogive Function) มากที่สุดเท่าที่จะเป็นไปได้ มีค่า 1.70

ตอนที่ 4 มโนทัศน์ของขนาดอิทธิพล

4.1 การรายงานขนาดอิทธิพลของงานวิจัยในปัจจุบัน

ปัจจุบันการนำเสนองานวิจัยมักเกิดคำถามใน 3 ประเด็นหลัก ประเด็นแรก ได้แก่ งานวิจัยชิ้นนี้มีวัตถุประสงค์/ความเหมาะสมและมีขอบเขตของงานเพียงใด ประเด็นที่สอง คือสิ่งที่ได้ออกแบบไว้ตามวิธีวิทยาการมีความแกร่งน่าเชื่อถือที่จะสนับสนุนข้อค้นพบจากงานชิ้นนั้นได้เพียงใด และประเด็นที่สาม คือมีความรู้ใดเป็นข้อค้นพบเพิ่มเติมจากสิ่งที่เป็นที่ทราบกันดีอยู่แล้วบ้าง (Hudson, 2009) ซึ่ง Hudson (2009) ได้อ้างอิงข้อแนะนำจากนักสถิติว่าก่อนที่จะเสนองานวิจัยสิ่งเริ่มต้นที่สำคัญที่ต้องนำเสนอคือประเด็นที่เกี่ยวข้องกับขนาดของกลุ่มตัวอย่าง อำนาจการทดสอบและขนาดอิทธิพลเมื่อได้ศึกษา เอกสารและงานวิจัยที่เกี่ยวข้องแล้วยังไม่สามารถนำสิ่งเหล่านั้นมาใช้ประโยชน์ได้ เท่ากับว่าเป็นผลที่ล้มเหลวจากการศึกษานำร่อง นักวิจัยหลายคนยังสับสนเกี่ยวกับการรายงานขนาดอิทธิพลอยู่มาก จากการวิจัยทางด้านจิตวิทยา พบว่าการรายงานขนาดอิทธิพลแต่ละประเภทมีความชัดเจนทั้งความหมายและกิจกรรมที่เกิดขึ้น จำแนกประเภทของการรายงานขนาดอิทธิพลได้เป็น 3 ประเภทคือ ขนาดอิทธิพลจากอำนาจการทดสอบ ขนาดอิทธิพลจากการสังเคราะห์งานวิจัย และขนาดอิทธิพลจากผลของรายงานผลที่เกิดจากการวิจัยแต่ละประเภทมีเป้าหมาย ดังนี้

1) ขนาดอิทธิพลจากอำนาจการทดสอบ ขนาดอิทธิพลจากอำนาจการทดสอบเป็นการอธิบายผลนัยสำคัญทางการปฏิบัติซึ่ง Cohen (1988) แบ่งขนาดอิทธิพลออกเป็น 3 ขนาด คือ ขนาดเล็ก ขนาดกลางและขนาดใหญ่

2) ขนาดอิทธิพลจากการสังเคราะห์งานวิจัย ขนาดอิทธิพลจากการสังเคราะห์งานวิจัยเน้นไปที่การศึกษาเชิงประจักษ์ (empirical studies) และมุ่งเน้นการนำเสนอ ชายแดนความรู้ (The state of knowledge) ของเรื่องที่สนใจและมุ่งไปที่ประเด็นที่สำคัญที่งานวิจัยยังไม่สามารถแก้ได้ พร้อมกับพิจารณาถึงประเด็นที่ยังไม่ทราบที่จะทำให้ได้ข้อมูลใหม่ในปริมาณที่มากที่สุด

3) ขนาดอิทธิพลจากรายงานผลที่เกิดจากการวิจัย โดยเน้นไปที่การศึกษาหลักฐานเชิงประจักษ์เป็นหลัก

การทำหน้าที่ต่างกันของข้อสอบ ประกอบด้วย 2 ประเด็นหลักที่ควรมีการรายงานขนาดอิทธิพลร่วมด้วย คือ ประเด็นแรกเมื่อพบว่าสถิติทดสอบมี นัยสำคัญทางสถิติ จึงจำเป็นต้องรายงานผลการวัดขนาดของอิทธิพลขณะนี้ถือว่าเป็นสิ่งที่จำเป็นเพราะเมื่อกลุ่มตัวอย่างที่มีขนาดเล็กมากๆ มักจะไม่มีนัยสำคัญเท่ากับกรณีที่ศึกษากับกลุ่มตัวอย่างที่มีขนาดใหญ่ที่จะเกิดการมีนัยสำคัญง่ายกว่า เมื่อกลุ่มตัวอย่างขนาดใหญ่มากๆ การศึกษาขนาดอิทธิพลจะเป็นตัวช่วยทำให้เข้าใจความสัมพันธ์ระหว่างตัวแปรซึ่งบางครั้งพบว่าถึงแม้ค่าสถิติทดสอบจะมีนัยสำคัญทางสถิติแต่ก็ไม่มีความหมายอะไรมากนัก ประเด็นที่สอง Zumbo และ Hubley (2003) เป็นผู้สนับสนุนให้นักวิจัยมีการรายงานเกี่ยวกับขนาดอิทธิพลทั้งที่มีนัยสำคัญทางสถิติและไม่มีนัยสำคัญทางสถิติ

4.2 ความหมายของขนาดอิทธิพล

การใช้คำศัพท์เกี่ยวกับขนาดอิทธิพล (Effect Size) มีทั้งกลุ่มที่ใช้คำว่า ความเข้มของอิทธิพล (Effect Magnitude) หรือความเข้มของอิทธิพลโดยตรง (ชยุตม์ ภิรมย์สมบัติ , 2547) ซึ่งการนิยามในรูปของขนาดอิทธิพลและ/หรือสถิติอื่นๆ ในกลุ่มความสัมพันธ์ สามารถวิเคราะห์ความหมายและนิยามดังนี้ กลุ่มที่กล่าวถึงสัดส่วนหรืออัตราส่วนทางคณิตศาสตร์ ได้แก่ Jacob Cohen (1969 อ้างถึงใน นงลักษณ์ วิรัชชัย, 2542) เป็นผู้พัฒนาสูตรการประมาณค่าขนาดอิทธิพลในยุคแรก ได้นิยามในเชิงปฏิบัติว่าขนาดอิทธิพลเป็นอัตราส่วนระหว่างผลต่างของค่าเฉลี่ยจากกลุ่มทดลองและกลุ่มควบคุมกับค่าส่วนเบี่ยงเบนมาตรฐานรวม Fidler and Thompson (2001) กล่าวว่าขนาดอิทธิพลเป็นความแตกต่างมาตรฐานของค่าเฉลี่ยระหว่างกลุ่มทดสอบกับกลุ่มควบคุม (standardized effect size) หรือ ความผันแปรของตัวแปรตามที่สามารถอธิบายหรือทำนายได้ด้วยตัวแปรต้น กลุ่มที่กล่าวถึงความสัมพันธ์ระหว่างตัวแปร Kirk (1996) กล่าวว่าขนาดอิทธิพลเป็นขนาด/ระดับความสัมพันธ์ของตัวแปรต้นและตัวแปรตาม Hair et al. (1998) กล่าวว่าขนาดอิทธิพลเป็นค่าประมาณของระดับของปรากฏการณ์ที่ศึกษาว่ามีอยู่หรือเกิดขึ้นในประชากร Trusty, Thompson and Petrocelli (2004) กล่าวว่าขนาดอิทธิพลเป็นค่าที่ใช้วัดอำนาจหรือระดับของความสัมพันธ์ นงลักษณ์ วิรัชชัย (2542) กล่าวว่า ขนาดอิทธิพลเป็นค่าสถิติที่บอกปริมาณผลของตัวแปรจัดกระทำที่มีต่อตัวแปรตามในการวิจัยเชิงทดลอง จากการศึกษาความหมายของผู้ที่ให้คำนิยามไว้ จึงสรุปความหมาย “ขนาดอิทธิพล” ว่าเป็นค่าสถิติที่ใช้วัดเพื่อบอกระดับความสัมพันธ์ระหว่างอิทธิพลของตัวแปรต้นที่มีต่อตัวแปรตาม

4.3 การตัดสินใจทางการวิจัย

การตัดสินใจทางการวิจัยมีในกรณีของการมีนัยสำคัญทางสถิติ กับ การมีนัยสำคัญทางปฏิบัติ (Chohen, 1988) การที่ผู้วิจัยจะตัดสินใจว่ายอมรับสมมติฐานทางการวิจัยที่ตั้งไว้หรือไม่ต้องทำการทดสอบสมมติฐานทางสถิติคือ H_0 (Null Hypothesis) H_1 (Alternative Hypothesis) แล้วพิจารณาจากระดับนัยสำคัญทางสถิติที่กำหนดไว้ การวิจัยทางสังคมศาสตร์นิยมกำหนดนัยสำคัญทางสถิติเท่ากับ .05

เมื่อค่าสถิติที่คำนวณได้ตกอยู่ในพื้นที่วิกฤต สรุปได้ว่าปฏิเสธ (reject) H_o และยอมรับ (accept) H_1 ว่าเป็นจริง นั่นคือตัดสินใจยอมรับสมมติฐานทางการวิจัยที่ตั้งไว้ แต่ถ้าค่าสถิติที่คำนวณได้ไม่ตกอยู่ในพื้นที่วิกฤต สรุปว่ายอมรับ (retain) H_o แต่ไม่ได้หมายความว่า H_o เป็นจริง เพียงแต่ไม่มีหลักฐานที่พอเพียงในการปฏิเสธ H_o นั่นคือตัดสินใจไม่ยอมรับสมมติฐานทางการวิจัยที่ตั้งไว้ การนำนัยสำคัญทางสถิติ (Statistical Significance) มาใช้ในการทดสอบสมมติฐานทางสถิติ เพื่อตัดสินใจยอมรับหรือไม่ยอมรับสมมติฐานทางการวิจัยพบว่ามีข้อที่ควรพิจารณาอยู่ 2 ประการคือ 1) ความแตกต่างอย่างมีนัยสำคัญทางสถิติ สามารถเปลี่ยนแปลงไปตามขนาดของกลุ่มตัวอย่าง 2) ความแตกต่างอย่างมีนัยสำคัญทางสถิติไม่ได้บ่งบอกถึงปริมาณความแตกต่างของค่าเฉลี่ยของประชากร 2 กลุ่ม เช่น นัยสำคัญทางสถิติที่ระดับ.01 และ .05 ไม่ได้บ่งบอกว่าปริมาณความแตกต่างของค่าเฉลี่ยที่นัยสำคัญทางสถิติที่ระดับ .01 จะมีปริมาณความแตกต่างของค่าเฉลี่ยมากกว่าที่นัยสำคัญทางสถิติที่ระดับ .05 จึงได้มีการนำนัยสำคัญทางปฏิบัติ (Practical Significance) มาร่วมใช้ในการพิจารณาซึ่งสามารถบอกถึงปริมาณความแตกต่างของค่าเฉลี่ยว่ามีมากน้อยเท่าไร เพื่อช่วยตัดสินใจว่าปริมาณความแตกต่างของค่าเฉลี่ยที่พบมากพอต่อการนำไปใช้ในทางปฏิบัติหรือไม่ และนัยสำคัญทางปฏิบัติไม่ขึ้นอยู่กับขนาดของกลุ่มตัวอย่าง จึงทำให้นัยสำคัญทางปฏิบัติมีความแน่นอนไม่ว่าจะได้มาจากกลุ่มตัวอย่างขนาดเล็กหรือขนาดใหญ่ นัยสำคัญทางปฏิบัติที่รู้จักกันทั่วไปคือขนาดอิทธิพล เพราะเป็นสัดส่วนระหว่างความแตกต่างของค่าเฉลี่ย 2 กลุ่มต่อส่วนเบี่ยงเบนมาตรฐานการประมาณค่าขนาดอิทธิพลโดย Jacob Cohen (1969 อ้างถึงใน นงลักษณ์ วิรัชชัย, 2542) การคำนวณค่าอัตราส่วนระหว่างผลต่างของค่าเฉลี่ยจากกลุ่มทดลองและกลุ่มควบคุมกับค่าส่วนเบี่ยงเบนมาตรฐานรวม (Kirk, 1996) ในทางปฏิบัติขนาดอิทธิพลจากนิยามของ Jacob Cohen เรียกว่า ผลต่างค่ามาตรฐานระหว่างคะแนนเฉลี่ย (standardized mean differences) ดังสมการต่อไปนี้

$$\delta = \frac{\mu_\varepsilon - \mu_C}{\sigma_{pooled}}$$

โดยที่ δ แทน ขนาดอิทธิพล

μ_ε แทน ค่าเฉลี่ยของประชากรกลุ่มทดลอง

μ_C แทน ค่าเฉลี่ยของประชากรกลุ่มควบคุม

σ_{pooled} แทน ส่วนเบี่ยงเบนมาตรฐานรวมของกลุ่มทดสองและกลุ่มควบคุม

สถิติแต่ละชนิดมีความเหมาะสมในการประมาณค่าขนาดอิทธิพลต่างกัน นักสถิติหลายท่านได้เสนอความเห็นเกี่ยวกับการประมาณค่าขนาดอิทธิพลที่เหมาะสมสำหรับงานวิจัยเชิงปริมาณทั่วไปๆไปเอาไว้ การประมาณค่าขนาดอิทธิพลควรใช้สถิติในกลุ่มความสัมพันธ์หรือความแปรปรวนที่ถูกอธิบาย เพราะสามารถนำมาใช้ได้กับการวิเคราะห์เชิงปริมาณที่ใช้โมเดลเชิงเส้นทั่วไปและสามารถประยุกต์ใช้ได้กับงานวิจัยเชิงปริมาณทั้งที่เป็นเชิงทดลองและไม่ใช่เชิงทดลอง เนื่องจากเป้าหมายหนึ่งของการวิจัยเชิงปริมาณส่วนใหญ่ต้องการศึกษาความสัมพันธ์หรือความผันแปรร่วมกันระหว่างกลุ่มตัวแปรนั่นเอง

สำหรับการแปลความหมายของขนาดอิทธิพลเกณฑ์ของ Cohen ตรงกับเกณฑ์ของการแปลความหมายสัมประสิทธิ์สหสัมพันธ์ (Kirk; 1996) คือ 0.2 หมายถึง มีอิทธิพลหรือมีความสัมพันธ์ต่ำ 0.5 หมายถึง มีอิทธิพลหรือมีความสัมพันธ์ปานกลาง 0.8 หมายถึง มีอิทธิพลหรือมีความสัมพันธ์สูง

4.4 ขนาดอิทธิพลกับการศึกษาการทำหน้าที่ต่างกันของข้อสอบ

เมื่อตรวจพบว่าข้อสอบทำหน้าที่ต่างกัน คำถามที่มักจะตามมาคือ “แล้วจะอธิบายการทำหน้าที่ต่างกันของข้อสอบข้อนี้อย่างไร?” การตอบคำถามดังกล่าวค่อนข้างเฉพาะเจาะจง เนื่องจากเป็นเนื้อหาระหว่างขนาดอิทธิพลกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่วิเคราะห์ภายใต้ข้อมูลที่ใช้โมเดลเป็นฐาน (model-based data) การแปลความหมายของค่าพารามิเตอร์ตามแบบแผนโมเดลการตอบสนองข้อสอบมีวัตถุประสงค์เพื่อการอธิบายในระดับหน่วยของการวัดโดยมากเป็นการวิเคราะห์แบบสอบหรือแบบสอบถามทางด้านการศึกษาและทางจิตวิทยา ดังนั้นในการวิเคราะห์เพื่อทำการตรวจสอบความแตกต่างในการตอบสนองข้อสอบ เช่น การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบที่เปรียบเทียบในพารามิเตอร์ของข้อสอบรายข้อการแสดงผลหรือขนาดของมันจึงมีความซับซ้อนมากที่สุด ผู้อ่านที่อ่านผลจากรายงานที่เกี่ยวข้องกับทฤษฎีการตอบสนองข้อสอบจำเป็นมากที่จะต้องเข้าใจความหมายของการแปลผลค่าพารามิเตอร์และมุ่งการตีความเพื่อทำความเข้าใจไปในส่วนของการนำเสนอผล

4.4.1 The Threshold (b) Parameter ค่าพารามิเตอร์ b เป็นหน่วยของตัวแปรแฝง (latent variable) จัดเป็นคุณลักษณะของสิ่งที่สนใจศึกษาค่าของ b=0 บ่งบอกถึงระดับค่าเฉลี่ยของคุณลักษณะพื้นฐานของประชากรค่าของ b=1.00 เป็นค่าเบี่ยงเบนมาตรฐานของคุณลักษณะพื้นฐานของประชากร พารามิเตอร์ b แตกต่างกันไปตามหน่วยของความเบี่ยงเบนมาตรฐานการเปรียบเทียบค่าพารามิเตอร์ b พยายามหลีกเลี่ยงปัญหาที่เกิดขึ้นจากการแปลความหมายระหว่างสัดส่วนของตัวเลขที่อาจคลาดเคลื่อน เช่น พารามิเตอร์ b ที่แตกต่างระหว่าง .98 และ .99 ผลที่ได้มีความแตกต่างกันมากกว่าพารามิเตอร์ b ที่แตกต่างระหว่าง .50 และ .51 อย่างไรก็ตามการทำความเข้าใจค่าพารามิเตอร์ในทฤษฎีการตอบสนองข้อสอบ (IRT) ต้องเข้าใจถึงความสัมพันธ์ของพารามิเตอร์ b ว่าไม่ปรากฏเป็นรูปเชิงเส้นตรง จำเป็นต้องแปลความหมายร่วมกับผลของพารามิเตอร์ a ตารางที่ 5 แสดงความสัมพันธ์ระหว่างพารามิเตอร์ b และสัดส่วนของการตอบสนองเชิงบวก ในโมเดล 2 พารามิเตอร์หรือสัดส่วนของการตอบสนองใน k ระดับมีสมการเป็นดังนี้

$$P+ = \int_{-\infty}^{\infty} T(u)\,\phi(\theta)\,d\,\theta$$

โดยที่ $T(u)$ แทน การตอบสนองในระดับของ u

$\phi(\theta)$ แทน การกระจายของประชากรตามคุณลักษณะที่เกิดจากการวัด

ตารางที่ 2.5 แสดงสัดส่วนของระดับการตอบสนอง k ระดับ ตามช่วงระดับความสามารถของ b ใช้ค่า b ระหว่าง ±3.00 เมื่อกำหนดค่า a เป็น .50, 1.00 และ 2.00

b	a		
	.50	1.00	2.00
-3.00	0.81	0.93	0.99
-2.75	0.79	0.91	0.98
-2.50	0.77	0.89	0.97
-2.25	0.74	0.87	0.95
-2.00	0.72	0.84	0.93
-1.75	0.70	0.81	0.91
-1.50	0.67	0.78	0.87
-1.25	0.64	0.74	0.83
-1.00	0.62	0.70	0.78
-0.75	0.59	0.65	0.72
-0.50	0.56	0.60	0.65
-0.25	0.53	0.55	0.58
0.00	0.50	0.50	0.50
0.25	0.47	0.45	0.42
0.50	0.44	0.40	0.35
0.75	0.41	0.35	0.28
1.00	0.38	0.30	0.22
1.25	0.36	0.26	0.17
1.50	0.33	0.22	0.13
1.75	0.30	0.19	0.09
2.00	0.28	0.16	0.07
2.25	0.26	0.13	0.05
2.50	0.23	0.11	0.03
2.75	0.21	0.09	0.02
3.00	0.19	0.07	0.01

จากตารางที่ 2.5 เป็นค่าปกติที่ได้จากการสังเกตของความชัน กำหนดช่วงห่างของพารามิเตอร์ b ช่วงละ .25 จะเห็นว่าพารามิเตอร์ a ค่อนข้างแตกต่างกันเล็กน้อย สัดส่วนของความแตกต่างอยู่ระหว่างช่วง .02-.05 ความแตกต่างดังกล่าวอาจจะสามารถหรืออาจจะไม่มีความสำคัญต่อการใช้งานจริงในทางปฏิบัติสำหรับใช้เป็นสารสนเทศในแบบสอบหรือแบบสอบถามในอีกด้านหนึ่งถ้ามีความแตก ต่างของพารามิเตอร์ b ที่ .05 หรือที่เกี่ยวข้องกับความแตกต่างขนาดใหญ่ในสัดส่วนของการตอบสนองทางบวกของ 2 พารามิเตอร์ หรือสัดส่วนของการตอบสนองใน k เมื่อพารามิเตอร์ b แตกต่างกันกัน .50

4.4.2 การทดสอบระดับนัยสำคัญทางสถิติสำหรับการทำหน้าที่ต่างกันของข้อสอบ (Test of Significance for DIF) ด้วยสถิติไค -สแควร์ ในวิธีถดถอยโลจิสติก การทดสอบค่านัยสำคัญทางสถิติสำหรับการทำหน้าที่ต่างกันของข้อสอบได้มีการกำหนดวิธีการศึกษาเป็นรูปแบบโมเดลที่เป็นระดับขั้น (Hierarchy) โดยมีระดับขั้นสำหรับทดสอบ 3 ระดับ (Zumbo and Bruno, 1999) คือระดับที่ 1 เป็นระดับขั้นการกำหนดเงื่อนไขของตัวแปร ระดับที่ 2 การนำกลุ่มตัวแปรเข้าสู่สมการ และระดับที่ 3 เป็นการศึกษาปฏิสัมพันธ์ในสมการสำหรับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก ทดสอบค่านัยสำคัญทางสถิติไค -สแควร์ (Chi-square: χ^2) ค่าสัมประสิทธิ์ของวิธีถดถอยโลจิสติก ประมาณค่าโดยวิธี maximum likelihood เมื่อก่อนโมเดลประมาณค่าในอดีตพิจารณาค่าจากความเพียงพอหรือความสัมพันธ์ส่วนประกอบของตัวแปรที่ต้องการตรวจสอบ การตรวจสอบความแตกต่างระหว่าง ระดับที่ 1 กับระดับที่ 2 เป็นค่าการเปลี่ยนแปลงที่เกิดจากการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป การตรวจสอบความแตกต่างระหว่าง ระดับที่ 2 กับระดับที่ 3 เป็นค่าการเปลี่ยนแปลงที่เกิดจากการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป การตรวจสอบความแตกต่างใช้สถิติ G^2 ส่วนการทดสอบนัยสำคัญใช้สถิติ χ^2 ที่มี df=1 เมื่อ Swaminatha และ Rogers (1990) เสนอวิธีการทดสอบนัยสำคัญใช้สถิติ χ^2 ที่มี df=2 ซึ่งยอมให้มีทั้งการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและแบบอเนกรูป ซึ่งสามารถตรวจสอบพร้อมๆ กันได้

4.4.3 ขนาดอิทธิพลในการทำหน้าที่ต่างกันของข้อสอบ

การวัดขนาดอิทธิพลในการทำหน้าที่ต่างกันของข้อสอบ พิจารณาจากตารางที่ 2.6

ตารางที่ 2.6 ค่า R^2 ของการวัดขนาดอิทธิพลในการทำหน้าที่ต่างกันของข้อสอบ(Zumbo, 1999)

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 2.6 พบว่า การคำนวณ R^2 สำหรับการวัดขนาดอิทธิพลในการทำหน้าที่ต่างกันของข้อสอบในข้อสอบที่มีรูปแบบการให้คะแนนแบบพหุวิภาคมีเพียงวิธีเดียว คือ R^2 ที่เป็นลำดับขั้น (R-squared for ordinal) ข้อสอบที่มีรูปแบบการให้คะแนนแบบทวิวิภาคมีการคำนวณ R^2 สามวิธี ประกอบด้วย

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

2) วิธีการคำนวณ R^2 ของ WLS (Weighted-least-squares Squared) เป็นกลุ่มนามบัญญัติวิธีการนี้เหมาะสำหรับใช้ประโยชน์กับกลุ่มตัวอย่างที่มีตัวแปรเป็นลำดับขั้นภายในกลุ่ม

3) วิธีการคำนวณ R^2 สำหรับการจัดลำดับข้อมูลหรือ R^2 ที่เป็นลำดับขั้น (R-squared for ordinal) เป็นกลุ่มจัดลำดับ เหมาะกับกลุ่มตัวแปรที่อยู่ในงานวิจัยประเภทพฤติกรรมศาสตร์และสังคมศาสตร์ เป็นตัวแปรที่มีความเป็นคุณลักษณะแฝงที่มีความต่อเนื่อง

ตอนที่ 5 เอกสารและงานวิจัยที่เกี่ยวข้อง

5.1 งานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบในประเทศ

การนำเสนอในส่วนของเอกสารและงานวิจัยที่เกี่ยวข้องได้เน้นการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบทวิวิภาค (Dichotomously) แบบพหุวิภาค (Polytomously) ที่มุ่งศึกษาความเป็นแบบสอบที่เป็นมิติเดียว (uni-dimensional) และพหุมิติ (Multidimensional) มีรายละเอียดดังนี้

สุรศักดิ์ อมรรัตนศักดิ์ (2531) ได้ทำการศึกษาเปรียบเทียบผลของวิธีวิเคราะห์หาความลำเอียงของข้อสอบ 4 วิธี คือ วิธีวิเคราะห์ความแปรปรวน วิธีแปลงค่าความยากง่ายของข้อสอบ ธีโค้งลักษณะข้อทดสอบที่มีพารามิเตอร์ 1 ตัว และวิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 3 ตัว เมื่อใช้วิเคราะห์หาความลำเอียงต่อเพศของข้อสอบที่ใช้สอบแข่งขันเพื่อบรรจุเป็นข้าราชการ 4 ฉบับ และใช้ผลวิจัยเพศชายและเพศหญิง ชนิดละ 1,170 คน ผู้วิจัยวิเคราะห์หาดัชนีความลำเอียงของข้อสอบแล้วหาสัมประสิทธิ์สหสัมพันธ์ระหว่างวิธีการวิเคราะห์ทั้ง 4 วิธี และเปรียบเทียบความแตกต่างของผลการคัดเลือกก่อนและหลังการศึกษาความลำเอียงของข้อสอบตามวิธีการคิดคะแนนรวมที่แตกต่างกัน 6 วิธี ในด้านจำนวนผู้ที่ได้รับการคัดเลือก สัดส่วนของชายต่อหญิงที่ได้รับการคัดเลือกและความเที่ยงของแบบสอบ ผลการวิจัยพบว่าวิธีวิเคราะห์หาความลำเอียงแต่ละวิธีพบข้อสอบที่มีความลำเอียงต่างกันโดยวิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 3 ตัว พบข้อสอบที่มีความลำเอียงจำนวนมากที่สุดและค่าสัมประสิทธิ์สหสัมพันธ์ของดัชนีที่บ่งบอกความลำเอียงของข้อทดสอบทั้ง 4 วิธีสูงมากคือมีค่า r_{xy} ระหว่าง 0.754 - 0.992 สำหรับการใช้คะแนนดิบและคะแนนรวมแบบต่างกัน 5 วิธี มีจำนวนผู้ได้รับการคัดเลือกแตกต่างกันประมาณร้อยละ 4-24 ส่วนการให้คะแนนมาตรฐานทีปกติรวมกับคะแนนแปลงแบบอื่นๆ 4 วิธี มีผู้ได้รับการคัดเลือกแตกต่างกันร้อยละ 4-23 และเมื่อตัดข้อสอบที่มีความลำเอียงออกแล้ว พบว่าสัดส่วนหญิงและชายที่ได้รับการคัดเลือกมีความใกล้เคียงกันและค่าความเที่ยงของแบบสอบลดลงเล็กน้อย

กาญจนา วัธนสุนธร (2537) ได้พัฒนาเกณฑ์ตัดสินข้อสอบลำเอียงทางเพศด้วยข้อมูลเชิงประจักษ์ด้วยวิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) และวิธีซิบเทสท์ (SIBTEST) โดยใช้ข้อมูลการตอบข้อสอบคัดเลือกศึกษาต่อในระดับอุดมศึกษา ปีการศึกษา 2535 ความยาวของแบบสอบวิชาคณิตศาสตร์ 20 ,30,40 ข้อ และวิชาภาษาอังกฤษ 50 ,60,70,80 ข้อ ผู้สอบขนาด 100 ,200,400,600, 800,1,000 คน ผลการวิจัยพบว่ามีความไม่คงที่ข้ามขนาดผู้สอบและความยาวแบบสอบ ความสอดคล้องในการตรวจข้อสอบลำเอียงภายในวิธีเดียวกันข้ามขนาดผู้สอบค่อนข้างต่ำแต่จะสูงขึ้นเมื่อขนาดผู้สอบ 600 คนขึ้นไป ส่วนการวิเคราะห์ความลำเอียงของข้อสอบที่มีต่อเพศพบว่าข้อสอบวิชาภาษาอังกฤษลำเอียงเข้าข้างผู้หญิงส่วนข้อสอบวิชาภาษาคณิตศาสตร์ลำเอียงเข้าข้างผู้ชาย

เกษรหว่างจิตร (2539) ได้วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทลแฮนส์เซล โดยกลุ่มอ้างอิงและกลุ่มเปรียบเทียบจำแนกตามเพศ ภูมิลำเนา ประสบการณ์ในการสอบและสังกัดของสถานศึกษา ข้อมูลที่ใช้เป็นผลการตอบข้อสอบวิชาภาษาไทยของผู้สอบจำนวน 506 คน และผลการตอบข้อสอบวิชาภาษาอังกฤษของผู้สอบจำนวน 501 คน ในส่วนที่เป็นข้อสอบแบบเลือกตอบของศูนย์ทดสอบทางการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ผลการวิจัยพบว่าข้อสอบที่ทำหน้าที่ต่างกันส่วนมากจะเป็นข้อสอบที่มีค่าอำนาจจำแนกค่อนข้างต่ำทั้งสองวิชา เมื่อพิจารณาค่าความยากพบว่าข้อสอบที่ทำหน้าที่ต่างกันส่วนมากเป็นข้อสอบที่ง่ายมากสำหรับวิชาภาษาไทย ส่วนวิชาภาษาอังกฤษข้อสอบที่ทำหน้าที่ต่างกัน ส่วนมากเป็นข้อสอบที่ยากมาก อีกทั้งพบว่าส่วนมากเป็นข้อสอบที่ทำหน้าที่ต่างกันแบบ

อเนกรูป เมื่อจำแนกกลุ่มอ้างอิงและกลุ่มเปรียบเทียบตามเพศจะพบข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนมากที่สุด รองลงมาคือการจำแนกตามภูมิลำเนา สังกัดของสถานศึกษาและประสบการณ์ในการสอบ

จิติมา วรรณศรี (2539) เปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันระหว่างวิธีแมนเทล-แฮนส์เซล กับวิธีซิบเทสท์ โดยศึกษาจากข้อมูลจำลอง ปัจจัยที่ศึกษาได้แก่ความยาวแบบสอบ 3 ขนาด คือ 30, 60 และ 90 ข้อ ขนาดกลุ่มตัวอย่าง 3 ขนาดคือ 200, 600 และ 1000 คน โดยแต่ละขนาดมีอัตราส่วนระหว่างผู้สอบกลุ่มอ้างอิงกับกลุ่มเปรียบเทียบต่างกัน คือ 1:1, 1:0.9, 1:0.75 และ 1:0.5 ผลการวิจัยพบว่าวิธีแมนเทล -แฮนส์เซล กับวิธีซิบเทสท์มีประสิทธิภาพเท่าเทียมกันในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ทุกขนาดกลุ่มตัวอย่างและทุกอัตราส่วนภายใต้ความยาวแบบสอบเดียวกัน และเมื่อใช้แบบสอบที่มีความยาวปานกลาง (60 ข้อ) ทั้งสองวิธีสามารถตรวจสอบได้อย่างมีประสิทธิภาพที่สุด นอกจากนี้เมื่อใช้ขนาดกลุ่มตัวอย่างมากขึ้นจะสามารถตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันได้ถูกต้องมากขึ้น ส่วนมากวิธีซิบเทสท์มีอัตราความคลาด เคลื่อนประเภทที่ 1 มากกว่าวิธีแมนเทล-แฮนส์เซลเล็กน้อย

เรวดี อินทะสะระ (2539) ได้ศึกษาความเที่ยงตรงเชิงพยากรณ์ของแบบสอบคัดเลือกที่วิเคราะห์ความลำเอียงต่อเพศด้วยวิธีทฤษฎีการตอบสนองข้อสอบ วิธีแมนเทล-แฮนส์เซล และวิธีซิบเทสท์ พร้อมทั้งศึกษาการตัดสินผลการสอบที่คิดคะแนนมาตรฐานทีปกติและคะแนนน้ำหนักความ สามารถและสาเหตุของความลำเอียงของข้อสอบโดยศึกษาความลำเอียงของข้อสอบคัดเลือกเข้าศึกษาในชั้นปีที่ 1 รับตรงปีการศึกษา 2538 ของมหาวิทยาลัยสงขลานครินทร์ในวิชาเอกภาษาไทย ก วิชาสังคมศึกษา ก และวิชาภาษาอังกฤษ กข วิชาละ 8,127 คน (ชาย 2,722 คน หญิง 5,405 คน) วิชาภาษาไทย กข วิชาสังคมศึกษา กข และวิชาภาษาอังกฤษ กขค วิชาละ 5,415 คน (ชาย 1,451 คน หญิง 3,961 คน) ผลการวิจัยพบว่าวิธีการตรวจสอบความลำเอียงทั้ง 3 วิธีตัดสินจำนวนข้อสอบที่ลำเอียงแตกต่างกันในวิชาภาษาไทย ก ฉบับที่ 2 และวิชาสังคมศึกษา ก ฉบับที่ 1 ที่ระดับนัยสำคัญทางสถิติ ที่.05 นอกนั้นแตกต่างกันที่ระดับนัยสำคัญ.01 โดยวิธีทฤษฎีการตอบสนองข้อสอบตัดสินจำนวนข้อสอบ ที่ลำเอียงได้มากที่สุด ความสัมพันธ์ของลำดับที่ของการสอบไม่ว่าจะคิด คะแนนมาตรฐานปกติทีหรือคิดคะแนนน้ำหนักความสามารถและใช้ข้อสอบทั้งหมดหรือใช้เฉพาะข้อสอบที่ปราศจากความลำเอียงต่างมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

ญาณภัทร สีหะมงคล (2540) เปรียบเทียบความสอดคล้องของผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันระหว่างวิธี Lord's χ^2 วิธี Raju's Area Measures และวิธี Closed Interval Area เมื่อขนาดของกลุ่มตัวอย่าง ความยาวของแบบสอบและสัดส่วนจำนวนข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบต่างกัน ข้อมูลที่ใช้ในการศึกษาเป็นผลการสอบประเมินคุณภาพและความก้าวหน้าทางการศึกษาวิชาคณิตศาสตร์ของนักเรียนชั้นประถมศึกษาปีที่ 4 ปีการศึกษา 2536 สังกัดสำนักงาน การประถมศึกษาแห่งชาติจำนวน11,404 คน เครื่องมือที่ใช้เป็นแบบสอบแบบเลือกตอบ จำนวน80 ข้อ ผลการวิจัย พบว่าจำนวน

ข้อสอบที่ทำหน้าที่ต่างกันจากการตรวจสอบด้วยวิธีการทั้งสามวิธี แตกต่างกันเมื่อขนาดของกลุ่มตัวอย่างและความยาวของแบบสอบต่างกัน ส่วนความสัมพันธ์ระหว่างวิธีการทั้งสามมีค่าสัมประสิทธิ์สหสัมพันธ์ค่อนข้างสูงมากและมีนัยสำคัญทางสถิติเกือบทุกเงื่อนไขของการศึกษาและสำหรับความสอดคล้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบส่วนมากจะมีค่าปานกลางถึงต่ำเกือบทุกเงื่อนไขของการศึกษา

พรรณี จิตมาศ (2540) ได้วิเคราะห์ความลำเอียงต่อเพศของแบบสอบคณิตศาสตร์โจทย์ปัญหาที่ผู้วิจัยสร้างขึ้นด้วยวิธีวิเคราะห์ 3 วิธี คือวิธีแปลงค่าความยาก วิธี แมนเทล -แฮนส์เซล และวิธีซิบเทสท์แต่ละขนาดของกลุ่มผู้สอบ 500 และ 1,000 คนโดยเปรียบเทียบจำนวนข้อที่มีความลำเอียงและเปรียบเทียบค่าความเชื่อมั่นแบบครึ่งฉบับของแบบสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2539 ของโรงเรียนสังกัดกรมสามัญศึกษาส่วนกลาง จำนวน 2,200 คน ซึ่งเลือกมาโดยการสุ่มแบบแบ่งชั้น มีขนาดของโรงเรียนเป็นชั้นและโรงเรียนเป็นหน่วยการสุ่ม เครื่องมือที่ใช้คือแบบสอบคณิตศาสตร์โจทย์ปัญหาเรื่องสมการอัตราส่วนและร้อยละผู้วิจัยสร้างขึ้นเองเป็นแบบเลือกตอบชนิด 5 ตัวเลือก 40 ข้อ ผลการวิจัยพบว่ากลุ่มตัวอย่างขนาด 500 คน วิธีซิบเทสท์พบข้อสอบที่มีความลำเอียงมากที่สุดและ วิธีแปลงค่าความยากพบข้อสอบที่มีความลำเอียงน้อยที่สุด โดยจำนวนข้อสอบที่มีความลำเอียงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติทุกวิธีวิเคราะห์และเมื่อวิเคราะห์จากขนาด 1,000 คน วิธีแมนเทล -แฮนส์เซล พบข้อสอบมีความลำเอียงมากที่สุด วิธีแปลงค่าความยากไม่พบข้อสอบที่ลำเอียง โดยจำนวนข้อสอบที่ลำเอียงจากการวิเคราะห์ด้วยวิธีแปลงค่าความยากกับวิธีแมนเทล-แฮนส์เซลและวิธีแปลงค่าความยากกับ วิธีซิบเทสท์แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 นอกนั้นมีค่าไม่แตกต่างกัน

รัชรินทร์ มุคดา (2540) ศึกษาเปรียบเทียบประสิทธิภาพระหว่างวิธีแมนเทล -แฮนส์เซล กับวิธีถดถอยโลจิสติก ในการตรวจสอบการทำหน้าที่ต่างกันของแบบสอบอเนกรูปในกรณีที่จัดกลุ่มความสามารถ ค่าความยากของข้อสอบ ค่าอำนาจจำแนกของข้อสอบต่างกัน โดยศึกษาจากข้อมูลที่จำลองขึ้นด้วยโปรแกรม IRTDATA และเงื่อนไขที่ศึกษา คือกลุ่มความสามารถผู้สอบ 3 ระดับ ได้แก่ กลุ่มผู้สอบที่มีความสามารถสูง ปานกลางและต่ำ ค่าความยากของข้อสอบ 3 ระดับ ได้แก่ กลุ่มข้อสอบที่มีความยากสูง ปานกลาง และต่ำ ค่าอำนาจจำแนกของข้อสอบ 3 ระดับ ได้แก่ กลุ่มข้อสอบที่มีอำนาจจำแนกสูง ปานกลางและต่ำ ผลการวิจัยพบว่า วิธีแมนเทล -แฮนส์เซลกับวิธีถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันของแบบสอบอเนกรูปเท่ากันในทุกกลุ่มผู้สอบและในกลุ่มผู้สอบที่มีความสามารถสูงข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันแบบอเนกรูปมากที่สุดเป็นข้อสอบที่มีความยากสูงอำนาจจำแนกสูง ในกลุ่มผู้สอบที่มีความสามารถปานกลางเป็นข้อสอบที่มีค่าความยากปานกลางค่าอำนาจจำแนกสูง กลุ่มผู้สอบที่มีความสามารถต่ำเป็นข้อสอบที่มีความยากต่ำค่าอำนาจจำแนกสูง

เสรี ชัดแช้ม (2540) ศึกษาเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป ระหว่างวิธีแมนเทล-แฮนส์เซลแบบปกติกับวิธีแมนเทล-แฮนส์เซลแบบแบ่ง กลุ่มความสามารถของผู้สอบและความยากของข้อสอบใช้วิธี IRT เป็นเกณฑ์ ศึกษาจากข้อมูลผลการตอบแบบสอบวัดความสามารถในการอ่านภาษาไทยของนักเรียนชั้นมัธยมศึกษาปีที่ 1 สังกัดกรมสามัญศึกษาจังหวัดชลบุรี จำนวน 1,200 คน กลุ่มผู้สอบจำแนกตามเพศ ผลการวิจัยพบว่าวิธีแมนเทล -แฮนส์เซลแบบแบ่งกลุ่มความสามารถของผู้สอบและความยากของข้อสอบสามารถตรวจพบข้อสอบทำหน้าที่ต่างกันแบบอเนกรูปได้สอดคล้องกับวิธี IRT และตรวจพบข้อสอบทำหน้าที่ต่างกันมากกว่าวิธีแมนเทล -แฮนส์เซลแบบปกติ ข้อสอบที่ตรวจพบส่วนใหญ่เป็นข้อสอบยากปานกลางและข้อสอบที่มีค่าความยากซึ่งมีลักษณะข้อสอบของกลุ่มผู้สอบสองกลุ่มตัดกันบริเวณใกล้ๆ จุดกลางของช่วงความสามารถ

นพมาศ พิพัฒนสุข (2541) ได้ศึกษาเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีแมนเทล-แฮนส์เซล กับวิธีถดถอยโลจิสติก ในแบบสอบพหุมิติ เมื่อใช้เกณฑ์จับคู่เปรียบเทียบแตกต่างกัน 3 เกณฑ์ ได้แก่ คะแนนรวมคะแนนแบบสอบย่อยและคะแนนหลายแบบสอบย่อย โดยเก็บข้อมูลจากแบบสอบวัดความสามารถทางคณิตศาสตร์ที่ผู้วิจัยสร้างขึ้น ผลการวิจัยพบว่าวิธีแมนเทล-แฮนส์เซลมีประสิทธิภาพมากกว่าวิธีถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบชนิดพหุมิติเมื่อใช้เกณฑ์จับคู่คะแนนรวมและมีประสิทธิภาพไม่แตกต่างกันเมื่อใช้เกณฑ์จับคู่คะแนนแบบสอบย่อยและวิธีถดถอยโลจิสติกเมื่อใช้เกณฑ์การจับคู่เปรียบเทียบคะแนนหลายแบบสอบย่อยมีความเหมาะสมในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบชนิดพหุมิติ

นิคม กีรติวากูร (2542) ได้เปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีการวิเคราะห์องค์ประกอบจำกัด (RFA) วิธีแมนเทล-แฮนส์เซล และวิธีการตอบสนองข้อสอบ แบบ 2 พารามิเตอร์โดยเทียบกับเกณฑ์ที่กำหนด ศึกษาจากข้อมูลจำลอง ปัจจัยที่ศึกษาได้แก่ ขนาดกลุ่มตัวอย่าง 2 ขนาด คือ ขนาดเล็ก (300 คน) และขนาด ใหญ่ (1000 คน) ค่าความยาวแบบสอบ 2 ขนาด คือ แบบสอบสั้น (25 ข้อ) และแบบสอบยาว (75 ข้อ) ค่าความยากของข้อสอบแบ่งออกเป็น 3 ระดับคือ กลุ่มข้อสอบที่มีความยากสูง ปานกลางและต่ำ ขนาดความลำเอียงของข้อสอบแบ่งออกเป็น 2 ขนาดคือ กลุ่มข้อสอบที่มีความลำเอียงสูงและต่ำ ผลการวิจัยพบว่า วิธี RFA มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูงที่สุดโดยวิธีแมนเทล-แฮนส์เซลมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูงภายใต้เงื่อนไข ข้อสอบที่มีความยากต่ำอำนาจจำแนกสูง วิธี IRT แบบ 2 พารามิเตอร์ มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูงภายใต้เงื่อนไขข้อสอบที่มีความยากต่ำ วิธี IRT มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีแมนเทล-แฮนส์เซล และวิธี RFA

อารี วัชรโสตถิกุล (2543) ได้ศึกษาเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้รูปแบบต่างกัน คือ รูปแบบคะแนนรวมทั้งฉบับแยกตามเนื้อหา และแยกตามระดับพฤติกรรมด้วย

วิธีการตรวจสอบต่างกัน คือ วิธีซิบเทสท์และวิธีถดถอยโลจิสติก แล้วทำการคัดเลือกข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบสอบเพื่อเปรียบเทียบค่าความเชื่อมั่น ผลการวิจัยพบว่า จำนวนข้อสอบที่ทำหน้าที่ต่างกันโดยใช้วิธีการตรวจสอบต่างกัน ซึ่งแตกต่างกันในรูปแบบรวมทั้งฉบับ ส่วนรูปแบบแยกตามเนื้อหาและแยกตามระดับพฤติกรรมไม่แตกต่างกัน

ทองอยู่สาระ (2543) ได้ศึกษาเปรียบเทียบอำนาจการตรวจสอบและจำแนกผิดพลาดในการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบสม่ำเสมอและแบบไม่สม่ำเสมอ ระหว่างวิธีแมนเทล -แฮนส์เซลและวิธีถดถอยโลจิสติก โดยใช้ความยาวแบบสอบและขนาดกลุ่มตัวอย่างต่างกัน ผลการวิจัยพบว่า อำนาจการตรวจสอบและการจำแนกผิดพลาดในการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันทั้งแบบสม่ำเสมอและแบบไม่สม่ำเสมอ ระหว่างวิธีแมนเทล-แฮนส์เซลและวิธีถดถอยโลจิสติกภายใต้ความยาวแบบสอบและขนาดกลุ่มตัวอย่างสูงปานกลางและต่ำที่ศึกษาเกือบทุกเงื่อนไขไม่แตกต่างกัน ส่วนความยาวของแบบสอบไม่มีผลต่ออำนาจการตรวจสอบและการจำแนก ผิดพลาดในทั้ง2 วิธี แต่ขนาดของกลุ่มตัวอย่างมีผลต่ออำนาจการตรวจสอบด้วยวิธีแมนเทลแฮนส์เซล และวิธีถดถอยโลจิสติกเกือบทุกเงื่อนไข เมื่อขนาดกลุ่มตัวอย่างเพิ่มขึ้นอำนาจการตรวจสอบจะเพิ่มขึ้น แต่ขนาดของกลุ่มตัวอย่างไม่มีผลต่อการจำแนกผิดพลาดในเกือบทุกเงื่อนไข

วลีมาศ แซ่อึ้ง (2543) เปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีถดถอยโลจิสติก ข้อมูลที่ใช้ในการศึกษาจำลองภายใต้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ชนิดกำหนดค่าการเดา(c)คงที่แล้วจัดกระทำข้อมูลตามปัจจัย 4 ตัว คือ (1) ลักษณะของข้อสอบที่มีค่าความยาก (b) และอำนาจจำแนก(a) ระดับต่ำ ปานกลาง และสูง จำนวน 9 ลักษณะ (2) ความยาวของแบบสอบ 2 ระดับ (3) สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบ 3 ระดับและ (4) ขนาดกลุ่มตัวอย่าง6 ระดับ รวมข้อมูลที่ศึกษาทั้งหมดจำนวน324 เงื่อนไข แล้วนำข้อมูลของแต่ละเงื่อนไขมาคำนวณค่าอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่1ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป ผลการวิจัยพบว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปของวิธีซิปเทสท์ปรับใหม่และวิธีถดถอยโลจิสติกมีค่าเท่าเทียมกันภายใต้เกือบทุกเงื่อนไขและทั้งสองวิธีดังกล่าวมีอำนาจการทดสอบสูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลภายใต้เกือบทุกเงื่อนไข อัตราความคลาดเคลื่อน ประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปของวิธี ซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซลและวิธีถดถอยโลจิสติก มีค่าอยู่ภายในเกณฑ์ของอัตราความคลาดเคลื่อนประเภทที่ 1 ที่ระดับ 10% ภายใต้เกือบทุกเงื่อนไข

รักชนก ยี่สุ่นศรี (2544) ทำการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบสำหรับกลุ่มผู้สอบเมื่อจำแนกตามเพศและสถานที่ตั้งทางภูมิศาสตร์และโรงเรียนที่จบการศึกษาเพื่อเปรียบเทียบความเที่ยง ความตรงและฟังก์ชันสารสนเทศของแบบสอบระหว่างแบบสอบฉบับก่อนและหลังตัดข้อสอบ

ที่ DIF ข้อมูลจากการตอบแบบสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษา วิชาภาษาอังกฤษและคณิตศาสตร์ ปีการศึกษา 2543 ครั้งที่ 1 และเลือกศึกษาใน ส่วนที่เป็นข้อสอบแบบหลายตัวเลือกจากกลุ่มตัวอย่างที่เป็นผู้เข้าสอบจำนวน4,000 คน และ3,600 คน ตามลำดับ ผลการวิจัย พบว่าแบบสอบวิชาภาษาอังกฤษทำหน้าที่ต่างกันตามเพศและสถานที่ตั้งตามภูมิศาสตร์ ส่วนแบบสอบวิชาคณิตศาสตร์ทำหน้าที่ต่างกันตามเพศของผู้สอบ และข้อสอบทำหน้าที่ต่างกันตามเพศของผู้สอบมากที่สุดทั้งสองวิชา นอกจากนี้เมื่อเปรียบเทียบคุณภาพของแบบสอบฉบับก่อนและหลังการตัดข้อสอบที่พบว่าทำหน้าที่ต่างกัน พบว่าในด้านความตรงจะไม่ต่างกัน แต่แบบสอบฉบับหลังจากตัดข้อสอบที่พบว่าทำหน้าที่ต่างกัน ส่วนใหญ่จะมีค่าความเที่ยงลดลงและมีค่าฟังก์ชันสารสนเทศมากขึ้น ส่วนความสัมพันธ์ระหว่างตำแหน่งของคะแนนรวมของผู้สอบก่อนและหลังจากการตัดข้อสอบที่พบว่าทำหน้าที่ต่างกันทั้งในกรณีที่ตัดทุกข้อและตัดในบางข้อพบว่าทุกกรณีมีความสัมพันธ์ในทางบวกซึ่งกันและกันอย่างมีนัยสำคัญ

สิริรัตน์ วิภาสศิลป์ (2545) ศึกษาการเปรียบเทียบวิธีซิบเทสท์ และวิธีดีเอฟไอที (DFIT) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ หมวดข้อสอบและแบบสอบจากข้อมูลการตอบข้อสอบที่ใช้ความสามารถหลายมิติ เครื่องมือคือแบบวัดผลสัมฤทธิ์ทางคณิตศาสตร์ เงื่อนไขความยาว 30 , 40 และ 50 ข้อ กลุ่มตัวอย่างเป็นกลุ่มอย่างเทียม โดยสุ่มกลุ่มตัวอย่างแบบใส่คืนมีขนาด 50, 100, 200, 500 และ 1,000 คน วิเคราะห์ข้อมูลด้วยวิธีซิบเทสท์และวิธีดีเอฟไอที ผลการวิจัยพบว่า 1) เมื่อแบบสอบประกอบด้วยข้อสอบ 30, 40 และ 50 ข้อ กลุ่มตัวอย่างมีขนาด 50 , 100 และ 200 คน ส่งผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีซิบเทสท์ ไม่แตกต่างกัน 2) กลุ่มตัวอย่างขนาด 500 และ 1,000 คน ส่งผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีซิบเทสท์ สูงกว่ากลุ่มตัวอย่างขนาด 50, 100 และ 200 คน 3) เมื่อแบบสอบที่ประกอบด้วยข้อสอบ 30, 40 และ 50 ข้อ กลุ่มตัวอย่างขนาด 50 , 100 , 200, 500 และ 1,000 คน ส่งผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีดีเอฟไอที ไม่แตกต่างกัน และ 4) การตรวจสอบการทำหน้าที่ต่างกันของหมวดข้อสอบ พบว่าเมื่อแบบสอบที่ประกอบด้วยข้อสอบ 30 ข้อ วิธีซิบเทสท์มีความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของหมวดข้อสอบมากกว่าวิธีดีเอฟไอที เมื่อกลุ่มตัวอย่างมีขนาด 1,000 คน เมื่อแบบสอบประกอบด้วยข้อสอบ 40 ข้อ วิธีซิบเทสท์ มีความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของหมวดข้อสอบมากกว่าวิธีดีเอฟไอที เมื่อกลุ่มตัวอย่างมีขนาด 500 คน เมื่อแบบสอบประกอบด้วยข้อสอบ 50 ข้อ ไม่พบความแตกต่างระหว่างวิธีการทั้งสอง

สุมาลี แกล้วทนงค์ (2547) ศึกษาสาเหตุของการทำหน้าที่ต่างกันของข้อสอบสาระการเรียนรู้ภาษาไทยและสาระการเรียนรู้สังคมศึกษา ศาสนาและวัฒนธรรมต่างกัน กลุ่มตัวอย่างคือนักเรียนชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2546 สำนักงานเขตพื้นที่การศึกษาสงขลา พัทลุง ตรังและสตูล จำนวน 1,320 คน วิเคราะห์ข้อมูลด้วยวิธีแมนเทล-แฮนส์เซล และวิธีซิบเทสท์ ผลการวิจัยพบว่าแบบสอบกลุ่มสาระการเรียนรู้ภาษาไทย 3 ฉบับ 120 ข้อ ข้อสอบที่ทำหน้าที่ต่างกันด้านเพศ 9 ข้อ ด้านภาษาพูด 15 ข้อ

ด้านเชื้อชาติ 28 ข้อ และกลุ่มสาระการเรียบรู้สังคมศึกษา ศาสนาและวัฒนธรรม มีข้อสอบที่ทำหน้าที่ต่างกันด้านเพศ 22 ข้อ ด้านภาษาพูด 52 ข้อ และด้านเชื้อชาติ 20 ข้อ โดยข้อสอบที่ทำหน้าที่ต่างกันด้านเพศ ส่วนใหญ่มีสาเหตุมาจากเนื้อหาและภาษาที่ใช้ในแบบสอบ ข้อสอบที่ทำหน้าที่ต่างกันด้านภาษาพูดส่วนใหญ่มีสาเหตุมาจากการใช้คำศัพท์เฉพาะและบริบททางวัฒนธรรมและข้อสอบที่ทำหน้าที่ต่างกันด้านเชื้อชาติ ส่วนใหญ่มีสาเหตุมาจากบริบททางภาษาและบริบททางวัฒนธรรม

ปิยะทิพย์ ตินวร (2549) ได้เปรียบเทียบประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบพหุมิติ ระหว่างวิธีการวิเคราะห์องค์ประกอบจำกัดกับวิธีถดถอยโลจิสติก จำนวน 18 เงื่อนไข คือ ขนาดกลุ่มตัวอย่าง 3 ขนาด (2000, 1000 และ 300 คน) ความยาวของแบบสอบ 3 ขนาด (40, 30 และ 20 ข้อ) และเกณฑ์การจับคู่ 2 เกณฑ์ (คะแนนรวมทั้งฉบับและคะแนนแบบสอบย่อย) เมื่อใช้ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยวิธีซิบเทสท์ เป็นเกณฑ์สำหรับการเปรียบเทียบประสิทธิภาพ กลุ่มตัวอย่างเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ปีการศึกษา 2546 ที่เข้าสอบวัดผลสัมฤทธิ์ทางการเรียนระดับชาติ วิชาภาษาไทย จำนวน 2,000 คน ผลการวิจัยพบว่า วิธีถดถอยโลจิสติกมีประสิทธิภาพไม่แตกต่างกันกับวิธีวิเคราะห์องค์ประกอบจำกัด ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบพหุมิติจาก 2 เงื่อนไข และวิธีถดถอยโลจิสติกมีประสิทธิภาพมากกว่าวิธีวิเคราะห์องค์ประกอบจำกัดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบพหุมิติ 16 เงื่อนไข

อุทัยวรรณ สายพัฒนะ (2547) ศึกษาเปรียบเทียบประสิทธิภาพของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบที่มีการให้คะแนนแบบหลายค่า ในเงื่อนไขความยาวของแบบสอบและกลุ่มตัวอย่างที่แตกต่างกันระหว่างวิธี GMH กับวิธี Polytomous SIBTEST ผลการวิจัย 1) วิธี GMH เมื่อแบบสอบประกอบด้วยข้อสอบ 40 30 และ 20 ข้อ กลุ่มตัวอย่างขนาด 1,000, 500 และ 250 คน ส่งผลต่อความถูกต้องและความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแตกต่างกัน ยกเว้น กรณีแบบสอบประกอบด้วยข้อสอบ 30 ข้อและ 20 ข้อ กลุ่มตัวอย่างขนาด 500 และ 250 คน ส่งผลต่อความผิดพลาดในการตรวจสอบ DIF ไม่แตกต่างกัน 2) วิธี Polytomous SIBTEST เมื่อแบบสอบประกอบด้วยข้อสอบ 40, 30 และ 20 ข้อ กลุ่มตัวอย่างขนาด 1,000, 500 และ 250 คน ส่งผลต่อความถูกต้องและความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยแตกต่างกัน ยกเว้นกรณีแบบสอบประกอบด้วยข้อสอบ 20 ข้อ กลุ่มตัวอย่างขนาด 1,000 และ 500 คนส่งผลต่อความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบไม่แตกต่างกัน และเมื่อพิจารณาการเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี GMH และวิธี Polytomous SIBTEST ในทุกเงื่อนไขความยาวของแบบสอบ กลุ่มตัวอย่างทุกขนาดส่งผลต่อประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของวิธีการทั้งสองไม่แตกต่างกัน ยกเว้นกรณีกลุ่มตัวอย่างขนาด 1,000 คน

ของแบบสอบขนาด 20 ข้อ วิธี Polytomous SIBTEST มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูงกว่าวิธี GMH

อรินทร์ น่วมถนอม (2549) ศึกษาการเปรียบเทียบวิธีโพลีซิบเทสท์ (Poly-SIBTEST) วิธีถดถอยโลจิสติกแบบจัดอันดับ (Ordinal Logistic regression) และวิธีถดถอยโลจิสติกแบบจัดอันดับหลายมิติ (Multidimensional Ordinal Logistic regression) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า โดยการจำลองข้อมูลภายใต้โมเดลพาเชียลเครดิตทั่วไปแบบหลายมิติ จำลองผลการตอบจากแบบสอบที่วัดความสามารถ 2 มิติจำนวน 40 ข้อ ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่ต่างกัน 2 รูปแบบ สัดส่วนของข้อสอบทำหน้าที่ต่างกัน 3 ขนาด ความแตกต่างของการแจกแจงความสามารถ 3 ระดับ และขนาดตัวอย่าง 4 ขนาด รวมข้อมูลที่ใช้ศึกษา 72 เงื่อนไข วิเคราะห์ข้อมูลในแต่ละเงื่อนไขด้วยวิธีโพลีซิบเทสท์ วิธีถดถอยโลจิสติกแบบจัดอันดับและวิธีถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ผลการวิจัยพบว่า วิธีถดถอยโลจิสติกแบบจัดอันดับและวิธีถดถอยโลจิสติกแบบจัดอันดับหลายมิติ มีอัตราความถูกต้องใกล้เคียงกัน การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและแบบอเนกรูปทั้งสองวิธีมีอัตราความถูกต้องสูงกว่าวิธีโพลีซิบเทสท์ ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป สัดส่วนของข้อสอบทำหน้าที่ต่างกันในแบบสอบไม่มีผลต่อวิธีโพลีซิบเทสท์ วิธีถดถอยโลจิสติกแบบจัดอันดับหลายมิติแต่มีผลต่อวิธีถดถอยโลจิสติกแบบจัดอันดับ เมื่อความแตกต่างของการแจกแจงความสามารถเพิ่มขึ้น วิธีโพลีซิบเทสท์สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ที่สูงเกินปกติได้ดีกว่าวิธีอื่นและเมื่อขนาดตัวอย่างเพิ่มขึ้นมีผลทำให้ทุกวิธีมีอัตราความถูกต้องเพิ่มขึ้นเกือบทุกเงื่อนไขที่ศึกษา

สุทธิพร ศุรธรณี (2550) ศึกษาความสามารถในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบตามวิธีตัวแบบวางนัยทั่วไประดับลดหลั่น กำหนดให้ผลตอบสนองข้อสอบแต่ละข้อเป็นแบบทวิภาคซึ่งได้ขยายตัวแบบของคามาตะ โดยศึกษาความแกร่งของตัวแบบคามาตะ ในการตรวจสอบ DIF ในกรณีที่ข้อสมมติเกี่ยวกับการแจกแจงของผู้สอบไม่เป็นปกติ โดยกำหนดให้ความสามารถของ ผู้สอบมีการแจกแจงแบบฟิชเชอร์ทิปเปทท์ ที่ค่าความสามารถมีค่าได้ตั้งแต่ $-\infty$ ถึง ∞ แต่ความเบ้เท่ากับ 1.14 ซึ่งสุทธิพรตัดเฉพาะค่าที่อยู่ระหว่าง -3 ถึง 3 เท่านั้นเพื่อที่ให้ช่วงความเชื่อมั่นเหมือนการแจกแจงแบบปกติมาตรฐาน การศึกษาจะใช้การจำลองข้อมูลด้วยเทคนิคมอนติคาล์โดยใช้โปรแกรม R เวอร์ชัน 2.01 จำลองจำนวนผู้สอบ 1,000 คน เป็นกลุ่มอ้างอิง 500 คน และกลุ่มเปรียบเทียบ 500 คน แล้วตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยวิธีการ HGLM โดยใช้โปรแกรมสำเร็จรูป HLM เวอร์ชัน 6.0 ซึ่งแต่ละกรณีทำซ้ำจำนวน 100 ครั้ง แต่ละครั้งจะคำนวณค่าประมาณของ DIF ค่า RMSE กำลังทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ผลการวิจัยพบว่า กรณีที่ความสามารถของผู้สอบทั้งสองกลุ่มมีการแจกแจงแบบปกติ พบความแตกต่างระหว่างค่าเฉลี่ยของความสามารถกลุ่มอ้างอิงและกลุ่มเปรียบเทียบไม่มีผล

ต่อค่า RMSE ค่าเฉลี่ยประมาณของ DIF กำลังการทดสอบและอัตราความผิดพลาดประเภทที่ 1 วิธีการตรวจสอบที่ละข้อจะขึ้นอยู่กับสัดส่วนของข้อสอบที่มี DIF โดยที่ค่า RMSE จะมีค่ามากขึ้นเมื่อสัดส่วนของผู้สอบที่มี DIF สูงขึ้น แต่ในทางกลับกันสัดส่วนของข้อสอบที่มี DIF ไม่มีผลต่อวิธีการตรวจสอบทุกข้อพร้อมกันและวิธีการตรวจสอบทีละข้อยังมีค่า RMSE น้อยกว่าวิธีการตรวจสอบทุกข้อพร้อมกันกรณีที่มี DIF จำนวน 1 และ 3 ข้อ แต่กรณีที่มีข้อสอบมี DIF จำนวน 6 ข้อวิธีการตรวจสอบที่ละข้อยังมีค่า RMSE มากกว่าวิธีการตรวจสอบทุกข้อพร้อมกันและค่าเฉลี่ยประมาณของ DIF จะขึ้นอยู่กับวิธีการที่ใช้ในการตรวจสอบมี DIF โดยที่วิธีการตรวจสอบทีละข้อให้ค่าเฉลี่ยต่ำกว่าค่าจริง ขณะที่วิธีการตรวจสอบทุกข้อพร้อมกันให้ค่าเฉลี่ยทั้งต่ำกว่าและสูงกว่าค่าจริง ส่วนกำลังการทดสอบและอัตราส่วนของข้อสอบที่มี DIF ซึ่งวิธีตรวจสอบทีละข้อมีกำลังการทดสอบสูงกว่าวิธีการตรวจสอบทุกข้อพร้อมกัน กรณีความสามารถของผู้สอบทั้งสองกลุ่มที่การแจกแจงแบบฟิชเชอร์ทิปเปทท์ พบความแตกต่างระหว่างค่าเฉลี่ยของค่าประมาณของ DIF และกำลังการทดสอบแต่มีผลต่อค่า RMSE และอัตราความผิดพลาดประเภทที่ 1 โดยที่วิธีการรวจสอบทีละข้อมีค่าเฉลี่ยของค่าประมาณ DIF ต่ำกว่าค่าจริงและมีค่ากำลังการทดสอบสูงกว่าวิธีการตรวจสอบทุกข้อพร้อมกัน เมื่อเปรียบเทียบทั้งกรณีที่ความสามารถของผู้สอบมีการแจกแจงแบบปกติและการแจกแจงแบบฟิชเชอร์ทิปเปทท์ พบว่ากำลังการทดสอบกรณีที่ความสามารถของผู้สอบมีการแจกแจงแบบปกติมีค่ามากกว่ากรณีที่ความสามารถของผู้สอบมีการแจกแจงแบบฟิชเชอร์ทิปเปทท์ในกรณีที่ข้อสอบมี DIF จำนวน 6 ข้อ สำหรับวิธีการตรวจสอบทีละข้อมีค่า RMSE มากกว่าวิธีการตรวจสอบทุกข้อพร้อมกันในกรณีที่มีความแตกต่างระหว่างค่าเฉลี่ยของความสามารถของผู้สอบทั้งกลุ่มและกรณีที่มีข้อสอบมี DIF จำนวน 6 ข้อ

ตารางที่ 2.7 สรุปงานวิจัยที่ได้ศึกษาเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบในประเทศไทยตั้งแต่อดีตถึงปัจจุบัน

พ.ศ.	ผู้วิจัย	ประเด็นที่ศึกษา	องค์ความรู้ที่ได้รับ
2531	สุรศักดิ์ อมรรัตนศักดิ์	เปรียบเทียบผลของวิธีวิเคราะห์หาความลำเอียงของข้อสอบ 4 วิธี 1) วิธีวิเคราะห์ความแปรปรวน 2) วิธีแปลงค่าความยากง่ายของข้อสอบ 3) วิธีโค้งลักษณะของข้อทดสอบที่มีพารามิเตอร์ 1 ตัว 4) วิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 3 ตัว	วิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 3 ตัว พบข้อทดสอบที่มีความลำเอียงจำนวนมากที่สุดและยังพบว่าค่าสัมประสิทธิ์สหสัมพันธ์ของดัชนีที่บ่งบอกความลำเอียงของข้อทดสอบทั้ง 4 วิธีสูงมาก
2537	กาญจนา วัธนสุนธร	พัฒนาเกณฑ์ตัดสินข้อสอบลำเอียงทางเพศด้วยข้อมูลเชิงประจักษ์ด้วยวิธีแมนเทล-แฮนส์เซล และวิธีซิบเทสท์	มีความไม่คงที่ข้ามขนาดผู้สอบและความยาวแบบสอบ ความสอดคล้องในการตรวจข้อสอบลำเอียงภายในวิธีเดียวกันข้ามขนาดผู้สอบค่อนข้างต่ำแต่จะสูงขึ้นเมื่อขนาดผู้สอบ 600 คนขึ้นไป การวิเคราะห์ความลำเอียงของข้อสอบที่มีต่อเพศ พบว่าข้อสอบวิชาภาษาอังกฤษลำเอียงเข้าข้างผู้หญิง ส่วนข้อสอบวิชาภาษาคณิตศาสตร์ลำเอียงเข้าข้างผู้ชาย
2539	เกษร หว่างจิตร	วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซล โดยกลุ่มอ้างอิงและกลุ่มเปรียบเทียบจำแนกตามเพศ ภูมิลำเนา ประสบการณ์ในการสอบและสังกัดของสถานศึกษา	ข้อสอบที่ทำหน้าที่ต่างกันส่วนมากจะเป็นข้อสอบที่มีค่าอำนาจจำแนกค่อนข้างต่ำและเมื่อจำแนกตามเพศจะพบข้อสอบที่ทำหน้าที่ต่างกันมีจำนวนมากที่สุด
2539	จิติมา วรรณศรี	เปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันระหว่างวิธี 1) วิธีแมนเทล-แฮนส์เซล 2) วิธีซิบเทสท์	แบบสอบที่มีความยาวปานกลาง ทั้ง 2 วิธีสามารถตรวจสอบได้อย่างมีประสิทธิภาพที่สุดเมื่อใช้ขนาดกลุ่มตัวอย่างมากขึ้นจะสามารถตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันได้ถูกต้องมากขึ้น

ตารางที่ 2.7 (ต่อ)

พ.ศ.	ผู้วิจัย	ประเด็นที่ศึกษา	องค์ความรู้ที่ได้รับ
2539	เรวดี อินทะสะระ	ศึกษาความเที่ยงตรงเชิงพยากรณ์ของแบบสอบคัดเลือกที่วิเคราะห์ความลำเอียงต่อเพศด้วย 1) วิธี IRT 2) วิธีแมนเทล-แฮนส์เซล และ 3) วิธีซิบเทสท์	วิธีการตรวจสอบความลำเอียงทั้ง 3 วิธีตัดสินจำนวนข้อสอบที่ลำเอียงแตกต่างกัน โดยวิธี IRT ตัดสินจำนวนข้อสอบที่ลำเอียงได้มากที่สุด
2540	ญาณภัทร สีหะมงคล	เปรียบเทียบความสอดคล้องของผลการตรวจสอบขอสอบที่ทำหน้าที่ต่างกันระหว่าง 1) วิธี Lord's χ^2 2) วิธี Raju's Area Measures 3) วิธี Closed Interval Area	จำนวนข้อสอบที่ทำหน้าที่ต่างกันจากการตรวจสอบด้วยวิธีการทั้งสามวิธี แตกต่างกันเมื่อขนาดของกลุ่มตัวอย่างและความยาวของแบบสอบต่างกัน
2540	พรรณี จิตมาศ	วิเคราะห์ความลำเอียงต่อเพศของแบบสอบคณิตศาสตร์ ด้วย 3 วิธี คือ 1) วิธีแปลงค่าความยาก 2) วิธีแมนเทล-แฮนส์เซล 3) วิธีซิบเทสท์	วิเคราะห์จากกลุ่มตัวอย่างขนาด 500 คน วิธีซิบเทสท์ พบข้อสอบที่มีความลำเอียงมากที่สุดและเมื่อวิเคราะห์จากกลุ่มผู้สอบขนาด 1,000 คน วิธีแมนเทล-แฮนส์เซล พบข้อสอบมีความลำเอียงมากที่สุด
2540	รัชรินทร์ มุคดา	ศึกษาการเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของแบบสอบอเนกรูปในกรณีที่จัดกลุ่มความสามารถ ค่าความยากของข้อสอบ ค่าอำนาจจำแนกของข้อสอบต่างกัน ระหว่าง 1) วิธีแมนเทล-แฮนส์เซล 2) วิธีถดถอยโลจิสติก	วิธีแมนเทล-แฮนส์เซล กับวิธีถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันของแบบสอบอเนกรูปเท่ากันในทุกกลุ่มผู้สอบ
2540	เสรี ชัดแช้ม	ศึกษาเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธี 1) วิธีแมนเทล-แฮนส์เซลแบบปกติ 2) วิธีแมนเทล-แฮนส์เซลแบบแบ่งกลุ่มความสามารถของผู้สอบและความยากของข้อสอบ	วิธีแมนเทล-แฮนส์เซลแบบแบ่งกลุ่มความสามารถของผู้สอบและความยากของข้อสอบ สามารถตรวจพบข้อสอบทำหน้าที่ต่างกันแบบอเนกรูปได้สอดคล้องกับวิธี IRT และตรวจพบ DIF มากกว่าวิธีแมนเทล-แฮนส์เซล แบบปกติ

ตารางที่ 2.7 (ต่อ)

พ.ศ.	ผู้วิจัย	ประเด็นที่ศึกษา	องค์ความรู้ที่ได้รับ
2541	นพมาศ พิพัฒนสุข	เปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบพหุมิติ เมื่อใช้เกณฑ์จับคู่เปรียบเทียบแตกต่างระหว่าง 1) วิธีแมนเทล-แฮนส์เซล 2) วิธีถดถอยโลจิสติก	วิธีแมนเทล-แฮนส์เซลมีประสิทธิภาพมากกว่าวิธีถดถอยโลจิสติกในการตรวจสอบในแบบสอบชนิดพหุมิติเมื่อใช้เกณฑ์จับคู่คะแนนรวม และมีประสิทธิภาพไม่แตกต่างกันเมื่อใช้เกณฑ์จับคู่คะแนนแบบสอบย่อย
2542	นิคม กีรติวากูร	เปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี 1) วิธี RFA 2) วิธีแมนเทล-แฮนส์เซล 3) วิธี IRT แบบ 2 พารามิเตอร์	วิธี RFA มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูงที่สุดและวิธี IRT มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีแมนเทล-แฮนส์เซลและวิธี RFA ตามลำดับ
2543	อารี วัชรโสตถิกุล	เปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้รูปแบบและวิธีการตรวจสอบต่างกัน ระหว่าง 1) วิธีซิบเทสท์ 2) วิธีถดถอยโลจิสติก	จำนวนข้อสอบที่ทำหน้าที่ต่างกันโดยใช้วิธีการตรวจสอบต่างกันแตกต่างกันในรูปแบบรวมทั้งฉบับ ส่วนรูปแบบแยกตามเนื้อหาและแยกตามระดับพฤติกรรมไม่แตกต่างกัน
2543	ทองอยู่ สาระ	เปรียบเทียบอำนาจการตรวจสอบและจำแนกผิดพลาดในการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบสม่ำเสมอและแบบไม่สม่ำเสมอระหว่าง 1) วิธีแมนเทล-แฮนส์เซล 2) วิธีถดถอยโลจิสติก	ความยาวของแบบสอบไม่มีผลต่ออำนาจการตรวจสอบและการจำแนก ผิดพลาดในทั้ง 2 วิธี แต่เมื่อขนาดกลุ่มตัวอย่างเพิ่มขึ้นอำนาจการตรวจสอบจะเพิ่มขึ้น
2543	วลีมาศ แซ่อึ้ง	เปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป ระหว่าง 1) วิธีซิปเทสท์ปรับใหม่ 2) วิธีซิบเทสท์ 3) วิธีแมนเทล-แฮนส์เซล (4) วิธีถดถอยโลจิสติก	อำนาจการทดสอบในการตรวจสอบข้อสอบแบบอเนกรูปของวิธีซิปเทสท์ปรับใหม่และวิธีถดถอยโลจิสติกมีค่า เท่าเทียมกันภายใต้เกือบทุกเงื่อนไขและทั้งสองวิธีมีอำนาจการทดสอบสูงกว่าวิธีซิปเทสท์ และวิธี MH ภายใต้เกือบทุกเงื่อนไข ส่วนอัตรา

ตารางที่ 2.7 (ต่อ)

พ.ศ.	ผู้วิจัย	ประเด็นที่ศึกษา	องค์ความรู้ที่ได้รับ
			ความคลาดเคลื่อนประเภทที่ 1 ทั้ง 4 วิธี มีค่าอยู่ภายในเกณฑ์
2544	รักชนก ยี่สุ่นศรี	วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ สำหรับกลุ่มผู้สอบเมื่อจำแนกตามเพศและสถานที่ตั้งทางภูมิศาสตร์และโรงเรียนที่จบการศึกษา	แบบสอบวิชาภาษาอังกฤษทำหน้าที่ต่างกันตามเพศและสถานที่ตั้งตามภูมิศาสตร์ ส่วนแบบสอบวิชาคณิตศาสตร์ทำหน้าที่ต่างกันตามเพศของผู้สอบและข้อสอบทำหน้าที่ต่างกันตามเพศของผู้สอบมากที่สุดทั้งสองวิชา
2545	สิริรัตน์ วิภาสศิลป์	ศึกษาการเปรียบเทียบการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ หมวดข้อสอบและแบบสอบ จากข้อมูลการตอบข้อสอบที่ใช้ความสามารถหลายมิติ ระหว่าง 1) วิธีซิบเทสท์ 2) วิธีดีเอฟไอที (DFIT)	1)เมื่อแบบสอบประกอบด้วยข้อสอบ 30, 40 และ 50 ข้อ กลุ่มตัวอย่างมีขนาด 50, 100 และ 200 คน ส่งผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีซิบเทสท์ไม่แตกต่างกัน 2) กลุ่มตัวอย่างขนาด 500 และ 1 ,000 คน ส่งผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีซิบเทสท์ สูงกว่ากลุ่มตัวอย่างขนาด 50, 100 และ 200 คน 3) เมื่อแบบสอบที่ มีข้อสอบ 30 , 40 และ 50 ข้อ กลุ่มตัวอย่างขนาด 50 , 100 , 200, 500 และ 1,000 คน ส่งผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีดีเอฟไอที ไม่แตกต่างกัน 4) การตรวจสอบการทำหน้าที่ต่างกันของหมวดข้อสอบ พบว่าเมื่อแบบสอบที่ประกอบด้วยข้อสอบ 30 ข้อ วิธีซิบเทสท์ มี

ตารางที่ 2.7 (ต่อ)

พ.ศ.	ผู้วิจัย	ประเด็นที่ศึกษา	องค์ความรู้ที่ได้รับ
			ความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของหมวดข้อสอบมากกว่าวิธีดีเอฟไอที เมื่อกลุ่มตัวอย่างมีขนาด 1 ,000 คน เมื่อแบบสอบประกอบด้วยข้อสอบ 40 ข้อ วิธีซิบเทสท์ มีความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของหมวดข้อสอบมากกว่าวิธีดีเอฟไอที เมื่อกลุ่มตัวอย่างมีขนาด 500 คน เมื่อแบบสอบประกอบด้วยข้อสอบ 50 ข้อ ไม่พบความแตกต่างระหว่างวิธีการทั้งสอง
2547	สุมาลี แกล้วทนงค์	ศึกษาสาเหตุของการทำหน้าที่ต่างกันของข้อสอบสาระการเรียนรู้ภาษาไทยและสาระการเรียนรู้สังคมศึกษา ศาสนาและวัฒนธรรมต่างกัน ระหว่าง 1) วิธีแมนเทล-แฮนส์เซล 2) วิธีซิบเทสท์	ข้อสอบที่ทำหน้าที่ต่างกันด้านเพศ ส่วนใหญ่มีสาเหตุมาจากเนื้อหาและภาษาที่ใช้ในแบบสอบ ข้อสอบที่ทำหน้าที่ต่างกันด้านภาษาพูดส่วนใหญ่มีสาเหตุมาจากการใช้คำศัพท์เฉพาะและบริบททางวัฒนธรรมและข้อสอบที่ทำหน้าที่ต่างกันด้านเชื้อชาติ ส่วนใหญ่มีสาเหตุมาจากบริบททางภาษาและบริบททางวัฒนธรรม
2547	อุทัยวรรณ สายพัฒนะ	ศึกษาการเปรียบเทียบประสิทธิภาพของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบที่มีการให้คะแนนแบบหลายค่า ในเงื่อนไขความยาวของแบบสอบและกลุ่มตัวอย่างที่แตกต่างกัน ระหว่าง 1) วิธี GMH 2) วิธี Polytomous SIBTEST	การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างขนาดของกลุ่มตัวอย่างของแต่ละวิธีการ 1) วิธี GMH เมื่อแบบสอบประกอบด้วยข้อสอบ 30 ข้อ และ 20 ข้อ กลุ่มตัวอย่างขนาด 500 และ 250 คน ส่งผลต่อความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

ตารางที่ 2.7 (ต่อ)

พ.ศ.	ผู้วิจัย	ประเด็นที่ศึกษา	องค์ความรู้ที่ได้รับ
			ไม่แตกต่างกันนอกนั้นแตกต่างกัน 2)วิธี Polytomous SIBTEST เมื่อแบบสอบประกอบด้วยข้อสอบ 20 ข้อ กลุ่มตัวอย่างขนาด 1,000 และ 500 คน ส่งผลต่อความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบไม่แตกต่างกัน นอกนั้นแตกต่างกัน 3) การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี GMH และวิธี Polytomous SIBTEST ในทุกเงื่อนไขความยาวของแบบสอบ กลุ่มตัวอย่างทุกขนาดส่งผลต่อประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของวิธีการทั้งสองไม่แตกต่างกัน ยกเว้นกรณีกลุ่มตัวอย่างขนาด 1,000 คนของแบบสอบ ที่ยาง 20 ข้อ วิธี Polytomous SIBTEST มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูงกว่าวิธี GMH
2549	ปิยะทิพย์ ตินวร	เปรียบเทียบประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบพหุมิติ ระหว่าง 1) การวิเคราะห์องค์ประกอบจำกัด และ 2) วิธีถดถอยโลจิสติก	วิธีถดถอยโลจิสติกมีประสิทธิภาพไม่แตกต่างกันกับวิธีวิเคราะห์องค์ประกอบจำกัดและวิธีถดถอยโลจิสติกมีประสิทธิภาพมากกว่าวิธีวิเคราะห์องค์ประกอบจำกัด
2549	อรินทร์ น่วมถนอม	ศึกษาการเปรียบเทียบการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า โดยการจำลองข้อมูลภายใต้โมเดลพาเชียลเครดิตทั่วไปแบบหลายมิติ ภายใต้ปัจจัยที่แปรเปลี่ยน 4	ทั้งสามวิธีมีอัตราความถูกต้องใกล้เคียงกันในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป (Uniform DIF) และแบบอเนกรูป(Nonuniform DIF) ทั้งสองวิธีมีอัตราความถูกต้อง

ตารางที่ 2.7 (ต่อ)

พ.ศ.	ผู้วิจัย	ประเด็นที่ศึกษา	องค์ความรู้ที่ได้รับ
		ปัจจัย ระหว่าง 1) วิธีโพลีซิบเทสท์ 2) วิธีถดถอยโลจิสติกแบบจัดอันดับ (Ordinal Logistic regression) 3) วิธีถดถอยโลจิสติกแบบจัดอันดับหลายมิติ (Multidimensional Ordinal Logistic regression)	สูงกว่าวิธีโพลีซิบเทสท์ ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปสัดส่วนของข้อสอบทำหน้าที่ต่างกันในแบบสอบไม่มีผลต่อวิธีโพลีซิบเทสท์ วิธีถดถอยโลจิสติกแบบจัดอันดับหลายมิติ แต่มีผลต่อวิธีถดถอยโลจิสติกแบบจัดอันดับเมื่อความแตกต่างของการแจกแจงความสามารถเพิ่มขึ้นวิธีโพลีซิบเทสท์ สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ที่สูงเกินปกติได้ดีกว่าวิธีอื่น นอกจากนี้ เมื่อขนาดตัวอย่างเพิ่มขึ้นมีผลทำให้ทุกวิธีมีอัตราความถูกต้องเพิ่มขึ้นเกือบทุกเงื่อนไขที่ศึกษา
2550	สุทธิพร ศุรธรณี	ศึกษาความสามารถในการตรวจสอบ DIF ตามวิธีตัวแบบวางนัยทั่วไประดับลดหลั่น ผลตอบข้อสอบเป็นแบบทวิภาคขยายตัวแบบของคามาตะ โดยศึกษาความแกร่งของตัวแบบคามาตะ การตรวจสอบ DIF ในกรณีที่ข้อสมมติเกี่ยวกับการแจกแจงของผู้สอบไม่เป็นปกติ กำหนดให้ความสามารถของผู้สอบมีการแจกแจงแบบฟิชเชอร์ทิปเปทท์ ค่าความสามารถมีค่าได้ตั้งแต่ $-\infty$ ถึง ∞ ความเบ้เท่ากับ 1.14 ซึ่งตัดเฉพาะค่าที่อยู่ระหว่าง -3 ถึง 3 เท่านั้นเพื่อที่ให้ช่วงความเชื่อมั่นเหมือนกับการแจกแจงแบบปกติมาตรฐาน จำลองข้อมูลด้วยเทคนิคมอนติคาล์โดยใช้โปรแกรม R เวอร์ชัน 2.01 ผู้สอบ 1000 คน แบ่งเป็นกลุ่มอ้างอิง 500 กลุ่มเปรียบเทียบ 500 คน	กรณีที่ความสามารถของผู้สอบทั้งสองกลุ่มมีการแจกแจงแบบปกติ ความแตกต่างระหว่างค่าเฉลี่ยของความสามารถกลุ่มอ้างอิงและกลุ่มเปรียบเทียบไม่มีผลต่อค่า RMSE ค่าเฉลี่ยประมาณของ DIF กำลังการทดสอบและอัตราความผิดพลาดประเภทที่ 1 วิธีการตรวจสอบที่ละข้อจะขึ้นอยู่กับสัดส่วนของข้อสอบที่มี DIF โดยที่ค่า RMSE จะมีค่ามากขึ้นเมื่อสัดส่วนของผู้สอบที่มี DIF สูงขึ้น กลับกัน สัดส่วนของข้อสอบที่มี DIF ไม่มีผลต่อวิธีการตรวจสอบทุกข้อพร้อมกันและวิธีการตรวจสอบทีละข้อยังมีค่า RMSE น้อยกว่าวิธีการตรวจสอบทุกข้อพร้อมกันกรณี

ตารางที่ 2.7 (ต่อ)

พ.ศ.	ผู้วิจัย	ประเด็นที่ศึกษา	องค์ความรู้ที่ได้รับ
		ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยวิธี HGLM ใช้โปรแกรมสำเร็จรูป HLM เวอร์ชัน 6.0 ทำซ้ำ 100 ครั้ง คำนวณค่าประมาณของ DIF ค่า RMSE กำลังทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1	ที่มี DIF 1 และ 3 ข้อ แต่กรณีที่มี DIF 6 ข้อ วิธีการตรวจสอบที่ละข้อยังมีค่า RMSE มากกว่าวิธีการตรวจสอบทุกข้อพร้อมกัน และค่าเฉลี่ยประมาณ DIF ขึ้นอยู่กับวิธีการที่ใช้ในการตรวจสอบ DIF โดยที่วิธีการตรวจสอบทีละข้อให้ค่าเฉลี่ยต่ำกว่าค่าจริง ขณะที่วิธีการตรวจสอบทุกข้อพร้อมกันให้ค่าเฉลี่ยทั้งต่ำกว่าและสูงกว่าค่าจริง

หมายเหตุ : การสังเคราะห์งานวิจัยทั้งในประเทศและต่างประเทศอยู่ในหัวข้อที่ 5.3

5.2 งานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบต่างประเทศ

Swaminathan and Rogers (1990) ได้เปรียบเทียบวิธีถดถอยโลจิสติก (Logistic Regression) กับวิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) กลุ่มตัวอย่างขนาด 250 และ 500 คน ความยาว 3 ขนาด 40, 60, 80 ข้อ ผลการวิจัยพบว่าวิธีถดถอยโลจิสติกและวิธีแมนเทล-แฮนส์เซล ให้ผลการตรวจสอบการทำหน้าที่ต่างกันแบบเอกรูปได้ถูกต้องร้อยละ 70 กรณีกลุ่มตัวอย่าง 250 คน ตรวจสอบได้ร้อยละ 100 กรณีกลุ่มตัวอย่าง 500 คน ในทุกความยาวของแบบสอบ สำหรับการตรวจสอบการทำหน้าที่ต่างกันแบบอเนกรูป พบว่าวิธีแมนเทล-แฮนส์เซลสามารถตรวจสอบได้เล็กน้อย ส่วนวิธีถดถอยโลจิสติกสามารถตรวจสอบได้ถูกต้องร้อยละ 50 กรณีกลุ่มตัวอย่างน้อยและข้อสอบสั้นและถูกต้องร้อยละ 75 กรณีแบบสอบยาวและกลุ่มตัวอย่างขนาดใหญ่

Mazor et al (1992) ศึกษาผลกระทบของขนาดกลุ่มตัวอย่างที่มีต่อการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซล ศึกษาจากข้อมูลจำลอง กลุ่มตัวอย่างที่ใช้มี 5 ขนาด คือ 100, 200, 500, 1,000 และ 2,000 คน ความยาวแบบสอบ 75 ข้อ ผลการวิจัยพบว่าเมื่อขนาดกลุ่มตัวอย่างเท่ากับ 500 คนหรือน้อยกว่าจะสามารถระบุข้อสอบที่ทำหน้าที่ต่างกันได้ถูกต้องน้อยกว่าร้อยละ 50 และเมื่อขนาดกลุ่มตัวอย่างเท่ากับ 2,000 คน สามารถระบุข้อสอบที่ทำหน้าที่ต่างกันได้ถูกต้องร้อยละ 70 ถึงร้อยละ 75 และได้กล่าวว่า ข้อสอบที่ไม่สามารถตรวจสอบพบหรือระบุว่าทำหน้าที่ต่างกันได้เนื่องมาจากข้อสอบเหล่านั้นมีความยากมากหรือมีความยากต่างกันเพียงเล็กน้อยระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบอีกทั้งเป็นข้อที่มีค่าอำนาจจำแนกต่ำ

Rogersand Swaminathan (1993) ศึกษาวิธีถดถอยโลจิสติก กับวิธีแมนเทล-แฮนส์เซล เพื่อเปรียบเทียบการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยจำลองข้อมูลเพื่อศึกษาการกระจายของสถิติทดสอบและประสิทธิภาพการทำหน้าที่ต่างกันแบบเอกรูปและแบบอเนกรูป การศึกษากระจายของสถิติทดสอบเป็นการศึกษาปัจจัยที่แปรเปลี่ยน คือ ขนาดของกลุ่มตัวอย่าง 250 และ 500 คน ความเหมาะสมของข้อมูลกับโมเดล ค่าความยากและอำนาจจำแนกของข้อสอบ ความยาวของแบบสอบ 40 ข้อ ผลการวิจัยพบว่าการกระจายของสถิติทดสอบทั้งสองวิธีน่าพอใจ ยกเว้นการกระจายของสถิติวิธีถดถอยโลจิสติกไม่เป็นไปตามที่คาดไว้ ในกรณีข้อสอบยากมากและอำนาจจำแนกสูง ด้านประสิทธิภาพการตรวจสอบทั้งสองวิธีมีเท่ากันในการตรวจสอบการทำหน้าที่ต่าง กันแบบเอกรูปวิธีถดถอยโลจิสติกตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูปได้ดีในกรณีที่ข้อสอบมีความยากปานกลางและอำนาจจำแนกสูง ส่วนวิธีแมนเทล-แฮนส์เซล ตรวจสอบข้อสอบที่มีความยากปานกลางได้น้อยมากแต่สามารถตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปได้ดีในกรณีที่ข้อสอบง่ายมากหรือข้อสอบที่ยากมาก

Mazor et al., (1994) ศึกษาการใช้วิธีแมนเทล-แฮนส์เซล ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการจำลองข้อมูลปัจจัยที่ศึกษา คือความยาวของแบบสอบ 2 ขนาด คือ 20 , 40 ข้อ ภายใต้เงื่อนไขข้อสอบที่ทำหน้าที่ต่างกันเนื่องจากฟังก์ชันการตอบข้อสอบแปรเปลี่ยนไป โดยกำหนด

ฟังก์ชันการตอบของกลุ่มอ้างอิงคงที่ (a=1.0, b=0, c=0.2) ส่วนกลุ่มเปรียบเทียบกำหนดค่าอำนาจจำแนกไว้ 3 ระดับ ค่าความยากที่ต่างกัน 3 ระดับ และค่าโอกาสในการเดา 2 ระดับ กลุ่มตัวอย่างกลุ่มละ 500 คน ผลการวิจัยพบว่าภายใต้เงื่อนไขข้อสอบที่ DIF ค่าความยาก ค่าอำนาจจำแนก และค่าโอกาสในการเดา การกระจายความสามารถและปฏิสัมพันธ์ระหว่างการกระจายความสามารถ กับค่าพารามิเตอร์ของแบบสอบมีผลต่อการประมาณค่า α_{MH} และข้อสอบที่ทำหน้าที่ต่างกันส่วนใหญ่เป็นข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูปมากกว่าแบบอเนกรูป

Narayanan and Swaminathan (1996) เปรียบเทียบประสิทธิภาพระหว่างวิธีแมนเทล-แฮนส์เซล วิธีถดถอยโลจิสติกและวิธีโครซิบท์ (CRO-SIB) ในการตรวจสอบการทำหน้าที่ต่างกันแบบอเนกรูป โดยการจำลองข้อมูลภายใต้เงื่อนไข คือ 1) กลุ่มตัวอย่างขนาด 500 และ 1,000 คน ในกลุ่มอ้างอิง และ 200 , 500 คน ในกลุ่มเปรียบเทียบ 2) การกระจายความ สามารถระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบเท่ากันและไม่เท่ากัน 3) สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 3 ระดับ คือ 0% ,10% และ 20% 4) ขนาดของข้อสอบที่ทำหน้าที่ต่างกันหรือพื้นที่รวมแตกต่างระหว่างโค้งคุณลักษณะข้อสอบ 2 กลุ่ม 4 ระดับ คือ 0.4, 0.6, 0.8, และ 1.0 และ 5) ค่าโอกาสในการเดากำหนดเท่ากันที่ 0.2 และความยาวของแบบสอบเป็น 40 ข้อทุกเงื่อนไข ผลการวิจัยพบว่า วิธีถดถอยโลจิสติกและวิธีโครซิบท์ให้ผลใกล้เคียงกันในการตรวจสอบการทำหน้าที่ต่างกันแบบอเนกรูปและ 2 วิธีตรวจ จับการทำหน้าที่ต่างกันได้ดีกว่าวิธีแมนเทล-แฮนส์เซล ปัจจัยที่ส่งผลต่อการตรวจสอบการทำหน้าที่ต่าง กันแบบอเนกรูป ประกอบด้วย ขนาดกลุ่มตัวอย่างเมื่อเพิ่มขนาดกลุ่มตัวอย่างทั้ง 3 วิธีสามารถตรวจสอบได้มากขึ้น การกระจายความสามารถระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบแบบเท่ากันทำให้ตรวจ สอบได้มากขึ้น พื้นที่ความแตกต่างระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบเพิ่มขึ้นจาก 0.4 เป็น 1.0 ทั้ง 3 วิธี สามารถตรวจสอบได้มากขึ้น ข้อสอบที่พบว่าทำหน้าที่ต่างกันด้วยวิธีถดถอยโลจิสติกและวิธีโครซิบท์ส่วนใหญ่เป็นข้อสอบที่มีค่าความยากต่ำ ค่าอำนาจจำแนกสูง ส่วนวิธีแมนเทล-แฮนส์เซล ตรวจ สอบข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปได้ดีเฉพาะกรณีข้อสอบยากและข้อสอบง่ายซึ่งโค้งลักษณะข้อสอบ (ICC) ของผู้สอบ 2 กลุ่มตัดกันที่ระดับความสามารถสูงหรือความสามารถต่ำเท่านั้น

Roussos and Stout (1996) ศึกษากลุ่มตัวอย่างขนาดเล็กที่มีต่อความคลาดเคลื่อนชนิดที่ 1 ของวิธีซิบเทสท์ และวิธีแมนเทล-แฮนส์เซล ศึกษาครั้งแรกใช้ขนาดของกลุ่มตัวอย่าง 100, 200, 300 และ 1,000 คน ความแตกต่างของค่าเฉลี่ยการกระจายความสามารถระหว่างกลุ่มเป็น 0 , 0.5 และ 1.0 ข้อสอบจำนวน 25 ข้อ ผลการวิจัยพบว่า ค่าสถิติของวิธีซิบเทสท์และวิธีแมนเทล-แฮนส์เซล มีแนวโน้มที่จะมีความคลาดเคลื่อนชนิดที่ 1 เพิ่มขึ้น เมื่อขนาดกลุ่มตัวอย่างมีความแตกต่างของค่าเฉลี่ยระหว่างกลุ่มเพิ่มขึ้น ส่วนการศึกษาครั้งที่ 2 ใช้ขนาดของกลุ่มตัวอย่าง 500, 1,000 และ 3,000 คน ความแตกต่างของค่าเฉลี่ยการกระจายความสามารถระหว่างกลุ่มเป็น 0 และ 1.0 ค่าอำนาจจำแนก 3 ระดับ ค่าความยาก 5 ระดับ ค่าโอกาสในการเดา 3 ระดับ พบว่าเมื่อความแตกต่างของค่าเฉลี่ยการกระจายความสามารถ

เป็น 1.0 ทำให้ความคลาดเคลื่อนชนิดที่ 1 เพิ่มขึ้นทุกวิธี

Chang et al. (1996) ได้เปรียบเทียบประสิทธิภาพของวิธีซิปเทสท์ปรับใหม่กับวิธี Mantel และวิธี Standardized Mean Difference (SMD) โดยใช้การจำลองข้อมูล ในการศึกษา 2 ครั้ง การศึกษาครั้งแรกใช้การจำลองข้อมูลจากการสอบ NAEP ซึ่ง Zwick et al. (1993) ได้ศึกษาไว้ประกอบด้วยข้อสอบ 24 ข้อ เป็นข้อสอบที่ให้คะแนนแบบ 2 ค่า จำนวน 20 ข้อ และข้อสอบที่ให้คะแนนแบบหลายค่าแบบ 4 ลำดับขั้นคะแนน คือ 0, 1, 2, 3 จำนวน 4 ข้อ ส่วนวิธี Mantel และ SMD จะมีทั้งหมด 25 ข้อ ในการจำลองการตอบสนองข้อสอบใช้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ เงื่อนไขที่ศึกษาทั้งหมด 54 เงื่อนไขเป็นการแจกแจงความสามารถของกลุ่มสนใจ 2 แบบ คือ N(0,1) กับ N(-1,1) ทำการศึกษากับข้อสอบ 27 ข้อ ได้จากประเภทของ DIF 4 ประเภท คือ Constant DIF, Low-shift DIF, High-shift DIF และ Balanced DIF, ขนาดของ DIF 2 ขนาด คือ .1 และ .25, พารามิเตอร์ความยาก 3 ค่าและข้อสอบที่เป็น null DIF อีก 3 ข้อ แต่ละเงื่อนไขกระทำซ้ำ (replications) 600 ครั้ง ผลการศึกษาพบว่าวิธี SIBTEST ตรวจสอบ DIF ได้ดีแต่วิธี Mantel และ SMD ดีกว่าวิธี SIBTEST เล็กน้อย ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธี Mantel และ SMD เป็น .049 และ .046 ตามลำดับ ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธี SIBTEST สูงกว่าเล็กน้อย .063 วิธีการทั้ง 3 สามารถตรวจสอบ DIF แบบทิศทางเดียวได้ดี การศึกษาครั้งที่สองเป็นการศึกษาที่เพิ่มเติมจากการศึกษาครั้งแรก โดยรวมข้อสอบที่ศึกษากับอำนาจจำแนกที่แตกต่างกันส่วนระดับการเปลี่ยนแปลงค่าความยากเป็น . 25 ปริมาณ DIF ในข้อสอบที่ศึกษาขึ้นอยู่กับค่าพารามิเตอร์ความยากและอำนาจจำแนกผลการ ศึกษาพบว่าเมื่อพารามิเตอร์อำนาจจำแนกมีค่ามากขึ้น อัตราการปฏิเสธสมมติฐานที่เป็นกลางหรืออำนาจการทดสอบสูงขึ้นปริมาณ DIF กับพารามิเตอร์อำนาจจำแนกโดยตรงซึ่งเป็นไปตามที่คาดหวังเมื่อกำหนดการเปลี่ยนแปลงพารามิเตอร์ความยากไว้คงที่ วิธี SIBTEST สามารถควบคุมผลกระทบที่ก่อให้เกิดอัตราความคลาดเคลื่อนประเภทที่ 1 ได้เหนือกว่าวิธี Mantel และวิธี SMD ภายใต้เงื่อนไขทั่วๆ ไปของการตรวจสอบ DIF เมื่อขนาดของกลุ่มตัวอย่างเพิ่มขึ้น อำนาจในการตรวจสอบของวิธี Mantel และ SMD จะเร็วในขณะที่วิธี SIBTEST มีความคงที่พอควร

French and Miller (1996) ได้ศึกษาความเป็นไปได้ของการใช้วิธีถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนแบบหลายค่าโดยศึกษาจากแบบสอบที่จำลองขึ้น 25 ข้อ แต่ละข้อมี 4 ลำดับขั้นคะแนน คือ ตั้งแต่ 0 ถึง 3 คะแนน ซึ่งในแบบสอบจะมีข้อสอบข้อเดียวที่จำลองพารามิเตอร์ของข้อสอบให้แตกต่างกันเพื่อสร้างเงื่อนไข 3 เงื่อนไขสำหรับการทำหน้าที่ต่างกันแบบอเนกรูปและมี 1 เงื่อนไขสำหรับการทำหน้าที่ต่างกันแบบเอกรูป ส่วนคะแนนสอบถูกจำลองขึ้นโดยใช้โมเดล GPCM (generalized partial credit model) ของ Muraki (1992) ข้อมูลที่ได้จะถูกนำมากำหนดรหัสใหม่ (recoded) ให้เป็น multiple dichotomies โดยใช้โมเดลตามแนวคิดของ Agresti (1990) 3 โมเดล คือโมเดลโลจิทของอัตราส่วนที่ต่อเนื่องกัน โมเดลโลจิทแบบสะสม และโมเดลโลจิทของลำดับขั้นที่ติดกัน เพื่อใช้กับวิธีถดถอยโลจิสติกในการกำหนดรหัสให้ใช้กับรูปแบบการตอบแบบหลายค่า

(multinomial response models) ผลการวิจัยพบว่า การเปลี่ยนแปลงขนาดของกลุ่มตัวอย่างมีผลต่ออำนาจในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อกลุ่มตัวอย่างมีขนาดเล็กลงอำนาจในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบลดลง วิธีโมเดลโลจิทของอัตราส่วนที่ต่อเนื่องกันและแบบสะสมมีอำนาจในการตรวจสอบสูงสุดในทุกๆครั้งของการถดถอยแต่ต่างกันที่วิธีโมเดลโลจิทของอัตราส่วนที่ต่อเนื่องกันมีการสูญเสียข้อมูลในการกำหนดรหัสครั้งที่ 2 และ 3 ในขณะที่วิธีโมเดลโลจิทแบบสะสมยังคงเก็บข้อมูลทั้งหมดไว้ในทุกครั้งของการกำหนดรหัส ส่วนวิธีโมเดลโลจิทของลำดับขั้นที่ติดกัน มีอำนาจการตรวจสอบของการถดถอยครั้งแรกต่ำแต่มีอำนาจเพียงพอในการตรวจสอบของการถดถอยครั้งที่ 2 และ 3 วิธีกามีการสูญเสียข้อมูลในทุกครั้งของการถดถอยและมีอำนาจในการตรวจสอบ DIF ของข้อสอบต่ำกว่าอีก 2 วิธี แต่มีประสิทธิภาพในการเปรียบเทียบระหว่างแต่ละลำดับขั้นคะแนนโดยตรงซึ่งมีประโยชน์ในการค้นหาตำแหน่งของการทำหน้าที่ต่างกันของข้อสอบภายในข้อสอบ นอกจากนั้นพบว่าเมื่อพารามิเตอร Nอำนาจจำแนกของข้อสอบยิ่งแตกต่างกันมาก อำนาจในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปยิ่งเพิ่มขึ้นและพบว่าการทำให้ข้อมูลแบบหลายค่ากลายเป็นข้อมูลแบบ 2 ค่า ตามหลักการกำหนดรหัสของทั้ง 3 วิธี มีผลต่ออำนาจในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเนื่องจากเกิดการสูญหายของข้อมูล

Flowers et al. (1997) ได้ศึกษาการบรรยาย DFIT ของข้อสอบที่ให้คะแนนแบบ Polytomous และประเมินรวมถึงเปรียบเทียบการทำ DFIT ในการแผ่ขยายของขั้นตอน SIBTEST และ Lord's chi-square ใช้ข้อมูลที่จำลองขึ้นมา กลุ่มตัวอย่าง 500 และ 1,000 คน การกระจายของกลุ่มเปรียบเทียบ (N(0,1) และ N(-1,1)) จำนวนข้อสอบที่ทำหน้าที่ต่างกัน (0%, 10%, และ 20%) ค่า DIF ที่มากที่สุดและค่าพารามิเตอร์ a ได้รับการประเมิน ค่าความคลาดเคลื่อนประเภทที่ 1 จะได้รับผลกระทบจากจำนวนข้อสอบที่ทำหน้าที่ต่างกัน ค่า DIF ที่มากที่สุดและค่าของพารามิเตอร์ a การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้รับผลกระทบจากทุกปัจจัย การคำนวณ DFIT และ Lord's chi-square จำเป็นต้องใช้การประมาณค่าความสามารถ คือขั้นตอนปรับเทียบใช้โปรแกรมคอมพิวเตอร์ PARSCALE 2 วิธี Maximum marginal likelihood และ EM algorithm การประมาณค่าสัมประสิทธิ์การปรับเทียบใช้โปรแกรม EQUATE 2.0 การคำนวณ DFIT ใช้โปรแกรม FORTRAN โดย Raju (1995) และการคำนวณ Lord's chi-square ใช้โปรแกรม FORTRAN ที่เขียนโดย Kim (1993) ส่วนการตรวจสอบ DIF โดยวิธี SIBTEST ใช้โปรแกรม PSIBTEST ผลการวิจัยพบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ใกล้เคียงกับระดับแอลฟายกเว้นเมื่อจำนวนข้อสอบที่มี DIF 20% และจำนวน DIF ที่มากสุดมีค่าสูง ปัจจัยที่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 คือค่าของพารามิเตอร์ หาก a ต่ำ อัตราความคลาดเคลื่อนประเภทที่ 1 จะสูง การทำหน้าที่ต่างกันของข้อสอบได้รับผลกระทบมาจากทุกปัจจัยในการศึกษาครั้งนี้

Oshima, Raju and Flowers (1997) ศึกษาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบแบบพหุมิติ (Multidimensional) โดยใช้กรอบแนวคิดของวิธีดีเอฟไอที (DFIT) ในการศึกษานี้ใช้ข้อมูลจำลอง

จากโมเดลโลจิสติกแบบพหุมิติ แบบ 2 พารามิเตอร์ (M2PL) จำนวน 40 ข้อ ศึกษาภายใต้เงื่อนไข รูปแบบการทำหน้าที่ต่างกัน 2 รูปแบบ คือ แบบเอกรูปและแบบอเนกรูป ทิศทางการทำหน้าที่ต่างกัน 2 รูปแบบ คือ แบบทิศทางเดียว (Unidirectional) และ 2 ทิศทางที่สมดุลกันและการแจกแจงความสามารถที่แตกต่างกันระหว่างกลุ่มอ้างอิง (reference group) และ กลุ่มเปรียบเทียบ (focal group) 2 รูปแบบ คือ (0,1) และ (1,1) ผลการวิจัย พบว่าข้อสอบที่จำลองขึ้นไม่ทำหน้าที่ต่างกัน (No DIF) ยังพบว่าเมื่อค่าความยากของทั้ง 2 มิติแตกต่างกันจะทำให้ค่าดัชนี CDIF และค่าดัชนี NCDIF มีค่าเพิ่มมากขึ้น หากค่าความยากของทั้ง 2 มิติแตกต่างกันในทิศทางตรงกันข้าจำทำให้ค่าดัชนี CDIF มีค่าเท่ากัน แต่หากค่าความยากของทั้ง 2 มิติแตกต่างกันแต่เป็นไปในทิศทางเดียวกันจะทำให้ค่าดัชนี NCDIF จะมีค่ามากขึ้น

Oshima, Raju, Flowers and Slinde (1998) ศึกษาสาเหตุการทำหน้าที่ต่างกันในกลุ่มข้อสอบ (Differential Bundle Functioning: DFIT-DBF) โดยแบ่งข้อสอบออกเป็นกลุ่มๆ แตกต่างกันตามลำดับขั้นของการเรียนรู้ เครื่องมือที่ใช้เป็นแบบวัดทักษะการอ่าน Metropolitan Achievement Tests จำนวน 55 ข้อ กลุ่มตัวอย่างเป็นนักเรียนเกรด 4 จำนวน 1,000 คน ตัวแปรที่ใช้แบ่งกลุ่มตัวอย่าง คือ ตัวแปรเพศ และตัวแปรเศรษฐานะทางสังคม จากนั้นคำนวณค่าดัชนี bundle-CDIF, bundle-NCDIF และ bundle-DTF ด้วยวิธีดีเอฟไอที (DFIT) ผลการวิจัยพบว่าเมื่อเปรียบเทียบระหว่างเพศหญิงและเพศชาย เมื่อตัดข้อสอบข้อ 22 และข้อ 25 ซึ่งเป็นข้อที่ตรวจสอบพบว่าเกิดการทำหน้าที่ต่างกันออกจากแบบสอบตรวจสอบการทำหน้าที่ต่างกันของแบบสอบ (DTF) แบบสอบไม่ทำหน้าที่ต่างกัน นั่นคือดัชนี DTF ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ เมื่อเปรียบเทียบเศษฐานะทางสังคม ไม่พบข้อสอบที่ทำหน้าที่ต่างกัน (No DIF) และเมื่อแบ่งวิเคราะห์ตามกลุ่มข้อสอบ พบว่ากลุ่มข้อสอบที่ 5 มีค่าดัชนี NCDIF สูงที่สุดโดยเข้าข้างเพศชายมากกว่าเพศหญิงแต่ค่าดัชนี bundle NCDIF มีค่าไม่แตกต่างกันเมื่อแบ่งกลุ่มตามเศรษฐานะทางสังคม

Kim (2000) ศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการทดสอบอัตราส่วนไลด์ลิฮู้ด-แมนเทล และวิธีแมนเทล-แฮนส์เซล แบบทั่วไป (GMH) โดยใช้ข้อมูลจากการประเมินโรงเรียนระดับอนุบาลของรัฐจอเจียร์ แบ่งขนาดของกลุ่มตัวอย่างที่ศึกษาออกเป็น 4 ขนาด คือกลุ่มตัวอย่างขนาด 105,731 คน กลุ่มตัวอย่างขนาด 10,000 คน กลุ่มตัวอย่างขนาด 1,000 คน และกลุ่มตัวอย่างขนาด 100 คน ข้อคำถามที่นำมาตรวจสอบการทำหน้าที่ต่างกันให้คะแนนมากกว่า 2 ค่า มีวัตถุประสงค์เพื่อศึกษาความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการทดสอบอัตราส่วนไลด์ลิฮู้ด วิธีแมนเทลและวิธีแมนเทล-แฮนส์เซลแบบทั่วไป ผลการวิจัยพบว่า ทั้ง 3 วิธี ให้ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ดีเมื่อกลุ่มตัวอย่างขนาด 100 คน ข้อค้นพบที่สำคัญคือการใช้กลุ่มตัวอย่างที่มีขนาดใหญ่เกินไปจะไม่มีประโยชน์ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

Penfield (2001) ศึกษาการทำหน้าที่ต่างกันของข้อสอบในหลายกลุ่มด้วยวิธีแมนเทล-แฮนส์เซล 3 แบบ โดยมีวัตถุประสงค์เพื่อศึกษาขนาดของความคลาดเคลื่อนชนิดที่ 1 เมื่อกลุ่มตัวอย่างที่ศึกษ าหมี

ลายกลุ่ม เงื่อนไขที่ศึกษา คือ 1) ขนาดของกลุ่มตัวอย่าง กลุ่มอ้างอิง 1 กลุ่มและกลุ่มเปรียบเทียบ 1 กลุ่ม กลุ่มอ้างอิง 1 กลุ่มและกลุ่มเปรียบเทียบ 2 กลุ่ม กลุ่มอ้างอิง 1 กลุ่มและกลุ่มเปรียบ เทียบ 3 กลุ่ม กลุ่มอ้างอิง 1 กลุ่มและกลุ่มเปรียบเทียบ 4 กลุ่ม และ 2) การแจกแจงความสามารถของกลุ่มเปรียบเทียบ โดยเปรียบเทียบด้วยวิธีแมนเทล-แฮนส์เซลแบบไคสแควร์ ที่ไม่ปรับระดับของ α วิธีแมนเทล-แฮนส์เซลแบบไคสแควร์ที่ไม่ปรับระดับของ α ด้วย Bonferroni และวิธีแมนเทล-แฮนส์เซลแบบทั่วไป ผลการวิจัยพบว่า วิธีแมนเทล-แฮนส์เซลแบบทั่วไป ดีที่สุดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในหลายกลุ่ม

Walker and Beretvas (2001) ศึกษาการสืบสอบเชิงประจักษ์กระบวนทัศน์ของการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแบบพหุมิติ:การอธิบายทางพุทธิปัญญาสำหรับการทำหน้าที่ต่างกันของข้อสอบ เพื่อเปรียบเทียบผลของการศึกษาการทำหน้าที่ต่างกันของข้อสอบแบบเอกมิติ (Unidimensional) กับการทำหน้าที่ต่างกันของข้อสอบแบบพหุมิติ (Multidimensional) ว่าการศึกษาแบบใดจะสอดคล้องกับข้อมูลเชิงประจักษ์มากกว่ากัน ระหว่าง วิธีโพลีซิบเทสท์ (poly-SIBTEST) และวิธี LISREL ผลการวิจัยพบว่า การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแบบพหุมิติ สอดคล้องกับข้อมูลเชิงประจักษ์มากกว่าการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแบบเอกมิติ

Gierl, Bisanz, Bisanz and Boughton (2003) ศึกษาการระบุเนื้อหาและทักษะทางพุทธิปัญญาที่ทำให้เกิดความแตกต่างทางเพศที่มีต่อวิชาคณิตศาสตร์ โดยใช้กระบวนทัศน์ของการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแบบพหุมิติ (Multidimensional DIF) กลุ่มตัวอย่างเป็นนักเรียนระดับเกรด 9 จำนวน 12,000 คน เป็นชาย 6,000 คน และหญิง จำนวน 6,000 คน เครื่องมือที่ใช้เป็นแบบวัดผลสัมฤทธิ์ทางคณิตศาสตร์แบบเลือกตอบจำนวน 45 ข้อ และ แบบเติมคำตอบจำนวน 10 ข้อ โดยข้อสอบทุกข้อเป็นข้อสอบที่ตรวจให้คะแนนแบบทวิวิภาค โดยผู้วิจัยใช้กรอบแนวคิดทางคณิตศาสตร์แบบปรับปรุงของ Gallagher (2000) ประกอบด้วย 1) การแก้ปัญหาทางคณิตศาสตร์ในสถานการณ์ที่ไม่คุ้นเคย 2) การแก้ปัญหาคณิตศาสตร์ในสถานการณ์ที่คุ้นเคย 3) ทักษะความจำ และ 4) มิติสัมพันธ์ กำหนดความสามารถทางคณิตศาสตร์เป็นมิติความสามารถที่ 1 และกำหนดความสามารถทางพุทธิปัญญาเป็นมิติความสามารถที่ 2 จากนั้นวิเคราะห์ข้อมูลด้วยวิธีซิบเทสท์และวิธีดิมเทสท์ (DIMTEST) โดยวิธีซิบเทสท์ในการหาขนาดของการเกิดการทำหน้าที่ต่างกันของข้อสอบ ส่วนวิธีดิมเทสท์ใช้ในการวิเคราะห์มิติของแบบสอบ ผลการวิจัยพบว่าข้อสอบบางข้อเข้าข้างนักเรียนชาย ส่วนข้อสอบบางข้อเข้าข้างนักเรียนหญิง นักเรียนชายทำคะแนนในส่วนของมิติสัมพันธ์ (spatial) ได้ดีกว่านักเรียนหญิง ในขณะที่นักเรียนหญิงทำคะแนนในส่วนของทักษะความจำ (memorization) ได้ดีกว่านักเรียนชาย

Cohen and Bolt (2005) ได้วิเคราะห์โมเดลแบบผสมในการทำหน้าที่ต่างกันของข้อสอบตามเพศและใช้แบบสอบแบบผสมตามแนวคิด IRT มีวัตถุประสงค์เพื่อ 1) เพื่อสำรวจมิติที่ทำให้เกิดการทำหน้าที่ต่างกันของข้อสอบ 2) ศึกษาคุณลักษณะของข้อสอบที่มีความเกี่ยวข้องกับมิติที่ทำให้เกิดการทำหน้าที่ต่างกันของข้อสอบ 3) เปรียบเทียบคุณลักษณะทางวิชาการของผู้เรียนกับคุณลักษณะที่ปรากฏใน

ชั้นเรียน แบบสอบที่ใช้เป็นแบบสอบจัดตำแหน่งวิชาคณิตศาสตร์ แบบเลือกตอบ 32 ข้อ กลุ่มตัวอย่าง 1,000 คน แบ่งเป็นเพศชาย 500 คน และ เพศหญิง 500 คน วิเคราะห์ข้อมูลด้วยวิธีการทดสอบอัตราส่วนไลด์ลิฮู้ด (Likelihood Ratio Test) ด้วยโปรแกรม Multilog ผลการวิจัย พบว่ามีข้อสอบที่ทำหน้าที่ต่างกันจำนวน 5 ข้อ โดยข้อสอบจำนวน 4 ข้อ เข้าข้างเพศชายและข้อสอบอีก 1 ข้อ เข้าข้างเพศหญิง

Su and Wang (2005) ได้จำลองข้อมูลในการสืบสอบปัจจัยที่มีผลต่อการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธีแมนเทล วิธีแมนเทล-แฮนส์เซลทั่วไป (Generalized Mantel-Haenzel) และวิธี Logistic Discriminant Function Analysis (LDFA) ได้นำมาใช้ในการประเมินการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาค ผลการวิจัย พบว่าความสำคัญของคะแนนจับคู่คือการวัดพื้นที่เฉลี่ยระหว่างโค้งลักษณะข้อสอบ 2 โค้งของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบซึ่งมีความ สำคัญมากกว่าจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้ง 3 วิธีมีการควบคุมค่าอัตราความคลาดเคลื่อนประเภทที่ 1 ได้อย่างดี วิธีแมนเทลและวิธี Logistic Discriminant Function Analysis มีอำนาจการตรวจสอบสูงกว่าวิธี Logistic Discriminant Function Analysis

Finch (2005) ได้เปรียบเทียบโมเดล MIMIC กับวิธีซิบเทสท์ วิธีแมนเทล-แฮนส์เซลและวิธีการทดสอบอัตราส่วนไลด์ลิฮู้ดแบบ IRT (Likelihood Ratio Test) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการจำลองข้อมูลจำนวนผู้สอบและจำนวนข้อสอบด้วยวิธีมอลติคาร์โล ผลการวิจัยพบว่าโมเดล MIMIC ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ดีในกรณีที่ข้อสอบมีจำนวน 50 ข้อ แบบ 2 พารามิเตอร์ และโมเดล MIMIC สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้สูงในกรณีที่ข้อสอบมีจำนวน 20 ข้อ แบบ 3 พารามิเตอร์โลจิสติก ส่วนความคลาดเคลื่อนชนิดที่ 1 มีค่าต่ำสุดในวิธีแมนเทล-แฮนส์เซล นอกจากนี้ยังได้ข้อค้นพบว่าวิธีซิบเทสท์ ให้ผลคล้ายวิธีแมนเทล-แฮนส์เซล แต่มีขนาดของความคลาดเคลื่อนชนิดที่ 1 สูงกว่า

Lei et al. (2006) ได้ศึกษาเปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์จำลองข้อมูลการทำหน้าที่ต่างกันของข้อสอบทั้งแบบมีทิศทางและไม่มีทิศทาง ภายใต้เงื่อนไขที่ศึกษา คือกลุ่มตัวอย่างที่แตกต่างกัน เงื่อนไขที่ 1 จำนวนกลุ่มตัวอย่าง 1 ,000 คน แบ่งเป็นกลุ่มอ้างอิง 500 คน และกลุ่มเปรียบเทียบ 500 คน ส่วนเงื่อนไขที่ 2 จำนวนกลุ่มตัวอย่าง 1 ,000 คนเท่ากัน แบ่งเป็นกลุ่มอ้างอิง 900 คน และกลุ่มเปรียบเทียบ 100 คน การแจกแจงความสามารถที่แตกต่างกันพิจารณาจากการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็น 0 และ ส่วนเบี่ยงเบนมาตรฐานเป็น 1 ส่วนเงื่อนไขข้อสอบแบ่งเป็น 3 แบบ คือ ไม่เกิดทำหน้าที่ต่างกันของข้อสอบ เกิดการทำหน้าที่ต่างกันของข้อสอบแบบมีทิศทาง (Unidirectional DIF) เกิดการทำหน้าที่ต่างกันของข้อสอบแบบไม่มีทิศทาง (Nondirectional DIF) วิเคราะห์ข้อมูลด้วยวิธีถดถอยโลจิสติก วิธีการทดสอบอัตราส่วนไลด์ลิฮู้ดแบบ IRT (IRT Likelihood Ratio Test) และวิธีแคทซิบ (CATSIB) ผลการวิจัยพบว่า วิเคราะห์ข้อมูลด้วยวิธีถดถอยโลจิสติก และวิธีการทดสอบอัตราส่วนไลด์ลิฮู้ดแบบ IRT ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้งแบบมีทิศทาง

และแบบไม่มีทิศทางได้ดีเท่าๆกัน ทั้ง 2 วิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ดีกว่าวิธีแคทซิบ ในขณะที่วิธีแคทซิบตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปได้ดีกว่าแบบอเนกรูป

Park (2006) ศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้านภาษาและเพศในการทดสอบการเขียนความเรียง MELAB ซึ่งเป็นการวัดความสามารถทางภาษาอังกฤษของรัฐมิชิแกน ประเทศสหรัฐอเมริกา ประกอบด้วยการวัดทักษะการอ่าน การฟังและไวยากรณ์ กลุ่มตัวอย่างมีเวลาเขียนเรียงความ 30 นาที โดยเลือกเขียนเรียงความ 1 หัวข้อ จาก 2 หัวข้อที่กำหนดให้ กลุ่มตัวอย่างในการศึกษาครั้งนี้มีจำนวน 2,269 คน เป็นเพศชาย 686 คน และเป็นเพศหญิง 1 ,583 คน วิเคราะห์ข้อมูลด้วยวิธีถดถอยโลจิสติก แบบ 3 ขั้นตอนในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป และแบบอเนกรูป พบว่า ไม่เกิดการทำหน้าที่ต่างกันของแบบสอบ MELAB

Penfield (2006) ได้ศึกษาการประมาณค่าอิทธิพลของการทำหน้าที่ต่างกันในการวัดการทำหน้าที่ต่างกันของแบบสอบ (DTF) โดยไม่คิดเครื่องหมายในแบบสอบแบบผสม (mixed format test) โดยการจำลองข้อมูล ประกอบด้วยข้อสอบที่ตรวจให้คะแนนแบบทวิวิภาค (dichotomously) และข้อสอบที่ตรวจให้คะแนนแบบหลายค่า (polytomously) 4 ตัวเลือก การวิเคราะห์ข้อมูลแบ่งออกเป็น 2 กรณี คือ กรณีแรก แบบสอบที่ประกอบด้วยข้อสอบที่ตรวจให้คะแนนแบบทวิวิภาค 20 ข้อ และข้อสอบที่ตรวจให้คะแนนแบบหลายค่า 8 ข้อ กรณีที่ 2 แบบสอบที่ประกอบด้วยข้อสอบที่ตรวจให้คะแนนแบบทวิวิภาค 8 ข้อ และข้อสอบที่ตรวจให้คะแนนแบบหลายค่า 12 ข้อ โดยข้อสอบ 2 ค่า วิเคราะห์แบบ 3 พารามิเตอร์ ส่วนข้อสอบหลายค่าวิเคราะห์ด้วยวิธีแมนเทล-แฮนส์เซลแบบทั่วไป (GMH) กลุ่มตัวอย่างจำนวน 1 ,000 คน แบ่งเป็นกลุ่มอ้างอิงจำนวน 500 คน และกลุ่มเปรียบเทียบจำนวน 500 คน โดยพิจารณาจากการแจกแจงแบบปกติที่มีส่วนเบี่ยงเบนมาตรฐานเป็น 1 และค่าเฉลี่ยขึ้นอยู่กับเงื่อนไข 40 เงื่อนไข (2 ระดับของค่าเฉลี่ยการแจกแจงความสามารถ x2 ชนิดของแบบสอบ x2 พารามิเตอร์โอกาสในการเดา x5 ขนาดอิทธิพลของการทำหน้าที่ต่างกัน) ผลการวิจัย พบว่าแบบสอบที่มีข้อสอบที่ตรวจให้คะแนนแบบทวิวิภาคจำนวนมากจะส่งผลต่อความลำเอียงทางลบแต่แบบสอบที่มีข้อที่ตรวจให้คะแนนแบบหลายค่าจำนวนมากจะส่งผลต่อความลำเอียงทางบวกเพียงเล็กน้อยเท่านั้น

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

ต่างกันของข้อสอบจัดให้เห็นแง่มุมความแตกต่างของวิธีแบบดั้งเดิมในการวัดประเด็นการวิจัยทางวัฒนธรรมและความเป็นอยู่ที่ดี การวิเคราะห์ IRT แสดงให้เห็นความแตกต่างของค่าเฉลี่ยระหว่างกลุ่มชาวจีนและชาวอเมริกัน ข้อสอบที่มีความลำเอียงจะให้คะแนนโดยมีน้ำหนักน้อยกว่า ดังนั้นความแตกต่างของค่าเฉลี่ยที่ค้นพบก่อนหน้าระหว่างกลุ่มชาวจีนและชาวอเมริกันอาจจะไม่ค้นพบได้ง่ายในข้อสอบที่มีความลำเอียง สุดท้ายการวิเคราะห์ IRT จัดหาข้อมูลที่เกี่ยวข้องกับแนวคิดของความพึงพอใจในชีวิต การวิจัยครั้งนี้แสดงให้เห็นความสำคัญและประโยชน์ของการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบในโครงสร้างอื่นๆ ผู้วิจัยหวังว่า การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและโมเดล IRT อื่นๆ จะเป็นประโยชน์ในอนาคตในหัวข้อการวิจัยอื่นๆที่เกี่ยวข้องในเรื่องวัฒนธรรมและบุคลิกภาพ

Stark et al. (2006) ได้พัฒนาและทดสอบแผนการร่วมที่ใช้ในการระบุการทำหน้าที่ต่างกันของข้อสอบรวมถึงการทดสอบอัตราส่วนไลค์ลิฮูด (Likelihood Ratio Test) ซึ่งสามารถใช้ได้ทั้งการวิเคราะห์องค์ประกอบเชิงยืนยันและทฤษฎีการตอบสนองข้อสอบ โดยใช้ข้อมูลจำลองในการตรวจสอบความตรงของทั้งสองวิธี IRT ตั้งบนพื้นฐานของวิธี Likelihood Ratio Test และ CFA ตั้งบนพื้นฐานของวิธี mean and covariance structures (MACS) โดยใช้แบบวัดมิติเดียวจำนวน 15 ข้อคำถาม มีตัวแปรที่เกี่ยวข้อง 8 ตัว จำนวนเงื่อนไขที่ใช้จำลองข้อมูลคือ 320 ($2^6 + 2^8$) แต่ละเงื่อนไขจะมีการทำซ้ำ 50 ครั้ง วิเคราะห์วิธี MACS โดยใช้โปรแกรมลิสเรล 8 และการทดสอบด้วยทฤษฎีตอบสนองข้อสอบด้วยวิธี Likelihood Ratio Test ใช้พื้นฐานของโมเดล graded response โดยใช้โปรแกรมคอมพิวเตอร์ MULTILOG ผลการวิจัยพบว่า IRT วิธี Likelihood Ratio Test ให้ผลดีกว่าวิธี MACS ในขณะที่กลุ่มตัวอย่างมีขนาดเล็ก ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบวัดมิติเดียวเมื่อใช้กลุ่มตัวอย่างขนาดเล็ก การวิเคราะห์ MACS ให้ผลดีกว่า หากในกรณีกลุ่มตัวอย่างมีขนาดใหญ่ และข้อมูลเป็นแบบ dichotomous มิติเดียว การตรวจสอบการทำหน้าที่ต่างกันของแบบสอบโดยใช้วิธี IRT ให้ผลดีกว่าอย่างไรก็ตามวิธี IRT หลายๆ วิธีจะมีความแกร่งในการทดสอบ (robust) หากมีการฝ่าฝืนข้อตกลงเบื้องต้นในเรื่องความเป็นเอกมิติ

Kim, Chosen and Kim (2007) ศึกษาการทำหน้าที่ต่างกันของขนาดอิทธิพลของข้อสอบที่ตรวจให้คะแนนแบบหลายค่า (polytomas) กลุ่มตัวอย่างขนาดใหญ่ (N=105,731) เพื่อเปรียบเทียบถึงความสอดคล้องตามวิธีการตรวจสอบทั้ง 5 วิธี ได้แก่ วิธีการทดสอบอัตราส่วนไลด์ลิฮู้ดแบบ IRT วิธีถดถอยโลจิสติก วิธีการทดสอบอัตราส่วนไลด์ลิฮู้ด (Likelihood Ratio Test) วิธีแมนเทล และวิธีแมนเทล-แฮนส์เซลแบบทั่วไป (GMH) โดยใช้โปรแกรม MULTILOG และโปรแกรม IRTLRDIF วิเคราะห์ด้วยวิธีการทดสอบอัตราส่วนไลด์ลิฮู้ดแบบ IRT โปรแกรม SAS ใช้ในการวิเคราะห์ด้วยวิธีถดถอยโลจิสติกและวิธีการทดสอบอัตราส่วนไลด์ลิฮู้ด ส่วนวิธีแมนเทลและวิธีแมนเทล-แฮนส์เซลแบบทั่วไปเขียนโปรแกรมด้วยภาษาฟอร์เทรน ผลการวิจัยพบว่า สามารถตรวจพบข้อสอบที่ทำหน้าที่ต่างกันจากทั้ง 5 วิธี ข้อค้นพบที่สำคัญ คือการใช้กลุ่มตัวอย่างที่มีขนาดใหญ่เกินไปจะไม่มีประโยชน์ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

Elosua and Jauregui (2007) ได้ศึกษาแหล่งของการทำหน้าที่ต่างกันของข้อสอบที่ส่งผลต่อการแปลแบบสอบการศึกษาในครั้งนี้มีวัตถุประสงค์เพื่อหาแหล่งของการทำหน้าที่ต่างกันของข้อสอบที่ส่งผลต่อการแปลแบบสอบ โดยจำแนกบนพื้นฐานของเกณฑ์ทางภาษาและวัฒนธรรม 4 แบบ คือ 1) ความเกี่ยวข้องทางวัฒนธรรม 2) ปัญหาการแปล 3) ไวยากรณ์ และ 4) การตีความหมายคำ การศึกษาในครั้งนี้ใช้ข้อคำถามที่ตรวจให้คะแนนแบบ 2 ค่า เกี่ยวกับภาษาจำนวน 53 ข้อ ซึ่งเป็นข้อคำถามที่สร้างเป็นภาษาสเปนจากนั้นแปลเป็นภาษาบาสต์ และมีการแปลย้อนกลับอีกครั้ง (back-translation) กลุ่มตัวอย่างที่ศึกษาเป็นนักเรียนอายุระหว่า 9-11 ปี 1 ,048 คน แบ่งเป็นกลุ่มอ้างอิง 498 คน และกลุ่มเปรียบเทียบ 550 คน วิเคราะห์ข้อมูลด้วยวิธีแมนเทล-แฮนส์เซล และจากความเห็นของผู้เชี่ยวชาญ (expert judgment) พบว่า เกณฑ์ทั้ง 4 แบบ คือ 1) ความเกี่ยวข้องทางวัฒนธรรม (cultural relevance) 2) ปัญหาการแปล (translation problems) 3) ไวยากรณ์ (grammar) 4) ตีความหมายคำ (semantic differences) ส่งผลต่อการทำหน้าที่ต่างกันของข้อสอบทั้งสิ้นและวิธีแมนเทล-แฮนส์เซล ตรวจสอบพบว่าการทำหน้าที่ต่างกันของข้อสอบได้ทั้งสิ้น 32 ข้อ ผู้เชี่ยวชาญตรวจสอบพบว่าการทำหน้าที่ต่างกันของข้อสอบได้ 28 ข้อ และมีข้อคำถามที่ทั้งผู้เชี่ยวชาญและวิธีแมนเทล-แฮนส์เซล ตรวจสอบพบว่าเกิดการทำหน้าที่ต่างกันของข้อสอบได้ตรงกันจำนวน 22 ข้อ และมีแหล่งของการทำหน้าที่ต่างกันของข้อสอบทั้งสิ้น 29 แหล่ง

Walker, Zhang and Surber (2008) ศึกษาการใช้กรอบแนวคิดกระบวนทัศน์ของการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแบบพหุมิติในการตัดสินความสามารถในการอ่านที่ส่งผลต่อความสามารถทางคณิตศาสตร์ วิเคราะห์ข้อมูลด้วยโปรแกรม NOHARM ผลการวิจัยพบว่า ความสามารถในการอ่านส่งผลต่อความสามารถทางคณิตศาสตร์ในทางบวก นั่นคือนักเรียนที่มีความสามารถในการอ่านสูงจะสามารถทำคะแนนในส่วนของคณิตศาสตร์ได้สูงด้วยและมีนักเรียนเพียงส่วนหนึ่งเท่านั้นที่มีความสามารถในการอ่านสูงแต่ทำคะแนนในส่วนของคณิตศาสตร์ได้ไม่ค่อยดี

Marie (2009) ศึกษาการทำหน้าที่ต่างกันของข้อสอบของแบบวัดความสามารถระดับสูง Mastery Tests เปรียบเทียบของสามวิธีโดยใช้ข้อมูลจริง เพื่อเปรียบเทียบวิธีลอกลิเนียร์ โมเดลโลจิสติกรีเกรสชันและวิธีแมนเทลแฮนส์เซลในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบวัดความสามารถระดับสูง ข้อมูลที่ใช้ในการวิเคราะห์เป็นผลการสอบจากเครื่องมือ Swedish theory driving license test (SDLT) และ mastery test ประด้วยข้อสอบจำนวน 65 ข้อ ในระดับยากผู้เข้าร่วมต้องทำข้อสอบได้อย่างน้อย 52 ข้อขึ้นไปถึงจะผ่านการทดสอบและจากผู้เข้าสอบ 5,404 คนและสุ่มคัดเลือกข้อสอบมา 15 ข้อที่ครอบคลุมหลักสูตรเพื่อนำมาตรวจสอบ DIF สถิติในการตรวจสอบ DIF คือ ลอกลิเนียร์โมเดล (LLM) โลจิสติกรีเกรสชัน และวิธีแมนเทลแฮนส์เซล ใช้โปรแกรม R package (R-Development-Core-Team, 2007) ผลการวิจัยพบว่ามีความสัมพันธ์กันค่อนข้างสูงเกี่ยวกับขนาดของการทำหน้าที่ต่างกันของข้อสอบการใช้วิธีแมนเทลแฮนส์เซลให้ผลที่เหมือนกันกับทั้งสองวิธี วิธีโลจิสติกรีเกรสชันและลอกลิเนียร์โมเดล

ให้ผลที่สอดคล้องกันพอสมควร ส่วนวิธีลอกลิเนียร์โมเดลจะมีประโยชน์ในการให้ค่าช่วงคะแนนในการสอบที่แน่นอนซึ่งถือเป็นสิ่งที่น่าสนใจเป็นพิเศษในแบบวัดความ สามารถระดับสูงนี้ ซึ่งในการทดสอบคะแนนส่วนนี้วิธีการโลจิสติกรีเกรสชันและวิธีแมนเทลแฮนส์เซลให้ผลลัพธ์ที่แตกต่างกัน

Nilufer and Paul De Boeck, (2009) ศึกษารูปแบบ DIF ของข้อสอบที่มีผลการตอบซับซ้อนโดยใช้ยุทธวิธีในการออกแบบแบบสอบ จุดหมายของการศึกษาเพื่อนำเสนอวิธีการสร้างรูปแบบของการตอบ สนองข้อมูลพหุมิติกับกลุ่มโครงสร้างที่เกี่ยวข้องและปัจจัยหลักของกระบวนการประมาณค่าระดับพารา มิเตอร์ของข้อสอบถูกขยายเพื่อรวมผลกระทบของมิติของแบบสอบและปัจจัยจากกลุ่ม ความแตกต่างในสมรรถนะของการทำข้อสอบนอกเหนือจากการประเมินผลจากกลุ่ม การจำแนกความแตกต่างของการเกิดการทำหน้าที่ต่างกันของข้อสอบใน 2 ระดับ ข้อมูลที่ใช้ในการวิเคราะห์เป็นข้อมูลจริงจากการสุ่มนักเรียนประถมศึกษาเกรด 3, 4 ระดับละ 269 คน แบบสอบที่ใช้เป็นแบบเขียนตอบเกี่ยวกับคำศัพท์ที่กำหนดให้ สถิติที่ใช้ในการทดสอบ DIF ใช้ประมาณการปรับเหมาะของวิธีถดถอยโลจิสติกภายใต้วิธีการของทฤษฎีการทดสอบแนวใหม่ ผลการวิจัยพบว่า การให้ตัวอย่างประกอบนี้เป็นการนำเสนอการใช้มาตรวัดความเชี่ยวชาญหรือชำนาญในการสะกดคำของชาวดัสช์ โดยดำเนินการจากสองกลุ่มย่อยคือ ปฏิสัมพันธ์ระหว่างด้านกลุ่มกับข้อสอบและปฏิสัมพันธ์ระหว่างกลุ่มกับข้อสอบในแต่ละด้านโมเดลหลักโดยเฉพาะของข้อสอบแต่ละข้อ

Gómez-Benito, Hidalgo and Padilla, (2009) ศึกษาประสิทธิภาพของขนาดอิทธิพลในการพัฒนาวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก โดยการจำลองข้อมูลในปัจจัยที่แปลเปลี่ยน 5 ปัจจัย คือ รูปแบบของการทำหน้าที่ต่างกันของข้อสอบ ขน าดอิทธิพลของการทำหน้าที่ต่าง กันของข้อสอบ จำนวนข้อสอบที่เกิดการทำหน้าที่ต่างกันในแบบสอบแต่ละฉบับ ขนาดกลุ่มอ้างอิงต่อกลุ่มเปรียบเทียบและความยาวของข้อสอบทั้งฉบับ ศึกษา 225 เงื่อนไข วิธีการตรวจสอบการทำหน้าที่ต่างกันเลือกใช้วิธีถดถอยโลจิสติกภายใต้โมเดลทฤษฎีการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ ผลการวิจัยพบว่า ขนาดอิทธิพลที่เหมาะสมโดยพิจารณาเปรียบ เทียบค่า R^2 จากเกณฑ์ Jodoin and Gierl (2001) ได้ศึกษาประสิทธิภาพของการวัดขนาดอิทธิพลในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ พบว่า การวัดขนาดอิทธิพลโดยสถิติ R^2 ร่วมกับการทดสอบนัยสำคัญจะได้ค่าที่ลดลงจนเกือบจะเป็นศูนย์ของการสรุปผิดว่าข้อสอบทำหน้าที่ต่างกัน (DIF) ทั้งที่ความเป็นจริงข้อสอบไม่ได้ทำหน้าที่ต่าง กัน (No DIF) (False Positive: FP) โดยเมื่อข้อสอบยิ่งมีความยาวมากขึ้น FP ยิ่งใกล้ศูนย์และในทางกลับกันการทดสอบนัยสำคัญของสถิติเพียงอย่างเดียวจะทำให้ได้ FP สูงกว่าเล็กน้อยหรือใกล้เคียงจากค่าปกติทั่วไป อย่างไรก็ตามการวัดขนาดอิทธิพลโดยสถิติ R^2 ให้ผลของอำนาจการทดสอบ ($1-\beta$) ที่ต่ำลงจากการทดสอบนัยสำคัญ ซึ่งผลการวิจัยสนับสนุนให้ศึกษาการวัดขนาดอิทธิพลโดยสถิติ R^2 ร่วมกับการทดสอบนัยสำคัญทางสถิติจะทำให้ได้สารสนเทศมากยิ่งขึ้น

ตารางที่ 2.8 สรุปงานวิจัยที่ได้ศึกษาเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบในต่างประเทศถึงปัจจุบัน

ปี ค.ศ.	ผู้วิจัย	ประเด็นที่ศึกษา	องค์ความรู้ที่ได้รับ
1990	Swaminathan and Rogers	เปรียบเทียบการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีถดถอยโลจิสติก กับ วิธีแมนเทล-แฮนส์เซล	วิธีถดถอยโลจิสติกและวิธีแมนเทล-แฮนส์เซล ให้ผลการตรวจสอบการทำหน้าที่ต่างกันแบบเอกรูปตรวจสอบได้ถูกต้องร้อยละ 100 กรณีกลุ่มตัวอย่าง 500 คน ในทุกความยาวของแบบสอบ สำหรับการตรวจสอบการทำหน้าที่ต่างกันแบบอเนกรูป พบว่า วิธีแมนเทล-แฮนส์เซล ตรวจสอบได้เล็กน้อย ส่วนวิธีถดถอยโลจิสติก ตรวจสอบถูกต้องร้อยละ 50 กรณีกลุ่มตัวอย่างน้อย ข้อสอบสั้นถูกต้องร้อยละ 75 กรณีแบบสอบยาวและกลุ่มตัวอย่างขนาดใหญ่
1992	Mazor et al	ศึกษาผลกระทบของขนาดกลุ่มตัวอย่างที่มีต่อการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซล	1) เมื่อตัวอย่างขนาดใหญ่ขึ้นจะสามารถระบุข้อสอบที่ทำหน้าที่ต่างกันได้ถูกต้องได้มากขึ้น 2) ข้อสอบที่ไม่สามารถตรวจสอบพบหรือระบุว่าทำหน้าที่ต่างกันได้เนื่องมาจากข้อสอบเหล่านั้นมีความยากมากหรือมีความยากต่างกันเพียงเล็กน้อยระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบอีกทั้งเป็นข้อที่มีค่าอำนาจจำแนกต่ำ
1993	Rogers and Swaminathan	เปรียบเทียบการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่าง วิธีถดถอยโลจิสติก กับ วิธีแมนเทล-แฮนส์เซล	1) การกระจายของสถิติทดสอบทั้งสองวิธีน่าพอใจยกเว้นการกระจายของสถิติวิธีถดถอยโลจิสติก ไม่เป็นไปตามที่คาดไว้ในกรณีข้อสอบยากมาก,อำนาจจำแนกสูง 2) ด้านประสิทธิภาพการตรวจสอบ พบว่าทั้งสองวิธีมีประสิทธิภาพในการตรวจสอบเท่ากัน และ 3) การตรวจสอบการทำหน้าที่ต่างกันแบบเอกรูปวิธีถดถอย

ปี ค.ศ.	ผู้วิจัย	ประเด็นที่ศึกษา	องค์ความรู้ที่ได้รับ
			โลจิสติก ตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูปได้ดีในกรณีที่ข้อสอบมีความยากปานกลางและอำนาจจำแนกสูง ส่วนวิธีแมนเทล-แฮนส์เซล ตรวจสอบข้อสอบที่มีความยากปานกลางได้น้อยมาก แต่สามารถตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปได้ดีในกรณีที่ข้อสอบง่ายมากหรือข้อสอบที่ยากมาก
1994	Mazor and al	ศึกษาการใช้วิธีแมนเทล-แฮนส์เซล ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ	ภายใต้เงื่อนไขข้อสอบที่ทำหน้าที่ต่างกัน ค่าความยาก ค่าอำนาจจำแนก ค่าโอกาสในการเดา การกระจายความสามารถและปฏิสัมพันธ์ระหว่างการกระจายความสามารถกับค่าพารามิเตอร์ของแบบสอบมีผลต่อการประมาณค่า α_{MH} และข้อสอบที่ทำหน้าที่ต่างกันส่วนใหญ่เป็นข้อสอบDIFแบบเอกรูปมากกว่าแบบอเนกรูป
1996	NarayananandSwaminathan	ศึกษาเปรียบเทียบประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันแบบอเนกรูประหว่าง 1) วิธีแมนเทล-แฮนส์เซล 2) วิธีถดถอยโลจิสติก 3) วิธีโครซิบท์ (CRO-SIB)	วิธีถดถอยโลจิสติกและวิธีโครซิบท์ ให้ผลในการตรวจสอบการทำหน้าที่ต่างกันแบบอเนกรูปใกล้เคียงกันและทั้ง 2 วิธีตรวจจับการทำหน้าที่ต่างกันได้ดีกว่าวิธีแมนเทล-แฮนส์เซล ปัจจัยที่ส่งผลต่อ DIF แบบอเนกรูป คือขนาดกลุ่มตัวอย่าง เมื่อเพิ่มขนาดกลุ่มตัวอย่างทั้ง 3 วิธีสามารถตรวจสอบได้มากขึ้นการกระจายความสามารถระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบแบบเท่ากันทำให้ตรวจสอบได้มากขึ้น พื้นที่ความแตกต่างระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบเพิ่มขึ้นจาก 0.4 เป็น 1.0 ทั้ง 3 วิธี สามารถตรวจสอบได้มากขึ้น ข้อสอบที่พบว่าทำหน้าที่ต่างกันด้วยวิธีถดถอยโลจิสติกและวิธีโคร

ปี ค.ศ.	ผู้วิจัย	ประเด็นที่ศึกษา	องค์ความรู้ที่ได้รับ
			ซิบท์ ส่วนใหญ่เป็นข้อสอบที่มีค่าความยากต่ำ ค่าอำนาจจำแนกสูง ส่วนวิธีแมนเทล-แฮนส์เซล ตรวจสอบข้อสอบที่DIFแบบอเนกรูปได้ดีเฉพาะกรณีข้อสอบยากและข้อสอบง่ายซึ่งโค้งลักษณะข้อสอบ (ICC) ของผู้สอบ 2 กลุ่มตัดกันที่ระดับความสามารถสูงหรือความสามารถต่ำเท่านั้น
1996	Roussos and Stout	ศึกษาผลของกลุ่มตัวอย่างขนาดเล็กที่มีต่อความคลาดเคลื่อนนิดที่ 1 ระหว่าง วิธีซิบเทสท์ กับ วิธีแมนเทล-แฮนส์เซล	ค่าสถิติของวิธีซิบเทสท์ และวิธีแมนเทล-แฮนส์เซล มีแนวโน้มที่จะมีความคลาดเคลื่อนชนิดที่ 1 เพิ่มขึ้นเมื่อขนาดกลุ่มตัวอย่างมีความแตกต่างของค่าเฉลี่ยระหว่างกลุ่มเพิ่มขึ้น ส่วนการศึกษาครั้งที่ 2 กลุ่มตัวอย่าง 500 , 1,000 และ 3,000 คน ความแตกต่างของค่าเฉลี่ยการกระจายความสามารถระหว่างกลุ่มเป็น 0 และ 1.0 ค่าอำนาจจำแนก 3 ระดับ ค่าความยาก 5 ระดับ ค่าโอกาสในการเดา 3 ระดับ พบว่าเมื่อความแตกต่างของค่าเฉลี่ยการกระจายความสามารถเป็น 1.0 ความคลาดเคลื่อนชนิดที่ 1 เพิ่มขึ้นทั้ง 2 วิธี
1996	Chang et al.	ได้เปรียบเทียบประสิทธิภาพของวิธีซิปเทสท์ปรับใหม่กับวิธี Mantel และวิธี Standardized Mean Difference (SMD)	วิธี SIBTEST ตรวจสอบ DIF ได้ดี แต่วิธี Mantel และ SMD ดีกว่า SIBTEST เล็กน้อย วิธี SIBTEST สามารถควบคุมผลกระทบที่ก่อให้เกิดอัตราความคลาดเคลื่อนประเภทที่ 1 ได้เหนือกว่าวิธี Mantel และวิธี SMD ภายใต้เงื่อนไขทั่วๆ ไปของการตรวจสอบ DIF เมื่อขนาดของกลุ่มตัวอย่างเพิ่มขึ้น อำนาจในการตรวจสอบของวิธี Mantel และ SMD จะเพิ่มขึ้นอย่างรวดเร็ว

ปี ค.ศ.	ผู้วิจัย	ประเด็นที่ศึกษา	องค์ความรู้ที่ได้รับ
1996	French and Miller	ศึกษาความเป็นไปได้ของการใช้วิธีถดถอยโลจิสติก ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบพหุวิภาค	เมื่อกลุ่มตัวอย่างมีขนาดเล็กลง อำนาจในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบลดลงและเมื่อพารามิเตอร์อำนาจจำแนกของข้อสอบยิ่งแตกต่างกันมาก อำนาจในการตรวจสอบการทหน้าที่ต่างกันของข้อสอบแบบอเนกรูปยิ่งเพิ่มขึ้น
1997	Flowers et al.	ศึกษาการบรรยาย DFIT ของข้อสอบที่ให้คะแนนแบบ Polytomous และประเมินรวมถึงเปรียบเทียบการทำ DFIT ในการแผ่ขยายของขั้นตอน SIBTEST และ Lord's chi-square	อัตราความคลาดเคลื่อนประเภทที่ 1 ใกล้เคียงกับระดับแอลฟา ยกเว้นเมื่อจำนวนข้อสอบที่มี DIF 20% และจำนวน DIF ที่มากที่สุดมีค่าสูง ปัจจัยที่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 คือค่าของพารามิเตอร์
1997	Oshima, Raju and Flowers	ศึกษาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบแบบพหุมิติ (Multidimensional DIF) โดยใช้กรอบแนวคิดของวิธีดีเอฟไอที (DFIT) ในการศึกษานี้ใช้ข้อมูลจำลองจากโมเดลโลจิสติกแบบพหุมิติ แบบ 2 พารามิเตอร์ (M2PL)	ข้อสอบที่จำลองขึ้นไม่ทำหน้าที่ต่างกัน (No DIF) นอกจากนี้ยังพบว่า เมื่อค่าความยากของทั้ง 2 มิติแตกต่างกันจะทำให้ค่าดัชนี CDIF และค่าดัชนี NCDIF มีค่าเพิ่มมากขึ้นหากค่าความยากของทั้ง 2 มิติแตกต่างกันในทิศทางตรงกันข้ามจะทำให้ค่าดัชนี CDIF มีค่าเท่ากันแต่หากค่าความยากของทั้ง 2 มิติแตกต่างกันแต่เป็นไปในทิศทางเดียวกันจะทำให้ค่าดัชนี NCDIF จะมีค่าเพิ่มขึ้น
1998	Oshima, Raju, Flowers and Slinde	ศึกษาสาเหตุของการทำหน้าที่ต่างกันของกลุ่มข้อสอบ (Differential Bundle Functioning: DFIT-DBF) โดยแบ่งข้อสอบออกเป็นกลุ่มๆ ที่แตกต่างกัน	เมื่อเปรียบเทียบระหว่างเพศหญิงและเพศชาย เมื่อตัดข้อสอบซึ่งเป็นข้อที่ตรวจสอบพบว่าเกิดการทำหน้าที่ต่างกัน (DIF) ออกจากแบบสอบแล้วตรวจสอบการทำหน้าที่ต่างกันของแบบสอบ (DTF) แบบสอบไม่ทำหน้าที่ต่างกัน นั่นคือ ดัชนี DTF ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ เมื่อเปรียบเทียบเศษฐานะทางสังคม พบว่าไม่พบ

ปี ค.ศ.	ผู้วิจัย	ประเด็นที่ศึกษา	องค์ความรู้ที่ได้รับ
			ข้อสอบที่ทำหน้าที่ต่างกัน (No DIF) และเมื่อแบ่งวิเคราะห์ตามกลุ่มข้อสอบ พบว่ากลุ่มข้อสอบที่ 5 มีค่าดัชนี NCDIF สูงที่สุดโดยเข้าข้างเพศชายมากกว่าเพศหญิงแต่ค่าดัชนี bundle NCDIF มีค่าไม่แตกต่างกันเมื่อแบ่งกลุ่มตามเศรษฐานะทางสังคม
2000	Kim	ศึกษาเปรียบเทียบการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่าง 1) วิธีการทดสอบอัตราส่วนไลด์ลิฮู้ด (Likelihood Ratio Test) 2) วิธีแมนเทล 3) วิธีแมนเทล-แฮนส์เซลแบบทั่วไป (GMH)	ทั้ง 3 วิธี ให้ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ดีเมื่อกลุ่มตัวอย่างขนาด 100 คน ยังมีข้อค้นพบที่สำคัญคือการใช้กลุ่มตัวอย่างที่มีขนาดใหญ่เกินไปจะไม่มีประโยชน์ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ
2001	Penfield	ศึกษาการทำหน้าที่ต่างกันของข้อสอบในหลายกลุ่มด้วยวิธีแมนเทล-แฮนส์เซล 3 แบบ โดยมีวัตถุประสงค์เพื่อศึกษาขนาดของความคลาดเคลื่อนชนิดที่ 1 เมื่อมีกลุ่มตัวอย่างที่ศึกษาพร้อมกันหลายกลุ่ม เปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบด้วย 1) วิธีแมนเทล-แฮนส์เซลแบบไคสแควร์ที่ไม่ปรับระดับของ α 2) วิธีแมนเทล-แฮนส์เซลแบบไคสแควร์ที่ไม่ปรับระดับของ α ด้วย Bonferroni 3) วิธีแมนเทล-แฮนส์เซลแบบทั่วไป	วิธีแมนเทล-แฮนส์เซลแบบทั่วไปดีที่สุดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในหลายกลุ่ม
2001	Walker and Beretvas	ศึกษาการสืบสอบเชิงประจักษ์กระบวนทัศน์ของการ	การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแบบพหุมิติ สอดคล้อง

ปี ค.ศ.	ผู้วิจัย	ประเด็นที่ศึกษา	องค์ความรู้ที่ได้รับ
		วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแบบพหุมิติ : การอธิบายทางพุทธิปัญญาสำหรับการทำหน้าที่ต่างกันของข้อสอบ เป็นการศึกษาเพื่อเปรียบเทียบผลการทำหน้าที่ต่างกันของข้อสอบแบบเอกมิติ (Unidimensional) กับผลการทำหน้าที่ต่างกันของข้อสอบแบบพหุมิติ (Multidimensional) ว่าแบบใดจะสอดคล้องกับข้อมูลเชิงประจักษ์มากกว่ากัน ระหว่าง 1) วิธีโพลีซิบเทสท์ และ 2) วิธี LISREL	กับข้อมูลเชิงประจักษ์มากกว่าการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแบบเอกมิติ
2002	Bolt	เปรียบเทียบการตรวจจับการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบหลายค่าด้วยสถิติพาราเมติกและนันพาราเมตริก เป็นการศึกษาโดยการจำลองข้อมูลด้วยเทคนิคมอลติคาร์โล ระหว่าง 1) วิธี GRM 2) วิธี GRM-LR 3) วิธี GRM-DFIT	1) วิธี GRM ให้ผลที่สอดคล้องกับข้อมูลเชิงประจักษ์มากที่สุด 2) วิธี GRM-LR ให้ค่าความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีอื่นๆ 3) วิธี GRM-DFIT ให้ค่าความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าวิธีอื่น
2003	Gierl, Bisanz, Bisanz and Boughton	ศึกษาการระบุเนื้อหาและทักษะทางพุทธิปัญญาที่ทำให้เกิดความแตกต่างทางเพศที่มีต่อวิชาคณิตศาสตร์ โดยใช้กระบวนทัศน์ของการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแบบพหุมิติ (Multidimensional DIF) ในข้อสอบที่ให้คะแนนแบบ 2 ค่า กำหนดความสามารถทางคณิตศาสตร์เป็นมิติความสามารถที่ 1 ความสามารถทางพุทธิปัญญาเป็นมิติความสามารถ	ข้อสอบบางข้อเข้าข้างนักเรียนชาย ส่วนข้อสอบบางข้อเข้าข้างนักเรียนหญิง โดยนักเรียนชายทำคะแนนในส่วนของมิติสัมพันธ์ (spatial) ได้ดีกว่านักเรียนหญิงในขณะที่นักเรียนหญิงทำคะแนนในส่วนของทักษะความจำ (memorization) ได้ดีกว่านักเรียนชาย

ปี ค.ศ.	ผู้วิจัย	ประเด็นที่ศึกษา	องค์ความรู้ที่ได้รับ
		ที่ 2 วิเคราะห์ข้อมูลด้วยวิธีซิบเทสท์และ DIMTEST โดยวิธีซิบเทสท์ในการตรวจสอบเพื่อหาขนาดของการเกิดการทำหน้าที่ต่างกันของข้อสอบ ส่วนวิธีดิมเทสท์ใช้ในการวิเคราะห์มิติของแบบสอบ	
2005	Cohen and Bolt	วิเคราะห์โมเดลแบบผสมในการทำหน้าที่ต่างกันของข้อสอบโดยวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบตามเพศและใช้แบบสอบแบบผสม (mixed format test) ตามแนวคิด IRT ด้วยวิธีการทดสอบอัตราส่วนไลด์ลิฮู้ด (Likelihood Ratio Test) ด้วยโปรแกรม Multilog	พบการทำหน้าที่ต่างกันของข้อสอบโดยมีข้อสอบที่ทำหน้าที่ต่างกันจำนวน 5 ข้อ โดยข้อสอบจำนวน 4 ข้อ เข้าข้างเพศชายและข้อสอบอีก 1 ข้อ เข้าข้างเพศหญิง
2005	Su and Wang	จำลองข้อมูลในการสืบสอบปัจจัยที่มีผลต่อการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธีแมนเทล วิธีแมนเทล -แฮนส์เซลทั่วไป วิธี Logistic Discriminant Function Analysis (LDFA)	ทั้ง 3 วิธีมีการควบคุมค่าอัตราความคลาดเคลื่อนประเภทที่ 1 ได้เป็นอย่างดีวิธี Mantel และวิธี LDFA มีอำนาจการตรวจสอบสูงกว่าวิธี LDFA
2005	Finch	ศึกษาเปรียบเทียบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โมเดล MIMIC โดยการจำลองข้อมูลจำนวนผู้สอบและจำนวนข้อสอบด้วยวิธีมอลติคาร์โล ระหว่าง 1) วิธีซิบเทสท์ 2) วิธีแมนเทล-แฮนส์เซล 3) วิธีการทดสอบอัตราส่วนไลด์ลิฮู้ด แบบ	โมเดล MIMIC ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ดีในกรณีที่ข้อสอบมีจำนวน 50 ข้อ แบบ 2 พารามิเตอร์ และโมเดล MIMIC สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้สูงในกรณีที่ข้อสอบมีจำนวน 20 ข้อ แบบ 3 พารามิเตอร์โลจิสติก ส่วนความคลาดเคลื่อนชนิดที่ 1 มีค่าต่ำสุดในวิธีแมนเทล-แฮนส์เซล

ปี ค.ศ.	ผู้วิจัย	ประเด็นที่ศึกษา	องค์ความรู้ที่ได้รับ
		IRT (IRT Likelihood Ratio Test)	นอกจากนี้ยังได้ข้อค้นพบว่า วิธีซิบเทสท์ ให้ผลคล้ายวิธีแมนเทล-แฮนส์เซลแต่มีขนาดของความคลาดเคลื่อนชนิดที่ 1 สูงกว่า
2006	Lei and al	ศึกษาการเปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ การจำลองข้อมูลการทำหน้าที่ต่างกันของข้อสอบทั้งแบบมีทิศทางและไม่มีทิศทางระหว่าง 1) วิธีถดถอยโลจิสติก 2) วิธีการทดสอบอัตราส่วนไลด์ลิฮู้ดแบบ IRT 3) วิธีแคทซิบ	วิธีถดถอยโลจิสติกและวิธีการทดสอบอัตราส่วนไลด์ลิฮู้ดแบบ IRT ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้งแบบมีทิศทาง และแบบไม่มีทิศทางได้ดีเท่ากันและทั้ง 2 วิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ดีกว่าวิธีแคทซิบ ในขณะที่วิธีแคทซิบตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบมีทิศทางได้ดีกว่าแบบไม่มีทิศทาง
2006	Park	ตรวจสอบ DIF แบบเอกรูปและแบบอเนกรูป จากข้อสอบด้านภาษาและเพศในการทดสอบการเขียนความเรียง MELAB วัดความสามารถภาษาอังกฤษ รัฐมิชิแกน สหรัฐอเมริกา ด้า การวัดทักษะการอ่าน การฟังและไวยากรณ์ด้วยวิธีถดถอยโลจิสติกแบบ 3 ขั้นตอน	ไม่เกิดการทำหน้าที่ต่างกันของแบบสอบ MELAB
2006	Penfield	ศึกษาการประมาณค่าอิทธิพลของการทำหน้าที่ต่างกันของแบบสอบ (DTF) โดยไม่คิดเครื่องหมายในแบบสอบแบบผสม (mixed format test) ศึกษาจำลองข้อมูล ประกอบด้วยข้อสอบที่ตรวจให้คะแนนแบบ 2 ค่า และแบบหลายค่า 4 ตัวเลือก การวิเคราะห์ข้อมูล	แบบสอบที่มีข้อสอบที่ตรวจให้คะแนนแบบ 2 ค่า จำนวนมากจะส่งผลต่อความลำเอียงทางลบแต่แบบสอบที่มีข้อที่ตรวจให้คะแนนแบบหลายค่าจำนวนมากจะส่งผลต่อความลำเอียงทางบวกเพียงเล็กน้อยเท่านั้น

ปี ค.ศ.	ผู้วิจัย	ประเด็นที่ศึกษา	องค์ความรู้ที่ได้รับ
		แบ่งออกเป็น 2 กรณี คือ แบบสอบที่ประกอบด้วยข้อสอบที่ตรวจให้คะแนนแบบ 2 ค่า 20 ข้อ และ แบบหลายค่าจำนวน 8 ข้อ กรณีที่ 2 แบบสอบที่ตรวจให้คะแนนแบบ 2 ค่า 8 ข้อ และข้อสอบที่ให้คะแนนแบบหลายค่า 12 ข้อ โดยข้อสอบ 2 ค่า วิเคราะห์แบบ 3 พารามิเตอร์ ส่วนข้อสอบหลายค่าวิเคราะห์ด้วยวิธีแมนเทล-แฮนส์เซลแบบทั่วไป (GMH) กลุ่มตัวอย่างจำนวน 1,000 คน แบ่งเป็นกลุ่มอ้างอิง 500 คน และกลุ่มเปรียบเทียบ 500 คน พิจารณาจากการแจกแจงแบบปกติที่มีส่วนเบี่ยงเบนมาตรฐานเป็น1 และค่าเฉลี่ยขึ้นอยู่กับเงื่อนไขทั้งหมด 40 เงื่อนไขที่แตกต่างกัน (2 ระดับค่าเฉลี่ยการแจกแจงความสามารถ x2 ชนิดของแบบสอบ x2 พารามิเตอร์โอกาสในการเดา x5 ขนาดอิทธิพลของการทำหน้าที่ต่างกัน)	
2006	Oishi	ตรวจสอบความเท่าเทียมของการวัดความพึงพอใจด้วยแบบวัดความพึงพอใจในชีวิตระหว่างกลุ่มตัวอย่างชาวอเมริกันและชาวจีนโดยใช้ Multigroup Structural Equation Modeling (SEM), Multiple indicator multiple cause model (MIMIC) ทฤษฎีการตอบสนองข้อสอบ	การวิเคราะห์ IRT แสดงให้เห็นความแตกต่างของค่าเฉลี่ยระหว่างกลุ่มชาวจีนและชาวอเมริกัน

ปี ค.ศ.	ผู้วิจัย	ประเด็นที่ศึกษา	องค์ความรู้ที่ได้รับ
2006	Stark et al.	ได้พัฒนาและทดสอบแผนการร่วมที่ใช้ในการระบุการทำหน้าที่ต่างกันของข้อสอบ โดยใช้วิธี MACS และวิธี LR โดยใช้ข้อมูลจำลองในการตรวจสอบความเที่ยงตรงของทั้งสองวิธี	IRT วิธี LR ให้ผลดีกว่าวิธี MACS ในขณะที่กลุ่มตัวอย่างมีขนาดเล็ก ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบวัดมิติเดียว เมื่อใช้กลุ่มตัวอย่างขนาดเล็ก การวิเคราะห์ MACS ให้ผลดีกว่า
2007	Kim, Chosen and Kim	ศึกษาการทำหน้าที่ต่างกันของขนาดอิทธิพลของข้อสอบที่ตรวจให้คะแนนแบบหลายค่า ใช้กลุ่มตัวอย่างขนาดใหญ่ (N=105,731) เพื่อเปรียบเทียบถึงความสอดคล้องตามวิธี 1) วิธีการทดสอบอัตราส่วนไลด์ลิฮู้ดแบบ IRT 2) วิธีถดถอยโลจิสติก 3) วิธีการทดสอบอัตราส่วนไลด์ลิฮู้ด 4) วิธีแมนเทล 5) วิธีแมนเทล-แฮนส์เซลแบบทั่วไป	ตรวจพบการทำหน้าที่ต่างกันของข้อสอบทั้ง 10 ข้อจากทั้ง 5 วิธีและได้ข้อค้นพบที่สำคัญ คือการใช้กลุ่มตัวอย่างที่มีขนาดใหญ่เกินไปจะไม่มีประโยชน์ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ
2007	Elosua and Jauregui	ศึกษาแหล่งของการทำหน้าที่ต่างกันของข้อสอบที่ส่งผลต่อการแปลแบบสอบ การศึกษาในครั้งนี้มีวัตถุประสงค์เพื่อหาแหล่งของการทำหน้าที่ต่างกันของข้อสอบที่ส่งผลต่อการแปลแบบสอบในข้อคำถามที่ตรวจให้คะแนนแบบ 2 ค่า วิเคราะห์ข้อมูลด้วยวิธีแมนเทล-แฮนส์เซล และจากความเห็นของผู้เชี่ยวชาญ (expert judgment)	เกณฑ์ทั้ง 4 แบบ คือ ความเกี่ยวข้องทางวัฒนธรรม (cultural relevance) ปัญหาในการแปล (translation problems) ไวยากรณ์ (grammar) และการตีความหมายคำ (semantic differences) ส่งผลต่อการทำหน้าที่ต่างกันของข้อสอบ วิธีแมนเทล-แฮนส์เซล ตรวจพบ 32 ข้อ ผู้เชี่ยวชาญตรวจสอบ พบ 28 ข้อ และมีข้อคำถามที่ทั้งผู้เชี่ยวชาญและวิธีแมนเทล-แฮนส์เซล ตรวจสอบพบว่าเกิด DIF ตรงกัน 22 ข้อ การทำหน้าที่ต่างกันของข้อสอบมี 29 แหล่ง

ปี ค.ศ.	ผู้วิจัย	ประเด็นที่ศึกษา	องค์ความรู้ที่ได้รับ
2008	Walker Zhang and Surber	ศึกษาการใช้กรอบแนวคิดกระบวนทัศน์ของการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแบบพหุมิติในการตัดสินผลความสามารถในการอ่านที่ส่งผลต่อความสามารถทางคณิตศาสตร์ วิเคราะห์ข้อมูลด้วยโปรแกรม NOHARM	ความสามารถในการอ่านส่งผลต่อความสามารถทางคณิตศาสตร์ในทางบวก นั่นคือนักเรียนที่มีความสามารถในการอ่านสูงจะสามารถทำคะแนนในส่วนของคณิตศาสตร์ได้สูงด้วยและมีนักเรียนเพียงส่วนหนึ่งเท่านั้นที่มีความสามารถในการอ่านสูงแต่ทำคะแนนในส่วนของคณิตศาสตร์ได้ไม่ค่อยดี
2009	Marie Wiberg	ศึกษาการทำหน้าที่ต่างกันของข้อสอบของแบบวัดความสามารถระดับสูง Mastery Tests ทำการเปรียบเทียบ 3 วิธีโดยใช้ข้อมูลจริง เพื่อต้องการเปรียบเทียบวิธีการลอกลิเนียร์ โมเดลโลจิสติกรีเกรสชันและวิธีแมนเทลแฮนส์เซลสถิติในการตรวจสอบ DIF 1) ลอกลิเนียร์โมเดล(LLM) 2) วิธีถดถอยโลจิสติก 3) วิธีแมนเทล-แฮนส์เซลใช้โปรแกรม R package (R-Development-Core-Team, 2007)	การตรวจสอบDIF ในแบบวัดความสามารถระดับสูงข้อมูลที่ใช้ในการวิเคราะห์เป็นผลการสอบจากเครื่องมือ Swedish theory driving license test (SDLT) และ mastery test ประด้วยข้อสอบจำนวน 65 ข้อ ในระดับยากซึ่งผู้เข้าร่วมต้องทำข้อสอบได้อย่างน้อย 52 ข้อขึ้นไปถึงจะผ่านการทดสอบและจากผู้เข้าสอบ 5404 คนและสุ่มคัดเลือกข้อสอบมา 15 ข้อที่ครอบคลุมหลักสูตร เพื่อนำมาตรวจสอบ DIF พบว่า มีความสัมพันธ์กันค่อนข้างสูงเกี่ยวกับขนาดของการเกิดการทำหน้าที่ต่างกันของข้อสอบ วิธีโลจิสติกรีเกรสชันและลอกลิเนียร์โมเดลให้ผลที่สอดคล้องกันพอสมควร ส่วนวิธีลอกลิเนียร์โมเดลจะมีประโยชน์ในการให้ค่าช่วงคะแนนในการสอบที่แน่นอนซึ่งถือเป็นสิ่งที่น่าสนใจเป็นพิเศษในแบบวัดความสามารถระดับสูงนี้ ซึ่งในการทดสอบคะแนนส่วนนี้วิธีการโลจิสติกรีเกรสชันและวิธีแมนเทลแฮนส์เซลให้ผลลัพธ์ที่แตกต่างกัน
2009	Nilufer and Paul De	ศึกษารูปแบบ DIF จากข้อมูลที่มีผลการตอบซับซ้อนโดยใช้ยุทธวิธีในการออกแบบแบบสอบ สถิติที่ใช้ใน	จุดหมายของการศึกษา เพื่อเสนอวิธีการสร้างรูปแบบของการตอบสนองข้อมูลพหุมิติกับกลุ่มโครงสร้างที่เกี่ยวข้องและปัจจัยหลัก

ปี ค.ศ.	ผู้วิจัย	ประเด็นที่ศึกษา	องค์ความรู้ที่ได้รับ
		การทดสอบ DIF ใช้ประมาณการปรับเหมาะของวิธีถดถอยโลจิสติก ภายใต้วิธีการของทฤษฎีการทดสอบแนวใหม่ (IRT Approach)	ของกระบวนการประมาณค่าระดับพารามิเตอร์เพื่อขยายรวมผลกระทบของมิติของแบบสอบและปัจจัยจากกลุ่ม ความแตกต่างในสมรรถนะของการทำข้อสอบนอกเหนือจากการประเมินผลจากกลุ่ม จำแนกความแตกต่างของการเกิดการทำหน้าที่ต่างกันของข้อสอบ 2 ระดับ ใช้ข้อมูลจริงจากการสุ่มนักเรียนประถมศึกษาเกรด 3, 4 ระดับละ 269 คน ใช้แบบเขียนตอบคำศัพท์ที่กำหนดให้ ผลวิจัยพบว่าตัวอย่างประกอบนี้เป็นการนำเสนอการใช้มาตรวัดความเชี่ยวชาญหรือชำนาญในการสะกดคำของชาวดัสช์ จากสองกลุ่มย่อยคือปฏิสัมพันธ์ระหว่างกลุ่มกับข้อสอบและปฏิสัมพันธ์ระหว่างกลุ่มกับข้อสอบในแต่ละด้านโมเดลหลักโดยเฉพาะข้อสอบแต่ละข้อ
2009	Gómez-Benito, Hidalgo and Padilla	ประสิทธิภาพของขนาดอิทธิพลในวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก จำลองข้อมูล 5 ปัจจัย คือรูปแบบของการทำหน้าที่ต่างกันของข้อสอบ ขนดอิทธิพลของการทำหน้าที่ต่างกันของข้อสอบ จำนวนข้อสอบที่เกิดการทำหน้าที่ต่างกันในแบบสอบแต่ละฉบับ ขนาดกลุ่มอ้างอิงต่อกลุ่มเปรียบเทียบและความยาวของข้อสอบทั้งฉบับ เงื่อนไขที่ทำการศึกษาจำนวน 225 เงื่อนไข	วิธีการตรวจสอบการทำหน้าที่ต่างกันเลือกใช้วิธีถดถอยโลจิสติกภายใต้โมเดลทฤษฎีการตอบสนองข้อสอบ 2 พารามิเตอร์ ผลการวิจัยสนับสนุนให้ศึกษาการวัดขนาดอิทธิพลโดยสถิติ R^2 ร่วมกับการทดสอบนัยสำคัญทางสถิติจะทำให้ได้สารสนเทศมากยิ่งขึ้น

5.3 สรุปประเด็นปัญหาที่พบเกี่ยวกับการศึกษาการทำหน้าที่ต่างกันของข้อสอบ

การวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้งในและต่างประเทศมีการพัฒนาและปรับปรุงอย่างต่อเนื่องโดยอาศัยวิธีการที่มีอยู่พัฒนาสู่วิธีการใหม่หรืออาศัยวิธีที่ใช้ทฤษฎีทางการสอบแบบดั้งเดิม ขยายไปสู่วิธีที่ใช้ทฤษฎีการตอบสนองข้อสอบ อาศัยรูปแบบการตรวจให้คะแนนแบบทวิวิภาค ขยายไปสู่รูปแบบการตรวจให้คะแนนแบบพหุวิภาค การเลือกใช้วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบก็เพื่อให้ผลการตรวจสอบมีประสิทธิภาพและเกิดประสิทธิผลต่อข้อสอบ การศึกษางานวิจัยที่เกี่ยวข้องทั้งในและต่างประเทศ พบว่า ปัจจัยสำคัญที่ส่งผลต่อประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ คือ ความยาวของแบบสอบ (test length) และขนาดของกลุ่มตัวอย่าง (Sample size) สำหรับปัจจัยด้านอื่นๆ มีการศึกษาร่วมได้แก่ รูปแบบของข้อสอบทำหน้าที่ต่างกัน (Form of DIF) สัดส่วนของข้อสอบทำหน้าที่ต่างกัน (Proportion of DIF) ความแตกต่างของการแจกแจงความสามารถ (Ability distribution differences) เป็นต้น การศึกษาเอกสารงานวิจัยทั้งในและต่างประเทศที่ผ่านมามุ่งปลายทางไปยังประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกัน สามารถสรุปประเด็นที่มีการศึกษาเปรียบเทียบเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบที่น่าสนใจซึ่งสอดคล้องกับการศึกษาของ Merie (2007) ดังนี้

5.3.1 การเปรียบเทียบประสิทธิภาพจากผลการตรวจสอบระหว่างสถิติที่แตกต่างกันบนฐานมิติด้านสถิติพาราเมตริกกับสถิตินันพาราเมตริก

การเปรียบเทียบประสิทธิภาพจากผลการตรวจสอบระหว่างสถิติที่แตกต่างกันบนฐานมิติด้านสถิติพาราเมตริกกับสถิตินันพาราเมตริก (Parametric vs. non-parametric) โดยนักการศึกษาเลือกใช้สถิติมาตรวจสอบตามข้อตกลงเบื้องต้นที่สอดคล้องกับข้อมูลที่ต้องการจะศึกษาเรื่องของความเข้มงวดในด้านข้อตกลงเบื้องต้นหรือการกระจายของกลุ่มตัวอย่างต้องเลือกใช้สถิติพาราเมตริก เพราะประการแรกสำหรับการเลือกใช้สถิติในกลุ่มพาราเมตริก คือ การรักษาข้อตกลงเบื้องต้นของสถิติอย่างเข้มงวด หากมีการละเมิดข้อตกลงเบื้องต้นของสถิติก็ควรเลือกใช้สถิตินันพาราเมตริกจะมีความเหมาะสมกว่าการใช้สถิติพาราเมตริก เช่นวิธี LR ที่มีข้อตกลงที่เข้มงวดเกี่ยวกับความสัมพันธ์ระหว่างความน่าจะเป็นในการตอบข้อสอบถูกและคะแนนที่สังเกตได้ต้องอยู่ในรูปเชิงเส้น Embretson and Reise (2000) และ Lord (1980) จึงกล่าวได้ว่าการศึกษาที่เกี่ยวข้องกับการตรวจสอบ DIF ต้องสัมพันธ์กับวิธีการและการเลือกใช้สถิติสำหรับมาตรวจสอบ การศึกษา DIF ในประเทศไทย ช่วงระยะเวลา 20 กว่าปีที่ผ่านมาพบว่าส่วนใหญ่เน้นความสำคัญต่อการเปรียบเทียบประสิทธิภาพของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างสถิติที่ใช้ตรวจสอบโดยเปรียบเทียบวิธีการระหว่างมิติสถิติพาราเมตริกกับมิติสถิตินันพาราเมตริก ร่วมกับ มิติทฤษฎีการทดสอบแนวใหม่กับมิติทฤษฎีทางการสอบแบบดั้งเดิม งานวิจัยส่วนมากที่พบนั้นจะทำการศึกษากับข้อมูลที่สังเกตได้เป็นหลักมากกว่าการศึกษาโดยจำลองข้อมูล เริ่มต้นเมื่อ พ.ศ. 2531 โดยสุรศักดิ์ อมรรัตนศักดิ์ ทำวิจัยเกี่ยวกับประเด็นความลำเอียงของข้อสอบ

(สอดคล้องกับนักวิจัยต่างประเทศที่เริ่มต้นศึกษาใน ค.ศ. 1990 คือ Swaminathan and Rogers; Mazor et al, 1992 เป็นต้น) ต่อมาก็มี กาญจนา วัธนสุนธร (2537) และ เกษรหว่างจิตร (2539) ได้ศึกษาในประเด็นที่คล้ายคลึงกันคือเน้นข้อมูลจริง มีรูปแบบการให้คะแนนแบบทวิวิภาค เป็นการเปรียบเทียบประสิทธิภาพด้านความถูกต้องและความคลาดเคลื่อนที่ใกล้เคียงกัน (งานวิจัยที่พบว่าศึกษาเปรียบเทียบระหว่างวิธีได้แก่งานของนักการศึกษาเหล่านี้ คือ จิติมา วรรณศรี, 2539; เรวดี อินทะสะระ, 2539; ญาณภัทร สีหะมงคล, 2540 ; พรรณี จิตมาศ , 2540; รัชรินทร์ มุคดา, 2540; เสรี ชัดแช้ม, 2540; นพมาศ พิพัฒนสุข, 2541; นิคม กีรติวากูร, 2542; อารี วัชรโสตถิกุล, 2543; ทองอยู่ สาระ, 2543; วลีมาศ แซ่อึ้ง, 2543; รักชนก ยี่สุ่นศรี, 2544; สิริรัตน์ วิภาสศิลป์, 2545 ; สุมาลี แกล้วทนงค์, 2547 ; ปิยะทิพย์ ตินวร, 2549; อุทัยวรรณ สายพัฒนะ, 2547; อรินทร์ น่วมถนอม, 2549 ส่วนต่างประเทศได้แก่)

5.3.2 การศึกษาการทำหน้าที่ต่างกันของข้อสอบโดยการจับคู่ตัวแปรที่เป็นคะแนนที่สังเกตได้ กับคะแนนของตัวแปรแฝง

การศึกษาการทำหน้าที่ต่างกันของข้อสอบโดยการจับคู่ตัวแปร (Matching variable) ที่เป็นคะแนนที่สังเกตได้ (observed score) กับคะแนนของตัวแปรแฝง (latent score) เกี่ยวข้องกับความสัมพันธ์ในลักษณะของการจับคู่ระหว่างตัวแปรที่เป็นคะแนนที่สังเกตได้ กับคะแนนของตัวแปรแฝงบนข้อตกลงของทฤษฎีทางการสอบแบบดั้งเดิมกับทฤษฎีการตอบสนองข้อสอบ เกณฑ์นี้ค่อนข้างมีอิทธิพลต่อผลการตรวจสอบหรือผลของข้อค้นพบ มีลักษณะเป็นการเปรียบเทียบระหว่างผลของการตรวจสอบโดยวิธีที่แตกต่างบนฐานของเกณฑ์ที่ใช้ในการเปรียบเทียบ วิธีการที่ถูกนำมาใช้นั้นต้องมีข้อตกลงเบื้องต้นเกี่ยวข้องกับการจับคู่ระหว่างตัวแปรว่าสามารถนำมาใช้ในการคำนวณค่าได้หรือไม่ วิธีการที่ใช้คะแนนที่สังเกตได้ในการจับคู่ตัวแปร ได้แก่ วิธีแมนเทล-แฮนส์เซล วิธีแมนเทล-แฮนส์เซลทั่วไป วิธีดัชนีมาตรฐาน การปรับให้เป็นมาตรฐานด้วยน้ำหนักตัวประกอบ หรือ STND (Standardization) วิธี Logistic Regression (LR) วิธี LLM (Log Linear Models) และวิธีไค-สแควร์ หรือ χ^2 (Chi-square methods) เช่นการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีของแมนเทล -แฮนส์เซลทั่วไป จะใช้คะแนนรวมจากการทดสอบแทนคุณลักษณะภายในหรือความสามารถของผู้สอบแบบสอบที่มีความยาวมากกว่าย่อมมีความน่าเชื่อถือมากกว่าเพราะการจับคู่เปรียบเทียบระหว่างผู้สอบมีความถูกต้องมากขึ้น กลุ่มวิธีการที่ใช้คะแนนของตัวแปรแฝง ได้แก่ วิธีการที่อาศัยทฤษฎีการตอบสนองข้อสอบ (IRT Method) วิธีการอาศัยสถิตินันพาราเมตริก SIBTEST (the non-parametric SIBTEST) และวิธีการผสม (Mixed effect models) (งานวิจัยที่พบว่าศึกษาเปรียบเทียบระหว่างวิธีได้แก่งานของนักการศึกษาเหล่านี้ คือ สุรศักดิ์ อมรรัตนศักดิ์, 2531; กาญจนา วัธนสุนธร, 2537; เกษรหว่างจิตร, 2539; พรรณี จิตมาศ, 2540; เสรี ชัดแช้ม, 2540; รักชนก ยี่สุ่นศรี, 2544; ปิยะทิพย์ ตินวร, 2549)

5.3.3 การศึกษาการทำหน้าที่ต่างกันของข้อสอบโดยรูปแบบทวิวิภาคกับพหุวิภาค

การศึกษาการทำหน้าที่ต่างกันของข้อสอบโดยรูปแบบทวิวิภาคกับพหุวิภาค (Dichotomously vs. Polytomously) เป็นการศึกษาการทำหน้าที่ต่างกันของข้อสอบที่พิจารณาเรื่องรูปแบบของการให้คะแนนในแบบสอบเป็นหลัก มีความยืดหยุ่นเกี่ยวกับรูปแบบของข้อสอบว่าเป็นการให้คะแนนแบบทวิวิภาค หรือแบบพหุวิภาค วิธีการที่ถูกนำมาใช้นั้นต้องมีข้อตกลงเบื้องต้นเกี่ยวข้องกับคะแนนของแบบสอบว่าสามารถนำมาใช้ในการคำนวณค่าได้ เช่น การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบใน วิธี MH, LR, SIBTEST, LRT, general IRT-LR, LLM และวิธีการผสม (Mixed effect models) แรกเริ่มนั้นการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี MH ไม่ได้มีไว้สำหรับตรวจสอบข้อสอบที่ให้คะแนนแบบพหุวิภาคแต่มีการมาปรับขยายสูตรเพื่อศึกษาเกี่ยวกับข้อสอบที่ให้คะแนนแบบพหุวิภาค ภายหลัง (Zwick, Donoghue and Grimo, 1993) ต่อมาวิธี LR ได้ถูกนำมาปรับใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีรูปแบบการให้คะแนนแบบพหุวิภาค (Camilli and Congdon, 1999) การศึกษา DIF ที่ให้คะแนนแบบพหุวิภาคในประเทศ คือ อุทัยวรรณ สายพัฒนะ (2547), อรินทร์ น่วมถนอม (2549) สำหรับประเทศไทยการสอบในระดับชาติยังคงเน้นรูปแบบของการตรวจให้คะแนนแบบทวิวิภาค นักการศึกษาไทยจึงยังคงเน้นศึกษา DIF ในข้อสอบที่ให้คะแนนแบบทวิวิภาคเป็นส่วนใหญ่

5.3.4 การศึกษาการทำหน้าที่ต่างกันของข้อสอบพิจารณาการวัด /หรือการทดสอบ DIF

การวัด และ/หรื อการทดสอบ DIF (Measure and/or test DIF) การศึกษาประเด็นของการวัด (Measure) นี้เหมือนการคำนวณ หาขนาดหรือปริมาณการเกิด DIF ว่ามีขนาดใหญ่หรือเพียงเล็กน้อยจนแทบจะไม่เกิดข้อแตกต่าง ส่วนการทดสอบ (test DIF) เป็นการตรวจสอบเพื่อมุ่งคำตอบว่าข้อสอบ DIF หรือ NO-DIF เกณฑ์นี้เกี่ยวข้องกับวิธีการที่ถือว่าสามารถทั้งตรวจสอบและวัด DIF วิธีการที่ยอดนิยมที่นักการศึกษาทั้งไทยและต่างประเทศใช้ในการศึกษาคือ วิธีถดถอยโลจิสติก ซึ่งเป็นวิธี ที่สามารถตรวจสอบและหาขนาดของ DIF ได้จากคะแนนที่สังเกตได้มาใช้ในการศึกษา นักการศึกษาสามารถคำนวณค่าได้โดยไม่ยุ่งยากมากนัก จึงค่อนข้างเป็นที่นิยมในต่างประเทศก็มีการศึกษาอย่างต่อเนื่อง เช่น Swaminathanand Rogers (1990); RogersandSwaminathan (1993); NarayananandSwaminathan (1996); FrenchandMiller (1996); Leiandal (2006); Park (2006); Kim, ChosenandKim (2007); Gómez-Benito, HidalgoandPadilla (2009) เป็นต้น

5.3.5 การศึกษารูปแบบของการทำหน้าที่ต่างกันของข้อสอบ เอกรูป/ อเนกรูป

รูปแบบของการทำหน้าที่ต่างกันของข้อสอบที่เป็นแบบเอกรูป (Uniform DIF) กับแบบอเนกรูป (non-uniform DIF) ผลจากการศึกษาในประเด็นนี้ Swaminathan and Rogers (1990) ได้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้งแบบเอกรูป และแบบอเนกรูป เมื่อใช้แบบสอบที่มีความยาวมากขึ้นทำให้ความถูกต้องในการตรวจสอบของวิธีแมนเทล -แฮนส์เซล และวิธีถดถอยโลจิสติกเพิ่มมากขึ้น ยกเว้นใน

กรณีการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปของวิธีแมนเทล-แฮนส์เซล ต่อมาในปี 1993 ได้ศึกษาอีกครั้งให้ผลที่ขัดแย้งกับผลที่ศึกษาในปี 1990 คือ ความยาวของแบบสอบไม่มีผลต่ออำนาจการทดสอบ (power rate) ของวิธีแมนเทล-แฮนส์เซลและวิธีถดถอยโลจิสติกยกเว้นในการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปของวิธีถดถอยโลจิสติก งานวิจัยต่างประเทศส่วนใหญ่จะศึกษารูปแบบทั้งสองควบคู่กันไป ส่วนในไทยนักการศึกษาสนใจรูปแบบอเนกรูปมากกว่า (เสรี ชัดแช้ม , 2540; ทองอยู่ สาระ , 2543; วลีมาศ แซ่อึ้ง, 2543 และ สิริรัตน์ วิภาสศิลป์, 2545 รูปแบบเอกรูปมีศึกษา คือ รัชรินทร์ มุคดา , 2540 และ ทองอยู่ สาระ, 2543)

5.3.6 การศึกษาการทำหน้าที่ต่างกันของข้อสอบโดยพิจารณาความยาวของแบบสอบ

ความยาวของแบบสอบ (Test Length) เป็นตัวแปรที่เกี่ยวข้องกับการสอบทุกครั้ง เมื่อจะตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ จำนวนข้อสอบเป็นตัวแปรที่สำคัญเพราะ จำนวนข้อในแบบสอบส่งผลกระทบต่อความถูกต้องในการจับคู่เปรียบเทียบกลุ่มผู้สอบ นักวิจัยต่างประเทศที่ศึกษาเกี่ยวกับความยาวในแบบสอบที่มีผลต่อการทำหน้าที่ต่างกันของข้อสอบ คือ Swaminathan and Rogers, 1992; Mazor et al, 1992 นักการศึกษาในประเทศไทยที่เน้นการศึกษาเปรียบ เทียบในประเด็นจำนวนข้อในแบบสอบ คือ กาญจนา วัธนสุนธร (2537) ญาณภัทร สีหะมงคล (2540), นิคม กีรติวากูร (2542), ทองอยู่ สาระ (2543), วลีมาศ แซ่อึ้ง (2543), รักชนก ยี่สุ่นศรี (2544), สิริรัตน์ วิภาสศิลป์ (2545), อุทัยวรรณ สายพัฒนะ (2547) และปิยะทิพย์ ตินวร (2549)

5.3.7 การศึกษาการทำหน้าที่ต่างกันของข้อสอบโดยมุ่งไปที่ประเด็นขนาดกลุ่มตัวอย่าง

ขนาดของกลุ่มตัวอย่าง (Sample Size) เป็นตัวแปรหนึ่งที่เกี่ยวข้องกับการสอบทุกครั้ง เมื่อจำนวนผู้สอบต่างกันส่งก็จะส่งผลในการตรวจสอบต่างกันเพราะมีความเฉียบคมหรือไวในการตรวจสอบต่างกัน Kim (2000) ทำการศึกษากับข้อสอบที่ให้คะแนนแบบพหุวิภาค พบว่าการใช้กลุ่มตัวอย่างที่มีขนาดใหญ่เกินจะไม่มีประโยชน์ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ Kim, Chosen, Alagoz and Kim (2007) ศึกษาอีกครั้งกับข้อสอบที่ให้คะแนนแบบพหุวิภาค พบว่าการใช้กลุ่มตัวอย่างขนาดใหญ่จะมีความไวในการตรวจพบการทำหน้าที่ต่างกันค่อนข้างสูงแม้ว่าจะเลือกใช้วิธีใดมาใช้ตรวจสอบก็ทำให้ได้สารสนเทศว่าข้อสอบทำหน้าที่ต่างกัน ทำให้เกิดประเด็นว่าการใช้กลุ่มตัวอย่างที่มีขนาดใหญ่ไม่เป็นประโยชน์ต่อการศึกษาการทำหน้าที่ต่างกันของข้อสอบเท่าที่ควรแต่ก็เกิดข้อเสนอแนะว่าเมื่อนำกลุ่มตัวอย่างที่มีขนาดใหญ่มาตรวจสอบการทำหน้าที่ต่างกันก็ควรศึกษาความเข้มของขนาดอิทธิพล French and Miller (1996) ได้ศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติกให้คะแนนแบบพหุวิภาค พบว่า เมื่อกลุ่มตัวอย่างมีขนาดเล็กลงอำนาจในการตรวจ สอบจะลดลง ดังนั้นกลุ่มตัวอย่างที่มีขนาดเหมาะสมก็ขึ้นอยู่กับเงื่อนไขที่แตกต่างกันในประเด็นสัดส่วนระหว่างจำนวนผู้สอบกับจำนวนข้อสอบกับสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันก็จะเกิดความไม่คงที่ในการตัดสินข้อสอบที่ทำ

หน้าที่ต่างกัน การสรุปจำนวนกลุ่มตัวอย่างที่เหมาะสมมีข้อสนับสนุนดังนี้ Swaminathan and Rogers (1990) พบว่าเมื่อเพิ่มขนาดของกลุ่มตัวอย่างจะมีผลทำให้อำนาจการตรวจสอบของวิธีถดถอยโลจิสติกเพิ่มขึ้นเกือบทุกเงื่อนไข Mazor et al., (1992) Narayanan and Swaminathan (1994) ได้ผลที่สอดคล้องกันคือเมื่อเพิ่มขนาดกลุ่มตัวอย่างมากขึ้นความถูกต้องในการตรวจสอบจะสูงขึ้น (Hill, 1990 cited in Mazor et al., 1992) สำหรับวิธีแมนเทล-แฮนส์เซล (MH) ควรใช้กลุ่มตัวอย่างขนาดระหว่าง 100 ถึง 300 คนสำหรับกลุ่มใดกลุ่มหนึ่งหรือทั้งสองกลุ่มก็เพียงพอในขณะที่ Mazor et al., (1992) เสนอว่าใช้กลุ่มตัวอย่าง 200 คน ก็เพียงพอและไม่ควรน้อยกว่านี้ Narayanan and Swaminathan (1994) เสนอว่าวิธีแมนเทล-แฮนส์เซล และวิธีซิบเทสท์ ขนาดตัวอย่างกลุ่มละ 300 คน ก็เพียงพอที่จะตรวจสอบได้อย่างมีประสิทธิภาพ จิติมา วรรณศรี (2539) พบว่าเมื่อขนาดกลุ่มตัวอย่าง 200 และ 600 คน วิธีแมนเทล-แฮนส์เซลและวิธีซิบเทสท์สามารถตรวจสอบได้ถูกต้องร้อยละ 50 แต่ถ้าขนาดของกลุ่มตัวอย่างเพิ่มขึ้นเป็น 1,000 คน สามารถตรวจสอบได้ถูกต้อง 100% (นักการศึกษาไทยที่ศึกษาการทำหน้าที่ต่างกันของข้อสอบในประเด็นนี้ คือ สุรศักดิ์ อมรรัตนศักดิ์, 2531; กาญจนา วัธนสุนธร, 2537; ญาณภัทร สีหะมงคล, 2540; พรรณี จิตมาศ, 2540; นิคม กีรติวากูร, 2542; ทองอยู่สาระ, 2543; วลีมาศ แซ่อึ้ง, 2543; สิริรัตน์ วิภาสศิลป์, 2545; ปิยะทิพย์ ตินวร, 2549; อุทัยวรรณ สายพัฒนะ, 2547; อรินทร์ น่วมถนอม, 2549.)

5.3.8 การศึกษาการทำหน้าที่ต่างกันของข้อสอบโดยมุ่งไปที่ประเด็นคะแนนจุดตัด

คะแนนจุดตัด (Handle the cut-off score) หรือเกณฑ์การจำแนก พบในงานที่ศึกษาเกี่ยวกับขนาดอิทธิพล (effect size) เพราะต้องมีการกำหนดเกณฑ์การตัดสินขนาดของการเกิดการทำหน้าที่ต่างกันของข้อสอบ งานวิจัยที่พบคือ Gómez-Benito, Hidalgo and Padilla (2009) ประเภทของข้อมูลเชิงประจักษ์ (empirical data) กับข้อมูลจำลอง (simulation data) การศึกษาในประเด็นการทำหน้าที่ต่างกันของข้อสอบส่วนใหญ่ พบว่าใช้ศึกษาตรวจสอบในข้อมูลจริง ส่วนงานวิจัยที่ศึกษาเกี่ยวกับข้อมูลจำลองในประเทศได้แก่ จิติมา วรรณศรี, 2539; นิคม กีรติวากูร, 2542; วลีมาศ แซ่อึ้ง, 2543 และ อรินทร์ น่วมถนอม, 2549 เป็นต้น และมีการศึกษาที่เกี่ยวข้องกับ มิติของข้อสอบระหว่างแบบสอบที่มีมิติเดียว (uni-dimensional) กับแบบสอบที่มีหลายมิติ (multidimensional) การศึกษาในแบบสอบที่มีหลายมิติของประเทศไทย คือ นพมาศ พิพัฒนสุข, 2541; สิริรัตน์ วิภาสศิลป์ , 2545; ปิยะทิพย์ ตินวร , 2549 และ อรินทร์ น่วมถนอม, 2549.

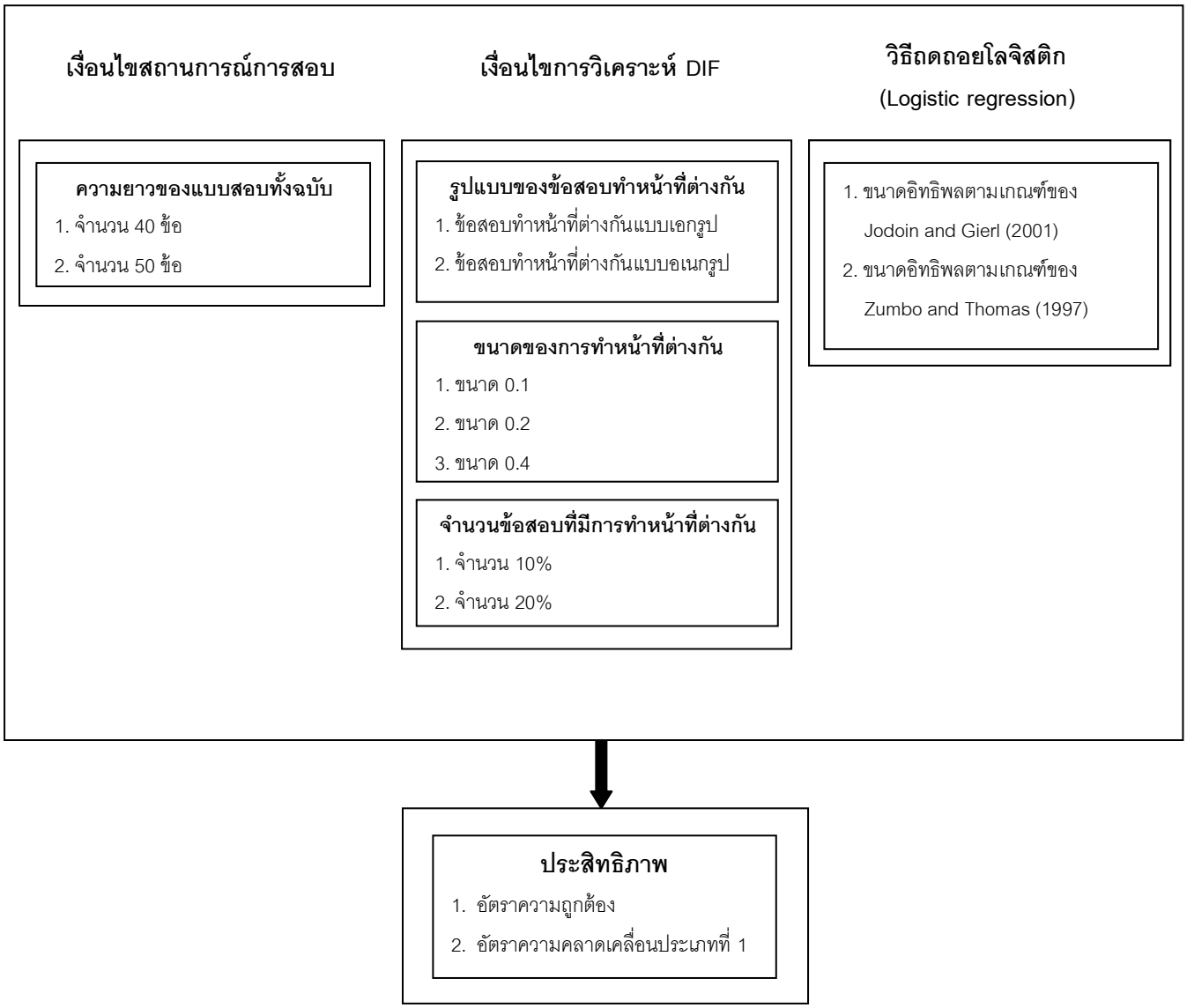
5.3.9 การศึกษาการทำหน้าที่ต่างกันของข้อสอบโดยมุ่ง พิจารณา ประสิทธิภาพ ของการตรวจสอบการทำหน้าที่ต่างกัน

ตัวแปรตามในการศึกษาที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบส่วนใหญ่คือประสิทธิภาพ (Efficiency) พิจารณาจากค่าอำนาจการทดสอบ (power rate) หรืออัตราความถูกต้อง (correct identification) ความคลาดเคลื่อนประเภทที่ 1 (type I error) และอัตราความคลาดเคลื่อน

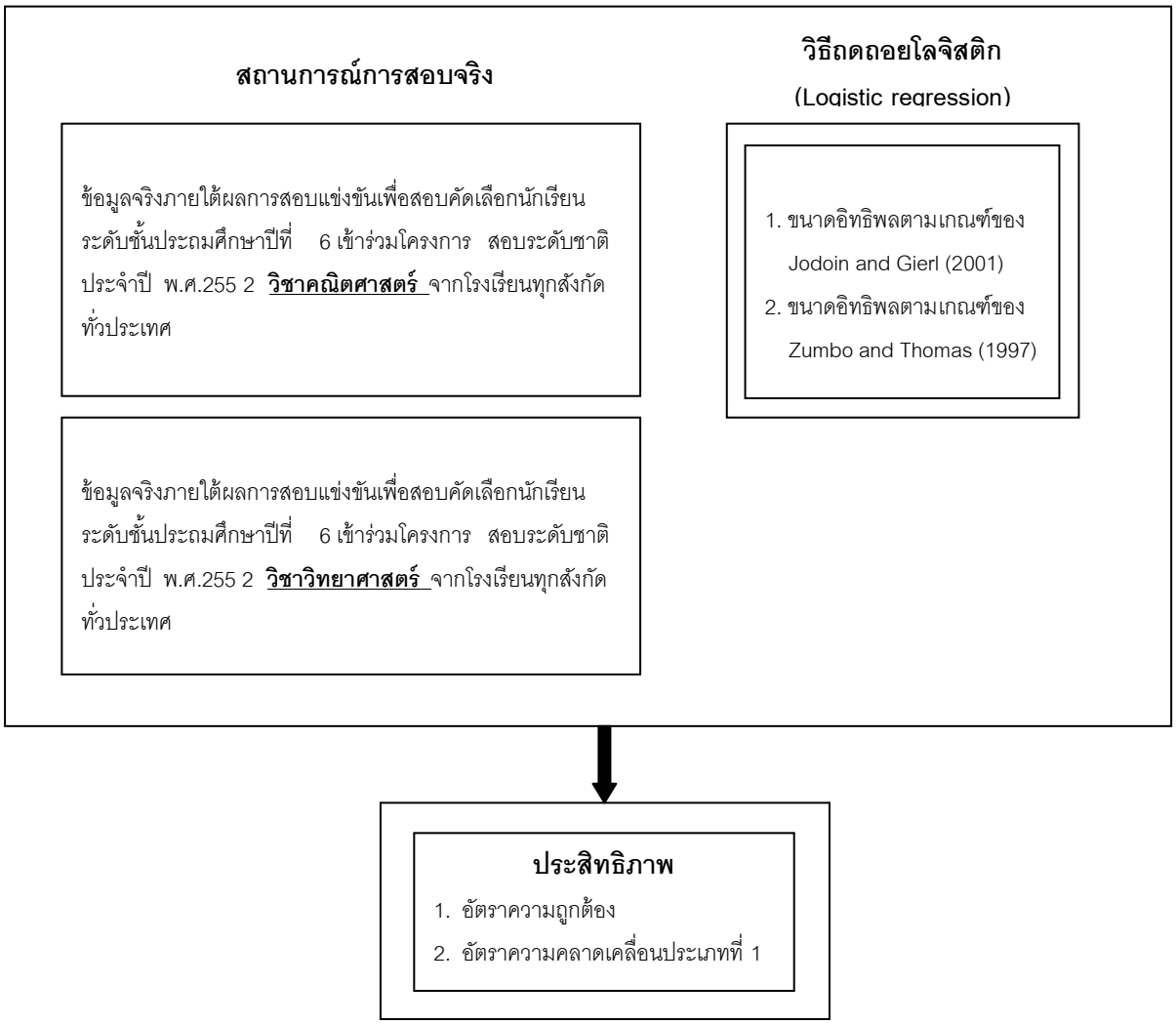
ประเภทที่ 2 (type II error) ผลการศึกษางานวิจัยที่ผ่านมาพบว่า มีปัจจัยบางตัวที่มีผลต่อความถูกต้องแม่นยำในการตรวจ สอบการทำหน้าที่ต่างกันของข้อสอบเช่น ความยาวของแบบสอบขนาดกลุ่มตัวอย่าง ลักษณะของ ข้อสอบ การแจกแจงค่าความสามารถ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน เป็นต้น (สุรศักดิ์ อมรรัตนศักดิ์, 2531; กาญจนา วัธนสุนทร , 2537; เกษร หว่างจิตร, 2539 ; จิติมา วรรณศรี , 2539; เรวดี อินทะสะระ , 2539; รัชรินทร์ มุคดา, 2540 ; ญาณภัทร สีหะมงคล , 2540; พรรณี จิตมาศ , 2540; เสรี ชัดแช้ม , 2540; นพมาศ พิพัฒนสุข , 2541; นิคม กีรติวากูร , 2542; อารี วัชรโสตถิกุล , 2543; ทองอยู่ สาระ, 2543; วลีมาศ แซ่อึ้ง , 2543; รักชนก ยี่สุ่นศรี , 2544; สิริรัตน์ วิภาสศิลป์ , 2545; สุมาลี แกล้วทนงค์ , 2547; ปิยะทิพย์ ตินวร , 2549; อุทัยวรรณ สายพัฒนะ , 2547; อรินทร์ น่วมถนอม , 2549; Swaminathan and Rogers, 1990; Cohen and Kim, 1993; Rogers and Swaminathan, 1993; Uttara and Milsap, 1994; Narayanan and Swaminathan, 1994, 1996; Flowers, Claudia et al., 1997; Stark et al., 2006) เนื่องจากค่าอำนาจการทดสอบ (power of test) กับอัตราความคลาดเคลื่อนประเภทที่ 2 เป็นค่าดัชนีที่มีสเกลที่มีค่าผกผันกัน ดังนั้นการพิจารณาประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจึงสามารถพิจารณาได้จากดัชนีบ่งชี้คุณภาพ 2 ตัว คืออัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจพบการทำหน้าที่ต่างการของข้อสอบก็เพียงพอที่จะได้สาระครบ (ศิริชัย กาญจนวาสี, 2550)

5.4 กรอบแนวคิดการวิจัย

การศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องเกี่ยวกับแนวคิดทฤษฎีเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบทำให้สามารถสรุปกรอบแนวคิดในการวิจัยได้ดังภาพที่ 2.5 และภาพที่ 2.6 ที่เป็นกรอบแนวคิดในการวิจัยกรณีการศึกษาการจำลองข้อมูลและกรณีการศึกษาข้อมูลเชิงประจักษ์ต่อไปนี้



ภาพที่ 2.4 กรอบแนวคิดในการวิจัย กรณีศึกษาการจำลองข้อมูล



ภาพที่ 2.5 กรอบแนวคิดในการวิจัย กรณีศึกษาข้อมูลเชิงประจักษ์

บทที่ 3

วิธีดำเนินการวิจัย

การศึกษาการวัดขนาดอิทธิพลและผลของประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในวิธีถดถอยโลจิสติก สำหรับข้อสอบที่มีรูปแบบการตรวจให้คะแนนแบบทวิวิภาค กรณีข้อมูลจำลองและข้อมูลเชิงประจักษ์ มีวัตถุประสงค์เฉพาะ ดังนี้

1. เพื่อเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีรูปแบบการให้คะแนนแบบทวิวิภาค โดยการจำลอง ข้อมูลในวิธีถดถอยโลจิสติก ระหว่างขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl และขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas **ภายใต้เงื่อนไขเดียวกัน** ของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน ขนาดของการทำหน้าที่ต่างกัน จำนวนข้อสอบที่ทำหน้าที่ต่างกัน และความยาวของแบบสอบทั้งฉบับ

2. เพื่อเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีรูปแบบการให้คะแนนแบบทวิวิภาค โดยการจำลอง ข้อมูลในวิธีถดถอยโลจิสติก ด้วยขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl และขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas **ภายใต้เงื่อนไขต่างกัน**ของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน ขนาดของการทำหน้าที่ต่างกัน จำนวนข้อสอบที่ทำหน้าที่ต่างกัน และความยาวของแบบสอบทั้งฉบับ

3. เพื่อเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีรูปแบบการให้คะแนนแบบทวิวิภาค โดยข้อมูลเชิงประจักษ์ในวิธีถดถอยโลจิสติก ระหว่างขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl และขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas

ผู้วิจัยนำเสนอวิธีดำเนินการวิจัยเป็น 5 ตอน ดังนี้

ตอนที่ 1 ขั้นตอนการวิจัย

ตอนที่ 2 การจัดกระทำข้อมูลตามปัจจัยที่ศึกษา

ตอนที่ 3 การจำลองข้อมูล

ตอนที่ 4 การวิเคราะห์ข้อมูล

ตอนที่ 5 การเปรียบเทียบประสิทธิภาพของผลการตรวจสอบ

ตอนที่ 1 ขั้นตอนการวิจัย

การจำลองข้อมูล (Simulation data) ใช้วิธีการทางคณิตศาสตร์และเทคโนโลยี เพื่อกำหนดสถานการณ์ตามเงื่อนไขที่ซับซ้อน ทำให้ศึกษาการทำหน้าที่ต่างกันของข้อสอบได้หลายเงื่อนไข ขณะที่การศึกษาข้อมูลโดยทั่วไปไม่สามารถดำเนินการได้อย่างครบถ้วนและสมบูรณ์ (Harwell, 1996) ข้อมูลเชิงประจักษ์ (Empirical data) จาก “โครงการ สอบระดับชาติของสถาบันแห่งหนึ่ง ประจำปี พ.ศ.255 2 มีรูปแบบการตรวจให้คะแนนแบบทวิวิภาค ระดับชั้นประถมศึกษาปีที่ 6 จากโรงเรียนทุกสังกัด วิชาวิทยาศาสตร์ และวิชาคณิตศาสตร์ ” วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบตามสภาพที่เกิดขึ้นจริง ใช้วิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) เป็นวิธีเกณฑ์ในการตรวจสอบ ถ้าข้อสอบข้อใดตรวจพบว่าทำหน้าที่ต่างกันสอดคล้องกับวิธีเกณฑ์ก็ถือว่าข้อสอบข้อนั้นตรวจสอบการทำหน้าที่ต่างกันได้ถูกต้อง ข้อจำกัดของการศึกษาข้อมูลเชิงประจักษ์ คือผู้วิจัยไม่สามารถจัดกระทำกับข้อมูลให้เกิดเงื่อนไขต่างๆ เหมือนกับเงื่อนไขที่ศึกษาโดยการจำลองข้อมูล ขั้นตอนการวิจัยมีการดำเนินการ ดังนี้

1.1 ศึกษาค้นคว้าเกี่ยวกับมโนทัศน์ของทฤษฎีทางการสอบแบบดั้งเดิม มโนทัศน์ของทฤษฎีการตอบสนองข้อสอบ มโนทัศน์ของการทำหน้าที่ต่างกันของข้อสอบ ขนาดอิทธิพลในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและงานวิจัยที่เกี่ยวข้องจากเอกสาร หนังสือ วารสารวิชาการและงานวิจัย

1.2 ศึกษาโปรแกรมในการวิเคราะห์ข้อมูล ได้แก่ โปรแกรม WinGen โปรแกรม MULTILOG-MG โปรแกรม R และ โปรแกรม SPSS

1.3 ศึกษาแนวคิดและแนวทางในการวิเคราะห์ข้อมูล

1.3.1 ข้อมูลจำลอง

1) การจำลองข้อมูล ใช้โมเดลภายใต้ทฤษฎีการตอบสนองข้อสอบชนิด 2 พารามิเตอร์ จำลองการตอบข้อสอบที่มีโครงสร้างวัดความสามารถเอกมิติ ข้อสอบทุกข้อให้คะแนนแบบทวิวิภาค กำหนดให้มีผู้สอบจำนวน 2,000 คน (แบ่งสองกลุ่มเท่ากันระหว่างกลุ่มอ้างอิงกับกลุ่มเปรียบเทียบ) ผลการตอบข้อสอบ กำหนดภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน 2 รูปแบบ ขนาดของการทำหน้าที่ต่างกัน 3 ขนาด จำนวนข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด และความยาวของแบบสอบทั้งฉบับ 2 ขนาด มีข้อมูลที่ศึกษาทั้งสิ้น 24 เงื่อนไข (2 รูปแบบ × 3 ขนาด × 2 ขนาด × 2 ขนาด) ทุกเงื่อนไขจำลองข้อมูลซ้ำ 25 ครั้ง

2) ตรวจสอบข้อมูลที่จำลองขึ้นว่าเป็นไปตามเงื่อนไขของปัจจัยที่แปรเปลี่ยนหรือไม่ พร้อมกับวิเคราะห์คุณภาพเบื้องต้นของเครื่องมือ

3) วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก วิเคราะห์ขนาดอิทธิพล (measure of effect size) จากข้อสอบที่ ทำหน้าที่ต่างกัน ตามเกณฑ์ Jodoin and Gierl และเกณฑ์ Zumbo and Thomas

4) คำนวณอัตราความถูกต้อง (correct identification) และอัตราความคลาดเคลื่อนประเภทที่ 1 (Type I error rate)

5) ศึกษาประสิทธิภาพ ของการระบุการทำหน้าที่ต่างกันของข้อสอบจากการตรวจสอบ ภายใต้วิธีถดถอยโลจิสติก ระหว่าง ขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl และขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas พิจารณาประสิทธิภาพจากอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1

6) เปรียบเทียบประสิทธิภาพการตรวจสอบ

7) สรุปผลการวิเคราะห์ของกรณีข้อมูลจำลอง

1.3.2 ข้อมูลเชิงประจักษ์

1) เตรียมข้อมูลเชิงประจักษ์ " โดยพิจารณาความสมบูรณ์ครบถ้วนของข้อมูล จาก "โครงการ สอบระดับชาติของสถาบันแห่งหนึ่ง ประจำปี พ.ศ.2552 มีรูปแบบการตรวจให้คะแนนแบบทวิวิภาค ระดับชั้นประถมศึกษาปีที่ 6 จากโรงเรียนทุกสังกัด วิชาวิทยาศาสตร์ และวิชาคณิตศาสตร์

2) วิเคราะห์สถิติพื้นฐาน ค่าพารามิเตอร์ข้อสอบ

3) ตรวจสอบความเป็นเอกมิติของแบบสอบ

4) วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติกแล้วตัดสินขนาดอิทธิพลของการทำหน้าที่ต่างกันของข้อสอบตามเกณฑ์ Jodoin and Gierl และตามเกณฑ์ Zumbo and Thomas พร้อมกับวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบในข้อมูลเชิงประจักษ์ ด้วยวิธีการวัดพื้นที่ของ Raju เป็นเกณฑ์ในการตรวจสอบ

5) คำนวณอัตราความถูกต้อง (correct identification) และอัตราความคลาดเคลื่อนประเภทที่ 1 (Type I error rate)

6) ศึกษาประสิทธิภาพ (อัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1) ของการระบุการทำหน้าที่ต่างกันของข้อสอบจากการตรวจสอบภายใต้วิธีถดถอยโลจิสติก ระหว่าง ขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl และขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas

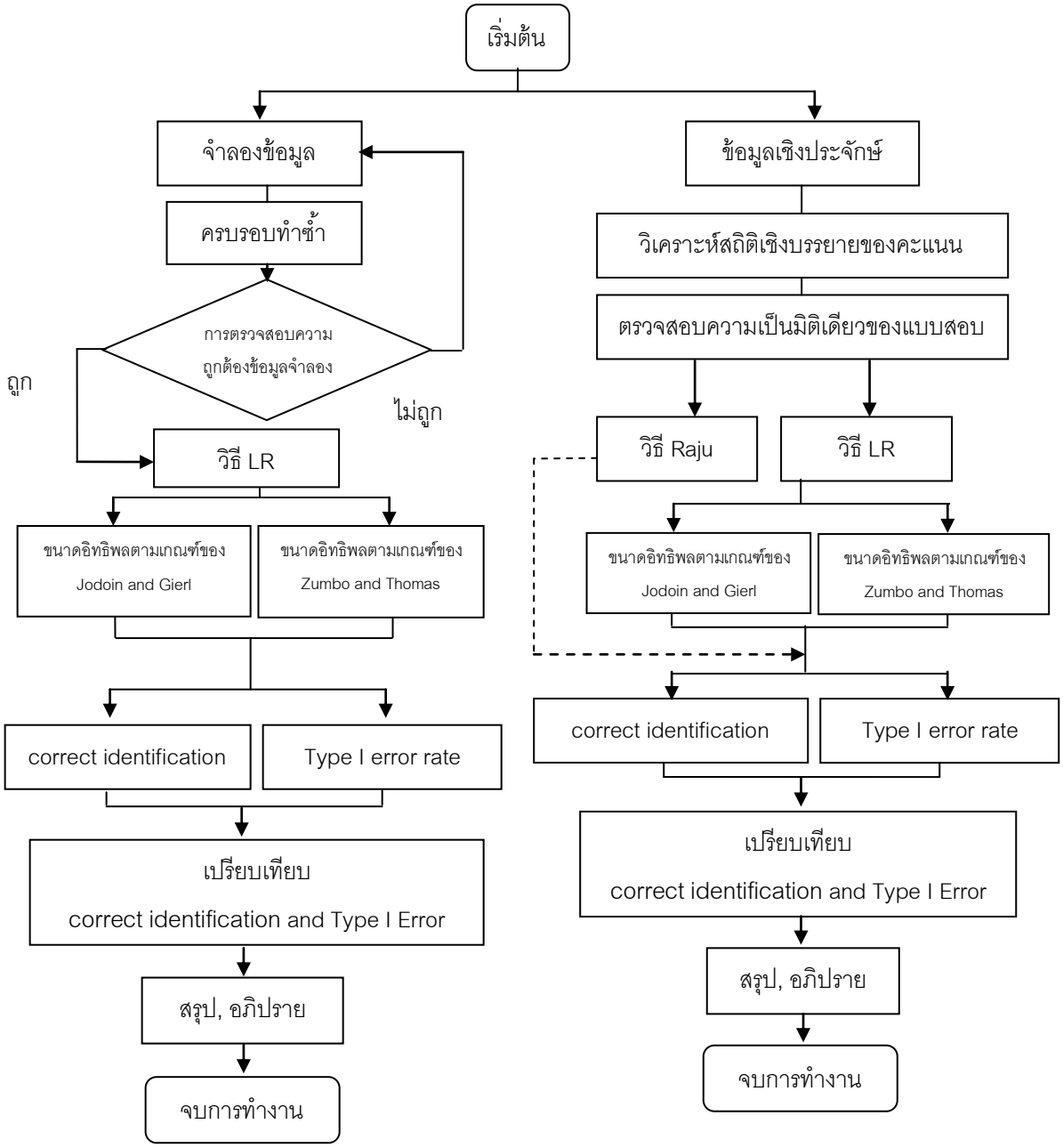
7) เปรียบเทียบประสิทธิภาพการตรวจสอบ

8) สรุปผลการวิเคราะห์ของกรณีข้อมูลเชิงประจักษ์

1.4 สรุปผลการวิเคราะห์ อภิปรายผล

1.5 เขียนรายงานการวิจัย

สรุปขั้นตอนการดำเนินการศึกษา ตามภาพที่ 3.1



ภาพที่ 3.1 ขั้นตอนการดำเนินการศึกษา

ตอนที่ 2 การจัดกระทำข้อมูลตามปัจจัยที่ศึกษา

2.1 **จำลองข้อมูล** (Simulation data)

จำลองผลการตอบข้อสอบภายใต้โมเดลของทฤษฎีการตอบสนองข้อสอบ แบบ 2 พารามิเตอร์ จำนวน 40 และ 50 ข้อ จำลองการตอบแบบสอบที่มีโครงสร้างวัดความสามารถเอกมิติ ข้อสอบทุกข้อมี 2 รายการตอบ คือ 0 และ 1 ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน ขนาดของการทำหน้าที่ต่างกัน จำนวนข้อสอบที่ทำหน้าที่ต่างกัน และความยาวของแบบสอบทั้งฉบับ มีข้อมูลที่ต้องจัดกระทำเพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจำนวน 24 เงื่อนไข (2 × 3 × 2 × 2) ทุกเงื่อนไขจำลองข้อมูลซ้ำ 25 ครั้ง

สำหรับโมเดล ของทฤษฎีการตอบสนองข้อสอบ แบบ 2 พารามิเตอร์ เป็นโมเดลที่ใช้ข้อตกลงเกี่ยวกับความเป็นเอกมิติ (Unidimensionality) และทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนแบบทวิวิภาค (dichotomously IRT Model) มีโค้งลักษณะข้อสอบ (item characteristic curves: ICC) เขียนในรูปฟังก์ชันปกติสะสม (Normal Ogive Function) และฟังก์ชันโลจิส (Logistic Function) (ศิริชัย กาญจนวาสี, 2550) ดังนี้

ฟังก์ชันทางคณิตศาสตร์ของโมเดลการตอบสนองข้อสอบแบบ 2 พารามิเตอร์

ฟังก์ชันปกติสะสม (Normal Ogive Function)

$$P_i(\theta) = \int_{-\infty}^{a_1(\theta - b_i)} \frac{1}{\sqrt{2\pi}} e^{-Z^{2}/2} dZ$$

ฟังก์ชันโลจิส (Logistic Function)

$$P_i(\theta) = \frac{1}{1 + e^{-Da_1(\theta - b_i)}}$$

โดยที่ θ คือ ระดับความสามารถของผู้ตอบข้อสอบใดๆ ที่ประมาณได้จากโมเดลตามทฤษฎีการตอบสนองข้อสอบ ปรับให้เป็นคะแนนมาตรฐานที่มีค่าเฉลี่ยเป็น0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 ซึ่ง θ มีพิสัยระหว่าง $\pm\infty$ แต่ในทางปฏิบัติส่วนใหญ่ใช้ค่า θ ระหว่าง ± 3

$P_i(\theta)$ คือ ความน่าจะเป็นที่ผู้ตอบข้อสอบที่ได้มาจากการสุ่มและมีความสามารถ θ ตอบคำถามข้อที่ I ได้ถูกต้อง

b_i คือ ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i ซึ่งเป็นค่าที่แสดงตำแหน่งของโค้งคุณลักษณะ ข้อสอบ (ICC) ตามแกนนอนบนเสกล θ ณ จุดที่โค้งมีความชันมากที่สุด หรือที่เรียกว่าจุดเปลี่ยนโค้งหรือที่ตำแหน่งต่อไปนี้

สำหรับโมเดล 1 และ 2 พารามิเตอร์ $b_i = \theta$ ที่ $P_i(\theta) = 0.5$

สำหรับโมเดล 3 พารามิเตอร์ $b_i = \theta$ ที่ $P_i(\theta) = (1+c_i)/2$

ค่า b_i มีพิสัยระหว่าง $\pm\infty$ แต่ในทางปฏิบัติส่วนใหญ่ใช้ค่า b_i ระหว่าง ±2.5

a_i คือ ค่าอำนาจจำแนกของข้อสอบข้อที่ i ซึ่งเป็นสัดส่วนต่อความชันของโค้งคุณลักษณะข้อสอบ (ICC) ณ จุดเปลี่ยนโค้ง หรือที่จุด $\theta = b_i$ ค่า a_i มีค่าสูงแสดงว่าข้อสอบข้อนั้นมีความชันที่มีค่าสูงจึงจำแนกผู้ที่มีความสามารถแตกต่างกันได้ดี ค่า a_i มีพิสัยระหว่าง $\pm\infty$ แต่ในทางปฏิบัติส่วนใหญ่ใช้ค่า a_i ระหว่าง +0.5 ถึง +2.5

e คือ ค่าคงที่ของลอการิทึมธรรมชาติ มีค่าประมาณ 2.71828

D คือ ค่าองค์ประกอบการปรับเสกลให้โลจิสติกฟังก์ชัน มีค่าใกล้เคียงกับฟังก์ชันปกติสะสม (Normal Ogive Function) มากที่สุดเท่าที่จะเป็นไปได้ มีค่า 1.70

การศึกษาครั้งนี้ จำลองข้อมูลโดย ใช้โปรแกรม WinGen (Windows software that generates IRT parameters and item responses) โดย Kyung T. Han (2007) จำลองผลการตอบข้อสอบโปรแกรม WinGen ดังกล่าวสามารถศึกษาในส่วนที่เกี่ยวข้องกับ

1) โปรแกรม WinGen สนับสนุนโมเดล IRT ต่างๆ ทั้งเอกมิติและพหุมิติของโมเดล ทฤษฎีการตอบสนองข้อสอบ (IRT) ที่ใช้กันอย่างแพร่หลาย (1) โมเดล IRT ที่ให้คะแนนแบบทวิวิภาค ชนิด 1, 2 และ 3 พารามิเตอร์ (2) โมเดลที่ไม่อิงพารามิเตอร์ (3) โมเดล IRT ที่ให้คะแนนแบบพหุวิภาค เช่น partial credit model, generalized partial credit model, graded response model, rating scale model, และ nominal response model และ (4) โมเดลแบบพหุมิติ

2) โปรแกรม WinGen สามารถสร้างเซตของค่าพารามิเตอร์ข้อสอบ (item parameters) และเซตของพารามิเตอร์ความสามารถของผู้สอบ (examinee ability parameters) เพื่อสร้างข้อมูลการตอบสนองข้อสอบตามการแจกแจงได้หลายชนิด

3) โปรแกรม WinGen ง่ายต่อการใช้งานและการเข้าถึงรายละเอียดของโปรแกรม การใช้งานบนโปรแกรม Windows มีแนวทางสำหรับการจำลองข้อมูล คือ (1) สร้างหรือการอ่านในค่าพารามิเตอร์ความสามารถของผู้สอบ (2) การสร้างหรือการอ่านค่าพารามิเตอร์ข้อสอบและ (3) การจำลองข้อมูลการตอบสนองข้อสอบ

4) โปรแกรม WinGen ขึ้นอยู่กับสภาพแวดล้อมของคอมพิวเตอร์ล่าสุด WinGen ได้รับการพัฒนาบน Microsoft .NET frameworks 2.0 คอมพิวเตอร์ที่ได้รับการพัฒนาซอฟต์แวร์ ล่าสุด โปรแกรม

สามารถทำงานบน 32 bit ชุด Windows (เช่น Windows XP) หรือ 64 bit ชุด Windows (เช่น Windows Vista) และมีการเพิ่มประสิทธิภาพการใช้งานระบบแหล่งข้อมูลได้อย่างมีประสิทธิภาพเพื่อให้โปรแกรมสามารถจัดการกับข้อมูลที่มีขนาดใหญ่โดยใช้เวลาน้อย สามารถใช้งานซอฟแวร์นี้ได้อย่างง่ายดายและถูกแปลงเป็นซอฟต์แวร์บนเว็บ

5) โปรแกรม WinGen ให้เครื่องมือวิจัยที่มีประสิทธิภาพ เพื่อการวิจัยต่างๆ ใน WinGen ข้อมูลการจำลองแบบสามารถจำลองได้ถึง 1,000,000 ชุด และไฟล์คำสั่ง syntax files สำหรับโปรแกรม IRT อื่นๆ เช่น PARSCALE (Muraki and Bock, 2003) BILOG-MG (Zimowski, Muraki, Mislevy, and Bock, 2003) และ MULTILOG (Thissen, 2003) เช่น การจำลองข้อสอบที่ทำหน้าที่ต่างกัน (DIF)

2.2 การตรวจสอบความถูกต้องของข้อมูลตามเงื่อนไขของปัจจัยที่ศึกษา

การตรวจสอบความถูกต้องของข้อมูลที่ได้จากการจำลองข้อมูลตามเงื่อนไขในแบบสอบจาก Harwell, Hsu and Kirisci (1996) ที่ศึกษาเกี่ยวกับการทดลองซ้ำพบว่าภายใต้บริบทโมเดลของ IRT เป็นฐาน เมื่อใช้กลุ่มตัวอย่าง 1,000 คน ขึ้นไป ควรมีการทำซ้ำอย่างน้อย 20 ครั้ง สอดคล้องกับผลการศึกษาของ Gómez-Benito และคณะ (2009) ซึ่งใช้ข้อมูลจำลองในการวิจัยโดยมีการการทดลองซ้ำ 25 ครั้ง ผลการตอบข้อสอบที่สร้างขึ้นตามเงื่อนไขปัจจัยหลักซึ่งเป็นตัวแปรที่ต้องการศึกษาโดยใช้ The simulation algorithm สำหรับเมตริกซ์การตอบสนองรายข้อเริ่มต้นที่การกำหนดรูปแบบการแจกแจงความสามารถ (ability distribution) และกำหนดค่าพารามิเตอร์ของข้อสอบ (item parameter) จำลองข้อมูลโดยใช้โปรแกรม WinGen (Windows software that generates IRT parameters and item responses) แต่ละเงื่อนไขของการทดลองซ้ำ 25 ครั้ง

การตรวจสอบความถูกต้องของข้อมูลจำลองว่าเป็นไปตามเงื่อนไขของปัจจัยที่กำหนดขึ้น โดยใช้โปรแกรม DIFAS (Differential item Functioning analysis system) โดย Randall D. Penfield (2007) หลังจากนั้นจึงนำข้อมูลดังกล่าวมาวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบตามขั้นตอน

ตอนที่ 3 การจำลองข้อมูล

จำลองข้อมูลตามเงื่อนไขปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ด้วย โปรแกรม WinGen (Windows software that generates IRT parameters and item responses) โดย Kyung T. Han (2007) จำลองข้อมูลเป็นกลุ่มข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป กลุ่มข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป และกลุ่มข้อสอบที่ไม่ทำหน้าที่ต่างกัน ตามขั้นตอนดังนี้

3.1 กำหนดจำนวนผู้สอบเท่ากับ 2,000 คน ให้มีผลการตอบข้อสอบแต่ละข้อเป็นแบบทวิวิภาค ความสามารถของผู้สอบมีการแจกแจงแบบปกติ

3.2 กำหนดขนาดของการทำหน้าที่ต่างกัน คือ 0.1, 0.2 และ 0.4

3.3 กำหนดรูปแบบของข้อสอบทำหน้าที่ต่างกัน และความยาวของแบบสอบทั้งฉบับ

3.3.1 การจำลองข้อมูลกรณีข้อสอบ 40 ข้อ

1) สุ่มตัวอย่างข้อสอบมา 40 ข้อ มีสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 10 มี 4 ข้อ (กำหนดเป็นข้อที่ 1 ถึงข้อที่ 4) ที่เป็นกลุ่มข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป ที่เหลือเป็นกลุ่มข้อสอบที่ทำหน้าที่ไม่ต่างกัน มี 36 ข้อ (กำหนดเป็นข้อที่ 5 ถึงข้อที่ 40) ให้ข้อสอบที่ถูกกำหนดให้ทำหน้าที่ต่างกันนั้นมีขนาดของการทำหน้าที่ต่างกัน เท่ากับ 0.1, 0.2 และ 0.4 ตามลำดับ

2) สุ่มตัวอย่างข้อสอบมา 40 ข้อ โดยมีสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็น ร้อยละ 10 มี 4 ข้อ (กำหนดเป็นข้อที่ 1 ถึงข้อที่ 4) ที่เป็นกลุ่มรูปแบบของข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป ที่เหลือเป็นข้อสอบกลุ่มข้อสอบที่ทำหน้าที่ไม่ต่างกัน มี 36 ข้อ (กำหนดเป็นข้อที่ 5 ถึงข้อที่ 40) ให้ข้อสอบที่ถูกกำหนดให้ทำหน้าที่ต่างกันนั้นมีขนาดของการทำหน้าที่ต่างกัน เท่ากับ 0.1, 0.2 และ 0.4 ตามลำดับ

3) สุ่มตัวอย่างข้อสอบมา 40 ข้อ โดยมีสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 20 มี 8 ข้อ (กำหนดเป็นข้อที่ 1 ถึงข้อที่ 8) เป็นกลุ่มข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป ที่เหลือเป็นข้อสอบกลุ่มข้อสอบที่ทำหน้าที่ไม่ต่างกัน มี 32 ข้อ (กำหนดเป็นข้อที่ 9 ถึงข้อที่ 40) ให้ข้อสอบที่ถูกกำหนดให้ทำหน้าที่ต่างกันนั้นมีขนาดของการทำหน้าที่ต่างกัน เท่ากับ 0.1, 0.2 และ 0.4 ตามลำดับ

4) สุ่มตัวอย่างข้อสอบมา 40 ข้อ โดยมีสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 20 มี 8 ข้อ (กำหนดเป็นข้อที่ 1 ถึงข้อที่ 8) ที่เป็นกลุ่มรูปแบบของข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป ที่เหลือเป็นข้อสอบกลุ่มข้อสอบที่ทำหน้าที่ไม่ต่างกัน มี 32 ข้อ (กำหนดเป็นข้อที่ 9 ถึงข้อที่ 40) ให้ข้อสอบที่ถูกกำหนดให้ทำหน้าที่ต่างกันนั้นมีขนาดของการทำหน้าที่ต่างกัน เท่ากับ 0.1, 0.2 และ 0.4 ตามลำดับ

3.3.2 การจำลองข้อมูลกรณีข้อสอบ 50 ข้อ

1) สุ่มตัวอย่างข้อสอบมา 50 ข้อ โดยมีสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 10 มี 5 ข้อ กำหนดให้เป็นข้อที่ 1 ถึงข้อที่ 5 เป็นกลุ่มข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป เหลือเป็นกลุ่มข้อสอบที่ทำหน้าที่ไม่ต่างกัน มี 45 ข้อ (กำหนดเป็นข้อที่ 6 ถึงข้อที่ 50) ให้ข้อสอบที่ถูกกำหนดให้ทำหน้าที่ต่างกันนั้นมีขนาดของการทำหน้าที่ต่างกัน เท่ากับ 0.1, 0.2 และ 0.4 ตามลำดับ

2) สุ่มตัวอย่างข้อสอบมา 50 ข้อ โดยมีสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 10 มี 5 ข้อ (กำหนดเป็นข้อที่ 1 ถึงข้อที่ 5) ที่เป็นกลุ่มรูปแบบของข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป ที่เหลือเป็นข้อสอบกลุ่มข้อสอบที่ทำหน้าที่ไม่ต่างกัน มี 45 ข้อ (กำหนดเป็นข้อที่ 6 ถึงข้อที่ 50) ให้ข้อสอบที่ถูกกำหนดให้ทำหน้าที่ต่างกันนั้นมีขนาดของการทำหน้าที่ต่างกัน เท่ากับ 0.1, 0.2 และ 0.4 ตามลำดับ

3) สุ่มตัวอย่างข้อสอบมา 50 ข้อ โดยมีสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 20 มี 10 ข้อ (กำหนดเป็นข้อที่ 1 ถึงข้อที่ 10) เป็นกลุ่มข้อสอบที่ทำหน้าที่ต่างกันแบบเอก

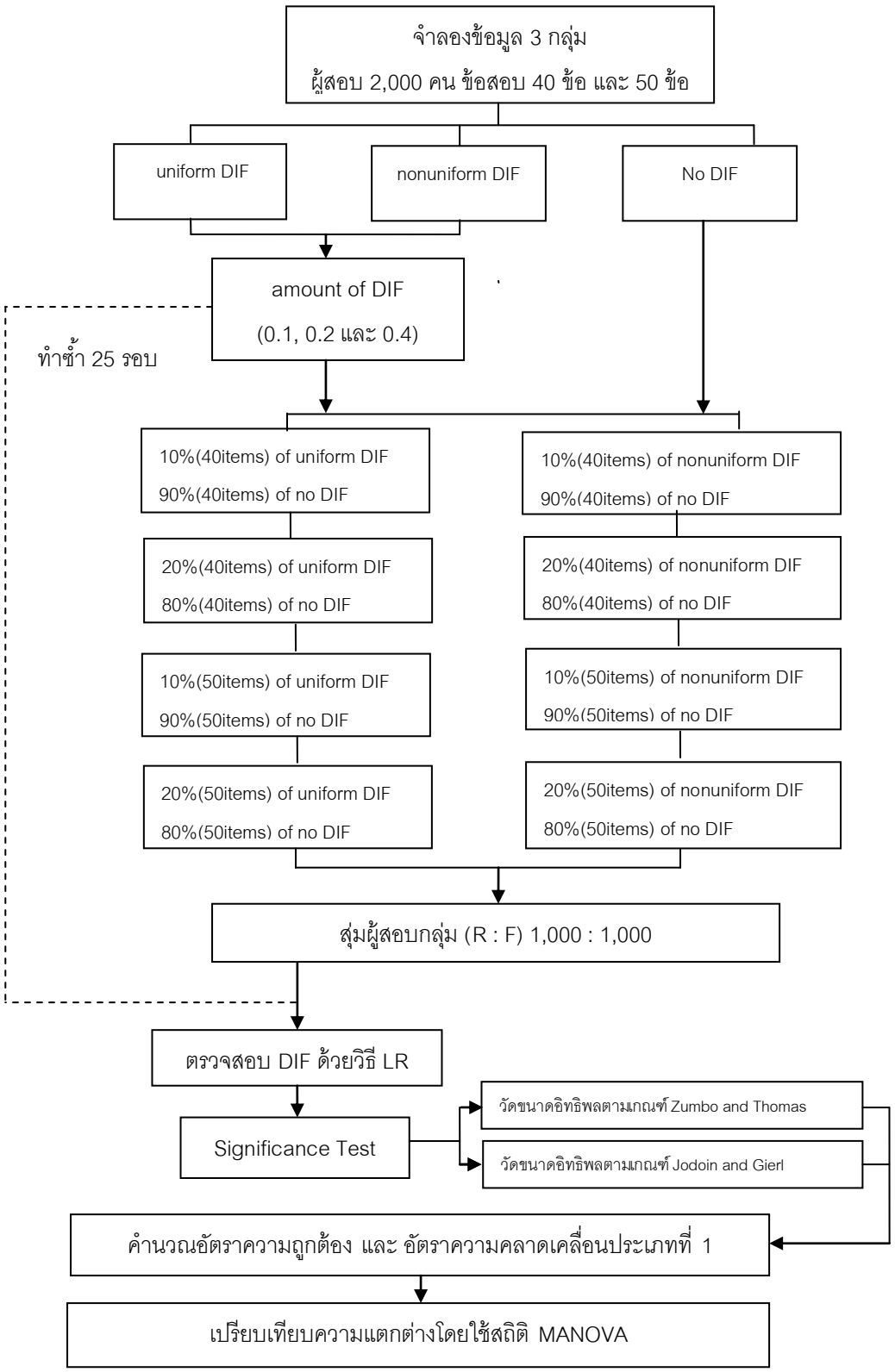
รูป ที่เหลือเป็นข้อสอบกลุ่มข้อสอบที่ทำหน้าที่ไม่ต่างกัน มี 40 ข้อ (กำหนดเป็นข้อที่ 11 ถึงข้อที่ 50) ให้ข้อสอบที่ถูกกำหนดให้ทำหน้าที่ต่างกันนั้นมีขนาดของการทำหน้าที่ต่างกัน เท่ากับ 0.1, 0.2 และ 0.4 ตามลำดับ

4) สุ่มตัวอย่างข้อสอบมา 50 ข้อ โดยมีสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 20 มี 10 ข้อ (กำหนดเป็นข้อที่ 1 ถึงข้อที่ 10) เป็นกลุ่มรูปแบบของข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป ที่เหลือเป็นข้อสอบกลุ่มข้อสอบที่ทำหน้าที่ไม่ต่างกัน มี 40 ข้อ (กำหนดเป็นข้อที่ 11 ถึงข้อที่ 50) ให้ข้อสอบที่ถูกกำหนดให้ทำหน้าที่ต่างกันนั้นมีขนาดของการทำหน้าที่ต่างกัน 0.1, 0.2 และ 0.4 ตามลำดับ

3.4 ขนาดกลุ่มตัวอย่าง (sample size) สุ่มขนาดกลุ่มตัวอย่างกลุ่มอ้างอิงต่อกลุ่มเปรียบเทียบ (R: F) ในสัดส่วนที่เท่าๆ กัน คือมีจำนวนผู้สอบจำนวน 2,000 คนคงที่ แบ่งออกเป็นผู้สอบกลุ่มอ้างอิง (R) จำนวน 1,000 คน และผู้สอบกลุ่มเปรียบเทียบ (F) จำนวน 1,000 คน

3.5 ทำซ้ำ 25 รอบ ในแต่ละเงื่อนไข

การจำลองข้อมูลดังภาพที่ 3.2 และ แผนผังของการจำลองข้อมูล ดังภาพที่ 3.3



ภาพที่ 3.2 ขั้นตอนการจำลองข้อมูล

รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน	ความยาวของแบบสอบทั้งฉบับ	จำนวนข้อสอบที่ทำหน้าที่ต่างกัน (คิดเป็นร้อยละ)	ขนาดของการทำหน้าที่ต่างกัน	เงื่อนไข
อเนกรูป	40 ข้อ	10%	0.1	เงื่อนไขที่ 1
			0.2	เงื่อนไขที่ 2
			0.4	เงื่อนไขที่ 3
		20%	0.1	เงื่อนไขที่ 4
			0.2	เงื่อนไขที่ 5
			0.4	เงื่อนไขที่ 6
	50 ข้อ	10%	0.1	เงื่อนไขที่ 7
			0.2	เงื่อนไขที่ 8
			0.4	เงื่อนไขที่ 9
		20%	0.1	เงื่อนไขที่ 10
			0.2	เงื่อนไขที่ 11
			0.4	เงื่อนไขที่ 12
เอกรูป	40 ข้อ	10%	0.1	เงื่อนไขที่ 13
			0.2	เงื่อนไขที่ 14
			0.4	เงื่อนไขที่ 15
		20%	0.1	เงื่อนไขที่ 16
			0.2	เงื่อนไขที่ 17
			0.4	เงื่อนไขที่ 18
	50 ข้อ	10%	0.1	เงื่อนไขที่ 19
			0.2	เงื่อนไขที่ 20
			0.4	เงื่อนไขที่ 21
		20%	0.1	เงื่อนไขที่ 22
			0.2	เงื่อนไขที่ 23
			0.4	เงื่อนไขที่ 24

ภาพที่ 3.3 แผนผังของการจำลองข้อมูล

ตอนที่ 4 การวิเคราะห์ข้อมูล

การวิเคราะห์เพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค ด้วยวิธีถดถอยโลจิสติก ผู้วิจัยดำเนินการดังนี้

4.1 การวิเคราะห์ค่าพารามิเตอร์ใช้โมเดลภายใต้ทฤษฎีการตอบสนองข้อสอบ โดย ใช้โปรแกรม MULTILOG - MG ที่มีคุณสมบัติในการประมาณค่าพารามิเตอร์ของข้อสอบและผู้สอบ และมีความเหมาะสมสำหรับแบบแผนการตอบข้อสอบที่ มีการ ให้คะแนนแบบทวิวิภาค พัฒนาขึ้นโดย David Thissen โมเดลที่เหมาะสมสำหรับนำมาวิเคราะห์ ด้วยโปรแกรม MULTILOG - MG ได้แก่ โมเดลที่มีการตรวจให้คะแนนแบบ ทวิวิภาค Grade responses model (Samejima, 1969) Nominal responses model (Bock, 1972) และ Multiple–choice model (Thissen and Steinberg, 1984) มีข้อจำกัดที่ การวิเคราะห์ไม่สามารถแสดงผลทั้งพารามิเตอร์ของข้อสอบและพารามิเตอร์ความสามารถของผู้สอบได้พร้อมกันภายใต้ คำสั่งให้ประมวลผล (run) เพียงครั้งเดียว ผลการวิเคราะห์จะแสดงพารามิเตอร์ของข้อสอบหรือพารามิเตอร์ความสามารถของผู้สอบขึ้นอยู่กับการเลือกประเภทของการวิเคราะห์

4.2 ตรวจสอบความเป็นเอกมิติของแบบสอบในข้อมูลเชิงประจักษ์ โดยการวิเคราะห์องค์ประกอบ (Factor Analysis) ด้วยโปรแกรม SPSS

4.3 การวิเคราะห์คุณภาพของข้อสอบใช้โปรแกรม SPSS การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติกใช้การเขียนคำสั่งและประมวลผลด้วยโปรแกรม R รายละเอียดดังต่อไปนี้

4.3.1 การวิเคราะห์สมการถดถอยโลจิสติก เพื่อศึกษาผลของตัวแปรต้นที่มีต่อตัวแปรตาม ซึ่งตัวแปรตามเป็นได้ทั้ง ตัวแปรทวิวิภาค (Dichotomous variable) หรือตัวแปรพหุวิภาค (Polytomous variable) ความสัมพันธ์ระหว่างตัวแปรต้นกับโอกาสของการเกิดเหตุการณ์ของตัวแปรตามไม่ใช่ความสัมพันธ์เชิงเส้นตรงแต่มีลักษณะเป็นฟังก์ชันโลจิสติก (Logistic Function)

โมเดลของการวิเคราะห์ด้วยวิธีถดถอยโลจิสติก คือ

$$P(y=1/X)=\frac{e^{z}}{1+e^{z}}$$

เมื่อ $P(y=1/X)$ แทน ความน่าจะเป็นในการตอบข้อสอบถูกของผู้สอบ

Z แทน ผลรวมเชิงเส้นของตัวแปรทำนาย

โมเดลสมการถดถอยโลจิสติก คือ

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

เมื่อ X แทน ระดับความสามารถของผู้สอบเป็นคะแนนสอบ

G แทน สมาชิกของกลุ่มผู้สอบคือกลุ่มอ้างอิงหรือกลุ่มเปรียบเทียบ

XG แทน ปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับกลุ่มผู้สอบ

β_0 แทน ค่า Intercept (จุดตัดแกน)

β_1 แทน พารามิเตอร์ที่เกี่ยวข้องกับความสามารถที่แตกต่างรายข้อ

β_2 แทน พารามิเตอร์ที่เกี่ยวข้องกับความสามารถที่แตกต่างระหว่างกลุ่มรายข้อ

β_3 แทน พารามิเตอร์ที่เกี่ยวข้องกับปฏิสัมพันธ์ระหว่างกลุ่มกับความสามารถ

ระดับความสามารถของผู้สอบถูกกำหนดตามคะแนนรวมของแบบสอบ (Total test score) การเป็นสมาชิกของกลุ่มผู้สอบ คือ กลุ่มอ้างอิง (R) หรือกลุ่มเปรียบเทียบ (F) การตัดสินข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป (Uniform DIF) หรืออเนกรูป (nonuniform DIF) พิจารณาจากค่าพารามิเตอร์ของ β_2 กับ β_3 จากโมเดลดังกล่าวข้างต้น (Gómez-Benito, Hidalgo and Padilla, 2009) กล่าวคือ

ถ้า $\beta_2 = \beta_3 = 0$ แสดงว่าข้อสอบทำหน้าที่ไม่ต่างกัน

ถ้า $\beta_2 \neq 0$ และ $\beta_3 = 0$ แสดงว่าข้อสอบทำหน้าที่ต่างกันแบบเอกรูป

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

สำหรับการพิจารณาการเป็นสมาชิกของกลุ่ม กำหนดให้ G=1 เมื่อผู้สอบอยู่ในกลุ่มอ้างอิง G=0 เมื่อผู้สอบอยู่ในกลุ่มเปรียบเทียบ ใช้สถิติ χ^2 ทดสอบสมมติฐาน การวิจัยครั้งนี้ผู้วิจัยใช้เกณฑ์ตัดสินข้อสอบที่ทำหน้าที่ต่างกันที่ระดับนัยสำคัญ .05

4.3.2 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซล พัฒนาโดย Mantel-Haenszel (Camilli and Shepard, 1994; citing Mantel-Haenszel, 1959) เดิมทดสอบอัตราส่วนเปรียบเทียบด้วยไค -สแควร์ ต่อมาฮอนแลนด์ (Holland, 1985; citing Holland and Thayer, 1988) ได้นำวิธีแมนเทล -แฮนส์เซล มาประยุกต์ใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับหน่วยงานการบริการทดสอบทางการศึกษาของประเทศสหรัฐอเมริกา เป็นวิธีที่ได้รับความนิยมเป็นที่ยอมรับจากนักวิจัยอย่างกว้างขวางหลักการคำนวณสามารถนำไปใช้ ตรวจสอบการทำหน้าที่

ต่างกันของข้อสอบได้ง่าย ขั้นตอนการคำนวณไม่ซับซ้อน มีการทดสอบทางสถิติแบบนอนพาราเมตริก ไม่จำเป็นต้องใช้โมเดลประมาณค่า (ศิริชัย กาญจนวาสี, 2550)

หลักการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยเปรียบเทียบผลการตอบข้อสอบระหว่างผู้สอบกลุ่มเปรียบเทียบกับกลุ่มอ้างอิงในทุกระดับความสามารถของผู้สอบกลุ่มย่อยสองกลุ่มที่มีระดับความสามารถเท่ากันใช้คะแนนรวมของการสอบเป็นเกณฑ์การจับคู่กลุ่มผู้สอบได้ผลวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบเป็นรายข้อ เมื่อจับคู่กลุ่มผู้สอบแล้วจะนำข้อมูลผลการตอบข้อสอบมาจัดลงในตารางการณ์จรแบบ 2 x 2 (กลุ่มผู้สอบ 2 กลุ่ม x ผลการตอบ 2 แบบ) โดยที่ตารางการณ์จร 1 ตาราง แทนคะแนนรวม 1 ระดับ ดังนั้น ถ้ามีคะแนนรวมของกลุ่มผู้สอบทั้งสิ้น k ระดับ จะต้องสร้างตารางการณ์จรแบบ 2 x 2 ทั้งหมด k ตาราง สำหรับตารางการณ์จรแบบ 2 x 2 ของข้อสอบแต่ละข้อที่มีคะแนนรวมระดับ j ใช้สถิติแมนเทล-แฮนส์เซลไค-สแควร์ ที่ระดับชั้นความเป็นอิสระเท่ากับ 1 (df = 1) ในการทดสอบนัยสำคัญของสมมติฐาน

ตอนที่ 5 การเปรียบเทียบประสิทธิภาพของผลการตรวจสอบ DIF

ประสิทธิภาพของการตรวจสอบ พิจารณาจากอัตราความถูกต้อง (correct identification) และอัตราความคลาดเคลื่อนประเภทที่ 1 (Type I Error rate) ซึ่งการวิเคราะห์อัตราความถูกต้อง (1-β) กับความคลาดเคลื่อนประเภทที่ 2 (β) และการวิเคราะห์ระดับความเชื่อมั่น (1-α) กับความคลาดเคลื่อนประเภทที่ 1 (α) เป็นค่าดัชนีที่มีเสกลผกผันกัน ดังนั้นจึงพิจารณาเพียงอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 เท่านั้น (ศิริชัย กาญจนวาสี, 2550)

5.1 อัตราความถูกต้อง (correct identification: CI) คำนวณจากจำนวนของข้อสอบที่ตรวจสอบได้ถูกต้องว่าทำหน้าที่ต่างกัน ต่อจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งหมดในแบบสอบ คำนวณเป็นค่าร้อยละ

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

เมื่อ CI แทน อัตราความถูกต้อง

n_1 แทน จำนวนข้อสอบที่ตรวจสอบได้ถูกต้องว่า DIF

N_1 แทน จำนวนข้อสอบที่ DIF ทั้งหมดที่ตรวจสอบด้วยวิธีเกณฑ์

5.2 อัตราความคลาดเคลื่อนประเภทที่ 1 (type I error rate: TE) เป็นการระบุผิดพลาดว่าข้อสอบทำหน้าที่ต่างกัน (False Positive: FP) ทั้งที่ความเป็นจริงข้อสอบทำหน้าที่ไม่ต่างกัน คำนวณจาก

สัดส่วนของจำนวนข้อสอบที่ตรวจสอบผิดพลาดว่าทำหน้าที่ต่างกันทั้งที่ในความเป็นจริงข้อสอบไม่ได้ทำหน้าที่ต่างกัน คำนวณเป็นค่าร้อยละ

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

เมื่อ E_1 แทน อัตราความคลาดเคลื่อนประเภทที่ 1

n_2 แทน จำนวนข้อสอบที่ระบุผิดว่า DIF

N_2 แทน จำนวนข้อสอบที่ไม่ DIF ทั้งหมดที่ตรวจสอบด้วยวิธีเกณฑ์

5.3 การเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1

การเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ตัวแปรอิสระ 5 ตัว คือ วิธีการตรวจสอบ รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน ขนาดของการทำหน้าที่ต่างกัน จำนวนข้อสอบที่ทำหน้าที่ต่างกัน และ ความยาวของแบบสอบทั้งฉบับ ทดสอบความแตกต่างของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีการตรวจสอที่มีผลต่ออัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 โดยสถิติการวิเคราะห์ความแปรปรวนหลายตัวแปร (Multivariate analysis of variance; MANOVA) ที่ระดับนัยสำคัญ .001 โดยกำหนดการวิเคราะห์ให้มีปฏิสัมพันธ์ระหว่างตัวแปรอิสระไม่เกินอันดับที่สอง ถ้าผลการทดสอบมีนัยสำคัญทางสถิติจะทดสอบผลระหว่างกลุ่ม (Test of between-subjects effects) ของตัวแปรตามที่ระดับนัยสำคัญ .001 แล้วทดสอบผลย่อย (Simple effect) ภายใต้ตัวแปรที่ศึกษา ระดับนัยสำคัญ .001 และทดสอบภายหลังด้วยวิธีของเชฟเฟ (Scheffé) โดยใช้ระดับนัยสำคัญระดับเดียวกับการทดสอบผลย่อย

ผลการทดสอบนัยสำคัญจากการวิเคราะห์ความแปรปรวนหลายตัวแปร พิจารณาเฉพาะประเด็นที่ตอบคำถามการวิจัย คือ ผลการทดสอบปฏิสัมพันธ์สองทางและสามทาง ระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย กับวิธีการตรวจสอบ การทดสอบผลย่อยภายใต้กรอบการวิเคราะห์ ดังนี้

1. เมื่อผลการทดสอบปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย กับ วิธีการตรวจสอบ มีนัยสำคัญทางสถิติจะทดสอบผลย่อย 2 กรณี ดังนี้

1.1 ทดสอบผลย่อยของวิธีการตรวจสอบ ในแต่ละเงื่อนไขของปัจจัยที่แปรเปลี่ยน

1.2 ทดสอบผลย่อยของปัจจัยที่แปรเปลี่ยน ในแต่ละวิธีการตรวจสอบ

2. เมื่อผลการทดสอบปฏิสัมพันธ์สามทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย กับ วิธีการตรวจสอบมีนัยสำคัญทางสถิติจะทดสอบผลย่อย 2 กรณี ดังนี้

2.1 ทดสอบผลย่อยของวิธีการตรวจสอบ ในแต่ละเงื่อนไขของปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน

2.2 ทดสอบผลย่อยของปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน ในแต่ละวิธีการตรวจสอบ

5.4 กำหนดเกณฑ์ที่ใช้เปรียบเทียบประสิทธิภาพในการทำหน้าที่ต่างกันของข้อสอบ ภายใต้วิธีถดถอยโลจิสติก ระหว่างขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl และขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas วิธีใดเกิดอัตราความถูกต้องในการตรวจสอบสูงและเกิดอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำ แสดงว่ามีประสิทธิภาพสูงสุดในการทำหน้าที่ต่างกันของข้อสอบ

บทที่ 4

ผลการวิเคราะห์ข้อมูล

การวิจัยครั้งนี้มุ่งศึกษาประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้านอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการจำลองข้อมูลและข้อมูลเชิงประจักษ์ภายใต้วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl (2001) และการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas (1997) มีวัตถุประสงค์เฉพาะของการวิจัย ดังนี้

1) เพื่อเปรียบเทียบอัตราคมคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีรูปแบบการให้คะแนนแบบทวิวิภาคโดยการจำลอง ข้อมูล ในวิธีถดถอยโลจิสติ กระหว่างการวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl และเกณฑ์ Zumbo and Thomas **ภายใต้เงื่อนไขเดียวกัน**ของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน ขนาดของการทำหน้าที่ต่างกัน จำนวนข้อสอบที่ทำหน้าที่ต่างกันและความยาวของแบบสอบทั้งฉบับ

2) เพื่อเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีรูปแบบการให้คะแนนแบบทวิวิภาค โดยการจำลอง ข้อมูลในวิธีถดถอยโลจิสติก ด้วยการวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl และเกณฑ์ Zumbo and Thomas **ภายใต้เงื่อนไขต่างกัน**ของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน ขนาดของการทำหน้าที่ต่างกัน จำนวนข้อสอบที่ทำหน้าที่ต่างกัน และความยาวของแบบสอบทั้งฉบับ

3) เพื่อเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีรูปแบบการให้คะแนนแบบทวิวิภาค โดยข้อมูลเชิงประจักษ์ในวิธีถดถอยโลจิสติ ก ระหว่างการวัด ขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl และเกณฑ์ Zumbo and Thomas

ข้อมูลเชิงประจักษ์ นำมาจาก “โครงการ สอบระดับชาติของสถาบันแห่งหนึ่ง ประจำปี พ.ศ. 2552 มีรูปแบบการตรวจให้คะแนนแบบทวิวิภาค ระดับชั้นประถมศึกษาปีที่ 6 จากโรงเรียนทุกสังกัด วิชาวิทยาศาสตร์ และวิชาคณิตศาสตร์ ”

ผู้วิจัยขอนำเสนอผลการวิเคราะห์ข้อมูลเป็น 4 ตอน ดังนี้

ตอนที่ 1 การคำนวณค่าเฉลี่ยของอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1

ตอนที่ 2 ผลการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างการวัดขนาดอิทธิพลตามเกณฑ์ที่ศึกษา

ตอนที่ 3 สรุปผลการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1

ตอนที่ 4 ผลการศึกษาในกรณีข้อมูลเชิงประจักษ์

เพื่อความสะดวกในการนำเสนอผลการวิเคราะห์ข้อมูล ผู้วิจัยจึงขอกำหนดสัญลักษณ์ ดังนี้

LR หมายถึง วิธีถดถอยโลจิสติก

LRs หมายถึง ผล การทดสอบระดับนัยสำคัญ

LRz หมายถึง ผลการวัดขนาดอิทธิพล เกณฑ์ Zumbo and Thomas

LRj หมายถึง ผลการวัดขนาดอิทธิพล เกณฑ์ Jodoin and Gierl

TYPE หมายถึง รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน

TYPE1 หมายถึง รูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป

TYPE2 หมายถึง รูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป

%DIF หมายถึง จำนวนข้อสอบที่ทำหน้าที่ต่างกัน

10%DIF หมายถึง จำนวนข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 10

20%DIF หมายถึง จำนวนข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 20

Length หมายถึง ความยาวของแบบสอบทั้งฉบับ

Length40 หมายถึง ความยาวของแบบสอบทั้งฉบับ 40 ข้อ

Length50 หมายถึง ความยาวของแบบสอบทั้งฉบับ 50 ข้อ

Amount หมายถึง ขนาดของการทำหน้าที่ต่างกันของข้อสอบ

Amount 0.1 หมายถึง ขนาดของการทำหน้าที่ต่างกันของข้อสอบขนาดเล็กน้อย

Amount 0.2 หมายถึง ขนาดของการทำหน้าที่ต่างกันของข้อสอบขนาดปานกลาง

Amount 0.4 หมายถึง ขนาดของการทำหน้าที่ต่างกันของข้อสอบขนาดใหญ่

ตอนที่ 1 การคำนวณค่าเฉลี่ยของอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1

การศึกษาข้อมูล จำลอง ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน จำนวนทั้งสิ้น 24 เงื่อนไข รายละเอียดดังตารางที่ 4.1

ตารางที่ 4.1 ภาพรวมของการจำลองข้อมูลจำแนกตามปัจจัยและเงื่อนไขของปัจจัยที่แปรเปลี่ยน

ปัจจัยที่ศึกษา	เงื่อนไข
รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน (DIF type)	1. แบบเอกรูป (uniform)
	2. แบบอเนกรูป (nonuniform)
ขนาดของการทำหน้าที่ต่างกัน (Amount of DIF)	1. ขนาด 0.1
	2. ขนาด 0.2
	3. ขนาด 0.4
จำนวนข้อสอบที่ทำหน้าที่ต่างกัน (the number of items with DIF)	1. จำนวน 10%
	2. จำนวน 20%
ความยาวของแบบสอบทั้งฉบับ (Test length)	1. ความยาว 40 ข้อ
	2. ความยาว 50 ข้อ

ผู้วิจัยนำเสนอค่าพารามิเตอร์ของข้อมูลจากการจำลองตามเงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน มี 2 เงื่อนไข คือ แบบเอกรูป และแบบอเนกรูป ปัจจัยขนาดของการทำหน้าที่ต่างกัน มี 3 เงื่อนไข คือ ขนาด 0.1, 0.2 และ 0.4 ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน มี 2 เงื่อนไข คือทั้งฉบับคิดเป็นร้อยละ 10 และร้อยละ 20 ปัจจัยความยาวของแบบสอบทั้งฉบับ มี 2 เงื่อนไข คือ ความยาว 40 ข้อ และ 50 ข้อ มีข้อมูลทั้งหมดที่ต้องจัดกระทำเพื่อตรวจสอบการทำหน้าที่ต่างกันจำนวน 24 เงื่อนไข (2 × 3 × 2 × 2) ทุกเงื่อนไขจำลองข้อมูลซ้ำ 25 ครั้ง

1.1 การตรวจสอบคุณภาพเบื้องต้นของข้อมูลจำลอง

การจำลองข้อมูล DIF แบบอเนกรูปในแบบสอบที่ยาว 40 ข้อ พารามิเตอร์ผู้สอบ (θ) มีค่าระหว่าง -4.083 ถึง 3.615 พารามิเตอร์ a มีค่าระหว่าง 0.537 ถึง 1.675 และพารามิเตอร์ b มีค่าระหว่าง -2.910 ถึง 2.355 ในแบบสอบที่ยาว 50 ข้อ พารามิเตอร์ของผู้สอบ (θ) มีค่าระหว่าง -4.269 ถึง 4.113 พารามิเตอร์ a มีค่าระหว่าง 0.550 ถึง 2.024 และพารามิเตอร์ b มีค่าระหว่าง -3.126 ถึง 2.453

การจำลองข้อมูล DIF แบบเอกรูปในแบบสอบที่ยาว 40 ข้อ พารามิเตอร์ผู้สอบ (θ) มีค่าระหว่าง -4.655 ถึง 3.435 พารามิเตอร์ a มีค่าระหว่าง 0.587 ถึง 1.701 และพารามิเตอร์ b มีค่าระหว่าง -2.272 ถึง 2.843 สำหรับแบบสอบที่ยาว 50 ข้อ ค่าพารามิเตอร์ของผู้สอบ (θ) มีค่าระหว่าง -4.463 ถึง 3.472 ค่าพารามิเตอร์ a มีค่าระหว่าง 0.518 ถึง 1.774 และค่าพารามิเตอร์ b มีค่าระหว่าง -2.961 ถึง 3.459

ค่าพารามิเตอร์ในทฤษฎีการตอบสนองข้อสอบ คือพารามิเตอร์ของผู้สอบ (person parameter) หรือคุณลักษณะของผู้สอบ (θ) ส่วนใหญ่มีค่าอยู่ในช่วง -3.0 ถึง +3.0 กรณีที่แบบสอบที่มีข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 10 ที่มีค่าพารามิเตอร์ของผู้สอบ (θ) ค่อนข้างต่ำ ส่วนพารามิเตอร์ข้อสอบ (Item parameter) คือความยาก (b) และอำนาจจำแนก รายละเอียดดังตารางที่ 4.2

ตารางที่ 4.2 ผลการตรวจสอบคุณภาพข้อมูลจำลองตามจำแนกตามปัจจัยที่ศึกษา 24 เงื่อนไข

ปัจจัยที่แปรเปลี่ยน				**ค่าพารามิเตอร์ของผู้เข้าสอบและข้อสอบ** (min, max)		
DIF type	Test length	%DIF	Amount of DIF	θ	Parameter a	Parameter b
อเนกรูป	40	10	0.1	-3.564 , 2.968	0.743 , 1.657	-2.346 , 2.277
อเนกรูป	40	10	0.2	-3.475 , 3.256	0.609 , 1.554	-1.965 , 2.223
อเนกรูป	40	10	0.4	-4.083 , 3.615	0.701 , 1.675	-2.014 , 2.355
อเนกรูป	40	20	0.1	-3.955 , 3.322	0.537 , 1.465	-2.036 , 2.290
อเนกรูป	40	20	0.2	-3.696 , 3.495	0.628 , 1.476	-2.910 , 2.057
อเนกรูป	40	20	0.4	-3.484 , 3.546	0.718 , 1.389	-1.614 , 1.458
อเนกรูป	50	10	0.1	-3.231 , 2.714	0.618 , 1.596	-2.807 , 2.453
อเนกรูป	50	10	0.2	-4.269 , 3.245	0.694 , 2.024	-1.528 , 2.302
อเนกรูป	50	10	0.4	-3.784 , 4.113	0.550 , 1.536	-2.208 , 2.422
อเนกรูป	50	20	0.1	-3.157 , 2.639	0.742 , 1.542	-2.127 , 1.829
อเนกรูป	50	20	0.2	-3.075 , 2.628	0.587 , 1.476	-3.126 , 2.042
อเนกรูป	50	20	0.4	-3.542 , 3.149	0.637 , 1.640	-2.089 , 1.605
เอกรูป	40	10	0.1	-3.058 , 2.939	0.742 , 1.612	-1.843 , 2.506
เอกรูป	40	10	0.2	-3.283 , 3.424	0.705 , 1.439	-2.272 , 1.789
เอกรูป	40	10	0.4	-3.425 , 3.100	0.587 , 1.701	-1.987 , 1.959
เอกรูป	40	20	0.1	-3.158 , 3.147	0.633 , 1.511	-2.149 , 1.844
เอกรูป	40	20	0.2	-3.655 , 3.364	0.731 , 1.605	-1.679 , 1.927
เอกรูป	40	20	0.4	-3.289 , 3.435	0.697 , 1.672	-2.133 , 2.843
เอกรูป	50	10	0.1	-3.035 , 3.081	0.518 , 1.443	-1.982 , 1.289
เอกรูป	50	10	0.2	-4.463 , 3.089	0.668 , 1.553	-1.328 , 3.459
เอกรูป	50	10	0.4	-2.916 , 2.609	0.637 , 1.413	-2.193 , 1.910
เอกรูป	50	20	0.1	-3.425 , 3.472	0.730 , 1.774	-1.966 , 2.551
เอกรูป	50	20	0.2	-3.430 , 2.809	0.700 , 1.670	-1.978 , 2.635
เอกรูป	50	20	0.4	-3.012 , 3.219	0.713 , 1.542	-2.961 , 1.595

หมายเหตุ *: DIF type รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน; Test Length ความยาวของแบบสอบทั้งฉบับ; %DIF จำนวนข้อสอบที่ทำหน้าที่ต่างกัน; Amount of DIF ขนาดของการทำหน้าที่ต่างกัน*

1.2 ประสิทธิภาพด้านอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1

ค่าเฉลี่ยร้อยละของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ทุกเงื่อนไขการตรวจสอบประสิทธิภาพการทำหน้าที่ต่างกันของข้อสอบภายใต้วิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl และเกณฑ์ Zumbo and Thomas ยึดเงื่อนไขของ ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ผลวิเคราะห์ อัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 รายละเอียดดังตารางที่ 4.3

ตารางที่ 4.3 ร้อยละเฉลี่ย ($\overline{X}$) และส่วนเบี่ยงเบนมาตรฐาน (SD) ร้อยละของอัตราความถูกต้อง (correct identification) และอัตราความคลาดเคลื่อนประเภทที่ 1 (type I error rate) ในภาพรวม

DIF type	Test length	%DIF	Amount of DIF	อัตราความถูกต้อง						อัตราความคลาดเคลื่อนประเภทที่ 1					
				LRs		LRz		LRj		LRs		LRz		LRj	
				$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD
อเนกรูป	40	10	0.1	40.00	25.50	3.00	8.12	0.00	0.00	1.78	2.34	0.11	0.54	0.00	0.00
			0.2	70.00	15.81	3.00	14.70	33.00	22.05	6.33	4.34	0.22	1.09	0.11	0.54
			0.4	100.00	0.00	18.00	11.22	62.00	12.49	6.89	4.17	1.22	1.38	0.22	0.75
		20	0.1	89.50	9.80	0.00	0.00	15.00	7.91	7.13	5.19	0.13	0.61	0.13	0.61
			0.2	92.00	8.57	0.00	0.00	37.00	9.00	5.13	3.52	0.00	0.00	0.38	1.02
			0.4	95.50	6.00	1.00	3.39	45.50	9.27	15.00	4.24	0.25	0.85	0.50	1.15
	50	10	0.1	86.40	17.64	0.00	0.00	8.80	9.93	6.58	3.11	0.00	0.00	0.09	0.44
			0.2	98.40	5.43	0.00	0.00	37.60	16.32	5.60	2.89	0.00	0.00	0.27	0.72
			0.4	99.20	3.92	4.80	8.54	63.20	12.24	7.29	4.00	0.53	0.95	0.36	1.03
		20	0.1	18.80	10.32	2.00	4.00	0.00	0.00	0.50	1.00	0.50	1.00	0.00	0.00
			0.2	90.80	8.91	0.00	0.00	46.80	10.09	7.50	3.46	0.00	0.00	0.10	0.49
			0.4	94.80	5.74	4.00	4.90	58.40	12.22	11.40	4.95	1.00	1.22	0.30	0.81

หมายเหตุ*: DIF type รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน; Test Length ความยาวของแบบสอบทั้งฉบับ; %DIF สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน; Amount of DIF ขนาดของการทำหน้าที่ต่างกัน; LRs การทดสอบระดับนัยสำคัญ; LRz การวัดขนาดอิทธิพล เกณฑ์ Zumbo and Thomas LRj การวัดขนาดอิทธิพล เกณฑ์ Jodoin and Gierl*

ตารางที่ 4.3 (ต่อ)

DIF type	Test length	%DIF	Amount of DIF	อัตราความถูกต้อง						อัตราความคลาดเคลื่อนประเภทที่ 1					
				LRs		LRz		LRj		LRs		LRz		LRj	
				$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD
เอกรูป	40	10	0.1	21.00	24.17	0.00	0.00	0.00	0.00	4.33	3.61	0.00	0.00	0.11	0.54
			0.2	22.00	22.72	2.00	6.78	2.00	9.80	5.44	5.58	0.22	0.75	0.22	0.75
			0.4	74.00	22.89	4.00	9.17	33.00	19.65	6.44	3.83	0.33	0.90	0.44	1.02
		20	0.1	17.50	10.61	0.00	0.00	0.00	0.00	1.75	2.81	0.00	0.00	0.25	0.85
			0.2	32.50	13.69	0.00	0.00	1.00	3.39	7.50	5.15	0.00	0.00	0.38	1.02
			0.4	32.50	16.58	1.50	4.06	2.00	4.58	6.50	3.53	0.38	1.02	0.50	1.15
	50	10	0.1	10.40	15.09	0.00	0.00	0.00	0.00	5.33	3.14	0.18	0.60	0.00	0.00
			0.2	25.60	17.45	1.60	5.43	2.40	6.50	4.53	3.23	0.18	0.60	0.27	0.72
			0.4	91.20	12.75	2.40	6.50	32.00	18.76	6.76	3.90	0.27	0.72	0.36	0.81
		20	0.1	18.00	9.80	0.00	0.00	0.00	0.00	1.80	2.06	0.00	0.00	0.20	0.68
			0.2	30.80	12.94	0.80	2.71	1.20	3.25	5.70	4.09	0.50	1.00	0.30	0.81
			0.4	84.00	8.00	0.80	2.71	6.40	7.42	9.20	4.62	0.50	1.00	0.40	0.92

หมายเหตุ*: DIF type รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน; Test Length ความยาวของแบบสอบทั้งฉบับ; %DIF สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน; Amount of DIF ขนาดของการทำหน้าที่ต่างกัน; LRs การทดสอบระดับนัยสำคัญ; LRz การวัดขนาดอิทธิพล เกณฑ์ Zumbo and Thomas, LRj การวัดขนาดอิทธิพล เกณฑ์ Jodoin and Gierl*

จากตารางที่ 4.3 การศึกษาภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ผลการวิเคราะห์อัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า

1.2.1 อัตราความถูกต้อง

การทดสอบระดับนัยสำคัญ ร้อยละเฉลี่ยสูงสุดคิดเป็นร้อยละ 100.00 ภายใต้เงื่อนไขรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป ความยาวของแบบสอบทั้งฉบับ 40 ข้อ จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับ คิดเป็นร้อยละ 10 มีขนาดของการทำหน้าที่ต่างกันขนาด 0.4 ส่วนร้อยละเฉลี่ยต่ำสุด คิดเป็นร้อยละ 3.00 ภายใต้เงื่อนไขรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป ความยาวของแบบสอบทั้งฉบับ 40 ข้อ จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับ คิดเป็นร้อยละ 10 มีขนาดของการทำหน้าที่ต่างกันขนาด 0.1

การวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas ร้อยละเฉลี่ยสูงสุดคิดเป็นร้อยละ 18.00 ภายใต้เงื่อนไขรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป ที่มีความยาวของแบบสอบทั้งฉบับ 40 ข้อ มีจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10 มีขนาดของการทำหน้าที่ต่างกันขนาด 0.4 ส่วนร้อยละเฉลี่ยต่ำสุดคิดเป็นร้อยละ 0.00 โดยพบในข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูปและอเนกรูปด้วยจำนวนตามเงื่อนไขเท่าๆ กัน

การวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl ร้อยละเฉลี่ยสูงสุด คิดเป็นร้อยละ 63.20 ภายใต้เงื่อนไขรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป ที่มีความยาวของแบบสอบทั้งฉบับ 50 ข้อ จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับ คิดเป็นร้อยละ 10 มีขนาดของการทำหน้าที่ต่างกันขนาด 0.4 ส่วนร้อยละเฉลี่ยต่ำสุด คิดเป็นร้อยละ 0.00 พบในข้อสอบที่ทำหน้าที่ต่างกัน แบบเอกรูปมากกว่าแบบอเนกรูป (อัตราความถูกต้องคิดเป็นร้อยละ 0.00 เมื่อ DIF มีขนาด 0.1 ทุกเงื่อนไขของข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป)

1.2.2 อัตราความคลาดเคลื่อนประเภทที่ 1

การทดสอบระดับนัยสำคัญ ร้อยละเฉลี่ยสูงสุด ของอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 15.00 ภายใต้เงื่อนไขรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป ที่มีความยาวของแบบสอบทั้งฉบับ 40 ข้อ จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับ คิดเป็นร้อยละ 20 มีขนาดของการทำหน้าที่ต่างกันขนาด 0.4 ส่วนร้อยละเฉลี่ยต่ำสุด คิดเป็นร้อยละ 1.75 ภายใต้เงื่อนไขรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป มีความยาวของแบบสอบทั้งฉบับเป็น 40 ข้อ มีจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับ คิดเป็นร้อยละ 20 มีขนาดของการทำหน้าที่ต่างกันขนาด 0.1

การวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas ร้อยละเฉลี่ยสูงสุดของอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 1.22 ภายใต้เงื่อนไขรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป ที่มีความยาวของแบบสอบทั้งฉบับ 40 ข้อ จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับ คิดเป็นร้อยละ 10 มีขนาดของการทำหน้าที่ต่างกันขนาด 0.4 ส่วนร้อยละเฉลี่ยต่ำสุด คือร้อยละ 0.00

ขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl เมื่อข้อสอบทำหน้าที่ต่างกันทั้งแบบเอกรูปและแบบอเนกรูป พบว่า ข้อสอบที่ทำหน้าที่ต่างกัน ที่มีขนาด 0.1 จะไม่เกิด อัตราความคลาดเคลื่อน

ประเภทที่ 1 ในบางกรณี คิดเป็นร้อยละ 0.00 ส่วนร้อยละเฉลี่ยสูงสุด ของอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 0.50 ภายใต้เงื่อนไขรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป ที่มีความยาวของแบบสอบทั้งฉบับเป็น 40 ข้อ จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับ คิดเป็นร้อยละ 20 มีขนาดของการทำหน้าที่ต่างกันขนาด 0.4 และพบภายใต้เงื่อนไขรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป ที่มีความยาวของแบบสอบทั้งฉบับเป็น 40 ข้อ จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับ คิดเป็นร้อยละ 20 มีขนาดของการทำหน้าที่ต่างกันขนาด 0.4

1.2.3 อัตราความถูกต้องและ อัตราความคลาดเคลื่อนประเภทที่ 1 จำแนกตาม ปัจจัยและวิธีการ

ค่าเฉลี่ยร้อยละของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้วิธีถดถอยโลจิสติก ในภาพรวม จำแนกตามปัจจัยและวิธีการที่ศึกษา ปรากฏผลตามตารางที่ 4.4 – 4.5

ตารางที่ 4.4 ค่าเฉลี่ยร้อยละของอัตราความถูกต้องในทุกวิธีที่ศึกษาภายใต้วิธีถดถอยโลจิสติก

ปัจจัยที่แปรเปลี่ยน	อัตราความถูกต้อง					
	LRs		LRz		LRj	
	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD
รูปแบบของ DIF						
อเนกรูป	81.28	6.59	2.98	4.12	33.94	6.61
เอกรูป	38.29	5.14	1.09	3.05	6.67	6.69
ความยาว(ข้อ)						
40	57.21	7.71	2.71	4.92	19.21	7.09
50	62.37	4.40	1.37	2.78	21.40	6.10
ร้อยละข้อสอบที่ DIF						
10%	61.52	8.03	3.23	4.76	22.83	7.49
20%	58.06	2.99	0.84	1.90	17.78	4.20
ขนาดของ DIF						
0.1	37.70	6.08	0.63	2.82	2.98	3.89
0.2	57.76	5.20	0.93	4.92	20.13	5.99
0.4	83.90	6.98	4.56	2.97	37.81	5.29
เฉลี่ยทั้งหมด	59.79	1.54	2.04	1.05	20.30	1.17

ค่าเฉลี่ยร้อยละของอัตราความคลาดเคลื่อนประเภทที่ 1 จำแนกตามวิธีที่ศึกษา เฉพาะในภาพรวมของการเกิดการทำหน้าที่ต่างกันของข้อสอบตามเงื่อนไขของปัจจัยที่แปรเปลี่ยน ดังตารางที่ 4.5

ตารางที่ 4.5 ค่าเฉลี่ย ร้อยละ ของอัตราความคลาดเคลื่อนประเภทที่ 1 จำแนกตามวิธีที่ศึกษาภายใต้วิธีถดถอยโลจิสติก

ปัจจัยที่แปรเปลี่ยน	อัตราความคลาดเคลื่อนประเภทที่ 1					
	LRs		LRz		LRj	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
รูปแบบของ DIF						
อเนกรูป	6.76	1.04	0.33	0.47	0.21	0.33
เอกรูป	5.44	0.94	0.21	0.41	0.29	0.28
ความยาว(ข้อ)						
40	6.19	0.92	0.24	0.48	0.27	0.31
50	6.02	1.04	0.31	0.45	0.22	0.32
ร้อยละข้อสอบที่ DIF						
10%	5.61	0.80	0.27	0.43	0.20	0.32
20%	6.59	1.23	0.27	0.51	0.29	0.31
ขนาดของ DIF						
0.1	3.65	1.14	0.12	0.42	0.10	0.32
0.2	5.97	0.89	0.14	0.45	0.25	0.18
0.4	8.69	0.42	0.56	0.20	0.39	0.14
เฉลี่ยทั้งหมด	6.10	0.22	0.27	0.09	0.25	0.06

จากตารางที่ 4.4 - 4.5 สรุปผลค่าเฉลี่ยร้อยละของ อัตราความถูกต้องและ อัตราความคลาดเคลื่อนประเภทที่ 1 จำแนกตามปัจจัยและวิธีการที่ศึกษาดังนี้

1) รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน

ผลการตรวจสอบ ในวิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ พบว่า ภายใต้ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป มีค่าเฉลี่ยของอัตราความถูกต้องคิดเป็นร้อยละ 81.28 และอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 6.76 ตามลำดับ ภายใต้ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป มีค่าเฉลี่ยของอัตราความถูกต้องคิดเป็นร้อยละ 38.29 และอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 5.44

ผลการตรวจสอบในวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas พบว่า ภายใต้ปัจจัย รูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป มีค่าเฉลี่ยของอัตราความถูกต้อง คิดเป็นร้อยละ 2.98 และอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 0.33

ตามลำดับ ภายใต้ปัจจัย รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน แบบเอกรูป มีค่าเฉลี่ยของอัตราความถูกต้องคิดเป็นร้อยละ 1.09 และอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 0.21

ผลการตรวจสอบในวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl พบว่า ภายใต้ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป มีค่าเฉลี่ยของอัตราความถูกต้อง คิดเป็นร้อยละ 33.94 และอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 0.21 ตามลำดับ ภายใต้ปัจจัย รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน แบบเอกรูป มีค่าเฉลี่ยของอัตราความถูกต้องคิดเป็นร้อยละ 6.67 และอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 0.29

2) ขนาดของการทำหน้าที่ต่างกัน

ผลการตรวจสอบ ในวิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ พบว่า ภายใต้ปัจจัยขนาดของการทำหน้าที่ต่างกันขนาด 0.1 มีค่าเฉลี่ยของอัตราความถูกต้อง คิดเป็นร้อยละ 37.7 และอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 3.65 ตามลำดับ ภายใต้ปัจจัย ขนาดของการทำหน้าที่ต่างกันขนาด 0.2 มีค่าเฉลี่ยของอัตราความถูกต้อง คิดเป็นร้อยละ 57.76 และอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 5.97 ภายใต้ปัจจัย ขนาดของการทำหน้าที่ต่างกันขนาด 0.4 มีค่าเฉลี่ยของอัตราความถูกต้อง คิดเป็นร้อยละ 83.9 และอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 8.69

ผลการตรวจสอบในวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas พบว่า ภายใต้ปัจจัย ขนาดของการทำหน้าที่ต่างกันขนาด 0.1 มีค่าเฉลี่ยของอัตราความถูกต้อง คิดเป็นร้อยละ 0.63 และอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 0.12 ตามลำดับ ภายใต้ปัจจัยขนาดของการทำหน้าที่ต่างกันขนาด 0.2 มีค่าเฉลี่ยของอัตราความถูกต้อง คิดเป็นร้อยละ 0.93 และอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 0.14 ภายใต้ปัจจัยขนาดของการทำหน้าที่ต่างกันขนาด 0.4 มีค่าเฉลี่ยของอัตราความถูกต้อง คิดเป็นร้อยละ 4.56 และอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 0.56

ผลการตรวจสอบในวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl พบว่า ภายใต้ปัจจัยขนาดของการทำหน้าที่ต่างกันขนาด 0.1 มีค่าเฉลี่ยของอัตราความถูกต้อง คิดเป็นร้อยละ 2.98 และอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 0.10 ตามลำดับ ภายใต้ปัจจัยขนาดของการทำหน้าที่ต่างกันขนาด 0.2 มีค่าเฉลี่ยของอัตราความถูกต้อง คิดเป็นร้อยละ 20.13 และอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 0.25 ภายใต้ปัจจัย ขนาดของการทำหน้าที่ต่างกันขนาด 0.4 ค่าเฉลี่ยของอัตราความถูกต้อง คิดเป็นร้อยละ 37.81 และอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 0.39

3) จำนวนข้อสอบที่ทำหน้าที่ต่างกัน

ผลการตรวจสอบ ในวิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ พบว่า ภายใต้ปัจจัย จำนวนข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 10 มีค่าเฉลี่ยของอัตราความถูกต้องคิดเป็นร้อยละ 61.52 และอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 5.61 ตามลำดับ ภายใต้ปัจจัย จำนวนข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 20 มีค่าเฉลี่ยของอัตราความถูกต้องคิดเป็นร้อยละ 58.06 และอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 6.59

ผลการตรวจสอบในวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas พบว่า ภายใต้ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 10 มีค่าเฉลี่ยของอัตราความถูกต้องคิดเป็นร้อยละ 3.23 และอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 0.27 ตามลำดับ ภายใต้ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 20 มีค่าเฉลี่ยของอัตราความถูกต้องคิดเป็นร้อยละ 0.84 และอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 0.27

ผลการตรวจสอบในวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl พบว่า ภายใต้ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 10 มีค่าเฉลี่ยของอัตราความถูกต้องคิดเป็นร้อยละ 22.83 และอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 0.20 ตามลำดับ ภายใต้ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 20 มีค่าเฉลี่ยของอัตราความถูกต้องคิดเป็นร้อยละ 17.78 และอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 0.29

4) ความยาวของแบบสอบ

ผลการตรวจสอบ ในวิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ พบว่า ภายใต้ปัจจัย ความยาวของแบบสอบทั้งฉบับ 40 ข้อ มีค่าเฉลี่ยของอัตราความถูกต้อง คิดเป็นร้อยละ 57.21 และอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 6.19 ตามลำดับ ภายใต้ปัจจัย ความยาวของแบบสอบทั้งฉบับ 50 ข้อ มีค่าเฉลี่ยของอัตราความถูกต้อง คิดเป็นร้อยละ 62.37 และอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 6.12

ผลการตรวจสอบในวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas พบว่า ภายใต้ปัจจัย ความยาวของแบบสอบทั้งฉบับ 40 ข้อ มีค่าเฉลี่ยของอัตราความถูกต้องคิดเป็นร้อยละ 2.71 และอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 0.24 ตามลำดับ ภายใต้ปัจจัยความยาวของแบบสอบทั้งฉบับ 50 ข้อ มีค่าเฉลี่ยของอัตราความถูกต้องคิดเป็นร้อยละ 1.37 และอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 0.31

ผลการตรวจสอบในวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl พบว่า ภายใต้ปัจจัยความยาวของแบบสอบทั้งฉบับ 40 ข้อ มีค่าเฉลี่ยของอัตราความถูกต้องคิดเป็นร้อยละ 19.21 และอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 0.27 ตามลำดับ ภายใต้

ปัจจัยความยาวของแบบสอบทั้งฉบับ 50 ข้อ มีค่าเฉลี่ยของอัตราความถูกต้องคิดเป็นร้อยละ 21.40 และอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นร้อยละ 0.22

ตอนที่ 2 ผลการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างการวัดขนาดอิทธิพลตามเกณฑ์ที่ศึกษา

2.1 การวิเคราะห์ความแปรปรวนหลายตัวแปรของอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้วิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl กับเกณฑ์ Zumbo and Thomas เมื่อได้ผลการทำหน้าที่ต่างกันของข้อสอบจึงนำผลการตรวจสอบดังกล่าวมาคำนวณอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธีการแล้วนำมาเปรียบเทียบกันโดยเทคนิคการวิเคราะห์ ความแปรปรวน พหุ (Multivariate analysis of variance; MANOVA) ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ได้แก่ ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน 2 เงื่อนไข คือ แบบเอกรูปและแบบอเนกรูป ปัจจัยขนาดของการทำหน้าที่ต่างกัน 3 เงื่อนไข คือ ขนาด 0.1, 0.2 และขนาด 0.4 ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 2 เงื่อนไข คือ ทั้งฉบับคิดเป็นร้อยละ 10 และ 20 และปัจจัยความยาวของแบบสอบทั้งฉบับ 2 เงื่อนไข คือ ความยาว 40 และ 50 ข้อ พิจารณาประสิทธิภาพ จากผลการตรวจสอบ ที่มีอัตราความถูกต้องในการตรวจสอบสูงและมีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบต่ำ ซึ่งแสดงถึงประสิทธิภาพในการตรวจสอบสูงสุด ถือเป็นเงื่อนไขที่ต้องการ (อุทัยวรรณ สายพัฒนะ, 2547) ดังภาพประกอบ 4.1

		อัตราความถูกต้องในการตรวจสอบ	
		สูง	ต่ำ
อัตราความคลาดเคลื่อนประเภท ที่ 1 ในการตรวจสอบ	สูง	ไม่ต้องการ	ไม่ต้องการ
	ต่ำ	ต้องการ	ไม่ต้องการ

ภาพที่ 4.1 เกณฑ์การพิจารณาประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

การวิเคราะห์ความแตกต่างโดยใช้สถิติการวิเคราะห์ความแปรปรวนพหุ (MANOVA) ที่ระดับนัยสำคัญ .001 มีตัวแปรตาม 2 ตัว คือ อัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ส่วนตัวแปรอิสระ มี 5 ตัว คือ วิธีการ ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดย วิธีถดถอยโลจิสติก ใน 2 วิธีที่ศึกษา และปัจจัยที่แปรเปลี่ยน 4 ปัจจัยดังกล่าวข้างต้น ถ้าผลการทดสอบมีนัยสำคัญทางสถิติแล้วจะทดสอบผลระหว่างกลุ่ม (Test of between-subjects effects) ของตัวแปรตามแต่ละตัวที่ระดับนัยสำคัญ .001 มีรายละเอียดดังนี้

2.1.1 ผลการวิเคราะห์ความแปรปรวนหลายตัวแปร

ผู้วิจัยทำการตรวจสอบข้อตกลงเบื้องต้นของสถิติวิเคราะห์ ผลการวิเคราะห์แสดงดังตารางที่ 4.6

ตารางที่ 4.6 การทดสอบ Box's Test และ Bartlett's Test

Box's M	474.516	**Bartlett's Test of Sphericity**	
F	6.128	Likelihood Ratio	.000
df1	75	Approx. Chi-Square	4042.159
df2	339098.420	Df	2
Sig.	.000	Sig.	.000

จากตารางที่ 4.6 ผลการตรวจสอบข้อตกลงเบื้องต้นของสถิติวิเคราะห์ พบว่าเมตริกซ์ความแปรปรวน -ความแปรปรวนร่วมของตัวแปรตาม คือ อัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ต่างกัน ระหว่างกลุ่มอย่างมีนัยสำคัญทางสถิติ ถือว่าละเมิดข้อตกลงเบื้องต้นทางสถิติของการวิเคราะห์ความแปรปรวนพหุ ที่กำหนดให้เมตริกซ์ความแปรปรวน -ความแปรปรวนร่วมของทุกกลุ่มต้องเท่ากัน (Box's M = 474.516, df = 75 และ 339098.420, Sig. = .000)

เนื่องจากว่าการทดสอบด้วยวิธี Box's M ค่อนข้างมีความไวต่อการละเมิดข้อตกลงเบื้องต้น งานวิจัยของ Holloway and Dunn (1967) Hakstain, Roed and Linn (1979) และ Olson (1974) (อ้างใน พัชรี จันทร์เพ็ง , 2551) พบว่าการที่เมตริกซ์ค่าแปรปรวนร่วมไม่เท่ากันจะไม่มีผลกระทบต่อระดับนัยสำคัญในแต่ละกลุ่มย่อยเท่ากัน (ขนาดของข้อมูลในแต่ละกลุ่มย่อยของการศึกษาครั้งนี้เท่ากัน) ซึ่งการทดสอบความแปรปรวนพหุด้วย F-test มีความแกร่งเพียงพอเมื่อมีการละเมิดข้อตกลงเบื้องต้นดังกล่าว (พัชรี จันทร์เพ็ง, 2551; ณัฏภรณ์ หลาวทอง และ สกล ซื่อธนาพรกุล, 2550)

การทดสอบ Bartlett's Test of Sphericity เป็นการทดสอบข้อตกลงเบื้องต้นของความเป็นเอกพันธ์ของความแปรปรวน เนื่องจากเป็นการตรวจสอบความสัมพันธ์ระหว่างประสิทธิภาพด้าน อัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ว่ามีความสัมพันธ์แตกต่างจากเมตริกซ์เอกลักษณ์หรือไม่ ผลการวิเคราะห์พบว่าตัวแปรตาม 2 ตัวแปรมีความสัมพันธ์กันจึงสามารถวิเคราะห์ความแปรปรวนพหุได้ (พัชรี จันทร์เพ็ง , 2551) (Likelihood Ratio =.000, Approx. Chi-Square = 4042.159, df = 2, Sig= .000) ผลการวิเคราะห์ความแปรปรวนพหุที่มีนัยสำคัญทางสถิติ กำหนดโมเดลของกา รวิเคราะห์ให้มีปฏิสัมพันธ์ระหว่างตัวแปรอิสระไม่เกินอันดับสอง ผู้วิจัยพิจารณาผลการทดสอบปฏิสัมพันธ์สองทางและสามทางระหว่างเงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย กับ วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบใน การเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ตัวแปรอิสระ 5 ตัว คือ วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ กับ เงื่อนไขของ ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ได้ผลการวิเคราะห์ความแปรปรวนพหุ ดังตารางที่ 4.7

ตารางที่ 4.7 การเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ในวิธีถดถอยโลจิสติก ระหว่างการวัดขนาด อิทธิพล ตาม เกณฑ์ ที่ศึกษา ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย

Effect	Value Pillai's Trace	F	p
Intercept	.640	1021.948	.000
LR	.566	750.528	.000
TYPE	.450	470.244	.000
%DIF	.055	33.450	.000
LENGTH	.001	.372	.689
AMOUNT	.476	179.749	.000
LR * TYPE	.397	379.321	.000
LR * %DIF	.009	5.205	.006
LR * LENGTH	.016	9.312	.000
LR * AMOUNT	.394	141.239	.000
LR * TYPE * %DIF	.033	19.702	.000
LR * TYPE * LENGTH	.006	3.326	.036
LR * %DIF * LENGTH	.012	7.063	.001
LR * TYPE * AMOUNT	.176	55.599	.000
LR * %DIF * AMOUNT	.061	18.167	.000
LR * LENGTH * AMOUNT	.012	3.526	.007

จากตารางที่ 4.7 เนื่องจากมีการละเมิดข้อตกลงเบื้องต้นเกี่ยวกับความเท่ากันของเมตริกซ์ความแปรปรวน -ความแปรปรวนร่วมของ กลุ่มตัวอย่าง จึงเลือกใช้ค่าสถิติ Pillai's Trace ซึ่งมีความแกร่ง (Robustness) มากกว่า พบว่า ปฏิสัมพันธ์สองทาง ระหว่าง วิธีการตรวจสอบการทำหน้าที่ ต่างกันของข้อสอบภายใต้วิธีถดถอยโลจิสติก กับเงื่อนไขของปัจจัยที่แปรเปลี่ยน 3 คู่ มีอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ค่า sig เท่ากับ .000 ซึ่งน้อยกว่าระดับนัยสำคัญที่กำหนด หมายความว่ามีนัยสำคัญทางสถิติ ใน 3 เงื่อนไขย่อย คือ 1) วิธีการตรวจสอบ กับ ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน 2) วิธีการตรวจสอบ กับ ปัจจัยความยาวของแบบสอบทั้งฉบับ และ 3) วิธีการตรวจสอบ กับ ปัจจัย ขนาดของ การทำหน้าที่ต่างกันของข้อสอบ ส่วนปฏิสัมพันธ์สามทางระหว่างวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้วิธีถดถอยโลจิสติกกับเงื่อนไขของปัจจัย ที่แปรเปลี่ยน 4 คู่ มีอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 คือ 1) วิธีการตรวจสอบ กับ ปัจจัยรูปแบบของข้อสอบ

ที่ทำหน้าที่ต่างกัน และ ปัจจัย จำนวนข้อสอบที่ทำหน้าที่ต่างกัน 2) วิธีการตรวจสอบ กับ ปัจจัย จำนวนข้อสอบที่ทำหน้าที่ต่างกัน และ ปัจจัย ความยาวของแบบสอบทั้งฉบับ 3) วิธีการตรวจสอบ กับ ปัจจัย รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน และ ปัจจัย ขนาดของ การทำหน้าที่ต่างกันของข้อสอบ และ 4) วิธีการตรวจสอบ กับ ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน และปัจจัย ขนาดของ การทำหน้าที่ต่างกันของข้อสอบ

2.1.2 ผลการทดสอบระหว่างกลุ่ม

การวิเคราะห์ความแปรปรวนพหุที่มีนัยสำคัญทางสถิติ พิจารณาเฉพาะกรณีที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ของปฏิสัมพันธ์สองทางและสามทางระหว่างวิธีการ กับ เงื่อนไขปัจจัยที่แปรเปลี่ยน 7 เงื่อนไขย่อย แล้วทดสอบระหว่างกลุ่ม (Test of between-subjects effects) ของอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 เปรียบเทียบแยกทีละตัวแปรตาม ดังตารางที่ 4.8

ตารางที่ 4.8 ผลการทดสอบระหว่างกลุ่ม ภายใต้เงื่อนไขปฏิสัมพันธ์สองทาง ระหว่าง วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบกับปัจจัยที่แปรเปลี่ยน

แหล่งของความแปรปรวน	ประสิทธิภาพ	F	P
LR * TYPE	CI	659.394	.000
	TE	4.919	.027
LR * LENGTH	CI	12.777	.000
	TE	1.679	.195
LR * AMOUNT	CI	326.611	.000
	TE	3.416	.033
LR * TYPE * %DIF	CI	36.638	.000
	TE	.017	.896
LR * %DIF * LENGTH	CI	5.572	.018
	TE	12.115	.001
LR * TYPE * AMOUNT	CI	110.300	.000
	TE	3.200	.041
LR * %DIF * AMOUNT	CI	33.354	.000
	TE	.181	.835

จากตารางที่ 4.8 สรุปผลการวิเคราะห์ปฏิสัมพันธ์สองทางและสามทางระหว่างวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและเงื่อนไขปัจจัยที่แปรเปลี่ยน ดังนี้

1) ปฏิสัมพันธ์สองทางระหว่างวิธีการตรวจสอบ กับ ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน มีผลต่ออัตราความถูกต้องอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่อ อัตราความคลาดเคลื่อนประเภทที่ 1

2) ปฏิสัมพันธ์สองทางระหว่างวิธีการตรวจสอบ กับ ปัจจัยความยาวของแบบสอบทั้งฉบับ มีผลต่ออัตราความถูกต้องอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1

3) ปฏิสัมพันธ์สองทางระหว่างวิธีการตรวจสอบ กับ ปัจจัยขนาดของ การทำหน้าที่ต่างกันของข้อสอบ มีผลต่ออัตราความถูกต้องอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่อ อัตราความคลาดเคลื่อนประเภทที่ 1

4) ปฏิสัมพันธ์สามทางระหว่างวิธีการตรวจสอบ กับ ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน และ ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน มีผลต่ออัตราความถูกต้องอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1

5) ปฏิสัมพันธ์ สามทางระหว่างวิธีการตรวจสอบ กับ ปัจจัย จำนวนข้อสอบที่ทำหน้าที่ต่างกัน และ ปัจจัยความยาวของแบบสอบทั้งฉบับ มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่ออัตราความถูกต้อง

6) ปฏิสัมพันธ์สามทางระหว่างวิธีการตรวจสอบ กับ ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน และ ปัจจัย ขนาดของ การทำหน้าที่ต่างกันของข้อสอบ มีผลต่ออัตราความถูกต้องอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1

7) ปฏิสัมพันธ์ สามทางระหว่างวิธีการตรวจสอบ กับ ปัจจัย จำนวนข้อสอบที่ทำหน้าที่ต่างกัน และ ปัจจัย ขนาดของ การทำหน้าที่ต่างกันของข้อสอบ มีผลต่ออัตราความถูกต้องอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1

ผลการทดสอบระหว่างกลุ่มของประสิทธิภาพด้านอัตราความถูกต้อง พบว่ามีนัยสำคัญทางสถิติ 6 เงื่อนไข ส่วนประสิทธิภาพด้านอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่ามีนัยสำคัญทางสถิติ เพียง 1 เงื่อนไข ผู้วิจัยจึงทำการทดสอบความแปรปรวนผลย่อยต่อไป

2.2 การวิเคราะห์ความแปรปรวนการทดสอบผลย่อยของอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 พร้อมทั้งการเปรียบเทียบภายหลัง

การทดสอบความแปรปรวนผลย่อย (Simple effect) ของอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 พร้อมการเปรียบเทียบภายหลังเป็นการนำผลจากการทดสอบปฏิสัมพันธ์สองทางและสามทางระหว่างวิธีการตรวจสอบการทำหน้าที่ ต่างกันของข้อสอบ กับ ปัจจัยที่แปรเปลี่ยนที่มีนัยสำคัญทางสถิติมาวิเคราะห์ความแปรปรวน โดยการทดสอบความแปรปรวนผลย่อยที่ระดับนัยสำคัญ

.001 เมื่อพบว่ามีนัยสำคัญทางสถิติจะ เปรียบเทียบความแตกต่างรายคู่โดยใช้วิธีของเชฟเฟ (Scheffé) การทดสอบความแปรปรวนผลย่อย ดำเนินการทดสอบใน 7 เงื่อนไขย่อย ดังนี้

เงื่อนไขย่อยที่ 1 การทดสอบความแปรปรวนผลย่อย จากปฏิสัมพันธ์สองทาง ระหว่างวิธีการตรวจสอบ กับ ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน

เงื่อนไขย่อยที่ 2 การทดสอบความแปรปรวนผลย่อย จากปฏิสัมพันธ์สองทาง ระหว่างวิธีการตรวจสอบ กับ ปัจจัยความยาวของแบบสอบทั้งฉบับ

เงื่อนไขย่อยที่ 3 การทดสอบความแปรปรวนผลย่อย จากปฏิสัมพันธ์สองทาง ระหว่างวิธีการตรวจสอบ กับ ปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ

เงื่อนไขย่อยที่ 4 การทดสอบความแปรปรวนผลย่อย จากปฏิสัมพันธ์ สามทาง ระหว่างวิธีการตรวจสอบ กับ ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน และ ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน

เงื่อนไขย่อยที่ 5 การทดสอบความแปรปรวนผลย่อย จากปฏิสัมพันธ์ส ามทาง ระหว่างวิธีการตรวจสอบ กับ ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน และ ปัจจัยความยาวของแบบสอบทั้งฉบับ

เงื่อนไขย่อยที่ 6 การทดสอบความแปรปรวนผลย่อย จากปฏิสัมพันธ์ส ามทาง ระหว่างวิธีการตรวจสอบ กับ ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน และ ปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ

เงื่อนไขย่อยที่ 7 การทดสอบความแปรปรวนผลย่อย จากปฏิสัมพันธ์ส ามทาง ระหว่างวิธีการตรวจสอบ กับ ปัจจัย จำนวนข้อสอบที่ทำหน้าที่ต่างกัน และ ปัจจัย ขนาดของ การทำหน้าที่ต่างกันของข้อสอบ

การนำเสนอผลการวิเคราะห์ในแต่ละเงื่อนไขย่อย มีประเด็นการนำเสนอ ดังนี้ 1) ภายใต้เงื่อนไขเดียวกัน คือ การเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบภายใต้วิธีการที่ศึกษา โดยพิจารณาเฉพาะเงื่อนไขแต่ละระดับของปัจจัยที่แปรเปลี่ยน 2) ภายใต้เงื่อนไขต่างกัน คือ การเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบภายใต้ปัจจัยที่ศึกษา โดยพิจารณาเฉพาะวิธีการตรวจสอบแต่ละวิธีการ

2.2.1 **การทดสอบความแปรปรวนผลย่อย เงื่อนไขย่อยที่** 1

การทดสอบความแปรปรวนผลย่อยโดยพิจารณาผล จากปฏิสัมพันธ์สองทางระหว่างวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ กับ ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน ที่มีนัยสำคัญทางสถิติ ดังตารางที่ 4.9

ตารางที่ 4.9 การทดสอบความแปรปรวนผลย่อยระหว่างวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (LR) กับปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน (TYPE)

แหล่งของความแปรปรวน	CI		TE	
	F	p	F	P
ภายใต้เงื่อนไขเดียวกัน				
LR ที่ TYPE1	413.915	.000	-	-
LRz (Mean=2.983, SD=8.238)				
LRj (Mean=33.942, SD=25.036)				
LR ที่ TYPE2	38.970	.000	-	-
LRz (Mean=1.092, SD= 4.576)				
LRj (Mean= 6.667, SD= 14.776)				
ภายใต้เงื่อนไขต่างกัน				
TYPE ที่ LRz	12.088	.001	-	-
TYPE1 (Mean= 2.983, SD= 8.238)				
TYPE2 (Mean= 1.092, SD= 4.576)				
TYPE ที่ LRj	264.082	.000	-	-
TYPE1 (Mean= 33.942, SD= 25.036)				
TYPE2 (Mean= 6.667, SD= 14.776)				

จากตารางที่ 4.9 การทดสอบความแปรปรวนผลย่อยระหว่างวิธีการตรวจสอบ กับ ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน

1) การทดสอบความแปรปรวนผลย่อย เงื่อนไขย่อยที่ 1 ภายใต้เงื่อนไขเดียวกัน

การทดสอบความแปรปรวนผลย่อย ของวิธีการ ตรวจสอบการทำหน้าที่ต่างกัน ภายใต้เงื่อนไขเดียวกัน คือ การเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบภายใต้วิธีการที่ศึกษา โดยพิจารณาเฉพาะ เงื่อนไข แต่ละระดับ ของปัจจัยที่แปรเปลี่ยน ดังนี้ 1) การเปรียบเทียบ ค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบภายใต้วิธีถดถอยโลจิสติก โดยพิจารณาเฉพาะรูปแบบของข้อสอบที่ ทำหน้าที่ต่างกัน แบบอเนกรูป และ 2) การเปรียบเทียบ ค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบภายใต้วิธีถดถอยโลจิสติก เฉพาะรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป พบว่า

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง เงื่อนไขปัจจัยรูปแบบของ ข้อสอบที่ ทำหน้าที่ต่างกันทั้งแบบอเนกรูปและแบบเอกรูป ภายใต้วิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพลทั้ง 2 เกณฑ์ มีอัตราความถูกต้อง แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ที่ระดับ .001 การวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl มีค่าเฉลี่ยอัตราความถูกต้องสูงกว่าเกณฑ์ Zumbo and Thomas

2) การทดสอบความแปรปรวนผลย่อย เงื่อนไขย่อยที่ 1 ภายใต้เงื่อนไขต่างกัน

การทดสอบความแปรปรวนผลย่อย ของปัจจัยรูปแบบของข้อสอบที่ ทำหน้าที่ต่างกันภายใต้เงื่อนไขต่างกัน คือ การเปรียบเทียบ ค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบภายใต้ปัจจัยที่ศึกษา โดยพิจารณาเฉพาะวิธีการตรวจสอบแต่ละวิธีการ ดังนี้ 1) กรณีการเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน พิจารณาเฉพาะวิธีการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas และ 2) กรณีการเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน พิจารณาเฉพาะ วิธีการวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl พบว่า ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง วิธีถดถอยโลจิสติก โดย การวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas ภายใต้รูปแบบของข้อสอบที่ทำหน้าที่ต่างกันทั้ง 2 ลักษณะ มีอัตราความถูกต้อง แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 โดยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป มีค่าเฉลี่ยอัตราความถูกต้องสูงกว่ารูปแบบของข้อสอบที่ ทำหน้าที่ต่างกัน แบบเอกรูป วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl ภายใต้รูปแบบของข้อสอบที่ทำหน้าที่ต่างกันทั้ง 2 ลักษณะ มีอัตราความถูกต้อง แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 โดยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป มีค่าเฉลี่ยอัตราความถูกต้อง สูงกว่ารูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป

2.2.2 การทดสอบความแปรปรวนผลย่อย เงื่อนไขย่อยที่ 2

การทดสอบความแปรปรวนผลย่อย พิจารณาผลจาก ปฏิสัมพันธ์สองทางระหว่างวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ กับปัจจัย ความยาวของแบบสอบทั้งฉบับ ที่มีนัยสำคัญทางสถิติ ดังตารางที่ 4.10

ตารางที่ 4.10 การทดสอบความแปรปรวนผลย่อยระหว่าง วิธีการตรวจสอบการทำหน้าที่ ต่างกันของข้อสอบ (LR) กับ ปัจจัยความยาวของแบบสอบทั้งฉบับ (LENGTH)

แหล่งของความแปรปรวน	CI		TE	
	F	p	F	P
ภายใต้เงื่อนไขเดียวกัน				
LR ที่ Length40	130.297	.000	-	-
LRz (Mean= 2.708, SD= 8.397)				
LRj (Mean=19.208, SD=23.587)				

ตารางที่ 4.10 (ต่อ)

แหล่งของความแปรปรวน	CI		TE	
	F	p	F	P
LR ที่ Length50	177.381	.000	-	-
LRz (Mean=1.367, SD=4.38135)				
LRj (Mean=21.400, SD=25.68213)				
ภายใต้เงื่อนไขต่างกัน				
LENGTH ที่ LRz	6.020	.014	-	-
LENGTH40 (Mean= 2.708, SD=8.397)				
LENGTH50 (Mean=1.367, SD=4.381)				
LENGTH ที่ LRj	1.185	.277	-	-
LENGTH40 (Mean=19.208, SD=23.587)				
LENGTH50 (Mean= 21.400,SD= 25.682)				

จากตารางที่ 4.10 การทดสอบความแปรปรวนผลย่อยระหว่าง วิธีการตรวจสอบการทำหน้าที่ต่าง กันของข้อสอบ กับ ปัจจัยความยาวของแบบสอบทั้งฉบับ สรุปได้ดังนี้

1) การทดสอบความแปรปรวนผลย่อย เงื่อนไขย่อยที่ 2 ภายใต้เงื่อนไขเดียวกัน

การทดสอบความแปรปรวนผลย่อย ของวิธีการ ตรวจสอบการทำหน้าที่ต่างกัน ภายใต้เงื่อนไขเดียวกัน คือ การเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบภายใต้วิธีการที่ศึกษา โดยพิจารณาเฉพาะ เงื่อนไข แต่ละระดับ ของปัจจัยที่แปรเปลี่ยน ดังนี้ 1) การเปรียบเทียบ ค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบภายใต้วิธีถดถอยโลจิสติก โดยพิจารณาเฉพาะ ความยาวของแบบสอบทั้งฉบับ 40 ข้อ และ 2) การเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบภายใต้ วิธีถดถอยโลจิสติก โดยพิจารณาเฉพาะความยาวของแบบสอบทั้งฉบับ 50 ข้อ

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง เงื่อนไขปัจจัย ความยาวของแบบสอบทั้งฉบับ 40 และ 50 ข้อ ภายใต้วิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพลทั้ง 2 เกณฑ์ มีอัตราความถูกต้อง แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ผลการเปรียบเทียบขนาดความยาวของแบบสอบทั้ง 2 ฉบับ มีความสอดคล้องกันกล่าวคือ ภายใต้วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl มีค่าเฉลี่ยอัตราความถูกต้องสูงกว่าตามเกณฑ์ Zumbo and Thomas

2) การทดสอบความแปรปรวนผลย่อย เงื่อนไขย่อยที่ 2 ภายใต้เงื่อนไขต่างกัน

การทดสอบความแปรปรวนผลย่อยใน ปัจจัย ความยาวของแบบสอบทั้งฉบับ ภายใต้เงื่อนไขต่างกัน คือ การเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1

ของข้อสอบภายใต้ปัจจัยที่ศึกษา โดยพิจารณา เฉพาะวิธีการตรวจสอบแต่ละวิธีการ ดังนี้ 1) กรณีการเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้องละอัตราความคลาดเคลื่อนประเภทที่ 1 ของ ความยาวของแบบสอบทั้งฉบับ โดยพิจารณาเฉพาะ การวัดขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas และ 2) กรณีการเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของความยาวของแบบสอบทั้งฉบับ โดยพิจารณาเฉพาะการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl พบว่า

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง วิธีถดถอยโลจิสติกโดย การวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas และการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl ให้ผลสอดคล้องกันคือ ภายใต้ความยาวของแบบสอบทั้งฉบับ ทั้ง 2 ขนาด มีอัตราความถูกต้อง ไม่แตกต่างกัน

2.2.3 การทดสอบความแปรปรวนผลย่อย เงื่อนไขย่อยที่ 3

การทดสอบความแปรปรวนผลย่อย พิจารณาผลจาก ปฏิสัมพันธ์สองทางระหว่างวิธีการตรวจสอบการทำหน้าที่ ต่างกันของข้อสอบ กับปัจจัย ขนาดของ การทำหน้าที่ต่างกันของข้อสอบ ที่มีนัยสำคัญทางสถิติ ดังตารางที่ 4.11

ตารางที่ 4.11 การทดสอบความแปรปรวนผลย่อยระหว่างวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (LR) กับ ปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ (AMOUNT)

แหล่งของความแปรปรวน	CI		TE	
	F	p	F	P
ภายใต้เงื่อนไขเดียวกัน				
LR ที่ Amount 0.1	18.144	.000	-	-
LRz (Mean=.625, SD=3.397)				
LRj (Mean=2.975, SD=7.024)				
LR ที่ Amount 0.2	138.495	.000	-	-
LRz (Mean=.925, SD=6.222)				
LRj (Mean=20.125, SD=22.218)				
LR ที่ Amount 0.4	293.309	.000		
LRz (Mean=4.563, SD=8.730)				
LRj (Mean=37.813, SD=26.031)				
ภายใต้เงื่อนไขต่างกัน				
AMOUNT ที่ LRz	22.791	.000	-	-
AMOUNT0.1(Mean=.625, SD=3.397)				
AMOUNT0.2(Mean=.925, SD=6.222)				
AMOUNT0.4(Mean=4.563, SD=8.730)				

ตารางที่ 4.11 (ต่อ)

แหล่งของความแปรปรวน	CI		TE	
	F	p	F	P
AMOUNT ที่ LRj	149.157	.000	-	-
AMOUNT0.1(Mean= 2.975, SD=7.024)				
AMOUNT0.2(Mean= 20.125, SD=22.218)				
AMOUNT0.4(Mean= 37.813, SD=26.031)				

จากตารางที่ 4.11 การทดสอบความแปรปรวนผลย่อยระหว่างวิธีการตรวจสอบ กับ ปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ สรุปได้ดังนี้

1) การทดสอบความแปรปรวนผลย่อย เงื่อนไขย่อยที่ 3 ภายใต้เงื่อนไขเดียวกัน

การทดสอบความแปรปรวนผลย่อย ของวิธีการ ตรวจสอบการทำหน้าที่ต่างกัน ภายใต้เงื่อนไขเดียวกัน คือ การเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบภายใต้วิธีการที่ศึกษา โดยพิจารณาเฉพาะเงื่อนไข แต่ละระดับ ของปัจจัยที่แปรเปลี่ยน ดังนี้ 1) การเปรียบเทียบ ค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบภายใต้ วิธี ถดถอยโลจิสติก โดยพิจารณาเฉพาะ ขนาดของ การทำหน้าที่ต่างกันของข้อสอบ เป็น 0.1 2) การเปรียบเทียบ ค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบภายใต้วิธีถดถอยโลจิสติก โดยพิจารณาเฉพาะขนาดของการทำหน้าที่ต่างกันของข้อสอบ เป็น 0.2 และ 3) การเปรียบเทียบ ค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบภายใต้ วิธี ถดถอยโลจิสติก โดยพิจารณาเฉพาะ ขนาดของ การทำหน้าที่ต่างกันของข้อสอบ เป็น 0.4 ตามลำดับ

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง ทุกเงื่อนไขปัจจัย ขนาดของ การทำหน้าที่ต่างกันของข้อสอบเป็น 0.1, 0.2 และ 0.4 ภายใต้วิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพลทั้ง 2 เกณฑ์ มี อัตราความถูกต้อง แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ที่ระดับ .001 ซึ่งผลการเปรียบเทียบทั้ง 3 ขนาดของการทำหน้าที่ต่างกันของข้อสอบ มีความสอดคล้องกันกล่าวคือ ภายใต้วิธีถดถอยโลจิสติกโดยการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl มีค่าเฉลี่ยอัตราความถูกต้อง สูงกว่าการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas

2) การทดสอบความแปรปรวนผลย่อย เงื่อนไขย่อยที่ 3 ภายใต้เงื่อนไขต่างกัน

การทดสอบความแปรปรวนผลย่อยใน ปัจจัยขนาดของ การทำหน้าที่ต่างกันของข้อสอบภายใต้เงื่อนไขต่างกัน คือ การเปรียบเทียบ ค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบภายใต้ปัจจัยที่ศึกษา โดยพิจารณา เฉพาะวิธีการตรวจสอบแต่ละวิธีการ ดังนี้ 1) กรณีการเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของขนาด

ของการทำหน้าที่ต่างกันของข้อสอบ พิจารณาเฉพาะ การวัดขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas และ 2) กรณีการเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ขนาดของการทำหน้าที่ต่างกันของข้อสอบ โดยพิจารณาเฉพาะการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl พบว่า ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง วิธีถดถอยโลจิสติกโดยการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas ภายใต้ขนาดของการทำหน้าที่ต่างกันของข้อสอบ ทั้ง 3 ขนาด มีอัตราความถูกต้อง แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ที่ระดับ .001 เมื่อนำ ผลการทดสอบที่มีนัยสำคัญทางสถิติไปเปรียบเทียบรายคู่ โดยใช้วิธีของเชฟเฟ (Scheffé) ได้ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้องภายใต้เงื่อนไขของ ปัจจัยขนาดของ การทำหน้าที่ต่างกันของข้อสอบ ขนาด 0.1 ขนาด 0.2 และขนาด 0.4 ซึ่งในคู่ของการเปรียบเทียบค่าเฉลี่ย ของอัตราความถูกต้อง มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 คือระหว่างขนาดของการทำหน้าที่ต่างกันของข้อสอบ ขนาด 0.1 กับ ขนาด 0.4 และขนาดของการทำหน้าที่ต่างกันของข้อสอบ ขนาด 0.2 กับ ขนาด 0.4

วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl ภายใต้ขนาดของการทำหน้าที่ต่างกันของข้อสอบ 3 ขนาด มีอัตราความถูกต้องแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 เมื่อนำ ผลการทดสอบที่มีนัยสำคัญทางสถิติไปเปรียบเทียบรายคู่ โดยใช้วิธีของเชฟเฟ (Scheffé) ได้ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้องภายใต้เงื่อนไขของ ปัจจัยขนาดของ การทำหน้าที่ต่างกันของข้อสอบ ขนาด 0.1 ขนาด 0.2 และขนาด 0.4 ในทุกคู่ของการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้องมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 คือระหว่างขนาดของ การทำหน้าที่ต่างกันของข้อสอบ ขนาด 0.1 กับ ขนาด 0.4 และขนาดของ การทำหน้าที่ต่างกันของข้อสอบ ขนาด 0.1 กับ ขนาด 0.2

2.2.4 **การทดสอบความแปรปรวนผลย่อย เงื่อนไขย่อยที่** 4

การทดสอบความแปรปรวนผลย่อย พิจารณาผลจาก ปฏิสัมพันธ์ สาม ทางระหว่างวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ กับ ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน และปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน ที่มีนัยสำคัญทางสถิติดังตารางที่ 4.12

ตารางที่ 4.12 การทดสอบความแปรปรวนผลย่อยระหว่างวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (LR) กับ ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน (TYPE) และปัจจัยจำนวนข้อสอบ ที่ทำหน้าที่ต่างกัน (%DIF)

แหล่งของความแปรปรวน	CI		TE	
	F	p	F	p
ภายใต้เงื่อนไขเดียวกัน				
LR ที่ (TYPE1 × 10%DIF)	144.648	.000	-	-

ตารางที่ 4.12 (ต่อ)

แหล่งของความแปรปรวน	CI		TE	
	F	p	F	p
LRz (Mean=4.800, SD=10.897)				
LRj (Mean=34.100, SD=27.775)				
LR ที่ (TYPE1 × 20%DIF)	321.072	.000	-	-
LRz (Mean=1.167, SD=3.285)				
LRj (Mean=33.783, SD=22.050)				
LR ที่ (TYPE2 × 10%DIF)	36.363	.000	-	-
LRz (Mean=1.667, SD=5.983)				
LRj (Mean=11.567, SD=19.196)				
LR ที่ (TYPE2 × 20%DIF)	8.735	.003	-	-
LRz (Mean=.517, SD=2.358)				
LRj (Mean=1.767, SD=4.612)				
ภายใต้เงื่อนไขต่างกัน				
(TYPE×%DIF) ที่ LRz	5.413	.020		
(TYPE1×10%DIF) (Mean=4.800, SD=10.897)				
(TYPE1×20%DIF) (Mean=1.167, SD=3.285)			-	-
(TYPE2×10%DIF) (Mean=1.667, SD=5.983)				
(TYPE2×20%DIF) (Mean=.517, SD=2.358)				
(TYPE×%DIF) ที่ LRj	8.188	.004	-	-
(TYPE1×10%DIF) (Mean=34.100, SD=27.776)				
(TYPE1×20%DIF) (Mean=33.783, SD=22.050)				
(TYPE2×10%DIF) (Mean=11.567, SD=19.196)				
(TYPE2×20%DIF) (Mean=1.767, SD=4.612)				

จากตารางที่ 4.12 การทดสอบความแปรปรวนผลย่อยระหว่างวิธีการตรวจสอบ กับ ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันและปัจจัยจำนวนข้อสอบ ที่ทำหน้าที่ต่างกัน สรุปได้ดังนี้

1) การทดสอบความแปรปรวนผลย่อย เงื่อนไขย่อยที่ 4 ภายใต้เงื่อนไขเดียวกัน

การทดสอบความแปรปรวนผลย่อย ของวิธีการ ตรวจสอบการทำหน้าที่ต่างกัน ภายใต้เงื่อนไขเดียวกัน คือ การเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบภายใต้วิธีการที่ศึกษา โดยพิจารณาเฉพาะ เงื่อนไข แต่ละระดับ ของปัจจัยที่แปรเปลี่ยน ดังนี้ เป็นการเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบ

ภายใต้วิธีถดถอยโลจิสติก โดยพิจารณาเฉพาะ รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน แบบอเนกรูปที่มีจำนวนข้อสอบที่ ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10 และอีก 3 กรณีที่ทำการทดสอบความแปรปรวนผลย่อยภายใต้วิธีถดถอยโลจิสติก คือ 1) พิจารณาเฉพาะรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป ที่มีจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 20 2) พิจารณาเฉพาะ รูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป ที่มีจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10 และ 3) พิจารณาเฉพาะรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน แบบเอกรูปที่มี จำนวนข้อสอบ ที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 20 มีผลการเปรียบเทียบค่าเฉลี่ยมีดังนี้

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง เงื่อนไขรูปแบบของข้อสอบที่ทำหน้าที่ต่าง กันแบบอเนกรูป ที่มี จำนวนข้อที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 10 ภายใต้วิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพลทั้ง 2 เกณฑ์ มีอัตราความถูกต้องแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 การวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl มีค่าเฉลี่ยอัตราความถูกต้อง สูงกว่าเกณฑ์ Zumbo and Thomas

เงื่อนไขรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน แบบอเนกรูป ที่มี จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 20 ภายใต้วิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพลทั้ง 2 เกณฑ์ มีอัตราความถูกต้อง แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ที่ระดับ .001 ซึ่งการวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl มีค่าเฉลี่ยอัตราความถูกต้องสูงกว่าเกณฑ์ Zumbo and Thomas

เงื่อนไข รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน แบบเอกรูป ที่มี จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10 ภายใต้วิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพลทั้ง 2 เกณฑ์ มีอัตราความถูกต้อง แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ที่ระดับ .001 ซึ่งการวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl มีค่าเฉลี่ยอัตราความถูกต้องสูงกว่าเกณฑ์ Zumbo and Thomas

เงื่อนไข รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน แบบเอกรูป ที่มี จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 20 ภายใต้วิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพลทั้ง 2 เกณฑ์ มีอัตราความถูกต้องไม่แตกต่างกัน

2) การทดสอบความแปรปรวนผลย่อย เงื่อนไขย่อยที่ 4 ภายใต้เงื่อนไขต่างกัน

การทดสอบความแปรปรวนผลย่อยใน ปัจจัยรูปแบบของข้อสอบที่ ทำหน้าที่ต่างกัน กับปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน ภายใต้เงื่อนไขต่างกัน คือ การเปรียบเทียบ ค่าเฉลี่ยอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบภายใต้ปัจจัยที่ศึกษา โดยพิจารณา เฉพาะวิธีการตรวจสอบแต่ละวิธีการ ดังนี้ 1) กรณีการเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน กับปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน โดยพิจารณาเฉพาะ การวัดขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas และ 2) กรณีการเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของรูปแบบของข้อสอบ

ที่ทำหน้าที่ต่างกัน กับปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน โดยพิจารณาเฉพาะ การวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl พบว่า

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง วิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas และตามเกณฑ์ Jodoin and Gierl ภายใต้รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน กับปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน (ทั้งฉบับคิดเป็นร้อยละ 10 และ 20) มีอัตราความถูกต้องไม่แตกต่างกัน

2.2.5 การทดสอบความแปรปรวนผลย่อย เงื่อนไขย่อยที่ 5

การทดสอบความแปรปรวนผลย่อย พิจารณาผลจาก ปฏิสัมพันธ์ สามทางระหว่างวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ กับ ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน และ ปัจจัยความยาวของแบบสอบทั้งฉบับ ที่มีนัยสำคัญทางสถิติดังตารางที่ 4.13

ตารางที่ 4.13 การทดสอบความแปรปรวนผลย่อยระหว่าง วิธีการตรวจสอบการทำหน้าที่ ต่างกันของข้อสอบ (LR) กับ ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน (%DIF) และปัจจัย ความยาวของแบบสอบทั้งฉบับ (LENGTH)

แหล่งของความแปรปรวน	CI		TE	
	F	p	F	p
ภายใต้เงื่อนไขเดียวกัน				
LR ที่ (10%DIF × Length40) LRz (Mean=.352, SD=.981) LRj (Mean=.185, SD=.695)	-	-	2.882	.091
LR ที่ (10%DIF × Length50) LRz (Mean=.193, SD=.627) LRj (Mean=.222, SD=.717)	-	-	.145	.703
LR ที่ (20%DIF × Length40) LRz (Mean=.125, SD=.614) LRj (Mean=.354, SD=.994)	-	-	5.769	.017
LR ที่ (20%DIF × Length50) LRz (Mean=.417, SD=.933) LRj (Mean=.217, SD=.706)	-	-	4.373	.037

ตารางที่ 4.13 (ต่อ)

แหล่งของความแปรปรวน	CI		TE	
	F	p	F	p
ภายใต้เงื่อนไขต่างกัน				
(%DIF×LENGTH) ที่ LRz	-	-	17.012	.000
(10%DIF×LENGTH40) (Mean=.352, SD=.981)				
(10%DIF×LENGTH50) (Mean=.193, SD=.627)				
(20%DIF×LENGTH40) (Mean=.125, SD=.614)				
(20%DIF×LENGTH50) (Mean=.417, SD=.935)				
(%DIF×LENGTH) ที่ LRj	-	-	.005	.944
(10%DIF×LENGTH40) (Mean=.185, SD=.695)				
(10%DIF×LENGTH50) (Mean=.222, SD=.717)				
(20%DIF×LENGTH40) (Mean=.354, SD=.994)				
(20%DIF×LENGTH50) (Mean=.217, SD=.706)				

จากตารางที่ 4.13 การทดสอบความแปรปรวนผลย่อยระหว่าง วิธีการตรวจสอบ กับ ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน และ ปัจจัยความยาวของแบบสอบทั้งฉบับ สรุปได้ดังนี้

1) การทดสอบความแปรปรวนผลย่อย เงื่อนไขย่อยที่ 5 ภายใต้เงื่อนไขเดียวกัน

การทดสอบความแปรปรวนผลย่อย ของวิธีการ ตรวจสอบการทำหน้าที่ต่างกัน ภายใต้เงื่อนไขเดียวกัน คือ การเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบภายใต้วิธีการที่ศึกษา โดยพิจารณาเฉพาะ เงื่อนไข แต่ละระดับ ของปัจจัยที่แปรเปลี่ยน ดังนี้ กรณีเปรียบเทียบ ค่าเฉลี่ยอัตราความถูกต้อง และ อัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบภายใต้วิธีถดถอยโลจิสติก โดยพิจารณาเฉพาะ แบบสอบที่มี ความยาวของแบบสอบทั้งฉบับ 40 ข้อ จำนวนข้อสอบที่ทำหน้าที่ต่างกัน ที่มีจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10 และอีก 3 กรณีที่ทำการทดสอบความแปรปรวนผลย่อยภายใต้วิธีถดถอยโลจิสติก คือ (1) พิจารณาเฉพาะ แบบสอบที่มีความยาวของแบบสอบทั้งฉบับ 40 ข้อ จำนวนข้อสอบที่ทำหน้าที่ต่างกัน ที่มีจำนวนข้อสอบที่ ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 20 (2) พิจารณาเฉพาะแบบสอบที่มีความยาวของแบบสอบทั้งฉบับ 50 ข้อ จำนวนข้อสอบที่ทำหน้าที่ต่างกัน ที่มีจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10 และ (3) พิจารณาเฉพาะแบบสอบที่มีความยาวของแบบสอบทั้งฉบับ 50 ข้อ จำนวนข้อสอบที่ทำหน้าที่ต่างกัน ที่มีจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 20 ผลการเปรียบเทียบค่าเฉลี่ยมีดังนี้

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้วิธีถดถอยโลจิสติก โดยพิจารณาเฉพาะ ความ

ยาวของแบบสอบทั้งฉบับ 40 ข้อ และ 50 ข้อ กับจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่มีจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10 และร้อยละ 20 พบว่าในทุกเงื่อนไขดังกล่าว มี อัตรา ความคลาดเคลื่อนประเภทที่ 1 ไม่แตกต่างกัน

2) การทดสอบความแปรปรวนผลย่อย เงื่อนไขย่อยที่ 5 ภายใต้เงื่อนไขต่างกัน

การทดสอบความแปรปรวนผลย่อยในปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน กับ ปัจจัยความยาวของแบบสอบทั้งฉบับ ภายใต้เงื่อนไขต่างกัน คือ การเปรียบเทียบ ค่าเฉลี่ยอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบ ภายใต้ปัจจัยที่ศึกษา โดยพิจารณา เฉพาะวิธีการตรวจสอบแต่ละวิธีการ ดังนี้ 1) กรณีการเปรียบเทียบ ค่าเฉลี่ยอัตราความถูกต้อง และ อัตราความคลาดเคลื่อนประเภทที่ 1 ของจำนวนข้อสอบที่ทำหน้าที่ต่างกัน กับปัจจัยความยาวของแบบสอบทั้งฉบับ โดยพิจารณาเฉพาะการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas และ 2) กรณีการเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของ จำนวนข้อสอบที่ทำหน้าที่ต่างกัน กับปัจจัยความยาวของแบบสอบทั้งฉบับ โดยพิจารณาเฉพาะ การวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl พบว่า

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas ภายใต้ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10 และร้อยละ 20 กับปัจจัยความยาวของแบบสอบทั้งฉบับ ทั้ง 2 ขนาด 40 และ 50 ข้อ มีอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 โดยจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10 มีค่าเฉลี่ยของ อัตรา ความคลาดเคลื่อนประเภทที่ 1 สูงกว่าร้อยละ 20 จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 20 มีประสิทธิภาพที่ดีและความยาวของแบบสอบทั้งฉบับ 50 ข้อ มีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าความยาวของแบบสอบทั้งฉบับ 40 ข้อ

วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl ภายใต้ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด ทั้งฉบับคิดเป็นร้อยละ 10 และร้อยละ 20 กับปัจจัยความยาวของแบบสอบทั้งฉบับ ทั้ง 2 ขนาด 40 และ 50 ข้อ มีอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่แตกต่างกัน

2.2.6 การทดสอบความแปรปรวนผลย่อย เงื่อนไขย่อยที่ 6

การทดสอบความแปรปรวนผลย่อย พิจารณาผลจาก ปฏิสัมพันธ์ สาม ทางระหว่างวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ กับปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน และ ปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบที่มีนัยสำคัญทางสถิติ ดังตารางที่ 4.14

ตารางที่ 4.14 การทดสอบความแปรปรวนผลย่อยระหว่าง วิธีการตรวจสอบการทำหน้าที่ ต่างกันของข้อสอบ (LR) กับ ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน (TYPE) และปัจจัย ขนาดของการทำหน้าที่ต่างกันของข้อสอบ (AMOUNT)

แหล่งของความแปรปรวน	CI		TE	
	F	p	F	p
ภายใต้เงื่อนไขเดียวกัน				
LR ที่ (TYPE1 × AMOUNT0.1)	21.302	.000	-	-
LRz (Mean=1.250, SD=4.734)				
LRj (Mean= 5.950, SD=9.016)				
LR ที่ (TYPE1 × AMOUNT0.2)	450.285	.000	-	-
LRz (Mean=.750, SD=7.500)				
LRj (Mean=38.600, SD=16.184)				
LR ที่ (TYPE1 × AMOUNT0.4)	876.796	.000	-	-
LRz (Mean=6.950, SD=10.117)				
LRj (Mean=57.275, SD=13.657)				
LR ที่ (TYPE2 × AMOUNT0.1)	.000	.000	-	-
LRz (Mean=.000, SD=.000)				
LRj (Mean=.000, SD=.000)				
LR ที่ (TYPE2 × AMOUNT0.2)	.485	.487	-	-
LRz (Mean=1.100, SD=4.637)				
LRj (Mean=1.650, SD=6.388)				
LR ที่ (TYPE2 × AMOUNT0.4)	58.175	.000	-	-
LRz (Mean=2.175, SD=6.273)				
LRj (Mean=18.350, SD=20.258)				
ภายใต้เงื่อนไขต่างกัน				
(TYPE× AMOUNT) ที่ LRz	8.528	.000	-	-
(TYPE1×AMOUNT0.1) (Mean=1.250, SD=4.734)				
(TYPE1×AMOUNT0.2) (Mean=.750, SD=7.500)				
(TYPE1×AMOUNT0.4) (Mean=6.950, SD=10.117)				
(TYPE2×AMOUNT0.1) (Mean=.000, SD=.000)				
(TYPE2×AMOUNT0.2) (Mean=1.100, SD=4.637)				
(TYPE2×AMOUNT0.4) (Mean=2.175, SD=6.273)				

ตารางที่ 4.14 (ต่อ)

แหล่งของความแปรปรวน	CI		TE	
	F	p	F	p
(TYPE× AMOUNT) ที่ LRj	104.612	.000	-	-
(TYPE1×AMOUNT0.1) (Mean=5.950, SD=9.016)				
(TYPE1×AMOUNT0.2) (Mean=38.600, SD=16.184)				
(TYPE1×AMOUNT0.4) (Mean=57.275, SD=13.657)				
(TYPE2×AMOUNT0.1) (Mean=.000, SD=.000)				
(TYPE2×AMOUNT0.2) (Mean=1.650, SD=6.388)				
(TYPE2×AMOUNT0.4) (Mean=18.350, SD=20.258)				

จากตารางที่ 4.14 การทดสอบความแปรปรวนผลย่อยระหว่าง วิธีการตรวจสอบ กับ ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน และ ปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ สรุปได้ดังนี้

1) การทดสอบความแปรปรวนผลย่อย เงื่อนไขย่อยที่ 6 ภายใต้เงื่อนไขเดียวกัน

การทดสอบความแปรปรวนผลย่อย ของวิธีการ ตรวจสอบการทำหน้าที่ต่างกัน ภายใต้เงื่อนไขเดียวกัน คือ การเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบภายใต้วิธีการที่ศึกษา โดยพิจารณาเฉพาะ เงื่อนไข แต่ละระดับ ของปัจจัยที่แปรเปลี่ยน ดังนี้ กรณี เปรียบเทียบ ค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบภายใต้วิธีถดถอยโลจิสติก โดยพิจารณาเฉพาะแบบสอบที่มี รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน แบบอเนกรูป ที่มีขนาดของการทำหน้าที่ต่างกันของข้อสอบขนาด 0.1 และอีก 5 กรณีที่ทำการทดสอบความแปรปรวนผลย่อยภายใต้วิธีถดถอยโลจิสติก คือ (1) พิจารณาเฉพาะแบบสอบที่มีรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป ที่มีขนาดของการทำหน้าที่ต่างกันของข้อสอบขนาด 0.2 (2) พิจารณาเฉพาะแบบสอบที่มีรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป ที่มีขนาดของ การทำหน้าที่ต่างกันของข้อสอบขนาด 0.4 (3) พิจารณาเฉพาะแบบสอบที่มีรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป ที่มีขนาดของการทำหน้าที่ต่างกันของข้อสอบขนาด 0.1 (4) พิจารณาเฉพาะแบบสอบที่มีรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป ที่มีขนาดของการทำหน้าที่ต่างกันของข้อสอบขนาด 0.2 และ (5) พิจารณาเฉพาะแบบสอบที่มีรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป ที่มีขนาดของ การทำหน้าที่ต่างกันของข้อสอบขนาด 0.4 มีผลการเปรียบเทียบค่าเฉลี่ยมีดังนี้

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง ผลการเปรียบเทียบค่าเฉลี่ยของ อัตราความถูกต้องภายใต้วิธีถดถอยโลจิสติก โดยพิจารณาเฉพาะรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน แบบอเนกรูปและเอกรูป กับขนาดของการทำหน้าที่ต่างกันของข้อสอบ 3 ขนาด คือ 0.1, 0.2 และ 0.4 พบว่ามีเพียงกรณีที่ รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน แบบเอกรูป ที่มีขนาดของ การทำหน้าที่ต่างกันของ

ข้อสอบขนาด 0.2 ที่ไม่พบความแตกต่าง นอกนั้นทุกเงื่อนไขดังกล่าว มี อัตราความถูกต้อง แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ซึ่งทุกกรณีที่ศึกษาพบว่าการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl มีค่าเฉลี่ยอัตราความถูกต้องสูงกว่าเกณฑ์ Zumbo and Thomas สอดคล้องกันในทุกกรณี

2) การทดสอบความแปรปรวนผลย่อย เงื่อนไขย่อยที่ 6 ภายใต้เงื่อนไขต่างกัน

การทดสอบความแปรปรวนผลย่อยใน ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน กับปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ ภายใต้เงื่อนไขต่างกัน คือ การเปรียบเทียบ ค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบ ภายใต้ปัจจัยที่ศึกษา โดยพิจารณาเฉพาะวิธีการตรวจสอบแต่ละวิธีการ ดังนี้ 1) กรณีการเปรียบเทียบ ค่าเฉลี่ยอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน กับ ปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ โดยพิจารณาเฉพาะ การวัดขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas และ 2) กรณีการเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน กับ ปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ โดยพิจารณาเฉพาะการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl พบว่า

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง วิธีถดถอยโลจิสติก โดย การวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas ภายใต้รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน แบบอเนกรูปและเอกรูป กับ ปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ 3 ขนาด คือ 0.1, 0.2 และ 0.4 มีอัตราความถูกต้อง แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ที่ระดับ .001 โดยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป ให้ค่าเฉลี่ยของอัตราความถูกต้องสูงกว่ารูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป และปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ ที่มีขนาด 0.4 ให้ค่าเฉลี่ยของ อัตราความถูกต้องสูงกว่าขนาดของการทำหน้าที่ต่างกันของข้อสอบที่มีขนาด 0.1 และ 0.2

วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl ภายใต้รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน แบบอเนกรูปและเอกรูป กับ ปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ 3 ขนาด คือ 0.1, 0.2 และ 0.4 มีอัตราความถูกต้องแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 โดยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน แบบอเนกรูป ให้ค่าเฉลี่ยของ อัตราความถูกต้องสูงกว่ารูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป และปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบที่มีขนาด 0.4 ให้ค่าเฉลี่ยของอัตราความถูกต้องสูงกว่าขนาด 0.1

2.2.7 การทดสอบความแปรปรวนผลย่อย เงื่อนไขย่อยที่ 7

การทดสอบความแปรปรวนผลย่อย พิจารณาผลจาก ปฏิสัมพันธ์ สามทางระหว่างวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ กับ ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน และ ปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบที่มีนัยสำคัญทางสถิติ ดังตารางที่ 4.15

ตารางที่ 4.15 การทดสอบความแปรปรวนผลย่อยระหว่าง วิธีการตรวจสอบการทำหน้าที่ ต่างกันของข้อสอบ (LR) กับ ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน (%DIF) และปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ (AMOUNT)

แหล่งของความแปรปรวน	CI		TE	
	F	p	F	p
ภายใต้เงื่อนไขเดียวกัน				
LR ที่ (10%DIF × AMOUNT0.1)	6.630	.058	-	-
LRz (Mean=.750, SD=4.286)				
LRj (Mean=2.200, SD=6.289)				
LR ที่ (10%DIF × AMOUNT0.2)	16.714	.000	-	-
LRz (Mean=1.650, SD=8.647)				
LRj (Mean=18.750, SD=22.455)				
LR ที่ (10%DIF × AMOUNT0.4)	50.502	.000	-	-
LRz (Mean=7.300, SD=11.019)				
LRj (Mean=47.550, SD=22.196)				
LR ที่ (20%DIF × AMOUNT0.1)	93.314	.000	-	-
LRz (Mean=.5000, SD=2.19043)				
LRj (Mean=3.7500, SD=7.64176)				
LR ที่ (20%DIF × AMOUNT0.2)	263.812	.000	-	-
LRz (Mean=.200, SD=1.407)				
LRj (Mean=21.500, SD=22.005)				
LR ที่ (20%DIF × AMOUNT0.4)	99.212	.000	-	-
LRz (Mean=1.825, SD=4.080)				
LRj (Mean=28.075, SD=26.036)				
ภายใต้เงื่อนไขต่างกัน				
(%DIF× AMOUNT) ที่ LRz	9.442	.000	-	-
(10%DIF ×AMOUNT0.1) (Mean=.750, SD=4.286)				
(10%DIF ×AMOUNT0.2) (Mean=1.650, SD=8.647)				
(10%DIF ×AMOUNT0.4) (Mean=7.300, SD=11.019)				
(20%DIF ×AMOUNT0.1) (Mean=.500, SD=2.190)				
(20%DIF ×AMOUNT0.2) (Mean=.200, SD=1.407)				
(20%DIF ×AMOUNT0.4) (Mean=1.825, SD=4.080)				

ตารางที่ 4.15 (ต่อ)

แหล่งของความแปรปรวน	CI		TE	
	F	p	F	p
(%DIF × AMOUNT) ที่ LRj	20.768	.000	-	-
(10%DIF ×AMOUNT0.1) (Mean=2.200, SD=6.289)				
(10%DIF ×AMOUNT0.2) (Mean=18.750, SD=22.455)				
(10%DIF ×AMOUNT0.4) (Mean=47.550, SD=22.196)				
(20%DIF ×AMOUNT0.1) (Mean=3.750,SD=7.642)				
(20%DIF ×AMOUNT0.2) (Mean=21.500, SD=22.005)				
(20%DIF ×AMOUNT0.4) (Mean=28.075, SD=26.036)				

จากตารางที่ 4.15 การทดสอบความแปรปรวนผลย่อยระหว่าง วิธีการตรวจสอบ กับ ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน และ ปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ สรุปได้ดังนี้

1) การทดสอบความแปรปรวนผลย่อย เงื่อนไขย่อยที่ 7 ภายใต้เงื่อนไขเดียวกัน

การทดสอบความแปรปรวนผลย่อย ของวิธีการ ตรวจสอบการทำหน้าที่ต่างกัน ภายใต้เงื่อนไขเดียวกัน คือ การเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบภายใต้วิธีการที่ศึกษา โดยพิจารณาเฉพาะ เงื่อนไข แต่ละระดับ ของปัจจัยที่แปรเปลี่ยน ดังนี้ กรณี การเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบภายใต้วิธีถดถอยโลจิสติก โดยพิจารณาเฉพาะแบบสอบที่มีปัจจัย จำนวนข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 10 ที่มีขนาดของการทำหน้าที่ต่างกันของข้อสอบขนาด 0.1 และอีก 5 กรณีที่ทำการทดสอบความแปรปรวนผลย่อยภายใต้วิธีถดถอยโลจิสติก คือ (1) พิจารณาเฉพาะ แบบสอบที่มี จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10 มีขนาดของ การทำหน้าที่ต่างกันของข้อสอบ ขนาด 0.2 (2) พิจารณาเฉพาะแบบสอบที่มีจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10 มีขนาดของการทำหน้าที่ต่างกันของข้อสอบ ขนาด 0.4 (3) พิจารณาเฉพาะ แบบสอบที่มี จำนวนข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 20 มีขนาดของ การทำหน้าที่ต่างกันของข้อสอบ ขนาด 0.1 (4) พิจารณาเฉพาะแบบสอบที่มีจำนวนข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 20 มีขนาดของการทำหน้าที่ต่างกันของข้อสอบขนาด 0.2 และ (5) พิจารณาเฉพาะแบบสอบที่มี จำนวนข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 20 มีขนาดของ การทำหน้าที่ต่างกันของข้อสอบ ขนาด 0.4 ผลการเปรียบเทียบค่าเฉลี่ยมีดังนี้

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง ผลการเปรียบเทียบค่าเฉลี่ยของ อัตราความถูกต้องภายใต้วิธีถดถอยโลจิสติก โดยพิจารณาเฉพาะจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10 และร้อยละ 20 กับขนาดของการทำหน้าที่ต่างกันของข้อสอบ 3 ขนาด คือ 0.1, 0.2 และ

0.4 พบว่าทุกกรณี มี อัตรา ความถูกต้อง แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ที่ระดับ .001 ทุกกรณี ที่ศึกษาพบว่า การวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl มีค่าเฉลี่ยอัตราความถูกต้อง สูงกว่า เกณฑ์ Zumbo and Thomas สอดคล้องกันทุกกรณี

2) การทดสอบความแปรปรวนผลย่อย เงื่อนไขย่อยที่ 7 ภายใต้เงื่อนไขต่างกัน

การทดสอบความแปรปรวนผลย่อยในปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน กับ ปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ ภายใต้เงื่อนไขต่างกัน คือการเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบภายใต้ปัจจัยที่ศึกษา โดยพิจารณา เฉพาะวิธีการตรวจสอบแต่ละวิธีการ ดังนี้ 1) กรณีการเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของจำนวนข้อสอบที่ทำหน้าที่ต่างกัน กับ ปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ โดยพิจารณาเฉพาะการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas และ 2) กรณีการเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของจำนวนข้อสอบที่ทำหน้าที่ต่างกันกับปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ โดยพิจารณาเฉพาะ การวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl พบว่า

ผลการเปรียบเทียบค่าเฉลี่ยของอัตรา ความถูกต้อง วิธีถดถอยโลจิสติก โดย การวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas ภายใต้จำนวนข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 10 และร้อยละ 20 กับ ปัจจัยขนาดของ การทำหน้าที่ต่างกันของข้อสอบ 3 ขนาด คือ 0.1, 0.2 และ 0.4 มีอัตรา ความถูกต้อง แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ที่ระดับ .001 โดยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 10 ให้ค่าเฉลี่ยของ อัตรา ความถูกต้องสูงกว่า จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 20 และปัจจัยขนาดของ การทำหน้าที่ต่างกันของข้อสอบ ที่มีขนาด 0.4 ให้ค่าเฉลี่ยของ อัตรา ความถูกต้องสูงกว่า ขนาดของ การทำหน้าที่ต่างกันของข้อสอบ ที่มีขนาด 0.1 และ 0.2

วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl ภายใต้ จำนวนข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 10 และร้อยละ 20 กับปัจจัยขนาดของ การทำหน้าที่ต่างกันของข้อสอบ 3 ขนาด คือ 0.1, 0.2 และ 0.4 มีอัตราความถูกต้องแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 โดยจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10 ให้ค่าเฉลี่ยของ อัตราความถูกต้องสูงกว่าจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 20 และปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ ที่มีขนาด 0.4 ให้ค่าเฉลี่ยของ อัตรา ความถูกต้องสูงกว่า ขนาดของ การทำหน้าที่ต่างกันของข้อสอบที่มีขนาด 0.1 และ 0.2

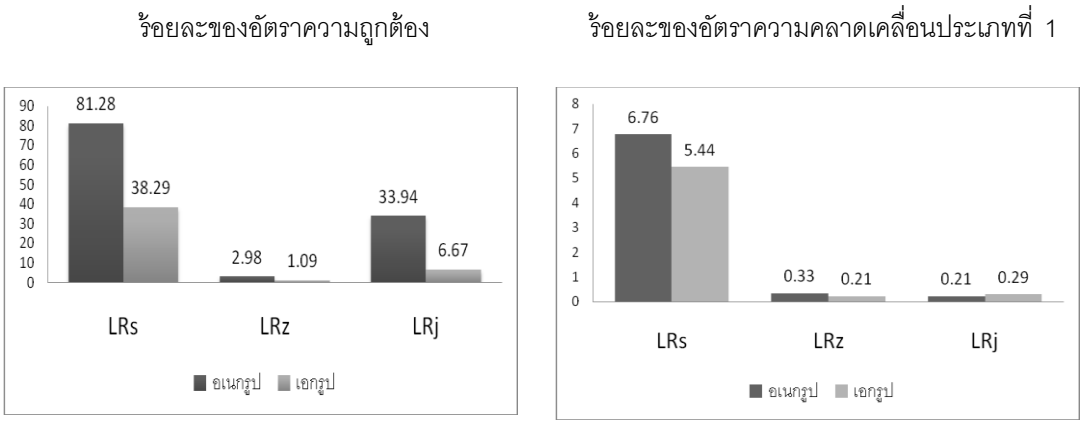
ตอนที่ 3 สรุปผลการเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1

การวัดขนาดอิทธิพลและผลของประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในวิธีถดถอยโลจิสติก สำหรับข้อสอบที่มีรูปแบบการตรวจให้คะแนนแบบทวิวิภาค กรณีข้อมูลจำลองและข้อมูลเชิงประจักษ์ ภายใต้เงื่อนไขของปัจจัยหลัก 4 ปัจจัย สรุปได้ดังนี้

3.1 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

3.1.1 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ตามปัจจัยที่ศึกษา

สรุปผลในรูปค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ปัจจัยที่ศึกษา คือ ภายใต้วิธีถดถอยโลจิสติก ระหว่างการทดสอบระดับนัยสำคัญ (LR_S) ขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl (LR_J) และขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas (LR_Z) แสดงดังภาพประกอบ 4.2–4.5



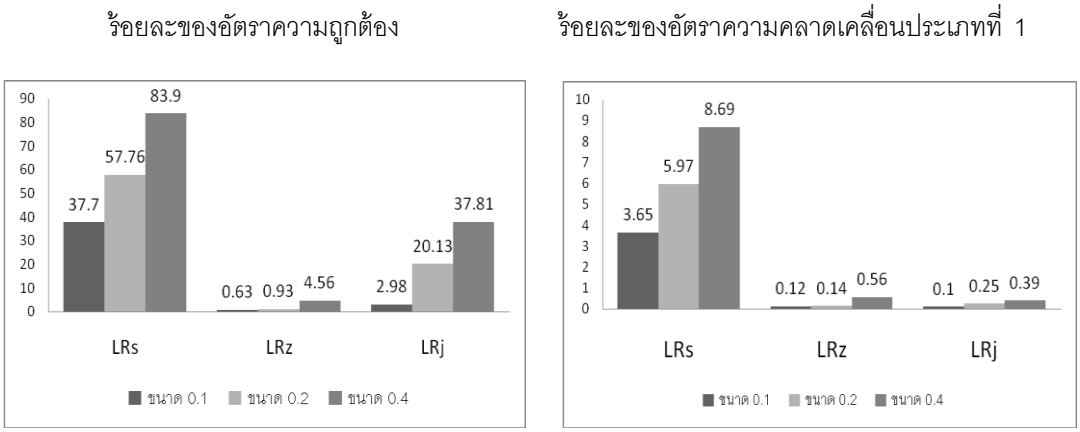
ภาพที่ 4.2 อัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1
ภายใต้ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน 2 เงื่อนไข

จากภาพประกอบ 4.2 แสดงอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ตามลำดับ ภายใต้ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน 2 เงื่อนไข คือ ข้อสอบทำหน้าที่ ต่างกันแบบเอกรูป และข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป สรุปได้ดังนี้ เงื่อนไขข้อสอบทำหน้าที่ต่างกันแบบอเนกรูปภายใต้วิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ มีค่าเฉลี่ยอัตราความถูกต้องสูงสุดคิดเป็นร้อยละ 81.28 ส่วนค่าเฉลี่ยอัตราความถูกต้องต่ำที่สุดอยู่ในเงื่อนไขข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป โดยขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas มีค่าต่ำสุดคิดเป็นร้อยละ 1.09

เงื่อนไขข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป โดยการทดสอบระดับนัยสำคัญ มีค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุดคิดเป็นร้อยละ 6.76 ส่วนค่าเฉลี่ยอัตราความคลาดเคลื่อน

ประเภท ที่ 1 ต่ำที่สุดอยู่ในเงื่อนไขข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป โดยการวัดขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas ค่าต่ำสุดคิดเป็นร้อยละ 0.21 และอยู่ในเงื่อนไขข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป โดยการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl มีค่าต่ำสุด คิดเป็นร้อยละ 0.21 อยู่ในเงื่อนไขข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป

สรุปปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน พบว่า ร้อยละเฉลี่ยของอัตราความถูกต้องในรูปแบบข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูปให้ค่าที่ต่ำกว่าแบบอเนกรูป เมื่อขนาดของ การทำหน้าที่ต่างกันแบบเอกรูปยิ่งมากค่าร้อยละเฉลี่ยของ อัตราความถูกต้องจะมีค่าสูงขึ้นด้วย ค่าร้อยละเฉลี่ยของอัตราความถูกต้องในรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป จะมีค่าสูงขึ้นเมื่อขนาดของการทำหน้าที่ต่างกันสูงขึ้นจากขนาด 0.1 เป็น 0.4 เมื่อความยาวข้อสอบมีค่าเพิ่มขึ้นจาก 40 เป็น 50 ข้อ และเมื่อแบบสอบที่มีข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10 ของความยาวทั้งฉบับ ร้อยละเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ จากขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas มีค่าต่ำสุดอยู่ในเงื่อนไขข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป สำหรับขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl มีค่าต่ำสุดอยู่ในเงื่อนไขข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป

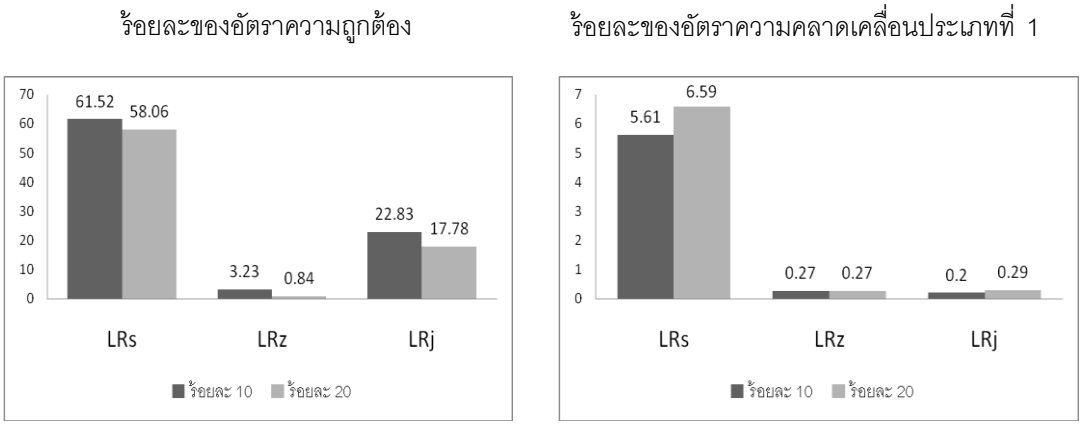


ภาพที่ 4.3 อัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1
ภายใต้ปัจจัยขนาดของการทำหน้าที่ต่างกัน 3 เงื่อนไข

จากภาพประกอบ 4.3 แสดงอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ตามลำดับ ภายใต้ปัจจัย ขนาดของการทำหน้าที่ต่างกัน 3 เงื่อนไข คือ ขนาด 0.1, 0.2 และ 0.4 ดังนี้ เงื่อนไขขนาดของการทำหน้าที่ต่างกัน 0.4 ภายใต้วิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ มีค่าเฉลี่ยอัตราความถูกต้องสูงสุด คิดเป็นร้อยละ 83.90 ส่วนค่าเฉลี่ยอัตราความถูกต้องต่ำที่สุดอยู่ในเงื่อนไขขนาดของการทำหน้าที่ต่างกัน 0.1 โดยขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas คิดเป็นร้อยละ 0.63 เงื่อนไข ขนาดของการทำหน้าที่ต่างกัน 0.4 จากการตรวจสอบ DIF ภายใต้วิธีถดถอย

โลจิสติก โดยการทดสอบระดับนัยสำคัญ มีค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุดคิดเป็นร้อยละ 8.69 ส่วนค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำที่สุดอยู่ในเงื่อนไข ขนาดของการทำหน้าที่ต่างกัน 0.1 โดยการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl มีค่าต่ำสุดคิดเป็นร้อยละ 0.10

สรุปปัจจัยขนาดของการทำหน้าที่ต่างกัน พบว่า ร้อยละเฉลี่ยของอัตราความถูกต้อง เมื่อมีขนาดของการทำหน้าที่ต่างกันสูงขึ้นก็ส่งผลให้อัตราความถูกต้องสูงขึ้นตามไปด้วยในทุกวิธีที่ศึกษา สามารถเรียงลำดับอัตราความถูกต้องสูงสุดคือวิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ ขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl และสุดท้ายคือ ขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas อัตราความคลาดเคลื่อนประเภทที่ 1 ที่คำนวณได้จากการตรวจสอบด้วยวิธี ถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ ให้ค่าร้อยละเฉลี่ยสูงกว่าการวัดขนาดอิทธิพลทั้งสองเกณฑ์ในทุกระดับของขนาดของการทำหน้าที่ต่างกัน เมื่อพิจารณาเฉพาะการวัดขนาดอิทธิพล พบว่ามีความสอดคล้องกัน คือ ขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas และขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl มีอัตราความคลาดเคลื่อนประเภทที่1 ต่ำที่สุดเมื่อขนาดของการทำหน้าที่ต่างกันเป็น 0.1 และสูงสุดเมื่อขนาดของการทำหน้าที่ต่างกันเป็น 0.4

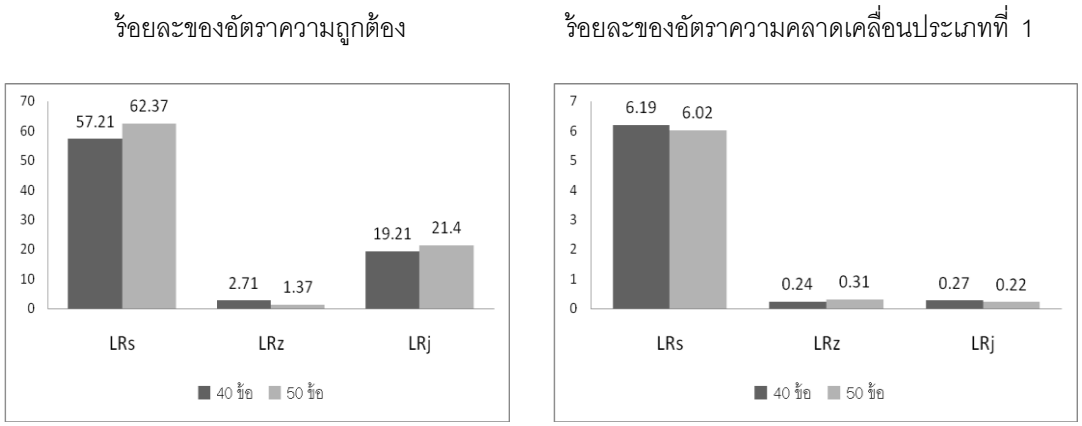


ภาพที่ 4.4 อัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่1
ภายใต้ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน2 เงื่อนไข

จากภาพประกอบ 4.4 แสดงอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ตามลำดับ ภายใต้ปัจจัย จำนวนข้อสอบที่ทำหน้าที่ต่างกัน 2 เงื่อนไข คือ จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10 และทั้งฉบับคิดเป็นร้อยละ 20 ดังนี้ เงื่อนไข จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับร้อยละ 10 จากการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ มีค่าเฉลี่ยอัตราความถูกต้องสูงสุดคิดเป็นร้อยละ 61.52 ส่วนค่าเฉลี่ยอัตราความถูกต้องต่ำสุดอยู่ในเงื่อนไข จำนวนข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับร้อยละ 20 โดยการวัดขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas มีค่าต่ำสุดคิดเป็นร้อยละ 0.84

เงื่อนไข ขนาดของการทำหน้าที่ต่างกัน 0.4 จากการตรวจสอบ DIF ภายใต้วิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ มีค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุดคิดเป็นร้อยละ 8.69 ส่วนค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำที่สุดอยู่ในเงื่อนไข ขนาดของการทำหน้าที่ต่างกัน 0.1 โดยการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl มีค่าต่ำสุดคิดเป็นร้อยละ 0.10

สรุปปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน พบว่า แบบสอบที่มีข้อสอบทำหน้าที่ต่างกันทั้งฉบับร้อยละ 10 ภายใต้เงื่อนไขนี้พบว่าค่าร้อยละเฉลี่ยของอัตราความถูกต้องสูงกว่าแบบสอบที่มีข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 20 วิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญและการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas และเกณฑ์ Jodoin and Gierl ให้ผลสอดคล้องตรงกัน อัตราความคลาดเคลื่อนประเภทที่ ที่คำนวณได้จากการตรวจสอบด้วยวิธี ถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญมีค่าสูงกว่าขนาดอิทธิพล จากการวัดขนาดอิทธิพลทั้ง 2 เกณฑ์ โดย อัตราความคลาดเคลื่อนประเภทที่ ตามเกณฑ์ Zumbo and Thomas มีค่าต่ำสุดในเงื่อนไขข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 20 สำหรับขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl มีค่าต่ำสุดในเงื่อนไขข้อสอบทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10



ภาพที่ 4.5 อัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1
ภายใต้ปัจจัยความยาวของแบบสอบทั้งฉบับ 2 เงื่อนไข

จากภาพที่ 4.5 แสดงอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ตามลำดับ ภายใต้ปัจจัยความยาวของแบบสอบทั้งฉบับ 2 เงื่อนไข คือความยาว 40 และ 50 ข้อ สรุปได้ว่าเงื่อนไข ความยาวของแบบสอบทั้งฉบับ 50 ข้อ จากการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้วิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ มีค่าเฉลี่ยอัตราความถูกต้องสูงสุด คิดเป็นร้อยละ 62.37 ส่วนค่าเฉลี่ยอัตราความถูกต้องต่ำสุดอยู่ในเงื่อนไขความยาวของแบบสอบทั้งฉบับ 50 ข้อ โดยการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas คิดเป็นร้อยละ 1.37

เงื่อนไขความยาวของแบบสอบทั้งฉบับ 40 ข้อ จากการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้วิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ มีค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุดคิดเป็นร้อยละ 6.19 ส่วนค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำที่สุดอยู่ในเงื่อนไขความยาวแบบสอบทั้งฉบับ 50 ข้อ โดยการวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl มีค่าต่ำสุดคิดเป็นร้อยละ 0.22

สรุปปัจจัยความยาวของแบบสอบทั้งฉบับ พบว่า ปัจจัยความยาวของแบบสอบทั้งฉบับมีนัยสำคัญทางสถิติในกรณีเดียวเท่านั้นเมื่อข้อสอบมีจำนวนข้อ เพิ่มขึ้น ผลจากวิธีตรวจสอบ ภายใต้วิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญกับการวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl ให้ผลตรงกันคือ มีร้อยละเฉลี่ยของ อัตราความถูกต้องสูงขึ้น ร้อยละเฉลี่ยของ อัตราความคลาดเคลื่อนประเภทที่1 ในแบบสอบที่มีความยาว40 ข้อ สูงกว่าความยาว50 ข้อ ให้ผลเป็นไปในทิศทางเดียวกัน ในทุกวิธีตรวจสอบ

3.1.2 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ตามวิธีการตรวจสอบที่ศึกษา

สรุปผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สรุปผลในรูปค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ ภายใต้วิธีการที่ศึกษาดังนี้

1. วิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ ร้อยละเฉลี่ยสูงสุดของอัตราความถูกต้องจากการตรวจสอบโดยการทดสอบระดับนัยสำคัญ คือภายใต้เงื่อนไขรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปที่มีความยาวของแบบสอบทั้งฉบับเป็น 40 ข้อ จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10 มีขนาดของการทำหน้าที่ต่างกันขนาดใหญ่ ส่วนร้อยละเฉลี่ย ของอัตราความถูกต้องต่ำสุด คือ ภายใต้เงื่อนไขรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป ความยาวของแบบสอบทั้งฉบับเป็น 40 ข้อ จำนวนข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 10 มีขนาดของการทำหน้าที่ต่างกันขนาดเล็กน้อย สรุปได้ว่าเมื่อข้อสอบเกิดการทำหน้าที่ต่างกันแบบอเนกรูป ส่งผลให้อัตราความถูกต้องสูงกว่าข้อสอบทำหน้าที่ต่างกันแบบเอกรูป ซึ่งการทำหน้าที่ต่างกันขนาดใหญ่กว่าจะพบอัตราความถูกต้องสูงกว่าการทำหน้าที่ต่างกันที่มีขนาดเล็ก หากในแบบสอบมีจำนวนข้อที่ทำหน้าที่ต่างกันหลายข้อจะส่งผลให้อัตราความถูกต้องลดลง เมื่อพิจารณาในทุกวิธีที่ศึกษา พบว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ มีอัตราความถูกต้องสูงที่สุด ร้อยละเฉลี่ยสูงสุดของอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้เงื่อนไขรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปที่มีความยาวของแบบสอบทั้งฉบับเป็น 40 ข้อ จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 20 มีขนาดของการทำหน้าที่ต่างกันขนาดใหญ่ ส่วนร้อยละเฉลี่ยต่ำสุดของอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้เงื่อนไขรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป มีความยาวของแบบสอบทั้งฉบับเป็น 40 ข้อ มีจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 20 มีขนาดของการทำหน้าที่ต่างกันขนาดเล็กน้อย สรุปได้ว่า เมื่อข้อสอบทำหน้าที่ต่างกันทั้งแบบ

เอกรูปและแบบอเนกรูป ในการทำหน้าที่ต่างกันที่มีขนาดใหญ่จะพบอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าการทำหน้าที่ต่างกันที่มีขนาดเล็กน้อยในทุกเงื่อนไข

2. ขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas ร้อยละเฉลี่ยสูงสุดของ อัตราความถูกต้อง คือ ภายใต้เงื่อนไขรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปที่มีความยาวของแบบสอบทั้งฉบับ 40 ข้อ มีจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10 มีขนาดของการทำหน้าที่ต่างกันขนาดใหญ่ ส่วนร้อยละเฉลี่ยต่ำสุดของอัตราความถูกต้องพบในการทำหน้าที่ต่างกันแบบเอกรูปและอเนกรูปด้วยจำนวนตามเงื่อนไขเท่าๆ กัน สรุปได้ว่า การตรวจสอบการทำหน้าที่ต่างกันด้วยวิธีถดถอยโลจิสติกตามเกณฑ์ Zumbo and Thomas มีผลของอัตราความถูกต้องคล้าย ผลที่ได้จากการตรวจสอบโดย การทดสอบระดับนัยสำคัญ นั่นคือการทำหน้าที่ต่างกันที่มีขนาดใหญ่พบอัตราความถูกต้องสูงกว่าการทำหน้าที่ต่างกันที่มีขนาดเล็ก และการทำหน้าที่ต่างกันแบบอเนกรูปที่มีขนาดปานกลางจะมีอัตราความถูกต้องสูงกว่าการทำหน้าที่ต่างกันแบบอเนกรูปที่มีขนาดเล็ก ร้อยละเฉลี่ยสูงสุดของอัตราความคลาดเคลื่อนประเภทที่ 1 คือ ภายใต้เงื่อนไขรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป ที่มีความยาวของแบบสอบทั้งฉบับเป็น 40 ข้อ จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10 มีขนาดของการทำหน้าที่ต่างกันขนาดใหญ่ ส่วนร้อยละเฉลี่ยต่ำสุดพบว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ในเงื่อนไขการทำหน้าที่ต่างกันแบบอเนกรูปและเอกรูปมีค่าเท่ากัน

3. ขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl ร้อยละเฉลี่ยสูงสุดของ อัตราความถูกต้อง คือ ภายใต้เงื่อนไขรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป ที่มีความยาวของแบบสอบทั้งฉบับเป็น 50 ข้อ จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10 มีขนาดของการทำหน้าที่ต่างกันขนาดใหญ่ ส่วนร้อยละเฉลี่ยต่ำสุดของ อัตราความถูกต้อง พบการทำหน้าที่ต่างกันแบบเอกรูปมากกว่าแบบอเนกรูป สรุปได้ว่าเมื่อข้อสอบทำหน้าที่ต่างกันแบบเอกรูป การตรวจสอบการทำหน้าที่ต่างกัน ด้วยวิธีถดถอยโลจิสติก ตามเกณฑ์ Jodoin and Gierl จะไม่สามารถตรวจพบข้อสอบที่ทำหน้าที่ต่างกันกรณีการทำหน้าที่ต่างกัน มีขนาดเล็กน้อย และเมื่อการทำหน้าที่ต่างกันมีขนาดใหญ่ อัตราความถูกต้องจะสูงขึ้น เมื่อข้อสอบทำหน้าที่ต่างกันทั้งแบบเอกรูปและแบบอเนกรูป การทำหน้าที่ต่างกันที่มีขนาดเล็กน้อยจะไม่เกิด อัตราความคลาดเคลื่อนประเภทที่ 1 ส่วนร้อยละเฉลี่ยสูงสุดของอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้เงื่อนไขรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปที่มีความยาวของแบบสอบทั้งฉบับเป็น 40 ข้อ จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 20 มีขนาดของการทำหน้าที่ต่างกันขนาดใหญ่และพบภายใต้เงื่อนไขรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูปที่มีความยาวของแบบสอบทั้งฉบับเป็น 40 ข้อ จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 20 มีขนาดของการทำหน้าที่ต่างกันขนาดใหญ่

3.2 การเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีถดถอยโลจิสติก ระหว่าง การวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl และเกณฑ์ Zumbo and Thomas ภายใต้ปัจจัยที่ศึกษา

3.2.1 ผลกา รทดสอบปัจจัยที่ศึกษาที่มีผลต่อ อัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

ผลการทดสอบปัจจัยที่ศึกษาได้แก่ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ที่มีผลต่อประสิทธิภาพด้านอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl และเกณฑ์ Zumbo and Thomas โดยการทดสอบผลย่อย (Simple effect) ดังตารางที่ 4.16

ตารางที่ 4.16 ปัจจัยที่มีผลต่ออัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ในวิธีถดถอยโลจิสติก ของการวัดขนาดอิทธิพล 2 เกณฑ์

ปัจจัยที่แปรเปลี่ยน	การวัดขนาดอิทธิพล	
	CI	TE
TYPE	Sig	N-Sig
LENGTH	Sig	N-Sig
AMOUNT	Sig	N-Sig
TYPE * AMOUNT	Sig	N-Sig
TYPE * %DIF	Sig	N-Sig
LENGTH * %DIF	N-Sig	Sig
AMOUNT * %DIF	Sig	N-Sig

Sig = มีนัยสำคัญ, N-Sig = ไม่มีนัยสำคัญ

การทดสอบผลย่อยตามตารางที่ 4.16 สรุปได้ว่า ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันขนาดของการทำหน้าที่ต่างกัน ปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน ระหว่าง รูปแบบของข้อสอบที่ทำหน้าที่ต่างกันกับขนาดของการทำหน้าที่ต่างกัน และระหว่าง จำนวนข้อสอบที่ทำหน้าที่ต่างกัน กับขนาดของการทำหน้าที่ต่างกัน มีผลต่ออัตราความถูกต้องของวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ของ Zumbo and Thomas อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ยกเว้นความยาวของแบบสอบทั้งฉบับและปฏิสัมพันธ์ระหว่างรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน กับจำนวนข้อสอบที่ทำหน้าที่ต่างกัน ในขณะเดียวกันปฏิสัมพันธ์ระหว่างความยาวของแบบสอบทั้งฉบับกับจำนวน

ข้อสอบที่ทำหน้าที่ต่างกัน มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ของ Zumbo and Thomas อย่างมีนัยสำคัญทางสถิติที่ระดับ .001

ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันขนาดของการทำหน้าที่ต่างกัน ปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยนระหว่างรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันกับขนาดของการทำหน้าที่ต่างกัน และระหว่างจำนวนข้อสอบที่ทำหน้าที่ต่างกัน กับขนาดของการทำหน้าที่ต่างกัน มีผลต่ออัตราความถูกต้องของวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ของ Jodoin and Gierl อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ยกเว้น ความยาวของแบบสอบทั้งฉบับ และปฏิสัมพันธ์ระหว่าง รูปแบบของข้อสอบที่ทำหน้าที่ต่างกันกับจำนวนข้อสอบที่ทำหน้าที่ต่างกัน ในขณะเดียวกันปฏิสัมพันธ์ระหว่างความยาวของแบบสอบทั้งฉบับกับจำนวนข้อสอบที่ทำหน้าที่ต่างกัน ไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ของ Jodoin and Gierl

3.2.2 ผลการเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีถดถอยโลจิสติก ระหว่าง การวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl และเกณฑ์ Zumbo and Thomas ภายใต้ปัจจัยที่ศึกษา

ผลการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีถดถอยโลจิสติก ระหว่าง การวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl และขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย (รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน 2 รูปแบบ ขนาดของการทำหน้าที่ต่างกัน 3 ขนาด จำนวนข้อสอบที่มีการทำหน้าที่ต่างกัน 2 ขนาด และความยาวของแบบสอบทั้งฉบับ 2 ขนาด) และปฏิสัมพันธ์สองทางระหว่างปัจจัย ที่แปรเปลี่ยน 4 ปัจจัย สรุปได้ดังนี้

สรุปผลการศึกษาตามวัตถุประสงค์การวิจัย ข้อที่ 1

การเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl และขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas **ภายใต้เงื่อนไขเดียวกัน** ของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ดังตารางที่ 4.17

ตารางที่ 4.17 สรุปผลการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl และเกณฑ์ Zumbo and Thomas **ภายใต้เงื่อนไขเดียวกัน**

ปัจจัย	การวัดขนาดอิทธิพล 2 เกณฑ์ ระหว่างเกณฑ์ Jodoin and Gierl และ เกณฑ์ Zumbo and Thomas	
	อัตราความถูกต้อง	อัตราความคลาดเคลื่อนประเภทที่ 1
TYPE		
อเนกรูป	Jodoin and Gierl ให้ผลที่สูงกว่า	ไม่มีการเปรียบเทียบ
เอกรูป	Jodoin and Gierl ให้ผลที่สูงกว่า	ไม่มีการเปรียบเทียบ
LENGTH		
40 ข้อ	Jodoin and Gierl ให้ผลที่สูงกว่า	ไม่มีการเปรียบเทียบ
50 ข้อ	Jodoin and Gierl ให้ผลที่สูงกว่า	ไม่มีการเปรียบเทียบ
AMOUNT		
ขนาด DIF 0.1	Jodoin and Gierl ให้ผลที่สูงกว่า	ไม่มีการเปรียบเทียบ
ขนาด DIF 0.2	Jodoin and Gierl ให้ผลที่สูงกว่า	ไม่มีการเปรียบเทียบ
ขนาด DIF 0.4	Jodoin and Gierl ให้ผลที่สูงกว่า	ไม่มีการเปรียบเทียบ
TYPE * AMOUNT		
อเนกรูป กับ ขนาด DIF 0.1	Jodoin and Gierl ให้ผลที่สูงกว่า	ไม่มีการเปรียบเทียบ
อเนกรูป กับ ขนาด DIF 0.2	Jodoin and Gierl ให้ผลที่สูงกว่า	ไม่มีการเปรียบเทียบ
อเนกรูป กับ ขนาด DIF 0.4	Jodoin and Gierl ให้ผลที่สูงกว่า	ไม่มีการเปรียบเทียบ
เอกรูป กับ ขนาด DIF 0.1	ไม่แตกต่างอย่างมีนัยสำคัญ	ไม่มีการเปรียบเทียบ
เอกรูป กับ ขนาด DIF 0.2	ไม่แตกต่างอย่างมีนัยสำคัญ	ไม่มีการเปรียบเทียบ
เอกรูป กับ ขนาด DIF 0.4	Jodoin and Gierl ให้ผลที่สูงกว่า	ไม่มีการเปรียบเทียบ
TYPE * %DIF		
อเนกรูป กับ 10%DIF	Jodoin and Gierl ให้ผลที่สูงกว่า	ไม่มีการเปรียบเทียบ
อเนกรูป กับ 20%DIF	Jodoin and Gierl ให้ผลที่สูงกว่า	ไม่มีการเปรียบเทียบ
เอกรูป กับ 10%DIF	Jodoin and Gierl ให้ผลที่สูงกว่า	ไม่มีการเปรียบเทียบ
เอกรูป กับ 20%DIF	Jodoin and Gierl ให้ผลที่สูงกว่า	ไม่มีการเปรียบเทียบ

ตารางที่ 4.17 (ต่อ)

ปัจจัย	การวัดขนาดอิทธิพล 2 เกณฑ์ ระหว่างเกณฑ์ Jodoin and Gierl และ เกณฑ์ Zumbo and Thomas	
	อัตราความถูกต้อง	อัตราความคลาดเคลื่อนประเภทที่ 1
LENGTH * %DIF		
40 ข้อ กับ 10%DIF	ไม่มีการเปรียบเทียบ	ไม่แตกต่างอย่างมีนัยสำคัญ
40 ข้อ กับ 20%DIF	ไม่มีการเปรียบเทียบ	ไม่แตกต่างอย่างมีนัยสำคัญ
50 ข้อ กับ 10%DIF	ไม่มีการเปรียบเทียบ	ไม่แตกต่างอย่างมีนัยสำคัญ
50 ข้อ กับ 20%DIF	ไม่มีการเปรียบเทียบ	ไม่แตกต่างอย่างมีนัยสำคัญ
AMOUNT * %DIF		
ขนาด DIF 0.1 กับ 10%DIF	Jodoin and Gierl ให้ผลที่สูงกว่า	ไม่มีการเปรียบเทียบ
ขนาด DIF 0.2 กับ 10%DIF	Jodoin and Gierl ให้ผลที่สูงกว่า	ไม่มีการเปรียบเทียบ
ขนาด DIF 0.4 กับ 10%DIF	Jodoin and Gierl ให้ผลที่สูงกว่า	ไม่มีการเปรียบเทียบ
ขนาด DIF 0.1 กับ 20%DIF	Jodoin and Gierl ให้ผลที่สูงกว่า	ไม่มีการเปรียบเทียบ
ขนาด DIF 0.2 กับ 20%DIF	Jodoin and Gierl ให้ผลที่สูงกว่า	ไม่มีการเปรียบเทียบ
ขนาด DIF 0.4 กับ 20%DIF	Jodoin and Gierl ให้ผลที่สูงกว่า	ไม่มีการเปรียบเทียบ

จากตารางที่ 4.17 พบว่า วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl มีอัตราความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สูงกว่าเกณฑ์ Zumbo and Thomas และภายใต้ปฏิสัมพันธ์ระหว่างปัจจัยที่แปรเปลี่ยนเกือบทุกเงื่อนไข

สรุปผลการศึกษาตามวัตถุประสงค์การวิจัย ข้อที่ 2

การเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl และขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas **ภายใต้เงื่อนไขต่างกัน** ของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ดังตารางที่ 4.18

ตารางที่ 4.18 สรุปผลการเปรียบเทียบอัตราความถูกต้อง (CI) และอัตราความคลาดเคลื่อนประเภทที่ 1 (TE) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพลตามเกณฑ์Jodoin and Gierl และเกณฑ์ Zumbo and Thomas **ภายใต้เงื่อนไขต่างกัน**

ปัจจัย		การวัดขนาดอิทธิพล 2 เกณฑ์			
		Zumbo and Thomas		Jodoin and Gierl	
		CI	TE	CI	TE
TYPE					
อเนกรูป	เอกรูป	อเนกรูป**สูงกว่า**เอกรูป	ไม่มีการเปรียบเทียบ	อเนกรูป**สูงกว่า**เอกรูป	ไม่มีการเปรียบเทียบ
LENGTH					
40 ข้อ	50 ข้อ	ไม่แตกต่างอย่างมีนัยสำคัญ	ไม่มีการเปรียบเทียบ	ไม่แตกต่างอย่างมีนัยสำคัญ	ไม่มีการเปรียบเทียบ
AMOUNT					
DIF0.1	DIF0.2	0.2 **สูงกว่า** 0.1	ไม่มีการเปรียบเทียบ	0.2 **สูงกว่า** 0.1	ไม่มีการเปรียบเทียบ
	DIF0.4	0.4 **สูงกว่า** 0.1	ไม่มีการเปรียบเทียบ	0.4 **สูงกว่า** 0.1	ไม่มีการเปรียบเทียบ
TYPE * %DIF					
อเนกรูป กับ 10%DIF	อเนกรูป กับ 20%DIF	ไม่แตกต่างอย่างมีนัยสำคัญ	ไม่มีการเปรียบเทียบ	ไม่แตกต่างอย่างมีนัยสำคัญ	ไม่มีการเปรียบเทียบ
	เอกรูป กับ 10%DIF	ไม่แตกต่างอย่างมีนัยสำคัญ	ไม่มีการเปรียบเทียบ	ไม่แตกต่างอย่างมีนัยสำคัญ	ไม่มีการเปรียบเทียบ
	เอกรูป กับ 20%DIF	ไม่แตกต่างอย่างมีนัยสำคัญ	ไม่มีการเปรียบเทียบ	ไม่แตกต่างอย่างมีนัยสำคัญ	ไม่มีการเปรียบเทียบ
LENGTH * %DIF					
40 ข้อ กับ 10% DIF	40 ข้อ กับ 20%DIF	ไม่มีการเปรียบเทียบ	40 ข้อ กับ 10% DIF **สูงกว่า** 40 ข้อ กับ 20%DIF	ไม่มีการเปรียบเทียบ	40 ข้อ กับ 10% DIF**ต่ำกว่า**
	50 ข้อ กับ 10%DIF	ไม่มีการเปรียบเทียบ	40 ข้อ กับ 10% DIF **สูงกว่า** 50 ข้อ กับ 10%DIF	ไม่มีการเปรียบเทียบ	40 ข้อ กับ 10% DIF **ต่ำกว่า** 50 ข้อ กับ 10%DIF

ตารางที่ 4.18 (ต่อ)

ปัจจัย		การวัดขนาดอิทธิพล 2 เกณฑ์			
		Zumbo and Thomas		Jodoin and Gierl	
		CI	TE	CI	TE
LENGTH * %DIF					
40 ข้อ กับ 10% DIF	50 ข้อ กับ 20%DIF	ไม่มีการเปรียบเทียบ	40 ข้อ กับ 10% DIF **ต่ำกว่า** 50 ข้อ กับ 20%DIF	ไม่มีการเปรียบเทียบ	40 ข้อ กับ 10% DIF **ต่ำกว่า** 50 ข้อ กับ 20%DIF
TYPE * AMOUNT					
อเนกรูป กับ DIF0.1	อเนกรูป กับ DIF0.2	อเนกรูป กับ DIF0.1 **สูงกว่า** อเนกรูป กับ DIF0.2	ไม่มีการเปรียบเทียบ	อเนกรูป กับ DIF0.1 **ต่ำกว่า** อเนกรูป กับ DIF0.2	ไม่มีการเปรียบเทียบ
	อเนกรูป กับ DIF0.4	อเนกรูป กับ DIF0.1 **ต่ำกว่า** อเนกรูป กับ DIF0.4	ไม่มีการเปรียบเทียบ	อเนกรูป กับ DIF0.1 **ต่ำกว่า** อเนกรูป กับ DIF0.4	ไม่มีการเปรียบเทียบ
	เอกรูป กับ DIF0.1	อเนกรูป กับ DIF0.1 **สูงกว่า** เอกรูป กับ DIF0.1	ไม่มีการเปรียบเทียบ	อเนกรูป กับ DIF0.1 **สูงกว่า** เอกรูป กับ DIF0.1	ไม่มีการเปรียบเทียบ
	เอกรูป กับ DIF0.2	อเนกรูป กับ DIF0.1 **สูงกว่า** เอกรูป กับ DIF0.2	ไม่มีการเปรียบเทียบ	อเนกรูป กับ DIF0.1 **สูงกว่า** เอกรูป กับ DIF0.2	ไม่มีการเปรียบเทียบ
	เอกรูป กับ DIF0.4	อเนกรูป กับ DIF0.1 **ต่ำกว่า** เอกรูป กับ DIF0.4	ไม่มีการเปรียบเทียบ	อเนกรูป กับ DIF0.1 **ต่ำกว่า** เอกรูป กับ DIF0.4	ไม่มีการเปรียบเทียบ

ตารางที่ 4.18 (ต่อ)

ปัจจัย		การวัดขนาดอิทธิพล 2 เกณฑ์			
		Zumbo and Thomas		Jodoin and Gierl	
		CI	TE	CI	TE
AMOUNT * %DIF					
DIF0.1 กับ 10%DIF	DIF0.2 กับ 10%DIF	DIF0.1 กับ 10%DIF **ต่ำกว่า** DIF0.2 กับ 10%DIF	ไม่มีการเปรียบเทียบ	DIF0.1 กับ 10%DIF **ต่ำกว่า** DIF0.2 กับ 10%DIF	ไม่มีการเปรียบเทียบ
	DIF0.4 กับ 10%DIF	DIF0.1 กับ 10%DIF **สูงกว่า** DIF0.4 กับ 10%DIF	ไม่มีการเปรียบเทียบ	DIF0.1 กับ 10%DIF **ต่ำกว่า** DIF0.4 กับ 10%DIF	ไม่มีการเปรียบเทียบ
	DIF0.1 กับ 20%DIF	DIF0.1 กับ 10%DIF **ต่ำกว่า** DIF0.1 กับ 20%DIF	ไม่มีการเปรียบเทียบ	DIF0.1 กับ 10%DIF **ต่ำกว่า** DIF0.1 กับ 20%DIF	ไม่มีการเปรียบเทียบ
	DIF0.2 กับ 20%DIF	DIF0.1 กับ 10%DIF **ต่ำกว่า** DIF0.2 กับ 20%DIF	ไม่มีการเปรียบเทียบ	DIF0.1 กับ 10%DIF **ต่ำกว่า** DIF0.2 กับ 20%DIF	ไม่มีการเปรียบเทียบ
	DIF0.4 กับ 20%DIF	DIF0.1 กับ 10%DIF **สูงกว่า** DIF0.4 กับ 20%DIF	ไม่มีการเปรียบเทียบ	DIF0.1 กับ 10%DIF **ต่ำกว่า** DIF0.4 กับ 20%DIF	ไม่มีการเปรียบเทียบ

จากตารางที่ 4.18 พบว่า ข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปมีอัตราความถูกต้องจากการวัดขนาดอิทธิพลทั้ง 2 เกณฑ์สูงกว่าแบบอเนกรูป แบบสอบที่มีจำนวนข้อสอบทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 20 มีอัตราความถูกต้องจากการวัดขนาดอิทธิพลทั้ง 2 เกณฑ์สูงกว่าในแบบสอบที่มีจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10 และเมื่อขนาดอิทธิพลของข้อสอบที่การทำหน้าที่ต่างกันเพิ่มขึ้น มีผลทำให้อัตราความถูกต้อง จากการวัด ขนาดอิทธิพล ทั้ง 2 เกณฑ์เพิ่มขึ้นภายใต้เกือบทุกเงื่อนไข

ตอนที่ 4 การเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในวิธีถดถอยโลจิสติ ก ระหว่าง ขนาดอิทธิพล ตาม เกณฑ์ Jodoin and Gierl และเกณฑ์ Zumbo and Thomas กรณีศึกษาข้อมูลเชิงประจักษ์

การนำเสนอผลการวิเคราะห์ข้อมูลเชิงประจักษ์ เริ่มต้นด้วยสถิติเชิงบรรยายของคะแนนจากข้อมูล “โครงการ สอบระดับชาติของสถาบันแห่งหนึ่ง ประจำปี พ.ศ.255 2 มีรูปแบบการตรวจให้คะแนนแบบทวิวิภาค ระดับชั้นประถมศึกษาปีที่ 6 จากโรงเรียนทุกสังกัด วิชาวิทยาศาสตร์ และวิชาคณิตศาสตร์ ” มีรายละเอียดดังต่อไปนี้

4.1 การเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ในแบบสอบวิชาคณิตศาสตร์

4.1.1 การวิเคราะห์ข้อมูลด้านสถิติเชิงบรรยายของคะแนนจากแบบสอบ

การวิเคราะห์คุณภาพเบื้องต้นและ ค่าความเที่ยง แบบความสอดคล้องภายในโดยสูตรสัมประสิทธิ์แอลฟาของครอนบาค (Alpha’s Cronbach) ในวิชาคณิตศาสตร์ จำนวน 40 ข้อ รายละเอียดดังตารางที่ 4.19

ตารางที่ 4.19 สถิติเชิงบรรยายของคะแนนจากแบบสอบรายวิชาคณิตศาสตร์

กลุ่มนักเรียน	N (ร้อยละ)	Max	Min	Range	$\overline{X}$	Median	Mode	SD
ในเขตอำเภอเมือง	68,640 (55.70)	35	0	35	1.580	1.00	0	2.778
นอกเขตอำเภอเมือง	54,527 (44.30)	33	0	33	0.890	0.00	0	1.555
รวม	123,167 (100.00)	35	0	35	1.30	1.00	0	2.386

จากตารางที่ 4.19 ผลการสอบของนักเรียนที่เข้าสอบทั้งประเทศจำนวน 123,167 คน จากแบบสอบ**วิชาคณิตศาสตร์** ภาพรวมข้ อสอบ 40 ข้อ คะแนนเต็ม 40 คะแนน ค่าคะแนนสูงสุด 35 คะแนน ค่าคะแนนต่ำสุด 0 คะแนน ค่าเฉลี่ยของคะแนนเท่ากับ 1.30 คะแนน ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 2.386 ผู้เข้าสอบตามสังกัด โรงเรียนที่ตั้งในเขตอำเภอเมือง จำนวน 68,640 คน คิดเป็นร้อยละ 55.70 ของผู้เข้าสอบทั่วประเทศ ค่าคะแนนสูงสุด 35 คะแนน ค่าคะแนนต่ำสุด 0 คะแนน ค่าเฉลี่ยของคะแนนเท่ากับ 1.580 คะแนน ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 2.778 ผู้เข้าสอบตามสังกัดโรงเรียนที่ตั้งนอกเขตอำเภอเมือง จำนวน 54,527 คน คิดเป็นร้อยละ 44.30 ของผู้เข้าสอบทั่วประเทศ ค่าคะแนนสูงสุด 33 คะแนน ค่าคะแนนต่ำสุด 0 คะแนน ค่าเฉลี่ยของคะแนนเท่ากับ 0.890 คะแนน ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.555

ค่าความเที่ยงแบบความสอดคล้องภายในของแบบสอบโดยสูตรสัมประสิทธิ์แอลฟา ของครอนบาค (Alpha's Cronbach) รายละเอียดดังตารางที่ 4.20

ตารางที่ 4.20 ค่าความเที่ยงแบบความสอดคล้องภายในของแบบสอบโดยสูตรสัมประสิทธิ์แอลฟา ของครอนบาค วิชาคณิตศาสตร์ จำแนกตามกลุ่มผู้สอบ

กลุ่มผู้สอบ	จำนวนผู้เข้าสอบ (ร้อยละ)	ค่าความเที่ยง
ในเขตอำเภอเมือง	68,640 (55.70)	.8213
นอกเขตอำเภอเมือง	54,527 (44.30)	.6293
รวม	123,167 (100.00)	.7887

หมายเหตุ: ***ในเขตอำเภอเมือง*** *(กลุ่มอ้างอิง : reference groups),* ***นอกเขตอำเภอเมือง*** *(กลุ่มเปรียบเทียบ : focal groups)*

จากตารางที่ 4.19 แบบสอบวิชาคณิตศาสตร์ จำนวน 40 ข้อ คะแนนเต็ม 40 คะแนน มีค่าความเที่ยงแบบความสอดคล้องภายใน 0.7887 วิเคราะห์คุณภาพรายข้อ ตามทฤษฎีทางการสอบแบบดั้งเดิม ประกอบด้วยค่าความยาก (p) และค่าอำนาจจำแนก (r) จากจำนวนผู้เข้าสอบ จำนวน 123,167 คน ใช้เทคนิคการแบ่งกลุ่มสูงกลุ่มต่ำ (27%) ปรากฏผลดังตารางที่ 4.21

ตารางที่ 4.21 ค่าความยาก (p) และอำนาจจำแนก (r) ของข้อสอบรายข้อ วิชาคณิตศาสตร์ (40 ข้อ)

ข้อสอบ	คุณภาพ		การแปลความหมาย	ข้อสอบ	คุณภาพ		การแปลความหมาย
	(p)	(r)	(p), (r)		(p)	(r)	(p), (r)
1	0.03	0.05	ยากมาก จำแนกไม่ดี	11	0.04	0.08	ยากมาก จำแนกไม่ดี
2	0.01	0.01	ยากมาก จำแนกไม่ดี	12	0.05	0.09	ยากมาก จำแนกไม่ดี
3	0.02	0.04	ยากมาก จำแนกไม่ดี	13	0.14	0.28	ยากมาก จำแนกไม่ดี
4	0.10	0.20	ยากมาก จำแนกไม่ดี	14	0.06	0.11	ยากมาก จำแนกไม่ดี
5	0.01	0.02	ยากมาก จำแนกไม่ดี	15	0.04	0.07	ยากมาก จำแนกไม่ดี
6	0.04	0.09	ยากมาก จำแนกไม่ดี	16	0.05	0.09	ยากมาก จำแนกไม่ดี
7	0.04	0.08	ยากมาก จำแนกไม่ดี	17	0.01	0.03	ยากมาก จำแนกไม่ดี
8	0.01	0.02	ยากมาก จำแนกไม่ดี	18	0.03	0.05	ยากมาก จำแนกไม่ดี
9	0.06	0.13	ยากมาก จำแนกไม่ดี	19	0.05	0.10	ยากมาก จำแนกไม่ดี
10	0.02	0.04	ยากมาก จำแนกไม่ดี	20	0.03	0.05	ยากมาก จำแนกไม่ดี

ตารางที่ 4.21 (ต่อ)

ข้อสอบ	คุณภาพ		การแปลความหมาย	ข้อสอบ	คุณภาพ		การแปลความหมาย
	(p)	(r)	(p), (r)		(p)	(r)	(p), (r)
21	0.07	0.14	ยากมาก จำแนกไม่ดี	31	0.06	0.12	ยากมาก จำแนกไม่ดี
22	0.03	0.06	ยากมาก จำแนกไม่ดี	32	0.01	0.01	ยากมาก จำแนกไม่ดี
23	0.02	0.04	ยากมาก จำแนกไม่ดี	33	0.07	0.14	ยากมาก จำแนกไม่ดี
24	0.19	0.37	ยากมาก จำแนกไม่ดี	34	0.07	0.15	ยากมาก จำแนกไม่ดี
25	0.12	0.25	ยากมาก จำแนกไม่ดี	35	0.01	0.01	ยากมาก จำแนกไม่ดี
26	0.04	0.07	ยากมาก จำแนกไม่ดี	36	0.08	0.15	ยากมาก จำแนกไม่ดี
27	0.07	0.13	ยากมาก จำแนกไม่ดี	37	0.01	0.03	ยากมาก จำแนกไม่ดี
28	0.00	0.01	ยากมาก จำแนกไม่ดี	38	0.01	0.02	ยากมาก จำแนกไม่ดี
29	0.05	0.10	ยากมาก จำแนกไม่ดี	39	0.00	0.00	ยากมาก จำแนกไม่ดี
30	0.04	0.08	ยากมาก จำแนกไม่ดี	40	0.01	0.01	ยากมาก จำแนกไม่ดี

จากตารางที่ 4.21 เมื่อวิเคราะห์ข้อสอบรายข้อตามทฤษฎีทางการสอบแบบดั้งเดิม ในแบบสอบวิชาคณิตศาสตร์ จำนวน 40 ข้อ ผู้เข้าสอบจำนวน 123,167 คน มีค่าความยาก (p) ระหว่าง 0.00 ถึง 0.19 และมีค่าอำนาจจำแนก (r) ระหว่าง 0.00 ถึง 0.37 เมื่อพิจารณาคุณภาพรายข้อตามเกณฑ์สำหรับคัดเลือกข้อสอบเข้าคลังข้อสอบหรือข้อที่มีคุณภาพที่ดีเพื่อนำไปใช้สอบนั้นพบว่าแบบสอบคณิตศาสตร์มีค่าความยากที่ยากมากและยังจำแนกได้ไม่ชัดเจนนัก

การวิเคราะห์คุณภาพรายข้อของแบบสอบ ตามทฤษฎีการตอบสนองข้อสอบ ใช้โปรแกรม MULTILOG (version 7.03) วิเคราะห์ค่าพารามิเตอร์อำนาจจำแนก (a) และค่าพารามิเตอร์ความยาก (b) ของข้อสอบรายข้อ วิชาคณิตศาสตร์ ปรากฏผลดังตารางที่ 4.22

ตารางที่ 4.22 ค่าพารามิเตอร์อำนาจจำแนก (a) และความยาก (b) ของข้อสอบรายข้อ วิชาคณิตศาสตร์

ข้อสอบ	(b)	(a)	ข้อสอบ	(b)	(a)
1	1.94	3.11	8	2.13	0.50
2	3.60	1.45	9	2.43	0.52
3	2.72	1.74	10	3.66	1.44
4	1.70	2.09	11	4.58	0.70
5	2.57	2.78	12	2.01	1.98
6	1.88	2.52	13	1.97	1.30
7	2.06	2.63	14	1.81	2.85

ตารางที่ 4.22 (ต่อ)

ข้อสอบ	(b)	(a)	ข้อสอบ	(b)	(a)
15	2.07	2.72	28	3.67	1.62
16	1.82	3.24	29	3.07	1.04
17	2.44	2.25	30	3.38	1.14
18	1.87	3.43	31	3.12	1.08
19	4.23	0.68	32	3.91	1.23
20	2.01	3.49	33	1.73	2.81
21	2.32	1.55	34	2.71	1.17
22	2.01	2.56	35	2.99	2.31
23	2.42	1.71	36	1.71	2.77
24	1.33	1.92	37	4.45	1.13
25	2.05	1.24	38	4.31	1.21
26	2.01	3.03	39	2.31	0.54
27	3.29	0.81	40	2.58	0.56
Mean (b) = 5.12 SD (b) = 0.87					
Mean (a) = 3.55 SD (a) = 0.90					

จากตารางที่ 4.21 เมื่อวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ พบว่า แบบสอบวิชาคณิตศาสตร์ จำนวน 40 ข้อ ผู้เข้าสอบจำนวน 123,167 คน มีพารามิเตอร์ความยาก (b) ระหว่าง 1.33 ถึง 4.58 มีพารามิเตอร์อำนาจจำแนก (a) ระหว่าง 0.50 ถึง 3.49

4.1.2 การตรวจสอบความเป็นเอกมิติของแบบสอบวิชาคณิตศาสตร์

การตรวจสอบความเป็นเอกมิติของแบบสอบก็เพื่อตรวจสอบข้อตกลงพื้นฐานที่สำคัญว่าแบบสอบนั้นวัดคุณลักษณะแฝง (latent trait) ที่ต้องการศึกษาเพียงคุณลักษณะเดียวตามทฤษฎีการตอบสนองข้อสอบ การละเลยต่อข้อตกลงเบื้องต้นข้อนี้อาจนำไปสู่การสรุปผลการศึกษาที่ผิดพลาดดังนั้นจึงตรวจสอบความเป็นเอกมิติของแบบสอบด้วยการวิเคราะห์องค์ประกอบ (Factor Analysis)

การตรวจสอบความเป็นเอกมิติสามารถดำเนินการโดยการวิเคราะห์องค์ประกอบ (Factor Analysis) ว่ามีกี่องค์ประกอบ เริ่มต้นจากการวิเคราะห์สหสัมพันธ์ระหว่างข้อสอบในวิชาคณิตศาสตร์และวิทยาศาสตร์ที่จะนำมาใช้ในการวิเคราะห์องค์ประกอบถือเป็นการวิเคราะห์หาความสัมพันธ์ระหว่างข้อสอบทั้งหมดโดยใช้สัมประสิทธิ์สหสัมพันธ์ของเพียร์สันแล้วพิจารณาร้อยละของความแปรปรวน

การวิเคราะห์สหสัมพันธ์ระหว่างข้อสอบในวิชาคณิตศาสตร์ที่จะนำมาใช้ในการวิเคราะห์องค์ประกอบถือเป็นการวิเคราะห์หาความสัมพันธ์ระหว่างข้อสอบทั้งหมดได้ค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปร 780 คู่ มีค่าสัมประสิทธิ์สหสัมพันธ์ที่แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และ .05 จำนวน 717 คู่ คิดเป็นร้อยละ 91.923 ของค่าสัมประสิทธิ์สหสัมพันธ์ทั้งหมด สัมประสิทธิ์สหสัมพันธ์ส่วนใหญ่มีทิศทางบวก ขนาดน้อยถึงปานกลาง มีค่าพิสัยตั้งแต่ -.002 ถึง .398

เมื่อพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างข้อสอบรายข้อ โดยพิจารณาความสัมพันธ์ของข้อสอบ 40 ข้อ (H_0: ข้อสอบทั้ง 40 ข้อ ไม่มีความสัมพันธ์กัน / H_1: ข้อสอบมีความสัมพันธ์กัน) ผลการทดสอบค่าสถิติ Bartlett's test of sphericity ซึ่งเป็นค่าสถิติทดสอบสมมติฐานว่าเมทริกซ์สหสัมพันธ์นั้นเป็นหรือไม่เป็นเมทริกซ์เอกลักษณ์ (identity metrix) พบว่ามีค่าเท่ากับ 480821.656 ($p<.01$) แสดงให้เห็นว่าเมทริกซ์สหสัมพันธ์นี้มีความแตกต่างจากเมทริกซ์เอกลักษณ์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 สอดคล้องกับค่าดัชนี Kaiser-Meyer-Olkin (KMO) ซึ่งมีค่าใกล้ 1 (.943) แสดงให้เห็นว่าข้อสอบต่างมีความสัมพันธ์กันมากมีความเหมาะสมที่จะนำไปใช้ในการวิเคราะห์องค์ประกอบ ปรากฏผลการวิเคราะห์ดังตารางที่ 4.23

ตารางที่ 4.23 ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation matrix), KMO, ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานของแบบสอบวิชาคณิตศาสตร์ (n=123,167 คน)

ข้อ	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}	X_{19}	X_{20}
X_1	1.000																			
X_2	0.076**	1.000																		
X_3	0.221**	0.033**	1.000																	
X_4	0.198**	0.073**	0.101**	1.000																
X_5	0.156**	0.096**	0.103**	0.186**	1.000															
X_6	0.304**	0.042**	0.162**	0.217**	0.128**	1.000														
X_7	0.300**	0.099**	0.132**	0.264**	0.288**	0.194**	1.000													
X_8	-0.014**	0.002	-0.011**	-0.008**	-0.002	0.000	-0.005	1.000												
X_9	-0.024**	0.044**	-0.017**	0.114**	0.111**	-0.017**	0.147**	0.003	1.000											
X_{10}	-0.014**	0.076**	-0.013**	0.171**	0.217**	-0.007*	0.252**	0.007*	0.212**	1.000										
X_{11}	0.055**	0.032**	0.040**	0.061**	0.080**	0.044**	0.078**	0.009**	0.034**	0.055**	1.000									
X_{12}	0.266**	0.043**	0.147**	0.139**	0.095**	0.227**	0.152**	-0.002	-0.021**	-0.009**	0.052**	1.000								
X_{13}	0.133**	0.042**	0.066**	0.166**	0.107**	0.129**	0.155**	-0.011**	0.082**	0.111**	0.043**	0.115**	1.000							
X_{14}	0.230**	0.080**	0.116**	0.251**	0.235**	0.205**	0.311**	-0.003	0.130**	0.206**	0.071**	0.184**	0.158**	1.000						
X_{15}	0.209**	0.098**	0.124**	0.225**	0.279**	0.165**	0.325**	0.001	0.126**	0.219**	0.087**	0.156**	0.130**	0.323**	1.000					
X_{16}	0.250**	0.089**	0.129**	0.269**	0.280**	0.223**	0.371**	-0.003	0.136**	0.229**	0.078**	0.170**	0.146**	0.366**	0.382**	1.000				
X_{17}	0.288**	0.058**	0.194**	0.151**	0.164**	0.264**	0.200**	-0.003	-0.012**	-0.005	0.055**	0.217**	0.076**	0.181**	0.204**	0.204**	1.000			
X_{18}	0.344**	0.062**	0.185**	0.197**	0.145**	0.339**	0.230**	-0.004	-0.016**	-0.002	0.055**	0.255**	0.118**	0.240**	0.224**	0.271**	0.343**	1.000		
X_{19}	0.100**	0.014**	0.067**	0.041**	0.051**	0.089**	0.053**	0.001	-0.011**	-0.019**	0.021**	0.085**	0.022**	0.059**	0.058**	0.073**	0.102**	0.110**	1.000	
X_{20}	0.245**	0.105**	0.142**	0.248**	0.296**	0.199**	0.351**	-0.007*	0.129**	0.225**	0.088**	0.176**	0.141**	0.323**	0.363**	0.381**	0.240**	0.273**	0.084**	1.000
X_{21}	0.112**	0.058**	0.075**	0.153**	0.150**	0.107**	0.172**	-0.001	0.081**	0.120**	0.060**	0.114**	0.108**	0.190**	0.173**	0.183**	0.114**	0.134**	0.041**	0.212**
X_{22}	0.242**	0.101**	0.135**	0.262**	0.298**	0.204**	0.385**	-0.003	0.144**	0.232	0.071**	0.166**	0.142**	0.330**	0.358**	0.406**	0.202**	0.287**	0.076**	0.398**
X_{23}	0.262**	0.048**	0.155**	0.120**	0.118**	0.211**	0.175**	0.004	-0.010**	-0.007**	0.043**	0.184**	0.070**	0.170**	0.183**	0.196**	0.245**	0.294**	0.107**	0.217**
X_{24}	0.175**	0.058**	0.095**	0.235**	0.152**	0.174**	0.236**	-0.014**	0.118**	0.163**	0.064**	0.145**	0.181**	0.248**	0.207**	0.245**	0.138**	0.190**	0.036**	0.234**
X_{25}	0.146**	0.054**	0.077**	0.151**	0.140**	0.127**	0.190**	-0.009**	0.077**	0.123**	0.054**	0.118**	0.118**	0.186**	0.179**	0.195**	0.116**	0.143**	0.048**	0.205**
X_{26}	0.247**	0.093**	0.126**	0.243**	0.276**	0.197**	0.356**	-0.009**	0.139**	0.234**	0.076**	0.174**	0.144**	0.319**	0.328**	0.357**	0.208**	0.239**	0.067**	0.380**
X_{27}	0.138**	0.017**	0.086**	0.062**	0.047**	0.121**	0.078**	-0.003	-0.026**	-0.025**	0.016**	0.110**	0.039**	0.076**	0.079**	0.093**	0.113**	0.154**	0.066**	0.089**

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

ตารางที่ 4.23 (ต่อ)

ข้อ	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}	X_{19}	X_{20}
X_{28}	0.082**	0.079**	0.062**	0.097**	0.185**	0.076**	0.157**	-0.005	0.057**	0.117**	0.043**	0.055**	0.055**	0.116**	0.135**	0.151**	0.109**	0.110**	0.021**	0.173**
X_{29}	0.139**	0.027**	0.099**	0.077**	0.058**	0.123**	0.089**	0.000	-0.020**	-0.018**	0.029**	0.115**	0.043**	0.079**	0.091**	0.095**	0.134**	0.153**	0.067**	0.103**
X_{30}	0.096**	0.047**	0.065**	0.098**	0.136**	0.083**	0.134**	-0.005	0.046**	0.086**	0.041**	0.072**	0.061**	0.117**	0.139**	0.135**	0.080**	0.089**	0.031**	0.160**
X_{31}	0.059**	0.023**	0.038**	0.094**	0.067**	0.063**	0.083**	0.001	0.043**	0.055**	0.025**	0.043**	0.075**	0.078**	0.072**	0.080**	0.082**	0.056**	0.008**	0.083**
X_{32}	0.105**	0.038**	0.080**	0.046**	0.079**	0.076**	0.065**	0.002	-0.001	-0.006*	0.019**	0.076**	0.027**	0.058**	0.071**	0.068**	0.108**	0.100**	0.050**	0.074**
X_{33}	0.233**	0.080**	0.118**	0.279**	0.237**	0.219**	0.351**	-0.003	0.140**	0.236**	0.065**	0.162**	0.174**	0.304**	0.296**	0.347**	0.229**	0.255**	0.047**	0.321**
X_{34}	0.086**	0.030**	0.051**	0.107**	0.086**	0.080**	0.122**	-0.010**	0.056**	0.077**	0.030**	0.065**	0.083**	0.106**	0.108**	0.116**	0.055**	0.080**	0.029**	0.110**
X_{35}	0.151**	0.132**	0.113**	0.149**	0.328**	0.097**	0.271**	-0.002	0.087**	0.192**	0.077**	0.094**	0.083**	0.198**	0.280**	0.262**	0.180**	0.144**	0.050**	0.316**
X_{36}	0.230**	0.082**	0.115**	0.261**	0.227**	0.203**	0.315**	-0.006**	0.138**	0.225**	0.073**	0.180**	0.177**	0.322**	0.296**	0.339**	0.172**	0.245**	0.051**	0.344**
X_{37}	0.078**	0.058**	0.058**	0.076**	0.148**	0.058**	0.109**	0.015**	0.039**	0.086**	0.036**	0.048**	0.040**	0.092**	0.127**	0.106**	0.070**	0.069**	0.023**	0.135**
X_{38}	0.110**	0.035**	0.077**	0.045**	0.072**	0.081**	0.067**	0.005	0.000	-0.008**	0.014**	0.081**	0.022**	0.063**	0.077**	0.069**	0.126**	0.117**	0.051**	0.085**
X_{39}	-0.003	0.001	0.001	0.003	0.004	-0.001	0.007*	0.002	0.001	0.007*	0.004	0.001	0.003	-0.001	0.006	0.005	-0.001	0.002	-0.005	0.004
X_{40}	0.023**	0.023**	0.013**	0.023**	0.036**	0.014**	0.035**	0.001	0.014**	0.030**	0.018**	0.016**	0.024**	0.025**	0.043**	0.036**	0.015**	0.021**	0.007**	0.036**
$\bar{X}$	0.017	0.004	0.014	0.073	0.008	0.029	0.025	0.012	0.050	0.013	0.037	0.032	0.111	0.037	0.024	0.030	0.008	0.017	0.046	0.017
S.D.	0.130	0.060	0.118	0.260	0.087	0.168	0.157	0.110	0.218	0.114	0.189	0.176	0.314	0.188	0.154	0.172	0.090	0.130	0.210	0.130

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

ตารางที่ 4.23 (ต่อ)

ข้อ	X_{21}	X_{22}	X_{23}	X_{24}	X_{25}	X_{26}	X_{27}	X_{28}	X_{29}	X_{30}	X_{31}	X_{32}	X_{33}	X_{34}	X_{35}	X_{36}	X_{37}	X_{38}	X_{39}	X_{40}
X_{21}	1.000																			
X_{22}	0.196**	1.000																		
X_{23}	0.104**	0.213**	1.000																	
X_{24}	0.181**	0.252**	0.125**	1.000																
X_{25}	0.121**	0.204**	0.114**	0.179**	1.000															
X_{26}	0.199**	0.381**	0.186**	0.241**	0.248**	1.000														
X_{27}	0.048**	0.095**	0.121**	0.066**	0.061**	0.089**	1.000													
X_{28}	0.085**	0.169**	0.075**	0.082**	0.074**	0.150**	0.044**	1.000												
X_{29}	0.061**	0.093**	0.136**	0.088**	0.063**	0.087**	0.106**	0.049**	1.000											
X_{30}	0.087**	0.150**	0.077**	0.116**	0.088**	0.137**	0.041**	0.085**	0.057**	1.000										
X_{31}	0.069**	0.083**	0.038**	0.118**	0.062**	0.068**	0.026**	0.049**	0.039**	0.052**	1.000									
X_{32}	0.047**	0.085**	0.103**	0.042**	0.041**	0.076**	0.044**	0.049**	0.055**	0.043**	0.030**	1.000								
X_{33}	0.168**	0.361**	0.161**	0.272**	0.191**	0.345**	0.079**	0.152**	0.086**	0.123**	0.107**	0.071**	1.000							
X_{34}	0.086**	0.119**	0.067**	0.137**	0.085**	0.113**	0.037**	0.050**	0.057**	0.074**	0.064**	0.037**	0.130**	1.000						
X_{35}	0.147**	0.318**	0.129**	0.121**	0.127**	0.259**	0.049**	0.235**	0.073**	0.141**	0.071**	0.092**	0.231**	0.078**	1.000					
X_{36}	0.198**	0.358**	0.176**	0.293**	0.206**	0.348**	0.081**	0.128**	0.096**	0.133**	0.105**	0.061**	0.364**	0.135**	0.219**	1.000				
X_{37}	0.078**	0.125**	0.053**	0.065**	0.063**	0.127**	0.027**	0.099**	0.035**	0.071**	0.043**	0.048**	0.099**	0.044**	0.177**	0.101**	1.000			
X_{38}	0.042**	0.088**	0.105**	0.037**	0.039**	0.080**	0.061**	0.044**	0.054**	0.045**	0.021**	0.077**	0.061**	0.038**	0.112**	0.063**	0.045**	1.000		
X_{39}	0.006*	0.007*	0.001	0.003	0.004	0.007*	0.004	-0.002	-0.005	-0.002	0.000	-0.003	0.005	-0.002	0.005	0.003	0.002	-0.003	1.000	
X_{40}	0.022**	0.051**	0.015**	0.023**	0.018**	0.035**	0.009**	0.036**	0.010**	0.017**	0.020**	0.005**	0.032**	0.018**	0.061**	0.032**	0.023**	0.013**	0.003	1.000
$\bar{X}$	0.051	0.018	0.016	0.143	0.104	0.024	0.060	0.003	0.039	0.032	0.046	0.005	0.046	0.058	0.004	0.053	0.011	0.007	0.001	0.006
S.D.	0.219	0.133	0.125	0.350	0.305	0.154	0.237	0.058	0.193	0.176	0.209	0.071	0.210	0.235	0.064	0.224	0.104	0.081	0.039	0.074

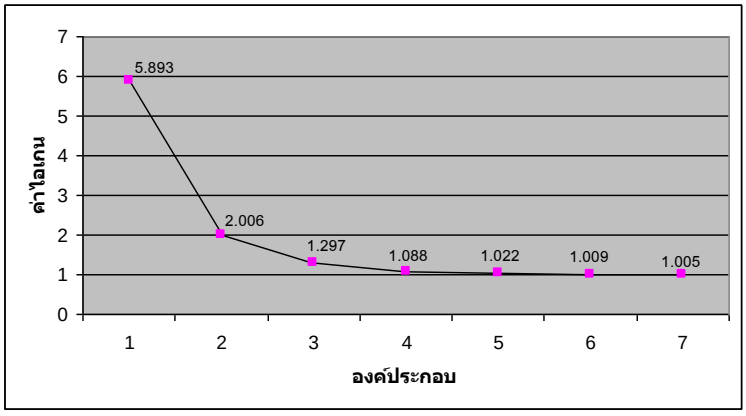
Bartlett's Test of Sphericity = 480821.656 df = 780 P = .000 Kaiser-Meyer-Olkin Measurement of Sampling Adequacy = .943

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

การวิเคราะห์องค์ประกอบหลัก (Principle Componant Analysis) หมุนแกนด้วยวิธีแวริแมกซ์ (Varimax) ค่าไอเกน (Eigen Value) และร้อยละของความแปรปรวน ดังตารางที่ 4.24

ตารางที่4.24 ค่าไอเกนและร้อยละของความแปรปรวนขององค์ประกอบของแบบสอบวิชาคณิตศาสตร์

องค์ประกอบ	ค่าไอเกน	ร้อยละของความแปรปรวน
1	5.893	14.732
2	2.006	5.015
3	1.297	3.242
4	1.088	2.719
5	1.022	2.555
6	1.009	2.523
7	1.005	2.513



ภาพที่ 4.6 ผลการตรวจสอบความเป็นเอกมิติของแบบสอบวิชาคณิตศาสตร์ จำนวน 40 ข้อ

ตารางที่ 4.24 เมื่อพิจารณาค่าไอเกนจากการวิเคราะห์ผลแบบสอบวิชาคณิตศาสตร์ จำนวน 40 ข้อ พบว่า ค่าไอเกนขององค์ประกอบที่ 1 มีค่าสูงสุด(5.893) มีค่าสูงกว่าค่าไอเกนขององค์ประกอบที่ 2 (2.006) ประมาณ 2.938 เท่า ส่วนค่าไอเกนขององค์ประกอบอื่นๆ ที่เหลือพบว่ามีค่าใกล้เคียงกัน เมื่อพิจารณาค่าร้อยละของความแปรปรวน พบว่า องค์ประกอบที่ 1 มีค่าร้อยละของความแปรปรวนค่อนข้างสูง (14.732) การพิจารณาค่าไอเกน (eigen value) ที่เสนอโดย Lord และ Novick (1968) ถ้าผลการวิเคราะห์พบว่ามีค่าไอเกนตัวเดียวหรืออาจหลายตัวก็ตามแต่ตัวแรกมีค่ามากกว่าตัวอื่นอย่างเห็นได้ชัด สามารถสรุปได้ว่าเครื่องมือชุดนั้นมีความเป็นเอกมิติซึ่ง ใกล้เคียงกับเกณฑ์ของ Reckase (อ้างถึงใน Raju, 1993; อุทัยวรรรณสายพัฒนะ, 2547) ที่เสนอว่าค่าร้อยละของความแปรปรวนควรมีค่าไม่น้อยกว่าร้อยละ 20 หรือค่าไอเกนขององค์ประกอบที่1 ต้องมีความแตกต่างจากค่าไอเกนขององค์ประกอบอื่นอย่างเด่นชัดจึงจะถือว่าแบบสอบมีความเป็นเอกมิติ เมื่อพิจารณาค่าไอเกนและค่าร้อยละของความแปรปรวนของ

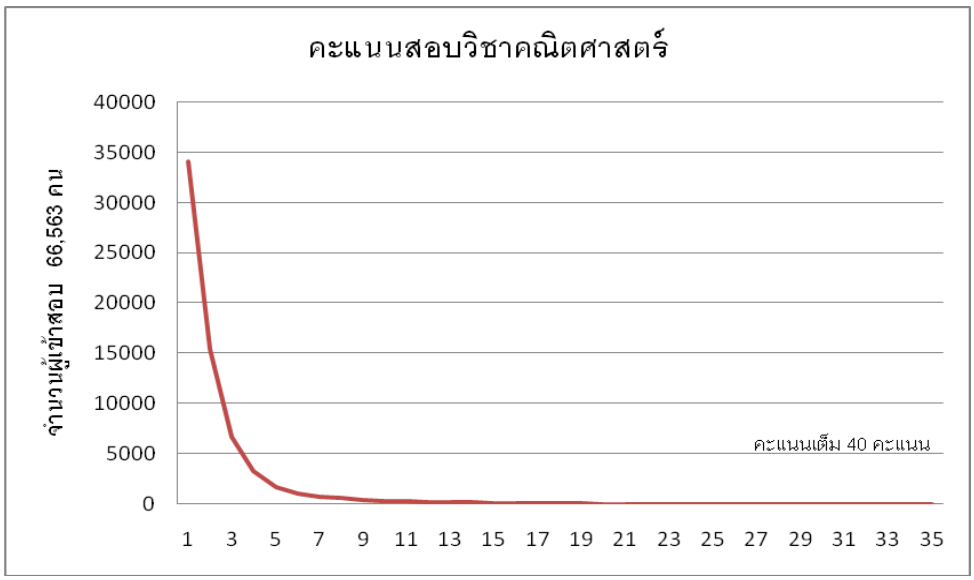
องค์ประกอบที่ 1 และภาพที่ 4.2 ประกอบกันแล้วในทางปฏิบัติถือได้ว่าแบบทดสอบฉบับนี้มีความเป็นเอกมิติสามารถนำข้อมูลไปวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ

4.1.3 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ เนื่องจากมีข้อสังเกตเกี่ยวกับผลการตอบข้อสอบของนักเรียนส่วนใหญ่ที่ไม่มีคะแนนจากการสอบ การนำผลการตอบข้อสอบลักษณะดังกล่าวไปคำนวณการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ อาจให้ผลลัพธ์ที่คลาดเคลื่อนไปจากความเป็นจริง ผู้วิจัยจึงตัดกรณีที่ผู้เข้าสอบได้คะแนนการสอบ 0 คะแนนออก แล้วนำผลการตอบที่เหลือมาคำนวณ ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์

เมื่อจำแนกคะแนนการสอบของนักเรียนทั้งหมดพบว่าส่วนใหญ่ได้คะแนนน้อย มีนักเรียนที่ได้คะแนน 0 คะแนน เป็นจำนวนมากถึง 56,604 คน คิดเป็นร้อยละ 45.96 จากจำนวนผู้เข้าสอบทั้งหมด

การแจกแจงความถี่ของคะแนนสอบ แสดงผลดังภาพที่ 4.6



ภาพที่ 4.6 กราฟแสดงคะแนนการสอบของผู้เข้าสอบวิชาคณิตศาสตร์

1) วิธีแมนเทล-แฮนส์เซล

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซลในแบบสอบวิชาคณิตศาสตร์ จำนวน 40 ข้อ ผู้เข้าสอบมีจำนวน 123,167 คน ตัดกรณีผู้เข้าสอบที่ได้ 0 คะแนน จำนวน 56,604 คน ออกจากการคำนวณ คิดเป็นร้อยละ 49.56 ของผู้เข้าสอบทั้งหมด มีข้อมูลที่นำมาใช้วิเคราะห์จริงจำนวน 66,563 คน ได้ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบดังตารางที่ 4.25

ตารางที่ 4.25 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซล วิชาคณิตศาสตร์

ข้อสอบ	ค่าสถิติ	p-value	ข้อสอบ	ค่าสถิติ	p-value
1	4.0729	0.0436 *	21	7.684	0.0056 **
2	2.3086	0.1287	22	0.3545	0.5516
3	12.4732	0.0004 ***	23	76.4393	0.0000 ***
4	51.8426	0.0000 ***	24	130.8514	0.0000 ***
5	14.763	0.0001 ***	25	0.1226	0.7262
6	30.7875	0.0000 ***	26	19.6687	0.0000 ***
7	102.6501	0.0000 ***	27	431.3287	0.0000 ***
8	293.2368	0.0000 ***	28	9.3312	0.0023 **
9	177.3265	0.0000 ***	29	227.4586	0.0000 ***
10	241.9711	0.0000 ***	30	0.3414	0.559
11	59.4471	0.0000 ***	31	4.241	0.0395 *
12	37.3454	0.0000 ***	32	26.6862	0.0000 ***
13	248.4409	0.0000 ***	33	47.1145	0.0000 ***
14	19.8277	0.0000 ***	34	0.9543	0.3286
15	14.714	0.0001 ***	35	1.5283	0.2164
16	1.2619	0.2613	36	116.1168	0.0000 ***
17	136.9272	0.0000 ***	37	7.0677	0.0078 **
18	41.7603	0.0000 ***	38	24.853	0.0000 ***
19	325.7855	0.0000 ***	39	0.0009	0.9762
20	18.4509	0.0000 ***	40	5.2701	0.0217 *

****P<.001*

จากตารางที่ 4.24 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซล ของแบบสอบที่มีรูปแบบการให้คะแนนแบบทวิวิภาควิชาคณิตศาสตร์ 40 ข้อ พิจารณาค่า p-value หากข้อสอบข้อใดมีนัยสำคัญ (Significance) หมายความว่าข้อสอบข้อนั้นทำหน้าที่ต่างกัน ซึ่งจากการตรวจสอบพบว่าข้อสอบ 32 ข้อทำหน้าที่ต่างกัน มีข้อที่ไม่ทำหน้าที่ต่างกัน 8 ข้อ คือ ข้อที่ 2, 16, 22, 25, 30, 34, 35 และข้อที่ 39

2) วิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ (significance test)

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้วิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ วิชาคณิตศาสตร์ 40 ข้อ ใช้คะแนนรวมทั้งฉบับ รายละเอียดดังตารางที่ 4.26

ตารางที่ 4.26 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้วิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ ในวิชาคณิตศาสตร์

ข้อ	กลุ่มผู้สอบ (g)		ปฏิสัมพันธ์ระหว่างกลุ่มกับความสามารถ (g by x)		ข้อ	กลุ่มผู้สอบ (g)		ปฏิสัมพันธ์ระหว่างกลุ่มกับความสามารถ (g by x)	
	B	sig	B	sig		B	sig	B	sig
1	-.122	.038*	-.164	.000*	21	-.170	.000*	-.021	.034*
2	-.328	.004*	-.003	.864	22	-.346	.000*	-.028	.072
3	.138	.013	-.089	.000*	23	.434	.000*	-.061	.000*
4	-.296	.000*	-.105	.000*	24	-.270	.000*	-.127	.000*
5	-.627	.000*	-.015	.331	25	-.028	.201	-.003	.753
6	.037	.373	-.148	.000*	26	-.423	.000*	.022	.107
7	-.790	.000*	-.047	.001*	27	.530	.000*	-.065	.000*
8	.910	.000*	.046	.087	28	-.624	.000*	.003	.853
9	-.507	.000*	.063	.000*	29	.455	.000*	-.075	.000*
10	-1.739	.000*	.036	.029	30	-.011	.770	.005	.625
11	.222	.000*	.007	.447	31	-.031	.305	-.065	.000*
12	.119	.002*	-.108	.000*	32	.475	.000*	-.002	.888
13	-.419	.000*	-.080	.000*	33	-.410	.000*	-.037	.007*
14	-.359	.000*	-.055	.000*	34	-.111	.000*	-.052	.000*
15	-.280	.000*	-.016	.230	35	-.293	.049*	.013	.502
16	-.067	.167	.004	.756	36	-.503	.000*	-.065	.000*
17	.714	.000*	-.097	.000*	37	.197	.001*	.05	.000*
18	.131	.025*	.166	.000*	38	.440	.000*	.003	.803
19	.535	.000*	-.035	.000*	39	-.007	.966	.016	.817
20	-.534	.000*	-.038	.013*	40	.169	.039*	.041	.077

**p< .05*

จากตารางที่ 4.25 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ภายใต้วิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ ของแบบสอบวิชาคณิตศาสตร์ เมื่อใช้คะแนนรวมทั้งฉบับเป็นเกณฑ์จับคู่ การตรวจสอบ DIF พิจารณาค่า p-value หากข้อสอบข้อใดมีนัยสำคัญ หมายความว่าข้อสอบข้อนั้นทำหน้าที่ต่างกัน มีข้อสอบที่ทำหน้าที่ต่างกันมีทั้งหมด 36 ข้อ

ค่านัยสำคัญของกลุ่มผู้เข้าสอบ (g) ถ้าข้อสอบข้อใดมีนัยสำคัญที่ระดับ .05 หมายความว่าข้อสอบข้อนั้นทำหน้าที่ต่างกันแบบเอกรูป ผลจากการตรวจสอบพบข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป จำนวน 13 ข้อ คือ ข้อ 2, 5, 8, 10, 11, 15, 22, 26, 28, 32, 35, 38 และข้อ 40

ข้อสอบที่ทำหน้าที่ไม่ต่างกันมี 4 ข้อ คือ ข้อ 16, 25, 30 และข้อ 39

ค่านัยสำคัญของปฏิสัมพันธ์ระหว่างกลุ่มผู้เข้าสอบกับความสามารถ (g by x) ถ้าข้อสอบข้อใดมีนัยสำคัญที่ระดับ .05 หมายความว่าข้อสอบข้อนั้นทำหน้าที่ต่างกันแบบอเนกรูป ผลจากการตรวจสอบ พบว่า มีข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป จำนวน 23 ข้อ คือ ข้อ 1, 3, 4, 6, 7, 9, 12, 13, 14, 17, 18, 19, 20, 21, 23, 24, 27, 29, 31, 33, 34, 36 และข้อ 37

3) การวัดขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas วิชาคณิตศาสตร์ 40 ข้อ ดังตารางที่ 4.27

ตารางที่ 4.27 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติกโดยการวัดขนาดอิทธิพลในวิชาคณิตศาสตร์ ตามเกณฑ์ Zumbo and Thomas

ข้อ	ขนาดอิทธิพล		ข้อ	ขนาดอิทธิพล		ข้อ	ขนาดอิทธิพล		ข้อ	ขนาดอิทธิพล	
	R^2	Amount of DIF		R^2	Amount of DIF		R^2	Amount of DIF		R^2	Amount of DIF
1	.0000	*	11	.0453	*	21	.0000	*	31	.0000	*
2	.0110	*	12	.0000	*	22	.0000	*	32	.0402	*
3	.0000	*	13	.0000	*	23	.0000	*	33	.0000	*
4	.0000	*	14	.0000	*	24	.0000	*	34	.0000	*
5	.0000	*	15	.0000	*	25	.0000	*	35	.0000	*
6	.0000	*	16	.0000	*	26	.0000	*	36	.0000	*
7	.0000	*	17	.0000	*	27	.0000	*	37	.0202	*
8	.6390	***	18	.0000	*	28	.0158	*	38	.0398	*
9	.0000	*	19	.2536	**	29	.0000	*	39	.0014	*
10	.0000	*	20	.0000	*	30	.0000	*	40	.0335	*

*Effect size (Nagelkerke's R^2): Zumbo and Thomas (ZT): * .00-.13 negligible effect, **.13- .26 moderate effect, ***<.26 large effect*

จากตารางที่ 4.27 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก โดยการวัดขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas วิชาคณิตศาสตร์ จำนวน 40 ข้อ พบว่า ข้อสอบทุกข้อทำหน้าที่ต่างกันโดยมีขนาดของการทำหน้าที่ต่างกัน 3 ขนาด คือ ขนาดเล็ก ขนาดปานกลาง และขนาดใหญ่ จำแนกได้ดังนี้ ข้อสอบที่ มีขนาดอิทธิพลของการ ทำหน้าที่ต่างกันขนาดเล็กน้อยจนแทบจะไม่มีเลย (*.00< R^2 <.13*) มีจำนวน 38 ข้อ คือ ข้อ 1, 2, 3, 4, 6, 7, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39

และข้อ 40 ส่วนข้อสอบที่ มีขนาดอิทธิพลของการ ทำหน้าที่ต่างกันขนาดปานกลาง ($.13< R^2 <.26$) คือ ข้อ 19 และข้อสอบที่มีขนาดอิทธิพลของการทำหน้าที่ต่างกันขนาดใหญ่ ($R^2<.26$) คือ ข้อ 8

4) การวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl ในแบบสอบวิชาคณิตศาสตร์ จำนวน 40 ข้อ ดังตารางที่ 4.28

ตารางที่ 4.28 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติกโดยการวัดขนาดอิทธิพลในวิชาคณิตศาสตร์ ตามเกณฑ์ Jodoin and Gierl

ข้อ	ขนาดอิทธิพล		ข้อ	ขนาดอิทธิพล		ข้อ	ขนาดอิทธิพล		ข้อ	ขนาดอิทธิพล	
	R^2	Amount of DIF		R^2	Amount of DIF		R^2	Amount of DIF		R^2	Amount of DIF
1	.0000	*	11	.0453	**	21	.0000	*	31	.0000	*
2	.0110	*	12	.0000	*	22	.0000	*	32	.0402	**
3	.0000	*	13	.0000	*	23	.0000	*	33	.0000	*
4	.0000	*	14	.0000	*	24	.0000	*	34	.0000	*
5	.0000	*	15	.0000	*	25	.0000	*	35	.0000	*
6	.0000	*	16	.0000	*	26	.0000	*	36	.0000	*
7	.0000	*	17	.0000	*	27	.0000	*	37	.0202	*
8	.6390	***	18	.0000	*	28	.0158	*	38	.0398	*
9	.0000	*	19	.2536	***	29	.0000	*	39	.0014	*
10	.0000	*	20	.0000	*	30	.0000	*	40	.0335	*

*Effect size (Nagelkerke's R^2): Jodoign and Gierl (JG): * .00-.035 negligible effect, **.0351- .07 moderate effect, ***<.071 large effect*

จากตารางที่ 4.27 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ภายใต้วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl พบว่า ข้อสอบทุกข้อทำหน้าที่ต่างกัน โดยมีขนาดของการทำหน้าที่ต่างกัน 3 ขนาด คือ ขนาดเล็กน้อย ขนาดปานกลาง และขนาดใหญ่ จำแนกได้ดังนี้ ข้อสอบที่ทำหน้าที่ต่างกันขนาดเล็กน้อยจนแทบจะไม่มีเลย ($.00< R^2 <.035$) มีจำนวน 36 ข้อ คือ ข้อ 1, 2, 3, 4, 5, 6, 7, 9, 10, 12, 13, 14, 15, 16, 17, 18, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 33, 34, 35, 36, 37, 38, 39 และข้อ 40 ข้อสอบที่มีขนาดอิทธิพลของการทำหน้าที่ต่างกันขนาดปานกลาง ($.351< R^2 <.07$) มี 2 ข้อ คือ ข้อ 11 และข้อ 32 ข้อสอบที่มีขนาดอิทธิพลของการทำหน้าที่ต่างกันขนาดใหญ่ ($.071 < R^2$) มีจำนวน 2 ข้อ คือ ข้อ 8 และข้อ 19

5) สรุปผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์

สรุปผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ (LR_S) กับการวัดขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas (LR_Z) และเกณฑ์ Jodoin and Gierl (LR_J) ดังตารางที่ 4.29

ตารางที่ 4.29 สรุปผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

แบบสอบ	ความยาว (ข้อ)	จำนวนข้อสอบ DIF (ข้อ)				จำนวนข้อสอบ NO DIF (ข้อ)			
		MH	LR_S	LR_J	LR_Z	MH	LR_S	LR_J	LR_Z
คณิตศาสตร์	40	32	36	4	2	8	4	36	38

จากตารางที่ 4.29 ผลการตรวจสอบเปรียบเทียบผลภายใต้วิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพล 2 เกณฑ์ ในแบบสอบวิชาคณิตศาสตร์ พบข้อสอบที่ทำหน้าที่ต่างกันในปริมาณใกล้เคียงกันกล่าวคือ การทดสอบระดับนัยสำคัญตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 36 ข้อ การวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl พบข้อสอบ DIF ที่มีขนาดปานกลางขึ้นไปจำนวน 4 ข้อและตรงกับผลการตรวจด้วยวิธีเกณฑ์ทุกข้อ การวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas พบข้อสอบที่ทำหน้าที่ต่างกันที่มีขนาดปานกลางขึ้นไปจำนวน 2 ข้อ ผลสรุปของจำนวนการตรวจพบการทำหน้าที่ต่างกันของข้อสอบมีรายละเอียดดังตารางที่ 4.30

ตารางที่ 4.30 จำนวนข้อของการเกิดและไม่เกิดการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์

วิธีแมนเทล-แฮนส์เซล		LR_S		LR_J		LR_Z	
		DIF	NO DIF	DIF	NO DIF	DIF	NO DIF
ข้อสอบทำหน้าที่ต่างกัน	(DIF)	32	0	4	28	2	30
ข้อสอบไม่ทำหน้าที่ต่างกัน	(NO DIF)	4	4	0	8	0	8

***หมายเหตุ** : ข้อสอบทั้งฉบับมีจำนวน 40 ข้อ*

จากตารางที่การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญกับวิธีแมนเทล-แฮนส์เซลซึ่งเป็นวิธีเกณฑ์ พบข้อร่วมของการทำหน้าที่ต่างกันระหว่างสองวิธีดังกล่าว จำนวน 32 ข้อ จากข้อสอบ 40 ข้อ คิดเป็นร้อยละ 80.00

ระหว่างวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตาม เกณฑ์ Jodoin and Gierl กับวิธีแมนเทล-แฮนส์เซลซึ่งเป็นวิธีเกณฑ์ พบข้อร่วมของการทำหน้าที่ต่างกันระหว่างสองวิธีดังกล่าว จำนวน 4 ข้อ จากข้อสอบ 40 ข้อ คิดเป็นร้อยละ 10.00

ระหว่างวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas กับวิธีแมนเทล-แฮนส์เซลซึ่งเป็นวิธีเกณฑ์ พบข้อร่วมของการทำหน้าที่ต่างกันระหว่างสองวิธีดังกล่าว จำนวน 2 ข้อ จากข้อสอบ 40 ข้อ คิดเป็นร้อยละ 5.00

4.1.4 ประสิทธิภาพด้านอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1

พิจารณาประสิทธิภาพของอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของแบบสอบวิชาคณิตศาสตร์ รายละเอียดดังตารางที่ 4.31

ตารางที่ 4.31 ร้อยละของอัตราความถูกต้องและร้อยละของอัตราความคลาดเคลื่อนประเภทที่ 1 จำแนกตามการตรวจสอบการทำหน้าที่ต่างกันของแบบสอบตามวิธีที่ศึกษา ในวิชาคณิตศาสตร์

ร้อยละของอัตราความถูกต้องของการวัดขนาดอิทธิพล		ร้อยละของอัตราความคลาดเคลื่อนประเภทที่ 1 ของการวัดขนาดอิทธิพล	
Jodoin and Gierl	Zumbo and Thomas	Jodoin and Gierl	Zumbo and Thomas
12.50	6.25	0.00	0.00

จากตารางที่ 4.31 ร้อยละของอัตราความถูกต้อง วิชาคณิตศาสตร์ โดยการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl คิดเป็นร้อยละ 12.50 โดยการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas คือ คิดเป็นร้อยละ 6.25

ร้อยละของอัตราความคลาดเคลื่อนประเภทที่ 1 โดยการวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl คิดเป็นร้อยละ 0.00 โดยการวัดขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas คิดเป็นร้อยละ 0.00

เกณฑ์การพิจารณาประสิทธิภาพ ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีอัตราความถูกต้องในการตรวจสอบของข้อสอบสูง และมีอัตราความ คลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบของข้อสอบต่ำ แสดงถึงประสิทธิภาพในการตรวจสอบสูงสุด ถือเป็นเงื่อนไขที่ต้องการ (รายละเอียดตอนที่ 1 ภาพประกอบ 4.1) นั่นคือ ผลจากการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบวิชาคณิตศาสตร์ของวิธีการตรวจสอบทุกวิธีที่ศึกษาเป็นไปตามเกณฑ์เงื่อนไขที่ต้องการ ผลประสิทธิภาพ พบว่า เกณฑ์ Jodoin and Gierl ให้ค่าร้อยละของอัตราความถูกต้อง ในการตรวจสอบข้อสอบสูงกว่า ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบข้อสอบคิดเป็นร้อยละที่เท่ากัน จึงตัดสินผลของการเปรียบเทียบประสิทธิภาพระหว่าง 2 เกณฑ์ดังกล่าวได้ว่าภายใต้วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl มีประสิทธิภาพดีกว่าการวัดขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas

4.2 การเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ในแบบสอบวิชาวิทยาศาสตร์

4.2.1 การวิเคราะห์ข้อมูลด้านสถิติเชิงบรรยายของคะแนนจากแบบสอบ

ผลการวิเคราะห์คุณภาพเบื้องต้นจากข้อมูลของ “โครงการ สอบระดับชาติของสถาบันแห่งหนึ่ง ประจำปี พ.ศ.255 2 มีรูปแบบการตรวจให้คะแนนแบบทวิวิภาค ระดับชั้นประถมศึกษาปีที่ 6 จากโรงเรียนทุกสังกัด ” ในวิชาวิทยาศาสตร์ จำนวน 50 ข้อ ดังตารางที่ 4.32

ตารางที่ 4.32 สถิติเชิงบรรยายของคะแนนจากแบบสอบรายวิชาวิทยาศาสตร์

	N (ร้อยละ)	Max	Min	Range	$\overline{X}$	Median	Mode	SD
ในเขตอำเภอเมือง	62,533 (56.50)	48	0	48	19.65	18.00	15	7.592
นอกเขตอำเภอเมือง	48,076 (43.50)	48	0	48	16.82	16.00	14	6.045
รวม	110,609 (100.00)	48	0	48	18.42	17.00	14	7.097

แบบสอบ**วิชาวิทยาศาสตร์** ภาพรวมข้อสอบ 50 ข้อ คะแนนเต็ม 50 คะแนน ผู้เข้าสอบจำนวน 110,609 คน ค่าคะแนนสูงสุด 48 คะแนน ค่าคะแนนต่ำสุด 0 คะแนน ค่าเฉลี่ยของคะแนนเท่ากับ 18.42 คะแนน ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 7.097 ผู้เข้าสอบตามสังกัด โรงเรียนที่ตั้งในเขตอำเภอเมือง จำนวน 62,533 คน คิดเป็นร้อยละ 56.50 ของผู้เข้าสอบทั่วประเทศ ค่าคะแนนสูงสุด 48 คะแนน ค่าคะแนนต่ำสุด 0 คะแนน ค่าเฉลี่ยของคะแนนเท่ากับ 19.65 คะแนน ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 7.592 ผู้เข้าสอบตามสังกัดโรงเรียนที่ตั้งนอกเขตอำเภอเมือง จำนวน 48,076 คน คิดเป็นร้อยละ 43.50 ของผู้เข้าสอบทั่วประเทศค่าคะแนนสูงสุด 48 คะแนน ค่าคะแนนต่ำสุด 0 คะแนน ค่าเฉลี่ยของคะแนนเท่ากับ 16.82 คะแนน ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 6.045

พิจารณา ค่าความเที่ยงแบบความสอดคล้องภายในของแบบสอบโดยสูตรสัมประสิทธิ์แอลฟาของครอนบาค (Alpha’s Cronbach) แบบสอบ วิชาวิทยาศาสตร์ รายละเอียดดังตารางที่ 4.33

ตารางที่ 4.33 ค่าความเที่ยงแบบความสอดคล้องภายในของแบบสอบโดยสูตรสัมประสิทธิ์แอลฟาของครอนบาค จำแนกตามวิชาวิทยาศาสตร์และกลุ่มผู้สอบ

กลุ่มผู้สอบ	จำนวนผู้เข้าสอบ (ร้อยละ)	ค่าความเที่ยง
ในเขตอำเภอเมือง	62,533 (56.50)	.8276
นอกเขตอำเภอเมือง	48,076 (43.50)	.7293
รวม	110,609 (100.00)	.8037

หมายเหตุ: ***ในเขตอำเภอเมือง*** *(กลุ่มอ้างอิง : reference groups),* ***นอกเขตอำเภอเมือง*** *(กลุ่มเปรียบเทียบ : focal groups)*

จากตารางที่ 4.33 แบบสอบวิชาวิทยาศาสตร์ จำนวน 50 ข้อ คะแนนเต็ม 50 คะแนน มีค่าความเที่ยงแบบความสอดคล้องภายใน 0.8037 วิเคราะห์คุณภาพรายข้อ ตามทฤษฎีทางการสอบแบบดั้งเดิม ประกอบด้วยค่าความยาก (p) และค่าอำนาจจำแนก (r) จากจำนวนผู้เข้าสอบ จำนวน 110,609 คน ใช้เทคนิคการแบ่งกลุ่มสูงกลุ่มต่ำ (27%) ปรากฏผลดังตารางที่ 4.34

ตารางที่ 4.34 ค่าความยาก (p) และอำนาจจำแนก (r) ของข้อสอบรายข้อ วิชาวิทยาศาสตร์ (50 ข้อ)

ข้อสอบ	คุณภาพ		การแปลความหมาย	ข้อสอบ	คุณภาพ		การแปลความหมาย
	(p)	(r)	(p), (r)		(p)	(r)	(p), (r)
1	0.42	0.32	ยากปานกลาง, จำแนกดี	26	0.25	0.21	ยาก, จำแนกดี
2	0.18	-0.01	ยากมาก, จำแนกไม่ดี	27	0.38	0.28	ยากปานกลาง, จำแนกดี
3	0.61	0.17	ค่อนข้างง่าย,จำแนกไม่ดี	28	0.21	0.11	ยาก, จำแนกไม่ดี
4	0.57	0.58	ยากปานกลาง, จำแนกดีมาก	29	0.49	0.40	ยากปานกลาง, จำแนกดีมาก
5	0.46	0.49	ยากปานกลาง, จำแนกดีมาก	30	0.49	0.38	ยากปานกลาง, จำแนกดี
6	0.39	0.26	ยากปานกลาง, จำแนกดี	31	0.25	0.23	ยาก, จำแนกดี
7	0.24	0.12	ยาก, จำแนกไม่ดี	32	0.37	0.39	ยากปานกลาง, จำแนกดี
8	0.39	0.22	ยากปานกลาง, จำแนกดี	33	0.27	0.19	ยาก, จำแนกไม่ดี
9	0.32	0.25	ยากปานกลาง, จำแนกดี	34	0.28	0.21	ยาก, จำแนกดี
10	0.28	0.15	ยาก, จำแนกไม่ดี	35	0.45	0.49	ยากปานกลาง, จำแนกดีมาก
11	0.48	0.42	ยากปานกลาง,จำแนกดีมาก	36	0.49	0.50	ยากปานกลาง, จำแนกดีมาก
12	0.29	0.22	ยาก, จำแนกดี	37	0.55	0.56	ค่อนข้างง่าย, จำแนกดีมาก
13	0.26	0.22	ยาก, จำแนกดี	38	0.22	0.02	ยาก, จำแนกไม่ดี
14	0.21	0.01	ยาก, จำแนกไม่ดี	39	0.53	0.62	ค่อนข้างง่าย, จำแนกดีมาก
15	0.54	0.66	ค่อนข้างง่าย, จำแนกดีมาก	40	0.53	0.56	ค่อนข้างง่าย, จำแนกดีมาก
16	0.36	0.41	ยากปานกลาง, จำแนกดีมาก	41	0.41	0.45	ยากปานกลาง, จำแนกดีมาก
17	0.29	0.18	ยาก, จำแนกไม่ดี	42	0.66	0.50	ค่อนข้างง่าย, จำแนกดีมาก
18	0.21	0.11	ยาก, จำแนกไม่ดี	43	0.47	0.41	ยากปานกลาง, จำแนกดีมาก
19	0.48	0.47	ยากปานกลาง, จำแนกดีมาก	44	0.41	0.45	ยากปานกลาง, จำแนกดีมาก
20	0.20	0.08	ยาก, จำแนกไม่ดี	45	0.25	0.03	ยาก, จำแนกไม่ดี
21	0.48	0.42	ยากปานกลาง, จำแนกดีมาก	46	0.57	0.53	ค่อนข้างง่าย, จำแนกดีมาก
22	0.63	0.47	ค่อนข้างง่าย, จำแนกดีมาก	47	0.24	0.06	ยาก, จำแนกไม่ดี
23	0.47	0.42	ยากปานกลาง, จำแนกดีมาก	48	0.48	0.40	ยากปานกลาง, จำแนกดีมาก
24	0.32	0.34	ยาก, จำแนกดี	49	0.30	0.09	ยาก, จำแนกไม่ดี
25	0.27	0.29	ยาก, จำแนกดี	50	0.45	0.32	ยากปานกลาง, จำแนกดี

จากตารางที่ 4.34 ผลการ วิเคราะห์ข้อสอบรายข้อ มีค่าความยาก (p) ระหว่าง 0.18 ถึง 0.66 มีข้อสอบที่มีคุณภาพเหมาะสม จำนวน 36 ข้อ วิเคราะห์คุณภาพรายข้อของแบบสอบ ตามทฤษฎีการตอบสนองข้อสอบใช้โปรแกรมMULTILOG (version 7.03) ดังตารางที่ 4.35

ตารางที่ 4.35 ค่าพารามิเตอร์อำนาจจำแนก (a) และความยาก (b) ของข้อสอบรายข้อวิชาวิทยาศาสตร์

ข้อสอบ	(b)	(a)	ข้อสอบ	(b)	(a)
1	0.84	0.51	26	2.71	0.51
2	4.33	0.31	27	1.33	0.46
3	-3.35	0.17	28	2.72	0.13
4	-0.23	1.33	29	0.15	0.71
5	0.37	1.02	30	0.12	0.62
6	1.32	0.43	31	2.43	0.57
7	4.66	0.28	32	0.97	0.84
8	1.46	0.32	33	2.90	0.35
9	1.91	0.46	34	2.87	0.35
10	4.12	0.24	35	0.38	1.00
11	0.25	0.76	36	0.25	1.05
12	2.49	0.42	37	-0.05	1.29
13	2.79	0.44	38	4.11	0.30
14	4.26	0.25	39	-0.03	1.53
15	-0.06	1.73	40	0.00	1.22
16	1.03	0.91	41	0.66	0.88
17	3.40	0.28	42	-0.67	1.31
18	2.08	0.10	43	0.28	0.74
19	0.21	0.89	44	0.63	0.92
20	2.92	0.21	45	4.05	0.24
21	0.29	0.76	46	-0.16	1.27
22	-0.61	1.04	47	2.91	0.10
23	0.32	0.74	48	0.31	0.79
24	1.40	0.77	49	2.91	0.11
25	1.96	0.69	50	0.41	0.49
Mean (b) = 1.41 SD (b) = 1.66					
Mean (a) = 0.66 SD (a) = 0.41					

จากตารางที่ 4.35 เมื่อวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ แบบสอบ วิชาวิทยาศาสตร์ จำนวน 50 ข้อ ผู้เข้าสอบจำนวน 110,609 คน มีค่าพารามิเตอร์ความยาก (b) ระหว่าง -3.35 ถึง 4.66 และมีค่าพารามิเตอร์อำนาจจำแนกระหว่าง 0.10 ถึง 1.73

4.2.2 การตรวจสอบความเป็นเอกมิติของแบบสอบวิชาวิทยาศาสตร์

การตรวจสอบความเป็นเอกมิติของแบบสอบเพื่อตรวจสอบข้อตกลงพื้นฐานที่สำคัญว่าแบบสอบนั้นวัดคุณลักษณะแฝง (latent trait) ที่ต้องการศึกษาเพียงคุณลักษณะเดียวตามทฤษฎีการตอบสนองข้อสอบ การละเลยต่อข้อตกลงเบื้องต้นข้อนี้อาจจะนำไปสู่การสรุปผลการศึกษาที่ผิดพลาด ดังนั้นจึงทำการตรวจสอบความเป็นเอกมิติของแบบสอบทั้งสองวิชาดังกล่าวด้วยการวิเคราะห์องค์ประกอบ (Factor Analysis) ของแบบสอบ การตรวจสอบความเป็นเอกมิติสามารถดำเนินการโดยการวิเคราะห์องค์ประกอบว่ามีกี่องค์ประกอบ เริ่มต้นจากการวิเคราะห์สหสัมพันธ์ระหว่างข้อสอบในวิชาคณิตศาสตร์และวิทยาศาสตร์ที่จะนำมาใช้ในการวิเคราะห์องค์ประกอบถือเป็นการวิเคราะห์หาความสัมพันธ์ระหว่างข้อสอบทั้งหมดโดยใช้สัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน แล้วพิจารณาร้อยละของความแปรปรวน

การวิเคราะห์สหสัมพันธ์ระหว่างข้อสอบในวิชาวิทยาศาสตร์ที่จะนำมาใช้ในการวิเคราะห์องค์ประกอบ ถือเป็นการวิเคราะห์หาความสัมพันธ์ระหว่างข้อสอบทั้งหมดโดยใช้สัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน ได้ค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปร 1,219 คู่ มีค่าสัมประสิทธิ์สหสัมพันธ์ที่แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และ .05 จำนวน 1,151 คู่ คิดเป็นร้อยละ 94.421 ของค่าสัมประสิทธิ์สหสัมพันธ์ทั้งหมด สัมประสิทธิ์สหสัมพันธ์ส่วนใหญ่มีทิศทางบวก ขนาดต่ำถึงปานกลาง มีค่าพิสัยตั้งแต่ -.001 ถึง .290 รายละเอียดดังตารางที่ 4.50

เมื่อพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างข้อสอบรายข้อจากตารางที่ 6 โดยพิจารณาความสัมพันธ์ของข้อสอบทั้ง 50 ข้อ (H_0 : ข้อสอบทั้ง 50 ข้อ ไม่มีความสัมพันธ์กัน / H_1 : ข้อสอบมีความสัมพันธ์กัน) ผลการทดสอบค่าสถิติ Bartlett's test of sphericity ซึ่งเป็นค่าสถิติทดสอบสมมติฐานว่าเมทริกซ์สหสัมพันธ์นั้นเป็นหรือไม่เป็นเมทริกซ์เอกลักษณ์ (identity metrix) พบว่ามีค่าเท่ากับ 387449.118 ($p<.01$) แสดงให้เห็นว่าเมทริกซ์สหสัมพันธ์นี้มีความแตกต่างจากเมทริกซ์เอกลักษณ์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 สอดคล้องกับค่าดัชนี Kaiser-Meyer-Olkin (KMO) ซึ่งมีค่าใกล้ 1 (.934) แสดงให้เห็นว่าข้อสอบต่างมีความสัมพันธ์กันมากมีความเหมาะสมที่จะนำไปใช้ในการวิเคราะห์องค์ประกอบ ปรากฏผลการวิเคราะห์ดังตารางที่ 4.36

ตารางที่ 4.36 ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation matrix), KMO, ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานของข้อสอบวิชาวิทยาศาสตร์ (n=110,609 คน)

ข้อ	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}	X_{19}	X_{20}
X_1	1.000																			
X_2	-0.083**	1.000																		
X_3	0.018**	-0.016**	1.000																	
X_4	0.129**	-0.032**	0.067**	1.000																
X_5	0.086**	-0.073**	0.043**	0.218**	1.000															
X_6	0.051**	0.001	0.004	0.099	0.120**	1.000														
X_7	0.006*	0.011**	-0.033**	0.029**	0.034**	0.119**	1.000													
X_8	0.033**	0.006*	0.004	0.073**	0.067**	0.065**	0.085**	1.000												
X_9	0.057**	-0.005	0.009**	0.095**	0.101**	0.040**	0.064**	0.048**	1.000											
X_{10}	0.026**	-0.001	-0.004	0.042**	0.046**	0.048**	0.051**	0.048**	0.043**	1.000										
X_{11}	0.076**	-0.023**	0.024**	0.179**	0.151**	0.104**	0.080**	0.060**	0.086**	0.063**	1.000									
X_{12}	0.046**	-0.002	-0.003	0.090**	0.099**	0.066**	0.063**	0.048**	0.057**	0.042**	0.110**	1.000								
X_{13}	0.046**	-0.006	-0.006	0.091**	0.093**	0.058**	0.048**	0.043**	0.049**	0.033**	0.075**	0.053**	1.000							
X_{14}	-0.022**	0.010**	-0.007*	-0.053**	-0.046**	-0.014**	-0.009**	-0.007*	-0.017**	-0.009**	-0.026**	-0.020**	0.164**	1.000						
X_{15}	0.152**	-0.051**	0.045**	0.304**	0.260**	0.109**	0.048**	0.070**	0.117**	0.054**	0.211**	0.107**	0.110**	-0.048**	1.000					
X_{16}	0.097**	-0.018**	0.016**	0.186**	0.184**	0.110**	0.089**	0.076**	0.104**	0.074**	0.140**	0.117**	0.125**	0.008**	0.185**	1.000				
X_{17}	0.026**	-0.001	0.003	0.048**	0.054**	0.037**	0.033**	0.034**	0.033**	0.031**	0.052**	0.030**	0.034**	0.011**	0.041**	0.027**	1.000			
X_{18}	0.014**	-0.009**	0.006	0.035**	0.017**	-0.003	-0.007*	0.007*	0.009	0.000*	0.007*	0.007**	0.013	-0.003**	0.046**	0.014	-0.006**	1.000		
X_{19}	0.101**	-0.029**	0.030**	0.199**	0.177**	0.067**	0.035**	0.052**	0.081**	0.036**	0.132**	0.078**	0.076**	-0.040**	0.238**	0.153**	0.060**	0.041**	1.000	
X_{20}	0.028**	0.016**	-0.024**	0.025**	0.046**	0.055**	0.078**	0.042**	0.047**	0.045**	0.036**	0.058**	0.054**	0.005	0.026**	0.077**	0.027**	-0.007*	-0.038**	1.000
X_{21}	0.093**	-0.020**	0.019**	0.149**	0.159**	0.082**	0.054**	0.055**	0.076**	0.048**	0.116**	0.067**	0.067**	-0.022**	0.195**	0.154**	0.046**	0.012**	0.137**	0.049**
X_{22}	0.098**	-0.046**	0.041**	0.265**	0.170**	0.074**	0.015**	0.053**	0.075**	0.043**	0.143**	0.068**	0.074**	-0.045**	0.243**	0.153**	0.042**	0.021**	0.163**	0.011**
X_{23}	0.091**	-0.025**	0.023**	0.150**	0.160**	0.071**	0.034**	0.050**	0.072**	0.038**	0.107**	0.069**	0.063**	-0.034**	0.200**	0.148**	0.043**	0.012**	0.143**	0.041**
X_{24}	0.090**	-0.013**	0.007*	0.160**	0.174**	0.092**	0.080**	0.063**	0.082**	0.063**	0.116**	0.116**	0.102**	-0.029**	0.188**	0.174**	0.041**	0.006	0.141**	0.097**
X_{25}	0.075**	-0.007*	0.000	0.151**	0.141**	0.086**	0.076**	0.051**	0.081**	0.067**	0.107**	0.102**	0.090**	-0.025**	0.165**	0.153**	0.043**	0.007*	0.123**	0.081**
X_{26}	0.050**	0.008*	-0.018**	0.117**	0.110**	0.082**	0.097**	0.053**	0.068**	0.073**	0.098**	0.095**	0.085**	-0.021**	0.119**	0.130**	0.030**	-0.007*	0.093**	0.088**

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

ตารางที่ 4.36 (ต่อ)

ข้อ	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}	X_{19}	X_{20}
X_{27}	0.054**	-0.010**	0.010**	0.115**	0.104**	0.062**	0.059**	0.040**	0.055**	0.028**	0.074**	0.063**	0.056**	-0.027**	0.131**	0.109**	0.030**	0.010**	0.088**	0.044**
X_{28}	0.005	0.001	0.002	0.018**	0.018**	0.030**	0.019**	0.010**	0.011**	-0.007*	0.031**	0.010**	0.015**	-0.005**	0.028**	0.028**	0.013**	-0.002	0.018**	0.014**
X_{29}	0.081**	-0.023**	0.030**	0.138**	0.126**	0.059**	0.019**	0.046**	0.063**	0.033**	0.096**	0.048**	0.055**	-0.024**	0.174**	0.127**	0.053**	0.014**	0.118**	0.008**
X_{30}	0.067**	-0.029**	0.038**	0.129**	0.111**	0.046**	0.015**	0.038**	0.053**	0.016**	0.103**	0.033**	0.035**	-0.025**	0.168**	0.099**	0.053**	0.005	0.110**	0.004
X_{31}	0.067**	-0.003	0.001	0.107**	0.111**	0.074**	0.067**	0.054**	0.074**	0.055**	0.086**	0.079**	0.069**	-0.016**	0.133**	0.143**	0.041**	-0.001	0.092**	0.080**
X_{32}	0.091**	-0.018**	0.026**	0.167**	0.168**	0.087**	0.074**	0.067**	0.087**	0.054**	0.126**	0.083**	0.081**	-0.028**	0.213**	0.175**	0.062**	0.006*	0.138**	0.058**
X_{33}	0.038**	-0.002	0.010**	0.075**	0.082**	0.050**	0.037**	0.035**	0.034**	0.031**	0.051**	0.046**	0.046**	-0.018**	0.078**	0.073**	0.023**	0.004	0.057**	0.035**
X_{34}	0.035**	-0.013**	0.005	0.077**	0.067**	0.045**	0.025**	0.032**	0.039**	0.023**	0.055**	0.037**	0.034**	-0.022**	0.087**	0.065**	0.027**	0.006*	0.065**	0.027**
X_{35}	0.098**	-0.025**	0.018**	0.206**	0.172**	0.082**	0.056**	0.065**	0.087**	0.054**	0.122**	0.068**	0.078**	-0.030**	0.226**	0.162**	0.050**	0.030**	0.153**	0.039**
X_{36}	0.108**	-0.027**	0.027**	0.205**	0.206**	0.089**	0.061**	0.073**	0.094**	0.047**	0.157**	0.094**	0.089**	-0.040**	0.245**	0.186**	0.058**	0.013**	0.174**	0.057**
X_{37}	0.120**	-0.037**	0.031**	0.223**	0.205**	0.087**	0.047**	0.069**	0.093**	0.045**	0.156**	0.077**	0.090**	-0.042**	0.268**	0.184**	0.066**	0.019**	0.182**	0.034**
X_{38}	0.007*	0.017**	-0.016**	-0.002	0.007*	0.025**	0.056**	0.027**	0.024**	0.020**	0.002	0.033**	0.019**	0.007*	-0.015**	0.026**	0.016**	-0.006*	0.001	0.051**
X_{39}	0.124**	-0.040**	0.044**	0.240**	0.220**	0.091**	0.042**	0.065**	0.101**	0.041**	0.164**	0.080**	0.083**	-0.045**	0.306**	0.196**	0.070**	0.023**	0.192**	0.022**
X_{40}	0.113**	-0.029**	0.033**	0.198**	0.190**	0.081**	0.042**	0.068**	0.092**	0.049**	0.146**	0.078**	0.076**	-0.041**	0.257**	0.178**	0.069**	0.007*	0.171**	0.033**
X_{41}	0.103**	-0.028**	0.022**	0.166**	0.168**	0.074**	0.041**	0.054**	0.080**	0.033**	0.123**	0.069**	0.067**	-0.028**	0.217**	0.151**	0.047**	0.011**	0.147**	0.038**
X_{42}	0.098**	-0.034**	0.050**	0.197**	0.152**	0.055**	0.015**	0.060**	0.076**	0.028**	0.142**	0.051**	0.050**	-0.036**	0.238**	0.139**	0.062**	0.015**	0.154**	-0.004
X_{43}	0.089**	-0.021**	0.026**	0.159**	0.138**	0.062**	0.037**	0.045**	0.073**	0.020**	0.110**	0.064**	0.069**	-0.026**	0.191**	0.132**	0.041**	0.008**	0.128**	0.031**
X_{44}	0.092**	-0.021**	0.023**	0.162**	0.164**	0.084**	0.057**	0.071**	0.086**	0.043**	0.132**	0.072**	0.074**	-0.029**	0.211**	0.163**	0.069**	0.008**	0.145**	0.041**
X_{45}	-0.004	0.012**	-0.010**	-0.010**	-0.005	0.009**	0.020**	0.012**	0.014**	0.013**	-0.006*	0.011**	0.008**	0.011**	-0.013**	0.006**	0.007*	-0.007*	-0.008**	0.037**
X_{46}	0.112**	-0.036**	0.039**	0.212**	0.191**	0.077**	0.039**	0.064**	0.091**	0.037**	0.145**	0.077**	0.073**	-0.040**	0.254**	0.168**	0.063**	0.022**	0.175**	0.024**
X_{47}	0.024**	0.011**	-0.015**	0.025**	0.053**	0.050**	0.054**	0.033**	0.038**	0.023**	0.023**	0.039**	0.040**	-0.005	0.027**	0.057**	0.019**	-0.006	0.030**	0.057**
X_{48}	0.086**	-0.016**	0.022**	0.162**	0.147**	0.075**	0.050**	0.052**	0.073**	0.038**	0.114**	0.072**	0.067**	-0.030**	0.186**	0.130**	0.048**	0.014**	0.130**	0.037**
X_{49}	0.018**	0.004	-0.004	0.025**	0.035**	0.033**	0.041**	0.024**	0.028**	0.016**	0.031**	0.027**	0.023**	-0.004	0.026**	0.054**	0.018**	-0.009**	0.030**	0.042**
X_{50}	0.050**	-0.014**	0.018**	0.089**	0.083**	0.047**	0.027**	0.044**	0.046**	0.026**	0.077**	0.035**	0.032**	-0.017**	0.114**	0.091**	0.051**	0.001	0.077**	0.018**
$\bar{X}$	0.404	0.194	0.634	0.551	0.420	0.369	0.219	0.389	0.305	0.278	0.458	0.272	0.240	0.236	0.503	0.312	0.279	0.225	0.458	0.192
S.D.	0.491	0.396	0.482	0.497	0.494	0.482	0.414	0.487	0.460	0.448	0.498	0.445	0.427	0.425	0.500	0.463	0.449	0.418	0.498	0.394

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

ตารางที่ 4.36 (ต่อ)

ข้อ	X_{21}	X_{22}	X_{23}	X_{24}	X_{25}	X_{26}	X_{27}	X_{28}	X_{29}	X_{30}	X_{31}	X_{32}	X_{33}	X_{34}	X_{35}	X_{36}	X_{37}	X_{38}	X_{39}	X_{40}
X_{21}	1.000																			
X_{22}	0.135**	1.000																		
X_{23}	0.134**	0.076**	1.000																	
X_{24}	0.148**	0.135**	0.163**	1.000																
X_{25}	0.106**	0.122**	0.109**	0.173**	1.000															
X_{26}	0.085**	0.085**	0.064**	0.138**	0.177**	1.000														
X_{27}	0.083**	0.080**	0.070**	0.103**	0.090**	0.017**	1.000													
X_{28}	0.029**	0.020**	0.023**	0.034**	0.032**	0.041**	0.076**	1.000												
X_{29}	0.115**	0.118**	0.119**	0.091**	0.090**	0.051**	0.071**	0.017**	1.000											
X_{30}	0.095**	0.109**	0.098**	0.067**	0.061**	0.030**	0.051**	-0.015**	0.131**	1.000										
X_{31}	0.104**	0.089**	0.096**	0.130**	0.119**	0.096**	0.069**	0.022**	0.084**	0.051**	1.000									
X_{32}	0.136**	0.134**	0.129**	0.148**	0.135**	0.102**	0.099**	0.027**	0.124**	0.088**	0.074**	1.000								
X_{33}	0.056**	0.071**	0.060**	0.084**	0.069**	0.063**	0.033**	0.008**	0.040**	0.028**	0.048**	0.073**	1.000							
X_{34}	0.051**	0.062**	0.055**	0.062**	0.054**	0.046**	0.032**	0.015**	0.048**	0.051**	0.009**	0.016**	0.072**	1.000						
X_{35}	0.151**	0.154**	0.136**	0.128**	0.124**	0.093**	0.099**	0.012**	0.147**	0.114**	0.110**	0.171**	0.081**	0.097**	1.000					
X_{36}	0.155**	0.162**	0.149**	0.163**	0.144**	0.115**	0.107**	0.021**	0.132**	0.125**	0.118**	0.178**	0.043**	0.086**	0.203**	1.000				
X_{37}	0.158**	0.177**	0.159**	0.152**	0.131**	0.093**	0.105**	0.017**	0.162**	0.148**	0.110**	0.175**	0.068**	0.082**	0.253**	0.146**	1.000			
X_{38}	0.009**	-0.001	0.001	0.039**	0.031**	0.049**	0.017**	0.000	-0.011**	-0.016**	0.038**	0.017**	0.016**	0.009**	-0.005	-0.016**	-0.105**	1.000		
X_{39}	0.176**	0.199**	0.184**	0.154**	0.130**	0.090**	0.107**	0.022**	0.185**	0.169**	0.118**	0.193**	0.066**	0.085**	0.222**	0.243**	0.290**	0.014**	1.000	
X_{40}	0.158**	0.165**	0.152**	0.135**	0.123**	0.097**	0.092**	0.021**	0.165**	0.156**	0.113**	0.174**	0.057**	0.072**	0.198**	0.214**	0.252**	-0.017**	0.251**	1.000
X_{41}	0.128**	0.134**	0.139**	0.128**	0.114**	0.082**	0.082**	0.015**	0.122**	0.117**	0.095**	0.143**	0.060**	0.066**	0.161**	0.179**	0.200**	-0.005	0.225**	0.202**
X_{42}	0.127**	0.170**	0.133**	0.085**	0.085**	0.040**	0.077**	0.014**	0.161**	0.166**	0.086**	0.147**	0.038**	0.063**	0.176**	0.174**	0.225**	-0.038**	0.280**	0.243**
X_{43}	0.114**	0.125**	0.110**	0.105**	0.098**	0.072**	0.080**	0.017**	0.110**	0.102**	0.084**	0.121**	0.040**	0.047**	0.141**	0.155**	0.172**	0.002	0.185**	0.159**
X_{44}	0.132**	0.130**	0.125**	0.118**	0.110**	0.085**	0.085**	0.025**	0.134**	0.118**	0.105**	0.155**	0.057**	0.069**	0.177**	0.180**	0.205**	0.012**	0.231**	0.215**
X_{45}	0.002	-0.003	0.003	0.005	0.004	0.012**	0.011**	0.002	-0.006*	-0.001	0.026**	0.003	0.008*	0.005	-0.009**	-0.003	-0.008*	0.025**	-0.012**	-0.012**
X_{46}	0.141**	0.172**	0.150**	0.130**	0.113**	0.087**	0.100**	0.012**	0.150**	0.144**	0.097**	0.164**	0.056**	0.069**	0.200**	0.213**	0.255**	-0.014**	0.277**	0.246**

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

ตารางที่ 4.36 (ต่อ)

ข้อ	X_{21}	X_{22}	X_{23}	X_{24}	X_{25}	X_{26}	X_{27}	X_{28}	X_{29}	X_{30}	X_{31}	X_{32}	X_{33}	X_{34}	X_{35}	X_{36}	X_{37}	X_{38}	X_{39}	X_{40}
X_{47}	0.040**	0.019**	0.033**	0.068**	0.052**	0.061**	0.035**	0.012**	0.012**	-0.003	0.052**	0.051**	0.031**	0.022**	0.036**	0.048**	0.026**	0.043**	0.009**	0.016**
X_{48}	0.107**	0.133**	0.116**	0.117**	0.106**	0.082**	0.086**	0.014**	0.110**	0.099**	0.092**	0.129**	0.052**	0.055**	0.149**	0.160**	0.187**	0.001	0.191**	0.173**
X_{49}	0.035**	0.016**	0.029**	0.046**	0.036**	0.047**	0.027**	0.008**	0.014**	0.013**	0.047**	0.036**	0.021**	0.014**	0.024**	0.041**	0.027**	0.027**	0.022**	0.030**
X_{50}	0.071**	0.072**	0.070**	0.056**	0.057**	0.045**	0.033**	0.014**	0.086**	0.091**	0.062**	0.088**	0.025**	0.042**	0.089**	0.093**	0.121**	-0.003	0.141**	0.130**
$\bar{X}$	0.449	0.627	0.448	0.279	0.233	0.221	0.365	0.216	0.475	0.484	0.219	0.330	0.268	0.277	0.417	0.443	0.505	0.220	0.496	0.495
S.D.	0.497	0.484	0.497	0.449	0.423	0.415	0.481	0.411	0.499	0.500	0.414	0.470	0.443	0.448	0.493	0.497	0.500	0.414	0.500	0.500

Bartlett's Test of Sphericity = 387449.118 df = 1225 P = .000 Kaiser-Meyer-Olkin Measurement of Sampling Adequacy = .934

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

ตารางที่ 4.36 (ต่อ)

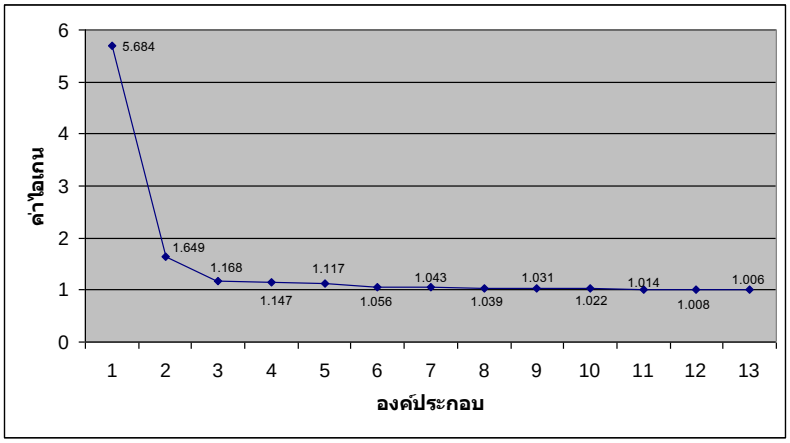
ข้อ	X_{41}	X_{42}	X_{43}	X_{44}	X_{45}	X_{46}	X_{47}	X_{48}	X_{49}	X_{50}
X_{41}	1.000									
X_{42}	0.185**	1.000								
X_{43}	0.123**	0.198**	1.000							
X_{44}	0.178**	0.127**	0.146**	1.000						
X_{45}	-0.009**	0.008**	-0.048**	0.003	1.000					
X_{46}	0.192**	0.253**	0.154**	0.251**	-0.138**	1.000				
X_{47}	0.040**	-0.026**	0.029**	0.037**	-0.011**	-0.125**	1.000			
X_{48}	0.139**	0.168**	0.125**	0.153**	-0.024**	0.183**	-0.076**	1.000		
X_{49}	0.028**	0.008**	0.022**	0.033**	0.029**	-0.004	0.010**	-0.140**	1.000	
X_{50}	0.095**	0.131**	0.083**	0.119**	-0.001	0.132**	0.002	0.140**	0.035**	1.000
$\bar{X}$	0.377	0.659	0.456	0.378	0.263	0.532	0.228	0.447	0.301	0.453
S.D.	0.485	0.474	0.498	0.485	0.440	0.499	0.420	0.497	0.459	0.498

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

การวิเคราะห์องค์ประกอบหลัก (Principle Componant Analysis) หมุนแกนด้วยวิธีแวริแมกซ์ (Varimax) ค่าไอเกน (Eigen Value) และร้อยละของความแปรปรวน ดังตารางที่ 4.37

ตารางที่4.37 ค่าไอเกนและร้อยละของความแปรปรวนขององค์ประกอบในแบบสอบวิชาวิทยาศาสตร์

องค์ประกอบ	ค่าไอเกน	ร้อยละของความแปรปรวน
1	5.684	11.368
2	1.649	3.298
3	1.168	2.336
4	1.147	2.294
5	1.117	2.234
6	1.056	2.112
7	1.043	2.087
8	1.039	2.078
9	1.031	2.062
10	1.022	2.044
11	1.014	2.027
12	1.008	2.017
13	1.006	2.011



ภาพที่ 4.7 ผลการตรวจสอบความเป็นเอกมิติของแบบสอบวิชาวิทยาศาสตร์ จำนวน 50 ข้อ

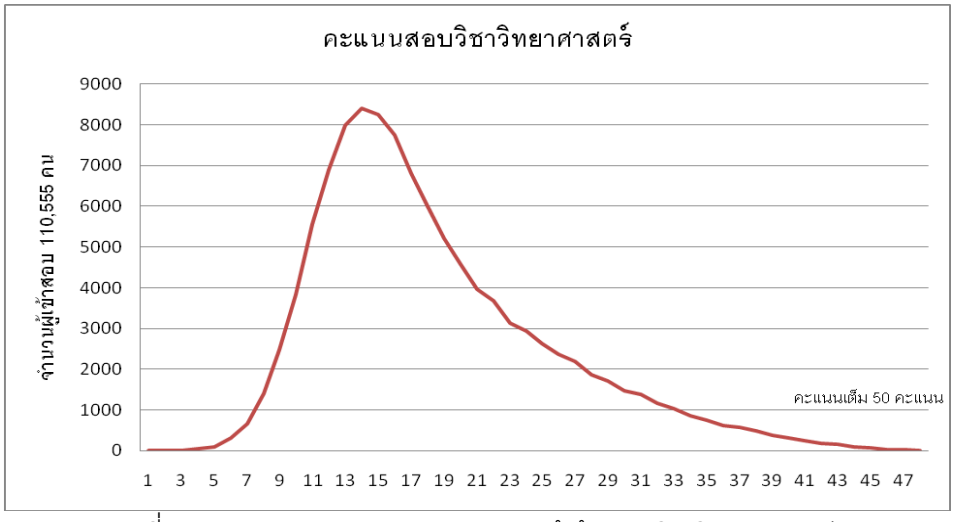
จากตารางที่ 4.37 เมื่อพิจารณาค่าไอเกนจากการวิเคราะห์ผลแบบสอบวิชาวิทยาศาสตร์ จำนวน 50 ข้อ พบว่า ค่าไอเกนขององค์ประกอบที่ 1 มีค่าสูงสุด (5.684) มีค่าสูงกว่าค่าไอเกนขององค์ประกอบที่ 2 (1.649) ประมาณ 3.447 เท่า ซึ่งองค์ประกอบที่ 1 มีค่าไอเกนแตกต่างจากไอเกนของ

องค์ประกอบอื่นอย่างไม่เด่นชัดนัก ส่วนค่าไอเกนขององค์ประกอบอื่นๆ ที่เหลือพบว่ามีค่าใกล้เคียงกัน เมื่อพิจารณาค่าร้อยละของความแปรปรวน พบว่า องค์ประกอบที่ 1 มีค่าร้อยละของความแปรปรวนค่อนข้างสูง (11.368) การพิจารณาค่าไอเกน (eigen value) ซึ่งเสนอโดย Lord และ Novick (1968) ถ้าผลการวิเคราะห์พบว่ามีค่าไอเกนตัวเดียวหรืออาจหลายตัวก็ตามแต่ตัวแรกมีค่ามากกว่าตัวอื่นๆ อย่างเห็นได้ชัด สามารถสรุปได้ว่าเครื่องมือชุดนั้นมีความเป็นเอกมิติ เกณฑ์ของ Reckase (อ้างถึงใน Raju, 1993; อุทัยวรรรณ สายพัฒนะ, 2547) ที่เสนอว่าค่าร้อยละของความแปรปรวนควรมีค่าไม่น้อยกว่าร้อยละ 20 หรือค่าไอเกนขององค์ประกอบที่1 ต้องมีความแตกต่างจากค่าไอเกนขององค์ประกอบอื่นอย่างเด่นชัดจึงจะถือว่าแบบสอบมีความเป็นเอกมิติเมื่อพิจารณาค่าไอเกนและค่าร้อยละของความแปรปรวนขององค์ประกอบที่ 1 และภาพที่ 4.3 ประกอบกันในทางปฏิบัติถือได้ว่าแบบทดสอบฉบับนี้มีความเป็นเอกมิติสามารถนำข้อมูลไปวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบได้

4.2.3 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาวิทยาศาสตร์

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาวิทยาศาสตร์ เนื่องจากมีข้อสังเกตเกี่ยวกับผลการตอบข้อสอบของนักเรียนส่วนใหญ่ที่ไม่มีคะแนนจากการสอบ การนำผลการตอบข้อสอบลักษณะดังกล่าวไปคำนวณการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ อาจให้ผลลัพธ์ที่คลาดเคลื่อนไปจากความเป็นจริง ผู้วิจัยจึงตัดกรณีที่ผู้เข้าสอบได้คะแนนการสอบ 0 คะแนนออก แล้วนำผลการตอบที่เหลือมาคำนวณ ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาวิทยาศาสตร์

เมื่อจำแนกคะแนนการสอบของนักเรียนทั้งหมดพบว่าส่วนใหญ่ได้คะแนนน้อย มีนักเรียนที่ได้คะแนน 0 คะแนน เป็นจำนวน 54 คน คิดเป็นร้อยละ 0.048 จากจำนวนผู้เข้าสอบทั้งหมด การแจกแจงความถี่ของคะแนนสอบไม่นำนักเรียนที่ได้คะแนน 0 คะแนนมาแสดงผล รายละเอียดดังภาพที่ 4.7



ภาพที่ 4.8 กราฟแสดงคะแนนการสอบของผู้เข้าสอบวิชาวิทยาศาสตร์

1) วิธีแมนเทล-แฮนส์เซล

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีแมนเทล -แฮนส์เซล ในแบบสอบวิทยาศาสตร์ 50 ข้อ รายละเอียดดังตารางที่ 4.38

ตารางที่ 4.38 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซล วิชาวิทยาศาสตร์

ข้อสอบ	ค่าสถิติ	p-value	ข้อสอบ	ค่าสถิติ	p-value
1	4.7101	0.0300 *	26	56.7542	0.0000 ***
2	48.743	0.0000 ***	27	11.0183	0.0009 ***
3	3.2829	0.0700	28	19.8176	0.0000 ***
4	223.3776	0.0000 ***	29	0.0712	0.7895
5	0.0000	0.9997	30	0.314	0.5752
6	57.1037	0.0000 ***	31	6.6159	0.0101 *
7	60.5255	0.0000 ***	32	23.8247	0.0000 ***
8	112.1921	0.0000 ***	33	342.9173	0.0000 ***
9	0.6966	0.4039	34	29.9951	0.0000 ***
10	62.9202	0.0000 ***	35	17.6077	0.0000 ***
11	6.2267	0.0126 *	36	1.6047	0.2052
12	1.4591	0.2271	37	4.4374	0.0352 *
13	5.3919	0.0202 *	38	17.3437	0.0000 ***
14	85.8293	0.0000 ***	39	53.74	0.0000 ***
15	381.0867	0.0000 ***	40	16.7049	0.0000 ***
16	0.0926	0.7609	41	1.6906	0.1935
17	59.3484	0.0000 ***	42	28.7507	0.0000 ***
18	6.0886	0.0136 *	43	91.5224	0.0000 ***
19	39.0238	0.0000 ***	44	2.8327	0.0924
20	0.9544	0.3286	45	12.9844	0.0003 ***
21	0.0027	0.9585	46	5.2077	0.0225 *
22	44.4778	0.0000 ***	47	6.0209	0.0141 *
23	2.4648	0.1164	48	0.0151	0.9022
24	0.969	0.3249	49	8.3755	0.0038 **
25	20.7474	0.0000 ***	50	28.5725	0.0000 ***

****P<.001, **P<.01, *P<.05*

จากตารางที่ 4.38 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธีแมนเทล-แฮนส์เซล วิชาวิทยาศาสตร์ 50 ข้อ พิจารณาค่า p-value หากข้อสอบข้อใดมีนัยสำคัญ หมายความว่าข้อสอบข้อนั้นทำหน้าที่ต่างกัน พบข้อที่ทำหน้าที่ต่างกัน 35 ข้อ ได้แก่ ข้อ 1, 2, 4, 6, 7, 8, 10, 11, 13, 14, 15, 17, 18, 19, 22, 25, 26, 27, 28, 31, 32, 33, 34, 35, 37, 38, 39, 40, 42, 43, 45, 46, 47, 49 และข้อ 50 ข้อที่ทำหน้าที่ไม่ต่างกันมี 15 ข้อ ได้แก่ ข้อ 3, 5, 9, 12, 16, 20, 21, 23, 24, 29, 30, 36, 41, 44, และข้อ 48

2) วิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ (significance test)

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้วิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ วิชาวิทยาศาสตร์ 50 ข้อ ใช้คะแนนรวมทั้งฉบับ รายละเอียดดังตารางที่ 4.39

ตารางที่ 4.39 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ภายใต้วิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ ในวิชาวิทยาศาสตร์

ข้อ	กลุ่มผู้สอบ (g)		ปฏิสัมพันธ์ระหว่างกลุ่มกับความสามารถ (g by x)		ข้อ	กลุ่มผู้สอบ (g)		ปฏิสัมพันธ์ระหว่างกลุ่มกับความสามารถ (g by x)	
	B	sig	B	sig		B	sig	B	sig
1	-.028	.035*	-.009	.000*	21	.002	.856	.009	.000*
2	.120	.000*	.005	.042*	22	-.091	.000*	-.004	.185
3	.030	.019*	-.030	.000*	23	-.021	.108	-.002	.336
4	-.210	.000*	-.003	.279	24	.019	.213	.015	.000*
5	-.003	.809	.004	.098	25	-.068	.000*	.010	.000*
6	.108	.000*	.009	.000*	26	-.113	.000*	.032	.000*
7	.132	.000*	.026	.000*	27	-.040	.003*	.013	.000*
8	.148	.000*	.004	.028*	28	.072	.000*	-.007	.001*
9	-.007	.634	-.004	.034*	29	.006	.624	-.008	.001*
10	.120	.000*	.008	.000*	30	-.006	.656	-.023	.000*
11	-.032	.017*	-.010	.000*	31	-.034	.034*	.025	.000*
12	.025	.080	.011	.000*	32	-.070	.000*	.007	.002*
13	.039	.009*	.007	.002*	33	.274	.000*	-.006	.002*
14	.147	.000*	-.015	.002*	34	.081	.000*	-.006	.002*
15	-.291	.000*	.006	.065	35	-.057	.000*	-.003	.216
16	-.002	.899	.023	.000*	36	-.022	.115	.017	.000*
17	.115	.000*	-.001	.561	37	-.036	.010*	.012	.000*
18	.041	.006*	-.019	.000*	38	.081	.000*	.022	.000*
19	-.085	.000*	-.003	.271	39	-.109	.000*	.009	.003*
20	.032	.044*	.030	.000*	40	-.060	.000*	.004	.121

ตารางที่ 4.39 (ต่อ)

ข้อ	กลุ่มผู้สอบ (g)		ปฏิสัมพันธ์ระหว่างกลุ่มกับความสามารถ (g by x)		ข้อ	กลุ่มผู้สอบ (g)		ปฏิสัมพันธ์ระหว่างกลุ่มกับความสามารถ (g by x)	
	B	sig	B	sig		B	sig	B	sig
41	-.019	.163	-.006	.005*	46	-.037	.037*	.007	.009*
42	-.079	.000*	-.006	.069	47	.048	.002*	.034	.000*
43	-.125	.000*	.001	.646	48	-.002	.907	.005	.014*
44	.024	.085	.004	.095	49	.051	.000*	.017	.000*
45	.064	.000*	.014	.000*	50	.075	.000*	-.009	.000*

*$p < .05$

จากตารางที่ 4.39 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ภายใต้วิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ ของแบบสอบวิชาวิทยาศาสตร์ เมื่อใช้คะแนนรวมทั้งฉบับเป็นเกณฑ์จับคู่ การตรวจสอบ DIF พิจารณาค่า p-value หากข้อสอบข้อใดมีนัยสำคัญ หมายความว่าข้อสอบข้อนั้นทำหน้าที่ต่างกัน มีข้อสอบที่ทำหน้าที่ต่างกันมีทั้งหมด 47 ข้อ

ค่านัยสำคัญของกลุ่มผู้เข้าสอบ (g) ถ้าข้อสอบข้อใดมีนัยสำคัญที่ระดับ .05 หมายความว่าข้อสอบข้อนั้นทำหน้าที่ต่างกันแบบเอกรูป ผลจากการตรวจสอบพบข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป จำนวน 9 ข้อ คือ ข้อ 4, 15, 17, 19, 22, 35, 40, 42 และข้อ 43

ข้อสอบที่ทำหน้าที่ไม่ต่างกันมี 3 ข้อ คือ ข้อ 5, 23 และข้อ 44

ค่านัยสำคัญของปฏิสัมพันธ์ระหว่างกลุ่มผู้เข้าสอบกับความสามารถ (g by x) ถ้าข้อสอบข้อใดมีนัยสำคัญที่ระดับ .05 หมายความว่าข้อสอบข้อนั้นทำหน้าที่ต่างกันแบบอเนกรูป ผลจากการตรวจสอบ พบว่ามีข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป จำนวน 38 ข้อ คือ ข้อ 1, 2, 3, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 18, 20, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 36, 37, 38 และข้อ 39

3) การวัดขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas (1997)

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas วิชาวิทยาศาสตร์ 50 ข้อ ดังตารางที่ 4.40

ตารางที่ 4.40 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติกโดยการวัดขนาดอิทธิพลในวิชาวิทยาศาสตร์ ตามเกณฑ์ของ Zumbo and Thomas

ข้อ	ขนาดอิทธิพล		ข้อ	ขนาดอิทธิพล		ข้อ	ขนาดอิทธิพล		ข้อ	ขนาดอิทธิพล	
	R^2	Amount of DIF		R^2	Amount of DIF		R^2	Amount of DIF		R^2	Amount of DIF
1	.0000	*	14	.1105	*	27	.0000	*	40	.0000	*
2	.0862	*	15	.0000	*	28	.0000	*	41	.0000	*
3	.0000	*	16	.0000	*	29	.0000	*	42	.0000	*
4	.0000	*	17	.0000	*	30	.0000	*	43	.0000	*
5	.0000	*	18	.0000	*	31	.0000	*	44	.0000	*
6	.0000	*	19	.0000	*	32	.0000	*	45	.0000	*
7	.0000	*	20	.0000	*	33	.0000	*	46	.0000	*
8	.0000	*	21	.0000	*	34	.0000	*	47	.0000	*
9	.0000	*	22	.0000	*	35	.0000	*	48	.0000	*
10	.0000	*	23	.0000	*	36	.0000	*	49	.0000	*
11	.0000	*	24	.0000	*	37	.0000	*	50	.0000	*
12	.0000	*	25	.0000	*	38	.0000	*			
13	.0000	*	26	.0000	*	39	.0000	*			

*Effect size (Nagelkerke's R^2): Zumbo and Thomas (ZT): * 0.00-0.13 negligible effect, **0.13- 0.26 moderate effect, ***<0.26 large effect*

จากตารางที่ 4.40 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก โดยการวัดขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas วิชาวิทยาศาสตร์ จำนวน 50 ข้อ พบว่า ข้อสอบทุกข้อทำหน้าที่ต่างกันโดยมีขนาดอิทธิพลของการทำหน้าที่ต่างกันขนาด เดียว คือ ขนาดเล็กจนแทบจะไม่มีเลย (*.00< R^2 <.13*) ทั้ง 50 ข้อ

4) การวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl (2001)

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl ในแบบสอบวิชาวิทยาศาสตร์ จำนวน 50 ข้อ ดังตารางที่ 4.41

ตารางที่ 4.41 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติกโดยการวัดขนาดอิทธิพลในวิชาคณิตศาสตร์ ตามเกณฑ์ Jodoin and Gierl

ข้อ	ขนาดอิทธิพล		ข้อ	ขนาดอิทธิพล		ข้อ	ขนาดอิทธิพล		ข้อ	ขนาดอิทธิพล	
	R^2	Amount of DIF		R^2	Amount of DIF		R^2	Amount of DIF		R^2	Amount of DIF
1	.0000	*	14	.1105	***	27	.0000	*	40	.0000	*
2	.0862	***	15	.0000	*	28	.0000	*	41	.0000	*
3	.0000	*	16	.0000	*	29	.0000	*	42	.0000	*
4	.0000	*	17	.0000	*	30	.0000	*	43	.0000	*
5	.0000	*	18	.0000	*	31	.0000	*	44	.0000	*
6	.0000	*	19	.0000	*	32	.0000	*	45	.0000	*
7	.0000	*	20	.0000	*	33	.0000	*	46	.0000	*
8	.0000	*	21	.0000	*	34	.0000	*	47	.0000	*
9	.0000	*	22	.0000	*	35	.0000	*	48	.0000	*
10	.0000	*	23	.0000	*	36	.0000	*	49	.0000	*
11	.0000	*	24	.0000	*	37	.0000	*	50	.0000	*
12	.0000	*	25	.0000	*	38	.0000	*			
13	.0000	*	26	.0000	*	39	.0000	*			

*Effect size (Nagelkerke's R^2): Jodoign and Gierl (JG): * 0.00-0.035 negligible effect, **0.0351- 0.07 moderate effect, ***0.071-1.00 large effect*

จากตารางที่ 4.41 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl พบว่า ข้อสอบทุกข้อทำหน้าที่ต่างกัน โดยมีขนาดของการทำหน้าที่ต่างกัน 2 ขนาด คือ ขนาดเล็กน้อย และขนาดใหญ่ จำแนกได้ ว่าข้อสอบที่ทำหน้าที่ต่างกันขนาดเล็กน้อยจนแทบจะไม่มีเลย ($.00< R^2 <.035$) มีจำนวน 48 ข้อ ยกเว้นข้อ 2 และข้อ 14 ที่มีขนาดอิทธิพลของการทำหน้าที่ต่างกันขนาดใหญ่ ($.071 < R^2$)

5) สรุปผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาวิทยาศาสตร์

สรุปผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ (LR_S) กับการวัดขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas (LR_Z) และเกณฑ์ Jodoin and Gierl (LR_J) ดังตารางที่ 4.42

ตารางที่ 4.42 สรุปผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาวิทยาศาสตร์

แบบสอบ	ความยาว (ข้อ)	จำนวนข้อสอบ DIF (ข้อ)				จำนวนข้อสอบ NO DIF (ข้อ)			
		MH	LR_S	LR_J	LR_Z	MH	LR_S	LR_J	LR_Z
วิทยาศาสตร์	50 ข้อ	35	47	2	0	15	3	48	50

จากตารางที่ 4.42 ผลการตรวจสอบเปรียบเทียบผลภายใต้วิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพล 2 เกณฑ์ ในแบบสอบวิชา วิทยาศาสตร์ พบว่า การทดสอบระดับนัยสำคัญตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 47 ข้อ การวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl พบข้อสอบ DIF ที่มีขนาดปานกลางขึ้นไปจำนวน 2 ข้อ การวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas ไม่พบข้อสอบที่ทำหน้าที่ต่างกัน ผลสรุปของจำนวนการตรวจพบการทำหน้าที่ต่างกันของข้อสอบมีรายละเอียดดังตารางที่ 4.43

ตารางที่ 4.43 จำนวนข้อของการเกิดและไม่เกิดการทำหน้าที่ต่างกันของข้อสอบของวิชาวิทยาศาสตร์

วิธีแมนเทล-แฮนส์เซล		LR_S		LR_J		LR_Z	
		DIF	NO DIF	DIF	NO DIF	DIF	NO DIF
ข้อสอบทำหน้าที่ต่างกัน	(DIF)	35	0	2	33	0	35
ข้อสอบไม่ทำหน้าที่ต่างกัน	(NO DIF)	12	3	0	15	0	15

หมายเหตุ *: ข้อสอบทั้งฉบับมีจำนวน 50 ข้อ*

จากตารางที่ 4.43 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญกับวิธีแมนเทล-แฮนส์เซลซึ่งเป็นวิธีเกณฑ์ พบข้อร่วมของการทำหน้าที่ต่างกันระหว่างสองวิธีดังกล่าว จำนวน 35 ข้อ จากข้อสอบ 50 ข้อ คิดเป็นร้อยละ 70.00

ระหว่างวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตาม เกณฑ์ Jodoin and Gierl กับวิธีแมนเทล-แฮนส์เซลซึ่งเป็นวิธีเกณฑ์ พบข้อร่วมของการทำหน้าที่ต่างกันระหว่างสองวิธีดังกล่าว จำนวน 2 ข้อ จากข้อสอบ 50 ข้อ คิดเป็นร้อยละ 4.00

ระหว่างวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas กับวิธีแมนเทล-แฮนส์เซลซึ่งเป็นวิธีเกณฑ์ ไม่พบข้อร่วมของการทำหน้าที่ต่างกันระหว่างสองวิธีดังกล่าว จากข้อสอบ 50 ข้อ คิดเป็นร้อยละ 0.00

4.2.4 อัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1

สำหรับผลการวิเคราะห์อัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 มีรายละเอียดดังตารางที่ 4.44

ตารางที่ 4.44 ร้อยละของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 จำแนกตามการตรวจสอบการทำหน้าที่ต่างกันของแบบสอบตามวิธีที่ศึกษา วิชาวิทยาศาสตร์

ร้อยละของอัตราความถูกต้อง		ร้อยละของอัตราความคลาดเคลื่อนประเภทที่ 1	
Jodoin and Gierl	Zumbo and Thomas	Jodoin and Gierl	Zumbo and Thomas
5.71	0.00	0.00	0.00

จากตารางที่ 4.44 ร้อยละของอัตราความถูกต้อง วิชา วิทยาศาสตร์ โดยการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl คิดเป็นร้อยละ 12.50 โดยการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas คือ คิดเป็นร้อยละ 6.25

ร้อยละของอัตราความคลาดเคลื่อนประเภทที่ 1 โดยการวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl คิดเป็นร้อยละ 0.00 โดยการวัดขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas คิดเป็นร้อยละ 0.00

เกณฑ์การพิจารณาประสิทธิภาพจากผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีอัตราความถูกต้องในการตรวจสอบของข้อสอบสูง และมีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบของข้อสอบต่ำแสดงถึงประสิทธิภาพในการตรวจสอบสูงสุด ถือเป็นเงื่อนไขที่ต้องการ (รายละเอียดตอนที่ 1 ภาพประกอบ 4.1) ผลจากการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบวิชา วิทยาศาสตร์ ของทุกวิธีที่ศึกษาเป็นไปตามเกณฑ์เงื่อนไขที่ต้องการ ผลประสิทธิภาพพบว่า เกณฑ์ Jodoin and Gierl ให้ค่าร้อยละของอัตราความถูกต้อง ในการตรวจสอบข้อสอบ สูงกว่าเกณฑ์ Zumbo and Thomas ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบข้อสอบคิดเป็นร้อยละที่เท่ากัน ตัดสินผลการเปรียบเทียบประสิทธิภาพระหว่าง 2 เกณฑ์ได้ว่าภายใต้วิธีถดถอยโลจิสติกโดยการวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl มีประสิทธิภาพดีกว่าการวัดขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas

4.3 ประโยชน์ของการพิจารณาขนาดอิทธิพลร่วมกับการนำผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบไปใช้เพื่อการตัดสินใจ

วัตถุประสงค์หลักของการศึกษาครั้งนี้คือ การศึกษาประสิทธิภาพ การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ภายใต้วิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl

และเกณฑ์ Zumbo and Thomas การศึกษาครั้งนี้ศึกษาประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเป็นหลัก ยังมิได้เน้นการตรวจสอบเพื่อนำผลไปปรับปรุงข้อสอบ หลักการสำคัญประการหนึ่งในการศึกษาการตรวจสอบคือการเลือกวิธีการตรวจสอบ ซึ่งถือว่ามีความสำคัญมากเนื่องจากนักการศึกษาต้องพิจารณาที่ความเหมาะสมของข้อมูลที่จะศึกษา เช่น รูปแบบข้อสอบ รูปแบบการตรวจให้คะแนน จำนวนผู้เข้าสอบและความยาวแบบสอบ

ผู้วิจัยขอเสนอตัวอย่างผลการตรวจสอบในแบบสอบวิชาคณิตศาสตร์ 40 ข้อ พบว่าการศึกษาภายใต้วิธี ถดถอยโลจิสติก นั้น การทดสอบระดับนัยสำคัญ ให้ประสิทธิภาพการตรวจสอบที่ดีที่สุด แต่ผู้วิจัยไม่ได้มุ่งเอาผลจากการทดสอบระดับนัยสำคัญ ดังกล่าวมาร่วมเปรียบเทียบเพราะต้องการศึกษาประสิทธิภาพของการวัดขนาดอิทธิพลจากเกณฑ์ 2 เกณฑ์ เป็นหลัก เพื่อแสดงให้เห็นถึงประโยชน์ของการพิจารณาขนาดอิทธิพลร่วมกับการนำผลการตรวจสอบเพื่อการตัดสินใจ รายละเอียดดังตารางที่ 4.45

ตารางที่ 4.45 ผลการตรวจสอบ DIF ระหว่าง การทดสอบระดับนัยสำคัญ (LR_S) กับการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl (LR_J)

ข้อสอบ	การทดสอบระดับนัยสำคัญ				ผล DIF (LR_S)	ขนาดอิทธิพลเกณฑ์ Jodoin and Gierl		การพิจารณาข้อสอบ (LR_S) ร่วมกับ (LR_J)
	กลุ่มผู้สอบ (g)		ปฏิสัมพันธ์ระหว่างกลุ่มกับความสามารถ(g by x)			R^2	ผลขนาดอิทธิพล (LR_J)	
	B	sig	B	sig				
1	-.122	.038*	-.164	.000*	DIF	.0000	เล็กมากๆ	คงข้อสอบไว้
2	-.328	.004*	-.003	.864	DIF	.0110	เล็กมากๆ	คงข้อสอบไว้
3	.138	.013	-.089	.000*	DIF	.0000	เล็กมากๆ	คงข้อสอบไว้
4	-.296	.000*	-.105	.000*	DIF	.0000	เล็กมากๆ	คงข้อสอบไว้
5	-.627	.000*	-.015	.331	DIF	.0000	เล็กมากๆ	คงข้อสอบไว้
6	.037	.373	-.148	.000*	DIF	.0000	เล็กมากๆ	คงข้อสอบไว้
7	-.790	.000*	-.047	.001*	DIF	.0000	เล็กมากๆ	คงข้อสอบไว้
8	.910	.000*	.046	.087	DIF	.6390	ขนาดใหญ่	ตัดสินใจว่าปรับปรุง/แก้ไขหรือสร้างใหม่
9	-.507	.000*	.063	.000*	DIF	.0000	เล็กมากๆ	คงข้อสอบไว้
10	-1.739	.000*	.036	.029	DIF	.0000	เล็กมากๆ	คงข้อสอบไว้
11	.222	.000*	.007	.447	DIF	.0453	ขนาดปานกลาง	ต้องปรับปรุง/แก้ไข
12	.119	.002*	-.108	.000*	DIF	.0000	เล็กมากๆ	คงข้อสอบไว้
13	-.419	.000*	-.080	.000*	DIF	.0000	เล็กมากๆ	คงข้อสอบไว้
14	-.359	.000*	-.055	.000*	DIF	.0000	เล็กมากๆ	คงข้อสอบไว้
15	-.280	.000*	-.016	.230	DIF	.0000	เล็กมากๆ	คงข้อสอบไว้
16	-.067	.167	.004	.756	No DIF	.0000	เล็กมากๆ	คงข้อสอบไว้
17	.714	.000*	-.097	.000*	DIF	.0000	เล็กมากๆ	คงข้อสอบไว้
18	.131	.025*	.166	.000*	DIF	.0000	เล็กมากๆ	คงข้อสอบไว้
19	.535	.000*	-.035	.000*	DIF	.2536	ขนาดใหญ่	ตัดสินใจว่าปรับปรุง/แก้ไขหรือสร้างใหม่
20	-.534	.000*	-.038	.013*	DIF	.0000	เล็กมากๆ	คงข้อสอบไว้
21	-.170	.000*	-.021	.034*	DIF	.0000	เล็กมากๆ	คงข้อสอบไว้
22	-.346	.000*	-.028	.072	DIF	.0000	เล็กมากๆ	คงข้อสอบไว้

ตารางที่ 4.45 (ต่อ)

ข้อสอบ	การทดสอบระดับนัยสำคัญ				ผล DIF (LR_S)	ขนาดอิทธิพลเกณฑ์ Jodoin and Gierl		การพิจารณาข้อสอบ (LR_S) ร่วมกับ (LR_J)
	กลุ่มผู้สอบ (g)		ปฏิสัมพันธ์ระหว่างกลุ่มกับความสามารถ(g by x)			R^2	ผลขนาดอิทธิพล (LR_J)	
	B	sig	B	sig				
23	.434	.000*	-.061	.000*	DIF	.0000	เล็กมากๆ	คงข้อสอบไว้
24	-.270	.000*	-.127	.000*	DIF	.0000	เล็กมากๆ	คงข้อสอบไว้
25	-.028	.201	-.003	.753	No DIF	.0000	เล็กมากๆ	คงข้อสอบไว้
26	-.423	.000*	.022	.107	DIF	.0000	เล็กมากๆ	คงข้อสอบไว้
27	.530	.000*	-.065	.000*	DIF	.0000	เล็กมากๆ	คงข้อสอบไว้
28	-.624	.000*	.003	.853	DIF	.0158	เล็กมากๆ	คงข้อสอบไว้
29	.455	.000*	-.075	.000*	DIF	.0000	เล็กมากๆ	คงข้อสอบไว้
30	-.011	.770	.005	.625	No DIF	.0000	เล็กมากๆ	คงข้อสอบไว้
31	-.031	.305	-.065	.000*	DIF	.0000	เล็กมากๆ	คงข้อสอบไว้
32	.475	.000*	-.002	.888	DIF	.0402	ขนาดปานกลาง	ต้องปรับปรุง/แก้ไข
33	-.410	.000*	-.037	.007*	DIF	.0000	เล็กมากๆ	คงข้อสอบไว้
34	-.111	.000*	-.052	.000*	DIF	.0000	เล็กมากๆ	คงข้อสอบไว้
35	-.293	.049*	.013	.502	DIF	.0000	เล็กมากๆ	คงข้อสอบไว้
36	-.503	.000*	-.065	.000*	DIF	.0000	เล็กมากๆ	คงข้อสอบไว้
37	.197	.001*	.05	.000*	DIF	.0202	เล็กมากๆ	คงข้อสอบไว้
38	.440	.000*	.003	.803	DIF	.0398	เล็กมากๆ	คงข้อสอบไว้
39	-.007	.966	.016	.817	No DIF	.0014	เล็กมากๆ	คงข้อสอบไว้
40	.169	.039*	.041	.077	DIF	.0335	เล็กมากๆ	คงข้อสอบไว้

*$p < .05$ **ผลขนาดอิทธิพล**: .00-.035 = DIF ขนาดเล็กมากๆ; .0351- .07 = DIF ขนาดปานกลาง; <.071 = DIF ขนาดใหญ่

จากตารางที่ 4.45 แบบสอบคณิตศาสตร์ที่มีความยาว 40 ข้อ เมื่อศึกษาผลของการทำข้อสอบจากเด็กนักเรียน 2 กลุ่ม คือ นักเรียนในเขต อำเภอเมือง (กลุ่มอ้างอิง : reference groups) กับนักเรียนนอกเขตอำเภอเมือง (กลุ่มเปรียบเทียบ : focal groups) ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ พบว่าวิธีการทดสอบระดับนัยสำคัญพบข้อที่ทำหน้าที่ต่างกัน 36 ข้อ ซึ่งทราบแต่เพียงว่าข้อสอบเกิดการทำหน้าที่ต่างกันแต่ไม่ทราบสารสนเทศอื่น นั่นก็ถือว่ายังไม่เพียงพอสำหรับการตัดสินใจว่าจะตัดสินข้อสอบอย่างไร จะตัดข้อสอบข้อนั้นออกจากแบบสอบหรือคงข้อสอบข้อนั้นไว้ในแบบสอบเพื่อเก็บเข้าธนาคารข้อสอบต่อไป การวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl พบข้อที่ทำหน้าที่ต่างกัน 4 ข้อ หากใช้เกณฑ์การตัดสินใจ โดยพิจารณาจากวิธีการหลักที่ศึกษาร่วมกับการวัดขนาดอิทธิพลที่มีขนาดปานกลางและขนาดใหญ่ ตามเกณฑ์ดังกล่าวในการตัดสินข้อสอบจะมีข้อสอบที่ทำหน้าที่ต่างกันในระดับที่ต้องพิจารณาอย่างจริงจังเพียง 4 ข้อ ต้องดำเนินการปรับปรุงหรือแก้ไขที่ตัวข้อสอบ ผลที่เกิดขึ้นเกิดผลดีด้านการบริหารจัดการข้อสอบ เกิดความประหยัดทั้งเวลา แรงงาน มันสมองและทุนทรัพย์ ดังนั้นจึงควรสนับสนุนให้พิจารณาขนาดอิทธิพลของการทำหน้าที่ต่างกันของข้อสอบร่วมกับการตรวจสอบโดยวิธีการหลักเลือกศึกษา เพื่อให้เกิดความรอบคอบต่อการตัดสินข้อสอบและแบบสอบ

บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

การวิจัยครั้งนี้มีวัตถุประสงค์เฉพาะของการวิจัย คือ 1) เพื่อเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่มีรูปแบบการให้คะแนนแบบทวิวิภาคโดยการจำลองข้อมูลในวิธี ถดถอยโลจิสติก ระหว่างขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl และขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas **ภายใต้เงื่อนไขเดียวกัน** ของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน ขนาดของการทำหน้าที่ต่างกัน จำนวนข้อสอบที่ทำหน้าที่ต่างกัน ความยาวของแบบสอบทั้งฉบับ 2) เพื่อเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีรูปแบบการให้คะแนนแบบทวิวิภาคโดยการจำลองข้อมูลในวิธี ถดถอยโลจิสติก ด้วยขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl และขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas **ภายใต้เงื่อนไขต่างกัน**ของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และ 3) เพื่อเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่มีรูปแบบการให้คะแนนแบบทวิวิภาคโดยข้อมูลเชิงประจักษ์ในวิธีถดถอยโลจิสติก ระหว่างขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl และขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas

ผู้วิจัยตั้งสมมติฐานการวิจัย 3 ข้อ คือ 1) วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas และการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีรูปแบบการให้คะแนนแบบทวิวิภาค ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน4 ปัจจัยและปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย มีอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกัน 2) ปัจจัยที่แปรเปลี่ยน 4 ปัจจัยและปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย มีผลทำให้อัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่มีรูปแบบการให้คะแนนแบบทวิวิภาค ในวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas แตกต่างกัน และโดยการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl แตกต่างกัน และ 3) วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas และการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่มีรูปแบบการให้คะแนนแบบทวิวิภาค ภายใต้ข้อสอบของข้อมูลเชิงประจักษ์ มีผลทำให้อัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกัน

การศึกษาข้อมูลจำลอง ศึกษาเงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน 2 รูปแบบ ขนาดของการทำหน้าที่ต่างกัน 3 ขนาด จำนวนข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด และความยาวของแบบสอบทั้งฉบับ 2 ขนาด รวม 24 เงื่อนไข (2×3×2×2) แต่ละเงื่อนไขทำซ้ำเงื่อนไขละ 25 ครั้ง การวิเคราะห์ข้อมูลใช้โปรแกรม WinGen จำลองข้อมูลให้ได้ข้อมูลที่มีความเหมาะสม และเพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบตามวิธีที่ศึกษาตรวจสอบความถูกต้องของข้อมูลจำลองโดยใช้โปรแกรมDIFAS ใช้โปรแกรมMULTILOG และโปรแกรมSPSS เพื่อหาคุณภาพพื้นฐานของรูปแบบการตอบข้อสอบ

ผู้วิจัยวิเคราะห์ข้อมูลเพื่อตอบคำถามการวิจัยตามสมมติฐาน ด้วยการเปรียบเทียบความแตกต่างของอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ในวิธีถดถอยโลจิสติก ภายใต้ปัจจัยที่ศึกษาด้วยการวิเคราะห์ความแปรปรวนหลายตัวแปร (Multivariate analysis of variance; MANOVA) ที่ระดับนัยสำคัญ .001 กำหนดการวิเคราะห์ให้มีปฏิสัมพันธ์ระหว่างตัวแปรอิสระไม่เกินอันดับที่สอง ถ้าผลการทดสอบมีนัยสำคัญทางสถิติจะทดสอบผลระหว่างกลุ่ม (Test of between-subjects effects) ของตัวแปรตามที่ระดับนัยสำคัญ .001 แล้วทดสอบผลย่อย (Simple effect) ภายใต้ตัวแปรที่ศึกษา ที่ระดับนัยสำคัญ.001 และทดสอบภายหลังด้วยวิธีของเชฟเฟ (Scheffé) ใช้ระดับนัยสำคัญระดับเดียวกับการทดสอบผลย่อย ตัวแปรตาม มี 2 ตัว คือ อัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ส่วนตัวแปรอิสระมี 5 ตัว คือ 1) การวัดขนาดอิทธิพล ภายใต้วิธีถดถอยโลจิสติก โดยใช้เกณฑ์ขนาดอิทธิพล 2 วิธี 2) รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน 3) ขนาดของการทำหน้าที่ต่างกัน 4) จำนวนข้อสอบที่ทำหน้าที่ต่างกัน และ 5) ความยาวของแบบสอบทั้งฉบับ

การวิเคราะห์ความแปรปรวนพหุ ได้ผลการทดสอบปัจจัยที่ศึกษาที่มีผลต่อ อัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่1 ของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย มี 3 ปัจจัยที่มีอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ได้แก่ ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน และปัจจัย ขนาดของ การทำหน้าที่ต่างกันของข้อสอบ ส่วนปัจจัยความยาวของแบบสอบทั้งฉบับ ไม่พบความแตกต่าง การวิเคราะห์ความแปรปรวนพหุได้ผลการทดสอบปัจจัยที่ศึกษาที่มีผลต่ออัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่1 ของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ภายใต้ ปฏิสัมพันธ์สองทางระหว่างวิธีการตรวจสอบกับ เงื่อนไขของปัจจัยที่แปรเปลี่ยน พบว่า มีอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ใน 3 เงื่อนไขย่อย ได้แก่ 1) วิธีการตรวจสอบ กับปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน 2) วิธีการตรวจสอบ กับ ปัจจัยความยาวของแบบสอบทั้งฉบับ 3) วิธีการตรวจสอบ กับ ปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ และการ วิเคราะห์ความแปรปรวนพหุ ได้ผลการทดสอบปัจจัยที่ศึกษาที่มีผลต่อ อัตราความถูกต้องและอัตราความคลาดเคลื่อน

ประเภทที่ 1 ของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ภายใต้ปฏิสัมพันธ์สามทาง ระหว่างวิธีการตรวจสอบ กับ เงื่อนไขของปัจจัยที่แปรเปลี่ยน พบว่า มีอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ใน 4 เงื่อนไขย่อย คือ 1) วิธีการตรวจสอบ กับ ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน และ ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 2) วิธีการตรวจสอบ กับ ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน และ ปัจจัยความยาวของแบบสอบทั้งฉบับ 3) วิธีการตรวจสอบ กับปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน และ ปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ 4) วิธีการตรวจสอบ กับ ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน และ ปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ

ผลการทดสอบปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ในข้อมูลจำลอง ที่มีผลต่อประสิทธิภาพด้านอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีถดถอยโลจิสติก ระหว่างขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl และขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas โดยการทดสอบผลย่อย (Simple effect) สรุปได้ใน 2 ประเด็น อันนำไปสู่การตอบคำถามการวิจัย คือ 1) ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน ขนาดของการทำหน้าที่ต่างกัน ปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ย น ระหว่าง รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน กับ ขนาดของการทำหน้าที่ต่างกัน และ ระหว่างจำนวนข้อสอบที่ทำหน้าที่ต่างกัน กับ ขนาดของการทำหน้าที่ต่างกัน มีผลต่ออัตราความถูกต้องของ วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ของ Zumbo and Thomas อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ยกเว้น ความยาวของแบบสอบทั้งฉบับ และ ปฏิสัมพันธ์ระหว่าง รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน กับจำนวนข้อสอบที่ทำหน้าที่ต่างกัน ในขณะเดียวกันปฏิสัมพันธ์ระหว่างความยาวของแบบสอบทั้งฉบับ กับ จำนวนข้อสอบที่ทำหน้าที่ต่างกัน มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ของ Zumbo and Thomas อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 และ 2) ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน ขนาดของการทำหน้าที่ต่างกัน ปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน ระหว่างรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน กับ ขนาดของการทำหน้าที่ต่างกัน และ ระหว่าง จำนวนข้อสอบที่ทำหน้าที่ต่างกัน กับขนาดของการทำหน้าที่ต่างกัน มีผลต่ออัตราความถูกต้องของ วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ของ Jodoin and Gierl อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ยกเว้น ความยาวของแบบสอบทั้งฉบับ และ ปฏิสัมพันธ์ระหว่าง รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน กับ จำนวนข้อสอบที่ทำหน้าที่ต่างกัน ในขณะเดียวกันปฏิสัมพันธ์ระหว่างความยาวของแบบสอบทั้งฉบับ กับ จำนวนข้อสอบที่ทำหน้าที่ต่างกัน ไม่ มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ของ Jodoin and Gierl

สรุปผลการวิจัย

สรุปผลการวิจัย จำแนกตามวัตถุประสงค์การวิจัย ดังนี้

1. **ผลการเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการ ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ภายใต้เงื่อนไขเดียวกัน ของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย**

เมื่อนำผลการวิเคราะห์ความแปรปรวนพหุที่มีนัยสำคัญทางสถิติมาทำการทดสอบระหว่างกลุ่มเพื่อเปรียบเทียบ ประสิทธิภาพด้านอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 โดยเปรียบเทียบแยกทีละตัวแปรตาม ผลการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่มีรูปแบบการให้คะแนนแบบทวิวิภาค โดยการจำลองข้อมูล ในวิธีถดถอยโลจิสติก ระหว่างขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl และขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas **ภายใต้เงื่อนไขเดียวกัน** (การเปรียบเทียบ ค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบภายใต้ วิธีการที่ศึกษา โดยพิจารณาในแต่ละระดับของเงื่อนไขปัจจัยที่แปรเปลี่ยน) ของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน ขนาดของการทำหน้าที่ต่างกัน จำนวนข้อสอบที่ทำหน้าที่ต่างกัน และความยาวของแบบสอบทั้งฉบับ ผลสรุปการเปรียบเทียบตามวัตถุประสงค์การวิจัย ดังต่อไปนี้

1.1 วิธีการตรวจสอบ กับ ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน

การเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบภายใต้วิธีถดถอยโลจิสติก ภายใต้เงื่อนไขเดียวกัน พิจารณาเฉพาะแต่ละเงื่อนไขของรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน คือ เงื่อนไขแบบอเนกรูปและเงื่อนไขแบบอเนกรูป พบว่า เงื่อนไขปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันทั้งแบบอเนกรูปและแบบเอกรูป ภายใต้วิธีถดถอยโลจิสติก ระหว่างขนาดอิทธิพลทั้ง 2 เกณฑ์ มีอัตราความถูกต้องแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 โดยขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl มีค่าเฉลี่ยอัตราความถูกต้องสูงกว่า ขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่พบความแตกต่างอย่างมีนัยสำคัญ

1.2 วิธีการตรวจสอบ กับ ปัจจัยความยาวของแบบสอบทั้งฉบับ

การเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบภายใต้วิธีถดถอยโลจิสติก ภายใต้เงื่อนไขเดียวกัน พิจารณาเฉพาะแต่ละเงื่อนไขของปัจจัยความยาวของแบบสอบทั้งฉบับ 40 ข้อ และ 50 ข้อ พบว่า ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้องเงื่อนไขปัจจัยความยาวของแบบสอบทั้งฉบับ 40 ข้อ และ 50 ข้อ ภายใต้วิธีถดถอยโลจิสติกระหว่างขนาดอิทธิพลทั้ง 2 เกณฑ์ มีอัตราความถูกต้อง แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ผลการเปรียบเทียบขนาดความยาวของแบบสอบทั้ง 2 ฉบับ มีความสอดคล้องกันกล่าวคือ ภายใต้วิธีถดถอย

โลจิสติกโดยขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl มีค่าเฉลี่ยอัตราความถูกต้องสูงกว่า ขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่พบความแตกต่างอย่างมีนัยสำคัญ

1.3 วิธีการตรวจสอบ กับ ปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ

การเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบภายใต้วิธีถดถอยโลจิสติก ภายใต้เงื่อนไขเดียวกัน พิจารณา เฉพาะ แต่ละเงื่อนไขของ ขนาดของการทำหน้าที่ต่างกันของข้อสอบ 0.1, 0.2 และ 0.4 พบว่า ผลการเปรียบเทียบทุกเงื่อนไขปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ ภายใต้วิธีถดถอยโลจิสติก ระหว่าง การใช้เกณฑ์ขนาดอิทธิพล 2 วิธี มีอัตราความถูกต้องแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ผลการเปรียบเทียบมีความสอดคล้องกัน คือขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl มีค่าเฉลี่ยอัตราความถูกต้องสูงกว่า ขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่พบความแตกต่างอย่างมีนัยสำคัญ

1.4 วิธีการตรวจสอบ กับ ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน และ ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน

การเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบภายใต้วิธีถดถอยโลจิสติก ภายใต้เงื่อนไขเดียวกัน พิจารณาเฉพาะแต่ละเงื่อนไขของรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน และจำนวนข้อสอบที่ทำหน้าที่ต่างกัน พบว่า 1) รูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป ที่มีจำนวนข้อที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10 ระหว่างการ ใช้เกณฑ์ขนาดอิทธิพล 2 วิธี มีอัตราความถูกต้องแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 การวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl มีค่าเฉลี่ยอัตราความถูกต้องสูงกว่า เกณฑ์ Zumbo and Thomas 2) รูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป ที่มีจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 20 ระหว่างขนาดอิทธิพลทั้ง 2 เกณฑ์ มีอัตราความถูกต้องแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 การวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl มีค่าเฉลี่ยอัตราความถูกต้องสูงกว่า เกณฑ์ Zumbo and Thomas 3) รูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป ที่มีจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10 ระหว่างการ ใช้เกณฑ์ขนาดอิทธิพล 2 วิธี มีอัตราความถูกต้องแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ซึ่งการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl มีค่าเฉลี่ยอัตราความถูกต้องสูงกว่า เกณฑ์ Zumbo and Thomas 4) เงื่อนไข รูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป ที่มีจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 20 ระหว่างการใช้เกณฑ์ขนาดอิทธิพล 2 วิธี มีอัตราความถูกต้องไม่แตกต่างกันอย่างมีนัยสำคัญ

1. 5 วิธีการตรวจสอบ กับ ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน และ ปัจจัยความยาวของแบบสอบทั้งฉบับ

การเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบภายใต้ วิธีถดถอยโลจิสติก ภายใต้เงื่อนไขเดียวกัน พิจารณาเฉพาะแต่ละเงื่อนไขของ จำนวนข้อสอบที่ทำหน้าที่ต่างกัน และ ความยาวของแบบสอบทั้งฉบับ พบว่าค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้วิธีถดถอยโลจิสติก โดยพิจารณาเฉพาะความยาวของแบบสอบทั้งฉบับ 40 ข้อ และ 50 ข้อ กับจำนวนข้อสอบที่ทำหน้าที่ต่างกัน ที่มีจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับ คิดเป็นร้อยละ 10 และร้อยละ 20 พบว่าในทุกเงื่อนไขดังกล่าว มี อัตราความคลาดเคลื่อนประเภทที่ 1 ไม่แตกต่างกันอย่างมีนัยสำคัญ

1.6 วิธีการตรวจสอบ กับ ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน และ ปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ

การเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบภายใต้วิธีถดถอยโลจิสติก ภายใต้เงื่อนไขเดียวกัน พิจารณาเฉพาะแต่ละเงื่อนไขของรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน และ ขนาดของ การทำหน้าที่ต่างกันของข้อสอบ พบว่า ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้องภายใต้วิธีถดถอยโลจิสติก โดยพิจารณาเฉพาะ รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน (แบบอเนกรูปและเอกรูป) กับขนาดของการทำหน้าที่ต่างกันของข้อสอบ (3 ขนาด คือ 0.1, 0.2 และ 0.4) พบว่ามีเพียงกรณีที่ไม่พบความแตกต่าง คือ รูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป ที่มีขนาดของ การทำหน้าที่ต่างกันของข้อสอบขนาด 0.2 นอกนั้นทุกเงื่อนไข มี อัตราความถูกต้องแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ซึ่งทุกกรณีการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl มีค่าเฉลี่ยอัตราความถูกต้องสูงกว่าเกณฑ์ Zumbo and Thomas

1.7 วิธีการตรวจสอบ กับ ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน และปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ

การเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบภายใต้วิธีถดถอยโลจิสติก ภายใต้เงื่อนไขเดียวกัน พิจารณาเฉพาะแต่ละเงื่อนไขของรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน และ ขนาดของ การทำหน้าที่ต่างกันของข้อสอบ พบว่า ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้องภายใต้วิธีถดถอยโลจิสติก โดยพิจารณาเฉพาะจำนวนข้อสอบที่ทำหน้าที่ต่างกัน (ทั้งฉบับคิดเป็นร้อยละ 10 และร้อยละ 20) กับขนาดของ การทำหน้าที่ต่างกันของข้อสอบ (3 ขนาด คือ 0.1, 0.2 และ 0.4) พบว่าทุกกรณีมีอัตราความถูกต้องแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 โดยการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl มีค่าเฉลี่ยอัตราความถูกต้องสูงกว่าเกณฑ์ Zumbo and Thomas สอดคล้องกันทุกกรณี

2. ผลการเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ<u>ภายใต้เงื่อนไขต่างกัน</u> ของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย

เมื่อนำผลการวิเคราะห์ความแปรปรวนพหุที่มีนัยสำคัญทางสถิติมาทำการทดสอบระหว่างกลุ่มเพื่อเปรียบเทียบ ประสิทธิภาพด้านอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 โดยเปรียบเทียบแยกทีละตัวแปรตาม ผลการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่มีรูปแบบการให้คะแนนแบบทวิวิภาค โดยการจำลองข้อมูล ในวิธีถดถอยโลจิสติก ระหว่างขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl และขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas **<u>ภายใต้เงื่อนไขต่างกัน</u>** (การเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบภายใต้ปัจจัยที่ศึกษา โดยพิจารณาในแต่ละวิธีการตรวจสอบ) ของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน ขนาดของการทำหน้าที่ต่างกัน จำนวนข้อสอบที่ทำหน้าที่ต่างกัน และความยาวของแบบสอบทั้งฉบับ ได้ผลสรุปการ เปรียบเทียบ ตามวัตถุประสงค์การวิจัย ดังต่อไปนี้

2.1 วิธีการตรวจสอบ กับ ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน

การเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบตามปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน โดยพิจารณาเฉพาะทีละวิธีการตรวจสอบ พบว่าผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง วิธีถดถอยโลจิสติก โดยขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas ภายใต้รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน ทั้ง 2 ลักษณะมีอัตราความถูกต้องแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 โดยรูปแบบอเนกรูป มีค่าเฉลี่ยอัตราความถูกต้องสูงกว่ารูปแบบเอกรูป ส่วนวิธีถดถอยโลจิสติก โดยขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl ภายใต้รูปแบบของข้อสอบที่ทำหน้าที่ต่างกันทั้ง 2 ลักษณะ มีอัตราความถูกต้องแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 โดยรูปแบบอเนกรูป มีค่าเฉลี่ยอัตราความถูกต้องสูงกว่ารูปแบบเอกรูป ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่พบความแตกต่างอย่างมีนัยสำคัญ

2.2 วิธีการตรวจสอบ กับ ปัจจัยความยาวของแบบสอบทั้งฉบับ

การเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบตามปัจจัยความยาวของแบบสอบทั้งฉบับ โดยพิจารณา เฉพาะทีละวิธีการตรวจสอบ พบว่าผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง วิธีถดถอยโลจิสติกโดยขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas และขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl ให้ผลสอดคล้องกันคือภายใต้ ความยาวของแบบสอบทั้งฉบับ ทั้ง 2 ขนาด มีอัตราความถูกต้องไม่แตกต่างกันอย่างมีนัยสำคัญ ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่พบความแตกต่างอย่างมีนัยสำคัญ

2.3 วิธีการตรวจสอบ กับ ปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ

การเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบตามปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ โดยพิจารณาเฉพาะทีละวิธีการตรวจสอบ พบว่า ขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas ภายใต้ขนาดของการทำหน้าที่ต่างกันของข้อสอบ 3 ขนาด มีอัตราความถูกต้องแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 เมื่อนำผลการทดสอบที่มีนัยสำคัญทางสถิติไปเปรียบเทียบรายคู่ โดยใช้วิธีของเชฟเฟ (Scheffé) พบว่าในคู่ของ ขนาด 0.1 กับ ขนาด 0.4 และคู่ของขนาด 0.2 กับ ขนาด 0.4 มีค่าเฉลี่ยของอัตราความถูกต้องแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ส่วนการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl ภายใต้ขนาดของการทำหน้าที่ต่างกันของข้อสอบ 3 ขนาด มีอัตราความถูกต้องแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 เมื่อนำผลการทดสอบที่มีนัยสำคัญทางสถิติไปเปรียบเทียบรายคู่ โดยใช้วิธีของเชฟเฟ พบว่าในทุกคู่ของ ขนาด 0.1 ขนาด 0.2 และขนาด 0.4 การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้องมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่พบความแตกต่างอย่างมีนัยสำคัญ

2.4 วิธีการตรวจสอบ กับ ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน และ ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน

การเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบตามรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน และจำนวนข้อสอบที่ทำหน้าที่ต่างกัน โดยพิจารณาเฉพาะทีละวิธีการตรวจสอบ พบว่า ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง ระหว่างขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas และตามเกณฑ์ Jodoin and Gierl ภายใต้รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน กับปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 10 และ 20 มีอัตราความถูกต้องไม่แตกต่างกันอย่างมีนัยสำคัญ

2.5 วิธีการตรวจสอบ กับ ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน และ ปัจจัยความยาวของแบบสอบทั้งฉบับ

การเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบตามจำนวนข้อสอบที่ทำหน้าที่ต่างกัน และ ความยาวของแบบสอบทั้งฉบับ โดยพิจารณาเฉพาะทีละวิธีการตรวจสอบ พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas ภายใต้ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน ทั้ง 2 ขนาด ทั้งฉบับคิดเป็นร้อยละ 10 และร้อยละ 20 กับปัจจัยความยาวของแบบสอบทั้งฉบับ ทั้ง 2 ขนาด 40 และ 50 ข้อ มีอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 โดยจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10 ให้ค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 20 จำนวนข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 20 ให้ประสิทธิภาพที่ดีกว่า และความยาวของแบบสอบทั้งฉบับ 50 ข้อ ให้ค่าเฉลี่ย

ของอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่า ความยาวของแบบสอบทั้งฉบับ 40 ข้อ ส่วนการเปรียบเทียบในวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl ภายใต้ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน ทั้ง 2 ขนาด ทั้งฉบับคิดเป็นร้อยละ 10 และร้อยละ 20 กับปัจจัยความยาวของแบบสอบทั้งฉบับ ทั้งขนาด 40 และ 50 ข้อ มีอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่แตกต่างกันอย่างมีนัยสำคัญ

2.6 วิธีการตรวจสอบ กับ ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน และ ปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ

การเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบตามรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน และ ขนาดของการทำหน้าที่ต่างกันของข้อสอบ โดยพิจารณาเฉพาะทีละวิธีการตรวจสอบ พบว่า 1) ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas ภายใต้รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน แบบอเนกรูปและเอกรูป กับ ปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ 3 ขนาด คือ 0.1, 0.2 และ 0.4 มีอัตราความถูกต้อง แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 โดยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป ให้ค่าเฉลี่ยของอัตราความถูกต้องสูงกว่ารูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป และปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบที่มีขนาด 0.4 ให้ค่าเฉลี่ยของอัตราความถูกต้อง สูงกว่าขนาดของการทำหน้าที่ต่างกันของข้อสอบ ที่มีขนาด 0.1 และ 0.2 และ 2) วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl ภายใต้รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน แบบอเนกรูปและเอกรูป กับ ปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ 3 ขนาดคือ 0.1, 0.2 และ 0.4 มีอัตราความถูกต้อง แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 โดยรูปแบบอเนกรูปให้ค่าเฉลี่ยของอัตราความถูกต้องสูงกว่า รูปแบบเอกรูป และปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบที่มีขนาด 0.4 ให้ค่าเฉลี่ยของอัตราความถูกต้องสูงกว่าขนาดของการทำหน้าที่ต่างกันของข้อสอบที่มีขนาด 0.1

2.7 วิธีการตรวจสอบ กับ ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน และปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ

การเปรียบเทียบค่าเฉลี่ยอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบตามรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน และ ขนาดของการทำหน้าที่ต่างกันของข้อสอบ โดยพิจารณาเฉพาะทีละวิธีการตรวจสอบ พบว่า 1) วิธีถดถอยโลจิสติกโดยการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas ภายใต้จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10 และร้อยละ 20 กับ ปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ 3 ขนาด คือ 0.1, 0.2 และ 0.4 มีอัตราความถูกต้อง แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 โดยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 10 ให้ค่าเฉลี่ยของอัตราความถูกต้องสูงกว่าจำนวนข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 20 และปัจจัยขนาดของ การทำหน้าที่ต่างกันของข้อสอบ ที่มีขนาด 0.4 ให้ค่าเฉลี่ยของ อัตราความ

ถูกต้องสูงกว่าขนาดของการทำหน้าที่ต่างกันของข้อสอบ ที่มีขนาด 0.1 และ 0.2 และ 2) วิธีถดถอยโลจิสติกโดยการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl ภายใต้จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10 และร้อยละ 20 กับปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบ 3 ขนาดคือ 0.1, 0.2 และ 0.4 มีอัตราความถูกต้อง แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 โดยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 10 ให้ค่าเฉลี่ยของอัตราความถูกต้องสูงกว่าทั้งฉบับคิดเป็นร้อยละ 20 และปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบที่มีขนาด 0.4 ให้ค่าเฉลี่ยของ อัตราความถูกต้องสูงกว่าขนาด 0.1 และ 0.2

3. **ผลการเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่มีรูปแบบการให้คะแนนแบบทวิวิภาค โดยข้อมูลเชิงประจักษ์ ในวิธี ถดถอยโลจิสติก ระหว่างขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl และขนาดอิทธิพลตามเกณฑ์** Zumbo and Thomas

สรุป 3 ประเด็น ดังรายละเอียดต่อไปนี้

3.1 การตรวจสอบคุณภาพและสถิติเบื้องต้น

สถิติเบื้องต้นและคุณภาพของแบบสอบ**วิชาคณิตศาสตร์** ข้อสอบ 40 ข้อ คะแนนเต็ม 40 คะแนน ผลจากสถิติพื้นฐานบอกได้ว่าในภาพรวมผู้สอบส่วนใหญ่ได้คะแนนน้อยเพราะมีค่าเฉลี่ยของคะแนนต่ำมากเมื่อเทียบกับคะแนนเต็ม ซึ่ง ผลการสอบ ในภาพรวม ทั้งประเทศมีช่วงห่างของคะแนนสูงมาก คะแนนสูงสุด เป็น นักเรียนสังกัด โรงเรียนที่ตั้งในเขตอำเภอเมือง มีค่า สูงกว่าคะแนนสูงสุดของนักเรียนสังกัด โรงเรียนที่ตั้งนอกเขตอำเภอเมือง ข้อมูลชุดนี้มีความโด่งมากกว่าโค้งปกติจึงสรุปได้ว่าผู้สอบส่วนใหญ่ได้คะแนนน้อยและมีคนส่วนน้อยได้คะแนนสูง ค่าความเที่ยงแบบความสอดคล้องภายในคำนวณโดยสูตรสัมประสิทธิ์แอลฟาของครอนบาคเป็น 0.7887 เมื่อวิเคราะห์คุณภาพรายข้อ พบว่า วิชาคณิตศาสตร์ 40 ข้อ ผู้เข้าสอบ 123,167 คน มีค่าความยาก ระหว่าง 0.00 ถึง 0.19 และมีค่าอำนาจจำแนก ระหว่าง 0.00 ถึง 0.37 คุณภาพรายข้อตามทฤษฎีการตอบสนองข้อสอบ มีพารามิเตอร์ความยาก ระหว่าง 1.33 ถึง 4.58 พารามิเตอร์อำนาจจำแนก ระหว่าง 0.50 ถึง 3.49 สรุปได้ว่ามีอำนาจจำแนกที่ดีและข้อสอบมีความยากมาก แบบสอบ**วิชาวิทยาศาสตร์** ข้อสอบ 50 ข้อ คะแนนเต็ม 50 คะแนน ผลจากสถิติพื้นฐานบอกได้ว่าในภาพรวมผู้สอบส่วนใหญ่ได้คะแนนน้อยเพราะมีค่าเฉลี่ยของคะแนนต่ำมากเมื่อเทียบกับคะแนนเต็ม ซึ่งผลการสอบในภาพรวมทั้งประเทศมีช่วงห่างของคะแนนสูงมากเช่นเดียวกันกับวิชาคณิตศาสตร์ คะแนนสูงสุดของนักเรียนที่เข้าสอบตามสังกัดโรงเรียนที่ตั้งในเขตอำเภอเมือง เท่ากับคะแนนสูงสุดของนักเรียนที่เข้าสอบตามสังกัดโรงเรียนที่ตั้งนอกเขตอำเภอเมือง ข้อมูลชุดนี้มีความโด่งมากกว่าโค้งปกติ จึงสรุปได้ว่าผู้สอบส่วนใหญ่ได้คะแนนน้อยหรือมีคนส่วนน้อยได้คะแนนสูง ค่าความเที่ยง แบบความสอดคล้องภายในคำนวณโดยสูตรสัมประสิทธิ์แอลฟาของครอนบาค เป็น 0.8037 คุณภาพรายข้อของแบบสอบ วิชาวิทยาศาสตร์ 50 ข้อ ผู้เข้าสอบ 110,609 คน มีค่าความยาก ระหว่าง 0.18 ถึง 0.66 และมีค่าอำนาจจำแนก ระหว่าง -0.01 ถึง 0.66 คุณภาพรายข้อตามทฤษฎีการตอบสนอง

ข้อสอบ พารามิเตอร์ความยาก ระหว่าง -3.35 ถึง 4.66 และพารามิเตอร์อำนาจจำแนก ระหว่าง 0.10 ถึง 1.73 สรุปได้ว่ามีอำนาจจำแนกที่ดีและข้อสอบมีความยากมาก เมื่อตรวจสอบความเป็นเอกมิติของแบบสอบ ด้วยการวิเคราะห์องค์ประกอบ วิเคราะห์หาความสัมพันธ์ระหว่างข้อสอบทั้งหมดโดยใช้สัมประสิทธิ์สหสัมพันธ์ของเพียร์สันแล้วพิจารณาร้อยละของความแปรปรวน รวมถึงการพิจารณาค่าไอเกน (eigen value) ซึ่งเสนอโดย Lord และ Novick (1968) ถ้าค่าไอเกนตัวเดียวหรือหลายตัวแต่ตัวแรกมีค่ามากกว่าตัวอื่นๆ อย่างเห็นได้ชัด สามารถสรุปได้ว่าเครื่องมือชุดนั้นมีความเป็นมิติเดียว ผลการตรวจสอบความเป็นมิติเดียวของแบบสอบทั้งสองวิชาพบว่าทั้งวิชาคณิตศาสตร์และวิชาวิทยาศาสตร์มีความเป็นมิติเดียว ซึ่งเป็นไปตามข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ

3.2 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

เมื่อจำแนกคะแนนการสอบของนักเรียนทั้งหมดพบว่าส่วนใหญ่ได้คะแนนน้อย มีนักเรียนที่ได้คะแนน 0 คะแนน เป็นจำนวนมากถึง 56,604 คน คิดเป็นร้อยละ 45.96 จากจำนวนผู้เข้าสอบทั้งหมด ดังนั้น ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ เนื่องจากมีข้อสังเกตเกี่ยวกับผลการตอบข้อสอบของนักเรียนส่วนใหญ่ที่ไม่มีคะแนนจากการสอบ การนำผลการตอบข้อสอบลักษณะดังกล่าวไปคำนวณการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ อาจให้ผลลัพธ์ที่คลาดเคลื่อนไปจากความเป็นจริง ผู้วิจัยจึงตัดกรณีที่ผู้เข้าสอบได้คะแนนการสอบ 0 คะแนนออก แล้วนำผลการตอบที่เหลือมาคำนวณผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์

3.2.1 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ จำแนกตามวิธีที่ศึกษา

ผลการตรวจสอบโดย วิธีแมนเทล -แฮนส์เซล วิชาคณิตศาสตร์ มีข้อมูลที่นำมาใช้วิเคราะห์จริงจำนวน 66,563 คน 40 ข้อ พบข้อสอบทำหน้าที่ต่างกัน 32 ข้อ และวิชาวิทยาศาสตร์ 50 ข้อ มีข้อมูลที่นำมาใช้วิเคราะห์จริงจำนวน 110,555 คน พบข้อที่ทำหน้าที่ต่างกันจำนวน 35 ข้อ

ผลการตรวจสอบโดย วิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญวิชาคณิตศาสตร์ ข้อสอบ 40 ข้อ พบข้อสอบทำหน้าที่ต่างกัน 36 ข้อ เป็นแบบเอกรูป 13 ข้อ แบบอเนกรูป 23 ข้อ ส่วนข้อสอบที่ทำหน้าที่ไม่ต่างกันมี 4 ข้อ วิชาวิทยาศาสตร์ ข้อสอบ 50 ข้อ พบข้อสอบทำหน้าที่ต่างกัน 47 ข้อ เป็นแบบเอกรูป 9 ข้อ แบบอเนกรูป 38 ข้อ ส่วนข้อสอบที่ทำหน้าที่ไม่ต่างกันมี 3 ข้อ

ผลการตรวจสอบโดยการวัด ขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas วิชาคณิตศาสตร์ ข้อสอบ 40 ข้อ พบว่าข้อสอบที่มีขนาดอิทธิพลของการทำหน้าที่ต่างกันขนาดเล็กน้อยจนแทบไม่มีเลย ($0.00< R^2 <0.13$) มี 38 ข้อ ข้อสอบที่มีขนาดอิทธิพลของการทำหน้าที่ต่างกันขนาดปานกลาง ($0.13< R^2 <0.26$) มี 1 ข้อ ข้อสอบที่มีขนาดอิทธิพลของการทำหน้าที่ต่างกันขนาดใหญ่ ($R^2<0.26$) มี 1 ข้อ วิชาวิทยาศาสตร์ ข้อสอบ 50 ข้อ พบว่าทุกข้อ มีขนาดอิทธิพลของการ ทำหน้าที่ต่างกันขนาดเล็กน้อยจนแทบจะไม่มีเลย ($0.00< R^2 <0.13$) ทั้ง 50 ข้อ

ตรวจสอบโดยการวัด ขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl วิชาคณิตศาสตร์ ข้อสอบ 40 ข้อ พบว่าข้อสอบที่มีขนาดอิทธิพลของการทำหน้าที่ต่างกันขนาดเล็กน้อยจนแทบจะไม่มีเลย

(*0.00< R^2 <0.035*) มี 36 ข้อ ข้อสอบที่มีขนาดอิทธิพลของการทำหน้าที่ต่างกันขนาดปานกลาง (*0.351< R^2 <0.07*) มี 2 ข้อ ข้อสอบที่มีขนาดอิทธิพลของการทำหน้าที่ต่างกันขนาดใหญ่ (*0.071 < R^2*) มีจำนวน 2 ข้อ วิชาวิทยาศาสตร์ ข้อสอบ 50 ข้อ พบว่าข้อสอบที่ทำหน้าที่ต่างกันขนาดเล็กน้อยจนแทบไม่มีเลย (*.00< R^2 <.035*) มีจำนวน 48 ข้อ ข้อสอบที่ทำหน้าที่ต่างกันขนาดใหญ่ (*.071 < R^2*) มี 2 ข้อ

สรุปผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจากวิธีที่ศึกษาเทียบกับวิธีเกณฑ์ **วิชาคณิตศาสตร์** การวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl พบข้อสอบ ทำหน้าที่ต่างกัน ที่มีขนาดปานกลางขึ้นไปจำนวน 4 ข้อ และตรงกับผลการตรวจด้วยวิธีเกณฑ์ทุกข้อ ขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas ไม่พบข้อสอบที่ทำหน้าที่ต่างกัน 1) ผลการตรวจสอบระหว่างวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตาม เกณฑ์ Zumbo and Thomas กับวิธีเกณฑ์ พบข้อร่วมของการทำหน้าที่ต่างกันระหว่างสองวิธีดังกล่าว จำนวน 2 ข้อ จาก 40 ข้อ คิดเป็นร้อยละ 5.00 2) ผลการตรวจสอบระหว่างวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตาม เกณฑ์ Jodoin and Gierl กับวิธีเกณฑ์ พบข้อร่วมของการทำหน้าที่ต่างกันระหว่างสองวิธีดังกล่าว จำนวน 4 ข้อ จาก 40 ข้อ คิดเป็นร้อยละ 10.00

สรุปผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจากวิธีที่ศึกษาเทียบกับวิธีเกณฑ์ **วิชาวิทยาศาสตร์** การวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl พบข้อสอบทำหน้าที่ต่างกันที่มีขนาดปานกลางขึ้นไปจำนวน 2 ข้อ ในขณะที่การวัดขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas ไม่พบข้อร่วมของการทำหน้าที่ต่างกันระหว่างสองวิธีดังกล่าว 1) ผลการตรวจสอบระหว่างวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตาม เกณฑ์ Jodoin and Gierl กับวิธีเกณฑ์ พบข้อร่วมของการทำหน้าที่ต่างกันระหว่างสองวิธีดังกล่าว จำนวน 2 ข้อ จาก 50 ข้อ คิดเป็นร้อยละ 4.00

3.2.2 ประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

วิชาคณิตศาสตร์ มีร้อยละของอัตราความถูกต้องของ การวัดขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas เป็นร้อยละ 6.25 และอัตราความถูกต้อ งของขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl เป็นร้อยละ 12.5 ร้อยละของอัตราความคลาดเคลื่อนประเภทที่ 1 ของการวัด ขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas และตามเกณฑ์ Jodoin and Gierl มีค่าเท่ากันเป็น ร้อยละ 0.00 สามารถตัดสินได้ว่า วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl มีประสิทธิภาพดีกว่าขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas เนื่องจาก มีร้อยละของอัตราความถูกต้องในการตรวจสอบสูงกว่า ในขณะที่อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าเท่ากัน

วิชาวิทยาศาสตร์ มีร้อยละของอัตราความถูกต้องของการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas เป็นร้อยละ 0.00 และอัตราความถูกต้อ งของขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl เป็นร้อยละ 5.71 ร้อยละของอัตราความคลาดเคลื่อนประเภทที่ 1 ของการวัด ขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas และตามเกณฑ์ Jodoin and Gierl มีค่าเท่ากันเป็น ร้อยละ 0.00 สามารถตัดสินได้ว่า วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl มี

ประสิทธิภาพดีกว่าขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas เนื่องจาก มีร้อยละของอัตราความถูกต้องในการตรวจสอบสูงกว่า ในขณะที่อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าเท่ากัน

การศึกษาในข้อมูลเชิงประจักษ์จากแบบสอบทั้ง 2 ฉบับ ให้ผลที่สอดคล้องกัน คือ วิธีถดถอยโลจิสติก โดยขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl มีประสิทธิภาพดีกว่าขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas

3.3 ความสอดคล้องของการศึกษากรณีข้อมูลเชิงประจักษ์กับข้อมูลจำลอง

การวิจัยครั้งนี้ เน้นศึกษาประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้วิธีถดถอยโลจิสติก ผลจากแบบสอบวิชาคณิตศาสตร์ พบว่า ประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl ให้ประสิทธิภาพการตรวจสอบที่ดีที่สุด

ผู้วิจัยนำผลการตรวจสอบขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl ในวิชาคณิตศาสตร์มานำเสนอเพื่อให้เกิดสารสนเทศอันนำไปสู่การตัดสินใจเพิ่มเติม พบว่า ผลการวัดระดับนัยสำคัญพบข้อที่ทำหน้าที่ต่างกัน 36 ข้อ ถ้ายึดสารสนเทศนี้เป็นหลัก ทำให้ต้องตัดข้อสอบออกจากแบบสอบเป็นจำนวนมาก สารสนเทศดังกล่าวบอกให้ทราบเพียงว่าข้อสอบเกิดการทำหน้าที่ต่างกันแต่ไม่รู้สารสนเทศอื่นเลย นั่นก็ถือว่ายังไม่เพียงพอสำหรับการตัดสินใจว่าจะตัดข้อสอบออกจากแบบสอบหรือคงข้อสอบข้อนั้นไว้ในแบบสอบ ส่วนการวัด ขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl พบข้อที่ทำหน้าที่ต่างกันที่มีขนาดอิทธิพลของการทำหน้าที่ต่างกันขนาดปานกลาง จำนวน 2 ข้อ พบข้อที่ทำหน้าที่ต่างกันที่มีขนาดอิทธิพลของการทำหน้าที่ต่างกันขนาดใหญ่จำนวน 2 ข้อ ซึ่งทั้ง 4 ข้อ ให้ผลตรงกับผลการวัดระดับนัยสำคัญว่าข้อสอบทำหน้าที่ต่างกัน หากพิจารณาผลการตรวจสอบด้วย วิธีถดถอยโลจิสติก โดยการวัดระดับนัยสำคัญ ร่วมกับผลการตรวจสอบด้วยการวัดขนาดอิทธิพลที่มีอิทธิพลระดับขนาดปานกลางขึ้นไป พบว่าต้องพิจารณาดำเนินการปรับปรุงข้อสอบแล้วตรวจสอบคุณภาพใหม่จากแบบสอบเพียง 4 ข้อ จากผลการศึกษาดังกล่าวจึงเกิดผลดีในด้านการบริหารจัดการข้อสอบ ประหยัดเวลา แรงงาน และทุนทรัพย์ ดังนั้นจึงควรสนับสนุนให้พิจารณาขนาดอิทธิพลของการทำหน้าที่ต่างกันของข้อสอบร่วมกับการตรวจสอบโดยวิธีการหลักที่เลือกศึกษา เพื่อเกิดความรอบคอบต่อการตัดสินข้อสอบและแบบสอบ

อภิปรายผลการวิจัย

การวิจัยครั้งนี้ศึกษาการวัดขนาดอิทธิพลและผลของประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในวิธีถดถอยโลจิสติก สำหรับข้อสอบที่มีรูปแบบการตรวจให้คะแนนแบบทวิวิภาค กรณีข้อมูลจำลองและข้อมูลเชิงประจักษ์ อภิปรายใน3 ประเด็น ดังนี้

1. **ประสิทธิภาพ การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ภายใต้ วิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพลตามเกณฑ์** Jodoin and Gierl **และ** Zumbo and Thomas

การศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ภายใต้วิธีถดถอยโลจิสติก ซึ่งเป็นวิธีตรวจสอบที่ถูกออกแบบมาสำหรับตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันทั้งแบบเอกรูปและแบบอเนกรูป การตรวจสอบด้วยวิธีนี้มีแนวโน้มว่าความคลาดเคลื่อนประเภทที่ 1 จะเพิ่มขึ้น ผลที่เกิดขึ้นนี้ อาจมีส่วนในการแสดงผลที่คลาดเคลื่อนเกี่ยวกับการศึกษาการทำหน้าที่ต่างกันของข้อสอบ นำไปสู่การใช้ทรัพยากรข้อสอบที่ไม่มีประสิทธิภาพ เมื่อการวัดขนาดอิทธิพลถูกพัฒนาขึ้นจากวิธีถดถอยโลจิสติก การตัดสินขนาดของอิทธิพลมีเกณฑ์ที่ใช้สถิติ R^2 สำหรับตัดสิน 2 เกณฑ์ คือ เกณฑ์ของ Zumbo และ Thomas (1997) และเกณฑ์ของ Jodoin และ Gierl (2001) โดย Zumbo และ Thomas เสนอการจัดขนาดอิทธิพลเป็น 3 ระดับ คือ ข้อสอบที่ทำหน้าที่ต่างกันขนาดเล็กน้อย มีค่าความแตกต่าง $\Delta R^2 < 0.13$ ข้อสอบที่ทำหน้าที่ต่างกัน ขนาดปานกลาง มีค่าความแตกต่าง $0.13 \leq \Delta R^2 \leq 0.26$ และข้อสอบที่ทำหน้าที่ต่างกันขนาดใหญ่ มีค่าความแตกต่าง $\Delta R^2 > 0.26$ โดยทั้งข้อสอบที่ทำหน้าที่ต่างกันขนาดปานกลางและขนาดใหญ่จำเป็นต้องให้สถิติ G^2 มีนัยสำคัญทางสถิติ

Jodoin และ Gierl เสนอการจัดขนาดอิทธิพลเป็น 3 ระดับ เช่นเดียวกับ Zumbo และ Thomas แต่มีรายละเอียดที่แตกต่าง คือข้อสอบที่ทำหน้าที่ต่างกัน ขนาดเล็กน้อย มีค่าความแตกต่าง $\Delta R^2 < 0.035$ ข้อสอบที่ทำหน้าที่ต่างกันขนาดปานกลาง มีค่าความแตกต่าง $0.035 \leq \Delta R^2 \leq 0.07$ และข้อสอบที่ทำหน้าที่ต่างกัน ขนาดใหญ่ มีค่าความแตกต่าง $\Delta R^2 > 0.07$ ผู้วิจัยได้นำเกณฑ์ตัดสินขนาดอิทธิพล 2 เกณฑ์ ดังกล่าวมาศึกษาภายใต้สถานการจำลอง ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน ขนาดของการทำหน้าที่ต่างกัน จำนวนข้อสอบที่ทำหน้าที่ต่างกัน และ ความยาวของแบบสอบทั้งฉบับและศึกษาปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ผลพบว่า

1.1 ปัจจัย รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน มีผลต่ออัตราความถูกต้องของ วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas และเกณฑ์ Jodoin and Gierl อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ผลของอัตราความถูกต้องของ การวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl มีค่าสูงกว่าอัตราความถูกต้องของ การวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas ทั้งข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปและแบบเอกรูป พบว่า ผลของอัตราความถูกต้องของการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl มีค่าสูงกว่าอัตราความถูกต้องของการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas ผลดังกล่าวนี้เป็นไปตามสมมติฐานข้อที่ 1 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่พบความแตกต่างอย่างมีนัยสำคัญทางสถิติ

1.2 ปัจจัย ขนาดของการทำหน้าที่ต่างกัน มีผลต่ออัตราความถูกต้องของ วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas และเกณฑ์ Jodoin and Gierl อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 โดยผลของอัตราความถูกต้องของ การวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl มีค่าสูงกว่าอัตราความถูกต้องของ การวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and

Thomas ทุกขนาดของการทำหน้าที่ต่างกัน ได้แก่ 0.1, 0.2 และ 0.4 พบว่าผลของอัตราความถูกต้องของการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl มีค่าสูงกว่าอัตราความถูกต้องของ การวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas ผลดังกล่าวนี้เป็นไปตามสมมติฐานข้อที่ 1 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่พบความแตกต่างอย่างมีนัยสำคัญทางสถิติ

1.3 ปัจจัยความยาวของแบบสอบทั้งฉบับ จากการศึกษาครั้งนี้พบว่าไม่มีผลต่ออัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ทั้งนี้อาจจะเป็นเพราะกำหนดความยาวของแบบสอบใกล้เคียงกันมากระหว่าง 40 กับ 50 ข้อ จึงไม่มีผลของความแตกต่างอย่างมีนัยสำคัญทางสถิติ

1.4 ปฏิสัมพันธ์สองทาง ของปัจจัยที่แปรเปลี่ยน ระหว่าง รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน กับ ขนาดของการทำหน้าที่ต่างกัน มีผลต่ออัตราความถูกต้องของวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas และเกณฑ์ Jodoin and Gierl อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ผลของอัตราความถูกต้องของการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl มีค่าสูงกว่าอัตราความถูกต้องของ การวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas โดยปฏิสัมพันธ์สองทางระหว่างระหว่างรูปแบบอเนกรูปกับทุกขนาดของการทำหน้าที่ต่างกัน ได้แก่ 0.1, 0.2 และ 0.4 พบว่าผลของอัตราความถูกต้องของการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl มีค่าสูงกว่าอัตราความถูกต้องของการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่พบความแตกต่างอย่างมีนัยสำคัญทางสถิติ ส่วนปฏิสัมพันธ์สองทางระหว่าง ระหว่างรูปแบบเอกรูป กับ ขนาดของการทำหน้าที่ต่างกัน 0.4 เท่านั้น ที่พบว่า ผลของอัตราความถูกต้องของ การวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl มีค่าสูงกว่าอัตราความถูกต้องของการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas ผลดังกล่าวนี้เป็นไปตามสมมติฐานข้อที่ 1 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่พบความแตกต่างอย่างมีนัยสำคัญทางสถิติ

1.5 ปฏิสัมพันธ์สองทางของปัจจัยที่แปรเปลี่ยนระหว่าง จำนวนข้อสอบที่ทำหน้าที่ต่างกัน กับ ขนาดของการทำหน้าที่ต่างกัน มีผลต่ออัตราความถูกต้องของ วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas และเกณฑ์ Jodoin and Gierl อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ผลของอัตราความถูกต้องของการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl มีค่าสูงกว่าอัตราความถูกต้องของการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas โดยปฏิสัมพันธ์สองทางระหว่างทุกเงื่อนไขของจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10 กับทุกขนาดของการทำหน้าที่ต่างกัน ได้แก่ 0.1, 0.2 และ 0.4 และ ทุกเงื่อนไขของ จำนวนข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 20 กับทุกขนาดของการทำหน้าที่ต่างกัน ได้แก่ 0.1, 0.2 และ 0.4 พบว่า ผลของอัตราความถูกต้องของการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl มีค่าสูงกว่าอัตราความถูกต้องของการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas ผลดังกล่าวนี้เป็นไปตามสมมติฐานข้อที่ 1 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่พบความแตกต่างอย่างมีนัยสำคัญทางสถิติ

1.6 ปฏิสัมพันธ์สองทาง ของปัจจัยที่แปรเปลี่ยน ระหว่าง รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน กับ จำนวนข้อสอบที่ทำหน้าที่ต่างกัน มีผลต่ออัตราความถูกต้องของวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas และเกณฑ์ Jodoin and Gierl อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ผลของอัตราความถูกต้องของการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl มีค่าสูงกว่าอัตราความถูกต้องของการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas โดยปฏิสัมพันธ์สองทางระหว่างทุกเงื่อนไขของรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน แบบอเนกรูป กับ จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10 และร้อยละ 20 และทุกเงื่อนไขของรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป กับจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10 และร้อยละ 20 พบว่า ผลของอัตราความถูกต้องของการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl มีค่าสูงกว่าอัตราความถูกต้องของการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas ผลดังกล่าวนี้เป็นไปตามสมมติฐานข้อที่ 1 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่พบความแตกต่างอย่างมีนัยสำคัญทางสถิติ

ผลการศึกษาเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ในวิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl และการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas ข้างต้น พบว่า การวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl ให้อัตราความถูกต้องสูงกว่า การวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas ในเกือบทุกเงื่อนไข ที่เป็นเช่นนี้อาจเนื่องมาจากเกณฑ์ของ Zumbo and Thomas ไม่ได้มีความละเอียดอ่อนมากนัก ช่วงการแบ่งขนาดมีความกว้างกว่าเกณฑ์ของ Jodoin and Gierl (2001) มีความ สอดคล้องกับแนวทาง Zieky (1993) โดยสถาบันบริการทดสอบทางการศึกษา (Educational Testing Service: ETS) ที่แบ่งขนาดอิทธิพลในการประเมินการทำหน้าที่ต่างกันของข้อสอบโดยค่าเฉลี่ยตามวิธีตรวจสอบของ Mantel-Haenszel จึงให้ความถูกต้องที่มากกว่า สอดคล้องกับการศึกษาของGómez-Benito และคณะ (2009) ที่ศึกษาประสิทธิภาพของขนาดอิทธิพลในการพัฒนาวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก ศึกษาประสิทธิภาพของขนาดอิทธิพลในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ พบว่าขนาดอิทธิพลโดยสถิติ R^2 ให้ผลของอำนาจการทดสอบต่ำ กว่าผลจาการ ทดสอบระดับนัยสำคัญ ผลการวิจัยสนับสนุนให้ศึกษาขนาดอิทธิพลโดยสถิติ R^2 ร่วมกับการทดสอบระดับนัยสำคัญทางสถิติจะทำให้ได้สารสนเทศมากยิ่งขึ้น ดังนั้น ในการตรวจสอบการทำหน้าที่ ต่างกันของข้อสอบ ที่วัดความสามารถมิติเดียว และให้คะแนน แบบสอง ค่าภายใต้วิธีถดถอยโลจิสติก จึงมีประสิทธิภาพในการตรวจสอบ การทำหน้าที่ต่างกันของข้อสอบ ที่มีขนาดอิทธิพลทั้งขนาดเล็ก ขนาดกลาง และขนาดใหญ่ ได้อย่างมีประสิทธิภาพ

2. ประสิทธิภาพ การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ของ วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลของการทำหน้าที่ต่างกัน ภายใต้ปัจจัยที่ศึกษา

2.1 รูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปมีอัตราความถูกต้องสูงกว่า รูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป และรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป เมื่อวัด

ขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl ให้ค่าอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำ กว่าการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas ผลดังกล่าวมีนัยสำคัญเป็นไปตามสมมติฐานข้อที่ 2 เมื่อพิจารณาเทียบอัตราความถูกต้องระหว่างผลการตรวจสอบจากการทดสอบระดับนัยสำคัญ กับ ผลการตรวจสอบจากการวัดขนาดอิทธิพลตามเกณฑ์ทั้ง 2 เกณฑ์ พบว่า เกณฑ์ Jodoin and Gierl ให้ค่าอัตราความถูกต้องสูงกว่าเกณฑ์ Zumbo and Thomas ผลของอัตราความคลาดเคลื่อนประเภทที่1 ระหว่างผลการตรวจสอบจากการทดสอบระดับนัยสำคัญ กับ ผลการตรวจสอบจากการวัดขนาดอิทธิพลตามเกณฑ์ทั้ง 2 เกณฑ์ พบว่า รูปแบบการทำหน้าที่ต่างกันแบบอเนกรูปตามเกณฑ์ Jodoin and Gierl ให้ค่าอัตราความคลาดเคลื่อนประเภทที่1 ต่ำกว่าเกณฑ์ Zumbo and Thomas แต่รูปแบบการทำหน้าที่ต่างกันแบบเอกรูปตามเกณฑ์ Jodoin and Gierl ให้ค่าอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าเกณฑ์ Zumbo and Thomas

งานวิจัยที่ผ่านมามีการศึกษาเกี่ยวกับรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันมากพอสมควรเช่น วลีมาศ แซ่อึ้ง(2543) พบว่าวิธีถดถอยโลจิสติกสามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้งแบบอเนกรูปและแบบเอกรูปได้อย่างมีประสิทธิภาพ Swaminathan and Rogers (1990) พบว่าวิธีการถดถอยโลจิสติกสามารถใช้โมเดลทดสอบผลของปฏิสัมพันธ์ระหว่างระดับความสามารถกับการเป็นสมาชิกของกลุ่มผู้สอบทำให้สามารถตรวจสอบได้ทั้งข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูปและแบบอเนกรูป Rogers and Swaminathan (1993) ที่เปรียบเทียบการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีถดถอยโลจิสติก กับ วิธีแมนเทล-แฮนส์เซล แล้วพบว่าการตรวจสอบการทำหน้าที่ต่างกันแบบเอกรูปในวิธีถดถอยโลจิสติกตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูปได้ดีในกรณีที่ข้อสอบมีความยากปานกลางและ เมื่ออำนาจจำแนกสูง จะตรวจสอบข้อสอบที่มีความยากปานกลางได้น้อยมาก แต่สามารถตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปได้ดีในกรณีที่ข้อสอบง่ายหรือยากมาก French and Miller (1996) ศึกษาความเป็นไปได้ของการใช้วิธีถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบพหุวิภาค แล้วพบว่า เมื่อกลุ่มตัวอย่างขนาดเล็กลง อำนาจในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจะลดลงและเมื่อพารามิเตอร์อำนาจจำแนกของข้อสอบยิ่งแตกต่างกันมากอำนาจในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปยิ่งเพิ่มขึ้น การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่วัดความสามารถมิติเดียวและให้คะแนนแบบสองค่าภายใต้วิธีถดถอยโลจิสติก จึงมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้งแบบอเนกรูปและแบบเอกรูปได้อย่างมีประสิทธิภาพตามปัจจัยที่เกี่ยวข้องอื่นด้วย เมื่อวิธีถดถอยโลจิสติกสามารถตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันในรูปแบบเอกรูปและอเนกรูปได้อย่างมีประสิทธิภาพ ดังนั้น ภายใต้รูปแบบการทำหน้าที่ต่างกัน ในวิธีถดถอยโลจิสติก โดย การวัดขนาดอิทธิพล จึงมีความเหมาะสมในการนำผลการตรวจสอบไปตัดสินใจร่วมกับการตรวจสอบประสิทธิภาพด้วย วิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ

2.2 ขนาดของการทำหน้าที่ต่างกัน มีผลต่อประสิทธิภาพการตรวจสอบภายใต้วิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl และการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas ผลดังกล่าวมีนัยสำคัญเป็นไปตามสมมติฐานข้อ 2 โดยพบว่าเมื่อขนาดของการทำหน้าที่ต่างกันมีขนาดใหญ่อัตราความถูกต้องจะมีค่าสูง เมื่อขนาดของการทำหน้าที่ต่างกันมีขนาดเล็กอัตราความคลาดเคลื่อนประเภทที่ 1 จะมีค่าต่ำ ถือว่าเป็นผลดีต่อการตรวจสอบประสิทธิภาพ

2.3 จำนวนข้อสอบที่ทำหน้าที่ต่างกัน มีผลต่อประสิทธิภาพการตรวจสอบภายใต้วิธีถดถอยโลจิสติก ระหว่างขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl และขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ผลดังกล่าวมีนัยสำคัญเป็นไปตามสมมติฐานข้อ 2 เมื่อจำนวนข้อสอบที่ทำหน้าที่ต่างกันเพิ่มขึ้น 10% (จาก 10% ถึง 20%) มีผลทำให้อัตราความถูกต้องลดลงในช่วง 3% ถึง 5% และมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นในช่วง 0.1% ถึง 2% ผลการศึกษาดังกล่าว Narayanan and Swaminathan (1996) ที่พบว่าเมื่อสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันเพิ่มขึ้นจาก 10% ถึง 20% มีผลทำให้วิธีถดถอยโลจิสติกมีอัตราความถูกต้องลดลงและมีอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้น Oshima and Miller (1992) พบว่าในข้อสอบวัดความสามารถหลายมิติเมื่อสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันเพิ่มขึ้น มีผลทำให้วิธีการวัดพื้นที่แบบคิดเครื่องหมาย วิธีการวัดพื้นที่แบบไม่คิดเครื่องหมาย วิธีผลรวมของกำลังสองแบบคิดเครื่องหมาย ผลรวมของกำลังสองแบบไม่คิดเครื่องหมาย มีอัตราความถูกต้องลดลง ที่เป็นเช่นนี้ สอดคล้องกับ Narayanan and Swaminathan (1994) ที่กล่าวว่าผลดังกล่าวอาจเป็นเพราะ เมื่อสัดส่วนของข้อสอบ ที่ทำหน้าที่ต่างกันเพิ่มขึ้นมีผลทำให้อัตราความถูกต้องของวิธีถดถอยโลจิสติกลดลง มีสาเหตุมาจากปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน ทั้ง 2 ขนาด ทั้งฉบับคิดเป็นร้อยละ 10 และร้อยละ 20 ที่เพิ่มขึ้น จึงมีผลทำให้ค่าประมาณความสามารถมีความเชื่อมั่นต่ำลงซึ่งจะมีผลทำให้เกณฑ์การจับคู่ความสามารถที่ใช้ในการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันขาดความแม่นยำ จึงมีผลทำให้อัตราความถูกต้องของวิธีการตรวจสอบลดลง ดังนั้น ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่วัดความสามารถมิติเดียว และให้คะแนนแบบสองค่าภายใต้วิธีถดถอยโลจิสติก ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 10 และ 20 จึงมีผลต่อประสิทธิภาพของวิธีถดถอยโลจิสติก

2.4 ความยาวของแบบสอบทั้งฉบับ มีผลต่อประสิทธิภาพการตรวจสอบ ภายใต้วิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl และเกณฑ์ Zumbo and Thomas อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ผลดังกล่าวมีนัยสำคัญเป็นไปตามสมมติฐานข้อ 2 อัตราความถูกต้องของการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl สอดคล้องกับผลการทดสอบระดับนัยสำคัญกล่าวคืออัตราความถูกต้องมีค่าสูงขึ้นเมื่อจำนวนข้อสอบเพิ่มขึ้นและเมื่อจำนวนข้อเพิ่มขึ้นทั้ง 2 วิธีดังกล่าวก็ให้ผลอัตราความคลาดเคลื่อนประเภทที่ 1 ลดลงสอดคล้องกันด้วย สาเหตุที่เลือกแบบสอบที่มีจำนวน 40 และ 50 ข้อ เนื่องจากสอดคล้องกับผลการศึกษาของ Narayanan and Swaminathan

(1994,1996) ที่พบว่าการจัดกระทำกับข้อมูลในด้านความยาวของเครื่องมืออาจไม่ต้องกำหนดเงื่อนไขที่หลากหลาย เนื่องจากที่ระดับความยาว 40 ข้อนั้นแม้จะเป็นตัวแทนของการทดสอบผลสัมฤทธิ์ทางการเรียนสั้นๆ แต่มีความน่าเชื่อถือที่ได้มาตรฐานและสอดคล้องกับผลการศึกษาของ จิติมา วรรณศรี (2539) ญาณภัทร สีหะมงคล (2540) ปิยะทิพย์ ตินวร (2549) และ Kim and Cohen (1998) ที่พบว่าข้อสอบที่มีความยาวปานกลางขึ้นไปจะส่งผลต่อประสิทธิภาพในการตรวจสอบมากที่สุด อีกทั้งเป็นระดับความยาวที่เหมาะสมกับการนำไปใช้เก็บข้อมูลจริง ผลดังกล่าวมีนัยสำคัญเป็นไปตามสมมติฐานข้อ 1-2 ซึ่งจะเห็นว่าความยาวของแบบสอบทั้งฉบับ ที่เลือกศึกษา ในครั้งนี้คือจำนวน 40 และ 50 ข้อ มีความเหมาะสมสอดคล้องกับ Swaminathan and Rogers (1990) ที่ว่าในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและแบบอเนกรูป เมื่อใช้แบบสอบที่มีความยาวมากขึ้น มีผลทำให้อำนาจการทดสอบวิธีแมนเทล-แฮนส์เซล และอำนาจการทดสอบของวิธีถดถอยโลจิสติกมีค่ามากขึ้น Rogers and Swaminathan (1993) พบว่าความยาวของแบบสอบไม่มีผลต่ออำนาจการทดสอบของวิธีแมนเทล -แฮนส์เซลและวิธีถดถอยโลจิสติกยกเว้นในกรณีแบบอเนกรูปของวิธีถดถอยโลจิสติก

3. **ประสิทธิภาพ การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ภายใต้ วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl และ Zumbo and Thomas ในข้อมูลเชิงประจักษ์**

ประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ภายใต้ วิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl และเกณฑ์ Zumbo and Thomas มีผลการตรวจสอบจากวิธีแมนเทล-แฮนส์เซลเป็นเกณฑ์สำหรับเปรียบเทียบประสิทธิภาพกับผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบตามวิธีที่ศึกษา ผลการศึกษาในข้อมูลเชิงประจักษ์ในภาพรวมจากแบบสอบวิชาคณิตศาสตร์และวิชาวิทยาศาสตร์ให้ผลที่สอดคล้องกัน

การตรวจสอบประสิทธิภาพด้านอัตราความถูกต้องโดยการวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl ให้ผลประสิทธิภาพที่ดีกว่าการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas โดยมีความถูกต้องสูงเมื่อแบบสอบมีจำนวนข้อสอบมากขึ้น อัตราความคลาดเคลื่อนประเภทที่ 1 ก็จะลดน้อยลง พิจารณาในแต่ละประเด็นย่อย ดังนี้

3.1 ด้านคุณภาพของข้อสอบ ค่าความเที่ยง แบบความสอดคล้องภายในโดยสูตรสัมประสิทธิ์แอลฟาของครอนบาคในแบบสอบวิชาคณิตศาสตร์ มีค่า 0.7887 และวิชาวิทยาศาสตร์ มีค่า 0.8037 จัดว่า มีค่าคุณภาพด้าน ความเที่ยงในระดับสูง คะแนนดิบของแบบสอบวิชาคณิตศาสตร์ในภาพรวมของประเทศคะแนนผลการสอบมีช่วงห่างของคะแนนสูงมาก โดยรวมแล้วเด็กนักเรียนเก่งของทั้งสองสังกัดมีความสามารถวิชาคณิตศาสตร์ไม่แตกต่างกันมากนัก นักเรียนที่เข้าสอบตามสังกัด โรงเรียนที่ตั้งในเขตอำเภอเมืองมีคะแนนสูงสุดสูงกว่า นักเรียนที่เข้าสอบตามสังกัด โรงเรียนที่ตั้งนอกเขตอำเภอเมือง ส่วนแบบสอบวิชาวิทยาศาสตร์ในภาพรวมของประเทศคะแนนผลการสอบมีช่วงห่างของคะแนนสูงมากเช่นเดียวกันกับวิชาคณิตศาสตร์ โดยรวมแล้วเด็กนักเรียนเก่งของทั้งสองสังกัดมีความสามารถวิชา

วิทยาศาสตร์ไม่แตกต่างกันมากนัก ซึ่งผลคะแนนดิบนี้พอจะบอกได้ว่านักเรียนที่เข้าสอบมีความสามารถแตกต่างกันอย่างมากหรือน้อยเท่านั้น อย่างไรก็ตามผลของคะแนนดิบนี้ยังไม่สามารถสรุปความสามารถของนักเรียนได้เนื่องจากเป็นคะแนนที่ค่อนข้างหยาบ การวิเคราะห์ คุณภาพรายข้อของแบบสอบ ตามทฤษฎีทางการสอบแบบดั้งเดิม ค่าความยากและค่าอำนาจจำแนกของข้อสอบในแบบสอบทั้ง 2 ฉบับ เกณฑ์โดยทั่วไปที่ใช้คัดเลือกข้อสอบที่มีคุณภาพ มีความยากระหว่าง 0.20–0.80 และอำนาจจำแนกตั้งแต่ 0.20 ขึ้นไป ในแบบสอบวิชาคณิตศาสตร์ ข้อสอบมีความยากมากถึงยากมากที่สุด อำนาจจำแนกบางข้อดีในระดับที่ใช้ได้ข้อสอบส่วนใหญ่ไม่มีอำนาจจำแนก ส่วนแบบสอบวิชาวิทยาศาสตร์มีคุณภาพรายข้อคล้ายคลึงกันกับแบบสอบวิชาคณิตศาสตร์ มีความยากมากถึงยากมากที่สุด อำนาจจำแนกบางข้อดีในระดับที่ใช้ได้ส่วนบางข้อไม่มีค่าอำนาจจำแนกและบางข้ออำนาจจำแนกมีค่าเป็นลบ

3.2 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ผลการตรวจสอบซึ่งใช้วิธีการวัดพื้นที่ของราชูเป็นวิธีเกณฑ์และเปรียบเทียบผลภายใต้วิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพล 2 เกณฑ์ แบบสอบ**วิชาคณิตศาสตร์** พบข้อสอบที่ทำหน้าที่ต่างกันในปริมาณใกล้เคียงกัน ระหว่าง การทดสอบระดับนัยสำคัญ ตรวจพบข้อสอบที่ทำหน้าที่ต่างกันตรงกับผลการตรวจด้วยวิธีเกณฑ์ในทุกข้อแต่ไม่ครบตามวิธีเกณฑ์ตรวจพบ การวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl พบข้อสอบที่ทำหน้าที่ต่างกันที่มีขนาดอิทธิพลปานกลางขึ้นไป สูงกว่า การวัดขนาดอิทธิพลตาม เกณฑ์ Zumbo and Thomas และตรงกับผลการตรวจด้วยวิธีเกณฑ์ทุกข้อ แบบสอบ **วิชาวิทยาศาสตร์** พบว่า การทดสอบระดับนัยสำคัญตรวจพบข้อสอบที่ทำหน้าที่ต่างกันตรงกับเกณฑ์สูงที่สุด รองลงมาคือการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl ซึ่งตรวจพบข้อสอบที่ทำหน้าที่ต่างกันที่มีขนาดปานกลางขึ้นไป เพียงครึ่งหนึ่งของการทดสอบระดับนัยสำคัญในขณะที่การวัดขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas ตรวจพบข้อสอบที่ทำหน้าที่ต่างกันที่มีขนาดปานกลางขึ้นไปเพียงหนึ่งในสี่ของเกณฑ์ Jodoin and Gierl

3.3 การเปรียบเทียบประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ซึ่งความยาวของแบบสอบทั้ง2 ฉบับใกล้เคียงกันคือ40 และ 50 ข้อ เมื่อคุณภาพรายข้อของแบบสอบทั้ง 2 ฉบับมีระดับความยากมากและมีอำนาจจำแนกไม่ดี และผู้สอบในแต่ละฉบับมีจำนวนสูงมาก ผลการเปรียบเทียบประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ นี้สอดคล้องกับการศึกษาในข้อมูลจำลองกล่าวคือ ภายใต้วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl ให้ผลอัตราความถูกต้องที่สูงกว่าการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas ผลนี้สอดคล้องกันในแบบสอบทั้ง 2 ฉบับและมีอัตราความคลาดเคลื่อนประเภทที่ 1 เท่ากันในแบบสอบทั้ง 2 ฉบับ

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1.1 ผลการศึกษาข้อมูลจำลอง พบว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัยโดยภาพรวมทุกเงื่อนไข คือ รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน ขนาดของการทำหน้าที่ต่างกัน จำนวนข้อสอบที่ทำหน้าที่ต่างกัน และ ความยาวของแบบสอบทั้งฉบับ ด้วยวิธีถดถอยโลจิสติกโดยการวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl มีประสิทธิภาพการตรวจสอบดีกว่า การวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas ผลจากการศึกษาในข้อมูลจำลองนี้สอดคล้องกับข้อมูลเชิงประจักษ์ในแบบสอบทั้ง 2 วิชาคือ วิชาคณิตศาสตร์และวิชาวิทยาศาสตร์ ดังนั้น หากจะศึกษาขนาดอิทธิพลของการทำหน้าที่ต่างกันของข้อสอบจึงควรเลือกใช้การวัดขนาดอิทธิพลตามเกณฑ์ของ Jodoin and Gierl เพื่อได้สารสนเทศเกี่ยวกับขนาดของการทำหน้าที่ต่างกันของข้อสอบได้อย่างเหมาะสม หากมองในด้านการได้ประโยชน์ของหน่วยงานที่จัดสอบ รูปแบบการสอบที่มีความยากสูงมีอำนาจจำแนกที่ดีการตัดสินโดยเกณฑ์ของ Zumbo and Thomas กลับเป็นผลดีต่อหน่วยงาน เนื่องจากตรวจไม่พบข้อสอบที่ทำหน้าที่ต่างกัน หรือพบเป็นจำนวนข้อที่น้อย ซึ่งในความเป็นจริงสิ่งที่เราต้องการคือการไม่เกิดการทำหน้าที่ต่างกันของข้อสอบ แต่เมื่อตรวจพบก็เป็นหน้าที่ของนักการศึกษาที่จะดำเนินการบางอย่างเกี่ยวกับการปรับปรุงเพื่อให้ข้อสอบข้อนั้นๆ มีคุณภาพต่อไป

1.2 ผลการศึกษาข้อมูลจำลองในปัจจัย ความยาวของแบบสอบ พบว่า มีผลต่ออัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ ภายใต้การตรวจสอบประสิทธิภาพด้วยวิธีถดถอยโลจิสติกโดยการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl กล่าวคือ อัตราความถูกต้องในแบบสอบที่ยาวกว่าจะมีค่าอัตราความถูกต้องสูงกว่าและอัตราความคลาดเคลื่อนประเภทที่ 1 ในแบบสอบที่ยาวกว่าจะมีค่าอัตราความคลาดเคลื่อนประเภทที่ ต่ำกว่าในแบบสอบที่สั้น เมื่อเพิ่มความยาวของแบบสอบจะทำให้อัตราความถูกต้องมีค่าเพิ่มขึ้นและอัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าลดลง นอกเหนือจากอิทธิพลหลักของความยาวของแบบสอบ แล้ว ปฏิสัมพันธ์ระหว่างความยาวของแบบสอบ กับ จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับ ยังมีผลต่อประสิทธิภาพการตรวจสอบ กล่าวคือ หากในแบบสอบมีความยาวมากกว่าและมีข้อสอบที่ทำหน้าที่ต่างกันหลายข้อจะทำให้ประสิทธิภาพการตรวจสอบด้วยวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl ให้ผลดี ในทางปฏิบัติไม่สามารถทราบจำนวนข้อที่ทำหน้าที่ต่างกันได้จนกว่าจะนำผลมาตรวจสอบโดยวิธีการทางสถิติ ดังนั้น หากจะใช้วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl มาศึกษาจึงควรเลือกใช้แบบสอบที่มีความยาวที่สุดที่มีความเหมาะสมภายใต้บริบทของเวลาในการสอบและธรรมชาติของวิชาที่สอบ

1.3 ผลการศึกษาข้อมูลจำลองในปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน พบว่า มีผลต่ออัตราความถูกต้อง ภายใต้การตรวจสอบประสิทธิภาพด้วยวิธีถดถอยโลจิสติกโดยการวัดขนาดอิทธิพล

ตามเกณฑ์ Jodoin and Gierl และการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas ซึ่งรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป มีอัตราความถูกต้องสูงกว่ารูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป และอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบที่ทำหน้าที่ต่างกัน แบบอเนกรูป การตรวจสอบด้วยวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl ให้ค่าต่ำสุด นอกเหนือจากอิทธิพลหลักของรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน แล้ว ยังมี (1) ปฏิสัมพันธ์ระหว่างรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน กับ ขนาดของการทำหน้าที่ต่างกัน ที่มีผลต่อประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ หากข้อสอบทำหน้าที่ต่างกันแบบอเนกรูปที่มีขนาดของการทำหน้าที่ต่างกันขนาดใหญ่ ผลของอัตราความถูกต้องตามเกณฑ์ Jodoin and Gierl และเกณฑ์Zumbo and Thomas จะสูงกว่าข้อสอบทำหน้าที่ต่างกันแบบอเนกรูปที่มีขนาดของการทำหน้าที่ต่างกันขนาดเล็ก และข้อสอบทำหน้าที่ต่างกันแบบเอกรูปที่มีขนาดของการทำหน้าที่ต่างกันขนาดเล็กกับปานกลาง ดังนั้น จึงควรนำวิธีถดถอยโลจิสติกโดยการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl มาตรวจสอบประสิทธิภาพการทำหน้าที่ต่างกันของข้อสอบร่วมกับการตรวจสอบประสิทธิภาพด้วยวิธีถดถอยโลจิสติกโดยการทดสอบระดับนัยสำคัญ ภายใต้อิทธิพลหลักของรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน ไม่ว่าจะเป็นแบบอเนกรูปหรือแบบเอกรูปและปฏิสัมพันธ์ระหว่างรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน กับ ขนาดของการทำหน้าที่ต่างกัน

1.4 ผลการศึกษาข้อมูลจำลอง ในปัจจัย ขนาดของการทำหน้าที่ต่างกัน พบว่า ขนาดอิทธิพล ของข้อสอบที่ทำหน้าที่ต่างกัน ขนาดใหญ่ จะมีอัตราความถูกต้องสูงกว่า ภายใต้การตรวจสอบประสิทธิภาพด้วยวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl เมื่อขนาดอิทธิพลของข้อสอบที่ทำหน้าที่ต่างกันมีขนาดเล็กและใหญ่ ตามเกณฑ์ Jodoin and Gierl จะมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าตามเกณฑ์ Zumbo and Thomas นอกจากอิทธิพลหลักของขนาดของการทำหน้าที่ต่างกันแล้ว ปฏิสัมพันธ์ระหว่างขนาดของการทำหน้าที่ต่างกัน กับรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน (รายละเอียดในข้อที่ 1.3) และปฏิสัมพันธ์ระหว่างขนาดของการทำหน้าที่ต่างกัน กับจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับ ยังมีผลต่อประสิทธิภาพการตรวจสอบ กล่าวคือ หากในแบบสอบที่มีขนาดของการทำหน้าที่ต่างกันขนาดใหญ่ และมีจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละที่สูงจะทำให้ประสิทธิภาพการตรวจสอบด้วยวิธีถดถอยโลจิสติกโดยการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl ให้ผลที่สูงกว่า ดังนั้น จึงควรนำวิธีถดถอยโลจิสติกโดยการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl มาตรวจสอบประสิทธิภาพการทำหน้าที่ต่างกันของข้อสอบร่วมกับการตรวจสอบประสิทธิภาพด้วยวิธีถดถอยโลจิสติกโดยการทดสอบระดับนัยสำคัญ ภายใต้อิทธิพลหลักของขนาดของการทำหน้าที่ต่างกัน และปฏิสัมพันธ์ระหว่าง ขนาดของการทำหน้าที่ต่างกัน กับรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน และจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับ

1.5 ผลการศึกษาข้อมูลเชิงประจักษ์ พบว่า ลักษณะรูปแบบการตอบข้อสอบมีผลต่อคะแนนการตอบ เช่น แบบสอบวิชาคณิตศาสตร์ มีรูปแบบการให้คะแนนแบบทวิวิภาค มีการกำหนดรูปแบบการตอบที่ไม่ใช่แบบเลือกตอบข้อถูกหรือข้อผิดแต่เป็นการตอบแบบปลายเปิดโดยให้ฝนคำตอบลงในกระดาษคำตอบที่ตรวจด้วยระบบคอมพิวเตอร์ ผู้วิจัยสังเกตพบว่ามีข้อมูลการตอบที่คลาดเคลื่อนหลายประเด็น อาทิ ผู้เข้าสอบฝนคำตอบที่ถูกต้องแต่ลงช่องผิดจึงทำให้ไม่ได้คะแนนในข้อนั้น หรือ ผู้เข้าสอบฝนคำตอบไม่ครบหลักที่ต้องตอบทั้งที่ความจริงสามารถคำนวณได้ถูกต้องจึงทำให้ไม่ได้คะแนนในข้อนั้น หรือผู้เข้าสอบไม่ตอบหรือไม่ฝนคำตอบใดๆ จึงทำให้ไม่ได้คะแนนในข้อนั้น เป็นต้น ซึ่งเป็นความคลาดเคลื่อนนอกเหนือจากสิ่งที่ต้องการจะวัดจากผู้เข้าสอบ สิ่งเหล่านี้ส่งผลต่อคะแนนรวมที่ผู้เข้าสอบแต่ละคนจะได้ ทำให้อัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบคลาดเคลื่อน ดังนั้น ต้องออกแบบและกำหนดรูปแบบการตอบให้มีความชัดเจนและเหมาะสมกับวัยของผู้เข้าสอบ และเป็นที่น่าสังเกตว่าข้อสอบในวิชาคณิตศาสตร์ค่อนข้างตรวจเจอการทำหน้าที่ต่างกันสูง ทั้งนี้เพราะความยากและธรรมชาติของวิชาที่สอบตลอดจนผลที่เกิดกับตัวผู้เข้าสอบหลังจากการทำข้อสอบเสร็จสิ้น เพราะการสอบที่นักเรียนสมัครใจเข้าร่วมโดยไม่มีผลต่อคะแนนในชั้นเรียนเด็กย่อมไม่แสดงความสามารถอย่างเต็มที่ กรณีนี้ทำให้การนำผลมาวิเคราะห์ด้วยทฤษฎีการทดสอบแนวใหม่ไม่มีความเหมาะสมเนื่องจากไม่มีความเป็นอิสระระหว่างระดับความสามารถของกลุ่มผู้สอบกับโอกาสของการตอบถูก

2. ข้อเสนอในการวิจัยครั้งต่อไป

2.1 การศึกษาข้อมูลจำลอง นอกจากปัจจัยที่แปรเปลี่ยนที่ผู้วิจัยศึกษา4 ปัจจัย คือ รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน ขนาดของการทำหน้าที่ต่างกัน จำนวนข้อสอบที่ทำหน้าที่ต่างกัน และความยาวของแบบสอบทั้งฉบับ ยังมีปัจจัยอื่นที่คาดว่าน่าจะมีผลต่ออัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่1 ซึ่งสามารถนำมาเป็นตัวแปรหลักในการศึกษาครั้งต่อไปอีก อาทิ ขนาดของกลุ่มตัวอย่างที่ใช้ในการศึกษา ความแตกต่างเกี่ยวกับการแจกแจงความสามารถของผู้สอบ ความยากของข้อสอบ เหล่านี้เป็นต้น และการศึกษาข้อมูลเชิงประจักษ์ที่มีขนาดกลุ่มตัวอย่างเป็นจำนวนมาก ผลที่ได้จากการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการ ทดสอบระดับนัยสำคัญ อาจส่งผลให้เกิดความคลาดเคลื่อนในการทดสอบเพราะโอกาสที่จะพบความแตกต่างอย่างมีนัยสำคัญค่อนข้างสูง นักวิจัยควรพิจารณาเงื่อนไขเกี่ยวกับจำนวนกลุ่มตัวอย่างและเงื่อนไขเกี่ยวกับความยาวของแบบสอบเพื่อให้เกิดความเหมาะสมและรัดกุม

2.2 การศึกษาในข้อมูลเชิงประจักษ์ พบว่า ลักษณะของข้อสอบมีผลต่ออัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการที่ศึกษา กล่าวคือ เมื่อข้อสอบส่วนใหญ่ในแบบสอบวิชาคณิตศาสตร์ มีลักษณะความยากมากจะให้อัตราความถูกต้องจากการตรวจสอบภายใต้วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพล ตามเกณฑ์ Jodoin and Gierl มีค่าสูงกว่าการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas เมื่อพิจารณาอำนาจจำแนก

ของข้อสอบฉบับดังกล่าวซึ่งพบว่า ข้อสอบจำแนกเด็กเก่งกับเด็กอ่อนไม่ ดี กลับทำให้อัตราความคลาดเคลื่อนประเภทที่1 มีค่าต่ำมากทั้ง2 เกณฑ์ที่ศึกษา ในขณะที่ลักษณะข้อสอบส่วนใหญ่ในแบบสอบวิชาวิทยาศาสตร์ มีความยากระดับปานกลางขึ้นไป จะให้อัตราความถูกต้องจากการตรวจสอบภายใต้วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl มีค่าสูงกว่าการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas ซึ่งสอดคล้องกับในวิชาคณิตศาสตร์ เมื่อพิจารณาอำนาจจำแนกของข้อสอบวิชาวิทยาศาสตร์ซึ่งพบว่าข้อสอบจำแนกเด็กเก่งกับเด็กอ่อนได้ระดับดีถึงระดับดีมาก กลับให้อัตราความคลาดเคลื่อนประเภทที่1 มีค่าสูงมากทั้ง 2 เกณฑ์ที่ศึกษา ซึ่งในการศึกษาข้อมูลจำลองไม่มีปัจจัยที่เกี่ยวกับระดับความยากและอำนาจจำแนกจากการตอบข้อสอบ ดังนั้นการวิจัยครั้งต่อไปจึงควรพิจารณาปัจจัยดังกล่าวด้วย

2.3 เนื่องจากการศึกษาตัวแปรด้านปัจจัยที่แปรเปลี่ยน4 ปัจจัยหลักในข้อมูลจำลอง พบว่านอกเหนือจากอิทธิพลของปัจจัยหลักที่มีผลต่ออัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ยังมีปฏิสัมพันธ์ระหว่างปัจจัยที่แปรเปลี่ยนอีกด้วยที่มีผลต่ออัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ดังนั้นเมื่อศึกษาในข้อมูลเชิงประจักษ์ ควรมีการศึกษาเพิ่มเติมเกี่ยวกับตัวแปรอื่น เช่น ศาสนา ภูมิลำเนา กลุ่มอายุ เพศ และความรู้ในวิชาอื่นๆ นอกเหนือจากที่ศึกษาในครั้งนี้ รวมถึงปฏิสัมพันธ์ระหว่างตัวแปรอื่นที่สนใจศึกษาด้วย และควรมุ่งศึกษาเพิ่มในประเด็นความเป็นพหุมิติ ศึกษาเพิ่มในประเด็นรูปแบบของการตรวจให้คะแนนแบบพหุมิติ ตลอดจนการตรวจสอบในชุดข้อสอบ เน้นศึกษาเชิงลึกของข้อสอบกรณีแบบเลือกตอบที่ตัวเลือกและตัวลวง เพื่อให้เกิดความหลากหลายและสอดคล้องต่อสถานการณ์ที่น่าจะมีโอกาสเกิดขึ้นในความเป็นจริง

2.4 เมื่อนักวิจัยต้องการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ต้อง ไม่เน้นเพียง เพื่อตัดสินใจตัดข้อสอบออกไปจากแบบสอบเท่านั้นต่ควรหาสารสนเทศอื่นมาประกอบการตัดสินข้อสอบอย่างรัดกุม เนื่องจากการสอบแต่ละครั้งโดยเฉพาะการสอบระดับชาติเป็นเรื่องที่ต้องมีการลงทุนสูงในหลายด้านและใช้ทุนทรัพย์ค่อนข้างสูง การออกแบบบริหารจัดการการสอบเป็นเรื่องที่ต้องตั้งอยู่บนหลักของความคุ้มค่า เกิดประสิทธิผลและมีประสิทธิภาพสูงสุด ดังนั้นนักการศึกษาต้องอาศัยหลักฐานที่มีน้ำหนักเพียงพอจึงควรมีการพิจารณาตัดสินการทำหน้าที่ต่างกันของข้อสอบโดยวิธีการอื่นๆ ที่หลากหลาย วิธีการที่น่าสนใจคือการอาศัยความชำนาญด้านเนื้อหาและด้านวัดผลประเมินผลของ ผู้เชี่ยวชาญมาตัดสินการทำหน้าที่ต่างกันของข้อสอบร่วมกับการใช้หลักฐานในทางสถิติเพื่อเป็นการตรวจสอบความถูกต้องของผลการตัดสินการทำหน้าที่ของข้อสอบให้เกิดความน่าเชื่อถือและมีความยุติธรรมมากที่สุดในการตัดสินข้อสอบ

รายการอ้างอิง

ภาษาไทย

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

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

จิติมา วรรณศรี. (2539). การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซลกับวิธีซิปเทสท์ เมื่อความยาวของแบบทดสอบ ขนาดกลุ่มตัวอย่าง และอัตราส่วน ของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบต่างกัน. วิทยานิพนธ์ปริญญามหาบัณฑิต. ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

ชยุตม์ ภิรมย์สมบัติ. (2547). คุณสมบัติของตัวประมาณค่าความเข้มของอิทธิพล: การเปรียบเทียบระหว่างทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบ.วิทยานิพนธ์ปริญญามหาบัณฑิต. ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

โชติกา ภาษีผล. (2554). การสร้างและพัฒนาเครื่องมือในการวัดและประเมินผลการศึกษา สำนักพิมพ์ : คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

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

ทองอยู่ สาระ. (2543). การเปรียบเทียบอำนาจการตรวจสอบและการจำแนกผิดพลาดในการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบสม่ำเสมอและแบบไม่สม่ำเสมอระหว่างวิธีแมนเทลแฮนส์เซลและวิธีการถดถอยโลจิสติคโดยใช้ความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่างต่างกัน ปริญญานิพนธ์ กศ.ม.(การวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

นงลักษณ์ วิรัชชัย. (2542). การวิเคราะห์อภิมาน: META-ANALYSIS. ปทุมวัน. กรุงเทพมหานคร. โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

นพมาศ พิพัฒนสุข. (2541). การเปรียบเทียบประสิทธิภาพระหว่างวิธีแมนเทล-เฮลส์เซลกับวิธีถดถอยโลจิสติก ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อใช้เกณฑ์จับคู่เปรียบเทียบ

แตกต่างกันในแบบสอบชนิดพหุมิติ. วิทยานิพนธ์ปริญญามหาบัณฑิต. ภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

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

บุญธรรม กิจปรีดาบริสุทธิ์. (2543). รวมบทความ การวิจัย การวัดและประเมินผล. พิมพ์ครั้งที่ 2 กรุงเทพฯ: โรงพิมพ์ศรีอนันต์.

ปิยะทิพย์ ตินวร. (2549). การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบพหุมิติ: การเปรียบเทียบประสิทธิภาพระหว่างวิธีการวิเคราะห์องค์ประกอบจำกัดกับวิธีถดถอยโลจิสติก. วิทยานิพนธ์ปริญญาหาบัณฑิต. ภาควิชาวิจัยและวัดผลการศึกษาคณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา.

พรรณี จินตมาศ. (2540). การเปรียบเทียบผลการวิเคราะห์ความลำเอียงของข้อสอบโดยใช้ขนาด กลุ่มผู้สอบและวิธีวิเคราะห์ต่างกัน. วิทยานิพนธ์ปริญญามหาบัณฑิต. สาขาวิชาการวัดผลการศึกษา. บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ.

รัชรินทร์ มุคดา. (2540). การเปรียบเทียบประสิทธิภาพระหว่างวิธีแมนเทล-แฮนส์เซลกับวิธีถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปในกรณีการจัดกลุ่มความสามารถ ค่าความยากของข้อสอบ และค่าอำนาจจำแนกของข้อสอบต่างกัน.วิทยานิพนธ์ปริญญามหาบัณฑิต. ภาควิชาวิจัยและจิตวิทยาการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

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

เรวดี อินทะสะระ. (2539). ผลการตรวจสอบความลำเอียงของข้อสอบต่อการศึกษาความเที่ยงตรงเชิงพยากรณ์ของแบบทดสอบคัดเลือกที่คิดคะแนนต่างกัน ปริญญานิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา). บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ.

วลีมาศ แซ่อึ้ง. (2543). การเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อน ประเภทที่1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติก. วิทยานิพนธ์ปริญญาครุศาสตร์บัณฑิต. ภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

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

ศิริชัย กาญจนวาสี. (2548). ทฤษฎีการทดสอบแบบดั้งเดิม (CLASSICAL TEST THEORIES). (พิมพ์ครั้งที่ 5). กรุงเทพมหานคร. โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนวาสี. (2550). ทฤษฎีการทดสอบแนวใหม่ (MODERN TEST THEORIES). (พิมพ์ครั้งที่ 3). กรุงเทพมหานคร. โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

สิริรัตน์ วิภาสศิลป์. (2545). การเปรียบเทียบสิธีซิปเทสท์และดีเอฟไอทีในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบ จากข้อมูลการตอบข้อสอบที่ใช้ความสามารถหลายมิติ. ปริญญานิพนธ์ กศ.ด.(การทดสอบและวัดผลการศึกษา). บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ.

สุทธิพร สุรธณี. (2550). การศึกษาความสามารถในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ตามตัวแบบเชิงเส้นวางนัยทั่วไประดับลดหลั่นิทยานิพนธ์ปริญญามหาบัณฑิต. คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์.

สุมาลี แกล้วทนงค์. (2547). สาเหตุการทำหน้าที่ต่างกันของข้อสอบสาระการเรียนรู้ภาษาไทย และสาระการเรียนรู้สังคมศึกษา ศาสนาและวัฒนธรรม วิทยานิพนธ์ปริญญามหาบัณฑิต. ภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

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

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

เสรี ชัดแช้ม.(2540). วิธีการทางสถิติที่ใช้ตรวจสอบข้อสอบทำหน้าที่ต่างกัน *วารสารมหาวิทยาลัยบูรพา, 2*(1), 41-53.

สำนักงานสถิติแห่งชาติ. (2555). http://statstd.nso.go.th/sitesearch/Display.aspx?id=A9D42FDAF354DE3906B81FE7FB6AFD66andq= (ข้อมูล ณ วันที่ 4 พฤษภาคม 2555)

อรินทร์ น่วมถนอม. (2549). การเปรียบเทียบวิธีโพลี-ซิปเทสท์วิธีการถดถอยโลจิสติกแบบจัดอันดับและวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า.ปริญญานิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา). บัณฑิตวิทยาลัย. มหาวิทยาลัยศรีนครินทรวิโรฒ.

อารี วัชรโสตถิกุล. (2543). การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้รูปแบบและวิธีการแตกต่างกัน วิทยานิพนธ์ปริญญาการศึกษามหาบัณฑิต. สาขาวิชาการวัดผลการศึกษา. บัณฑิตวิทยาลัย. มหาวิทยาลัยศรีนครินทรวิโรฒ.

อุทัยวรรณ สายพัฒนะ. (2547). การเปรียบเทียบประสิทธิภาพของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่าวิธี GMH และวิธี Polytomous SIBTEST. วิทยานิพนธ์ดุษฎีบัณฑิต. ภาควิชาการทดสอบและวัดผลทางการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.

อุทุมพร จามรมาน. (2540). ทฤษฎีการวัดทางจิตวิทยา. กรุงเทพมหานคร: ฟันนี่พับลิชชิ่ง.

ภาษาอังกฤษ

Ackerman, T.A. (1992). A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. Journal of Educational Measurement 29(1): 67-91.

Agresti, A. (1990). Categorical data analysis. New York: Wiley.

Allen, M.J. and Yen, W.M. (1979). Introduction to Measurement Theory. Monterey, CA: Brooks/Cole Publications,Inc.

Angoff, W.H. (1993). Perspectives on differential item functioning methodology. In P.W.

Baker, F.B. (1977). “Advance in Item Analysis,” Review of Education Research. 47: 151-178.

Bolt, D.M. (2002). A Monte Carlo comparison of parametric and nonparametric polytomous DIF detection methods. Applied Measurement in Education 15(2): 113-141.

Camilli, G. and Congdon, P. (1999). Application of a method of estimating DIF for polytomous test items. Journal of Educational and Behavioral Statistics, 24(4), 323-341.

Camilli, G. and Shepard, L.A. (1994). Methods for identifying biased test items. Thousand Oaks, CA: Sage Publications. (Vol. 4)

Chang, H., Mazzeo, J. and Roussos, L. (1996l). Detecting DIF for polytomously scored items: An adaptation of the SIBTEST procedure. Journal of Educational Measurement 33(3): 333-353.

Clauser, R.E. and Mazor, K.M. (1998). Using statistical procedures to identify differentially functioning test items. Educational Measurement: Issues and Practice 17(1): 31-44.

Cohen, A.S. and Bolt, D.M. (2005). A mixture model analysis of differential item functioning. Journal of Educational Measurement 42(2), 133-148.

Cohen, A.S. and Kim, S. H. (1993). A comparison of Lord's chi-square and Raju's area measures in detection of DIF. Applied Psychological Measurement 17(1): 39-52.

Cohen, A.S., Kim, S.H., Wollack, J.A. (1996). An Investigation of the Likelihood Ratio Test for Detection of Differential Item Functioning. Applied Psychological Measurement 20(1), 15-26.

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Douglas, A., Roussos, A. and Stout, F. (1996). “Item-Bundle DIF Hypothesis Testing: Identifying Suspect Bundles and Assessing Their Differential Functioning”, Journal of ducational Measurement 33(4): 465-484

Dorans, N.J. and Kulick, E. (1986). "Demonstrating the Utility of the Standardization Approach to Assessing Unexpected Differential Item Performance on the Scholastic Aptitude Test," Journal of Educational Measurement 23(4): 355-368.

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

Gómez-Benito, J.Hidalgo, M.D. and Padilla, J.L.,(2009). Efficacy of Effect Size Measures in Logistic Regression An Application for Detecting DIF. Methodology: European Journal of Research Methods for the Behavioral and Social Sciences. Vol 5(1), 2009, 18-25.

Huseyin H. Y. and Giray Berberoˇglu, (2009). Judgmental and Statistical DIF Analyses of the PISA-2003 Mathematics Literacy Items. International Journal of Testing, 9: 108–121.

Hambleton, R.K. and Cook, L.L. (1977). Latent Trait Model and Their Use in The Analysis of Education Test Data. Journal of Educational Measurement 14(2): 75-96.

Hambleton, R.K. and Swaminathan, H. (1985). Item Response Theory. Boston: Kluwer-Nijhoff.

Hambleton, R.K., Swaminatan, H. and Roger, H. (1991). Fundamentals of Item Response Theory. California: Sage Publications: Hambleton, 1991.

Harwell, M., Stone, C.A., Hsu, T-C, and Kirisci, L. (1996). Monte Carlo studies in Item Response Theory. Applied psychological measurement 20(2): 101-125.

Holland, W.P. and Thayer, D.T. (1988). Differential item performance and the Mantel-Haenszel procedure In Test validity. H. Wainer; and H. I. Braun. pp. 129-145. Hillsdale, NJ: Erlbaum.

Holland, W.P. and Wainer H. (1993). Differential Item Functioning. Hillsdale, New Jersey: Lowrence Erlbaum Associates. pp. 3-23.

Hudson, Z. (2009). Sample size, power and effect size- What all researchers need to know. Physica! Therapy in Sport 10(2009): 43-44.

Hulin, C.L., Drasgow, F. and Parsons, C.K., (1983). Item Response Theory: Application to psychological measurement. Homewood, I.L: Dow Jones: Irwin.

Jodoin, M.G.,and 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, 329 – 349.

Kederman, H. (1990). Item bias detection using loglinear IRT. Psychometricka, 54, 681-697.

Kim, S.H.; and Cohen, A.S. (1991). A comparison of two area measures for detecting differential item functioning. Applied Psychological Measurement 15(3): 269-278.

Kim, S.H., Chosen, A.S. and Kim, S. (2007). DIF Detection and Effect Size Measures for Polytomously Scored Items. Journal of Educational Measurement Summer, Vol. 44, No. 2, pp. 93-116.

Kirk, R.E. (1996). Practical significance: a concept whose time has com. Educational and Psychological Measurement 56: 746-759.

Kristjansson, E., Aylesworth, R., Mcdowell, I. and Zombo, B.D. (2005). A Comparison of you methods for detecting differential item functioning in ordered response items. Educational and Psychological Measurement, 65, 935-953.

Kyung T. H. (2007). User's Manual for WinGen: Windows Software that Generates IRT Model Parameters and Item Responses. Center for Educational Assessment Research Report No. 642. Amherst, MA: University of Massachusetts, Centerfor Educational Assessment.

Lee, Y.W., Breland, H., Muraki, E. (2004). Comparability of TOEFL CBT Writing Prompts for Different Native Language Groups (TOEFL Research Rep. No. RR-77). Princeton, NJ Educational Testing Service.

Li, H. and Stout, W. (1996). A new procedure for detecting crossing DIF. Psychometrika. 61(4): 647-677.

Linn, R.L., et al. (1981). An investigation of item bias in a test of reading comprehension. Applied Psychological Measurement 5: 159-173.

Lord, F.M. (1980). "Application of Item Response Theory to Practical Testing Problem." Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Lord, F.M. and Novick, M.R. (1968). Statistical theories of mental test score. Massachusetts: Addison-Wesley.

Luc, (2009). Investigating Gender Differential Item Functioning Across Countries and Test Languages for PISA Science Items. International Journal of Testing, 9: 122–133.

Marie, W. et al., (2007). "Measuring and Detecting Differential Item Functioning in Criterion Referenced Licensing Test: A Theoretic Comparison of Methods". Umea°, May 7-8.

Marie W., (2009). Differential Item Functioning in Mastery Tests: A Comparison of Three Methods Using Real Data. International Journal of Testing, 9: 41–59.

Mazor, K.M., Kanjee, A. and Clauser, B.E. (1995). Using logistic regression and the Mantel-Haenszel with multiple ability estimates to detect differential item functioning. Journal of Educational Measurement, 32, 131-144.

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

Mazor, F.M., Clauser, B.E. and Hambleton, R.K. (1994). Identification of nonuniform differential item functioning using a variation of the Mantel-Haenszel procedure. Educational and Psychological Measurement 54, 284-291.

Mellenbergh, G.J. (1982). "Contingency table models for assessing item bias. Journal of Educational Statistics 7, 105 - 118.

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

Muraki, E. (1992). A generalized partial credit model: Application of an EM Algorithm. Applied Psychological Measurement 16(2): 159-176.

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

Narayanan, P. and Swaminathan, H. (1996). Identification of items that show nonuniform DIF. Applied Psychological Measurement 20(3): 257-274.

Nilufer K. and Paul D. B., (2009). Modeling DIF in Complex Response Data Using Test Design Strategies. International Journal of Testing, 9: 151–166.

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

Oshima, T.C., Raju, N.S. and Flowers, C.P. (1997). Development and demonstration of multidimensional IRT-based internal measures of differential functioning of items and tests. Journal of Educational Measurement 34(3): 253-272.

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

Park, T. (2006). Detecting DIF across Different Language and Gender Groups in the MELAB Essay Test using the Logistic Regression Method. Spaan Fellow Working Papers iin Second or Foreign Language Assessment. Volume 4.

Penfield, R.D. (2005). DIFAS: Differential Item Functioning Analysis System. Applied Psychological Measurement Vol. 29, No. 2, pp. 150-151

Penfield, R.D. (2007). DIFAS: Differential Item Functioning Analysis System. User's Manual.

Penfield, R.D. (2001). Assessing differential item functioning across multiple groups: A comparison of three Mantel-Haenszel procedures. Applied Measurement in Education, 14, 235-259.

Penfield, R.D. and Algina, J. (2006). A generalized DIF effect variance estimator for measuring global differential test functioning in mixed format tests. Journal of Educational Measurement 43, 295-312.

Popham, W.A. (1981). Modern Education Measurement. Engwood Cliffs, NJ: Prentice-Hall.

Potenza, M.T. and Dorans, N.J. (1995). "DIF assessment for polytomously scored items: A framework for classification and evaluation". Applied Psychological Measurement 19(2), 211-237.

Procedure. In P.W. Wainerand H.T. Braun (eds.). Test Validity, pp. 129-145. Hillsdale, NJ: Lawrence Erlbaum Associates.

Raju, N.S. (1988). The area between two item characteristic curves. Psychometrika 53, 495-502.

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.

Raju, S. et al. (1993). "An Empirical Comparison of the Area Methods, Lord's Chi-Square Test, and the Mantel- Haenszel Techinque for Assessing Differential Item Functioning." Educational and Psychological Measurement 53: 301-314.

Raju, S., van der Linden, J. and Fleer F. (1995). "IRT-based Internal Measures of Differential Functioning of Items and Tests", Applied Psychological Measurement 19(4): 353-368.

Rogers, H.J. and Swaminathan, H. (1993). A comparison of logistic regression and Mantel-Haenszel procedures for detecting differential item functioning. Applied Psychological Measurement 17(2): 105-116.

Roussos, L. and Stout, W. (1996b). A multidimensionality-based DIF analysis paradigm. Applied Psychological Measurement 20(4): 355-371.

Rudner, L.M. (1977). An evaluation of select approaches for biased item identification. Unpublished doctoral dissertation, Catholic University of America, Washing DC.

Rudner, L.W., Getson, P.R., and Knight, D.L. (1980). A monte carlo comparison of seven biased item detection techniques. Journal of Educational Measurement 17(1): 1-10.

Ryan, Katherine E. and Chiu, S. (2001). "An Examination of Item Context Effects, DIF and Gender DIF", Applied Measurement in Education 14(1): 73-90.

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

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

Shen, L. (1999). A Multilevel Assessment of Differential Item Functioning. Paper Presented at the annual meeting of American Educational Research Association, Montreal, Quebec, Canada.

Shepard, L.A., Camilli, G. and Williams, D.M. (1984). Accounting for statistical artifacts in item bias research. Journal of Educational Statistics 9(2): 93-128.

Shumacker, R.E. (2005). http://www.appliedmeasurementassociates.com/White%20Papers/TEST% 20BIAS%20AND%20DIFFERENTIAL%20ITEM%20FUNCTIONING.pdf. [วันที่ 11 มิถุนายน 2551]

Stark, S., Chernyshenko, O.S. and Drasgow, F. (2006). Detecting Differential Item Functioning With Comfirmatory Factor Analysis and Item Response Theory: Toward a Unified Strategy. Journal of Applied Psychology 91: 1292-1306.

Su, Y.H., and Wang, W.C. (2005). Efficiency of the Mantel, generalized Mantel-Haenszel, and logistic discriminant function analysis methods in detecting differential item functioning for polytomous item. Applied Measurement in Education 18, 313-350.

Swaminatha, H. and Rogers, H.J. (1990). Detecting differencetial item functioning using logistic regression procedure. Journal of Educational Measurement 27, 361-370.

Thissen, D. (2001). IRTLRDIF v.2.0b [Computer Program]. University of North Carolina at Chapel Hall: L.L. Thurstone Psychometric Laboratory.

Thissen, D. and Steinberg, L. (2006). Using Effect Sizes for Research Reporting: Examples Using Item Response Theory to Analyze Differential Item Functioning. Journal of Psychological Methods 11(4): 402-415.

Thissen, D., Steinberg, L. and Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In Differential item functioning. W. P. Holland; and H. Wainer. pp. 67-113. Hillsdale, NJ: Erlbaum.

Trusty, J., Thompson, B. and Petrocelli, V.j. (2004). Practical guide for reporting effect size in quantitative research in the journal of counseling and development. Journal of Counseling and Development, 82: 107-110.

Wang, W.C., and Su, Y.H. (2004). Factors influencing the Mantel and Generalized Mantel-Haenszel Methods for the assessment of differential item functioning in polytomous items. Journal of Applied Psychological Measurement 28(6), 450-480.

Wiberg, M. (2007). Measuring and Detecting Differential Item Functioning in Criterion-Referenced Licensing Test. EM No, 60.

Zwick, R., Donoghue, J.R., and Grimo, A. (1993). Assessing differential item functioning in performance tasks. Journal of Educational Measurement 30, 233-251.

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

Zumbo, B.D. (1999). A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type item scores. Ottawa, Canada: Directorate of Human Resources Research and Evaluation, Department of National Defense. Retrieved from http://www.edu.ubc.ca/faculty/zumbo/DIF/index.html

Zumbo, B.D. (2005). “A comparison of four methods for detecting differential item functioning in ordered response items”. Journal of Educational and Psychological Measurement 65(6): 935-953.

Zumbo, B.D., and Hubley, A.M. (2003). Item bias. In Roci’o Fern’andez-Ballesteros (Ed.), Encyclopedia of psychological assessment (pp. 505-509). Thousand Oaks, CA: Sage.

Zumbo, B.D., and Thomas, D.R. (1997). A measure of effect size for a model-based approach for studying DIF. Prince George, Canada: University of Northern British Columbia, Edgeworht Laboratory for Quantitative Behavioral Science.

ภาคผนวก

ภาคผนวก ก

ค่าพารามิเตอร์ข้อสอบรายข้อ ภายใต้ทฤษฎีการตอบสนองข้อสอบ

ชนิด 2 พารามิเตอร์ (two-parameter)

ภาคผนวก ก

ค่าพารามิเตอร์ข้อสอบรายข้อ ภายใต้ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory model) ชนิด 2 พารามิเตอร์ (two-parameter)

ตาราง ก-1 ค่าพารามิเตอร์ข้อสอบรายข้อ ภายใต้ทฤษฎีการตอบสนองข้อสอบ ชนิด 2 พารามิเตอร์ (two-parameter) จำแนกตามแบบสอบแต่ละฉบับ

แบบสอบ (รหัส)	ข้อที่	A	B	ข้อที่	A	B	ข้อที่	A	B	ข้อที่	A	B
ฉบับที่ 1	1d*	1.063	-0.890	11	0.944	-1.754	21	0.743	-2.297	31	1.626	0.001
(N_u1_k40d0.1nd4)	2d*	1.057	-0.720	12	1.003	1.093	22	1.273	0.949	32	1.126	-0.188
	3d*	0.786	-0.337	13	1.030	0.431	23	0.844	0.773	33	0.976	1.840
	4d*	1.118	0.594	14	0.751	-0.060	24	0.985	0.207	34	1.198	-1.518
	5	1.428	-2.346	15	0.944	-0.589	25	1.070	-0.711	35	1.657	-0.303
	6	0.853	-0.151	16	0.806	0.847	26	0.877	-0.646	36	1.528	0.443
	7	1.038	0.047	17	0.783	-0.326	27	1.062	0.261	37	0.990	0.847
	8	0.907	0.646	18	1.083	1.566	28	0.904	1.792	38	0.868	1.288
	9	1.120	1.612	19	1.158	0.093	29	1.535	1.573	39	0.836	-1.607
	10	0.913	2.277	20	1.009	-0.999	30	0.901	1.125	40	1.058	-0.788
ฉบับที่ 2	1d*	0.869	0.846	11	0.900	0.072	21	0.875	0.443	31	1.302	1.773
(N_u 2_k40d0.2nd4)	2d*	1.095	0.83	12	0.734	-0.543	22	1.186	1.776	32	0.828	-1.018
	3d*	1.015	-0.837	13	0.959	1.516	23	1.136	1.013	33	1.075	-0.659
	4d*	0.934	0.004	14	1.030	-0.417	24	1.422	-1.003	34	0.960	-1.280

ตาราง ก-1 (ต่อ)

แบบสอบ (รหัส)	ข้อที่	A	B	ข้อที่	A	B	ข้อที่	A	B	ข้อที่	A	B
	5	0.731	-0.47	15	1.234	-1.316	25	0.937	-0.145	35	0.741	-1.952
	6	1.138	0.207	16	1.077	2.223	26	1.065	0.270	36	0.997	-0.695
	7	1.311	0.761	17	1.047	0.416	27	0.895	-1.266	37	0.675	1.327
	8	1.191	1.002	18	0.764	0.330	28	0.896	0.069	38	0.891	-0.209
	9	1.335	1.706	19	0.910	0.414	29	1.270	-0.417	39	0.609	-0.747
	10	0.807	-0.675	20	0.922	1.641	30	1.554	-0.209	40	1.189	-1.965
ฉบับที่ 3	1d*	0.864	-0.372	11	1.675	-0.936	21	0.962	-0.781	31	1.186	-0.543
(N_u3_k40d0.4nd4)	2d*	1.071	-1.347	12	0.920	-1.319	22	1.102	0.009	32	1.231	-0.287
	3d*	0.849	1.496	13	1.218	-0.825	23	1.125	-0.170	33	0.935	-0.464
	4d*	0.701	-0.126	14	1.400	0.084	24	1.293	0.181	34	0.893	0.611
	5	0.977	0.598	15	1.337	2.069	25	0.924	0.665	35	1.153	-0.072
	6	0.911	0.424	16	0.965	0.238	26	1.372	-0.960	36	1.408	0.357
	7	0.829	0.503	17	0.892	2.355	27	1.088	-2.014	37	1.186	-0.114
	8	0.837	-1.236	18	1.117	-1.609	28	1.317	-1.105	38	1.286	0.614
	9	0.723	0.927	19	0.712	0.347	29	1.073	0.067	39	1.122	0.041
	10	1.364	2.013	20	0.942	-0.665	30	0.790	-1.124	40	1.281	-1.186
ฉบับที่ 4	1d*	0.897	-0.895	11	1.440	1.134	21	0.988	-0.159	31	0.899	0.276
(N_u4_k40d0.1nd8)	2d*	1.352	0.126	12	0.798	0.568	22	1.220	-0.414	32	1.080	-0.478
	3d*	1.063	1.965	13	0.832	1.389	23	1.020	-0.384	33	1.005	-0.851
	4d*	1.300	-0.185	14	1.294	1.129	24	0.906	0.021	34	0.919	0.042

ตาราง ก-1 (ต่อ)

แบบสอบ (รหัส)	ข้อที่	A	B	ข้อที่	A	B	ข้อที่	A	B	ข้อที่	A	B
	5d*	1.476	-0.530	15	0.896	-1.002	25	1.203	-2.835	35	0.760	-1.454
	6d*	1.059	1.013	16	1.071	-2.910	26	0.948	0.407	36	0.628	-1.731
	7d*	0.904	-2.561	17	0.876	1.064	27	1.026	-1.086	37	1.029	0.860
	8d*	1.166	1.713	18	0.821	2.057	28	0.812	0.653	38	1.052	-0.409
	9	1.129	-0.492	19	1.060	0.694	29	1.022	-0.028	39	1.373	0.950
	10	0.971	0.177	20	1.036	-2.301	30	1.002	-1.325	40	1.319	0.878
ฉบับที่ 5	1d*	1.146	0.028	11	0.904	0.979	21	0.766	-1.386	31	1.410	0.582
(N_u5_k40d0.2nd8)	2d*	0.796	0.524	12	0.784	-0.352	22	0.896	0.388	32	1.022	-0.456
	3d*	1.280	-1.529	13	1.105	-1.764	23	0.957	1.684	33	1.284	-0.478
	4d*	0.873	-0.960	14	0.729	-1.631	24	1.072	-0.553	34	0.895	2.290
	5d*	0.806	-0.522	15	0.863	-2.036	25	0.954	1.138	35	0.800	-0.364
	6d*	0.850	-1.071	16	1.023	0.955	26	0.729	0.067	36	1.260	1.147
	7d*	0.892	0.408	17	0.997	0.418	27	0.946	1.172	37	0.966	0.077
	8d*	1.102	-0.715	18	0.973	-0.891	28	0.912	0.523	38	0.547	0.771
	9	1.465	0.574	19	0.938	-0.028	29	0.920	0.422	39	0.926	0.547
	10	1.040	0.824	20	0.976	0.528	30	0.537	-0.155	40	0.973	-0.323
ฉบับที่ 6	1d*	0.907	-0.268	11	0.921	-1.542	21	1.111	0.379	31	1.193	0.593
(N_u6_k40d0.4nd8)	2d*	1.156	-0.581	12	0.832	0.935	22	0.822	0.131	32	1.050	0.760
	3d*	0.786	0.287	13	0.810	0.707	23	1.155	-0.294	33	0.988	0.667
	4d*	1.277	-1.614	14	0.924	-0.686	24	1.313	0.666	34	1.005	-0.045
	5d*	0.927	0.543	15	0.973	-0.254	25	1.170	-0.896	35	1.163	0.688

ตาราง ก-1 (ต่อ)

แบบสอบ (รหัส)	ข้อที่	A	B	ข้อที่	A	B	ข้อที่	A	B	ข้อที่	A	B
	6d*	0.814	-1.169	16	1.115	1.000	26	1.249	-0.652	36	1.139	1.074
	7d*	0.859	0.789	17	1.176	1.276	27	0.718	-0.254	37	0.870	-0.256
	8d*	0.924	-1.152	18	0.835	0.242	28	0.944	0.847	38	0.997	0.436
	9	0.857	-0.683	19	1.194	1.458	29	1.389	-0.362	39	1.137	1.288
	10	0.818	0.708	20	1.304	0.098	30	0.783	0.103	40	1.093	0.502
ฉบับที่ 7	1d*	1.266	-1.268	14	1.217	1.464	27	1.106	0.986	40	1.343	-0.323
(N_u7_k50d0.1nd5)	2d*	0.899	-0.848	15	0.915	-0.553	28	0.981	0.696	41	1.018	2.265
	3d*	1.075	-0.103	16	0.964	1.239	29	1.422	1.619	42	1.121	-0.321
	4d*	1.287	0.310	17	0.888	-0.583	30	1.028	-0.663	43	0.776	2.453
	5d*	0.981	-1.323	18	0.955	1.271	31	0.782	2.004	44	0.965	-0.834
	6	0.903	2.123	19	1.291	0.522	32	0.968	0.384	45	0.933	2.032
	7	0.959	0.454	20	0.880	0.937	33	1.136	-0.650	46	1.043	-2.807
	8	0.955	-0.770	21	0.866	0.643	34	0.948	-0.566	47	1.100	-0.745
	9	0.837	-1.455	22	1.079	-1.599	35	0.618	-0.886	48	1.423	1.946
	10	1.053	-0.503	23	0.714	1.927	36	0.963	-1.790	49	1.022	0.895
	11	1.178	-0.706	24	0.687	-1.359	37	1.091	-1.045	50	0.914	0.558
	12	0.869	-0.304	25	0.932	-1.004	38	1.106	1.140			
	13	1.596	0.540	26	0.892	0.425	39	1.108	0.133			
ฉบับที่ 8	1d*	2.024	0.915	14	1.207	0.632	27	0.964	0.078	40	0.960	1.600
(N_u8_k50d0.2nd5)	2d*	1.118	0.399	15	0.923	-0.848	28	1.152	1.114	41	0.710	-0.578
	3d*	0.914	1.166	16	1.218	-1.076	29	0.908	2.302	42	1.152	0.328

ตาราง ก-1 (ต่อ)

แบบสอบ (รหัส)	ข้อที่	A	B	ข้อที่	A	B	ข้อที่	A	B	ข้อที่	A	B
	4d*	1.179	0.403	17	0.933	0.537	30	0.694	0.008	43	0.773	-0.100
	5d*	0.884	-1.256	18	0.822	-1.528	31	1.066	0.018	44	1.497	-0.234
	6	0.862	-1.138	19	1.286	0.347	32	0.928	1.152	45	1.219	0.365
	7	0.720	-0.607	20	0.908	0.173	33	1.314	-0.690	46	0.937	-0.134
	8	0.945	1.119	21	0.877	0.697	34	0.760	-0.164	47	0.869	0.335
	9	1.175	-0.993	22	1.214	0.703	35	0.855	0.872	48	1.544	-0.922
	10	0.979	2.055	23	0.794	1.054	36	0.826	-0.228	49	1.223	1.596
	11	1.271	-0.188	24	1.039	-0.663	37	0.875	0.415	50	1.079	0.544
	12	1.011	-1.451	25	0.959	1.403	38	1.019	1.223			
	13	1.088	-0.318	26	1.035	-1.368	39	0.996	-0.015			
ฉบับที่ 9	1d*	0.806	0.934	14	1.071	-1.596	27	0.811	-1.565	40	1.052	-0.119
(N_u9_k50d0.4nd5)	2d*	0.602	0.253	15	0.649	0.974	28	1.145	-0.343	41	0.816	0.245
	3d*	0.550	-0.464	16	0.990	-2.208	29	1.012	-1.238	42	0.969	-0.308
	4d*	1.050	-0.869	17	0.810	-0.728	30	0.637	0.011	43	1.233	-1.516
	5d*	0.722	-0.933	18	0.753	0.366	31	1.017	0.885	44	1.182	1.052
	6	1.093	1.349	19	1.192	0.688	32	0.948	-0.751	45	0.933	1.193
	7	1.276	-0.487	20	1.136	0.126	33	0.879	1.293	46	0.830	1.027
	8	0.960	0.122	21	0.875	0.315	34	0.955	-0.583	47	1.112	0.546
	9	0.944	1.319	22	0.791	1.717	35	1.134	1.623	48	0.969	-0.048
	10	1.065	0.307	23	1.049	1.227	36	1.025	0.864	49	1.179	-0.019
	11	1.257	-0.532	24	1.047	0.381	37	0.994	1.025	50	0.858	2.422

ตาราง ก-1 (ต่อ)

แบบสอบ (รหัส)	ข้อที่	A	B	ข้อที่	A	B	ข้อที่	A	B	ข้อที่	A	B
	12	1.312	-0.533	25	0.987	-0.169	38	1.536	-0.623			
	13	0.754	0.171	26	0.615	0.742	39	0.932	-0.265			
ฉบับที่ 10	1d*	1.172	1.684	14	0.689	1.065	27	1.081	0.125	40	1.081	0.125
(N_u10_k50d0.1nd10)	2d*	1.065	1.746	15	0.904	0.163	28	1.394	-0.011	41	1.394	-0.011
	3d*	0.692	0.391	16	0.785	0.482	29	1.059	-0.761	42	1.059	-0.761
	4d*	1.199	-0.034	17	1.156	2.042	30	0.725	-0.077	43	0.725	-0.077
	5d*	0.991	0.328	18	1.081	-1.517	31	0.980	-0.121	44	0.980	-0.121
	6d*	1.079	-0.163	19	1.222	-0.889	32	0.857	-0.535	45	0.857	-0.535
	7d*	0.639	1.458	20	1.035	-0.347	33	1.153	1.137	46	1.153	1.137
	8d*	1.182	-0.483	21	1.476	0.276	34	1.161	-1.057	47	1.161	-1.057
	9d*	0.922	1.418	22	0.714	1.879	35	1.312	-1.920	48	1.312	-1.920
	10d*	0.727	-0.224	23	1.296	0.416	36	1.091	0.025	49	1.091	0.025
	11	0.728	-0.708	24	1.109	-0.063	37	1.293	-0.444	50	1.293	-0.444
	12	0.587	1.571	25	0.796	0.363	38	1.323	1.321			
	13	1.099	1.373	26	0.944	-0.663	39	0.796	0.308			
ฉบับที่ 11	1d*	1.542	-0.369	14	0.785	-0.795	27	1.298	0.209	40	0.742	1.325
(N_u11_k50d0.2nd10)	2d*	0.863	-1.264	15	0.802	1.336	28	0.894	-0.458	41	1.174	-1.304
	3d*	0.807	1.484	16	1.161	-0.506	29	1.035	-0.563	42	0.885	0.108
	4d*	1.138	0.196	17	1.030	-0.543	30	1.339	0.413	43	0.750	1.306
	5d*	0.900	0.807	18	0.816	-1.443	31	1.130	-0.466	44	0.855	-0.389
	6d*	1.464	-0.580	19	0.929	1.465	32	0.904	0.115	45	1.184	0.843

ตาราง ก-1 (ต่อ)

แบบสอบ (รหัส)	ข้อที่	A	B	ข้อที่	A	B	ข้อที่	A	B	ข้อที่	A	B
	7d*	0.987	0.442	20	1.406	0.183	33	0.964	-2.127	46	1.200	1.329
	8d*	0.927	0.382	21	1.417	0.933	34	0.781	-0.129	47	1.050	-0.785
	9d*	0.950	-1.534	22	1.515	-0.843	35	1.181	0.110	48	0.930	-0.397
	10d*	1.108	0.056	23	1.071	-1.677	36	0.920	0.611	49	1.159	0.459
	11	1.148	-0.303	24	1.233	-0.156	37	1.001	0.998	50	1.199	0.372
	12	0.846	1.405	25	1.223	-1.324	38	0.937	0.290			
	13	0.893	1.391	26	0.907	-0.187	39	0.909	1.829			
ฉบับที่ 12	1d*	0.932	0.111	14	1.376	0.359	27	0.811	1.378	40	1.128	-0.122
(N_u12_k50d0.4nd10)	2d*	0.941	0.218	15	1.120	1.605	28	1.012	-2.089	41	0.949	0.372
	3d*	1.401	0.309	16	1.193	-1.947	29	0.702	-0.301	42	1.341	-0.557
	4d*	1.102	-0.088	17	1.019	-0.551	30	1.128	0.225	43	1.114	-0.555
	5d*	0.801	1.317	18	1.002	-0.464	31	1.047	-1.908	44	1.226	-1.032
	6d*	0.904	-1.430	19	0.925	-0.731	32	0.832	1.292	45	1.161	0.969
	7d*	1.270	-0.968	20	1.134	-0.731	33	0.965	0.285	46	0.914	-0.937
	8d*	0.909	-1.766	21	1.199	0.338	34	0.811	0.714	47	0.781	-0.529
	9d*	0.772	-0.571	22	1.088	0.523	35	0.813	0.281	48	0.771	0.989
	10d*	0.899	0.198	23	1.163	-0.743	36	1.036	-1.205	49	0.912	0.695
	11	1.035	-0.760	24	1.640	1.295	37	0.696	-0.376	50	0.926	-1.861
	12	1.549	0.573	25	1.014	-0.190	38	0.637	1.125			
	13	0.766	0.106	26	0.863	-1.936	39	1.096	0.310			
ฉบับที่ 13 (U1_k40d0.1nd4)	1d*	1.175	0.795	11	0.958	1.044	21	0.757	-1.843	31	1.013	-0.028

ตาราง ก-1 (ต่อ)

แบบสอบ (รหัส)	ข้อที่	A	B	ข้อที่	A	B	ข้อที่	A	B	ข้อที่	A	B
	2d*	1.334	-0.527	12	0.903	2.506	22	0.852	-0.918	32	1.033	0.407
	3d*	0.962	1.134	13	1.612	0.950	23	0.979	1.476	33	1.281	-1.507
	4d*	1.018	-0.062	14	0.863	-0.104	24	0.900	-0.318	34	1.152	0.712
	5	1.076	-0.301	15	0.997	1.491	25	0.742	-0.948	35	0.912	-0.754
	6	1.257	0.614	16	1.206	-0.897	26	0.876	-0.261	36	1.053	2.363
	7	0.912	2.059	17	0.845	-0.170	27	0.995	-0.114	37	1.104	0.666
	8	0.958	0.621	18	1.012	0.899	28	1.026	1.257	38	1.012	0.520
	9	0.907	1.502	19	1.205	1.742	29	0.903	-0.826	39	1.076	0.564
	10	0.932	1.905	20	1.117	-0.326	30	0.790	-0.904	40	0.996	-1.465
ฉบับที่ 14	1d*	0.794	0.123	11	0.985	-0.980	21	1.100	-0.937	31	1.018	-0.832
(U2_k40d0.2nd4)	2d*	0.818	-1.072	12	0.914	0.391	22	1.094	0.698	32	0.800	-1.383
	3d*	1.216	-2.272	13	1.102	-0.387	23	1.076	-0.152	33	1.075	0.130
	4d*	0.963	0.478	14	1.294	-0.086	24	1.147	0.502	34	1.159	0.136
	5	0.972	-0.132	15	1.095	-0.893	25	1.176	-1.622	35	0.705	0.119
	6	0.779	-1.663	16	1.150	-0.349	26	0.976	0.974	36	1.131	-0.053
	7	0.879	-0.116	17	0.939	-0.291	27	1.123	-0.029	37	1.090	-0.925
	8	0.914	1.077	18	1.110	1.118	28	1.375	-1.894	38	1.075	1.789
	9	0.964	-0.504	19	1.439	1.106	29	0.800	-0.166	39	1.267	0.751
	10	1.059	-1.077	20	0.922	-0.191	30	0.940	1.697	40	1.333	-0.304
ฉบับที่ 15	1d*	0.660	-1.778	11	0.660	-1.778	21	0.880	1.042	31	1.035	0.913
(U3_k40d0.4nd4)	2d*	0.856	-0.418	12	0.856	-0.418	22	0.808	0.157	32	1.015	-0.927

ตาราง ก-1 (ต่อ)

แบบสอบ (รหัส)	ข้อที่	A	B	ข้อที่	A	B	ข้อที่	A	B	ข้อที่	A	B
	3d*	0.830	-0.450	13	0.830	-0.450	23	0.869	0.030	33	1.060	1.959
	4d*	0.697	-1.105	14	0.697	-1.105	24	0.857	-1.463	34	1.341	0.041
	5	0.871	-0.504	15	0.871	-0.504	25	0.587	-0.905	35	1.012	0.532
	6	1.309	-1.570	16	1.309	-1.570	26	0.627	0.834	36	0.814	-1.189
	7	0.874	1.235	17	0.874	1.235	27	1.118	-0.906	37	1.701	-0.805
	8	0.970	-1.381	18	0.970	-1.381	28	1.018	-0.561	38	0.787	1.171
	9	0.896	-0.448	19	0.896	-0.448	29	0.873	-0.750	39	0.971	0.033
	10	0.825	0.183	20	0.825	0.183	30	1.078	-1.032	40	1.042	-0.122
ฉบับที่ 16	1d*	1.275	-0.739	11	1.131	1.039	21	1.183	-0.201	31	0.865	-1.387
(U4_k40d0.1nd8)	2d*	0.633	-1.064	12	0.965	-1.797	22	1.478	-0.445	32	0.893	0.735
	3d*	0.775	-1.234	13	1.009	-2.149	23	1.511	-0.079	33	1.090	1.123
	4d*	1.034	-0.688	14	1.069	1.844	24	0.861	1.276	34	0.961	0.459
	5d*	1.175	-1.509	15	1.064	0.348	25	1.058	-1.399	35	1.059	0.473
	6d*	0.869	0.298	16	0.898	-0.705	26	1.096	-1.123	36	1.037	-0.837
	7d*	0.858	0.060	17	1.279	0.605	27	0.980	-0.133	37	1.357	1.318
	8d*	1.035	0.262	18	0.953	-1.497	28	1.066	-0.287	38	0.769	-0.803
	9	1.001	-2.025	19	1.066	0.586	29	0.858	0.575	39	1.130	-0.695
	10	0.927	-0.510	20	0.810	0.951	30	1.158	0.506	40	0.888	-1.418
ฉบับที่ 17	1d*	1.443	-0.350	11	1.439	-0.885	21	1.343	-0.671	31	1.194	-0.125
(U5_k40d0.2nd8)	2d*	0.746	-0.448	12	1.152	-0.539	22	1.220	1.417	32	0.987	1.265
	3d*	1.046	-0.442	13	0.970	1.395	23	0.916	-0.851	33	0.731	1.927

ตาราง ก-1 (ต่อ)

แบบสอบ (รหัส)	ข้อที่	A	B	ข้อที่	A	B	ข้อที่	A	B	ข้อที่	A	B
	4d*	1.073	-0.204	14	1.001	-0.156	24	0.790	0.517	34	1.052	1.191
	5d*	1.125	-1.247	15	0.933	-0.568	25	0.849	0.852	35	1.088	0.615
	6d*	0.962	0.548	16	0.767	-1.470	26	1.121	-1.679	36	1.261	0.978
	7d*	1.223	-0.064	17	0.811	-0.143	27	1.026	0.550	37	0.894	0.660
	8d*	1.054	0.149	18	0.889	0.404	28	0.814	-1.257	38	0.794	0.073
	9	1.171	0.553	19	1.199	0.361	29	1.605	-0.224	39	1.253	0.495
	10	0.824	-0.820	20	1.156	-0.989	30	0.829	-1.548	40	1.013	1.150
ฉบับที่ 18	1d*	1.046	0.883	11	1.051	0.405	21	0.879	-0.004	31	1.483	2.843
(U6_k40d0.4nd8)	2d*	1.341	0.785	12	0.919	-0.783	22	1.117	0.268	32	1.193	0.349
	3d*	0.964	0.305	13	0.893	-2.133	23	0.879	0.928	33	0.742	-0.687
	4d*	1.017	-0.580	14	0.732	0.796	24	1.236	0.241	34	1.081	-0.188
	5d*	0.947	-0.187	15	0.931	-0.443	25	0.697	-0.118	35	1.125	-1.530
	6d*	0.910	-1.286	16	1.070	0.830	26	0.906	-0.519	36	0.886	-0.854
	7d*	1.046	0.784	17	1.091	-0.424	27	1.672	0.762	37	0.944	0.609
	8d*	1.408	0.148	18	1.136	-0.580	28	1.002	-1.563	38	1.009	0.427
	9	1.146	1.102	19	0.900	0.788	29	0.877	1.552	39	0.847	0.863
	10	1.409	0.511	20	0.871	-0.861	30	1.014	-0.146	40	0.950	1.139
ฉบับที่ 19	1d*	1.018	0.643	14	1.160	0.260	27	1.307	-1.334	40	0.702	-1.982
(U7_k50d0.1nd5)	2d*	0.951	-0.511	15	0.954	-0.447	28	1.039	-1.403	41	1.164	-1.204
	3d*	0.815	0.266	16	0.971	0.754	29	1.060	-0.256	42	0.873	0.454
	4d*	1.110	-0.633	17	1.237	-0.890	30	1.026	-0.352	43	1.227	-1.224

ตาราง ก-1 (ต่อ)

แบบสอบ (รหัส)	ข้อที่	A	B	ข้อที่	A	B	ข้อที่	A	B	ข้อที่	A	B
	5d*	1.443	0.006	18	0.958	1.289	31	0.958	0.924	44	0.696	-0.252
	6	0.518	-0.950	19	0.815	-0.383	32	0.963	-1.056	45	1.232	0.562
	7	1.314	-0.559	20	0.921	-1.158	33	1.138	1.103	46	0.983	-0.446
	8	1.180	0.257	21	0.846	0.049	34	1.047	0.058	47	1.195	0.759
	9	1.185	-0.294	22	1.220	1.243	35	1.128	-0.059	48	0.820	0.637
	10	0.878	0.295	23	0.874	-1.147	36	1.109	-1.833	49	1.034	-0.371
	11	0.952	-1.090	24	0.805	-1.043	37	1.040	-0.921	50	1.166	0.484
	12	0.679	0.642	25	1.018	-0.391	38	1.120	0.236			
	13	0.812	0.056	26	1.000	0.251	39	0.971	0.510			
ฉบับที่ 20	1d*	1.051	0.822	14	1.364	0.963	27	0.772	-0.047	40	1.012	-0.057
(U8_k50d0.2nd5)	2d*	1.065	-0.849	15	1.106	-0.199	28	1.234	-0.710	41	0.794	0.628
	3d*	1.105	-0.724	16	0.735	-0.842	29	1.081	-0.837	42	1.304	0.161
	4d*	1.153	0.511	17	1.164	-0.756	30	1.004	0.736	43	1.065	-0.591
	5d*	0.902	-1.032	18	0.714	0.424	31	0.917	0.808	44	0.804	0.629
	6	0.822	0.903	19	1.030	0.098	32	0.837	0.526	45	1.251	-0.326
	7	1.163	-0.502	20	0.798	-0.278	33	1.398	-0.416	46	0.741	1.571
	8	1.068	0.001	21	0.668	0.072	34	0.915	-0.469	47	0.998	-0.776
	9	1.090	-1.328	22	0.959	0.331	35	0.884	-0.979	48	0.838	0.734
	10	0.872	-0.370	23	0.820	-0.247	36	1.161	-1.100	49	1.553	0.111
	11	0.744	1.908	24	0.776	-0.440	37	0.836	0.871	50	1.130	-1.053
	12	0.760	3.459	25	1.083	0.232	38	1.125	0.197			

ตาราง ก-1 (ต่อ)

แบบสอบ (รหัส)	ข้อที่	A	B	ข้อที่	A	B	ข้อที่	A	B	ข้อที่	A	B
	13	0.935	-0.502	26	0.939	1.018	39	0.827	-0.609			
ฉบับที่ 21	1d*	0.958	-1.300	14	1.189	0.563	27	1.020	1.082	40	0.894	0.679
(U9_k50d0.4nd5)	2d*	0.983	0.953	15	0.969	0.748	28	0.901	0.272	41	1.413	-0.297
	3d*	1.279	0.514	16	1.135	1.282	29	1.121	-0.715	42	0.971	-0.618
	4d*	1.284	1.615	17	0.926	-0.545	30	0.861	0.471	43	1.155	-1.014
	5d*	0.885	0.316	18	0.902	-2.193	31	1.022	1.275	44	0.958	0.844
	6	0.936	0.188	19	0.910	0.978	32	0.769	-0.702	45	0.902	0.556
	7	0.918	0.669	20	1.120	-0.689	33	1.054	0.679	46	0.951	-0.291
	8	1.005	-1.189	21	1.402	1.056	34	0.784	-0.731	47	0.925	-1.262
	9	0.742	-2.018	22	0.720	-0.997	35	1.098	-0.617	48	1.041	0.349
	10	1.026	0.212	23	1.104	0.405	36	1.059	0.676	49	0.637	1.480
	11	0.986	-0.424	24	1.143	-0.457	37	0.662	-1.198	50	1.132	-0.983
	12	1.169	-0.073	25	0.791	0.571	38	1.013	0.020			
	13	0.980	-2.021	26	1.054	1.910	39	0.897	0.605			
ฉบับที่ 22	1d*	0.898	-0.386	14	1.188	-0.039	27	1.107	1.234	40	1.447	-0.883
(U10_k50d0.1nd10)	2d*	1.399	1.618	15	1.011	1.262	28	0.811	-0.353	41	1.271	-0.545
	3d*	1.774	-1.966	16	0.981	1.189	29	0.770	1.112	42	1.048	0.329
	4d*	0.812	0.719	17	0.849	0.566	30	1.052	0.044	43	1.001	-0.588
	5d*	1.226	0.454	18	1.440	-0.842	31	1.053	-0.057	44	0.808	0.951
	6d*	0.793	0.019	19	1.039	-0.332	32	1.122	0.867	45	0.955	0.905
	7d*	1.020	0.592	20	1.296	-0.345	33	1.275	1.239	46	1.283	-0.001

ตาราง ก-1 (ต่อ)

แบบสอบ (รหัส)	ข้อที่	A	B	ข้อที่	A	B	ข้อที่	A	B	ข้อที่	A	B
	8d*	0.730	-1.138	21	1.171	0.081	34	0.781	0.356	47	1.179	0.974
	9d*	1.054	-0.042	22	0.938	-1.467	35	0.983	0.331	48	1.244	1.006
	10d*	1.169	0.516	23	0.909	2.473	36	1.504	-0.290	49	0.850	-0.420
	11	1.254	-0.017	24	0.816	-0.610	37	1.104	-1.040	50	1.299	-0.272
	12	1.220	-1.346	25	1.228	-0.811	38	1.196	0.755			
	13	1.322	0.294	26	1.129	2.551	39	0.894	-0.130			
ฉบับที่ 23	1d*	0.751	-0.749	14	1.036	0.021	27	0.933	-0.372	40	0.871	-0.930
(U11_k50d0.2nd10)	2d*	1.054	-0.025	15	1.234	-1.741	28	1.181	-0.385	41	0.998	0.070
	3d*	1.481	0.341	16	1.266	2.635	29	1.176	1.275	42	1.088	-0.747
	4d*	0.884	1.165	17	0.829	0.271	30	0.771	-0.084	43	0.702	-0.464
	5d*	1.211	-0.222	18	0.808	0.233	31	1.505	1.507	44	1.253	0.467
	6d*	1.087	-0.403	19	1.574	-0.967	32	1.188	0.026	45	0.829	-1.105
	7d*	0.880	-0.236	20	1.129	0.163	33	0.700	0.152	46	1.017	-1.978
	8d*	1.143	-1.304	21	0.968	-0.008	34	0.987	0.304	47	1.425	-1.465
	9d*	1.074	-1.869	22	0.852	0.433	35	1.360	0.810	48	1.325	0.940
	10d*	0.736	0.244	23	0.892	0.034	36	1.046	-0.553	49	1.522	-0.368
	11	1.670	-1.034	24	1.020	-0.404	37	0.736	0.945	50	1.210	0.732
	12	1.091	0.856	25	0.855	-1.004	38	1.141	-0.072			
	13	0.801	0.302	26	1.008	1.169	39	0.824	-0.240			
ฉบับที่ 24	1d*	0.873	0.698	14	1.082	0.887	27	1.047	-0.731	40	1.140	-1.528
(U12_k50d0.4nd10)	2d*	0.817	-0.367	15	0.976	-1.108	28	1.118	-0.238	41	0.886	-1.050

ตาราง ก-1 (ต่อ)

แบบสอบ (รหัส)	ข้อที่	A	B	ข้อที่	A	B	ข้อที่	A	B	ข้อที่	A	B
	3d*	1.025	0.437	16	1.140	-2.861	29	0.872	-0.289	42	1.148	0.895
	4d*	1.013	-0.663	17	0.867	-0.791	30	0.785	-0.355	43	1.170	-1.395
	5d*	1.339	0.262	18	1.357	1.595	31	0.866	-0.599	44	1.009	-0.412
	6d*	1.320	-0.469	19	0.883	-0.977	32	0.946	0.255	45	0.840	-0.760
	7d*	1.351	-0.531	20	0.883	-0.557	33	1.305	0.668	46	0.735	0.245
	8d*	0.994	-0.482	21	0.949	0.557	34	1.157	0.491	47	0.936	-1.353
	9d*	1.388	-0.357	22	1.020	-0.982	35	0.826	0.873	48	0.954	1.084
	10d*	1.187	-1.354	23	1.068	0.065	36	1.089	-0.053	49	0.713	0.071
	11	1.101	-0.689	24	1.005	-0.581	37	0.860	-1.231	50	0.895	-1.140
	12	1.001	1.242	25	1.084	-0.429	38	1.542	0.279			
	13	1.326	-2.838	26	0.986	-2.961	39	0.879	0.704			

d* หมายถึง ข้อที่ถูกกำหนดให้ทำหน้าที่ต่างกัน

ตัวอย่างผลการวิเคราะห์ข้อมูล

WinGen Sysntax
24/6/2554 9:38:18
normal
1000
normal
0
1
2
2PLM
Lognormal
0
0.2
Normal
0
1
rep
25
C:\Documents and Settings\Compaq\My Documents\n1_1000k40d0.1nd4\n1.wgr

1 2	PLM	2 1.063 -0.890	21 2	PLM	2 0.743 -2.297
2 2	PLM	2 1.057 -0.720	22 2	PLM	2 1.273 0.949
3 2	PLM	2 0.786 -0.337	23 2	PLM	2 0.844 0.773
4 2	PLM	2 1.118 0.594	24 2	PLM	2 0.985 0.207
5 2	PLM	2 1.428 -2.346	25 2	PLM	2 1.070 -0.711
6 2	PLM	2 0.853 -0.151	26 2	PLM	2 0.877 -0.646
7 2	PLM	2 1.038 0.047	27 2	PLM	2 1.062 0.261
8 2	PLM	2 0.907 0.646	28 2	PLM	2 0.904 1.792
9 2	PLM	2 1.120 1.612	29 2	PLM	2 1.535 1.573
10 2	PLM	2 0.913 2.277	30 2	PLM	2 0.901 1.125
11 2	PLM	2 0.944 -1.754	31 2	PLM	2 1.626 0.001
12 2	PLM	2 1.003 1.093	32 2	PLM	2 1.126 -0.188
13 2	PLM	2 1.030 0.431	33 2	PLM	2 0.976 1.840
14 2	PLM	2 0.751 -0.060	34 2	PLM	2 1.198 -1.518
15 2	PLM	2 0.944 -0.589	35 2	PLM	2 1.657 -0.303
16 2	PLM	2 0.806 0.847	36 2	PLM	2 1.528 0.443
17 2	PLM	2 0.783 -0.326	37 2	PLM	2 0.990 0.847
18 2	PLM	2 1.083 1.566	38 2	PLM	2 0.868 1.288
19 2	PLM	2 1.158 0.093	39 2	PLM	2 0.836 -1.607
20 2	PLM	2 1.009 -0.999	40 2	PLM	2 1.058 -0.788

ภาคผนวก ข

<u>ตัวอย่าง</u> ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติกโดยการ ทดสอบระดับนัยสำคัญ (significance test) และการวัดขนาดอิทธิพล (measure of effect size) ภายใต้ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory model) ชนิด 2 พารามิเตอร์ (two-parameter)

ภาคผนวก ข

ตัวอย่าง ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติกโดยการ ทดสอบระดับนัยสำคัญ (significance test) และการวัดขนาดอิทธิพล (measure of effect size) ภายใต้ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory model) ชนิด 2 พารามิเตอร์ (two-parameter)

ตอนที่ ข-1 ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก กรณี DIF แบบอเนกรูป, ความยาวของแบบสอบทั้งฉบับเป็น 40 ข้อ, ขนาด DIF 0.1, 10 %DIF

1> difLogistic(c1c1[,1:40],group=c1c1[,41],focal.name=2) Detection of both types of Differential Item Functioning using Logistic regression method, without item purification and with LRT DIF statistic Logistic regression DIF statistic:

Stat.	P-value									
X1	3.0730	0.2151	X1.10	3.3781	0.1847	X1.21	0.8250	0.6620	X1.32	9.9745	0.0068 **
X1.1	0.9522	0.6212	X1.11	7.4912	0.0236 *	X1.22	0.0218	0.9892	X1.33	0.1661	0.9203
X1.2	1.1975	0.5495	X1.12	0.4768	0.7879	X1.23	2.9314	0.2309	X1.34	0.2011	0.9043
X1.3	3.0964	0.2126	X1.13	0.9520	0.6213	X1.24	1.4404	0.4866	X1.35	0.6022	0.7400
X1.4	0.5900	0.7445	X1.14	6.4431	0.0399 *	X1.25	5.3658	0.0684 .	X1.36	0.9154	0.6327
X1.5	2.8541	0.2400	X1.15	0.5850	0.7464	X1.26	0.6613	0.7184	X1.37	3.7088	0.1565
X1.6	2.0713	0.3550	X1.16	1.1786	0.5547	X1.27	0.4122	0.8137	X1.38	1.7527	0.4163
X1.7	1.5678	0.4566	X1.17	1.7035	0.4267	X1.28	3.7065	0.1567	X1.39	0.7743	0.6790
X1.8	0.2362	0.8886	X1.18	0.5133	0.7736	X1.29	1.1373	0.5663			
X1.9	1.1936	0.5506	X1.19	2.4975	0.2869	X1.30	2.9033	0.2342			
			X1.20	1.9143	0.3840	X1.31	0.7316	0.6936			

Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Detection threshold: 5.9915
(significance level: 0.05)
Items detected as DIF items:
X1.11
X1.14
X1.32

Effect size (Nagelkerke's R^2): Effect size code: 'A': negligible effect 'B': moderate effect 'C': large effect

R^2	ZT	JG										
X1	0.0021	A	A	X1.9	0.0034	A	A	X1.19	0.0027	A	A	X1.29	0.0015	A	A	X1.39	0.0008	A	A
X1.1	0.0006	A	A	X1.10	0.0059	A	A	X1.20	0.0051	A	A	X1.30	0.0000	A	A				
X1.2	0.0008	A	A	X1.11	0.0078	A	A	X1.21	0.0007	A	A	X1.31	0.0006	A	A				
X1.3	0.0020	A	A	X1.12	0.0004	A	A	X1.22	0.0000	A	A	X1.32	0.0189	A	A				
X1.4	0.0019	A	A	X1.13	0.0012	A	A	X1.23	0.0029	A	A	X1.33	0.0002	A	A				
X1.5	0.0028	A	A	X1.14	0.0059	A	A	X1.24	0.0013	A	A	X1.34	0.0001	A	A				
X1.6	0.0018	A	A	X1.15	0.0008	A	A	X1.25	0.0062	A	A	X1.35	0.0004	A	A				
X1.7	0.0018	A	A	X1.16	0.0013	A	A	X1.26	0.0006	A	A	X1.36	0.0011	A	A				
X1.8	0.0004	A	A	X1.17	0.0024	A	A	X1.27	0.0008	A	A	X1.37	0.0060	A	A				
				X1.18	0.0004	A	A	X1.28	0.0046	A	A	X1.38	0.0036	A	A				

Effect size codes:
Zumbo & Thomas (ZT):
0 'A' 0.13 'B' 0.26 'C' 1
Jodoign & Gierl (JG):
0 'A' 0.035 'B' 0.07 'C' 1

2> difLogistic(c1c2[,1:40],group=c1c2[,41],focal.name=2) Detection of both types of Differential Item Functioning using Logistic regression method, without item purification and with LRT DIF statistic Logistic regression DIF statistic:

Stat.	P-value									
X1	2.2928	0.3178	X1.11	5.5931	0.0610 .	X1.23	1.9926	0.3692	X1.35	0.4293	0.8068
X1.1	3.7108	0.1564	X1.12	0.7258	0.6957	X1.24	1.0453	0.5929	X1.36	0.7058	0.7027
X1.2	0.4866	0.7840	X1.13	1.9848	0.3707	X1.25	8.3856	0.0151 *	X1.37	0.2247	0.8938
X1.3	1.3376	0.5123	X1.14	1.5380	0.4635	X1.26	0.7969	0.6714	X1.38	0.1297	0.9372
X1.4	7.3180	0.0258 *	X1.15	9.0931	0.0106 *	X1.27	0.6653	0.7170	X1.39	0.9626	0.6180
X1.5	0.0653	0.9679	X1.16	0.7957	0.6718	X1.28	4.5180	0.1045			
X1.6	4.9530	0.0840 .	X1.17	1.8583	0.3949	X1.29	6.6389	0.0362 *			
X1.7	3.6887	0.1581	X1.18	2.2007	0.3328	X1.30	1.8956	0.3876			
X1.8	4.3510	0.1135	X1.19	8.6836	0.0130 *	X1.31	4.1433	0.1260			
X1.9	0.4884	0.7833	X1.20	1.9449	0.3782	X1.32	2.4246	0.2975			
X1.10	5.6936	0.0580 .	X1.21	1.4142	0.4931	X1.33	4.6664	0.0970 .			
			X1.22	3.8269	0.1476	X1.34	0.1952	0.9070			

Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Detection threshold: 5.9915
(significance level: 0.05)
Items detected as DIF items:
X1.4
X1.15
X1.19
X1.25
X1.29

Effect size (Nagelkerke's R^2): Effect size code: 'A': negligible effect 'B': moderate effect 'C': large effect

R^2	ZT	JG										
X1	0.0015	A	A	X1.9	0.0013	A	A	X1.19	0.0114	A	A	X1.29	0.0088	A	A	X1.39	0.0009	A	A
X1.1	0.0024	A	A	X1.10	0.0098	A	A	X1.20	0.0052	A	A	X1.30	0.0010	A	A				
X1.2	0.0003	A	A	X1.11	0.0070	A	A	X1.21	0.0013	A	A	X1.31	0.0032	A	A				
X1.3	0.0008	A	A	X1.12	0.0007	A	A	X1.22	0.0047	A	A	X1.32	0.0044	A	A				
X1.4	0.0228	A	A	X1.13	0.0023	A	A	X1.23	0.0019	A	A	X1.33	0.0069	A	A				
X1.5	0.0001	A	A	X1.14	0.0016	A	A	X1.24	0.0010	A	A	X1.34	0.0001	A	A				
X1.6	0.0041	A	A	X1.15	0.0114	A	A	X1.25	0.0083	A	A	X1.35	0.0003	A	A				
X1.7	0.0036	A	A	X1.16	0.0009	A	A	X1.26	0.0007	A	A	X1.36	0.0007	A	A				
X1.8	0.0069	A	A	X1.17	0.0032	A	A	X1.27	0.0014	A	A	X1.37	0.0003	A	A				
				X1.18	0.0016	A	A	X1.28	0.0063	A	A	X1.38	0.0002	A	A				

Effect size codes:
Zumbo & Thomas (ZT):
0 'A' 0.13 'B' 0.26 'C' 1
Jodoign & Gierl (JG):
0 'A' 0.035 'B' 0.07 'C' 1

3> difLogistic(c1c3[,1:40],group=c1c3[,41],focal.name=2) Detection of both types of Differential Item Functioning using Logistic regression method, without item purification and with LRT DIF statistic Logistic regression DIF statistic:

```
     Stat.  P-value
X1    2.0419 0.3603
X1.1  2.9672 0.2268
X1.2  0.9033 0.6366
X1.3  2.3830 0.3038
X1.4  0.5376 0.7643
X1.5  7.2085 0.0272  *
X1.6  2.9265 0.2315
X1.7  0.6825 0.7109
X1.8  3.4636 0.1770
X1.9  5.3130 0.0702  .
X1.10 2.6970 0.2596
X1.11 1.6266 0.4434
X1.12 0.1347 0.9349
X1.13 1.9839 0.3709
X1.14 1.8161 0.4033
X1.15 2.7056 0.2585
X1.16 0.4271 0.8077
X1.17 2.4728 0.2904
X1.18 1.3967 0.4974
X1.19 0.1098 0.9466
X1.20 0.7076 0.7020
X1.21 0.0949 0.9537
X1.22 3.6756 0.1592
X1.23 2.8358 0.2422
X1.24 1.5184 0.4680
X1.25 0.5675 0.7529
X1.26 1.4757 0.4781
X1.27 2.9354 0.2305
X1.28 0.0738 0.9638
X0    2.5156 0.2843
X1.29 3.1201 0.2101
X1.30 2.7191 0.2568
X1.31 4.9036 0.0861  .
X1.32 4.8412 0.0889  .
X1.33 0.3967 0.8201
X1.34 0.7424 0.6899
X1.35 1.3565 0.5075
X1.36 1.9765 0.3722
X1.37 8.1264 0.0172  *
X1.38 0.1549 0.9255
```

Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Detection threshold: 5.9915
(significance level: 0.05)
Items detected as DIF items:
X1.5
X1.37

Effect size (Nagelkerke's R^2): Effect size code: 'A': negligible effect 'B': moderate effect 'C': large effect

```
     R^2  ZT JG
X1    0.0013 A  A
X1.1  0.0020 A  A
X1.2  0.0006 A  A
X1.3  0.0015 A  A
X1.4  0.0017 A  A
X1.5  0.0076 A  A
X1.6  0.0026 A  A
X1.7  0.0007 A  A
X1.8  0.0050 A  A
X1.9  0.0151 A  A
X1.10 0.0056 A  A
X1.11 0.0018 A  A
X1.12 0.0001 A  A
X1.13 0.0025 A  A
X1.14 0.0020 A  A
X1.15 0.0038 A  A
X1.16 0.0005 A  A
X1.17 0.0034 A  A
X1.18 0.0011 A  A
X1.19 0.0001 A  A
X1.20 0.0019 A  A
X1.21 0.0001 A  A
X1.22 0.0043 A  A
X1.23 0.0027 A  A
X1.24 0.0014 A  A
X1.25 0.0006 A  A
X1.26 0.0011 A  A
X1.27 0.0047 A  A
X1.28 0.0001 A  A
X0    0.0040 A  A
X1.29 0.0016 A  A
X1.30 0.0019 A  A
X1.31 0.0082 A  A
X1.32 0.0059 A  A
X1.33 0.0002 A  A
X1.34 0.0004 A  A
X1.35 0.0013 A  A
X1.36 0.0034 A  A
X1.37 0.0153 A  A
X1.38 0.0001 A  A
```

Effect size codes:
Zumbo & Thomas (ZT):
0 'A' 0.13 'B' 0.26 'C' 1
Jodoign & Gierl (JG):
0 'A' 0.035 'B' 0.07 'C' 1

4> difLogistic(c1c4[,1:40],group=c1c4[,41],focal.name=2) Detection of both types of Differential Item Functioning using Logistic regression method, without item purification and with LRT DIF statistic Logistic regression DIF statistic:

```
     Stat.  P-value
X1    1.7072 0.4259
X1.1  2.3383 0.3106
X1.2  1.2259 0.5418
X1.3  1.2404 0.5378
X1.4  0.1820 0.9130
X1.5  5.2986 0.0707  .
X1.6  2.3059 0.3157
X1.7  0.3201 0.8521
X0    0.5803 0.7482
X1.8  3.7917 0.1502
X1.9  0.5199 0.7711
X1.10 0.5875 0.7455
X1.11 1.6817 0.4313
X1.12 1.2581 0.5331
X1.13 2.0396 0.3607
X1.14 1.7747 0.4118
X1.15 0.3254 0.8498
X1.16 1.1584 0.5603
X1.17 3.5439 0.1700
X1.18 3.3936 0.1833
X1.19 2.9067 0.2338
X1.20 0.1192 0.9422
X1.21 4.8545 0.0883  .
X1.22 0.1501 0.9277
X1.23 1.9402 0.3790
X1.24 0.8285 0.6608
X1.25 0.7527 0.6864
X1.26 0.7964 0.6715
X1.27 0.4623 0.7936
X1.28 6.9420 0.0311  *
X1.29 7.4436 0.0242  *
X1.30 0.1774 0.9151
X0.1  2.1683 0.3382
X1.31 1.0428 0.5937
X1.32 0.2925 0.8639
X1.33 1.2398 0.5380
X1.34 8.2271 0.0163  *
X1.35 1.7177 0.4236
X1.36 6.0676 0.0481  *
X1.37 5.8039 0.0549  .
```

Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Detection threshold: 5.9915
(significance level: 0.05)
Items detected as DIF items:
X1.28
X1.29
X1.34
X1.36

Effect size (Nagelkerke's R^2): Effect size code: 'A': negligible effect 'B': moderate effect 'C': large effect

```
     R^2  ZT JG
X1    0.0012 A  A
X1.1  0.0015 A  A
X1.2  0.0008 A  A
X1.3  0.0008 A  A
X1.4  0.0006 A  A
X1.5  0.0056 A  A
X1.6  0.0019 A  A
X1.7  0.0004 A  A
X0    0.0008 A  A
X1.8  0.0111 A  A
X1.9  0.0009 A  A
X1.10 0.0006 A  A
X1.11 0.0014 A  A
X1.12 0.0014 A  A
X1.13 0.0021 A  A
X1.14 0.0024 A  A
X1.15 0.0004 A  A
X1.16 0.0018 A  A
X1.17 0.0024 A  A
X1.18 0.0034 A  A
X1.19 0.0072 A  A
X1.20 0.0001 A  A
X1.21 0.0058 A  A
X1.22 0.0001 A  A
X1.23 0.0016 A  A
X1.24 0.0009 A  A
X1.25 0.0006 A  A
X1.26 0.0016 A  A
X1.27 0.0006 A  A
X1.28 0.0092 A  A
X1.29 0.0038 A  A
X1.30 0.0001 A  A
X0.1  0.0043 A  A
X1.31 0.0014 A  A
X1.32 0.0002 A  A
X1.33 0.0008 A  A
X1.34 0.0082 A  A
X1.35 0.0023 A  A
X1.36 0.0124 A  A
X1.37 0.0057 A  A
```

Effect size codes:
Zumbo & Thomas (ZT):
0 'A' 0.13 'B' 0.26 'C' 1
Jodoign & Gierl (JG):
0 'A' 0.035 'B' 0.07 'C' 1

5> Detection of both types of Differential Item Functioning using Logistic regression method, without item purification and with LRT DIF statistic

Logistic regression DIF statistic:

	Stat.	P-value										
X1	4.9562	0.0839	.	X1.10	0.1597	0.9233	X1.21	0.4295	0.8068	X1.32	3.2841	0.1936
X1.1	0.6622	0.7181		X1.11	1.5748	0.4550	X1.22	4.5077	0.1050	X1.33	1.3857	0.5001
X1.2	3.1463	0.2074		X1.12	0.1934	0.9078	X1.23	3.5055	0.1733	X1.34	0.3063	0.8580
X1.3	0.3285	0.8485		X1.13	0.6582	0.7196	X1.24	0.6039	0.7394	X1.35	1.1783	0.5548
X1.4	0.0633	0.9688		X1.14	0.4617	0.7939	X1.25	6.1739	0.0456 *	X1.36	1.8054	0.4055
X1.5	0.7099	0.7012		X1.15	0.3388	0.8442	X1.26	2.1582	0.3399	X1.37	4.3627	0.1129
X1.6	0.4599	0.7946		X1.16	5.0272	0.0810 .	X1.27	0.9942	0.6083	X1.38	0.6755	0.7134
X1.7	1.4640	0.4810		X1.17	3.5025	0.1736	X1.28	0.3851	0.8248	X1.39	0.5519	0.7588
X1.8	15.5313	0.0004 ***		X1.18	0.6410	0.7258	X1.29	2.2104	0.3311			
X1.9	7.5670	0.0227 *		X1.19	2.8808	0.2368	X1.30	0.9697	0.6158			
				X1.20	2.6003	0.2725	X1.31	2.6078	0.2715			

Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Detection threshold: 5.9915
(significance level: 0.05)

Items detected as DIF items:
X1.8
X1.9
X1.25

Effect size (Nagelkerke's R^2): Effect size code: 'A': negligible effect 'B': moderate effect 'C': large effect

	R^2	ZT	JG											
X1	0.0034	A	A	X1.9	0.0269	A	A	X1.19	0.0030	A	A	X1.29	0.0030	A A
X1.1	0.0004	A	A	X1.10	0.0003	A	A	X1.20	0.0067	A	A	X1.30	0.0005	A A
X1.2	0.0021	A	A	X1.11	0.0018	A	A	X1.21	0.0004	A	A	X1.31	0.0023	A A
X1.3	0.0002	A	A	X1.12	0.0002	A	A	X1.22	0.0050	A	A	X1.32	0.0062	A A
X1.4	0.0002	A	A	X1.13	0.0009	A	A	X1.23	0.0028	A	A	X1.33	0.0022	A A
X1.5	0.0008	A	A	X1.14	0.0005	A	A	X1.24	0.0005	A	A	X1.34	0.0002	A A
X1.6	0.0004	A	A	X1.15	0.0005	A	A	X1.25	0.0068	A	A	X1.35	0.0007	A A
X1.7	0.0014	A	A	X1.16	0.0057	A	A	X1.26	0.0019	A	A	X1.36	0.0020	A A
X1.8	0.0238	A	A	X1.17	0.0048	A	A	X1.27	0.0022	A	A	X1.37	0.0067	A A
				X1.18	0.0005	A	A	X1.28	0.0005	A	A	X1.38	0.0012	A A

X1.39 0.0006 A A

Effect size codes:
Zumbo & Thomas (ZT):
0 'A' 0.13 'B' 0.26 'C' 1
Jodoign & Gierl (JG):
0 'A' 0.035 'B' 0.07 'C' 1

6>Detection of both types of Differential Item Functioning using Logistic regression method, without item purification and with LRT DIF statistic

Logistic regression DIF statistic:

	Stat.	P-value										
X1.1	0.4864	0.7841		X0	0.9076	0.6352	X1.19	1.1960	0.5499	X1.29	0.2835	0.8679
X1.2	5.8851	0.0527	.	X1.10	1.7003	0.4273	X1.20	5.1224	0.0772 .	X1.30	1.8678	0.3930
X1.3	0.3726	0.8300		X1.11	0.1889	0.9099	X1.21	1.7000	0.4274	X1.31	3.0195	0.2210
X1.4	1.8017	0.4062		X1.12	3.8568	0.1454	X1.22	0.7982	0.6709	X1.32	0.8135	0.6658
X1.5	1.1609	0.5597		X1.13	0.0961	0.9531	X1.23	1.4747	0.4784	X1.33	2.9549	0.2282
X1.6	0.4951	0.7807		X1.14	1.6893	0.4297	X1.24	1.3039	0.5210	X1.34	0.3444	0.8418
X1.7	1.0266	0.5985		X1.15	2.8752	0.2375	X1.25	1.3440	0.5107	X1.35	4.2205	0.1212
X1.8	2.8447	0.2411		X1.16	0.6434	0.7249	X1.26	0.5027	0.7778	X1.36	0.4383	0.8032
X1.9	2.4520	0.2935		X1.17	5.1880	0.0747 .	X1.27	1.0670	0.5865	X1.37	0.1617	0.9223
				X1.18	1.1198	0.5713	X1.28	0.3399	0.8437	X1.38	2.6654	0.2638

Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Detection threshold: 5.9915
(significance level: 0.05)

Items detected as DIF items:
No DIF item detected

Effect size (Nagelkerke's R^2): Effect size code: 'A': negligible effect 'B': moderate effect 'C': large effect

	R^2	ZT	JG											
X1.1	0.0003	A	A	X0	0.0028	A	A	X1.19	0.0014	A	A	X1.29	0.0004	A A
X1.2	0.0037	A	A	X1.10	0.0034	A	A	X1.20	0.0138	A	A	X1.30	0.0009	A A
X1.3	0.0003	A	A	X1.11	0.0002	A	A	X1.21	0.0015	A	A	X1.31	0.0024	A A
X1.4	0.0012	A	A	X1.12	0.0035	A	A	X1.22	0.0011	A	A	X1.32	0.0019	A A
X1.5	0.0036	A	A	X1.13	0.0001	A	A	X1.23	0.0015	A	A	X1.33	0.0043	A A
X1.6	0.0005	A	A	X1.14	0.0017	A	A	X1.24	0.0012	A	A	X1.34	0.0002	A A
X1.7	0.0009	A	A	X1.15	0.0043	A	A	X1.25	0.0016	A	A	X1.35	0.0026	A A
X1.8	0.0032	A	A	X1.16	0.0008	A	A	X1.26	0.0005	A	A	X1.36	0.0004	A A
X1.9	0.0034	A	A	X1.17	0.0086	A	A	X1.27	0.0020	A	A	X1.37	0.0002	A A
				X1.18	0.0008	A	A	X1.28	0.0005	A	A	X1.38	0.0044	A A

X1.39 0.0022 A A

Effect size codes:
Zumbo & Thomas (ZT):
0 'A' 0.13 'B' 0.26 'C' 1
Jodoign & Gierl (JG):
0 'A' 0.035 'B' 0.07 'C' 1

7> Detection of both types of Differential Item Functioning using Logistic regression method, without item purification and with LRT DIF statistic

Logistic regression DIF statistic:

	Stat.	P-value													
X1.1	1.3422	0.5112	X1.13	0.9545	0.6205	X1.26	4.4120	0.1101	X1.39	5.0541	0.0799	.
X1.2	11.1674	0.0038	**	X1.14	1.6175	0.4454	X1.27	0.3884	0.8235	X1.40	1.2162	0.5444
X1.3	3.5751	0.1674		X1.15	5.3750	0.0681	.	X1.28	0.8722	0.6466			
X1.4	0.5829	0.7472		X1.16	1.1281	0.5689		X1.29	2.9961	0.2236			
X1.5	8.5982	0.0136	*	X1.17	3.0014	0.2230		X1.30	5.6282	0.0600	.		
X1.6	1.1809	0.5541		X1.18	1.1621	0.5593		X1.31	1.2547	0.5340			
X1.7	0.7889	0.6740		X1.19	9.0707	0.0107	*	X1.32	2.4289	0.2969			
X1.8	0.1668	0.9200		X1.20	2.7344	0.2548		X1.33	0.4129	0.8135			
X1.9	7.5746	0.0227	*	X1.21	2.5863	0.2744		X1.34	1.6582	0.4364			
X1.10	5.7681	0.0559	.	X1.22	1.1449	0.5641		X1.35	1.4408	0.4866			
X1.11	6.5361	0.0381	*	X1.23	2.4987	0.2867		X1.36	8.0985	0.0174	*		
X1.12	2.2879	0.3186		X1.24	3.8291	0.1474		X1.37	1.4301	0.4892			
				X1.25	4.3286	0.1148		X1.38	3.6437	0.1617			

Signif. codes:

0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Detection threshold: 5.9915

(significance level: 0.05)

Items detected as DIF items:

X1.2
X1.5
X1.9
X1.11
X1.19
X1.36

Effect size (Nagelkerke's R^2): Effect size code: 'A': negligible effect 'B': moderate effect 'C': large effect

	R^2	ZT	JG																
X1.1	0.0009	A	A	X1.10	0.0155	A	A	X1.20	0.0028	A	A	X1.30	0.0074	A	A	X1.40	0.0013	A	A
X1.2	0.0076	A	A	X1.11	0.0110	A	A	X1.21	0.0063	A	A	X1.31	0.0007	A	A				
X1.3	0.0023	A	A	X1.12	0.0028	A	A	X1.22	0.0010	A	A	X1.32	0.0020	A	A				
X1.4	0.0004	A	A	X1.13	0.0009	A	A	X1.23	0.0030	A	A	X1.33	0.0008	A	A				
X1.5	0.0295	A	A	X1.14	0.0019	A	A	X1.24	0.0035	A	A	X1.34	0.0025	A	A				
X1.6	0.0011	A	A	X1.15	0.0052	A	A	X1.25	0.0043	A	A	X1.35	0.0008	A	A				
X1.7	0.0006	A	A	X1.16	0.0012	A	A	X1.26	0.0049	A	A	X1.36	0.0048	A	A				
X1.8	0.0002	A	A	X1.17	0.0035	A	A	X1.27	0.0003	A	A	X1.37	0.0016	A	A				
X1.9	0.0122	A	A	X1.18	0.0017	A	A	X1.28	0.0019	A	A	X1.38	0.0048	A	A				
				X1.19	0.0067	A	A	X1.29	0.0038	A	A	X1.39	0.0094	A	A				

Effect size codes:

Zumbo & Thomas (ZT):
0 'A' 0.13 'B' 0.26 'C' 1
Jodoign & Gierl (JG):
0 'A' 0.035 'B' 0.07 'C' 1

8> Detection of both types of Differential Item Functioning using Logistic regression method, without item purification and with LRT DIF statistic

Logistic regression DIF statistic:

	Stat.	P-value													
X1.1	2.1273	0.3452	X1.11	2.1411	0.3428	X1.22	2.6417	0.2669	X1.33	1.0707	0.5855	
X1.2	3.2700	0.1950	X1.12	0.9835	0.6116	X1.23	1.9740	0.3727	X1.34	6.1047	0.0472	*
X1.3	3.7962	0.1499	X1.13	3.5801	0.1670	X1.24	0.7972	0.6712	X1.35	0.3152	0.8542	
X1.4	0.1716	0.9178	X1.14	1.9060	0.3856	X1.25	3.6253	0.1632	X1.36	0.6949	0.7065	
X1.5	6.6610	0.0358	*	X1.15	4.6171	0.0994	.	X1.26	0.4322	0.8056	X1.37	5.3832	0.0678	.
X1.6	3.5051	0.1733	X1.16	0.5500	0.7596	X1.27	2.9380	0.2302	X1.38	3.6324	0.1626	
X1.7	0.7903	0.6736	X1.17	1.2792	0.5275	X1.28	0.3332	0.8466	X1.39	0.1149	0.9442	
X1.8	0.8073	0.6679	X1.18	2.2687	0.3216	X1.29	0.6809	0.7114	X1.40	0.9461	0.6231	
X1.9	3.5145	0.1725	X1.19	0.1840	0.9121	X1.30	0.7034	0.7035				
X1.10	9.5391	0.0085	**	X1.20	1.3560	0.5076	X1.31	1.5795	0.4539			
				X1.21	1.8080	0.4049	X1.32	5.2411	0.0728	.		

Signif. codes:

0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Detection threshold: 5.9915

(significance level: 0.05)

Items detected as DIF items:

X1.5
X1.10
X1.34

Effect size (Nagelkerke's R^2): Effect size code: 'A': negligible effect 'B': moderate effect 'C': large effect

	R^2	ZT	JG																
X1.1	0.0013	A	A	X1.10	0.0223	A	A	X1.20	0.0017	A	A	X1.30	0.0008	A	A	X1.40	0.0009	A	A
X1.2	0.0021	A	A	X1.11	0.0042	A	A	X1.21	0.0052	A	A	X1.31	0.0008	A	A				
X1.3	0.0023	A	A	X1.12	0.0012	A	A	X1.22	0.0022	A	A	X1.32	0.0043	A	A				
X1.4	0.0001	A	A	X1.13	0.0033	A	A	X1.23	0.0024	A	A	X1.33	0.0021	A	A				
X1.5	0.0207	A	A	X1.14	0.0018	A	A	X1.24	0.0007	A	A	X1.34	0.0087	A	A				
X1.6	0.0037	A	A	X1.15	0.0046	A	A	X1.25	0.0035	A	A	X1.35	0.0002	A	A				
X1.7	0.0007	A	A	X1.16	0.0006	A	A	X1.26	0.0005	A	A	X1.36	0.0004	A	A				
X1.8	0.0009	A	A	X1.17	0.0016	A	A	X1.27	0.0024	A	A	X1.37	0.0054	A	A				
X1.9	0.0048	A	A	X1.18	0.0031	A	A	X1.28	0.0006	A	A	X1.38	0.0047	A	A				
				X1.19	0.0001	A	A	X1.29	0.0009	A	A	X1.39	0.0002	A	A				

Effect size codes:

Zumbo & Thomas (ZT):
0 'A' 0.13 'B' 0.26 'C' 1
Jodoign & Gierl (JG):
0 'A' 0.035 'B' 0.07 'C' 1

9>Detection of both types of Differential Item Functioning using Logistic regression method, without item purification and with LRT DIF statistic

Logistic regression DIF statistic:

	Stat.	P-value										
X1.1	0.0142	0.9929	X1.10	10.2431	0.0060	**	X1.21	0.6643	0.7174	X1.31	3.1848	0.2034
X1.2	1.0314	0.5971	X1.11	4.2439	0.1198		X1.22	1.6078	0.4476	X1.32	0.3732	0.8298
X1.3	1.1808	0.5541	X1.12	6.5995	0.0369	*	X1.23	1.7055	0.4262	X1.33	4.4137	0.1100
X1.4	0.2688	0.8742	X1.13	2.8102	0.2453		X1.24	3.8851	0.1433	X1.34	4.8538	0.0883 .
X1.5	0.1941	0.9075	X1.14	0.3783	0.8277		X1.25	2.0188	0.3644	X1.35	1.0300	0.5975
X1.6	0.9633	0.6178	X1.15	0.4372	0.8036		X1.26	0.6144	0.7355	X1.36	3.2680	0.1951
X1.7	0.5979	0.7416	X1.16	3.4062	0.1821		X1.27	7.7821	0.0204	*	X1.37	0.0942	0.9540
X1.8	0.7845	0.6755	X1.17	1.7289	0.4213		X0.1	0.0609	0.9700	X1.38	1.1288	0.5687
X1.9	2.3577	0.3076	X1.18	2.6263	0.2690		X1.28	3.0097	0.2220			
X0	4.7399	0.0935 .	X1.19	3.1036	0.2119		X1.29	0.2887	0.8656			
			X1.20	2.8254	0.2435		X1.30	0.3527	0.8383			

Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Detection threshold: 5.9915
(significance level: 0.05)
Items detected as DIF items:
X1.10
X1.12
X1.27

Effect size (Nagelkerke's R^2): Effect size code: 'A': negligible effect 'B': moderate effect 'C': large effect

	R^2	ZT	JG												
X1.1	0.0000	A	A	X0	0.0124	A	A	X1.19	0.0032	A	A	X1.28	0.0037	A	A	X1.38	0.0011	A	A
X1.2	0.0006	A	A	X1.10	0.0181	A	A	X1.20	0.0081	A	A	X1.29	0.0002	A	A				
X1.3	0.0008	A	A	X1.11	0.0046	A	A	X1.21	0.0006	A	A	X1.30	0.0003	A	A				
X1.4	0.0002	A	A	X1.12	0.0054	A	A	X1.22	0.0019	A	A	X1.31	0.0064	A	A				
X1.5	0.0006	A	A	X1.13	0.0034	A	A	X1.23	0.0016	A	A	X1.32	0.0005	A	A				
X1.6	0.0009	A	A	X1.14	0.0004	A	A	X1.24	0.0036	A	A	X1.33	0.0025	A	A				
X1.7	0.0005	A	A	X1.15	0.0006	A	A	X1.25	0.0021	A	A	X1.34	0.0030	A	A				
X1.8	0.0008	A	A	X1.16	0.0040	A	A	X1.26	0.0005	A	A	X1.35	0.0010	A	A				
X1.9	0.0036	A	A	X1.17	0.0024	A	A	X1.27	0.0147	A	A	X1.36	0.0051	A	A				
				X1.18	0.0019	A	A	X0.1	0.0001	A	A	X1.37	0.0002	A	A				

Effect size codes:
Zumbo & Thomas (ZT):
0 'A' 0.13 'B' 0.26 'C' 1
Jodoign & Gierl (JG):
0 'A' 0.035 'B' 0.07 'C' 1

10> Detection of both types of Differential Item Functioning using Logistic regression method, without item purification and with LRT DIF statistic

Logistic regression DIF statistic:

	Stat.	P-value										
X1.1	0.2254	0.8934	X1.10	0.0829	0.9594		X1.21	1.5019	0.4719	X1.31	0.8921	0.6401
X1.2	4.2468	0.1196	X1.11	2.5139	0.2845		X1.22	3.3447	0.1878	X1.32	0.0744	0.9635
X1.3	0.5366	0.7647	X1.12	0.1575	0.9243		X1.23	2.6423	0.2668	X1.33	0.3662	0.8327
X1.4	3.2798	0.1940	X1.13	2.3968	0.3017		X1.24	3.0247	0.2204	X1.34	0.1226	0.9406
X1.5	0.3864	0.8243	X1.14	4.4984	0.1055		X1.25	0.6464	0.7238	X0.2	3.3780	0.1847
X1.6	0.1999	0.9049	X1.15	1.3845	0.5005		X1.26	2.8628	0.2390	X1.35	6.3505	0.0418 *
X1.7	7.7339	0.0209 *	X1.16	1.9822	0.3712		X0.1	0.7367	0.6919	X1.36	0.0915	0.9553
X1.8	3.4037	0.1823	X1.17	0.6105	0.7369		X1.27	4.5665	0.1019	X1.37	2.6875	0.2609
X1.9	0.8629	0.6496	X1.18	0.1039	0.9494		X1.28	1.2729	0.5292			
X0	0.8928	0.6399	X1.19	1.0822	0.5821		X1.29	0.1715	0.9178			
			X1.20	7.7940	0.0203	*	X1.30	0.1016	0.9505			

Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Detection threshold: 5.9915
(significance level: 0.05)
Items detected as DIF items:
X1.7
X1.20
X1.35

Effect size (Nagelkerke's R^2): Effect size code: 'A': negligible effect 'B': moderate effect 'C': large effect

	R^2	ZT	JG												
X1.1	0.0002	A	A	X0	0.0027	A	A	X1.19	0.0011	A	A	X1.28	0.0017	A	A	X1.37	0.0027	A	A
X1.2	0.0025	A	A	X1.10	0.0002	A	A	X1.20	0.0244	A	A	X1.29	0.0001	A	A				
X1.3	0.0003	A	A	X1.11	0.0028	A	A	X1.21	0.0013	A	A	X1.30	0.0001	A	A				
X1.4	0.0020	A	A	X1.12	0.0001	A	A	X1.22	0.0037	A	A	X1.31	0.0018	A	A				
X1.5	0.0011	A	A	X1.13	0.0032	A	A	X1.23	0.0023	A	A	X1.32	0.0001	A	A				
X1.6	0.0002	A	A	X1.14	0.0045	A	A	X1.24	0.0027	A	A	X1.33	0.0002	A	A				
X1.7	0.0061	A	A	X1.15	0.0020	A	A	X1.25	0.0007	A	A	X1.34	0.0001	A	A				
X1.8	0.0041	A	A	X1.16	0.0022	A	A	X1.26	0.0023	A	A	X0.2	0.0035	A	A				
X1.9	0.0014	A	A	X1.17	0.0009	A	A	X0.1	0.0016	A	A	X1.35	0.0096	A	A				
				X1.18	0.0001	A	A	X1.27	0.0061	A	A	X1.36	0.0001	A	A				

Effect size codes:
Zumbo & Thomas (ZT):
0 'A' 0.13 'B' 0.26 'C' 1
Jodoign & Gierl (JG):
0 'A' 0.035 'B' 0.07 'C' 1

11>Detection of both types of Differential Item Functioning using Logistic regression method, without item purification and with LRT DIF statistic

Logistic regression DIF statistic:

	Stat.	P-value										
X1.1	1.1987	0.5492	X1.10	0.6242	0.7319	X1.20	0.2062	0.9020	X1.29	5.3973	0.0673	.
X1.2	1.6431	0.4397	X1.11	0.5060	0.7765	X1.21	0.8030	0.6693	X1.30	0.8441	0.6557	
X1.3	0.0993	0.9515	X1.12	2.0454	0.3596	X1.22	1.3948	0.4979	X1.31	3.9918	0.1359	
X1.4	0.5958	0.7424	X1.13	0.4161	0.8122	X0	0.0078	0.9961	X1.32	0.8193	0.6639	
X1.5	2.2272	0.3284	X1.14	0.6125	0.7362	X1.23	1.6310	0.4424	X1.33	3.9655	0.1377	
X1.6	0.6079	0.7379	X1.15	4.8265	0.0895 .	X1.24	1.3485	0.5095	X1.34	0.5023	0.7779	
X1.7	5.1660	0.0755 .	X1.16	2.7789	0.2492	X1.25	9.4622	0.0088 **	X1.35	0.3690	0.8315	
X1.8	0.4684	0.7912	X1.17	2.8413	0.2416	X1.26	5.5400	0.0627 .	X1.36	1.7241	0.4223	
X1.9	0.3153	0.8541	X1.18	5.6020	0.0607 .	X1.27	1.3408	0.5115	X1.37	0.7334	0.6930	
			X1.19	2.1191	0.3466	X1.28	0.5491	0.7599	X1.38	2.8681	0.2383	
												X1.39 4.8191 0.0899 .

Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Detection threshold: 5.9915
(significance level: 0.05)

Items detected as DIF items:
X1.25

Effect size (Nagelkerke's R^2): Effect size code: 'A': negligible effect 'B': moderate effect 'C': large effect

	R^2	ZT	JG										
X1.1	0.0008	A	A	X1.10	0.0020	A	A	X1.20	0.0002	A	A	X1.29	0.0070 A A
X1.2	0.0011	A	A	X1.11	0.0010	A	A	X1.21	0.0022	A	A	X1.30	0.0004 A A
X1.3	0.0001	A	A	X1.12	0.0024	A	A	X1.22	0.0012	A	A	X1.31	0.0030 A A
X1.4	0.0004	A	A	X1.13	0.0004	A	A	X0	0.0000	A	A	X1.32	0.0016 A A
X1.5	0.0065	A	A	X1.14	0.0008	A	A	X1.23	0.0014	A	A	X1.33	0.0056 A A
X1.6	0.0006	A	A	X1.15	0.0047	A	A	X1.24	0.0013	A	A	X1.34	0.0002 A A
X1.7	0.0044	A	A	X1.16	0.0035	A	A	X1.25	0.0104	A	A	X1.35	0.0002 A A
X1.8	0.0005	A	A	X1.17	0.0034	A	A	X1.26	0.0046	A	A	X1.36	0.0019 A A
X1.9	0.0005	A	A	X1.18	0.0075	A	A	X1.27	0.0029	A	A	X1.37	0.0010 A A
				X1.19	0.0016	A	A	X1.28	0.0007	A	A	X1.38	0.0051 A A
													X1.39 0.0045 A A

Effect size codes:
Zumbo & Thomas (ZT):
0 'A' 0.13 'B' 0.26 'C' 1
Jodoign & Gierl (JG):
0 'A' 0.035 'B' 0.07 'C' 1

12>Detection of both types of Differential Item Functioning using Logistic regression method, without item purification and with LRT DIF statistic

Logistic regression DIF statistic:

	Stat.	P-value										
X1.1	2.2187	0.3298	X1.11	0.6871	0.7092	X1.22	3.7159	0.1560	X1.32	0.8393	0.6573	
X1.2	1.3732	0.5033	X1.12	4.1171	0.1276	X1.23	2.4752	0.2901	X1.33	0.3624	0.8343	
X1.3	1.2457	0.5364	X1.13	0.0809	0.9603	X1.24	0.5015	0.7782	X1.34	3.3677	0.1857	
X1.4	3.6107	0.1644	X1.14	1.4195	0.4918	X1.25	0.2772	0.8706	X1.35	1.9382	0.3794	
X1.5	0.8055	0.6685	X1.15	1.6501	0.4382	X1.26	0.2495	0.8827	X1.36	2.0155	0.3650	
X1.6	4.7546	0.0928 .	X1.16	6.9732	0.0306 *	X1.27	2.5634	0.2776	X1.37	3.6774	0.1590	
X1.7	0.1611	0.9226	X1.17	3.0044	0.2226	X1.28	5.3774	0.0680 .	X1.38	8.4878	0.0144 *	
X1.8	1.7708	0.4126	X1.18	5.2914	0.0710 .	X0	0.1413	0.9318	X1.39	2.6099	0.2712	
X1.9	1.7215	0.4229	X1.19	0.2651	0.8758	X1.29	8.5786	0.0137 *				
X1.10	1.5161	0.4686	X1.20	0.9803	0.6125	X1.30	3.0524	0.2174				
			X1.21	0.0144	0.9928	X1.31	10.2995	0.0058 **				

Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Detection threshold: 5.9915
(significance level: 0.05)

Items detected as DIF items:
X1.16
X1.29
X1.31
X1.38

Effect size (Nagelkerke's R^2): Effect size code: 'A': negligible effect 'B': moderate effect 'C': large effect

	R^2	ZT	JG										
X1.1	0.0017	A	A	X1.10	0.0054	A	A	X1.20	0.0011	A	A	X1.29	0.0109 A A
X1.2	0.0008	A	A	X1.11	0.0012	A	A	X1.21	0.0000	A	A	X1.30	0.0016 A A
X1.3	0.0008	A	A	X1.12	0.0052	A	A	X1.22	0.0033	A	A	X1.31	0.0077 A A
X1.4	0.0023	A	A	X1.13	0.0001	A	A	X1.23	0.0029	A	A	X1.32	0.0016 A A
X1.5	0.0025	A	A	X1.14	0.0019	A	A	X1.24	0.0005	A	A	X1.33	0.0005 A A
X1.6	0.0047	A	A	X1.15	0.0016	A	A	X1.25	0.0002	A	A	X1.34	0.0017 A A
X1.7	0.0001	A	A	X1.16	0.0091	A	A	X1.26	0.0003	A	A	X1.35	0.0012 A A
X1.8	0.0019	A	A	X1.17	0.0034	A	A	X1.27	0.0023	A	A	X1.36	0.0021 A A
X1.9	0.0027	A	A	X1.18	0.0074	A	A	X1.28	0.0108	A	A	X1.37	0.0051 A A
				X1.19	0.0002	A	A	X0	0.0002	A	A	X1.38	0.0148 A A
													X1.39 0.0026 A A

Effect size codes:
Zumbo & Thomas (ZT):
0 'A' 0.13 'B' 0.26 'C' 1
Jodoign & Gierl (JG):
0 'A' 0.035 'B' 0.07 'C' 1

13>Detection of both types of Differential Item Functioning using Logistic regression method, without item purification and with LRT DIF statistic

Logistic regression DIF statistic:

	Stat.	P-value										
X1.1	3.6847	0.1584	X1.10	1.9419	0.3787	X1.21	1.2172	0.5441	X1.32	1.3931	0.4983	
X1.2	0.3069	0.8578	X1.11	0.6621	0.7182	X1.22	1.4531	0.4836	X1.33	1.2730	0.5291	
X1.3	4.0734	0.1305	X1.12	0.2460	0.8843	X1.23	0.7101	0.7011	X1.34	0.5546	0.7578	
X1.4	2.9384	0.2301	X1.13	1.7154	0.4241	X1.24	2.0796	0.3535	X1.35	0.5523	0.7587	
X1.5	2.6501	0.2658	X1.14	3.1606	0.2059	X1.25	0.7084	0.7018	X1.36	0.6073	0.7381	
X1.6	1.6191	0.4451	X1.15	1.4915	0.4744	X1.26	0.0153	0.9924	X1.37	5.7350	0.0568	.
X1.7	0.0760	0.9627	X1.16	6.5955	0.0370	*	X1.27	0.1767	0.9155	X1.38	0.4501	0.7985
X1.8	1.0651	0.5871	X1.17	1.4226	0.4910	X1.28	0.1238	0.9400	X1.39	0.1005	0.9510	
X1.9	0.5021	0.7780	X1.18	2.0284	0.3627	X1.29	0.0963	0.9530				
X0	5.4242	0.0664	.	X1.19	0.8032	0.6692	X1.30	6.4685	0.0394	*		
			X1.20	0.9987	0.6069	X1.31	0.7719	0.6798				

Signif. codes:

0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Detection threshold: 5.9915

(significance level: 0.05)

Items detected as DIF items:

X1.16

X1.30

Effect size (Nagelkerke's R^2): Effect size code: 'A': negligible effect 'B': moderate effect 'C': large effect

	R^2	ZT	JG										
X1.1	0.0026	A	A	X0	0.0150	A	A	X1.19	0.0008	A	A	X1.29	0.0001	A	A	X1.39	0.0001	A	A
X1.2	0.0002	A	A	X1.10	0.0032	A	A	X1.20	0.0029	A	A	X1.30	0.0035	A	A				
X1.3	0.0030	A	A	X1.11	0.0008	A	A	X1.21	0.0012	A	A	X1.31	0.0006	A	A				
X1.4	0.0019	A	A	X1.12	0.0003	A	A	X1.22	0.0018	A	A	X1.32	0.0028	A	A				
X1.5	0.0085	A	A	X1.13	0.0021	A	A	X1.23	0.0006	A	A	X1.33	0.0018	A	A				
X1.6	0.0018	A	A	X1.14	0.0029	A	A	X1.24	0.0018	A	A	X1.34	0.0003	A	A				
X1.7	0.0001	A	A	X1.15	0.0021	A	A	X1.25	0.0008	A	A	X1.35	0.0003	A	A				
X1.8	0.0013	A	A	X1.16	0.0073	A	A	X1.26	0.0000	A	A	X1.36	0.0007	A	A				
X1.9	0.0009	A	A	X1.17	0.0020	A	A	X1.27	0.0004	A	A	X1.37	0.0081	A	A				
				X1.18	0.0016	A	A	X1.28	0.0002	A	A	X1.38	0.0009	A	A				

Effect size codes:

Zumbo & Thomas (ZT):

0 'A' 0.13 'B' 0.26 'C' 1

Jodoign & Gierl (JG):

0 'A' 0.035 'B' 0.07 'C' 1

14> Detection of both types of Differential Item Functioning using Logistic regression method, without item purification and with LRT DIF statistic

Logistic regression DIF statistic:

	Stat.	P-value										
X1.1	1.0586	0.5890	X1.10	2.1632	0.3390	X1.19	1.7443	0.4181	X1.29	2.5835	0.2748	
X1.2	3.5249	0.1716	X1.11	2.5325	0.2819	X1.20	0.3838	0.8254	X1.30	1.0580	0.5892	
X1.3	1.1879	0.5521	X1.12	2.8174	0.2445	X1.21	0.2305	0.8911	X1.31	1.6054	0.4481	
X1.4	0.4976	0.7797	X0	0.2079	0.9013	X1.22	0.0303	0.9850	X1.32	4.5006	0.1054	
X1.5	0.0395	0.9804	X1.13	1.9611	0.3751	X1.23	0.0928	0.9547	X1.33	0.3434	0.8422	
X1.6	1.1743	0.5559	X1.14	0.6030	0.7397	X1.24	0.9594	0.6190	X1.34	1.3898	0.4991	
X1.7	1.4846	0.4760	X1.15	1.1670	0.5579	X1.25	0.7233	0.6965	X1.35	3.3537	0.1870	
X1.8	0.4891	0.7831	X1.16	2.6904	0.2605	X1.26	0.5966	0.7421	X1.36	0.1139	0.9446	
X1.9	0.5900	0.7445	X1.17	4.0297	0.1333	X1.27	7.4308	0.0243	*	X1.37	4.6837	0.0961	.
			X1.18	2.1253	0.3455	X1.28	0.3324	0.8469	X1.38	0.4462	0.8000	

Signif. codes:

0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Detection threshold: 5.9915

(significance level: 0.05)

Items detected as DIF items:

X1.27

Effect size (Nagelkerke's R^2): Effect size code: 'A': negligible effect 'B': moderate effect 'C': large effect

	R^2	ZT	JG										
X1.1	0.0008	A	A	X1.10	0.0057	A	A	X1.19	0.0020	A	A	X1.29	0.0040	A	A	X1.39	0.0006	A	A
X1.2	0.0021	A	A	X1.11	0.0043	A	A	X1.20	0.0010	A	A	X1.30	0.0006	A	A				
X1.3	0.0007	A	A	X1.12	0.0032	A	A	X1.21	0.0002	A	A	X1.31	0.0011	A	A				
X1.4	0.0003	A	A	X0	0.0002	A	A	X1.22	0.0000	A	A	X1.32	0.0082	A	A				
X1.5	0.0002	A	A	X1.13	0.0023	A	A	X1.23	0.0001	A	A	X1.33	0.0005	A	A				
X1.6	0.0013	A	A	X1.14	0.0006	A	A	X1.24	0.0008	A	A	X1.34	0.0007	A	A				
X1.7	0.0012	A	A	X1.15	0.0015	A	A	X1.25	0.0008	A	A	X1.35	0.0020	A	A				
X1.8	0.0006	A	A	X1.16	0.0030	A	A	X1.26	0.0005	A	A	X1.36	0.0001	A	A				
X1.9	0.0009	A	A	X1.17	0.0057	A	A	X1.27	0.0134	A	A	X1.37	0.0065	A	A				
				X1.18	0.0016	A	A	X1.28	0.0004	A	A	X1.38	0.0009	A	A				

Effect size codes:

Zumbo & Thomas (ZT):

0 'A' 0.13 'B' 0.26 'C' 1

Jodoign & Gierl (JG):

0 'A' 0.035 'B' 0.07 'C' 1

15>Detection of both types of Differential Item Functioning using Logistic regression method, without item purification and with LRT DIF statistic

Logistic regression DIF statistic:

Stat.	P-value									
X1.1	0.9932	0.6086	X1.10	1.0082	0.6041	X1.21	3.1696	0.2050	X1.32	3.6278	0.1630
X1.2	0.4373	0.8036	X1.11	0.6586	0.7194	X1.22	0.1673	0.9197	X1.33	1.2384	0.5384
X1.3	2.2056	0.3319	X1.12	1.7065	0.4260	X1.23	5.3831	0.0678 .	X1.34	2.7169	0.2571
X1.4	0.4847	0.7848	X1.13	0.5547	0.7578	X1.24	7.0143	0.0300 *	X1.35	0.6697	0.7154
X1.5	0.4606	0.7943	X1.14	0.8908	0.6406	X1.25	1.8488	0.3968	X1.36	3.2722	0.1947
X1.6	5.1782	0.0751 .	X1.15	4.9658	0.0835 .	X1.26	0.1375	0.9336	X1.37	2.9572	0.2280
X1.7	1.2437	0.5370	X1.16	5.1534	0.0760 .	X1.27	1.2715	0.5295	X1.38	3.4491	0.1782
X1.8	6.8744	0.0322 *	X1.17	0.8367	0.6581	X1.28	1.6688	0.4341	X1.39	4.1274	0.1270
X1.9	1.6326	0.4421	X1.18	1.3418	0.5112	X1.29	1.4080	0.4946			
X0	0.7985	0.6708	X1.19	3.5235	0.1717	X1.30	1.5879	0.4521			
			X1.20	3.3250	0.1897	X1.31	2.9921	0.2240			

Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1
Detection threshold: 5.9915
(significance level: 0.05)
Items detected as DIF items:
X1.8
X1.24

Effect size (Nagelkerke's R^2): Effect size code: 'A': negligible effect 'B': moderate effect 'C': large effect

R^2	ZT	JG										
X1.1	0.0007	A	A	X0	0.0027	A	A	X1.19	0.0036	A	A	X1.29	0.0018	A	A	X1.39	0.0037	A	A
X1.2	0.0003	A	A	X1.10	0.0022	A	A	X1.20	0.0094	A	A	X1.30	0.0008	A	A				
X1.3	0.0016	A	A	X1.11	0.0007	A	A	X1.21	0.0028	A	A	X1.31	0.0024	A	A				
X1.4	0.0003	A	A	X1.12	0.0015	A	A	X1.22	0.0002	A	A	X1.32	0.0081	A	A				
X1.5	0.0014	A	A	X1.13	0.0007	A	A	X1.23	0.0045	A	A	X1.33	0.0016	A	A				
X1.6	0.0052	A	A	X1.14	0.0009	A	A	X1.24	0.0063	A	A	X1.34	0.0016	A	A				
X1.7	0.0011	A	A	X1.15	0.0070	A	A	X1.25	0.0019	A	A	X1.35	0.0004	A	A				
X1.8	0.0065	A	A	X1.16	0.0068	A	A	X1.26	0.0001	A	A	X1.36	0.0037	A	A				
X1.9	0.0025	A	A	X1.17	0.0012	A	A	X1.27	0.0023	A	A	X1.37	0.0047	A	A				
				X1.18	0.0011	A	A	X1.28	0.0021	A	A	X1.38	0.0062	A	A				

Effect size codes:
Zumbo & Thomas (ZT):
0 'A' 0.13 'B' 0.26 'C' 1
Jodoign & Gierl (JG):
0 'A' 0.035 'B' 0.07 'C' 1

16>Detection of both types of Differential Item Functioning using Logistic regression method, without item purification and with LRT DIF statistic

Logistic regression DIF statistic:

Stat.	P-value									
X1.1	1.1196	0.5713	X1.10	0.3840	0.8253	X1.20	1.5372	0.4637	X0	3.4319	0.1798	X1.39	2.5033	0.2860
X1.2	1.7966	0.4073	X1.11	2.5335	0.2817	X1.21	0.6286	0.7303	X1.30	0.3512	0.8390			
X1.3	0.4477	0.7994	X1.12	3.6651	0.1600	X1.22	0.1049	0.9489	X1.31	2.5383	0.2811			
X1.4	0.1580	0.9240	X1.13	2.0561	0.3577	X1.23	0.3410	0.8432	X1.32	3.7435	0.1539			
X1.5	0.1251	0.9394	X1.14	0.7194	0.6979	X1.24	1.1163	0.5723	X1.33	0.2350	0.8891			
X1.6	1.8913	0.3884	X1.15	4.4108	0.1102	X1.25	2.5991	0.2727	X1.34	3.1925	0.2027			
X1.7	3.6045	0.1649	X1.16	3.4538	0.1778	X1.26	1.8947	0.3878	X1.35	0.0633	0.9689			
X1.8	0.9776	0.6134	X1.17	0.9450	0.6235	X1.27	2.5076	0.2854	X1.36	1.8338	0.3997			
X1.9	0.3901	0.8228	X1.18	3.8945	0.1427	X1.28	0.1084	0.9473	X1.37	1.9528	0.3767			
			X1.19	1.1127	0.5733	X1.29	2.3339	0.3113	X1.38	0.3995	0.8189			

Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1
Detection threshold: 5.9915
(significance level: 0.05)
Items detected as DIF items:
No DIF item detected

Effect size (Nagelkerke's R^2): Effect size code: 'A': negligible effect 'B': moderate effect 'C': large effect

R^2	ZT	JG										
X1.1	0.0007	A	A	X1.10	0.0012	A	A	X1.20	0.0017	A	A	X0	0.0043	A	A	X1.39	0.0027	A	A
X1.2	0.0012	A	A	X1.11	0.0047	A	A	X1.21	0.0018	A	A	X1.30	0.0002	A	A				
X1.3	0.0003	A	A	X1.12	0.0045	A	A	X1.22	0.0001	A	A	X1.31	0.0019	A	A				
X1.4	0.0001	A	A	X1.13	0.0017	A	A	X1.23	0.0004	A	A	X1.32	0.0078	A	A				
X1.5	0.0004	A	A	X1.14	0.0009	A	A	X1.24	0.0009	A	A	X1.33	0.0003	A	A				
X1.6	0.0019	A	A	X1.15	0.0051	A	A	X1.25	0.0025	A	A	X1.34	0.0016	A	A				
X1.7	0.0030	A	A	X1.16	0.0047	A	A	X1.26	0.0022	A	A	X1.35	0.0000	A	A				
X1.8	0.0013	A	A	X1.17	0.0012	A	A	X1.27	0.0020	A	A	X1.36	0.0020	A	A				
X1.9	0.0006	A	A	X1.18	0.0056	A	A	X1.28	0.0002	A	A	X1.37	0.0029	A	A				
				X1.19	0.0008	A	A	X1.29	0.0028	A	A	X1.38	0.0006	A	A				

Effect size codes:
Zumbo & Thomas (ZT):
0 'A' 0.13 'B' 0.26 'C' 1
Jodoign & Gierl (JG):
0 'A' 0.035 'B' 0.07 'C' 1

17>Detection of both types of Differential Item Functioning using Logistic regression method, without item purification and with LRT DIF statistic

Logistic regression DIF statistic:

Stat. P-value	X0 3.2129 0.2006	X1.19 4.9287 0.0851 .	X1.29 1.8252 0.4015	X1.38 1.6446 0.4394
X1.1 1.5067 0.4708	X1.10 0.6221 0.7327	X1.20 0.6441 0.7247	X1.30 0.9671 0.6166	
X1.2 0.1615 0.9224	X1.11 3.8913 0.1429	X1.21 1.9185 0.3832	X1.31 1.2200 0.5433	**Signif. codes:**
X1.3 2.4782 0.2896	X1.12 1.0350 0.5960	X1.22 2.0658 0.3560	X0.1 0.7619 0.6832	0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
X1.4 0.9470 0.6228	X1.13 1.5483 0.4611	X1.23 0.1489 0.9282	X1.32 2.2823 0.3194	' 1
X1.5 3.6046 0.1649	X1.14 0.1863 0.9111	X1.24 1.8657 0.3934	X1.33 3.8960 0.1426	Detection threshold: 5.9915
X1.6 4.7162 0.0946 .	X1.15 2.9218 0.2320	X1.25 0.5575 0.7567	X1.34 0.0197 0.9902	(significance level: 0.05)
X1.7 0.3700 0.8311	X1.16 1.5088 0.4703	X1.26 0.5923 0.7437	X1.35 1.9437 0.3784	**Items detected as DIF items:**
X1.8 2.4413 0.2950	X1.17 1.4494 0.4845	X1.27 0.7554 0.6854	X1.36 0.3510 0.8390	No DIF item detected
X1.9 3.4682 0.1766	X1.18 0.0664 0.9673	X1.28 3.1121 0.2110	X1.37 2.2288 0.3281	

Effect size (Nagelkerke's R^2): Effect size code: 'A': negligible effect 'B': moderate effect 'C': large effect

R^2 ZT JG	X0 0.0114 A A	X1.19 0.0058 A A	X1.29 0.0024 A A	X1.38 0.0015 A A
X1.1 0.0011 A A	X1.10 0.0012 A A	X1.20 0.0018 A A	X1.30 0.0005 A A	
X1.2 0.0001 A A	X1.11 0.0047 A A	X1.21 0.0018 A A	X1.31 0.0009 A A	**Effect size codes:**
X1.3 0.0017 A A	X1.12 0.0010 A A	X1.22 0.0025 A A	X0.1 0.0015 A A	Zumbo & Thomas (ZT):
X1.4 0.0006 A A	X1.13 0.0019 A A	X1.23 0.0001 A A	X1.32 0.0029 A A	0 'A' 0.13 'B' 0.26 'C' 1
X1.5 0.0105 A A	X1.14 0.0002 A A	X1.24 0.0016 A A	X1.33 0.0020 A A	Jodoign & Gierl (JG):
X1.6 0.0049 A A	X1.15 0.0043 A A	X1.25 0.0006 A A	X1.34 0.0000 A A	0 'A' 0.035 'B' 0.07 'C' 1
X1.7 0.0003 A A	X1.16 0.0018 A A	X1.26 0.0005 A A	X1.35 0.0021 A A	
X1.8 0.0028 A A	X1.17 0.0022 A A	X1.27 0.0014 A A	X1.36 0.0005 A A	
X1.9 0.0051 A A	X1.18 0.0000 A A	X1.28 0.0041 A A	X1.37 0.0038 A A	

18>Detection of both types of Differential Item Functioning using Logistic regression method, without item purification and with LRT DIF statistic

Logistic regression DIF statistic:

Stat. P-value	X1.11 0.8869 0.6418	X1.23 11.9427 0.0026 **	X1.34 5.6738 0.0586 .	**Signif. codes:**
X1.1 0.6721 0.7146	X1.12 0.4807 0.7863	X1.24 0.8812 0.6437	X1.35 0.7476 0.6881	0 '***' 0.001 '**'
X1.2 2.0096 0.3661	X1.13 0.4100 0.8146	X1.25 0.6679 0.7161	X1.36 0.5135 0.7736	0.01 '*' 0.05 '.' 0.1 ' ' 1
X1.3 1.7004 0.4273	X1.14 2.4448 0.2945	X1.26 4.1559 0.1252	X1.37 7.0285 0.0298 *	Detection threshold: 5.9915
X1.4 0.0024 0.9988	X1.15 2.4433 0.2947	X0.1 0.0642 0.9684	X1.38 1.2958 0.5232	(significance level: 0.05)
X1.5 6.2063 0.0449 *	X1.16 1.0929 0.5790	X1.27 1.2447 0.5367		**Items detected as DIF items:**
X1.6 1.2447 0.5367	X1.17 2.4202 0.2982	X1.28 2.9866 0.2246		X1.5
X1.7 0.6403 0.7261	X1.18 1.0417 0.5940	X1.29 0.7801 0.6770		X1.23
X1.8 5.7865 0.0554 .	X1.19 1.4356 0.4878	X1.30 0.5179 0.7719		X1.31
X1.9 0.1132 0.9450	X1.20 0.7527 0.6864	X1.31 6.0243 0.0492 *		X1.37
X0 0.7511 0.6869	X1.21 0.4463 0.8000	X1.32 0.0985 0.9519		
X1.10 1.9516 0.3769	X1.22 1.7733 0.4120	X1.33 1.6434 0.4397		

Effect size (Nagelkerke's R^2): Effect size code: 'A': negligible effect 'B': moderate effect 'C': large effect

R^2 ZT JG	X0 0.0026 A A	X1.19 0.0016 A A	X1.28 0.0040 A A	X1.38 0.0012 A A
X1.1 0.0004 A A	X1.10 0.0033 A A	X1.20 0.0023 A A	X1.29 0.0004 A A	
X1.2 0.0013 A A	X1.11 0.0010 A A	X1.21 0.0004 A A	X1.30 0.0004 A A	**Effect size codes:**
X1.3 0.0012 A A	X1.12 0.0005 A A	X1.22 0.0020 A A	X1.31 0.0102 A A	Zumbo & Thomas (ZT):
X1.4 0.0000 A A	X1.13 0.0005 A A	X1.23 0.0101 A A	X1.32 0.0001 A A	0 'A' 0.13 'B' 0.26 'C' 1
X1.5 0.0212 A A	X1.14 0.0023 A A	X1.24 0.0008 A A	X1.33 0.0009 A A	Jodoign & Gierl (JG):
X1.6 0.0013 A A	X1.15 0.0030 A A	X1.25 0.0007 A A	X1.34 0.0038 A A	0 'A' 0.035 'B' 0.07 'C' 1
X1.7 0.0005 A A	X1.16 0.0012 A A	X1.26 0.0033 A A	X1.35 0.0008 A A	
X1.8 0.0069 A A	X1.17 0.0032 A A	X0.1 0.0001 A A	X1.36 0.0007 A A	
X1.9 0.0002 A A	X1.18 0.0008 A A	X1.27 0.0017 A A	X1.37 0.0141 A A	

19>Detection of both types of Differential Item Functioning using Logistic regression method, without item purification and with LRT DIF statistic

Logistic regression DIF statistic:

Stat. P-value				
X1.1 0.5533 0.7583	X1.10 1.8868 0.3893	X1.20 1.1161 0.5723	X1.30 1.0243 0.5992	X1.39 0.4169 0.8118
X1.2 1.5311 0.4651	X1.11 0.6965 0.7059	X1.21 2.7360 0.2546	X1.31 2.7035 0.2588	
X1.3 0.6978 0.7054	X1.12 2.0216 0.3639	X1.22 2.4350 0.2960	X1.32 1.2970 0.5228	**Signif. codes:**
X1.4 5.1909 0.0746 .	X1.13 3.7716 0.1517	X1.23 0.0023 0.9988	X0 1.5572 0.4590	0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
X1.5 3.1924 0.2027	X1.14 0.5335 0.7659	X1.24 2.2769 0.3203	X1.33 1.6229 0.4442	' ' 1
X1.6 1.8815 0.3903	X1.15 0.5161 0.7726	X1.25 0.9376 0.6258	X1.34 1.1095 0.5742	Detection threshold: 5.9915
X1.7 0.0717 0.9648	X1.16 3.5161 0.1724	X1.26 2.6288 0.2686	X1.35 1.3370 0.5125	(significance level: 0.05)
X1.8 0.2507 0.8822	X1.17 1.2419 0.5374	X1.27 0.8477 0.6545	X1.36 3.3737 0.1851	**Items detected as DIF items:**
X1.9 3.1469 0.2073	X1.18 1.5814 0.4535	X1.28 1.6797 0.4318	X1.37 0.9297 0.6282	No DIF item detected
	X1.19 0.9906 0.6094	X1.29 1.1386 0.5659	X1.38 1.7381 0.4194	

Effect size (Nagelkerke's R^2): Effect size code: 'A': negligible effect 'B': moderate effect 'C': large effect

R^2 ZT JG				
X1.1 0.0004 A A	X1.10 0.0054 A A	X1.20 0.0011 A A	X1.30 0.0016 A A	X1.39 0.0004 A A
X1.2 0.0009 A A	X1.11 0.0012 A A	X1.21 0.0079 A A	X1.31 0.0014 A A	
X1.3 0.0005 A A	X1.12 0.0027 A A	X1.22 0.0023 A A	X1.32 0.0010 A A	**Effect size codes:**
X1.4 0.0033 A A	X1.13 0.0030 A A	X1.23 0.0000 A A	X0 0.0027 A A	Zumbo & Thomas (ZT):
X1.5 0.0109 A A	X1.14 0.0006 A A	X1.24 0.0019 A A	X1.33 0.0024 A A	0 'A' 0.13 'B' 0.26 'C' 1
X1.6 0.0021 A A	X1.15 0.0005 A A	X1.25 0.0009 A A	X1.34 0.0006 A A	Jodoign & Gierl (JG):
X1.7 0.0001 A A	X1.16 0.0045 A A	X1.26 0.0026 A A	X1.35 0.0008 A A	0 'A' 0.035 'B' 0.07 'C' 1
X1.8 0.0003 A A	X1.17 0.0014 A A	X1.27 0.0006 A A	X1.36 0.0036 A A	
X1.9 0.0046 A A	X1.18 0.0021 A A	X1.28 0.0035 A A	X1.37 0.0015 A A	
	X1.19 0.0008 A A	X1.29 0.0013 A A	X1.38 0.0034 A A	

20>Detection of both types of Differential Item Functioning using Logistic regression method, without item purification and with LRT DIF statistic

Logistic regression DIF statistic:

Stat. P-value				
X1.1 2.9932 0.2239	X1.10 3.8109 0.1488	X1.20 1.1763 0.5553	X0 2.2951 0.3174	X1.38 0.7830 0.6761
X1.2 0.4621 0.7937	X1.11 0.3107 0.8561	X1.21 1.4155 0.4927	X1.30 0.2684 0.8744	
X1.3 0.4370 0.8037	X1.12 1.7376 0.4195	X1.22 1.0308 0.5973	X1.31 0.8660 0.6485	**Signif. codes:**
X1.4 1.2458 0.5364	X1.13 1.3083 0.5199	X1.23 3.0213 0.2208	X0.1 1.7537 0.4161	0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
X1.5 2.9826 0.2251	X1.14 0.3784 0.8276	X1.24 0.0746 0.9634	X1.32 2.1058 0.3489	' ' 1
X1.6 1.8801 0.3906	X1.15 1.3854 0.5002	X1.25 0.2920 0.8641	X1.33 1.8248 0.4016	Detection threshold: 5.9915
X1.7 4.0411 0.1326	X1.16 1.9471 0.3777	X1.26 0.6744 0.7138	X1.34 0.3904 0.8227	(significance level: 0.05)
X1.8 0.3453 0.8414	X1.17 0.4456 0.8003	X1.27 2.9072 0.2337	X1.35 1.9388 0.3793	**Items detected as DIF items:**
X1.9 3.7203 0.1556	X1.18 0.6155 0.7351	X1.28 2.3060 0.3157	X1.36 6.5914 0.0370 *	X1.36
	X1.19 0.1918 0.9086	X1.29 3.4059 0.1821	X1.37 0.8830 0.6431	

Effect size (Nagelkerke's R^2): Effect size code: 'A': negligible effect 'B': moderate effect 'C': large effect

R^2 ZT JG				
X1.1 0.0020 A A	X1.10 0.0093 A A	X1.20 0.0015 A A	X0 0.0030 A A	X1.38 0.0007 A A
X1.2 0.0003 A A	X1.11 0.0006 A A	X1.21 0.0037 A A	X1.30 0.0001 A A	
X1.3 0.0003 A A	X1.12 0.0020 A A	X1.22 0.0009 A A	X1.31 0.0007 A A	**Effect size codes:**
X1.4 0.0008 A A	X1.13 0.0012 A A	X1.23 0.0036 A A	X0.1 0.0034 A A	Zumbo & Thomas (ZT):
X1.5 0.0101 A A	X1.14 0.0005 A A	X1.24 0.0001 A A	X1.32 0.0031 A A	0 'A' 0.13 'B' 0.26 'C' 1
X1.6 0.0023 A A	X1.15 0.0014 A A	X1.25 0.0003 A A	X1.33 0.0010 A A	Jodoign & Gierl (JG):
X1.7 0.0033 A A	X1.16 0.0026 A A	X1.26 0.0007 A A	X1.34 0.0002 A A	0 'A' 0.035 'B' 0.07 'C' 1
X1.8 0.0004 A A	X1.17 0.0006 A A	X1.27 0.0024 A A	X1.35 0.0021 A A	
X1.9 0.0061 A A	X1.18 0.0008 A A	X1.28 0.0044 A A	X1.36 0.0101 A A	
	X1.19 0.0001 A A	X1.29 0.0044 A A	X1.37 0.0017 A A	

21>Detection of both types of Differential Item Functioning using Logistic regression method, without item purification and with LRT DIF statistic

Logistic regression DIF statistic:

Stat.	P-value									
X1.1	3.8064	0.1491	X0	1.8529	0.3959	X1.19	3.4804	0.1755	X1.29	4.2685	0.1183	X1.39	5.1363	0.0767	.
X1.2	4.8606	0.0880	.	X1.10	3.2237	0.1995	X1.20	0.2242	0.8939	X1.30	0.2707	0.8734			
X1.3	0.8925	0.6400	X1.11	5.5870	0.0612	.	X1.21	0.1014	0.9506	X1.31	1.7584	0.4151	**Signif. codes:**		
X1.4	0.9465	0.6230	X1.12	2.5652	0.2773		X1.22	0.7500	0.6873	X1.32	0.4087	0.8152	0 '***' 0.001 '**' 0.01 '*' 0.05 '.'		
X1.5	2.2946	0.3175	X1.13	0.5880	0.7453		X1.23	1.5147	0.4689	X1.33	0.9925	0.6088	0.1 ' ' 1		
X1.6	2.1807	0.3361	X1.14	0.5727	0.7510		X1.24	1.3434	0.5108	X1.34	1.3878	0.4996	Detection threshold: 5.9915		
X1.7	4.2078	0.1220	X1.15	0.8265	0.6615		X1.25	0.1583	0.9239	X1.35	1.9864	0.3704	(significance level: 0.05)		
X1.8	1.5752	0.4549	X1.16	2.5097	0.2851		X1.26	0.0255	0.9873	X1.36	0.3142	0.8546	**Items detected as DIF items:**		
X1.9	0.0176	0.9912	X1.17	1.2377	0.5386		X1.27	0.1082	0.9474	X1.37	1.9162	0.3836	X1.28		
			X1.18	1.6247	0.4438		X1.28	6.3267	0.0423	*	X1.38	0.7983	0.6709		

Effect size (Nagelkerke's R^2): Effect size code: 'A': negligible effect 'B': moderate effect 'C': large effect

R^2	ZT	JG													
X1.1	0.0026	A	A	X0	0.0057	A	A	X1.19	0.0041	A	A	X1.29	0.0055	A	A	X1.39	0.0049	A	A
X1.2	0.0031	A	A	X1.10	0.0049	A	A	X1.20	0.0007	A	A	X1.30	0.0001	A	A				
X1.3	0.0006	A	A	X1.11	0.0063	A	A	X1.21	0.0001	A	A	X1.31	0.0014	A	A	**Effect size codes:**			
X1.4	0.0006	A	A	X1.12	0.0024	A	A	X1.22	0.0009	A	A	X1.32	0.0007	A	A	Zumbo & Thomas (ZT):			
X1.5	0.0079	A	A	X1.13	0.0007	A	A	X1.23	0.0015	A	A	X1.33	0.0014	A	A	0 'A' 0.13 'B' 0.26 'C' 1			
X1.6	0.0022	A	A	X1.14	0.0006	A	A	X1.24	0.0011	A	A	X1.34	0.0007	A	A	Jodoign & Gierl (JG):			
X1.7	0.0035	A	A	X1.15	0.0011	A	A	X1.25	0.0002	A	A	X1.35	0.0013	A	A	0 'A' 0.035 'B' 0.07 'C' 1			
X1.8	0.0017	A	A	X1.16	0.0028	A	A	X1.26	0.0000	A	A	X1.36	0.0003	A	A				
X1.9	0.0000	A	A	X1.17	0.0018	A	A	X1.27	0.0002	A	A	X1.37	0.0026	A	A				
				X1.18	0.0013	A	A	X1.28	0.0084	A	A	X1.38	0.0016	A	A				

22>Detection of both types of Differential Item Functioning using Logistic regression method, without item purification and with LRT DIF statistic

Logistic regression DIF statistic:

Stat.	P-value														
X1.1	1.2653	0.5312	X0	5.9174	0.0519	.	X1.18	0.3820	0.8261	X1.28	1.0405	0.5944	X1.38	2.6802	0.2618
X1.2	0.5713	0.7515	X1.10	1.6140	0.4462		X1.19	2.9062	0.2338	X1.29	2.5288	0.2824			
X1.3	3.6708	0.1596	X1.11	0.5558	0.7574		X1.20	0.0316	0.9843	X1.30	0.9751	0.6141	**Signif. codes:**		
X1.4	0.9232	0.6303	X1.12	1.0961	0.5781		X1.21	2.1090	0.3484	X1.31	1.0114	0.6031	0 '***' 0.001 '**' 0.01 '*' 0.05 '.'		
X1.5	3.1257	0.2095	X1.13	0.7952	0.6719		X1.22	0.7250	0.6959	X1.32	4.9861	0.0827	.	0.1 ' ' 1	
X1.6	1.9644	0.3745	X1.14	2.8112	0.2452		X1.23	1.5800	0.4538	X1.33	0.8180	0.6643	Detection threshold: 5.9915		
X1.7	2.2826	0.3194	X1.15	1.0502	0.5915		X1.24	2.9360	0.2304	X1.34	0.3167	0.8535	(significance level: 0.05)		
X1.8	0.6651	0.7171	X1.16	2.4348	0.2960		X1.25	1.0964	0.5780	X1.35	0.8776	0.6448	**Items detected as DIF items:**		
X1.9	0.3880	0.8237	X0.1	1.7943	0.4077		X1.26	0.6873	0.7092	X1.36	0.2309	0.8910	No DIF item detected		
			X1.17	0.8811	0.6437		X1.27	4.5755	0.1015	X1.37	2.3704	0.3057			

Effect size (Nagelkerke's R^2): Effect size code: 'A': negligible effect 'B': moderate effect 'C': large effect

R^2	ZT	JG													
X1.1	0.0009	A	A	X0	0.0176	A	A	X1.18	0.0004	A	A	X1.28	0.0015	A	A	X1.38	0.0028	A	A
X1.2	0.0003	A	A	X1.10	0.0030	A	A	X1.19	0.0090	A	A	X1.29	0.0012	A	A				
X1.3	0.0022	A	A	X1.11	0.0006	A	A	X1.20	0.0000	A	A	X1.30	0.0007	A	A	**Effect size codes:**			
X1.4	0.0006	A	A	X1.12	0.0011	A	A	X1.21	0.0023	A	A	X1.31	0.0019	A	A	Zumbo & Thomas (ZT):			
X1.5	0.0112	A	A	X1.13	0.0009	A	A	X1.22	0.0007	A	A	X1.32	0.0075	A	A	0 'A' 0.13 'B' 0.26 'C' 1			
X1.6	0.0020	A	A	X1.14	0.0030	A	A	X1.23	0.0013	A	A	X1.33	0.0004	A	A	Jodoign & Gierl (JG):			
X1.7	0.0021	A	A	X1.15	0.0014	A	A	X1.24	0.0032	A	A	X1.34	0.0002	A	A	0 'A' 0.035 'B' 0.07 'C' 1			
X1.8	0.0007	A	A	X1.16	0.0027	A	A	X1.25	0.0009	A	A	X1.35	0.0010	A	A				
X1.9	0.0006	A	A	X0.1	0.0028	A	A	X1.26	0.0013	A	A	X1.36	0.0003	A	A				
				X1.17	0.0006	A	A	X1.27	0.0059	A	A	X1.37	0.0044	A	A				

23>Detection of both types of Differential Item Functioning using Logistic regression method, without item purification and with LRT DIF statistic

Logistic regression DIF statistic:

Stat.	P-value									
X1.1	0.2183	0.8966	X1.11	0.5326	0.7662	X1.22	4.8138	0.0901 .	X1.33	0.4149 0.8127
X1.2	0.1178	0.9428	X1.12	2.5720	0.2764	X1.23	2.1462	0.3419	X1.34	0.4961 0.7803
X1.3	3.1522	0.2068	X1.13	0.3379	0.8445	X1.24	1.7656	0.4136	X1.35	0.4889 0.7831
X1.4	0.4824	0.7857	X1.14	1.2793	0.5275	X1.25	1.0627	0.5878	X1.36	5.3262 0.0697 .
X1.5	6.0364	0.0489 *	X1.15	0.8999	0.6377	X1.26	0.5836	0.7469	X1.37	0.7789 0.6774
X1.6	1.7901	0.4086	X1.16	0.0454	0.9775	X1.27	0.8713	0.6468	X1.38	3.1016 0.2121
X1.7	0.6453	0.7242	X1.17	1.8202	0.4025	X1.28	5.0981	0.0782 .	X1.39	3.4060 0.1821
X1.8	0.8957	0.6390	X1.18	3.5265	0.1715	X1.29	1.7855	0.4095	X1.40	1.8665 0.3933
X1.9	0.8486	0.6542	X1.19	1.1460	0.5638	X1.30	1.4199	0.4917		
X1.10	0.3539	0.8378	X1.20	0.0323	0.9840	X1.31	6.6575	0.0358 *		
			X1.21	2.9394	0.2300	X1.32	1.7262	0.4219		

Signif. codes:

0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Detection threshold: 5.9915

(significance level: 0.05)

Items detected as DIF items:

X1.5

X1.31

Effect size (Nagelkerke's R^2): Effect size code: 'A': negligible effect 'B': moderate effect 'C': large effect

R^2	ZT	JG										
X1.1	0.0001	A A	X1.10	0.0011	A A	X1.20	0.0000	A A	X1.30	0.0018	A A	X1.40 0.0019 A A
X1.2	0.0001	A A	X1.11	0.0009	A A	X1.21	0.0079	A A	X1.31	0.0036	A A	
X1.3	0.0020	A A	X1.12	0.0031	A A	X1.22	0.0044	A A	X1.32	0.0012	A A	
X1.4	0.0003	A A	X1.13	0.0003	A A	X1.23	0.0025	A A	X1.33	0.0008	A A	
X1.5	0.0173	A A	X1.14	0.0016	A A	X1.24	0.0015	A A	X1.34	0.0007	A A	
X1.6	0.0019	A A	X1.15	0.0009	A A	X1.25	0.0011	A A	X1.35	0.0002	A A	
X1.7	0.0005	A A	X1.16	0.0001	A A	X1.26	0.0007	A A	X1.36	0.0033	A A	
X1.8	0.0009	A A	X1.17	0.0019	A A	X1.27	0.0007	A A	X1.37	0.0008	A A	
X1.9	0.0014	A A	X1.18	0.0052	A A	X1.28	0.0111	A A	X1.38	0.0047	A A	
			X1.19	0.0009	A A	X1.29	0.0022	A A	X1.39	0.0066	A A	

Effect size codes:

Zumbo & Thomas (ZT):

0 'A' 0.13 'B' 0.26 'C' 1

Jodoign & Gierl (JG):

0 'A' 0.035 'B' 0.07 'C' 1

24>Detection of both types of Differential Item Functioning using Logistic regression method, without item purification and with LRT DIF statistic

Logistic regression DIF statistic:

Stat.	P-value									
X1.1	3.0173	0.2212	X1.11	1.7970	0.4072	X1.22	1.1644	0.5587	X1.33	0.4251 0.8085
X1.2	8.0753	0.0176 *	X1.12	4.1690	0.1244	X1.23	1.1004	0.5768	X1.34	0.8419 0.6564
X1.3	2.8099	0.2454	X1.13	2.8400	0.2417	X1.24	0.5927	0.7435	X1.35	1.8299 0.4005
X1.4	3.7913	0.1502	X1.14	5.4562	0.0653 .	X1.25	0.7626	0.6830	X1.36	1.4654 0.4806
X1.5	3.3813	0.1844	X1.15	2.8851	0.2363	X1.26	9.5966	0.0082 **	X1.37	1.3344 0.5131
X1.6	4.4760	0.1067	X1.16	0.9835	0.6116	X1.27	0.7695	0.6806	X1.38	2.4556 0.2929
X1.7	0.6748	0.7136	X1.17	0.4566	0.7959	X1.28	0.0175	0.9913	X1.39	8.2141 0.0165 *
X1.8	1.2095	0.5462	X1.18	0.6584	0.7195	X1.29	0.3955	0.8206	X1.40	1.6591 0.4362
X1.9	4.3819	0.1118	X1.19	1.1409	0.5653	X1.30	0.4083	0.8153		
X1.10	3.4538	0.1778	X1.20	1.0721	0.5850	X1.31	1.8975	0.3872		
			X1.21	0.7560	0.6852	X1.32	4.4051	0.1105		

Signif. codes:

0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Detection threshold: 5.9915

(significance level: 0.05)

Items detected as DIF items:

X1.2

X1.26

X1.39

Effect size (Nagelkerke's R^2): Effect size code: 'A': negligible effect 'B': moderate effect 'C': large effect

R^2	ZT	JG										
X1.1	0.0021	A A	X1.10	0.0110	A A	X1.20	0.0013	A A	X1.30	0.0006	A A	X1.40 0.0015 A A
X1.2	0.0048	A A	X1.11	0.0038	A A	X1.21	0.0021	A A	X1.31	0.0009	A A	
X1.3	0.0019	A A	X1.12	0.0048	A A	X1.22	0.0010	A A	X1.32	0.0030	A A	
X1.4	0.0025	A A	X1.13	0.0025	A A	X1.23	0.0014	A A	X1.33	0.0008	A A	
X1.5	0.0141	A A	X1.14	0.0062	A A	X1.24	0.0006	A A	X1.34	0.0011	A A	
X1.6	0.0053	A A	X1.15	0.0026	A A	X1.25	0.0007	A A	X1.35	0.0010	A A	
X1.7	0.0005	A A	X1.16	0.0012	A A	X1.26	0.0110	A A	X1.36	0.0009	A A	
X1.8	0.0013	A A	X1.17	0.0005	A A	X1.27	0.0006	A A	X1.37	0.0014	A A	
X1.9	0.0070	A A	X1.18	0.0009	A A	X1.28	0.0000	A A	X1.38	0.0036	A A	
			X1.19	0.0008	A A	X1.29	0.0005	A A	X1.39	0.0138	A A	

Effect size codes:

Zumbo & Thomas (ZT):

0 'A' 0.13 'B' 0.26 'C' 1

Jodoign & Gierl (JG):

0 'A' 0.035 'B' 0.07 'C' 1

25>Detection of both types of Differential Item Functioning using Logistic regression method, without item purification and with LRT DIF statistic

Logistic regression DIF statistic:

	Stat.	P-value
X1.1	0.5678	0.7528
X1.2	2.4360	0.2958
X1.3	0.1785	0.9146
X1.4	0.6242	0.7319
X1.5	4.4053	0.1105
X1.6	4.2642	0.1186
X1.7	1.3594	0.5068
X1.8	1.9560	0.3761
X1.9	1.7146	0.4243
X0	0.5231	0.7699
X1.10	1.1839	0.5533
X1.11	0.1942	0.9074
X1.12	1.3651	0.5053
X1.13	0.7066	0.7024
X1.14	1.1397	0.5656
X1.15	0.3542	0.8377
X1.16	2.1303	0.3447
X1.17	0.4541	0.7969
X1.18	3.4216	0.1807
X1.19	0.6029	0.7397
X1.20	1.0734	0.5847
X1.21	0.2177	0.8969
X1.22	3.4755	0.1759
X1.23	4.0333	0.1331
X1.24	0.6544	0.7209
X1.25	0.7614	0.6834
X1.26	3.4460	0.1785
X1.27	2.5085	0.2853
X1.28	1.4212	0.4914
X1.29	1.4331	0.4884
X1.30	6.7550	0.0341 *
X1.31	0.5876	0.7454
X1.32	2.7578	0.2519
X1.33	0.1498	0.9278
X1.34	7.6173	0.0222 *
X1.35	2.4067	0.3002
X1.36	4.0549	0.1317
X1.37	0.6157	0.7350
X1.38	4.9968	0.0822 .
X1.39	0.1018	0.9504

Signif. codes:

0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Detection threshold: 5.9915

(significance level: 0.05)

Items detected as DIF items:

X1.30

X1.34

Effect size (Nagelkerke's R^2): Effect size code: 'A': negligible effect 'B': moderate effect 'C': large effect

	R^2	ZT	JG
X1.1	0.0004	A	A
X1.2	0.0015	A	A
X1.3	0.0001	A	A
X1.4	0.0004	A	A
X1.5	0.0122	A	A
X1.6	0.0043	A	A
X1.7	0.0011	A	A
X1.8	0.0021	A	A
X1.9	0.0027	A	A
X0	0.0016	A	A
X1.10	0.0020	A	A
X1.11	0.0002	A	A
X1.12	0.0013	A	A
X1.13	0.0009	A	A
X1.14	0.0012	A	A
X1.15	0.0005	A	A
X1.16	0.0025	A	A
X1.17	0.0007	A	A
X1.18	0.0025	A	A
X1.19	0.0006	A	A
X1.20	0.0027	A	A
X1.21	0.0002	A	A
X1.22	0.0041	A	A
X1.23	0.0032	A	A
X1.24	0.0006	A	A
X1.25	0.0009	A	A
X1.26	0.0031	A	A
X1.27	0.0040	A	A
X1.28	0.0017	A	A
X1.29	0.0017	A	A
X1.30	0.0036	A	A
X1.31	0.0005	A	A
X1.32	0.0050	A	A
X1.33	0.0002	A	A
X1.34	0.0040	A	A
X1.35	0.0014	A	A
X1.36	0.0042	A	A
X1.37	0.0009	A	A
X1.38	0.0094	A	A
X1.39	0.0001	A	A

Effect size codes:

Zumbo & Thomas (ZT):

0 'A' 0.13 'B' 0.26 'C' 1

Jodoign & Gierl (JG):

0 'A' 0.035 'B' 0.07 'C' 1

ภาคผนวก ค

Print out ผลการเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ และการวัดขนาดอิทธิพล

1. t-Test การเปรียบเทียบอัตราความถูกต้องภายใต้เงื่อนไขของรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน 2 รูปแบบ (แบบเอกรูปและแบบอเนกรูป)

วิธีถดถอยโลจิสติกโดยการทดสอบระดับนัยสำคัญ

Group Statistics

	TYPEDIF	N	Mean	Std. Deviation	Std. Error Mean
LR	Nonuniform	300	77.1000	34.75196	2.00641
	Uniform	300	36.5250	32.85959	1.89715

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
LR	Equal variances assumed	.188	.665	14.694	598	.000	40.5750	2.76131	35.15195	45.99805
	Equal variances not assumed			14.694	596.135	.000	40.5750	2.76131	35.15192	45.99808

การวัดขนาดอิทธิพล Zumbo and Thomas

Group Statistics

	TYPEDIF	N	Mean	Std. Deviation	Std. Error Mean
RQZ	Nonuniform	300	2.7333	7.93918	.45837
	Uniform	300	.0000	.00000	.00000

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
RQZ	Equal variances assumed	162.303	.000	5.963	598	.000	2.7333	.45837	1.83312	3.63354
	Equal variances not assumed			5.963	299.000	.000	2.7333	.45837	1.83130	3.63537

การวัดขนาดอิทธิพล Jodoin and Gierl

Group Statistics

	TYPEDIF	N	Mean	Std. Deviation	Std. Error Mean
RQJ	Nonuniform	300	33.9417	25.03561	1.44543
	Uniform	300	6.3417	14.73044	.85046

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
RQJ	Equal variances assumed	155.004	.000	16.457	598	.000	27.6000	1.67707	24.30634	30.89366
	Equal variances not assumed			16.457	483.866	.000	27.6000	1.67707	24.30476	30.89524

2. t-Test การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้เงื่อนไขของรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน 2 รูปแบบ

<u>วิธีถดถอยโลจิสติกโดยการทดสอบระดับนัยสำคัญ</u>

Group Statistics

	TYPEDIF	N	Mean	Std. Deviation	Std. Error Mean
NLR	Nonuniform	300	15.0260	25.57284	1.47645
	Uniform	300	5.8732	4.24806	.24526

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
NLR	Equal variances assumed	79.177	.000	6.115	598	.000	9.1528	1.49668	6.21338	12.09216
	Equal variances not assumed			6.115	315.489	.000	9.1528	1.49668	6.20803	12.09750

การวัดขนาดอิทธิพล Zumbo and Thomas

Group Statistics

	TYPEDIF	N	Mean	Std. Deviation	Std. Error Mean
NRQZ	Nonuniform	300	5.4202	17.92122	1.03468
	Uniform	300	.0000	.00000	.00000

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
NRQZ	Equal variances assumed	130.970	.000	5.238	598	.000	5.4202	1.03468	3.38811	7.45222
	Equal variances not assumed			5.238	299.000	.000	5.4202	1.03468	3.38399	7.45635

การวัดขนาดอิทธิพล Jodoin and Gierl

Group Statistics

	TYPEDIF	N	Mean	Std. Deviation	Std. Error Mean
NRQJ	Nonuniform	300	.1736	.70614	.04077
	Uniform	300	.0695	.42097	.02430

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
NRQJ	Equal variances assumed	19.652	.000	2.195	598	.029	.1042	.04746	.01095	.19738
	Equal variances not assumed			2.195	487.698	.029	.1042	.04746	.01091	.19743

3. t-Test การเปรียบเทียบอัตราความถูกต้องภายใต้เงื่อนไขของจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด (จำนวน 10% และ จำนวน 20%)

วิธีถดถอยโลจิสติกโดยการทดสอบระดับนัยสำคัญ

Group Statistics

	PERS	N	Mean	Std. Deviation	Std. Error Mean
LR	10	300	58.4333	40.27139	2.32507
	20	300	55.1917	38.55274	2.22584

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
LR	Equal variances assumed	.271	.603	1.007	598	.314	3.2417	3.21875	-3.07975	9.56309
	Equal variances not assumed			1.007	596.866	.314	3.2417	3.21875	-3.07978	9.56311

การวัดขนาดอิทธิพล Zumbo and Thomas

Group Statistics

	PERS	N	Mean	Std. Deviation	Std. Error Mean
RQZ	10	300	2.1500	7.73479	.44657
	20	300	.5833	2.39175	.13809

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
RQZ	Equal variances assumed	49.680	.000	3.352	598	.001	1.5667	.46743	.64866	2.48467
	Equal variances not assumed			3.352	355.661	.001	1.5667	.46743	.64739	2.48594

การวัดขนาดอิทธิพล Jodoin and Gierl

Group Statistics

	PERS	N	Mean	Std. Deviation	Std. Error Mean
RQJ	10	300	22.6333	26.46823	1.52814
	20	300	17.6500	22.64720	1.30754

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
RQJ	Equal variances assumed	11.704	.001	2.478	598	.013	4.9833	2.01119	1.03348	8.93318
	Equal variances not assumed			2.478	584.030	.014	4.9833	2.01119	1.03329	8.93337

4. t-Test การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้เงื่อนไขของจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด

วิธีถดถอยโลจิสติกโดยการทดสอบระดับนัยสำคัญ

Group Statistics

	PERS	N	Mean	Std. Deviation	Std. Error Mean
NLR	10	300	13.6000	25.84200	1.49199
	20	300	7.2992	5.13006	.29618

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
NLR	Equal variances assumed	66.565	.000	4.142	598	.000	6.3008	1.52110	3.31348	9.28819
	Equal variances not assumed			4.142	322.530	.000	6.3008	1.52110	3.30830	9.29337

การวัดขนาดอิทธิพล Zumbo and Thomas

Group Statistics

	PERS	N	Mean	Std. Deviation	Std. Error Mean
NRQZ	10	300	5.3889	17.92794	1.03507
	20	300	.0313	.31195	.01801

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
NRQZ	Equal variances assumed	129.335	.000	5.175	598	.000	5.3576	1.03523	3.32444	7.39069
	Equal variances not assumed			5.175	299.181	.000	5.3576	1.03523	3.32032	7.39482

การวัดขนาดอิทธิพล Jodoin and Gierl

Group Statistics

	PERS	N	Mean	Std. Deviation	Std. Error Mean
NRQJ	10	300	.1055	.49993	.02886
	20	300	.1376	.65639	.03790

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
NRQJ	Equal variances assumed	1.916	.167	-.672	598	.502	-.0320	.04764	-.12559	.06152
	Equal variances not assumed			-.672	558.552	.502	-.0320	.04764	-.12560	.06154

5. t-Test การเปรียบเทียบอัตราความถูกต้องภายใต้เงื่อนไขของความยาวของแบบสอบทั้งฉบับ 2 ขนาด (40 ข้อ และ 50 ข้อ)

วิธีถดถอยโลจิสติกโดยการทดสอบระดับนัยสำคัญ

Group Statistics

	K	N	Mean	Std. Deviation	Std. Error Mean
LR	40	300	53.1250	38.66130	2.23211
	50	300	60.5000	39.89220	2.30318

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
LR	Equal variances assumed	1.872	.172	-2.299	598	.022	-7.3750	3.20733	-13.67399	-1.07601
	Equal variances not assumed			-2.299	597.414	.022	-7.3750	3.20733	-13.67401	-1.07599

การวัดขนาดอิทธิพล Zumbo and Thomas

Group Statistics

	K	N	Mean	Std. Deviation	Std. Error Mean
RQZ	40	300	1.8333	7.35601	.42470
	50	300	.9000	3.49725	.20191

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
RQZ	Equal variances assumed	17.433	.000	1.985	598	.048	.9333	.47025	.00978	1.85688
	Equal variances not assumed			1.985	427.597	.048	.9333	.47025	.00904	1.85763

การวัดขนาดอิทธิพล Jodoin and Gierl

Group Statistics

	K	N	Mean	Std. Deviation	Std. Error Mean
RQJ	40	300	19.0833	23.65516	1.36573
	50	300	21.2000	25.77027	1.48785

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
RQJ	Equal variances assumed	4.042	.045	-1.048	598	.295	-2.1167	2.01963	-6.08310	1.84977
	Equal variances not assumed			-1.048	593.667	.295	-2.1167	2.01963	-6.08316	1.84983

6. t-Test การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้เงื่อนไขของความยาวของแบบสอบทั้งฉบับ 2 ขนาด

วิธีถดถอยโลจิสติกโดยการทดสอบระดับนัยสำคัญ

Group Statistics

	K	N	Mean	Std. Deviation	Std. Error Mean
NLR	40	300	14.4171	25.78318	1.48859
	50	300	6.4821	4.20510	.24278

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
NLR	Equal variances assumed	82.200	.000	5.261	598	.000	7.9350	1.50826	4.97290	10.89717
	Equal variances not assumed			5.261	314.895	.000	7.9350	1.50826	4.96749	10.90258

การวัดขนาดอิทธิพล Zumbo and Thomas

Group Statistics

	K	N	Mean	Std. Deviation	Std. Error Mean
NRQZ	40	300	5.4202	17.92122	1.03468
	50	300	.0000	.00000	.00000

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
NRQZ	Equal variances assumed	130.970	.000	5.238	598	.000	5.4202	1.03468	3.38811	7.45222
	Equal variances not assumed			5.238	299.000	.000	5.4202	1.03468	3.38399	7.45635

การวัดขนาดอิทธิพล Jodoin and Gierl

Group Statistics

	K	N	Mean	Std. Deviation	Std. Error Mean
NRQJ	40	300	.1311	.65050	.03756
	50	300	.1120	.50789	.02932

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
NRQJ	Equal variances assumed	.729	.394	.401	598	.689	.0191	.04765	-.07448	.11268
	Equal variances not assumed			.401	564.778	.689	.0191	.04765	-.07449	.11269

ภาคผนวก ง

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิธีแมนเทล-แฮนส์เซล วิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ และการวัดขนาดอิทธิพล ในข้อมูลเชิงประจักษ์

วิธีแมนเทล-แฮนส์เซล
แบบสอบวิชาคณิตศาสตร์ 40 ข้อ

Detection of Differential Item Functioning using Mantel-Haenszel method with continuity correction and without item purification
Mantel-Haenszel Chi-square statistic:

	Stat.	P-value
m1	4.0729	0.0436 *
m2	2.3086	0.1287
m3	12.4732	0.0004 ***
m4	51.8426	0.0000 ***
m5	14.7630	0.0001 ***
m6	30.7875	0.0000 ***
m7	102.6501	0.0000 ***
m8	293.2368	0.0000 ***
m9	177.3265	0.0000 ***
m10	241.9711	0.0000 ***
m11	59.4471	0.0000 ***
m12	37.3454	0.0000 ***
m13	248.4409	0.0000 ***
m14	19.8277	0.0000 ***
m15	14.7140	0.0001 ***
m16	1.2619	0.2613
m17	136.9272	0.0000 ***
m18	41.7603	0.0000 ***
m19	325.7855	0.0000 ***
m20	18.4509	0.0000 ***
m21	7.6840	0.0056 **
m22	0.3545	0.5516
m23	76.4393	0.0000 ***
m24	130.8514	0.0000 ***
m25	0.1226	0.7262
m26	19.6687	0.0000 ***
m27	431.3287	0.0000 ***

m28 9.3312 0.0023 **
m29 227.4586 0.0000 ***
m30 0.3414 0.5590
m31 4.2410 0.0395 *
m32 26.6862 0.0000 ***
m33 47.1145 0.0000 ***
m34 0.9543 0.3286
m35 1.5283 0.2164
m36 116.1168 0.0000 ***
m37 7.0677 0.0078 **
m38 24.8530 0.0000 ***
m39 0.0009 0.9762
m40 5.2701 0.0217 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Detection threshold: 3.8415 (significance level: 0.05)
Items detected as DIF items:

m1
m3
m4
m5
m6
m7
m8
m9
m10
m11
m12
m13
m14
m15
m17
m18
m19

m20
m21
m23
m24
m26
m27
m28
m29
m31
m32
m33
m36
m37
m38
m40

Effect size (ETS Delta scale):
Effect size code:
'A': negligible effect
'B': moderate effect
'C': large effect

	alphaMH	deltaMH	
m1	1.1276	-0.2822	A
m2	0.8315	0.4336	A
m3	1.2123	-0.4523	A
m4	0.8148	0.4812	A
m5	0.6667	0.9527	A
m6	1.2643	-0.5511	A
m7	0.5458	1.4229	B
m8	2.5040	-2.1571	C
m9	0.6587	0.9811	A
m10	0.2421	3.3328	C
m11	1.2879	-0.5946	A
m12	1.2600	-0.5431	A

m13 0.7058 0.8188 A
m14 0.8237 0.4559 A
m15 0.8135 0.4851 A
m16 1.0585 -0.1337 A
m17 2.5224 -2.1742 C
m18 1.4656 -0.8984 A
m19 1.6805 -1.2198 B
m20 0.7257 0.7534 A
m21 0.9169 0.2039 A
m22 0.9569 0.1037 A
m23 1.6079 -1.1161 B
m24 0.7840 0.5720 A
m25 0.9920 0.0189 A
m26 0.7739 0.6025 A
m27 1.7172 -1.2707 B
m28 0.6357 1.0645 B
m29 1.6183 -1.1312 B
m30 1.0228 -0.0529 A
m31 1.0655 -0.1491 A
m32 1.5754 -1.0681 B
m33 0.7659 0.6267 A
m34 0.9726 0.0653 A
m35 0.8249 0.4523 A
m36 0.6735 0.9288 A
m37 1.1823 -0.3936 A
m38 1.4723 -0.9090 A
m39 1.0077 -0.0181 A
m40 1.2153 -0.4583 A
Effect size codes: 0 'A' 1.0 'B' 1.5 'C'
(for absolute values of 'deltaMH')

Output was not captured!

วิธีแมนเทล-แฮนส์เซล
แบบสอบวิชาวิทยาศาสตร์ 50 ข้อ

Detection of Differential Item Functioning using Mantel-Haenszel method with continuity correction and without item purification
Mantel-Haenszel Chi-square statistic:

	Stat.	P-value
s1	4.7101	0.0300 *
s2	48.7430	0.0000 ***
s3	3.2829	0.0700 .
s4	223.3776	0.0000 ***
s5	0.0000	0.9997
s6	57.1037	0.0000 ***
s7	60.5255	0.0000 ***
s8	112.1921	0.0000 ***
s9	0.6966	0.4039
s10	62.9202	0.0000 ***
s11	6.2267	0.0126 *
s12	1.4591	0.2271
s13	5.3919	0.0202 *
s14	85.8293	0.0000 ***
s15	381.0867	0.0000 ***
s16	0.0926	0.7609
s17	59.3484	0.0000 ***
s18	6.0886	0.0136 *
s19	39.0238	0.0000 ***
s20	0.9544	0.3286
s21	0.0027	0.9585
s22	44.4778	0.0000 ***
s23	2.4648	0.1164
s24	0.9690	0.3249
s25	20.7474	0.0000 ***
s26	56.7542	0.0000 ***
s27	11.0183	0.0009 ***

```
s28   19.8176    0.0000 ***
s29    0.0712    0.7895
s30    0.3140    0.5752
s31    6.6159    0.0101 *
s32   23.8247    0.0000 ***
s33 342.9173    0.0000 ***
s34   29.9951    0.0000 ***
s35   17.6077    0.0000 ***
s36    1.6047    0.2052
s37    4.4374    0.0352 *
s38   17.3437    0.0000 ***
s39   53.7400    0.0000 ***
s40   16.7049    0.0000 ***
s41    1.6906    0.1935
s42   28.7507    0.0000 ***
s43   91.5224    0.0000 ***
s44    2.8327    0.0924 .
s45   12.9844    0.0003 ***
s46    5.2077    0.0225 *
s47    6.0209    0.0141 *
s48    0.0151    0.9022
s49    8.3755    0.0038 **
s50   28.5725    0.0000 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Detection threshold: 3.8415 (significance level: 0.05)
Items detected as DIF items:

 s1
 s2
 s4
 s6
 s7
 s8
 s10
```

s11
s13
s14
s15
s17
s18
s19
s22
s25
s26
s27
s28
s31
s32
s33
s34
s35
s37
s38
s39
s40
s42
s43
s45
s46
s47
s49
s50

Effect size (ETS Delta scale):
Effect size code:
'A': negligible effect
'B': moderate effect
'C': large effect

	alphaMH	deltaMH	
s1	0.9718	0.0673	A
s2	1.1157	-0.2572	A
s3	1.0240	-0.0558	A
s4	0.8113	0.4915	A
s5	1.0001	-0.0002	A
s6	1.1065	-0.2378	A
s7	1.1295	-0.2862	A
s8	1.1486	-0.3256	A
s9	0.9882	0.0278	A
s10	1.1190	-0.2641	A
s11	0.9672	0.0784	A
s12	1.0178	-0.0416	A
s13	1.0360	-0.0832	A
s14	1.1445	-0.3172	A
s15	0.7509	0.6732	A
s16	0.9954	0.0109	A
s17	1.1152	-0.2563	A
s18	1.0377	-0.0870	A
s19	0.9191	0.1983	A
s20	1.0163	-0.0379	A
s21	1.0008	-0.0018	A
s22	0.9122	0.2160	A
s23	0.9792	0.0494	A
s24	1.0152	-0.0353	A
s25	0.9298	0.1710	A
s26	0.8859	0.2847	A
s27	0.9564	0.1049	A
s28	1.0704	-0.1599	A
s29	1.0036	-0.0085	A
s30	0.9926	0.0174	A
s31	0.9593	0.0975	A
s32	0.9317	0.1661	A
s33	1.3051	-0.6257	A
s34	1.0813	-0.1837	A

```
s35  0.9433  0.1371 A
s36  0.9824  0.0416 A
s37  0.9709  0.0694 A
s38  1.0657 -0.1496 A
s39  0.9000  0.2475 A
s40  0.9447  0.1338 A
s41  0.9819  0.0430 A
s42  0.9263  0.1798 A
s43  0.8808  0.2982 A
s44  1.0241 -0.0559 A
s45  1.0527 -0.1207 A
s46  0.9690  0.0740 A
s47  1.0382 -0.0880 A
s48  1.0017 -0.0041 A
s49  1.0407 -0.0938 A
s50  1.0718 -0.1630 A

Effect size codes: 0 'A' 1.0 'B' 1.5 'C'
 (for absolute values of 'deltaMH')

Output was not captured!
```

วิธีถดถอยโลจิสติก
แบบสอบวิชาคณิตศาสตร์ 40 ข้อ

mathRun<-difLogistic(math[,2:41],group=math[,1],focal.name=1) > mathRun
Detection of both types of Differential Item Functioning using Logistic regression method, without item purification and with LRT DIF statistic
Logistic regression DIF statistic:

	Stat.	P-value	
m1	160.0883	0.0000	***
m2	8.5307	0.0140	*
m3	67.0478	0.0000	***
m4	192.0916	0.0000	***
m5	39.8798	0.0000	***
m6	165.1333	0.0000	***
m7	206.1968	0.0000	***
m8	291.5168	0.0000	***
m9	322.5792	0.0000	***
m10	491.4645	0.0000	***
m11	47.2726	0.0000	***
m12	112.6677	0.0000	***
m13	455.6780	0.0000	***
m14	91.0845	0.0000	***
m15	29.8240	0.0000	***
m16	2.0182	0.3645	
m17	124.4544	0.0000	***
m18	157.5843	0.0000	***
m19	352.8206	0.0000	***
m20	60.9956	0.0000	***
m21	35.0509	0.0000	***
m22	27.8462	0.0000	***
m23	90.1816	0.0000	***
m24	250.8351	0.0000	***
m25	1.7330	0.4204	
m26	59.4412	0.0000	***
m27	460.5161	0.0000	***

```
m28  21.1432    0.0000 ***
m29 259.5377    0.0000 ***
m30   0.3241    0.8504
m31  49.5612    0.0000 ***
m32  28.4547    0.0000 ***
m33 123.2594    0.0000 ***
m34  49.4540    0.0000 ***
m35   4.4707    0.1070
m36 218.3892    0.0000 ***
m37  25.3635    0.0000 ***
m38  32.0623    0.0000 ***
m39   0.0574    0.9717
m40   7.8060    0.0202 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Detection threshold: 5.9915 (significance level: 0.05)
Items detected as DIF items:

 m1
 m2
 m3
 m4
 m5
 m6
 m7
 m8
 m9
 m10
 m11
 m12
 m13
 m14
 m15
 m17
 m18
```

m19
m20
m21
m22
m23
m24
m26
m27
m28
m29
m31
m32
m33
m34
m36
m37
m38
m40

Effect size (Nagelkerke's R^2): Effect size code:
'A': negligible effect
'B': moderate effect
'C': large effect

	R^2	ZT	JG
m1	0.0000	A	A
m2	0.0110	A	A
m3	0.0000	A	A
m4	0.0000	A	A
m5	0.0000	A	A
m6	0.0000	A	A
m7	0.0000	A	A
m8	0.6390	C	C
m9	0.0000	A	A
m10	0.0000	A	A

m11 0.0453 A B
m12 0.0000 A A
m13 0.0000 A A
m14 0.0000 A A
m15 0.0000 A A
m16 0.0000 A A
m17 0.0000 A A
m18 0.0000 A A
m19 0.2536 B C
m20 0.0000 A A
m21 0.0000 A A
m22 0.0000 A A
m23 0.0000 A A
m24 0.0000 A A
m25 0.0000 A A
m26 0.0000 A A
m27 0.0000 A A
m28 0.0158 A A
m29 0.0000 A A
m30 0.0000 A A
m31 0.0000 A A
m32 0.0402 A B
m33 0.0000 A A
m34 0.0000 A A
m35 0.0000 A A
m36 0.0000 A A
m37 0.0202 A A
m38 0.0398 A B
m39 0.0014 A A
m40 0.0335 A A

Effect size codes:
 Zumbo & Thomas (ZT): 0 'A' 0.13 'B' 0.26 'C' 1
 Jodoign & Gierl (JG): 0 'A' 0.035 'B' 0.07 'C' 1
 Output was not captured!

วิธีถดถอยโลจิสติก
แบบสอบวิชาวิทยาศาสตร์ 50 ข้อ

> sciRun Detection of both types of Differential Item Functioning using Logistic regression method, without item purification and with LRT DIF statistic Logistic regression DIF statistic:

	Stat.	P-value
s1	24.5010	0.0000 ***
s2	63.1341	0.0000 ***
s3	213.3186	0.0000 ***
s4	225.4068	0.0000 ***
s5	2.7864	0.2483
s6	86.6615	0.0000 ***
s7	221.3989	0.0000 ***
s8	132.4727	0.0000 ***
s9	4.7315	0.0939 .
s10	89.6727	0.0000 ***
s11	23.5399	0.0000 ***
s12	33.0182	0.0000 ***
s13	16.2003	0.0003 ***
s14	147.7314	0.0000 ***
s15	395.6337	0.0000 ***
s16	94.7373	0.0000 ***
s17	66.7604	0.0000 ***
s18	91.6465	0.0000 ***
s19	41.1809	0.0000 ***
s20	183.6273	0.0000 ***
s21	14.1865	0.0008 ***
s22	45.5323	0.0000 ***
s23	3.5039	0.1734
s24	46.5749	0.0000 ***
s25	39.1346	0.0000 ***
s26	256.7667	0.0000 ***
s27	52.7246	0.0000 ***

```
s28  33.0153    0.0000 ***
s29  11.8090    0.0027 **
s30 110.2148    0.0000 ***
s31 130.4967    0.0000 ***
s32  33.2438    0.0000 ***
s33 370.9312    0.0000 ***
s34  42.1933    0.0000 ***
s35  18.6925    0.0001 ***
s36  46.5193    0.0000 ***
s37  26.4022    0.0000 ***
s38 130.9098    0.0000 ***
s39  66.3371    0.0000 ***
s40  21.0893    0.0000 ***
s41   9.7711    0.0076 **
s42  34.0746    0.0000 ***
s43  88.5208    0.0000 ***
s44   5.7510    0.0564 .
s45  66.9368    0.0000 ***
s46  14.1330    0.0009 ***
s47 270.5992    0.0000 ***
s48   6.0456    0.0487 *
s49  88.6808    0.0000 ***
s50  52.6844    0.0000 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Detection threshold: 5.9915 (significance level: 0.05)
Items detected as DIF items:

 s1
 s2
 s3
 s4
 s6
 s7
 s8
```

s10
s11
s12
s13
s14
s15
s16
s17
s18
s19
s20
s21
s22
s24
s25
s26
s27
s28
s29
s30
s31
s32
s33
s34
s35
s36
s37
s38
s39
s40
s41
s42
s43
s45
s46

s47

s48

s49

s50

Effect size (Nagelkerke's R^2):

Effect size code:

'A': negligible effect

'B': moderate effect

'C': large effect

	R^2	ZT	JG
s1	0.0000	A	A
s2	0.0862	A	C
s3	0.0000	A	A
s4	0.0000	A	A
s5	0.0000	A	A
s6	0.0000	A	A
s7	0.0000	A	A
s8	0.0000	A	A
s9	0.0000	A	A
s10	0.0000	A	A
s11	0.0000	A	A
s12	0.0000	A	A
s13	0.0000	A	A
s14	0.1105	A	C
s15	0.0000	A	A
s16	0.0000	A	A
s17	0.0000	A	A
s18	0.0000	A	A
s19	0.0000	A	A
s20	NaN	?	?
s21	0.0000	A	A
s22	0.0000	A	A

```
s23 0.0000 A  A
s24 0.0000 A  A
s25 0.0000 A  A
s26 0.0000 A  A
s27 0.0000 A  A
s28 0.0000 A  A
s29 0.0000 A  A
s30 0.0000 A  A
s31 0.0000 A  A
s32 0.0000 A  A
s33 0.0000 A  A
s34 0.0000 A  A
s35 0.0000 A  A
s36 0.0000 A  A
s37 0.0000 A  A
s38    NaN ?  ?
s39 0.0000 A  A
s40 0.0000 A  A
s41 0.0000 A  A
s42 0.0000 A  A
s43 0.0000 A  A
s44 0.0000 A  A
s45 0.0000 A  A
s46 0.0000 A  A
s47    NaN ?  ?
s48 0.0000 A  A
s49 0.0000 A  A
s50 0.0000 A  A

Effect size codes:
 Zumbo & Thomas (ZT): 0 'A' 0.13 'B' 0.26 'C' 1
 Jodoign & Gierl (JG): 0 'A' 0.035 'B' 0.07 'C' 1
Output was not captured!
```

ภาคผนวก จ

ผลการวิเคราะห์ประสิทธิผลของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้านอัตราความถูกต้อง (correct identification) และอัตราความคลาดเคลื่อนประเภทที่ 1 (type I error rate) ของวิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ (significance test) และการวัดขนาดอิทธิพล (measure of effect size)

ภาคผนวก จ

ผลการวิเคราะห์ประสิทธิผลของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้านอัตราความถูกต้อง (correct identification) และอัตราความคลาดเคลื่อนประเภทที่ 1 (type I error rate) ของวิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ (significance test) และการวัดขนาดอิทธิพล (measure of effect size)

ตารางที่ ฉ-1 ร้อยละของอัตราความถูกต้อง (correct identification) จากการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ (significance test)

รหัส	รอบ1	รอบ2	รอบ3	รอบ4	รอบ5	รอบ6	รอบ7	รอบ8	รอบ9	รอบ10	รอบ11	รอบ12	รอบ13	รอบ14	รอบ15	รอบ16	รอบ17	รอบ18	รอบ19	รอบ20	รอบ21	รอบ22	รอบ23	รอบ24	รอบ25	mean	SD
N_U1_k40d0.1nd4	0.00	25.00	0.00	0.00	0.00	0.00	25.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	25.00	0.00	3.00	8.12
N_U2_k40d0.2nd4	75.00	50.00	50.00	100.00	75.00	50.00	75.00	100.00	75.00	75.00	75.00	50.00	50.00	75.00	75.00	75.00	50.00	50.00	100.00	75.00	75.00	75.00	75.00	50.00	75.00	70.00	15.81
N_U3_k40d0.4nd4	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	0.00
N_U4_k40d0.1nd8	100.00	87.50	87.50	100.00	100.00	87.50	75.00	87.50	87.50	87.50	100.00	100.00	100.00	87.50	100.00	87.50	100.00	75.00	100.00	100.00	75.00	100.00	87.50	100.00	87.50	92.00	8.57
N_U5_k40d0.2nd8	100.00	87.50	87.50	100.00	100.00	87.50	87.50	100.00	87.50	87.50	87.50	87.50	75.00	100.00	100.00	100.00	87.50	100.00	100.00	75.00	75.00	62.50	87.50	87.50	87.50	89.50	9.80
N_U6_k40d0.4nd8	100.00	100.00	100.00	87.50	100.00	100.00	100.00	100.00	87.50	100.00	100.00	87.50	100.00	100.00	87.50	100.00	87.50	100.00	100.00	87.50	87.50	87.50	100.00	100.00	87.50	95.50	6.00
N_U7_k50d0.1nd5	60.00	100.00	100.00	80.00	100.00	100.00	100.00	100.00	60.00	100.00	100.00	100.00	80.00	100.00	80.00	40.00	80.00	100.00	80.00	80.00	60.00	60.00	100.00	100.00	100.00	86.40	17.64
N_U8_k50d0.2nd5	100.00	100.00	100.00	100.00	100.00	100.00	80.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	80.00	100.00	98.40	5.43
N_U9_k50d0.4nd5	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	80.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	99.20	3.92
N_U10_k50d0.1nd10	100.00	100.00	100.00	80.00	100.00	100.00	90.00	100.00	80.00	100.00	70.00	80.00	100.00	80.00	90.00	90.00	90.00	90.00	80.00	90.00	90.00	100.00	80.00	100.00	90.00	90.80	8.91
N_U11_k50d0.2nd10	10.00	10.00	10.00	0.00	20.00	0.00	10.00	0.00	10.00	10.00	10.00	0.00	0.00	0.00	10.00	0.00	10.00	0.00	10.00	0.00	0.00	0.00	10.00	10.00	0.00	5.60	5.71
N_U12_k50d0.4nd10	90.00	90.00	100.00	100.00	100.00	100.00	100.00	100.00	80.00	90.00	90.00	90.00	90.00	100.00	90.00	100.00	100.00	100.00	90.00	100.00	100.00	100.00	90.00	90.00	90.00	94.80	5.74
U1_k40d0.1nd4	50.00	0.00	25.00	50.00	25.00	25.00	0.00	0.00	0.00	0.00	100.00	25.00	0.00	25.00	0.00	0.00	0.00	0.00	50.00	25.00	25.00	25.00	25.00	0.00	50.00	21.00	24.17
U2_k40d0.2nd4	25.00	0.00	75.00	25.00	0.00	0.00	25.00	75.00	25.00	25.00	50.00	50.00	25.00	25.00	0.00	0.00	25.00	25.00	0.00	50.00	0.00	0.00	0.00	25.00	0.00	22.00	22.72
U3_k40d0.4nd4	100.00	75.00	50.00	100.00	25.00	75.00	50.00	75.00	75.00	50.00	75.00	100.00	100.00	75.00	100.00	100.00	50.00	75.00	100.00	100.00	75.00	25.00	75.00	50.00	75.00	74.00	22.89
U4_k40d0.1nd8	0.00	0.00	0.00	12.50	0.00	0.00	12.50	0.00	0.00	0.00	12.50	12.50	0.00	0.00	0.00	12.50	37.50	25.00	0.00	0.00	0.00	0.00	0.00	0.00	12.50	5.50	9.41
U5_k40d0.2nd8	37.50	25.00	37.50	25.00	62.50	37.50	25.00	37.50	25.00	62.50	25.00	25.00	25.00	25.00	25.00	37.50	37.50	25.00	50.00	50.00	37.50	12.50	37.50	25.00	0.00	32.50	13.69
U6_k40d0.4nd8	62.50	25.00	25.00	25.00	12.50	25.00	75.00	12.50	50.00	25.00	12.50	37.50	37.50	25.00	25.00	12.50	37.50	37.50	37.50	37.50	12.50	50.00	25.00	62.50	25.00	32.50	16.58
U7_k50d0.1nd5	0.00	40.00	20.00	0.00	0.00	0.00	0.00	20.00	0.00	20.00	40.00	0.00	40.00	0.00	0.00	0.00	0.00	0.00	20.00	0.00	0.00	40.00	0.00	20.00	0.00	10.40	15.09
U8_k50d0.2nd5	20.00	60.00	20.00	20.00	0.00	20.00	20.00	0.00	20.00	60.00	0.00	60.00	40.00	0.00	20.00	20.00	40.00	20.00	40.00	20.00	40.00	20.00	20.00	40.00	20.00	25.60	17.45
U9_k50d0.4nd5	100.00	100.00	100.00	100.00	60.00	100.00	100.00	80.00	100.00	80.00	80.00	100.00	100.00	100.00	100.00	80.00	100.00	80.00	80.00	80.00	100.00	100.00	100.00	100.00	60.00	91.20	12.75
U10_k50d0.1nd10	0.00	0.00	0.00	30.00	20.00	0.00	0.00	20.00	0.00	10.00	10.00	0.00	0.00	20.00	30.00	30.00	0.00	20.00	0.00	0.00	0.00	10.00	0.00	10.00	10.00	8.80	10.70
U11_k50d0.2nd10	0.00	30.00	40.00	10.00	10.00	30.00	30.00	50.00	30.00	20.00	50.00	50.00	40.00	20.00	50.00	30.00	30.00	40.00	20.00	20.00	40.00	40.00	30.00	30.00	30.00	30.80	12.94
U12_k50d0.4nd10	80.00	80.00	80.00	80.00	80.00	90.00	90.00	90.00	90.00	100.00	80.00	70.00	100.00	70.00	80.00	90.00	70.00	90.00	90.00	90.00	80.00	80.00	90.00	80.00	80.00	84.00	8.00

หมายเหตุ : Nu:ข้อสอบ DIF แบบอเนกรูป, U:ข้อสอบ DIF แบบเอกรูป, k:ความยาวของแบบสอบทั้งฉบับเป็นข้อ, d:ขนาดของการทำหน้าที่ต่างกัน (amount of DIF), nd: จำนวนข้อสอบที่ DIF ในฉบับ

ตารางที่ ฉ-2 ร้อยละของอัตราความคลาดเคลื่อนประเภทที่ 1 (type I error rate) จากการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญ (significance test)

รหัส	รอบ1	รอบ2	รอบ3	รอบ4	รอบ5	รอบ6	รอบ7	รอบ8	รอบ9	รอบ10	รอบ11	รอบ12	รอบ13	รอบ14	รอบ15	รอบ16	รอบ17	รอบ18	รอบ19	รอบ20	รอบ21	รอบ22	รอบ23	รอบ24	รอบ25	mean	SD
N_u1_k40d0.1nd4	8.33	11.11	5.56	11.11	8.33	0.00	13.89	8.33	8.33	8.33	2.78	11.11	5.56	2.78	5.56	0.00	0.00	11.11	0.00	2.78	2.78	0.00	5.56	5.56	5.56	5.78	4.08
N_u2_k40d0.2nd4	97.22	100.00	97.22	97.22	97.22	100.00	97.22	100.00	100.00	97.22	97.22	97.22	94.44	97.22	100.00	100.00	100.00	100.00	100.00	100.00	100.00	94.44	94.44	100.00	97.22	98.22	1.91
N_u3_k40d0.4nd4	5.56	13.89	5.56	2.78	2.78	2.78	0.00	5.56	2.78	13.89	8.33	5.56	11.11	11.11	5.56	8.33	2.78	13.89	5.56	13.89	2.78	8.33	11.11	5.56	2.78	6.89	4.17
N_u4_k40d0.1nd8	0.00	3.13	3.13	6.25	6.25	3.13	9.38	3.13	6.25	6.25	0.00	6.25	0.00	3.13	0.00	9.38	6.25	6.25	6.25	9.38	6.25	6.25	3.13	3.13	15.63	5.13	3.52
N_u5_k40d0.2nd8	12.50	3.13	6.25	15.63	6.25	3.13	0.00	9.38	15.63	3.13	0.00	18.75	6.25	6.25	6.25	6.25	9.38	9.38	0.00	9.38	6.25	6.25	3.13	0.00	15.63	7.13	5.19
N_u6_k40d0.4nd8	12.50	15.63	15.63	12.50	15.63	12.50	12.50	15.63	12.50	15.63	12.50	18.75	15.63	15.63	21.88	12.50	12.50	15.63	6.25	12.50	15.63	9.38	15.63	21.88	28.13	15.00	4.24
N_u7_k50d0.1nd5	6.67	11.11	4.44	6.67	4.44	6.67	6.67	4.44	13.33	4.44	4.44	4.44	4.44	4.44	4.44	6.67	8.89	15.56	11.11	2.22	6.67	6.67	4.44	6.67	4.44	6.58	3.11
N_u8_k50d0.2nd5	2.22	6.67	6.67	0.00	8.89	8.89	11.11	6.67	6.67	2.22	6.67	4.44	6.67	0.00	4.44	6.67	6.67	6.67	2.22	4.44	6.67	2.22	6.67	4.44	11.11	5.60	2.89
N_u9_k50d0.4nd5	6.67	4.44	4.44	4.44	6.67	4.44	13.33	13.33	4.44	2.22	2.22	17.78	4.44	8.89	6.67	8.89	0.00	6.67	6.67	8.89	4.44	11.11	11.11	11.11	8.89	7.29	4.00
N_u10_k50d0.1nd10	5.00	12.50	2.50	5.00	5.00	7.50	7.50	10.00	7.50	5.00	7.50	5.00	5.00	7.50	10.00	12.50	2.50	15.00	7.50	7.50	15.00	5.00	2.50	7.50	10.00	7.50	3.46
N_u11_k50d0.2nd10	2.50	2.50	5.00	2.50	5.00	2.50	5.00	0.00	5.00	2.50	0.00	10.00	5.00	5.00	2.50	2.50	7.50	7.50	2.50	0.00	2.50	5.00	0.00	5.00	7.50	3.80	2.56
N_u12_k50d0.4nd10	7.50	15.00	25.00	12.50	17.50	20.00	7.50	7.50	10.00	7.50	15.00	10.00	7.50	17.50	10.00	10.00	15.00	5.00	5.00	5.00	15.00	7.50	12.50	10.00	10.00	11.40	4.95
U1_k40d0.1nd4	2.78	5.56	0.00	2.78	13.89	8.33	5.56	5.56	0.00	2.78	0.00	11.11	5.56	0.00	5.56	5.56	5.56	5.56	5.56	0.00	0.00	2.78	5.56	0.00	8.33	4.33	3.61
U2_k40d0.2nd4	8.33	0.00	25.00	8.33	8.33	2.78	5.56	0.00	0.00	2.78	11.11	0.00	8.33	5.56	0.00	8.33	2.78	8.33	0.00	2.78	0.00	2.78	11.11	11.11	2.78	5.44	5.58
U3_k40d0.4nd4	16.67	0.00	8.33	8.33	5.56	8.33	5.56	5.56	5.56	11.11	11.11	5.56	8.33	5.56	5.56	2.78	2.78	8.33	5.56	2.78	2.78	5.56	5.56	13.89	0.00	6.44	3.83
U4_k40d0.1nd8	3.13	3.13	3.13	3.13	3.13	3.13	9.38	9.38	0.00	0.00	3.13	3.13	6.25	0.00	3.13	12.50	3.13	9.38	6.25	9.38	3.13	6.25	0.00	3.13	9.38	4.63	3.44
U5_k40d0.2nd8	6.25	0.00	15.63	6.25	12.50	6.25	0.00	15.63	6.25	12.50	6.25	21.88	6.25	6.25	6.25	0.00	6.25	3.13	6.25	3.13	9.38	12.50	3.13	6.25	9.38	7.50	5.15
U6_k40d0.4nd8	12.50	6.25	9.38	3.13	3.13	3.13	9.38	0.00	3.13	6.25	3.13	3.13	0.00	9.38	6.25	6.25	12.50	9.38	9.38	3.13	9.38	9.38	9.38	9.38	6.25	6.50	3.53
U7_k50d0.1nd5	6.67	2.22	0.00	2.22	6.67	2.22	8.89	6.67	8.89	2.22	2.22	11.11	0.00	6.67	6.67	2.22	8.89	11.11	2.22	6.67	6.67	4.44	4.44	6.67	6.67	5.33	3.14
U8_k50d0.2nd5	2.22	4.44	2.22	11.11	8.89	4.44	2.22	0.00	2.22	2.22	2.22	8.89	0.00	4.44	6.67	8.89	4.44	8.89	4.44	4.44	8.89	6.67	0.00	0.00	4.44	4.53	3.23
U9_k50d0.4nd5	0.00	13.33	11.11	4.44	4.44	8.89	2.22	6.67	11.11	2.22	4.44	8.89	4.44	11.11	6.67	15.56	6.67	6.67	6.67	4.44	4.44	4.44	2.22	4.44	13.33	6.76	3.90
U10_k50d0.1nd10	2.50	7.50	5.00	2.50	0.00	5.00	2.50	5.00	2.50	7.50	7.50	0.00	2.50	2.50	5.00	10.00	2.50	7.50	7.50	2.50	0.00	2.50	7.50	0.00	5.00	4.10	2.82
U11_k50d0.2nd10	2.50	7.50	5.00	10.00	0.00	2.50	2.50	7.50	7.50	0.00	7.50	12.50	12.50	2.50	10.00	7.50	2.50	7.50	10.00	0.00	0.00	2.50	12.50	7.50	2.50	5.70	4.09
U12_k50d0.4nd10	10.00	12.50	12.50	5.00	7.50	10.00	5.00	12.50	7.50	15.00	7.50	7.50	2.50	10.00	20.00	10.00	20.00	10.00	10.00	2.50	5.00	7.50	2.50	5.00	12.50	9.20	4.62

หมายเหตุ : Nu:ข้อสอบ DIF แบบอเนกรูป, U:ข้อสอบ DIF แบบเอกรูป, k:ความยาวของแบบสอบทั้งฉบับเป็นข้อ, d:ขนาดของการทำหน้าที่ต่างกัน (amount of DIF), nd: จำนวนข้อสอบที่ DIF ในฉบับ

ตารางที่ ฉ-3 ร้อยละของอัตราความถูกต้อง (correct identification) จากการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพล (measure of effect size) ZT

รหัส	รอบ1	รอบ2	รอบ3	รอบ4	รอบ5	รอบ6	รอบ7	รอบ8	รอบ9	รอบ10	รอบ11	รอบ12	รอบ13	รอบ14	รอบ15	รอบ16	รอบ17	รอบ18	รอบ19	รอบ20	รอบ21	รอบ22	รอบ23	รอบ24	รอบ25	mean	SD
N_u1_k40d0.1nd4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N_u2_k40d0.2nd4	75.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.00	14.70
N_u3_k40d0.4nd4	25.00	25.00	0.00	25.00	25.00	25.00	25.00	0.00	25.00	25.00	0.00	25.00	25.00	25.00	25.00	0.00	25.00	25.00	0.00	0.00	0.00	25.00	25.00	25.00	25.00	18.00	11.22
N_u4_k40d0.1nd8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N_u5_k40d0.2nd8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N_u6_k40d0.4nd8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	12.50	12.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	3.39
N_u7_k50d0.1nd5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N_u8_k50d0.2nd5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N_u9_k50d0.4nd5	0.00	0.00	0.00	20.00	0.00	0.00	0.00	0.00	0.00	20.00	20.00	20.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	20.00	0.00	0.00	0.00	20.00	4.80	8.54
N_u10_k50d0.1nd10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	10.00	0.00	0.00	0.00	0.00	0.00	10.00	10.00	10.00	10.00	0.00	0.00	0.00	0.00	2.00	4.00
N_u11_k50d0.2nd10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N_u12_k50d0.4nd10	10.00	0.00	10.00	0.00	0.00	0.00	10.00	0.00	0.00	10.00	0.00	10.00	0.00	0.00	10.00	10.00	10.00	0.00	10.00	0.00	10.00	0.00	0.00	0.00	0.00	4.00	4.90
U1_k40d0.1nd4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U2_k40d0.2nd4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U3_k40d0.4nd4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U4_k40d0.1nd8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U5_k40d0.2nd8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U6_k40d0.4nd8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U7_k50d0.1nd5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U8_k50d0.2nd5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U9_k50d0.4nd5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U10_k50d0.1nd10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U11_k50d0.2nd10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U12_k50d0.4nd10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

หมายเหตุ : Nu:ข้อสอบ DIF แบบอเนกรูป, U:ข้อสอบ DIF แบบเอกรูป, k:ความยาวของแบบสอบทั้งฉบับเป็นข้อ, d:ขนาดของการทำหน้าที่ต่างกัน (amount of DIF), nd: จำนวนข้อสอบที่ DIF ในฉบับ

ตารางที่ ฉ-4 ร้อยละของอัตราความคลาดเคลื่อนประเภทที่ 1 (type I error rate) จากการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพล (measure of effect size) ZT

รหัส	รอบ1	รอบ2	รอบ3	รอบ4	รอบ5	รอบ6	รอบ7	รอบ8	รอบ9	รอบ10	รอบ11	รอบ12	รอบ13	รอบ14	รอบ15	รอบ16	รอบ17	รอบ18	รอบ19	รอบ20	รอบ21	รอบ22	รอบ23	รอบ24	รอบ25	mean	SD
N_u1_k40d0.1nd4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N_u2_k40d0.2nd4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N_u3_k40d0.4nd4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N_u4_k40d0.1nd8	0.00	3.13	0.00	0.00	0.00	0.00	0.00	3.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.13	0.00	0.38	1.02
N_u5_k40d0.2nd8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N_u6_k40d0.4nd8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N_u7_k50d0.1nd5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N_u8_k50d0.2nd5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N_u9_k50d0.4nd5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N_u10_k50d0.1nd10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N_u11_k50d0.2nd10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N_u12_k50d0.4nd10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U1_k40d0.1nd4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U2_k40d0.2nd4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U3_k40d0.4nd4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U4_k40d0.1nd8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U5_k40d0.2nd8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U6_k40d0.4nd8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U7_k50d0.1nd5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U8_k50d0.2nd5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U9_k50d0.4nd5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U10_k50d0.1nd10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U11_k50d0.2nd10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U12_k50d0.4nd10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

หมายเหตุ : Nu:ข้อสอบ DIF แบบอเนกรูป, U:ข้อสอบ DIF แบบเอกรูป, k:ความยาวของแบบสอบทั้งฉบับเป็นข้อ, d:ขนาดของการทำหน้าที่ต่างกัน (amount of DIF), nd: จำนวนข้อสอบที่ DIF ในฉบับ

ตารางที่ ฉ-5 ร้อยละของอัตราความถูกต้อง (correct identification) จากการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพล (measure of effect size) JG

รหัส	รอบ1	รอบ2	รอบ3	รอบ4	รอบ5	รอบ6	รอบ7	รอบ8	รอบ9	รอบ10	รอบ11	รอบ12	รอบ13	รอบ14	รอบ15	รอบ16	รอบ17	รอบ18	รอบ19	รอบ20	รอบ21	รอบ22	รอบ23	รอบ24	รอบ25	mean	SD
N_u1_k40d0.1nd4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N_u2_k40d0.2nd4	75.00	50.00	50.00	50.00	25.00	0.00	25.00	50.00	25.00	0.00	0.00	0.00	0.00	25.00	25.00	25.00	25.00	25.00	75.00	50.00	50.00	50.00	25.00	50.00	50.00	33.00	22.05
N_u3_k40d0.4nd4	50.00	75.00	50.00	75.00	50.00	75.00	50.00	75.00	50.00	50.00	50.00	75.00	75.00	50.00	75.00	50.00	50.00	50.00	75.00	75.00	75.00	75.00	50.00	75.00	50.00	62.00	12.49
N_u4_k40d0.1nd8	37.50	37.50	37.50	37.50	25.00	50.00	25.00	25.00	50.00	25.00	37.50	25.00	37.50	37.50	37.50	25.00	37.50	50.00	37.50	37.50	37.50	50.00	50.00	50.00	25.00	37.00	9.00
N_u5_k40d0.2nd8	25.00	25.00	12.50	12.50	0.00	12.50	12.50	25.00	25.00	12.50	12.50	12.50	12.50	12.50	12.50	12.50	12.50	12.50	37.50	12.50	12.50	12.50	12.50	25.00	0.00	15.00	7.91
N_u6_k40d0.4nd8	50.00	50.00	50.00	50.00	62.50	50.00	50.00	50.00	50.00	62.50	37.50	50.00	50.00	37.50	37.50	37.50	37.50	37.50	62.50	37.50	37.50	50.00	37.50	37.50	25.00	45.50	9.27
N_u7_k50d0.1nd5	20.00	0.00	20.00	0.00	20.00	20.00	0.00	20.00	0.00	20.00	0.00	0.00	0.00	0.00	20.00	20.00	0.00	20.00	0.00	20.00	0.00	0.00	0.00	20.00	0.00	8.80	9.93
N_u8_k50d0.2nd5	20.00	80.00	60.00	40.00	20.00	40.00	20.00	20.00	40.00	40.00	20.00	20.00	40.00	60.00	40.00	40.00	20.00	20.00	60.00	40.00	20.00	40.00	40.00	60.00	40.00	37.60	16.32
N_u9_k50d0.4nd5	60.00	60.00	60.00	60.00	60.00	60.00	60.00	60.00	80.00	80.00	80.00	40.00	60.00	60.00	60.00	100.00	60.00	40.00	60.00	60.00	60.00	60.00	80.00	60.00	60.00	63.20	12.24
N_u10_k50d0.1nd10	50.00	40.00	40.00	50.00	50.00	20.00	50.00	60.00	40.00	30.00	50.00	40.00	60.00	40.00	50.00	40.00	40.00	60.00	60.00	40.00	50.00	60.00	50.00	40.00	60.00	46.80	10.09
N_u11_k50d0.2nd10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N_u12_k50d0.4nd10	60.00	50.00	80.00	40.00	60.00	70.00	80.00	60.00	70.00	40.00	70.00	40.00	70.00	70.00	70.00	50.00	50.00	60.00	50.00	40.00	70.00	50.00	60.00	50.00	50.00	58.40	12.22
U1_k40d0.1nd4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U2_k40d0.2nd4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	50.00	0.00	0.00	0.00	0.00	0.00	2.00	9.80
U3_k40d0.4nd4	25.00	25.00	25.00	75.00	25.00	50.00	25.00	25.00	0.00	0.00	50.00	0.00	50.00	50.00	50.00	50.00	25.00	25.00	50.00	50.00	25.00	25.00	0.00	50.00	50.00	33.00	19.65
U4_k40d0.1nd8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U5_k40d0.2nd8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	12.50	0.00	12.50	0.00	0.00	0.00	0.00	0.00	0.00	1.00	3.39
U6_k40d0.4nd8	12.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.50	2.45
U7_k50d0.1nd5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U8_k50d0.2nd5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U9_k50d0.4nd5	40.00	40.00	20.00	20.00	0.00	40.00	0.00	40.00	60.00	40.00	60.00	20.00	0.00	40.00	40.00	40.00	20.00	20.00	60.00	60.00	40.00	40.00	0.00	20.00	40.00	32.00	18.76
U10_k50d0.1nd10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U11_k50d0.2nd10	0.00	0.00	0.00	10.00	0.00	0.00	10.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	10.00	0.00	0.00	0.00	0.00	1.20	3.25
U12_k50d0.4nd10	0.00	10.00	0.00	0.00	10.00	10.00	0.00	0.00	30.00	10.00	10.00	10.00	0.00	0.00	10.00	10.00	0.00	10.00	10.00	10.00	0.00	0.00	20.00	0.00	0.00	6.40	7.42

หมายเหตุ : Nu:ข้อสอบ DIF แบบอเนกรูป, U:ข้อสอบ DIF แบบเอกรูป, k:ความยาวของแบบสอบทั้งฉบับเป็นข้อ, d:ขนาดของการทำหน้าที่ต่างกัน (amount of DIF), nd: จำนวนข้อสอบที่ DIF ในฉบับ

ตารางที่ ฉ-6 ร้อยละของอัตราความคลาดเคลื่อนประเภทที่ 1 (type I error rate) จากการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพล (measure of effect size) JG

รหัส	รอบ1	รอบ2	รอบ3	รอบ4	รอบ5	รอบ6	รอบ7	รอบ8	รอบ9	รอบ10	รอบ11	รอบ12	รอบ13	รอบ14	รอบ15	รอบ16	รอบ17	รอบ18	รอบ19	รอบ20	รอบ21	รอบ22	รอบ23	รอบ24	รอบ25	mean	SD
N_u1_k40d0.1nd4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N_u2_k40d0.2nd4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.22	0.00	0.00	2.22	0.00	2.22	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.27	0.72
N_u3_k40d0.4nd4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.78	0.00	0.00	0.00	0.00	0.00	0.00	2.78	0.00	0.00	0.00	0.00	0.00	0.22	0.75
N_u4_k40d0.1nd8	0.00	0.00	0.00	0.00	3.13	0.00	3.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.13	0.00	0.00	3.13	0.00	0.00	0.00	0.00	0.00	6.25	0.75	1.60
N_u5_k40d0.2nd8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N_u6_k40d0.4nd8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N_u7_k50d0.1nd5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.22	0.00	0.00	2.22	0.00	2.22	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.27	0.72
N_u8_k50d0.2nd5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.22	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.44
N_u9_k50d0.4nd5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.22	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.44
N_u10_k50d0.1nd10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.49
N_u11_k50d0.2nd10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N_u12_k50d0.4nd10	0.00	0.00	2.50	0.00	2.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.30	0.81
U1_k40d0.1nd4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U2_k40d0.2nd4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U3_k40d0.4nd4	2.78	0.00	0.00	0.00	2.78	2.78	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.90
U4_k40d0.1nd8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U5_k40d0.2nd8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U6_k40d0.4nd8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U7_k50d0.1nd5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U8_k50d0.2nd5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U9_k50d0.4nd5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U10_k50d0.1nd10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.50	0.00	0.00	0.00	2.50	0.00	0.00	0.00	0.20	0.68
U11_k50d0.2nd10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U12_k50d0.4nd10	0.00	2.50	0.00	0.00	0.00	0.00	0.00	0.00	2.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.50	0.00	0.30	0.81

หมายเหตุ : Nu:ข้อสอบ DIF แบบอเนกรูป, U:ข้อสอบ DIF แบบเอกรูป, k:ความยาวของแบบสอบทั้งฉบับเป็นข้อ, d:ขนาดของการทำหน้าที่ต่างกัน (amount of DIF), nd: จำนวนข้อสอบที่ DIF ในฉบับ

ประวัติผู้เขียนวิทยานิพนธ์

นางสาว ธเกียรติกมล ทองงอก เกิดเมื่อวันที่ 4 กุมภาพันธ์ พ.ศ.2519 สำเร็จการศึกษาครุศาสตรบัณฑิต วิชาเอกคณิตศาสตร์ สถาบันราชภัฏสวนดุสิต ใน ปีการศึกษา 2541 สำเร็จการศึกษาศึกษาศาสตรมหาบัณฑิต สาขาการวัดผล การศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัย ศรีครินทรวิโรฒ ในปีการศึกษา 2545 เข้าศึกษาต่อในหลักสูตร ครุศาสตรดุษฎีบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย เมื่อปีการศึกษา 2550