

การเปรียบเทียบวิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติก
แบบจัดอันดับหลายมิติ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ
ที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า

ปริญญานิพนธ์
ของ
อรินทร์ น่วมถนนอม

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษาดุขฎีบัณฑิต สาขาวิชาการทดสอบและวัดผลการศึกษา
พฤษภาคม 2549

การเปรียบเทียบวิธีโพลี-ชิปเทสต์ วิธีการถอดถอยโลจิสติกแบบจัดอันดับ และวิธีการถอดถอยโลจิสติก
แบบจัดอันดับหลายมิติ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ
ที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า

ปริญญานิพนธ์

ของ

อรินทร์ น่วมถนอม

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษาดุริยบัณฑิต สาขาวิชาการทดสอบและวัดผลการศึกษา

พฤษภาคม 2549

ลิขสิทธิ์เป็นของมหาวิทยาลัยศรีนครินทรวิโรฒ

การเปรียบเทียบวิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติก
แบบจัดอันดับหลายมิติ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ
ที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า

บทคัดย่อ

ของ

อรินทร์ น่วมถนนอม

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษาดุษฎีบัณฑิต สาขาวิชาการทดสอบและวัดผลการศึกษา

พฤษภาคม 2549

อรินทร์ น่วมถนนม. (2549). การเปรียบเทียบวิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ในการตรวจสอบการทำหน้าที่เป็ยเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า. ปรินญานิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. คณะกรรมการควบคุม: รองศาสตราจารย์ ดร.บุญเชิด ภิญโญอนันตพงษ์, รองศาสตราจารย์ ดร.ผจจจิต อินทสุวรรณ, ผู้ช่วยศาสตราจารย์ ดร.สุรัชย์ มีชาญ.

การศึกษาค้นคว้าครั้งนี้มีจุดมุ่งหมายเพื่อเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ในการตรวจสอบการทำหน้าที่เป็ยเบนของข้อสอบที่วัดความสามารถหลายมิติ และให้คะแนนหลายค่า ในการศึกษาครั้งนี้ใช้การจำลองข้อมูลภายใต้โมเดลพหุคูณเชิงเส้นแบบหลายมิติ จำลองผลการตอบจากแบบทดสอบที่วัดความสามารถสองมิติจำนวน 40 ข้อ ข้อสอบทุกข้อมีรายการตอบ 5 รายการ ให้คะแนนเป็น 1, 2, 3, 4 หรือ 5 คะแนน ข้อมูลดังกล่าวจำลองภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เป็ยเบน 2 รูปแบบ สัดส่วนของข้อสอบทำหน้าที่เป็ยเบน 3 ขนาด ความแตกต่างของการแจกแจงความสามารถ 3 ระดับ และขนาดตัวอย่าง 4 ขนาด รวมข้อมูลที่ใช้ศึกษาทั้งหมด 72 เงื่อนไข ในแต่ละเงื่อนไขจำลองซ้ำ 50 ครั้ง แล้ววิเคราะห์ข้อมูลในแต่ละเงื่อนไขด้วยวิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ในการวิเคราะห์การทำหน้าที่เป็ยเบนของข้อสอบทั้งหมดใช้ระดับนัยสำคัญ .05 ผลการวิจัยสรุปได้ดังนี้

วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ มีอัตราความถูกต้องใกล้เคียงกัน ในการตรวจสอบการทำหน้าที่เป็ยเบนของข้อสอบที่เป็นรูปแบบเดียวกันและที่ไม่เป็นรูปแบบเดียวกัน ทั้งสองวิธีมีอัตราความถูกต้องสูงกว่าวิธีโพลี-ชิปเทสท์ ในการตรวจสอบการทำหน้าที่เป็ยเบนของข้อสอบที่ไม่เป็นรูปแบบเดียวกัน สัดส่วนของข้อสอบทำหน้าที่เป็ยเบนในแบบทดสอบไม่มีผลต่อวิธีโพลี-ชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ แต่มีผลต่อวิธีการถดถอยโลจิสติกแบบจัดอันดับ เมื่อความแตกต่างของการแจกแจงความสามารถเพิ่มขึ้น วิธีโพลี-ชิปเทสท์สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ที่สูงเกินปกติได้ดีกว่าวิธีอื่น นอกจากนี้ เมื่อขนาดตัวอย่างเพิ่มขึ้นมีผลทำให้ทุกวิธีมีอัตราความถูกต้องเพิ่มขึ้น ภายใต้เงื่อนไขทุกเงื่อนไข

โดยภาพรวม วิธีโพลี-ชิปเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ภายในขอบเขตที่กำหนด ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าขอบเขตที่กำหนด ผลการศึกษาค้นคว้าครั้งนี้เสนอแนะให้วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติใช้ระดับนัยสำคัญ .02 เพื่อให้สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้

A COMPARISON OF THE POLY-SIBTEST, ORDINAL LOGISTIC REGRESSION, AND
MULTIDIMENSIONAL ORDINAL LOGISTIC REGRESSION PROCEDURES
IN THE DETECTION OF DIFFERENTIAL ITEM FUNCTIONING FOR
MULTIDIMENSIONAL POLYTOMOUS SCORED ITEMS

AN ABSTRACT
BY
ARIN NAUMTANOM

Presented in Partial Fulfillment of the Requirements for the
Doctor of Education Degree in Testing and Measurement
at Srinakharinwirot University

May 2006

Arin Naumtanom. (2006). *A Comparison of the Poly-SIBTEST, Ordinal Logistic Regression, and Multidimensional Ordinal Logistic Regression Procedures in the Detection of Differential Item Functioning for Multidimensional Polytomous Scored Items*. Dissertation, Ed.D. (Testing and Measurement). Bangkok: Graduate School, Srinakharinwirot University. Advisor Committee: Assoc. Prof. Dr. Boonchird Pinyoanuntapong, Assoc. Prof. Dr. Pachongchit Intasuwan, Asst. Prof. Dr. Surachai Meechan.

The purpose of this research was to compare the power and Type I error rates of the Poly-SIBTEST, ordinal logistic regression (OLR), and multidimensional ordinal logistic regression (MOLR) procedures in the detection of differential item functioning (DIF) for multidimensional polytomous scored items. In this study, data were simulated under the multidimensional generalized partial credit model, and responses were simulated from two-dimensional test with 40 items. All items were in five response categories scoring as 1, 2, 3, 4, or 5. These data were simulated under a variety of four factors: two levels forms of DIF, three levels of ability distribution differences, three levels of proportion of DIF items, and four levels of sample sizes. A total of 72 conditions were studied. The data were replicated 50 times for each condition. Each data set was analyzed using the Poly-SIBTEST, OLR, and MOLR procedures. The significance level of .05 was applied in all DIF analyses.

Results of the study were as follows:

The power rates of OLR and MOLR procedures in detecting uniform and nonuniform DIF were equal. They were found to be higher than those of the Poly-SIBTEST procedure in detecting nonuniform DIF. Proportion of DIF items in the test did not have any effect on the Poly-SIBTEST and MOLR procedures, but affected the OLR procedure. When the ability distribution differences increased, the Poly-SIBTEST procedure had better control of Type I error rate inflation than other procedures. In addition, the power rates for all three procedures increased when sample sizes were increased under almost all conditions.

In general, the Type I error rates of the Poly-SIBTEST procedure were within the nominal limits. They were higher than nominal limits for the OLR and MOLR procedures. The results of this study suggested the MOLR procedure to use a significance level of .02 to control the Type I error rates.

ปริญญานิพนธ์
เรื่อง

การเปรียบเทียบวิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติก
แบบจัดอันดับหลายมิติ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ
ที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า

ของ
อรินทร์ น่วมถนนอม

ได้รับอนุมัติจากบัณฑิตวิทยาลัยให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาการศึกษาดุสิตบัณฑิต สาขาวิชาการทดสอบและวัดผลการศึกษา
ของมหาวิทยาลัยศรีนครินทรวิโรฒ

..... คณบดีบัณฑิตวิทยาลัย
(ผู้ช่วยศาสตราจารย์ ดร.เพ็ญสิริ จีระเดชากุล)
วันที่ เดือน พ.ศ. 2549

..... ประธานควบคุมปริญญานิพนธ์
(รองศาสตราจารย์ ดร.บุญเชิด ภิญโญนนตพงษ์)

..... กรรมการควบคุมปริญญานิพนธ์
(รองศาสตราจารย์ ดร.ผจงจิต อินทสุวรรณ)

..... กรรมการควบคุมปริญญานิพนธ์
(ผู้ช่วยศาสตราจารย์ ดร.สุรัชย์ มีชาญ)

..... กรรมการที่แต่งตั้งเพิ่มเติม
(ผู้ช่วยศาสตราจารย์ ดร.นิคม ดังคะพิภพ)

..... กรรมการที่แต่งตั้งเพิ่มเติม
(รองศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี)

ประกาศคุณูปการ

ปริญญานิพนธ์ฉบับนี้สำเร็จลุล่วงไปด้วยดี เป็นเพราะผู้วิจัยได้รับความกรุณาอย่างดียิ่งจากรองศาสตราจารย์ ดร.บุญเชิด ภิญโญนนตพงษ์ ประธานกรรมการควบคุมปริญญานิพนธ์ รองศาสตราจารย์ ดร.ผจงจิต อินทสุวรรณ และผู้ช่วยศาสตราจารย์ ดร.สุรัชย์ มีชาญ กรรมการควบคุมปริญญานิพนธ์ ท่านทั้งสามได้เสียสละเวลาอันมีค่าในการให้คำปรึกษา คำแนะนำ และข้อคิดเห็น ตลอดจนแก้ไขข้อบกพร่องต่างๆ อีกทั้งให้กำลังใจแก่ผู้วิจัยตลอดมา ผู้วิจัยรู้สึกซาบซึ้งและขอกราบขอบพระคุณเป็นอย่างสูงมา ณ โอกาสนี้

ขอกราบขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.นิคม ตั้งคะพิภพ และรองศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี กรรมการที่แต่งตั้งเพิ่มเติม ที่กรุณาให้คำแนะนำ และข้อเสนอแนะที่มีคุณค่า ทำให้ปริญญานิพนธ์ฉบับนี้มีความสมบูรณ์ยิ่งขึ้น

ขอกราบขอบพระคุณ รองศาสตราจารย์ ดร.บุญเชิด ภิญโญนนตพงษ์ และอาจารย์ ดร.สุพร เข้มเฮง อาจารย์ที่ปรึกษานิพนธ์ ซึ่งให้การดูแลด้วยความห่วงใย และให้คำแนะนำอย่างดียิ่งตลอดระยะเวลาที่ศึกษาในมหาวิทยาลัย

ขอขอบพระคุณ ศาสตราจารย์ Dr.Eiji Muraki แห่งศูนย์บริการทดสอบทางการศึกษา ประเทศสหรัฐอเมริกา ที่กรุณาอนุเคราะห์โปรแกรมคอมพิวเตอร์สำเร็จรูป RESGEN สำหรับจำลองข้อมูล และศาสตราจารย์ Dr.Bruno D. Zumbo แห่งมหาวิทยาลัยบริติช โคลัมเบีย ประเทศแคนาดา ที่กรุณาให้ดาวน์โหลด SYNTAX ของโปรแกรม SPSS สำหรับใช้วิเคราะห์วิธีการถดถอยโลจิสติกแบบจัดอันดับ

ขอกราบขอบพระคุณ ดร.สุวัฒน์ วิวัฒนานนท์ ผู้อำนวยการสถานศึกษา โรงเรียนบางกะปิ ที่สนับสนุนให้ผู้วิจัยมีโอกาสศึกษาต่ออย่างเต็มที่ ขอขอบพระคุณมหาวิทยาลัยศรีนครินทรวิโรฒ ที่ให้ทุนสนับสนุนส่วนหนึ่งในการวิจัยครั้งนี้ ขอขอบคุณเพื่อนๆ ทุกคน ที่ให้กำลังใจด้วยดีตลอดมา

ขอกราบขอบพระคุณ ครู-อาจารย์ทุกท่านทั้งในอดีตและปัจจุบัน ที่ได้อบรมสั่งสอนและประสิทธิ์ประสาทวิชาความรู้ จนกระทั่งผู้วิจัยประสบความสำเร็จในการศึกษา และหน้าที่การงานเป็นอย่างดี

ขอกราบขอบพระคุณ มารดา และบุคคลในครอบครัวที่คอยให้กำลังใจ และปรารถนาจะเห็นความสำเร็จของผู้วิจัยเป็นอย่างยิ่ง คุณค่าและประโยชน์ของปริญญานิพนธ์ฉบับนี้ ผู้วิจัยขอมอบเป็นเครื่องบูชาพระคุณ บิดา มารดา ครู-อาจารย์ทุกท่าน

อรินทร์ น่วมถนอม

งานวิจัยนี้ได้รับทุนอุดหนุนการวิจัย

จาก

งบประมาณแผ่นดิน ประจำปีงบประมาณ 2549

มหาวิทยาลัยศรีนครินทรวิโรฒ

สารบัญ

บทที่	หน้า
1 บทนำ.....	1
ภูมิหลัง.....	1
ความมุ่งหมายของการวิจัย.....	15
ความสำคัญของการวิจัย.....	16
ขอบเขตของการวิจัย.....	17
นิยามศัพท์เฉพาะ.....	19
กรอบแนวคิดในการวิจัย.....	22
สมมติฐานในการวิจัย.....	28
2 เอกสารและงานวิจัยที่เกี่ยวข้อง.....	29
นิยามของการทำหน้าที่เบี่ยงเบนของข้อสอบ.....	29
รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน.....	33
แนวคิดและหลักการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ.....	41
วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ.....	60
งานวิจัยที่เกี่ยวข้อง.....	86
3 วิธีดำเนินการวิจัย.....	110
การจัดกระทำตัวแปร.....	110
การจำลองข้อมูล.....	122
การวิเคราะห์ข้อมูล.....	136
การศึกษานำร่อง.....	140

สารบัญ (ต่อ)

บทที่

หน้า

4 ผลการวิเคราะห์ข้อมูล.....	144
การคำนวณค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อน	
ประเภทที่ 1.....	145
การวิเคราะห์ความแปรปรวนหลายตัวแปรของอัตราความถูกต้อง และ	
อัตราความคลาดเคลื่อนประเภทที่ 1.....	170
การวิเคราะห์ความแปรปรวนการทดสอบผลย่อยของอัตราความถูกต้อง และ	
อัตราความคลาดเคลื่อนประเภทที่ 1 พร้อมทั้งการเปรียบเทียบภายหลัง.....	176
สรุปผลการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อน	
ประเภทที่ 1.....	211
 5 สรุป อภิปรายผล และข้อเสนอแนะ.....	225
สรุปผลการวิจัย.....	228
อภิปรายผลการวิจัย.....	261
ข้อเสนอแนะ.....	279
 บรรณานุกรม.....	284
 ภาคผนวก.....	293
 ประวัติย่อผู้วิจัย.....	259

บัญชีตาราง

ตาราง

หน้า

- 1 ทิศทางของข้อสอบ พารามิเตอร์อำนาจจำแนกของข้อสอบในแต่ละมิติ
และพารามิเตอร์อำนาจจำแนกของข้อสอบแบบหลายมิติ สำหรับกลุ่มอ้างอิง
และกลุ่มสนใจ ซึ่งใช้จำลองข้อสอบทำหน้าที่ไม่เบี่ยงเบน..... 126
- 2 พารามิเตอร์รายการของข้อสอบ และพารามิเตอร์ความยากของข้อสอบแบบ
หลายมิติ สำหรับกลุ่มอ้างอิงและกลุ่มสนใจ ซึ่งใช้จำลองข้อสอบ
ทำหน้าที่ไม่เบี่ยงเบน..... 128
- 3 ทิศทางของข้อสอบ พารามิเตอร์อำนาจจำแนกของข้อสอบในแต่ละมิติ
และพารามิเตอร์อำนาจจำแนกของข้อสอบแบบหลายมิติ สำหรับกลุ่มอ้างอิง
และกลุ่มสนใจ ซึ่งใช้จำลองข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน..... 130
- 4 พารามิเตอร์รายการของข้อสอบ และพารามิเตอร์ความยากของข้อสอบแบบ
หลายมิติ สำหรับกลุ่มอ้างอิง ซึ่งใช้จำลองข้อสอบทำหน้าที่เบี่ยงเบน
ที่เป็นรูปแบบเดียวกัน..... 130
- 5 พารามิเตอร์รายการของข้อสอบ และพารามิเตอร์ความยากของข้อสอบแบบ
หลายมิติ สำหรับกลุ่มสนใจ ซึ่งใช้จำลองข้อสอบทำหน้าที่เบี่ยงเบน
ที่เป็นรูปแบบเดียวกัน..... 131
- 6 ทิศทางของข้อสอบ พารามิเตอร์อำนาจจำแนกของข้อสอบในแต่ละมิติ
และพารามิเตอร์อำนาจจำแนกของข้อสอบแบบหลายมิติ สำหรับกลุ่มอ้างอิง
ซึ่งใช้จำลองข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน..... 133
- 7 ทิศทางของข้อสอบ พารามิเตอร์อำนาจจำแนกของข้อสอบในแต่ละมิติ
และพารามิเตอร์อำนาจจำแนกของข้อสอบแบบหลายมิติ สำหรับกลุ่มสนใจ
ซึ่งใช้จำลองข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน..... 133
- 8 พารามิเตอร์รายการของข้อสอบ และพารามิเตอร์ความยากของข้อสอบแบบ
หลายมิติ สำหรับกลุ่มอ้างอิง ซึ่งใช้จำลองข้อสอบทำหน้าที่เบี่ยงเบน
ที่ไม่เป็นรูปแบบเดียวกัน..... 134

บัญชีตาราง (ต่อ)

ตาราง

หน้า

9	พารามิเตอร์รายการของข้อสอบ และพารามิเตอร์ความยากของข้อสอบแบบ หลายมิติ สำหรับกลุ่มสนใจ ซึ่งใช้จำลองข้อสอบทำหน้าที่เบี่ยงเบน ที่ไม่เป็นรูปแบบเดียวกัน.....	134
10	ค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของ วิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติก แบบจัดอันดับหลายมิติ ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย.....	141
11	ค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของ วิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติก แบบจัดอันดับหลายมิติ ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัยหลัก.....	146
12	ค่าเฉลี่ยของอัตราความถูกต้อง (P) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ของวิธีการตรวจสอบ ภายใต้ปฏิสัมพันธ์แบบสองทางระหว่างปัจจัยรูปแบบของ ข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน.....	148
13	ค่าเฉลี่ยของอัตราความถูกต้อง (P) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ของวิธีการตรวจสอบ ภายใต้ปฏิสัมพันธ์แบบสองทางระหว่างปัจจัยรูปแบบของ ข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถ...	149
14	ค่าเฉลี่ยของอัตราความถูกต้อง (P) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ของวิธีการตรวจสอบ ภายใต้ปฏิสัมพันธ์แบบสองทางระหว่างปัจจัยรูปแบบของ ข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยขนาดตัวอย่าง.....	150
15	ค่าเฉลี่ยของอัตราความถูกต้อง (P) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ของวิธีการตรวจสอบ ภายใต้ปฏิสัมพันธ์แบบสองทางระหว่างปัจจัยสัดส่วนของ ข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถ...	152
16	ค่าเฉลี่ยของอัตราความถูกต้อง (P) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ของวิธีการตรวจสอบ ภายใต้ปฏิสัมพันธ์แบบสองทางระหว่างปัจจัยสัดส่วนของ ข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยขนาดตัวอย่าง.....	153

บัญชีตาราง (ต่อ)

ตาราง

หน้า

17	ค่าเฉลี่ยของอัตราความถูกต้อง (P) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ของวิธีการตรวจสอบ ภายใต้ปฏิสัมพันธ์แบบสองทางระหว่างปัจจัยความแตกต่างของการแจกแจงความสามารถ กับปัจจัยขนาดตัวอย่าง.....	155
18	ค่าเฉลี่ยของอัตราความถูกต้อง (P) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ของวิธีการตรวจสอบ ภายใต้ปฏิสัมพันธ์แบบสามทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถ.....	157
19	ค่าเฉลี่ยของอัตราความถูกต้อง (P) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ของวิธีการตรวจสอบ ภายใต้ปฏิสัมพันธ์แบบสามทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยขนาดตัวอย่าง.....	159
20	ค่าเฉลี่ยของอัตราความถูกต้อง (P) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ของวิธีการตรวจสอบ ภายใต้ปฏิสัมพันธ์แบบสามทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ปัจจัยความแตกต่างของการแจกแจงความสามารถ กับปัจจัยขนาดตัวอย่าง.....	161
21	ค่าเฉลี่ยของอัตราความถูกต้อง (P) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ของวิธีการตรวจสอบ ภายใต้ปฏิสัมพันธ์แบบสามทางระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ปัจจัยความแตกต่างของการแจกแจงความสามารถ กับปัจจัยขนาดตัวอย่าง.....	163
22	ค่าเฉลี่ยของอัตราความถูกต้อง (P) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ของวิธีการตรวจสอบ ภายใต้ปฏิสัมพันธ์แบบสี่ทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ปัจจัยความแตกต่างของการแจกแจงความสามารถ กับปัจจัยขนาดตัวอย่าง.....	166
23	การเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย.....	171

บัญชีตาราง (ต่อ)

ตาราง

หน้า

24	การทดสอบผลระหว่างกลุ่มของอัตราความถูกต้อง และอัตราความคลาดเคลื่อน ประเภทที่ 1 ภายใต้เงื่อนไขปฏิสัมพันธ์สองทางและสามทาง ระหว่างปัจจัยที่ แปรเปลี่ยน 4 ปัจจัย กับวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ.....	174
25	การทดสอบผลย่อยของปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน และ วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ.....	178
26	การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อน ประเภทที่ 1 ระหว่างวิธีการตรวจสอบสามวิธี ภายใต้เงื่อนไขเดียวกัน ของปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน.....	295
27	การทดสอบผลย่อยของปัจจัยรูปสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และ วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ.....	180
28	การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อน ประเภทที่ 1 ของวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขต่างกันของ ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน.....	295
29	การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อน ประเภทที่ 1 ระหว่างวิธีการตรวจสอบสามวิธี ภายใต้เงื่อนไขเดียวกัน ของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน.....	296
30	การทดสอบผลย่อยของปัจจัยความแตกต่างของการแจกแจงความสามารถ ระหว่างกลุ่มผู้สอบ และวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ.....	182
31	การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อน ประเภทที่ 1 ของแต่ละวิธีการตรวจสอบ ภายใต้เงื่อนไขต่างกัน ของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน.....	297
32	การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อน ประเภทที่ 1 ระหว่างวิธีการตรวจสอบสามวิธี ภายใต้เงื่อนไขเดียวกัน ของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน.....	298

บัญชีตาราง (ต่อ)

ตาราง

หน้า

33 การทดสอบผลย่อยของปัจจัยขนาดตัวอย่าง และวิธีการตรวจสอบการทำหน้าที่ เบี่ยงเบนของข้อสอบ.....	185
34 การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อน ประเภทที่ 1 ของแต่ละวิธีการตรวจสอบ ภายใต้เงื่อนไขต่างกัน ของปัจจัยขนาดตัวอย่าง.....	299
35 การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อน ประเภทที่ 1 ระหว่างวิธีการตรวจสอบสามวิธี ภายใต้เงื่อนไขเดียวกัน ของปัจจัยขนาดตัวอย่าง.....	300
36 การทดสอบผลย่อยของปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของข้อสอบ ทำหน้าที่เบี่ยงเบน กับปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และ วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ.....	187
37 การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อน ประเภทที่ 1 ของแต่ละวิธีการตรวจสอบ ภายใต้เงื่อนไขต่างกัน ของปฏิสัมพันธ์ระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน.....	301
38 การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อน ประเภทที่ 1 ระหว่างวิธีการตรวจสอบสามวิธี ภายใต้เงื่อนไขเดียวกัน ของปฏิสัมพันธ์ระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถ.....	303
39 การทดสอบผลย่อยของปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของข้อสอบ ทำหน้าที่เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถ และ วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ.....	191
40 การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อน ประเภทที่ 1 ของแต่ละวิธีการตรวจสอบ ภายใต้เงื่อนไขต่างกันของ ปฏิสัมพันธ์ระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถ.....	304

บัญชีตาราง (ต่อ)

ตาราง

หน้า

- 41 การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อน
ประเภทที่ 1 ระหว่างวิธีการตรวจสอบสามวิธี ภายใต้เงื่อนไขเดียวกันของ
ปฏิสัมพันธ์ระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เพียงแบบ
กับปัจจัยความแตกต่างของการแจกแจงความสามารถ..... 306
- 42 การทดสอบผลย่อยของปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของข้อสอบทำ
หน้าที่เพียงแบบ กับปัจจัยขนาดตัวอย่าง และวิธีการตรวจสอบการทำหน้าที่
เพียงแบบของข้อสอบ..... 195
- 43 การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อน
ประเภทที่ 1 ของแต่ละวิธีการตรวจสอบ ภายใต้เงื่อนไขต่างกันของ
ปฏิสัมพันธ์ระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เพียงแบบ
กับปัจจัยขนาดตัวอย่าง..... 307
- 44 การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อน
ประเภทที่ 1 ระหว่างวิธีการตรวจสอบสามวิธี ภายใต้เงื่อนไขเดียวกันของ
ปฏิสัมพันธ์ระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เพียงแบบ
กับปัจจัยขนาดตัวอย่าง..... 310
- 45 การทดสอบผลย่อยของปฏิสัมพันธ์สองทางระหว่างปัจจัยสัดส่วนของข้อสอบทำ
หน้าที่เพียงแบบ กับปัจจัยความแตกต่างของการแจกแจงความสามารถ และ
วิธีการตรวจสอบการทำหน้าที่เพียงแบบของข้อสอบ..... 200
- 46 การเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ของ
วิธีการถดถอยโลจิสติกแบบจัดอันดับ ภายใต้เงื่อนไขต่างกันของ
ปฏิสัมพันธ์ระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เพียงแบบ
กับปัจจัยความแตกต่างของการแจกแจงความสามารถ..... 312
- 47 การเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่าง
วิธีการตรวจสอบสามวิธี ภายใต้เงื่อนไขเดียวกันของปฏิสัมพันธ์ระหว่าง
ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เพียงแบบ กับปัจจัยความแตกต่าง
ของการแจกแจงความสามารถ..... 314

บัญชีตาราง (ต่อ)

ตาราง

หน้า

48	การทดสอบผลย่อยของปฏิสัมพันธ์สองทางระหว่างปัจจัยสัดส่วนของข้อสอบ ทำหน้าที่เพียงเบน กับปัจจัยขนาดตัวอย่าง และวิธีการตรวจสอบการทำหน้าที่ เพียงเบนของข้อสอบ.....	202
49	การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อน ประเภทที่ 1 ของแต่ละวิธีการตรวจสอบ ภายใต้เงื่อนไขต่างกันของปฏิสัมพันธ์ ระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เพียงเบน กับปัจจัยขนาดตัวอย่าง.....	316
50	การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อน ประเภทที่ 1 ระหว่างวิธีการตรวจสอบสามวิธี ภายใต้เงื่อนไขเดียวกันของ ปฏิสัมพันธ์ระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เพียงเบน กับปัจจัยขนาดตัวอย่าง.....	321
51	การทดสอบผลย่อยของปฏิสัมพันธ์สองทางระหว่างปัจจัยความแตกต่างของ การแจกแจงความสามารถ กับปัจจัยขนาดตัวอย่าง และวิธีการตรวจสอบ การทำหน้าที่เพียงเบนของข้อสอบ.....	206
52	การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อน ประเภทที่ 1 ของแต่ละวิธีการตรวจสอบ ภายใต้เงื่อนไขต่างกันของปฏิสัมพันธ์ ระหว่างปัจจัยความแตกต่างของการแจกแจงความสามารถ กับปัจจัยขนาดตัวอย่าง.....	323
53	การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อน ประเภทที่ 1 ระหว่างวิธีการตรวจสอบสามวิธี ภายใต้เงื่อนไขเดียวกันของ ปฏิสัมพันธ์ระหว่างปัจจัยความแตกต่างของการแจกแจงความสามารถ กับปัจจัยขนาดตัวอย่าง.....	330
54	สรุปปัจจัยที่มีผลต่ออัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และ วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ.....	216

บัญชีตาราง (ต่อ)

ตาราง

หน้า

- 55 สรุปผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง ระหว่างวิธีโพลี-ชิปเทสท์
วิธีการถอดรอยโลจิสติกแบบจัดอันดับ และวิธีการถอดรอยโลจิสติกแบบ
จัดอันดับหลายมิติ ภายใต้ทุกเงื่อนไขของปัจจัยที่ศึกษา..... 218
- 56 สรุปผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1
ระหว่างวิธีโพลี-ชิปเทสท์ วิธีการถอดรอยโลจิสติกแบบจัดอันดับ และ
วิธีการถอดรอยโลจิสติกแบบจัดอันดับหลายมิติ ภายใต้ทุกเงื่อนไข
ของปัจจัยที่ศึกษา..... 219
- 57 อัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของ
วิธีการถอดรอยโลจิสติกแบบจัดอันดับหลายมิติ ที่ระดับนัยสำคัญ 5 ระดับ..... 222

บัญชีภาพประกอบ

ภาพประกอบ

หน้า

1	ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF) ภายใต้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ (3PLM).....	34
2	ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) ภายใต้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ (3PLM).....	34
3	ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF) ภายใต้โมเดลเกรดเรสพอน (GRM).....	36
4	ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) ภายใต้โมเดลเกรดเรสพอน (GRM).....	36
5	ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF) ภายใต้โมเดลพหุคูณเครดิตทั่วไป (GPCM).....	37
6	ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) ภายใต้โมเดลพหุคูณเครดิตทั่วไป (GPCM).....	37
7	ข้อสอบทำหน้าที่เบี่ยงเบนรูปแบบคงที่ (Constant DIF) ภายใต้โมเดลพหุคูณเครดิต (PCM).....	38
8	ข้อสอบทำหน้าที่เบี่ยงเบนแบบเปลี่ยนขนาดที่ระดับความสามารถต่ำ (Low-shift DIF) ภายใต้โมเดลพหุคูณเครดิต (PCM).....	39
9	ข้อสอบทำหน้าที่เบี่ยงเบนแบบเปลี่ยนขนาดที่ระดับความสามารถสูง (High-shift DIF) ภายใต้โมเดลพหุคูณเครดิต (PCM).....	39
10	ข้อสอบทำหน้าที่เบี่ยงเบนแบบสมดุล (Balanced DIF) ภายใต้โมเดลพหุคูณเครดิตทั่วไป (PCM).....	40
11	ฟังก์ชันการตอบข้อสอบ (IRF) ภายใต้โมเดลโลจิสติกแบบ 3 พารามิเตอร์มิติเดียว (U3PLM).....	48
12	ผิวหน้าการตอบข้อสอบ (IRS) ภายใต้โมเดลโลจิสติกแบบ 3 พารามิเตอร์หลายมิติ (M3PLM).....	52

บัญชีภาพประกอบ (ต่อ)

ภาพประกอบ

หน้า

13	ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF) ภายใต้โมเดลพาเซียนเคเรดิททั่วไปแบบหลายมิติ (MGPCM).....	112
14	ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) ภายใต้โมเดลพาเซียนเคเรดิททั่วไปแบบหลายมิติ (MGPCM).....	113
15	เส้นโครงร่างของการแจกแจงความสามารถแบบสองมิติของกลุ่มอ้างอิงและ กลุ่มสนใจ เมื่อค่าเฉลี่ยของความสามารถแตกต่างกัน 0.5SD.....	116
16	เส้นโครงร่างของการแจกแจงความสามารถแบบสองมิติของกลุ่มอ้างอิงและ กลุ่มสนใจ เมื่อค่าเฉลี่ยของความสามารถแตกต่างกัน 1.0SD.....	116
17	การแจกแจงความสามารถแบบสองมิติ เมื่อค่าเฉลี่ยของความสามารถเท่ากับ 0.0...	117
18	การแจกแจงความสามารถแบบสองมิติ เมื่อค่าเฉลี่ยของความสามารถเท่ากับ -0.5..	117
19	การแจกแจงความสามารถแบบสองมิติ เมื่อค่าเฉลี่ยของความสามารถเท่ากับ -1.0..	118
20	แผนผังของการจำลองข้อมูล.....	120
21	เวกเตอร์ของข้อสอบ 3 ข้อ.....	124
22	แสดงอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน 2 รูปแบบ.....	211
23	แสดงอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 3 ขนาด.....	212
24	แสดงอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ปัจจัยความแตกต่างของการแจกแจงความสามารถ 3 ระดับ.....	213
25	แสดงอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ปัจจัยขนาดตัวอย่าง 4 ขนาด.....	214
26	แสดงอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ปัจจัยขนาดตัวอย่าง 3 ขนาด.....	223
27	แสดงอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ปัจจัยความแตกต่างของ การแจกแจงความสามารถ 3 ระดับ.....	224

บทที่ 1

บทนำ

ภูมิหลัง

“ความยุติธรรมของแบบทดสอบ” (Test Fairness) เริ่มมีการศึกษาอย่างจริงจังตั้งแต่ทศวรรษ 1960 ในประเทศสหรัฐอเมริกา ทั้งนี้เนื่องจากสังคมอเมริกันมีความแตกต่างในเรื่องของวัฒนธรรมระหว่างชนกลุ่มน้อยและชนกลุ่มใหญ่ ทำให้ผลการทดสอบไม่เท่าเทียมกัน โดยเฉพาะนักเรียนที่เป็นชนกลุ่มน้อย เป็นต้นว่า นักเรียนอเมริกันผิวดำ นักเรียนอเมริกันเชื้อสายสเปน นักเรียนอเมริกันเชื้อสายเอเชีย และนักเรียนอเมริกันพื้นเมือง นักเรียนในกลุ่มดังกล่าวมักเสียเปรียบนักเรียนผิวขาวซึ่งเป็นชนกลุ่มใหญ่ เนื่องจากข้อสอบส่วนมากวัดความรู้นอกเหนือวัฒนธรรมของชนกลุ่มน้อย ความไม่ยุติธรรมของการตอบข้อสอบเกิดจาก “**ความลำเอียงของข้อสอบ หรือ แบบทดสอบ**” (Item bias/Test bias) ทำให้ชนกลุ่มน้อยเสียเปรียบชนกลุ่มใหญ่ ความแตกต่างของผลการตอบข้อสอบระหว่างกลุ่มดังกล่าว เป็นผลมาจากความแตกต่างของวัฒนธรรมซึ่งเป็นภูมิหลังของผู้สอบมากกว่าความสามารถของผู้สอบ (Angoff. 1993: 3) ดังนั้นความลำเอียงของข้อสอบหรือแบบทดสอบ จึงเป็นประเด็นสำคัญประเด็นหนึ่งของการทดสอบทางการศึกษาและจิตวิทยา ซึ่งจำเป็นต้องมีการตรวจสอบความลำเอียงของข้อสอบ เช่นเดียวกับการตรวจสอบความเที่ยงตรง ความเชื่อมั่น และคุณลักษณะของข้อสอบที่เคยปฏิบัติกันมา

นักวิจัยในปัจจุบันจึงให้ความสนใจในประเด็นดังกล่าวมากยิ่งขึ้น เพราะในวงการศึกษาและหน่วยงานต่างๆ ได้นำผลการทดสอบมาใช้พิจารณาตัดสินใจในเรื่องต่างๆ มากยิ่งขึ้น ดังเช่น การคัดเลือกเพื่อเข้าศึกษาต่อ การคัดเลือกเพื่อบรรจุงาน การคัดเลือกเพื่อเลื่อนขั้นหรือตำแหน่ง การออกใบอนุญาตหรือใบรับรอง เป็นต้น แบบทดสอบที่นำมาใช้ควรเป็นแบบทดสอบมาตรฐานที่มีความเที่ยงตรง ความเชื่อมั่น ความยากเหมาะสม ตลอดจนจำแนกความสามารถของผู้สอบได้ นอกจากนี้ยังต้องคำนึงถึงความยุติธรรมต่อผู้สอบ แบบทดสอบที่ขาดความยุติธรรมจะเกิดความลำเอียง โดยข้อสอบหรือแบบทดสอบอาจเข้าข้างผู้สอบกลุ่มใดกลุ่มหนึ่ง ซึ่งมีผลทำให้ผู้สอบกลุ่มนั้นได้เปรียบในการตอบข้อสอบ ส่วนผู้สอบอีกกลุ่มหนึ่งเสียเปรียบ ทั้งๆ ที่ผู้สอบสองกลุ่มมีระดับความสามารถเท่ากัน ความลำเอียงของแบบทดสอบดังกล่าว ถือเป็นความคลาดเคลื่อนอย่างเป็นระบบ (Systematic error) (Camilli; & Shepard. 1994: 8) ดังนั้นการตรวจสอบความลำเอียงของ

ข้อสอบจึงมีความจำเป็นสำหรับนักพัฒนาแบบทดสอบ เพื่อใช้เป็นหลักฐานที่แสดงความยุติธรรมของแบบทดสอบ (AERA/APA/NCME. 1999: 74) แล้วคัดเลือกข้อสอบที่มีความลำเอียงออกจากแบบทดสอบ กระบวนการตรวจสอบดังกล่าวถือเป็นส่วนหนึ่งของการพัฒนาแบบทดสอบ

ในปัจจุบันนักวิจัยใช้คำว่า “**การทำหน้าที่เบี่ยงเบนของข้อสอบ**” (Differential item functioning) หรือเรียกว่า “DIF” แทน “**ความลำเอียงของข้อสอบ**” (Item bias) เนื่องจากคำว่า “**ความลำเอียงของข้อสอบ**” ที่ใช้กันมาในอดีตมีสองความหมาย คือ ความหมายในเชิงสถิติและความหมายในเชิงสังคม ความหมายทั้งสองเป็นประเด็นที่มีการโต้เถียงว่ามีความขัดแย้งกัน ต่อมาได้มีผู้เสนอให้ใช้คำว่า “**การทำหน้าที่เบี่ยงเบนของข้อสอบ**” ซึ่งเน้นหลักฐานเชิงประจักษ์ที่ได้จากการทดสอบทางสถิติ โดยแยกออกจากการประเมินที่มีความหมายทางสังคม (Angoff. 1993: 4-5; Hambleton; Swaminathan; & Rogers. 1991: 109) คำดังกล่าวมีความเหมาะสมมากกว่าในการอธิบายสารสนเทศเชิงสถิติ นักวิจัยจึงนำมาใช้กันอย่างแพร่หลายจนถึงทุกวันนี้ ส่วนการกำหนดความหมายของการทำหน้าที่เบี่ยงเบนของข้อสอบ นักวิจัยส่วนมากให้ความหมายในทำนองเดียวกัน สามารถสรุปได้ว่า “**การทำหน้าที่เบี่ยงเบนของข้อสอบเกิดขึ้นเมื่อ ผู้สอบจากกลุ่มที่แตกต่างกัน และมีการจับคู่ความสามารถที่แบบทดสอบหรือข้อสอบต้องการวัดเท่ากัน มีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องไม่เท่ากัน**” (Camilli. 1993: 397; Dorans; & Holland. 1993: 37; Clauser; & Mazor. 1998: 40; Gierl; Gotzmann; & Boughton. 2004: 242; Li; & Stout. 1996: 647; Potenza; & Dorans. 1995: 23; Raju; & Ellis. 2002: 156; Zumbo. 1999: 12)

กรอบแนวคิดเชิงทฤษฎีในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบจะเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มผู้สอบสองกลุ่ม โดยกำหนดให้กลุ่มหนึ่งเป็น “**กลุ่มสนใจ**” (Focal group) หรือ “**กลุ่มรอง**” (Minority group) ซึ่งเป็นกลุ่มเป้าหมายที่ต้องการศึกษา และอีกกลุ่มหนึ่งเป็น “**กลุ่มอ้างอิง**” (Reference group) หรือ “**กลุ่มหลัก**” (Majority group) ซึ่งเป็นกลุ่มที่ใช้เปรียบเทียบกับกลุ่มสนใจ ข้อสอบที่ใช้ในการตรวจสอบเรียกว่า “**ข้อสอบศึกษา**” (Studied item) เมื่อข้อสอบทำหน้าที่เบี่ยงเบน กลุ่มอ้างอิงเป็นกลุ่มที่คาดว่าข้อสอบเข้าข้าง ส่วนกลุ่มสนใจเป็นกลุ่มที่คาดว่าข้อสอบไม่เข้าข้าง (Donoghue; Holland; & Thayer. 1993: 137-138; Roussos; & Stout. 1996a: 216) การทำหน้าที่เบี่ยงเบนของข้อสอบแสดงถึงข้อสอบในแบบทดสอบแตกต่างกันในเชิงสถิติ ซึ่งเป็นความแตกต่างของผลการตอบข้อสอบที่ศึกษาระหว่างกลุ่มอ้างอิงและกลุ่มสนใจ ภายหลังจากการควบคุมความแตกต่างระดับความสามารถจริงระหว่างกลุ่มผู้สอบสองกลุ่ม

การควบคุมความแตกต่างสามารถทำได้โดยการเปรียบเทียบผลการตอบข้อสอบภายใต้ตัวแปรการจับคู่ (Matching variable) วิธีที่ใช้ทฤษฎีการทดสอบแบบมาตรฐานเดิม (Classical test theory; CTT) มักใช้คะแนนรวมของแบบทดสอบเป็นเกณฑ์การจับคู่ ส่วนวิธีที่ใช้ทฤษฎีการตอบข้อสอบ (Item response theory; IRT) จะใช้ค่าประมาณระดับความสามารถของผู้สอบเป็นเกณฑ์การจับคู่ ถ้าความแตกต่างของผลการตอบข้อสอบระหว่างกลุ่มผู้สอบเกิดขึ้นจากการแจกแจงความสามารถของผู้สอบทั้งหมด ซึ่งไม่ได้มีการควบคุมตัวแปรดังกล่าวจะเรียกว่า **“ผลกระทบของข้อสอบ”** (Item impact) (Ackerman. 1992: 69; Camilli. 1992: 131; Dorans; & Holland. 1993: 37; Donoghue; Holland; & Thayer. 1993: 141)

ขนาดและทิศทางของการทำหน้าที่เบี่ยงเบนของข้อสอบสามารถแปรเปลี่ยนตามระดับความสามารถของผู้สอบ ดังจะเห็นได้จากเมลเลนเบิร์ก (Mellenbergh. 1982: 110) ได้จำแนกข้อสอบทำหน้าที่เบี่ยงเบนออกเป็นสองรูปแบบ คือ **“ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน”** (Uniform DIF) และ **“ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน”** (Non-uniform DIF) โดยใช้โมเดลล็อก-ลิเนียร์ (Log-linear model) ทดสอบนัยสำคัญภายใต้ข้อสอบที่ให้คะแนนสองค่า (Dichotomous) ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกันเกิดขึ้นเมื่อไม่มีปฏิสัมพันธ์ระหว่างระดับความสามารถ กับการเป็นสมาชิกของกลุ่ม หรืออธิบายได้ว่าความน่าจะเป็นในการตอบข้อสอบถูกของผู้สอบกลุ่มหนึ่งมากกว่าอีกกลุ่มหนึ่งอย่างคงที่ทุกระดับความสามารถ ซึ่งภายใต้ทฤษฎีการตอบข้อสอบสามารถพิจารณาได้จากฟังก์ชันการตอบข้อสอบ (Item response functions; IRFs) ระหว่างกลุ่มผู้สอบที่ขนานกัน ส่วนข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกันเกิดขึ้นเมื่อมีปฏิสัมพันธ์ระหว่างระดับความสามารถ กับการเป็นสมาชิกของกลุ่ม แสดงว่าความแตกต่างของความน่าจะเป็นในการตอบข้อสอบถูกของผู้สอบสองกลุ่มไม่เหมือนกันทุกระดับความสามารถ ซึ่งพิจารณาได้จากฟังก์ชันการตอบข้อสอบระหว่างกลุ่มผู้สอบที่ไม่ขนานกัน (Li; & Stout. 1996: 647; Swaminathan; & Rogers. 1990: 361) สำหรับรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนสองรูปแบบที่กล่าวมานั้น ในงานวิจัยของซีลลีและสเตาท์ (Shealy; & Stout. 1993) ได้ใช้คำว่า **“ข้อสอบทำหน้าที่เบี่ยงเบนที่มีทิศทางเดียวกัน”** (Unidirectional DIF) และ **“ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่มีทิศทางเดียวกัน”** (Nonunidirectional DIF) แทนชื่อทั้งสอง ต่อมา ลิและสเตาท์ (Li; & Stout. 1996: 648) ได้ใช้คำใหม่ว่า **“ข้อสอบทำหน้าที่เบี่ยงเบนแบบไม่ตัดกัน”** (Non-crossing DIF) และ **“ข้อสอบทำหน้าที่เบี่ยงเบนแบบตัดกัน”** (Crossing DIF) แต่คำดังกล่าวไม่เป็นที่ยอมรับใช้กัน นอกจากนี้ยังมีการศึกษาการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนน

หลายค่า (Polytomous) (Zwick; Donoghue; & Grima. 1993) เช่น รูปแบบคงที่ (Constant DIF) รูปแบบเปลี่ยนขนาดที่ระดับความสามารถต่ำ (Low-shift DIF) รูปแบบเปลี่ยนขนาดที่ระดับความสามารถสูง (High-shift DIF) และรูปแบบสมดุล (Balanced DIF) เป็นต้น

วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบได้มีการพัฒนามาตลอดตั้งแต่อดีตจนถึงปัจจุบัน แนวคิดในระยะแรกมุ่งเน้นการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถมิติเดียวและให้คะแนนสองค่า (Unidimensional dichotomous) โดยวิเคราะห์ภายใต้ทฤษฎีการทดสอบแบบมาตรฐานเดิม ตามปกติใช้คะแนนรวมของแบบทดสอบ ซึ่งเป็นคะแนนที่ได้จากการสอบเป็นเกณฑ์การจับคู่ภายใน (Internal matching criterion) ต่อมา มีการนำทฤษฎีการตอบข้อสอบมาประยุกต์ใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ โดยไม่ใช้คะแนนรวมของผู้สอบเหมือนกับทฤษฎีการทดสอบแบบมาตรฐานเดิม แต่ใช้ค่าประมาณระดับความสามารถของผู้สอบ ซึ่งเป็นตัวแปรคุณลักษณะแฝงเป็นเกณฑ์การจับคู่ วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบภายใต้สองทฤษฎีดังกล่าวมีจุดเริ่มต้นและพัฒนาการมาเป็นลำดับ พอที่จะกล่าวเป็นสังเขปดังนี้

กลุ่มวิธีที่ใช้ทฤษฎีการทดสอบแบบมาตรฐานเดิม (Classical test theory methods)
 ในยุคเริ่มต้น คาร์เดลและคอฟฟ์แมน (Angoff. 1993: 5; citing Cardall; & Coffman. 1964) ได้นำ**วิธีการวิเคราะห์ความแปรปรวน** (Analysis of variance; ANOVA) มาประยุกต์ใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ด้วยการทดสอบปฏิสัมพันธ์ระหว่างข้อสอบ กับกลุ่มผู้สอบ ถ้าผลการทดสอบมีนัยสำคัญแสดงว่าแบบทดสอบทำหน้าที่เบี่ยงเบน ส่วนการตัดสินใจการทำหน้าที่เบี่ยงเบนของข้อสอบต้องวิเคราะห์ภายหลังต่อไป การตรวจสอบด้วยวิธีนี้เป็นจุดเริ่มต้นของการพัฒนาวิธีอื่นๆ ดังเช่น แองกอฟฟ์ (Angoff. 1993: 5; citing Angoff. 1972) ได้เสนอ**วิธีแปลงค่าความยากของข้อสอบ** (Transformed item difficulty; TID) แนวคิดของวิธีนี้ใช้การทดสอบปฏิสัมพันธ์ระหว่างข้อสอบ กับกลุ่มผู้สอบ ซึ่งคล้ายกับเทคนิคการวิเคราะห์ความแปรปรวน แต่ตรวจสอบด้วยการลงจุดเดลต้า (Delta-plot) ระหว่างกลุ่มผู้สอบ ซึ่งคำนวณมาจากค่าความยากของข้อสอบ (p-values) ในแต่ละกลุ่มผู้สอบ แล้วแปลงเป็นคะแนนมาตรฐาน จากนั้นพิจารณาค่าเดลต้าของข้อสอบข้อใดที่เบี่ยงเบนผิดปกติไปจากเส้นแกนหลัก (Major axis) เพื่อตัดสินใจการทำหน้าที่เบี่ยงเบนของข้อสอบ การตรวจสอบด้วยวิธีนี้มักมีปัญหาเกี่ยวกับความแตกต่างของความสามารถระหว่างกลุ่มผู้สอบ ซึ่งเกิดขึ้นจากการปะปนกันของความแตกต่างของค่าความยากกับค่าอำนาจจำแนกของข้อสอบ ถ้าข้อสอบมีค่าอำนาจจำแนกสูงแล้วความแตกต่างในค่าความยาก

ของข้อสอบจะมีค่าสูงผิดปกติ ส่วนข้อสอบที่มีค่าอำนาจจำแนกต่ำแล้วความแตกต่างในค่าความยากของข้อสอบจะมีค่าต่ำมาก ความแตกต่างในการคำนวณค่าความยากดังกล่าวมีผลทำให้การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบเกิดความคลาดเคลื่อนประเภทที่ 1 และความคลาดเคลื่อนประเภทที่ 2 (Angoff. 1993: 9-10; Camilli; & Shepard. 1994: 27-28)

เชาเนแมน (Scheuneman. 1979) ได้เสนอ**วิธีไค-กำลังสอง** โดยพัฒนาเป็นอิสระกับวิธีที่ใช้ทฤษฎีการตอบข้อสอบ แต่มีเทคนิคการตรวจสอบคล้ายคลึงกัน หลักการตรวจสอบจะใช้ตารางการถ่วง (Contingency table) โดยแบ่งคะแนนรวมของแบบทดสอบออกเป็นช่วงย่อยๆ ประมาณ 3-5 ช่วง คำนวณสัดส่วนของผู้สอบที่ตอบข้อสอบถูกในแต่ละช่วงคะแนน แล้วทดสอบด้วยสถิติไค-กำลังสอง ต่อมาเบเกอร์ (Baker. 1981: 59-62) ได้วิพากษ์วิจารณ์การตรวจสอบของเชาเนแมนว่า ให้ผลการตรวจสอบไม่สอดคล้องกัน เนื่องจากอิทธิพลของขนาดตัวอย่างระหว่างกลุ่มผู้สอบและการแจกแจงสุ่มไม่เหมือนกับไค-กำลังสอง ดังนั้นจึงไม่มีคุณสมบัติเพียงพอที่จะนำไปใช้ทดสอบนัยสำคัญทางสถิติ จากข้อวิจารณ์ดังกล่าว เชาเนแมน (Scheuneman. 1981) ได้ปรับแก้โดยนำผลการตอบข้อสอบผิดเข้าร่วมการวิเคราะห์ด้วย ซึ่งเรียกว่า **“สูตรไค-กำลังสองแบบเต็มรูป”** (Full chi-square) ทำให้สามารถแก้ไขปัญหาดังกล่าวได้ จุดเด่นของวิธีไค-กำลังสองคือ มีกระบวนการตรวจสอบคล้ายกับวิธีที่ใช้ทฤษฎีการตอบข้อสอบ โดยมีการจับคู่ความสามารถของกลุ่มผู้สอบก่อนที่จะเปรียบเทียบผลการตอบข้อสอบ ทำให้สามารถแก้ปัญหาความแตกต่างความสามารถของผู้สอบได้ แต่ก็ยังไม่สมบูรณ์เพราะเป็นการเปรียบเทียบฟังก์ชันการตอบข้อสอบอย่างหยาบๆ ถ้าแบ่งช่วงคะแนนกว้างเกินไปอาจทำให้ระดับความสามารถระหว่างกลุ่มผู้สอบไม่เท่าเทียมกัน ดังนั้นผลการตรวจสอบอาจคลาดเคลื่อนจากความเป็นจริงได้

จากแนวคิดการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยเทคนิคไค-กำลังสองทำให้นักวิจัยหลายคนได้คิดค้นและพัฒนาวิธีการตรวจสอบให้มีความเหมาะสมมากยิ่งขึ้น ดังเช่นวิธีล็อก-ลิเนียร์ (Log-linear) (Mellenbergh. 1982) วิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) (Holland; & Thayer. 1988) วิธีการทำให้เป็นมาตรฐาน (Standardization; STD) (Dorans; & Kulick. 1986) วิธีการถดถอยโลจิสติก (Logistic regression; LR) (Swaminathan; & Rogers. 1990) เป็นต้น สำหรับ**วิธีล็อก-ลิเนียร์** เมลเลนเบิร์ก (Mellenbergh. 1982) ได้นำโมเดลล็อก-ลิเนียร์มาประยุกต์ใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ โดยใช้ตารางการถ่วงแบบสามทาง คือ ระดับความสามารถ กลุ่มผู้สอบ และผลการตอบข้อสอบ แล้วทดสอบด้วยสถิติอัตราส่วนความน่าจะเป็น (Likelihood ratio) วิธีล็อก-ลิเนียร์มีจุดเด่นตรงที่สามารถใช้โมเดลตรวจสอบการทำ

หน้าที่เบี่ยงเบนของข้อสอบ ทั้งที่เป็นรูปแบบเดียวกัน (Uniform DIF) และที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) ส่วนข้อจำกัดของวิธีล็อก-ลิเนียร์ คือ ในการวิเคราะห์จะทำการคำนวณแบบวนซ้ำ (Iterative algorithm) ซึ่งวิเคราะห์ยากกว่าวิธีอื่นๆ ที่กล่าวมา ทำให้เสียเวลาและค่าใช้จ่ายมากกว่า ส่วน**วิธีแมนเทล-แฮนส์เซล** (Mantel-Haenszel; MH) ฮอลแลนด์และทาเยอร์ (Holland; & Thayer. 1988) ได้นำสถิติแมนเทล-แฮนส์เซล ซึ่งนิยมใช้ในงานวิจัยทางการแพทย์มาปรับขยายเพื่อใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หลักการตรวจสอบจะเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มสนใจที่ระดับความสามารถเดียวกัน โดยนำข้อมูลผลการตอบมาวิเคราะห์ในตารางการถัวกันของเดียวกับวิธีล็อก-ลิเนียร์ จากนั้นประมาณค่าอัตราส่วนแต้มต่อรวม (Common odds ratio) ระหว่างกลุ่มในแต่ละระดับคะแนน แล้วทดสอบนัยสำคัญโดยใช้สถิติไค-กำลังสองแบบแมนเทล-แฮนส์เซล จุดเด่นของวิธีแมนเทล-แฮนส์เซล คือ ใช้คะแนนรวมของแบบทดสอบในแต่ละหนึ่งหน่วยของช่วงคะแนนเป็นตัวแปรการจับคู่ ทำให้หลีกเลี่ยงปัญหาเกี่ยวกับการยุบมาตราของคะแนนเป็นกลุ่มคะแนน ซึ่งเกิดขึ้นในวิธีไค-กำลังสองแบบดั้งเดิม นอกจากนี้ยังสามารถตรวจสอบอย่างง่าย ใช้ตัวอย่างขนาดเล็ก และไม่จำเป็นต้องใช้โปรแกรมคำนวณที่ซับซ้อน แต่มีข้อจำกัดที่ไม่สามารถตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) (Li; & Stout. 1996: 650)

ดูแรนส์และคูลิค (Dorans; & Kulick. 1986) ได้พัฒนา**วิธีการทำให้เป็นมาตรฐาน** (Standardization; STD) ซึ่งมีแนวคิดในการตรวจสอบคล้ายกับวิธีที่ใช้ทฤษฎีการตอบข้อสอบ โดยเปรียบเทียบโค้งการถดถอยของข้อสอบในแบบทดสอบ (Item-test regression curve) ระหว่างกลุ่มผู้สอบในแต่ละระดับคะแนน ทั้งยังมีความสัมพันธ์กับวิธีแมนเทล-แฮนส์เซล เว้นแต่จะไม่ใช้สถิติทดสอบนัยสำคัญ ศูนย์บริการทดสอบทางการศึกษา (Educational Testing Service; ETS) ของสหรัฐอเมริกาใช้ทั้งสองวิธีควบคู่กัน โดยวิธีแมนเทล-แฮนส์เซลใช้แสดงหลักฐานทางสถิติของการทำหน้าที่เบี่ยงเบนของข้อสอบในคลังข้อสอบ ส่วนวิธีการทำให้เป็นมาตรฐานใช้เป็นสารสนเทศส่วนหนึ่งในการตัดสินใจเกี่ยวกับข้อสอบที่ตรวจพบว่าทำหน้าที่เบี่ยงเบน (Donoghue; Holland; & Thayer. 1993: 137) สำหรับ**วิธีการถดถอยโลจิสติก** (Logistic regression; LR) สวามินาธานและโรเจอร์ส (Swaminathan; & Rogers. 1990) ได้ดัดแปลงมาจากวิธีล็อก-ลิเนียร์และวิธีแมนเทล-แฮนส์เซล อีกทั้งกระบวนการตรวจสอบใช้หลักเดียวกับวิธีที่ใช้ทฤษฎีการตอบข้อสอบ จุดเด่นของวิธีนี้คือ ใช้โมเดลการถดถอยโลจิสติก ซึ่งสามารถทดสอบอิทธิพลของปฏิสัมพันธ์ระหว่างระดับความสามารถ กับการเป็นสมาชิกของกลุ่มผู้สอบ ทำให้มีประสิทธิภาพในการตรวจสอบการทำ

หน้าที่เบี่ยงเบนของข้อสอบ ทั้งที่เป็นรูปแบบเดียวกัน (Uniform DIF) และที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) (Li; & Stout. 1996: 650) นอกจากนี้ยังใช้ตัวแปรความสามารถแบบต่อเนื่อง ทำให้มีความถูกต้องและแม่นยำในการตรวจสอบมากกว่าวิธีล็อก-ลิเนียร์ที่ใช้ตัวแปรความสามารถแบบไม่ต่อเนื่อง ส่วนข้อด้อยของวิธีนี้เหมือนกับวิธีล็อก-ลิเนียร์

กลุ่มวิธีที่ใช้ทฤษฎีการตอบข้อสอบ (Item response theory methods) วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในกลุ่มนี้ จำแนกออกเป็นสองกลุ่มใหญ่ คือ กลุ่มหนึ่งใช้การเปรียบเทียบค่าประมาณพารามิเตอร์ของข้อสอบระหว่างกลุ่มผู้สอบ ส่วนอีกกลุ่มหนึ่งใช้การเปรียบเทียบค่าประมาณฟังก์ชันการตอบข้อสอบระหว่างกลุ่มผู้สอบ โดยการวัดพื้นที่ระหว่างฟังก์ชันการตอบข้อสอบ แนวคิดและหลักการตรวจสอบในกลุ่มแรกใช้**การเปรียบเทียบค่าประมาณพารามิเตอร์ของข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มสนใจ โดยการทดสอบนัยสำคัญทางสถิติ** ในยุคเริ่มต้น ไรท์ มีด และดราบา (Millsap; & Everson. 1993: 36; citing Wright; Mead; & Draba. 1976) ได้ใช้สถิติดราบา (Draba statistic) ทดสอบความแตกต่างของค่าพารามิเตอร์ความยากของข้อสอบระหว่างกลุ่มผู้สอบ ภายใต้โมเดลของราสช์ (Rasch model) ต่อมา ลอร์ด (Lord. 1980) ได้พัฒนาอีกขั้นหนึ่ง โดยใช้สถิติโฮลเทลลิง (Holtelling statistic) ทดสอบความแตกต่างของค่าพารามิเตอร์ของข้อสอบระหว่างกลุ่มผู้สอบ ภายใต้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ วิธีนี้นิยมเรียกว่า **“การทดสอบไค-กำลังสองของลอร์ด”** ส่วนอีกทางเลือกหนึ่งจะทดสอบความแตกต่างของพารามิเตอร์ของข้อสอบระหว่างกลุ่มผู้สอบ โดยใช้สถิติอัตราส่วนความน่าจะเป็น (Likelihood ratio statistic) (Thissen; Steinberge; & Wainer. 1993) ทดสอบความเหมาะสมระหว่างโมเดลสองโมเดล ซึ่งแต่ละโมเดลประกอบด้วยค่าพารามิเตอร์ของข้อสอบ

แนวคิดและหลักการตรวจสอบในกลุ่มหลังใช้**การวัดพื้นที่ของฟังก์ชันการตอบข้อสอบ (Item response function; IRFs) ระหว่างกลุ่มอ้างอิงและกลุ่มสนใจโดยตรง** แล้วตัดสินการทำหน้าที่เบี่ยงเบนของข้อสอบโดยนำดัชนีที่คำนวณได้ไปเปรียบเทียบกับเกณฑ์ที่กำหนดขึ้น หรือใช้การทดสอบทางสถิติ การพัฒนาของวิธีในกลุ่มนี้เริ่มต้นจาก กรีนและดราเปอร์ (Rudner. 1977: 16; citing Green; & Draper. 1972) ได้ประยุกต์ใช้สัดส่วนของการตอบข้อสอบถูกในแต่ละระดับคะแนนรวมแทนความน่าจะเป็นในการตอบข้อสอบถูก และใช้คะแนนรวมแทนความสามารถของผู้สอบ แล้วนำไปลงจุดของแต่ละกลุ่มผู้สอบแยกกัน จากนั้นจะเปรียบเทียบจุดเหล่านั้นเพื่อตัดสินข้อสอบข้อใดทำหน้าที่เบี่ยงเบน ต่อมา รัดเนอร์ (Rudner. 1977: 32) ได้นำวิธีการดังกล่าวมาพัฒนาเป็นดัชนีการวัดพื้นที่แบบไม่คิดเครื่องหมาย ภายใต้โมเดลโลจิสติกแบบ 3 พารามิเตอร์

หลังจากนั้นได้มีนักวิจัยหลายคนได้พัฒนาสูตรการวัดพื้นที่ เช่น วิธีการวัดพื้นที่ของลินน์และคนอื่นๆ (Linn; et al. 1981) วิธีการวัดพื้นที่ของเซพาร์ด คามิลลี และวิลเลียมส์ (Shepard; Camilli; & Williams. 1984) วิธีการวัดพื้นที่ของราชู (Raju. 1990) วิธีการวัดพื้นที่ของคิมและโคเฮ็น (Kim; & Cohen. 1991) เป็นต้น ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบภายใต้ทฤษฎีการตอบข้อสอบ นอกจากจะใช้แนวคิดและหลักการสองแนวทางดังกล่าวแล้ว ราชู แวนเดอร์ลินเด็น และเฟลีย์ (Raju; van der Linden; & Fleer. 1995) ได้ใช้กรอบแนวคิดภายใต้ทฤษฎีการตอบข้อสอบทั่วไป (General IRT-base) ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบและแบบทดสอบ (DIF/DTF) ซึ่งเรียกว่า “วิธีดีเอฟไอที” (DFIT)

ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ นักวิจัยส่วนมากให้การยอมรับวิธีในกลุ่มทฤษฎีการตอบข้อสอบมากกว่าวิธีในกลุ่มทฤษฎีการทดสอบแบบมาตรฐานเดิม ทั้งนี้เนื่องจากการประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบข้อสอบไม่แปรเปลี่ยนตามกลุ่มผู้สอบ (Hambleton; Swaminathan; & Rogers. 1991: 18) ทำให้ความแตกต่างของความสามารถระหว่างกลุ่มผู้สอบไม่เป็นปัญหาในการตรวจสอบ อีกทั้งค่าประมาณพารามิเตอร์ของข้อสอบ ได้แก่ ค่าความยาก ค่าอำนาจจำแนก และค่าการเดาของข้อสอบมีการปะปนกันน้อยกว่าการใช้ทฤษฎีการทดสอบแบบมาตรฐานเดิม (Camilli; & Shepard. 1994: 47) ทำให้ความคลาดเคลื่อนในการตรวจสอบน้อยลง สำหรับวิธีการตรวจสอบโดยใช้ทฤษฎีการตอบข้อสอบที่นักวิจัยนิยมนำไปใช้อย่างกว้างขวาง ได้แก่ วิธีไค-กำลังสองของลอร์ด วิธีอัตราส่วนความน่าจะเป็น วิธีการวัดพื้นที่ของราชู และวิธีดีเอฟไอที ทั้งนี้เนื่องจากสามารถนำไปประยุกต์ใช้ภายใต้โมเดลแบบ 1, 2 หรือ 3 พารามิเตอร์ มีการทดสอบนัยสำคัญทางสถิติ ทำให้ผลการตรวจสอบมีความถูกต้องและแม่นยำ ส่วนข้อจำกัดของวิธีดังกล่าวคือ ข้อมูลจะต้องเป็นไปตามข้อตกลงเบื้องต้นที่เข้มงวด การประมาณค่าพารามิเตอร์ภายใต้โมเดลแบบ 2 หรือ 3 พารามิเตอร์ จำเป็นต้องใช้ตัวอย่างขนาดใหญ่ (Clauser; & Mazor. 1998: 33) อีกทั้งมีกระบวนการตรวจสอบหลายขั้นตอน และการแปลผลค่อนข้างยาก นอกจากนี้การวิเคราะห์ต้องใช้โปรแกรมคอมพิวเตอร์ที่ซับซ้อนในการประมาณค่าพารามิเตอร์ ซึ่งมีราคาแพง ทำให้เสียเวลาและค่าใช้จ่ายมาก

จากข้อจำกัดของวิธีการตรวจสอบในกลุ่มทฤษฎีการตอบข้อสอบ จึงทำให้นักวิจัยหันมาพัฒนาวิธีการตรวจสอบโดยใช้ทฤษฎีการทดสอบแบบมาตรฐานเดิม แต่ใช้ฐานคิดและหลักการตรวจสอบของทฤษฎีการตอบข้อสอบ ทั้งนี้เพื่อให้สามารถตรวจสอบได้อย่างง่าย ๆ และใช้ตัวอย่างขนาดเล็ก แต่มีความถูกต้องและแม่นยำสูง ในการคำนวณไม่ยุ่งยากและซับซ้อน สามารถใช้

โปรแกรมคอมพิวเตอร์ที่มีอยู่ทั่วไป วิธีที่พัฒนาขึ้นและเป็นที่ยอมรับ เช่น วิธีการทำให้เป็นมาตรฐานวิธีแมนเทิล-แฮนส์เซล วิธีการถดถอยโลจิสติก เป็นต้น นอกจากนี้ ซีลลีและสเตาท์ (Shealy; & Stout. 1993) ได้พัฒนาวิธีซิปเทสท์ (SIBTEST) ซึ่งมีแนวคิดพื้นฐานมาจากทฤษฎีการตอบข้อสอบแบบหลายมิติ (Multidimensional IRT) โดยมีแนวคิดและหลักการตรวจสอบคล้ายกับวิธีการทำให้เป็นมาตรฐาน ในการตรวจสอบจะเปรียบเทียบความน่าจะเป็นในการตอบข้อสอบถูก ระหว่างกลุ่มผู้สอบที่มีคะแนนการจับคู่เท่ากัน ซึ่งจะแบ่งแบบทดสอบออกเป็นสองชุดย่อย คือ แบบทดสอบชุดย่อยที่สงสัยว่าทำหน้าที่เบี่ยงเบน (Suspect subtest) ประกอบด้วยข้อสอบที่ต้องการศึกษาหนึ่งข้อหรือหลายข้อ ส่วนอีกชุดหนึ่งเป็นแบบทดสอบที่มีความถูกต้อง (Valid subtest) ซึ่งนำไปใช้ในการจับคู่ความสามารถ ประกอบด้วยข้อสอบที่ไม่สงสัยว่าทำหน้าที่เบี่ยงเบนโดยแยกออกจากข้อสอบที่ศึกษาข้างต้น จุดเด่นของวิธีซิปเทสท์ คือ เป็นวิธีนันทพาราเมตริก (Nonparametric) ซึ่งไม่จำเป็นต้องใช้โมเดลประมาณค่า มีกระบวนการตรวจสอบที่ไม่ยุ่งยากซับซ้อน นอกจากนี้ยังใช้ตัวอย่างขนาดเล็ก ในกลุ่มอ้างอิงและกลุ่มสนใจถ้าใช้กลุ่มละ 300 คนก็มีความถูกต้องเพียงพอแล้ว (Sufficient power) (Narayanan; & Swaminathan. 1994: 315) ส่วนข้อจำกัดก็คือ ลักษณะเฉพาะของวิธีการตรวจสอบ สามารถตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกันแต่ไม่สามารถตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ไม่เป็นรูปแบบเดียวกัน (Li; & Stout. 1996: 650)

วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนสองค่า (Dichotomous) ได้มีการพัฒนาต่อเนื่องมาโดยตลอดดังที่กล่าวมาข้างต้น ในปัจจุบันนี้วงการทดสอบและวัดผลการศึกษาได้ให้ความสนใจการประเมินผลแนวใหม่ที่เรียกว่า **“วิธีการวัดประเมินการปฏิบัติ”** (Performance assessment methods) ดังเช่น การปฏิบัติภาระงาน การเขียนแฟ้มสะสมงาน การพิสูจน์ทางคณิตศาสตร์ และการทดลองทางวิทยาศาสตร์ เป็นต้น การทดสอบดังกล่าวยังจำเป็นต้องใช้แบบทดสอบเป็นเครื่องมือวัด แต่รูปแบบการให้คะแนนจากตอบผิดให้ 0 คะแนน และตอบถูกให้ 1 คะแนน ได้พัฒนามาเป็นการให้คะแนนหลายค่า (Polytomous) ดังนั้นนักวิจัยจึงหันมาสนใจพัฒนาวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนหลายค่า ซึ่งส่วนมากปรับเปลี่ยนหรือพัฒนามาจากวิธีการตรวจสอบเดิม ทั้งในกลุ่มที่ใช้ทฤษฎีการทดสอบแบบมาตรฐานเดิม และทฤษฎีการตอบข้อสอบ เช่น วิธีการทำให้เป็นมาตรฐานแบบให้คะแนนหลายค่า (Polytomous STND) (Potenza; & Dorans. 1995) วิธีแมนเทิล (Mantel) วิธีแมนเทิล-แฮนส์เซลทั่วไป (Generalized Mantel-Haenszel; GMH) (Zwick; Donoghue; & Grima. 1993)

วิธีการถดถอยโลจิสติกแบบจัดอันดับ (Ordinal logistic regression; OLR) (Zumbo. 1999) วิธีการวัดพื้นที่ (Cohen; Kim; & Baker. 1993) วิธีโพลี-ซิปเทสท์ (Poly-SIBTEST) (Chang; Mazzeo; & Roussos. 1996) วิธีดีเอฟไอที (DFIT) (Flowers; Oshima; & Raju. 1999) เป็นต้น

การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนสองค่า หรือหลายค่า ไม่ว่าจะใช้ทฤษฎีการทดสอบแบบมาตรฐานเดิม หรือทฤษฎีการตอบข้อสอบ มักอยู่บนเงื่อนไขของโมเดลการวัดความสามารถมิติเดียว การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบโดยใช้แนวคิดมิติเดียวได้ถูกนักวิจัยหลายคนวิพากษ์วิจารณ์ว่ามีจุดด้อยขาดความเหมาะสม (Ackerman. 1992: 67; Ackerman; & Evan. 1994: 329; Camilli. 1992: 130; Mazor; Hambleton; & Clauser. 1998: 358; Oshima; & Miller. 1992: 237; Oshima; Raju; & Flowers. 1997: 253; Stout; et al. 1997: 196) ทั้งนี้เพราะว่าในแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนทั่วไป ผู้สอบที่ตอบข้อสอบถูกต้องใช้ความสามารถในหลายมิติประกอบกัน ดังเช่น ในแบบทดสอบวิชาคณิตศาสตร์ ผู้สอบที่ตอบข้อสอบถูกจำเป็นต้องมีทักษะความสามารถทางคณิตศาสตร์และความเข้าใจในการอ่าน การตอบข้อสอบของผู้สอบแต่ละคนจะใช้ทักษะทั้งสองแตกต่างกัน ดังนั้นการใช้คะแนนรวมหรือค่าประมาณความสามารถบนฐานคิดแบบมิติเดียวมาจำแนกระดับความสามารถของผู้สอบ เพื่อจัดอันดับผู้สอบแต่ละคนจึงไม่ถูกต้อง ถ้านำผลการตอบดังกล่าวไปวิเคราะห์โดยใช้ทฤษฎีการตอบข้อสอบภายใต้มิติเดียวจะทำให้ข้อมูลไม่เหมาะสมกับโมเดล ซึ่งมีผลทำให้ฟังก์ชันการตอบข้อสอบระหว่างกลุ่มแตกต่างกัน แล้วสรุปว่า**ข้อสอบทำหน้าที่เบี่ยงเบน** (Oshima, & Miller. 1992: 237) ทำให้การสรุปเกิดความผิดพลาด เนื่องจากละเมิดข้อตกลงของโมเดลการวัดความสามารถมิติเดียว แบบทดสอบที่วัดคุณลักษณะซับซ้อน (Complex traits) จึงไม่เหมาะที่จะใช้คะแนนรวมเป็นเกณฑ์การจับคู่ เพื่อให้ผลการเปรียบเทียบมีความถูกต้องและแม่นยำควรใช้ความสามารถในทุกมิติเป็นเกณฑ์การจับคู่ นั่นคือ แก้ปัญหาโดยใช้โมเดลการวัดความสามารถหลายมิติ (Multidimensional)

แนวคิดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบโดยใช้โมเดลการวัดความสามารถของผู้สอบหลายมิติเริ่มชัดเจนขึ้น เมื่อซีลลีและสเตาท์ (Shealy; & Stout. 1993) ได้พัฒนาวิธีซิปเทสท์ (SIBTEST) ซึ่งมีพื้นฐานมาจากโมเดลของทฤษฎีการตอบข้อสอบแบบหลายมิติ โดยมีกรอบแนวคิดเชิงทฤษฎีว่า เมื่อการแจกแจงความสามารถของผู้สอบมีหลายมิติ ซึ่งเป็นผลมาจากความสามารถหลัก (Primary ability; θ) และความสามารถรอง (Secondary ability; η) ข้อสอบจะมีความไวต่อความสามารถทั้งสองแตกต่างกัน ถ้าการแจกแจงความสามารถรองแบบมีเงื่อนไขระหว่างกลุ่มแตกต่างกันจะมีผลทำให้ข้อสอบทำหน้าที่เบี่ยงเบน ในกรณีของแบบทดสอบสองมิติ

ความสามารถหลักถือว่าเป็นความสามารถเป้าหมาย (Target ability) ที่ข้อสอบต้องการวัด ส่วนความสามารถรองถือว่าเป็นความสามารถแทรกซ้อน (Nuisance ability) ที่ข้อสอบไม่ต้องการวัด แต่มีอิทธิพลต่อการตอบข้อสอบ

การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ในข้อสอบที่วัดความสามารถหลายมิติ ได้มีนักวิจัยหลายคนให้ความสนใจ ดังเช่น โอชิม่าและมิลเลอร์ (Oshima; & Miller. 1992) ได้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ในข้อสอบที่วัดความสามารถสองมิติ ภายใต้ทฤษฎีการตอบข้อสอบแบบหลายมิติ โดยใช้ดัชนีการวัดพื้นที่ 4 วิธี คือ ดัชนีการวัดพื้นที่แบบคิดเครื่องหมาย (SA) ดัชนีการวัดพื้นที่แบบไม่คิดเครื่องหมาย (UA) ดัชนีผลรวมกำลังสองแบบคิดเครื่องหมาย (SSOS) และดัชนีผลรวมกำลังสองแบบไม่คิดเครื่องหมาย (USOS) ศึกษาภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของความสามารถหลักและความสามารถแทรกซ้อน ผลการศึกษาพบว่า ข้อสอบแบบหลายมิติไม่จำเป็นต้องทำหน้าที่เบี่ยงเบน ยกเว้นค่าเฉลี่ยของความสามารถแทรกซ้อนแตกต่างกัน และเมื่อไม่มีความสัมพันธ์ระหว่างความสามารถหลักและความสามารถแทรกซ้อน ความแตกต่างของค่าเฉลี่ยของความสามารถหลักไม่ได้เป็นสาเหตุทำให้ข้อสอบทำหน้าที่เบี่ยงเบน โอชิม่า ราจู และเฟลาเวอร์ส (Oshima; Raju; & Flowers. 1997) ได้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบและแบบทดสอบโดยใช้วิธีดีเอฟไอที ภายใต้การจำลองข้อมูลแบบสองมิติ ศึกษาในเงื่อนไข 3 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ทิศทางของข้อสอบทำหน้าที่เบี่ยงเบน และความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่ม ผลการศึกษาพบว่า ภายใต้ 4 เงื่อนไข คือ (1) ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และมีทิศทางเดียวกัน (2) ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกันและมีสองทิศทางแบบสมดุล (3) ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกันและมีทิศทางเดียวกัน และ (4) ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกันและมีสองทิศทางแบบสมดุล พารามิเตอร์ของข้อสอบที่เป็นค่าจริงและค่าประมาณให้ผลสอดคล้องกันในทุกเงื่อนไข

สเตาท์และคนอื่นๆ (Stout; et al. 1997) ใช้วิธีมัลติซิป (MULTISIB) ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบแบบสองมิติ ภายใต้การจำลองข้อมูลที่วัดความสามารถในมิติหลักทั้งสองมิติ โดยศึกษาในเงื่อนไขพารามิเตอร์ของข้อสอบ คะแนนเกณฑ์การจับคู่ ความแตกต่างของการแจกแจงความสามารถ ทิศทางของข้อสอบ การปรับแก้ค่าการถดถอย และขนาดตัวอย่าง ผลการศึกษาพบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าอยู่ในระดับที่คาดหวัง และอัตราความถูกต้องมีค่าสูง เมเซอร์ แฮมเบิลตันและคลอเซอร์ (Mazor; Hambleton; & Clauser. 1998)

ได้ใช้วิธีการถดถอยโลจิสติก และวิธีแมนเทิล-แฮนส์เซลตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้การจำลองข้อมูลแบบสองมิติ แล้วศึกษาผลกระทบในแง่ของการใช้คะแนนเกณฑ์ การจับคู่ที่วัดความสามารถมิติเดียว ระหว่างคะแนนจากแบบทดสอบชุดย่อย และคะแนนรวมของ แบบทดสอบ ผลการศึกษาพบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ลดลง เมื่อใช้คะแนนใน แบบทดสอบชุดย่อยเป็นเกณฑ์การจับคู่ความสามารถ นพมาศ พิพัฒนสุข (2541) ได้เปรียบเทียบ ประสิทธิภาพระหว่างวิธีแมนเทิล-แฮนส์เซล กับวิธีการถดถอยโลจิสติก โดยศึกษาเกณฑ์การจับคู่ 3 เกณฑ์ คือ คะแนนรวมของแบบทดสอบ คะแนนแบบทดสอบชุดย่อย และคะแนนแบบทดสอบ หลายชุดย่อย ผลการศึกษาพบว่า วิธีการถดถอยโลจิสติกมีอัตราความถูกต้องเพิ่มขึ้น และอัตรา ความคลาดเคลื่อนประเภทที่ 1 ลดลง เมื่อใช้คะแนนแบบทดสอบหลายชุดย่อย สิริวิรัตน์ วิชาศิลป์ (2545) ได้เปรียบเทียบวิธีชิปเทสท์ และวิธีดีเอฟไอทีในการตรวจสอบการทำหน้าที่เบี่ยงเบนของ ข้อสอบ หมวดข้อสอบ และแบบทดสอบ จากข้อมูลการตอบข้อสอบที่ใช้ความสามารถหลายมิติ ภายใต้ปัจจัยความยาวของแบบทดสอบและขนาดตัวอย่าง ผลการศึกษาพบว่า อัตราความถูกต้อง ของวิธีชิปเทสท์สูงกว่าวิธีดีเอฟไอทีในทุกเงื่อนไข และความสอดคล้องในการตรวจสอบด้วยวิธีทั้ง สองมีค่าต่ำกว่าร้อยละ 10

จากที่กล่าวมาจะเห็นว่า วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบได้มีการคิดค้น และพัฒนามาอย่างต่อเนื่องไม่น้อยกว่าครึ่งศตวรรษ จากวิธีหนึ่งไปอีกวิธีหนึ่ง จากวิธีที่ใช้ทฤษฎี การทดสอบแบบมาตรฐานเดิมไปสู่วิธีที่ใช้ทฤษฎีการตอบข้อสอบ จากการตรวจสอบในข้อสอบที่ ให้คะแนนสองค่าไปสู่ข้อสอบที่ให้คะแนนหลายค่า จากการตรวจสอบในข้อสอบที่วัดความสามารถ มิติเดียวไปสู่ข้อสอบที่วัดความสามารถหลายมิติ ในปัจจุบันนี้ การแสดงหลักฐานเกี่ยวกับการทำ หน้าที่เบี่ยงเบนของข้อสอบเป็นประเด็นสำคัญประเด็นหนึ่งในการทดสอบทางการศึกษาและ จิตวิทยา ซึ่งมีความสำคัญไม่ยิ่งหย่อนไปกว่าการแสดงหลักฐานเกี่ยวกับความเที่ยงตรง ความ- เชื่อมั่น หรือคุณลักษณะของข้อสอบ จากการศึกษาแนวคิด ทฤษฎี หลักการ และงานวิจัยเกี่ยวกับ การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ มีประเด็นที่น่าสนใจ 3 ประเด็น ดังนี้

ประเด็นที่หนึ่ง แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนส่วนมากประกอบด้วยข้อสอบ ที่วัดความสามารถหลายมิติ ผู้สอบที่ตอบข้อสอบถูกจำเป็นต้องใช้ความสามารถหลายมิติ และใน แต่ละมิติก็ใช้ความสามารถแตกต่างกัน ปราภฏการณดังกล่าวสามารถเกิดขึ้นจริงในธรรมชาติ มากกว่าเงื่อนไขของความสามารถมิติเดียว ถ้าแบบทดสอบวัดความสามารถหลายมิติ แต่ใช้ฐาน คิดเพียงมิติเดียวจำแนกระดับความสามารถของผู้สอบ จะทำให้การจัดเรียงลำดับความสามารถ

ของผู้สอบเกิดความคลาดเคลื่อน การเลือกใช้โมเดลผิดพลาดเป็นสาเหตุหนึ่งที่ทำให้ข้อสอบทำหน้าที่เบี่ยงเบน (Ackerman. 1992: 67) ตัวอย่างเช่น แบบทดสอบฉบับหนึ่งวัดความสามารถสองมิติ กำหนดให้ θ_1 และ θ_2 แทนความสามารถในมิติที่หนึ่งและมิติที่สอง ตามลำดับ จำนวนครั้งหนึ่งของข้อสอบทั้งหมดวัดความสามารถ θ_1 ส่วนข้อสอบที่เหลืออีกครึ่งหนึ่งวัดความสามารถ θ_2 ถ้าผู้สอบกลุ่มหนึ่งตอบข้อสอบถูก 50% ของข้อสอบทั้งหมด แสดงว่าตอบถูกเฉพาะข้อสอบที่วัดความสามารถ θ_1 ทั้งหมด หรือตอบถูกเฉพาะข้อสอบที่วัดความสามารถ θ_2 ทั้งหมด หรือตอบถูกจากข้อสอบที่วัดความสามารถทั้งสองรวมกันก็ได้ เมื่อนำผลการตอบข้อสอบดังกล่าวมาวิเคราะห์การทำหน้าที่เบี่ยงเบนของข้อสอบบนฐานคิดมิติเดียว โดยใช้คะแนนรวมของแบบทดสอบเป็นเกณฑ์การจับคู่ (Matching criterion) ทำให้ผลการตรวจสอบไม่คงที่และสรุปผลผิดพลาด เพราะละเมิดข้อตกลงของความเป็นมิติเดียว ดังนั้นน่าจะใช้ความสามารถของผู้สอบในทุกมิติหลักเป็นเกณฑ์การจับคู่ เพื่อให้ผลการตรวจสอบมีความถูกต้องและแม่นยำ

นอกจากนี้การให้คะแนนข้อสอบมิใช่จำกัดเฉพาะรูปแบบตอบผิดได้ 0 หรือตอบถูกได้ 1 คะแนน เท่านั้น การให้คะแนนข้อสอบหลายค่าทำให้ได้สารสนเทศเกี่ยวกับแบบแผนการตอบข้อสอบได้ดีกว่าการให้คะแนนสองค่า ในปัจจุบันนี้การทดสอบทางการศึกษาและจิตวิทยาได้ให้ความสำคัญในประเด็นดังกล่าวมากขึ้น ดังนั้นข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า (Multidimensional polytomous) จึงเป็นประเด็นหนึ่งที่สนใจนำมาศึกษาครั้งนี้

ประเด็นที่สอง วิธีโพลี-ซิปเทสท์ (Poly-SIBTEST) ชาง มาซซีโอ และรูสโซ (Chang, Mazzeo, & Roussos. 1996) ปรับขยายมาจากวิธีซิปเทสท์ของชีลลีและสเตาท์ (Shealy; & Stout. 1993) เพื่อใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนหลายค่า วิธีนี้มีจุดเด่นหลายประการ เช่น ใช้เทคนิคการตรวจสอบแบบหลายมิติ โดยแยกข้อสอบที่ใช้เป็นเกณฑ์การจับคู่ออกจากข้อสอบที่ใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนอย่างชัดเจน ทำให้เกณฑ์มีความเที่ยงตรงและความเชื่อมั่นสูง คะแนนเกณฑ์การจับคู่ค่อนข้างเป็นคุณลักษณะแฝงมากกว่าคะแนนที่ได้จากการสอบ ทำให้มีความถูกต้องและแม่นยำ มีการคำนวณวนซ้ำหลายรอบ (Iterative algorithm) เพื่อคัดเลือกข้อสอบที่ทำหน้าที่เบี่ยงเบนออกจากคะแนนเกณฑ์การจับคู่ ทำให้เกณฑ์มีความบริสุทธิ์ (Purification) มีการปรับแก้ค่าการถดถอย (Regression correction) เพื่อลดความแตกต่างของค่าความสามารถเป้าหมายระหว่างกลุ่มผู้สอบ ทำให้สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ที่สูงเกินปกติได้ ส่วนข้อจำกัดของวิธีนี้ คือ เป็นวิธีที่ใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกัน (Uniform DIF) ซึ่งไม่สามารถตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF)

ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับ (Ordinal logistic regression) ซัมโบ (Zumbo. 1999) ได้ปรับขยายมาจากวิธีการถดถอยโลจิสติกของสวามินาธานและโรเจอร์ส (Swaminathan; & Rogers. 1990) เพื่อใช้ในการตรวจสอบการทำหน้าที่เปี่ยงเบนของข้อสอบที่ให้คะแนนหลายค่า โดยมีกรอบแนวคิดเชื่อมประสานระหว่างทฤษฎีการทดสอบแบบมาตรฐานเดิม และทฤษฎีการตอบข้อสอบ ข้อได้เปรียบของวิธีนี้ คือ ใช้โมเดลการถดถอยโลจิสติกทำนายความน่าจะเป็นในการตอบข้อสอบถูก ภายใต้คะแนนความสามารถแบบต่อเนื่องและการเป็นสมาชิกของกลุ่มผู้สอบ โมเดลดังกล่าวสามารถทดสอบอิทธิพลของปฏิสัมพันธ์ระหว่างระดับความสามารถ กับกลุ่มผู้สอบ ทำให้มีประสิทธิภาพในการตรวจสอบข้อสอบทำหน้าที่เปี่ยงเบน ทั้งที่เป็นรูปแบบเดียวกัน (Uniform DIF) และที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) ส่วนข้อจำกัดของวิธีนี้ คือ เป็นวิธีการตรวจสอบภายใต้กรอบแนวคิดที่วัดความสามารถมิติเดียว เมื่อนำมาตรวจสอบในข้อสอบที่วัดความสามารถหลายมิติจะใช้คะแนนรวมแทนความสามารถ ซึ่งคำนวณมาจากค่าเฉลี่ยของคะแนนในทุกมิติ คะแนนรวมดังกล่าวจึงเป็นตัวแทนมิติเดียว ทำให้ไม่สามารถจำแนกระดับความสามารถของผู้สอบในทุกมิติได้อย่างแท้จริง การใช้คะแนนรวมเป็นเกณฑ์การจับคู่จะทำให้ผลการตรวจสอบมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงเกินปกติ

ดังนั้นผู้วิจัยจึงปรับขยายวิธีการถดถอยโลจิสติกแบบจัดอันดับของซัมโบ โดยนำความสามารถในแต่ละมิติมาใช้เป็นเกณฑ์การจับคู่ ซึ่งคาดว่าจะแก้ปัญหาดังกล่าวได้ จากแนวคิดที่กล่าวมาข้างต้นมีประเด็นข้อสงสัยว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ (Multidimensional ordinal logistic regression; MOLR) ที่ผู้วิจัยนำเสนอครั้งนี้ มีประสิทธิภาพในการตรวจสอบการทำหน้าที่เปี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า หรือไม่ เพียงใด โดยศึกษาเปรียบเทียบกับวิธีโพลี-ชิปเทสต์และวิธีการถดถอยโลจิสติกแบบจัดอันดับ

ประเด็นที่สาม การเลือกใช้วิธีการตรวจสอบการทำหน้าที่เปี่ยงเบนของข้อสอบ เพื่อให้ผลการตรวจสอบมีประสิทธิภาพและประสิทธิผล ควรคำนึงถึงปัจจัยร่วมที่นำมาใช้ในการตรวจสอบ เช่น รูปแบบของข้อสอบทำหน้าที่เปี่ยงเบน (Form of DIF) สัดส่วนของข้อสอบทำหน้าที่เปี่ยงเบน (Proportion of DIF) ความแตกต่างของการแจกแจงความสามารถ (Ability distribution differences) และขนาดตัวอย่าง (Sample size) เป็นต้น ทั้งนี้จะได้เป็นสารสนเทศส่วนหนึ่งในการตัดสินใจเลือกใช้วิธีการตรวจสอบให้เหมาะสมมากที่สุด ภายใต้เงื่อนไขของปัจจัยที่นำมาตรวจสอบ ดังนั้นปัจจัยที่แปรเปลี่ยน 4 ปัจจัยดังกล่าวเป็นประเด็นที่สงสัยว่า มีผลต่อประสิทธิภาพในการตรวจสอบ

ของวิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ หรือไม่ ทั้งสามวิธีสามารถตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบได้ดีที่สุดในเงื่อนไขใด ระดับใด

จากประเด็นดังกล่าว ผู้วิจัยจึงมีความสนใจที่จะนำวิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ มาตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ซึ่งประกอบด้วย รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่าง ทั้งนี้เพื่อตอบคำถามในการวิจัยสองข้อ คือ (1) ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัย 4 ปัจจัย มีผลต่อประสิทธิภาพในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า ของวิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ หรือไม่ และ (2) วิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัย มีประสิทธิภาพในการตรวจสอบแตกต่างกันหรือไม่ ผลจากการศึกษาค้นคว้าครั้งนี้จะเป็นประโยชน์ต่อการทดสอบทางการศึกษาและจิตวิทยาต่อไป

ความมุ่งหมายของการวิจัย

จากประเด็นคำถามการวิจัยข้างต้น การวิจัยครั้งนี้จึงมีความมุ่งหมายสำคัญเพื่อศึกษาวิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า วิธีใดมีประสิทธิภาพในการตรวจสอบมากที่สุด ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ซึ่งประกอบด้วยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่าง โดยพิจารณาจากอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 สำหรับความมุ่งหมายของการวิจัยแยกเป็นความมุ่งหมายเฉพาะดังนี้

1. เพื่อเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า ด้วยวิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ภายใต้เงื่อนไขต่างกันของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ซึ่งประกอบด้วย รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่าง

2. เพื่อเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า ระหว่างวิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ภายใต้เงื่อนไขเดียวกันของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ซึ่งประกอบด้วย รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่าง

ความสำคัญของการวิจัย

ในการวิจัยครั้งนี้มีเป้าหมายในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า เมื่อใช้โมเดลแบบหลายมิติจะให้ผลการตรวจสอบแตกต่างไปจากการใช้โมเดลแบบมิติเดียวหรือไม่ อย่างไร โดยพิจารณาจากอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในครั้งนี้ผู้วิจัยได้ปรับขยายวิธีการถดถอยโลจิสติกแบบจัดอันดับเพื่อใช้กับข้อสอบแบบหลายมิติ โดยนำค่าประมาณความสามารถในทุกมิติเป็นเกณฑ์การจับคู่ ซึ่งกรอบแนวคิดเดิมที่ใช้กับข้อสอบแบบมิติเดียวจะใช้คะแนนรวมเป็นเกณฑ์การจับคู่ ส่วนวิธีโพลี-ชิปเทสต์ใช้คะแนนรวมของแบบทดสอบ ซึ่งค่อนข้างเป็นคุณลักษณะแฝงเป็นเกณฑ์การจับคู่ แล้วใช้เทคนิคการตรวจสอบแบบหลายมิติ

ผลการศึกษาครั้งนี้ สามารถนำสารสนเทศเกี่ยวกับวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ข้อสอบวัดความสามารถหลายมิติและให้คะแนนหลายค่า ไปใช้ประโยชน์ต่อการทดสอบทางการศึกษาและจิตวิทยา นักพัฒนาแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนสามารถเลือกวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่มีประสิทธิภาพ เหมาะสมกับข้อมูลที่ใช้วัดความสามารถ ซึ่งจะส่งผลให้การพัฒนาแบบทดสอบมีคุณภาพตามเกณฑ์มาตรฐาน และสามารถแปลความหมายของคะแนนจากการใช้แบบทดสอบมีความถูกต้องและชัดเจนมากยิ่งขึ้น

อีกทั้งยังทำให้ทราบถึง ปัจจัยที่มีผลต่อวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ เกี่ยวกับรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่าง ทำให้สามารถประเมินข้อดีและข้อจำกัดของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบภายใต้ปัจจัยดังกล่าว ซึ่งจะเป็นประโยชน์ในการเลือกใช้วิธีการตรวจสอบให้เหมาะสมกับลักษณะของข้อสอบและผู้สอบ เพื่อให้ได้สารสนเทศสูงสุดในการตรวจสอบ

นอกจากนี้ ยังเป็นแนวทางในการศึกษาการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า ภายใต้การจำลองข้อมูลโดยใช้ทฤษฎีการตอบข้อสอบหลายมิติ ซึ่งจะช่วยให้มีการคิดค้นและพัฒนาวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่มีประสิทธิภาพและประสิทธิผลต่อไป

ขอบเขตของการวิจัย

ขอบเขตของการวิจัยมีดังนี้

1. ข้อมูลที่ศึกษาเป็นข้อมูลจำลอง โดยใช้โมเดลพาเชียลเครดิตทั่วไปแบบหลายมิติ (Multidimensional generalized partial credit model) ภายใต้ทฤษฎีการตอบข้อสอบแบบหลายมิติ จำลองแบบทดสอบที่มีโครงสร้างวัดความสามารถสองมิติ (Two-dimensional) จำนวน 40 ข้อ โดยข้อสอบแต่ละข้อวัดความสามารถเป้าหมายทั้งสองมิติ กำหนดให้ θ_1 และ θ_2 แทนความสามารถในมิติที่หนึ่งและมิติที่สอง ตามลำดับ ข้อสอบทุกข้อมีรายการตอบ 5 รายการ โดยให้คะแนนเป็น 1, 2, 3, 4 หรือ 5 คะแนน ในการจำลองข้อมูลดังกล่าวจะจำลองผลการตอบข้อสอบ ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน 2 รูปแบบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 3 ขนาด ความแตกต่างของการแจกแจงความสามารถ 3 ระดับ และขนาดตัวอย่าง 4 ขนาด รวมข้อมูลทั้งหมดที่ต้องจัดกระทำเพื่อตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบจำนวน 72 เงื่อนไข ($2 \times 3 \times 3 \times 4$) ในแต่ละเงื่อนไขจำลองข้อมูลซ้ำ 50 ครั้ง

2. ตัวแปรที่ศึกษา ประกอบด้วย

- 2.1 ตัวแปรอิสระ มี 5 ตัวแปร ดังนี้

- 2.1.1 วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ มี 3 วิธี ดังนี้

- 1) วิธีโพลี-ชิปเทสท์
- 2) วิธีการถดถอยโลจิสติกแบบจัดอันดับ
- 3) วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ

2.1.2 รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน มี 2 รูปแบบ ดังนี้

- 1) ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน
- 2) ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน

2.1.3 สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน มี 3 ขนาด ดังนี้

- 1) จำนวน 10%
- 2) จำนวน 20%
- 3) จำนวน 30%

2.1.4 ความแตกต่างของการแจกแจงความสามารถ มี 3 ระดับ

- 1) การแจกแจงความสามารถระหว่างกลุ่มเท่ากัน (แตกต่างกัน 0.0SD)

$$(\bar{\theta}_{1R} - \bar{\theta}_{1F} = 0.0 \text{ และ } \bar{\theta}_{2R} - \bar{\theta}_{2F} = 0.0)$$

- 2) การแจกแจงความสามารถระหว่างกลุ่มแตกต่างกัน 0.5SD

$$(\bar{\theta}_{1R} - \bar{\theta}_{1F} = 0.5 \text{ และ } \bar{\theta}_{2R} - \bar{\theta}_{2F} = 0.5)$$

- 3) การแจกแจงความสามารถระหว่างกลุ่มแตกต่างกัน 1.0SD

$$(\bar{\theta}_{1R} - \bar{\theta}_{1F} = 1.0 \text{ และ } \bar{\theta}_{2R} - \bar{\theta}_{2F} = 1.0)$$

2.1.5 ขนาดตัวอย่าง ประกอบด้วยกลุ่มอ้างอิงและกลุ่มสนใจที่มีจำนวนผู้สอบในแต่ละกลุ่มเท่ากัน มี 4 ขนาด ดังนี้

- 1) จำนวน 250 คน
- 2) จำนวน 500 คน
- 3) จำนวน 1,000 คน
- 4) จำนวน 2,000 คน

2.2 ตัวแปรตาม มี 2 ตัวแปร ดังนี้

2.2.1 อัตราความถูกต้อง

2.2.2 อัตราความคลาดเคลื่อนประเภทที่ 1

3. การจำลองข้อสอบเกี่ยวกับปัจจัยความยาวของแบบทดสอบ (Test length) กำหนดให้คงที่ในทุกเงื่อนไขของการตรวจสอบ โดยจำลองความยาวของแบบทดสอบจำนวน 40 ข้อ ส่วนขนาดของข้อสอบทำหน้าที่เบี่ยงเบนไม่ได้เป็นปัจจัยที่ศึกษา จะกำหนดคงที่ตามเงื่อนไขรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กล่าวคือ ภายใต้เงื่อนไขของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน กำหนดพารามิเตอร์รายการของข้อสอบ (Item-category) ระหว่างกลุ่มอ้างอิงและกลุ่ม

สนใจมีค่าแตกต่างกัน 0.20 ส่วนพารามิเตอร์อำนาจจำแนกของข้อสอบ (Item discrimination) ในแต่ละมิติมีค่าคงที่ สำหรับภายใต้เงื่อนไขของข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน กำหนดพารามิเตอร์อำนาจจำแนกของข้อสอบในแต่ละมิติระหว่างกลุ่มอ้างอิงและกลุ่มสนใจมีค่าแตกต่างกัน 0.15 ในขณะที่พารามิเตอร์รายการของข้อสอบระหว่างกลุ่มมีค่าคงที่ ส่วนการจำลอง ภายใต้เงื่อนไขความแตกต่างของการแจกแจงความสามารถมี 3 กรณี โดยศึกษาเฉพาะความแตกต่างของค่าเฉลี่ยของความสามารถ θ_1 และ θ_2 ระหว่างกลุ่มอ้างอิงและกลุ่มสนใจเท่านั้น ซึ่งไม่ได้ศึกษาความแตกต่างของค่าความแปรปรวน และค่าสหสัมพันธ์ของค่าความสามารถ ดังนั้นในทุกเงื่อนไขจะกำหนดความแปรปรวนของค่าความสามารถในแต่ละกลุ่มมีค่าคงที่ เท่ากับ 1.0 ($S_{\theta_1}^2 = S_{\theta_2}^2 = 1.0$) และค่าสหสัมพันธ์ระหว่างค่าความสามารถมิติที่หนึ่ง กับมิติที่สองของแต่ละกลุ่มมีค่าคงที่ เท่ากับ 0.5 ($r_{\theta_1\theta_2} = 0.5$)

นิยามศัพท์เฉพาะ

1. **แบบทดสอบหลายมิติ (Multidimensional test)** หมายถึง แบบทดสอบที่มีโครงสร้างในการวัดความสามารถของผู้สอบที่มีลักษณะเด่นจำนวนสองมิติหรือมากกว่า ในการศึกษครั้งนี้ จำลองแบบทดสอบสองมิติ (Two-dimensional) ซึ่งเป็นการวัดความสามารถหลายมิติภายในข้อสอบ (Multidimensional within-item test) โดยข้อสอบแต่ละข้อวัดความสามารถเป้าหมายสองมิติที่มีน้ำหนักแตกต่างกัน กำหนดให้ θ_1 และ θ_2 แทนความสามารถหลักในมิติที่หนึ่งและมิติที่สองตามลำดับ ข้อสอบทุกข้อมีรายการตอบ 5 รายการ โดยให้คะแนนแบบหลายค่า (Polytomous) ผลการตอบในรายการ 1, 2, 3, 4 และ 5 ให้คะแนนเป็น 1, 2, 3, 4 และ 5 ตามลำดับ
2. **การทำหน้าที่เบี่ยงเบนของข้อสอบ (Differential item functioning; DIF)** หมายถึง ผู้สอบจากกลุ่มที่แตกต่างกัน และมีการจับคู่ความสามารถตามที่ข้อสอบต้องการวัดเท่ากัน มีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องไม่เท่ากัน
3. **กลุ่มอ้างอิง (Reference group; R)** หมายถึง กลุ่มผู้สอบที่คาดว่าจะได้เปรียบในการตอบข้อสอบเมื่อข้อสอบทำหน้าที่เบี่ยงเบน โดยมีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องมากกว่ากลุ่มสนใจ
4. **กลุ่มสนใจ (Focal group; F)** หมายถึง กลุ่มผู้สอบที่เป็นเป้าหมายของการศึกษา ซึ่งคาดว่าจะเสียเปรียบในการตอบข้อสอบเมื่อข้อสอบทำหน้าที่เบี่ยงเบน โดยมีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องน้อยกว่ากลุ่มอ้างอิง

5. **รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน (Form of DIF)** หมายถึง ขนาดและทิศทางของข้อสอบทำหน้าที่เบี่ยงเบน ซึ่งแปรเปลี่ยนตามความแตกต่างของระดับความสามารถของผู้สอบ ในทฤษฎีการตอบข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า สามารถพิจารณารูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนได้จากความแตกต่างของพารามิเตอร์อำนาจจำแนกของข้อสอบ (Item discrimination) และพารามิเตอร์รายการของข้อสอบ (Item-category) ระหว่างกลุ่มอ้างอิงและกลุ่มสนใจ ในการวิจัยครั้งนี้ศึกษารูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน 2 รูปแบบ ดังนี้

5.1 **ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF)** หมายถึง ความแตกต่างของความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจมีค่าคงที่ ทุกระดับของพารามิเตอร์ความสามารถ θ_1 และ θ_2 ในการวิจัยครั้งนี้กำหนดให้พารามิเตอร์รายการของข้อสอบระหว่างกลุ่มมีค่าแตกต่างกัน 0.20 ส่วนพารามิเตอร์อำนาจจำแนกของข้อสอบในแต่ละมิติระหว่างกลุ่มมีค่าเท่ากัน

5.2 **ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF)** หมายถึง ความแตกต่างของความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจมีค่าไม่เท่ากัน ทุกระดับของพารามิเตอร์ความสามารถ θ_1 และ θ_2 ในการวิจัยครั้งนี้กำหนดให้พารามิเตอร์อำนาจจำแนกของข้อสอบในแต่ละมิติระหว่างกลุ่มมีค่าแตกต่างกัน 0.15 ส่วนพารามิเตอร์รายการของข้อสอบระหว่างกลุ่มมีค่าเท่ากัน

6. **สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน (Proportion of DIF)** หมายถึง จำนวนข้อสอบที่ทำหน้าที่เบี่ยงเบน ต่อจำนวนข้อสอบทั้งหมดในแบบทดสอบ ในการวิจัยครั้งนี้ศึกษาสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 3 ขนาด คือ 10%, 20% และ 30%

7. **ความแตกต่างของการแจกแจงความสามารถ (Ability distribution differences)** หมายถึง ความสามารถในการมิติที่หนึ่งและมิติที่สอง (θ_1 และ θ_2) ของกลุ่มอ้างอิงและกลุ่มสนใจมีการแจกแจงแบบปกติสองตัวแปร (Bivariate normal distribution) แตกต่างกัน ในการวิจัยครั้งนี้ศึกษาความแตกต่างของการแจกแจงความสามารถ 3 ระดับ ดังนี้

7.1 **การแจกแจงความสามารถระหว่างกลุ่มเท่ากัน ($\bar{\theta}_{1R} - \bar{\theta}_{1F} = 0.0$ และ $\bar{\theta}_{2R} - \bar{\theta}_{2F} = 0.0$)** โดยกำหนดให้กลุ่มอ้างอิงและกลุ่มสนใจมีค่าเฉลี่ยของความสามารถเท่ากัน คือ 0.0 ความแปรปรวนของความสามารถในแต่ละกลุ่มมีค่าคงที่เท่ากับ 1.0 และค่าสหสัมพันธ์ระหว่างความสามารถในมิติที่หนึ่ง กับมิติที่สองของแต่ละกลุ่มมีค่าคงที่เท่ากับ 0.5

7.2 การแจกแจงความสามารถระหว่างกลุ่มแตกต่างกัน $0.5SD$ ($\bar{\theta}_{1R} - \bar{\theta}_{1F} = 0.5$ และ $\bar{\theta}_{2R} - \bar{\theta}_{2F} = 0.5$) โดยกำหนดให้กลุ่มอ้างอิงและกลุ่มสนใจมีค่าเฉลี่ยเท่ากับ 0.0 และ -0.5 ตามลำดับ ความแปรปรวนของความสามารถในแต่ละกลุ่มมีค่าคงที่เท่ากับ 1.0 และค่าสหสัมพันธ์ระหว่างความสามารถในมิติที่หนึ่ง กับมิติที่สองของแต่ละกลุ่มมีค่าคงที่เท่ากับ 0.5

7.3 การแจกแจงความสามารถระหว่างกลุ่มแตกต่างกัน $1.0SD$ ($\bar{\theta}_{1R} - \bar{\theta}_{1F} = 1.0$ และ $\bar{\theta}_{2R} - \bar{\theta}_{2F} = 1.0$) โดยกำหนดให้กลุ่มอ้างอิงและกลุ่มสนใจมีค่าเฉลี่ยเท่ากับ 0.0 และ -1.0 ตามลำดับ ความแปรปรวนของความสามารถในแต่ละกลุ่มมีค่าคงที่เท่ากับ 1.0 และค่าสหสัมพันธ์ระหว่างความสามารถในมิติที่หนึ่ง กับมิติที่สองของแต่ละกลุ่มมีค่าคงที่เท่ากับ 0.5

8. ขนาดตัวอย่าง (Sample size) หมายถึง จำนวนผู้สอบของกลุ่มอ้างอิงและกลุ่มสนใจในการวิจัยครั้งนี้ใช้จำนวนผู้สอบของกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน 4 ขนาด คือ จำนวน 250 คน, 500 คน, 1,000 คน และ 2,000 คน

9. วิธีโพลี-ซิปเทสต์ (Poly-SIBTEST) หมายถึง วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ในกลุ่มวิธีที่ใช้ทฤษฎีการตอบข้อสอบ รูปแบบนันทพาราเมตริก (Nonparametric form) ซึ่งพัฒนาโดยซาง มาซซซีโอ และรูสโซ (Chang; Mazzeo; & Roussos. 1996) เพื่อใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนหลายค่า ในการศึกษาครั้งนี้นำมาตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า โดยใช้คะแนนรวมของแบบทดสอบ ซึ่งค่อนข้างเป็นคุณลักษณะแฝงเป็นเกณฑ์การจับคู่ความสามารถ

10. วิธีการถดถอยโลจิสติกแบบจัดอันดับ (Ordinal logistic regression; OLR) หมายถึง วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ในกลุ่มวิธีที่ใช้ทฤษฎีการทดสอบมาตรฐานเดิม รูปแบบพาราเมตริก (Parametric form) ซึ่งพัฒนาโดยซัมโบ (Zumbo. 1999) เพื่อใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนหลายค่า ในการศึกษาครั้งนี้นำมาตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า โดยใช้คะแนนรวมของแบบทดสอบเป็นเกณฑ์การจับคู่ความสามารถ

11. วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ (Multidimensional ordinal logistic regression; MOLR) หมายถึง วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ซึ่งผู้วิจัยได้ปรับขยายมาจากวิธีการถดถอยโลจิสติกแบบจัดอันดับของซัมโบ (Zumbo. 1999) เพื่อใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า โดยใช้ค่าประมาณความสามารถสองมิติ (θ_1 และ θ_2) เป็นเกณฑ์การจับคู่ความสามารถ

12. อัตราความถูกต้อง (Power rate) หมายถึง จำนวนข้อสอบที่ตรวจสอบได้ถูกต้องว่าทำหน้าที่เบี่ยงเบน ต่อจำนวนข้อสอบที่ทำหน้าที่เบี่ยงเบนทั้งหมดในแบบทดสอบ โดยคำนวณในรูปร้อยละ

13. อัตราความคลาดเคลื่อนประเภทที่ 1 (Type I error rate) หมายถึง จำนวนข้อสอบที่ตรวจสอบผิดพลาดว่าทำหน้าที่เบี่ยงเบน ทั้งที่จริงข้อสอบทำหน้าที่ไม่เบี่ยงเบน ต่อจำนวนข้อสอบที่ทำหน้าที่ไม่เบี่ยงเบนทั้งหมดในแบบทดสอบ โดยคำนวณในรูปร้อยละ

กรอบแนวคิดในการวิจัย

การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ในข้อสอบที่ให้คะแนนแบบสองค่าหรือหลายค่ามีข้อจำกัดอยู่บนเงื่อนไขของข้อตกลงความเป็นมิติเดียวของแบบทดสอบ แนวคิดดังกล่าวได้ถูกนักวิจัยหลายคนวิพากษ์วิจารณ์ว่ามีจุดด้อยขาดความเหมาะสม (Ackerman. 1992: 67; Ackerman; & Evan. 1994: 329; Camilli. 1992: 130; Mazor; Hambleton; & Clauser. 1998: 358; Oshima; & Miller. 1992: 237; Oshima; Raju; & Flowers. 1997: 253; Stout; et al. 1997: 196) เนื่องจากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนทั่วไปมักวัดความสามารถที่มีลักษณะเด่นหลายความสามารถประกอบกัน ซึ่งความสามารถในแต่ละมิติมีอิทธิพลในการตอบข้อสอบ ดังนั้นการใช้คะแนนรวมของแบบทดสอบตามทฤษฎีการทดสอบแบบมาตรฐานเดิม หรือค่าประมาณความสามารถตามทฤษฎีการตอบข้อสอบ ภายใต้เงื่อนไขของข้อตกลงความเป็นมิติเดียวของแบบทดสอบ จึงไม่น่าเพียงพอที่จะนำมาใช้เป็นเกณฑ์การจับคู่ความสามารถของผู้สอบ ถ้าผู้สอบมาจากกลุ่มที่แตกต่างกันภายใต้การแจกแจงความสามารถหลายมิติ และข้อสอบสามารถจำแนกระดับความสามารถของผู้สอบในหลายมิติดังกล่าว การใช้แบบแผนการให้คะแนนความสามารถมิติเดียวจะมีผลทำให้ข้อสอบทำหน้าที่เบี่ยงเบน (Ackerman. 1992: 67) นั่นคือ เกิดความผิดพลาดของการเลือกใช้โมเดลสำหรับการวิเคราะห์ ซึ่งจะส่งผลให้การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงเกินปกติ ดังนั้นในแบบทดสอบที่วัดคุณลักษณะซับซ้อน (Complex traits) น่าจะใช้ความสามารถในทุกมิติเป็นเกณฑ์การจับคู่ เพื่อทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ลดลง

ในการวิจัยครั้งนี้ ผู้วิจัยได้นำวิธีการถดถอยโลจิสติกแบบจัดอันดับ (Ordinal logistic regression; OLR) ของซัมโบ (Zumbo. 1999) ที่ใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถมิติเดียวและให้คะแนนหลายค่า มาปรับขยายโมเดลเพื่อให้สามารถ

ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า โดยใช้ค่าประมาณความสามารถทั้งสองมิติเป็นเกณฑ์การจับคู่ สำหรับโมเดลการถดถอยโลจิสติกแบบจัดอันดับสามารถบรรยายในรูปโลจิสต์ (logit) แบบสองมิติดังนี้

$$\text{logit}[P(Y \leq j)] = \beta_{0j} + \beta_1\theta_1 + \beta_2\theta_2 + \beta_3G + \beta_4(\theta_1G) + \beta_5(\theta_2G) + \beta_6(\theta_1\theta_2G)$$

เมื่อ $P(Y \leq j)$ แทนความน่าจะเป็นสะสมของผลการตอบข้อสอบในรายการคำตอบที่ j หรือน้อยกว่า j สำหรับ θ_1 และ θ_2 แทนตัวแปรความสามารถในมิติที่หนึ่งและมิติที่สอง ตามลำดับ G แทนตัวแปรกลุ่มผู้สอบ θ_1G และ θ_2G แทนปฏิสัมพันธ์สองทางระหว่างความสามารถมิติที่หนึ่ง ความสามารถมิติที่สอง กับกลุ่มผู้สอบ $\theta_1\theta_2G$ แทนปฏิสัมพันธ์สามทางระหว่างความสามารถในมิติที่หนึ่ง ความสามารถในมิติที่สอง กับกลุ่มผู้สอบ β_{0j} แทนพารามิเตอร์ส่วนตัว $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ และ β_6 แทนพารามิเตอร์ที่เกี่ยวข้องกับตัวแปรแต่ละตัวในโมเดลดังกล่าว

ในการศึกษาครั้งนี้ ผู้วิจัยได้นำวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติที่ปรับขยายโมเดลดังกล่าว ไปเปรียบเทียบกับวิธีการถดถอยโลจิสติกแบบจัดอันดับ (Zumbo. 1999) และวิธีโพลี-ชิปเทสต์ (Chang; Mazzeo; & Roussos. 1996) ซึ่งใช้คะแนนรวมของแบบทดสอบเป็นเกณฑ์การจับคู่ความสามารถ โดยตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่าง กรอบแนวคิดเชิงทฤษฎีของการนำปัจจัยทั้ง 4 ดังกล่าวมาศึกษา มีรายละเอียดดังนี้

รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน

ปัจจัยแรกที่สนใจ คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน โดยทั่วไปนักวิจัยมักศึกษาในสองรูปแบบ คือ ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF) และข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) จากการศึกษาของนักวิจัยกลุ่มหนึ่ง (Li; & Stout. 1996; Narayanan; & Swaminathan. 1994, 1996; Rogers; & Swaminathan. 1993; Swaminathan; & Rogers. 1990; Whitmore; & Schumacker. 1999) ได้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนสองค่า ผลการศึกษาสรุปได้ว่า วิธีชิปเทสต์และวิธีแมนเทิล-แฮนส์เชลมีประสิทธิภาพในการตรวจสอบข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน ส่วนวิธีการถดถอย-

โลจิสติกมีประสิทธิภาพในการตรวจสอบข้อสอบทำหน้าที่เบี่ยงเบน ทั้งที่เป็นรูปแบบเดียวกันและที่ไม่เป็นรูปแบบเดียวกัน จากการศึกษาของนักวิจัยอีกกลุ่มหนึ่ง (Chang; Mazzeo; & Roussos. 1996; French; & Miller. 1996; Krisjansson; et al. 2005; Zwick; Donoghue; & Grima. 1993) ที่ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนหลายค่า โดยใช้วิธีแมนเทิล วิธีแมนเทิล-แฮนส์เซลทั่วไป วิธีความแตกต่างเฉลี่ยมาตรฐาน วิธีโพลี-ชิปเทสต์ และวิธีที่ใช้โมเดลการถดถอยโลจิสติก (Continuation ratio logits model, Cumulative logits model, Adjacent categories logits model, Unconstrained cumulative logits ordinal model) ผลการศึกษารูปได้ว่า วิธีแมนเทิล วิธีแมนเทิล-แฮนส์เซลทั่วไป วิธีความแตกต่างเฉลี่ยมาตรฐาน และวิธีโพลี-ชิปเทสต์ให้ผลสอดคล้องกับวิธีชิปเทสต์และวิธีแมนเทิล-แฮนส์เซล ส่วนวิธีที่ใช้โมเดลการถดถอยโลจิสติกให้ผลสอดคล้องกับการตรวจสอบในข้อสอบที่ให้คะแนนสองค่าเช่นเดียวกัน สาเหตุดังกล่าวอาจเนื่องมาจากการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนสองค่า วิธีชิปเทสต์และวิธีแมนเทิล-แฮนส์เซลใช้สถิติแบบคิดเครื่องหมาย (เครื่องหมายบวกและเครื่องหมายลบ) เมื่อข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ซึ่งสามารถพิจารณาได้จากฟังก์ชันการตอบข้อสอบ (Item response functions; IRFs) ระหว่างกลุ่มตัดกันตรงช่วงกึ่งกลางของความสามารถ รูปแบบของข้อสอบดังกล่าวทำให้เกิดทิศทางของข้อสอบทำหน้าที่เบี่ยงเบนเปลี่ยนแปลง ณ ตรงช่วงกึ่งกลางของความสามารถ ซึ่งมีผลทำให้ความแตกต่างของคะแนนส่วนหนึ่งที่เป็นค่าลบหักล้างกับคะแนนอีกส่วนหนึ่งที่เป็นค่าบวก (Cancellation) ดังนั้นวิธีชิปเทสต์และวิธีแมนเทิล-แฮนส์เซลจึงไม่สามารถตรวจพบรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนดังกล่าว ส่วนวิธีการถดถอยโลจิสติกสามารถใช้โมเดลทดสอบอิทธิพลของปฏิสัมพันธ์ระหว่างระดับความสามารถ กับการเป็นสมาชิกของกลุ่มผู้สอบ ทำให้สามารถตรวจพบข้อสอบทำหน้าที่เบี่ยงเบน ทั้งที่เป็นรูปแบบเดียวกันและที่ไม่เป็นรูปแบบเดียวกัน สำหรับการการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนหลายค่าก็สามารถอธิบายได้ในทำนองเดียวกัน ดังนั้นรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนเป็นปัจจัยที่สนใจนำมาศึกษาครั้งนี้

สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน

ปัจจัยที่สองที่สนใจ คือ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน เนื่องจากการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ จำเป็นต้องเปรียบเทียบความสามารถของผู้สอบ ภายใต้คะแนนเกณฑ์การจับคู่ที่มีความเชื่อมั่น (Clauser; & Mazor. 1988: 37; Potenza; & Dorans. 1995: 32) ถ้าในแบบทดสอบประกอบด้วยข้อสอบทำหน้าที่เบี่ยงเบนจะมีผลทำให้ค่าประมาณความสามารถ

มีความเชื่อมั่นต่ำ เกณฑ์การจับคู่จึงขาดความแม่นยำ ซึ่งจะมีผลทำให้อัตราความถูกต้องลดลง หรืออัตราความคลาดเคลื่อนประเภทที่ 1 สูงเกินไป จากผลการศึกษาของ เกียร์ล กอทซ์แมนน์ และบอทตัน (Gierl; Gotzmann; & Boughton. 2004) นารายานานและสวามินาธาน (Narayanan; & Swaminathan. 1994, 1996) โรเจอร์สและสวามินาธาน (Rogers; & Swaminathan. 1993) พบว่า สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน (ในช่วง 0% ถึง 20%) ไม่มีผลต่อวิธีชิปเทสต์และวิธีแมนเทิล-แฮนส์เชล แต่จะมีผลต่อวิธีการถดถอยโลจิสติก สาเหตุดังกล่าวเนื่องจากวิธีชิปเทสต์และวิธีแมนเทิล-แฮนส์เชลใช้กระบวนการวิเคราะห์แบบสองขั้นตอน ในขั้นตอนแรกจะคำนวณโดยใช้คะแนนรวมจากข้อสอบทุกข้อ และในขั้นตอนที่สองจะคำนวณซ้ำใหม่ แต่จะคัดเลือกข้อสอบที่ตรวจพบว่าทำหน้าที่เบี่ยงเบนในขั้นตอนแรกออกจากการคำนวณในขั้นตอนที่สอง ดังนั้นจึงทำให้คะแนนเกณฑ์การจับคู่มีความบริสุทธิ์ (Purification) ยิ่งในปัจจุบันวิธีชิปเทสต์ที่ใช้โปรแกรม SIBTEST ที่มีประสิทธิภาพสามารถวิเคราะห์แบบทวนซ้ำ (Iterative algorithm) โดยจะวิเคราะห์ซ้ำหลายครั้งจนกระทั่งได้เกณฑ์การจับคู่ที่ปลอดจากข้อสอบทำหน้าที่เบี่ยงเบน (DIF-free) (Clauser; & Mazor. 1998: 35) ส่วนวิธีการถดถอยโลจิสติกไม่มีกระบวนการตรวจสอบดังกล่าว ในการวิจัยครั้งนี้ผู้วิจัยสนใจศึกษาสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 3 ขนาด คือ 10%, 20% และ 30%

ความแตกต่างของการแจกแจงความสามารถ

ปัจจัยที่สามนี้เป็นประเด็นสำคัญที่ทำให้มีการพัฒนาเทคนิคการตรวจสอบตั้งแต่ยุคเริ่มต้นจนถึงปัจจุบัน กล่าวคือ **“การทำหน้าที่เบี่ยงเบนของข้อสอบ”** สามารถพิจารณาได้จากผู้สอบที่อยู่ต่างกลุ่มกัน มีความน่าจะเป็นในการตอบข้อสอบถูกแตกต่างกัน ภายหลังจากการจับคู่ความสามารถที่สนใจ ในขณะที่ความแตกต่างของผลการตอบข้อสอบที่ไม่ได้มีการจับคู่เรียกว่า **“ผลกระทบของข้อสอบ”** (Item impact) ซึ่งเกิดขึ้นจากความแตกต่างในการแจกแจงความสามารถของผู้สอบทั้งหมด (Camilli. 1992: 131; Clauser; & Mazor. 1998: 31; Dorans; & Holland. 1993: 37; Zumbo. 1999: 12) ดังนั้นถ้าใช้เทคนิคการตรวจสอบภายใต้โมเดลแบบมิติเดียวจะทำให้ผลการตรวจสอบเกิดความผิดพลาด ทั้งนี้เพราะไม่สามารถแยกออกจากกัน ระหว่างการทำหน้าที่เบี่ยงเบนของข้อสอบและผลกระทบของข้อสอบ แต่วิธีชิปเทสต์ที่ใช้เทคนิคการตรวจสอบภายใต้ทฤษฎีการตอบข้อสอบแบบหลายมิติ สามารถคำนวณคะแนนเกณฑ์ที่ใช้ในการจับคู่แยกจากข้อสอบที่ศึกษา ผลการตอบข้อสอบถูกจึงเป็นความแตกต่างของความสามารถแทรกซ้อน ซึ่งเป็นอิสระจากความสามารถเป้าหมาย ถึงแม้ว่าการแจกแจงความสามารถเป้าหมายระหว่างกลุ่มจะไม่

เท่ากัน ก็สามารถปรับแก้ค่าการถดถอย (Regression correction) เพื่อลดความแตกต่างของค่าความสามารถเป้าหมาย ทำให้สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ที่สูงเกินปกติได้ (Clausner; & Mazor. 1998: 35)

ส่วนวิธีการถดถอยโลจิสติกมีแนวคิดภายใต้การวัดแบบมิติเดียว เมื่อนำมาตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบแบบหลายมิติจะใช้คะแนนรวมแทนความสามารถของผู้สอบ ซึ่งเป็นค่าเฉลี่ยของคะแนนในหลายมิติ ทำให้ไม่สามารถแยกออกจากกันระหว่างความสามารถเป้าหมายและความสามารถแทรกซ้อน ดังนั้นความแตกต่างระหว่างความสามารถเป้าหมายยังคงมีอยู่ และแสดงผลในรูปค่าประมาณทางสถิติของการทำหน้าที่เบี่ยงเบนของข้อสอบ (Camilli. 1992: 131) ซึ่งจะทำให้ผลการตรวจสอบมีความคลาดเคลื่อนประเภทที่ 1 สูงเกินปกติ ในการศึกษาครั้งนี้ ผู้วิจัยได้ปรับขยายโมเดลการถดถอยโลจิสติก โดยนำค่าความสามารถทั้งสองมิติเป็นเกณฑ์การจับคู่ ซึ่งคาดว่าจะสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ที่สูงเกินปกติได้ ความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มที่ใช้ในการศึกษามี 3 ระดับ คือ การแจกแจงความสามารถระหว่างกลุ่มเท่ากัน (แตกต่างกัน 0.0SD) การแจกแจงความสามารถระหว่างกลุ่มแตกต่างกัน 0.5SD และการแจกแจงความสามารถระหว่างกลุ่มแตกต่างกัน 1.0SD ส่วนค่าความแปรปรวนของความสามารถในแต่ละกลุ่มกำหนดคงที่ เท่ากับ 1.0 และค่าสหสัมพันธ์ระหว่างความสามารถในมิติที่หนึ่งและมิติที่สองของแต่ละกลุ่มกำหนดคงที่ เท่ากับ 0.5

ขนาดตัวอย่าง

ปัจจัยสุดท้ายที่สนใจ คือ ขนาดตัวอย่าง ซึ่งมีความสำคัญในการทดสอบนัยสำคัญทางสถิติ วิธีการถดถอยโลจิสติกแบบจัดอันดับมีรูปแบบพาราเมตริก (Parametric form) ซึ่งจำเป็นต้องใช้ข้อตกลงของโมเดลอธิบายความสัมพันธ์ระหว่างผลการตอบข้อสอบกับตัวแปรการจับคู่ ในขณะที่วิธีโพลี-ชิปเทสที่มีรูปแบบนพาราเมตริก (Nonparametric form) ซึ่งไม่มีข้อตกลงดังกล่าว (Potenza; & Dorans. 1995: 24) ตามปกติแล้ววิธีที่มีรูปแบบพาราเมตริกมักใช้ขนาดตัวอย่างมากกว่าวิธีที่มีรูปแบบนพาราเมตริก ถึงแม้ว่าทั้งสองวิธีจะมีรูปแบบทางสถิติแตกต่างกัน แต่ทั้งสองวิธีใช้สถิติในการทดสอบภายใต้การแจกแจงแบบเชิงเส้นกำกับ (Asymptotic distribution) ซึ่งการแจกแจงดังกล่าวจำเป็นต้องใช้ขนาดตัวอย่างที่เพียงพอ เพื่อให้ผลการวิเคราะห์มีความถูกต้องจากการศึกษาของนักวิจัยหลายคน (Bolt. 2002; Chang; Mazzeo; & Roussos. 1996; French; & Miller. 1996; Gierl; Gotzmann; & Boughton. 2004; Narayanan; & Swaminathan. 1994, 1996;

Rogers; & Swaminathan. 1993; Roussos; & Stout. 1996a; Stout; et al. 1997; Swaminathan; & Rogers. 1990; Whitmore; & Schumacker. 1999) ให้ผลสอดคล้องกัน กล่าวคือ เมื่อขนาดตัวอย่างเพิ่มขึ้นมีผลทำให้วิธีในตระกูลซิปเทสท์ (SIBTEST, CRO-SIB, MULTISIB, Poly-SIBTEST) และวิธีในตระกูลโมเดลการถดถอยโลจิสติก (Logistic regression model, Continuation ratio logits model, Cumulative logits model, Adjacent categories logits model) มีอัตราความถูกต้องเพิ่มขึ้น ในการวิจัยครั้งนี้ศึกษาขนาดตัวอย่าง 4 ขนาด คือ 250, 500, 1,000 และ 2,000 คน โดยกลุ่มอ้างอิงกับกลุ่มสนใจมีขนาดตัวอย่างเท่ากัน

สำหรับความยาวของแบบทดสอบ ไม่ได้เป็นปัจจัยที่ใช้ในการศึกษา กำหนดคงที่เท่ากับ 40 ข้อ ในทุกเงื่อนไข ส่วนขนาดของข้อสอบทำหน้าที่เบี่ยงเบนไม่ได้เป็นปัจจัยที่ศึกษาเช่นกัน โดยกำหนดคงที่ตามรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่ศึกษา ดังนั้นในการวิเคราะห์การทำหน้าที่เบี่ยงเบนของข้อสอบจะต้องจำลองข้อมูลภายใต้การจัดกระทำปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน 2 รูปแบบ สัดส่วนของข้อสอบที่ทำหน้าที่เบี่ยงเบน 3 ขนาด ความแตกต่างของการแจกแจงความสามารถ 3 ระดับ และขนาดตัวอย่าง 4 ขนาด รวมข้อมูลที่ต้องจำลองทั้งหมด 72 เงื่อนไข ($2 \times 3 \times 3 \times 4$) ในแต่ละเงื่อนไขจำลองข้อมูลซ้ำ 50 ครั้ง

สมมติฐานในการวิจัย

จากกรอบแนวคิดในการวิจัยจะเห็นว่า การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในข้อสอบที่วัดความสามารถมิติเดียวและให้คะแนนหลายค่า โดยใช้วิธีโพลี-ซิปเทสท์และวิธีที่ใช้โมเดลการถดถอยโลจิสติกมีข้อได้เปรียบและข้อเสียเปรียบแตกต่างกัน ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่าง ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติเป็นวิธีใหม่ที่ผู้วิจัยเสนอในการศึกษานี้ ยังไม่มีข้อมูลสนับสนุนหรือคัดค้าน ดังนั้นผู้วิจัยจึงตั้งสมมติฐานดังนี้

1. ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ประกอบด้วย รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่าง มีผลทำให้อัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า ด้วยวิธีโพลี-ซิปเทสท์ แตกต่างกัน

2. ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ประกอบด้วย รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่าง มีผลทำให้อัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า ด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับแตกต่างกัน

3. ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ประกอบด้วย รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่าง มีผลทำให้อัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า ด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติแตกต่างกัน

4. วิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ประกอบด้วย รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่าง มีอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกัน

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

ในการวิจัยครั้งนี้ มีความมุ่งหมายเพื่อเปรียบเทียบวิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่าง ในการศึกษาดังกล่าว ผู้วิจัยได้ศึกษากรอบแนวคิด ทฤษฎี หลักการ รูปแบบ และวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ตลอดจนงานวิจัยต่างๆ ที่เกี่ยวข้อง เพื่อนำองค์ความรู้ที่ได้จากการศึกษาค้นคว้ามาประยุกต์ใช้กับการศึกษาครั้งนี้ แนวทางของการศึกษาค้นคว้า สามารถกำหนดเป็นกรอบการศึกษาใน 5 ประเด็น ดังนี้

1. นิยามของการทำหน้าที่เบี่ยงเบนของข้อสอบ
2. รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน
3. แนวคิดและหลักการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ
4. วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ
5. งานวิจัยที่เกี่ยวข้อง

1. นิยามของการทำหน้าที่เบี่ยงเบนของข้อสอบ

การศึกษาเกี่ยวกับ “ความยุติธรรมของแบบทดสอบ” (Test Fairness) มีมาตั้งแต่ทศวรรษ 1910 โดยเริ่มต้นจาก อัลเฟรด บินเน็ต (Camilli; & Shepard. 1994: 4-5; citing Binet; & Simon. 1916, 1973) ได้มีการโต้เถียงเกี่ยวกับแบบทดสอบวัดเชาว์ปัญญาที่ทำให้ผู้สอบขาดความยุติธรรม กล่าวคือ นักเรียนที่มีสถานะทางเศรษฐกิจและสังคมสูง มีผลการสอบดีกว่านักเรียนที่มีสถานะทางเศรษฐกิจและสังคมต่ำอย่างมาก บินเน็ตได้ปรับปรุงแก้ไขแบบทดสอบดังกล่าว โดยพิจารณาข้อสอบที่ทำให้นักเรียนขาดความยุติธรรม แล้วคัดออกจากแบบทดสอบ ต่อมาได้มีการศึกษาเรื่องความยุติธรรมของข้อสอบอย่างจริงจังตั้งแต่ทศวรรษ 1960 ในประเทศสหรัฐอเมริกา ซึ่งเป็นประเทศที่มีความแตกต่างของประชากรเกี่ยวกับเชื้อชาติ เผ่าพันธุ์ ศาสนา ภาษา วัฒนธรรม ประสิทธิภาพ สถานะทางเศรษฐกิจและสังคม หรือภูมิหลังอื่นๆ นักวิจัยได้ให้ความสำคัญใน

ประเด็นดังกล่าว เพราะว่าในวงการศึกษาระดับชั้นและหน่วยงานต่างๆ ได้มีการนำผลการทดสอบมาใช้พิจารณาตัดสินใจในเรื่องต่างๆ มากยิ่งขึ้น เช่น การคัดเลือกเพื่อเข้าศึกษาต่อ การคัดเลือกเพื่อบรรจุงาน การเลื่อนขั้นหรือตำแหน่ง การออกใบอนุญาตหรือใบรับรอง เป็นต้น แบบทดสอบที่นำมาใช้ควรเป็นแบบทดสอบมาตรฐานที่มีความเที่ยงตรง ความเชื่อมั่น ความยากเหมาะสม และจำแนกระดับความสามารถผู้สอบได้ นอกจากยังต้องคำนึงถึงความยุติธรรมต่อผู้สอบ เป้าหมายสำคัญของแบบทดสอบมาตรฐานเพื่อต้องการจำแนกระดับความสามารถของผู้สอบ แล้วจัดเรียงลำดับความสามารถของผู้สอบแต่ละคน (Ackerman. 1992: 67) ข้อสอบหรือแบบทดสอบที่ขาดความยุติธรรมจะเกิดความลำเอียง โดยข้อสอบหรือแบบทดสอบอาจเข้าข้างผู้สอบกลุ่มใดกลุ่มหนึ่งซึ่งทำให้ผู้สอบกลุ่มนั้นได้เปรียบในการตอบข้อสอบ โดยมีโอกาสตอบข้อสอบถูกได้มากกว่า ส่วนผู้สอบอีกกลุ่มหนึ่งเสียเปรียบ ทั้งๆ ที่ผู้สอบสองกลุ่มมีระดับความสามารถเท่ากัน ความแตกต่างของผลการตอบข้อสอบระหว่างกลุ่มดังกล่าวเป็นผลมาจากความแตกต่างของวัฒนธรรม ซึ่งเป็นภูมิหลังของผู้สอบมากกว่าความสามารถของผู้สอบ (Angoff, 1993: 3) ความลำเอียงของข้อสอบหรือแบบทดสอบ (Item bias/Test bias) เป็นความคลาดเคลื่อนอย่างเป็นระบบ (Systematic error) (Camilli; & Shepard. 1994: 8) ซึ่งจะส่งผลให้แบบทดสอบขาดความเที่ยงตรง

หนังสือ "Standards for Educational and Psychological Testing" (AERA/APA/NCME. 1999: 74-76) ได้อธิบายความยุติธรรมของแบบทดสอบไว้ว่า ความยุติธรรมของแบบทดสอบไม่ได้มีความหมายเดียว (Single meaning) ดังนั้นจึงไม่สามารถกำหนดเป็นนิยามที่มีความหมายเฉพาะ หนังสือ "Standards" ได้ให้ข้อสังเกตแนวคิดของความยุติธรรมของแบบทดสอบไว้ 4 ลักษณะ คือ

ความหมายแรก แบบทดสอบที่มีความยุติธรรมจะต้องปราศจากความลำเอียง ความลำเอียงเกิดขึ้นเมื่อแปลคะแนนของแบบทดสอบมีความหมายแตกต่างกัน ซึ่งเป็นผลมาจากความแตกต่างของสมาชิกในกลุ่ม **ความหมายที่สอง** แบบทดสอบที่มีความยุติธรรมจำเป็นต้องให้ผู้สอบทั้งหมดได้รับการปฏิบัติในกระบวนการทดสอบเท่ากัน ในความหมายนี้เป็นเงื่อนไขของการทดสอบที่ทำให้ผู้สอบมีโอกาสเท่ากัน เช่น รูปแบบของแบบทดสอบ วัสดุที่ใช้ในการฝึก ความสะดวกของผู้สอบที่ทุพพลภาพ การดำเนินการสอบ เป็นต้น **ความหมายที่สาม** แบบทดสอบที่มีความยุติธรรมจำเป็นต้องให้ผลของการทดสอบมีความเสมอภาค ซึ่งผู้สอบทุกคนมีโอกาสแสดงศักยภาพตามโครงสร้างของแบบทดสอบที่ไว้วัด เมื่อแบบทดสอบไม่ลำเอียงและผู้สอบได้รับการปฏิบัติเท่ากันในระหว่างการทดสอบ ดังนั้นผู้สอบที่มีความสามารถเท่ากันซึ่งวัดโดยแบบทดสอบดังกล่าวย่อมได้คะแนนสอบเท่ากัน และ**ความหมายสุดท้าย** แบบทดสอบที่มีความยุติธรรมจะต้องให้ผู้สอบมีโอกาสได้

เรียนรู้เนื้อหาของแบบทดสอบอย่างครอบคลุม ซึ่งแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนทั่วไป เปิดโอกาสให้ผู้สอบได้เรียนรู้ตามปกติ จากแนวคิดความยุติธรรมของแบบทดสอบดังกล่าวจะเห็นว่า ความหมายแรกและความหมายที่สามมีความสำคัญอย่างยิ่งต่อการทดสอบและวัดผลการศึกษา ซึ่งนักพัฒนาแบบทดสอบจำเป็นต้องตรวจสอบความลำเอียงของข้อสอบ เพื่อแสดงหลักฐานทาง สถิติ แล้วคัดเลือกข้อสอบที่มีความลำเอียงออกจากแบบทดสอบ กระบวนการตรวจสอบดังกล่าว ถือเป็นส่วนหนึ่งของการพัฒนาแบบทดสอบ ซึ่งมีความสำคัญไม่ยิ่งหย่อนไปกว่าการตรวจสอบ ความเที่ยงตรง ความเชื่อมั่น และคุณสมบัติของข้อสอบ

ในปัจจุบันนักวิจัยใช้คำว่า “การทำหน้าที่เบี่ยงเบนของข้อสอบ” (Differential item functioning) และ “การทำหน้าที่เบี่ยงเบนของแบบสอบ” (Differential test functioning) แทน คำว่า “ความลำเอียงของข้อสอบ” (Item bias) และ “ความลำเอียงของแบบทดสอบ” (Test bias) ตามลำดับ โดยเรียกสั้นๆ ว่า “DIF” และ “DTF” เนื่องจากความลำเอียงของข้อสอบหรือ แบบทดสอบมีความหมายเชิงสถิติและความหมายเชิงสังคม ซึ่งทำให้สับสนในการนำไปใช้ ต่อมา ได้มีผู้เสนอให้ใช้คำว่า การทำหน้าที่เบี่ยงเบนของข้อสอบหรือแบบทดสอบ ซึ่งมีความเหมาะสม มากกว่า สามารถนำไปประยุกต์ใช้ในการเปรียบเทียบความแตกต่างระหว่างกลุ่ม โดยมุ่งเน้นการ- ทดสอบทางสถิติ ซึ่งแตกต่างจากการประเมินโดยการตัดสินในความหมายเชิงสังคม (Angoff. 1993: 4-5; Hambleton; Swaminathan; & Rogers. 1991: 109) สำหรับความหมายของการทำหน้าที่ เบี่ยงเบนของข้อสอบได้มีนักวิจัยหลายคนกำหนดเป็นนิยาม ดังนี้

การทำหน้าที่เบี่ยงเบนของข้อสอบ หมายถึง ผู้สอบที่มีความสามารถระดับเดียวกัน แต่ มาจากกลุ่มผู้สอบที่แตกต่างกัน มีความน่าจะเป็นในการตอบข้อสอบถูกไม่เท่ากัน (Hambleton; Swaminathan; & Rogers. 1991: 110)

การทำหน้าที่เบี่ยงเบนของข้อสอบ หมายถึง การสังเกตได้ว่าข้อสอบแสดงคุณสมบัติทาง สถิติที่แตกต่างกัน ในผลการตอบข้อสอบจากกลุ่มผู้สอบแตกต่างกัน ภายหลังการควบคุมความ- แตกต่างในความสามารถของกลุ่มผู้สอบ (Angoff. 1993: 4-5)

การทำหน้าที่เบี่ยงเบนของข้อสอบ หมายถึง ความแตกต่างในการทำหน้าที่ของข้อสอบ ภายหลังการจับคู่ความสามารถของกลุ่มผู้สอบ ซึ่งเป็นความสามารถตามที่ข้อสอบต้องการวัด (Dorans; & Holland. 1993: 37)

การทำหน้าที่เบี่ยงเบนของข้อสอบ หมายถึง การเปรียบเทียบความสามารถระหว่าง ผู้สอบสองกลุ่มหรือมากกว่า ความน่าจะเป็นในการตอบข้อสอบถูกมีความเกี่ยวข้องกับการเป็น สมาชิกของกลุ่มผู้สอบ (Camilli. 1993: 397)

การทำหน้าที่เบี่ยงเบนของข้อสอบ หมายถึง ความแตกต่างในผลการตอบข้อสอบ ระหว่างกลุ่มผู้สอบที่เปรียบเทียบกัน โดยกลุ่มผู้สอบดังกล่าวมีการจับคู่ความสามารถที่วัดโดยแบบทดสอบ (Potenza; & Dorans. 1995: 23)

การทำหน้าที่เบี่ยงเบนของข้อสอบ หมายถึง ผู้สอบที่มีความสามารถระดับเดียวกัน (ความสามารถมิติเดียว หรือความสามารถหลายมิติที่วัดโดยแบบทดสอบ) แต่มาจากประชากรหลายกลุ่มที่แตกต่างกัน มีความน่าจะเป็นในการตอบข้อสอบถูกแตกต่างกัน (Roussos; & Stout. 1996a: 355)

การทำหน้าที่เบี่ยงเบนของข้อสอบ หมายถึง ข้อสอบมีความน่าจะเป็นในการตอบถูกแตกต่างกัน สำหรับผู้สอบที่ระดับความสามารถเดียวกัน แต่มาจากกลุ่มที่แตกต่างกัน (Stout; et al. 1997: 195)

การทำหน้าที่เบี่ยงเบนของข้อสอบ หมายถึง การเปรียบเทียบระดับความสามารถของผู้สอบ ซึ่งมาจากกลุ่มที่แตกต่างกัน มีความน่าจะเป็นในการตอบข้อสอบถูกแตกต่างกัน (Mazor; Hambleton; & Clauser. 1998: 357)

การทำหน้าที่เบี่ยงเบนของข้อสอบ หมายถึง ผู้สอบจากกลุ่มที่แตกต่างกัน มีความน่าจะเป็นที่จะประสบความสำเร็จในการตอบข้อสอบแตกต่างกัน ภายหลังการจับคู่ความสามารถตามที่ข้อสอบต้องการวัด (Clauser; & Mazor. 1998: 40)

การทำหน้าที่เบี่ยงเบนของข้อสอบ หมายถึง คุณสมบัติทางสถิติของข้อสอบ ซึ่งผู้สอบที่อยู่ต่างกลุ่มกันมีคะแนนรวมของแบบทดสอบเท่ากัน แต่มีคะแนนเฉลี่ยของข้อสอบแตกต่างกัน (AERA/APA/NCME. 1999: 175)

การทำหน้าที่เบี่ยงเบนของข้อสอบ หมายถึง ผู้สอบจากกลุ่มประชากรสองกลุ่มที่แตกต่างกัน ภายใต้อัตราความสามารถเดียวกัน มีความน่าจะเป็นที่จะประสบความสำเร็จในการตอบข้อสอบไม่เท่ากัน (Bolt. 2002: 115)

การทำหน้าที่เบี่ยงเบนของข้อสอบ หมายถึง ผู้สอบจากกลุ่มที่แตกต่างกัน ซึ่งมีระดับความสามารถเท่ากันตามที่ข้อสอบประเมิน มีความน่าจะเป็นในการตอบข้อสอบถูกแตกต่างกัน หรือมีคะแนนดิบที่คาดหวังจากการตอบข้อสอบแตกต่างกัน (Raju; & Ellis, 2002: 156)

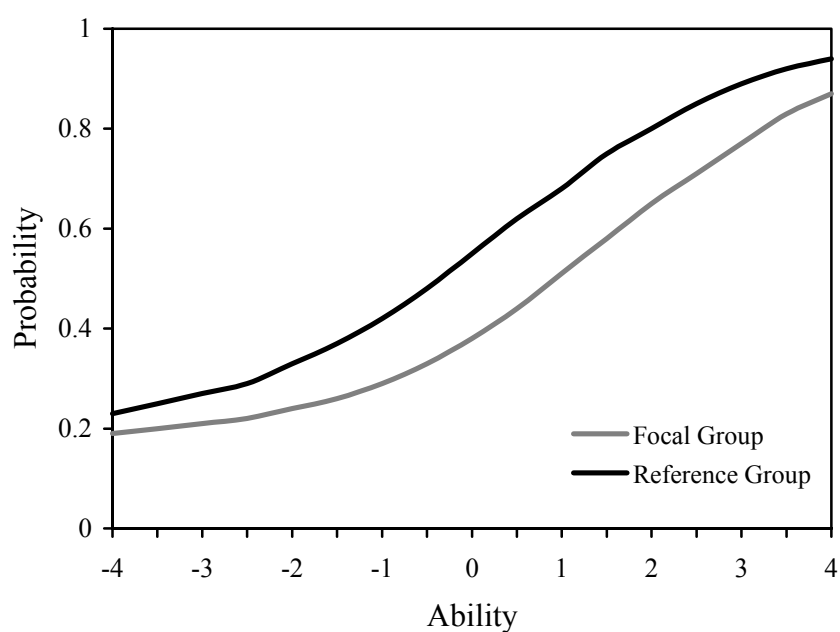
การทำหน้าที่เบี่ยงเบนของข้อสอบ หมายถึง ผู้สอบจากกลุ่มที่แตกต่างกัน มีความน่าจะเป็นในการตอบข้อสอบถูกแตกต่างกัน ภายหลังควบคุมความสามารถ (Gierl; Gotzmann; & Boughton. 2004: 242)

จากนิยามดังกล่าวสรุปได้ว่า **การทำหน้าที่เบี่ยงเบนของข้อสอบ หมายถึง ผู้สอบจากกลุ่มที่แตกต่างกัน และมีการจับคู่ความสามารถที่แบบทดสอบหรือข้อสอบ ต้องการวัดเท่ากัน มีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องไม่เท่ากัน** ดังนั้นในการศึกษาครั้งนี้ ผู้วิจัยใช้ความหมายดังกล่าวเป็นนิยามศัพท์เฉพาะในการวิจัย

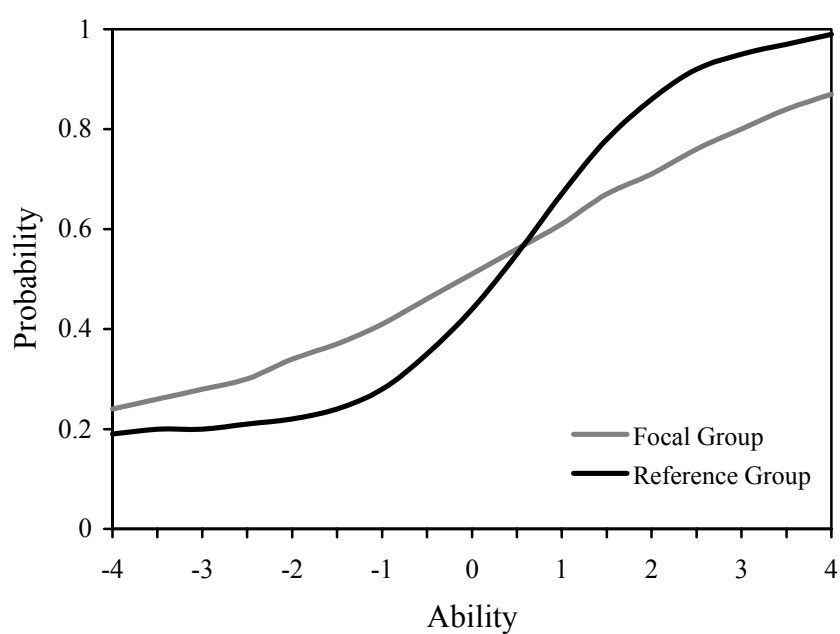
2. รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน

เมลเลนเบิร์ก (Mellenbergh. 1982) ได้ตรวจสอบการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนสองค่า (Dichotomous) โดยใช้โมเดลล็อก-ลิเนียร์ (Log-linear model) และโมเดลโลจิต (Logit model) ทำนายผลการตอบข้อสอบ ภายใต้ตารางการถัวจร (Contingency table) แบบสามมิติ ซึ่งประกอบด้วยปัจจัยสามปัจจัยหลัก คือ การเป็นสมาชิกของกลุ่ม (Group membership) ระดับความสามารถ (Ability level) และปฏิสัมพันธ์ (Interaction) ระหว่างการเป็นสมาชิกของกลุ่ม กับระดับความสามารถ ในการศึกษาได้จำแนกข้อสอบทำหน้าที่เบี่ยงเบนออกเป็นสองรูปแบบ คือ “ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน” (Uniform DIF) และ “ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน” (Nonuniform DIF) ข้อสอบทำหน้าที่เบี่ยงเบนรูปแบบแรกเกิดขึ้นเมื่อไม่มีปฏิสัมพันธ์ระหว่างระดับความสามารถกับการเป็นสมาชิกของกลุ่ม แสดงว่าความน่าจะเป็นในการตอบข้อสอบถูกของผู้สอบกลุ่มหนึ่งมากกว่าอีกกลุ่มหนึ่งอย่างคงที่ทุกระดับความสามารถ ส่วนข้อสอบทำหน้าที่เบี่ยงเบนรูปแบบหลังเกิดขึ้นเมื่อมีปฏิสัมพันธ์ระหว่างระดับความสามารถกับการเป็นสมาชิกของกลุ่ม แสดงว่าความแตกต่างของความน่าจะเป็นในการตอบข้อสอบถูกของผู้สอบสองกลุ่มไม่เหมือนกันทุกระดับความสามารถ

ในทฤษฎีการตอบข้อสอบ (Item response theory; IRT) สามารถพิจารณาปฏิสัมพันธ์ระหว่างระดับความสามารถกับการเป็นสมาชิกของกลุ่มได้จากความแตกต่างของค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ (Item discrimination) ระหว่างกลุ่มผู้สอบ กล่าวคือ ถ้าค่าพารามิเตอร์อำนาจจำแนกของข้อสอบระหว่างกลุ่มผู้สอบเท่ากัน แล้วฟังก์ชันการตอบข้อสอบ (Item response functions; IRFs) ระหว่างกลุ่มขนานกัน แสดงว่าข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน แต่ถ้าค่าพารามิเตอร์อำนาจจำแนกของข้อสอบระหว่างกลุ่มผู้สอบไม่เท่ากัน จะทำให้ฟังก์ชันการตอบข้อสอบระหว่างกลุ่มไม่ขนานกัน แสดงว่าข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Li; & Stout. 1996: 647) รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนทั้งสอง แสดงดังภาพประกอบ



ภาพประกอบ 1 ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF)
ภายใต้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ (3PLM)

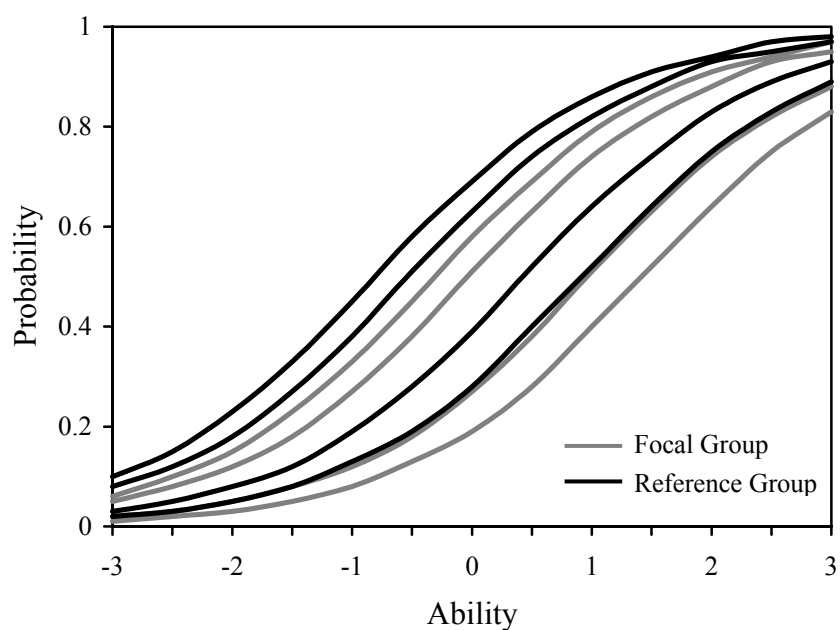


ภาพประกอบ 2 ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF)
ภายใต้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ (3PLM)

ที่มา: Hambleton; Swaminathan; & Rogers. (1991). Fundamental of Item Response Theory. p. 118-119.

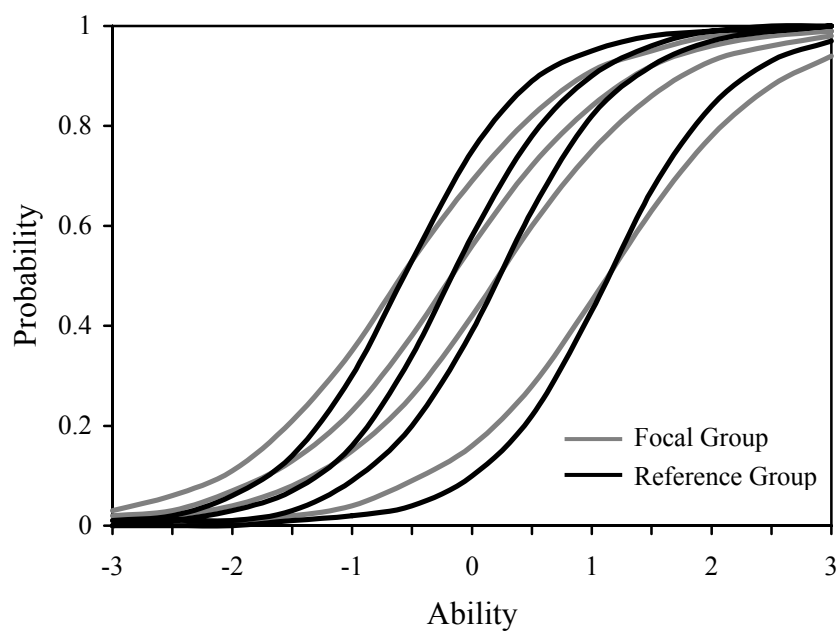
เมื่อพิจารณานิยามข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกันดังกล่าวจะเห็นว่า ฟังก์ชันการตอบข้อสอบระหว่างกลุ่มผู้สอบซึ่งไม่ขนานกันอาจจะตัดกันหรือไม่ตัดกันก็ได้ จากผล การศึกษาของสวามินาธานและโรเจอร์ส (Swaminathan; & Rogers. 1990: 369) พบว่า รูปแบบ ของข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกันสามารถจำแนกออกเป็นสองรูปแบบย่อย กล่าวคือ ในช่วงความสามารถตามมาตรฐานของทฤษฎีการตอบข้อสอบ เมื่อฟังก์ชันการตอบข้อสอบ ระหว่างกลุ่มผู้สอบตัดกันตรงกึ่งกลางของช่วงความสามารถดังกล่าว ข้อสอบทำหน้าที่เบี่ยงเบนที่ ไม่เป็นรูปแบบเดียวกันเมื่อวิเคราะห์ด้วยโมเดล ANOVA จะมีลักษณะ “ปฏิสัมพันธ์ไม่เป็นลำดับ” (Disordinal interaction) เมื่อฟังก์ชันการตอบข้อสอบระหว่างกลุ่มตัดกันนอกช่วงความสามารถ หรือฟังก์ชันการตอบข้อสอบระหว่างกลุ่มซึ่งไม่ขนานกันแต่ไม่ตัดกัน (สถานการณ์นี้อาจเกิดขึ้นเมื่อ ใช้โมเดลแบบ 3 พารามิเตอร์) ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกันมีลักษณะ “ปฏิสัมพันธ์เป็นลำดับ” (Ordinal interaction) ลิและสเตาท์ (Narayanan; & Swaminathan. 1996: 258; citing Li; & Stout. 1993) ได้เรียกรูปแบบของข้อสอบทั้งสองว่า **“ข้อสอบทำหน้าที่ เบี่ยงเบนที่ไม่มีทิศทางเดียวกัน” (Nonunidirectional DIF) และ “ข้อสอบทำหน้าที่เบี่ยงเบน ที่มีทิศทางเดียวกัน” (Unidirectional DIF) ตามลำดับ เพื่อไม่ให้เกิดความสับสนในการเรียกชื่อ ดังกล่าว ลิและสเตาท์ (Li; & Stout. 1996: 648) ได้ใช้คำใหม่ว่า “ข้อสอบทำหน้าที่เบี่ยงเบนแบบ ตัดกัน” (Crossing DIF) และ “ข้อสอบทำหน้าที่เบี่ยงเบนแบบไม่ตัดกัน” (Non-crossing DIF) แต่คำดังกล่าวไม่เป็นที่นิยมใช้กันส่วนมากจะใช้คำเดิมมากกว่า**

สำหรับเงื่อนไขของ “ปฏิสัมพันธ์” ดังกล่าวเป็นรูปแบบของการทำหน้าที่เบี่ยงเบนของ ข้อสอบให้คะแนนสองค่า (Dichotomous) ส่วนข้อสอบที่ให้คะแนนหลายค่า (Polytomous) จะมี ความสมบูรณ์มากกว่า ทั้งนี้เพราะว่าไม่เพียงจะเกิดปฏิสัมพันธ์ระหว่างระดับความสามารถกับ การเป็นสมาชิกของกลุ่มเท่านั้น แต่ยังมีตัวแปรที่สาม คือ ระดับคะแนนของข้อสอบเข้ามาเกี่ยวข้อง ด้วย ดังนั้นการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนหลายค่าจะเกิดขึ้นภายในรายการคะแนน ทั้งหมด (Score categories) (French; & Miller. 1996: 315) รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ที่ให้คะแนนหลายค่าจะมีลักษณะคล้ายกับรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่ให้คะแนนสอง ค่า ซึ่งสามารถแสดงด้วยฟังก์ชันการตอบขอบข่ายการ (Boundary response functions; BRFs) โดยใช้โมเดลเกรดเรสพอน (Graded response model; GRM) ดังภาพประกอบ 3 และ 4 สำหรับ ฟังก์ชันการตอบรายการข้อสอบ (Item-category response functions; ICRFs) โดยใช้โมเดล พาเชียลเครดิตทั่วไป (Generalized partial credit model; GPCM) แสดงดังภาพประกอบ 5 และ 6



ภาพประกอบ 3 ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF)

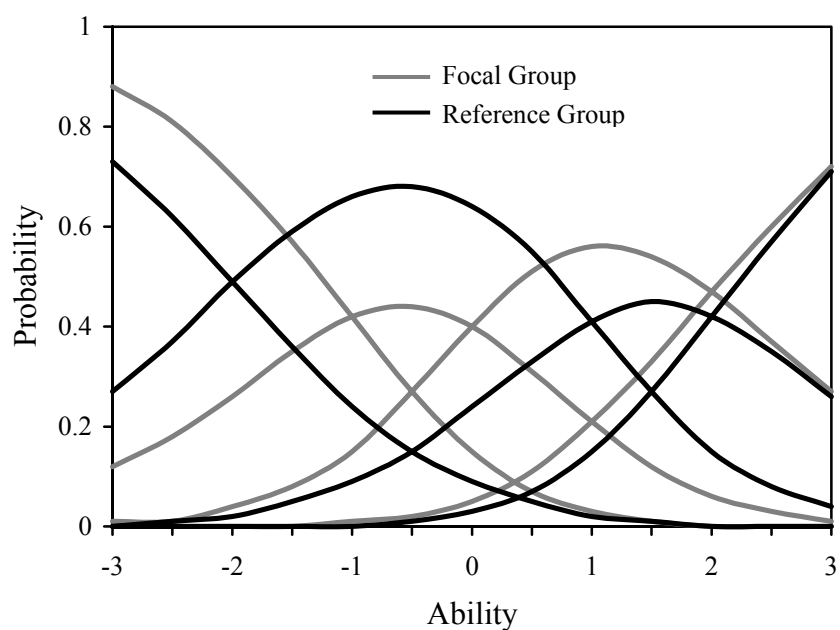
ภายใต้โมเดลเกรดเรสพอน (GRM)



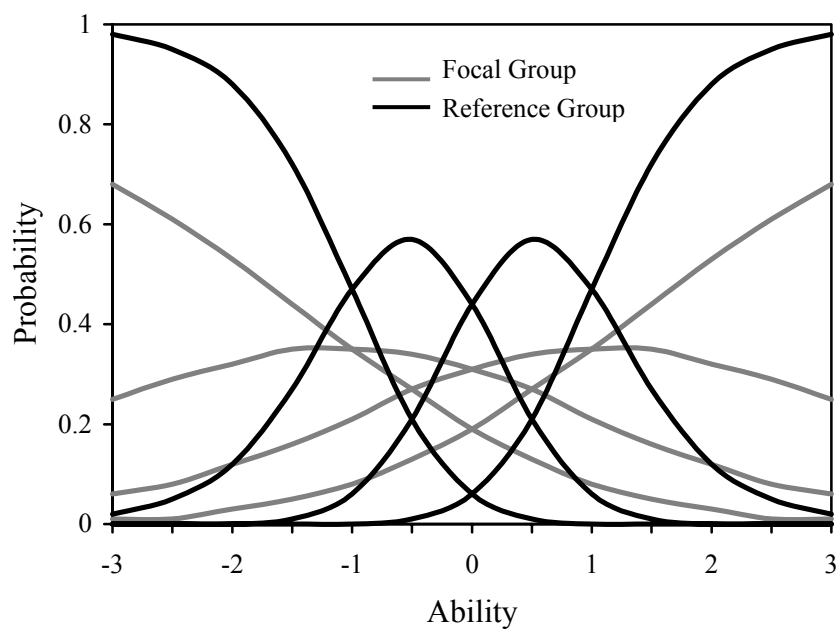
ภาพประกอบ 4 ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF)

ภายใต้โมเดลเกรดเรสพอน (GRM)

ที่มา: Cohen; Kim; & Baker. (1993). Detection of Differential Item Functioning in the Graded Response Model. p. 338.



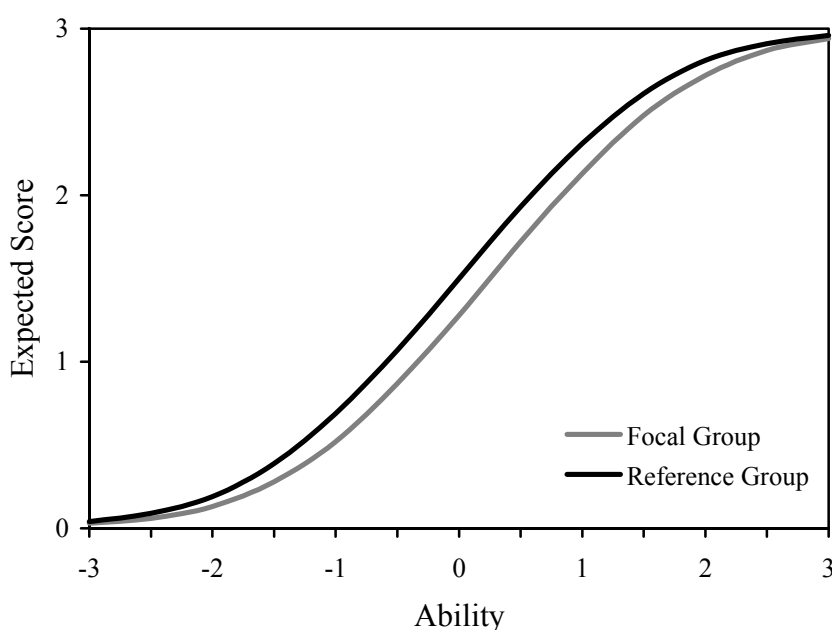
ภาพประกอบ 5 ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF)
ภายใต้โมเดลพหุเชิงเส้นเครดิตทั่วไป (GPCM)



ภาพประกอบ 6 ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF)
ภายใต้โมเดลพหุเชิงเส้นเครดิตทั่วไป (GPCM)

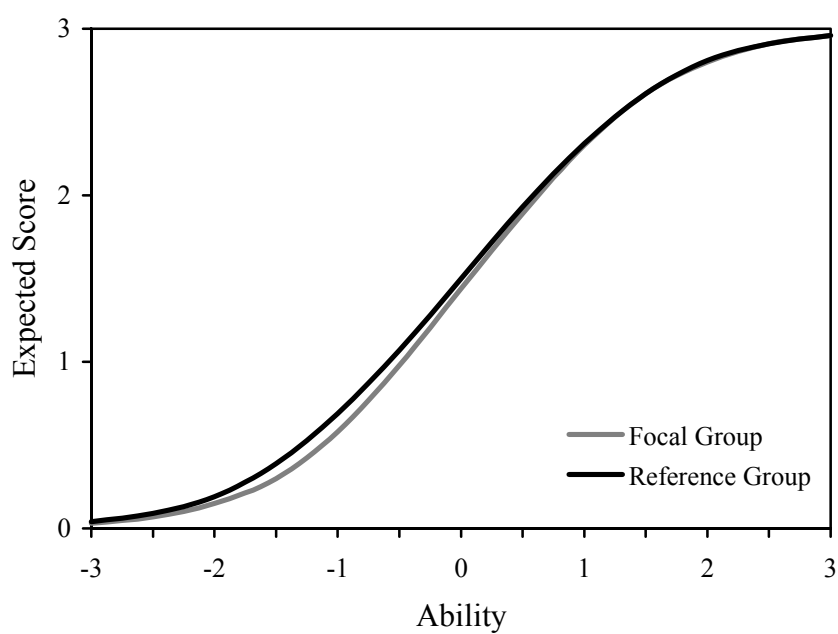
ที่มา: French; & Miller. (1996). Logistic Regression and Its Use in Detecting Differential Item Functioning in Polytomous Items. p. 322.

นอกจากนี้ ซวิก โดโนฮิว และกริมา (Zwick; Donoghue; & Grima. 1993: 243-244) ได้ศึกษารูปแบบของข้อสอบทำหน้าที่ย่อยแบ่งที่เป็ยเบนที่ให้คะแนนหลายค่า และได้กำหนดนิยามโดยใช้ฟังก์ชันความน่าจะเป็นสะสมของการตอบในรายการคะแนนตอบ (Score categories) ทุกรายการ ซึ่งมีรูปแบบแตกต่างกัน 4 รูปแบบ คือ (1) ข้อสอบทำหน้าที่ย่อยแบ่งแบบคงที่ (Constant DIF) เกิดขึ้นเมื่อค่าพารามิเตอร์ความยากในทุกรายการของข้อสอบสำหรับกลุ่มสนใจมีค่าสูงกว่ากลุ่มอ้างอิงในระดับคงที่ (2) ข้อสอบทำหน้าที่ย่อยแบ่งแบบเปลี่ยนขนาดที่ระดับความสามารถต่ำ (Low-shift DIF) เกิดขึ้นเมื่อค่าพารามิเตอร์ความยากในรายการต่ำสุดของข้อสอบสำหรับกลุ่มสนใจมีค่าสูงกว่ากลุ่มอ้างอิง ส่วนค่าพารามิเตอร์ความยากในรายการของข้อสอบที่เหลือสำหรับสองกลุ่มมีค่าเท่ากัน (3) ข้อสอบทำหน้าที่ย่อยแบ่งแบบเปลี่ยนขนาดที่ระดับความสามารถสูง (High-shift DIF) เกิดขึ้นเมื่อค่าพารามิเตอร์ความยากในรายการสูงสุดของข้อสอบสำหรับกลุ่มสนใจมีค่าสูงกว่ากลุ่มอ้างอิง ส่วนค่าพารามิเตอร์ความยากในรายการของข้อสอบที่เหลือสำหรับสองกลุ่มมีค่าเท่ากัน และ (4) ข้อสอบทำหน้าที่ย่อยแบ่งแบบสมดุล (Balanced DIF) เกิดขึ้นเมื่อค่าพารามิเตอร์ความยากในรายการต่ำสุดของข้อสอบสำหรับกลุ่มสนใจมีค่าสูงกว่ากลุ่มอ้างอิง และค่าพารามิเตอร์ความยากในรายการสูงสุดของข้อสอบสำหรับกลุ่มสนใจมีค่าต่ำกว่ากลุ่มอ้างอิง ส่วนค่าพารามิเตอร์ความยากในรายการของข้อสอบที่เหลือสำหรับสองกลุ่มมีค่าเท่ากัน รูปแบบของข้อสอบทั้ง 4 รูปแบบสามารถแสดงด้วยฟังก์ชันการตอบข้อสอบสะสมในทุกรายการคำตอบ ดังภาพประกอบ 7-10

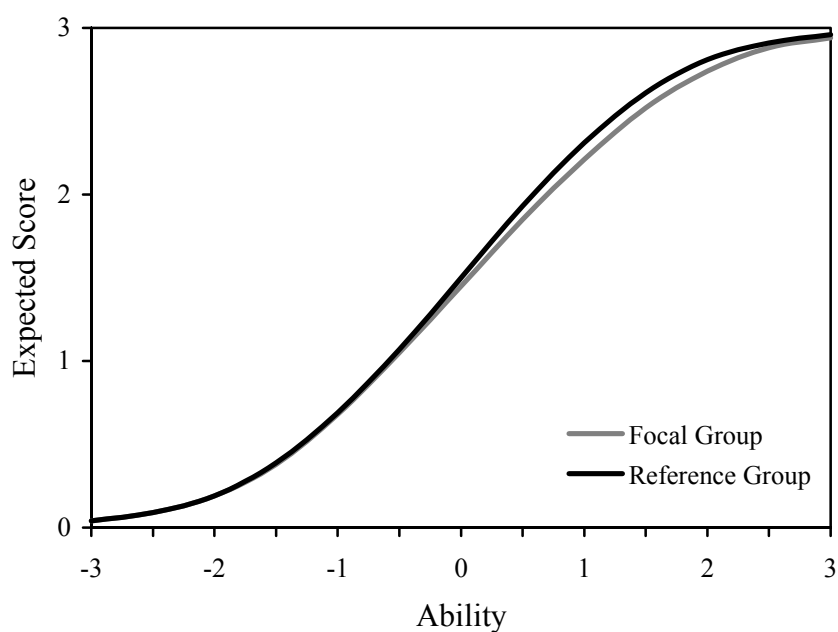


ภาพประกอบ 7 ข้อสอบทำหน้าที่ย่อยแบ่งแบบคงที่ (Constant DIF)

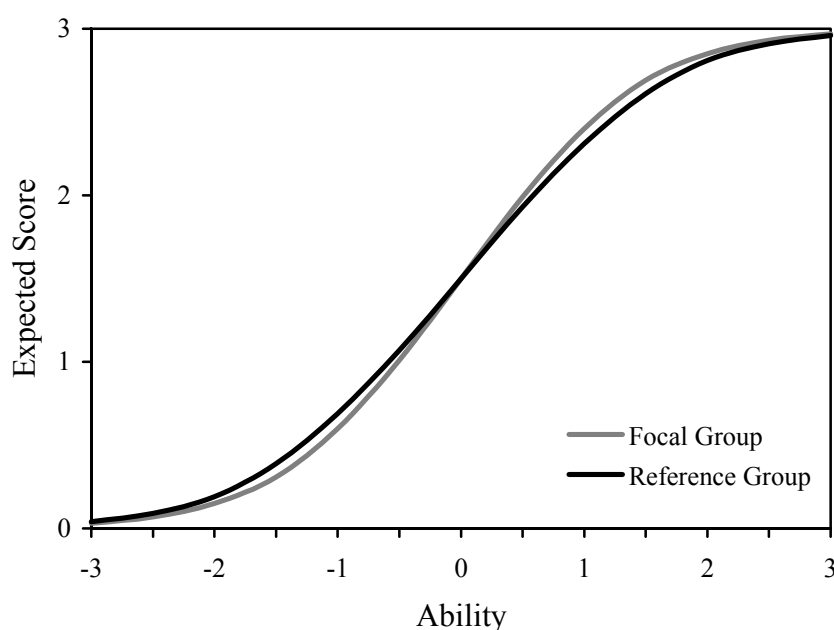
ภายใต้โมเดลพหุเชิงเส้นเครดิต (PCM)



ภาพประกอบ 8 ข้อสอบทำหน้าที่เบี่ยงเบนแบบเปลี่ยนขนาดที่ระดับความสามารถต่ำ
(Low-shift DIF) ภายใต้โมเดลพหุเชิงเส้นเครดิต (PCM)



ภาพประกอบ 9 ข้อสอบทำหน้าที่เบี่ยงเบนแบบเปลี่ยนขนาดที่ระดับความสามารถสูง
(High-shift DIF) ภายใต้โมเดลพหุเชิงเส้นเครดิต (PCM)



ภาพประกอบ 10 ข้อสอบทำหน้าที่เบี่ยงเบนแบบสมดุล (Balanced DIF)

ภายใต้โมเดลพหุเชิงเส้นเครดิต (PCM)

ที่มา: Chang; Mazzeo; & Roussos. (1996). Detecting DIF for Polytomously Scored Items: An Adaptation of the SIBTEST Procedure. p. 343.

จากภาพประกอบ 7-10 เมื่อเปรียบเทียบกับรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนในข้อสอบที่ให้คะแนนสองค่าจากภาพประกอบ 1 และ 2 จะเห็นว่า ข้อสอบทำหน้าที่เบี่ยงเบนแบบคงที่ (Constant DIF) ข้อสอบทำหน้าที่เบี่ยงเบนแบบเปลี่ยนขนาดที่ระดับความสามารถต่ำ (Low-shift DIF) และข้อสอบทำหน้าที่เบี่ยงเบนแบบเปลี่ยนขนาดที่ระดับความสามารถสูง (High-shift DIF) ก็คือ ข้อสอบทำหน้าที่เบี่ยงเบนที่มีทิศทางเดียวกัน (Unidirectional DIF) ส่วนข้อสอบทำหน้าที่เบี่ยงเบนแบบสมดุล (Balanced DIF) ก็คือ ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่มีทิศทางเดียวกัน (Nonunidirectional DIF) (Chang; Mazzeo; & Roussos. 1996: 342)

ในการศึกษาครั้งนี้ผู้วิจัยมีความสนใจรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน 2 รูปแบบ คือ ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF) และข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) ซึ่งจะนำมาจัดกระทำข้อมูลเพื่อใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า ภายใต้การจำลองข้อมูลโดยใช้ทฤษฎีการตอบข้อสอบหลายมิติ

3. แนวคิดและหลักการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ

นักวิจัยตั้งแต่อดีตจนถึงปัจจุบัน ได้พยายามyamคิดค้น ดัดแปลง ปรับปรุง และพัฒนาวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ เพื่อให้ผลการตรวจสอบมีความถูกต้องและแม่นยำมากที่สุด เริ่มต้นจากฐานคิดของการตรวจสอบในข้อสอบที่วัดความสามารถมิติเดียว (Unidimensional) จากข้อสอบที่ให้คะแนนสองค่า (Dichotomous) โดยใช้วิธีการตรวจสอบง่ายๆ ไม่ซับซ้อน จนถึงข้อสอบที่ให้คะแนนหลายค่า (Polytomous) และมีวิธีการตรวจสอบที่ซับซ้อนขึ้น ต่อมา มีการพัฒนาโมเดลการวัดความสามารถหลายมิติ (Multidimensional) ประกอบกับความเจริญก้าวหน้าของการพัฒนาโปรแกรมคอมพิวเตอร์ที่ใช้ในการคำนวณ ทำให้นักวิจัยหันมาพัฒนาวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติ ซึ่งถือว่ามีความสมบูรณ์ในการวัดมากกว่า

แนวคิดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ในข้อสอบที่ให้คะแนนสองค่า แบ่งออกเป็นสองกลุ่มใหญ่ๆ คือ กลุ่มวิธีที่ใช้ทฤษฎีการทดสอบแบบมาตรฐานเดิม (Classical test theory methods) และกลุ่มวิธีที่ใช้ทฤษฎีการตอบข้อสอบ (Item response theory methods) กลุ่มวิธีที่ใช้ทฤษฎีการทดสอบแบบมาตรฐานเดิม (CTT methods) จะใช้คะแนนที่ได้จากการสอบเป็นตัวแปรการจับคู่ (Matching variable) ความสามารถระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจ ตามปกติจะใช้คะแนนรวมของแบบทดสอบแทนระดับความสามารถของผู้สอบ วิธีที่สำคัญในกลุ่มนี้ เช่น วิธีการวิเคราะห์ความแปรปรวน (Analysis of variance; ANOVA) (Angoff. 1993; citing Cardall; & Coffman. 1964) วิธีแปลงค่าความยากของข้อสอบ (transformed item difficulty; TID) (Angoff. 1993; citing Angoff. 1972) วิธีไค-กำลังสอง (Chi-square) (Scheuneman. 1979, 1981) วิธีล็อก-ลิเนียร์ (Log-linear) (Mellenbergh. 1982) วิธีแมนเทิล-แฮนส์เซล (Mantel-Haenszel; MH) (Holland; & Thayer. 1988) วิธีการทำให้เป็นมาตรฐาน (Standardization; STD) (Dorans; & Kulick. 1986) วิธีการถดถอยโลจิสติก (Logistic regression; LR) (Swaminathan; & Rogers. 1990) เป็นต้น

สำหรับกลุ่มวิธีที่ใช้ทฤษฎีการตอบข้อสอบ (IRT methods) จะไม่ใช่คะแนนรวมเป็นเกณฑ์การจับคู่ แต่จะใช้ค่าประมาณความสามารถของผู้สอบ (θ) ซึ่งเป็นตัวแปรแฝง (Latent variable) เป็นเกณฑ์การจับคู่ วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในกลุ่มนี้สามารถจำแนกออกเป็นสองกลุ่มใหญ่ๆ คือ กลุ่มแรกใช้การเปรียบเทียบค่าประมาณพารามิเตอร์ของข้อสอบระหว่างกลุ่มผู้สอบ ได้แก่ วิธีเปลี่ยนค่าความยากของไรท์ มีด และดราบา (Hulin; Drasgow; &

Parsons. 1983; citing Wright; Mead; & Draba. 1976) วิธีไค-กำลังสองของลอร์ด (Lord. 1980) และวิธีอัตราส่วนความน่าจะเป็นของธีสเซน สไตน์เบอร์ก และวายนเนอร์ (Thissen; Steinberge; & Wainer. 1993) ส่วนอีกกลุ่มหนึ่งใช้การเปรียบเทียบค่าประมาณฟังก์ชันการตอบข้อสอบระหว่างกลุ่มผู้สอบ โดยการวัดพื้นที่ของฟังก์ชันการตอบข้อสอบ ได้แก่ วิธีการวัดพื้นที่ของรูดเนอร์ (Rudner. 1977) วิธีการวัดพื้นที่ของลินน์และคนอื่นๆ (Linn; et al. 1981) วิธีการวัดพื้นที่ของเซพาร์ด คามิลลี และวิลเลียมส์ (Shepard; Camilli; & Williams. 1984) วิธีการวัดพื้นที่ของราชู (Raju. 1990) วิธีการวัดพื้นที่ของคิมและโคเฮน (Kim; & Cohen. 1991) และวิธีดีเอฟไอที (Raju; van der Linden; & Fleer. 1995) เป็นต้น

ในภาพรวมของกลุ่มวิธีที่ใช้ทฤษฎีการทดสอบแบบมาตรฐานเดิมมีข้อได้เปรียบหลายประการ เช่น สามารถนำไปใช้ในเชิงปฏิบัติ มีกระบวนการตรวจสอบอย่างง่าย ไม่จำเป็นต้องใช้โปรแกรมคอมพิวเตอร์ที่ซับซ้อน สามารถแปลผลเข้าใจง่าย นอกจากนี้ยังใช้ตัวอย่างขนาดเล็ก ทำให้ประหยัดเวลาและเสียค่าใช้จ่ายไม่มาก ส่วนการตรวจสอบในกลุ่มนี้มีข้อจำกัดตรงที่ค่าสถิติของข้อสอบไม่คงที่ เช่น ค่าความยากและค่าอำนาจจำแนกของข้อสอบมีค่าแปรเปลี่ยนไปตามกลุ่มผู้สอบ ดัชนีที่คำนวณได้จึงเป็นความแตกต่างของกลุ่มผู้สอบ (Camilli; & Shepard. 1994: 38) ดังนั้นจึงทำให้การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบเกิดความคลาดเคลื่อนประเภทที่ 1 และประเภทที่ 2 สำหรับปัญหาดังกล่าวสามารถแก้ไขได้ถ้าใช้ทฤษฎีการตอบข้อสอบ เนื่องจากคุณสมบัติของความไม่แปรเปลี่ยน (Invariance) ของค่าพารามิเตอร์ของข้อสอบ (Hambleton; Swaminathan; & Rogers. 1991: 18) ค่าประมาณพารามิเตอร์ของข้อสอบจะไม่ขึ้นกับการแจกแจงความสามารถของกลุ่มผู้สอบ ดังนั้นฟังก์ชันการตอบข้อสอบจะมีค่าเท่ากันเมื่อผู้สอบมีระดับความสามารถเท่ากัน โดยไม่คำนึงว่าผู้สอบมาจากกลุ่มใด ทำให้สามารถเปรียบเทียบผลการตอบข้อสอบที่ระดับความสามารถเดียวกันได้ ซึ่งเป็นหลักการสำคัญของการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ วิธีการตรวจสอบภายใต้ทฤษฎีการตอบข้อสอบในปัจจุบันที่เป็นที่นิยม คือ วิธีไค-กำลังสองของลอร์ด วิธีอัตราส่วนความน่าจะเป็น วิธีการวัดพื้นที่ของราชู และวิธีดีเอฟไอที (Raju; & Ellis. 2002: 181) ทั้งสี่วิธีมีประสิทธิภาพในการตรวจสอบ สามารถประยุกต์ใช้ภายใต้โมเดลแบบ 1, 2 หรือ 3 พารามิเตอร์ ส่วนข้อจำกัดของวิธีดังกล่าว คือ มีรูปแบบพารามิเตอร์ (Parametric form) ซึ่งจำเป็นต้องมีข้อตกลงเกี่ยวกับโมเดลที่ใช้ประมาณค่าพารามิเตอร์ มีขั้นตอนการตรวจสอบที่ยุ่งยาก ต้องใช้โปรแกรมคอมพิวเตอร์ที่ซับซ้อนประมาณค่าพารามิเตอร์ ก่อนที่จะวิเคราะห์การทำหน้าที่เบี่ยงเบนของข้อสอบจะต้องแน่ใจว่าค่าพารามิเตอร์ของข้อสอบระหว่างกลุ่ม

ผู้สอบอยู่บนมาตรฐานเดียวกัน ข้อมูลจะต้องเป็นไปตามข้อตกลงเบื้องต้นของโมเดลที่เข้มงวด อีกทั้งจำเป็นต้องใช้ตัวอย่างขนาดใหญ่ เพื่อให้การประมาณค่าพารามิเตอร์มีความถูกต้อง โดยเฉพาะโมเดลแบบ 2 หรือ 3 พารามิเตอร์ (Clauser; & Mazor. 1998: 34)

จากข้อจำกัดของวิธีที่ใช้ทฤษฎีการตอบข้อสอบ จึงทำให้นักวิจัยหลายคนหันมาพัฒนาวิธีการตรวจสอบโดยใช้ทฤษฎีการทดสอบแบบมาตรฐานเดิม ซึ่งมีข้อได้เปรียบดังที่กล่าวมาข้างต้น ดังนั้นถ้าสามารถปรับปรุงเทคนิคการตรวจสอบ โดยใช้หลักการพื้นฐานของทฤษฎีการตอบข้อสอบ ก็สามารถตรวจสอบได้อย่างมีประสิทธิภาพเช่นเดียวกัน วิธีการตรวจสอบที่นักวิจัยให้การยอมรับอย่างกว้างขวาง ได้แก่ วิธีแมนเทิล-แฮนส์เซล วิธีการทำให้เป็นมาตรฐาน และวิธีการถดถอยโลจิสติก โดยสองวิธีแรกมีรูปแบบนันทพารามิเทริก (Nonparametric form) ซึ่งไม่มีข้อตกลงเกี่ยวกับโมเดลที่ใช้ในการประมาณค่า ส่วนวิธีหลังมีรูปแบบพารามิเทริก นอกจากนี้ ชีลลีและสแตท์ (Shealy; & Stout. 1993) ได้พัฒนาวิธีซิปเทสท์ (SIBTEST) โดยมีพื้นฐานมาจากโมเดลของทฤษฎีการตอบข้อสอบแบบหลายมิติ (Multidimensional IRT) วิธีนี้มีแนวคิดและกระบวนการตรวจสอบที่ต่างไปจากวิธีที่ใช้ทฤษฎีการทดสอบแบบมาตรฐานเดิม และทฤษฎีการตอบข้อสอบ จุดเด่นของวิธีซิปเทสท์คือ มีรูปแบบนันทพารามิเทริกที่ไม่จำเป็นต้องใช้โมเดลประมาณค่า มีขั้นตอนการตรวจสอบที่ไม่ยุ่งยาก และไม่จำเป็นต้องใช้ตัวอย่างขนาดใหญ่

วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่กล่าวมาเป็นวิธีการตรวจสอบภายใต้ข้อสอบที่ให้คะแนนสองค่า ต่อมาได้มีการปรับขยายสำหรับข้อสอบที่ให้คะแนนหลายค่า กลุ่มวิธีที่ใช้ทฤษฎีการทดสอบมาตรฐานเดิม เช่น วิธีแมนเทิล (Mantel) วิธีแมนเทิล-แฮนส์เซลทั่วไป (Generalized Mantel-Haenszel; GMH) (Zwick; Donoghue; & Grima. 1993) วิธีการทำให้เป็นมาตรฐานแบบให้คะแนนหลายค่า (Polytomous STND) (Potenza; & Dorans. 1995) วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก (Logistic discriminant function analysis) (Miller; & Spray. 1993) วิธีการถดถอยโลจิสติกแบบจัดอันดับ (Ordinal logistic regression; OLR) (Zumbo. 1999) สำหรับกลุ่มวิธีที่ใช้ทฤษฎีการตอบข้อสอบ เช่น วิธีการวัดพื้นที่ (Cohen; Kim; & Baker. 1993) วิธีโพลี-ซิปเทสท์ (Poly-SIBTEST) (Chang; Mazzeo; & Roussos. 1996) วิธีดีเอฟไอที (DFIT) (Flowers; Oshima; & Raju. 1999) เป็นต้น นอกจากนี้ยังมีการปรับขยายวิธีการตรวจสอบสำหรับข้อสอบที่วัดความสามารถหลายมิติ เช่น วิธีมัลติซิป (MULTISIB) (Stout; et al. 1997) วิธีดีเอฟไอที (DFIT) (Oshima; Raju; & Flowers. 1997) และวิธีการถดถอยโลจิสติก (Mazor; Hambleton; & Clauser. 1998) เป็นต้น

จากที่กล่าวมาจะเห็นว่า วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ใช้ทฤษฎีการทดสอบแบบมาตรฐานเดิมและวิธีที่ใช้ทฤษฎีการตอบข้อสอบ มีแนวคิดและหลักการคล้ายกัน เพราะต่างก็พัฒนาซึ่งกันและกัน ยกเว้นวิธีชิปเทสต์ของซีลลีและสเตาท์ (Shealy; & Stout. 1993) ที่ใช้กระบวนการค้นหาค้นใหม่ในการวิเคราะห์การทำหน้าที่เบี่ยงเบนของข้อสอบภายใต้ความสามารถแบบหลายมิติ (New multidimensionality-based DIF analysis paradigm) ซึ่งมีแนวคิดและกระบวนการตรวจสอบแตกต่างจากวิธีที่ใช้ทฤษฎีข้างต้น โดยเฉพาะกรอบการศึกษาเกี่ยวกับการจำลองข้อมูล ดังจะกล่าวรายละเอียดในหัวข้อต่อไป ในที่นี้จะกล่าวถึงแนวคิดและหลักการตรวจสอบการทำหน้าที่เบี่ยงเบนในสามประเด็น คือ กรอบแนวคิดทั่วไปในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ กรอบแนวคิดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบโดยใช้โมเดลของทฤษฎีการตอบข้อสอบ และกรอบแนวคิดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบโดยใช้โมเดลแบบหลายมิติของซีลลีและสเตาท์ สำหรับรายละเอียดมีดังนี้

3.1 กรอบแนวคิดทั่วไปในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ

จากกรอบแนวคิดของการทำหน้าที่เบี่ยงเบนของข้อสอบที่กำหนดเป็นนิยามข้างต้นกล่าวว่า **การทำหน้าที่เบี่ยงเบนของข้อสอบ หมายถึง ผู้สอบจากกลุ่มที่แตกต่างกัน และมีการจับคู่ความสามารถตามรูปแบบทดสอบหรือข้อสอบต้องการวัดเท่ากัน มีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องไม่เท่ากัน** นิยามดังกล่าวจะนำมากำหนดเป็นหลักการและแนวทางในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ โดยทั่วไปแล้ววิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ จะอยู่บนฐานคิดของทฤษฎีการทดสอบแบบมาตรฐานเดิม (CTT) และทฤษฎีการตอบข้อสอบ (IRT) แนวคิดและหลักการทั่วไปของสองทฤษฎีจะเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มผู้สอบสองกลุ่ม (อาจปรับขยายมากกว่าสองกลุ่ม) บนเงื่อนไขของการจับคู่ความสามารถของผู้สอบ โดยกำหนดให้ผู้สอบกลุ่มหนึ่งเป็น “กลุ่มสนใจ” (Focal group; F) และอีกกลุ่มหนึ่งเป็น “กลุ่มอ้างอิง” (Reference group; R) ผู้สอบกลุ่มแรกจะเป็นตัวแทนประชากรกลุ่มรอง (Minority group) ซึ่งเป็นกลุ่มเป้าหมายในการศึกษาการทำหน้าที่เบี่ยงเบนของข้อสอบ ส่วนผู้สอบกลุ่มหลังเป็นตัวแทนประชากรกลุ่มหลัก (Majority group) ซึ่งใช้เป็นกลุ่มเปรียบเทียบกับกลุ่มสนใจ (Donoghue; Holland; & Thayer. 1993: 137-138) ถ้าข้อสอบทำหน้าที่เบี่ยงเบนแล้วความน่าจะเป็นในการตอบข้อสอบถูกของผู้สอบสองกลุ่มจะไม่เท่ากัน โดยคาดว่าข้อสอบเข้าข้างผู้สอบกลุ่มอ้างอิงทำให้ได้เปรียบในการตอบข้อสอบ แต่ข้อสอบไม่เข้าข้างผู้สอบกลุ่มสนใจ

ทำให้เสียเปรียบในการตอบข้อสอบ ตามปกติแล้วในการกำหนดขนาดตัวอย่างมักให้กลุ่มอ้างอิงมีขนาดใหญ่กว่าหรือเท่ากับกลุ่มสนใจ สำหรับเกณฑ์ที่ใช้ในการจำแนกผู้สอบเป็นกลุ่มอ้างอิงและกลุ่มสนใจมีหลายลักษณะ เช่น เชื้อชาติ เผ่าพันธุ์ ศาสนา ภาษา เพศ วัฒนธรรม ประสบการณ์ ทักษะทางวิชาชีพสังคม หรือภูมิหลังอื่นๆ เป็นต้น ตัวอย่างเช่น ในการวิเคราะห์การทำหน้าที่เปี่ยงเบนของข้อสอบของศูนย์บริการทดสอบทางการศึกษา (Educational Testing Service; ETS) ของสหรัฐอเมริกาได้ใช้เชื้อชาติเป็นเกณฑ์กำหนดกลุ่มผู้สอบ โดยกำหนดให้ผู้สอบชาวอเมริกันผิวดำ ชาวอเมริกันเชื้อสายสเปน ชาวอเมริกันเชื้อสายเอเชีย และชาวอเมริกันพื้นเมืองเป็นกลุ่มสนใจ ส่วนชาวอเมริกันผิวขาวเป็นกลุ่มอ้างอิง (Dorans; & Holland. 1993: 39)

ข้อสอบที่ใช้ในการวิเคราะห์การทำหน้าที่เปี่ยงเบนของข้อสอบเรียกว่า **“ข้อสอบศึกษา”** (Studied item) ส่วนการเปรียบเทียบกลุ่มผู้สอบที่ความสามารถระดับเดียวกันเรียกว่า **“ตัวแปรการจับคู่”** (Matching variable) ความสามารถดังกล่าวเป็นความสามารถเป้าหมายตามที่ข้อสอบหรือแบบทดสอบต้องการวัด ซึ่งสอดคล้องกับนิยามที่กำหนดไว้ข้างต้น สาเหตุที่ต้องกำหนดเป็นความสามารถเป้าหมายตามที่แบบทดสอบหรือข้อสอบต้องการวัด ทั้งนี้เนื่องจากในการพัฒนาแบบทดสอบที่ใช้วัดทักษะ ความรู้ความสามารถอย่างสมบูรณ์จะต้องจำแนกลักษณะของเนื้อหาเฉพาะที่จะวัด ถ้าความแตกต่างของผลการตอบข้อสอบระหว่างผู้สอบสองกลุ่มที่เกิดขึ้น ภายหลังเงื่อนไขการจับคู่ความสามารถเฉพาะ แสดงว่าผลการตอบข้อสอบถูกจะขึ้นอยู่กับความสามารถอื่นๆ มากกว่า ซึ่งเป็นสาเหตุทำให้ข้อสอบทำหน้าที่เปี่ยงเบน (Clauser; & Mazor. 1998: 31) ตัวอย่างเช่น แบบทดสอบวิชาคณิตศาสตร์ การตอบข้อสอบถูกจะขึ้นกับความสามารถในการคิดคำนวณและความสามารถในการอ่าน ถ้าให้ความสามารถในการคิดคำนวณเป็นความสามารถเป้าหมายที่ต้องการวัดโดยข้อสอบ ซึ่งเป็นความสามารถที่ใช้ในการจับคู่ และถ้าผู้สอบกลุ่มหนึ่งมีความสามารถในการอ่านมากกว่าผู้สอบอีกกลุ่มหนึ่ง ความแตกต่างของผลการตอบข้อสอบระหว่างกลุ่มผู้สอบเป็นหลักฐานแสดงว่า **“ข้อสอบทำหน้าที่เปี่ยงเบน”** แต่ถ้าความแตกต่างในผลการตอบข้อสอบ ซึ่งไม่ได้มีการจับคู่ความสามารถจะเรียกว่า **“ผลกระทบของข้อสอบ”** (Item impact) ซึ่งเกิดขึ้นจากความแตกต่างของการแจกแจงความสามารถของผู้สอบทั้งหมด (Camilli. 1992: 131; Clauser; & Mazor. 1998: 31; Dorans; & Holland. 1993: 37; Zumbo. 1999: 12)

จากแนวคิดดังกล่าวจึงทำให้นักวิจัยให้ความสำคัญเกี่ยวกับมิติ (Dimensional) ของการวัด โดยเชื่อว่าการทำหน้าที่เปี่ยงเบนของข้อสอบมีอิทธิพลของมิติในการวัดเข้ามาเกี่ยวข้อง เช่น ในแบบทดสอบที่วัดความสามารถแบบสองมิติ ประกอบด้วยมิติหลัก (Primary dimension)

และมิติรอง (Secondary dimension) ในมิติหลักเป็นความสามารถบริสุทธิ์ที่ข้อสอบต้องการวัด เรียกว่า “ความสามารถเป้าหมาย” (Target ability) แทนด้วยสัญลักษณ์ θ ส่วนมิติรองอาจมีหลายความสามารถ ถ้าเป็นความสามารถตามที่ข้อสอบต้องการวัดเรียกว่า “ความสามารถเสริม” (Auxiliary ability) แต่ถ้าเป็นความสามารถตามที่ข้อสอบที่ไม่ต้องการวัดเรียกว่า “ความสามารถแทรกซ้อน” (Nuisance ability) แทนด้วยสัญลักษณ์ η (Ackerman. 1992: 69; Nandakumar. 1993: 295-296; Roussos; & Stout. 1996b: 356; Shealy; & Stout. 1993: 163) ข้อสอบที่วัดความสามารถนอกเหนือมิติหลักอย่างน้อยหนึ่งมิติเป็นสาเหตุทำให้ข้อสอบทำหน้าที่เบี่ยงเบน (Roussos; & Stout. 1996b: 356; citing Berk. 1982; Cronbach. 1990; Dorans; & Schmitt. 1989; Jensen. 1980; Lord. 1980; Messick. 1989; Scheuneman. 1982; Shepard. 1987; Wiley. 1990) ดังนั้นความสามารถในมิติรองเป็นสาเหตุที่ทำให้ข้อสอบทำหน้าที่เบี่ยงเบน ถ้าสาเหตุเกิดจากความสามารถเสริมถือว่าเป็นเรื่องที่ดีเพราะได้กำหนดไว้ในเนื้อหาของการทดสอบ แต่ถ้าสาเหตุเกิดจากความสามารถแทรกซ้อนถือว่าไม่ดี เพราะไม่ได้รวมอยู่ในบริบทที่ต้องการวัด จากกรอบแนวคิดทั่วไปดังกล่าว นักวิจัยจึงคิดค้นและพัฒนาโมเดลของตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ โดยสามารถจำแนกรอบแนวคิดออกเป็นสองแนวทางใหญ่ๆ ซึ่งมีรายละเอียดดังนี้

3.2 กรอบแนวคิดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบโดยใช้โมเดลของทฤษฎีการตอบข้อสอบ

วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ทฤษฎีการตอบข้อสอบ ได้พัฒนามาโดยตลอด ในยุคเริ่มต้นจะใช้โมเดลการวัดความสามารถมิติเดียว (Unidimensional) จนถึงปัจจุบันที่นักวิจัยให้ความสำคัญกับการวัดความสามารถหลายมิติ (Multidimensional) กรอบแนวคิดโดยใช้โมเดลของทฤษฎีการตอบข้อสอบเป็นฐานคิดของวิธีการตรวจสอบอื่นๆ ที่ใช้ทฤษฎีการทดสอบแบบมาตรฐานเดิม แม้กระทั่งแนวคิดของซีลลีและสเตาท์ (Shealy; & Stout. 1993) ที่ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบโดยใช้โมเดลแบบหลายมิติ ยังอาศัยกรอบแนวคิดของทฤษฎีการตอบข้อสอบ **หลักการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบโดยใช้ทฤษฎีการตอบข้อสอบจะเปรียบเทียบความน่าจะเป็นในการตอบข้อสอบถูกระหว่างผู้สอบที่มีระดับความสามารถเดียวกัน ซึ่งประเมินโดยใช้ข้อสอบ แต่มาจากกลุ่มผู้สอบที่แตกต่างกัน** กรอบความคิดดังกล่าวเป็นหลักการสำคัญเกี่ยวกับคุณสมบัติของความไม่แปรเปลี่ยน (Property of invariance) ของค่าพารามิเตอร์ของข้อสอบ (Hambleton; Swaminathan;

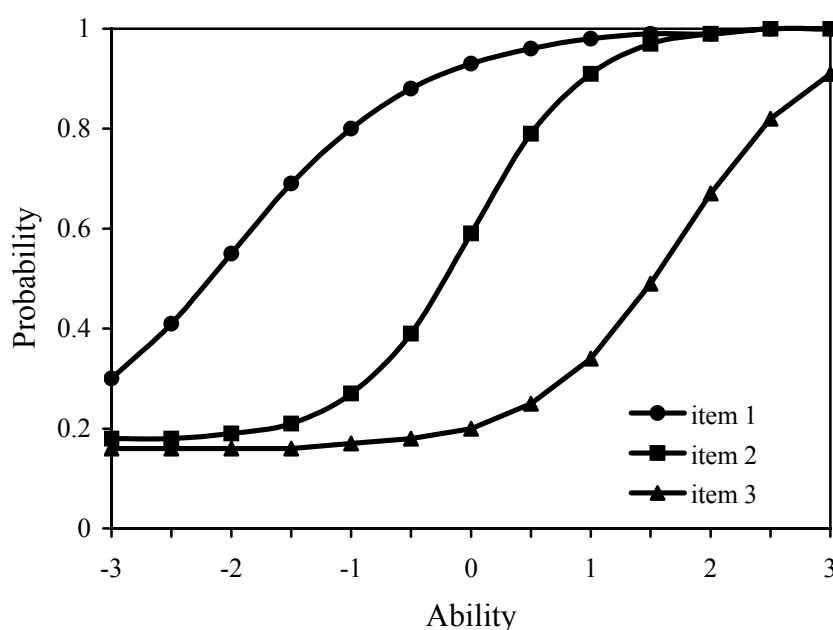
& Rogers. 1991: 18) พารามิเตอร์ของข้อสอบไม่ขึ้นกับการแจกแจงความสามารถของกลุ่มผู้สอบ ดังนั้นฟังก์ชันการตอบข้อสอบ (Item response function; IRF) จะไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบ ถ้าผู้สอบมีความสามารถระดับเดียวกันแล้วความน่าจะเป็นในการตอบข้อสอบถูกต้องจะเท่ากันโดยไม่คำนึงว่าผู้สอบมาจากกลุ่มอ้างอิงหรือกลุ่มสนใจ ในการวิเคราะห์การทำหน้าที่เบี่ยงเบนของข้อสอบภายใต้โมเดลการตอบข้อสอบ สามารถจำแนกออกเป็นสองรูปแบบ ตามลักษณะโมเดลของมิติความสามารถ ดังนี้

3.2.1 การทำหน้าที่เบี่ยงเบนของข้อสอบแบบมิติเดียว (Unidimensional DIF)

การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบโดยใช้ทฤษฎีการตอบข้อสอบในยุคเริ่มต้นได้พัฒนาวิธีการตรวจสอบภายใต้โมเดลแบบมิติเดียว วิธีการตรวจสอบที่พัฒนาขึ้นมีกรอบแนวคิดสองแนวทางใหญ่ๆ คือ กรอบแนวคิดแรกจะตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบโดยการเปรียบเทียบความแตกต่างของค่าพารามิเตอร์ของข้อสอบระหว่างกลุ่มผู้สอบที่มีความสามารถระดับเดียวกัน ส่วนอีกกรอบแนวคิดหนึ่งจะเปรียบเทียบความแตกต่างของพื้นที่โค้งลักษณะข้อสอบ (Item characteristic curves; ICCs) หรือฟังก์ชันการตอบข้อสอบ (Item response functions; IRFs) ระหว่างกลุ่มผู้สอบที่มีความสามารถระดับเดียวกัน ในการเปรียบเทียบความแตกต่างดังกล่าวมักใช้โมเดลโลจิสติกแบบ 1, 2 หรือ 3 พารามิเตอร์ สำหรับโมเดลโลจิสติกแบบ 3 พารามิเตอร์ (3PLM) สามารถคำนวณได้ดังนี้

$$P_i(\theta) = c_i + (1 - c_i) \frac{\exp[Da_i(\theta - b_i)]}{1 + \exp[Da_i(\theta - b_i)]} \quad (1)$$

เมื่อ $P_i(\theta)$ แทนความน่าจะเป็นของผู้สอบที่มีระดับความสามารถ θ จะตอบข้อสอบข้อที่ i ได้ถูกต้อง สำหรับ a_i แทนค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ (Item discrimination) b_i แทนค่าพารามิเตอร์ความยากของข้อสอบ (Item difficulty) และ c_i แทนค่าพารามิเตอร์การเดาของข้อสอบ (Item guessing) ส่วน D เป็นค่าปรับมาตรฐาน ปกติกำหนดให้มีค่าคงที่เท่ากับ 1 หรือ 1.7 สำหรับตัวอย่างฟังก์ชันการตอบข้อสอบของข้อสอบ 3 ข้อ ภายใต้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ มิติเดียว (Unidimensional 3-parameter logistic model; U3PLM) สามารถแสดงดังภาพประกอบ



ภาพประกอบ 11 ฟังก์ชันการตอบข้อสอบ (IRF)

ภายใต้โมเดลโลจิสติกแบบ 3 พารามิเตอร์มิติเดียว (U3PLM)

ที่มา: Hulin; Drasgow; & Parsons. (1983) Item Response Theory Application to Psychological Measurement. p. 49.

กรอบแนวคิดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ โดยการเปรียบเทียบความแตกต่างของค่าพารามิเตอร์ของข้อสอบระหว่างกลุ่ม ในยุคเริ่มต้น ไรท์ มีด และดราบา (Hulin; Drasgow; & Parsons. 1983: 168; citing Wright; Mead; & Draba. 1976) ได้เสนอใช้ “สถิติดราบา” (Draba statistic) ทดสอบความแตกต่างของค่าพารามิเตอร์ความยากของข้อสอบระหว่างกลุ่มผู้สอบ ภายใต้โมเดลของราสซ (Rasch model) โดยใช้สูตรดังนี้

$$z_i = \frac{\hat{b}_{iR} - \hat{b}_{iF}}{\sqrt{(\text{SE of } \hat{b}_{iR})^2 + (\text{SE of } \hat{b}_{iF})^2}} \quad (2)$$

เมื่อ $\hat{b}_{iR}$ และ $\hat{b}_{iF}$ แทนค่าประมาณความยากของข้อสอบข้อที่ i จากผู้สอบกลุ่มอ้างอิง และกลุ่มสนใจ ตามลำดับ SEof $\hat{b}_{iR}$ และ SEof $\hat{b}_{iF}$ แทนค่าความคลาดเคลื่อนมาตรฐานของค่าประมาณความยากของข้อสอบ จากผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจ ตามลำดับ ต่อมาลอร์ด (Lord. 1980: 223) ได้ใช้ “สถิติโฮลเทลลิง” (Holtelling statistic) ทดสอบความแตกต่างของค่าพารามิเตอร์ของข้อสอบ ภายใต้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ ดังนี้

$$\chi^2 = \mathbf{v}'_i \sum_i^{-1} \mathbf{v}_i \quad (3)$$

เมื่อ $\mathbf{v}'_i = [\hat{b}_{iR} - \hat{b}_{iF}, \hat{a}_{iR} - \hat{a}_{iF}]$ และ $\sum_i^{-1}$ แทนเมทริกซ์ค่าประมาณความแปรปรวน-ความแปรปรวนร่วมของ $(\hat{b}_{iR} - \hat{b}_{iF})$ และ $(\hat{a}_{iR} - \hat{a}_{iF})$ วิธีนี้มักนิยมเรียกว่า “วิธีการทดสอบโค-กำลังสองของลอร์ด” ส่วนการตรวจสอบอีกวิธีหนึ่งจะให้ “สถิติอัตราส่วนความน่าจะเป็น” (Likelihood ratio statistic) เปรียบเทียบความแตกต่างของค่าพารามิเตอร์ระหว่างกลุ่ม (Thissen; Steinberge; & Wainer. 1993: 73) โดยทดสอบความเหมาะสมระหว่างโมเดลสองโมเดล คือ โมเดลออกเมนต์ (Augmented; A) ประกอบด้วยพารามิเตอร์ของข้อสอบระหว่างกลุ่มผู้สอบที่มีค่าแตกต่างกัน และโมเดลคอมแพคต์ (Compact model; C) ประกอบด้วยพารามิเตอร์ของข้อสอบระหว่างกลุ่มผู้สอบที่มีค่าเท่ากัน สำหรับสถิติอัตราส่วนความน่าจะเป็นมีลักษณะดังนี้

$$G^2(df) = 2 \log \left[\frac{\text{likelihood}[A]}{\text{likelihood}[C]} \right] \quad (4)$$

ส่วนกรอบแนวคิดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ โดยการวัดพื้นที่ของฟังก์ชันการตอบข้อสอบจะเปรียบเทียบผลการตอบข้อสอบโดยตรง มีนักวิจัยหลายคนได้พัฒนาสูตรการวัดพื้นที่ เช่น วิธีการวัดพื้นที่ของรูดเนอร์ (Rudner. 1977) วิธีการวัดพื้นที่ของลินน์ และคนอื่นๆ (Linn; et al. 1981) วิธีการวัดพื้นที่ของเซพาร์ด คามิลลี และวิลเลียมส์ (Shepard; Camilli; & Williams. 1984) วิธีการวัดพื้นที่ของราจู (Raju. 1990) วิธีการวัดพื้นที่ของคิมและโคเฮน (Kim; & Cohen. 1991) เป็นต้น สำหรับวิธีการวัดพื้นที่แต่ละสูตรดังกล่าว วิธีการวัดพื้นที่ของราจู (Raju. 1990) มีความโดดเด่นมากที่สุด เนื่องจากสามารถคำนวณพื้นที่โดยการหาปริพันธ์แบบต่อเนื่อง (Continuous integration) ซึ่งเป็นวิธีทางคณิตศาสตร์ พร้อมทั้งใช้สถิติทดสอบนัยสำคัญ

ทำให้ผลการตรวจสอบมีความถูกต้องและแม่นยำสูง ราชูได้พัฒนาสูตรการวัดพื้นที่ในช่วงเปิดแบบคิดเครื่องหมาย (Exact signed area; ESA) และแบบไม่คิดเครื่องหมาย (Exact unsigned area; EUA) ภายใต้โมเดลโลจิสติกแบบ 1, 2 และ 3 พารามิเตอร์ สูตรทั่วไปในการคำนวณพื้นที่ของราชูเป็นดังนี้

$$ESA = \int_{-\infty}^{+\infty} [P_R(\theta) - P_F(\theta)] d\theta \quad (5)$$

$$EUA = \int_{-\infty}^{+\infty} |P_R(\theta) - P_F(\theta)| d\theta \quad (6)$$

เมื่อ $P_R(\theta)$ และ $P_F(\theta)$ แทนฟังก์ชันการตอบข้อสอบที่ระดับความสามารถ θ จากกลุ่มอ้างอิงและกลุ่มสนใจ ตามลำดับ ซึ่งคำนวณจากโมเดลในสมการ (1) ส่วนสูตรการวัดพื้นที่ซึ่งนักวิจัยนิยมนำมาใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หรือนำมาใช้คำนวณขนาดและพื้นที่ของข้อสอบทำหน้าที่เบี่ยงเบน เพื่อใช้ในการจำลองข้อมูล คือ สูตรการคำนวณพื้นที่แบบไม่คิดเครื่องหมาย ทั้งนี้เนื่องจากข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Non-uniform DIF) ฟังก์ชันการตอบข้อสอบระหว่างกลุ่มผู้สอบจะตัดกัน ถ้าตัดกันตรงช่วงกึ่งกลางของความสามารถแล้วคำนวณโดยใช้สูตรการวัดพื้นที่แบบคิดเครื่องหมายจะไม่สามารถตรวจพบข้อสอบที่มีรูปแบบดังกล่าวได้ เพราะว่าทิศทางของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นเครื่องหมายบวกและเครื่องหมายลบจะหักล้างกัน (Cancellation) สำหรับสูตรการวัดพื้นที่แบบคิดเครื่องหมายภายใต้โมเดลโลจิสติกแบบ 3 พารามิเตอร์มี 2 กรณี ดังนี้ (Raju. 1990: 201)

$$EUA = (1 - c)|\hat{b}_R - \hat{b}_F|, \text{ เมื่อ } \hat{a}_R = \hat{a}_F \quad (7)$$

$$EUA = (1 - \hat{c}) \left| \frac{2(\hat{a}_R - \hat{a}_F)}{D\hat{a}_R\hat{a}_F} \ln \left\{ 1 + \exp \left[\frac{D\hat{a}_R\hat{a}_F(\hat{b}_R - \hat{b}_F)}{\hat{a}_R - \hat{a}_F} \right] \right\} - (\hat{b}_R - \hat{b}_F) \right|, \text{ เมื่อ } \hat{a}_R \neq \hat{a}_F \quad (8)$$

สูตรการวัดพื้นที่ในกรณีที่ 1 เมื่อ $\hat{a}_R = \hat{a}_F$ ใช้คำนวณพื้นที่ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF) โดยกำหนดให้ความแตกต่างของค่าพารามิเตอร์ความยากของข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มสนใจมีค่าแปรเปลี่ยน ($b_{iR} \neq b_{iF}$) ส่วนค่าพารามิเตอร์

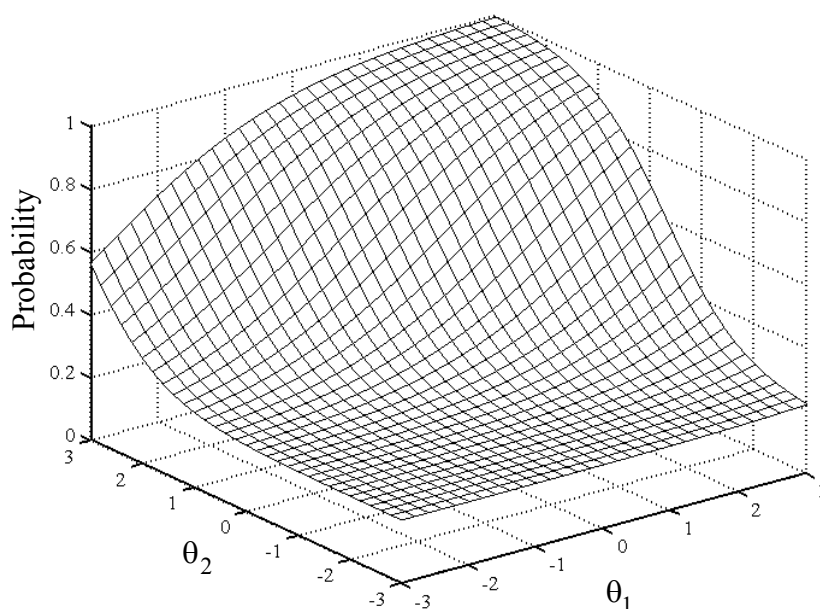
อำนาจจำแนกของข้อสอบระหว่างกลุ่มมีค่าเท่ากัน ($a_{iR} = a_{iF}$) และกรณีที่ 2 เมื่อ $\hat{a}_R \neq \hat{a}_F$ ใช้คำนวณพื้นที่ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) โดยกำหนดให้ความแตกต่างของค่าพารามิเตอร์อำนาจจำแนกของข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มสนใจมีค่าแปรเปลี่ยน ($a_{iR} \neq a_{iF}$) ส่วนค่าพารามิเตอร์ความยากของข้อสอบระหว่างกลุ่มจะมีค่าเท่ากันหรือแตกต่างกันก็ได้ ($b_{iR} = b_{iF}$ หรือ $b_{iR} \neq b_{iF}$) สำหรับค่าพารามิเตอร์การเดาของข้อสอบทั้งสองกรณีมักกำหนดให้มีค่าคงที่ สูตรดังกล่าวสามารถปรับเป็นโมเดลแบบ 1 หรือ 2 พารามิเตอร์ก็ได้ ซึ่งขึ้นอยู่กับเงื่อนไขของพารามิเตอร์ที่กำหนด จากแนวคิดและหลักการดังกล่าวนี้ โคเฮ็น คิม และ เบเกอร์ (Cohen; Kim; & Baker. 1993) ได้ปรับขยายสูตรการวัดพื้นที่ของราชูเพื่อใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนหลายค่า (Polytomous) โดยใช้โมเดลเกรดเรสพอน (Graded response model; GRM)

3.2.2 การทำหน้าที่เบี่ยงเบนของข้อสอบแบบหลายมิติ (Multidimensional DIF)

การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ในข้อสอบที่ให้คะแนนแบบสองค่าหรือหลายค่า มักใช้เงื่อนไขข้อตกลงเกี่ยวกับความเป็นมิติเดียวของแบบทดสอบ ดังเช่น ข้อสอบวัดความสามารถหลายมิติ แต่ให้คะแนนเสมือนวัดความสามารถมิติเดียว จะมีผลทำให้ข้อสอบทำหน้าที่เบี่ยงเบน (Ackerman. 1992: 67) ซึ่งมีผลทำให้การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบเกิดความคลาดเคลื่อนประเภทที่ 1 ปัญหาดังกล่าวเป็นผลมาจากการเลือกใช้โมเดลผิดพลาด (Model misspecification) ไม่เหมาะสมกับข้อมูล ถ้าข้อมูลวัดความสามารถหลายมิติจะต้องเลือกใช้โมเดลการวัดความสามารถหลายมิติ ซึ่งจะทำให้จำแนกระดับความสามารถของผู้สอบได้อย่างถูกต้องและแม่นยำ ในปัจจุบันนักวิจัยได้ให้ความสำคัญในประเด็นดังกล่าวมากยิ่งขึ้น ดังจะเห็นได้จากการศึกษาเกี่ยวกับการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในข้อมูลหลายมิติของนักวิจัยหลายคน (Ackerman. 1992; Camilli. 1992; Mazor; Hambleton; & Clauser. 1998; Oshima; & Miller. 1992; Oshima; Raju; & Flowers. 1997; Roussos; & Stout. 1996b; Shealy; & Stout. 1993; Stout; et al. 1997) ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของที่วัดความสามารถหลายมิติและให้คะแนนสองค่า ภายใต้ทฤษฎีการตอบข้อสอบแบบหลายมิตินักวิจัยส่วนมากนิยมใช้โมเดลโลจิสติกแบบ 2 หรือ 3 พารามิเตอร์หลายมิติที่มีการชดเชยค่าความสามารถ (Compensatory θ) สำหรับโมเดลโลจิสติกแบบ 3 พารามิเตอร์หลายมิติมีลักษณะดังนี้ (Reckase. 1997: 273)

$$P(U_{ij} = 1 | \mathbf{a}_i, d_i, c_i, \boldsymbol{\theta}_j) = c_i + (1 - c_i) \frac{\exp(\mathbf{a}_i' \boldsymbol{\theta}_j + d_i)}{1 + \exp(\mathbf{a}_i' \boldsymbol{\theta}_j + d_i)} \quad (9)$$

เมื่อ $P(U_{ij} = 1 | \mathbf{a}_i, d_i, c_i, \boldsymbol{\theta}_j)$ แทนความน่าจะเป็นของผู้สอบคนที่ j ที่มีเวกเตอร์ของความสามารถ $\boldsymbol{\theta}$ จะตอบข้อสอบข้อที่ i ได้ถูกต้อง $\mathbf{a}_i$ แทนเวกเตอร์ของพารามิเตอร์ ซึ่งเกี่ยวข้องกับอำนาจจำแนกของข้อสอบ d_i แทนสเกลาร์ของพารามิเตอร์ ซึ่งเกี่ยวข้องกับค่าความยากของข้อสอบ c_i แทนพารามิเตอร์การเดาของข้อสอบ $\boldsymbol{\theta}_j$ แทนเวกเตอร์ของพารามิเตอร์ความสามารถของผู้สอบคนที่ j สำหรับความน่าจะเป็นในการตอบข้อสอบดังกล่าว สามารถแสดงด้วยฟังก์ชันผิวหน้าการตอบข้อสอบ (Item response surface; IRS) ดังภาพประกอบ 12



ภาพประกอบ 12 ผิวหน้าการตอบข้อสอบ (IRS)

ภายใต้โมเดลโลจิสติกแบบ 3 พารามิเตอร์หลายมิติ (M3PLM)

ที่มา: Reckase. (1997). A Linear Logistic Multidimensional Model for Dichotomous Item Response Data. p. 274.

พารามิเตอร์ความยากของข้อสอบหลายมิติ (Multidimensional item difficulty; MDIFF) สามารถคำนวณดังนี้

$$\text{MDIFF}_i = -\frac{d_i}{\sqrt{\sum_{k=1}^m a_{ik}^2}} \quad (10)$$

เมื่อ a_{ik} แทนสมาชิกตัวที่ k ของเวกเตอร์ $\mathbf{a}_i$ โดยที่พารามิเตอร์ความยากของข้อสอบหลายมิติแทนระยะห่างจากจุดกำเนิดของสเปซความสามารถไปยังจุดที่ผิวหน้าการตอบข้อสอบ (Item response surface; IRS) ที่มีความชันมากที่สุด จุดดังกล่าวจะให้สารสนเทศสูงสุดเกี่ยวกับผู้สอบซึ่งถูกวัดด้วยข้อสอบข้อดังกล่าว เส้นที่เชื่อมระหว่างจุดที่ผิวหน้าการตอบข้อสอบไปยังจุดกำเนิดทำมุมกับมิติความสามารถที่ k เท่ากับ α_{ik} ซึ่งคำนวณขนาดของมุมดังนี้

$$\alpha_{ik} = \arccos \frac{a_{ik}}{\sqrt{\sum_{k=1}^m a_{ik}^2}} \quad (11)$$

ค่าของ α_{ik} แสดงทิศทางข้อสอบ (Item direction) ซึ่งใช้กำหนดน้ำหนักส่วนประกอบของความสามารถที่วัดโดยข้อสอบ ทิศทางข้อสอบแต่ละข้อจะทำมุมกับแกนความสามารถ θ_1 (แกนบวก) ในโมเดลแบบสองมิติ สมมติให้แกนความสามารถ θ_1 แทนข้อสอบวัดความสามารถ θ_1 และแกนความสามารถ θ_2 แทนข้อสอบวัดความสามารถ θ_2 ถ้าข้อสอบวัดเฉพาะความสามารถ θ_1 แล้ว α_{i1} เท่ากับ 0 องศา ถ้าข้อสอบวัดเฉพาะความสามารถ θ_2 แล้ว α_{i1} เท่ากับ 90 องศา แต่ถ้าข้อสอบวัดความสามารถ θ_1 และ θ_2 เท่ากัน แล้ว α_{i1} เท่ากับ 45 องศา ค่าของ α_{ik} มีค่าตั้งแต่ 0 ถึง 90 องศา ค่าจะมากหรือน้อยขึ้นอยู่กับข้อสอบที่วัดความสามารถทั้งสอง สำหรับพารามิเตอร์อำนาจจำแนกของข้อสอบหลายมิติ (Multidimensional item discrimination; MDISC) สามารถคำนวณดังนี้

$$\text{MDISC}_i = \sqrt{\sum_{k=1}^m a_{ik}^2} \quad (12)$$

พารามิเตอร์อำนาจจำแนกของข้อสอบหลายมิติถูกกำหนดโดยสมาชิกของเวกเตอร์ $\mathbf{a}_i$ สำหรับสมาชิกของเวกเตอร์ $\mathbf{a}_i$ ก็คือ พารามิเตอร์อำนาจจำแนกในโมเดลมิติเดียว พารามิเตอร์อำนาจจำแนกของข้อสอบหลายมิติมีความสัมพันธ์กับพารามิเตอร์ความยากของข้อสอบหลายมิติ ดังนี้

$$MDISC_i = -\frac{d}{MID_i} \quad (13)$$

สำหรับพารามิเตอร์การเดาของข้อสอบภายใต้หลายมิติมีความหมายเหมือนกับโมเดลโลจิสติก 3 พารามิเตอร์แบบมิติเดียว ซึ่งเป็นค่าที่บ่งชี้ถึงความน่าจะเป็นในการตอบข้อสอบถูกต้องสำหรับผู้สอบที่มีระดับความสามารถต่ำมาก ในทุกมิติของโมเดล

โอชิมะ ราจู และเฟลาเวอร์ส (Oshima; Raju; & Flowers. 1997) ได้ใช้กรอบแนวคิดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ตามทฤษฎีการตอบข้อสอบแบบมิติเดียวมาประยุกต์ใช้ในการตรวจสอบแบบหลายมิติ โดยใช้โมเดลโลจิสติก 2 พารามิเตอร์แบบหลายมิติที่มีการชดเชยค่าความสามารถจำลองข้อสอบแบบสองมิติ แล้วตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบโดยใช้วิธีดีเอฟไอที (DFIT) (Raju; van der Linden; & Fleer. 1995) โดยการคำนวณดัชนี 3 ประเภท คือ ดัชนีการทำหน้าที่เบี่ยงเบนของข้อสอบที่มีการชดเชย (Compensatory DIF; CDIF) ดัชนีการทำหน้าที่เบี่ยงเบนของข้อสอบไม่มีการชดเชย (Noncompensatory DIF; NCDIF) และดัชนีการทำหน้าที่เบี่ยงเบนของแบบทดสอบ (DTF) ดังนี้

$$CDIF = Cov(\hat{d}_i, \hat{D}) + \hat{\mu}_{\hat{d}_i} \hat{\mu}_{\hat{D}} \quad (14)$$

$$NCDIF = \hat{\sigma}_{\hat{d}_i}^2 + \hat{\mu}_{\hat{d}_i}^2 \quad (15)$$

$$DTF = \hat{\sigma}_{\hat{D}}^2 + \hat{\mu}_{\hat{D}}^2 \quad (16)$$

โดยที่

$$\hat{D} = \sum_{i=1}^k \hat{d}_i \quad (17)$$

$$\hat{d}_i = \hat{P}_{iF} - \hat{P}_{iR} \quad (18)$$

เมื่อ $\hat{P}_{iF}$ และ $\hat{P}_{iR}$ แทนความน่าจะเป็นในการตอบข้อสอบข้อที่ i ถูกต้องสำหรับผู้สอบกลุ่มสนใจและกลุ่มอ้างอิง ตามลำดับ ซึ่งประมาณค่ามาจากสมการ (9) [$\hat{d}_i$ ในสมการ (14) - (18) ไม่ใช่ d_i จากสมการ (9)] หลักการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถ

หลายมิติ จะเปรียบเทียบฟังก์ชันการตอบข้อสอบระหว่างกลุ่ม ทำนองเดียวกับการตรวจสอบที่วัดความสามารถมิติเดียว สมมติให้แบบทดสอบมีข้อสอบ k ข้อ และมีพารามิเตอร์ของข้อสอบสองชุด สำหรับกลุ่มอ้างอิงและกลุ่มสนใจ โดยที่พารามิเตอร์ของข้อสอบสองชุดดังกล่าวมีมาตร่วมกัน เมื่อเวกเตอร์ของความสามารถเท่ากัน แล้ว $\hat{P}_{iF}$ และ $\hat{P}_{iR}$ มีค่าแตกต่างกัน นั่นแสดงว่า “ข้อสอบทำหน้าที่ยั่งยืน” ส่วนการศึกษาเกี่ยวกับรูปแบบของข้อสอบทำหน้าที่ยั่งยืนภายใต้การจำลองข้อมูลแบบหลายมิติก็มีแนวคิดเช่นเดียวกัน กล่าวคือ ในเงื่อนไขของการจำลองข้อสอบทำหน้าที่ยั่งยืนที่เป็นรูปแบบเดียวกันจะกำหนดให้สเกลาร์ของพารามิเตอร์ d_i ระหว่างกลุ่มผู้สอบมีค่าแตกต่างกัน ในขณะที่เวกเตอร์ของพารามิเตอร์ $\mathbf{a}_i$ มีค่าเท่ากัน ส่วนในเงื่อนไขของข้อสอบทำหน้าที่ยั่งยืนที่ไม่เป็นรูปแบบเดียวกันจะกำหนดให้เวกเตอร์ของพารามิเตอร์ $\mathbf{a}_i$ ระหว่างกลุ่มผู้สอบมีค่าแตกต่างกัน ในขณะที่สเกลาร์ของพารามิเตอร์ d_i มีค่าเท่ากันหรือแตกต่างกันก็ได้

จากที่กล่าวมาจะเห็นได้ว่า *กรอบแนวคิดในการตรวจสอบการทำหน้าที่ยั่งยืนของข้อสอบโดยใช้โมเดลของทฤษฎีการตอบข้อสอบทั้งแบบมิติเดียวและหลายมิติจะอาศัยหลักการความไม่แปรเปลี่ยนของค่าพารามิเตอร์ของข้อสอบ* ซึ่งเป็นจุดแข็งตามข้อตกลงเบื้องต้นของทฤษฎีดังกล่าว

3.3 กรอบแนวคิดในการตรวจสอบการทำหน้าที่ยั่งยืนของข้อสอบโดยใช้โมเดลแบบหลายมิติของซีสลีและสเตาท์

ในการตรวจสอบการทำหน้าที่ยั่งยืนของข้อสอบตามแนวคิดของซีสลีและสเตาท์ (Shealy; & Stout. 1993) จะแบ่งกลุ่มผู้สอบออกเป็นสองกลุ่ม คือ กลุ่มอ้างอิงและกลุ่มสนใจ แล้วสุ่มตัวอย่างในแต่ละกลุ่มเพื่อดำเนินการสอบ โดยใช้แบบทดสอบจำนวน N ข้อ ในแบบทดสอบดังกล่าวประกอบด้วยข้อสอบส่วนหนึ่งที่ถูกตั้งข้อสงสัยว่าทำหน้าที่ยั่งยืน โดยเข้าข้างผู้สอบกลุ่มอ้างอิง ข้อสอบดังกล่าวเป็นเป้าหมายเพื่อศึกษาการทำหน้าที่ยั่งยืนของข้อสอบ ส่วนผลการตอบข้อสอบแทนด้วย $\mathbf{U} = (U_1, U_2, \dots, U_N)$ เมื่อ U_i แทนผลการตอบข้อสอบข้อที่ i ซึ่งให้คะแนนสองค่า ถ้าตอบถูกให้ 1 คะแนน และตอบผิดให้ 0 คะแนน สมมติให้สเปซความสามารถแฝงที่สมบูรณ์ (Complete latent space Θ) ของความสามารถผู้สอบมีหลายมิติ (Multidimensional) แทนด้วย $\{\theta = (\theta, \eta)\}$ เมื่อ θ แทนความสามารถเป้าหมายที่ต้องการวัด (Target ability) หรือเรียกว่าความสามารถหลัก (Primary ability) และเวกเตอร์ η แทนความสามารถแทรกซ้อนที่ไม่ต้องการวัด (Nuisance ability) หรือเรียกว่าความสามารถรอง (Secondary ability) ซึ่งมีอิทธิพล

ต่อการตอบข้อสอบ ส่วนการแจกแจงของ (θ, η) เลือกมาจากกลุ่มผู้สอบรวมทั้งหมดโดยการสุ่มแทนด้วย (Θ, η) ความสามารถทั้งสองมีข้อตกลงว่า ข้อสอบส่วนมากจะวัดความสามารถเป้าหมาย แต่อาจมีข้อสอบบางข้อที่วัดทั้งความสามารถเป้าหมายและความสามารถแทรกซ้อน ซึ่งมีความสามารถแทรกซ้อนเพียงความสามารถเดียว หรือมากกว่าหนึ่งความสามารถก็ได้ ถ้าความสามารถแทรกซ้อนมีอิทธิพลในการตอบข้อสอบมากกว่าความสามารถเป้าหมาย จะมีผลทำให้ข้อสอบทำหน้าที่เบี่ยงเบน (Stout; et al. 1997: 197)

ตัวอย่างเช่น แบบทดสอบวัดเหตุผลทางคณิตศาสตร์ ข้อสอบบางข้ออาจถามความรู้สำหรับผู้ชายเป็นพิเศษ เช่น ความรู้เกี่ยวกับกีฬาฟุตบอล ในขณะที่ข้อสอบบางข้ออาจถามความรู้เกี่ยวกับผู้หญิงโดยเฉพาะ เช่น ความรู้เกี่ยวกับงานเย็บปักถักร้อย จากสถานการณ์ดังกล่าวทักษะความรู้เกี่ยวกับเหตุผลทางคณิตศาสตร์เป็นความสามารถเป้าหมายที่ต้องการวัด (θ) ส่วนความรู้ทางด้านกีฬาฟุตบอล (η_1) และงานเย็บปักถักร้อย (η_2) เป็นความสามารถแทรกซ้อนที่ไม่ต้องการวัด

สำหรับข้อสอบที่ขึ้นอยู่กับการวัดความสามารถเป้าหมาย θ เพียงความสามารถเดียวแทนด้วยฟังก์ชันการตอบข้อสอบ $P_i(\theta)$ ข้อสอบดังกล่าวถือว่าเป็นข้อสอบที่ทำหน้าที่ไม่เบี่ยงเบน (No-DIF) ภายใต้ทฤษฎีการตอบข้อสอบ โดยใช้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ สามารถคำนวณฟังก์ชันดังกล่าวได้ดังนี้

$$P_i(\theta) = c_i + \frac{(1 - c_i)}{1 + \exp[-1.7a_{i\theta}(\theta - b_{i\theta})]}, \quad i = 1, \dots, N \quad (19)$$

ส่วนข้อสอบที่ขึ้นอยู่กับการวัดความสามารถเป้าหมาย θ และความสามารถแทรกซ้อน η แทนด้วย $P_i(\theta, \eta)$ ถือว่าเป็นข้อสอบทำหน้าที่เบี่ยงเบน (DIF) ถ้าเป็นข้อสอบที่วัดความสามารถสองมิติ ภายใต้ทฤษฎีการตอบข้อสอบโดยใช้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ สามารถคำนวณฟังก์ชันดังกล่าวได้ดังนี้

$$P_i(\theta, \eta) = c_i + \frac{(1 - c_i)}{1 + \exp[-1.7a_{i\theta}(\theta - b_{i\theta}) + a_{i\eta}(\eta - b_{i\eta})]}, \quad i = n + 1, \dots, N \quad (20)$$

ฟังก์ชันการตอบข้อสอบในสมการ (19) และ (20) เป็นโมเดลของทฤษฎีการตอบข้อสอบ ซึ่งมีข้อตกลงสามประการ (Li; & Stout. 1996: 650) คือ **ประการแรก** ค่าของฟังก์ชันการตอบข้อสอบไม่แปรเปลี่ยนตามกลุ่มผู้สอบ (Group invariance) ซึ่งค่า $P_i(\theta, \eta)$ จะเหมือนกันสำหรับ

ผู้สอบที่มีความสามารถ (θ, η) โดยไม่คำนึงว่าเป็นสมาชิกผู้สอบกลุ่มใด **ประการที่สอง** การเพิ่มขึ้นทางเดียว (Monotonicity) ของฟังก์ชันการตอบข้อสอบ สำหรับข้อสอบที่ขึ้นอยู่ความสามารถ θ และ η ค่า $P_i(\theta, \eta)$ จะเพิ่มขึ้นตามความสามารถ θ และ η ส่วนข้อสอบที่ขึ้นอยู่ความสามารถ θ เพียงความสามารถเดียว ค่า $P_i(\theta)$ จะเพิ่มขึ้นตามความสามารถ θ และ**ประการที่สาม** ความเป็นอิสระในการตอบข้อสอบ (Local independence) ของ U กำหนดในรูปความน่าจะเป็นดังนี้

$$P[U = (u_1, u_2, \dots, u_N) | \Theta = \theta, \eta = \eta] = \prod_{i=1}^N P_i(\theta, \eta)^{u_i} (1 - P_i(\theta, \eta))^{1-u_i} \quad (21)$$

เมื่อ U แทนผลการตอบข้อสอบที่ให้คะแนนสองค่า ถ้าตอบถูกให้ 1 คะแนน และตอบผิดให้ 0 คะแนน ซิลลีและสแตท์ (Shealy; & Stout. 1993: 164) ได้ใช้โมเดลทางคณิตศาสตร์ภายใต้บริบทของทฤษฎีการตอบข้อสอบที่วัดความสามารถหลายมิติ (Multidimensional IRT) อธิบายการทำหน้าที่เบี่ยงเบนของข้อสอบ โดยใช้ขอบของฟังก์ชันการตอบข้อสอบ (Marginal IRFs) ของกลุ่มผู้สอบ ดังนี้

$$P(X_i = 1 | \Theta = \theta) = \int P_i(\theta, \eta) f(\eta | \theta) d\eta \quad (22)$$

เมื่อ $P_i(\theta, \eta)$ แทนฟังก์ชันการตอบข้อสอบแบบสองมิติ (Two-dimensional) และ $f(\eta | \theta)$ แทนการแจกแจงความสามารถแทรกซ้อน η แบบมีเงื่อนไขของกลุ่มผู้สอบ เมื่อกำหนดค่าความสามารถเป้าหมาย θ คงที่ สำหรับค่าของ $P_i(\theta, \eta)$ จะมีค่าแปรเปลี่ยนตามความสามารถแทรกซ้อน η เมื่อกำหนดค่าความสามารถเป้าหมาย θ คงที่ สำหรับค่าของ $P(X_i = 1 | \Theta = \theta)$ ได้จากค่าเฉลี่ยของ $P_i(\theta, \eta)$ ในความสามารถแทรกซ้อน η ทั้งหมด เมื่อกำหนดค่าความสามารถเป้าหมาย θ คงที่ นั่นคือ $P(X_i = 1 | \Theta = \theta)$ เป็นฟังก์ชันการตอบข้อสอบที่วัดความสามารถแบบมิติเดียวเมื่อไม่ได้รับผลกระทบจากความสามารถแทรกซ้อน η ถ้า $f(\eta | \theta)$ ระหว่างกลุ่มมีค่าเท่ากันแล้วการทำหน้าที่เบี่ยงเบนของข้อสอบจะไม่เกิดขึ้น เพราะว่าผู้สอบที่มีความสามารถเป้าหมาย θ ระดับเดียวกันมีความน่าจะเป็นในการตอบข้อสอบถูกเท่ากัน (Ackerman. 1992: 76) ซิลลีและสแตท์ (Shealy; & Stout. 1993: 166-167) ได้คำนวณดัชนีการทำหน้าที่เบี่ยงเบนของข้อสอบ (β_{uni}) ดังนี้

$$\beta_{uni} = \int_{-\infty}^{+\infty} [T_R(\theta) - T_F(\theta)] f_F(\theta) d\theta \quad (23)$$

เมื่อ $T_R(\theta)$ และ $T_F(\theta)$ แทนขอบของฟังก์ชันการตอบข้อสอบจากกลุ่มอ้างอิงและกลุ่มสนใจ ตามลำดับ ส่วน $f_F(\theta)$ เป็นขอบของความหนาแน่นของความสามารถเป้าหมาย θ จากกลุ่มสนใจ ดัชนี β_{uni} คำนวณจากระยะห่างในแนวตั้งฉากกับแกนนอนระหว่างขอบฟังก์ชันการตอบข้อสอบจากข้อสอบที่ศึกษา ซึ่งวัดความสามารถเป้าหมาย θ ดัชนี β_{uni} ที่คำนวณได้ดังกล่าวเป็นค่าเฉลี่ยของความแตกต่างของความน่าจะเป็นในการตอบข้อสอบถูกต้องระหว่างกลุ่มผู้สอบสองกลุ่ม ซึ่งผลของการตอบข้อสอบถูกเกิดขึ้นจากประสบการณ์ที่ได้จากการเรียนรู้ หรือข้อสอบที่ศึกษาได้ ขจัดผลกระทบของข้อสอบออกแล้ว **ดังนั้นในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบตามแนวคิดของซีลลีและสเตาท์จะเปรียบเทียบความแตกต่างระหว่างค่าคาดหวังของการแจกแจงความสามารถแบบมีเงื่อนไขของ η เมื่อกำหนด θ คงที่ สำหรับกลุ่มอ้างอิงและกลุ่มสนใจ ถ้าความแตกต่างดังกล่าวไม่เท่ากับ 0 ทุกระดับความสามารถเป้าหมาย θ แสดงว่าข้อสอบทำหน้าที่เบี่ยงเบน**

สมมติให้การถดถอยของความสามารถแทรกซ้อน η บนความสามารถเป้าหมาย θ เป็นเส้นตรง ความแตกต่างระหว่างค่าคาดหวังของการแจกแจงแบบมีเงื่อนไขของ η เมื่อกำหนด θ คงที่ สามารถเขียนในรูปสัญลักษณ์ได้ดังนี้ (Ackerman, 1992: 78)

$$\begin{aligned} C_\beta &= E[\eta_R|\theta] - E[\eta_F|\theta] \\ &= (\mu_{\eta_R} - \mu_{\eta_F}) + \left(\rho_R \frac{\sigma_{\eta_R}}{\sigma_{\theta R}} \right) (\theta - \mu_{\theta R}) - \left(\rho_F \frac{\sigma_{\eta_F}}{\sigma_{\theta F}} \right) (\theta - \mu_{\theta F}) \end{aligned} \quad (24)$$

ปริมาณของค่า C_β คือ ขนาดของข้อสอบทำหน้าที่เบี่ยงเบน (Magnitude of DIF) โดยที่การแจกแจงร่วมกันของความสามารถทั้งสอง (θ, η) สมมติให้มีการแจกแจงแบบปกติสองตัวแปร (Bivariate normal) เมื่อ μ_θ และ μ_η แทนค่าเฉลี่ย σ_θ และ σ_η แทนส่วนเบี่ยงเบนมาตรฐาน และ ρ แทนค่าสหสัมพันธ์ ความแตกต่างระหว่างค่าคาดหวังเมื่อมีค่าไม่เท่ากับศูนย์จะทำให้ข้อสอบทำหน้าที่เบี่ยงเบน สาเหตุที่ทำให้ค่าคาดหวังดังกล่าวมีค่าไม่เท่ากับศูนย์ แอ็คเคอร์แมน (Ackerman; 1992: 78) สรุปไว้ 4 กรณี คือ **กรณีที่ 1** ค่าเฉลี่ยของความสามารถเป้าหมาย θ ระหว่างกลุ่มผู้สอบแตกต่างกัน **กรณีที่ 2** ค่าเฉลี่ยของความสามารถแทรกซ้อน η ระหว่างกลุ่มผู้สอบแตกต่างกัน **กรณีที่ 3** อัตราส่วนของความแปรปรวนความสามารถเป้าหมาย θ และความสามารถแทรกซ้อน η ระหว่างกลุ่มผู้สอบแตกต่างกัน และ**กรณีที่ 4** ค่าสหสัมพันธ์ของความสามารถเป้าหมาย θ และความสามารถแทรกซ้อน η ระหว่างกลุ่มผู้สอบแตกต่างกัน

รูสโซและสเตาท์ (Roussos; & Stout. 1996b: 362-365) ได้อธิบายสาเหตุที่ทำให้ข้อสอบทำหน้าที่เบี่ยงเบนมีสามกรณี คือ **กรณีที่ 1** ข้อสอบมีความไวต่อความสามารถแทรกซ้อน η และผู้สอบมีความแตกต่างกันเกี่ยวกับการแจกแจงแบบมีเงื่อนไขของความสามารถแทรกซ้อน $\eta[f_R(\eta|\theta) \neq f_F(\eta|\theta)]$ **กรณีที่ 2** ค่าเฉลี่ยของความสามารถแทรกซ้อน η มีค่าใกล้เคียงกัน ($\mu_{\eta_R} \approx \mu_{\eta_F}$) แต่ค่าเฉลี่ยของความสามารถเป้าหมาย θ แตกต่างกัน ($\mu_{\theta_R} > \mu_{\theta_F}$) ในขณะที่ความสามารถทั้งสองมีความสัมพันธ์กัน และ**กรณีที่ 3** สหสัมพันธ์ของความสามารถระหว่างกลุ่มมีค่าเท่ากัน ($\rho_R = \rho_F$) และส่วนเบี่ยงเบนมาตรฐานของความสามารถเป้าหมาย θ สำหรับกลุ่มสนใจน้อยกว่ากลุ่มอ้างอิง ($\sigma_{\theta_F} < \sigma_{\theta_R}$) ในขณะที่ส่วนเบี่ยงเบนมาตรฐานของความสามารถแทรกซ้อน η สำหรับกลุ่มสนใจมากกว่ากลุ่มอ้างอิง ($\sigma_{\eta_F} > \sigma_{\eta_R}$) ลักษณะดังกล่าวทำให้ข้อสอบทำหน้าที่เบี่ยงเบนแบบตัดกัน (Crossing DIF) นอกจากนี้ รูสโซและสเตาท์ (Roussos; & Stout. 1996b: 363-364) ได้อธิบายสาเหตุของข้อสอบที่ทำหน้าที่ไม่เบี่ยงเบนเนื่องจากข้อสอบไม่มีความไวต่อความสามารถแทรกซ้อน η ดังนั้นฟังก์ชันการตอบข้อสอบจะขึ้นอยู่กับความสามารถเป้าหมาย θ โดยเฉพาะ หรืออีกเหตุผลหนึ่ง คือ ข้อสอบมีความไวต่อความสามารถแทรกซ้อน η แต่มีความแตกต่างของการแจกแจงของความสามารถแทรกซ้อน η เมื่อเปรียบเทียบความสามารถเป้าหมาย θ

ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบตามแนวคิดของซีลลีและสเตาท์ (Shealy; & Stout. 1993) ภายใต้การจำลองข้อมูลโดยใช้ทฤษฎีการตอบข้อสอบจะใช้เงื่อนไขของการแจกแจงความสามารถเป้าหมาย θ และความสามารถแทรกซ้อน η ซึ่งสามารถพิจารณาได้จากสมการ (20) สำหรับการจำลองข้อมูลจะต้องพิจารณาโครงสร้างของความสามารถทั้งสอง ซึ่งประกอบด้วยค่าเฉลี่ยของความสามารถ ค่าความแปรปรวน-ความแปรปรวนร่วมของความสามารถ หรือค่าสหสัมพันธ์ระหว่างความสามารถ กล่าวคือ การจำลองข้อสอบให้ทำหน้าที่ไม่เบี่ยงเบนจะใช้โมเดลโลจิสติกแบบมิติเดียว ส่วนการจำลองข้อสอบให้ทำหน้าที่เบี่ยงเบนจะใช้โมเดลโลจิสติกแบบหลายมิติ โดยกำหนดขนาดและรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนของข้อสอบในสมการ (24) ถ้าจำลองข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF) กำหนดให้ค่าสหสัมพันธ์ระหว่างความสามารถเท่ากัน ($\rho_{\theta\eta_R} = \rho_{\theta\eta_F}$) แต่ถ้าจำลองข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) จะกำหนดให้ค่าสหสัมพันธ์ระหว่างความสามารถแตกต่างกัน ($\rho_{\theta\eta_R} \neq \rho_{\theta\eta_F}$) ดังนั้นการจำลองรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนของข้อสอบจะต้องควบคุมความแตกต่างของค่าสหสัมพันธ์ระหว่างความสามารถ (Li; & Stout. 1966: 657)

ในการศึกษาครั้งนี้ ผู้วิจัยใช้ข้อมูลจำลอง ภายใต้ทฤษฎีการตอบข้อสอบแบบหลายมิติ โดยใช้โมเดลพาเชียลเครดิตทั่วไปแบบหลายมิติ (Multidimensional generalized partial credit model) ที่ปรับขยายมาจากโมเดลพาเชียลเครดิตทั่วไปแบบมิติเดียวของมูรากิ (Muraki, 1992: 161)

4. วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ

ในการศึกษาครั้งนี้ ผู้วิจัยนำวิธีโพลี-ชิปเทสต์ (Poly-SIBTEST) (Chang; Mazzeo; & Roussos. 1996) และวิธีการถดถอยโลจิสติกแบบจัดอันดับ (Ordinal logistic regression; OLR) (Zumbo. 1999) ไปตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า ดังนั้นผู้วิจัยจะนำเสนอรายละเอียดของวิธีการตรวจสอบทั้งสองดังกล่าว โดยนำเสนอตั้งแต่วิธีที่พัฒนาเริ่มต้น คือ วิธีชิปเทสต์ (Shealy; & Stout. 1993) และวิธีการถดถอยโลจิสติก (Swaminathan; & Rogers. 1990) ซึ่งใช้ตรวจสอบในข้อมูลที่ให้คะแนนสองค่า รวมวิธีการตรวจสอบทั้งหมดที่จะนำเสนอมี 4 วิธี ซึ่งมีรายละเอียดดังนี้

4.1 วิธีชิปเทสต์ (SIBTEST)

ซีลลีและสตาท์ (Shealy; & Stout. 1993) ได้เสนอวิธี “Simultaneous item bias test” หรือเรียกสั้นๆ ว่า “วิธีชิปเทสต์” (SIBTEST) เพื่อใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบและแบบทดสอบ (Differential item/test functioning; DIF/DTF) ในข้อสอบที่ให้คะแนนสองค่า หลักการตรวจสอบด้วยวิธีชิปเทสต์มีแนวคิดคล้ายกับวิธีการทำให้เป็นมาตรฐาน (Standardization; STD) (Dorans; & Kulick. 1986) โดยพัฒนามาจากโมเดลความลำเอียงของแบบทดสอบภายใต้ทฤษฎีการตอบข้อสอบแบบหลายมิติ (Multidimensional IRT) มีรูปแบบนันทพารามิเทริก (Non-parametric form) ซึ่งไม่ต้องใช้ฟังก์ชันการตอบข้อสอบประมาณค่าความสามารถ วิธีชิปเทสต์เป็นวิธีที่ใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่มีทิศทางเดียวกัน (Unidirectional DIF) โดยเฉพาะ ดังนั้นจึงมีข้อจำกัดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ไม่มีทิศทางเดียวกัน (Nonunidirectional DIF) (Li; & Stout. 1996: 650; Shealy; & Stout. 1993: 162) ส่วนข้อได้เปรียบของวิธีชิปเทสต์ก็คือ สามารถคำนวณได้ง่าย เสียค่าใช้จ่ายไม่มาก และไม่จำเป็นต้องใช้ตัวอย่างขนาดใหญ่ สามารถใช้สถิติทดสอบทดสอบนัยสำคัญ เพื่อตัดสินการทำหน้าที่เบี่ยงเบนของข้อสอบครั้งละหนึ่งข้อ หรือมากกว่าหนึ่งข้อพร้อมกัน (Simultaneous) ผลการวิเคราะห์ทำให้

ทราบขนาดและทิศทางของการทำหน้าที่เบี่ยงเบนของข้อสอบ (Nandakumar. 1993: 295) นักวิจัยหลายคนได้นำวิธีชิปเทสต์มาศึกษาและปรับขยายเพื่อใช้ตรวจสอบในประเด็นต่างๆ เช่น แนนดาคุมาร์ (Nandakumar. 1993) ได้ศึกษาการทำหน้าที่เบี่ยงเบนของข้อสอบในสองกรณี คือ การขยายข้อสอบทำหน้าที่เบี่ยงเบน (DIF amplification) และการหักล้างข้อสอบทำหน้าที่เบี่ยงเบน (DIF cancellation) ลิและสเตาท์ (Li; & Stout. 1996) ได้พัฒนาวิธีครอสซิงชิปเทสต์ (Crossing SIBTEST) เพื่อใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบแบบตัดกัน รุสโซและสเตาท์ (Roussos; & Stout. 1996a) ได้ศึกษาอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีชิปเทสต์ในกลุ่มตัวอย่างขนาดเล็ก ชาง มาซซีโอ และรุสโซ (Chang; Mazzeo; & Roussos. 1996) ได้พัฒนาวิธีโพลี-ชิปเทสต์ (Poly-SIBTEST) เพื่อใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนหลายค่า สเตาท์และคนอื่นๆ (Stout; et al. 1997) ได้พัฒนาวิธีมัลติชิป (MULTISIB) เพื่อใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบแบบสองมิติ เป็นต้น

แนวคิดและหลักการ

ชีลลีและสเตาท์ (Shealy; & Stout. 1993: 163-164) ได้อธิบายการทำหน้าที่เบี่ยงเบนของข้อสอบ โดยใช้ขอบของฟังก์ชันการตอบข้อสอบ (Marginal IRFs) ของความสามารถเป้าหมายที่ต้องการวัด θ สำหรับกลุ่ม g (กลุ่มอ้างอิง หรือกลุ่มสนใจ) ดังนี้

$$M_{ig}(\theta) = E[P_i(\boldsymbol{\Theta}, \boldsymbol{\eta}) | \boldsymbol{\Theta} = \theta, G = g] \quad (1)$$

ถ้า $\boldsymbol{\eta} | \boldsymbol{\Theta} = \theta, G = g$ มีความหนาแน่นแบบมีเงื่อนไขของ $\boldsymbol{\eta}$ เมื่อกำหนดความสามารถ θ ของกลุ่ม g มีค่าคงที่ ซึ่งแทนด้วย $f_g(\boldsymbol{\eta} | \theta)$ ดังนั้นการกำหนดนิยามในสมการ (1) สามารถคำนวณได้ดังนี้

$$M_{ig}(\theta) = \int_{-\infty}^{+\infty} P_i(\theta, \boldsymbol{\eta}) f_g(\boldsymbol{\eta} | \theta) d\boldsymbol{\eta} \quad (2)$$

จากสมการ (2) ถ้าการแจกแจงแบบมีเงื่อนไขของความสามารถ $\boldsymbol{\eta}$ มีค่าเท่ากันสำหรับผู้สอบสองกลุ่ม แล้วข้อสอบจะทำหน้าที่ไม่เบี่ยงเบน (No-DIF) เพราะว่ามีความสามารถ θ เท่ากัน จะทำให้ความน่าจะเป็นในการตอบข้อสอบถูกเท่ากัน (Ackerman. 1992: 76) จากแนวคิดดังกล่าวสามารถนิยามการทำหน้าที่เบี่ยงเบนของข้อสอบ (DIF) โดยใช้ Marginal IRFs ได้ว่า “ถ้าฟังก์ชัน

การตอบข้อสอบของความสามารถเป้าหมายสำหรับกลุ่มอ้างอิงมีค่ามากกว่าฟังก์ชันการตอบข้อสอบสำหรับกลุ่มสนใจ แล้วข้อสอบจะทำหน้าที่เบี่ยงเบน โดยข้อสอบเข้าข้างกลุ่มอ้างอิง” ซึ่งแสดงในรูปสัญลักษณ์ทางคณิตศาสตร์ ดังนี้

$$M_{iR}(\theta) > M_{iF}(\theta) \quad (3)$$

ฟังก์ชันการตอบข้อสอบของแบบทดสอบที่ต้องการศึกษา สำหรับผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจ กำหนดดังนี้

$$M_{SR}(\theta) = \sum_{i=n+1}^N M_{iR}(\theta) \quad (4)$$

$$M_{SF}(\theta) = \sum_{i=n+1}^N M_{iF}(\theta) \quad (5)$$

เมื่อ $M_{SR}(\theta)$ และ $M_{SF}(\theta)$ แทนผลรวม Marginal IRFs ของข้อสอบที่ต้องการศึกษา ณ ระดับความสามารถ θ จากผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจ ตามลำดับ สำหรับปริมาณของการทำหน้าที่เบี่ยงเบนของข้อสอบ (Amount of DIF) สามารถคำนวณจากความแตกต่างระหว่าง $M_{SR}(\theta)$ และ $M_{SF}(\theta)$ ดังนี้

$$B(\theta) = M_{SR}(\theta) - M_{SF}(\theta) \quad (6)$$

ขนาดของความแตกต่างดังกล่าว แสดงถึงปริมาณของการทำหน้าที่เบี่ยงเบนของข้อสอบจากแบบทดสอบชุดย่อยที่ต้องการศึกษา ณ ระดับความสามารถ θ ซึ่งเข้าข้างกลุ่มอ้างอิง ซีลลีและสเตาท์ (Shealy; & Stout. 1993: 167) ได้คำนวณค่าเฉลี่ยของปริมาณการทำหน้าที่เบี่ยงเบนของข้อสอบที่มีทิศทางเดียวกัน (Unidirectional DIF) ดังนี้

$$\beta_{uni} = \int_{-\infty}^{+\infty} B(\theta) f_F(\theta) d\theta \quad (7)$$

เมื่อ β_{uni} แทนดัชนีการทำหน้าที่เบี่ยงเบนของข้อสอบที่มีทิศทางเดียวกัน และ $f_F(\theta)$ แทนฟังก์ชันความหนาแน่นความน่าจะเป็นของการแจกแจงความสามารถเป้าหมาย จากผู้สอบกลุ่มรวมทั้งหมด

กระบวนการตรวจสอบ

ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่มีทิศทางเดียวกันตามแนวคิดของ ซีลลีและสเตย์ (Shealy; & Stout. 1993) จะเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มอ้างอิง และกลุ่มสนใจ โดยใช้แบบทดสอบจำนวน N ข้อ แล้วแบ่งแบบทดสอบดังกล่าวออกเป็นสองชุด คือ แบบทดสอบชุดย่อยที่มีความถูกต้อง (Valid subtests) และแบบทดสอบชุดย่อยที่ใช้ในการศึกษา (Studied subtests) กล่าวคือ แบบทดสอบชุดแรกใช้ในการจับคู่เปรียบเทียบ (Matching subtests) ประกอบด้วยข้อสอบข้อที่ 1 ถึง n ซึ่งเป็นข้อสอบที่ไม่สงสัยว่าทำหน้าที่เบี่ยงเบน โดยวัดความสามารถเป้าหมาย θ เพียงความสามารถเดียว ส่วนแบบทดสอบชุดหลังเป็นส่วนที่เหลือจากชุดแรก ประกอบด้วยข้อสอบข้อที่ $n + 1$ ถึง N ข้อสอบดังกล่าวสงสัยว่าทำหน้าที่เบี่ยงเบน โดยวัดทั้งความสามารถเป้าหมาย θ และความสามารถแทรกซ้อน η

การทดสอบสมมติฐาน

ในการทดสอบสมมติฐานของการทำหน้าที่เบี่ยงเบนของข้อสอบที่มีทิศทางเดียวกัน เมื่อข้อสอบเข้าข้างกลุ่มอ้างอิง ค่าดัชนี β_{uni} มากำหนดสมมติฐานศูนย์ (H_0) และสมมติฐานทางเลือก (H_1) ดังนี้

$$\begin{aligned} H_0 : \beta_{uni} &= 0 \\ H_1 : \beta_{uni} &> 0 \end{aligned} \quad (8)$$

การทดสอบสมมติฐานการทำหน้าที่เบี่ยงเบนของข้อสอบที่มีทิศทางเดียวกันจะประมาณค่าดัชนี β_{uni} โดยคำนวณจากคะแนนของแบบทดสอบชุดย่อยที่ใช้ในการจับคู่เปรียบเทียบ และแบบทดสอบชุดย่อยที่ต้องการศึกษา ดังนี้

$$X = \sum_{i=1}^n U_i \quad (9)$$

$$Y = \sum_{i=n+1}^N U_i \quad (10)$$

เมื่อ

X แทน คะแนนรวมจากแบบทดสอบชุดย่อยที่ใช้ในการจับคู่เปรียบเทียบ

Y แทน คะแนนรวมจากแบบทดสอบชุดย่อยที่ต้องการศึกษา

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

คำนวณคะแนนเฉลี่ยจากผลการตอบข้อสอบในแบบทดสอบชุดย่อยที่ต้องการศึกษาของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจที่มีความสามารถระดับเดียวกัน แล้วนำคะแนนเฉลี่ยดังกล่าวมาจับคู่เปรียบเทียบ โดยพิจารณาได้จากคะแนนรวมที่เท่ากันของแบบทดสอบชุดย่อยที่ใช้ในการจับคู่เปรียบเทียบ ($X = k$) ดังนี้

$$\bar{Y}_{Rk} - \bar{Y}_{Fk} \quad ; \quad k = 0, 1, 2, \dots, n \quad (11)$$

เมื่อ $\bar{Y}_{Rk}$ และ $\bar{Y}_{Fk}$ แทนค่าเฉลี่ยของคะแนน Y จากการตอบแบบทดสอบชุดย่อยที่ใช้ในการจับคู่เปรียบเทียบ แล้วได้คะแนนรวม $X = k$ สำหรับผู้สอบในกลุ่มอ้างอิงและกลุ่มสนใจ ตามลำดับ คะแนนเฉลี่ยที่ใช้ในการเปรียบเทียบดังกล่าวอาจทำให้การตรวจสอบผิดพลาดจากความ-เป็นจริง กล่าวคือ เมื่อเกิดความแตกต่างของการแจกแจงค่าความสามารถ (Ability distribution) ของกลุ่มอ้างอิงและกลุ่มสนใจจะมีผลทำให้ $\bar{Y}_{Rk} - \bar{Y}_{Fk}$ มีค่าแตกต่างจาก 0 อย่างเป็นระบบ ทำให้ตรวจพบว่าข้อสอบทำหน้าที่เบี่ยงเบน ซึ่งความเป็นจริงแล้วข้อสอบทำหน้าที่ไม่เบี่ยงเบน ดังนั้นความแตกต่างของการแจกแจงค่าความสามารถของกลุ่มอ้างอิงและกลุ่มสนใจที่เกิดขึ้นสามารถปรับแก้ค่าการถดถอย (Regression correction) เพื่อกำจัดค่าที่สูงเกินปกติ (Inflate) (Shealy; & Stout. 1993: 169) สำหรับค่าเฉลี่ย $\bar{Y}_{Rk}$ และ $\bar{Y}_{Fk}$ ที่ปรับแก้แล้วแทนด้วย $\bar{Y}_{Rk}^*$ และ $\bar{Y}_{Fk}^*$ ตามลำดับ

ค่า $\bar{Y}_{Rk}^* - \bar{Y}_{Fk}^*$ เป็นความแตกต่างของผลการตอบข้อสอบในแบบทดสอบชุดย่อยที่ศึกษาระหว่างกลุ่มผู้สอบที่มีความสามารถระดับเดียวกัน ถ้า $\bar{Y}_{Rk}^* - \bar{Y}_{Fk}^* = 0$ ทุกคะแนน k แสดงว่าข้อสอบที่สงสัยในแบบทดสอบชุดย่อยทำหน้าที่ไม่เบี่ยงเบน (No-DIF) และถ้า $\bar{Y}_{Rk}^* - \bar{Y}_{Fk}^* > 0$ ทุกคะแนน k แสดงว่าข้อสอบที่สงสัยในแบบทดสอบชุดย่อยทำหน้าที่เบี่ยงเบนที่มีทิศทางเดียวกัน (Unidirectional DIF) โดยข้อสอบเข้าข้างกลุ่มอ้างอิง ถ้า $\bar{Y}_{Rk}^* - \bar{Y}_{Fk}^* < 0$ ทุกคะแนน k แสดงว่าข้อสอบที่สงสัยในแบบทดสอบชุดย่อยทำหน้าที่เบี่ยงเบนที่มีทิศทางเดียวกัน โดยข้อสอบเข้าข้างกลุ่มสนใจ สำหรับค่าความแตกต่างของผลการตอบข้อสอบดังกล่าว สามารถนำมาประมาณค่าในรูป $\hat{\beta}_{uni}$ ดังนี้

$$\hat{\beta}_{uni} = \sum_{k=0}^n \hat{P}_k (\bar{Y}_{Rk}^* - \bar{Y}_{Fk}^*) \quad (12)$$

เมื่อ $\hat{P}_k$ แทนสัดส่วนของผู้สอบทั้งหมด (กลุ่มอ้างอิงและกลุ่มสนใจ) ซึ่งตอบแบบทดสอบชุดย่อยที่ใช้จับคู่เปรียบเทียบ แล้วได้คะแนนรวม $X = k$ สัดส่วนของผู้สอบดังกล่าวสามารถเขียนในรูปสัญลักษณ์ ดังนี้

$$\hat{P}_k = \frac{(J_{Rk} + J_{Fk})}{\sum_{k=0}^n (J_{Rk} + J_{Fk})} \quad (13)$$

เมื่อ J_{Rk} และ J_{Fk} แทนจำนวนผู้สอบซึ่งตอบแบบทดสอบชุดย่อยที่ใช้จับคู่เปรียบเทียบ แล้วได้คะแนนรวม $X = k$ สำหรับกลุ่มอ้างอิงและกลุ่มสนใจ ตามลำดับ จากนั้นจึงนำค่าประมาณ $\hat{\beta}_{uni}$ ที่คำนวณในสมการ (12) มาทดสอบสมมติฐานศูนย์ (No-DIF) โดยใช้สถิติ B_{uni} ดังนี้

$$B_{uni} = \frac{\hat{\beta}_{uni}}{\hat{\sigma}(\hat{\beta}_{uni})} \quad (14)$$

$\hat{\sigma}(\hat{\beta}_{uni})$ เป็นค่าประมาณความคลาดเคลื่อนมาตรฐานของ B_{uni} คำนวณจาก

$$\hat{\sigma}(\hat{\beta}_{uni}) = \sqrt{\sum_{k=0}^n \hat{P}_k^2 \left[\frac{1}{J_{Rk}} \hat{\sigma}^2(Y|k, R) + \frac{1}{J_{Fk}} \hat{\sigma}^2(Y|k, F) \right]} \quad (15)$$

เมื่อ $\hat{\sigma}^2(Y|k, R)$ และ $\hat{\sigma}^2(Y|k, F)$ แทนค่าประมาณความแปรปรวนของคะแนนจากแบบทดสอบชุดย่อยที่ต้องการศึกษาในกลุ่มอ้างอิงและกลุ่มสนใจ ตามลำดับ สำหรับสถิติ B_{uni} มีการแจกแจงใกล้เคียงการแจกแจงแบบปกติมาตรฐาน $[N(0, 1)]$ เมื่อข้อสอบทำหน้าที่ไม่เบี่ยงเบน และถ้าผลการทดสอบพบว่า $B_{uni} > Z_\alpha$ อย่างมีนัยสำคัญที่ระดับ α โดยที่ $P[N(0, 1) > Z_\alpha] = \alpha$ แสดงว่า ปฏิเสธ H_0 นั่นคือ ข้อสอบทำหน้าที่เบี่ยงเบนที่มีทิศทางเดียวกัน (Unidirectional DIF) เมื่อ $B_{uni} > 0$ แสดงว่าข้อสอบเข้าข้างกลุ่มอ้างอิง และเมื่อ $B_{uni} < 0$ แสดงว่าข้อสอบเข้าข้างกลุ่มสนใจ รุสโซและสตาท์ (Roussos; & Stout. 1996a: 220) ได้เสนอเกณฑ์เพื่อใช้จำแนกขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบไว้ดังนี้

- ❑ DIF ระดับ A ขนาดเล็ก: ปฏิเสธสมมติฐานศูนย์ และ $|\hat{B}_{uni}| < 0.059$
- ❑ DIF ระดับ B ขนาดปานกลาง: ปฏิเสธสมมติฐานศูนย์ และ $0.059 \leq |\hat{B}_{uni}| < 0.088$
- ❑ DIF ระดับ C ขนาดใหญ่: ปฏิเสธสมมติฐานศูนย์ และ $|\hat{B}_{uni}| \geq 0.088$

การปรับแก้ค่าการถดถอย

การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบจะมีปัญหาในเชิงสถิติ ซึ่งเกิดจากความแตกต่างของการแจกแจงความสามารถเป้าหมายระหว่างกลุ่มอ้างอิงและกลุ่มสนใจ กล่าวคือ ถ้าการแจกแจงความสามารถเป้าหมายของผู้สอบกลุ่มอ้างอิงสูงกว่ากลุ่มสนใจจะเกิดเงื่อนไขที่เรียกว่า “ผลกระทบ” (Impact) ซึ่งทำให้สถิติ B_{uni} มีค่าเฟ้อ (Inflate) หรือมีค่าสูงเกินปกติ แล้วจะส่งผลให้การตรวจสอบเกิดความคลาดเคลื่อนประเภทที่ 1 (Type I error) เพราะในความเป็นจริงข้อสอบทำหน้าที่ไม่เบี่ยงเบนแต่ตรวจพบว่าข้อสอบทำหน้าที่เบี่ยงเบน ดังนั้นจึงมีความจำเป็นที่จะต้องปรับแก้ความแตกต่างของการแจกแจงความสามารถเป้าหมายโดยใช้การปรับแก้การถดถอย (Regression correction) เพื่อกำจัดอิทธิพลค่าเฟ้อของผลกระทบ ซึ่งจะแปลงค่า $\bar{Y}_{Rk}, \bar{Y}_{Fk}$ เป็น $\bar{Y}_{Rk}^*, \bar{Y}_{Fk}^*$ ที่ละคู่ ดังนั้น $\bar{Y}_{gk}^*$ เป็นตัวประมาณค่าอิทธิพลของค่าเฉลี่ยคะแนนจริงจากแบบทดสอบชุดย่อยที่ใช้ในการจับคู่เปรียบเทียบในกลุ่มย่อย k กลุ่ม ซึ่งแบ่งมาจากแต่ละกลุ่มของกลุ่มอ้างอิงและกลุ่มสนใจ โดยมีรายละเอียดดังนี้ (Shealy; & Stout. 1993: 190-193)

กำหนดให้ $X = k$ แทนคะแนนรวมซึ่งเป็นคะแนนสังเกตจากแบบทดสอบชุดย่อยที่ใช้ในการจับคู่เปรียบเทียบ $V_g(k)$ แทนค่าการถดถอยของคะแนนจริงจากแบบทดสอบชุดย่อยที่ใช้ในการจับคู่เปรียบเทียบซึ่งได้คะแนนเท่ากับ k ในการประมาณค่า $V_g(k)$ สมมติว่าเป็นเส้นตรงดังนี้

$$V_g(k) = \alpha + \beta k \quad (16)$$

ในการประมาณค่า $V_g(k)$ จะใช้ข้อตกลงของโมเดลคะแนนจริง ดังนี้

$$X = T + e \quad (17)$$

$$E(e) = 0 \text{ และ } cov(T, e) = 0 \quad (18)$$

เมื่อ X, T และ e แทนคะแนนสังเกต คะแนนจริง และคะแนนความคลาดเคลื่อนจากแบบทดสอบชุดย่อยที่ใช้ในการจับคู่เปรียบเทียบ ตามลำดับ แล้วใช้ทฤษฎีการถดถอยมาตรฐานประมาณค่า $V_g(k)$ ดังนี้

$$V_g(k) = ET + \left(\frac{\rho_{XT}\sigma_T}{\sigma_X} \right) (k - EX) \quad (19)$$

เมื่อ EX และ ET แทนค่าคาดหวังของ X และ T ในกลุ่ม g (R หรือ F) สำหรับ σ_X และ σ_T แทนส่วนเบี่ยงเบนมาตรฐานของ X และ T ในกลุ่ม g (R หรือ F) ส่วน ρ_{XT} แทนความสัมพันธ์ระหว่าง X และ T ในกลุ่ม g (R หรือ F) จากโมเดลคะแนนจริงสามารถคำนวณค่าความเชื่อมั่น ดังนี้

$$\frac{\rho_{XT}\sigma_T}{\sigma_X} = 1 - \frac{\sigma_e^2}{\sigma_X^2} \quad (20)$$

เมื่อ σ_e^2 และ σ_X^2 แทนความแปรปรวนของความคลาดเคลื่อนและความแปรปรวนของคะแนนสังเกตในกลุ่ม g (R หรือ F) ตามลำดับ จากสมการ (17) และ (18) แสดงว่า $ET = EX$ ดังนั้นสมการ (19) และ (20) สามารถเขียนใหม่ได้ดังนี้

$$V_g(k) = EX + \left(1 - \frac{\sigma_e^2}{\sigma_X^2}\right)(k - EX) \quad (21)$$

ค่าประมาณของ $V_g(k)$ สามารถคำนวณดังนี้

$$\hat{V}_g(k) = \bar{X}_g + \left[1 - \frac{\hat{\sigma}^2(e|g)}{\hat{\sigma}^2(X|g)}\right](k - \bar{X}_g) \quad (22)$$

โดยที่

$$\hat{\sigma}^2(X|g) = \frac{1}{(J_g - 1)} \sum_{j=1}^{J_g} (X_{gj} - \bar{X}_g)^2 \quad (23)$$

และ

$$\hat{\sigma}^2(e|g) = \sum_{i=1}^n \bar{U}_{ig}(1 - \bar{U}_{ig}) \quad (24)$$

เมื่อ

X_{gj} แทน คะแนนจากแบบทดสอบชุดย่อยที่ใช้ในการจับคู่เปรียบเทียบของผู้สอบ
คนที่ j ในกลุ่ม g

$\bar{X}_g$ แทน คะแนนเฉลี่ยจากแบบทดสอบชุดย่อยที่ใช้ในการจับคู่เปรียบเทียบของผู้สอบ
คนที่ j ในกลุ่ม g

$\bar{U}_{ig}$ แทน สัดส่วนการตอบข้อสอบถูกของผู้สอบกลุ่ม g ซึ่งตอบข้อสอบข้อที่ i
จากแบบทดสอบชุดย่อยที่ใช้ในการจับคู่เปรียบเทียบ

ค่า $\hat{V}_g(k)$ จากสมการ (22) เป็นค่าประมาณคะแนนจริงของความสามารถเป้าหมาย
สำหรับผู้สอบกลุ่ม g (R หรือ F) ที่ได้คะแนน $X = k$ จากแบบทดสอบชุดย่อยที่ใช้ในการจับคู่
เปรียบเทียบ ซึ่งสอดคล้องกับคะแนนจากแบบทดสอบชุดย่อยที่ต้องการศึกษา โดยใช้วิธีอนุกรม
ของเทเลอร์ (Taylor) ปรับแก้คะแนนดังนี้

$$\bar{Y}_{gk}^* = \bar{Y}_{gk} + \hat{M}_{gk}[\hat{V}(k) - \hat{V}_g(k)] \quad (25)$$

เมื่อ $\bar{Y}_{gk}$ แทนค่าเฉลี่ยของคะแนนสังเกตจากแบบทดสอบชุดย่อยที่ต้องการศึกษาของ
ผู้สอบกลุ่ม g (R หรือ F) ซึ่งได้คะแนน $X = k$ จากแบบทดสอบชุดย่อยที่ใช้ในการจับคู่เปรียบเทียบ
สำหรับ $\hat{V}(k)$ และ $\hat{M}_{gk}$ คำนวณดังนี้

$$\hat{V}(k) = \frac{\hat{V}_R(k) + \hat{V}_F(k)}{2} \quad (26)$$

$$\hat{M}_{gk} = \frac{\bar{Y}_{g,k+1} - \bar{Y}_{g,k-1}}{\hat{V}_g(k+1) - \hat{V}_g(k-1)} \quad (27)$$

เมื่อ $\bar{Y}_{g,k+1}$ และ $\bar{Y}_{g,k-1}$ แทนค่าเฉลี่ยของคะแนนสังเกตจากแบบทดสอบชุดย่อยที่
ต้องการศึกษาของผู้สอบกลุ่มย่อย จำนวน $k+1$ กลุ่มและ $k-1$ กลุ่ม ตามลำดับ $\hat{V}_g(k+1)$ และ
 $\hat{V}_g(k-1)$ แทนค่าประมาณของคะแนนจริงจากแบบทดสอบชุดย่อยที่ใช้ในการจับคู่เปรียบเทียบของ

ผู้สอบกลุ่มย่อย จำนวน $k + 1$ กลุ่มและ $k - 1$ กลุ่ม ตามลำดับ สำหรับค่า $\bar{Y}_{gk}^*$ ในสมการ (25) เป็นค่าประมาณของคะแนนจริงจากแบบทดสอบชุดย่อยที่ต้องการศึกษาของผู้สอบกลุ่มย่อย k ในกลุ่ม g (F หรือ R) ซึ่งสมมติว่าเท่ากับค่าประมาณคะแนนจริงจากแบบทดสอบชุดย่อยที่ใช้ในการจับคู่เปรียบเทียบ $\hat{V}(k)$ ของผู้สอบทั้งสองกลุ่ม

4.2 วิธีโพลี-ชิปเทสท์ (Poly-SIBTEST)

วิธีชิปเทสท์ (Shealy; & Stout. 1993) เป็นวิธีที่พัฒนาสำหรับการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนสองค่า ต่อมา ชาง มาซซีโอ และรูสโซ (Chang; Mazzeo; & Roussos. 1996) ได้ปรับขยายวิธีชิปเทสท์ (Modified SIBTEST) เพื่อให้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนหลายค่า วิธีที่พัฒนาใหม่นี้เรียกว่า “วิธีโพลี-ชิปเทสท์” (Poly-SIBTEST)

แนวคิดและหลักการ

นियามการทดสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนสองค่าของตัวแปรแฝง (Latent variable) กำหนดไว้ว่า เมื่อกำหนดให้ $E_R[Y | \theta]$ และ $E_F[Y | \theta]$ แทนการถดถอยของ Y บนตัวแปรแฝง θ ของกลุ่มอ้างอิงและกลุ่มสนใจ ตามลำดับ ข้อสอบจะไม่แสดงทำหน้าที่เบี่ยงเบน (Not exhibit DIF) ถ้าทุกค่าของความสามารถ θ ที่มี

$$E_R[Y | \theta] = E_F[Y | \theta] \quad (1)$$

และเมื่อกำหนด $E_R[Y | X]$ และ $E_F[Y | X]$ แทนการถดถอยของ Y บนคะแนนสังเกต (Observed score) ของกลุ่มอ้างอิงและกลุ่มสนใจ ตามลำดับ ข้อสอบจะไม่แสดงทำหน้าที่เบี่ยงเบน ถ้าทุกค่าของคะแนน X ที่มี

$$E_R[Y | X] = E_F[Y | X] \quad (2)$$

จากนियามการทดสอบสมมติฐานศูนย์ของการทำหน้าที่เบี่ยงเบนของข้อสอบ (Null-DIF) โดยใช้ตัวแปรแฝงและคะแนนสังเกตดังที่กล่าวมา มีงานวิจัยหลายเรื่องที่อภิปรายผลขัดแย้งกัน กล่าวคือ นักวิจัยกลุ่มหนึ่งเชื่อว่าการทดสอบโดยใช้วิธีตัวแปรแฝงและคะแนนสังเกตเท่าเทียมกัน

(Equivalent) แต่มีนักวิจัยอีกกลุ่มหนึ่งเชื่อว่าการทดสอบดังกล่าวไม่เท่าเทียมกัน อย่างไรก็ตาม จากการศึกษาของรูสโซและสเตาท์ (Chang; Mazzeo; & Roussos. 1966: 335; citing Roussos; & Stout. 1996a) ได้เสนอแนะว่า ในการนำนิยามทั้งสองกรณีมาใช้ควรพิจารณาแยกส่วนกัน

การนำนิยามของตัวแปรแฝงในสมการ (1) มาใช้ทดสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนหลายค่า สามารถกำหนดได้ว่า “ข้อสอบไม่แสดงทำหน้าที่เบี่ยงเบน (Not exhibit DIF) ถ้าการถดถอยของคะแนนข้อสอบบนตัวแปรแฝงเหมือนกัน สำหรับกลุ่มผู้สอบภายใต้ศึกษา” เมื่อกำหนดให้ Y แทนคะแนนของข้อสอบที่ต้องการศึกษา โดยเป็นคะแนนในรายการแบบจัดอันดับ (Ordered categories) ซึ่งมี $m + 1$ รายการ ($Y = k, 0 \leq k \leq m$) และ $P_{k,g}(\theta)$ แทนฟังก์ชันการตอบรายการของข้อสอบ (Item-category response function; ICRF) ที่ระดับคะแนน k ในกลุ่ม g ดังนั้นการถดถอยของคะแนนข้อสอบบนความสามารถ θ จะกำหนดในรูปผลรวมของฟังก์ชันการตอบรายการของข้อสอบแบบถ่วงน้ำหนัก ดังนี้

$$E_g[Y | \theta] = \sum_{k=1}^m kP_{k,g}(\theta) \quad (3)$$

สำหรับกรณีการให้คะแนนสองค่า การถดถอยดังกล่าวจะแสดงในรูปฟังก์ชันการตอบข้อสอบ (Item response function; IRF) จากสมการ (3) ถ้า $m = 1$ ข้อสอบที่ให้คะแนนหลายค่าจะเปลี่ยนเป็นแบบสองค่า ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนสองค่า จะเปรียบเทียบความแตกต่างของ IRFs ระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจทุกระดับความสามารถ θ โดยอาศัยนิยามจากสมการ (1) ส่วนการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนหลายค่าจะเปรียบเทียบความแตกต่างของ ICRFs เช่นเดียวกัน โดยสามารถเขียนในรูปสัญลักษณ์ได้ดังนี้

$$P_{kR}(\theta) = P_{kF}(\theta) , k = 1, 2, \dots, m \quad (4)$$

เมื่อ $P_{kR}(\theta)$ และ $P_{kF}(\theta)$ แทน ICRFs ที่ระดับคะแนน k ของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจ ตามลำดับ โดยที่ข้อสอบ 1 ข้อ จะมี ICRFs จำนวน m ฟังก์ชัน ดังนั้นนิยามจากสมการ (1) ของตัวแปรแฝงของข้อสอบที่ให้คะแนนสองค่า สามารถนำมาใช้กับข้อสอบที่ให้คะแนนหลายค่าแบบเรียงลำดับ

สำหรับวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนหลายค่า ซึ่งใช้คะแนนสังเกตในสมการ (2) ดังเช่น วิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel; MH) (Holland; & Thayer. 1988) และวิธีการทำให้เป็นมาตรฐาน (Standardization; STD) (Dorans; & Kulick. 1986) สามารถกำหนดเป็นนิยามได้ว่า “ข้อสอบไม่แสดงทำหน้าที่เบี่ยงเบน (Not exhibit DIF) ถ้าการถดถอยของคะแนนข้อสอบที่ให้คะแนนหลายค่า บนคะแนนสังเกตของแบบทดสอบที่ใช้ในการจับคู่เหมือนกัน สำหรับกลุ่มผู้สอบภายใต้ศึกษา”

กระบวนการตรวจสอบ

ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนหลายค่า โดยใช้วิธีโพลี-ชิปเทสท์จะใช้กรอบแนวคิดเช่นเดียวกับวิธีชิปเทสท์ต้นฉบับเดิม (Original SIBTEST) ที่ตรวจสอบในกรณีให้คะแนนสองค่า โดยการประยุกต์ใช้สถิติเพื่อทดสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนในรายการแบบจัดอันดับ (Ordered categories) ดังนี้

กำหนดให้

Y แทนคะแนนของข้อสอบที่ต้องการศึกษา ซึ่งให้คะแนนตามรายการแบบจัดอันดับจำนวน $m + 1$ รายการ ($Y = 0, 1, 2, \dots, m$)

$X_1, X_2, \dots, X_n$ แทนคะแนนของข้อสอบที่ใช้ในการจับคู่เปรียบเทียบ จำนวน n ข้อ

$m_1, m_2, \dots, m_n$ แทนคะแนนมากที่สุดที่เป็นไปได้ของ $X_1, X_2, \dots, X_n$ ตามลำดับ

X แทนคะแนนที่ใช้ในการจับคู่เปรียบเทียบ คำนวณจากสูตร

$$X = \sum_{j=1}^n X_j, \quad (5)$$

เมื่อ $X = 0, 1, 2, \dots, n_H$ โดยที่ n_H เป็นคะแนนของข้อสอบที่ใช้ในการจับคู่เปรียบเทียบซึ่งมีค่ามากที่สุดที่เป็นไปได้ สามารถคำนวณได้ดังนี้

$$n_H = \sum_{j=1}^n m_j \quad (6)$$

$\bar{Y}_{gk}$ แทนคะแนนเฉลี่ยของข้อสอบที่ต้องการศึกษาสำหรับผู้สอบทั้งหมดในกลุ่ม g (R หรือ F) ซึ่งได้คะแนน $X = k$ ถึงแม้ว่าวิธีชิปเทสท์พัฒนามาจากโมเดลของทฤษฎีการตอบข้อสอบ (IRT) แต่ในที่นี้จะนำทฤษฎีการทดสอบแบบมาตรฐานเดิม (CTT) มาอธิบายเพื่อทำความเข้าใจ

กระบวนการตรวจสอบของวิธีดังกล่าว ตามข้อตกลงของทฤษฎีการทดสอบมาตรฐานเดิมเกี่ยวกับคะแนน X กำหนดว่า $X = T + E$ เมื่อ T แทนคะแนนจริงของแบบทดสอบที่ใช้ในการจับคู่ ดังนั้นตัวแปร $E = X - T$ แทนความคลาดเคลื่อนในการวัด และสมมติว่ามีค่าเฉลี่ยเป็นศูนย์ทั้งสองกลุ่มเมื่อให้ $f_g(t)$ แทนความหนาแน่นของคะแนนจริงของแบบทดสอบที่ใช้ในการจับคู่ในกลุ่ม g (R หรือ F) ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบจะพิจารณาการถดถอยของคะแนนข้อสอบที่ต้องการศึกษาบนคะแนนจริงของแบบทดสอบที่ใช้ในการจับคู่ในกลุ่ม g (R หรือ F) ซึ่งการถดถอยดังกล่าวเขียนในรูปสัญลักษณ์ได้ว่า $E_g[Y | t] = [Y | T = t, G = g]$

นิยามของข้อสอบที่ต้องการศึกษาไม่แสดงทำหน้าที่เบี่ยงเบน (Not exhibit DIF) ตามทฤษฎีการทดสอบแบบมาตรฐานเดิมกำหนดได้ว่า “ถ้า $E_R[Y | t] = E_F[Y | t]$ สำหรับทุกค่าของคะแนนจริง t จากแบบทดสอบที่ใช้ในการจับคู่” นิยามดังกล่าวสมมูลกับนิยามที่กำหนดโดยใช้ตัวแปรแฝงตามทฤษฎีการตอบข้อสอบ (Chang; Mazzeo; & Roussos. 1966: 337; citing Chang; & Mazzeo. 1994) ดังนั้นสามารถนำนิยามดังกล่าวไปอธิบายการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนหลายค่าด้วยวิธีโพลี-ชิปเทสต์ ซึ่งมีรายละเอียดดังนี้

วิธีชิปเทสต์พัฒนามาจากโมเดลในทฤษฎีการตอบข้อสอบ ดังนั้นการกำหนดนิยามการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนสองค่าจะใช้ตัวแปรแฝง ซึ่งสามารถกำหนดได้ว่า “การทำหน้าที่เบี่ยงเบนของข้อสอบเกิดขึ้นเมื่อ $E_R[Y | \theta] \neq E_F[Y | \theta]$ ที่ระดับความสามารถ θ ” และปริมาณของการทำหน้าที่เบี่ยงเบนของข้อสอบสามารถวัดได้จาก $B_o = E_R[Y | \theta] - E_F[Y | \theta]$ เมื่อ $f_F(\theta)$ แทนความหนาแน่นของความสามารถ θ ในกลุ่มสนใจ ดังนั้นการทำหน้าที่เบี่ยงเบนของข้อสอบสามารถคำนวณได้จาก $\beta_o = \int B_o(\theta) f_F(\theta) d\theta$

จากนิยามการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนสองค่าโดยใช้ตัวแปรแฝง เมื่อเปลี่ยนตัวแปรแฝงเป็นคะแนนจริงตามทฤษฎีการทดสอบแบบมาตรฐานเดิม สามารถกำหนดนิยามการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนหลายค่าได้ว่า $B(t) = E_R[Y | t] - E_F[Y | t]$ โดยที่ดัชนีการทำหน้าที่เบี่ยงเบนของข้อสอบสามารถคำนวณจาก $f_F(t) \beta = \int B_o(t) f_F(t) dt$ ซึ่งสามารถประมาณค่าได้ดังนี้

$$d_k = \bar{Y}_{Rk} - \bar{Y}_{Fk} \quad , k = 0, 1, 2, \dots, n_H \quad (7)$$

เมื่อ $\bar{Y}_{Rk}$ และ $\bar{Y}_{Fk}$ แทนค่าเฉลี่ยของคะแนนข้อสอบที่ต้องการศึกษาของผู้สอบทั้งหมด ซึ่งได้คะแนน $X = k$ ของกลุ่มอ้างอิงและกลุ่มสนใจ ตามลำดับ ค่า $\bar{Y}_{Rk} - \bar{Y}_{Fk}$ เป็นความแตกต่างของผลการตอบข้อสอบที่ต้องการศึกษาระหว่างกลุ่มผู้สอบ ซึ่งมีคะแนนสังเกตจากแบบทดสอบที่ใช้ในการจับคู่เท่ากัน ถ้าสมมติว่าผู้สอบมีคะแนนสังเกตเท่ากับหรือใกล้เคียงคะแนนจริงจากแบบทดสอบที่ใช้ในการจับคู่ ซึ่งจะเป็นจริงได้ถ้าแบบทดสอบดังกล่าวมีความยาวมากพอเพื่อจะให้ความเชื่อมั่นสูง หรือกลุ่มผู้สอบที่ศึกษามีการแจกแจงความสามารถคล้ายคลึงกัน ดังนั้นสมการ (7) ถือได้ว่าเป็นความแตกต่างในคะแนนข้อสอบที่ระดับคะแนนจริงเท่ากัน ถ้าข้อสอบที่ศึกษาไม่มีคะแนนสังเกตที่ทำให้ข้อสอบทำหน้าที่เบี่ยงเบน แล้วคาดว่า $d_k \approx 0$ สำหรับการประมาณค่าดัชนีการทำหน้าที่เบี่ยงเบนของข้อสอบ สามารถคำนวณดังนี้

$$\hat{\beta} = \sum_{k=0}^{n_H} p_k d_k \quad (8)$$

โดยที่

$$p_k = \frac{N_{Rk} + N_{Fk}}{N} \quad (9)$$

เมื่อ p_k แทนสัดส่วนของผู้สอบทั้งหมด (กลุ่มอ้างอิงและกลุ่มสนใจ) ซึ่งตอบแบบทดสอบที่ใช้ในการจับคู่ $X_1, X_2, \dots, X_n$ แล้วได้คะแนน $X = k$ ต่อจากนั้นจะนำค่าประมาณ $\hat{\beta}$ มาทดสอบสมมติฐาน ดังนี้

การทดสอบสมมติฐาน

นำดัชนี $\hat{\beta}$ มาทดสอบสมมติฐานศูนย์ (No-DIF) โดยใช้สถิติ B ดังนี้

$$B = \frac{\hat{\beta}}{\hat{\sigma}(\hat{\beta})} \quad (10)$$

โดยที่

$$\hat{\sigma}(\hat{\beta}) = \sqrt{\sum_{k=0}^{n_H} P_k^2 \left[\frac{\hat{\sigma}^2(Y|k, R)}{N_{Rk}} + \frac{\hat{\sigma}^2(Y|k, F)}{N_{Fk}} \right]} \quad (11)$$

เมื่อ $\hat{\sigma}(\hat{\beta})$ แทนค่าประมาณความคลาดเคลื่อนมาตรฐานของ β และ $\hat{\sigma}^2(Y|k, g)$ แทนค่าประมาณความแปรปรวนของคะแนนจากแบบทดสอบที่ต้องการศึกษาในกลุ่ม g (R หรือ F) ซึ่งมีคะแนนรวมเท่ากับ k สำหรับสถิติ B มีการแจกแจงใกล้เคียงปกติมาตรฐาน $[N(0, 1)]$ เมื่อข้อสอบทำหน้าที่ไม่เบี่ยงเบน (No-DIF) ถ้าผลการทดสอบพบว่า $|B| > Z_{1-\frac{\alpha}{2}}$ อย่างมีนัยสำคัญที่ระดับ α แสดงว่า ปฏิเสธ H_0 นั่นคือ ข้อสอบทำหน้าที่เบี่ยงเบน (DIF) โดยเข้าข้างผู้สอบกลุ่มใดกลุ่มหนึ่ง

ในการทดสอบด้วยสถิติ B ดังกล่าวอาจมีความคลาดเคลื่อนประเภทที่ 1 สูง ซึ่งเกิดจากความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบ ความแตกต่างดังกล่าวจะส่งผลกระทบ (Impact) ต่อการทดสอบสถิติ B_{uni} โดยทำให้มีค่าเพ้อ (Inflate) หรือสูงผิดปกติ ซึ่งคล้ายกับปัญหาที่เกิดขึ้นในกรณีข้อสอบที่ให้คะแนนสองค่า ปัญหาดังกล่าวสามารถแก้ไขโดยการปรับแก้การถดถอย (Regression correction) ซึ่งมีวิธีการคำนวณคล้ายกับกรณีข้อสอบที่ให้คะแนนสองค่าดังที่กล่าวมาข้างต้น แต่จะแตกต่างตรงที่การประมาณโดยใช้การถดถอยเชิงเส้นของคะแนนจริงบนคะแนนสังเกตจากทฤษฎีการทดสอบแบบมาตรฐานเดิม ในกรณีข้อสอบให้คะแนนหลายค่าจะทำการคำนวณแอลฟาของครอนบาค (Cronbach's alpha) ซึ่งเป็นการประมาณค่าความชันของเส้นการถดถอยของแต่ละกลุ่ม แต่ในวิธีชิปเทสท์ต้นฉบับเดิมจะคำนวณด้วย K-R 20 (Kuder-Richardson formula 20) ส่วนการคำนวณอื่นๆ จะเหมือนเดิม

4.3 วิธีการถดถอยโลจิสติก (Logistic regression; LR)

สวามินาธานและโรเจอร์ส (Swaminathan; & Rogers. 1990) ได้นำโมเดลการถดถอยโลจิสติก (Logistic regression model) ไปประยุกต์ใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนสองค่า (Dichotomous) แนวคิดในการตรวจสอบดัดแปลงมาจากวิธีล็อก-ลิเนียร์และโมเดลโลจิสต์ (Logit model) ของเมลเลนเบิร์ก (Mellenbergh. 1982) รวมทั้งเชื่อมแนวคิดระหว่างวิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel; MH) ของฮอลแลนด์และทาเยอร์ (Holland; & Thayer. 1988) วิธีการทำให้เป็นมาตรฐาน (standardization; STD) ที่พัฒนาโดยดูแรนส์และคูลิค (Dorans; & Kulick. 1986) วิธีชิปเทสท์ของชีลลีและสเตาท์ (Shealy; & Stout. 1993) และวิธีที่ใช้ทฤษฎีการตอบข้อสอบ หลักการตรวจสอบจะใช้โมเดลการถดถอยโลจิสติกทำนายความน่าจะเป็นในการตอบข้อสอบถูก โดยใช้คะแนนรวมแทนระดับความสามารถซึ่งสมมติว่าเป็นตัวแปรความสามารถแบบต่อเนื่อง โมเดลดังกล่าวมีพจน์สำหรับทดสอบปฏิสัมพันธ์ระหว่างการเป็นสมาชิกของกลุ่มผู้สอบ (Group membership) กับระดับความสามารถ (Ability level) ดังนั้นจึงสามารถ

ตรวจสอบข้อสอบทำหน้าที่ย่อยเป็นรูปแบบเดียวกัน (Uniform DIF) และข้อสอบทำหน้าที่ย่อยเป็นรูปแบบเดียวกัน (Nonuniform DIF) โมเดลการถดถอยโลจิสติกมีความยืดหยุ่นสามารถนำไปปรับขยายเพื่อใช้ในการตรวจสอบการทำหน้าที่ย่อยของข้อสอบที่ให้คะแนนหลายค่า (Polytomous) (French; & Miller. 1996; Zumbo. 1999) และตรวจสอบการทำหน้าที่ย่อยของข้อสอบที่วัดความสามารถหลายมิติ (Multidimensional) (Mazor; Hambleton; & Clauser. 1998)

แนวคิดและหลักการ

ในการตรวจสอบการทำหน้าที่ย่อยของข้อสอบจะใช้โมเดลการถดถอยโลจิสติกทำนายความน่าจะเป็นในการตอบข้อสอบถูก ดังนี้

$$P(u = 1|\theta) = \frac{e^{(\beta_0 + \beta_1\theta)}}{1 + e^{(\beta_0 + \beta_1\theta)}} \quad (1)$$

เมื่อ

u แทน ผลการตอบข้อสอบ

θ แทน ความสามารถที่สังเกตได้ของผู้สอบ

β_0 แทน พารามิเตอร์ส่วนตัด (Intercept parameter)

β_1 แทน พารามิเตอร์ความชัน (Slope parameter)

สมการดังกล่าวเป็นโมเดลการถดถอยโลจิสติกแบบมาตรฐาน ซึ่งใช้ทำนายตัวแปรตามแบบสองค่าจากตัวแปรอิสระที่กำหนดให้ โมเดลการถดถอยโลจิสติกสามารถนำไปประยุกต์ใช้ในการตรวจสอบการทำหน้าที่ย่อยของข้อสอบ โดยสามารถแยกสมการกลุ่มผู้สอบสองกลุ่มที่สนใจได้ดังนี้

$$P(u_{ij} = 1 | \theta_{ij}) = \frac{e^{(\beta_{0j} + \beta_{1j}\theta_{ij})}}{1 + e^{(\beta_{0j} + \beta_{1j}\theta_{ij})}} \quad ; i = 1, \dots, n_j, j = 1, 2. \quad (2)$$

เมื่อ

u_{ij} แทน ผลการตอบข้อสอบของผู้สอบคนที่ i ในกลุ่ม j

θ_{ij} แทน ความสามารถของผู้สอบคนที่ i ในกลุ่ม j

β_{0j} แทน พารามิเตอร์ส่วนตัดในกลุ่ม j

β_{1j} แทน พารามิเตอร์ความชันในกลุ่ม j

จากนิยามของการทำหน้าที่เบี่ยงเบนของข้อสอบ (Differential item functioning; DIF) กำหนดไว้ว่า “ข้อสอบทำหน้าที่เบี่ยงเบน ถ้าผู้สอบมีความสามารถเท่ากัน แต่มาจากกลุ่มผู้สอบที่แตกต่างกัน มีความน่าจะเป็นในการตอบข้อสอบถูกไม่เท่ากัน” (Swaminathan; & Rogers. 1990: 363; citing Hambleton; & Swaminathan. 1985) จากนิยามดังกล่าวเมื่อพิจารณาโมเดลการถดถอยโลจิสติกสามารถสรุปได้ว่า ถ้า $\beta_{01} = \beta_{02}$ และ $\beta_{11} = \beta_{12}$ แล้วไค์การถดถอยโลจิสติกระหว่างกลุ่มผู้สอบเท่ากัน แสดงว่าข้อสอบที่ทำหน้าที่ไม่เบี่ยงเบน (No-DIF) ถ้า $\beta_{11} = \beta_{12}$ และ $\beta_{01} \neq \beta_{02}$ แล้วไค์การถดถอยโลจิสติกระหว่างกลุ่มผู้สอบขนานกันแต่ไม่ทับกัน แสดงว่าข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF) และถ้า $\beta_{01} = \beta_{02}$ และ $\beta_{11} \neq \beta_{12}$ แล้วไค์การถดถอยโลจิสติกระหว่างกลุ่มผู้สอบไม่ขนานกัน แสดงว่าข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF)

กระบวนการตรวจสอบ

ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบจะใช้โมเดลการถดถอยโลจิสติกทดสอบความแตกต่างของผลการตอบข้อสอบ ระหว่างกลุ่มผู้สอบที่ระดับความสามารถเดียวกัน โดยสามารถทดสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกัน และที่ไม่เป็นรูปแบบเดียวกันได้พร้อมกัน ดังนี้

$$P(u_{ij} = 1) = \frac{e^{z_{ij}}}{1 + e^{z_{ij}}} \quad (3)$$

เมื่อ

$$z_{ij} = \tau_0 + \tau_1\theta + \tau_2g + \tau_3(\theta g) \quad (4)$$

โมเดลในสมการ (4) สมมูลกับโมเดลในสมการ (1) โดยรวมพารามิเตอร์ของข้อสอบทำหน้าที่ย่อยเบนที่เป็นรูปแบบเดียวกัน และที่ไม่เป็นรูปแบบเดียวกันไว้ในโมเดลเดียวกัน สำหรับ τ_0 เป็นพารามิเตอร์ส่วนตัว τ_1, τ_2 และ τ_3 เป็นค่าสัมประสิทธิ์ของระดับความสามารถของผู้สอบ (θ) การเป็นสมาชิกของกลุ่มผู้สอบ (g) และปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับการเป็นสมาชิกของกลุ่มผู้สอบ (θg) ตามลำดับ ส่วนตัวแปร g ที่ใช้แทนสมาชิกของกลุ่มผู้สอบจะกำหนดรหัสเป็นตัวแปรดัมมี่ (Dummy variable) ดังนี้

$$g = \begin{cases} 1 & \text{ถ้าผู้สอบเป็นสมาชิกในกลุ่มอ้างอิง} \\ 0 & \text{ถ้าผู้สอบเป็นสมาชิกในกลุ่มสนใจ} \end{cases}$$

จากโมเดลการถดถอยโลจิสติกในสมการ (3) พจน์ θg เป็นผลคูณของตัวแปรอิสระ 2 ตัว คือ θ และ g ซึ่งกำหนดรหัสตัวแปรดังกล่าว สำหรับพารามิเตอร์ τ_2 และ τ_3 มีความเกี่ยวข้องกับพารามิเตอร์ของโมเดลในสมการ (2) ดังนี้

$$\tau_2 = \beta_{01} - \beta_{02} \quad (5)$$

$$\tau_3 = \beta_{11} - \beta_{12} \quad (6)$$

สำหรับสมการ (3) สามารถเขียนให้อยู่ในรูปล็อกธรรมชาติของอัตราส่วนต่อ (Odds ratio) ได้ดังนี้ (Zumbo, 1999: 23)

$$\ln \left[\frac{P}{1-P} \right] = \tau_0 + \tau_1 \theta + \tau_2 g + \tau_3 (\theta g) \quad (7)$$

ความสัมพันธ์ระหว่างวิธีการถดถอยโลจิสติกและวิธีแมนเทิล-แฮนส์เซล

สวามินาธานและโรเจอร์ส (Swaminathan; & Rogers, 1990: 363-364) ได้แสดงให้เห็นถึงความเกี่ยวข้องระหว่างวิธีการถดถอยโลจิสติกและวิธีแมนเทิล-แฮนส์เซล เมื่อกำหนดตัวแปรความสามารถแบบไม่ต่อเนื่อง m ระดับ (แบบทดสอบมี $m-1$ ข้อ) โดยนิยามตัวแปร X_k ($k = 1, 2, 3, \dots, m-1$) ดังนี้

$$X_k = \begin{cases} 1 & \text{ถ้าผู้สอบที่เป็นสมาชิกมีความสามารถระดับ } k \\ 0 & \text{ถ้าผู้สอบที่เป็นสมาชิกมีความสามารถระดับอื่นๆ} \end{cases}$$

ผู้สอบทุกคนที่มีความสามารถระดับ m ได้รับคะแนน 1 (หรือ 0) โดยกำหนดรหัสในสมการ (8) ดังนี้

$$z = \beta_0 + \sum_{k=1}^{m-1} \beta_k X_k + \tau g \quad (8)$$

จากสมการ (8) มีข้อสังเกตว่า ไม่มีพจน์ปฏิสัมพันธ์ระหว่าง X_k และ g สูตรดังกล่าวนี้สามารถเขียนในรูปล็อกของอัตราส่วนแอดมิตติวอดส์ดังนี้

$$\ln \left[\frac{P}{(1-P)} \right] = \beta_0 + \sum_{k=1}^{m-1} \beta_k X_k + \tau g \quad (9)$$

จากสมการ (9) $\tau = \ln \alpha$ โดยที่ α คือ อัตราส่วนแอดมิตติวอดส์ร่วม (Common odds ratio) ของวิธีแมนเทิล-แฮนส์เซล ซึ่งกำหนดโดยฮอลแลนด์และทาเยอร์ (Holland; & Thayer. 1988: 134) ส่วนการทดสอบสมมติฐาน $\tau = 0$ ของวิธีการถดถอยโลจิสติกจะสมมูลกับการทดสอบ $\alpha = 0$ ของวิธีแมนเทิล-แฮนส์เซล จากเหตุผลดังกล่าวแสดงว่าวิธีแมนเทิล-แฮนส์เซลมีพื้นฐานมาจากโมเดลการถดถอยโลจิสติก โดยมีลักษณะเฉพาะ คือ ตัวแปรความสามารถเป็นตัวแปรแบบไม่ต่อเนื่อง (Discrete variable) และไม่มีพจน์ปฏิสัมพันธ์ระหว่างตัวแปรกลุ่มผู้สอบ กับตัวแปรระดับความสามารถของผู้สอบ ส่วนวิธีการถดถอยโลจิสติกจะใช้ตัวแปรความสามารถแบบต่อเนื่อง (Continuous variable) และในโมเดลมีพจน์ปฏิสัมพันธ์ระหว่างตัวแปรกลุ่มผู้สอบ กับตัวแปรระดับความสามารถของผู้สอบ เมื่อพิจารณาโมเดลในแบบแผนการทดลองพบว่า โมเดลการถดถอยโลจิสติกที่คำนวณโดยใช้สมการ (2) และ (3) สอดคล้องกับการวิเคราะห์ของโมเดลความแปรปรวนร่วม (Analysis of covariance model) ส่วนโมเดลแมนเทิล-แฮนส์เซลที่คำนวณโดยใช้สมการ (8) สอดคล้องกับโมเดลแบบแผนบล็อกสุ่ม (Randomized block design model) ในกรณีของแบบแผนบล็อกสุ่มมีข้อตกลงว่า ไม่มีปฏิสัมพันธ์ระหว่างปัจจัย “บล็อก” กับปัจจัย “ทรีทเมนต์” อย่างไรก็ตาม โมเดลการถดถอยโลจิสติกและโมเดลแมนเทิล-แฮนส์เซลมีความคล้ายกัน

การประมาณค่าพารามิเตอร์

ในการประมาณค่าพารามิเตอร์ของข้อสอบจะใช้การประมาณค่าความน่าจะเป็นมากที่สุด (Maximum likelihood estimation; MLE) โดยใช้ฟังก์ชันในการคำนวณดังนี้ (Narayanan; & Swaminathan. 1996: 262)

$$L(u_{ij} | \theta) = \prod_{i=1}^N \prod_{j=1}^n P(u_{ij})^{u_{ij}} [1 - P(u_{ij})]^{1-u_{ij}} \quad (10)$$

เมื่อ N แทนขนาดตัวอย่าง n แทนความยาวของแบบทดสอบ $u_{ij} = 1$ และ $P(u_{ij})$ เป็นค่าที่คำนวณจากสมการ (3) ในการประมาณค่าพารามิเตอร์โดยใช้วิธีการประมาณค่าความน่าจะเป็นมากที่สุดมีการแจกแจงปกติหลายตัวแปรแบบเชิงเส้นกำกับ (Asymptotically multivariate normal) โดยมีเวกเตอร์เฉลี่ย τ และเมทริกซ์ความแปรปรวน-ความแปรปรวนร่วม Σ ดังนี้

$$\hat{\tau} \sim N(\tau, \Sigma) \quad (11)$$

เมื่อ $\hat{\tau}' = [\tau_0 \ \tau_1 \ \tau_2 \ \tau_3]$ สำหรับค่าความคลาดเคลื่อนมาตรฐานแบบเชิงเส้นกำกับของค่าประมาณ τ_s ($s = 0, \dots, 3$) สามารถคำนวณจากรากที่สองของสมาชิกในแนวเส้นทแยงมุมสี่ค่าของเมทริกซ์ Σ ดังนี้

$$SE(\hat{\tau}_s) = \sqrt{\Sigma_{ss}} \quad (12)$$

การทดสอบสมมติฐาน

การทดสอบสมมติฐานของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF) และข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) จะทดสอบพร้อมกันโดยกำหนดสมมติฐานศูนย์ $H_0 : \tau_2 = 0$ และ $H_0 : \tau_3 = 0$ และสมมติฐานทางเลือก $H_1 : \tau_2 \neq 0$ หรือ $H_1 : \tau_3 \neq 0$ พารามิเตอร์ τ_2 บ่งชี้ถึงความแตกต่างของค่าเฉลี่ยในผลการตอบข้อสอบของกลุ่มผู้สอบ และ τ_3 บ่งชี้ถึงปฏิสัมพันธ์ระหว่างระดับความสามารถ กับการเป็นสมาชิกของกลุ่มผู้สอบ ถ้า $\tau_2 \neq 0$ และ $\tau_3 = 0$ แสดงว่า ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน โดยเข้าข้าง

กลุ่มอ้างอิงเมื่อ $\tau_2 > 0$ หรือเข้าข้างกลุ่มสนใจเมื่อ $\tau_2 < 0$ แต่ถ้า $\tau_3 \neq 0$ ($\tau_2 = 0$ หรือ $\tau_2 \neq 0$ ก็ได้) แสดงว่า **ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน** เมื่อ $\tau_3 > 0$ ข้อสอบจะเข้าข้างกลุ่มอ้างอิงที่ระดับความสามารถสูงกว่า และเข้าข้างกลุ่มสนใจที่ระดับความสามารถต่ำกว่า ในทางตรงกันข้าม เมื่อ $\tau_3 < 0$ ข้อสอบจะเข้าข้างกลุ่มสนใจที่ระดับความสามารถสูงกว่า และเข้าข้างกลุ่มอ้างอิงที่ระดับความสามารถต่ำกว่า สำหรับสมมติฐานที่ใช้ทดสอบข้อสอบทำหน้าที่เบี่ยงเบนสองรูปแบบพร้อมกัน เป็นดังนี้

$$H_0 : C\tau = 0$$

$$H_1 : C\tau \neq 0 \quad (13)$$

เมื่อ

$$C = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (14)$$

สถิติที่ใช้ทดสอบสมมติฐานมีลักษณะดังนี้

$$\chi^2 = \hat{\tau}'C'(C\Sigma C')^{-1}C\hat{\tau} \quad (15)$$

สถิติการทดสอบดังกล่าวมีการแจกแจงแบบไค-กำลังสอง (χ^2) ที่ระดับชั้นของความเป็นอิสระเท่ากับ 2 ($df = 2$) เมื่อทดสอบนัยสำคัญทางสถิติโดยใช้สมการ (16) แล้วมีค่ามากกว่า $\chi^2_{\alpha:2}$ แสดงว่าปฏิเสธสมมติฐานของ**การทำหน้าที่ไม่เบี่ยงเบนของข้อสอบ** (No-DIF) โจดอยและเกียร์ล (Jodoin; & Gierl. 2001: 335) ได้เสนอดัชนี $R^2\Delta$ เพื่อใช้จำแนกขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบ โดยมีพื้นฐานจากค่าประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก (Weighted least squares estimate) ดัชนี $R^2\Delta$ มี 3 ระดับดังนี้

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

□ DIF ระดับ B ขนาดปานกลาง: ปฏิเสธสมมติฐานศูนย์ และ $0.035 \leq R^2\Delta < 0.070$

□ DIF ระดับ C ขนาดใหญ่: ปฏิเสธสมมติฐานศูนย์ และ $R^2\Delta \geq 0.070$

4.4 วิธีการถดถอยโลจิสติกแบบจัดอันดับ (Ordinal logistic regression; OLR)

นักวิจัยหลายคนได้นำโมเดลการถดถอยโลจิสติกไปประยุกต์ใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ในข้อสอบที่มีการให้คะแนนหลายค่า (Polytomous) ดังเช่น มิลเลอร์ และสเปร์ (Miller; & Spray. 1993) ได้เสนอการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก (Logistic discriminant function analysis) ส่วนเฟรนช์และมิลเลอร์ (French; & Miller. 1996) ได้ปรับขยายโมเดลการถดถอยโลจิสติก 3 โมเดล คือ โมเดลโลจิตของอัตราส่วนแบบต่อเนื่อง (Continuation ratio logits) โมเดลโลจิตแบบสะสม (Cumulative logits) และโมเดลโลจิตของรายการคะแนนที่อยู่ติดกัน (Adjacent categories logits) ต่อมาซัมโบ (Zumbo. 1999) ได้ใช้โมเดลการถดถอยโลจิสติกแบบจัดอันดับ (Ordinal logistic regression model) พร้อมกับเสนอแนวทางการทดสอบนัยสำคัญทางสถิติในการตัดสินข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และไม่เป็นรูปแบบเดียวกัน โดยใช้การทดสอบโมเดล 3 ขั้นตอน นอกจากนี้ยังได้พัฒนาดัชนีการวัดขนาดอิทธิพลของการทำหน้าที่เบี่ยงเบนของข้อสอบ (DIF effect size) โดยคำนวณค่า R^2 ที่ได้จากการทดสอบในแต่ละโมเดล

แนวคิดและหลักการ

การวิเคราะห์โดยใช้โมเดลการถดถอยโลจิสติกมีข้อได้เปรียบเหนือกว่าการถดถอยเชิงเส้น คือ ตัวแปรตามไม่เป็นตัวแปรต่อเนื่อง และตัวแปรอิสระกับตัวแปรตามไม่จำเป็นต้องมีความสัมพันธ์เชิงเส้น โมเดลการถดถอยโลจิสติกสามารถวิเคราะห์อิทธิพลของการเป็นสมาชิกของกลุ่มบนตัวแปรตาม โดยที่ความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตามจะเป็นเส้นตรงหรือไม่เป็นเส้นตรงก็ได้ ดังนั้นโมเดลการถดถอยโลจิสติกเหมาะสมที่จะนำมาประยุกต์ใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ไม่ว่าจะเป็นข้อมูลแบบสองค่าหรือหลายค่า

ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ โดยใช้โมเดลการถดถอยโลจิสติกแบบจัดอันดับ ตามกรอบแนวคิดของซัมโบ (Zumbo. 1999) จะประมาณค่าความน่าจะเป็นสะสม (Cumulative probability) ในรายการคำตอบที่ 1 ถึง $J - 1$ เมื่อ J เป็นจำนวนรายการคำตอบในมาตราแบบจัดอันดับ โดยที่ความน่าจะเป็นสะสมในรายการคำตอบสุดท้ายมีค่าเท่ากับ 1 ตัวอย่าง เช่น ข้อสอบมีรายการคำตอบ 3 รายการ โมเดลการถดถอยจะประมาณค่าความน่าจะเป็นสะสมในรายการที่ 1 ถึง 2 คือ ความน่าจะเป็นสะสมสำหรับการตอบในรายการที่ 1 หรือต่ำกว่า และความน่าจะเป็นสะสมสำหรับการตอบในรายการที่ 2 หรือต่ำกว่า โดยที่และความน่าจะเป็นสะสมในรายการที่ 3 ซึ่งเป็นรายการสุดท้ายมีค่าเท่ากับ 1 ดังนั้นจะคำนวณความน่าจะเป็นสะสมจำนวน $J - 1$ รายการ สำหรับโมเดลการถดถอยโลจิสติกแบบจัดอันดับ สามารถเขียนในรูปสัญลักษณ์ดังนี้

$$P(Y \leq j) = \frac{\exp(\alpha_j)}{1 + \exp(\alpha_j)} \quad (1)$$

เมื่อ $P(Y \leq j)$ แทนความน่าจะเป็นสะสมสำหรับการตอบในรายการที่ j หรือต่ำกว่า และ α_j เป็นพารามิเตอร์ส่วนตัวของความน่าจะเป็นสะสมในแต่ละรายการ โมเดลในสมการ (1) สามารถเขียนในรูปของล็อกได้ดังนี้

$$\log \left[\frac{P(Y \leq j)}{P(Y > j)} \right] = \alpha_j \quad (2)$$

หรือเขียนในรูปโลจิต (logit) ดังนี้

$$\text{logit}[P(Y \leq j)] = \alpha_j \quad (3)$$

โมเดลในสมการ (3) อยู่ในรูปโลจิต เมื่อโลจิตเป็นลอการิธึมธรรมชาติของอัตราส่วนแต้มต่อ (Odds ratio) สมการดังกล่าวมีเฉพาะพารามิเตอร์ส่วนตัวเท่านั้น เมื่อนำตัวแปรทำนายเพิ่มเข้าไปในโมเดล ได้แก่ คะแนนรวม (Total score) ซึ่งเป็นระดับความสามารถโดยสมมติว่าเป็นตัวแปรต่อเนื่อง ตัวแปรกลุ่มผู้สอบ (Group variable) และตัวแปรปฏิสัมพันธ์ระหว่างคะแนนรวม กับ กลุ่มผู้สอบ (Interaction variable) จะได้โมเดลใหม่ดังนี้

$$\text{logit}[P(Y \leq j)] = \alpha_j + b_1 X + b_2 G + b_3 (X * G) \quad (4)$$

เมื่อ

X แทน ระดับความสามารถของผู้สอบ

G แทน กลุ่มผู้สอบ

b_1 แทน พารามิเตอร์ที่เกี่ยวข้องกับระดับความสามารถของผู้สอบ (X)

b_2 แทน พารามิเตอร์ที่เกี่ยวข้องกับกลุ่มผู้สอบ (G)

b_3 แทน พารามิเตอร์ที่เกี่ยวข้องกับปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบ กับกลุ่มผู้สอบ ($X * G$)

กระบวนการตรวจสอบ

ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ โดยใช้วิธีการถดถอยโลจิสติกแบบจัดอันดับของซัมโบ (Zumbo. 1999) จะใช้กระบวนการตรวจสอบความเหมาะสมของข้อมูลกับโมเดล โดยจะทดสอบโมเดล 3 ขั้นตอน ดังนี้

ขั้นตอนที่หนึ่ง ทดสอบตัวแปรที่ใช้ในการจับคู่ความสามารถ (Matching variable) หรือตัวแปรเงื่อนไข (Conditioning variable) ในที่นี้คือคะแนนรวม ซึ่งเป็นระดับความสามารถของผู้สอบ (X) ในขั้นตอนนี้ถือว่าอิทธิพลของตัวแปรคะแนนรวมมีค่าเท่ากันในทุกการตอบ และสมมติให้ค่าประมาณของพารามิเตอร์ความชันมีเพียงค่าเดียว การนำตัวแปรดังกล่าวเข้าไปในสมการถดถอยมีลักษณะดังนี้

$$\text{logit}[P(Y \leq j)] = \alpha_j + b_1 X \quad (5)$$

ขั้นตอนที่สอง ทดสอบตัวแปรสมาชิกของกลุ่มผู้สอบ (G) ในที่นี้จะกำหนดรหัสเป็นตัวแปรดัมมี่ (Dummy variable) โดยให้กลุ่มอ้างอิงแทนด้วย 0 และกลุ่มสนใจแทนด้วย 1 แล้วเพิ่มตัวแปรดังกล่าวในสมการ (5) ของขั้นตอนที่ 1 ดังนี้

$$\text{logit}[P(Y \leq j)] = \alpha_j + b_1 X + b_2 G \quad (6)$$

ขั้นตอนที่สาม ทดสอบตัวแปรสมาชิกของกลุ่มผู้สอบ (G) และปฏิสัมพันธ์ระหว่างระดับความสามารถ กับสมาชิกของกลุ่มผู้สอบ ($X * G$) โดยเพิ่มพจน์ทั้งสองดังกล่าวในโมเดลสมการ (5) ของขั้นตอนที่ 1 ดังนี้

$$\text{logit}[P(Y \leq j)] = \alpha_j + b_1 X + b_2 G + b_3 (X * G) \quad (7)$$

การทดสอบสมมติฐาน

การทดสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกัน และที่ไม่เป็นรูปแบบเดียวกัน สามารถทดสอบได้พร้อมกัน โดยกำหนดสมมติฐานศูนย์ (H_0) และสมมติฐานทางเลือก (H_1) ดังนี้

$$\begin{aligned} H_0 : b_2 = 0 \text{ และ } b_3 = 0 \\ H_1 : b_2 \neq 0 \text{ หรือ } b_3 \neq 0 \end{aligned} \quad (8)$$

การทดสอบนัยสำคัญทางสถิติของการทำหน้าที่เบี่ยงเบนของข้อสอบ ตามกระบวนการทดสอบความเหมาะสมของข้อมูลกับโมเดล 3 ขั้นตอน โดยการนำตัวแปรเข้าทดสอบกับโมเดลตามลำดับขั้น ซึ่งได้กล่าวมาแล้วนั้น ในการตัดสินใจข้อสอบข้อใดทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน หรือที่ไม่เป็นรูปแบบเดียวกัน ซัมโบ (Zumbo. 1999: 26) ได้เสนอแนะให้เปรียบเทียบความแตกต่างระหว่างค่าไค-กำลังสองที่คำนวณได้ในขั้นตอนที่ 1 และ 3 ที่ระดับขั้นของความเป็นอิสระเท่ากับ 2 ($df = 2$) โดยมีค่า $p \leq .01$ สำหรับค่าของ df ดังกล่าวได้มาจากการเปรียบเทียบความแตกต่างระหว่าง df ของขั้นตอนทั้งสอง ในขั้นตอนที่ 1 มี $df = 1$ และในขั้นตอนที่ 3 มี $df = 3$ ดังนั้นจะทดสอบค่าไค-กำลังสองที่ระดับ $df = 2$ ซึ่งเป็นการทดสอบการทำหน้าที่เบี่ยงเบนของข้อสอบทั้งสองรูปแบบพร้อมกัน (Swaminathan; & Rogers. 1990: 365) ถ้าผลการทดสอบมีนัยสำคัญทางสถิติ โดยปฏิเสธสมมติฐานศูนย์ ($b_3 = 0$) นั้นแสดงว่าข้อสอบทำหน้าที่เบี่ยงเบน ซึ่งเป็นผลมาจากอิทธิพลของการเป็นสมาชิกของกลุ่มผู้สอบ และ/หรืออิทธิพลของปฏิสัมพันธ์ระหว่างการเป็นสมาชิกของกลุ่ม กับระดับความสามารถของผู้สอบ ส่วนการตัดสินใจข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน หรือที่ไม่เป็นรูปแบบเดียวกัน สามารถพิจารณาได้จากความแตกต่างของค่า R^2 ที่วิเคราะห์ได้ระหว่างช่วงแรก (ขั้นตอนที่ 1 กับ 2) และช่วงหลัง (ขั้นตอนที่ 2 กับ 3) ถ้าความแตกต่างในช่วงแรกมากกว่าช่วงหลัง สามารถสรุปได้ว่าข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน แต่ถ้าความแตกต่างในช่วงแรกน้อยกว่าช่วงหลังสรุปได้ว่าข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Zumbo. 1999: 30-31)

จากการศึกษาค้นคว้าวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ 4 วิธีข้างต้นสรุปได้ว่า วิธีชิปเทสต์ (SIBTEST) เป็นวิธีในกลุ่มที่ใช้ทฤษฎีการตอบข้อสอบ รูปแบบนันทวารเมทริก (Nonparametric form) ซึ่งซีลลีและสตาท์ (Shealy; & Stout. 1993) พัฒนาเพื่อใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนสองค่า ต่อมา ชาง มาซซีโอ และรูสโซ (Chang; Mazzeo; & Roussos. 1996) ได้ปรับขยายวิธีชิปเทสต์ ซึ่งเรียกว่า “วิธีโพลี-ชิปเทสต์” (Poly-SIBTEST) เพื่อใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนหลายค่า วิธีโพลี-ชิปเทสต์มีจุดเด่นหลายประการ เช่น คะแนนเกณฑ์การจับคู่ค่อนข้างเป็นคุณลักษณะแฝงมากกว่าคะแนนที่ได้จากการสอบ ทำให้มีความถูกต้องและแม่นยำ ใช้เทคนิคการตรวจสอบแบบหลายมิติ โดยแยกข้อสอบที่ใช้เป็นเกณฑ์การจับคู่ (Matching criterion) ออกจากข้อสอบที่ใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนอย่างชัดเจน ทำให้เกณฑ์มีความเที่ยงตรงและความเชื่อมั่น มีการคำนวณทวนซ้ำหลายรอบ (Iterative algorithm) เพื่อคัดเลือกข้อสอบที่ทำหน้าที่เบี่ยงเบนออกจากคะแนนเกณฑ์

การจับคู่ ทำให้เกณฑ์มีความบริสุทธิ์ (Purification) นอกจากนี้ยังมีการปรับแก้ค่าการถดถอย (Regression correction) เพื่อลดความแตกต่างของค่าความสามารถเป้าหมายระหว่างกลุ่มผู้สอบ ทำให้สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ที่สูงเกินปกติได้ นอกจากนี้ยังเป็นวิธีการตรวจสอบที่ไม่ยุ่งยากซับซ้อน ไม่จำเป็นต้องใช้ตัวอย่างขนาดใหญ่ ส่วนข้อจำกัดของวิธีนี้คือ เป็นวิธีที่ใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกัน (Uniform DIF) โดยเฉพาะ ซึ่งไม่สามารถตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF)

ส่วนวิธีการถดถอยโลจิสติก (Logistic regression; LR) เป็นวิธีที่ใช้ทฤษฎีการทดสอบแบบมาตรฐานเดิม รูปแบบพาราเมตริก (Parametric form) พัฒนาโดยสวามินาธานและโรเจอร์ส (Swaminathan; & Rogers. 1990) เพื่อใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถมิติเดียวและให้คะแนนสองค่า ต่อมา ซัมโบ (Zumbo. 1999) ได้ปรับขยายเป็น **“วิธีการถดถอยโลจิสติกแบบจัดอันดับ”** (Ordinal logistic regression; OLR) เพื่อใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนหลายค่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีจุดเด่นตรงที่ใช้โมเดลการถดถอยโลจิสติกทำนายความน่าจะเป็นในการตอบข้อสอบถูก ภายใต้คะแนนความสามารถแบบต่อเนื่องและการเป็นสมาชิกของกลุ่มผู้สอบ โมเดลดังกล่าวสามารถทดสอบอิทธิพลของปฏิสัมพันธ์ระหว่างระดับความสามารถกับกลุ่มผู้สอบ ทำให้มีประสิทธิภาพในการตรวจสอบข้อสอบทำหน้าที่เบี่ยงเบน ทั้งที่เป็นรูปแบบเดียวกัน (Uniform DIF) และที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) ส่วนข้อด้อยของวิธีนี้คือ เป็นวิธีการตรวจสอบภายใต้กรอบแนวคิดที่วัดความสามารถมิติเดียว เมื่อนำมาตรวจสอบในข้อสอบที่วัดความสามารถหลายมิติจะใช้คะแนนรวมแทนความสามารถ ซึ่งคำนวณมาจากค่าเฉลี่ยของคะแนนในทุกมิติ คะแนนรวมดังกล่าวจึงเป็นตัวแทนมิติเดียว ทำให้ไม่สามารถจำแนกระดับความสามารถของผู้สอบในทุกมิติได้อย่างแท้จริง การใช้คะแนนรวมเป็นเกณฑ์การจับคู่จะทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 สูงเกินปกติ

จากข้อจำกัดของวิธีการถดถอยโลจิสติกแบบจัดอันดับ ผู้วิจัยจึงมีความสนใจปรับขยายวิธีดังกล่าว เพื่อใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า (Multidimensional polytomous) โดยนำค่าความสามารถในทุกมิติเป็นเกณฑ์การจับคู่ เพื่อตรวจสอบว่า สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ที่สูงเกินปกติ ได้หรือไม่ ในการวิจัยครั้งนี้ใช้ชื่อวิธีที่ปรับใหม่ว่า **“วิธีการถดถอยโลจิสติกแบบจัดอันดับ**

หลายมิติ" (Multidimensional ordinal logistic regression; MOLR) ดังนั้นผู้วิจัยจึงนำวิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ มาเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า ภายใต้เงื่อนไขของปัจจัยที่ศึกษา โดยพิจารณาประสิทธิภาพในการตรวจสอบจากอัตราความถูกต้อง (Power rate) และอัตราความคลาดเคลื่อนประเภทที่ 1 (Type I error rate)

5. งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถมิติเดียวและให้คะแนนหลายค่า แบ่งออกเป็นสองส่วน คือ งานวิจัยต่างประเทศและงานวิจัยภายในประเทศ ขอบเขตของการศึกษาและผลการศึกษางานวิจัยดังกล่าว สรุปได้ดังนี้

5.1 งานวิจัยต่างประเทศ

โอชิมาและมิลเลอร์ (Oshima; & Miller. 1992: 237-248) ได้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ทฤษฎีการตอบข้อสอบแบบหลายมิติ โดยมีความมุ่งหมายในการศึกษาสองประการ คือ (1) ข้อสอบแบบหลายมิติไม่จำเป็นต้องทำหน้าที่เบี่ยงเบน ยกเว้นค่าเฉลี่ยของความสามารถรองแตกต่างกัน และ (2) เมื่อไม่มีความสัมพันธ์ระหว่างความสามารถหลักและความสามารถรอง แล้วความแตกต่างของความสามารถหลักมีสาเหตุทำให้ข้อสอบทำหน้าที่เบี่ยงเบน การศึกษาครั้งนี้จำลองข้อมูลภายใต้ทฤษฎีการตอบข้อสอบโดยใช้โมเดลโลจิสติก 2 พารามิเตอร์แบบหลายมิติที่มีการชดเชย (Compensatory multidimensional two-parameter logistic; M2PL) กำหนดให้ θ_1 เป็นความสามารถหลัก และ θ_2 เป็นความสามารถรอง โดยจำลองพารามิเตอร์ของข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มสนใจมีค่าเท่ากันในแต่ละเงื่อนไข พารามิเตอร์ความยากของข้อสอบแบบหลายมิติ (MID) มีการแจกแจงแบบปกติ โดยมีค่าเฉลี่ยเท่ากับ 0 และ SD เท่ากับ 1 มีค่าอยู่ในช่วง -2 ถึง 2 ส่วนพารามิเตอร์อำนาจจำแนกของข้อสอบแบบหลายมิติ (MDISC) มีการแจกแจงแบบ log-normal โดยมีค่าเฉลี่ยเท่ากับ 1.13 และ SD เท่ากับ .6 ในการจำลองข้อสอบมี 2 เงื่อนไข คือ ข้อสอบทำหน้าที่ไม่เบี่ยงเบน (No-DIF) และข้อสอบทำหน้าที่

เบี่ยงเบน (DIF) ในเงื่อนไข No-DIF กำหนดให้การแจกแจงความสามารถ θ_2 ระหว่างกลุ่มเท่ากัน ส่วนในเงื่อนไข DIF กำหนดให้การแจกแจงความสามารถ θ_2 ระหว่างกลุ่มไม่เท่ากัน การจำลองข้อมูลภายใต้ 2 เงื่อนไขดังกล่าวจะจำลองใน 2 ปัจจัย คือ **ปัจจัยแรก** ค่าเฉลี่ยของการแจกแจงความสามารถ θ_1 ระหว่างกลุ่มมี 2 ระดับ คือ $\bar{\theta}_{1F} - \bar{\theta}_{1R} = 0$ และ $\bar{\theta}_{1F} - \bar{\theta}_{1R} = .5$ ส่วน **ปัจจัยสอง** คือ ร้อยละของข้อสอบทำหน้าที่เบี่ยงเบน 3 ขนาด คือ 5%, 10% และ 20% ในทุกเงื่อนไขมีจำนวนข้อสอบ 40 ข้อ ดังนั้นจะต้องจำลองข้อมูลทั้งหมด 12 เงื่อนไข ($2 \times 2 \times 3$) ในแต่ละเงื่อนไขจำลองซ้ำ 10 ครั้ง โดยใช้ขนาดตัวอย่างกลุ่มละ 1,000 คน (กลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน) และให้ค่าสหสัมพันธ์ระหว่าง θ_1 และ θ_2 มีค่าเป็น 0 แล้วนำผลการจำลองข้อมูลดังกล่าวมาคำนวณดัชนีการทำหน้าที่เบี่ยงเบน 4 วิธี ภายใต้โมเดลโลจิสติกแบบ 2 พารามิเตอร์ คือ วิธีการวัดพื้นที่แบบคิดเครื่องหมาย (SA) วิธีการวัดพื้นที่แบบไม่คิดเครื่องหมาย (UA) วิธีผลรวมของกำลังสองแบบคิดเครื่องหมาย (SSOS) และวิธีผลรวมของกำลังสองแบบไม่คิดเครื่องหมาย (USOS)

ผลการศึกษาในเงื่อนไข No-DIF พบว่า ข้อสอบแบบหลายมิติไม่จำเป็นต้องทำหน้าที่เบี่ยงเบน เมื่อการแจกแจงความสามารถ θ_2 ของกลุ่มสนใจและกลุ่มอ้างอิงเท่ากัน ($\bar{\theta}_{2F} - \bar{\theta}_{2R} = 0$) แล้วความคลาดเคลื่อนเชิงบวกในข้อสอบแบบหลายมิติมีจำนวนไม่มากกว่าข้อสอบแบบมิติเดียว และมีแนวโน้มจะเท่ากันในเงื่อนไขของความแตกต่างระหว่างกลุ่มซึ่งเกิดจากความสามารถ θ_1 ($\bar{\theta}_{1F} - \bar{\theta}_{1R} = .5$) ส่วนผลการศึกษาในเงื่อนไข DIF พบว่า เมื่อการแจกแจงความสามารถ θ_2 ระหว่างกลุ่มแตกต่างกัน ($\bar{\theta}_{2F} - \bar{\theta}_{2R} = .5$) แล้วอัตราการตรวจสอบการทำหน้าที่เบี่ยงเบนในข้อสอบแบบหลายมิติสูงกว่าข้อสอบแบบมิติเดียว สำหรับเงื่อนไข DIF 5% และไม่มีความแตกต่างระหว่างกลุ่ม ($\bar{\theta}_{1F} - \bar{\theta}_{1R} = 0$) ทั้ง 4 วิธีสามารถตรวจสอบการทำหน้าที่เบี่ยงเบนในข้อสอบแบบหลายมิติได้ 100% ปัจจัยสัดส่วนของ DIF มีผลต่ออัตราการตรวจสอบ กล่าวคือ เมื่อสัดส่วนของ DIF เพิ่มขึ้น แล้วอัตราการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบลดลง ส่วนทิศทางของข้อสอบพบว่า เมื่อข้อสอบวัดความสามารถ θ_1 เพิ่มขึ้น (α_{11}) แล้วอัตราการตรวจสอบลดลง ซึ่งไม่เป็นไปตามที่คาดหวัง นอกจากนี้พารามิเตอร์ MDISC มีผลต่ออัตราการตรวจสอบ ถ้า MDISC มีค่าเพิ่มขึ้นแล้วอัตราการตรวจสอบการทำหน้าที่เบี่ยงเบนข้อสอบจะมีค่าเพิ่มขึ้นด้วย

โรเจอร์สและสวามินาทาน (Rogers; & Swaminathan. 1993: 105-116) ได้ศึกษาเปรียบเทียบวิธีการถดถอยโลจิสติกและวิธีแมนเทิล-แฮนส์เซล ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน โดยใช้ข้อมูลจำลองศึกษาใน 2 กรณี

คือ กรณี 1 ศึกษาการแจกแจงของสถิติที่ใช้ทดสอบระหว่างสองวิธี และกรณี 2 ศึกษาความถูกต้องในการตรวจสอบระหว่างสองวิธี สำหรับการ**ศึกษาในกรณี 1** จำลองข้อมูล 4 เงื่อนไข (2×2) คือ ความเหมาะสมของข้อมูลกับโมเดล 2 ระดับ (เหมาะสมและไม่เหมาะสม) และขนาดตัวอย่าง 2 ขนาด (250 คนต่อกลุ่มและ 500 คนต่อกลุ่ม) โดยจำลองความยาวของแบบทดสอบ 40 ข้อ และการแจกแจงความสามารถแบบปกติมาตรฐาน ในแต่ละเงื่อนไขจำลองซ้ำ 100 ครั้ง แล้วทดสอบการแจกแจงของสถิติทั้งสองโดยใช้สถิติ Kolmogorov-Smirnov ส่วน**การศึกษาในกรณี 2** จำลองข้อมูล 32 เงื่อนไข ($2 \times 2 \times 2 \times 2 \times 2$) คือ ความเหมาะสมของข้อมูลกับโมเดล 2 ระดับ (เหมาะสมและไม่เหมาะสม) ขนาดตัวอย่าง 2 ขนาด (250 คนต่อกลุ่มและ 500 คนต่อกลุ่ม) ความยาวของแบบทดสอบ 2 ขนาด (40 ข้อ และ 80 ข้อ) โค้งของการแจกแจงคะแนน 2 ระดับ (ปกติและเบ้ซ้าย) สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 2 ขนาด (15% และ 0%) ในแต่ละเงื่อนไขจะจำลองข้อสอบทำหน้าที่เบี่ยงเบน 2 รูปแบบ คือ ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และที่ไม่เป็นรูปแบบเดียวกัน โดยในแต่ละรูปแบบมีขนาดของข้อสอบทำหน้าที่เบี่ยงเบน 4 ขนาด คือ .2, .4, .6 และ .8 ในการจำลองข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกันมี 4 ประเภท คือ b ต่ำกับ a สูง, b ปานกลางกับ a ต่ำ, b ปานกลางกับ a สูง และ b สูงกับ a สูง ในขณะที่การจำลองข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกันมี 5 ประเภท คือ b ต่ำกับ a ต่ำ, b ปานกลางกับ a ต่ำ, b ปานกลางกับ a สูง, b สูงกับ a ต่ำ และ b ปานกลางกับ a สูง (แบบผสมโดย a และ b ระหว่าง 2 กลุ่มต่างกัน) ในแต่ละเงื่อนไขจำลองซ้ำ 20 ครั้ง แล้วคำนวณค่าเฉลี่ยร้อยละของอัตราความถูกต้องในการตรวจสอบ

ผลการ**ศึกษาในกรณี 1 การศึกษาการแจกแจงของสถิติ**พบว่า การแจกแจงของสถิติการถดถอยโลจิสติกและแมนเทิล-แฮนส์เซลเป็นไปตามที่คาดหวังเกือบทุกเงื่อนไข โดยในเงื่อนไขของข้อสอบที่มีค่าความยากและอำนาจจำแนกสูง การทดสอบความไม่เหมาะสมของโมเดล และขนาดตัวอย่าง 500 คนต่อกลุ่ม การแจกแจงของสถิติการถดถอยโลจิสติกไม่เป็นไปตามที่คาดไว้ ส่วน**ในกรณี 2 การศึกษาอัตราความถูกต้องของการตรวจสอบ**พบว่า วิธีถดถอยโลจิสติกมีอัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ไม่เป็นรูปแบบเดียวกันสูงกว่าวิธีแมนเทิล-แฮนส์เซล แต่ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกันทั้งสองวิธีมีอัตราความถูกต้องเท่ากัน สำหรับปัจจัยของขนาดตัวอย่าง ประเภทของข้อสอบ และขนาดของข้อสอบทำหน้าที่เบี่ยงเบนมีผลต่ออัตราความถูกต้องของสองวิธี เมื่อขนาดตัวอย่างและขนาดของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้น มีผลทำให้อัตราความถูกต้องของการตรวจสอบ

เพิ่มขึ้น ในเงื่อนไขของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน ซึ่งมีค่าอำนาจจำแนกสูงและค่าความยากปานกลาง และในเงื่อนไขของข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ซึ่งมีค่าอำนาจจำแนกสูงและค่าความยากปานกลางแบบผสม มีผลทำให้ทั้งสองวิธีมีอัตราความถูกต้องในการตรวจสอบสูงสุด ส่วนในเงื่อนไขของข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ซึ่งมีค่าความยากปานกลางและค่าอำนาจจำแนกต่ำหรือสูง มีผลทำให้วิธีแมนเทิล-แฮนส์เชลมีอัตราความถูกต้องต่ำสุด

นารายานานและสวามินาทาน (Narayanan; & Swaminathan. 1994: 315-328) ได้ศึกษาผลของการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกัน และไม่เป็นรูปแบบเดียวกัน ระหว่างวิธีแมนเทิล-แฮนส์เชลและวิธีชิปเทสต์ โดยเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 โดยจำลองข้อมูลภายใต้การจัดกระทำ 5 ปัจจัย คือ ขนาดตัวอย่าง 9 ขนาด (กลุ่มอ้างอิง 3 ขนาด จำนวน 300 คน 500 คน และ 1,000 คน $\times$ กลุ่มสนใจ 3 ขนาด จำนวน 100 คน 200 และ 300 คน) ความแตกต่างของการแจกแจงความสามารถ 3 ระดับ (แบบเท่ากัน แบบแตกต่างกัน .5SD และแบบแตกต่างกัน 1SD) สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 2 ขนาด (10% และ 20%) ขนาดของข้อสอบทำหน้าที่เบี่ยงเบน 4 ขนาด (.4, .6, .8, 1.0) และประเภทของข้อสอบ 6 ระดับ (b ต่ำกับ a ปานกลาง, b ต่ำกับ a สูง, b ปานกลางกับ a ต่ำ, b ปานกลางกับ a สูง, b สูงกับ a ต่ำ และ b สูงกับ a ปานกลาง) โดยจำลองความยาวของแบบทดสอบ 40 ข้อ ในทุกเงื่อนไข ภายใต้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ ซึ่งกำหนดค่า c ทุกข้อเท่ากับ .2 โดยใช้โปรแกรม DATAGEN ดังนั้นจำลองข้อมูลทั้งหมด 1,296 เงื่อนไข ($9 \times 3 \times 2 \times 4 \times 6$) โดยจำลองซ้ำ 100 ครั้งในแต่ละเงื่อนไข ในการตรวจสอบด้วยวิธีแมนเทิล-แฮนส์เชลและวิธีชิปเทสต์ ใช้โปรแกรม MHBIAS และ SIBTEST ตามลำดับ ที่ระดับนัยสำคัญ .05 และ .01

ผลการศึกษาพบว่า เมื่อการแจกแจงความสามารถระหว่างกลุ่มอ้างอิงและกลุ่มสนใจมีค่าเท่ากัน วิธีชิปเทสต์และวิธีแมนเทิล-แฮนส์เชลมีอัตราความถูกต้องเท่ากันในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกัน แต่ถ้การแจกแจงความสามารถมีค่าแตกต่างกันแล้ววิธีชิปเทสต์มีอัตราความถูกต้องสูงกว่าวิธีแมนเทิล-แฮนส์เชล ทั้งสองวิธีมีอัตราความถูกต้องเพียงพอเมื่อขนาดตัวอย่าง 300 คนในแต่ละกลุ่ม สำหรับปัจจัยเกี่ยวกับขนาดตัวอย่าง การแจกแจงความสามารถ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ขนาดของข้อสอบทำหน้าที่เบี่ยงเบน และประเภทของข้อสอบมีผลต่ออัตราความถูกต้องของวิธีชิปเทสต์และวิธีแมนเทิล-

แฮนส์เซลอย่างมีนัยสำคัญ ยกเว้นในเงื่อนไขการแจกแจงความสามารถไม่มีผลต่อวิธีแมนเทิล-แฮนส์เซล ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทิล-แฮนส์เซลมีค่าอยู่ภายในขอบเขตที่กำหนด แต่วิธีชิปเทสท์มีค่าสูงกว่าที่กำหนดเล็กน้อย ในขณะที่ปัจจัยขนาดตัวอย่าง และประเภทของข้อสอบไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งสองวิธี ถ้าการแจกแจงความสามารถระหว่างกลุ่มอ้างอิงและกลุ่มสนใจมีค่าเท่ากันแล้วอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีชิปเทสท์มีค่าสูงกว่าวิธีแมนเทิล-แฮนส์เซลเล็กน้อย แต่ถ้าความแตกต่างของการแจกแจงค่าความสามารถมีค่าเพิ่มขึ้นจะมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งสองวิธีมีค่าเพิ่มขึ้นด้วย

นารายานานและสวามินาทาน (Narayanan; & Swaminathan. 1996: 257-274) ได้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ไม่เป็นรูปแบบเดียวกัน โดยเปรียบเทียบ 3 วิธี คือ วิธีแมนเทิล-แฮนส์เซล วิธีโคร-ชิป และวิธีการถดถอยโลจิสติก สำหรับวิธีโคร-ชิปได้พัฒนามาจากวิธีชิปเทสท์เพื่อใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ไม่เป็นรูปแบบเดียวกันโดยเฉพาะข้อมูลที่ใช้ในการศึกษาเป็นข้อมูลจำลองภายใต้ปัจจัยที่แปรเปลี่ยน 384 เงื่อนไข ($4 \times 2 \times 3 \times 4 \times 4$) คือ ขนาดตัวอย่าง 4 ขนาด (กลุ่มอ้างอิง 2 ขนาด จำนวน 500 คน และ 1,000 คน $\times$ กลุ่มสนใจ 2 ขนาด จำนวน 200 คน และ 500 คน) สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 3 ขนาด (0%, 10% และ 20%) ความแตกต่างของการแจกแจงค่าความสามารถ 2 ระดับ (แบบเท่ากันและแบบไม่เท่ากัน) ขนาดของข้อสอบทำหน้าที่เบี่ยงเบน 4 ขนาด (.4, .6, .8 และ 1.0) ประเภทของข้อสอบ 4 ระดับ (b ต่ำกับ a สูง, b ปานกลางกับ a ต่ำ, b ปานกลางกับ a สูง และ b สูงกับ a ต่ำ) โดยกำหนดพารามิเตอร์ b ร่วมกันระหว่างกลุ่มอ้างอิงและกลุ่มสนใจ คือ ต่ำ ($b = -1.5$), ปานกลาง ($b = 0.0$) และสูง ($b = 1.5$) ส่วนพารามิเตอร์ a มีค่าแปรเปลี่ยน คือ ต่ำ ($.40 \leq a \leq .50$ ในกลุ่มอ้างอิงและ $.72 \leq a \leq 1.03$ ในกลุ่มสนใจ) และสูง ($.47 \leq a \leq .90$ ในกลุ่มอ้างอิงและ $1.68 \leq a \leq 2.01$ ในกลุ่มสนใจ) แล้วใช้โปรแกรม DATAGEN จำลองผลการตอบข้อสอบที่มีความยาว 40 ข้อ จำนวน 384 เงื่อนไข ในแต่ละเงื่อนไขจำลองซ้ำ 100 ครั้ง ภายใต้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ โดยกำหนดค่าการเดาเท่ากับ .20 ทุกข้อ ส่วนการวิเคราะห์การทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีแมนเทิล-แฮนส์เซลและวิธีการถดถอยโลจิสติกใช้โปรแกรม DICHODIF สำหรับวิธีโคร-ชิปใช้โปรแกรม CSIBTEST โดยทดสอบที่ระดับนัยสำคัญ .05 และ .01

ผลการศึกษาพบว่า ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ไม่เป็นรูปแบบเดียวกัน วิธีโคร-ชิป และวิธีการถดถอยโลจิสติกมีอัตราความถูกต้องเท่ากัน ภายใต้เกือบทุกเงื่อนไข ส่วนวิธีแมนเทิล-แฮนส์เชลไม่มีอัตราความถูกต้องในการตรวจสอบดังกล่าว และจากการทดสอบปัจจัยที่มีผลต่อวิธีทั้งสามพบว่า ขนาดตัวอย่าง ประเภทของข้อสอบ และขนาดของข้อสอบทำหน้าที่เบี่ยงเบนมีผลต่ออัตราความถูกต้องของทั้งสามวิธี กล่าวคือ เมื่อขนาดตัวอย่างและขนาดของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้นมีผลทำให้อัตราความถูกต้องของทั้งสามวิธีมีค่าเพิ่มขึ้น และเมื่อประเภทของข้อสอบมีค่าอำนาจจำแนกสูงและความยากต่ำแล้วอัตราความถูกต้องของทั้งสามวิธีมีค่าสูงสุด โดยที่ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนมีผลต่ออัตราความถูกต้องของวิธีการถดถอยโลจิสติก เมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้นแล้วอัตราความถูกต้องจะลดลง ส่วนปัจจัยการแจกแจงความสามารถมีผลต่ออัตราความถูกต้องของวิธีแมนเทิล-แฮนส์เชล และวิธีโคร-ชิป ถ้าความแตกต่างของการแจกแจงความสามารถเพิ่มขึ้นแล้วอัตราความถูกต้องจะลดลง สำหรับการตรวจสอบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีแมนเทิล-แฮนส์เชลมีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ภายในขอบเขตที่กำหนด แต่วิธีโคร-ชิปและวิธีการถดถอยโลจิสติกมีค่าสูงกว่าขอบเขตที่กำหนดในทุกเงื่อนไข โดยที่วิธีโคร-ชิปมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีการถดถอยโลจิสติกประมาณ 1%

ซาง มาซซีโอ และรูสโซ (Chang; Mazzeo; & Roussos. 1996: 333-353) ได้ปรับเปลี่ยนวิธีชิปเทสต์ (Modified SIBTEST) เพื่อใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ในข้อสอบที่ให้คะแนนหลายค่า โดยนำไปเปรียบเทียบกับวิธีความแตกต่างเฉลี่ยมาตรฐาน (Standardized mean difference; SMD) และวิธีแมนเทิล (Mantel) โดยศึกษาภายใต้การจำลองข้อมูล 2 กรณี คือ **การศึกษากรณี 1** ใช้ข้อมูลจากการศึกษาของซวิก โดโนฮิว และกริมา (Zwick; Donoghue; & Grima. 1993) โดยออกแบบการจำลองข้อมูล 54 เงื่อนไข ภายใต้ความแตกต่างของข้อสอบความสามารถของกลุ่มผู้สอบ และรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน แบบทดสอบที่ใช้จับคู่ (Matching test) ประกอบด้วยข้อสอบ 24 ข้อ โดยที่ข้อสอบ 20 ข้อแรกให้คะแนนสองค่า และอีก 4 ข้อให้คะแนนหลายค่า โดยมี 4 รายการตอบ (ให้คะแนน 0, 1, 2 หรือ 3) ข้อสอบที่ให้คะแนนสองค่าจำลองโดยใช้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ (3PLM) และข้อสอบให้คะแนนหลายค่าจำลองโดยใช้โมเดลพหุคูณเครดิต (PCM) ข้อสอบที่ใช้ในการศึกษาเป็นข้อสอบให้คะแนนหลายค่า (DIF & Null DIF) มี 3 ข้อ แต่ละข้อมี 4 รายการตอบ แล้วใช้โมเดล PCM จำลองผลการตอบข้อสอบ

DIF ระหว่างกลุ่มอ้างอิงและกลุ่มสนใจ โดยจำลองรูปแบบของข้อสอบ 4 รูปแบบ คือ แบบคงที่ (Constant DIF) แบบสมดุล (Balanced DIF) แบบเปลี่ยนขนาดที่ระดับความสามารถต่ำ (Low-shift DIF) และแบบเปลี่ยนขนาดที่ระดับความสามารถสูง (High-shift DIF) รูปแบบของข้อสอบดังกล่าวถูกกำหนดโดยพารามิเตอร์ความยากของข้อสอบระหว่างกลุ่มผู้สอบ โดยกลุ่มอ้างอิงมีค่าพารามิเตอร์ความยากของข้อสอบเท่ากับ $(-.91, .98, .21)$, $(-2.25, -1.80, 1.66)$ และ $(-.54, -2.11, .74)$ ตามลำดับ ส่วนพารามิเตอร์ความยากของกลุ่มสนใจคำนวณโดยใช้สูตร $b_{jiF} = b_{jiR} \pm C_{ji}$ เมื่อ $i = 1, 2, 3$ ในแต่ละรูปแบบของข้อสอบ DIF จะมีขนาดของ DIF 2 ขนาด คือ $C_{ji} = .10$ และ $C_{ji} = .25$ ดังนั้นข้อสอบที่ศึกษามีทั้งหมด 27 ข้อ คือ ข้อสอบ No-DIF 3 ข้อ และข้อสอบ DIF 24 ข้อ (ข้อสอบ DIF 3 ข้อ $\times$ รูปแบบของ DIF 4 แบบ $\times$ ขนาดของ DIF 2 ขนาด) สำหรับผลการตอบข้อสอบดังกล่าวจำลองภายใต้ขนาดตัวอย่างกลุ่มละ 500 คน โดยมีการแจกแจงความสามารถ 2 ระดับ คือ ไม่มีความแตกต่างของความสามารถระหว่างกลุ่ม และมีความแตกต่างของความสามารถระหว่างกลุ่ม โดยกลุ่มอ้างอิงมีการแจกแจงแบบ $N(0, 1)$ เสมอ แต่กลุ่มสนใจมีการแจกแจง 2 ระดับ คือ $N(0, 1)$ และ $N(-1, 1)$ ดังนั้นจะต้องจำลองข้อมูลทั้งหมด 54 เงื่อนไข (27×2) โดยจำลองซ้ำ 600 ครั้งในแต่ละเงื่อนไข ส่วน**การศึกษากรณี 2** ขยายจากการศึกษากรณี 1 โดยนำค่าพารามิเตอร์อำนาจจำแนกของข้อสอบเข้าร่วมศึกษาด้วย จำลองแบบทดสอบที่ใช้ในการจับคู่ 24 ข้อ ในข้อสอบ 20 ข้อแรกให้คะแนนสองค่า โดยมีค่าพารามิเตอร์เหมือนเดิม ซึ่งจำลองโดยใช้โมเดล 3PLM ส่วนข้อ 21-24 เป็นข้อสอบที่ให้คะแนนหลายค่า ซึ่งได้เพิ่มพารามิเตอร์อำนาจจำแนกเข้ามาวิเคราะห์ โดยนำมาจากผลการประมาณค่าพารามิเตอร์ที่เกิดขึ้นจริงของ Nation Assessment of Educational Progress (NAEP) แล้วจำลองโดยใช้โมเดลพาเชียลเครดิตทั่วไป (Generalized partial credit model; GPCM) ของมูรากิ (Muraki, 1992) สำหรับข้อสอบที่ใช้ในการศึกษามีค่าพารามิเตอร์ความยากของกลุ่มอ้างอิงเท่ากับ $-1, 0$, และ 1 เหมือนกันทุกข้อ ส่วนพารามิเตอร์อำนาจจำแนกของข้อสอบจะมีค่าแตกต่างกัน 11 ระดับ คือ 2.000, 1.500, 1.360, 1.120, 1.000, .869, .588, .330, .250, .230, และ .150 การจำลองข้อสอบดังกล่าวมี 2 เงื่อนไข คือ **เงื่อนไขที่หนึ่ง** Null-DIF พารามิเตอร์ของข้อสอบที่ศึกษาจะระหว่างกลุ่มอ้างอิงและกลุ่มสนใจมีค่าเท่ากัน และ**เงื่อนไขที่สอง** Constant DIF โดยกำหนดค่าความยากของข้อสอบแปรเปลี่ยนที่ระดับ 0.25 ส่วนการแจกแจงความสามารถมีเพียงเงื่อนไขเดียว คือ กลุ่มอ้างอิงและกลุ่มสนใจมีการแจกแจงความสามารถแบบ $N(0, 1)$ และแบบ $N(-1, 1)$ ตามลำดับ ดังนั้นเงื่อนไขของการทดลองมีทั้งหมด 22 เงื่อนไข (Null-DIF 11 ระดับ และ DIF 11 ระดับ) โดยในแต่ละเงื่อนไขจำลองซ้ำ 1,000 ครั้ง ภายใต้ขนาดตัวอย่าง

กลุ่มละ 500 คน และจำลองในเงื่อนไขของข้อสอบ $a = .230, .330, .588, 1.120$ และ 1.360 เมื่อขนาดตัวอย่างเพิ่มขึ้นเป็น 1,000 และ 2,000 โดยจำลองซ้ำ 1,000 ครั้งในแต่ละเงื่อนไข

การศึกษาในกรณี 1 พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทิลและวิธีความแตกต่างเฉลี่ยมาตรฐานมีค่าเท่ากับ 0.049 และ 0.046 ตามลำดับ ค่าดังกล่าวอยู่ภายในขอบเขตที่กำหนด ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีชิปเทสท์มีค่าเท่ากับ 0.063 ซึ่งถือว่ามีความเสี่ยงเล็กน้อย ในกรณีการแจกแจงความสามารถแบบไม่เท่ากันจะมีค่าสูงกว่าแบบเท่ากัน เมื่อพิจารณาอัตราความถูกต้องพบว่า ในทุกเงื่อนไขทั้งสามวิธีสามารถตรวจสอบการกำหนดที่เบี่ยงเบนของข้อสอบที่มีทิศทางเดียวกัน (Unidirectional DIF) ซึ่งประกอบด้วยรูปแบบ Constant DIF, Low-shift DIF และ High-shift DIF โดยอัตราความถูกต้องของสามวิธีแตกต่างกันเล็กน้อย แต่ทั้งสามวิธีไม่สามารถตรวจสอบ Balanced DIF เมื่อขนาดของ DIF มีค่าเพิ่มขึ้นจาก 0.10 เป็น 0.25 มีผลทำให้อัตราความถูกต้องของทั้งสามวิธีมีค่าเพิ่มขึ้นทุกเงื่อนไข สำหรับ**ผลการศึกษาในกรณี 2** พบว่า เมื่อขนาดตัวอย่าง 500 คน (แต่ละกลุ่ม) และพารามิเตอร์อำนาจจำแนกของข้อสอบที่ศึกษามีค่าเพิ่มขึ้น แล้วอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทิลและวิธีความแตกต่างเฉลี่ยมาตรฐานมีค่าสูงขึ้น โดยเพิ่มขึ้นถึง 40% ซึ่งต่างจากวิธีชิปเทสท์ ถึงแม้ว่าค่าอำนาจจำแนกจะยิ่งสูงขึ้น แต่อัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นเล็กน้อย เมื่อเพิ่มขนาดตัวอย่างเป็น 1,000 คนต่อกลุ่ม อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทิลและวิธีความแตกต่างเฉลี่ยมาตรฐานยิ่งเพิ่มมากขึ้นอย่างรวดเร็ว ส่วนของวิธีชิปเทสท์ค่อนข้างคงที่

เฟรนช์และมิลเลอร์ (French; & Miller. 1996: 315-332) ได้ใช้วิธีการถดถอยโลจิสติกตรวจสอบการกำหนดที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกัน (Uniform DIF) และที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) โดยใช้ข้อมูลจำลองภายใต้โมเดลพาเซียลเครดิตทั่วไป (GPCM) จำลองผลการตอบข้อสอบ 25 ข้อ ที่มี 4 รายการตอบ โดยให้คะแนนตั้งแต่ 0 ถึง 3 ในข้อสอบที่จำลองดังกล่าวมีเพียง 1 ข้อเป็นข้อสอบ DIF (ข้อ 25) ส่วนที่เหลืออีก 24 ข้อเป็นข้อสอบ No-DIF (ข้อ 1-24) การจำลองข้อสอบ DIF มี 4 เงื่อนไข คือ **เงื่อนไข 1 ถึง 3 จำลองข้อสอบ Nonuniform DIF** โดยกำหนดค่าพารามิเตอร์อำนาจจำแนก (a) มีค่าแปรเปลี่ยน ในเงื่อนไข 1, 2 และ 3 ของกลุ่มสนใจมีค่า a เท่ากับ 0.5 ทั้ง 3 เงื่อนไข ส่วนกลุ่มอ้างอิงมีค่า a เท่ากับ 1.0, 1.5 และ 2.0 ตามลำดับ ส่วนค่าพารามิเตอร์ความยาก (b) ในเงื่อนไข 1, 2 และ 3 มีค่าเท่ากันทั้ง 2 กลุ่ม คือ $b_1 = 0$, $b_2 = -1$, $b_3 = 0$ และ $b_4 = 1$ สำหรับ**เงื่อนไข 4 จำลองข้อสอบ Uniform DIF**

โดยกำหนดพารามิเตอร์ความยาก (b) มีค่าแปรเปลี่ยน ในกลุ่มสนใจมีค่าพารามิเตอร์ความยาก b_1, b_2, b_3 และ b_4 เท่ากับ 0, -1, 0 และ 1 ตามลำดับ ส่วนกลุ่มอ้างอิงมีพารามิเตอร์ความยาก b_1, b_2, b_3 และ b_4 เท่ากับ 0, -2, 1 และ 2 ตามลำดับ ในการจำลองข้อสอบ No-DIF กำหนดให้พารามิเตอร์ของข้อสอบระหว่าง 2 กลุ่มมีค่าเท่ากัน การจำลองขนาดตัวอย่างมี 2 ขนาด คือ 500 คนและ 1,000 คน โดยมีการแจกแจงความสามารถแบบปกติ $N(0, 1)$ ดังนั้นจะต้องจำลองข้อมูลทั้งหมด 4 เงื่อนไข (2×2) คือ รูปแบบของข้อสอบ DIF 2 รูปแบบ และขนาดตัวอย่าง 2 ขนาด โดยในแต่ละเงื่อนไขจำลองซ้ำ 100 ครั้ง แล้ววิเคราะห์การหาหน้าที่เบี่ยงเบนของข้อสอบโดยประยุกต์ใช้โมเดลการถดถอยโลจิสติก 3 โมเดล คือ โมเดลโลจิสติกของอัตราส่วนแบบต่อเนื่อง (Continuation ratio logits) โมเดลโลจิสติกแบบสะสม (Cumulative logits) และโมเดลโลจิสติกของรายการคำตอบที่อยู่ติดกัน (Adjacent categories logits) โมเดลทั้งสามมีการลงรหัสข้อมูลในแบบแผนการวิเคราะห์แตกต่างกัน สำหรับข้อมูลที่ให้คะแนนหลายค่าซึ่งมีรายการตอบ 3 รายการจะลงรหัสข้อมูลสำหรับการวิเคราะห์ 3 โมเดล โดยที่โมเดลโลจิสติกแบบสะสมจะนำข้อมูลไปใช้วิเคราะห์ในทุกโมเดล ดังนั้นจึงไม่สูญเสียข้อมูล (Loss of data) ในแบบแผนการวิเคราะห์ของแต่ละโมเดล ส่วนการประมาณค่าพารามิเตอร์ของโมเดลการถดถอยโลจิสติกใช้วิธีความน่าจะเป็นมากที่สุดของ Newton-Raphson แล้วทดสอบนัยสำคัญด้วยสถิติไค-กำลังสองที่ระดับนัยสำคัญ .002

ผลการศึกษาพบว่า ขนาดกลุ่มตัวอย่าง ค่าพารามิเตอร์ของข้อสอบ และการลงรหัสข้อมูลในแบบแผนของโมเดลการถดถอยโลจิสติก มีผลต่อความถูกต้องในการตรวจสอบการหาหน้าที่เบี่ยงเบนของข้อสอบ ในข้อสอบที่ให้คะแนนหลายค่า กล่าวคือ เมื่อขนาดตัวอย่างเพิ่มขึ้นจาก 500 คน เป็น 2,000 คน มีผลทำให้อัตราการตรวจสอบมีค่าเพิ่มขึ้น เมื่อเพิ่มขนาดตัวอย่างเป็น 2,000 คน วิธีโลจิสติกของอัตราส่วนแบบต่อเนื่องและวิธีโลจิสติกแบบสะสมมีอัตราความถูกต้องสูงสุด ภายใต้อัตราความถูกต้องของวิธีโลจิสติกของรายการคำตอบที่อยู่ติดกันมีค่าต่ำสุด ในโมเดลการถดถอยที่หนึ่ง แต่ในโมเดลการถดถอยที่สองและสามยังคงมีความถูกต้องเพียงพอ เมื่อขนาดตัวอย่างลดลงถึง 500 คน อัตราความถูกต้องในโมเดลแรกของวิธีโลจิสติกของอัตราส่วนแบบต่อเนื่องและวิธีโลจิสติกแบบสะสมมีค่าลดลง แต่วิธีโลจิสติกของรายการคำตอบที่อยู่ติดกันลดลงทุกโมเดล ดังนั้นวิธีโลจิสติกของอัตราส่วนแบบต่อเนื่องและวิธีโลจิสติกแบบสะสมให้ผลการตรวจสอบที่คล้ายกัน ซึ่งเป็นไปตามที่คาดหวัง ส่วนพารามิเตอร์อำนาจจำแนกของข้อสอบ เมื่อมีค่าเพิ่มขึ้นแล้ว ทำให้อัตราการตรวจสอบของทั้งสามวิธีมีค่าเพิ่มขึ้น และทั้งสามวิธีสามารถตรวจสอบการหาหน้าที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกัน และที่ไม่เป็นรูปแบบเดียวกัน

รูสโซและสเตาท์ (Roussos; & Stout. 1996a: 215-230) ได้จำลองข้อมูลเพื่อศึกษาผลของตัวอย่างขนาดเล็กและค่าพารามิเตอร์ของข้อสอบที่มีต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีชิปเทสต์และวิธีแมนเทิล-แฮนส์เซล โดยจำลองข้อมูลศึกษาใน 2 กรณี คือ **กรณี 1 ศึกษาตัวอย่างขนาดเล็ก** โดยใช้ขนาดตัวอย่าง 4 ขนาด (100 คน 200 คน 500 คน และ 1,000 คน) ขนาดตัวอย่างในกลุ่มอ้างอิงเท่ากับกลุ่มสนใจ และการแจกแจงความสามารถในกลุ่มอ้างอิงและกลุ่มสนใจแบบปกติ โดยมีความแปรปรวนเท่ากับ 1 และความแตกต่าง (d_T) ของการแจกแจงมีค่าเฉลี่ย 3 ระดับ (0, .5 และ 1) แล้วจำลองข้อมูลภายใต้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ โดยประมาณค่าพารามิเตอร์จากข้อสอบจริง จำนวน 25 ข้อ ซึ่งนำมาจากแบบทดสอบ Armed Services Vocational Aptitude Battery (ASVAB) ของมิลเลฟวี่และบ็อค (Mislevy; & Bock. 1984) ในแบบทดสอบประกอบด้วยข้อสอบที่ทำหน้าที่ไม่เบี่ยงเบน 1 ข้อ ซึ่งมีค่าพารามิเตอร์ $a = 1.32$, $b = .03$ และ $c = .25$ โดยคัดเลือกมาจากข้อสอบที่มีค่าพารามิเตอร์ใกล้เคียงกับค่าเฉลี่ยในแบบทดสอบ ($\bar{a} = 1.22$, $\bar{b} = .09$ และ $\bar{c} = .20$) สำหรับข้อมูลที่ศึกษาจำลอง 400 ครั้งในแต่ละเงื่อนไข แล้วใช้สถิติชิปเทสต์และแมนเทิล-แฮนส์เซลทดสอบอัตราความคลาดเคลื่อนประเภทที่ 1 ที่ระดับนัยสำคัญ .05 สำหรับ**กรณี 2 ศึกษาค่าพารามิเตอร์ของข้อสอบ** ใช้ตัวอย่างขนาดปานกลางถึงใหญ่ 3 ขนาด (500 คน 1,000 คน และ 3,000 คน) โดยขนาดตัวอย่างในกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน สำหรับความสามารถระหว่างกลุ่มดังกล่าวมีการแจกแจงแบบปกติ โดยมีความแปรปรวนเท่ากับ 1 และความแตกต่าง (d_T) ในค่าเฉลี่ยของการแจกแจง 2 ระดับ (0 และ 1.0) ส่วนแบบทดสอบเหมือนในกรณี 1 โดยเลือกค่าพารามิเตอร์อำนาจจำแนก 3 ระดับ (.4, 1.0, 2.5) พารามิเตอร์ความยาก 5 ระดับ (-1.5, -.5, 0, .5 และ 1.5) พารามิเตอร์การเดา 1 ระดับ (0.20) เมื่อ $d_T = 0$ และพารามิเตอร์การเดา 3 ระดับ (.20, .10 และ .05) เมื่อ $d_T = 1.00$ สำหรับขนาดตัวอย่างระหว่างกลุ่มสนใจและกลุ่มอ้างอิงจะใช้จำนวนเท่ากัน การศึกษาในกรณี 2 จำลองข้อมูล 100 ครั้งในแต่ละเงื่อนไข แล้วทดสอบด้วยสถิติชิปเทสต์และแมนเทิล-แฮนส์เซลที่ระดับ .05

ผลการศึกษากรณี 1 เมื่อศึกษาตัวอย่างขนาดเล็กพบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีชิปเทสต์และวิธีแมนเทิล-แฮนส์เซลมีค่าไม่แตกต่างกัน ยกเว้นในเงื่อนไขขนาดตัวอย่าง 100 และการแจกแจงความสามารถระหว่างกลุ่มอ้างอิงและกลุ่มสนใจไม่แตกต่างกัน ($d_T = 0$) อัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งสองวิธีมีค่าเพิ่มขึ้นเล็กน้อยเมื่อขนาดตัวอย่างและความแตกต่างการแจกแจงความสามารถเพิ่มขึ้น วิธีแมนเทิล-แฮนส์เซลมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าเล็กน้อยเมื่อ $d_T > 0$ สำหรับ**ผลการศึกษากรณี 2 เมื่อศึกษา**

ค่าพารามิเตอร์ของข้อสอบพบว่า เมื่อไม่มีความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มอ้างอิงและกลุ่มสนใจ ($d_T = 0$) แล้วอัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งสองวิธีมีค่าอยู่ในขอบเขตที่กำหนดในทุกเงื่อนไข โดยที่วิธีแมนเทิล-แฮนส์เชลดีกว่าเล็กน้อย เมื่อเพิ่มขนาดตัวอย่างแล้วอัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งสองวิธีมีค่าลดลง โดยมีค่าลดลงถึง 0 เมื่อเพิ่มขนาดตัวอย่างเป็น 3,000 คน แต่ถ้าความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มมีค่าเพิ่มขึ้น ($d_T = 1.0$) แล้วอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีชิปเทสท์จะมีค่าต่ำกว่าวิธีแมนเทิล-แฮนส์เชล ในทุกเงื่อนไขของการตรวจสอบ

โอชิมา ราจุ และเฟลาเวอร์ (Oshima; Raju; & Flowers. 1997: 253-272) ได้พัฒนาการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบและแบบทดสอบ โดยใช้การวัดภายในของทฤษฎีการตอบข้อสอบแบบหลายมิติ ซึ่งปรับขยายมาจากวิธี DFIT ที่ใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบและแบบทดสอบที่วัดความสามารถมิติเดียว ซึ่งพัฒนาโดยราจุ แวนเดอร์ลินเด็น และเฟลีย์ร์ (Raju; van der Linden; & Fleer. 1995) สามารถคำนวณดัชนีการทำหน้าที่เบี่ยงเบนของข้อสอบที่มีการชดเชย (Compensatory DIF; CDIF) ดัชนีการทำหน้าที่เบี่ยงเบนของข้อสอบที่ไม่มีการชดเชย (Noncompensatory DIF; NCDIF) และดัชนีการทำหน้าที่เบี่ยงเบนของแบบทดสอบ (DTF) ข้อมูลที่ใช้ในการศึกษาจำลองภายใต้โมเดลโลจิสติก 2 พารามิเตอร์แบบหลายมิติที่มีการชดเชย โดยจำลองแบบทดสอบที่วัดความสามารถหลายมิติภายในข้อสอบ (Multidimensional within-item test) ซึ่งมีโครงสร้างการวัดความสามารถเป้าหมายทั้งสองมิติ จำนวน 40 ข้อ ขนาดตัวอย่างกลุ่มละ 1,000 คน ภายใต้เงื่อนไขที่แปรเปลี่ยน 3 ปัจจัย คือ ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน กับที่ไม่เป็นรูปแบบเดียวกัน (Uniform versus Nonuniform DIF) ข้อสอบทำหน้าที่เบี่ยงเบนที่มีทิศทางเดียวกัน กับสองทิศทางแบบสมดุล (Unidirectional versus Balanced-bidirectional DIF) และความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มอ้างอิงและกลุ่มสนใจ ซึ่งมี 4 กรณี คือ **กรณี A** กลุ่มอ้างอิงและกลุ่มสนใจมีการแจกแจงความสามารถ θ_1 และ θ_2 แบบปกติ 2 ตัวแปร มีค่าเฉลี่ยเท่ากับ 0 ความแปรปรวนเท่ากับ 1 และความสัมพันธ์ระหว่างความสามารถ θ_1 และ θ_2 เท่ากับ 0 ($\rho = 0$) **กรณี B** คล้ายกับกรณี A แต่ $\rho = .5$ ส่วนใน**กรณี C** คล้ายกับกรณี B แต่ค่าเฉลี่ยของความสามารถ θ_2 ในกลุ่มอ้างอิงสูงกว่ากลุ่มสนใจ .5 และ**กรณี D** ทั้งสองกลุ่มมีค่าเฉลี่ยของการแจกแจงความสามารถเท่ากับ 0 ความแปรปรวนเท่ากับ 1 แต่ในกลุ่มอ้างอิงมี $\rho = .5$ และกลุ่มสนใจมี $\rho = 0$ ส่วนปัจจัยอื่นกำหนดคงที่ ได้แก่ ความยาวของแบบทดสอบ

40 ข้อ จำนวนของ DIF 4 ข้อ และขนาดของ DIF โดยกำหนดความแตกต่างของเวกเตอร์พารามิเตอร์อำนาจจำแนก (**a**) และสเกลาร์พารามิเตอร์ (**b**) เท่ากับ .3 และ .5 ตามลำดับ ซึ่งค่าคงที่ดังกล่าวเพิ่มขึ้นหรือลดลงขึ้นอยู่กับเงื่อนไขค่าพารามิเตอร์ของข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มสนใจ การประมาณค่าพารามิเตอร์ของข้อสอบใช้โปรแกรม NOHARM แล้วใช้โปรแกรม IPLING ของลีและโอชิม่า (Lee ; & Oshima. 1996) เปรียบเทียบมาตรฐานพารามิเตอร์ของข้อสอบระหว่างกลุ่มให้อยู่บนเมตริกซ์เดียวกัน จากนั้นจะใช้โปรแกรม DFIT วิเคราะห์การทำหน้าที่เบี่ยงเบนของข้อสอบและแบบทดสอบ โดยคำนวณดัชนี CDIF, NCDIF และ DTF ภายใต้อำนาจ 2 เงื่อนไข คือ จำนวนจากพารามิเตอร์ของข้อสอบที่เป็นค่าจริง (True item parameters) และจำนวนจากพารามิเตอร์ของข้อสอบที่เป็นค่าประมาณ (Estimated item parameters)

ผลการวิเคราะห์ DFIT ใน 4 เงื่อนไข คือ (1) ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกันและมีทิศทางเดียวกัน (2) ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกันและมีสองทิศทางแบบสมดุล (3) ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกันและมีทิศทางเดียวกัน และ (4) ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกันและมีสองทิศทางแบบสมดุล พบว่าพารามิเตอร์ของข้อสอบที่เป็นค่าจริงและค่าประมาณให้ผลสอดคล้องกันในทุกเงื่อนไข กล่าวคือ การตรวจสอบจากพารามิเตอร์ของข้อสอบที่เป็นค่าจริงทุกเงื่อนไขของการตรวจสอบ No-DIF (ข้อ 1-36) โดยใช้ดัชนี CDIF และ NCDIF ไม่พบความผิดพลาดเชิงบวก (False positive) ซึ่งเป็นไปตามที่คาดไว้ ส่วนการตรวจสอบ DIF (ข้อ 37-40) เมื่อใช้ดัชนี CDIF สามารถตรวจพบข้อสอบ DIF ในเงื่อนไข 1 มากกว่าเงื่อนไข 3 โดยในเงื่อนไข 1 สามารถตรวจพบได้ทั้งหมด ส่วนในเงื่อนไข 3 ตรวจพบได้เพียงข้อเดียว สำหรับในเงื่อนไข 2 และ 4 ไม่สามารถตรวจพบได้ทั้งหมด ซึ่งเป็นไปตามที่คาดหวัง เมื่อตรวจสอบด้วยดัชนี NCDIF ในเงื่อนไข 1 และ 2 สามารถตรวจพบ DIF ได้ทั้งหมด แต่ในเงื่อนไข 3 และ 4 สามารถตรวจพบได้ 2 ข้อเท่านั้น ส่วนผลการตรวจสอบจากพารามิเตอร์ของข้อสอบที่เป็นค่าประมาณ เมื่อตรวจสอบด้วยดัชนี CDIF และ NCDIF ในเงื่อนไขของ No-DIF ไม่พบความผิดพลาดเชิงบวก ยกเว้นในกรณี A โดยใช้ดัชนี NCDIF ส่วนการตรวจสอบในเงื่อนไขของ DIF โดยใช้ดัชนี CDIF สามารถตรวจพบในเงื่อนไข 1 กรณี A-D ได้ 2 ถึง 3 ข้อ จากทั้งหมด 4 ข้อ ส่วนในเงื่อนไข 2 ไม่สามารถตรวจพบ ยกเว้นกรณี B ที่ตรวจสอบผิดพลาด โดยตรวจพบ 1 ข้อ ส่วนในเงื่อนไข 3 และ 4 ตรวจพบ 2 ข้อกรณี A และ C (กรณีละ 1 ข้อ) ส่วนการตรวจสอบโดยใช้ดัชนี NCDIF สามารถตรวจพบกรณี A-D ในเงื่อนไข 1 ได้ 75% และตรวจพบในเงื่อนไข 2 ได้เกือบทั้งหมด ส่วนในเงื่อนไข 3 ให้ผลสอดคล้องกับการวิเคราะห์จากพารามิเตอร์จริง สำหรับเงื่อนไข 4 สามารถตรวจพบได้ในช่วง 75-100%

เฟลาเวอร์ โอชิมา และราชู (Flowers; Oshima; & Raju. 1999: 309-326) ได้ปรับขยายวิธี DFIT (Differential item/test functioning) ที่ใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ในข้อสอบที่ให้คะแนนสองค่า ซึ่งพัฒนาโดย ราชู แวนเดอร์ลินเด็น และเฟลีย์ร์ (Raju; van der Linden; & Fleer. 1995) เพื่อนำมาใช้ตรวจสอบในข้อสอบที่ให้คะแนนหลายค่า ข้อมูลที่ใช้ในการศึกษาเป็นข้อมูลจำลองภายใต้โมเดลเกรดเรสปอน (Graded Response Model; GRM) โดยนำค่าพารามิเตอร์จากการศึกษาของโคเฮ็นและคิม (Cohen; & Kim. 1993) และเฟลีย์ร์ (Fleer. 1993) มาปรับใหม่เพื่อให้เหมาะสมกับโมเดล GRM แล้วจำลองผลการตอบข้อสอบที่มี 5 รายการตอบ ข้อสอบทุกข้อให้คะแนน 5 ระดับ (0, 1, 2, 3 และ 4) และขนาดตัวอย่างกลุ่มละ 1,000 คน ภายใต้เงื่อนไขที่แปรเปลี่ยน 4 ปัจจัย คือ (1) ความยาวของแบบทดสอบ 2 ขนาด คือ 20 ข้อและ 40 ข้อ (2) ความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มอ้างอิงและกลุ่มสนใจ 2 ระดับ คือ แบบเท่ากันและแบบไม่เท่ากัน โดยแบบเท่ากันทั้ง 2 กลุ่มมีการแจกแจง $N(0, 1)$ ส่วนแบบไม่เท่ากันกลุ่มอ้างอิงมีการแจกแจง $N(0, 1)$ แต่กลุ่มสนใจมีการแจกแจง $N(-1, 1)$ (3) สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 4 ขนาด คือ 0%, 5%, 10% และ 20% (4) ทิศทางของข้อสอบทำหน้าที่เบี่ยงเบน 2 ลักษณะ คือ ข้อสอบทำหน้าที่เบี่ยงเบนที่มีทิศทางเดียวกัน (Unidirectional DIF) ซึ่งเข้าข้างกลุ่มอ้างอิงทุกข้อ และข้อสอบทำหน้าที่เบี่ยงเบนที่มีสองทิศทาง (Bidirectional DIF) ซึ่งเข้าข้างกลุ่มอ้างอิงและกลุ่มสนใจจำนวนเท่ากัน ยกเว้นในเงื่อนไข 5% ของแบบทดสอบขนาด 10 ข้อ ไม่มีการจำลองข้อสอบดังกล่าว และ (5) รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน 2 รูปแบบ คือ ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF) และที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) การจำลองรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนในเงื่อนไขสัดส่วนข้อสอบทำหน้าที่เบี่ยงเบน 20% จำลองข้อสอบ Uniform DIF และ Nonuniform DIF แต่ในเงื่อนไขอื่นๆ จำลองเฉพาะ Uniform DIF เท่านั้น รวมข้อมูลที่จำลองทั้งหมด 26 เงื่อนไข ในแต่ละเงื่อนไขจำลองซ้ำ 5 ครั้ง การประมาณค่าพารามิเตอร์ของข้อสอบและผู้สอบใช้โปรแกรม PASCAL ภายใต้โมเดล GRM แล้วใช้โปรแกรม EQUATE ปรับเทียบมาตรฐานประมาณพารามิเตอร์ของกลุ่มอ้างอิงให้อยู่บนเมทริกซ์ของกลุ่มสนใจ การคำนวณดัชนี DFIT ใช้โปรแกรม FORTRAN สำหรับดัชนี CDIF และ NCDIF จะคำนวณความถูกต้องเชิงบวก (True positive; TP) และความผิดพลาดเชิงบวก (False positive; FP) แล้วเปรียบเทียบภายใต้ 2 เงื่อนไข คือ เงื่อนไขจริง (True conditions) ซึ่งเป็นค่าพารามิเตอร์จริง และเงื่อนไขค่าประมาณ (Estimated conditions) ซึ่งเป็นค่าพารามิเตอร์จากการประมาณ

ผลการศึกษาพบว่า วิธี DFIT มีประสิทธิภาพในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (DIF) และแบบทดสอบ (DTF) ในข้อสอบที่ให้คะแนนหลายค่า ซึ่งเป็นข้อมูลจากการจำลองสำหรับความยาวของแบบทดสอบ ความแตกต่างของการแจกแจงความสามารถ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน และทิศทางของข้อสอบทำหน้าที่เบี่ยงเบนมีอิทธิพลต่อความถูกต้องเชิงบวกและความผิดพลาดเชิงบวกน้อยมาก ในทุกเงื่อนไขของการตรวจสอบ และเป็นไปตามที่คาดไว้เกี่ยวกับข้อสอบที่มี DIF จำนวนมากจะถูกตรวจพบมากกว่าข้อสอบที่มี DIF จำนวนน้อย ส่วนผลการตรวจสอบด้วยดัชนี CDIF ไม่คงที่เหมือนกับดัชนี NCDF ในทุกเงื่อนไข ซึ่งเหมือนกับการตรวจสอบในกรณีมิติเดียว (Fleer. 1993) และในกรณีหลายมิติ (Oshima; Raju; & Flowers. 1997) สำหรับการศึกษาครั้งนี้ ดัชนี CDIF ที่ตรวจสอบใน 2 เงื่อนไขมีความผันแปรไปจากที่คาดไว้ คือ เงื่อนไขความยาวของแบบทดสอบ 20 ข้อและการแจกแจงความสามารถแบบเท่ากัน มีความผิดพลาดในการตรวจสอบ 18% และในเงื่อนไขความยาวของแบบทดสอบ 50 ข้อและการแจกแจงความสามารถแบบไม่เท่ากันสามารถตรวจสอบได้ถูกต้องเพียง 50%

เกียร์ล กือทซ์แมนน์ และบอทตัน (Gierl; Gotzmann; & Boughton. 2004: 241-264) ได้ศึกษาผลของวิธีซีปเทสท์ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ เมื่อร้อยละของข้อสอบทำหน้าที่เบี่ยงเบนมีขนาดใหญ่ ข้อมูลที่ใช้ในการศึกษาเป็นข้อมูลจำลอง โดยจำลองผลการตอบข้อสอบจำนวน 40 ข้อ ภายใต้การจัดกระทำ 4 ตัวแปร ได้แก่ **ตัวแปร 1** สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 3 ขนาด คือ 20%, 30% และ 40% (8, 16 และ 24 ข้อตามลำดับ) และในแต่ละระดับดังกล่าวแบ่งออกเป็น 2 เงื่อนไข คือ 75/25 และ 50/50 ในเงื่อนไขแรก 75% ของข้อสอบมี DIF ขนาดปานกลาง และ 25% ของข้อสอบมี DIF ขนาดใหญ่ ในเงื่อนไขที่สอง 50% ของข้อสอบมี DIF ขนาดปานกลาง และ 50% ของข้อสอบมี DIF ขนาดใหญ่ ซึ่ง DIF ขนาดปานกลางมีค่าอยู่ในช่วง $.05 \leq |\hat{\beta}_{uni}| < .10$ และ DIF ขนาดใหญ่มีค่าอยู่ในช่วง $|\hat{\beta}_{uni}| \geq .10$ **ตัวแปร 2** ทิศทางของข้อสอบทำหน้าที่เบี่ยงเบน (Direction of DIF) 2 ลักษณะ คือ แบบสมดุล (Balanced) และแบบไม่สมดุล (Unbalanced) โดยแบบสมดุลมีข้อสอบ DIF เข้าข้างกลุ่มอ้างอิงและกลุ่มสนใจจำนวนเท่ากัน ส่วนแบบไม่สมดุลมีข้อสอบ DIF เข้าข้างกลุ่มอ้างอิงและกลุ่มสนใจจำนวนไม่เท่ากัน กล่าวคือ ในเงื่อนไข DIF 20%, 40% และ 60% เข้าข้างกลุ่มอ้างอิงจำนวน 6, 12 และ 18 ข้อตามลำดับ และเข้าข้างกลุ่มสนใจจำนวน 2, 4 และ 6 ข้อ ตามลำดับ **ตัวแปร 3** ขนาดตัวอย่าง 4 ขนาด คือ 500, 1,000, 1,500 และ 2,000 คนต่อ (กลุ่มอ้างอิงและกลุ่มสนใจ) และ **ตัวแปร 4**

ความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่ม 2 ระดับ คือ แบบเท่ากันและแบบไม่เท่ากัน ในกรณีเท่ากันทั้งสองกลุ่มมีการแจกแจงความสามารถแบบปกติ $N(0, 1)$ ส่วนในกรณีไม่เท่ากันกลุ่มอ้างอิงและกลุ่มสนใจมีการแจกแจงความสามารถ $N(0.025, 1)$ และ $N(-0.025, 1)$ ตามลำดับ รวมข้อมูลที่จำลองทั้งหมด 96 เงื่อนไข ($6 \times 2 \times 4 \times 2$) ในแต่ละเงื่อนไขจำลองซ้ำ 100 ครั้ง โดยใช้โปรแกรม DIFSIM จำลองผลการตอบภายใต้ 2 เงื่อนไข คือ **ในเงื่อนไข No-DIF** ใช้พารามิเตอร์ของข้อสอบจากการศึกษาของดราสโกว์ (Drasgow, 1987) จำนวน 32 ข้อ จำลองข้อสอบวัดเฉพาะความสามารถในมิติหลัก และให้ค่าพารามิเตอร์ของกลุ่มอ้างอิงกับกลุ่มสนใจเท่ากัน โดยใช้โมเดลโลจิสติก 3 พารามิเตอร์แบบมิติเดียว (Unidimensional 3-parameter logistic) ส่วน**ในเงื่อนไข DIF** ใช้พารามิเตอร์จากการศึกษาของเนนดากุมาร์ (Nandakumar, 1993) จำนวน 24 ข้อ จำลองข้อสอบวัดทั้งความสามารถในมิติหลักและมิติรอง ภายใต้โมเดลโลจิสติก 3 พารามิเตอร์แบบหลายมิติที่มีการชดเชย (Compensatory multidimensional) และกำหนดให้ค่าสหสัมพันธ์ระหว่างมิติในแต่ละกลุ่มเท่ากับ .5 ส่วนความแปรปรวนภายในมิติของแต่ละกลุ่มเท่ากับ 1 แล้วคำนวณอัตราการตรวจสอบ DIF 2 ชนิด คือ อัตราการตัดสินใจไม่ถูกต้อง (Incorrect decisions) และอัตราการตัดสินใจถูกต้อง (Correct decisions)

ผลการศึกษาพบว่า **ในเงื่อนไขสัดส่วน DIF 75/25 (Unbalanced DIF)** เมื่อข้อสอบทำหน้าที่เบี่ยงเบนจำนวนเล็กน้อย คือ 20% ประกอบด้วยข้อสอบเข้าข้างกลุ่มอ้างอิง 6 ข้อ และเข้าข้างกลุ่มสนใจ 2 ข้อ ไม่มีผลต่ออัตราการตรวจสอบ เมื่อสัดส่วนของการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ คือ 40% และ 60% อัตราการตัดสินใจไม่ถูกต้องลดลง แต่อัตราการตัดสินใจถูกต้องเพิ่มขึ้น โดยเพิ่มขึ้นสูงสุดในเงื่อนไขของขนาดตัวอย่างมากที่สุด สำหรับ**ในเงื่อนไขสัดส่วน DIF 50/50 (Balanced DIF)** เมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเล็กน้อย คือ 20% ให้ผลการตรวจสอบเช่นเดียวกัน และเมื่อสัดส่วนของการทำหน้าที่เบี่ยงเบนขนาดใหญ่ คือ 40% และ 60% อัตราการตัดสินใจไม่ถูกต้องลดลงในขณะที่ขนาดตัวอย่างเพิ่มขึ้น อย่างไรก็ตาม ในเงื่อนไข Unbalanced DIF เมื่อขนาดตัวอย่าง 500 คนต่อกลุ่ม แล้วอัตราการตัดสินใจไม่ถูกต้องมีค่าเกิน 10% ถึง 60% อัตราของการตัดสินใจถูกต้องมีแนวโน้มเพิ่มขึ้นเมื่อขนาดตัวอย่างเพิ่มขึ้นในเงื่อนไข Balanced DIF แต่ในในเงื่อนไข Unbalanced DIF มีค่าไม่แน่นอน อาจเพิ่มขึ้นหรือลดลงก็ได้ เมื่อเปรียบเทียบความแตกต่างระหว่างอัตราการตัดสินใจไม่ถูกต้องและอัตราการตัดสินใจถูกต้องที่เกิดขึ้นในเงื่อนไข Balanced DIF และ Unbalanced DIF พบว่า อัตราการตัดสินใจไม่ถูกต้องมีความคงที่สูงกว่า และอัตราการตัดสินใจถูกต้องมีความคงที่ต่ำกว่า เมื่อตรวจสอบในเงื่อนไข Unbalanced DIF

5.2 งานวิจัยภายในประเทศ

นพมาศ พิพัฒน์สุข (2541) ได้เปรียบเทียบประสิทธิภาพระหว่างวิธีแมนเทิล-แฮนส์เซลกับวิธีการถดถอยโลจิสติก ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อใช้เกณฑ์จับคู่เปรียบเทียบแตกต่างกันในแบบทดสอบหลายมิติ โดยใช้เกณฑ์การจับคู่ 3 เกณฑ์ คือ คะแนนรวมของแบบทดสอบ (Total test score) คะแนนแบบทดสอบชุดย่อย (Subtest score) และคะแนนแบบทดสอบหลายชุดย่อย (Multiple subtest scores) เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูลเป็นแบบทดสอบเลือกตอบชนิด 4 ตัวเลือก ประกอบด้วยแบบทดสอบชุดย่อย 2 ฉบับๆ ละ 40 ข้อ ซึ่งแต่ละฉบับวัดองค์ประกอบ 4 ด้าน คือ ความรู้ความเข้าใจทางคณิตศาสตร์ ความสามารถด้านการคิดคำนวณ ความสามารถในการพิจารณาผลลัพธ์อย่างสมเหตุสมผล ความสามารถด้านการแก้ปัญหา กลุ่มตัวอย่างเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 สังกัดสำนักงานการศึกษากรุงเทพมหานคร จำนวน 1,076 คน การวิเคราะห์ข้อมูลประกอบด้วยการวิเคราะห์องค์ประกอบเชิงยืนยันเพื่อตรวจสอบความเที่ยงตรงเชิงโครงสร้างโดยใช้โปรแกรม LISREL การวิเคราะห์ด้วยวิธีชิปเทสต์ (เป็นวิธีเกณฑ์สำหรับการเปรียบเทียบ) และวิธีแมนเทิล-แฮนส์เซล ภายใต้เกณฑ์คะแนนรวมและคะแนนแบบทดสอบชุดย่อยใช้โปรแกรม SIBTEST ส่วนวิธีการถดถอยโลจิสติกภายใต้คะแนนรวม คะแนนแบบทดสอบชุดย่อย และคะแนนแบบทดสอบหลายชุดย่อยใช้โปรแกรม SPSS แล้วคำนวณอัตราความถูกต้องและอัตราความคลาดเคลื่อนของวิธีแมนเทิล-แฮนส์เซลและวิธีการถดถอยโลจิสติก จากนั้นทดสอบความแตกต่างโดยใช้สถิติ Z

ผลการศึกษาพบว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทิล-แฮนส์เซลเมื่อใช้คะแนนรวมและคะแนนแบบทดสอบชุดย่อยเป็นเกณฑ์การจับคู่เปรียบเทียบสามารถตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนจำนวน 20% และ 18.67% ตามลำดับ ส่วนวิธีการถดถอยโลจิสติก เมื่อใช้คะแนนรวม คะแนนแบบทดสอบชุดย่อย และคะแนนแบบทดสอบหลายชุดย่อยเป็นเกณฑ์การจับคู่สามารถตรวจพบข้อสอบทำหน้าที่ต่างกัน จำนวน 26.67%, 22.67 % และ 17.33% ตามลำดับ ภายใต้เงื่อนไขเกณฑ์การจับคู่โดยใช้คะแนนรวมวิธีแมนเทิล-แฮนส์เซลมีประสิทธิภาพสูงกว่าวิธีการถดถอยโลจิสติก และเมื่อใช้คะแนนแบบทดสอบชุดย่อยทั้งสองวิธีมีประสิทธิภาพไม่แตกต่างกัน สำหรับวิธีการถดถอยโลจิสติกมีความเหมาะสมในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบหลายมิติ เมื่อใช้คะแนนแบบทดสอบหลายชุดย่อยเป็นเกณฑ์การจับคู่ความสามารถ

วลีมาศ แซ่อึ้ง (2543) ได้เปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป (ที่ไม่เป็นรูปแบบเดียวกัน) ระหว่างวิธีชิปเทสท์ปรับใหม่ วิธีชิปเทสท์ วิธีแมนเทล-แฮนส์เชล และวิธีการถดถอยโลจิสติก ข้อมูลที่ใช้ศึกษาเป็นข้อมูลจำลอง โดยใช้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ชนิดกำหนดค่าการเดาคงที่ (3PLM-c) จำลองผลการตอบข้อสอบภายใต้การจัดกระทำตัวแปรอิสระ 4 ตัว คือ ลักษณะของข้อสอบ 9 ระดับ (a ต่ำกับ b ต่ำ, a ต่ำกับ b ปานกลาง, a ต่ำกับ b สูง, a ปานกลางกับ b ต่ำ, a ปานกลางกับ b ปานกลาง, a ปานกลางกับ b สูง, a สูงกับ b ต่ำ, a สูงกับ b ปานกลาง และ a สูงกับ b สูง) ความยาวของแบบทดสอบ 2 ขนาด (30 ข้อ และ 60 ข้อ) สัดส่วนของข้อสอบทำหน้าที่ต่างกัน 3 ขนาด (5%, 10% และ 20%) และขนาดตัวอย่าง 6 ขนาด (จำนวนผู้สอบกลุ่มอ้างอิงต่อกลุ่มสนใจเท่ากับ 250 คนต่อ 250 คน, 500 คนต่อ 250 คน, 500 คนต่อ 500 คน, 1000 คนต่อ 250 คน, 1000 คนต่อ 500 คน และ 1000 คนต่อ 1000 คน) รวมข้อมูลที่จำลองทั้งหมด 324 เงื่อนไข ($9 \times 2 \times 3 \times 6$) โดยใช้โปรแกรม IRTDATA ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบมี 5 วิธี คือ วิธีการวัดพื้นที่ของราชู (เป็นวิธีเกณฑ์สำหรับการเปรียบเทียบ) ใช้โปรแกรม IRTDIF วิธีชิปเทสท์ปรับใหม่และวิธีชิปเทสท์ที่ใช้โปรแกรม SIBTEST วิธีแมนเทล-แฮนส์เชลและวิธีการถดถอยโลจิสติกใช้โปรแกรม MHDIF และ SPSS ตามลำดับ แล้วคำนวณอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธี ส่วนการเปรียบเทียบความแตกต่างใช้สถิติ ANOVA และ t-test

ผลการวิจัยพบว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป อำนาจการทดสอบของวิธีชิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าเท่าเทียมกันภายใต้เกือบทุกเงื่อนไข และทั้งสองวิธีมีอำนาจการทดสอบสูงกว่าวิธีชิปเทสท์และวิธีแมนเทล-แฮนส์เชลภายใต้เกือบทุกเงื่อนไข ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีชิปเทสท์ปรับใหม่ วิธีชิปเทสท์วิธีแมนเทล-แฮนส์เชล และวิธีการถดถอยโลจิสติกมีค่าอยู่ในเกณฑ์ที่ระดับ 10% ภายใต้เกือบทุกเงื่อนไข ในการทดสอบอิทธิพลของปัจจัย พบว่า อำนาจการทดสอบของวิธีชิปเทสท์ปรับใหม่วิธีแมนเทล-แฮนส์เชล และวิธีการถดถอยโลจิสติกมีค่าเพิ่มขึ้นในทุกเงื่อนไขของลักษณะข้อสอบที่มีค่าความยากปานกลาง ปัจจัยสัดส่วนของข้อสอบทำหน้าที่ต่างกันไม่มีผลต่ออำนาจการทดสอบของวิธีชิปเทสท์ปรับใหม่และวิธีชิปเทสท์ แต่มีผลต่อวิธีแมนเทล-แฮนส์เชลและวิธีการถดถอยโลจิสติก สำหรับปัจจัยขนาดตัวอย่างมีผลต่ออำนาจการทดสอบของทุกวิธี เมื่อขนาดตัวอย่างเพิ่มขึ้นแล้วอำนาจการทดสอบเพิ่มขึ้น ส่วนปัจจัยความยาวของแบบทดสอบไม่มีผลต่ออำนาจการทดสอบของทุกวิธี

ทองอยู่ สาระ (2543) ได้เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบแบบสมำเสมอ และแบบไม่สมำเสมอ ระหว่างวิธีแมนเทิล-แฮนส์เชลและวิธีการถดถอยโลจิสติก โดยใช้ความยาวของแบบทดสอบและขนาดตัวอย่างต่างกัน กลุ่มตัวอย่างที่ใช้ในการศึกษาเป็นนักเรียนระดับมัธยมศึกษาปีที่ 3 โรงเรียนสังกัดกรมสามัญศึกษา ส่วนกลาง กลุ่มที่ 3 จำนวน 3,242 คน เครื่องมือที่ใช้เป็นแบบทดสอบวัดความสามารถทางสมองตามแนวโครงสร้างของโอติส-เลนนอน ซึ่งเป็นแบบเลือกตอบชนิด 5 ตัวเลือก จำนวน 80 ข้อ วัดความสามารถทั่วไปสามด้าน คือ ความเข้าใจด้านภาษา เหตุผลด้านภาษา และเหตุผลด้านภาพ ในการศึกษาครั้งนี้ ผู้วิจัยได้สุ่มแบบทดสอบที่มีความยาว 3 ขนาด คือ 20, 40 และ 60 ข้อ และสุ่มขนาดตัวอย่างกลุ่มอ้างอิงต่อกลุ่มสนใจ 10 ขนาด คือ 300 : 100, 300 : 200, 300 : 300, 500 : 100, 500 : 200, 500 : 300, 1,000 : 100, 1,000 : 200, 1,000 : 300 และ 1,000 : 1,000 รวมเงื่อนไขทั้งหมด 30 เงื่อนไข (3×10)

ผลการศึกษาพบว่า ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบแบบสมำเสมอ และแบบไม่สมำเสมอ ภายใต้ปัจจัยความยาวของแบบทดสอบและขนาดตัวอย่าง วิธีแมนเทิล-แฮนส์เชลและวิธีการถดถอยโลจิสติกมีอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่แตกต่างกัน เกือบทุกเงื่อนไข ปัจจัยความยาวของแบบทดสอบไม่มีผลต่ออัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบทั้งสอง ส่วนปัจจัยขนาดตัวอย่างมีผลต่ออัตราความถูกต้องของทั้งสองวิธีเกือบทุกเงื่อนไข กล่าวคือ เมื่อขนาดตัวอย่างเพิ่มขึ้นมีผลทำให้อัตราความถูกต้องเพิ่มขึ้น ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ขนาดตัวอย่างไม่มีผลต่อทั้งสองวิธีเกือบทุกเงื่อนไข

สิริรัตน์ วิภาสศิลป์ (2545) ได้เปรียบเทียบวิธีชิปเทสต์และวิธีดีเอฟไอทีในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบ จากข้อมูลการตอบข้อสอบที่ใช้ความสามารถหลายมิติ ภายใต้ปัจจัยที่ศึกษา 2 ปัจจัย คือ (1) ความยาวของแบบทดสอบ 3 ขนาด คือ 30, 40 และ 50 ข้อ และ (2) ขนาดตัวอย่าง 5 ขนาด คือ 50, 100, 200, 500 และ 1,000 คน กลุ่มตัวอย่างที่ใช้ในการศึกษาได้มาจากการสุ่มแบบใส่คืนจากประชากรเทียม ซึ่งกำหนดจากนักเรียนชายและนักเรียนหญิง ชั้นมัธยมศึกษาปีที่ 1 จังหวัดนนทบุรี โดยสุ่มแต่ละขนาด 50 ครั้ง เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1 ที่ผู้วิจัยสร้างขึ้น ประกอบด้วยข้อสอบแบบเลือกตอบชนิด 5 ตัวเลือก จำนวน 50 ข้อ โดยมุ่งวัดความสามารถหลาย

มิติ โดยให้ความสามารถทางคณิตศาสตร์เป็นความสามารถหลัก ส่วนความสามารถด้านอื่นๆ เป็นความสามารถรอง เช่น ความสามารถในการอ่าน การแก้โจทย์ปัญหาที่ซับซ้อน ไหวพริบในการทำข้อสอบ เป็นต้น แล้วนำข้อสอบดังกล่าวให้ผู้เชี่ยวชาญตรวจสอบ ผลปรากฏว่า เป็นข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบนของข้อสอบต่อเพศชาย จำนวน 16 ข้อ จากนั้นจึงเก็บรวบรวมข้อมูลแล้วคัดเลือกข้อสอบตามสัดส่วนในตารางกำหนดข้อสอบเพื่อจัดเป็นแบบทดสอบที่มีความยาว 40 และ 30 ข้อ ในการวิเคราะห์มิติของแบบทดสอบจะวิเคราะห์องค์ประกอบเพื่อพิจารณาค่าไอเกนโดยใช้โปรแกรม SPSS ส่วนการประมาณค่าพารามิเตอร์ของข้อสอบภายใต้โมเดลแบบ 2 พารามิเตอร์ (M2PL) ใช้โปรแกรม NOHARM การปรับเทียบมาตรใช้โปรแกรม IPLINK การประมาณค่าความสามารถหลักใช้โปรแกรม BILOG สำหรับการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบใช้โปรแกรม SIBTEST และ DFIT แล้วนำผลการตรวจสอบไปวิเคราะห์อัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบด้วยวิธีเดียวกันและวิธีต่างกัน จากนั้นจึงเปรียบเทียบความแตกต่างโดยใช้สถิติ MANOVA และ Z-test

ผลการศึกษาในเงื่อนไขของข้อสอบพบว่า ปัจจัยความยาวของแบบทดสอบ 30, 40 และ 50 ข้อ และปัจจัยขนาดตัวอย่าง 50, 100 และ 200 คน ไม่มีผลต่ออัตราความถูกต้องของวิธีชิปเทสต์ ส่วนอัตราความถูกต้องของวิธีชิปเทสต์ เมื่อขนาดตัวอย่าง 500 และ 1,000 คน มีค่าสูงกว่าการตรวจสอบในขนาดตัวอย่าง 50, 100 และ 200 คน ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ก็มีค่าสูงกว่าด้วย สำหรับการตรวจสอบด้วยวิธีดีเอฟไอทีพบว่า ปัจจัยขนาดตัวอย่างทั้ง 5 ขนาด ไม่มีต่ออัตราความถูกต้องของวิธีดีเอฟไอที เมื่อเปรียบเทียบระหว่างวิธีชิปเทสต์และวิธีดีเอฟไอทีพบว่า อัตราความถูกต้องของวิธีชิปเทสต์มากกว่าวิธีดีเอฟไอทีในทุกเงื่อนไข และความสอดคล้องในการตรวจสอบด้วยวิธีทั้งสองมีค่าต่ำกว่าร้อยละ 10 ส่วนผลการศึกษาในเงื่อนไขของหมวดข้อสอบพบว่า อัตราความถูกต้องของวิธีชิปเทสต์มากกว่าวิธีดีเอฟไอทีใน 2 เงื่อนไข คือ ความยาวของแบบทดสอบ 30 ข้อและขนาดตัวอย่าง 1,000 คน กับความยาวของแบบทดสอบ 40 ข้อและขนาดตัวอย่าง 500 คน ส่วนผลการศึกษาในเงื่อนไขของแบบทดสอบพบว่า อัตราความถูกต้องของวิธีชิปเทสต์มากกว่าวิธีดีเอฟไอทีเมื่อความยาวของแบบทดสอบ 50 ข้อและขนาดตัวอย่าง 100, 200 และ 1,000 คน

อุทัยวรรณ สายพัฒนา (2547) ได้เปรียบเทียบประสิทธิภาพของผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ในแบบทดสอบที่มีการให้คะแนนหลายค่า ระหว่างวิธีแมนเทิล-แฮนส์เซลทั่วไป (GMH) และวิธีโพลี-ซิปเทสต์ (Poly-SIBTEST) โดยศึกษาภายใต้ปัจจัยที่แปรเปลี่ยน 2 ปัจจัย คือ (1) ความยาวของแบบทดสอบ 3 ขนาด คือ 20, 30 และ 40 ข้อ และ (2) ขนาดตัวอย่าง 3 ขนาด คือ 250, 500 และ 1,000 คน สำหรับกลุ่มตัวอย่างที่ใช้ในการศึกษาได้มาจากการสุ่มแบบใส่คืนจากประชากรเทียม ซึ่งกำหนดจากนักเรียนชายและนักเรียนหญิง ชั้นมัธยมศึกษาปีที่ 3 ในสังกัดกรมสามัญศึกษา จังหวัดพระนครศรีอยุธยา แต่ละขนาดตัวอย่างสุ่มจำนวน 10 ครั้ง เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ที่ผู้วิจัยสร้างขึ้นประกอบด้วยข้อสอบแบบเลือกตอบ ชนิด 5 ตัวเลือก จำนวน 40 ข้อ โดยผู้เชี่ยวชาญเป็นผู้พิจารณาการทำหน้าที่เบี่ยงเบนของข้อสอบ พบว่า เป็นข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบนโดยเมื่อต่อเพศชายจำนวน 8 ข้อ หลังจากเก็บรวบรวมข้อมูลแล้วคัดเลือกข้อสอบตามสัดส่วนในตารางกำหนดข้อสอบเพื่อจัดกระทำเป็นแบบทดสอบที่มีความยาว 30 และ 20 ข้อ ต่อจากนั้นจะตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบภายใต้เงื่อนไขต่างๆ โดยใช้วิธีแมนเทิล-แฮนส์เซลทั่วไป (GMH) และ วิธีโพลี-ซิปเทสต์ (Poly-SIBTEST) ส่วนวิธีพหุคูณเกรดทั่วไป (GPCM) ใช้เป็นเกณฑ์การเปรียบเทียบกับ 2 วิธีดังกล่าว ในการวิเคราะห์ข้อมูลประกอบด้วยการตรวจสอบความเป็นมิติเดียวของแบบทดสอบ โดยใช้โปรแกรม SPSS การประมาณค่าพารามิเตอร์ของข้อสอบภายใต้โมเดล GPCM ใช้โปรแกรม PARSCALE การวิเคราะห์วิธี GPCM (ไค-กำลังสองของลอร์ด) และวิธี GMH ใช้โปรแกรม MATLAB การวิเคราะห์วิธี Poly-SIBTEST ใช้โปรแกรม DIFPACK แล้วเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธี GMH และวิธี Poly-SIBTEST โดยใช้สถิติ MANOVA

ผลการวิจัยพบว่า ภายใต้ปัจจัยความยาวของแบบทดสอบและขนาดตัวอย่างส่งผลต่ออัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธี GMH ยกเว้นในเงื่อนไขความยาวของแบบทดสอบ 30 และ 20 ข้อ ขนาดตัวอย่าง 500 และ 250 คน ส่วนการตรวจสอบด้วยวิธี Poly-SIBTEST พบว่า ปัจจัยความยาวของแบบทดสอบและขนาดตัวอย่างมีผลต่ออัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ยกเว้นในเงื่อนไขความยาวของแบบทดสอบ 20 ข้อ และขนาดตัวอย่าง 1,000 และ 500 คน สำหรับการเปรียบเทียบประสิทธิภาพของสองวิธีพบว่า ในทุกเงื่อนไขของความยาวของแบบทดสอบและขนาดตัวอย่าง วิธี GMH และวิธี Poly-SIBTEST มีประสิทธิภาพในการตรวจสอบไม่แตกต่างกัน ยกเว้นในเงื่อนไขของขนาดตัวอย่าง 1,000 คน และความยาวของแบบทดสอบ 20 ข้อ วิธี Poly-SIBTEST มีประสิทธิภาพสูงกว่าวิธี GMH

จากการศึกษางานวิจัยต่างประเทศและภายในประเทศ ที่เกี่ยวข้องกับการตรวจสอบ การทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีโพลี-ชิปเทสต์และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ภายใต้เงื่อนไขของปัจจัยต่างๆ ซึ่งส่วนมากมักศึกษาภายใต้เงื่อนไขปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่าง สามารถสรุปได้ดังนี้

รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน (Form of DIF)

จากการศึกษางานวิจัยของทองอยู่ สารระ (2543) วลีมาศ แซ่ฉิ่ง (2543) นารายานานและสวามินาทาน (Narayanan; & Swaminathan. 1994, 1996) โรเจอร์สและสวามินาทาน (Rogers; & Swaminathan. 1993) ได้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถมิติเดียว และให้คะแนนสองค่า ผลการศึกษาสรุปรูปได้ว่า วิธีชิปเทสต์และวิธีแมนเทิล-แฮนส์เซลสามารถตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกัน แต่ไม่สามารถตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ไม่เป็นรูปแบบเดียวกัน ส่วนวิธีการถดถอยโลจิสติกสามารถตรวจสอบได้ทั้งที่เป็นรูปแบบเดียวกันและที่ไม่เป็นรูปแบบเดียวกัน จากการศึกษางานนักวิจัยของ ชาง มาซซีโอ และรูสโซ (Chang; Mazzeo; & Roussos. 1996) เฟรนช์และมิลเลอร์ (French; & Miller. 1996) ได้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถมิติเดียวและให้คะแนนหลายค่า ผลการศึกษาสรุปรูปได้ว่า วิธีโพลี-ชิปเทสต์ วิธีแมนเทิล วิธีแมนเทิล-แฮนส์เซลทั่วไป และวิธีความแตกต่างเฉลี่ยมาตรฐานให้ผลการตรวจสอบสอดคล้องกับวิธีชิปเทสต์และวิธีแมนเทิล-แฮนส์เซล ส่วนวิธีการถดถอยโลจิสติกให้ผลสอดคล้องกับการตรวจสอบในข้อสอบที่ให้คะแนนสองค่า เช่นเดียวกัน สาเหตุดังกล่าวอาจเนื่องมาจากวิธีชิปเทสต์และวิธีแมนเทิล-แฮนส์เซลใช้สถิติแบบคิดเครื่องหมาย (เครื่องหมายบวกและเครื่องหมายลบ) เมื่อข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกันในข้อสอบที่ให้คะแนนสองค่า ซึ่งสามารถพิจารณาได้จากฟังก์ชันการตอบข้อสอบ (Item response functions; IRFs) ระหว่างกลุ่มตัดกันตรงช่วงกึ่งกลางของความสามารถ ทำให้ทิศทางของข้อสอบทำหน้าที่เบี่ยงเบนเปลี่ยนแปลง ณ ตรงช่วงกึ่งกลางของความสามารถ ซึ่งมีผลทำให้ความแตกต่างของคะแนนส่วนหนึ่งที่เป็นค่าลบหักล้างกับคะแนนอีกส่วนหนึ่งที่เป็นค่าบวก ดังนั้นวิธีชิปเทสต์และวิธีแมนเทิล-แฮนส์เซลจึงไม่สามารถตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนที่มีรูปแบบดังกล่าว ส่วนวิธีการถดถอยโลจิสติกใช้โมเดลทดสอบอิทธิพลของปฏิสัมพันธ์ระหว่างระดับความสามารถ กับการเป็นสมาชิกของกลุ่มผู้สอบ ทำให้สามารถตรวจพบข้อสอบทำหน้าที่เบี่ยงเบน

ทั้งที่เป็นรูปแบบเดียวกันและที่ไม่เป็นรูปแบบเดียวกัน สำหรับการตรวจสอบในกรณีของข้อสอบที่ให้คะแนนหลายค่าก็สามารถอธิบายได้ในทำนองเดียวกัน

สัดส่วนของข้อสอบทำหน้าที่ต่างกัน (Proportion of DIF)

การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบจะต้องเปรียบเทียบความสามารถของผู้สอบ ภายใต้คะแนนเกณฑ์การจับคู่ที่มีความเชื่อมั่น (Clauser; & Mazor, 1988: 37; Potenza; & Dorans, 1995: 32) ถ้าในแบบทดสอบประกอบด้วยข้อสอบทำหน้าที่เบี่ยงเบนจะมีผลทำให้ค่าประมาณความสามารถมีความเชื่อมั่นต่ำ แล้วอัตราความถูกต้องของวิธีการตรวจสอบจะมีค่าลดลง หรืออัตราความคลาดเคลื่อนประเภทที่ 1 สูงเกินปกติ จากการศึกษางานวิจัยของเกียร์ล กอทซ์แมนน์ และบอทตัน (Gierl; Gotzmann; & Boughton. 2004) นารายานานและสวามินาทาน (Narayanan; & Swaminathan. 1994, 1996) โรเจอร์สและสวามินาทาน (Rogers; & Swaminathan. 1993) พบว่า สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน (ในช่วง 0% ถึง 20%) ไม่มีผลต่อวิธีชิปเทสท์และวิธีแมนเทล-แฮนส์เซล แต่มีผลต่อวิธีการถดถอยโลจิสติก สาเหตุดังกล่าวเนื่องจากวิธีชิปเทสท์และวิธีแมนเทล-แฮนส์เซลใช้กระบวนการวิเคราะห์แบบสองขั้นตอน **ในขั้นตอนแรก**จะคำนวณโดยใช้คะแนนรวมจากข้อสอบทุกข้อ และ**ในขั้นตอนที่สอง**จะคำนวณซ้ำใหม่ แต่คัดเลือกข้อสอบที่ตรวจพบว่าทำหน้าที่เบี่ยงเบนในขั้นตอนแรกออกจากการคำนวณในขั้นตอนที่สอง ดังนั้นจึงทำให้คะแนนเกณฑ์การจับคู่ความสามารถมีความบริสุทธิ์ (Purification) ส่วนวิธีการถดถอยโลจิสติกไม่มีกระบวนการตรวจสอบดังกล่าว

ความแตกต่างของการแจกแจงความสามารถ (Ability distribution differences)

“การทำหน้าที่เบี่ยงเบนของข้อสอบ” สามารถพิจารณาได้จากผู้สอบที่อยู่ต่างกลุ่มมีความน่าจะเป็นในการตอบข้อสอบถูกแตกต่างกัน ภายหลังจากการจับคู่ความสามารถที่สนใจ ในขณะที่ความแตกต่างของผลการตอบข้อสอบที่ไม่ได้มีการจับคู่ความสามารถเรียกว่า **“ผลกระทบของข้อสอบ”** (Item impact) ซึ่งเกิดขึ้นจากความแตกต่างในการแจกแจงความสามารถของผู้สอบทั้งหมด (Camilli, 1992: 131; Clauser; & Mazor, 1998: 31; Dorans; & Holland; 1993: 37; Zumbo, 1999: 12) จากการศึกษางานวิจัยของนารายานานและสวามินาทาน (Narayanan; & Swaminathan. 1994, 1996) พบว่า การแจกแจงความสามารถระหว่างกลุ่มที่มีความแตกต่างในช่วง 0.0SD ถึง 1.0SD ไม่มีผลต่ออัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1

ของวิธีชิปเทสต์ แต่มีผลต่อวิธีการถดถอยโลจิสติก กล่าวคือ เมื่อความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มเพิ่มขึ้น 1.0SD มีผลทำให้อัตราความถูกต้องลดลงประมาณ 4% และอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นประมาณ 4% สาเหตุดังกล่าวเนื่องมาจากวิธีชิปเทสต์ใช้เทคนิคการตรวจสอบแบบหลายมิติ สามารถคำนวณคะแนนเกณฑ์ที่ใช้ในการจับคู่แยกจากข้อสอบที่ศึกษา ผลการตอบข้อสอบถูกจึงเป็นความแตกต่างของความสามารถแทรกซ้อน ซึ่งเป็นอิสระจากความสามารถเป้าหมาย ทำให้ความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มไม่มีผลกระทบต่อประสิทธิภาพในการตรวจสอบ ถึงแม้ว่าการแจกแจงความสามารถเป้าหมายระหว่างกลุ่มไม่เท่ากัน ก็สามารถใช้เทคนิคการปรับแก้ค่าการถดถอย (Regression correction) เพื่อลดความแตกต่างของค่าความสามารถเป้าหมายระหว่างกลุ่มอ้างอิงและกลุ่มสนใจ ทำให้สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ที่สูงเกินปกติได้ (Clauser; & Mazor. 1998: 35) ส่วนวิธีการถดถอยโลจิสติกใช้เทคนิคการตรวจสอบภายใต้โมเดลวัดความสามารถมิติเดียว ทำให้ผลการตรวจสอบเกิดความผิดพลาด ทั้งนี้เพราะไม่สามารถแยกออกจากกัน ระหว่างการทำหน้าที่เบี่ยงเบนของข้อสอบและผลกระทบของข้อสอบ

ขนาดตัวอย่าง (Sample size)

ขนาดตัวอย่างมีผลต่อความถูกต้องของการทดสอบนัยสำคัญทางสถิติ (Power of the test) วิธีในตระกูลการถดถอยโลจิสติกมีรูปแบบพาราเมตริก (Parametric form) จำเป็นต้องใช้ข้อตกลงของโมเดลอธิบายความสัมพันธ์ระหว่างผลการตอบข้อสอบกับตัวแปรการจับคู่ ส่วนวิธีในตระกูลชิปเทสต์ แมนเทิล-แฮนส์เซล และการทำให้เป็นมาตรฐานมีรูปแบบนพาราเมตริก (Non-parametric form) ซึ่งไม่มีข้อตกลงดังกล่าว (Potenza; & Dorans. 1995: 24) ถึงแม้ว่าวิธีทั้งสองกลุ่มจะมีรูปแบบทางสถิติแตกต่างกัน แต่วิธีทั้งสองกลุ่มใช้สถิติในการทดสอบนัยสำคัญที่มีพื้นฐานของการแจกแจงแบบเชิงเส้นกำกับ (Asymptotic distribution) ซึ่งการแจกแจงดังกล่าวจำเป็นต้องใช้ขนาดตัวอย่างที่เพียงพอ จากผลการศึกษาของนักวิจัยหลายคน (ทองอยู่ สาระ. 2543; วลีมาศ แซ่จึ้ง. 2543; สิริวิรัตน์ วิภาสศิลป์. 2545; อุทัยวรรณ สายพัฒนะ. 2547; Chang; Mazzeo; & Roussos. 1996; Narayanan; & Swaminathan. 1994, 1996; Rogers; & Swaminathan. 1993; Roussos; & Stout. 1996a; Swaminathan; & Rogers. 1990; French; & Miller. 1996) ให้ผลสอดคล้องกัน กล่าวคือ ขนาดตัวอย่างมีผลต่ออัตราความถูกต้องของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบทั้งสองกลุ่ม เมื่อขนาดตัวอย่างเพิ่มขึ้นมีผลทำให้อัตราความถูกต้องของวิธีการตรวจสอบมีค่าเพิ่มขึ้น

จากการศึกษากรอบแนวคิดเชิงทฤษฎีเกี่ยวกับการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ตลอดจนงานวิจัยต่างๆ ที่เกี่ยวข้องพบว่า นักวิจัยตั้งแต่อดีตจนถึงปัจจุบัน พยายามคิดค้นและพัฒนาวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ เพื่อให้ผลการตรวจสอบมีความถูกต้องและแม่นยำมากที่สุด เริ่มต้นจากการตรวจสอบในข้อสอบที่วัดความสามารถมิติเดียวและให้คะแนนสองค่า แล้วพัฒนาต่อเนื่องมาจนถึงการตรวจสอบในข้อสอบที่ให้คะแนนหลายค่า ต่อมา มีการพัฒนาโมเดลการวัดความสามารถหลายมิติ และมีการพัฒนาโปรแกรมคอมพิวเตอร์ที่ใช้วิเคราะห์ค่าพารามิเตอร์ของข้อสอบที่วัดความสามารถหลายมิติ ตลอดจนสามารถประเมินมิติของแบบทดสอบ เช่น โปรแกรม NOHARM, TESTFACT, DIMTEST และ Poly-DIMTEST เป็นต้น ด้วยเหตุผลดังกล่าวจึงทำให้นักวิจัยหันมาพัฒนาเทคนิควิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติ ซึ่งเชื่อว่าจะมีความสมบูรณ์ในการวัดมากกว่า แต่การพัฒนา ยังจำกัดเฉพาะข้อสอบที่ให้คะแนนสองค่า จากการศึกษาวิจัยทั้งต่างประเทศและในประเทศ ยังไม่พบการตรวจสอบในข้อสอบที่ให้คะแนนหลายค่า ดังนั้นผู้วิจัยจึงมีความสนใจที่จะตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า (Multidimensional polytomous) โดยใช้วิธีโพลี-ซิปเทสท์ (Poly-SIBTEST) (Chang; Mazzeo; & Roussos. 1996) วิธีการถดถอยโลจิสติกแบบจัดอันดับ (Ordinal logistic regression; OLR) (Zumbo. 1999) และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ซึ่งเป็นวิธีที่ผู้วิจัยปรับขยายเพื่อใช้กับข้อมูลที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า โดยศึกษาภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ประกอบด้วย รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน สัดส่วนของข้อสอบที่ทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่าง

เนื่องจากการศึกษาครั้งนี้ ผู้วิจัยได้นำวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมาตรวจสอบเป็นครั้งแรก ดังนั้นจึงศึกษาภายใต้การจำลองข้อมูล ผลจากการวิจัยครั้งนี้จะได้สารสนเทศในเชิงวิชาการด้านการทดสอบและวัดผลการศึกษา ซึ่งสามารถนำไปใช้ประโยชน์ในเชิงปฏิบัติต่อไป

บทที่ 3

วิธีดำเนินการวิจัย

ในการวิจัยครั้งนี้ มีความมุ่งหมายเพื่อเปรียบเทียบวิธีโพลี-ซีปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่าง ในการดำเนินการวิจัยมีขั้นตอนดังนี้

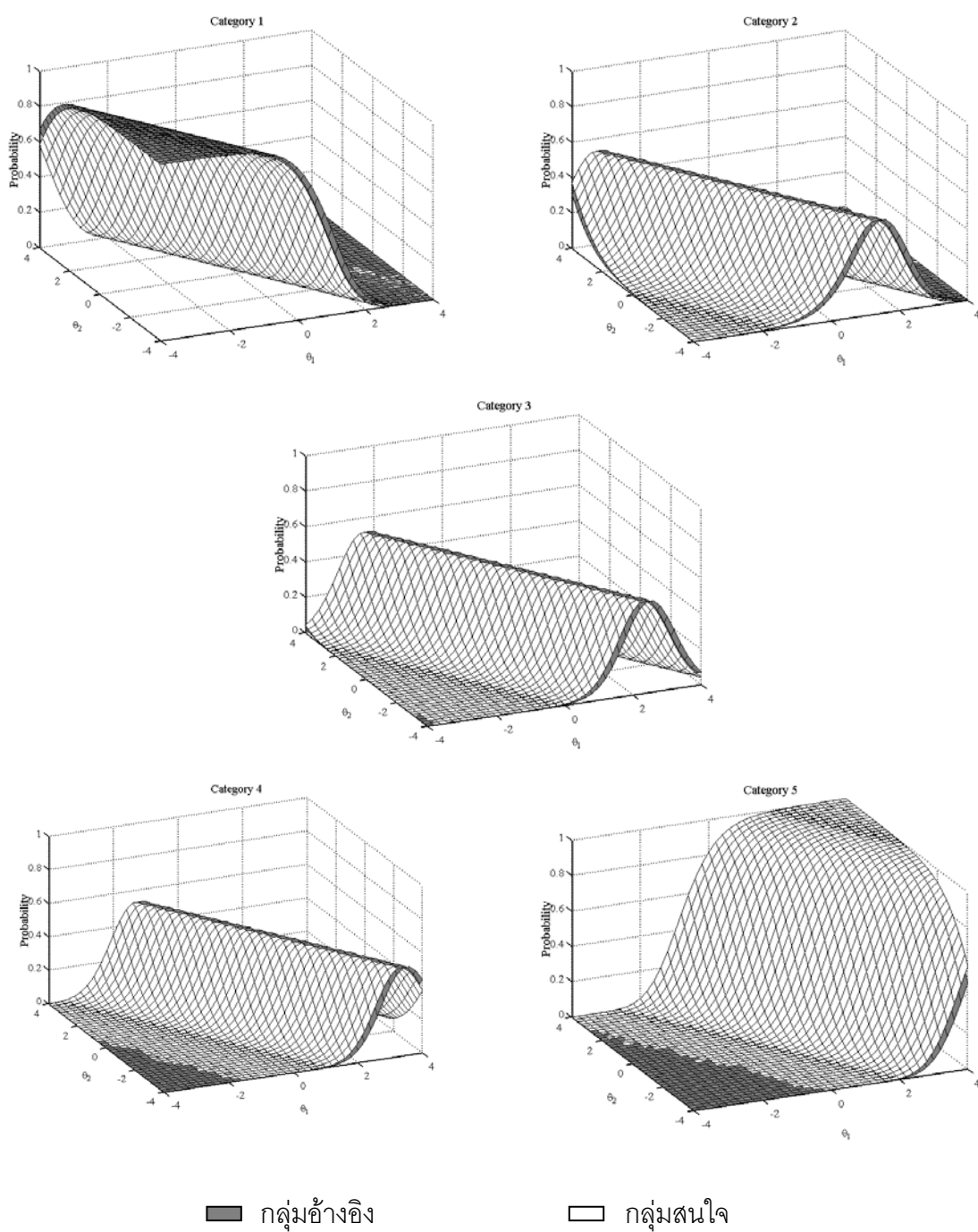
1. การจัดกระทำตัวแปร
2. การจำลองข้อมูล
3. การวิเคราะห์ข้อมูล
4. การศึกษานำร่อง

การจัดกระทำตัวแปร

การวิจัยครั้งนี้ศึกษาในสถานการณ์จำลอง โดยใช้ทฤษฎีการตอบข้อสอบแบบหลายมิติ (Multidimensional item response theory; MIRT) โมเดลพาเชียลเครดิตทั่วไปแบบหลายมิติ (Multidimensional generalized partial credit Model) จำลองข้อมูลภายใต้การจัดกระทำตัวแปรอิสระ 4 ตัวแปร คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน 2 รูปแบบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 3 ขนาด ความแตกต่างของการแจกแจงความสามารถ 3 ระดับ และขนาดตัวอย่าง 4 ขนาด โดยจำลองแบบทดสอบที่มีโครงสร้างแบบสองมิติ (Two-dimensional) จำนวน 40 ข้อ แบบทดสอบดังกล่าววัดความสามารถหลายมิติภายในข้อสอบ (Multidimensional within-item test) โดยข้อสอบแต่ละข้อวัดความสามารถเป้าหมายสองมิติที่มีน้ำหนักแตกต่างกัน ตัวอย่างเช่น แบบทดสอบวิชาคณิตศาสตร์ฉบับหนึ่งวัดความสามารถสองมิติ โดยให้มิติที่หนึ่งวัดความสามารถในการคิดคำนวณ (θ_1) และมิติที่สองวัดความสามารถในการแก้โจทย์ปัญหา (θ_2) เป็นต้น ข้อสอบแต่ละข้อมีรายการตอบ 5 รายการ ผลการตอบในรายการตอบ 1, 2, 3, 4 และ 5 ให้คะแนนเป็น 1, 2, 3, 4 และ 5 ตามลำดับ สำหรับการจัดกระทำตัวแปรอิสระทั้ง 4 ตัวแปรดังกล่าว ผู้วิจัยได้ออกแบบการวิจัยโดยมีรายละเอียดดังนี้

1. รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน (Form of DIF)

รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนจะขึ้นอยู่กับขนาดและทิศทางของข้อสอบทำหน้าที่เบี่ยงเบน ซึ่งแปรเปลี่ยนตามความแตกต่างของระดับความสามารถของผู้สอบ ในการวิจัยครั้งนี้ ศึกษา รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนสองรูปแบบ คือ ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF) และข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Non-uniform DIF) การจำลองข้อมูลในเงื่อนไขดังกล่าวใช้โมเดลพาเซียลเครดิตทั่วไปแบบหลายมิติ ภายใต้ทฤษฎีการตอบข้อสอบแบบหลายมิติ แนวคิดในการจำลองข้อมูลปรับขยายมาจากงานวิจัยของเฟรนช์และมิลเลอร์ (French; & Miller. 1996) ที่ศึกษาการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถมิติเดียวและให้คะแนนหลายค่า และงานวิจัยของโอชิมะ ราจู และเฟลาเวอร์ส (Oshima; Raju; & Flowers. 1997) ที่ศึกษาการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนสองค่า ในการศึกษานี้กำหนดความแตกต่างของพารามิเตอร์อำนาจจำแนกของข้อสอบในมิติที่หนึ่งและมิติที่สอง (a_{1i} และ a_{2i}) ระหว่างกลุ่มอ้างอิงและกลุ่มสนใจ เท่ากับ 0.15 และความแตกต่างของพารามิเตอร์รายการของข้อสอบ (d_{iv}) ระหว่างกลุ่มอ้างอิงและกลุ่มสนใจเท่ากับ 0.20 การกำหนดค่าดังกล่าวขึ้นอยู่กับเงื่อนไขของการจำลองรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กล่าวคือ การจำลองข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF) กำหนดให้พารามิเตอร์รายการของข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มสนใจมีค่าแปรเปลี่ยน ในขณะที่พารามิเตอร์อำนาจจำแนกของข้อสอบระหว่างกลุ่มมีค่าคงที่ ส่วนการจำลองข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) กำหนดให้พารามิเตอร์อำนาจจำแนกของข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มสนใจมีค่าแปรเปลี่ยน ในขณะที่พารามิเตอร์รายการของข้อสอบระหว่างกลุ่มมีค่าคงที่ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนทั้งสองรูปแบบสามารถแสดงด้วยฟังก์ชันผิวหน้าการตอบรายการ (Option response surfaces; ORSs) ซึ่งเป็นภาพสามมิติ ดังภาพประกอบ 13 และ 14 ส่วนการกำหนดค่าพารามิเตอร์ของข้อสอบทั้งสองรูปแบบ จะกล่าวรายละเอียดในหัวข้อการจำลองข้อมูล ซึ่งเป็นหัวข้อถัดไป



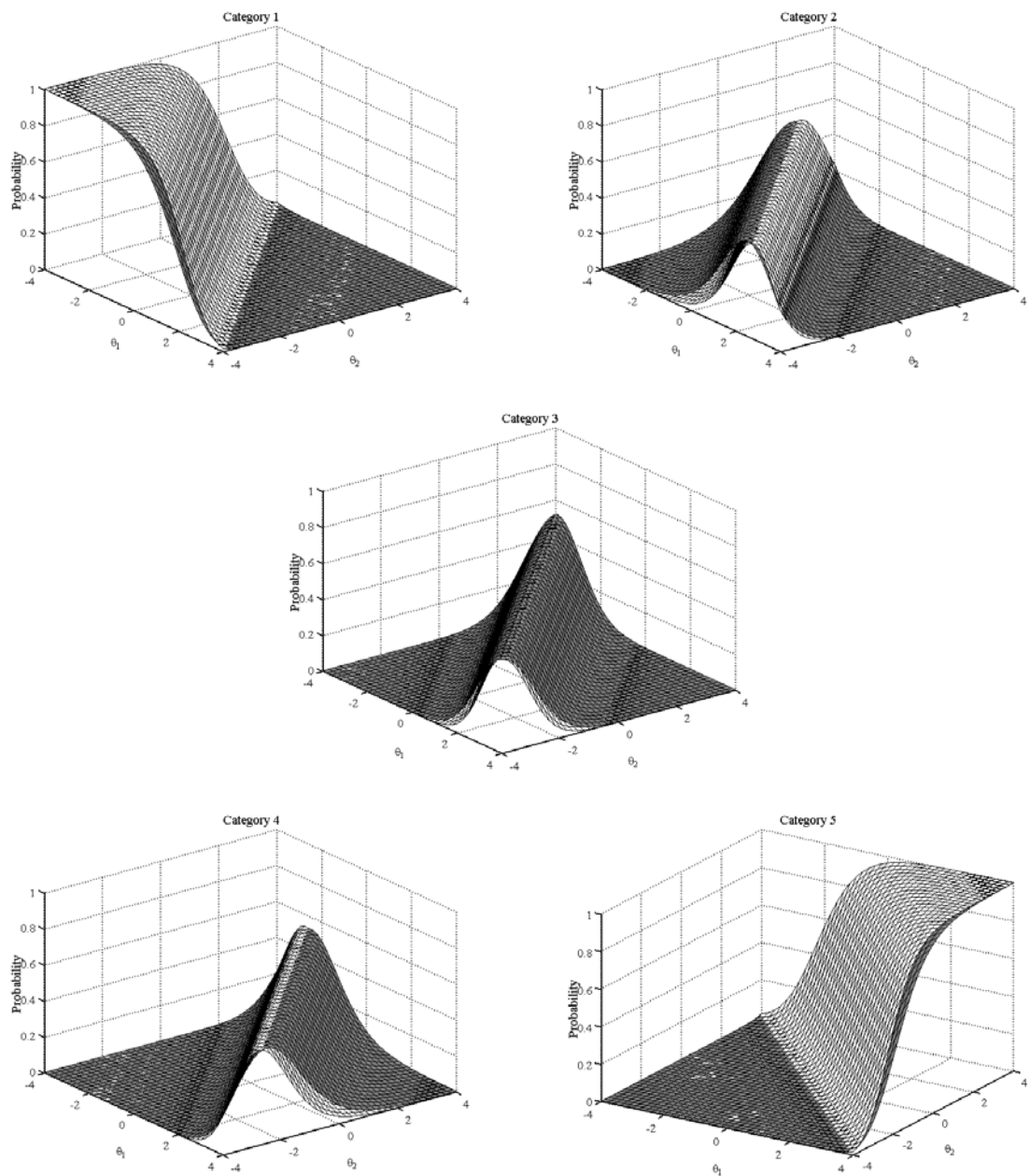
ภาพประกอบ 13 ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF)

ภายใต้โมเดลพหุคูณเชิงเส้นแบบหลายมิติ (MGPCM)

$$(a_{1R} = a_{1F} = 1.04, a_{2R} = a_{2F} = 0.60,$$

$$d_{2R} = 1.50, d_{3R} = 0.50, d_{4R} = -0.50, d_{5R} = -1.50,$$

$$d_{2F} = 1.30, d_{3F} = 0.30, d_{4F} = -0.70, d_{5F} = -1.70)$$



■ กลุ่มอ้างอิง

□ กลุ่มสนใจ

ภาพประกอบ 14 ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF)

ภายใต้โมเดลพหุคูณเชิงเส้นแบบหลายมิติ (MGPCM)

$$(a_{1R} = a_{2R} = 0.85, a_{1F} = a_{2F} = 0.70,$$

$$d_{2R} = d_{2F} = 1.50, d_{3R} = d_{3F} = 0.50, d_{4R} = d_{4F} = -0.50, d_{5R} = d_{5F} = -1.50)$$

2. สัดส่วนของการทำหน้าที่เบี่ยงเบนของข้อสอบ (Proportion of DIF)

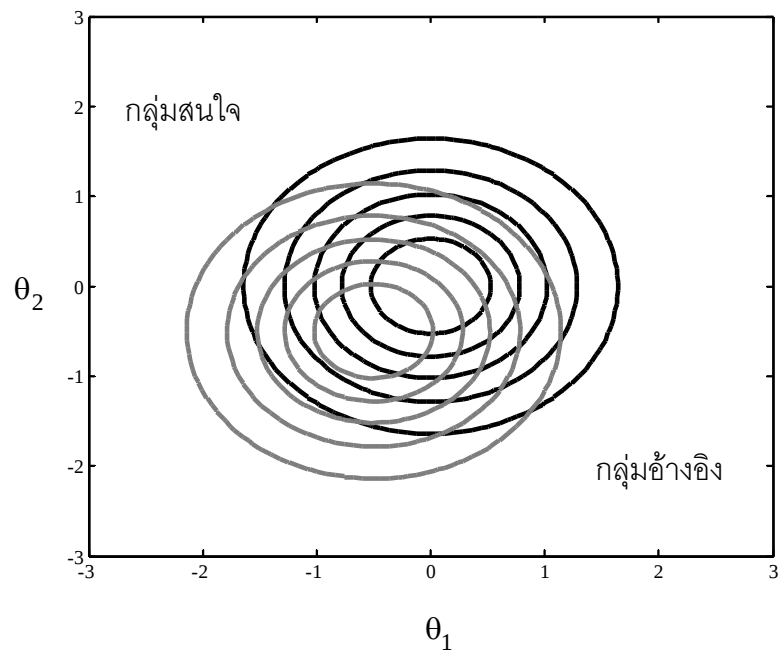
แบบทดสอบมาตรฐานวัดผลสัมฤทธิ์ทางการเรียน ถ้ามีข้อสอบทำหน้าที่เบี่ยงเบนตั้งแต่ 10% ถึง 15% ของจำนวนข้อสอบทั้งหมดถือว่าผิดปกติ (Narayanan; & Swaminathan. 1994: 319; citing Clauser. 1993) จากการศึกษาของดราสโกว์ (Oshima; & Miller. 1992: 239; citing Drasgow. 1987) ที่ได้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในแบบทดสอบหลายวิชา ซึ่งเป็นการทดสอบของวิทยาลัยอเมริกัน (American College Testing; ACT) ผลการศึกษาพบว่า มีข้อสอบทำหน้าที่เบี่ยงเบนตั้งแต่ 5% ถึง 29% ในการศึกษาปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน นักวิจัยหลายคน (วลิมาศ แซ่อิง. 2543; Flowers; Oshima; & Raju. 1999; Narayanan; & Swaminathan. 1994, 1996; Raju; van der Linden; & Fler. 1995; Rogers; & Swaminathan. 1993) มักศึกษาในช่วง 0% ถึง 20% ซึ่งสอดคล้องกับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนที่สามารถเกิดขึ้นกับข้อมูลจริง ในการวิจัยครั้งนี้ศึกษาสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 3 ขนาด คือ 10%, 20% และ 30% โดยใช้ความยาวของแบบทดสอบจำนวน 40 ข้อ ในทุกเงื่อนไข ดังนั้นในการจำลองแบบทดสอบแต่ละฉบับ ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน จำนวน 10%, 20% และ 30% จะต้องจัดกระทำข้อสอบทำหน้าที่ไม่เบี่ยงเบน กับข้อสอบทำหน้าที่เบี่ยงเบน จำนวน 36 ข้อกับ 4 ข้อ, 32 ข้อกับ 8 ข้อ และ 28 ข้อกับ 12 ข้อ ตามลำดับ

3. ความแตกต่างของการแจกแจงความสามารถ (Ability distribution differences)

โครงสร้างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบในโมเดลหลายมิติจะต้องพิจารณาถึงค่าเฉลี่ยและความแปรปรวนของความสามารถเป้าหมายในทุกมิติ ตลอดจนความแปรปรวนร่วมระหว่างความสามารถในทุกมิติ หรือค่าสหสัมพันธ์ระหว่างความสามารถในทุกมิติ ในการวิจัยครั้งนี้จำลองแบบทดสอบที่มีโครงสร้างวัดความสามารถแบบสองมิติ ซึ่งประกอบด้วยข้อสอบที่วัดความสามารถเป้าหมายทั้งสองมิติ โดยกำหนด θ_1 และ θ_2 แทนความสามารถในมิติที่หนึ่งและมิติที่สอง ตามลำดับ การจำลองค่าความสามารถ θ_1 และ θ_2 มีการแจกแจงแบบปกติสองตัวแปร (Bivariate normal distribution) ซึ่งมีลักษณะแตกต่างกัน 3 กรณี คือ **กรณีที่ 1** การแจกแจงความสามารถระหว่างกลุ่มผู้สอบเท่ากัน ($\bar{\theta}_{1R} - \bar{\theta}_{1F} = 0.0$ และ $\bar{\theta}_{2R} - \bar{\theta}_{2F} = 0.0$) โดยกำหนดให้กลุ่มอ้างอิงและกลุ่มสนใจมีค่าเฉลี่ยของความสามารถเท่ากัน คือ 0.0 **กรณีที่ 2** การแจกแจงความสามารถระหว่างกลุ่มผู้สอบแตกต่างกัน 0.5SD ($\bar{\theta}_{1R} - \bar{\theta}_{1F} = 0.5$ และ $\bar{\theta}_{2R} - \bar{\theta}_{2F} = 0.5$) โดยกำหนดให้กลุ่มอ้างอิงและกลุ่มสนใจมีค่าเฉลี่ยของความสามารถเท่ากับ 0.0 และ -0.5 ตามลำดับ

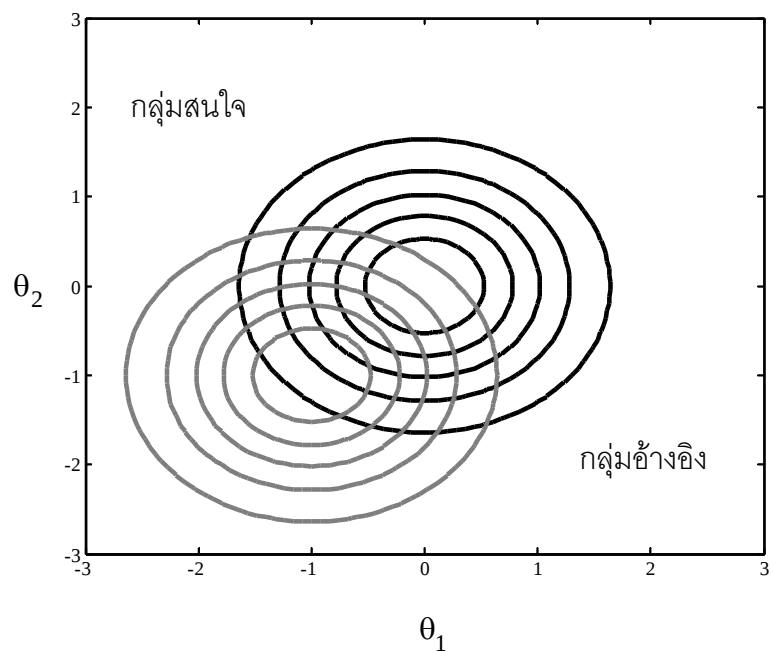
และกรณีที่ 3 การแจกแจงความสามารถระหว่างกลุ่มผู้สอบแตกต่างกัน 1.0SD ($\bar{\theta}_{1R} - \bar{\theta}_{1F} = 1.0$ และ $\bar{\theta}_{2R} - \bar{\theta}_{2F} = 1.0$) โดยกำหนดให้กลุ่มอ้างอิงและกลุ่มสนใจมีค่าเฉลี่ยของความสามารถเท่ากับ 0.0 และ -1.0 ตามลำดับ ในทุกกรณีจะกำหนดให้ค่าความแปรปรวนของค่าความสามารถในแต่ละกลุ่มมีค่าคงที่ เท่ากับ 1.0 และกำหนดค่าสหสัมพันธ์ระหว่างความสามารถ θ_1 และ θ_2 ในแต่ละกลุ่มมีค่าคงที่ เท่ากับ 0.50

ในการกำหนดค่าเฉลี่ย ความแปรปรวน และค่าสหสัมพันธ์ดังกล่าวจะต้องคำนึงถึงค่าที่สามารถเกิดขึ้นในข้อมูลจริง จากการศึกษาของลินน์และดราสโกว์ (Oshima; & Miller. 1992: 239; citing Linn; & Drasgow. 1987) พบว่า ความแตกต่างของค่าเฉลี่ยของความสามารถระหว่างผู้สอบกลุ่มผิวดำและกลุ่มผิวขาวเท่ากับหนึ่งส่วนเบี่ยงเบนมาตรฐาน (Standard deviation; SD) ซึ่งสอดคล้องกับความเห็นของสเตาท์และคนอื่นๆ (Stout; et al. 1997: 203) ที่กล่าวว่า ค่าเฉลี่ยของความสามารถเท่ากับ 1.0SD ถือว่าเป็นค่าประมาณที่มีค่ามากสุดในการนำมาประยุกต์ใช้ตามสภาพที่เป็นจริง ข้อสรุปดังกล่าวสอดคล้องกับการศึกษาของนักวิจัยหลายคน (Ackerman; & Evan. 1994; Bolt. 2002; Chang; Mazzeo; & Roussos. 1996; Flowers; Oshima; & Raju. 1999; Gierl; Gotzmann; & Boughton. 2004; Krisjansson; et al. 2005; Narayanan; & Swaminathan. 1994, 1996; Roussos; & Stout. 1996a; Stout; et al. 1997; Whitmore; & Schumacker. 1999) ที่กำหนดความแตกต่างของค่าเฉลี่ยของความสามารถระหว่างกลุ่มผู้สอบไม่เกินหนึ่ง 1.0SD ส่วนการศึกษาเกี่ยวกับค่าสหสัมพันธ์ระหว่างความสามารถ นักวิจัยหลายคน (Ackerman; & Evan. 1994; Mazor; Hambleton; & Clauser. 1998; Stout; et al. 1997) มักกำหนดค่าสหสัมพันธ์ในช่วง 0.3 ถึง 0.70 ซึ่งเป็นค่าที่เป็นไปได้ในข้อมูลจริง ในการกำหนดค่าเฉลี่ย ความแปรปรวน และค่าสหสัมพันธ์ดังกล่าวมีผลต่อความหนาแน่นของการแจกแจงความสามารถแบบปกติสองตัวแปร (θ_1 และ θ_2) ซึ่งสามารถแสดงด้วยภาพสองมิติและสามมิติ ดังภาพประกอบ 15-16 และ 17-19



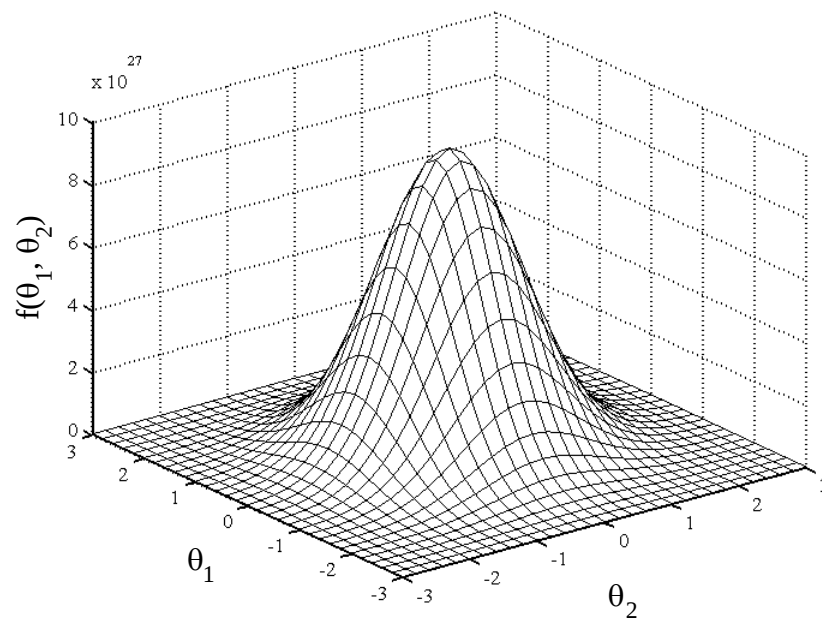
ภาพประกอบ 15 เส้นโค้งร่างของการแจกแจงความสามารถแบบสองมิติ
ของกลุ่มอ้างอิงและกลุ่มสนใจ เมื่อค่าเฉลี่ยของความสามารถแตกต่างกัน 0.5SD

$$(\bar{\theta}_{1R} = \bar{\theta}_{2R} = 0.0, \bar{\theta}_{1F} = \bar{\theta}_{2F} = -0.5, S_{\theta_1}^2 = S_{\theta_2}^2 = 1.0, r_{\theta_1\theta_2} = 0.5)$$



ภาพประกอบ 16 เส้นโค้งร่างของการแจกแจงความสามารถแบบสองมิติ
ของกลุ่มอ้างอิงและกลุ่มสนใจ เมื่อค่าเฉลี่ยของความสามารถแตกต่างกัน 1.0SD

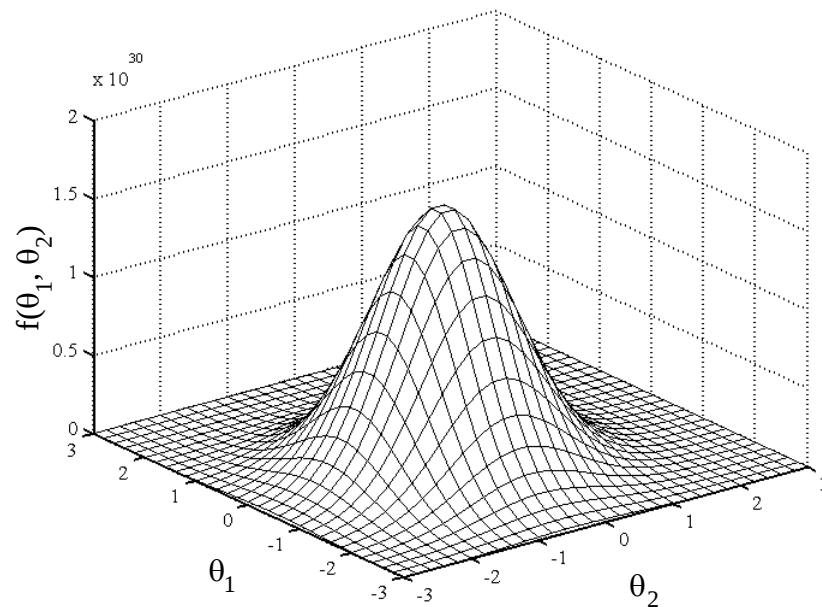
$$(\bar{\theta}_{1R} = \bar{\theta}_{2R} = 0.0, \bar{\theta}_{1F} = \bar{\theta}_{2F} = -1.0, S_{\theta_1}^2 = S_{\theta_2}^2 = 1.0, r_{\theta_1\theta_2} = 0.5)$$



ภาพประกอบ 17 การแจกแจงความสามารถแบบสองมิติ

เมื่อค่าเฉลี่ยของความสามารถเท่ากับ 0.0

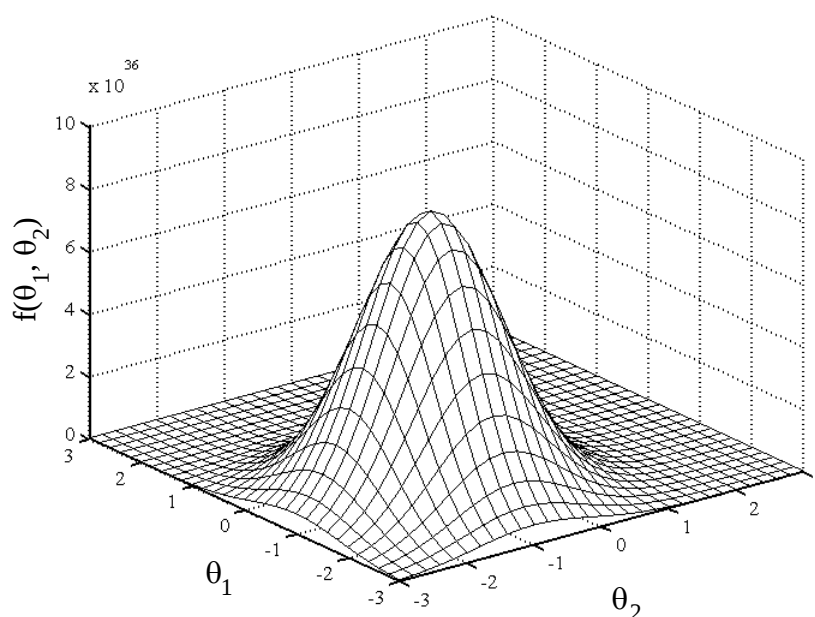
$$(\bar{\theta}_1 = \bar{\theta}_2 = 0.0, S_{\theta_1}^2 = S_{\theta_2}^2 = 1.0, r_{\theta_1\theta_2} = 0.5)$$



ภาพประกอบ 18 การแจกแจงความสามารถแบบสองมิติ

เมื่อค่าเฉลี่ยของความสามารถเท่ากับ -0.5

$$(\bar{\theta}_1 = \bar{\theta}_2 = -0.5, S_{\theta_1}^2 = S_{\theta_2}^2 = 1.0, r_{\theta_1\theta_2} = 0.5)$$



ภาพประกอบ 19 การแจกแจงความสามารถแบบสองมิติ

เมื่อค่าเฉลี่ยของความสามารถเท่ากับ -1.0

$$(\bar{\theta}_1 = \bar{\theta}_2 = -1.0, S_{\theta_1}^2 = S_{\theta_2}^2 = 1.0, r_{\theta_1\theta_2} = 0.5)$$

ขนาดตัวอย่าง (Sample size)

การศึกษานี้เกี่ยวข้องกับขนาดตัวอย่างที่ใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบจะแบ่งผู้สอบออกเป็นสองกลุ่ม คือ กลุ่มสนใจ (N_F) และกลุ่มอ้างอิง (N_R) สำหรับกลุ่มสนใจเป็นกลุ่มเป้าหมายที่ต้องการศึกษา ส่วนกลุ่มอ้างอิงเป็นกลุ่มเปรียบเทียบ เมื่อข้อสอบทำหน้าที่เบี่ยงเบนกลุ่มแรกคาดว่าจะเสียเปรียบในการตอบข้อสอบ ส่วนกลุ่มหลังคาดว่าจะได้เปรียบในการตอบข้อสอบ ตามปกติจะกำหนดขนาดตัวอย่างของกลุ่มอ้างอิงเท่ากับหรือมากกว่ากลุ่มสนใจในการศึกษาครั้งนี้กำหนดให้ขนาดของกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน เพราะในการจำลองข้อมูลภายใต้ทฤษฎีการตอบข้อสอบแบบหลายมิติ (Multidimensional IRT) ต้องการควบคุมการประมาณค่าพารามิเตอร์ความสามารถของแต่ละกลุ่มให้คงที่ ส่วนการกำหนดขนาดตัวอย่างเมื่อพิจารณางานวิจัยที่ตรวจสอบโดยใช้วิธีชิปเทสต์ วิธีโคร-ชิป วิธีมัลติชิป วิธีโพลี-ชิปเทสต์ และวิธีที่ใช้โมเดลการถดถอยโลจิสติกมักใช้ตัวอย่างที่มีขนาดเล็กจนถึงขนาดใหญ่ในช่วง 100 ถึง 3,000 คน (ทองอยู่สาระ. 2543; วลีมาศ แซ่ฮึง. 2543; อุทัยวรรณ สายพัฒนะ. 2547; Bolt. 2002; Chang; Mazzeo; & Roussos. 1996; French; & Miller. 1996; Gierl; Gotzmann; & Boughton. 2004; Krisjansson;

et al. 2005; Mazor; Hambleton; & Clauser. 1998; Narayanan; & Swaminathan. 1994, 1996; Roussos; & Stout. 1996a; Stout; et al. 1997; Whitmore; & Schumacker. 1999) ในการวิจัยครั้งนี้ศึกษาขนาดตัวอย่าง 4 ขนาด คือ $N_R = N_F = 250$, $N_R = N_F = 500$, $N_R = N_F = 1,000$ และ $N_R = N_F = 2,000$

ความยาวของแบบทดสอบไม่ได้เป็นปัจจัยที่ใช้ในการศึกษา โดยกำหนดคงที่เท่ากับ 40 ข้อ ในทุกเงื่อนไข ความยาวของแบบทดสอบดังกล่าวเป็นความยาวโดยเฉลี่ยของแบบทดสอบชุดย่อยในแบบทดสอบมาตรฐานวัดผลสัมฤทธิ์ทางการเรียน (Rogers; & Swaminathan. 1993: 106) ส่วนขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบไม่ได้เป็นปัจจัยที่ศึกษา โดยกำหนดคงที่ตามรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ดังนั้นในการวิเคราะห์การทำหน้าที่เบี่ยงเบนของข้อสอบจะต้องจำลองข้อมูลโดยการจัดกระทำปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน 2 รูปแบบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 3 ขนาด ความแตกต่างของการแจกแจงความสามารถ 3 ระดับ และขนาดตัวอย่าง 4 ขนาด รวมข้อมูลที่ต้องจำลองทั้งหมด 72 เงื่อนไข ($2 \times 3 \times 3 \times 4$) ในแต่ละเงื่อนไขจำลองข้อมูลซ้ำ 50 ครั้ง สำหรับแผนผังของการจำลองข้อมูลทั้ง 4 ปัจจัยแสดงในภาพประกอบ 20

รูปแบบของ DIF	สัดส่วนของ DIF	การแจกแจงความสามารถ	ขนาดตัวอย่าง
	10%	เท่ากัน	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1,000$
			$N_R = N_F = 2,000$
		แตกต่างกัน 0.5SD	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1,000$
			$N_R = N_F = 2,000$
		แตกต่างกัน 1.0SD	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1,000$
			$N_R = N_F = 2,000$
		เท่ากัน	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1,000$
			$N_R = N_F = 2,000$
ที่เป็นรูปแบบเดียวกัน	20%	แตกต่างกัน 0.5SD	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1,000$
			$N_R = N_F = 2,000$
		แตกต่างกัน 1.0SD	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1,000$
			$N_R = N_F = 2,000$
		เท่ากัน	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1,000$
			$N_R = N_F = 2,000$
	30%	แตกต่างกัน 0.5SD	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1,000$
			$N_R = N_F = 2,000$
		แตกต่างกัน 1.0SD	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1,000$
			$N_R = N_F = 2,000$
		เท่ากัน	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1,000$
			$N_R = N_F = 2,000$

ภาพประกอบ 20 แผนผังของการจำลองข้อมูล

รูปแบบของ DIF	สัดส่วนของ DIF	การแจกแจงความสามารถ	ขนาดตัวอย่าง
	10%	เท่ากัน	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1,000$
			$N_R = N_F = 2,000$
		แตกต่างกัน 0.5SD	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1,000$
			$N_R = N_F = 2,000$
		แตกต่างกัน 1.0SD	$N_R = N_F = 250$
			$N_R = N_F = 500$
	20%		$N_R = N_F = 1,000$
			$N_R = N_F = 2,000$
		เท่ากัน	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1,000$
			$N_R = N_F = 2,000$
		แตกต่างกัน 0.5SD	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1,000$
			$N_R = N_F = 2,000$
	30%		$N_R = N_F = 250$
			$N_R = N_F = 500$
		แตกต่างกัน 1.0SD	$N_R = N_F = 1,000$
			$N_R = N_F = 2,000$
		เท่ากัน	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1,000$
			$N_R = N_F = 2,000$
		แตกต่างกัน 0.5SD	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1,000$
			$N_R = N_F = 2,000$
		แตกต่างกัน 1.0SD	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1,000$
			$N_R = N_F = 2,000$

การจำลองข้อมูล

การศึกษานี้ใช้ข้อมูลจำลอง โดยจำลองข้อสอบภายใต้โมเดลพหุเชิงเส้นเครดิตทั่วไปแบบหลายมิติที่มีการชดเชยค่าความสามารถ (Compensatory multidimensional generalized partial credit model) จำนวน 40 ข้อ ในข้อสอบแต่ละข้อจะวัดความสามารถเป้าหมายทั้งสองมิติ กำหนดให้ θ_1 แทนความสามารถในมิติที่หนึ่ง และ θ_2 แทนความสามารถในมิติที่สอง ในแต่ละข้อมีรายการตอบ 5 รายการ ผลการตอบข้อสอบในรายการตอบ 1, 2, 3, 4 และ 5 ให้คะแนนเป็น 1, 2, 3, 4 และ 5 คะแนน ตามลำดับ สำหรับโมเดลพหุเชิงเส้นเครดิตทั่วไปแบบหลายมิติปรับขยายมาจากโมเดลพหุเชิงเส้นเครดิตทั่วไปแบบมิติเดียวของมูรากิ ดังนี้ (Muraki. 1992: 161)

$$P_{ix}(\theta_1, \theta_2) = \frac{\exp\left[\sum_{v=1}^x D(a_{1i}\theta_1 + a_{2i}\theta_2 + d_{iv})\right]}{\sum_{c=1}^5 \exp\left[\sum_{v=1}^c D(a_{1i}\theta_1 + a_{2i}\theta_2 + d_{iv})\right]} \quad (1)$$

เมื่อ

$P_{ix}(\theta_1, \theta_2)$ แทน ความน่าจะเป็นของผู้สอบที่มีความสามารถ θ_1 และ θ_2 จะได้รับ

คะแนน x ในการตอบข้อสอบข้อที่ i ($x = 1, 2, 3, 4$ และ 5)

a_{1i} และ a_{2i} แทน พารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่ i (Item discrimination)

ซึ่งเกี่ยวข้องกับพารามิเตอร์ความสามารถ θ_1 และ θ_2 ตามลำดับ

d_{iv} แทน พารามิเตอร์ในรายการที่ v ของข้อสอบข้อที่ i (Item-category)

($v = 1, 2, 3, 4$ และ 5) เมื่อ $v = 1$ แล้ว $d_{i1} = 0$

พารามิเตอร์อำนาจจำแนกของข้อสอบในแต่ละมิติ (a_{1i} และ a_{2i}) และพารามิเตอร์รายการของข้อสอบ (d_{iv}) สามารถนำมาคำนวณพารามิเตอร์อำนาจจำแนกของข้อสอบแบบหลายมิติ (Multidimensional item discrimination; MDISC) และพารามิเตอร์ความยากของข้อสอบแบบหลายมิติ (Multidimensional item difficulties; MDIFF) โดยใช้สูตรดังนี้

$$\alpha_i = \sqrt{a_{i1}^2 + a_{i2}^2} \quad (2)$$

$$\beta_{iv} = -\frac{d_{iv}}{\sqrt{a_{i1}^2 + a_{i2}^2}} \quad (3)$$

เมื่อ

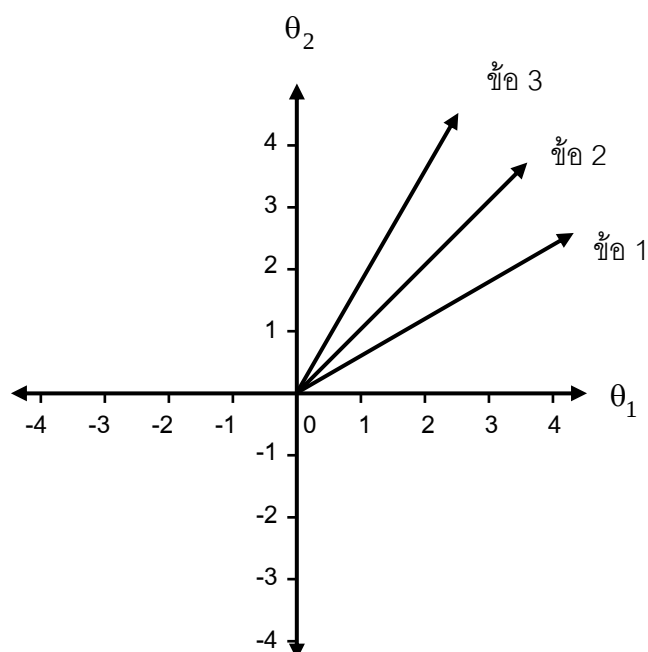
α_i แทน พารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่ i แบบหลายมิติ

β_{iv} แทน พารามิเตอร์ความยากในรายการที่ v ของข้อสอบข้อที่ i แบบหลายมิติ

ในการจำลองแบบทดสอบที่มีโครงสร้างของการวัดความสามารถแบบสองมิติสามารถกำหนดทิศทางของข้อสอบ (Item direction) เพื่อวัดความสามารถ θ_1 และ θ_2 ซึ่งอาจกำหนดให้ข้อสอบวัดความสามารถ θ_1 มากกว่า θ_2 หรือข้อสอบวัดความสามารถ θ_1 น้อยกว่า θ_2 หรือข้อสอบวัดความสามารถ θ_1 และ θ_2 เท่ากันก็ได้ ในการกำหนดทิศทางของข้อสอบดังกล่าวสามารถนำค่าพารามิเตอร์อำนาจจำแนกของข้อสอบในแต่ละมิติมาคำนวณขนาดของมุมในรูปโคไซน์ (Cosines) โดยใช้ฟังก์ชันตรีโกณมิติดังนี้

$$\alpha_{i1} = \arccos \frac{a_{i1}}{\sqrt{a_{i1}^2 + a_{i2}^2}} \quad (4)$$

เมื่อ α_{i1} แทนขนาดของมุมระหว่างข้อสอบข้อที่ i กับแกนความสามารถ θ_1 (แกนบวก) ค่าของ α_{i1} มีค่าตั้งแต่ 0 ถึง 90 องศา ขนาดของมุมดังกล่าวจะขึ้นอยู่กับระดับของข้อสอบที่วัดความสามารถ θ_1 และ θ_2 กล่าวคือ ถ้าข้อสอบวัดความสามารถเฉพาะ θ_1 แล้ว $\alpha_{i1} = 0$ องศา และถ้าข้อสอบวัดความสามารถเฉพาะ θ_2 แล้ว $\alpha_{i1} = 90$ องศา แต่ถ้าข้อสอบวัดความสามารถ θ_1 และ θ_2 เท่ากัน แล้ว $\alpha_{i1} = 45$ องศา การวัดความสามารถของข้อสอบในสองกรณีแรกเป็นการวัดแบบมิติเดียว เพราะวัดเฉพาะความสามารถ θ_1 หรือ θ_2 เพียงความสามารถเดียว ส่วนในกรณีหลังเป็นการวัดแบบสองมิติ โดยวัดทั้งความสามารถ θ_1 และ θ_2 สำหรับตัวอย่างของข้อสอบ 3 ข้อที่มีทิศทางขนาดของมุม (α_{i1}) 30, 45 และ 60 องศา แสดงดังภาพประกอบ 21



ภาพประกอบ 21 เวกเตอร์ของข้อสอบ 3 ข้อ

ในการศึกษาครั้งนี้ใช้โปรแกรม RESGEN (Muraki. 2000) จำลองผลการตอบข้อสอบ โดยใช้โมเดลในสมการ (1) โปรแกรมดังกล่าวสามารถกำหนดเงื่อนไขของการวิเคราะห์เกี่ยวกับโมเดลของการวิเคราะห์ จำนวนมิติของความสามารถ ค่าพารามิเตอร์ของข้อสอบ ค่าเฉลี่ยของความสามารถ ค่าความแปรปรวนของความสามารถ เมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมของความสามารถ เมทริกซ์สหสัมพันธ์ระหว่างความสามารถ จำนวนกลุ่มผู้สอบ และขนาดของกลุ่มผู้สอบ ผลการจำลองดังกล่าวจะได้คะแนนดิบของผลการตอบข้อสอบสำหรับผู้สอบแต่ละคน และค่าความสามารถของผู้สอบสองมิติ คือ θ_1 และ θ_2 แล้วนำผลการตอบข้อสอบดังกล่าวไปวิเคราะห์การทำหน้าที่เบี่ยงเบนของข้อสอบโดยใช้วิธีโพลิ-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ในการกำหนดเงื่อนไขของการจำลองข้อมูลโดยใช้โปรแกรม RESGEN เป็นไปตามเงื่อนไขของการจัดกระทำตัวแปรที่กล่าวมาแล้วข้างต้น ส่วนการกำหนดค่าพารามิเตอร์ของข้อสอบ สามารถจำแนกออกเป็นสามกรณี คือ ข้อสอบทำหน้าที่ไม่เบี่ยงเบน (No-DIF) ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF) และข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) มีรายละเอียดดังนี้

กรณีที่ 1 ข้อสอบทำหน้าที่ไม่เบี่ยงเบน (No-DIF)

การจำลองข้อสอบทำหน้าที่ไม่เบี่ยงเบน ภายใต้โมเดลพารามิเตอร์หลายมิติ มีเป้าหมายเพื่อให้ความน่าจะเป็นในการตอบข้อสอบถูกระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจ มีค่าเท่ากัน ทุกระดับความสามารถ θ_1 และ θ_2 ซึ่งสามารถทำได้โดยกำหนดให้พารามิเตอร์อำนาจจำแนกของข้อสอบ (a_{1i} และ a_{2i}) และพารามิเตอร์รายการของข้อสอบ (d_{iv}) ระหว่างกลุ่มผู้สอบ มีค่าเท่ากัน ค่าพารามิเตอร์ดังกล่าวแสดงในตาราง 1 และ 2 การกำหนดค่าพารามิเตอร์ดังกล่าว จะต้องคำนึงถึงค่าพารามิเตอร์อำนาจจำแนกของข้อสอบแบบหลายมิติ (α_i) และค่าพารามิเตอร์ความยากของข้อสอบแบบหลายมิติ (β_{iv}) ซึ่งสามารถคำนวณโดยใช้สมการ (2) และ (3) ที่กล่าวมาข้างต้น ทั้งนี้จะต้องเป็นค่าที่สามารถเกิดขึ้นในข้อมูลจริง หรือมีความเป็นไปได้ตามสภาพของการทดสอบจริง โดยส่วนมากแล้วค่าพารามิเตอร์ความยากของข้อสอบแบบหลายมิติจะมีค่าอยู่ในช่วง -2.0 และ 2.0 ส่วนพารามิเตอร์อำนาจจำแนกของข้อสอบแบบหลายมิติจะมีค่าอยู่ในช่วง 0.5 และ 2.5 (Oshima; & Miller. 1990: 276)

การกำหนดค่าพารามิเตอร์อำนาจจำแนกของข้อสอบในมิติที่หนึ่งและมิติที่สองจะมีผลต่อทิศทางของข้อสอบ (Item direction) ซึ่งทิศทางของข้อสอบจะเป็นตัวกำหนดระดับความยากของข้อสอบ ทั้งนี้เพราะว่าเมื่อข้อสอบมีขนาดของมุมต่างกัน จะมีผลทำให้วัดความสามารถในแต่ละมิติแตกต่างกันดังที่กล่าวมาข้างต้น ในที่นี้ได้ออกแบบทิศทางของข้อสอบเป็นระบบทั้ง 36 ข้อ ในข้อสอบทุกๆ 4 ข้อจะมีทิศทางของข้อสอบครบทั้ง 4 ระดับ คือ 15, 30, 60 และ 75 องศา ซึ่งทิศทางของข้อสอบขนาด 15 และ 30 องศาเป็นข้อสอบที่วัดความสามารถ θ_1 มากกว่า θ_2 ส่วนทิศทางของข้อสอบขนาด 60 และ 75 องศาเป็นข้อสอบที่วัดความสามารถ θ_2 มากกว่า θ_1 สำหรับขนาดของมุมสามารถคำนวณโดยใช้สมการ (4) ที่กล่าวมาข้างต้น ในการกำหนดทิศทางของข้อสอบ มุ่งหวังให้สอดคล้องกับปัจจัยเกี่ยวกับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 3 ขนาด คือ 10%, 20% และ 30% (4 ข้อ, 8 ข้อ และ 12 ข้อ ตามลำดับ) ซึ่งการจัดกระทำข้อสอบภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนดังกล่าว จะต้องตัดข้อสอบทำหน้าที่ไม่เบี่ยงเบน และเพิ่มข้อสอบทำหน้าที่เบี่ยงเบน การตัดข้อสอบหรือเพิ่มข้อสอบดังกล่าวจะต้องไม่ให้เกิดผลกระทบต่อเนื่องอื่นๆ ซึ่งมีผลต่ออัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ดังนั้นถ้าตัดข้อสอบหรือเพิ่มข้อสอบที่มีทิศทางอย่างมีระบบทั้ง 4 ระดับ จะทำให้สามารถควบคุมอิทธิพลเกี่ยวกับทิศทางของข้อสอบได้

ผลจากการออกแบบค่าพารามิเตอร์อำนาจจำแนกของข้อสอบและพารามิเตอร์รายการข้อสอบ ทำให้ได้พารามิเตอร์อำนาจจำแนกของข้อสอบแบบหลายมิติ (α_i) มีค่าอยู่ในช่วง 0.54 และ 2.00 โดยมีค่าเฉลี่ย ($\bar{\alpha}_i$) เท่ากับ 1.20 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.42 ส่วนค่าพารามิเตอร์ความยากของข้อสอบแบบหลายมิติ (β_{iv}) มีค่าอยู่ในช่วง -1.84 และ 1.84 โดยมีค่าเฉลี่ยในแต่ละรายการของข้อสอบ คือ $\bar{\beta}_{i2} = -1.16$, $\bar{\beta}_{i3} = -0.45$, $\bar{\beta}_{i4} = 0.68$ และ $\bar{\beta}_{i5} = 1.16$ ($\bar{\beta}_{i1} = 0$ ไม่นำมาพิจารณา) และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.40, 0.35, 0.26 และ 0.40 ตามลำดับ

ตาราง 1 ทิศทางของข้อสอบ พารามิเตอร์อำนาจจำแนกของข้อสอบในแต่ละมิติ และพารามิเตอร์อำนาจจำแนกของข้อสอบแบบหลายมิติ สำหรับกลุ่มอ้างอิงและกลุ่มสนใจ ซึ่งใช้จำลองข้อสอบทำหน้าที่ไม่เบี่ยงเบน

ข้อสอบ	α_{i1}	α_{i2}	a_{i1}	a_{i2}	α_i
1	15	75	0.67	0.18	0.69
2	30	60	0.59	0.34	0.68
3	60	30	0.41	0.71	0.82
4	75	15	0.24	0.90	0.93
5	15	75	0.82	0.22	0.85
6	30	60	0.50	0.29	0.58
7	60	30	0.61	1.06	1.22
8	75	15	0.32	1.19	1.23
9	15	75	0.60	0.16	0.62
10	30	60	0.47	0.27	0.54
11	60	30	0.49	0.85	0.98
12	75	15	0.48	1.79	1.85
13	15	75	1.42	0.38	1.47
14	30	60	1.26	0.73	1.46

ตาราง 1 (ต่อ)

ข้อสอบ	α_{i1}	α_{i2}	a_{i1}	a_{i2}	α_i
15	60	30	1.00	1.73	2.00
16	75	15	0.26	0.97	1.00
17	15	75	0.56	0.15	0.58
18	30	60	1.13	0.65	1.30
19	60	30	0.53	0.92	1.06
20	75	15	0.40	1.49	1.54
21	15	75	0.75	0.20	0.78
22	30	60	0.99	0.57	1.14
23	60	30	0.89	1.54	1.78
24	75	15	0.28	1.04	1.08
25	15	75	1.72	0.46	1.78
26	30	60	1.61	0.93	1.86
27	60	30	0.69	1.19	1.38
28	75	15	0.30	1.12	1.16
29	15	75	1.27	0.34	1.31
30	30	60	1.40	0.81	1.62
31	60	30	0.45	0.78	0.90
32	75	15	0.36	1.34	1.39
33	15	75	1.57	0.42	1.63
34	30	60	0.66	0.38	0.76
35	60	30	0.77	1.33	1.54
36	75	15	0.44	1.64	1.70
Mean			0.75	0.81	1.20
SD			0.42	0.49	0.42

ตาราง 2 พารามิเตอร์รายการของข้อสอบ และพารามิเตอร์ความยากของข้อสอบแบบหลายมิติ
สำหรับกลุ่มอ้างอิงและกลุ่มสนใจ ซึ่งใช้จำลองข้อสอบทำหน้าที่ไม่เป็นเชิงเบน

ข้อสอบ	d_{i2}	d_{i3}	d_{i4}	d_{i5}	β_{i2}	β_{i3}	β_{i4}	β_{i5}
1	1.00	0.50	-0.50	-1.00	-1.44	-0.72	0.72	1.44
2	1.00	0.00	-0.50	-1.00	-1.47	0.00	0.73	1.47
3	1.50	1.00	-1.00	-1.50	-1.83	-1.22	1.22	1.83
4	1.50	0.50	-1.00	-1.50	-1.61	-0.54	1.07	1.61
5	1.00	0.50	-0.50	-1.00	-1.18	-0.59	0.59	1.18
6	1.00	0.00	-0.50	-1.00	-1.73	0.00	0.87	1.73
7	1.50	1.00	-1.00	-1.50	-1.23	-0.82	0.82	1.23
8	1.50	0.50	-1.00	-1.50	-1.22	-0.41	0.81	1.22
9	1.00	0.50	-0.50	-1.00	-1.61	-0.81	0.81	1.61
10	1.00	0.00	-0.50	-1.00	-1.84	0.00	0.92	1.84
11	1.50	1.00	-1.00	-1.50	-1.53	-1.02	1.02	1.53
12	1.50	0.50	-1.00	-1.50	-0.81	-0.27	0.54	0.81
13	1.00	0.50	-0.50	-1.00	-0.68	-0.34	0.34	0.68
14	1.00	0.00	-0.50	-1.00	-0.69	0.00	0.34	0.69
15	1.50	1.00	-1.00	-1.50	-0.75	-0.50	0.5	0.75
16	1.50	0.50	-1.00	-1.50	-1.49	-0.50	1.00	1.49
17	1.00	0.50	-0.50	-1.00	-1.72	-0.86	0.86	1.72
18	1.00	0.00	-0.50	-1.00	-0.77	0.00	0.38	0.77
19	1.50	1.00	-1.00	-1.50	-1.41	-0.94	0.94	1.41
20	1.50	0.50	-1.00	-1.50	-0.97	-0.32	0.65	0.97
21	1.00	0.50	-0.50	-1.00	-1.29	-0.64	0.64	1.29
22	1.00	0.00	-0.50	-1.00	-0.88	0.00	0.44	0.88
23	1.50	1.00	-1.00	-1.50	-0.84	-0.56	0.56	0.84

ตาราง 2 (ต่อ)

ข้อสอบ	d_{i2}	d_{i3}	d_{i4}	d_{i5}	β_{i2}	β_{i3}	β_{i4}	β_{i5}
24	1.50	0.50	-1.00	-1.50	-1.39	-0.46	0.93	1.39
25	1.00	0.50	-0.50	-1.00	-0.56	-0.28	0.28	0.56
26	1.00	0.00	-0.50	-1.00	-0.54	0.00	0.27	0.54
27	1.50	1.00	-1.00	-1.50	-1.09	-0.73	0.73	1.09
28	1.50	0.50	-1.00	-1.50	-1.29	-0.43	0.86	1.29
29	1.00	0.50	-0.50	-1.00	-0.76	-0.38	0.38	0.76
30	1.00	0.00	-0.50	-1.00	-0.62	0.00	0.31	0.62
31	1.50	1.00	-1.00	-1.50	-1.67	-1.11	1.11	1.67
32	1.50	0.50	-1.00	-1.50	-1.08	-0.36	0.72	1.08
33	1.00	0.50	-0.50	-1.00	-0.62	-0.31	0.31	0.62
34	1.00	0.00	-0.50	-1.00	-1.31	0.00	0.66	1.31
35	1.50	1.00	-1.00	-1.50	-0.98	-0.65	0.65	0.98
36	1.50	0.50	-1.00	-1.50	-0.88	-0.29	0.59	0.88
Mean	1.25	0.50	-0.75	-1.25	-1.16	-0.45	0.68	1.16
SD	0.25	0.36	0.25	0.25	0.40	0.35	0.26	0.40

กรณีที่ 2 ข้อสอบเบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF)

การจำลองข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน ภายใต้โมเดลพหุคูณเชิงเส้นเครดิตทั่วไปแบบหลายมิติ มีเป้าหมายเพื่อให้ความแตกต่างของความน่าจะเป็นในการตอบข้อสอบถูกระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจมีค่าคงที่ ทุกระดับความสามารถ θ_1 และ θ_2 ในที่นี้กำหนดให้พารามิเตอร์อำนาจจำแนกของข้อสอบในมิติที่หนึ่งและมิติที่สอง (a_{1i} และ a_{2i}) ระหว่างกลุ่มมีค่าเท่ากัน ในขณะที่พารามิเตอร์รายการของข้อสอบ (d_{iv}) ระหว่างกลุ่มมีค่าแตกต่างกันเท่ากับ 0.20 โดยให้พารามิเตอร์รายการของข้อสอบในกลุ่มสนใจมีค่ามากกว่ากลุ่มอ้างอิงในทุกรายการ ในการกำหนดค่าพารามิเตอร์ดังกล่าวมีผลต่อรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน โดยทำให้

ตาราง 5 พารามิเตอร์รายการของข้อสอบ และพารามิเตอร์ความยากของข้อสอบแบบหลายมิติ สำหรับกลุ่มสนใจ ซึ่งใช้จำลองข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน

ข้อสอบ	d_{i2}	d_{i3}	d_{i4}	d_{i5}	β_{i2}	β_{i3}	β_{i4}	β_{i5}
1	1.30	0.30	-0.70	-1.70	-1.08	-0.25	0.58	1.42
2	1.30	0.30	-0.70	-1.70	-1.08	-0.25	0.58	1.42
3	1.30	0.30	-0.70	-1.70	-1.08	-0.25	0.58	1.42
4	1.30	0.30	-0.70	-1.70	-1.08	-0.25	0.58	1.42
Mean	1.30	0.30	-0.70	-1.70	-1.08	-0.25	0.58	1.42
SD	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

จากตาราง 3 จะเห็นว่า ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบในมิติที่หนึ่งและมิติที่สอง จำนวน 4 ข้อ กำหนดทิศทางของข้อสอบอย่างเป็นระบบ คือ 15, 30, 60 และ 75 องศา ซึ่งสอดคล้องกับปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนจำนวน 10% (4 ข้อ) ลักษณะดังกล่าว คล้ายกับการจำลองในกรณีข้อสอบทำหน้าที่ไม่เบี่ยงเบน ผลจากการออกแบบค่าพารามิเตอร์อำนาจจำแนกของข้อสอบในแต่ละมิติทำให้ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบแบบหลายมิติ (α_i) มีค่าเท่ากันทั้ง 4 ข้อ คือ 1.20 เมื่อพิจารณาค่าพารามิเตอร์ดังกล่าวพบว่า มีค่าเท่ากับค่าเฉลี่ยของพารามิเตอร์อำนาจจำแนกของข้อสอบแบบหลายมิติ ($\bar{\alpha}_i$) ซึ่งใช้ในการจำลองกรณี **ข้อสอบทำหน้าที่ไม่เบี่ยงเบน** แสดงว่าพารามิเตอร์อำนาจจำแนกของข้อสอบแบบหลายมิติทั้งสองกรณีดังกล่าวมีค่าใกล้เคียงกัน สำหรับปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนจำนวน 20% และ 30% (8 ข้อและ 12 ข้อ ตามลำดับ) จะใช้ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบชุดเดิม โดยเพิ่มตามสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน การออกแบบค่าพารามิเตอร์ดังกล่าวเพื่อหลีกเลี่ยงปัจจัยของค่าพารามิเตอร์ของข้อสอบ ที่อาจมีผลกระทบต่อการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ดังจะเห็นได้จากผลการศึกษาของ ชาง มาซซซีโอ และรูสโซ (Chang; Mazzeo; & Roussos. 1996: 347) ที่พบว่า เมื่อข้อสอบมีค่าอำนาจจำแนกเพิ่มขึ้น มีผลทำให้วิธีแมนเทล (Mantel) และวิธีความแตกต่างเฉลี่ยมาตรฐาน (Standardized mean difference; SMD) มีอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้น

ตาราง 4 และ 5 เป็นการออกแบบค่าพารามิเตอร์รายการของข้อสอบ (d_{iv}) สำหรับกลุ่มอ้างอิงและกลุ่มสนใจ ตามลำดับ โดยกำหนดให้มีค่าแตกต่างกัน 0.20 ค่าดังกล่าวเป็นขนาดของข้อสอบทำหน้าที่เบี่ยงเบน (Magnitude of DIF) ค่าพารามิเตอร์รายการของข้อสอบเกี่ยวข้องกับ

ค่าพารามิเตอร์ความยากของข้อสอบแบบหลายมิติ (β_{iv}) ดังสมการ (3) ที่กล่าวมาข้างต้น จากงานวิจัยของ โอชิมา ราจู และเฟลวเวอร์ส (Oshima; Raju; & Flowers. 1997) ที่ศึกษาการทำหน้าที่เบี่ยงเบนของข้อสอบ ในข้อสอบที่วัดความสามารถสองมิติและให้คะแนนสองค่า ได้กำหนดพารามิเตอร์ d_i ระหว่างกลุ่มมีค่าแตกต่างกัน 0.50 แต่ในการศึกษาครั้งนี้ได้กำหนดพารามิเตอร์ d_{iv} ระหว่างกลุ่มมีค่าแตกต่างกัน 0.20 ซึ่งมีค่าน้อยกว่า เนื่องจากในครั้งนี้ได้ศึกษาในข้อสอบที่วัดความสามารถสองมิติและให้คะแนนห้าค่า การคำนวณขนาดของข้อสอบทำหน้าที่เบี่ยงเบนจะต้องใช้ความน่าจะเป็นสะสมของการตอบข้อสอบทั้ง 5 รายการ ซึ่งมีขนาดมากกว่าเงื่อนไขของการตอบข้อสอบเพียงรายการเดียว (ให้คะแนนสองค่า) ดังนั้นในการกำหนดขนาดของข้อสอบทำหน้าที่เบี่ยงเบนจะต้องควบคุมไม่ให้มากผิดปกติ เพราะถ้ามีขนาดเพิ่มขึ้นจะส่งผลให้อัตราความถูกต้องของการตรวจสอบเพิ่มขึ้น (Narayanan; & Swaminathan. 1996: 263) นอกจากนี้ยังสอดคล้องกับงานวิจัยที่ศึกษาการทำหน้าที่เบี่ยงเบนของข้อสอบ ในข้อสอบที่วัดความสามารถมิติเดียวและให้คะแนนหลายค่า ซึ่งกำหนดขนาดของข้อสอบทำหน้าที่เบี่ยงเบนในช่วง 0.10 และ 0.25 (Chang; Mazzeo; & Roussos. 1996: 34; Zwick; Donoghue; & Grima. 1993: 243)

การกำหนดค่าพารามิเตอร์รายการของข้อสอบจำนวน 4 ข้อ ในแต่ละกลุ่มผู้สอบจะกำหนดให้มีค่าเท่ากันทั้ง 4 ข้อ ซึ่งมีเหตุผลทำนองเดียวกับการกำหนดค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ เพื่อควบคุมอิทธิพลของค่าพารามิเตอร์ที่อาจส่งผลกระทบต่อตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ จากการออกแบบค่าพารามิเตอร์อำนาจจำแนกของข้อสอบและพารามิเตอร์รายการของข้อสอบ ทำให้ได้ค่าพารามิเตอร์ความยากของข้อสอบแบบหลายมิติสำหรับ**กลุ่มอ้างอิง**มีค่าอยู่ในช่วง -1.25 และ 1.25 โดยมีค่าเฉลี่ยในแต่ละรายการของข้อสอบ คือ $\bar{\beta}_{iR2} = -1.25$, $\bar{\beta}_{iR3} = -0.42$, $\bar{\beta}_{iR4} = 0.42$ และ $\bar{\beta}_{iR5} = 1.25$ ($\bar{\beta}_{iR1} = 0$ ไม่นำมาพิจารณา) และมีส่วนเบี่ยงเบนมาตรฐานเท่ากันทุกรายการ คือ 0 สำหรับ**กลุ่มสนใจ**มีค่าอยู่ในช่วง -1.08 และ 1.42 โดยมีค่าเฉลี่ยในแต่ละรายการของข้อสอบ คือ $\bar{\beta}_{iF2} = -1.08$, $\bar{\beta}_{iF3} = -0.25$, $\bar{\beta}_{iF4} = 0.58$ และ $\bar{\beta}_{iF5} = 1.42$ ($\bar{\beta}_{iF1} = 0$ ไม่นำมาพิจารณา) และมีส่วนเบี่ยงเบนมาตรฐานเท่ากันทุกรายการ คือ 0

กรณีที่ 3 ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF)

การจำลองข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ภายใต้โมเดลพาเซียลเครดิตทั่วไปแบบหลายมิติ มีเป้าหมายเพื่อให้ความแตกต่างของความน่าจะเป็นในการตอบข้อสอบถูกระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจมีค่าแตกต่างกัน ทุกระดับความสามารถ θ_1 และ θ_2 ในที่นี้กำหนดให้พารามิเตอร์รายการของข้อสอบ (d_{iv}) ระหว่างกลุ่มมีค่าเท่ากัน ส่วนพารามิเตอร์อำนาจ

จำแนกของข้อสอบในมิติที่หนึ่งและมิติที่สอง (a_{1i} และ a_{2i}) ระหว่างกลุ่มมีค่าแตกต่างกันเท่ากับ 0.15 ค่าพารามิเตอร์ดังกล่าวแสดงในตาราง 6-9

ตาราง 6 ทิศทางของข้อสอบ พารามิเตอร์อำนาจจำแนกของข้อสอบในแต่ละมิติ และพารามิเตอร์อำนาจจำแนกของข้อสอบแบบหลายมิติ สำหรับกลุ่มอ้างอิง ซึ่งใช้จำลองข้อสอบทำหน้าที่ เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน

ข้อสอบ	α_{i1}	α_{i2}	a_{i1}	a_{i2}	α_i
1	45	45	0.85	0.85	1.20
2	45	45	0.85	0.85	1.20
3	45	45	0.70	0.70	0.99
4	45	45	1.00	1.00	1.41
Mean			0.85	0.85	1.20
SD			0.12	0.12	0.17

ตาราง 7 ทิศทางของข้อสอบ พารามิเตอร์อำนาจจำแนกของข้อสอบในแต่ละมิติ และพารามิเตอร์อำนาจจำแนกของข้อสอบแบบหลายมิติ สำหรับกลุ่มสนใจ ซึ่งใช้จำลองข้อสอบทำหน้าที่ เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน

ข้อสอบ	α_{i1}	α_{i2}	a_{i1}	a_{i2}	α_i
1	45	45	0.70	0.70	0.99
2	45	45	1.00	1.00	1.41
3	45	45	0.85	0.85	1.20
4	45	45	0.85	0.85	1.20
Mean			0.85	0.85	1.20
SD			0.12	0.12	0.17

ตาราง 8 พารามิเตอร์รายการของข้อสอบ และพารามิเตอร์ความยากของข้อสอบแบบหลายมิติ สำหรับกลุ่มอ้างอิง ซึ่งใช้จำลองข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน

ข้อสอบ	d_{i2}	d_{i3}	d_{i4}	d_{i5}	β_{i2}	β_{i3}	β_{i4}	β_{i5}
1	1.50	0.50	-0.50	-1.50	-1.25	-0.42	0.42	1.25
2	1.50	0.50	-0.50	-1.50	-1.25	-0.42	0.42	1.25
3	1.50	0.50	-0.50	-1.50	-1.52	-0.51	0.51	1.52
4	1.50	0.50	-0.50	-1.50	-1.06	-0.35	0.35	1.06
Mean	1.50	0.50	-0.50	-1.50	-1.27	-0.42	0.42	1.27
SD	0.00	0.00	0.00	0.00	0.19	0.06	0.06	0.19

ตาราง 9 พารามิเตอร์รายการของข้อสอบ และพารามิเตอร์ความยากของข้อสอบแบบหลายมิติ สำหรับกลุ่มสนใจ ซึ่งใช้จำลองข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน

ข้อสอบ	d_{i2}	d_{i3}	d_{i4}	d_{i5}	β_{i2}	β_{i3}	β_{i4}	β_{i5}
1	1.50	0.50	-0.50	-1.50	-1.52	-0.51	0.51	1.52
2	1.50	0.50	-0.50	-1.50	-1.06	-0.35	0.35	1.06
3	1.50	0.50	-0.50	-1.50	-1.25	-0.42	0.42	1.25
4	1.50	0.50	-0.50	-1.50	-1.25	-0.42	0.42	1.25
Mean	1.50	0.50	-0.50	-1.50	-1.27	-0.42	0.42	1.27
SD	0.00	0.00	0.00	0.00	0.19	0.06	0.06	0.19

ตาราง 6 และ 7 เป็นการออกแบบค่าพารามิเตอร์อำนาจจำแนกของข้อสอบในมิติที่หนึ่ง และมิติที่สองจำนวน 4 ข้อ สำหรับผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจ ตามลำดับ จะเห็นว่าทิศทางของข้อสอบมีขนาด 45 องศาเท่ากันทุกข้อ แสดงว่าข้อสอบดังกล่าววัดความสามารถในมิติที่หนึ่ง และมิติที่สองเท่ากัน ในข้อสอบข้อ 1 และ 2 กำหนดให้ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบทั้งสองมิติในกลุ่มอ้างอิงมีค่ามากกว่ากลุ่มสนใจอยู่ 0.15 ส่วนข้อ 3 และ 4 กำหนดค่าพารามิเตอร์

อำนาจจำแนกของข้อสอบตรงข้ามกับข้อ 1 และ 2 โดยใช้ค่าพารามิเตอร์ชุดเดิมแต่สลับกลุ่มผู้สอบ สำหรับตาราง 8 และ 9 เป็นการออกแบบค่าพารามิเตอร์รายการของข้อสอบจำนวน 4 ข้อ สำหรับผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจ ตามลำดับ โดยกำหนดให้ค่าพารามิเตอร์ในทุกรายการของข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มสนใจมีค่าเท่ากัน และใช้ค่าพารามิเตอร์ชุดเดียวกันทั้ง 4 ข้อ การกำหนดค่าพารามิเตอร์อำนาจจำแนกและพารามิเตอร์รายการของข้อสอบดังกล่าวมีผลต่อรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน โดยทำให้ฟังก์ชันความน่าจะเป็นของการตอบข้อสอบในทุก รายการ ซึ่งแทนด้วยผิวหน้าการตอบรายการ (Option response surfaces; ORSs) ระหว่างกลุ่มอ้างอิงและกลุ่มสนใจตัดกันอย่างสมดุลงាំងสองมิติ ในช่วงความสามารถ θ_1 หรือ θ_2 ในการกำหนดทิศทางของข้อสอบดังกล่าวทำให้สามารถควบคุมขนาดของข้อสอบทำหน้าที่เบี่ยงเบน และค่าพารามิเตอร์อำนาจจำแนกของข้อสอบแบบหลายมิติ ซึ่งมีผลต่ออัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1

สาเหตุที่ผู้วิจัยกำหนดค่าพารามิเตอร์อำนาจจำแนกของข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มสนใจมีค่าแตกต่างกัน 0.15 มีเหตุผลทำนองเดียวกับการกำหนดค่าพารามิเตอร์รายการของข้อสอบซึ่งได้กล่าวมาข้างต้น โดยที่งานวิจัยของโอชิมะ ราจู และเฟลโลเวอร์ส (Oshima; Raju; & Flowers. 1997) ที่ศึกษาการทำหน้าที่เบี่ยงเบนของข้อสอบในข้อสอบที่วัดความสามารถสองมิติ และให้คะแนนสองค่า ได้กำหนดพารามิเตอร์อำนาจจำแนกในแต่ละมิติมีค่าแตกต่างกัน 0.30 ผลการออกแบบค่าพารามิเตอร์ดังกล่าว ทำให้ได้ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบแบบหลายมิติ (α_i) สำหรับกลุ่มอ้างอิงและกลุ่มสนใจมีค่าอยู่ในช่วง 0.99 และ 1.41 โดยมีค่าเฉลี่ย ($\bar{\alpha}_i$) เท่ากับ 1.20 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.17 จะเห็นว่า ค่าเฉลี่ยของพารามิเตอร์อำนาจจำแนกของข้อสอบแบบหลายมิติดังกล่าวเท่ากับค่าเฉลี่ยที่จำลองข้อสอบในกรณีที่ 1 และ 2 ส่วนค่าพารามิเตอร์ความยากของข้อสอบแบบหลายมิติสำหรับกลุ่มอ้างอิงและกลุ่มสนใจมีค่าอยู่ในช่วง -1.52 และ 1.52 โดยมีค่าเฉลี่ยในแต่ละรายการของข้อสอบ คือ $\bar{\beta}_{i2} = -1.27$, $\bar{\beta}_{i3} = -0.42$, $\bar{\beta}_{i4} = 0.42$ และ $\bar{\beta}_{i5} = 1.27$ ($\bar{\beta}_{i1} = 0$ ไม่นำมาพิจารณา) และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.19, 0.06, 0.06 และ 0.19 ตามลำดับ

การวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูล มีรายละเอียดดังนี้

1. การวิเคราะห์การทำหน้าที่เบี่ยงเบนของข้อสอบ

การวิเคราะห์การทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีโพลี-ชิปเทสท์ (Poly-SIBTEST) วิธีการถดถอยโลจิสติกแบบจัดอันดับ (OLR) และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ (MOLR) จะใช้ผลการจำลองการตอบข้อสอบซึ่งเป็นคะแนนดิบมาวิเคราะห์ โดยวิธีแรกใช้โปรแกรม DIFPACK ส่วนสองวิธีหลังใช้โปรแกรม SPSS สำหรับวิธีโพลี-ชิปเทสท์และวิธีการถดถอยโลจิสติกแบบจัดอันดับจะนำผลการตอบข้อสอบรายข้อมาวิเคราะห์ และใช้คะแนนรวมของแบบทดสอบ แทนระดับความสามารถของผู้สอบ คะแนนรวมของวิธีโพลี-ชิปเทสท์ค่อนข้างเป็นค่าประมาณความสามารถมากกว่าคะแนนสอบ ส่วนคะแนนรวมของวิธีการถดถอยโลจิสติกแบบจัดอันดับ ถือว่าเป็นความสามารถแบบต่อเนื่อง จะเห็นว่าทั้งสองวิธีมีกรอบแนวคิดภายใต้ทฤษฎีการตอบข้อสอบ ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติจะนำผลการตอบรายข้อมาวิเคราะห์ เช่นเดียวกัน แต่จะใช้ค่าประมาณความสามารถทั้งสองมิติแทนระดับความสามารถของผู้สอบ ค่าประมาณความสามารถดังกล่าวได้มาจากการจำลองข้อมูล ภายใต้ทฤษฎีการตอบข้อสอบแบบหลายมิติและให้คะแนนหลายค่า ในการพิจารณาตัดสินการทำหน้าที่เบี่ยงเบนของข้อสอบของทั้งสามวิธีจะใช้ระดับนัยสำคัญ .05 สำหรับรายละเอียดของการวิเคราะห์ด้วยวิธีโพลี-ชิปเทสท์และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ผู้วิจัยนำเสนอไว้แล้วในบทที่ 2 ส่วนการวิเคราะห์ด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติจะกล่าวถึงดังต่อไปนี้

การวิเคราะห์การทำหน้าที่เบี่ยงเบนของข้อสอบโดยใช้วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ผู้วิจัยปรับขยายโมเดลการถดถอยโลจิสติกแบบจัดอันดับ (Zumbo, 1999: 57) ที่ใช้ในการวิเคราะห์กับข้อมูลมิติเดียว โดยเพิ่มพจน์ของความสามารถ และปฏิสัมพันธ์ระหว่างความสามารถ กับกลุ่มผู้สอบเพื่อวิเคราะห์กับข้อมูลสองมิติ ดังนี้

$$\log\left(\frac{P(Y \leq j)}{1 - P(Y \leq j)}\right) = \beta_{0j} + \beta_1\theta_1 + \beta_2\theta_2 + \beta_3G + \beta_4(\theta_1G) + \beta_5(\theta_2G) + \beta_6(\theta_1\theta_2G) \quad (1)$$

หรือเขียนในรูปโมเดลโลจิต (logit) ดังนี้

$$\text{logit}[P(Y \leq j)] = \beta_{0j} + \beta_1\theta_1 + \beta_2\theta_2 + \beta_3G + \beta_4(\theta_1G) + \beta_5(\theta_2G) + \beta_6(\theta_1\theta_2G) \quad (2)$$

เมื่อ

$P(Y \leq j)$ แทน ความน่าจะเป็นสะสมของผลการตอบข้อสอบในรายการคำตอบที่ j
หรือน้อยกว่า j ($j = 1, \dots, J - 1$ เมื่อ J แทนจำนวนรายการทั้งหมด)

θ_1 แทน ระดับความสามารถในมิติที่หนึ่ง

θ_2 แทน ระดับความสามารถในมิติที่สอง

G แทน กลุ่มผู้สอบ

$\theta_1 G$ แทน ปฏิสัมพันธ์สองทางระหว่างระดับความสามารถในมิติที่หนึ่ง กับกลุ่มผู้สอบ

$\theta_2 G$ แทน ปฏิสัมพันธ์สองทางระหว่างระดับความสามารถในมิติที่สอง กับกลุ่มผู้สอบ

$\theta_1 \theta_2 G$ แทน ปฏิสัมพันธ์สามทางระหว่างระดับความสามารถในมิติที่หนึ่ง ระดับความสามารถในมิติที่สอง กับกลุ่มผู้สอบ

β_{0j} แทน พารามิเตอร์ส่วนตัด $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ และ β_6 แทนพารามิเตอร์ที่เกี่ยวข้องกับพจน์ของตัวแปรแต่ละตัวในโมเดลดังกล่าว

การวิเคราะห์การทำหน้าที่เบี่ยงเบนของข้อสอบจะใช้โมเดลการถดถอยโลจิสติกแบบจัดอันดับทำนายความน่าจะเป็นสะสมของอัตราส่วนแต้มต่อ (Odds ratio) ในรายการคำตอบที่ 1 ถึง $J - 1$ เท่านั้น เพราะความน่าจะเป็นที่จะตอบในรายการสุดท้ายหรือต่ำกว่ามีค่าเป็น 1 ในการศึกษาครั้งนี้จำลองผลการตอบข้อสอบ 40 ข้อ ข้อสอบทุกข้อมีรายการตอบ 5 รายการ ในรายการ 1, 2, 3, 4 และ 5 ให้คะแนนเป็น 1, 2, 3, 4 และ 5 คะแนน ตามลำดับ ดังนั้นผลการตอบข้อสอบในแต่ละข้อเป็นตัวแปรตาม ส่วนค่าประมาณความสามารถในมิติที่หนึ่ง มิติที่สอง (θ_1 และ θ_2) และกลุ่มผู้สอบ (G) เป็นตัวแปรอิสระ ในการดำเนินการวิเคราะห์การทำหน้าที่เบี่ยงเบนของข้อสอบใช้การทดสอบความเหมาะสมของข้อมูลกับโมเดล โดยมีขั้นตอนการทดสอบ 3 ขั้นตอน ดังนี้

ขั้นตอนที่หนึ่ง ทดสอบตัวแปรที่ใช้ในการจับคู่ความสามารถ ในที่นี้ใช้ค่าประมาณความสามารถในมิติที่หนึ่งและมิติที่สองแทนความสามารถของผู้สอบ ในขั้นตอนนี้ถือว่าอิทธิพลของตัวแปรทำนายมีค่าเท่ากันในทุกรายการตอบ การนำตัวแปรดังกล่าวเข้าไปในสมการถดถอยมีลักษณะดังนี้

$$\text{logit}[P(Y \leq j)] = \beta_{0j} + \beta_1 \theta_1 + \beta_2 \theta_2 \quad (3)$$

ขั้นตอนที่สอง นำตัวแปรการเป็นสมาชิกของกลุ่มผู้สอบเข้าไปในสมการของขั้นตอนที่หนึ่ง ในที่นี้จะกำหนดรหัสเป็นตัวแปรดัมมี่ (Dummy variable) โดยให้กลุ่มอ้างอิงแทนด้วย 0 และกลุ่มสนใจแทนด้วย 1 การทดสอบความเหมาะสมของข้อมูลกับโมเดลในขั้นตอนที่สองจะรวมทั้งตัวแปรความสามารถและกลุ่มผู้สอบ โดยมีโมเดลที่ใช้ในการทดสอบดังนี้

$$\text{logit}[P(Y \leq j)] = \beta_{0j} + \beta_1\theta_1 + \beta_2\theta_2 + \beta_3G \quad (4)$$

ขั้นตอนที่สาม โมเดลในขั้นตอนนี้จะรวมตัวแปรที่จะใช้ทดสอบทั้งหมด คือ ตัวแปรความสามารถในมิติที่หนึ่ง ตัวแปรความสามารถในมิติที่สอง ตัวแปรกลุ่มผู้สอบ ตัวแปรปฏิสัมพันธ์ระหว่างความสามารถในมิติที่หนึ่ง กับกลุ่มผู้สอบ ตัวแปรปฏิสัมพันธ์ระหว่างความสามารถในมิติที่สอง กับกลุ่มผู้สอบ และตัวแปรปฏิสัมพันธ์ระหว่างความสามารถในมิติที่หนึ่ง ความสามารถในการมิติที่สอง กับกลุ่มผู้สอบ โมเดลของการทดสอบมีดังนี้

$$\text{logit}[P(Y \leq j)] = \beta_{0j} + \beta_1\theta_1 + \beta_2\theta_2 + \beta_3G + \beta_4(\theta_1G) + \beta_5(\theta_2G) + \beta_6(\theta_1\theta_2G) \quad (5)$$

ในการตัดสินใจการทำหน้าที่เบี่ยงเบนของข้อสอบ จะเปรียบเทียบความแตกต่างระหว่างค่าไค-ก้างสองที่คำนวณได้ในขั้นตอนที่หนึ่งและขั้นตอนที่สาม ที่ระดับชั้นของความเป็นอิสระเท่ากับ 4 ($df = 4$) เนื่องจากในขั้นตอนที่หนึ่งมี $df = 2$ และในขั้นตอนที่สามมี $df = 6$ ดังนั้นจะทดสอบค่าไค-ก้างสองที่ระดับ $df = 4$ (นำ df ในสองขั้นตอนมาลบกัน) ซึ่งเป็นการทดสอบพร้อมๆ กัน ระหว่างข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF) และข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) โดยการเปรียบเทียบความแตกต่างของค่า R^2 ระหว่างขั้นตอนที่หนึ่งกับขั้นตอนที่สอง และระหว่างขั้นตอนที่สองกับขั้นตอนที่สาม ถ้าความแตกต่างในช่วงแรกมากกว่าช่วงหลัง แสดงว่าข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน แต่ถ้าความแตกต่างในช่วงแรกน้อยกว่าช่วงหลัง แสดงว่าข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน

2. การคำนวณอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1

การคำนวณอัตราความถูกต้อง (Power rate) และอัตราความคลาดเคลื่อนประเภทที่ 1 (Type I error rate) ของวิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอย-โลจิสติกแบบจัดอันดับหลายมิติ ใช้สูตรการคำนวณดังนี้

$$P = \frac{n_1}{N_1} \times 100 \quad (6)$$

$$E_1 = \frac{n_2}{N_2} \times 100 \quad (7)$$

เมื่อ

P แทน อัตราความถูกต้อง

E_1 แทน อัตราความคลาดเคลื่อนประเภทที่ 1

n_1 แทน จำนวนข้อสอบที่ตรวจสอบได้ถูกต้องว่าทำหน้าที่เป็นอย่างดี

n_2 แทน จำนวนข้อสอบที่ตรวจสอบผิดพลาดว่าทำหน้าที่เป็นอย่างดี

N_1 แทน จำนวนข้อสอบทำหน้าที่เป็นอย่างดีในแบบทดสอบ

N_2 แทน จำนวนข้อสอบทำหน้าที่ไม่เป็นอย่างดีในแบบทดสอบ

3. การเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1

ในการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้วัย 5 ตัว คือ วิธีการตรวจสอบการทำหน้าที่เป็นอย่างดีของข้อสอบ ปัจจัยรูปแบบของข้อสอบทำหน้าที่เป็นอย่างดี ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เป็นอย่างดี ปัจจัยความแตกต่างของการแจกแจงความสามารถ และปัจจัยขนาดตัวอย่าง ใช้สถิติการวิเคราะห์ความแปรปรวนหลายตัวแปร (Multivariate analysis of variance; MANOVA) ที่ระดับนัยสำคัญ .001 โดยกำหนดโมเดลของการวิเคราะห์ให้มีปฏิสัมพันธ์ระหว่างตัวแปรอิสระไม่เกินอันดับที่สอง ถ้าผลการวิเคราะห์มีนัยสำคัญทางสถิติจะทดสอบผลระหว่างกลุ่ม (Test of between-subjects effects) ของตัวแปรตามแต่ละตัวที่ระดับนัยสำคัญ .001 จากนั้นจึงทดสอบผลย่อย (Simple effect) ภายใต้วัยที่ศึกษาที่ระดับ .001 แล้วทดสอบภายหลังด้วยวิธีของเชฟเฟ (Scheffé) โดยใช้ระดับนัยสำคัญระดับเดียวกับการทดสอบผลย่อย

ในการพิจารณาผลการทดสอบนัยสำคัญจากการวิเคราะห์ความแปรปรวนหลายตัวแปร จะพิจารณาเฉพาะประเด็นที่ตอบคำถามการวิจัย โดยพิจารณาผลการทดสอบปฏิสัมพันธ์สองทางและสามทาง ระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย กับวิธีการตรวจสอบการทำหน้าที่เป็นอย่างดีของข้อสอบ ส่วนการทดสอบผลย่อยจะยึดกรอบต่อไปนี้

3.1 เมื่อผลการทดสอบปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย กับวิธีการตรวจสอบมีนัยสำคัญทางสถิติ จะทดสอบผลย่อย 2 กรณี ดังนี้

กรณีที่ 1 การทดสอบผลย่อยของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ในแต่ละวิธีการตรวจสอบ

กรณีที่ 2 การทดสอบผลย่อยของวิธีการตรวจสอบ ในแต่ละเงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย

3.2 เมื่อผลการทดสอบปฏิสัมพันธ์สามทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย กับวิธีการตรวจสอบมีนัยสำคัญทางสถิติ จะทดสอบผลย่อย 2 กรณี ดังนี้

กรณีที่ 1 การทดสอบผลย่อยของปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ในแต่ละวิธีการตรวจสอบ

กรณีที่ 2 การทดสอบผลย่อยของวิธีการตรวจสอบ ในแต่ละเงื่อนไขของปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย

สำหรับเกณฑ์ที่ใช้ในการตัดสินประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ พิจารณาจากผลการตรวจสอบที่มีอัตราความถูกต้องสูง และอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำ ซึ่งมีค่าไม่เกิน 5% ($\alpha = .05$)

การศึกษานำร่อง

ในการวิจัยครั้งนี้ศึกษาในสถานการณ์จำลอง โดยใช้โปรแกรม RESGEN (Muraki. 2000) จำลองข้อมูลภายใต้โมเดลพหุเชิงเส้นเครดิตทั่วไปแบบหลายมิติ (Multidimensional generalized partial credit model) ในปัจจุบันนี้การวิเคราะห์ข้อมูลโดยใช้โมเดลดังกล่าวยังไม่เป็นที่แพร่หลายมากนัก ส่วนมากนักวิจัยนำมาใช้ศึกษาในเชิงทฤษฎีโดยการจำลองข้อมูล ดังนั้นจึงมีข้อจำกัดเกี่ยวกับโปรแกรมที่ใช้ในการวิเคราะห์ข้อมูลเชิงประจักษ์ ในครั้งนี้ผู้วิจัยทดลองจำลองข้อมูลแล้วตรวจสอบผลการจำลอง ซึ่งสามารถตรวจสอบได้เฉพาะการวิเคราะห์ข้อมูลแบบมิติเดียว โดยทดลองจำลองข้อมูลภายใต้โมเดลโลจิสติก (Logistic model) แบบ 1, 2 และ 3 พารามิเตอร์ โมเดลเกรดเรสปอน (Grade response model) และโมเดลพหุเชิงเส้นเครดิตทั่วไป (Generalized partial credit Model) แล้วนำผลการจำลองซึ่งเป็นข้อมูลดิบไปวิเคราะห์ภายใต้โมเดลดังกล่าว โดยใช้โปรแกรม BILOG-MG, MULTILOG และ PARSCALE ตามลำดับ ปรากฏว่า ค่าพารามิเตอร์ของข้อสอบที่กำหนดขึ้นเพื่อใช้จำลองข้อมูล สอดคล้องกับค่าพารามิเตอร์ที่ได้จากวิเคราะห์ นอกจากนี้ค่าเฉลี่ยและความแปรปรวนของค่าความสามารถก็สอดคล้องกันด้วย ผลการตรวจสอบเบื้องต้นแสดงให้เห็นว่า โปรแกรม RESGEN มีประสิทธิภาพในการจำลองข้อมูล

ผู้วิจัยทดลองศึกษานำร่อง โดยจำลองข้อมูล 72 เงื่อนไข ($2 \times 3 \times 3 \times 4$) ภายใต้รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน 2 รูปแบบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 3 ขนาด ความแตกต่างของการแจกแจงความสามารถ 3 ระดับ และขนาดตัวอย่าง 3 ขนาด โดยใช้ค่าพารามิเตอร์ของข้อสอบที่กำหนดไว้ในขั้นตอนการจำลองข้อมูล แล้วนำผลการจำลองข้อมูลไปวิเคราะห์การทำหน้าที่เบี่ยงเบนของข้อสอบ 3 วิธี คือ วิธีโพลี-ซิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ จากนั้นวิเคราะห์อัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ข้อมูล 72 เงื่อนไข ผลการวิเคราะห์แสดงดังตาราง 10

ตาราง 10 ค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ซิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย

ปัจจัย	วิธีการตรวจสอบ DIF					
	Poly-SIBTEST		OLR		MOLR	
	<i>P</i>	<i>E</i> ₁	<i>P</i>	<i>E</i> ₁	<i>P</i>	<i>E</i> ₁
รูปแบบของ DIF						
UDIF	74.42	2.39	78.01	19.08	79.51	8.04
NDIF	22.57	7.08	76.85	13.22	81.48	9.20
สัดส่วนของ DIF						
10%	47.92	6.25	79.17	14.58	84.38	9.26
20%	53.13	4.69	77.08	16.02	77.60	8.85
30%	44.44	3.27	76.04	17.86	79.51	7.74
ความแตกต่างของการแจกแจง θ						
0.0SD	42.19	3.88	78.82	8.88	82.81	6.17
0.5SD	49.13	4.62	82.99	11.44	82.64	7.93
1.0SD	54.17	5.71	70.49	28.13	76.04	11.75

ตาราง 10 (ต่อ)

ปัจจัย	วิธีการตรวจสอบ DIF					
	Poly-SIBTEST		OLR		MOLR	
	<i>P</i>	<i>E</i> ₁	<i>P</i>	<i>E</i> ₁	<i>P</i>	<i>E</i> ₁
ขนาดตัวอย่าง						
$N_R = N_F = 250$	24.31	8.23	51.62	8.97	46.53	7.01
$N_R = N_F = 500$	44.68	3.84	69.91	12.16	76.62	7.20
$N_R = N_F = 1,000$	60.65	3.93	90.28	15.85	98.84	8.34
$N_R = N_F = 2,000$	64.35	2.96	97.92	27.62	100.00	11.92
เฉลี่ยทั้งหมด	48.50	4.74	77.43	16.15	80.50	8.62

ผลการวิเคราะห์ตามตาราง 10 อัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ซิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติ และให้คะแนนหลายค่า ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่าง จำนวน 72 เงื่อนไข มีข้อสังเกต ดังนี้

1. วิธีโพลี-ซิปเทสท์มีค่าเฉลี่ยของอัตราความถูกต้องอยู่ในช่วงประมาณ 23% ถึง 74% โดยมีค่าเฉลี่ยสูงสุดในเงื่อนไขของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และมีค่าเฉลี่ยต่ำสุดในเงื่อนไขของข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าอยู่ในช่วงประมาณ 2% ถึง 7%

2. วิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยของอัตราความถูกต้องอยู่ในช่วงประมาณ 52% ถึง 92% โดยมีค่าเฉลี่ยสูงสุดในเงื่อนไขขนาดตัวอย่าง 2,000 คนต่อกลุ่ม (กลุ่มอ้างอิงและกลุ่มสนใจ) และมีค่าเฉลี่ยต่ำสุดในเงื่อนไขขนาดตัวอย่าง 250 คนต่อกลุ่ม ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าอยู่ในช่วงประมาณ 9% ถึง 28%

3. วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยของอัตราความถูกต้องอยู่ในช่วงประมาณ 47% ถึง 100% โดยมีค่าเฉลี่ยสูงสุดในเงื่อนไขขนาดตัวอย่าง 2,000 คนต่อกลุ่ม และมีค่าเฉลี่ยต่ำสุดในเงื่อนไขขนาดตัวอย่าง 250 คนต่อกลุ่ม ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าอยู่ในช่วงประมาณ 6% ถึง 12%

4. โดยภาพรวม วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องใกล้เคียงกัน ทั้งสองวิธีมีอัตราความถูกต้องสูงกว่าวิธีโพลี-ชิปเทสท์ เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าขอบเขตที่กำหนดในทุกเงื่อนไข (เกิน 5%) ส่วนวิธีโพลี-ชิปเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ภายในขอบเขตที่กำหนดเกือบทุกเงื่อนไข

บทที่ 4

ผลการวิเคราะห์ข้อมูล

การศึกษาวิจัยครั้งนี้ มีความมุ่งหมายเพื่อเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ซิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน 2 รูปแบบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 3 ขนาด ความแตกต่างของการแจกแจงความสามารถ 3 ระดับ และขนาดตัวอย่าง 4 ขนาด

ข้อมูลที่ใช้ในการศึกษาค้นครั้งนี้เป็นข้อมูลจำลอง ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย จำนวน 72 เงื่อนไข ($2 \times 3 \times 3 \times 4$) และในแต่ละเงื่อนไขจำลองซ้ำ 50 ครั้ง แล้วนำข้อมูลดังกล่าวมาตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบสามวิธี คือ วิธีโพลี-ซิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ สำหรับผลการวิเคราะห์ข้อมูลที่จะกล่าวต่อไป จำแนกออกเป็น 4 ตอน ดังนี้

1. การคำนวณค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1
 2. การวิเคราะห์ความแปรปรวนหลายตัวแปรของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1
 3. การวิเคราะห์ความแปรปรวนการทดสอบผลย่อยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 พร้อมทั้งการเปรียบเทียบภายหลัง
 4. สรุปผลการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1
- ในการนำเสนอผลการวิเคราะห์ข้อมูล เพื่อการสื่อสารเข้าใจได้ง่าย จึงกำหนดสัญลักษณ์ที่ใช้ในการวิจัย ดังนี้

DIF แทน การทำหน้าที่เบี่ยงเบนของข้อสอบ

MDIF แทน วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ

MDIF1 แทน วิธีโพลี-ซิปเทสท์ (Poly-SIBTEST)

MDIF2 แทน วิธีการถดถอยโลจิสติกแบบจัดอันดับ (OLR)

MDIF3 แทน วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ (MOLR)

FDIF แทน รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน

FDIF1 แทน ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (UDIF)

FDIF2 แทน ข้อสอบทำหน้าที่ย่อยแบบที่ไม่เป็นรูปแบบเดียวกัน (NDIF)

PDIF แทน สัดส่วนของข้อสอบทำหน้าที่ย่อยแบบ

PDIF1 แทน สัดส่วนของข้อสอบทำหน้าที่ย่อยแบบ 10%

PDIF2 แทน สัดส่วนของข้อสอบทำหน้าที่ย่อยแบบ 20%

PDIF3 แทน สัดส่วนของข้อสอบทำหน้าที่ย่อยแบบ 30%

$d(\theta)$ แทน ความแตกต่างของการแจกแจงความสามารถในมิติที่ 1 (θ_1) และมิติที่ 2 (θ_2) ระหว่างกลุ่มอ้างอิง (F) และกลุ่มสนใจ (R)

$d(\theta) 1$ แทน ความแตกต่างของการแจกแจงความสามารถ 0.0SD

$d(\theta) 2$ แทน ความแตกต่างของการแจกแจงความสามารถ 0.5SD

$d(\theta) 3$ แทน ความแตกต่างของการแจกแจงความสามารถ 1.0SD

SIZE แทน ขนาดตัวอย่าง [จำนวนผู้สอบกลุ่มอ้างอิง (N_R) ต่อจำนวนผู้สอบกลุ่มสนใจ (N_F)]

SIZE1 แทน ขนาดตัวอย่าง $N_R = N_F = 250$

SIZE2 แทน ขนาดตัวอย่าง $N_R = N_F = 500$

SIZE3 แทน ขนาดตัวอย่าง $N_R = N_F = 1,000$

SIZE4 แทน ขนาดตัวอย่าง $N_R = N_F = 2,000$

P แทน อัตราความถูกต้อง

E_1 แทน อัตราความคลาดเคลื่อนประเภทที่ 1

F แทน สถิติการทดสอบที่มีการแจกแจงแบบเอฟ

1. การคำนวณค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1

ผู้วิจัยได้นำข้อมูลที่ได้จากการตรวจสอบการทำหน้าที่ย่อยของวิธีโพลิ-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่ย่อยแบบ สัดส่วนของข้อสอบทำหน้าที่ย่อยแบบ ความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่าง ไปคำนวณอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 สำหรับผลการคำนวณแสดงในรูปค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบการทำหน้าที่ย่อยของข้อสอบสามวิธี ภายใต้เงื่อนไขของปัจจัยหลัก และปฏิสัมพันธ์ระหว่างปัจจัยดังนี้

1.1 ค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัยหลัก แสดงดังตาราง 11

ตาราง 11 ค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัยหลัก

ปัจจัย	วิธีการตรวจสอบ DIF					
	Poly-SIBTEST		OLR		MOLR	
	<i>P</i>	<i>E</i> ₁	<i>P</i>	<i>E</i> ₁	<i>P</i>	<i>E</i> ₁
รูปแบบของ DIF						
UDIF	76.65	2.58	76.96	20.09	79.00	8.81
NDIF	20.97	6.34	80.09	13.64	82.30	9.25
สัดส่วนของ DIF						
10%	50.58	5.32	81.58	14.00	80.98	9.04
20%	49.45	4.23	78.09	16.38	80.41	9.07
30%	46.40	3.84	75.90	20.20	80.58	8.98
ความแตกต่างของการแจกแจง θ						
0.0SD	41.83	3.63	79.72	10.06	82.44	7.03
0.5SD	47.14	3.80	79.78	10.31	81.75	7.72
1.0SD	57.45	5.96	76.08	30.22	77.77	12.33
ขนาดตัวอย่าง						
$N_R = N_F = 250$	27.21	7.60	49.86	8.36	48.60	7.47
$N_R = N_F = 500$	46.87	4.21	74.43	12.10	76.51	7.76
$N_R = N_F = 1,000$	58.63	2.96	92.06	18.11	97.67	9.11
$N_R = N_F = 2,000$	62.53	3.08	97.76	28.87	99.84	11.77
เฉลี่ยทั้งหมด	48.81	4.46	78.53	16.86	81.40	9.03

ผลการวิเคราะห์ตามตาราง 11 สรุปได้ดังนี้

ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยอัตราความถูกต้องสูงสุด ประมาณ 82.30% ในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน และในเงื่อนไขข้อสอบดังกล่าววิธีโพลี-ชิปเทสท์มีค่าเฉลี่ยต่ำสุด ประมาณ 20.97% เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน วิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยสูงสุด ประมาณ 20.09% และในเงื่อนไขเดียวกัน วิธีโพลี-ชิปเทสท์มีค่าเฉลี่ยต่ำสุด ประมาณ 2.58%

ภายใต้ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยอัตราความถูกต้องสูงสุด ประมาณ 81.58% ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% ส่วนวิธีโพลี-ชิปเทสท์มีค่าเฉลี่ยต่ำสุดประมาณ 46.40% ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยสูงสุด ประมาณ 20.20% ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% ส่วนวิธีโพลี-ชิปเทสท์มีค่าเฉลี่ยต่ำสุดประมาณ 3.84% ในเงื่อนไขเดียวกัน

ภายใต้ปัจจัยความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบ พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยอัตราความถูกต้องสูงสุด ประมาณ 82.44% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD และในเงื่อนไขเดียวกัน วิธีโพลี-ชิปเทสท์มีค่าเฉลี่ยต่ำสุด ประมาณ 41.83% เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยสูงสุด ประมาณ 30.22% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 1.0SD ส่วนวิธีโพลี-ชิปเทสท์มีค่าเฉลี่ยต่ำสุดประมาณ 3.63% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD

ภายใต้ปัจจัยขนาดตัวอย่าง พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยอัตราความถูกต้องสูงสุด ประมาณ 99.84% ในเงื่อนไขขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ส่วนวิธีโพลี-ชิปเทสท์มีค่าเฉลี่ยต่ำสุด ประมาณ 27.21% ในเงื่อนไขขนาดตัวอย่าง 250 คนต่อกลุ่ม เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยสูงสุด ประมาณ 28.87% ในเงื่อนไขขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ส่วนวิธีโพลี-ชิปเทสท์มีค่าเฉลี่ยต่ำสุดประมาณ 2.96% ในเงื่อนไขขนาดตัวอย่าง 1,000 คนต่อกลุ่ม

เมื่อพิจารณาผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า โดยภาพรวม พบว่า ภายใต้ทุกเงื่อนไขของปัจจัยที่แปรเปลี่ยน วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องโดยเฉลี่ยสูงสุด ประมาณ 81.40% รองลงมา ได้แก่ วิธีการถดถอยโลจิสติกแบบจัดอันดับ ประมาณ 78.53% และวิธีโพลี-ชิปเทสท์ ประมาณ 48.81% ตามลำดับ ในขณะที่เดียวกัน วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 โดยเฉลี่ยสูงสุด ประมาณ 16.86% รองลงมาได้แก่ วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ประมาณ 9.03% และวิธีโพลี-ชิปเทสท์ ประมาณ 4.46% ตามลำดับ

1.2 ค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ภายใต้ปฏิสัมพันธ์แบบสองทาง (Two-way interaction) ระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย แสดงดังตาราง 12-17

ตาราง 12 ค่าเฉลี่ยของอัตราความถูกต้อง (P) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ของวิธีการตรวจสอบ ภายใต้ปฏิสัมพันธ์แบบสองทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน

ปัจจัย		วิธีการตรวจสอบ DIF					
		Poly-SIBTEST		OLR		MOLR	
FDIF	PDIF	P	E_1	P	E_1	P	E_1
UDIF	10%	81.17	3.74	83.79	14.37	79.04	8.79
	20%	76.60	2.34	76.81	19.09	78.63	8.99
	30%	72.17	1.67	70.28	26.80	79.35	8.65
NDIF	10%	20.00	6.90	79.38	13.63	82.92	9.28
	20%	22.29	6.12	79.38	13.67	82.19	9.14
	30%	20.62	6.01	81.53	13.60	81.81	9.32

ผลการวิเคราะห์ตามตาราง 12 สรุปได้ดังนี้

ภายใต้ปัจจัยรูปแบบข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10%, 20% และ 30% พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยอัตราความถูกต้องสูงสุด ประมาณ 83.79% ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% แต่มีค่าเฉลี่ยต่ำสุด ประมาณ 70.28% ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยสูงสุด ประมาณ 26.80% ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% ส่วนวิธีโพลี-ชิปเทสที่มีค่าเฉลี่ยต่ำสุด ประมาณ 1.67% ในเงื่อนไขดังกล่าว

ภายใต้ปัจจัยรูปแบบข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10%, 20% และ 30% พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยอัตราความถูกต้องสูงสุด ประมาณ 82.92% ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% และในเงื่อนไขเดียวกัน วิธีโพลี-ชิปเทสที่มีค่าเฉลี่ยต่ำสุด ประมาณ 20% สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยสูงสุด ประมาณ 13.67% ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% ส่วนวิธีโพลี-ชิปเทสที่มีค่าเฉลี่ยต่ำสุด ประมาณ 6.01% ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30%

ตาราง 13 ค่าเฉลี่ยของอัตราความถูกต้อง (P) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ของวิธีการตรวจสอบ ภายใต้ปฏิสัมพันธ์แบบสองทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถ

ปัจจัย		วิธีการตรวจสอบ DIF					
		Poly-SIBTEST		OLR		MOLR	
FDIF	$d(\theta)$	P	E_1	P	E_1	P	E_1
UDIF	0.0SD	77.14	1.55	75.61	14.07	82.48	6.70
	0.5SD	77.30	2.12	76.75	13.94	79.72	7.74
	1.0SD	75.50	4.08	78.52	32.25	74.81	11.99
NDIF	0.0SD	6.53	5.71	83.82	6.05	82.40	7.36
	0.5SD	16.98	5.48	82.81	6.67	83.78	7.70
	1.0SD	39.41	7.84	73.65	28.19	80.73	12.68

ผลการวิเคราะห์ตามตาราง 13 สรุปได้ดังนี้

ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน ในเงื่อนไขกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถ 0.0SD, 0.5SD และ 1.0SD พบว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยอัตราความถูกต้องสูงสุด ประมาณ 82.48% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD แต่มีค่าเฉลี่ยต่ำสุด ประมาณ 74.81% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 1.0SD สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยสูงสุด ประมาณ 32.25% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 1.0SD ส่วนวิธีโพลี-ชิปเทสที่มีค่าเฉลี่ยต่ำสุด ประมาณ 1.55% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD

ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ในเงื่อนไขกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถ 0.0SD, 0.5SD และ 1.0SD พบว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยอัตราความถูกต้องสูงสุด ประมาณ 83.82% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD ส่วนวิธีโพลี-ชิปเทสที่มีค่าเฉลี่ยต่ำสุด ประมาณ 6.53% ในเงื่อนไขดังกล่าว เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยสูงสุด ประมาณ 28.19% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 1.0SD ส่วนวิธีโพลี-ชิปเทสที่มีค่าเฉลี่ยต่ำสุด ประมาณ 5.48% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.5SD

ตาราง 14 ค่าเฉลี่ยของอัตราความถูกต้อง (P) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ของวิธีการตรวจสอบ ภายใต้ปฏิสัมพันธ์แบบสองทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยขนาดตัวอย่าง

ปัจจัย		วิธีการตรวจสอบ DIF					
		Poly-SIBTEST		OLR		MOLR	
FDIF	SIZE	P	E_1	P	E_1	P	E_1
UDIF	250	37.52	6.29	43.05	9.17	45.20	7.33
	500	73.14	2.26	70.84	13.07	74.13	7.50
	1,000	96.15	1.14	94.35	21.83	97.00	9.12
	2,000	99.78	0.66	99.60	36.29	99.69	11.29

ตาราง 14 (ต่อ)

ปัจจัย		วิธีการตรวจสอบ DIF					
		Poly-SIBTEST		OLR		MOLR	
FDIF	SIZE	<i>P</i>	<i>E</i> ₁	<i>P</i>	<i>E</i> ₁	<i>P</i>	<i>E</i> ₁
NDIF	250	16.90	8.92	56.67	7.56	51.99	7.60
	500	20.60	6.17	78.01	11.14	78.89	8.02
	1,000	21.11	4.78	89.77	14.40	98.33	9.10
	2,000	25.28	5.50	95.93	21.44	100.00	12.26

ผลการวิเคราะห์ตามตาราง 14 สรุปได้ดังนี้

ภายใต้ปัจจัยรูปแบบข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน ในเงื่อนไขขนาดตัวอย่าง 250, 500, 1,000 และ 2,000 คนต่อกลุ่ม (กลุ่มอ้างอิง และกลุ่มสนใจ) พบว่า วิธีโพลี-ชิปเทสท์มีค่าเฉลี่ยอัตราความถูกต้องสูงสุด ประมาณ 99.78% ในเงื่อนไขขนาดตัวอย่าง 2,000 คนต่อกลุ่ม แต่มีค่าเฉลี่ยต่ำสุด ประมาณ 37.52% ในเงื่อนไขขนาดตัวอย่าง 250 คนต่อกลุ่ม เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยสูงสุด ประมาณ 36.29% ในเงื่อนไขขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ส่วนวิธีโพลี-ชิปเทสท์มีค่าเฉลี่ยต่ำสุด ประมาณ 0.66% ในเงื่อนไขดังกล่าว

ภายใต้ปัจจัยรูปแบบข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ในเงื่อนไขขนาดตัวอย่าง 250, 500, 1,000 และ 2,000 คนต่อกลุ่ม พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยอัตราความถูกต้องสูงสุด ประมาณ 100% ในเงื่อนไขขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ส่วนวิธีโพลี-ชิปเทสท์ มีค่าเฉลี่ยต่ำสุด ประมาณ 16.90% ในเงื่อนไขขนาดตัวอย่าง 250 คนต่อกลุ่ม เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยสูงสุด ประมาณ 21.44% ในเงื่อนไขขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ส่วนวิธีโพลี-ชิปเทสท์มีค่าเฉลี่ยต่ำสุด ประมาณ 4.78% ในเงื่อนไขขนาดตัวอย่าง 1,000 คนต่อกลุ่ม

ตาราง 15 ค่าเฉลี่ยของอัตราความถูกต้อง (P) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ของวิธีการตรวจสอบ ภายใต้ปฏิสัมพันธ์แบบสองทางระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เป็นเบเน กับปัจจัยความแตกต่างของการแจกแจงความสามารถ

ปัจจัย		วิธีการตรวจสอบ DIF					
		Poly-SIBTEST		OLR		MOLR	
PDIF	$d(\theta)$	P	E_1	P	E_1	P	E_1
10%	0.0SD	43.63	4.44	81.88	5.91	81.50	7.16
	0.5SD	47.63	4.39	83.56	7.24	82.25	7.72
	1.0SD	60.50	7.12	79.31	28.86	79.19	12.23
20%	0.0SD	42.75	3.51	79.56	9.79	82.63	7.66
	0.5SD	47.81	3.48	79.97	9.57	81.53	7.51
	1.0SD	57.78	5.71	74.75	29.79	77.06	12.02
30%	0.0SD	39.13	2.95	77.71	14.48	83.19	6.27
	0.5SD	45.98	3.54	75.81	14.11	81.48	7.93
	1.0SD	54.08	5.04	74.19	32.01	77.06	12.75

ผลการวิเคราะห์ตามตาราง 15 สรุปได้ดังนี้

ภายใต้ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เป็นเบเน 10% ในเงื่อนไขกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถ 0.0SD, 0.5SD และ 1.0SD พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยอัตราความถูกต้องสูงสุด ประมาณ 83.56% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.5SD ส่วนวิธีโพลี-ชิปเทสที่มีค่าเฉลี่ยต่ำสุด ประมาณ 43.63% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยสูงสุด ประมาณ 28.86% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 1.0SD ส่วนวิธีโพลี-ชิปเทสที่มีค่าเฉลี่ยต่ำสุด ประมาณ 4.39% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.5SD

ภายใต้ปัจจัยสัดส่วนของข้อสอบทำหน้าที่ยี่สิบเปอร์เซ็นต์ ในเงื่อนไขกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถ 0.0SD, 0.5SD และ 1.0SD พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยของอัตราความถูกต้องสูงสุด ประมาณ 82.63% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD และในเงื่อนไขเดียวกัน วิธีโพลี-ชิปเทสท์มีค่าเฉลี่ยต่ำสุด ประมาณ 42.75% เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยสูงสุด ประมาณ 29.79% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 1.0SD ส่วนวิธีโพลี-ชิปเทสท์มีค่าเฉลี่ยต่ำสุด ประมาณ 3.48% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.5SD

ภายใต้ปัจจัยสัดส่วนของข้อสอบทำหน้าที่ยี่สิบเปอร์เซ็นต์ ในเงื่อนไขกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถ 0.0SD, 0.5SD และ 1.0SD พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยของอัตราความถูกต้องสูงสุด ประมาณ 83.19% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD และในเงื่อนไขเดียวกันวิธีโพลี-ชิปเทสท์มีค่าเฉลี่ยต่ำสุด ประมาณ 39.13% เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยสูงสุด ประมาณ 32.01% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 1.0SD ส่วนวิธีโพลี-ชิปเทสท์มีค่าเฉลี่ยต่ำสุด ประมาณ 2.95% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD

ตาราง 16 ค่าเฉลี่ยของอัตราความถูกต้อง (P) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ของวิธีการตรวจสอบ ภายใต้ปฏิสัมพันธ์แบบสองทางระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่ยี่สิบเปอร์เซ็นต์ กับปัจจัยขนาดตัวอย่าง

ปัจจัย		วิธีการตรวจสอบ DIF					
		Poly-SIBTEST		OLR		MOLR	
PDIF	SIZE	P	E_1	P	E_1	P	E_1
10%	250	31.00	8.80	56.83	7.18	52.08	6.98
	500	51.33	4.84	78.67	10.29	74.25	7.75
	1,000	59.58	4.23	93.33	15.35	97.58	9.52
	2,000	60.42	3.40	97.50	23.20	100.00	11.90

ตาราง 16 (ต่อ)

ปัจจัย		วิธีการตรวจสอบ DIF					
		Poly-SIBTEST		OLR		MOLR	
PDIF	SIZE	P	E_1	P	E_1	P	E_1
20%	250	25.29	7.47	47.71	8.74	46.63	8.04
	500	48.08	3.95	75.83	11.99	76.75	7.95
	1,000	59.83	2.46	92.38	16.90	98.50	8.56
	2,000	64.58	3.05	96.46	27.91	99.75	11.71
30%	250	25.33	6.55	45.03	9.18	47.08	7.38
	500	41.19	3.85	68.78	14.04	78.53	7.58
	1,000	56.47	2.19	90.47	22.10	96.92	9.25
	2,000	62.58	2.79	99.33	35.49	99.78	11.71

ผลการวิเคราะห์ตามตาราง 16 สรุปได้ดังนี้

ภายใต้ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% ในเงื่อนไขขนาดตัวอย่าง 250, 500, 1,000 และ 2,000 คนต่อกลุ่ม (กลุ่มอ้างอิง และกลุ่มสนใจ) พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยอัตราความถูกต้องสูงสุด ประมาณ 100% ในเงื่อนไขขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ส่วนวิธีโพลี-ชิปเทสท์มีค่าเฉลี่ยต่ำสุด ประมาณ 31% ในเงื่อนไขขนาดตัวอย่าง 250 คนต่อกลุ่ม เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยสูงสุด ประมาณ 23.20% ในเงื่อนไขขนาดตัวอย่าง 2,000 คนต่อกลุ่ม และในเงื่อนไขเดียวกัน วิธีโพลี-ชิปเทสท์มีค่าเฉลี่ยต่ำสุด ประมาณ 3.40%

ภายใต้ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% ในเงื่อนไขขนาดตัวอย่าง 250, 500, 1,000 และ 2,000 คนต่อกลุ่ม พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยอัตราความถูกต้องสูงสุด ประมาณ 99.75% ในเงื่อนไขขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ส่วนวิธีโพลี-ชิปเทสท์มีค่าเฉลี่ยต่ำสุด ประมาณ 25.29% ในเงื่อนไขขนาดตัวอย่าง 250 คนต่อกลุ่ม เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยสูงสุด ประมาณ 27.91% ในเงื่อนไขขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ส่วนวิธีโพลี-ชิปเทสท์มีค่าเฉลี่ยต่ำสุด ประมาณ 2.46% ในเงื่อนไขขนาดตัวอย่าง 1,000 คนต่อกลุ่ม

ภายใต้ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เป็นเบส 30% ในเงื่อนไขขนาดตัวอย่าง 250, 500, 1,000 และ 2,000 คนต่อกลุ่ม พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยอัตราความถูกต้องสูงสุด ประมาณ 99.78% ในเงื่อนไขขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ส่วนวิธีโพลี-ชิปเทสที่มีค่าเฉลี่ยต่ำสุด ประมาณ 25.33% ในเงื่อนไขขนาดตัวอย่าง 250 คนต่อกลุ่ม เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยสูงสุด ประมาณ 35.49% ในเงื่อนไขขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ส่วนวิธีโพลี-ชิปเทสที่มีค่าเฉลี่ยต่ำสุด ประมาณ 2.19% ในเงื่อนไขขนาดตัวอย่าง 1,000 คนต่อกลุ่ม

ตาราง 17 ค่าเฉลี่ยของอัตราความถูกต้อง (P) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ของวิธีการตรวจสอบ ภายใต้ปฏิสัมพันธ์แบบสองทางระหว่างปัจจัยความแตกต่างของการแจกแจงความสามารถ กับปัจจัยขนาดตัวอย่าง

ปัจจัย		วิธีการตรวจสอบ DIF					
		Poly-SIBTEST		OLR		MOLR	
$d(\theta)$	SIZE	P	E_1	P	E_1	P	E_1
0.0SD	250	21.07	4.90	48.53	6.30	50.99	6.92
	500	41.32	3.59	74.89	9.06	79.46	6.37
	1,000	52.47	2.81	96.21	9.39	99.31	6.98
	2,000	52.47	3.21	99.24	15.49	100.00	7.84
0.5SD	250	25.64	6.36	50.22	5.98	50.33	7.00
	500	47.43	3.61	76.64	7.56	78.75	6.94
	1,000	55.17	2.64	94.15	10.63	98.18	8.00
	2,000	60.32	2.60	98.11	17.05	99.75	8.94
1.0SD	250	34.92	11.55	50.82	12.81	44.47	8.48
	500	51.86	5.44	71.75	19.69	71.32	9.96
	1,000	68.25	3.42	85.82	34.32	95.51	12.35
	2,000	74.79	3.42	95.94	54.06	99.78	18.54

ผลการวิเคราะห์ตามตาราง 17 สรุปได้ดังนี้

ภายใต้ปัจจัยกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถ 0.0SD ในเงื่อนไขขนาดตัวอย่าง 250, 500, 1,000 และ 2,000 คนต่อกลุ่ม (กลุ่มอ้างอิง และกลุ่มสนใจ) พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยอัตราความถูกต้องสูงสุด ประมาณ 100% ในเงื่อนไขขนาดตัวอย่าง 2,000 คน ส่วนวิธีโพลี-ชิปเทสที่มีค่าเฉลี่ยต่ำสุด ประมาณ 21.07% ในเงื่อนไขขนาดตัวอย่าง 250 คนต่อกลุ่ม เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับ มีค่าเฉลี่ยสูงสุด ประมาณ 15.49% ในเงื่อนไขขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ส่วนวิธีโพลี-ชิปเทสที่มีค่าเฉลี่ยต่ำสุด ประมาณ 2.81% ในเงื่อนไขขนาดตัวอย่าง 1,000 คนต่อกลุ่ม

ภายใต้ปัจจัยกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถ 0.5SD ในเงื่อนไขขนาดตัวอย่าง 250, 500, 1,000 และ 2,000 คนต่อกลุ่ม พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยอัตราความถูกต้องสูงสุด ประมาณ 99.75% ในเงื่อนไขขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ส่วนวิธีโพลี-ชิปเทสที่มีค่าเฉลี่ยต่ำสุด ประมาณ 25.64% ในเงื่อนไขขนาดตัวอย่าง 250 คนต่อกลุ่ม เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยสูงสุด ประมาณ 17.05% ในเงื่อนไขขนาดตัวอย่าง 2,000 คนต่อกลุ่ม และในเงื่อนไขเดียวกัน วิธีโพลี-ชิปเทสที่มีค่าเฉลี่ยต่ำสุดประมาณ 2.60%

ภายใต้ปัจจัยกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถ 1.0SD ในเงื่อนไขขนาดตัวอย่าง 250, 500, 1,000 และ 2,000 คนต่อกลุ่ม พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยอัตราความถูกต้องสูงสุด ประมาณ 99.78% ในเงื่อนไขขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ส่วนวิธีโพลี-ชิปเทสที่มีค่าเฉลี่ยต่ำสุด ประมาณ 34.92% ในเงื่อนไขขนาดตัวอย่าง 250 คนต่อกลุ่ม เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยสูงสุด ประมาณ 54.06% ในเงื่อนไขขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ส่วนวิธีโพลี-ชิปเทสที่มีค่าเฉลี่ยต่ำสุด ประมาณ 3.42% ในเงื่อนไขขนาดตัวอย่าง 1,000 และ 2,000 คนต่อกลุ่ม

1.3 ค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทส วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ภายใต้ปฏิสัมพันธ์แบบสามทาง (Three-way interaction) ระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย แสดงดังตาราง 18-21

ตาราง 18 ค่าเฉลี่ยของอัตราความถูกต้อง (P) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ของวิธีการตรวจสอบ ภายใต้ปฏิสัมพันธ์แบบสามทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถ

ปัจจัย			วิธีการตรวจสอบ DIF					
			Poly-SIBTEST		OLR		MOLR	
FDIF	PDIF	$d(\theta)$	P	E_1	P	E_1	P	E_1
UDIF	10%	0.0SD	81.63	2.28	83.13	6.54	83.00	6.75
		0.5SD	80.87	3.36	82.75	8.10	79.50	8.01
		1.0SD	81.00	5.57	85.50	28.49	74.63	11.61
	20%	0.0SD	76.13	1.31	75.38	12.78	81.19	6.97
		0.5SD	77.19	1.72	77.13	12.66	79.63	7.75
		1.0SD	76.50	4.00	77.94	31.84	75.06	12.25
	30%	0.0SD	73.67	1.07	68.33	22.89	83.25	6.37
		0.5SD	73.83	1.29	70.38	21.07	80.04	7.46
		1.0SD	69.00	2.66	72.13	36.43	74.75	12.11
NDIF	10%	0.0SD	5.62	6.60	80.63	5.28	80.00	7.57
		0.5SD	14.37	5.42	84.38	6.39	85.00	7.43
		1.0SD	40.00	8.68	73.13	29.24	83.75	12.85
	20%	0.0SD	9.37	5.70	83.75	6.80	84.06	8.36
		0.5SD	18.44	5.23	82.81	6.48	83.44	7.27
		1.0SD	39.06	7.42	71.56	27.73	79.06	11.80
	30%	0.0SD	4.58	4.82	87.08	6.07	83.13	6.16
		0.5SD	18.13	5.80	81.25	7.14	82.92	8.39
		1.0SD	39.17	7.41	76.25	27.59	79.37	13.39

ผลการวิเคราะห์ตามตาราง 18 สรุปได้ดังนี้

ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และความแตกต่างของการแจกแจงความสามารถ พบว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยอัตราความถูกต้องสูงสุด ประมาณ 85.50% ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% และความแตกต่างของการแจกแจงความสามารถ 1.0SD ส่วนวิธีโพลี-ชิปเทสที่มีค่าเฉลี่ยอัตราความถูกต้องต่ำสุด ประมาณ 69% ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% และความแตกต่างของการแจกแจงความสามารถ 1.0SD เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ประมาณ 36.43% ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% และความแตกต่างของการแจกแจงความสามารถ 1.0SD ส่วนวิธีโพลี-ชิปเทสที่มีค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ประมาณ 1.07% ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% และความแตกต่างของการแจกแจงความสามารถ 0.0SD

ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และความแตกต่างของการแจกแจงความสามารถ พบว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยอัตราความถูกต้องสูงสุด ประมาณ 87.08% ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% และความแตกต่างของการแจกแจงความสามารถ 0.0SD ส่วนในเงื่อนไขเดียวกัน วิธีโพลี-ชิปเทสที่มีค่าเฉลี่ยอัตราความถูกต้องต่ำสุด ประมาณ 4.58% เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุดประมาณ 29.24% ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% และความแตกต่างของการแจกแจงความสามารถ 1.0SD ส่วนวิธีโพลี-ชิปเทสที่มีค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ประมาณ 4.82% ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% และความแตกต่างของการแจกแจงความสามารถ 0.0SD

ตาราง 19 ค่าเฉลี่ยของอัตราความถูกต้อง (P) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ของวิธีการตรวจสอบ ภายใต้ปฏิสัมพันธ์แบบสามทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยขนาดตัวอย่าง

ปัจจัย			วิธีการตรวจสอบ DIF					
			Poly-SIBTEST		OLR		MOLR	
FDIF	PDIF	SIZE	P	E_1	P	E_1	P	E_1
UDIF	10%	250	42.83	7.69	54.50	7.13	45.83	7.02
		500	82.67	3.67	82.33	10.39	74.33	7.26
		1,000	99.17	2.44	98.33	15.61	96.00	9.22
		2,000	100.00	1.15	100.00	24.37	100.00	11.67
	20%	250	36.00	6.29	41.25	9.04	44.50	7.96
		500	73.25	1.96	71.25	12.52	72.67	7.46
		1,000	97.17	0.75	95.17	20.46	97.83	8.79
		2,000	100.00	0.37	99.58	34.35	99.50	11.75
	30%	250	33.72	4.88	33.39	11.33	45.28	7.02
		500	63.50	1.14	58.94	16.29	75.39	7.79
		1,000	92.11	0.21	89.56	29.43	97.17	9.33
		2,000	99.33	0.45	99.22	50.14	99.56	10.45
NDIF	10%	250	19.17	9.91	59.17	7.22	58.33	6.94
		500	20.00	6.02	75.00	10.19	74.17	8.24
		1,000	20.00	6.02	88.33	15.09	99.17	9.81
		2,000	20.83	5.65	95.00	22.04	100.00	12.13
	20%	250	14.58	8.65	54.17	8.44	48.75	8.12
		500	22.92	5.94	80.42	11.46	80.83	8.44
		1,000	22.50	4.17	89.58	13.33	99.17	8.33
		2,000	29.17	5.73	93.33	21.46	100.00	11.67

ตาราง 19 (ต่อ)

ปัจจัย			วิธีการตรวจสอบ DIF					
			Poly-SIBTEST		OLR		MOLR	
FDIF	PDIF	SIZE	<i>P</i>	<i>E</i> ₁	<i>P</i>	<i>E</i> ₁	<i>P</i>	<i>E</i> ₁
	30%	250	16.94	8.21	56.67	7.02	48.89	7.74
		500	18.89	6.55	78.61	11.79	81.67	7.38
		1,000	20.83	4.17	91.39	14.76	96.67	9.17
		2,000	25.83	5.12	99.44	20.83	100.00	12.98

ผลการวิเคราะห์ตามตาราง 19 สรุปได้ดังนี้

ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และขนาดตัวอย่าง พบว่า ทั้งสามวิธีมีค่าเฉลี่ยอัตราความถูกต้องสูงสุด ประมาณ 100% ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% และขนาดตัวอย่าง 2,000 คนต่อกลุ่ม (กลุ่มอ้างอิง และกลุ่มสนใจ) นอกจากนี้ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% และขนาดตัวอย่าง 2,000 คนต่อกลุ่ม วิธีโพลี-ชิปเทสท์มีค่าเฉลี่ยอัตราความถูกต้องสูงสุดด้วย (ประมาณ 100%) ส่วนในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% และขนาดตัวอย่าง 250 คนต่อกลุ่ม วิธีโพลี-ชิปเทสท์ มีค่าเฉลี่ยอัตราความถูกต้องต่ำสุด ประมาณ 33.72% เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ประมาณ 50.14% ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% และขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ส่วนวิธีโพลี-ชิปเทสท์มีค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ประมาณ 0.21% ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% และขนาดตัวอย่าง 1,000 คนต่อกลุ่ม

ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และขนาดตัวอย่าง พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยอัตราความถูกต้องสูงสุด ประมาณ 100% ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10%, 20% และ 30% ที่มีขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ส่วนวิธีโพลี-ชิปเทสท์มีค่าเฉลี่ยอัตราความถูกต้องต่ำสุด ประมาณ 14.58% ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% และขนาดตัวอย่าง 250 คนต่อกลุ่ม เมื่อพิจารณาอัตราความคลาดเคลื่อน

ประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ประมาณ 22.04% ในเงื่อนไขสัดส่วนของข้อสอบทำนายที่เบี่ยงเบน 10% และขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ส่วนวิธีโพลี-ชิปเทสที่มีค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ประมาณ 4.17% ในเงื่อนไขสัดส่วนของข้อสอบทำนายที่เบี่ยงเบน 20% และ 30% เมื่อใช้ขนาดตัวอย่าง 1,000 คนต่อกลุ่ม

ตาราง 20 ค่าเฉลี่ยของอัตราความถูกต้อง (P) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ของวิธีการตรวจสอบ ภายใต้ปฏิสัมพันธ์แบบสามทางระหว่างปัจจัยรูปแบบของข้อสอบทำนายที่เบี่ยงเบน ปัจจัยความแตกต่างของการแจกแจงความสามารถ กับปัจจัยขนาดตัวอย่าง

ปัจจัย			วิธีการตรวจสอบ DIF					
			Poly-SIBTEST		OLR		MOLR	
FDIF	$d(\theta)$	SIZE	P	E_1	P	E_1	P	E_1
UDIF	0.0SD	250	32.56	2.92	41.22	7.07	49.33	6.61
		500	77.78	1.52	67.56	9.58	81.69	5.80
		1,000	98.28	0.97	93.94	14.68	98.89	7.10
		2,000	99.94	0.81	99.72	24.96	100.00	7.28
	0.5SD	250	36.83	5.15	42.39	7.80	46.08	6.84
		500	75.28	1.69	71.75	8.98	75.56	6.97
		1,000	97.14	1.18	93.72	14.67	97.75	8.17
		2,000	99.94	0.47	99.14	24.31	99.50	9.00
	1.0SD	250	43.17	10.79	45.53	12.64	40.19	8.55
		500	66.36	3.56	73.22	20.63	65.14	9.73
		1,000	93.03	1.26	95.39	36.14	94.36	12.08
		2,000	99.44	0.70	99.94	59.60	99.56	17.59
NDIF	0.0SD	250	9.58	6.89	55.83	5.53	52.64	7.24
		500	4.86	5.66	82.22	8.53	77.22	6.94
		1,000	6.67	4.66	98.47	4.10	99.72	6.87
		2,000	5.00	5.62	98.75	6.03	100.00	8.40

ตาราง 20 (ต่อ)

ปัจจัย			วิธีการตรวจสอบ DIF					
			Poly-SIBTEST		OLR		MOLR	
FDIF	$d(\theta)$	SIZE	P	E_1	P	E_1	P	E_1
	0.5SD	250	14.44	7.57	58.06	4.17	54.58	7.15
		500	19.58	5.53	81.53	6.14	81.94	6.92
		1,000	13.19	4.11	94.58	6.60	98.61	7.83
		2,000	20.69	4.74	97.08	9.78	100.00	8.88
	1.0SD	250	26.67	12.31	56.11	12.98	48.75	8.41
		500	37.36	7.31	70.28	18.75	77.50	10.20
		1,000	43.47	5.59	76.25	32.50	96.67	12.62
		2,000	50.14	6.14	91.94	48.51	100.00	19.49

ผลการวิเคราะห์ตามตาราง 20 สรุปได้ดังนี้

ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เป็นรูปแบบเดียวกัน ในเงื่อนไขกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่าง พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยอัตราความถูกต้องสูงสุด ประมาณ 100% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD และขนาดตัวอย่าง 2,000 คนต่อกลุ่ม (กลุ่มอ้างอิง และกลุ่มสนใจ) ส่วนวิธีโพลี-ชิปเทสท์มีค่าเฉลี่ยอัตราความถูกต้องต่ำสุด ประมาณ 32.56% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD และขนาดตัวอย่าง 250 คนต่อกลุ่ม เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ประมาณ 59.60% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 1.0SD และขนาดตัวอย่าง 2,000 คนต่อกลุ่มส่วนวิธีโพลี-ชิปเทสท์มีค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ประมาณ 0.47% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.5SD และขนาดตัวอย่าง 2,000 คนต่อกลุ่ม

ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เป็นรูปแบบไม่เหมือนกัน ในเงื่อนไขกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่าง พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยอัตราความถูกต้องสูงสุด ประมาณ 100% ในเงื่อนไข

ความแตกต่างของการแจกแจงความสามารถ 0.0SD, 0.5SD และ 1.0SD เมื่อใช้ขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ส่วนวิธีโพลี-ชิปเทสต์มีค่าเฉลี่ยอัตราความถูกต้องต่ำสุด ประมาณ 4.86% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD และขนาดตัวอย่าง 500 คนต่อกลุ่ม เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ประมาณ 48.51% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 1.0SD และขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ส่วนวิธีโพลี-ชิปเทสต์มีค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ประมาณ 4.11% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.5SD และขนาดตัวอย่าง 1,000 คนต่อกลุ่ม

ตาราง 21 ค่าเฉลี่ยของอัตราความถูกต้อง (P) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ของวิธีการตรวจสอบ ภายใต้ปฏิสัมพันธ์แบบสามทางระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ปัจจัยความแตกต่างของการแจกแจงความสามารถ กับปัจจัยขนาดตัวอย่าง

ปัจจัย			วิธีการตรวจสอบ DIF					
			Poly-SIBTEST		OLR		MOLR	
PDIF	$d(\theta)$	SIZE	P	E_1	P	E_1	P	E_1
10%	0.0SD	250	23.25	6.25	53.00	4.92	54.00	6.17
		500	47.50	4.14	76.50	5.31	73.00	6.25
		1,000	52.50	4.17	98.00	6.22	99.00	7.58
		2,000	51.25	3.19	100.00	7.19	100.00	8.64
	0.5SD	250	25.00	6.53	55.50	4.06	53.25	7.06
		500	51.25	3.89	83.75	6.03	77.50	7.47
		1,000	55.50	3.89	96.25	7.31	98.25	7.25
		2,000	58.75	3.25	98.75	11.58	100.00	9.11
	1.0SD	250	44.75	13.61	62.00	12.56	49.00	7.72
		500	55.25	6.50	75.75	19.53	72.25	9.53
		1,000	70.75	4.64	85.75	32.53	95.50	13.72
		2,000	71.25	3.75	93.75	50.83	100.00	17.94

ตาราง 21 (ต่อ)

ปัจจัย			วิธีการตรวจสอบ DIF					
			Poly-SIBTEST		OLR		MOLR	
PDIF	$d(\theta)$	SIZE	P	E_1	P	E_1	P	E_1
20%	0.0SD	250	20.38	4.25	47.00	6.62	50.38	8.00
		500	41.13	4.06	76.00	9.44	80.38	6.94
		1,000	54.50	2.59	97.38	8.37	99.75	6.91
		2,000	55.00	3.12	97.88	14.72	100.00	8.81
	0.5SD	250	25.00	6.19	49.50	6.22	48.00	7.69
		500	48.12	2.94	79.75	7.41	79.75	6.72
		1,000	56.25	2.19	94.13	9.31	99.13	8.06
		2,000	61.87	2.59	96.50	15.34	99.25	7.56
	1.0SD	250	30.50	11.97	46.63	13.38	41.50	8.44
		500	55.00	4.84	71.75	19.12	70.13	10.19
		1,000	68.75	2.59	85.63	33.00	96.62	10.72
		2,000	76.87	3.44	95.00	53.66	100.00	18.75
30%	0.0SD	250	19.58	4.21	45.58	7.36	48.58	6.61
		500	35.33	2.57	72.17	12.43	85.00	5.93
		1,000	50.42	1.68	93.25	13.57	99.17	6.46
		2,000	51.17	3.32	99.83	24.57	100.00	6.07
	0.5SD	250	26.92	6.36	45.67	7.68	49.75	6.25
		500	42.92	4.00	66.42	9.25	79.00	6.64
		1,000	53.75	1.86	92.08	15.29	97.17	8.68
		2,000	60.33	1.96	99.08	24.21	100.00	10.14
	1.0SD	250	29.50	9.07	43.83	12.50	42.92	9.29
		500	45.33	4.96	67.75	20.43	71.58	10.18
		1,000	65.25	3.04	86.08	37.43	94.42	12.61
		2,000	76.25	3.07	99.08	57.68	99.33	18.93

ผลการวิเคราะห์ตามตาราง 21 สรุปได้ดังนี้

ภายใต้ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% ในเงื่อนไขกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่าง พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยอัตราความถูกต้องสูงสุด ประมาณ 100% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD, 0.5SD และ 1.0SD เมื่อใช้ขนาดตัวอย่าง 2,000 คนต่อกลุ่ม (กลุ่มอ้างอิงและกลุ่มสนใจ) นอกจากนี้ ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD และขนาดตัวอย่าง 2,000 คนต่อกลุ่ม วิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยอัตราความถูกต้องสูงสุดด้วย (ประมาณ 100%) ส่วนวิธีโพลี-ชิปเทสที่มีค่าเฉลี่ยอัตราความถูกต้องต่ำสุด ประมาณ 23.25% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD และขนาดตัวอย่าง 2,000 คนต่อกลุ่ม เมื่อพิจารณาค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ประมาณ 50.83% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 1.0SD และขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ส่วนวิธีโพลี-ชิปเทสที่มีค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ประมาณ 3.19% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD และขนาดตัวอย่าง 2,000 คนต่อกลุ่ม

ภายใต้ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% ในเงื่อนไขกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่าง พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยอัตราความถูกต้องสูงสุด ประมาณ 100% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD และ 1.0SD เมื่อใช้ขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ส่วนวิธีโพลี-ชิปเทสที่มีค่าเฉลี่ยอัตราความถูกต้องต่ำสุด ประมาณ 20.38% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD และขนาดตัวอย่าง 250 คนต่อกลุ่ม เมื่อพิจารณาค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ประมาณ 53.66% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 1.0SD และขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ส่วนวิธีโพลี-ชิปเทสที่มีค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ประมาณ 2.19% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.5SD และขนาดตัวอย่าง 1,000 คนต่อกลุ่ม

ภายใต้ปัจจัยสัดส่วนของข้อสอบทำหน้าที่ยี่สิบสาม 30% ในเงื่อนไขกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่าง พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยอัตราความถูกต้องสูงสุด ประมาณ 100% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD และ 0.5SD เมื่อใช้ขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ส่วนวิธีโพลี-ชิปเทสที่มีค่าเฉลี่ยอัตราความถูกต้องต่ำสุด ประมาณ 19.58% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD และขนาดตัวอย่าง 250 คนต่อกลุ่ม เมื่อพิจารณาค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ประมาณ 57.68% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 1.0SD และขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ส่วนวิธีโพลี-ชิปเทสที่มีค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ประมาณ 1.68% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD และขนาดตัวอย่าง 1,000 คนต่อกลุ่ม

1.4 ค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทส วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ภายใต้ปฏิสัมพันธ์แบบสี่ทาง (Four-way interaction) ระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย แสดงดังตาราง 22

ตาราง 22 ค่าเฉลี่ยของอัตราความถูกต้อง (P) และอัตราความคลาดเคลื่อนประเภทที่ 1 (E_1) ของวิธีการตรวจสอบ ภายใต้ปฏิสัมพันธ์แบบสี่ทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่ยี่สิบสาม ปัจจัยสัดส่วนของข้อสอบทำหน้าที่ยี่สิบสาม ปัจจัยความแตกต่างของการแจกแจงความสามารถ กับปัจจัยขนาดตัวอย่าง

ปัจจัย				วิธีการตรวจสอบ DIF					
				Poly-SIBTEST		OLR		MOLR	
FDIF	PDIF	$d(\theta)$	SIZE	P	E_1	P	E_1	P	E_1
UDIF	10%	0.0SD	250	36.50	3.89	53.50	5.11	50.50	6.50
			500	90.00	2.44	80.50	5.06	83.50	5.56
			1,000	100.00	1.94	98.50	7.17	98.00	7.11
			2,000	100.00	0.83	100.00	8.83	100.00	7.83

ตาราง 22 (ต่อ)

ปัจจัย				วิธีการตรวจสอบ DIF					
				Poly-SIBTEST		OLR		MOLR	
				P	E_1	P	E_1	P	E_1
FDIF	PDIF	0.5SD	250	40.00	6.11	51.00	5.61	46.50	6.89
			500	85.00	3.33	82.50	7.33	75.00	7.44
			1,000	98.50	2.78	97.50	7.94	96.50	8.11
			2,000	100.00	1.22	100.00	11.50	100.00	9.61
		1.0SD	250	52.00	13.06	59.00	10.67	40.50	7.67
			500	73.00	5.22	84.00	18.78	64.50	8.78
			1,000	99.00	2.61	99.00	31.72	93.50	12.44
			2,000	100.00	1.39	100.00	52.78	100.00	17.56
	UDIF	20%	250	29.50	2.87	39.00	6.37	47.00	7.25
			500	76.00	1.25	67.00	9.19	78.25	6.06
			1,000	99.00	0.81	96.00	13.31	99.50	6.62
			2,000	100.00	0.31	99.50	22.25	100.00	7.94
		0.5SD	250	35.00	5.19	41.50	7.44	47.25	7.56
			500	75.00	0.87	73.25	8.25	74.50	6.25
			1,000	98.75	0.62	94.50	13.00	98.25	8.31
			2,000	100.00	0.19	99.25	21.94	98.50	8.87
		1.0SD	250	43.50	10.81	43.25	13.31	39.25	9.06
			500	68.75	3.75	73.50	20.12	65.25	10.06
			1,000	93.75	0.81	95.00	35.06	95.75	11.44
			2,000	100.00	0.62	100.00	58.88	100.00	18.44
		30%	250	31.67	2.00	31.17	9.71	50.50	6.07
			500	67.33	0.86	55.17	14.50	83.33	5.79
			1,000	95.83	0.14	87.33	23.57	99.17	7.57
			2,000	99.83	1.29	99.67	43.79	100.00	6.07

ตาราง 22 (ต่อ)

ปัจจัย				วิธีการตรวจสอบ DIF					
				Poly-SIBTEST		OLR		MOLR	
				<i>P</i>	<i>E</i> ₁	<i>P</i>	<i>E</i> ₁	<i>P</i>	<i>E</i> ₁
FDIF	PDIF	0.5SD	250	35.50	4.14	34.67	10.36	44.50	6.07
			500	65.83	0.86	59.50	11.36	77.17	7.21
			1,000	94.17	0.14	89.17	23.07	98.50	8.07
			2,000	99.83	0.00	98.17	39.50	100.00	8.50
		1.0SD	250	34.00	8.50	34.33	13.93	40.83	8.93
			500	57.33	1.71	62.17	23.00	65.67	10.36
			1,000	86.33	0.36	92.17	41.64	93.83	12.36
			2,000	98.33	0.07	99.83	67.14	98.67	16.79
	10%	0.0SD	250	10.00	8.61	52.50	4.72	57.50	5.83
			500	5.00	5.83	72.50	5.56	62.50	6.94
			1,000	5.00	6.39	97.50	5.28	100.00	8.06
			2,000	2.50	5.56	100.00	5.56	100.00	9.44
		0.5SD	250	10.00	6.94	60.00	2.50	60.00	7.22
			500	17.50	4.44	85.00	4.72	80.00	7.50
			1,000	12.50	5.00	95.00	6.67	100.00	6.39
			2,000	17.50	5.28	97.50	11.67	100.00	8.61
		1.0SD	250	37.50	14.17	65.00	14.44	57.50	7.78
			500	37.50	7.78	67.50	20.28	80.00	10.28
			1,000	42.50	6.67	72.50	33.33	97.50	15.00
			2,000	42.50	6.11	87.50	48.89	100.00	18.33
	20%	0.0SD	250	11.25	5.62	55.00	6.87	53.75	8.75
			500	6.25	6.87	85.00	9.69	82.50	7.81
			1,000	10.00	4.37	98.75	3.44	100.00	7.19
			2,000	10.00	5.94	96.25	7.19	100.00	9.69

ตาราง 22 (ต่อ)

ปัจจัย				วิธีการตรวจสอบ DIF					
				Poly-SIBTEST		OLR		MOLR	
FDIF	PDIF	$d(\theta)$	SIZE	P	E_1	P	E_1	P	E_1
	0.5SD	250	250	15.00	7.19	57.50	5.00	48.75	7.81
			500	21.25	5.00	86.25	6.56	85.00	7.19
			1,000	13.75	3.75	93.75	5.62	100.00	7.81
			2,000	23.75	5.00	93.75	8.75	100.00	6.25
	1.0SD	250	250	17.50	13.13	50.00	13.44	43.75	7.81
			500	41.25	5.94	70.00	18.13	75.00	10.31
			1,000	43.75	4.37	76.25	30.94	97.50	10.00
			2,000	53.75	6.25	90.00	48.44	100.00	19.06
NDIF	30%	0.0SD	250	7.50	6.43	60.00	5.00	46.67	7.14
			500	3.33	4.29	89.17	10.36	86.67	6.07
			1,000	5.00	3.21	99.17	3.57	99.17	5.36
			2,000	2.50	5.36	100.00	5.36	100.00	6.07
		0.5SD	250	18.33	8.57	56.67	5.00	55.00	6.43
			500	20.00	7.14	73.33	7.14	80.83	6.07
			1,000	13.33	3.57	95.00	7.50	95.83	9.29
			2,000	20.83	3.93	100.00	8.93	100.00	11.79
	1.0SD	250	250	25.00	9.64	53.33	11.07	45.00	9.64
			500	33.33	8.21	73.33	17.86	77.50	10.00
			1,000	44.17	5.71	80.00	33.21	95.00	12.86
			2,000	54.17	6.07	98.33	48.21	100.00	21.07

ผลการวิเคราะห์ตามตาราง 22 สรุปได้ดังนี้

ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% ถึง 30% ความแตกต่างของการแจกแจงความสามารถ 0.0SD ถึง 1.0SD และขนาดตัวอย่าง 250 ถึง 2,000 คนต่อกลุ่ม (กลุ่มอ้างอิง และกลุ่มสนใจ) พบว่า วิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ มีค่าเฉลี่ยอัตราความถูกต้องอยู่ในช่วงประมาณ 29.50%-100%, 31.17%-100% และ 39.25%-100% ตามลำดับ ทั้งสามวิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในช่วงประมาณ 0%-13.06%, 5.06%-67.14% และ 5.56%-18.44% ตามลำดับ

ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% ถึง 30% ความแตกต่างของการแจกแจงความสามารถ 0.0SD ถึง 1.0SD และขนาดตัวอย่าง 250 ถึง 2,000 คนต่อกลุ่ม พบว่า วิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ มีค่าเฉลี่ยอัตราความถูกต้องอยู่ในช่วงประมาณ 2.50%-54.17%, 50%-100% และ 43.75%-100% ตามลำดับ ทั้งสามวิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในช่วงประมาณ 3.21%-14.17%, 2.50%-48.89% และ 5.36%-21.07% ตามลำดับ

2. การวิเคราะห์ความแปรปรวนหลายตัวแปรของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1

การวิเคราะห์ในตอนนี้ ผู้วิจัยได้นำผลการคำนวณอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ปัจจัยความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบ และปัจจัยขนาดตัวอย่าง ไปวิเคราะห์ความแตกต่าง โดยใช้สถิติการวิเคราะห์ความแปรปรวนหลายตัวแปร (Multivariate analysis of variance; MANOVA) ที่ระดับนัยสำคัญ .001 ซึ่งมีตัวแปรตาม 2 ตัว คือ อัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ส่วนตัวแปรอิสระมี 5 ตัว คือ วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ และปัจจัย 4 ปัจจัยดังที่กล่าวมาข้างต้น ถ้าผลการทดสอบมีนัยสำคัญทางสถิติแล้วจะทดสอบผลระหว่างกลุ่ม (Test of between-subjects effects) ของตัวแปรตามแต่ละตัว ที่ระดับนัยสำคัญ .001 ผลการวิเคราะห์มีรายละเอียดดังนี้

2.1 ผลการวิเคราะห์ความแปรปรวนหลายตัวแปร

ในการพิจารณาผลการวิเคราะห์ความแปรปรวนหลายตัวแปร (MANOVA) ที่มีนัยสำคัญทางสถิติ ผู้วิจัยจะพิจารณาเฉพาะประเด็นที่ตอบคำถามการวิจัย โดยพิจารณาผลการทดสอบปฏิสัมพันธ์สองทางและสามทาง ระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และวิธีการตรวจสอบการทำหน้าที่เปื่อยเบนของข้อสอบ ทั้งนี้เพื่อตอบคำถามว่า (1) ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัย มีผลต่ออัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ หรือไม่ และ (2) วิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ภายใต้เงื่อนไขเดียวกันของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัย มีอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกันหรือไม่ ดังนั้นในการวิเคราะห์ความแปรปรวนหลายตัวแปรจึงกำหนดโมเดลของการวิเคราะห์ให้มีปฏิสัมพันธ์ระหว่างตัวแปรอิสระไม่เกินอันดับสอง ผลการวิเคราะห์แสดงดังตาราง 23

ตาราง 23 การเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย

ผล	ค่าของสถิติ Wilks' Lambda	F
MDIF	0.31	4,333.70***
FDIF	0.74	1,864.54***
PDIF	0.97	69.76***
$d(\theta)$	0.62	1,431.56***
SIZE	0.34	2,561.01***
FDIF × MDIF	0.44	2,700.24***
PDIF × MDIF	0.94	90.27***
$d(\theta) \times MDIF$	0.64	680.37***

ตาราง 23 (ต่อ)

ผล	ค่าของสถิติ Wilks' Lambda	F
SIZE × MDIF	0.65	428.04***
FDIF × PDIF	0.97	94.35***
FDIF × $d(\theta)$	0.98	54.55***
FDIF × SIZE	0.86	280.21***
PDIF × $d(\theta)$	1.00	6.15***
PDIF × SIZE	0.98	22.33***
$d(\theta)$ × SIZE	0.86	135.43***
FDIF × PDIF × MDIF	0.94	90.63***
FDIF × $d(\theta)$ × MDIF	0.91	128.17***
FDIF × SIZE × MDIF	0.86	139.76***
PDIF × $d(\theta)$ × MDIF	0.99	6.81***
PDIF × SIZE × MDIF	0.97	12.08***
$d(\theta)$ × SIZE × MDIF	0.79	111.16***
FDIF × PDIF × $d(\theta)$	0.99	9.03***
FDIF × PDIF × SIZE	0.98	17.80***
FDIF × $d(\theta)$ × SIZE	0.99	8.56***
PDIF × $d(\theta)$ × SIZE	0.99	3.64***

***p < .001

ผลการวิเคราะห์ตามตาราง 23 การเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ตัวแปรอิสระ 5 ตัว คือ วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (MDIF) ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน (FDIF) ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน (PDIF) ปัจจัยความแตกต่างของการแจกแจงความสามารถ [$d(\theta)$] และปัจจัย

ขนาดตัวอย่าง (SIZE) โดยใช้การวิเคราะห์ความแปรปรวนหลายตัวแปร (MANOVA) ปรากฏผลดังนี้

วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (MDIF) ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน (FDIF) ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน (PDIF) ปัจจัยความแตกต่างของการแจกแจงความสามารถ $[d(\theta)]$ และปัจจัยขนาดตัวอย่าง (SIZE) มีผลต่ออัตราความถูกต้องหรืออัตราความคลาดเคลื่อนประเภทที่ 1 อย่างมีนัยสำคัญทางสถิติที่ระดับ .001

ปฏิสัมพันธ์สองทางในทุกเงื่อนไขของปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน (FDIF) ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน (PDIF) ปัจจัยความแตกต่างของการแจกแจงความสามารถ $[d(\theta)]$ ปัจจัยขนาดตัวอย่าง (SIZE) และวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (MDIF) มีผลต่ออัตราความถูกต้อง หรืออัตราความคลาดเคลื่อนประเภทที่ 1 อย่างมีนัยสำคัญทางสถิติที่ระดับ .001

ปฏิสัมพันธ์สามทางในทุกเงื่อนไขของปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน (FDIF) ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน (PDIF) ปัจจัยความแตกต่างของการแจกแจงความสามารถ $[d(\theta)]$ ปัจจัยขนาดตัวอย่าง (SIZE) และวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (MDIF) มีผลต่ออัตราความถูกต้อง หรืออัตราความคลาดเคลื่อนประเภทที่ 1 อย่างมีนัยสำคัญทางสถิติที่ระดับ .001

2.2 ผลการทดสอบผลระหว่างกลุ่ม

ผลการวิเคราะห์ความแตกต่างหลายตัวแปรตามตาราง 23 ผู้วิจัยสนใจผลการทดสอบที่มีนัยสำคัญเฉพาะเงื่อนไขปฏิสัมพันธ์สองทางและสามทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ดังนั้นจึงทดสอบผลระหว่างกลุ่ม (Test of between-subjects effects) ของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 เฉพาะเงื่อนไขดังกล่าว ผลการทดสอบแสดงดังตาราง 24

ตาราง 24 การทดสอบผลระหว่างกลุ่มของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้เงื่อนไขปฏิสัมพันธ์สองทางและสามทาง ระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย กับวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ

แหล่งของความแปรปรวน	ตัวแปรตาม	<i>F</i>
TDIF × MDIF	<i>P</i>	5,327.11***
	<i>E</i> ₁	796.63***
PDIF × MDIF	<i>P</i>	12.88***
	<i>E</i> ₁	164.91***
<i>d</i> (θ) × MDIF	<i>P</i>	213.04***
	<i>E</i> ₁	1,200.02***
SIZE × MDIF	<i>P</i>	81.49***
	<i>E</i> ₁	868.95***
FDIF × PDIF × MDIF	<i>P</i>	29.90***
	<i>E</i> ₁	144.76***
FDIF × <i>d</i> (θ) × MDIF	<i>P</i>	237.96***
	<i>E</i> ₁	14.45***
FDIF × SIZE × MDIF	<i>P</i>	162.60***
	<i>E</i> ₁	110.72***
PDIF × <i>d</i> (θ) × MDIF	<i>P</i>	2.53
	<i>E</i> ₁	10.78***
PDIF × SIZE × MDIF	<i>P</i>	8.02***
	<i>E</i> ₁	16.93***
<i>d</i> (θ) × SIZE × MDIF	<i>P</i>	8.33***
	<i>E</i> ₁	220.12***

****p* < .001

9) ปฏิสัมพันธ์สามทางระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน (PDIF) ปัจจัยปัจจัยขนาดตัวอย่าง (SIZE) กับวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (MDIF) มีผลต่ออัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 อย่างมีนัยสำคัญทางสถิติที่ระดับ .001

10) ปฏิสัมพันธ์สามทางระหว่างปัจจัยความแตกต่างของการแจกแจงความสามารถ $[d(\theta)]$ ปัจจัยขนาดตัวอย่าง (SIZE) กับวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (MDIF) มีผลต่ออัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 อย่างมีนัยสำคัญทางสถิติที่ระดับ .001

ดังนั้น การทดสอบผลระหว่างกลุ่มของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า มีนัยสำคัญทางสถิติจำนวน 9 และ 10 เงื่อนไข ตามลำดับ จากนั้นจึงทดสอบความแปรปรวนผลย่อย (Simple effect) ของอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ซึ่งจะเสนอผลการวิเคราะห์ในหัวข้อถัดไป

3. การวิเคราะห์ความแปรปรวนการทดสอบผลย่อยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 พร้อมทั้งการเปรียบเทียบภายหลัง

การวิเคราะห์ในตอนนี้ ผู้วิจัยนำผลการทดสอบปฏิสัมพันธ์สองทางและสามทาง ระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่มีนัยสำคัญทางสถิติ ไปวิเคราะห์ความแปรปรวนการทดสอบผลย่อย (Simple effect) ที่ระดับนัยสำคัญ .001 โดยทำการทดสอบใน 10 เงื่อนไข ดังนี้

เงื่อนไข 1 การทดสอบผลย่อยจากปฏิสัมพันธ์สองทาง ระหว่างปัจจัยรูปแบบของข้อสอบ ทำหน้าที่เบี่ยงเบน กับปัจจัยวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ($FDIF \times MDIF$)

เงื่อนไข 2 การทดสอบผลย่อยจากปฏิสัมพันธ์สองทาง ระหว่างปัจจัยสัดส่วนของข้อสอบ ทำหน้าที่เบี่ยงเบน กับปัจจัยวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ($PDIF \times MDIF$)

เงื่อนไข 3 การทดสอบผลย่อยจากปฏิสัมพันธ์สองทาง ระหว่างปัจจัยความแตกต่างของการแจกแจงความสามารถ และวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ $[d(\theta) \times MDIF]$

เงื่อนไข 4 การทดสอบผลย่อยจากปฏิสัมพันธ์สองทาง ระหว่างปัจจัยขนาดตัวอย่าง กับวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ($SIZE \times MDIF$)

เงื่อนไข 5 การทดสอบผลย่อยจากปฏิสัมพันธ์สามทาง ระหว่างปัจจัยรูปแบบของข้อสอบ ทำหน้าที่เบี่ยงเบน ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ($FDIF \times PDIF \times MDIF$)

เงื่อนไข 6 การทดสอบผลย่อยจากปฏิสัมพันธ์สามทาง ระหว่างปัจจัยรูปแบบของข้อสอบ ทำหน้าที่เบี่ยงเบน ปัจจัยความแตกต่างของการแจกแจงความสามารถ กับวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ [$FDIF \times d(\theta) \times MDIF$]

เงื่อนไข 7 การทดสอบผลย่อยจากปฏิสัมพันธ์สามทาง ระหว่างปัจจัยรูปแบบของข้อสอบ ทำหน้าที่เบี่ยงเบน ปัจจัยขนาดตัวอย่าง กับวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ [$FDIF \times SIZE \times MDIF$]

เงื่อนไข 8 การทดสอบผลย่อยจากปฏิสัมพันธ์สามทาง ระหว่างปัจจัยสัดส่วนของข้อสอบ ทำหน้าที่เบี่ยงเบน ปัจจัยความแตกต่างของการแจกแจงความสามารถ กับวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ [$PDIF \times d(\theta) \times MDIF$]

เงื่อนไข 9 การทดสอบผลย่อยจากปฏิสัมพันธ์สามทาง ระหว่างปัจจัยสัดส่วนของข้อสอบ ทำหน้าที่เบี่ยงเบน ปัจจัยขนาดตัวอย่าง กับวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ($PDIF \times SIZE \times MDIF$)

เงื่อนไข 10 การทดสอบผลย่อยจากปฏิสัมพันธ์สามทาง ระหว่างปัจจัยความแตกต่างของการแจกแจงความสามารถ ปัจจัยขนาดตัวอย่าง กับวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ [$d(\theta) \times SIZE \times MDIF$]

จากผลการทดสอบตามตาราง 24 พบว่า อัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 มีนัยสำคัญทางสถิติจำนวน 9 และ 10 เงื่อนไข ตามลำดับ ดังนั้นจึงนำผลการทดสอบภายใต้เงื่อนไขดังกล่าวไปทดสอบผลย่อย ถ้าผลการทดสอบมีนัยสำคัญทางสถิติจะทดสอบภายหลังโดยใช้วิธีของเชฟเฟ (Scheffé) ที่ระดับนัยสำคัญ .001 ผลการทดสอบมีรายละเอียดดังนี้

3.1 การทดสอบผลย่อยในเงื่อนไข 1

ผู้วิจัยนำผลการทดสอบปฏิสัมพันธ์สองทาง ระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ($FDIF \times MDIF$) ที่มีนัยสำคัญทางสถิติไปทดสอบผลย่อย (Simple effect) ดังตาราง 25

ตาราง 25 การทดสอบผลย่อยของปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน และวิธีการ-
ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ

แหล่งของความแปรปรวน	<i>F</i>	
	<i>P</i>	<i>E</i> ₁
ปัจจัยรูปแบบของ DIF		
FDIF ที่ MDIF1	4,516.54***	120.05***
FDIF ที่ MDIF2	9.29	353.75***
FDIF ที่ MDIF3	10.86	1.62
วิธีการตรวจสอบ DIF		
MDIF ที่ FDIF1	4.78	1,337.67***
MDIF ที่ FDIF2	3,527.10***	229.02***

****p* < .001

ผลการวิเคราะห์ตามตาราง 25 สรุปได้ดังนี้

3.1.1 ผลการทดสอบปัจจัยรูปแบบของ DIF

การทดสอบผลย่อยของปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ในแต่ละวิธีการ-
ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ พบว่า ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน
มีผลต่ออัตราความถูกต้องของวิธีโพลี-ชิปเทสท์อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผล
ต่อวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ส่วน
อัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า มีผลต่อวิธีโพลี-ชิปเทสท์ และวิธีการถดถอยโลจิสติก
แบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่อวิธีการถดถอยโลจิสติกแบบ
จัดอันดับหลายมิติ กล่าวคือ การตรวจสอบข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกันมีผล
ทำให้วิธีโพลี-ชิปเทสท์มีอัตราความถูกต้องสูงกว่าการตรวจสอบข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็น
รูปแบบเดียวกัน ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับ
หลายมิติสามารถตรวจสอบข้อสอบทำหน้าที่เบี่ยงเบนทั้งสองรูปแบบได้ใกล้เคียงกัน สำหรับอัตรา
ความคลาดเคลื่อนประเภทที่ 1 พบว่าการตรวจสอบในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็น
รูปแบบเดียวกันมีผลทำให้วิธีโพลี-ชิปเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าในเงื่อนไข

ข้อสอบที่ทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ซึ่งตรงข้ามกับวิธีการถดถอยโลจิสติกแบบจัดอันดับ ในขณะที่ไม่มีผลต่อวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ

3.1.2 ผลการทดสอบวิธีการตรวจสอบ DIF

การทดสอบผลย่อยของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ในแต่ละเงื่อนไขของปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน พบว่า วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน มีผลต่ออัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ยกเว้นอัตราความถูกต้องภายใต้เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน ดังนั้นจึงนำผลการทดสอบที่มีนัยสำคัญทางสถิติไปเปรียบเทียบรายคู่ โดยใช้วิธีของเชฟเฟ (Scheffé) (แสดงรายละเอียดของผลการวิเคราะห์ดังตาราง 26 ในภาคผนวก ก) ผลการเปรียบเทียบจำแนกออกเป็น 2 ลักษณะ ดังนี้

1) ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง ภายใต้เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนหลายรูปแบบ พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยอัตราความถูกต้องไม่แตกต่างกัน ทั้งสองวิธีมีค่าเฉลี่ยอัตราความถูกต้องสูงกว่าวิธีโพลี-ชิปทดสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ .001

2) ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ปรากฏว่า ให้ผลสอดคล้องกัน กล่าวคือ ในทุกคู่ของการเปรียบเทียบ ค่าเฉลี่ยอัตราความคลาดเคลื่อนของวิธีการตรวจสอบสามวิธีมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001

3.2 การทดสอบผลย่อยในเงื่อนไข 2

ผู้วิจัยนำผลการทดสอบปฏิสัมพันธ์สองทาง ระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน กับวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (PDIF × MDIF) ที่มีนัยสำคัญทางสถิติไปทดสอบผลย่อย (Simple effect) ดังตาราง 27

ตาราง 27 การทดสอบผลย่อยของปัจจัยรูปสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ

แหล่งของความแปรปรวน	<i>F</i>	
	<i>P</i>	<i>E</i> ₁
ปัจจัยสัดส่วนของ DIF		
PDIF ที่ MDIF1	6.43	6.47
PDIF ที่ MDIF2	11.25***	108.18***
PDIF ที่ MDIF3	.12	.02
วิธีการตรวจสอบ DIF		
MDIF ที่ PDIF1	430.84***	210.43***
MDIF ที่ PDIF2	407.78***	414.56***
MDIF ที่ PDIF3	470.96***	775.00***

****p* < .001

ผลการวิเคราะห์ตามตาราง 27 สรุปได้ดังนี้

3.2.1 ผลการทดสอบปัจจัยสัดส่วนของ DIF

การทดสอบผลย่อยของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ในแต่ละวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ พบว่า ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน มีผลต่ออัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่ออัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลีชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ดังนั้นจึงนำผลการทดสอบที่มีนัยสำคัญทางสถิติไปเปรียบเทียบความแตกต่างรายคู่โดยใช้วิธีของเชฟเฟ (Scheffé) (แสดงรายละเอียดของผลการวิเคราะห์ดังตาราง 28 ในภาคผนวก ก) ผลการเปรียบเทียบจำแนกออกเป็น 2 ลักษณะ ดังนี้

1) ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง ภายใต้วิธีการถดถอยโลจิสติกแบบจัดอันดับ พบว่า ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% มีค่าเฉลี่ยอัตราความถูกต้องสูงกว่าในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% อย่างมีนัยสำคัญทางสถิติที่

ระดับ .001 ส่วนค่าเฉลี่ยระหว่างคู่อื่นๆ ไม่แตกต่างกัน ผลดังกล่าวแสดงว่า เมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้น 20% (จาก 10% ถึง 30%) มีผลทำให้วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องลดลง แต่ไม่มีผลต่อวิธีโพลี-ชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ

2) ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้วิธีการถดถอยโลจิสติกแบบจัดอันดับ พบว่า ค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ในทุกเงื่อนไข ผลดังกล่าวแสดงว่า เมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้นในช่วง 10% ถึง 30% มีผลทำให้วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้น แต่ไม่มีผลต่อวิธีโพลี-ชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ

3.2.2 ผลการทดสอบวิธีการตรวจสอบ DIF

การทดสอบผลย่อยของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ในแต่ละเงื่อนไขของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน พบว่า วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10%, 20% และ 30% มีผลต่ออัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แสดงว่า ค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบอย่างน้อยหนึ่งคู่มีค่าแตกต่างกัน ดังนั้นจึงเปรียบเทียบความแตกต่างรายคู่ โดยใช้วิธีของเชฟเฟ (Scheffé) (แสดงรายละเอียดของผลการวิเคราะห์ดังตาราง 29 ในภาคผนวก ก) ผลการเปรียบเทียบจำแนกออกเป็น 2 ลักษณะ ดังนี้

1) ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% และ 20% ปรากฏว่าให้ผลสอดคล้องกัน กล่าวคือ ค่าเฉลี่ยอัตราความถูกต้องระหว่างวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติไม่แตกต่างกัน ทั้งสองวิธีมีค่าเฉลี่ยสูงกว่าวิธีโพลี-ชิปเทสท์อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ส่วนการเปรียบเทียบภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% พบว่า ค่าเฉลี่ยอัตราความถูกต้องของวิธีการตรวจสอบสามวิธีมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ในทุกคู่ของการเปรียบเทียบ

2) ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10%, 20% และ 30% พบว่าให้ผลสอดคล้องกัน กล่าวคือในทุกคู่ของการเปรียบเทียบ ค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบสามวิธีมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001

3.3 การทดสอบผลย่อยในเงื่อนไข 3

ผู้วิจัยนำผลการทดสอบปฏิสัมพันธ์สองทาง ระหว่างปัจจัยความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบ กับวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ [$d(\theta) \times \text{MDIF}$] ที่มีนัยสำคัญทางสถิติไปทดสอบผลย่อย (Simple effect) ดังตาราง 30

ตาราง 30 การทดสอบผลย่อยของปัจจัยความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบ และวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ

แหล่งของความแปรปรวน	<i>F</i>	
	<i>P</i>	<i>E</i> ₁
ปัจจัยความแตกต่างของ		
การแจกแจงความสามารถ		
$d(\theta)$ ที่ MDIF1	87.88***	25.64***
$d(\theta)$ ที่ MDIF2	6.24	2,041.91***
$d(\theta)$ ที่ MDIF3	8.85	126.89***
วิธีการตรวจสอบ DIF		
MDIF ที่ $d(\theta)$ 1	717.53***	157.86***
MDIF ที่ $d(\theta)$ 2	526.36***	163.57***
MDIF ที่ $d(\theta)$ 3	177.02***	2,413.78***

*** $p < .001$

ผลการวิเคราะห์ตามตาราง 30 สรุปได้ดังนี้

3.3.1 ผลการทดสอบปัจจัยความแตกต่างของการแจกแจงความสามารถ

การทดสอบผลย่อยของปัจจัยความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบ ในแต่ละวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ พบว่า ปัจจัยความแตกต่างของการแจกแจงความสามารถมีผลต่ออัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ยกเว้นอัตราความถูกต้องของวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ดังนั้นจึงนำผลการทดสอบที่มีนัยสำคัญไปเปรียบเทียบรายคู่ โดยใช้วิธีของเชฟเฟ (Scheffé) (แสดงรายละเอียดของผลการวิเคราะห์ดังตาราง 31 ในภาคผนวก ก) ผลการเปรียบเทียบจำแนกออกเป็น 2 ลักษณะดังนี้

1) ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้องภายใต้วิธีโพลี-ชิปเทสต์ พบว่า ภายใต้ทุกเงื่อนไขของการเปรียบเทียบ ค่าเฉลี่ยของอัตราความถูกต้องมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ผลดังกล่าวแสดงว่า เมื่อความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มเพิ่มขึ้นในช่วง 0.0SD ถึง 1.0SD มีผลทำให้อัตราความถูกต้องของวิธีโพลี-ชิปเทสต์เพิ่มขึ้น แต่ไม่มีผลต่อวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ

2) ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้วิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ พบว่าให้ผลสอดคล้องกัน กล่าวคือ ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 1.0SD มีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD และ 0.5SD อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ส่วนในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD และ 0.5SD ค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่แตกต่างกัน ผลดังกล่าวแสดงว่า เมื่อความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบเพิ่มขึ้น 1.0SD (จาก 0.0SD ถึง 1.0SD) มีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งสามวิธีเพิ่มขึ้น

3.3.2 ผลการทดสอบวิธีการตรวจสอบ DIF

การทดสอบผลย่อยของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ในแต่ละเงื่อนไขของปัจจัยความแตกต่างของการแจกแจงความสามารถ พบว่า วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้เงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD, 0.5SD และ 1.0SD มีผลต่ออัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แสดงว่า อัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบอย่างน้อยหนึ่งคู่มีค่าแตกต่างกัน ดังนั้นจึงเปรียบเทียบรายคู่ โดยใช้วิธีของเชฟเฟ (Scheffé) (แสดงรายละเอียดของผลการวิเคราะห์ดังตาราง 32 ในภาคผนวก ก) ผลการเปรียบเทียบจำแนกออกเป็น 2 ลักษณะ ดังนี้

1) ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง ภายใต้เงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD, 0.5SD และ 1.0SD พบว่าให้ผลสอดคล้องกัน กล่าวคือ ค่าเฉลี่ยอัตราความถูกต้องของวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ และวิธีการถดถอยโลจิสติกแบบจัดอันดับไม่แตกต่างกัน ทั้งสองวิธีมีค่าเฉลี่ยสูงกว่าวิธีโพลี-ชิปทดสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ .001

2) ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้เงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD, 0.5SD และ 1.0SD พบว่าให้ผลสอดคล้องกัน กล่าวคือ ค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบสามวิธีแตกต่างกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ในทุกคู่ของการเปรียบเทียบ

3.4 การทดสอบผลย่อยในเงื่อนไข 4

ผู้วิจัยนำผลการทดสอบปฏิสัมพันธ์สองทาง ระหว่างปัจจัยขนาดตัวอย่าง กับวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (SIZE × MDIF) ที่มีนัยสำคัญทางสถิติไปทดสอบผลย่อย (Simple effect) ดังตาราง 33

ตาราง 33 การทดสอบผลย่อยของปัจจัยขนาดตัวอย่าง และวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ

แหล่งของความแปรปรวน	<i>F</i>	
	<i>P</i>	<i>E</i> ₁
ปัจจัยขนาดตัวอย่าง		
SIZE ที่ MDIF1	406.53***	47.84***
SIZE ที่ MDIF2	749.55***	815.63***
SIZE ที่ MDIF3	916.70***	39.29***
วิธีการตรวจสอบ DIF		
MDIF ที่ SIZE1	261.69***	2.38
MDIF ที่ SIZE2	442.10***	158.96***
MDIF ที่ SIZE3	719.63***	591.01***
MDIF ที่ SIZE4	710.31***	1,751.17***

****p* < .001

ผลการวิเคราะห์ตามตาราง 33 สรุปได้ดังนี้

3.4.1 ผลการทดสอบปัจจัยขนาดตัวอย่าง

การทดสอบผลย่อยของปัจจัยขนาดตัวอย่าง ในแต่ละวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ พบว่า ปัจจัยขนาดตัวอย่าง 250, 500, 1000 และ 2,000 คนต่อกลุ่ม (กลุ่มอ้างอิง และกลุ่มสนใจ) มีผลต่ออัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ดังนั้นจึงเปรียบเทียบความแตกต่างรายคู่โดยใช้วิธีของเชฟเฟ (Scheffé) (แสดงรายละเอียดของผลการวิเคราะห์ดังตาราง 34 ในภาคผนวก ก) ผลการเปรียบเทียบจำแนกออกเป็น 2 ลักษณะ ดังนี้

1) ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง ภายใต้วิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ พบว่า ในทุกคู่ของการเปรียบเทียบภายใต้เงื่อนไขขนาดตัวอย่าง ค่าเฉลี่ยของอัตราความถูกต้องมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ยกเว้นการเปรียบเทียบเงื่อนไขขนาดตัวอย่าง 1,000 และ 2,000 คนต่อกลุ่ม ของวิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ผลการศึกษาดังกล่าวแสดงว่า เมื่อขนาดตัวอย่างเพิ่มขึ้นในช่วง 250 ถึง 2,000 คนต่อกลุ่ม มีผลทำให้อัตราความถูกต้องของวิธีการตรวจสอบทั้งสามวิธีมีค่าเพิ่มขึ้นเกือบทุกเงื่อนไข

2) ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้วิธีโพลี-ชิปเทสต์ พบว่า ค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ในเงื่อนไขขนาดตัวอย่าง 250 คนในแต่ละกลุ่ม มีค่าสูงกว่าเงื่อนไขขนาดตัวอย่าง 500, 1,000 และ 2,000 คนต่อกลุ่มอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ส่วนการเปรียบเทียบคู่อื่นๆ ไม่แตกต่างกัน สำหรับการเปรียบเทียบภายใต้วิธีการถดถอยโลจิสติกแบบจัดอันดับ พบว่า ค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ในทุกเงื่อนไข มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ส่วนการเปรียบเทียบภายใต้วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ พบว่า ค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างเงื่อนไข 250 ถึง 1,000 คนต่อกลุ่ม ไม่แตกต่างกัน ส่วนการเปรียบเทียบคู่อื่นๆ แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ผลดังกล่าวแสดงว่า เมื่อขนาดตัวอย่างเพิ่มขึ้นในช่วง 250 ถึง 2,000 คนต่อกลุ่ม มีผลทำให้วิธีโพลี-ชิปเทสต์มีอัตราความคลาดเคลื่อนประเภทที่ 1 ลดลงในบางเงื่อนไข แต่มีผลทำให้วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นในทุกเงื่อนไข ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นในบางเงื่อนไข

3.4.2 ผลการทดสอบวิธีการตรวจสอบ DIF

การทดสอบผลย่อยของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ในแต่ละเงื่อนไขของปัจจัยขนาดตัวอย่าง พบว่า วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้เงื่อนไขขนาดตัวอย่าง 250, 500, 1000 และ 2,000 คนต่อกลุ่ม มีผลต่ออัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ยกเว้นอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้เงื่อนไขขนาดตัวอย่าง 250 คนต่อกลุ่ม ดังนั้นจึงนำผลการทดสอบที่มีนัยสำคัญทางสถิติไปเปรียบเทียบรายคู่ โดยใช้วิธีของเชฟเฟ (Scheffé) (แสดงรายละเอียดของผลการวิเคราะห์ดังตาราง 35 ในภาคผนวก ก) ผลการเปรียบเทียบจำแนกออกเป็น 2 ลักษณะ ดังนี้

1) ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง ภายใต้เงื่อนไขขนาดตัวอย่าง 250, 500 และ 2,000 คนต่อกลุ่ม พบว่าให้ผลสอดคล้องกัน กล่าวคือ ค่าเฉลี่ยอัตราความถูกต้องของวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติไม่แตกต่างกัน ทั้งสองวิธีมีค่าเฉลี่ยสูงกว่าวิธีโพลี-ชิปทดสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ส่วนภายใต้เงื่อนไขขนาดตัวอย่าง 1,000 คนต่อกลุ่ม พบว่า ค่าเฉลี่ยอัตราความถูกต้องของทั้งสามวิธีมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ในทุกคู่ของการเปรียบเทียบ

2) ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้เงื่อนไขขนาดตัวอย่าง 500, 1,000 และ 2,000 คนต่อกลุ่ม พบว่าให้ผลสอดคล้องกัน กล่าวคือ ค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบสามวิธีมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ในทุกคู่ของการเปรียบเทียบ

3.5 การทดสอบผลย่อยในเงื่อนไข 5

ผู้วิจัยนำผลการทดสอบปฏิสัมพันธ์สามทาง ระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน กับวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบน ของข้อสอบ ($FDIF \times PDIF \times MDIF$) ที่มีนัยสำคัญทางสถิติไปทดสอบผลย่อย (Simple effect) ดังตาราง 36

ตาราง 36 การทดสอบผลย่อยของปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ

แหล่งของความแปรปรวน	F	
	P	E_1
ปฏิสัมพันธ์สองทางระหว่างปัจจัย		
FDIF กับ PDIF		
FDIF $\times$ PDIF ที่ MDIF1	12.06***	1.02
FDIF $\times$ PDIF ที่ MDIF2	30.27***	116.96***
FDIF $\times$ PDIF ที่ MDIF3	.27	.20

ตาราง 36 (ต่อ)

แหล่งของความแปรปรวน	<i>F</i>	
	<i>P</i>	<i>E</i> ₁
วิธีการตรวจสอบ DIF		
MDIF ที่ $FDIF1 \times PDIF1$	5.56	167.36***
MDIF ที่ $FDIF1 \times PDIF2$	1.21	420.39***
MDIF ที่ $FDIF1 \times PDIF3$	22.50***	994.11***
MDIF ที่ $FDIF2 \times PDIF1$	1,227.69***	68.94***
MDIF ที่ $FDIF2 \times PDIF2$	1,122.47***	85.39***
MDIF ที่ $FDIF2 \times PDIF3$	1,220.46***	85.57***

****p* < .001

ผลการวิเคราะห์ตามตาราง 36 สรุปได้ดังนี้

3.5.1 ผลการทดสอบปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของ DIF กับ ปัจจัยสัดส่วนของ DIF

การทดสอบผลย่อยของปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่ เบี่ยงเบน กับปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ในแต่ละวิธีการตรวจสอบการทำหน้าที่ เบี่ยงเบนของข้อสอบ พบว่า ปฏิสัมพันธ์ระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับ ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน มีผลต่ออัตราความถูกต้องของวิธีโพลี-ชิปเทสท์ และ วิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญที่ระดับ .001 แต่ไม่มีผลต่อวิธีการถดถอย-โลจิสติกแบบจัดอันดับหลายมิติ ในขณะเดียวกัน ปฏิสัมพันธ์ระหว่างปัจจัยดังกล่าว มีผลต่ออัตรา ความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญที่ระดับ .001 แต่ไม่มีผลต่อวิธีโพลี-ชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ดังนั้นจึงนำ ผลการทดสอบที่มีนัยสำคัญทางสถิติไปเปรียบเทียบความแตกต่างรายคู่ โดยใช้วิธีของเซฟเฟ (Scheffé) (แสดงรายละเอียดของผลการวิเคราะห์ดังตาราง 37 ในภาคผนวก ก) ผลการเปรียบเทียบ จำแนกออกเป็น 2 ลักษณะ ดังนี้

1) ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้องภายใต้วิธีโพลี-ชิปเทสท์ พบว่า ในเงื่อนไขรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน เมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้น 20% (จาก 10% ถึง 30%) มีผลทำให้ค่าเฉลี่ยของอัตราความถูกต้องเพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ส่วนในเงื่อนไขรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ปริมาณสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนในทุกเงื่อนไขไม่มีผลต่อค่าเฉลี่ยของอัตราความถูกต้อง เมื่อพิจารณาเงื่อนไขที่มีค่าเฉลี่ยแตกต่างกันมากที่สุด ตรงกับเงื่อนไขรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน กับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% ($P = 81.17\%$) และรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน กับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% ($P = 20.00\%$) ซึ่งแตกต่างกันประมาณ 61.17%

ส่วนผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้องภายใต้วิธีการถดถอยโลจิสติกแบบจัดอันดับ พบว่า ในเงื่อนไขรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน เมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้นในช่วง 10% ถึง 30% มีผลทำให้ค่าเฉลี่ยของอัตราความถูกต้องเพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ส่วนในเงื่อนไขรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ปริมาณสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนในทุกเงื่อนไขไม่มีผลต่อค่าเฉลี่ยของอัตราความถูกต้อง เมื่อพิจารณาเงื่อนไขที่มีค่าเฉลี่ยแตกต่างกันมากที่สุด ตรงกับเงื่อนไขรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน กับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% ($P = 83.79\%$) และรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน กับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% ($P = 70.28\%$) ซึ่งแตกต่างกันประมาณ 13.51%

2) ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้วิธีการถดถอยโลจิสติกแบบจัดอันดับ พบว่า ในเงื่อนไขรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน เมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้นในทุกเงื่อนไข (10%, 20% และ 30%) มีผลทำให้ค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ส่วนในเงื่อนไขรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ปริมาณสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนในทุกเงื่อนไขไม่มีผลต่อค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 เมื่อพิจารณาเงื่อนไขที่มีค่าเฉลี่ยแตกต่างกันมากที่สุด ตรงกับเงื่อนไข

รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน กับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% ($P = 26.80\%$) และรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน กับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% ($P = 13.60\%$) ซึ่งแตกต่างกันประมาณ 13.20%

3.5.2 ผลการทดสอบวิธีการตรวจสอบ DIF

การทดสอบผลย่อยของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ในแต่ละเงื่อนไขของปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน พบว่า วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้เงื่อนไขปฏิสัมพันธ์ระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน มีผลต่ออัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 อย่างมีนัยสำคัญที่ระดับ .001 ยกเว้นอัตราความถูกต้องภายใต้ปฏิสัมพันธ์ใน 2 เงื่อนไข คือ (1) รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน กับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% และ (2) รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน กับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% ดังนั้นจึงนำผลการทดสอบที่มีนัยสำคัญทางสถิติไปเปรียบเทียบรายคู่ โดยใช้วิธีของเซฟเฟ (Scheffé) (แสดงรายละเอียดของผลการวิเคราะห์ดังตาราง 38 ในภาคผนวก ก) ผลการเปรียบเทียบจำแนกออกเป็น 2 ลักษณะ ดังนี้

1) ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% ($FDIF1 \times PDIF3$) พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยอัตราความถูกต้องสูงกว่าวิธีโพลี-ชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ส่วนผลการเปรียบเทียบภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ในทุกเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ($FDIF2 \times PDIF1$, $FDIF2 \times PDIF2$ และ $FDIF2 \times PDIF3$) พบว่า ในทุกเงื่อนไข วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยของอัตราความถูกต้องไม่แตกต่างกัน ทั้งสองวิธีมีค่าเฉลี่ยของอัตราความถูกต้องสูงกว่าวิธีโพลี-ชิปเทสท์

2) ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1

ผลการเปรียบเทียบค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ทุกเงื่อนไขของปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ปรากฏว่าให้ผลสอดคล้องกัน กล่าวคือ ในทุกคู่ของการเปรียบเทียบ

ค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบสามวิธีมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001

3.6 การทดสอบผลย่อยในเงื่อนไข 6

ผู้วิจัยนำผลการทดสอบปฏิสัมพันธ์สามทาง ระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ปัจจัยความแตกต่างของการแจกแจงความสามารถ กับวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ [$FDIF \times d(\theta) \times MDIF$] ที่มีนัยสำคัญทางสถิติไปทดสอบผลย่อย (Simple effect) ดังตาราง 39

ตาราง 39 การทดสอบผลย่อยของปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถ และวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ

แหล่งของความแปรปรวน	<i>F</i>	
	<i>P</i>	<i>E</i> ₁
ปฏิสัมพันธ์สองทางระหว่างปัจจัย		
FDIF กับ $d(\theta)$		
FDIF $\times d(\theta)$ ที่ MDIF1	162.16***	.63
FDIF $\times d(\theta)$ ที่ MDIF2	25.42***	17.95***
FDIF $\times d(\theta)$ ที่ MDIF3	4.87	.71
วิธีการตรวจสอบ DIF		
MDIF ที่ FDIF1 $\times d(\theta)$ 1	13.43***	321.95***
MDIF ที่ FDIF1 $\times d(\theta)$ 2	2.58	284.26***
MDIF ที่ FDIF1 $\times d(\theta)$ 3	4.02	1,717.40***
MDIF ที่ FDIF2 $\times d(\theta)$ 1	2,018.84***	6.22
MDIF ที่ FDIF2 $\times d(\theta)$ 2	1,513.95***	9.96
MDIF ที่ FDIF2 $\times d(\theta)$ 3	504.12***	918.94***

*** $p < .001$

ผลการวิเคราะห์ตามตาราง 39 สรุปได้ดังนี้

3.6.1 ผลการทดสอบปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของ DIF กับ ปัจจัยความแตกต่างของการแจกแจงความสามารถ

การทดสอบผลย่อยของปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่ย่อย เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถ ในแต่ละวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ พบว่า ปฏิสัมพันธ์ระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่ย่อย เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่ม มีผลต่ออัตราความถูกต้องของวิธีโพลี-ชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่อวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ในขณะเดียวกัน ปฏิสัมพันธ์ระหว่างปัจจัยดังกล่าว มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่อวิธีโพลี-ชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ดังนั้นจึงนำผลการทดสอบที่มีนัยสำคัญทางสถิติไปเปรียบเทียบรายคู่โดยใช้วิธีของเซฟเฟ (Scheffé) (แสดงรายละเอียดของผลการวิเคราะห์ดังตาราง 40 ในภาคผนวก ก) ผลการเปรียบเทียบจำแนกออกเป็น 2 ลักษณะ ดังนี้

1) ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้องภายใต้วิธีโพลี-ชิปเทสท์ พบว่า ในเงื่อนไขรูปแบบของข้อสอบทำหน้าที่ย่อย เบี่ยงเบนที่เป็นรูปแบบเดียวกัน เมื่อกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถเพิ่มขึ้นในทุกเงื่อนไข (จาก 0.0SD ถึง 1.0SD) ไม่มีผลต่อค่าเฉลี่ยของอัตราความถูกต้อง ส่วนในเงื่อนไขรูปแบบของข้อสอบทำหน้าที่ย่อย เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน เมื่อกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถเพิ่มขึ้นในทุกเงื่อนไข มีผลทำให้ค่าเฉลี่ยของอัตราความถูกต้องเพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 เมื่อพิจารณาเงื่อนไขที่มีค่าเฉลี่ยแตกต่างกันมากที่สุด ตรงกับเงื่อนไขรูปแบบของข้อสอบทำหน้าที่ย่อย เบี่ยงเบนที่เป็นรูปแบบเดียวกัน กับความแตกต่างของการแจกแจงความสามารถ 0.5SD ($P = 77.30\%$) และรูปแบบของข้อสอบทำหน้าที่ย่อย เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน กับความแตกต่างของการแจกแจงความสามารถ 0.0SD ($P = 6.53\%$) ซึ่งแตกต่างกันประมาณ 70.80%

ส่วนผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้องภายใต้วิธีการถดถอยโลจิสติกแบบจัดอันดับ พบว่า ในเงื่อนไขรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน เมื่อกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถเพิ่มขึ้นในทุกเงื่อนไข ให้ผลสอดคล้องกับวิธีโพลี-ชิปเทสท์ ส่วนในเงื่อนไขรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน เมื่อความแตกต่างของการแจกแจงความสามารถเพิ่มขึ้น 0.5SD (จาก 0.0SD ถึง 0.5SD) ไม่มีผลต่อค่าเฉลี่ยของอัตราความถูกต้อง แต่ถ้าความแตกต่างของการแจกแจงความสามารถเพิ่มขึ้น 1.0SD (จาก 0.0SD ถึง 1.0SD) มีผลทำให้ค่าเฉลี่ยของอัตราความถูกต้องลดลงอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 เมื่อพิจารณาเงื่อนไขที่มีค่าเฉลี่ยแตกต่างกันมากที่สุด ตรงกับเงื่อนไขรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน กับความแตกต่างของการแจกแจงความสามารถ 0.0SD ($P = 83.82\%$) และรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ที่ไม่เป็นรูปแบบเดียวกัน กับความแตกต่างของการแจกแจงความสามารถ 0.0SD ($P = 73.65\%$) ซึ่งแตกต่างกันประมาณ 10.17%

2) ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้วิธีการถดถอยโลจิสติกแบบจัดอันดับ พบว่า ในเงื่อนไขรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน เมื่อกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถเพิ่มขึ้น 0.5SD (จาก 0.0SD ถึง 0.5SD) ไม่มีผลต่อค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 แต่ถ้ากลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถเพิ่มขึ้น 1.0SD (จาก 0.0SD ถึง 1.0SD) มีผลทำให้ค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 เมื่อพิจารณาเงื่อนไขที่มีค่าเฉลี่ยแตกต่างกันมากที่สุด ตรงกับเงื่อนไขรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน กับความแตกต่างของการแจกแจงความสามารถ 1.0SD ($P = 32.25\%$) และรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน กับความแตกต่างของการแจกแจงความสามารถ 0.0SD ($P = 6.05\%$) ซึ่งแตกต่างกันประมาณ 26.20%

3.6.2 ผลการทดสอบวิธีการตรวจสอบ DIF

การทดสอบผลย่อยของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ในแต่ละเงื่อนไขของปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบ พบว่า วิธีการตรวจสอบการทำ

หน้าที่เบี่ยงเบนของข้อสอบ ภายใต้เงื่อนไขปฏิสัมพันธ์ระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถ มีผลต่ออัตราความถูกต้องอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ยกเว้นอัตราความถูกต้องภายใต้ปฏิสัมพันธ์ใน 2 เงื่อนไข คือ (1) รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน กับความแตกต่างของการแจกแจงความสามารถ 0.5SD และ (2) รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน กับความแตกต่างของการแจกแจงความสามารถ 1.0SD ในขณะเดียวกัน ปฏิสัมพันธ์ระหว่างปัจจัยดังกล่าว มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ยกเว้นอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ปฏิสัมพันธ์ใน 2 เงื่อนไข คือ (1) รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน กับความแตกต่างของการแจกแจงความสามารถ 0.0SD และ (2) รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน กับความแตกต่างของการแจกแจงความสามารถ 0.5SD ดังนั้นจึงนำผลการทดสอบที่มีนัยสำคัญไปเปรียบเทียบรายคู่โดยใช้วิธีของเชฟเฟ (Scheffé) (แสดงรายละเอียดของผลการวิเคราะห์ดังตาราง 41 ในภาคผนวก ก) ผลการเปรียบเทียบจำแนกออกเป็น 2 ลักษณะ ดังนี้

1) ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD [$FDIF1 \times d(\theta) 1$] พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยอัตราความถูกต้องสูงกว่าวิธีโพลี-ชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ส่วนผลการเปรียบเทียบภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD และ 0.5SD [$FDIF2 \times d(\theta) 1$ และ $FDIF2 \times d(\theta) 2$] พบว่าให้ผลสอดคล้องกัน กล่าวคือ วิธีการถดถอยโลจิสติกแบบจัดอันดับและวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยของอัตราความถูกต้องไม่แตกต่างกัน ทั้งสองวิธีมีค่าเฉลี่ยของอัตราความถูกต้องสูงกว่าวิธีโพลี-ชิปเทสท์อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ส่วนในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 1.0SD [$FDIF2 \times d(\theta) 3$] พบว่า ทั้งสามวิธีมีค่าเฉลี่ยของอัตราความถูกต้องแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001

2) ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1

ผลการเปรียบเทียบค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่ย่อยแบ่งเป็นรูปแบบเดียวกัน ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบทุกเงื่อนไข [$FDIF1 \times d(\theta) 1$, $FDIF1 \times d(\theta) 2$ และ $FDIF1 \times d(\theta) 3$] พบว่าให้ผลสอดคล้องกัน กล่าวคือ ค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบสามวิธีมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ในทุกคู่ของการเปรียบเทียบ ส่วนภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่ย่อยแบ่งเป็นรูปแบบเดียวกัน ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบ 1.0SD [$FDIF2 \times d(\theta) 3$] พบว่าให้ผลสอดคล้องกับผลการเปรียบเทียบภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่ย่อยแบ่งเป็นรูปแบบเดียวกัน

3.7 การทดสอบผลย่อยในเงื่อนไข 7

ผู้วิจัยนำผลการทดสอบปฏิสัมพันธ์สามทาง ระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่ย่อยแบ่งเป็น ปัจจัยขนาดตัวอย่าง กับวิธีการตรวจสอบการทำหน้าที่ย่อยแบ่งของข้อสอบ [$FDIF \times SIZE \times MDIF$] ที่มีนัยสำคัญทางสถิติไปทดสอบผลย่อย (Simple effect) ดังตาราง 42

ตาราง 42 การทดสอบผลย่อยของปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่ย่อยแบ่ง กับปัจจัยขนาดตัวอย่าง และวิธีการตรวจสอบการทำหน้าที่ย่อยแบ่งของข้อสอบ

แหล่งของความแปรปรวน	F	
	P	E_1
ปฏิสัมพันธ์สองทางระหว่างปัจจัย		
FDIF กับ SIZE		
FDIF $\times$ SIZE ที่ MDIF1	588.82***	2.27
FDIF $\times$ SIZE ที่ MDIF2	69.46***	106.33***
FDIF $\times$ SIZE ที่ MDIF3	8.10***	.48

ตาราง 42 (ต่อ)

แหล่งของความแปรปรวน	<i>F</i>	
	<i>P</i>	<i>E</i> ₁
วิธีการตรวจสอบ DIF		
MDIF ที่ $FDIF1 \times SIZE1$	28.21***	11.76***
MDIF ที่ $FDIF1 \times SIZE2$	5.10	161.50***
MDIF ที่ $FDIF1 \times SIZE3$	3.28	602.23***
MDIF ที่ $FDIF1 \times SIZE4$	.01	1,849.38***
MDIF ที่ $FDIF2 \times SIZE1$	848.21***	3.31
MDIF ที่ $FDIF2 \times SIZE2$	2,002.80***	34.95***
MDIF ที่ $FDIF2 \times SIZE3$	3,216.58***	128.10***
MDIF ที่ $FDIF2 \times SIZE4$	3,168.92***	353.99***

***p < .001

ผลการวิเคราะห์ตามตาราง 42 สรุปได้ดังนี้

3.7.1 ผลการทดสอบปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของ DIF กับ ปัจจัยขนาดตัวอย่าง

การทดสอบผลย่อยของปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่ เบี่ยงเบน กับปัจจัยขนาดตัวอย่าง ในแต่ละวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ พบว่า ปฏิสัมพันธ์ระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยขนาดตัวอย่าง มีผลต่ออัตราความถูกต้องของวิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติอย่างมีนัยสำคัญที่ระดับ .001 ในขณะเดียวกัน ปฏิสัมพันธ์ระหว่างปัจจัยดังกล่าว มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญที่ระดับ .001 แต่ไม่มีผลต่อวิธีโพลี-ชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ดังนั้นจึงนำผลการทดสอบที่มีนัยสำคัญไปเปรียบเทียบรายคู่ โดยใช้วิธีของเชฟเฟ (Scheffé) (แสดงรายละเอียดของผลการวิเคราะห์ดังตาราง 43 ในภาคผนวก ก) ผลการเปรียบเทียบจำแนกเป็น 2 ลักษณะ ดังนี้

1) การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้องภายใต้วิธีโพลี-ชิปเทสท์ พบว่า ในเงื่อนไขรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และที่ไม่เป็นรูปแบบเดียวกัน เมื่อขนาดตัวอย่างเพิ่มขึ้นในทุกเงื่อนไข (250, 500, 1,000 และ 2,000 คนต่อกลุ่ม) มีผลทำให้ค่าเฉลี่ยของอัตราความถูกต้องเพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ยกเว้นใน 2 เงื่อนไข คือ (1) รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน เมื่อขนาดตัวอย่างเพิ่มขึ้นจาก 1,000 เป็น 2,000 คนต่อกลุ่ม และ (2) รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน เมื่อขนาดตัวอย่างเพิ่มขึ้นในช่วง 250 ถึง 1,000 คนต่อกลุ่ม และขนาดตัวอย่างเพิ่มขึ้นจาก 1,000 ถึง 2,000 คนต่อกลุ่ม เมื่อพิจารณาเงื่อนไขที่มีค่าเฉลี่ยแตกต่างกันมากที่สุด ตรงกับเงื่อนไขรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน กับขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ($P = 99.78\%$) และรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน กับขนาดตัวอย่าง 250 คนต่อกลุ่ม ($P = 16.90\%$) ซึ่งแตกต่างกันประมาณ 82.88%

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้องภายใต้วิธีการถดถอยโลจิสติกแบบจัดอันดับ พบว่า ในเงื่อนไขรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และที่ไม่เป็นรูปแบบเดียวกัน เมื่อขนาดตัวอย่างเพิ่มขึ้นในทุกเงื่อนไข มีผลทำให้ค่าเฉลี่ยของอัตราความถูกต้องเพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 เมื่อพิจารณาเงื่อนไขที่มีค่าเฉลี่ยแตกต่างกันมากที่สุด ตรงกับเงื่อนไขรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน กับขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ($P = 99.60\%$) และรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน กับขนาดตัวอย่าง 250 คนต่อกลุ่ม ($P = 56.67\%$) ซึ่งแตกต่างกันประมาณ 42.93%

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้องภายใต้วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ พบว่า ในเงื่อนไขรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และที่ไม่เป็นรูปแบบเดียวกัน เมื่อขนาดตัวอย่างเพิ่มขึ้นในทุกเงื่อนไข มีผลทำให้ค่าเฉลี่ยของอัตราความถูกต้องเพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ยกเว้นเมื่อขนาดตัวอย่างเพิ่มขึ้นจาก 1,000 เป็น 2,000 คนต่อกลุ่ม เมื่อพิจารณาเงื่อนไขที่มีค่าเฉลี่ยแตกต่างกันมากที่สุด ตรงกับเงื่อนไขรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน กับขนาดตัวอย่าง 250 คนต่อกลุ่ม ($P = 45.20\%$) และรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน กับขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ($P = 100\%$) ซึ่งแตกต่างกันประมาณ 54.80%

2) ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้วิธีการถดถอยโลจิสติกแบบจัดอันดับ พบว่า ในเงื่อนไขรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และที่ไม่เป็นรูปแบบเดียวกัน เมื่อขนาดตัวอย่างเพิ่มขึ้นในทุกเงื่อนไข มีผลทำให้ค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 เมื่อพิจารณาเงื่อนไขที่มีค่าเฉลี่ยแตกต่างกันมากที่สุด ตรงกับเงื่อนไขรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน กับขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ($P = 36.29\%$) และรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน กับขนาดตัวอย่าง 250 คนต่อกลุ่ม ($P = 7.56\%$) ซึ่งแตกต่างกันประมาณ 28.73%

3.7.2 ผลการทดสอบวิธีการตรวจสอบ DIF

การทดสอบผลย่อยของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ในแต่ละเงื่อนไขของปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยขนาดตัวอย่าง พบว่า วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้เงื่อนไขปฏิสัมพันธ์ระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยขนาดตัวอย่าง มีผลต่ออัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ยกเว้นภายใต้ปฏิสัมพันธ์ใน 4 เงื่อนไข คือ (1) อัตราความถูกต้องภายใต้รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน กับขนาดตัวอย่าง 500 คนต่อกลุ่ม (2) อัตราความถูกต้องภายใต้รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน กับขนาดตัวอย่าง 1,000 คนต่อกลุ่ม (3) อัตราความถูกต้องภายใต้รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน กับขนาดตัวอย่าง 2,000 คนต่อกลุ่ม และ (4) อัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน กับขนาดตัวอย่าง 250 คนต่อกลุ่ม ดังนั้นจึงนำผลการทดสอบที่มีนัยสำคัญทางสถิติไปเปรียบเทียบรายคู่ โดยใช้วิธีของเชฟเฟ (Scheffé) (แสดงรายละเอียดของผลการวิเคราะห์ดังตาราง 44 ในภาคผนวก ก) ผลการเปรียบเทียบจำแนกออกเป็น 2 ลักษณะ ดังนี้

1) ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน ในเงื่อนไขขนาดตัวอย่าง 250 คนต่อกลุ่ม (FDIF1 × SIZE1) พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับ

หลายมิติมีค่าเฉลี่ยอัตราความถูกต้องไม่แตกต่างกัน ทั้งสองวิธีมีค่าเฉลี่ยอัตราความถูกต้องสูงกว่าวิธีโพลี-ชิปเทสท์ อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ส่วนผลการเปรียบเทียบ ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ในทุกเงื่อนไขของขนาดตัวอย่าง ($FDIF2 \times SIZE1$, $FDIF2 \times SIZE2$, $FDIF2 \times SIZE3$ และ $FDIF2 \times SIZE4$) พบว่า เกือบทุกเงื่อนไข ทั้งสามวิธีมีค่าเฉลี่ยของอัตราความถูกต้องแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001

2) ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และไม่เป็นรูปแบบเดียวกัน ในเงื่อนไขของขนาดตัวอย่าง 7 เงื่อนไข คือ $FDIF1 \times SIZE1$, $FDIF1 \times SIZE2$, $FDIF1 \times SIZE3$, $FDIF1 \times SIZE4$, $FDIF2 \times SIZE2$, $FDIF2 \times SIZE3$ และ $FDIF2 \times SIZE4$ พบว่าทั้งสามวิธีมีค่าเฉลี่ยของอัตราความถูกต้องแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ยกเว้นใน 2 เงื่อนไข คือ (1) รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน กับขนาดตัวอย่าง 250 คนต่อกลุ่ม ($FDIF1 \times SIZE1$) และ (2) รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน กับขนาดตัวอย่าง 500 คนต่อกลุ่ม ($FDIF2 \times SIZE2$)

3.8 การทดสอบผลย่อยในเงื่อนไข 8

ผู้วิจัยนำผลการทดสอบปฏิสัมพันธ์สามทาง ระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ปัจจัยความแตกต่างของการแจกแจงความสามารถ กับวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ [$PDIF \times d(\theta) \times MDIF$] ที่มีนัยสำคัญทางสถิติไปทดสอบผลย่อย (Simple effect) ดังตาราง 45

ตาราง 45 การทดสอบผลย่อยของปฏิสัมพันธ์สองทางระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถ และวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ

แหล่งของความแปรปรวน	<i>F</i>	
	<i>P</i>	<i>E</i> ₁
ปฏิสัมพันธ์สองทางระหว่างปัจจัย		
PDIF กับ $d(\theta)$		
PDIF $\times d(\theta)$ ที่ MDIF1	-	.55
PDIF $\times d(\theta)$ ที่ MDIF2	-	10.33***
PDIF $\times d(\theta)$ ที่ MDIF3	-	1.78
วิธีการตรวจสอบ DIF		
MDIF ที่ PDIF1 $\times d(\theta)$ 1	-	9.76
MDIF ที่ PDIF1 $\times d(\theta)$ 2	-	10.07
MDIF ที่ PDIF1 $\times d(\theta)$ 3	-	678.89***
MDIF ที่ PDIF2 $\times d(\theta)$ 1	-	53.64***
MDIF ที่ PDIF2 $\times d(\theta)$ 2	-	50.48***
MDIF ที่ PDIF2 $\times d(\theta)$ 3	-	819.10***
MDIF ที่ PDIF3 $\times d(\theta)$ 1	-	185.31***
MDIF ที่ PDIF3 $\times d(\theta)$ 2	-	147.98***
MDIF ที่ PDIF3 $\times d(\theta)$ 3	-	1,014.19***

*** $p < .001$

ผลการวิเคราะห์ตามตาราง 45 สรุปได้ดังนี้

3.8.1 ผลการทดสอบปฏิสัมพันธ์สองทางระหว่างปัจจัยสัดส่วนของ DIF กับปัจจัยความแตกต่างของการแจกแจงความสามารถ

การทดสอบผลย่อยของปฏิสัมพันธ์สองทางระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถ ในแต่ละวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ พบว่า ปฏิสัมพันธ์ระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน

กับปัจจัยความแตกต่างของการแจกแจงความสามารถ มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่อวิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ดังนั้นจึงนำผลการทดสอบที่มีนัยสำคัญทางสถิติไปเปรียบเทียบรายคู่ โดยใช้วิธีของเซฟเฟ (Scheffé) (แสดงรายละเอียดของผลการวิเคราะห์ดังตาราง 46 ในภาคผนวก ก) ผลการเปรียบเทียบพบว่า ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10%, 20% และ 30% เมื่อกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถเพิ่มขึ้น 0.5SD (จาก 0.0SD ถึง 0.5SD) ไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับ แต่ถ้ากลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถเพิ่มขึ้น 1.0SD (จาก 0.0SD ถึง 1.0SD) มีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 เมื่อพิจารณาเงื่อนไขที่มีค่าเฉลี่ยแตกต่างกันมากที่สุด ตรงกับเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% กับกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถ 1.0SD ($P = 32.01\%$) และสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% กับกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถ 0.0SD ($P = 5.91\%$) ซึ่งแตกต่างกันประมาณ 26.10%

3.8.2 การทดสอบวิธีการตรวจสอบ DIF

การทดสอบผลย่อยของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ในแต่ละเงื่อนไขของปฏิสัมพันธ์สองทางระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบ พบว่า วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้เงื่อนไขของปฏิสัมพันธ์สองทางในทุกเงื่อนไขระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถ มีผลต่ออัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ยกเว้นอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ปฏิสัมพันธ์ใน 2 เงื่อนไข คือ (1) สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% กับความแตกต่างของการแจกแจงความสามารถ 0.0SD และ (2) สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% กับความแตกต่างของการแจกแจงความสามารถ 0.5SD ดังนั้นจึงนำผลการทดสอบที่มีนัยสำคัญไปเปรียบเทียบรายคู่ โดยใช้วิธีของเซฟเฟ (Scheffé) (แสดงรายละเอียดของผลการวิเคราะห์ดังตาราง 47 ในภาคผนวก ก) ผลการเปรียบเทียบพบว่า ค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ปัจจัยสัดส่วนข้อสอบทำหน้าที่เบี่ยงเบน และความแตกต่างของการแจกแจงความสามารถใน 7 เงื่อนไข คือ PDIF1 ×

$d(\theta)3$, PDIF2 $\times d(\theta)1$, PDIF2 $\times d(\theta)2$, PDIF2 $\times d(\theta)3$, PDIF3 $\times d(\theta)1$, PDIF3 $\times d(\theta)2$, PDIF $\times d(\theta)3$] พบว่า ค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบสามวิธี มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ยกเว้นใน 2 เงื่อนไข คือ (1) สัดส่วนข้อสอบ ทำหน้าที่เบี่ยงเบน 20% กับความแตกต่างของการแจกแจงความสามารถ 0.0SD [PDIF2 $\times d(\theta)1$] และ (2) สัดส่วนข้อสอบทำหน้าที่เบี่ยงเบน 20% กับความแตกต่างของการแจกแจงความสามารถ 0.5SD [PDIF2 $\times d(\theta)2$]

3.9 การทดสอบผลย่อยในเงื่อนไข 9

ผู้วิจัยนำผลการทดสอบปฏิสัมพันธ์สามทาง ระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ปัจจัยขนาดตัวอย่าง กับวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (PDIF $\times$ SIZE $\times$ MDIF) ที่มีนัยสำคัญทางสถิติไปทดสอบผลย่อย (Simple effect) ดังตาราง 48

ตาราง 48 การทดสอบผลย่อยของปฏิสัมพันธ์สองทางระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยขนาดตัวอย่าง และวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ

แหล่งของความแปรปรวน	<i>F</i>	
	<i>P</i>	<i>E</i> ₁
ปฏิสัมพันธ์สองทางระหว่างปัจจัย		
PDIF กับ SIZE		
PDIF $\times$ SIZE ที่ MDIF1	4.84***	.66
PDIF $\times$ SIZE ที่ MDIF2	6.47***	19.29***
PDIF $\times$ SIZE ที่ MDIF3	2.56	.66
วิธีการตรวจสอบ DIF		
MDIF ที่ PDIF1 $\times$ SIZE1	102.90***	3.49
MDIF ที่ PDIF1 $\times$ SIZE2	117.17***	26.08***
MDIF ที่ PDIF1 $\times$ SIZE3	235.94***	108.73***
MDIF ที่ PDIF1 $\times$ SIZE4	267.42***	346.92***

ตาราง 48 (ต่อ)

แหล่งของความแปรปรวน	<i>F</i>	
	<i>P</i>	<i>E</i> ₁
วิธีการตรวจสอบ DIF		
MDIF ที่ PDIF2 × SIZE1	86.97***	1.42
MDIF ที่ PDIF2 × SIZE2	144.46***	56.81***
MDIF ที่ PDIF2 × SIZE3	235.07***	184.58***
MDIF ที่ PDIF2 × SIZE4	205.32***	559.35***
MDIF ที่ PDIF3 × SIZE1	78.47***	6.35
MDIF ที่ PDIF3 × SIZE2	204.06***	93.39***
MDIF ที่ PDIF3 × SIZE3	256.99***	357.88***
MDIF ที่ PDIF3 × SIZE4	248.01***	1,004.08***

****p* < .001

ผลการวิเคราะห์ตามตาราง 48 สรุปได้ดังนี้

3.9.1 ผลการทดสอบปฏิสัมพันธ์สองทางระหว่างปัจจัยสัดส่วนของ DIF กับ ปัจจัยขนาดตัวอย่าง

การทดสอบผลย่อยของปฏิสัมพันธ์สองทางระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่ เบี่ยงเบน กับปัจจัยขนาดตัวอย่าง ในแต่ละวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ พบว่า ปฏิสัมพันธ์ระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยขนาดตัวอย่าง มีผลต่อวิธีโพลิ-ชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญที่ระดับ .001 แต่ไม่มีผลต่อวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ในขณะเดียวกัน ปฏิสัมพันธ์ระหว่าง ปัจจัยดังกล่าว มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัด อันดับอย่างมีนัยสำคัญที่ระดับ .001 แต่ไม่มีผลต่อวิธีโพลิ-ชิปเทสท์ และวิธีการถดถอยโลจิสติก แบบจัดอันดับหลายมิติ ดังนั้นจึงนำผลการทดสอบที่มีนัยสำคัญทางสถิติไปเปรียบเทียบรายคู่ โดยใช้วิธีของเชฟเฟ (Scheffé) (แสดงรายละเอียดของผลการวิเคราะห์ดังตาราง 49 ในภาคผนวก ก) ผลการเปรียบเทียบจำแนกเป็น 2 ลักษณะ ดังนี้

1) การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง

ผลการเปรียบเทียบภายใต้วิธีโพลี-ชิปเทสต์ พบว่า ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่ย่อยแบบ 10%, 20% และ 30% เมื่อขนาดตัวอย่างเพิ่มขึ้นในทุกเงื่อนไข (250, 500, 1,000 และ 2,000 คนต่อกลุ่ม) มีผลทำให้ค่าเฉลี่ยของอัตราความถูกต้องเพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ยกเว้นใน 2 เงื่อนไข คือ (1) สัดส่วนของข้อสอบทำหน้าที่ย่อยแบบ 10% เมื่อขนาดตัวอย่างเพิ่มขึ้นในช่วง 500 ถึง 2,000 คนต่อกลุ่ม และ (2) สัดส่วนของข้อสอบทำหน้าที่ย่อยแบบ 20% และ 30% เมื่อขนาดตัวอย่างเพิ่มจาก 1,000 เป็น 2,000 คนต่อกลุ่ม เมื่อพิจารณาเงื่อนไขที่มีค่าเฉลี่ยแตกต่างกันมากที่สุด ตรงกับเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่ย่อยแบบ 20% กับขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ($P = 64.58\%$) และสัดส่วนของข้อสอบทำหน้าที่ย่อยแบบ 20% กับขนาดตัวอย่าง 250 คนต่อกลุ่ม ($P = 25.29\%$) ซึ่งแตกต่างกันประมาณ 39.29%

ผลการเปรียบเทียบภายใต้วิธีการถดถอยโลจิสติกแบบจัดอันดับ พบว่า ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่ย่อยแบบ 10%, 20% และ 30% เมื่อขนาดตัวอย่างเพิ่มขึ้นในทุกเงื่อนไข มีผลทำให้ค่าเฉลี่ยของอัตราความถูกต้องเพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ยกเว้นเมื่อขนาดตัวอย่างเพิ่มจาก 1,000 เป็น 2,000 คนต่อกลุ่ม เมื่อพิจารณาเงื่อนไขที่มีค่าเฉลี่ยแตกต่างกันมากที่สุด ตรงกับเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่ย่อยแบบ 30% กับขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ($P = 99.33\%$) และสัดส่วนของข้อสอบทำหน้าที่ย่อยแบบ 30% กับขนาดตัวอย่าง 250 คนต่อกลุ่ม ($P = 45.03\%$) ซึ่งแตกต่างกันประมาณ 54.30%

2) ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1

ผลการเปรียบเทียบภายใต้วิธีการถดถอยโลจิสติกแบบจัดอันดับ พบว่า ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่ย่อยแบบ 10%, 20% และ 30% เมื่อขนาดตัวอย่างเพิ่มขึ้นในทุกเงื่อนไข (250, 500, 1,000 และ 2,000 คนต่อกลุ่ม) มีผลทำให้ค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ยกเว้นใน 2 เงื่อนไข คือ (1) สัดส่วนของข้อสอบทำหน้าที่ย่อยแบบ 10% เมื่อขนาดตัวอย่างเพิ่มขึ้นจาก 250 เป็น 500 คนต่อกลุ่ม และ (2) สัดส่วนของข้อสอบทำหน้าที่ย่อยแบบ 20% เมื่อขนาดตัวอย่างเพิ่มขึ้นจาก 250 เป็น 500 คนต่อกลุ่ม เมื่อพิจารณาเงื่อนไขที่มีค่าเฉลี่ยแตกต่างกันมากที่สุด ตรงกับเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่ย่อยแบบ 30% กับขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ($P = 35.49\%$) และสัดส่วนของข้อสอบทำหน้าที่ย่อยแบบ 10% กับขนาดตัวอย่าง 250 คนต่อกลุ่ม ($P = 7.18\%$) ซึ่งแตกต่างกันประมาณ 28.31%

3.9.2 การทดสอบวิธีการตรวจสอบ DIF

การทดสอบผลย่อยของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ในแต่ละเงื่อนไขของปฏิสัมพันธ์สองทางระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยขนาดตัวอย่าง พบว่า วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้เงื่อนไขปฏิสัมพันธ์ระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยขนาดตัวอย่าง มีผลต่ออัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ยกเว้นอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ปฏิสัมพันธ์ใน 3 เงื่อนไข คือ (1) สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% กับขนาดตัวอย่าง 250 คนต่อกลุ่ม (2) สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% กับขนาดตัวอย่าง 250 คนต่อกลุ่ม และ (3) สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% กับขนาดตัวอย่าง 250 คนต่อกลุ่ม ดังนั้นจึงนำผลการทดสอบที่มีนัยสำคัญไปเปรียบเทียบรายคู่ โดยใช้วิธีของเชฟเฟ (Scheffé) (แสดงรายละเอียดของผลการวิเคราะห์ดังตาราง 50 ในภาคผนวก ก) ผลการเปรียบเทียบจำแนกออกเป็น 2 ลักษณะ ดังนี้

1) ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง ภายใต้ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% ในเงื่อนไขขนาดตัวอย่าง 250, 500, 1,000 และ 2,000 คนต่อกลุ่ม ($PDIF1 \times SIZE1$, $PDIF1 \times SIZE2$, $PDIF1 \times SIZE3$ และ $PDIF1 \times SIZE4$) พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยของอัตราความถูกต้องไม่แตกต่างกัน ทั้งสองวิธีมีค่าเฉลี่ยของอัตราความถูกต้องสูงกว่าวิธีโพลี-ชิปทดสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ในทุกเงื่อนไข

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง ภายใต้ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% ในเงื่อนไขขนาดตัวอย่าง 250, 500, 1,000 และ 2,000 คนต่อกลุ่ม ($PDIF2 \times SIZE1$, $PDIF2 \times SIZE2$, $PDIF2 \times SIZE3$ และ $PDIF2 \times SIZE4$) พบว่าให้ผลสอดคล้องกับภายใต้ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10%

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง ภายใต้ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% ในเงื่อนไขขนาดตัวอย่าง 250, 1,000 และ 2,000 คนต่อกลุ่ม ($PDIF3 \times SIZE1$, $PDIF3 \times SIZE3$ และ $PDIF3 \times SIZE4$) พบว่าให้ผลสอดคล้องกับภายใต้ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% ส่วนในเงื่อนไขขนาดตัวอย่าง 500 คนต่อกลุ่ม ($PDIF3 \times SIZE2$) ทั้งสามวิธีมีค่าเฉลี่ยของอัตราความถูกต้องแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001

2) ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ปัจจัยสัดส่วนข้อสอบทำหน้าที่เปี่ยงเบน และขนาดตัวอย่างใน 9 เงื่อนไข คือ PDIF1 $\times$ SIZE2, PDIF1 $\times$ SIZE3, PDIF1 $\times$ SIZE4, PDIF2 $\times$ SIZE2, PDIF2 $\times$ SIZE3, PDIF2 $\times$ SIZE4, PDIF3 $\times$ SIZE2, PDIF3 $\times$ SIZE3 และ PDIF3 $\times$ SIZE4 พบว่า ทั้งสามวิธีมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ยกเว้นในเงื่อนไขสัดส่วนข้อสอบทำหน้าที่เปี่ยงเบน 10% และขนาดตัวอย่าง 500 คนต่อกลุ่ม (PDIF1 $\times$ SIZE2)

3.10 การทดสอบผลย่อยในเงื่อนไข 10

ผู้วิจัยนำผลการทดสอบปฏิสัมพันธ์สามทาง ระหว่างปัจจัยความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบ ปัจจัยขนาดตัวอย่าง กับวิธีการตรวจสอบการทำหน้าที่เปี่ยงเบนของข้อสอบ [$d(\theta) \times \text{SIZE} \times \text{MDIF}$] ที่มีนัยสำคัญทางสถิติไปทดสอบผลย่อย (Simple effect) ดังตาราง 51

ตาราง 51 การทดสอบผลย่อยของปฏิสัมพันธ์สองทางระหว่างปัจจัยความแตกต่างของการแจกแจงความสามารถ กับปัจจัยขนาดตัวอย่าง และวิธีการตรวจสอบการทำหน้าที่เปี่ยงเบนของข้อสอบ

แหล่งของความแปรปรวน	<i>F</i>	
	<i>P</i>	<i>E</i> ₁
ปฏิสัมพันธ์สองทางระหว่างปัจจัย		
<i>d</i>(θ) กับ SIZE		
<i>d</i> (θ) $\times$ SIZE ที่ MDIF1	4.61***	15.66***
<i>d</i> (θ) $\times$ SIZE ที่ MDIF2	4.21***	443.62***
<i>d</i> (θ) $\times$ SIZE ที่ MDIF3	2.15	31.52***
วิธีการตรวจสอบ DIF		
MDIF ที่ <i>d</i> (θ) 1 $\times$ SIZE1	153.64***	7.22
MDIF ที่ <i>d</i> (θ) 1 $\times$ SIZE2	241.58***	50.44***

ตาราง 51 (ต่อ)

แหล่งของความแปรปรวน	<i>F</i>	
	<i>P</i>	<i>E</i> ₁
วิธีการตรวจสอบ DIF		
MDIF ที่ $d(\theta)$ 1 × SIZE3	382.08***	74.76***
MDIF ที่ $d(\theta)$ 1 × SIZE4	412.76***	259.74***
MDIF ที่ $d(\theta)$ 2 × SIZE1	112.71***	1.77
MDIF ที่ $d(\theta)$ 2 × SIZE2	170.67***	30.53***
MDIF ที่ $d(\theta)$ 2 × SIZE3	314.36***	111.90***
MDIF ที่ $d(\theta)$ 2 × SIZE4	277.16***	353.93***
MDIF ที่ $d(\theta)$ 3 × SIZE1	35.69***	33.46***
MDIF ที่ $d(\theta)$ 3 × SIZE2	71.89***	358.34***
MDIF ที่ $d(\theta)$ 3 × SIZE3	106.38***	1,706.91***
MDIF ที่ $d(\theta)$ 3 × SIZE4	100.86***	4,561.82***

*** $p < .001$

ผลการวิเคราะห์ตามตาราง 51 สรุปได้ดังนี้

3.10.1 ผลการทดสอบปฏิสัมพันธ์สองทางระหว่างความแตกต่างของการแจกแจงความสามารถ กับปัจจัยขนาดตัวอย่าง

การทดสอบผลย่อยของปฏิสัมพันธ์สองทางระหว่างปัจจัยความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบ กับปัจจัยขนาดตัวอย่าง ในแต่ละวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ พบว่า ปฏิสัมพันธ์สองทางระหว่างปัจจัยความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบ กับปัจจัยขนาดตัวอย่าง มีผลต่ออัตราความถูกต้องของวิธีโพลีชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่อวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ในขณะเดียวกัน ปฏิสัมพันธ์ระหว่างปัจจัยดังกล่าว มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งสามวิธีอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ดังนั้นจึงนำผลการทดสอบที่มีนัยสำคัญไปเปรียบเทียบความแตกต่างรายคู่โดยใช้วิธีของเชฟเฟ (Scheffé) (แสดงรายละเอียดของผลการวิเคราะห์ดังตาราง 52 ในภาคผนวก ก) ผลการเปรียบเทียบจำแนกเป็น 2 ลักษณะ ดังนี้

1) การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง

ผลการเปรียบเทียบภายใต้วิธีโพลี-ชิปเทสต์ พบว่า ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD, 0.5SD และ 1.0SD เมื่อขนาดตัวอย่างเพิ่มขึ้นในทุกเงื่อนไข (250, 500, 1,000 และ 2,000 คนต่อกลุ่ม) มีผลทำให้ค่าเฉลี่ยของอัตราความถูกต้องเพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ยกเว้นใน 2 เงื่อนไข คือ (1) ความแตกต่างของการแจกแจงความสามารถ 0.0SD, 0.5SD และ 1.0SD เมื่อขนาดตัวอย่างเพิ่มขึ้นจาก 1,000 เป็น 2,000 คนต่อกลุ่ม และ (2) ความแตกต่างของการแจกแจงความสามารถ 0.5SD เมื่อขนาดตัวอย่างเพิ่มขึ้นจาก 500 เป็น 1,000 คนต่อกลุ่ม เมื่อพิจารณาเงื่อนไขที่มีค่าเฉลี่ยแตกต่างกันมากที่สุด ตรงกับเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.5SD กับขนาดตัวอย่าง 250 คนต่อกลุ่ม ($P = 74.79\%$) และความแตกต่างของการแจกแจงความสามารถ 0.5SD กับขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ($P = 21.07\%$) ซึ่งแตกต่างกันประมาณ 53.72%

ส่วนผลการเปรียบเทียบภายใต้วิธีการถดถอยโลจิสติกแบบจัดอันดับ พบว่า ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD, 0.5SD และ 1.0SD เมื่อขนาดตัวอย่างเพิ่มขึ้นในทุกเงื่อนไขมีผลทำให้ค่าเฉลี่ยของอัตราความถูกต้องเพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ยกเว้นในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD และ 0.5SD เมื่อขนาดตัวอย่างเพิ่มขึ้นจาก 1,000 เป็น 2,000 คนต่อกลุ่ม เมื่อพิจารณาเงื่อนไขที่มีค่าเฉลี่ยแตกต่างกันมากที่สุด ตรงกับเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD กับขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ($P = 99.24\%$) และความแตกต่างของการแจกแจงความสามารถ 0.0SD กับขนาดตัวอย่าง 250 คนต่อกลุ่ม ($P = 48.53\%$) ซึ่งแตกต่างกันประมาณ 50.71%

2) ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1

ภายใต้วิธีโพลี-ชิปเทสต์ ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD ขนาดตัวอย่างทุกระดับไม่มีผลต่อค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ส่วนในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.5SD และ 1.0SD เมื่อขนาดตัวอย่างเพิ่มขึ้น 3 ระดับ คือ จาก 250 เป็น 500, 1,000 และ 2,000 คนต่อกลุ่ม มีผลทำให้ค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ลดลงอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 เมื่อพิจารณาเงื่อนไขที่มีค่าเฉลี่ยแตกต่างกันมากที่สุด ตรงกับเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 1.0SD กับขนาดตัวอย่าง 250 คนต่อกลุ่ม ($P = 11.55\%$) และความแตกต่างของการแจกแจงความสามารถ 0.5SD กับขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ($P = 2.60\%$) ซึ่งแตกต่างกันประมาณ 8.95%

ภายใต้วิธีการถดถอยโลจิสติกแบบจัดอันดับ พบว่า ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD, 0.5SD และ 1.0SD เมื่อขนาดตัวอย่างเพิ่มขึ้นใน ช่วง 500 ถึง 2,000 คนต่อกลุ่ม มีผลทำให้ค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ยกเว้นใน 2 เงื่อนไข คือ (1) ความแตกต่างของการแจกแจงความสามารถ 0.0SD เมื่อขนาดตัวอย่างเพิ่มขึ้นจาก 500 เป็น 1,000 คนต่อกลุ่ม และ (2) ความแตกต่างของการแจกแจงความสามารถ 0.5SD เมื่อขนาดตัวอย่างเพิ่มขึ้นจาก 250 เป็น 500 คนต่อกลุ่ม เมื่อพิจารณาเงื่อนไขที่มีค่าเฉลี่ยแตกต่างกันมากที่สุด ตรงกับเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 1.0SD กับขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ($P = 54.06\%$) และความแตกต่างของการแจกแจงความสามารถ 0.5SD กับขนาดตัวอย่าง 250 คนต่อกลุ่ม ($P = 5.98\%$) ซึ่งแตกต่างกันประมาณ 48.08%

ภายใต้วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ พบว่า ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD และ 0.5SD ขนาดตัวอย่างทุกระดับไม่มีผลต่อค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ส่วนในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 1.0SD เมื่อขนาดตัวอย่างเพิ่มขึ้นในช่วง 250 ถึง 2,000 คนต่อกลุ่ม มีผลต่อทำให้ค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้น ยกเว้นใน 2 เงื่อนไข คือ (1) เมื่อขนาดตัวอย่างเพิ่มขึ้นจาก 250 เป็น 500 คนต่อกลุ่ม และ (2) เมื่อขนาดตัวอย่างเพิ่มขึ้นจาก 500 เป็น 1,000 คนต่อกลุ่ม เมื่อพิจารณาเงื่อนไขที่มีค่าเฉลี่ยแตกต่างกันมากที่สุด ตรงกับเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 1.0SD กับขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ($P = 18.54\%$) และความแตกต่างของการแจกแจงความสามารถ 0.0SD กับขนาดตัวอย่าง 250 คนต่อกลุ่ม ($P = 6.37\%$) ซึ่งแตกต่างกันประมาณ 12.17%

3.10.2 ผลการทดสอบวิธีการตรวจสอบ DIF

การทดสอบผลย่อยของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ในแต่ละเงื่อนไขของปฏิสัมพันธ์สองทางระหว่างปัจจัยความแตกต่างของการแจกแจงความสามารถ กับปัจจัยขนาดตัวอย่าง พบว่า วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้เงื่อนไขปฏิสัมพันธ์ระหว่างปัจจัยความแตกต่างของการแจกแจงความสามารถ กับปัจจัยขนาดตัวอย่าง มีผลต่ออัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 อย่างมีนัยสำคัญที่ระดับ .001 ยกเว้นอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ปฏิสัมพันธ์ใน 2 เงื่อนไข คือ (1) ความแตกต่างของการแจกแจงความสามารถ 0.0SD กับขนาดตัวอย่าง 250 คนต่อกลุ่ม และ (2) ความแตกต่าง

ของการแจกแจงความสามารถ 0.5SD กับขนาดตัวอย่าง 250 คนต่อกลุ่ม ดังนั้นจึงนำผลการทดสอบที่มีนัยสำคัญไปเปรียบเทียบรายคู่ โดยใช้วิธีของเชฟเฟ (Scheffé) (แสดงรายละเอียดของผลการวิเคราะห์ดังตาราง 53 ในภาคผนวก ก) ผลการเปรียบเทียบจำแนกเป็น 2 ลักษณะ ดังนี้

1) ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง

ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง ภายใต้ปัจจัยความแตกต่างของการแจกแจงความสามารถ 0.0SD, 0.5SD และ 1.0SD ในเงื่อนไขขนาดตัวอย่าง 250, 500, 1,000 และ 2,000 คนต่อกลุ่ม [$d(\theta) 1 \times \text{SIZE1}$, $d(\theta) 1 \times \text{SIZE2}$, $d(\theta) 1 \times \text{SIZE3}$, $d(\theta) 1 \times \text{SIZE4}$, $d(\theta) 2 \times \text{SIZE1}$, $d(\theta) 2 \times \text{SIZE2}$, $d(\theta) 2 \times \text{SIZE3}$, $d(\theta) 2 \times \text{SIZE4}$, $d(\theta) 3 \times \text{SIZE1}$, $d(\theta) 3 \times \text{SIZE2}$, $d(\theta) 3 \times \text{SIZE3}$ และ $d(\theta) 3 \times \text{SIZE4}$] พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยของอัตราความถูกต้องไม่แตกต่างกัน ทั้งสองวิธีมีค่าเฉลี่ยของอัตราความถูกต้องสูงกว่าวิธีโพลี-ชิปเทสอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ยกเว้นในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 1.0SD และขนาดตัวอย่าง 250 คนต่อกลุ่ม [$d(\theta) 3 \times \text{SIZE1}$] กับความแตกต่างของการแจกแจงความสามารถ 1.0SD และขนาดตัวอย่าง 1,000 คนต่อกลุ่ม [$d(\theta) 3 \times \text{SIZE3}$]

2) ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1

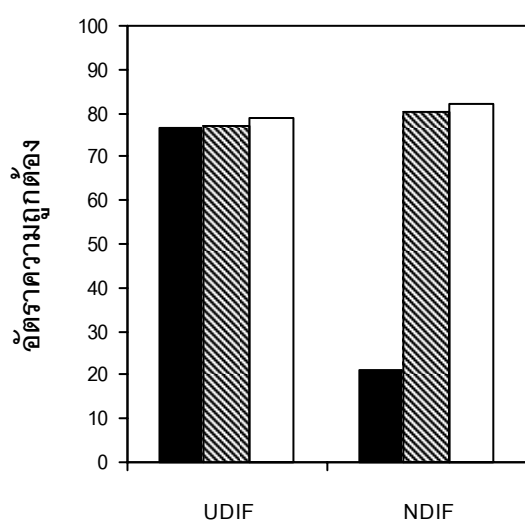
ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ปัจจัยความแตกต่างของการแจกแจงความสามารถ 0.0SD, 0.5SD และ 1.0SD ในเงื่อนไขขนาดตัวอย่าง 250, 500, 1,000 และ 2,000 คนต่อกลุ่ม จำนวน 10 เงื่อนไข คือ $d(\theta) 1 \times \text{SIZE2}$, $d(\theta) 1 \times \text{SIZE3}$, $d(\theta) 1 \times \text{SIZE4}$, $d(\theta) 2 \times \text{SIZE2}$, $d(\theta) 2 \times \text{SIZE3}$, $d(\theta) 2 \times \text{SIZE4}$, $d(\theta) 3 \times \text{SIZE1}$, $d(\theta) 3 \times \text{SIZE2}$, $d(\theta) 3 \times \text{SIZE3}$ และ $d(\theta) 3 \times \text{SIZE4}$] พบว่า ทั้งสามวิธีมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ยกเว้นใน 2 เงื่อนไข คือ (1) ความแตกต่างของการแจกแจงความสามารถ 0.5SD และขนาดตัวอย่าง 500 คนต่อกลุ่ม [$d(\theta) 2 \times \text{SIZE2}$] กับ (2) ความแตกต่างของการแจกแจงความสามารถ 1.0D และขนาดตัวอย่าง 250 คนต่อกลุ่ม [$d(\theta) 3 \times \text{SIZE1}$]

4. สรุปผลการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1

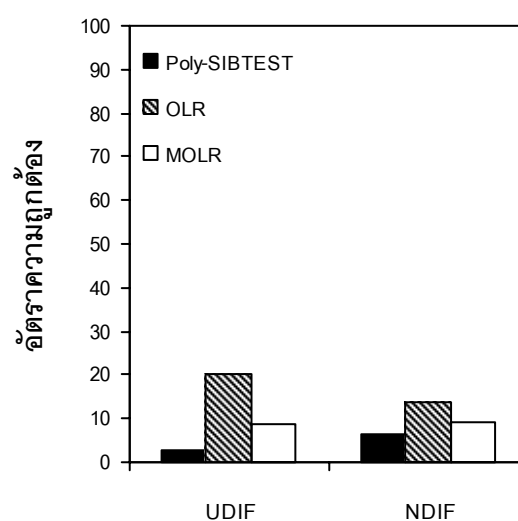
ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า โดยใช้วิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย สรุปได้ดังนี้

4.1 ผลการคำนวณค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัยหลัก

ผลการคำนวณค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ปัจจัยหลัก 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่าง แสดงดังภาพประกอบ 22-24



ก. อัตราความถูกต้อง



ข. อัตราความคลาดเคลื่อนประเภทที่ 1

ภาพประกอบ 22 แสดงอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1

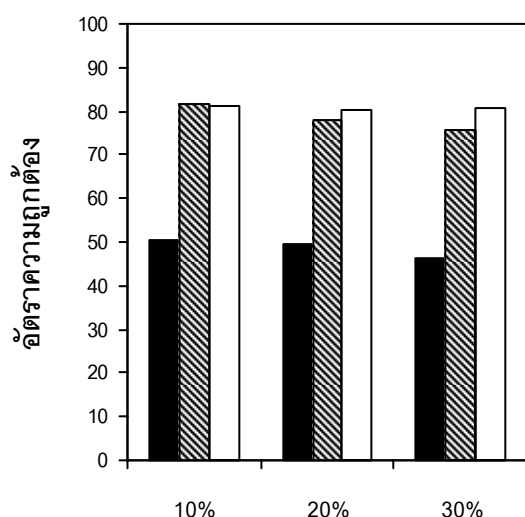
ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน 2 รูปแบบ

ภาพประกอบ 22 ก และ 22 ข แสดงอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ตามลำดับ ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน 2 รูปแบบ คือ ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (UDIF) และข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (NDIF) สรุปได้ดังนี้

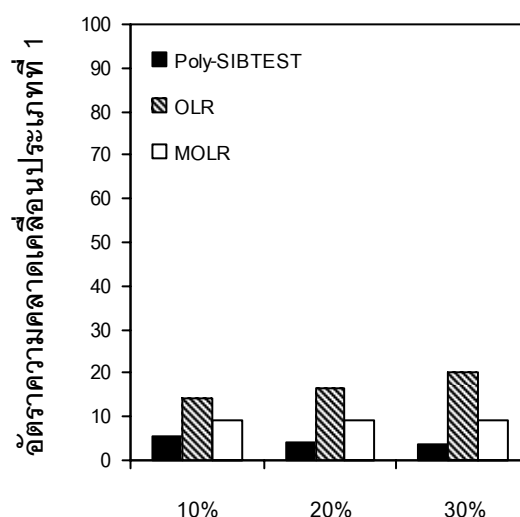
ผลการตรวจสอบโดยใช้วิธีโพลี-ซิป/เทสต์ พบว่า ภายใต้ปัจจัยข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน มีค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณ 76.65% และ 2.58% ตามลำดับ ส่วนภายใต้ปัจจัยข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน มีค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณ 20.97% และ 6.34% ตามลำดับ

ผลการตรวจสอบโดยใช้วิธีการถดถอยโลจิสติกแบบจัดอันดับ พบว่า ภายใต้ปัจจัยข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน มีค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณ 76.96% และ 20.09% ตามลำดับ ส่วนภายใต้ปัจจัยข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน มีค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณ 80.09% และ 13.64% ตามลำดับ

ผลการตรวจสอบโดยใช้วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ พบว่า ภายใต้ปัจจัยข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน มีค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณ 79.00% และ 8.81% ตามลำดับ ส่วนภายใต้ปัจจัยข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน มีค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณ 82.30% และ 9.25% ตามลำดับ



ก. อัตราความถูกต้อง



ข. อัตราความคลาดเคลื่อนประเภทที่ 1

ภาพประกอบ 23 แสดงอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1

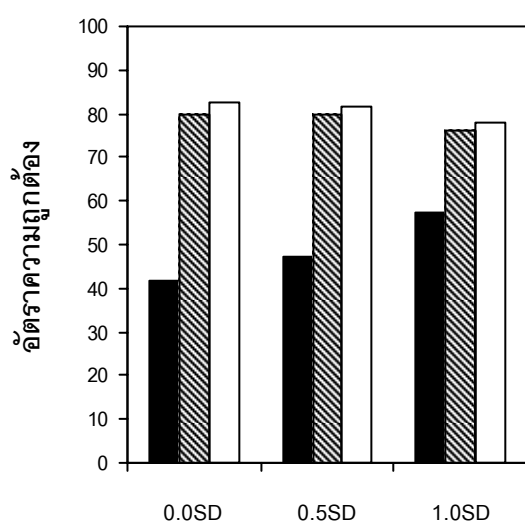
ภายใต้ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 3 ขนาด

ภาพประกอบ 23 ก และ 23 ข แสดงอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ตามลำดับ ภายใต้ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 3 ขนาด คือ 10%, 20% และ 30% สรุปได้ดังนี้

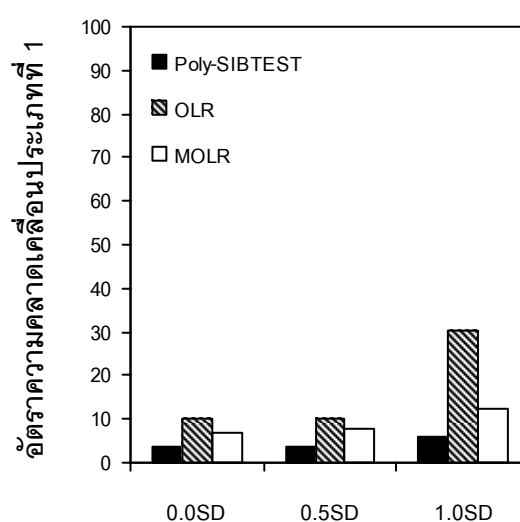
ผลการตรวจสอบโดยใช้วิธีโพลี-ซิป/เทสต์ พบว่า ค่าเฉลี่ยของอัตราความถูกต้องมีค่าอยู่ในช่วง 46.40%-50.58% และค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าอยู่ในช่วง 3.84%-5.32% เมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้นในช่วง 10% ถึง 30% ค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ลดลงในทุกเงื่อนไข

ผลการตรวจสอบโดยใช้วิธีการถดถอยโลจิสติกแบบจัดอันดับ พบว่า ค่าเฉลี่ยของอัตราความถูกต้องมีค่าอยู่ในช่วง 75.90%-81.58% และค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าอยู่ในช่วง 14.00%-20.20% เมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้นในช่วง 10% ถึง 30% ค่าเฉลี่ยของอัตราความถูกต้องลดลงในทุกเงื่อนไข และค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นในทุกเงื่อนไข

ผลการตรวจสอบโดยใช้วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ พบว่า ค่าเฉลี่ยของอัตราความถูกต้องมีค่าอยู่ในช่วง 80.41%-80.98% และค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าอยู่ในช่วง 8.98%-9.07% เมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้นในช่วง 10% ถึง 30% ค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ค่อนข้างคงที่



ก. อัตราความถูกต้อง



ข. อัตราความคลาดเคลื่อนประเภทที่ 1

ภาพประกอบ 24 แสดงอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1

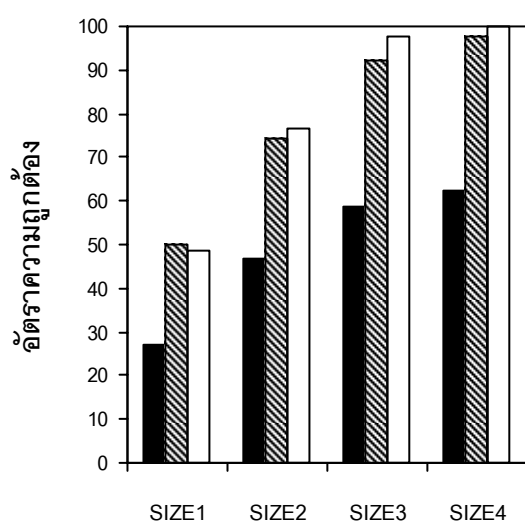
ภายใต้ปัจจัยความแตกต่างของการแจกแจงความสามารถ 3 ระดับ

ภาพประกอบ 24 ก และ 24 ข แสดงอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ตามลำดับ ภายใต้ปัจจัยความแตกต่างของการแจกแจงความสามารถ 3 ระดับ คือ 0.0SD, 0.5SD และ 1.0SD สรุปได้ดังนี้

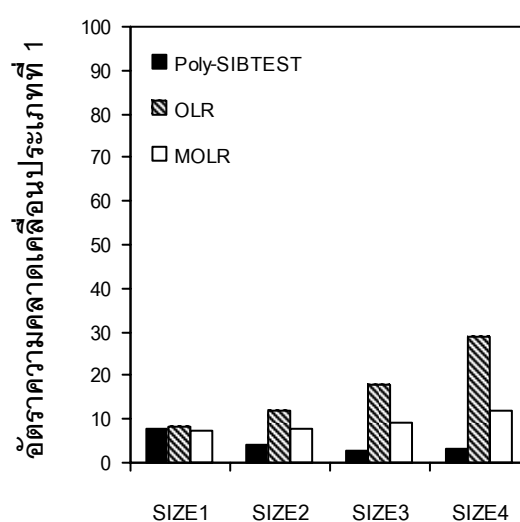
ผลการตรวจสอบโดยใช้วิธีโพลี-ซิป/เทสต์ พบว่า ค่าเฉลี่ยของอัตราความถูกต้องมีค่าอยู่ในช่วง 41.83%-57.45% และค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าอยู่ในช่วง 3.63%-5.96% เมื่อความแตกต่างของการแจกแจงความสามารถเพิ่มขึ้นในช่วง 0.0SD ถึง 1.0SD ค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นในทุกเงื่อนไข

ผลการตรวจสอบโดยใช้วิธีการถดถอยโลจิสติกแบบจัดอันดับ พบว่า ค่าเฉลี่ยของอัตราความถูกต้องมีค่าอยู่ในช่วง 79.72%-76.08% และค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าอยู่ในช่วง 10.06%-30.22% เมื่อความแตกต่างของการแจกแจงความสามารถเพิ่มขึ้นในช่วง 0.0SD ถึง 1.0SD ค่าเฉลี่ยของอัตราความถูกต้องลดลงเกือบทุกเงื่อนไข และค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นในทุกเงื่อนไข

ผลการตรวจสอบโดยใช้วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ พบว่า ค่าเฉลี่ยของอัตราความถูกต้องมีค่าอยู่ในช่วง 77.77%-82.44% และค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าอยู่ในช่วง 7.03%-12.33% เมื่อความแตกต่างของการแจกแจงความสามารถเพิ่มขึ้นในช่วง 0.0SD ถึง 1.0SD ค่าเฉลี่ยของอัตราความถูกต้องลดลงในทุกเงื่อนไข และค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นในทุกเงื่อนไข



ก. อัตราความถูกต้อง



ข. อัตราความคลาดเคลื่อนประเภทที่ 1

ภาพประกอบ 25 แสดงอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1

ภายใต้ปัจจัยขนาดตัวอย่าง 4 ขนาด

ภาพประกอบ 25 ก และ 25 ข แสดงอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ตามลำดับ ภายใต้ปัจจัยขนาดตัวอย่าง 4 ขนาด คือ 250, 500, 1,000 และ 2,000 คน ในแต่ละกลุ่ม (กลุ่มอ้างอิง และกลุ่มเปรียบเทียบ) สรุปได้ดังนี้

ผลการตรวจสอบโดยใช้วิธีโพลี-ชิปเทสต์ พบว่า ค่าเฉลี่ยของอัตราความถูกต้องมีค่าอยู่ในช่วง 27.21%-62.53 และค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าอยู่ในช่วง 2.96%-7.60% เมื่อขนาดตัวอย่างเพิ่มขึ้น ค่าเฉลี่ยของอัตราความถูกต้องเพิ่มขึ้นในทุกเงื่อนไข และค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ลดลงเกือบทุกเงื่อนไข

ผลการตรวจสอบโดยใช้วิธีการถดถอยโลจิสติกแบบจัดอันดับ พบว่า ค่าเฉลี่ยของอัตราความถูกต้องมีค่าอยู่ในช่วง 49.86%-97.76 และค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าอยู่ในช่วง 8.36%-28.87% เมื่อขนาดตัวอย่างเพิ่มขึ้น ค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นในทุกเงื่อนไข

ผลการตรวจสอบโดยใช้วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ พบว่า ค่าเฉลี่ยของอัตราความถูกต้องมีค่าอยู่ในช่วง 48.60%-99.84 และค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าอยู่ในช่วง 7.47%-11.77% เมื่อขนาดตัวอย่างเพิ่มขึ้น ค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นในทุกเงื่อนไข

4.2 ผลการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ปัจจัยที่ศึกษา

ในตอนนี้ ผู้วิจัยนำผลการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ทุกเงื่อนไขของการตรวจสอบมาสรุปในภาพรวม โดยสรุปในสองประเด็นตามความมุ่งหมายของการวิจัย คือ (1) การทดสอบปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ที่มีผลต่ออัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ และ (2) การเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ผลการเปรียบเทียบตามในสองประเด็นดังกล่าวมีรายละเอียดดังนี้

4.2.1 ผลการทดสอบปัจจัยที่ศึกษาที่มีผลต่ออัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบการทำหน้าที่เป็ยงเบนของข้อสอบ

ผลการทดสอบปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ที่มีผลต่ออัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ โดยการทดสอบผลย่อย (Simple effect) สรุปได้ดังตาราง 54

ตาราง 54 สรุปปัจจัยที่มีผลต่ออัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ

ปัจจัย	วิธีการตรวจสอบ DIF					
	Poly-SIBTEST		OLR		MOLR	
	<i>P</i>	<i>E</i> ₁	<i>P</i>	<i>E</i> ₁	<i>P</i>	<i>E</i> ₁
FDIF	s	s	ns	s	ns	ns
PDIF	ns	ns	s	s	ns	ns
<i>d</i> (θ)	s	s	ns	s	ns	s
SIZE	s	s	s	s	s	s
FDIF × PDIF	s	ns	s	s	ns	ns
FDIF × <i>d</i> (θ)	s	ns	s	s	ns	ns
FDIF × SIZE	s	ns	s	s	s	ns
PDIF × <i>d</i> (θ)	-	ns	-	s	-	ns
PDIF × SIZE	s	ns	s	s	ns	ns
<i>d</i> (θ) × SIZE	s	s	s	s	ns	s

s = มีนัยสำคัญ, ns = ไม่มีนัยสำคัญ

การทดสอบผลย่อยตามตาราง 54 สรุปได้ดังนี้

ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถ ขนาดตัวอย่าง ปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย มีผลต่ออัตราความถูกต้องของวิธีโพลี-ชิป/เทสท์อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ยกเว้นปฏิสัมพันธ์ระหว่างสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน กับความแตกต่างของการแจกแจงความสามารถ ในขณะเดียวกัน ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถ ขนาดตัวอย่าง ปฏิสัมพันธ์ระหว่างความแตกต่างของการแจกแจงความสามารถ กับขนาดตัวอย่างมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิป/เทสท์อย่างมีนัยสำคัญทางสถิติที่ระดับ .001

ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ขนาดตัวอย่าง ปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย มีผลต่ออัตราความถูกต้องของวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ยกเว้นปฏิสัมพันธ์ระหว่างสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน กับความแตกต่างของการแจกแจงความสามารถ ในขณะเดียวกัน ทุกปัจจัยที่ศึกษามีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .001

ปัจจัยขนาดตัวอย่าง ปฏิสัมพันธ์ระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนและขนาดตัวอย่าง มีผลต่ออัตราความถูกต้องของวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ในขณะเดียวกัน ปัจจัยความแตกต่างของการแจกแจงความสามารถ ขนาดตัวอย่าง ปฏิสัมพันธ์ระหว่างความแตกต่างของการแจกแจงความสามารถ กับขนาดตัวอย่าง มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติอย่างมีนัยสำคัญทางสถิติที่ระดับ .001

4.2.2 ผลการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยที่ศึกษา

ผลการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย โดยใช้วิธีของเชฟเฟ (Scheffé) สรุปได้ดังตาราง 55 และ 56

ตาราง 55 สรุปผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง ระหว่างวิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ภายใต้ทุกเงื่อนไขของปัจจัยที่ศึกษา

การเปรียบเทียบ	วิธีการตรวจสอบ DIF			ผลการทดสอบนัยสำคัญ		
	Poly-SIB	OLR	MOLR	1	2	3
1	1	-1	0	ns	s	s
2	1	0	-1	s	s	s
3	0	1	-1	s	ns	s
จำนวนเงื่อนไข				2	37	9

Poly-SIB = Poly-SIBTEST

1, -1, 0 = สัมประสิทธิ์ของค่าเฉลี่ย; s = มีนัยสำคัญ, ns = ไม่มีนัยสำคัญ

ผลสรุปตามตาราง 55 แสดงการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง ระหว่างวิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ภายใต้เงื่อนไขของปัจจัยที่ศึกษาจำนวน 48 เงื่อนไข (การทดสอบผลย่อยที่มีนัยสำคัญ) สามารถจำแนกผลการทดสอบที่สอดคล้องกันเป็น 3 กลุ่ม ดังนี้

กลุ่ม 1 การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้องให้ผลสอดคล้องกัน จำนวน 2 เงื่อนไข กล่าวคือ วิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยของอัตราความถูกต้องไม่แตกต่างกัน ส่วนคู่อื่นๆ ที่เหลือมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001

กลุ่ม 2 การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้องให้ผลสอดคล้องกัน จำนวน 37 เงื่อนไข กล่าวคือ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยของอัตราความถูกต้องไม่แตกต่างกัน ส่วนคู่อื่นๆ ที่เหลือมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001

กลุ่ม 3 การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้องให้ผลสอดคล้องกัน จำนวน 9 เงื่อนไข กล่าวคือ วิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ มีค่าเฉลี่ยของอัตราความถูกต้องแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ในทุกคู่ของการเปรียบเทียบ

ผลการทดสอบใน 3 กลุ่ม แสดงว่า การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง ภายใต้ปัจจัยที่ศึกษาให้ผลการทดสอบสอดคล้องกันมากที่สุด จำนวน 37 เงื่อนไข ดังนั้นสามารถสรุปในภาพรวมได้ว่า ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ได้แก่ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่าง รวมทั้งปฏิสัมพันธ์สองทางระหว่าง 4 ปัจจัยดังกล่าว **วิธีการถดถอยโลจิสติกแบบจัดอันดับและวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องใกล้เคียงกัน ทั้งสองวิธีมีอัตราความถูกต้องสูงกว่าวิธีโพลี-ชิปเทสท์อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ภายใต้เงื่อนไขส่วนมาก**

ตาราง 56 สรุปผลการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ภายใต้ทุกเงื่อนไขของปัจจัยที่ศึกษา

การเปรียบเทียบ	วิธีการตรวจสอบ DIF			ผลการทดสอบนัยสำคัญ				
	Poly-SIB	OLR	MOLR	1	2	3	4	5
1	1	-1	0	s	ns	s	s	s
2	1	0	-1	ns	s	ns	s	s
3	0	1	-1	ns	s	s	ns	s
จำนวนเงื่อนไข				1	1	1	4	47

Poly-SIB = Poly-SIBTEST

1, -1, 0 = สัมประสิทธิ์ของค่าเฉลี่ย; s = มีนัยสำคัญ, ns = ไม่มีนัยสำคัญ

ผลสรุปตามตาราง 56 แสดงการเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ภายใต้เงื่อนไขของปัจจัยที่ศึกษาจำนวน 54 เงื่อนไข (การทดสอบผลย่อยที่มีนัยสำคัญ) สามารถจำแนกผลการทดสอบที่สอดคล้องกันเป็น 5 กลุ่ม ดังนี้

กลุ่ม 1 การเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 จำนวน 1 เงื่อนไข พบว่า วิธีโพลี-ชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ส่วนคู่อื่นๆ ที่เหลือไม่แตกต่างกัน

กลุ่ม 2 การเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 จำนวน 1 เงื่อนไข พบว่า วิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่แตกต่างกัน ส่วนคู่อื่นๆ ที่เหลือมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001

กลุ่ม 3 การเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 จำนวน 1 เงื่อนไข พบว่า วิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่แตกต่างกัน ส่วนคู่อื่นๆ ที่เหลือมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001

กลุ่ม 4 การเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ให้ผลสอดคล้องกันจำนวน 4 เงื่อนไข กล่าวคือ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่แตกต่างกัน ส่วนคู่อื่นๆ ที่เหลือมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001

กลุ่ม 5 การเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ให้ผลสอดคล้องกันจำนวน 47 เงื่อนไข กล่าวคือ วิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ในทุกคู่ของการเปรียบเทียบ

ผลการทดสอบใน 5 กลุ่ม แสดงว่า การเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ปัจจัยที่ศึกษา ให้ผลการทดสอบสอดคล้องกันมากที่สุด จำนวน 47 เงื่อนไข ดังนั้นสามารถสรุปในภาพรวมได้ว่า ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ได้แก่ รูปแบบของข้อสอบ ทำหน้าที่เบี่ยงเบน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความ-สามารถ และขนาดตัวอย่าง รวมทั้งปฏิสัมพันธ์สองทางระหว่าง 4 ปัจจัยดังกล่าว **อัตราความ-คลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบสามวิธีมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ภายใต้เกือบทุกเงื่อนไข**

เมื่อพิจารณาอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ทุกเงื่อนไข พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ และวิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องสูงใกล้เคียงกัน แต่วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ เมื่อนำอัตราความคลาดเคลื่อนประเภทที่ 1 ไปเปรียบเทียบกับเกณฑ์ ($\alpha = .05$) พบว่า วิธีการถดถอยโลจิสติก

แบบจัดอันดับหลายมิติมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าขอบเขตที่กำหนด (Nominal limits) ประมาณ 4% ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าขอบเขตที่กำหนดประมาณ 12% นั้นแสดงว่า **ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีประสิทธิภาพสูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ**

ส่วนวิธีโพลี-ชิปเทสท์สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้มีค่าอยู่ภายในขอบเขตที่กำหนดเกือบทุกเงื่อนไขของการตรวจสอบ (ประมาณ 4%) แต่เมื่อพิจารณาอัตราความถูกต้องพบว่า มีค่าต่ำกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ผลดังกล่าวชี้ให้เห็นว่าวิธีโพลี-ชิปเทสท์มีข้อได้เปรียบที่มีความผิดพลาดต่ำในการตรวจสอบ แต่มีข้อเสียเปรียบที่ไม่สามารถตรวจพบข้อสอบที่ทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) ซึ่งมีผลทำให้อัตราความถูกต้องโดยภาพรวมค่อนข้างต่ำ (ประมาณ 49%) สำหรับวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีข้อได้เปรียบที่สามารถตรวจพบข้อสอบที่ทำหน้าที่เบี่ยงเบน ทั้งที่เป็นรูปแบบเดียวกัน (Uniform DIF) และที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) ทำให้อัตราความถูกต้องโดยภาพรวมค่อนข้างสูง (ประมาณ 81%) แต่ผลการตรวจสอบมีความผิดพลาดสูงกว่าเกณฑ์นั้นแสดงว่า ยังไม่สามารถสรุปประสิทธิภาพของสองวิธีดังกล่าวได้อย่างชัดเจน ดังนั้นจำเป็นต้องปรับแก้ (Adjustment) ระดับนัยสำคัญของวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติที่ใช้ในการตัดสินใจการทำหน้าที่เบี่ยงเบนของข้อสอบให้มีค่าลดลง เพื่อให้ให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าลดลงในระดับที่ยอมรับได้ แล้วจึงพิจารณาอัตราความถูกต้องอีกครั้งหนึ่ง ดังนั้นผู้วิจัยจึงปรับระดับนัยสำคัญให้มีค่าลดลง 4 ระดับ คือ .04, .03, .02 และ .01 ตามลำดับ แล้วนำข้อมูลเดิมมาพิจารณาการทำหน้าที่เบี่ยงเบนของข้อสอบ โดยใช้เกณฑ์ตัดสินใจที่ระดับนัยสำคัญ 4 ระดับดังกล่าว ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบทั้ง 5 ระดับ แสดงดังตาราง 57

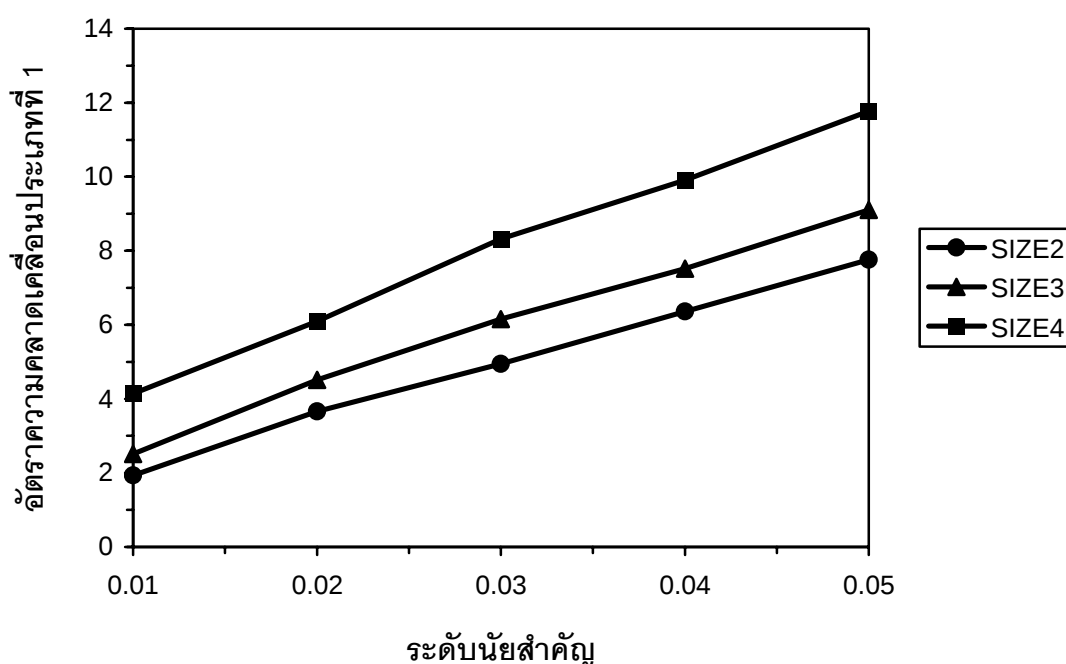
ตาราง 57 อัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ที่ระดับนัยสำคัญ 5 ระดับ

ปัจจัย	ระดับนัยสำคัญ (α)									
	.05		.04		.03		.02		.01	
	<i>P</i>	<i>E</i> ₁	<i>P</i>	<i>E</i> ₁	<i>P</i>	<i>E</i> ₁	<i>P</i>	<i>E</i> ₁	<i>P</i>	<i>E</i> ₁
รูปแบบของ DIF										
UDIF	79.00	8.81	77.14	7.27	74.75	5.86	71.55	4.30	66.49	2.46
NDIF	82.30	9.25	80.74	7.65	78.46	6.16	76.13	4.62	70.93	2.65
สัดส่วนของ DIF										
10%	80.98	9.04	79.15	7.42	76.71	5.94	73.42	4.41	68.06	2.59
20%	80.41	9.07	78.74	7.44	76.23	5.98	73.98	4.34	69.16	2.41
30%	80.58	8.98	78.94	7.53	76.88	6.12	74.12	4.63	68.91	2.66
ความแตกต่างของการแจกแจง θ										
0.0SD	82.44	7.03	80.94	5.62	78.61	4.38	76.26	3.22	70.88	1.84
0.5SD	81.75	7.72	79.96	6.24	78.14	4.95	74.89	3.47	69.87	1.81
1.0SD	77.77	12.33	75.92	10.53	73.07	8.71	70.37	6.69	65.38	4.02
ขนาดตัวอย่าง										
$N_R = N_F = 250$	48.60	7.47	44.94	6.05	40.08	4.64	34.57	3.28	25.13	1.64
$N_R = N_F = 500$	76.51	7.76	73.59	6.36	70.13	4.95	66.12	3.66	58.25	1.93
$N_R = N_F = 1,000$	97.67	9.11	97.41	7.52	96.46	6.15	94.94	4.51	91.95	2.51
$N_R = N_F = 2,000$	99.84	11.77	99.83	9.91	99.76	8.32	99.74	6.09	99.50	4.14
เฉลี่ยทั้งหมด	81.40	9.03	78.94	7.46	76.61	6.01	73.84	4.44	68.71	2.56

ผลการวิเคราะห์ตามตาราง 57 พบว่า เมื่อระดับนัยสำคัญลดลงจาก .05 ถึง .01 มีผลทำให้อัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าลดลงอย่างต่อเนื่อง ตามระดับนัยสำคัญที่ลดลง เมื่อเปรียบเทียบผลการตัดสินใจทำนายที่เบี่ยงเบนที่ระดับนัยสำคัญ .02 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับ

หลายมิติ กับที่ระดับนัยสำคัญ .05 ของวิธีโพลี-ชิปเทสต์ (ตาราง 11) พบว่า ในภาพรวมทั้งสองวิธี มีอัตราความคลาดเคลื่อนประเภทที่ 1 ใกล้เคียงกัน (ประมาณ 4%) แต่วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องสูงกว่าวิธีโพลี-ชิปเทสต์ (ประมาณ 25%) แสดงว่า **ที่ระดับนัยสำคัญดังกล่าว วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีประสิทธิภาพสูงกว่าวิธีโพลี-ชิปเทสต์ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า**

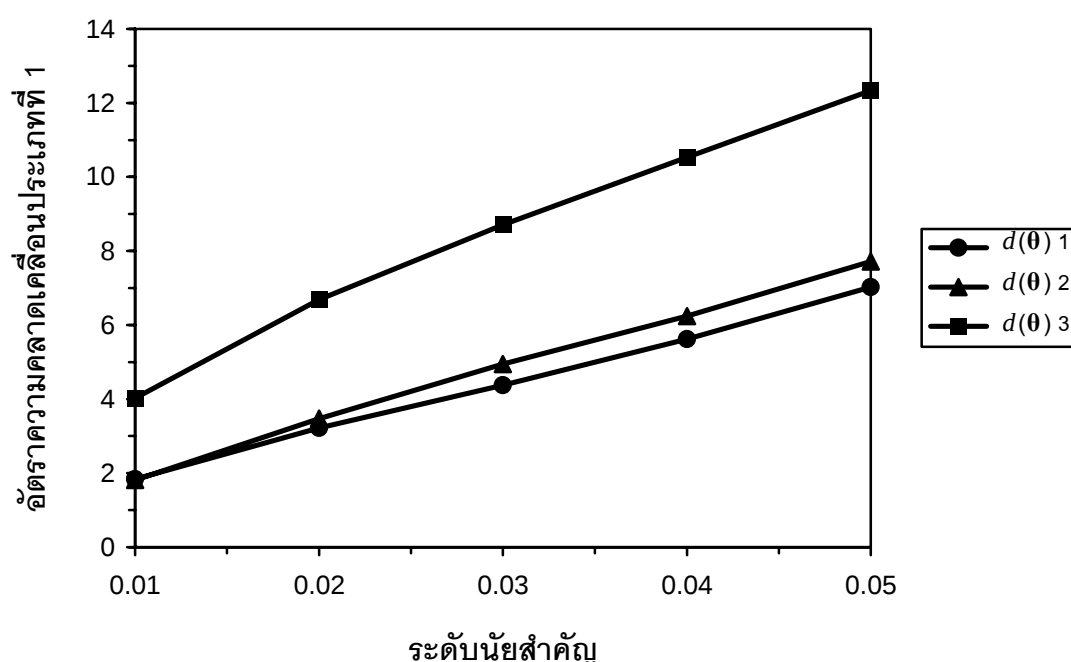
ถ้านำวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติไปใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในข้อมูลจริง จะต้องกำหนดระดับนัยสำคัญให้เหมาะสมกับปัจจัยขนาดตัวอย่าง และปัจจัยความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบ เพื่อให้ผลการตรวจสอบมีอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่เกิน 5% ดังนั้นผู้วิจัยจึงนำผลการวิเคราะห์จากตาราง 53 มาปรับระดับนัยสำคัญให้เหมาะสมกับปัจจัยดังกล่าว ผลการปรับระดับนัยสำคัญทั้งสองปัจจัย แสดงดังภาพประกอบ 26 และ 27



ภาพประกอบ 26 แสดงอัตราความคลาดเคลื่อนประเภทที่ 1

ภายใต้ปัจจัยขนาดตัวอย่าง 3 ขนาด

ภาพประกอบ 26 แสดงผลการปรับระดับนัยสำคัญภายใต้ขนาดตัวอย่าง 500, 1,000 และ 2,000 คนต่อกลุ่ม (กลุ่มอ้างอิง และกลุ่มสนใจ) พบว่า ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า ด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ เมื่อใช้ขนาดตัวอย่าง 500, 1,000 และ 2,000 คนต่อกลุ่ม สามารถกำหนดระดับนัยสำคัญ .03, .023 และ .014 ตามลำดับ ซึ่งจะทำให้ผลการตรวจสอบมีอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่เกิน 5% สำหรับขนาดตัวอย่าง 250 คนต่อกลุ่ม ไม่ได้นำมาปรับระดับนัยสำคัญ เพราะผลการตรวจสอบที่ระดับนัยสำคัญ .05 มีอัตราความถูกต้องค่อนข้างต่ำ



ภาพประกอบ 27 แสดงอัตราความคลาดเคลื่อนประเภทที่ 1
ภายใต้ปัจจัยความแตกต่างของการแจกแจงความสามารถ 3 ระดับ

ภาพประกอบ 27 แสดงผลการปรับระดับนัยสำคัญภายใต้ความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบ พบว่า ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า ด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ เมื่อกลุ่มผู้สอบมีการแจกแจงความสามารถแตกต่างกัน 0.0SD, 0.5SD และ 1.0SD สามารถกำหนดระดับนัยสำคัญ .035, .03 และ .014 ตามลำดับ ซึ่งจะทำให้ผลการตรวจสอบมีอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่เกิน 5%

บทที่ 5

สรุป อภิปรายผล และข้อเสนอแนะ

ในการวิจัยครั้งนี้ มีความมุ่งหมายเพื่อศึกษาประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่าง โดยใช้ข้อมูลจำลอง ก่อนที่จะสรุปผลการวิจัย อภิปรายผล และให้ข้อเสนอแนะ เพื่อให้เห็นกรอบของการวิจัยจะขอกล่าวถึงความมุ่งหมาย สมมติฐาน และวิธีดำเนินการวิจัย โดยสังเขปดังนี้

ความมุ่งหมายของการวิจัย

การวิจัยครั้งนี้ มีความมุ่งหมายเฉพาะดังนี้

1. เพื่อเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า ด้วยวิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ภายใต้เงื่อนไขต่างกันของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ซึ่งประกอบด้วย รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่าง
2. เพื่อเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า ระหว่างวิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ภายใต้เงื่อนไขเดียวกันของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ซึ่งประกอบด้วย รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่าง

สมมติฐานของการวิจัย

สมมติฐานในการวิจัยมีดังนี้

1. ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ประกอบด้วย รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่าง มีผลทำให้อัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า ด้วยวิธีโพลี-ชิปเทสต์ แตกต่างกัน
2. ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ประกอบด้วย รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่าง มีผลทำให้อัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า ด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับ แตกต่างกัน
3. ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ประกอบด้วย รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่าง มีผลทำให้อัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า ด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ แตกต่างกัน
4. วิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ประกอบด้วย รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่าง มีอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกัน

วิธีดำเนินการวิจัย

ในการดำเนินการวิจัย สรุปได้ดังนี้

1. ข้อมูลที่ศึกษาเป็นข้อมูลจำลอง โดยใช้โมเดลพาเชียลเครดิตทั่วไปแบบหลายมิติ (Multidimensional generalized partial credit Model) ภายใต้ทฤษฎีการตอบข้อสอบแบบหลายมิติ จำลองแบบทดสอบที่มีโครงสร้างวัดความสามารถสองมิติ (Two-dimensional) จำนวน 40 ข้อ ซึ่งเป็นการวัดความสามารถหลายมิติภายในข้อสอบ (Multidimensional within-item test) โดยข้อสอบแต่ละข้อวัดความสามารถเป้าหมายสองมิติที่มีน้ำหนักแตกต่างกัน กำหนดให้ θ_1 และ θ_2 แทนความสามารถหลักในมิติที่หนึ่งและมิติที่สอง ตามลำดับ ข้อสอบทุกข้อมีรายการตอบ 5 รายการ ผลการตอบข้อสอบในรายการ 1, 2, 3, 4 และ 5 ให้คะแนนเป็น 1, 2, 3, 4 และ 5 คะแนน ตามลำดับ ในการจำลองข้อมูลใช้โปรแกรม RESGEN จำลองผลการตอบข้อสอบภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ (1) รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน 2 รูปแบบ; ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (2) สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 3 ขนาด; 10%, 20% และ 30% (3) ความแตกต่างของการแจกแจงความสามารถ 3 ระดับ; 0.0SD, 0.5SD และ 1.0SD และ (4) ขนาดตัวอย่าง 4 ขนาด; 250, 500, 1,000 และ 2,000 คนในแต่ละกลุ่ม รวมข้อมูลทั้งหมดที่ต้องจัดกระทำเพื่อตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ จำนวน 72 เงื่อนไข ($2 \times 3 \times 3 \times 4$) และในแต่ละเงื่อนไขจำลองข้อมูลซ้ำ 50 ครั้ง

2. นำข้อมูลที่ได้จากการจำลองทั้งหมดมาตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบสามวิธี คือ วิธีโพลี-ซิปเทสท์ (Poly-SIBTEST) วิธีการถดถอยโลจิสติกแบบจัดอันดับ (Ordinal logistic regression; OLR) และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ (Multidimensional ordinal logistic regression; MOLR) โดยวิธีแรกใช้โปรแกรม DIFPACK ส่วนสองวิธีหลังใช้โปรแกรม SPSS แล้วตัดสินใจข้อสอบทำหน้าที่เบี่ยงเบนที่ระดับนัยสำคัญ .05 จากนั้นนำผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบทั้งหมดมาคำนวณอัตราความถูกต้อง (Power rate; P) และอัตราความคลาดเคลื่อนประเภทที่ 1 (Type I error rate; E_1)

3. การเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ 3 วิธี ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย โดยใช้สถิติการวิเคราะห์ความแปรปรวนหลายตัวแปร (Multivariate analysis of variance; MANOVA) ซึ่งมีตัวแปรอิสระ 5 ตัว คือ วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ปัจจัยรูปแบบของ

ข้อสอบทำหน้าที่เบี่ยงเบน ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ปัจจัยความแตกต่างของการแจกแจงความสามารถ และปัจจัยขนาดตัวอย่าง ส่วนตัวแปรตามมี 2 ตัว คือ อัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการศึกษาครั้งนี้ กำหนดโมเดลของการวิเคราะห์ให้มีปฏิสัมพันธ์ระหว่างตัวแปรอิสระไม่เกินอันดับสอง (Second-order) ถ้าผลการวิเคราะห์มีนัยสำคัญทางสถิติจะทดสอบผลระหว่างกลุ่ม (Test of between-subjects effects) ของตัวแปรตามแต่ละตัวต่อจากนั้นจะทดสอบผลย่อย (Simple effect) ภายใต้อัตราความถูกต้องสูง แล้วทดสอบภายหลังโดยใช้วิธีของเชฟเฟ (Scheffé) สำหรับเกณฑ์ที่ใช้ในการตัดสินใจประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ พิจารณาจากผลการตรวจสอบที่มีอัตราความถูกต้องสูง และอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำ ซึ่งมีค่าไม่เกิน 5% ($\alpha = .05$)

สรุปผลการวิจัย

การสรุปผลการวิจัยแบ่งออกเป็น 3 ตอน คือ (1) ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้อัตราความถูกต้องสูง (2) ผลการทดสอบปัจจัยที่ศึกษาที่มีผลต่ออัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ และ (3) ผลการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้อัตราความถูกต้องสูง ดังรายละเอียดต่อไปนี้

1. ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้อัตราความถูกต้องสูง

การสรุปผลการวิจัยในตอนนี้ เป็นผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า โดยใช้วิธีการตรวจสอบสามวิธี คือ วิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ภายใต้อัตราความถูกต้องสูงของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบ และขนาดตัวอย่าง สำหรับผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ สรุปในรูปค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้อัตราความถูกต้องสูง 4 ปัจจัย และปฏิสัมพันธ์ระหว่างปัจจัยมีดังนี้

1.1 ผลการตรวจสอบภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องสูงสุดในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน คิดเป็นค่าเฉลี่ยประมาณ 82.30% และในเงื่อนไขข้อสอบดังกล่าววิธีโพลี-ชิปเทสที่มีอัตราความถูกต้องต่ำสุด คิดเป็นค่าเฉลี่ยประมาณ 20.97% เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน คิดเป็นค่าเฉลี่ยประมาณ 13.64% และในเงื่อนไขเดียวกัน วิธีโพลี-ชิปเทสที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด คิดเป็นค่าเฉลี่ยประมาณ 2.58%

1.2 ผลการตรวจสอบภายใต้ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องสูงสุดในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% คิดเป็นค่าเฉลี่ยประมาณ 81.58% ส่วนในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% วิธีโพลี-ชิปเทสที่มีอัตราความถูกต้องต่ำสุด คิดเป็นค่าเฉลี่ยประมาณ 46.40% เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด คิดเป็นค่าเฉลี่ยประมาณ 20.20% และในเงื่อนไขเดียวกัน วิธีโพลี-ชิปเทสที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด คิดเป็นค่าเฉลี่ยประมาณ 3.84%

1.3 ผลการตรวจสอบภายใต้ปัจจัยความแตกต่างของการแจกแจงความสามารถ

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบ พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องสูงสุดในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD คิดเป็นค่าเฉลี่ยประมาณ 82.44% และในเงื่อนไขเดียวกัน วิธีโพลี-ชิปเทสที่มีอัตราความถูกต้องต่ำสุด คิดเป็นค่าเฉลี่ยประมาณ 41.83% เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 1.0SD คิดเป็นค่าเฉลี่ยประมาณ 30.22% ส่วนในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD วิธีโพลี-ชิปเทสที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด คิดเป็นค่าเฉลี่ยประมาณ 3.63%

1.4 ผลการตรวจสอบภายใต้ปัจจัยขนาดตัวอย่าง

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยขนาดตัวอย่าง พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องสูงสุด ในเงื่อนไขขนาดตัวอย่าง 2,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 99.84% ส่วนวิธีโพลี-ชิปเทสที่มีอัตราความถูกต้องต่ำสุด ในเงื่อนไขขนาดตัวอย่าง 250 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 27.21% เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ในเงื่อนไขขนาดตัวอย่าง 2,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 28.87% ส่วนวิธีโพลี-ชิปเทสที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ในเงื่อนไขขนาดตัวอย่าง 1,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 2.96%

1.5 ผลการตรวจสอบภายใต้ปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยรูปแบบข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนในช่วง 10% ถึง 20% พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องสูงสุด ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% คิดเป็นค่าเฉลี่ยประมาณ 83.79% แต่มีค่าต่ำสุดในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% คิดเป็นค่าเฉลี่ยประมาณ 70.28% เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% คิดเป็นค่าเฉลี่ยประมาณ 26.80% และในเงื่อนไขเดียวกัน วิธีโพลี-ชิปเทสที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด คิดเป็นค่าเฉลี่ยประมาณ 1.67%

ส่วนผลการตรวจสอบภายใต้ปัจจัยรูปแบบข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน และสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนในช่วง 10% ถึง 20% พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องสูงสุด ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% คิดเป็นค่าเฉลี่ยประมาณ 82.92% และในเงื่อนไขเดียวกัน วิธีโพลี-ชิปเทสที่มีอัตราความถูกต้องต่ำสุด คิดเป็นค่าเฉลี่ยประมาณ 20% เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% คิดเป็นค่าเฉลี่ยประมาณ 13.67% ส่วนวิธีโพลี-ชิปเทสที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% คิดเป็นค่าเฉลี่ยประมาณ 6.01%

1.6 ผลการตรวจสอบภายใต้ปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถ

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถ ในช่วง 0.0SD ถึง 1.0SD พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องสูงสุด ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD คิดเป็นค่าเฉลี่ยประมาณ 82.48% แต่ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 1.0SD มีอัตราความถูกต้องต่ำสุด คิดเป็นค่าเฉลี่ยประมาณ 74.81% สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 1.0SD วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด คิดเป็นค่าเฉลี่ยประมาณ 32.25% ส่วนในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD วิธีโพลี-ชิปเทสที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด คิดเป็นค่าเฉลี่ยประมาณ 1.55%

ส่วนผลการตรวจสอบภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน และกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถในช่วง 0.0SD ถึง 1.0SD พบว่า ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องสูงสุด คิดเป็นค่าเฉลี่ยประมาณ 83.82% และในเงื่อนไขเดียวกันวิธีโพลี-ชิปเทสที่มีอัตราความถูกต้องต่ำสุด คิดเป็นค่าเฉลี่ยประมาณ 6.53% สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 1.0SD คิดเป็นค่าเฉลี่ยประมาณ 28.19% ส่วนในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.5SD วิธีโพลี-ชิปเทสที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด คิดเป็นค่าเฉลี่ยประมาณ 5.48%

1.7 ผลการตรวจสอบภายใต้ปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยขนาดตัวอย่าง

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยรูปแบบข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และขนาดตัวอย่างในช่วง 250 ถึง 2,000 คนต่อกลุ่ม พบว่าวิธีโพลี-ชิปเทสที่มีอัตราความถูกต้องสูงสุด ในเงื่อนไขขนาดตัวอย่าง 2,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 99.78% ส่วนในเงื่อนไขขนาดตัวอย่าง 250 คนต่อกลุ่มมีอัตราความถูกต้อง

ต่ำสุด คิดเป็นค่าเฉลี่ยประมาณ 37.52% เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ในเงื่อนไข ขนาดตัวอย่าง 2,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 36.29% และในเงื่อนไขเดียวกัน วิธีโพลี-ชิปเทสที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด คิดเป็นค่าเฉลี่ยประมาณ 0.66%

ส่วนผลการตรวจสอบ ภายใต้ปัจจัยรูปแบบข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน และขนาดตัวอย่างในช่วง 250 ถึง 2,000 คนต่อกลุ่ม พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องสูงสุด ในเงื่อนไขขนาดตัวอย่าง 2,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 100% ส่วนวิธีโพลี-ชิปเทสที่มีอัตราความถูกต้องต่ำสุด ในเงื่อนไขขนาดตัวอย่าง 250 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 16.90% สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ในเงื่อนไขขนาดตัวอย่าง 2,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 21.44% ส่วนวิธีโพลี-ชิปเทสที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ในเงื่อนไขขนาดตัวอย่าง 1,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 4.78%

1.8 ผลการตรวจสอบภายใต้ปฏิสัมพันธ์สองทางระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถ

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% และความแตกต่างของการแจกแจงความสามารถในช่วง 0.0SD ถึง 1.0SD พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องสูงสุด ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.5SD คิดเป็นค่าเฉลี่ยประมาณ 83.56% ส่วนวิธีโพลี-ชิปเทสที่มีอัตราความถูกต้องต่ำสุด ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD คิดเป็นค่าเฉลี่ยประมาณ 43.63% เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 1.0SD คิดเป็นค่าเฉลี่ยประมาณ 28.86% ส่วนวิธีโพลี-ชิปเทสที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.5SD คิดเป็นค่าเฉลี่ยประมาณ 4.39%

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% และความแตกต่างของการแจกแจงความสามารถในช่วง 0.0SD ถึง 1.0SD พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องสูงสุด ในเงื่อนไขความ-

แตกต่างของการแจกแจงความสามารถ 0.0SD คิดเป็นค่าเฉลี่ยประมาณ 82.63% และในเงื่อนไขเดียวกัน วิธีโพลี-ชิปเทสที่มีอัตราความถูกต้องต่ำสุด คิดเป็นค่าเฉลี่ยประมาณ 42.75% สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 1.0SD คิดเป็นค่าเฉลี่ยประมาณ 29.79% ส่วนวิธีโพลี-ชิปเทสที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.5SD คิดเป็นค่าเฉลี่ยประมาณ 3.48%

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% และความแตกต่างของการแจกแจงความสามารถในช่วง 0.0SD ถึง 1.0SD พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องสูงสุด ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD คิดเป็นค่าเฉลี่ยประมาณ 83.19% และในเงื่อนไขเดียวกัน วิธีโพลี-ชิปเทสที่มีอัตราความถูกต้องต่ำสุด คิดเป็นค่าเฉลี่ยประมาณ 39.13% สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 1.0SD คิดเป็นค่าเฉลี่ยประมาณ 32.01% ส่วนวิธีโพลี-ชิปเทสที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD คิดเป็นค่าเฉลี่ยประมาณ 2.95%

1.9 ผลการตรวจสอบภายใต้ปฏิสัมพันธ์สองทางระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยขนาดตัวอย่าง

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% และขนาดตัวอย่างในช่วง 250 ถึง 2,000 คนต่อกลุ่ม พบว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องสูงสุด ในเงื่อนไขขนาดตัวอย่าง 2,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 100% ส่วนวิธีโพลี-ชิปเทสที่มีอัตราความถูกต้องต่ำสุด ในเงื่อนไขขนาดตัวอย่าง 250 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 31% เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ในเงื่อนไขขนาดตัวอย่าง 2,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 23.20% และในเงื่อนไขเดียวกัน วิธีโพลี-ชิปเทสที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด คิดเป็นค่าเฉลี่ยประมาณ 3.40%

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยสัดส่วนของข้อสอบทำ
หน้าที่เบี่ยงเบน 20% และขนาดตัวอย่างในช่วง 250 ถึง 2,000 คนต่อกลุ่ม พบว่าวิธีการถดถอย-
โลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องสูงสุด ในเงื่อนไขขนาดตัวอย่าง 2,000 คน
ต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 99.75% ส่วนวิธีโพลี-ชิปเทสที่มีอัตราความถูกต้องต่ำสุด ใน
เงื่อนไขขนาดตัวอย่าง 250 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 25.29% สำหรับอัตราความ-
คลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อน
ประเภทที่ 1 สูงสุด ในเงื่อนไขขนาดตัวอย่าง 2,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 27.91%
ส่วนวิธีโพลี-ชิปเทสที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ในเงื่อนไขขนาดตัวอย่าง 1,000
คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 2.46%

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยสัดส่วนของข้อสอบทำ
หน้าที่เบี่ยงเบน 30% และขนาดตัวอย่างในช่วง 250 ถึง 2,000 คนต่อกลุ่ม พบว่าวิธีการถดถอย-
โลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องสูงสุด ในเงื่อนไขขนาดตัวอย่าง 2,000 คน
ต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 99.78% ส่วนวิธีโพลี-ชิปเทสที่มีอัตราความถูกต้องต่ำสุด ใน
เงื่อนไขขนาดตัวอย่าง 250 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 25.33% สำหรับอัตราความ-
คลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อน
ประเภทที่ 1 สูงสุด ในเงื่อนไขขนาดตัวอย่าง 2,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 35.49%
ส่วนวิธีโพลี-ชิปเทสที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ในเงื่อนไขขนาดตัวอย่าง 1,000
คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 2.19%

1.10 ผลการตรวจสอบภายใต้ปฏิสัมพันธ์สองทางระหว่างปัจจัยความแตกต่างของการแจกแจงความสามารถ กับปัจจัยขนาดตัวอย่าง

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยความแตกต่างของ
การแจกแจงความสามารถ 0.0SD และขนาดตัวอย่างในช่วง 250 ถึง 2,000 คนต่อกลุ่ม พบว่า
วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องสูงสุด ในเงื่อนไขขนาดตัวอย่าง
2,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 100% ส่วนในเงื่อนไขขนาดตัวอย่าง 250 คนต่อกลุ่ม
วิธีโพลี-ชิปเทสที่มีอัตราความถูกต้องต่ำสุด คิดเป็นค่าเฉลี่ยประมาณ 21.07% สำหรับอัตราความ-
คลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อน
ประเภทที่ 1 สูงสุด ในเงื่อนไขขนาดตัวอย่าง 2,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 15.49%
ส่วนวิธีโพลี-ชิปเทสที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ในเงื่อนไขขนาดตัวอย่าง 1,000
คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 2.81%

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยความแตกต่างของการแจกแจงความสามารถ $0.5SD$ และขนาดตัวอย่างในช่วง 250 ถึง 2,000 คนต่อกลุ่ม พบว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องสูงสุด ในเงื่อนไขขนาดตัวอย่าง 2,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 99.75% ส่วนในเงื่อนไขขนาดตัวอย่าง 250 คนต่อกลุ่ม วิธีโพลี-ชิปเทสที่มีอัตราความถูกต้องต่ำสุด คิดเป็นค่าเฉลี่ยประมาณ 25.64% เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ในเงื่อนไขขนาดตัวอย่าง 2,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 17.05% และในเงื่อนไขเดียวกัน วิธีโพลี-ชิปเทสที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด คิดเป็นค่าเฉลี่ยประมาณ 2.60%

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยความแตกต่างของการแจกแจงความสามารถ $1.0SD$ และขนาดตัวอย่างในช่วง 250 ถึง 2,000 คนต่อกลุ่ม พบว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องสูงสุด ในเงื่อนไขขนาดตัวอย่าง 2,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 99.78% ส่วนในเงื่อนไขขนาดตัวอย่าง 250 คนต่อกลุ่ม วิธีโพลี-ชิปเทสที่มีอัตราความถูกต้องต่ำสุด คิดเป็นค่าเฉลี่ยประมาณ 34.92% เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ในเงื่อนไขขนาดตัวอย่าง 2,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 54.06% ส่วนวิธีโพลี-ชิปเทสที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ในเงื่อนไขขนาดตัวอย่าง 1,000 คน และ 2,000 ต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 3.42%

1.11 ผลการตรวจสอบภายใต้ปฏิสัมพันธ์สามทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถ

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% ถึง 30% และความแตกต่างของการแจกแจงความสามารถ $0.0SD$ ถึง $1.0SD$ พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องสูงสุด ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% และความแตกต่างของการแจกแจงความสามารถ $1.0SD$ คิดเป็นค่าเฉลี่ยประมาณ 85.50% ในขณะที่วิธีโพลี-ชิปเทสที่มีอัตราความถูกต้องต่ำสุด ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่

เบี่ยงเบน 30% และความแตกต่างของการแจกแจงความสามารถ 1.0SD คิดเป็นค่าเฉลี่ยประมาณ 69% เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% และความแตกต่างของการแจกแจงความสามารถ 1.0SD คิดเป็นค่าเฉลี่ยประมาณ 36.43% ในขณะที่ วิธีโพลี-ชิปเทสที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% และความแตกต่างของการแจกแจงความสามารถ 0.0SD คิดเป็นค่าเฉลี่ยประมาณ 1.07%

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% ถึง 30% และความแตกต่างของการแจกแจงความสามารถ 0.0SD ถึง 1.0SD พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องสูงสุด ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% และความแตกต่างของการแจกแจงความสามารถ 0.0SD คิดเป็นค่าเฉลี่ยประมาณ 87.08% และในเงื่อนไขเดียวกัน วิธีโพลี-ชิปเทสที่มีอัตราความถูกต้องต่ำสุด คิดเป็นค่าเฉลี่ยประมาณ 4.58% เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% และความแตกต่างของการแจกแจงความสามารถ 1.0SD คิดเป็นค่าเฉลี่ยประมาณ 29.24% ในขณะที่วิธีโพลี-ชิปเทสที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% และความแตกต่างของการแจกแจงความสามารถ 0.0SD คิดเป็นค่าเฉลี่ยประมาณ 4.82%

1.12 ผลการตรวจสอบภายใต้ปฏิสัมพันธ์สามทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยขนาดตัวอย่าง

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% ถึง 30% และขนาดตัวอย่าง 250 ถึง 2,000 คนต่อกลุ่ม พบว่า ทั้งสามวิธีมีอัตราความถูกต้องสูงสุด ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% และขนาดตัวอย่าง 2,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 100% นอกจากนี้ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20%

และขนาดตัวอย่าง 2,000 คนต่อกลุ่ม วิธีโพลี-ชิปเทสต์มีค่าเฉลี่ยอัตราความถูกต้องสูงสุดด้วย ส่วนในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เพียงเบน 30% และขนาดตัวอย่าง 250 คนต่อกลุ่ม วิธีโพลี-ชิปเทสต์ มีอัตราความถูกต้องต่ำสุด คิดเป็นค่าเฉลี่ยประมาณ 33.72% สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เพียงเบน 30% และขนาดตัวอย่าง 2,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 50.14% ส่วนวิธีโพลี-ชิปเทสต์มีค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เพียงเบน 30% และขนาดตัวอย่าง 1,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 0.21%

ผลการตรวจสอบการทำหน้าที่เพียงเบนของข้อสอบ ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เพียงเบนที่ไม่เป็นรูปแบบเดียวกัน ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เพียงเบน 10% ถึง 30% และขนาดตัวอย่าง 250 ถึง 2,000 คนต่อกลุ่ม พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องสูงสุด ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เพียงเบน 10%, 20% และ 30% ที่มีขนาดตัวอย่าง 2,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 100% ส่วนวิธีโพลี-ชิปเทสต์มีอัตราความถูกต้องต่ำสุด ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เพียงเบน 20% และขนาดตัวอย่าง 250 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 14.58% เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เพียงเบน 10% และขนาดตัวอย่าง 2,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 22.04% ส่วนวิธีโพลี-ชิปเทสต์มีค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เพียงเบน 20% และ 30% เมื่อใช้ขนาดตัวอย่าง 1,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 4.17%

1.13 ผลการตรวจสอบภายใต้ปฏิสัมพันธ์สามทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เพียงเบน ปัจจัยความแตกต่างของการแจกแจงความสามารถ กับปัจจัยขนาดตัวอย่าง

ผลการตรวจสอบการทำหน้าที่เพียงเบนของข้อสอบ ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เพียงเบนที่เป็นรูปแบบเดียวกัน ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD ถึง 1.0SD และขนาดตัวอย่าง 250 ถึง 2,000 คนต่อกลุ่ม พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องสูงสุด ในเงื่อนไขความแตกต่างของการแจกแจง

ความสามารถ 0.0SD และขนาดตัวอย่าง 2,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 100% ส่วนวิธีโพลี-ชิปเทสที่มีอัตราความถูกต้องต่ำสุด ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD และขนาดตัวอย่าง 250 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 32.56% สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุดในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 1.0SD และขนาดตัวอย่าง 2,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 59.60% ส่วนวิธีโพลี-ชิปเทสที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.5SD และขนาดตัวอย่าง 2,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 0.47%

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD ถึง 1.0SD และขนาดตัวอย่าง 250 ถึง 2,000 คนต่อกลุ่ม พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องสูงสุด ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD, 0.5SD และ 1.0SD เมื่อใช้ขนาดตัวอย่าง 2,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 100% ส่วนวิธีโพลี-ชิปเทสที่มีอัตราความถูกต้องต่ำสุด ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD และขนาดตัวอย่าง 500 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 4.86% สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 1.0SD และขนาดตัวอย่าง 2,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 48.51% ส่วนวิธีโพลี-ชิปเทสที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.5SD และขนาดตัวอย่าง 1,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 4.11%

1.14 ผลการตรวจสอบภายใต้ปฏิสัมพันธ์สามทางระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ปัจจัยความแตกต่างของการแจกแจงความสามารถ กับปัจจัยขนาดตัวอย่าง

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD ถึง 1.0SD และขนาดตัวอย่าง 250 ถึง 2,000 คนต่อกลุ่ม พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องสูงสุด ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD,

0.5SD และ 1.0SD เมื่อใช้ขนาดตัวอย่าง 2,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 100% และในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD และขนาดตัวอย่าง 2,000 คนต่อกลุ่ม วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องสูงสุดด้วย ส่วนวิธีโพลี-ชิปเทสต์มีอัตราความถูกต้องต่ำสุด ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD และขนาดตัวอย่าง 250 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 23.25% เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 1.0SD และขนาดตัวอย่าง 2,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 50.83% ส่วนวิธีโพลี-ชิปเทสต์มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD และขนาดตัวอย่าง 2,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 3.19%

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยสัดส่วนของข้อสอบ ทำหน้าที่เบี่ยงเบน 20% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD ถึง 1.0SD และขนาดตัวอย่าง 250 ถึง 2,000 คนต่อกลุ่ม พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องสูงสุด ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD และ 1.0SD เมื่อใช้ขนาดตัวอย่าง 2,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 100% ในขณะที่วิธีโพลี-ชิปเทสที่มีอัตราความถูกต้องต่ำสุด ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD และขนาดตัวอย่าง 250 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 20.38% เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 1.0SD และขนาดตัวอย่าง 2,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 53.66% ส่วนวิธีโพลี-ชิปเทสที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.5SD และขนาดตัวอย่าง 1,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 2.19%

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยสัดส่วนของข้อสอบ ทำหน้าที่เบี่ยงเบน 30% ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD ถึง 1.0SD และขนาดตัวอย่าง 250 ถึง 2,000 คนต่อกลุ่ม พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลาย มิติมีอัตราความถูกต้องสูงสุด ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD และ 0.5SD เมื่อใช้ขนาดตัวอย่าง 2,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 100% ในขณะที่วิธีโพลี-ชิปเทสต์มีอัตราความถูกต้องต่ำสุด ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD

และขนาดตัวอย่าง 250 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 19.58% เมื่อพิจารณาค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 1.0SD และขนาดตัวอย่าง 2,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 57.68% ส่วนวิธีโพลี-ชิปเทสท์มีค่าเฉลี่ยอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ในเงื่อนไขความแตกต่างของการแจกแจงความสามารถ 0.0SD และขนาดตัวอย่าง 1,000 คนต่อกลุ่ม คิดเป็นค่าเฉลี่ยประมาณ 1.68%

1.15 ผลการตรวจสอบภายใต้ปฏิสัมพันธ์สี่ทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ปัจจัยความแตกต่างของการแจกแจงความสามารถ กับปัจจัยขนาดตัวอย่าง

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยรูปแบบของข้อสอบ ทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% ถึง 30% ความแตกต่างของการแจกแจงความสามารถ 0.0SD ถึง 1.0SD และขนาดตัวอย่าง 250 ถึง 2,000 คนต่อกลุ่ม พบว่า วิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยอัตราความถูกต้องอยู่ในช่วงประมาณ 29.50%-100%, 31.17%-100% และ 39.25%-100% ตามลำดับ และทั้งสามวิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในช่วงประมาณ 0%-13.06%, 5.06%-67.14% และ 5.56%-18.44% ตามลำดับ

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยรูปแบบของข้อสอบ ทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% ถึง 30% ความแตกต่างของการแจกแจงความสามารถ 0.0SD ถึง 1.0SD และขนาดตัวอย่าง 250 ถึง 2,000 คนต่อกลุ่ม พบว่า วิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยอัตราความถูกต้องอยู่ในช่วงประมาณ 2.50%-54.17%, 50%-100% และ 43.75%-100% ตามลำดับ และทั้งสามวิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในช่วงประมาณ 3.21%-14.17%, 2.50%-48.89% และ 5.36%-21.07% ตามลำดับ

เมื่อพิจารณาผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า โดยภาพรวมภายใต้ทุกเงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องโดยเฉลี่ยสูงสุดประมาณ 81% ซึ่งมีอัตราความถูกต้องสูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีโพลี-

ชิปเทสท์ ประมาณ 3% และ 33% ตามลำดับ เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีโพลี-ชิปเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 โดยเฉลี่ยต่ำสุด ประมาณ 4% ซึ่งเกือบทุกเงื่อนไขมีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ภายในขอบเขตที่กำหนด (Nominal limits) ซึ่งไม่เกิน 5% ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 โดยเฉลี่ยสูงสุด ประมาณ 17% ซึ่งสูงกว่าขอบเขตที่กำหนด ประมาณ 12% ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าขอบเขตที่กำหนด ประมาณ 4%

2. ผลการทดสอบปัจจัยที่ศึกษาที่มีผลต่ออัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ

การสรุปผลการวิจัยในตอนนี้ เป็นผลการทดสอบปัจจัยที่มีผลต่ออัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ โดยสรุปผลภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่าง รวมทั้งเงื่อนไขของปฏิสัมพันธ์สองทางระหว่างปัจจัยดังกล่าว ดังนั้นการสรุปผลในตอนนี้ได้มาจากการทดสอบผลย่อย (Simple effect) และการเปรียบเทียบภายหลังโดยใช้วิธีของเชฟเฟ (Scheffé) ซึ่งมีผลการทดสอบอัตราความถูกต้องจำนวน 9 เงื่อนไข และผลการทดสอบอัตราความคลาดเคลื่อนประเภทที่ 1 จำนวน 10 เงื่อนไข ดังรายละเอียดต่อไปนี้

2.1 ผลการทดสอบปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน

ผลการทดสอบปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ภายใต้วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ พบว่า ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนมีผลต่ออัตราความถูกต้องของวิธีโพลี-ชิปเทสท์อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่อวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ กล่าวคือ การตรวจสอบในข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน มีผลทำให้วิธีโพลี-ชิปเทสท์มีอัตราความถูกต้องสูงกว่าการตรวจสอบในข้อสอบที่ทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ สามารถตรวจสอบข้อสอบทำหน้าที่เบี่ยงเบนได้ทั้งสองรูปแบบ

ตัวอย่างเช่น การตรวจสอบในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน วิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ มีอัตราความถูกต้องประมาณ 20.97%, 80.09% และ 82.30% ตามลำดับ ส่วนการตรวจสอบในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน ทั้งสามวิธีมีอัตราความถูกต้องประมาณ 76.65%, 76.96% และ 79.00% ตามลำดับ ดังนั้นอัตราความถูกต้องของวิธีโพลี-ชิปเทสต์เพิ่มขึ้นประมาณ 55.68% ส่วนอัตราความถูกต้องของวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติลดลงประมาณ 3.13% และ 3.30% ตามลำดับ

ส่วนผลการทดสอบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ กล่าวคือ การตรวจสอบในข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน มีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสต์มีค่าต่ำกว่าการตรวจสอบในข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ซึ่งตรงข้ามกับวิธีการถดถอยโลจิสติกแบบจัดอันดับ ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ เมื่อตรวจสอบในข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน มีอัตราความคลาดเคลื่อนประเภทที่ 1 ใกล้เคียงกัน

ตัวอย่างเช่น การตรวจสอบในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน วิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ มีอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณ 6.34%, 13.64% และ 9.25% ตามลำดับ เมื่อตรวจสอบในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน ทั้งสามวิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณ 2.58%, 20.09% และ 8.81% ตามลำดับ ดังนั้นวิธีโพลี-ชิปเทสต์มีอัตราความคลาดเคลื่อนประเภทที่ 1 ลดลง ประมาณ 3.76% และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความคลาดเคลื่อนประเภทที่ 1 ลดลงเล็กน้อย ประมาณ 0.44% ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้น ประมาณ 6.45%

2.2 ผลการทดสอบปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน

ผลการทดสอบปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ภายใต้วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ พบว่า ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนมีผลต่ออัตราความถูกต้องของวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่ออัตราความถูกต้องของวิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ กล่าวคือ เมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้น 20% (จาก 10% ถึง 30%) มีผลทำให้วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องลดลง แต่ไม่มีผลต่อวิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ

ตัวอย่างเช่น การตรวจสอบในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% อัตราความถูกต้องของวิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าประมาณ 50.58%, 81.58% และ 80.98% ตามลำดับ ส่วนในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% อัตราความถูกต้องของสามวิธีดังกล่าวมีค่าประมาณ 46.40%, 75.90% และ 80.58% ดังนั้น ถ้าสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้น 20% แล้ววิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องลดลง ประมาณ 4.18% และ 5.68% ตามลำดับ ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องลดลงเล็กน้อย ประมาณ 0.40%

ส่วนผลการทดสอบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ กล่าวคือ เมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้นในช่วง 10% ถึง 30% มีผลทำให้วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้น แต่ไม่มีผลกระทบต่อวิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ

ตัวอย่างเช่น การตรวจสอบในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ มีค่าประมาณ 5.32%, 14.00% และ 9.04% ตามลำดับ ส่วนในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% อัตราความคลาดเคลื่อนประเภทที่ 1 ของสามวิธีดังกล่าวมีค่าประมาณ 3.84%, 20.20% และ 8.98% ตามลำดับ ดังนั้น ถ้า

สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้น 20% แล้ววิธีโพลี-ชิปเทสที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ลดลง ประมาณ 1.48% ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้น ประมาณ 6.20% ในขณะที่วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิตีมีอัตราความคลาดเคลื่อนประเภทที่ 1 ลดลงเล็กน้อย ประมาณ 0.06%

2.3 ผลการทดสอบปัจจัยความแตกต่างของการแจกแจงความสามารถ

ผลการทดสอบปัจจัยความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบ ภายใต้วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ พบว่า ปัจจัยความแตกต่างของการแจกแจงความสามารถมีผลต่ออัตราความถูกต้องของวิธีโพลี-ชิปเทสอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่อวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิตี กล่าวคือ เมื่อความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มเพิ่มขึ้น ในช่วง 0.0SD ถึง 1.0SD มีผลทำให้อัตราความถูกต้องของวิธีโพลี-ชิปเทสเพิ่มขึ้น แต่ไม่มีผลต่อวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิตี

ตัวอย่างเช่น เมื่อไม่มีความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบ วิธีโพลี-ชิปเทส วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิตี มีอัตราความถูกต้องประมาณ 41.83%, 79.72% และ 82.44% ตามลำดับ เมื่อกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถเพิ่มขึ้น 1.0SD (จาก 0.0SD ถึง 1.0SD) ทั้งสามวิธีมีอัตราความถูกต้องประมาณ 57.45%, 76.08% และ 77.77% ตามลำดับ ดังนั้น ถ้ากลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถเพิ่มขึ้น 1.0SD แล้วอัตราความถูกต้องของวิธีโพลี-ชิปเทสเพิ่มขึ้น ประมาณ 15.62% ส่วนอัตราความถูกต้องของวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิตีลดลง ประมาณ 3.46% และ 4.67% ตามลำดับ

ส่วนผลการทดสอบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ปัจจัยความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทส วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิตีอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 กล่าวคือ เมื่อความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบเพิ่มขึ้น 1.0SD (จาก 0.0SD ถึง 1.0SD) มีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งสามวิธีเพิ่มขึ้น

ตัวอย่างเช่น เมื่อไม่มีความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบวิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ มีอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณ 3.63%, 10.06% และ 7.03% ตามลำดับ เมื่อความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบเพิ่มขึ้น 1.0SD (จาก 0.0SD ถึง 1.0SD) ทั้งสามวิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณ 5.96%, 30.22% และ 12.33% ตามลำดับ ดังนั้น ถ้ากลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถเพิ่มขึ้น 1.0SD แล้ววิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ มีอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้น ประมาณ 2.33%, 20.16% และ 5.3% ตามลำดับ

2.4 ผลการทดสอบปัจจัยขนาดตัวอย่าง

ผลการทดสอบปัจจัยขนาดตัวอย่าง ภายใต้วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ พบว่า ปัจจัยขนาดตัวอย่างมีผลต่ออัตราความถูกต้องของวิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 กล่าวคือ เมื่อขนาดตัวอย่างเพิ่มขึ้นเกือบทุกเงื่อนไข มีผลทำให้อัตราความถูกต้องของวิธีการตรวจสอบทั้งสามวิธีมีค่าเพิ่มขึ้น

ตัวอย่างเช่น การตรวจสอบในเงื่อนไขขนาดตัวอย่าง 250 คนต่อกลุ่ม อัตราความถูกต้องของวิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ มีค่าประมาณ 27.21%, 49.86% และ 48.60% ตามลำดับ ส่วนการตรวจสอบในเงื่อนไขขนาดตัวอย่าง 2,000 คนต่อกลุ่ม อัตราความถูกต้องของสามวิธีมีค่าประมาณ 62.53%, 97.76% และ 99.84% ตามลำดับ ดังนั้น ถ้าขนาดตัวอย่างเพิ่มขึ้น 1,750 คนต่อกลุ่ม แล้วอัตราความถูกต้องของวิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ มีค่าเพิ่มขึ้นประมาณ 35.32%, 47.90% และ 51.24% ตามลำดับ

ส่วนผลการทดสอบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ปัจจัยขนาดตัวอย่างมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 กล่าวคือ เมื่อขนาดตัวอย่างเพิ่มขึ้นมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสท์ ลดลงในบางเงื่อนไข แต่มีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับเพิ่มขึ้นในทุกเงื่อนไข ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นในบางเงื่อนไข

ตัวอย่างเช่น การตรวจสอบในเงื่อนไขขนาดตัวอย่าง 250 ต่อกลุ่ม วิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ มีอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณ 7.60%, 8.36% และ 7.47% ตามลำดับ ส่วนในเงื่อนไขขนาดตัวอย่าง 2,000 ต่อกลุ่ม ทั้งสามวิธีดังกล่าวมีอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณ 3.08%, 28.87% และ 11.77% ตามลำดับ ดังนั้น ถ้าขนาดตัวอย่างเพิ่มขึ้น 1,750 คนต่อกลุ่ม มีผลทำให้วิธีโพลี-ชิปเทสต์มีอัตราความคลาดเคลื่อนประเภทที่ 1 ลดลง ประมาณ 4.52% ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้น ประมาณ 20.51% และ 4.30% ตามลำดับ

2.5 ผลการทดสอบปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่ย่อยกับปัจจัยสัดส่วนของข้อสอบทำหน้าที่ย่อย

ผลการทดสอบปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่ย่อยกับปัจจัยสัดส่วนของข้อสอบทำหน้าที่ย่อย ภายใต้วิธีการตรวจสอบการทำหน้าที่ย่อยของข้อสอบ พบว่า ปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่ย่อยกับปัจจัยสัดส่วนของข้อสอบทำหน้าที่ย่อย มีผลต่ออัตราความถูกต้องของวิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่ออัตราความถูกต้องของวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ กล่าวคือ การตรวจสอบในข้อสอบทำหน้าที่ย่อยที่เป็นรูปแบบเดียวกัน เมื่อสัดส่วนของข้อสอบทำหน้าที่ย่อยเพิ่มขึ้นในช่วง 10% ถึง 30% มีผลให้อัตราความถูกต้องของวิธีโพลี-ชิปเทสต์ลดลงในบางเงื่อนไข และมีผลให้อัตราความถูกต้องของวิธีการถดถอยโลจิสติกแบบจัดอันดับลดลงในทุกเงื่อนไข แต่ไม่มีผลต่อวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ส่วนการตรวจสอบในข้อสอบทำหน้าที่ย่อยที่ไม่เป็นรูปแบบเดียวกัน ปริมาณสัดส่วนของข้อสอบทำหน้าที่ย่อยในช่วง 10% ถึง 30% ไม่มีผลต่ออัตราความถูกต้องของทั้งสามวิธี

ตัวอย่างเช่น การตรวจสอบในเงื่อนไขข้อสอบทำหน้าที่ย่อยที่เป็นรูปแบบเดียวกัน และมีสัดส่วนของข้อสอบทำหน้าที่ย่อย 10% อัตราความถูกต้องของวิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าประมาณ 81.17%, 83.79% และ 79.04% ตามลำดับ ส่วนการตรวจสอบในเงื่อนไขข้อสอบทำหน้าที่ย่อยที่ไม่เป็นรูปแบบเดียวกัน และมีสัดส่วนของข้อสอบทำหน้าที่ย่อย 30% ทั้งสามวิธีมีอัตราความ-

ถูกต้องประมาณ 72.17%, 70.28% และ 79.35% ตามลำดับ ดังนั้น ในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน เมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนลดลง 20% มีผลทำให้อัตราความถูกต้องของวิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับเพิ่มขึ้น ประมาณ 9.00% และ 13.51% ตามลำดับ ส่วนอัตราความถูกต้องของวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติลดลงเล็กน้อย ประมาณ 0.31%

ส่วนผลการทดสอบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ปฏิสัมพันธ์สองทางของปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ กล่าวคือ การตรวจสอบในข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน เมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้นในช่วง 10% ถึง 30% มีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับเพิ่มขึ้น แต่ไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ส่วนการตรวจสอบในข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ปริมาณสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนในช่วง 10% ถึง 30% ไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งสามวิธี

ตัวอย่างเช่น การตรวจสอบในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และมีสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% อัตราความถูกต้องของวิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ มีค่าประมาณ 20.00%, 79.38% และ 82.92% ตามลำดับ ส่วนการตรวจสอบในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และมีสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% ทั้งสามวิธีมีอัตราความถูกต้องประมาณ 20.62%, 81.53% และ 81.81% ตามลำดับ ดังนั้น ในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน เมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนลดลง 20% มีผลทำให้อัตราความถูกต้องของวิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับเพิ่มขึ้นเล็กน้อย ประมาณ 0.79% และ 0.03% ตามลำดับ ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องลดลงเล็กน้อย ประมาณ 0.04%

2.6 ผลการทดสอบปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของข้อสอบทำ หน้าที่เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถ

ผลการทดสอบปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถ ภายใต้วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ พบว่า ปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถ มีผลต่ออัตราความถูกต้องของวิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่ออัตราความถูกต้องของวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ กล่าวคือ การตรวจสอบในข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน เมื่อกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถเพิ่มขึ้นในช่วง 0.0SD ถึง 1.0SD มีผลทำให้วิธีโพลี-ชิปเทสต์มีอัตราความถูกต้องเพิ่มขึ้นในทุกเงื่อนไข และมีผลทำให้วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องลดลงบางเงื่อนไข แต่ไม่มีผลต่ออัตราความถูกต้องของวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ส่วนการตรวจสอบในข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน กลุ่มผู้สอบที่มีความแตกต่างของการแจกแจงความสามารถในช่วง 0.0SD ถึง 1.0SD ไม่มีผลต่ออัตราความถูกต้องของทั้งสามวิธี

ตัวอย่างเช่น การตรวจสอบในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน เมื่อไม่มีความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบ อัตราความถูกต้องของวิธีโพลี-ชิปเทสต์วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ มีค่าประมาณ 6.53%, 83.82% และ 82.40% ตามลำดับ เมื่อตรวจสอบในข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน และมีความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่ม 1.0SD (จาก 0.0SD ถึง 1.0SD) อัตราความถูกต้องของสามวิธีดังกล่าวมีค่าประมาณ 39.41%, 10.17% และ 1.67% ตามลำดับ ดังนั้น ในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน เมื่อความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มเพิ่มขึ้น 1.0SD แล้ว อัตราความถูกต้องของวิธีโพลี-ชิปเทสต์เพิ่มขึ้น ประมาณ 32.88% แต่อัตราความถูกต้องของวิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าลดลง ประมาณ 10.17% ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องลดลงเล็กน้อย ประมาณ 1.67%

ผลการทดสอบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถ มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ กล่าวคือ ในข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน หรือข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน เมื่อความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบเพิ่มขึ้น 1.0SD (จาก 0.0 SD ถึง 1.0SD) มีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับเพิ่มขึ้น แต่ไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ

ตัวอย่างเช่น การตรวจสอบในเงื่อนไขข้อสอบที่ทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน เมื่อไม่มีความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบ อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ มีค่าประมาณ 1.55%, 14.07% และ 6.70% ตามลำดับ เมื่อตรวจสอบในข้อสอบที่ทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถ 1.0SD (จาก 0.0SD ถึง 1.0SD) อัตราความคลาดเคลื่อนประเภทที่ 1 ของสามวิธีมีค่าประมาณ 4.08%, 32.25% และ 11.99% ตามลำดับ ดังนั้น ในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน เมื่อความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบเพิ่มขึ้น 1.0SD มีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับเพิ่มขึ้น ประมาณ 18.18% ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติเพิ่มขึ้นเล็กน้อย ประมาณ 2.53% และ 5.29% ตามลำดับ

2.7 ผลการทดสอบปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยขนาดตัวอย่าง

ผลการทดสอบปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยขนาดตัวอย่าง ภายใต้วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ปรากฏว่า ปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยขนาดตัวอย่าง

มีผลต่ออัตราความถูกต้องของวิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 กล่าวคือ การตรวจสอบในข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน หรือข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน เมื่อขนาดตัวอย่างเพิ่มขึ้นในช่วง 250 ถึง 2,000 คนต่อกลุ่ม มีผลทำให้ อัตราความถูกต้องของวิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติเพิ่มขึ้นเกือบทุกเงื่อนไข ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องเพิ่มขึ้นในทุกเงื่อนไข

ตัวอย่างเช่น การตรวจสอบในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และมีขนาดตัวอย่าง 250 คนต่อกลุ่ม อัตราความถูกต้องของวิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ มีค่าประมาณ 37.52%, 43.05% และ 45.20% ตามลำดับ ส่วนการตรวจสอบในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และมีขนาดตัวอย่าง 2,000 คนต่อกลุ่ม อัตราความถูกต้องของสามวิธีมีค่าประมาณ 99.78%, 99.60% และ 99.69% ตามลำดับ ดังนั้น ในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน เมื่อขนาดตัวอย่างเพิ่มขึ้น 1,750 คนต่อกลุ่ม มีผลทำให้วิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ มีอัตราความถูกต้องเพิ่มขึ้นประมาณ 62.26%, 56.55% และ 54.49% ตามลำดับ

สำหรับผลการทดสอบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยขนาดตัวอย่าง มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ และวิธีโพลี-ชิปเทสต์ กล่าวคือ การตรวจสอบในข้อสอบที่ทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน หรือข้อสอบที่ทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน เมื่อขนาดตัวอย่างเพิ่มขึ้นในช่วง 250 ถึง 2,000 คนต่อกลุ่ม มีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับเพิ่มขึ้นในทุกเงื่อนไข แต่ไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ

ตัวอย่างเช่น การตรวจสอบในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และมีขนาดตัวอย่าง 250 คนต่อกลุ่ม อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ มีค่าประมาณ 6.29%, 9.17% และ 7.33% ตามลำดับ ส่วนการตรวจสอบในเงื่อนไขข้อสอบทำหน้าที่

เบี่ยงเบนเป็นรูปแบบเดียวกัน และมีขนาดตัวอย่าง 2,000 คนต่อกลุ่ม ทั้งสามวิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณ 0.66%, 36.29% และ 11.29% ตามลำดับ ดังนั้น ในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน เมื่อขนาดตัวอย่างเพิ่มขึ้น 1,750 คนต่อกลุ่ม มีผลทำให้วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้น ประมาณ 27.12% ส่วนวิธีโพลี-ชิปเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 ลดลงเล็กน้อย ประมาณ 5.63% ในขณะที่วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นเล็กน้อย ประมาณ 3.96%

2.8 ผลการทดสอบปฏิสัมพันธ์สองทางระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถ

ผลการทดสอบปฏิสัมพันธ์สองทางระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถ ภายใต้วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ปรากฏว่า ปฏิสัมพันธ์สองทางระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถ มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับ กล่าวคือ ในแบบทดสอบที่มีสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10%, 20% หรือ 30% เมื่อความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบเพิ่มขึ้น 1.0SD (จาก 0.0SD ถึง 1.0SD) มีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับเพิ่มขึ้น แต่ไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ

ตัวอย่างเช่น ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% เมื่อไม่มีความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบ วิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ มีอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณ 4.44%, 5.91% และ 7.16% ตามลำดับ ส่วนในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% ที่มีกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถ 1.0SD (จาก 0.0SD ถึง 1.0SD) อัตราความคลาดเคลื่อนประเภทที่ 1 ของสามวิธีมีค่าประมาณ 7.12%, 28.86% และ 12.23% ตามลำดับ ดังนั้น ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% เมื่อกลุ่มผู้สอบ

มีความแตกต่างของการแจกแจงความสามารถเพิ่มขึ้น 1.0SD มีผลทำให้วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้น ประมาณ 22.95% ส่วนวิธีโพลี-ชิปเทสท์และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิตีมีอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นเล็กน้อย ประมาณ 2.68% และ 5.07% ตามลำดับ

2.9 ผลการทดสอบปฏิสัมพันธ์สองทางระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยขนาดตัวอย่าง

ผลการทดสอบปฏิสัมพันธ์สองทางระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และปัจจัยขนาดตัวอย่าง ภายใต้วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ปรากฏว่าปฏิสัมพันธ์สองทางระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยขนาดตัวอย่าง มีผลต่ออัตราความถูกต้องของวิธีโพลี-ชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่อวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิตี กล่าวคือ ในแบบทดสอบที่มีสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10%, 20% และ 30% เมื่อขนาดตัวอย่างเพิ่มขึ้นในช่วง 250 ถึง 1,000 คนต่อกลุ่ม มีผลทำให้วิธีโพลี-ชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องเพิ่มขึ้นเกือบทุกเงื่อนไข

ตัวอย่างเช่น การตรวจสอบด้วยวิธีโพลี-ชิปเทสท์ ในเงื่อนไขปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% และมีขนาดตัวอย่าง 500, 1,000 และ 2,000 คนต่อกลุ่มมีอัตราความถูกต้องประมาณ 51.33% 59.58% และ 60.42% ตามลำดับ แสดงว่าเมื่อขนาดตัวอย่างเพิ่มขึ้น 500 คน (จาก 500 เป็น 1,000 คนต่อกลุ่ม) มีอัตราความถูกต้องเพิ่มขึ้นประมาณ 8.25% แต่ถ้ามีขนาดตัวอย่างเพิ่มขึ้น 1,500 (จาก 500 เป็น 2,000 คนต่อกลุ่ม) มีอัตราความถูกต้องเพิ่มขึ้นประมาณ 9.09%

ส่วนผลการทดสอบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ปฏิสัมพันธ์สองทางระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และปัจจัยขนาดตัวอย่าง มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิตี กล่าวคือ ในแบบทดสอบที่มีสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10%, 20% และ 30% เมื่อขนาดตัวอย่างเพิ่มขึ้นในช่วง 250 ถึง 2,000 คนต่อกลุ่ม มีผลทำให้วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นเกือบทุกเงื่อนไข แต่ไม่มีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิตีเพิ่มขึ้นหรือลดลง

ตัวอย่างเช่น การตรวจสอบในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10%, 20% และ 30% เมื่อใช้ขนาดตัวอย่าง 250 คนต่อกลุ่ม วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณ 7.18%, 8.74% และ 9.18% ตามลำดับ แต่ถ้าใช้ขนาดตัวอย่าง 2,000 คนต่อกลุ่ม อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าประมาณ 23.20%, 27.91% และ 35.49% ตามลำดับ ดังนั้น ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10%, 20% และ 30% เมื่อขนาดตัวอย่างเพิ่มขึ้น 1,750 คนต่อกลุ่ม มีผลทำให้วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้น ประมาณ 16.02%, 19.17% และ 26.31% ตามลำดับ

2.10 ผลการทดสอบปฏิสัมพันธ์สองทางระหว่างปัจจัยความแตกต่างของการแจกแจงความสามารถ กับปัจจัยขนาดตัวอย่าง

ผลการทดสอบปฏิสัมพันธ์สองทางระหว่างปัจจัยความแตกต่างของการแจกแจงความสามารถ กับปัจจัยขนาดตัวอย่าง ภายใต้วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ พบว่า ปฏิสัมพันธ์สองทางระหว่างปัจจัยความแตกต่างของการแจกแจงความสามารถ และปัจจัยขนาดตัวอย่าง มีผลต่ออัตราความถูกต้องของวิธีโพลี-ชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับ อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่อวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ กล่าวคือ ในเงื่อนไขกลุ่มผู้สอบที่มีความแตกต่างของการแจกแจงความสามารถ 0.0SD, 0.5SD และ 1.0SD เมื่อขนาดตัวอย่างเพิ่มขึ้นในช่วง 250 ถึง 1,000 คนต่อกลุ่ม มีผลทำให้วิธีโพลี-ชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องเพิ่มขึ้นเกือบทุกเงื่อนไข

ตัวอย่างเช่น การตรวจสอบด้วยวิธีโพลี-ชิปเทสท์ ในเงื่อนไขกลุ่มผู้สอบที่มีความแตกต่างของการแจกแจงความสามารถ 0.0SD และมีขนาดตัวอย่าง 500, 1,000 และ 2,000 คนต่อกลุ่มมีอัตราความถูกต้องประมาณ 41.32% 52.47% และ 52.47% ตามลำดับ แสดงว่าเมื่อขนาดตัวอย่างเพิ่มขึ้น 500 คน (จาก 500 เป็น 1,000 คนต่อกลุ่ม) มีอัตราความถูกต้องเพิ่มขึ้นประมาณ 11.15% และถ้ามีขนาดตัวอย่างเพิ่มขึ้น 1,500 คน (จาก 500 เป็น 2,000 คนต่อกลุ่ม) มีอัตราความถูกต้องเพิ่มขึ้นเท่าเดิม

ส่วนผลการทดสอบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ปฏิสัมพันธ์สองทางระหว่างปัจจัยความแตกต่างของการแจกแจงความสามารถ กับปัจจัยขนาดตัวอย่าง มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งสามวิธีอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 กล่าวคือ ในกลุ่มผู้สอบที่มีความแตกต่างของการแจกแจงความสามารถ 0.0SD, 0.5SD หรือ 1.0SD เมื่อ

ขนาดตัวอย่างเพิ่มขึ้นในช่วง 250 ถึง 2,000 คนต่อกลุ่ม มีผลทำให้วิธีโพลี-ชิปเทสต์มีอัตราความคลาดเคลื่อนประเภทที่ 1 ลดลงในบางเงื่อนไข และมีผลทำให้วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นเกือบทุกเงื่อนไข ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นในบางเงื่อนไข

ตัวอย่างเช่น ในกลุ่มผู้สอบที่มีความแตกต่างของการแจกแจงความสามารถ 1.0SD และมีขนาดตัวอย่าง 250 คนต่อกลุ่ม วิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณ 11.55%, 12.81% และ 8.48% ตามลำดับ และในกลุ่มผู้สอบที่มีความแตกต่างของการแจกแจงความสามารถเท่าเดิม แต่มีขนาดตัวอย่าง 2,000 คนต่อกลุ่ม อัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งสามวิธีมีค่าประมาณ 3.04%, 55.11% และ 18.35% ตามลำดับ ดังนั้น ในเงื่อนไขกลุ่มผู้สอบที่มีความแตกต่างของการแจกแจงความสามารถ 1.0SD เมื่อขนาดตัวอย่างเพิ่มขึ้น 1,750 คนต่อกลุ่ม มีผลทำให้วิธีโพลี-ชิปเทสต์มีอัตราความคลาดเคลื่อนประเภทที่ 1 ลดลง ประมาณ 8.13% ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้น ประมาณ 41.25% และ 10.06% ตามลำดับ

3. ผลการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยที่ศึกษา

การสรุปผลการวิจัยในตอนนี้เป็นผลการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ผลการเปรียบเทียบภายใต้ทุกเงื่อนไขสามารถสรุปได้ดังนี้

ผลการเปรียบเทียบอัตราความถูกต้อง ระหว่างวิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบ และขนาดตัวอย่าง รวมทั้งปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัยดังกล่าว พบว่า ภายใต้เงื่อนไข

ส่วนมาก วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ มีอัตราความถูกต้องใกล้เคียงกัน ทั้งสองวิธีมีอัตราความถูกต้องสูงกว่าวิธีโพลี-ชิปเทสท์อย่างมีนัยสำคัญทางสถิติที่ระดับ .001

ส่วนผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ภายใต้เกือบทุกเงื่อนไข วิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ มีอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ

เมื่อพิจารณาค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ทุกเงื่อนไข พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยของอัตราความถูกต้องประมาณ 79% และมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณ 17% ในขณะที่วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณ 81% และ 9% ตามลำดับ จะเห็นว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องเพียงพอ (Sufficient power) ตามเกณฑ์ของโคเฮน (Cohen, 1988: 5) ที่กำหนดอัตราความถูกต้องมากกว่า 80% ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องใกล้เคียงกับเกณฑ์ดังกล่าว ถึงแม้ว่าทั้งสองวิธีจะมีอัตราความถูกต้องใกล้เคียงกัน แต่เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 ของสองวิธีพบว่ามีค่าแตกต่างกัน กล่าวคือ วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ โดยเฉลี่ยประมาณ 8% ดังนั้นสามารถสรุปได้ว่า **ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีประสิทธิภาพในการตรวจสอบสูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ**

ส่วนผลการตรวจสอบของวิธีโพลี-ชิปเทสท์ พบว่า มีค่าเฉลี่ยของอัตราความถูกต้องประมาณ 49% ซึ่งมีค่าต่ำกว่าเกณฑ์ของโคเฮนค่อนข้างมาก เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า เกือบทุกเงื่อนไขมีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ภายในเกณฑ์ที่กำหนด (Nominal limits) โดยมีค่าเฉลี่ยประมาณ 4% ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าขอบเขตที่กำหนดประมาณ 4% ผลดังกล่าวชี้ให้เห็นว่า **ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า วิธีโพลี-ชิปเทสท์มีความผิดพลาดในระดับที่ยอมรับได้ แต่ยังมีความถูกต้องไม่เพียงพอในการตรวจสอบ ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ**

ถึงแม้ว่ามีความถูกต้องเพียงพอ แต่ยังคงมีความคลาดเคลื่อนประเภทที่ 1 เกินกว่าระดับที่กำหนด นั่นคือ ยังไม่สามารถสรุปได้อย่างชัดเจนว่าวิธีใดมีประสิทธิภาพในการตรวจสอบดีกว่ากัน ดังนั้นจำเป็นต้องปรับแก้ (Adjustment) ระดับนัยสำคัญของวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติที่ใช้ในการตัดสินใจการทำหน้าที่เบี่ยงเบนของข้อสอบให้มีค่าลดลง เพื่อให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าลดลงในระดับที่ยอมรับได้ แล้วจึงพิจารณาอัตราความถูกต้องอีกครั้งหนึ่ง

ในการศึกษาครั้งนี้ใช้ระดับนัยสำคัญ .05 ตัดสินการทำหน้าที่เบี่ยงเบนของข้อสอบ ดังนั้นผู้วิจัยจึงปรับระดับนัยสำคัญให้ลดลง 4 ระดับ คือ .04, .03, .02 และ .01 แล้วนำข้อมูลเดิมมาตรวจสอบใหม่โดยใช้ระดับนัยสำคัญดังกล่าว ผลการตรวจสอบพบว่า ค่าเฉลี่ยของอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีค่าลดลงอย่างต่อเนื่อง ตามระดับนัยสำคัญที่ลดลง เมื่อเปรียบเทียบผลการตัดสินใจการทำหน้าที่เบี่ยงเบนของข้อสอบที่ระดับนัยสำคัญ .02 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ กับที่ระดับนัยสำคัญ .05 ของวิธีโพลี-ชิปเทสท์ พบว่า โดยภาพรวมทั้งสองวิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 ใกล้เคียงกัน (ประมาณ 4%) แต่วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องสูงกว่าวิธีโพลี-ชิปเทสท์ ประมาณ 25% **นั่นแสดงว่า ที่ระดับนัยสำคัญดังกล่าว วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีประสิทธิภาพสูงกว่าวิธีโพลี-ชิปเทสท์ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า**

ในการนำวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ไปใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า จะต้องกำหนดระดับนัยสำคัญให้เหมาะสมกับขนาดตัวอย่าง ตลอดจนความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบ เพื่อให้อัตราความคลาดเคลื่อนประเภทที่ 1 ที่เกิดขึ้นจริงมีค่าอยู่ภายในขอบเขตที่กำหนด จากผลการปรับระดับนัยสำคัญ พบว่า ขนาดตัวอย่าง 500, 1,000 และ 2,000 คนต่อกลุ่ม สามารถใช้ระดับนัยสำคัญ .03, .023 และ .014 ตามลำดับ ส่วนขนาดตัวอย่าง 250 คนต่อกลุ่ม ไม่ได้นำมาปรับระดับนัยสำคัญ เพราะมีขนาดเล็กเกินไป ซึ่งมีอัตราความถูกต้องค่อนข้างต่ำ สำหรับความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบ 0.0SD, 0.5SD และ 1.0SD สามารถใช้ระดับนัยสำคัญ .035, .03 และ .014 ตามลำดับ

สรุปผลการวิจัยในภาพรวม

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่าง สามารถสรุปโดยภาพรวมได้ดังนี้

1. ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนมีผลต่ออัตราความถูกต้องของวิธีโพลี-ชิปเทสต์ แต่ไม่มีผลต่อวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ กล่าวคือ วิธีโพลี-ชิปเทสต์สามารถตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกัน แต่ไม่สามารถตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ไม่เป็นรูปแบบเดียวกัน ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติสามารถตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ทั้งที่เป็นรูปแบบเดียวกันและที่ไม่เป็นรูปแบบเดียวกัน

เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนมีผลต่อวิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับ แต่ไม่มีผลต่อวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ เมื่อตรวจสอบในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน วิธีโพลี-ชิปเทสต์มีอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้น ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 ลดลง เมื่อเทียบกับการตรวจสอบในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน

2. ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนมีผลต่ออัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับ แต่ไม่มีผลต่อวิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ เมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้น 20% (จาก 10% ถึง 30%) มีผลทำให้อัตราความถูกต้องของวิธีการถดถอยโลจิสติกแบบจัดอันดับลดลงประมาณ 6% และมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นประมาณ 6%

3. ปัจจัยความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบมีผลต่ออัตราความถูกต้องของวิธีโพลี-ชิปเทสต์ แต่ไม่มีผลต่อวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ เมื่อกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถเพิ่มขึ้นในช่วง 0.0SD ถึง 1.0SD มีผลทำให้วิธีโพลี-ชิปเทสต์มีอัตราความถูกต้องเพิ่มขึ้นในช่วง 5% ถึง 16%

เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ปัจจัยความแตกต่างของการแจกแจงความสามารถมีผลต่อทั้งสามวิธี โดยมีผลต่อวิธีการถดถอยโลจิสติกแบบจัดอันดับมากที่สุด เมื่อกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถเพิ่มขึ้น 1.0SD (จาก 0.0SD ถึง 1.0SD) มีผลทำให้วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นประมาณ 20%

4. ปัจจัยขนาดตัวอย่างมีผลต่ออัตราความถูกต้องของทั้งสามวิธี เมื่อขนาดตัวอย่างเพิ่มขึ้นในทุกเงื่อนไข (จาก 250 ถึง 2,000 คนต่อกลุ่ม) มีผลทำให้อัตราความถูกต้องของทั้งสามวิธีเพิ่มขึ้น ยกเว้นเมื่อเพิ่มขนาดตัวอย่างจาก 1,000 ถึง 2,000 คนต่อกลุ่มของวิธีโพลี-ชิปเทสต์และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ

เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ปัจจัยขนาดตัวอย่างมีผลต่อทั้งสามวิธี โดยมีผลต่อวิธีการถดถอยโลจิสติกแบบจัดอันดับมากที่สุด เมื่อขนาดตัวอย่างเพิ่มขึ้นในทุกเงื่อนไข มีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นในช่วง 4% ถึง 21%

5. ปฏิสัมพันธ์ระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน มีผลต่ออัตราความถูกต้องของวิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับ แต่ไม่มีผลต่อวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ภายใต้ปัจจัยข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน เมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้น 20% (จาก 10% ถึง 30%) มีผลทำให้วิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องลดลงประมาณ 9% และ 14% ตามลำดับ

เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ปฏิสัมพันธ์ระหว่างปัจจัยดังกล่าวมีผลต่อวิธีการถดถอยโลจิสติกแบบจัดอันดับ แต่ไม่มีผลต่อวิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ภายใต้ปัจจัยข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน เมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้นในช่วง 10% ถึง 30% มีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับเพิ่มขึ้นในช่วง 5% ถึง 12%

6. ปฏิสัมพันธ์ระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบ มีผลต่ออัตราความถูกต้องของวิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับ แต่ไม่มีผลวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน เมื่อกลุ่มผู้สอบ

มีความแตกต่างของการแจกแจงความสามารถเพิ่มขึ้น 1.0SD (จาก 0.0SD ถึง 1.0SD) มีผลทำให้วิธีโพลี-ชิปเทสที่มีอัตราความถูกต้องเพิ่มขึ้น ประมาณ 33% ในขณะที่วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องลดลงประมาณ 10%

เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ปฏิสัมพันธ์ระหว่างปัจจัยดังกล่าวมีผลต่อวิธีการถดถอยโลจิสติกแบบจัดอันดับ แต่ไม่มีผลต่อวิธีโพลี-ชิปเทส และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และที่ไม่เป็นรูปแบบเดียวกัน เมื่อความแตกต่างของการแจกแจงความสามารถเพิ่มขึ้น 1.0SD (จาก 0.0SD ถึง 1.0SD) มีผลทำให้วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้น ประมาณ 19% และ 22% ตามลำดับ

7. ปฏิสัมพันธ์ระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยขนาดตัวอย่าง มีผลต่ออัตราความถูกต้องของทั้งสามวิธี ภายใต้ปัจจัยข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และที่ไม่เป็นรูปแบบเดียวกัน เมื่อขนาดตัวอย่างเพิ่มขึ้นในช่วง 250 ถึง 2,000 คนต่อกลุ่มมีผลทำให้อัตราความถูกต้องของทั้งสามวิธีเพิ่มขึ้นเกือบทุกเงื่อนไข

เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ปฏิสัมพันธ์ระหว่างปัจจัยดังกล่าวมีผลต่อวิธีการถดถอยโลจิสติกแบบจัดอันดับ แต่ไม่มีผลต่อวิธีโพลี-ชิปเทส และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ภายใต้ปัจจัยข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และที่ไม่เป็นรูปแบบเดียวกัน เมื่อขนาดตัวอย่างเพิ่มขึ้นในช่วง 250 ถึง 2,000 คนต่อกลุ่ม มีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับเพิ่มขึ้นในทุกเงื่อนไข

8. ปฏิสัมพันธ์ระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถ ไม่มีผลต่ออัตราความถูกต้องของทั้งสามวิธี เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า มีผลต่อวิธีการถดถอยโลจิสติกแบบจัดอันดับ แต่ไม่มีผลต่อวิธีโพลี-ชิปเทส และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ภายใต้สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10%, 20% และ 30% เมื่อกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถเพิ่มขึ้น 1.0SD (จาก 0.0SD ถึง 1.0SD) มีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับเพิ่มขึ้นประมาณ 34%, 20% และ 18% ตามลำดับ

9. ปฏิสัมพันธ์ระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยขนาดตัวอย่าง มีผลต่ออัตราความถูกต้องของวิธีโพลี-ชิปเทส และวิธีการถดถอยโลจิสติกแบบจัดอันดับ แต่ไม่มีผลต่อวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ภายใต้สัดส่วนของข้อสอบทำหน้าที่

เบี่ยงเบน 10%, 20% หรือ 30% เมื่อขนาดตัวอย่างเพิ่มขึ้นในช่วง 250 ถึง 1,000 คนต่อกลุ่ม มีผลทำให้อัตราความถูกต้องเพิ่มขึ้นเกือบทุกเงื่อนไข

เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ปฏิสัมพันธ์ระหว่างปัจจัยดังกล่าว มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับ แต่ไม่มีผลต่อวิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ภายใต้ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10%, 20% หรือ 30% เมื่อขนาดตัวอย่างเพิ่มขึ้นในช่วง 250 ถึง 2,000 คนต่อกลุ่ม มีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับเพิ่มขึ้นในช่วง 5% ถึง 26%

10. ปฏิสัมพันธ์ระหว่างปัจจัยความแตกต่างของการแจกแจงความสามารถ กับปัจจัยขนาดตัวอย่างมีผลต่ออัตราความถูกต้องของวิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับ แต่ไม่มีผลต่อวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ภายใต้กลุ่มผู้สอบที่มีความแตกต่างของการแจกแจงความสามารถ 0.0SD, 0.5SD หรือ 1.0SD เมื่อขนาดตัวอย่างเพิ่มขึ้นในช่วง 250 ถึง 1,000 คนต่อกลุ่ม มีผลทำให้อัตราความถูกต้องเพิ่มขึ้นเกือบทุกเงื่อนไข

เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ปฏิสัมพันธ์ระหว่างปัจจัยดังกล่าว มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งสามวิธี โดยมีผลต่อวิธีการถดถอยโลจิสติกแบบจัดอันดับมากที่สุด ภายใต้กลุ่มผู้สอบที่มีความแตกต่างของการแจกแจงความสามารถ 0.0SD, 0.5SD หรือ 1.0SD เมื่อขนาดตัวอย่างเพิ่มขึ้นในช่วง 250 ถึง 2,000 คนต่อกลุ่ม มีผลทำให้วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นในช่วง 7% ถึง 41%

11. ผลการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติ และให้คะแนนหลายค่า ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบ และขนาดตัวอย่าง รวมทั้งปฏิสัมพันธ์สองทางระหว่าง 4 ปัจจัย พบว่า ภายใต้เงื่อนไขส่วนมาก วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ และวิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องใกล้เคียงกัน ทั้งสองวิธีมีอัตราความถูกต้องสูงกว่าวิธีโพลี-ชิปเทสต์ เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับ

และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิตีมีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณ 17% และ 9% ตามลำดับ ซึ่งมีค่าสูงกว่าขอบเขตที่กำหนด ส่วนวิธีโพลี-ชิปเทสท์มีค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ภายในขอบเขตที่กำหนด ภายใต้เกือบทุกเงื่อนไข (ประมาณ 4%)

อภิปรายผลการวิจัย

การวิจัยครั้งนี้มีเป้าหมายเพื่อตอบคำถามในสองประเด็นหลัก คือ (1) ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์ระหว่างปัจจัย มีผลต่อประสิทธิภาพในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิตี หรือไม่ และ (2) วิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิตี ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย และปฏิสัมพันธ์ระหว่างปัจจัย มีประสิทธิภาพในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบแตกต่างกันหรือไม่ จากผลการวิจัยที่สรุปมาข้างต้นมีประเด็นสำคัญที่นำมาอภิปรายผล ดังนี้

1. ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิตีและให้คะแนนหลายค่า พบว่า ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนมีผลต่ออัตราความถูกต้องของวิธีโพลี-ชิปเทสท์อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่อวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิตี ในขณะเดียวกัน มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่อวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิตี ผลดังกล่าวที่มีนัยสำคัญเป็นไปตามสมมติฐานข้อ 1-3 จะเห็นว่า ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนมีผลต่อประสิทธิภาพในการตรวจสอบของวิธีโพลี-ชิปเทสท์ค่อนข้างมาก กล่าวคือ ในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน วิธีโพลี-ชิปเทสท์มีอัตราความถูกต้องต่ำที่สุด ประมาณ 21% ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิตีมีอัตราความถูกต้องสูงใกล้เคียงกัน คือ ประมาณ 80% และ 82% ตามลำดับ ส่วนในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน ทั้งสามวิธีสามารถตรวจสอบได้ถูกต้องใกล้เคียงกัน โดยวิธีโพลี-ชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความ-

ถูกต้องประมาณ 77% ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องประมาณ 79% นอกจากนี้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนยังส่งผลกระทบต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับจะเห็นว่า เมื่อตรวจสอบในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน วิธีโพลี-ชิปเทสต์มีอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นประมาณ 3% ส่วนวิธีการถดถอยโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 ลดลงประมาณ 6% เมื่อเทียบกับการตรวจสอบในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน สำหรับวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความคลาดเคลื่อนประเภทที่ 1 ใกล้เคียงกันทั้งสองรูปแบบ

ผลดังกล่าว สอดคล้องกับผลการศึกษาของ ทองอยู่ สาระ (2543) วลีมาศ แซ่ฮึง (2543) โรเจอร์สและสวามินาทาน (Rogers; & Swaminathan. 1993) นารายานานและสวามินาทาน (Narayanan; & Swaminathan. 1994, 1996) วิทเมอร์และชูแม็คเกอร์ (Whitmore; & Schumacker. 1999) ที่ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถมิติเดียวและให้คะแนนสองค่า และผลการศึกษาของ ชาง มาซซีโอ และรูสโซ (Chang, Mazzeo, & Roussos. 1996) เฟรนช์และมิลเลอร์ (French; & Miller. 1996) คริสต์แจนส์สันและคนอื่นๆ (Krisjansson; et al. 2005) ที่ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถมิติเดียวและให้คะแนนหลายค่า ซึ่งพบว่า วิธีชิปเทสต์ และวิธีโพลี-ชิปเทสต์มีอัตราความถูกต้องสูงในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกัน แต่มีอัตราความถูกต้องต่ำในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ไม่เป็นรูปแบบเดียวกัน ส่วนวิธีที่ใช้โมเดลการถดถอยโลจิสติก (Continuation ratio logits model, Cumulative logits model, Adjacent categories logits model, Unconstrained cumulative logits ordinal model) สามารถตรวจสอบข้อสอบทำหน้าที่เบี่ยงเบนทั้งที่เป็นรูปแบบเดียวกันและที่ไม่เป็นรูปแบบเดียวกันได้อย่างมีประสิทธิภาพ

สาเหตุที่ทำให้วิธีโพลี-ชิปเทสต์ไม่มีประสิทธิภาพในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ไม่เป็นรูปแบบเดียวกัน เพราะว่าวิธีโพลี-ชิปเทสต์ปรับขยายมาจากวิธีชิปเทสต์ ซึ่งเป็นวิธีที่ใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกันโดยเฉพาะ (Li; & Stout. 1996: 650) โดยใช้สถิติแบบคิดเครื่องหมาย (เครื่องหมายบวกและเครื่องหมายลบ) ซึ่งคล้ายกับวิธีแมนเทิล-แฮนส์เซล วิธีแมนเทิล วิธีแมนเทิล-แฮนส์เซลทั่วไป และวิธีความแตกต่างเฉลี่ยมาตรฐาน สำหรับข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ในข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า สามารถพิจารณาได้จากฟังก์ชันผิวหน้าการตอบรายการ

(Option response surfaces; ORSs) ระหว่างกลุ่มอ้างอิงและกลุ่มสนใจตัดกันในช่วงระดับความสามารถของผู้สอบ ซึ่งคล้ายกับฟังก์ชันการตอบข้อสอบ (Item response functions; IRFs) ในข้อสอบที่วัดความสามารถมิติเดียวและให้คะแนนสองค่า ข้อมูลที่ใช้ในการศึกษาครั้งนี้ ผู้วิจัยออกแบบค่าพารามิเตอร์อำนาจจำแนกและพารามิเตอร์รายการของข้อสอบ เพื่อให้ฟังก์ชันผิวหน้าการตอบรายการตัดกันอย่างสมดุลง ในช่วงระดับความสามารถ θ_1 หรือ θ_2 รูปแบบของข้อสอบดังกล่าวทำให้ทิศทางของข้อสอบทำหน้าที่เบี่ยงเบนเปลี่ยนแปลง ณ ตรงช่วงตัดกัน ซึ่งมีผลทำให้ความแตกต่างของคะแนนส่วนหนึ่งที่เป็นค่าลบ กับคะแนนอีกส่วนหนึ่งที่เป็นค่าบวกหักล้างกัน (Cancellation) (Narayanan; & Swaminathan. 1996: 271) ดังนั้นวิธีโพลี-ชิปเทสที่จึงไม่สามารถตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน

ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ปรับขยายมาจากวิธีการถดถอยโลจิสติก (Swaminathan; & Rogers. 1990) สามารถใช้โมเดลทดสอบผลของปฏิสัมพันธ์ระหว่างระดับความสามารถ กับการเป็นสมาชิกของกลุ่มผู้สอบ ทำให้สามารถตรวจสอบได้ทั้งข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และที่ไม่เป็นรูปแบบเดียวกัน (Clauser; & Mazor. 1998: 35; Narayanan; & Swaminathan. 1996: 257; Rogers; & Swaminathan, 1993: 105) ถึงแม้ว่าทั้งสองวิธีจะมีอัตราความถูกต้องใกล้เคียงกัน แต่วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ดีกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ เพราะว่าข้อมูลที่ใช้ศึกษาครั้งนี้เป็นข้อมูลที่วัดความสามารถสองมิติ วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติใช้ค่าประมาณความสามารถทั้งสองมิติเป็นเกณฑ์การจับคู่ (Matching criterion) ทำให้ข้อมูลเหมาะสมกับโมเดลที่ใช้วิเคราะห์ จึงสามารถตรวจสอบได้อย่างถูกต้องและแม่นยำ ซึ่งต่างจากวิธีการถดถอยโลจิสติกแบบจัดอันดับที่ใช้คะแนนรวมของแบบทดสอบเป็นเกณฑ์การจับคู่ ทำให้ผลการตรวจสอบไม่คงที่ เพราะละเมิดข้อตกลงความเป็นมิติเดียว ผลการตัดสินใจข้อสอบทำหน้าที่เบี่ยงเบนจึงมีความผิดพลาดสูง ซึ่งส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 สูงเกินปกติ

ดังนั้น ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีประสิทธิภาพในการตรวจสอบทั้งข้อสอบที่ทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และข้อสอบที่ทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ส่วนวิธีโพลี-ชิปเทสที่มีประสิทธิภาพในการตรวจสอบเฉพาะข้อสอบที่ทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน

2. ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า พบว่า ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนมีผลต่ออัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่อวิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ผลดังกล่าวที่มีนัยสำคัญเป็นไปตามสมมติฐานข้อ 1-3 จะเห็นว่า ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนมีผลกระทบต่อประสิทธิภาพในการตรวจสอบของวิธีการถดถอยโลจิสติกแบบจัดอันดับ กล่าวคือ เมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้น 20% (จาก 10% ถึง 30%) มีผลทำให้อัตราความถูกต้องลดลงในช่วง 4% ถึง 6% และมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นในช่วง 2% ถึง 6% ส่วนอีกสองวิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 ค่อนข้างคงที่ วิธีโพลี-ชิปเทสต์สามารถควบคุมได้ดีที่สุด โดยมีอัตราความคลาดเคลื่อนประเภทที่ 1 ในช่วง 4% ถึง 5% ซึ่งมีค่าอยู่ในขอบเขตที่กำหนด (Nominal limits) ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณ 9% ซึ่งมีค่าสูงกว่าขอบเขตที่กำหนดไม่มาก

ผลการศึกษาวิธีโพลี-ชิปเทสต์ดังกล่าว สอดคล้องกับผลการศึกษาของ วลีมาศ แซ่อึ้ง (2543) เกียร์ล กอทท์แมนน์ และบอทตัน (Gierl; Gotzmann; & Boughton. 2004) ที่พบว่า สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนในช่วง 10% ถึง 20% ไม่มีผลต่อวิธีชิปเทสต์ ส่วนผลการศึกษาวิธีการถดถอยโลจิสติกแบบจัดอันดับสอดคล้องกับผลการศึกษาของนารายานานและสวามินาทาน (Narayanan; & Swaminathan. 1996) ที่พบว่า เมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้นจาก 10% ถึง 20% มีผลทำให้วิธีการถดถอยโลจิสติกมีอัตราความถูกต้องลดลง และมีอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้น นอกจากนี้ยังสอดคล้องกับผลการศึกษาของ โอชิม่าและมิลเลอร์ (Oshima; & Miller. 1992) ที่พบว่า ในข้อสอบวัดความสามารถหลายมิติ เมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้น มีผลทำให้วิธีการวัดพื้นที่แบบคิดเครื่องหมาย (SA) วิธีการวัดพื้นที่แบบไม่คิดเครื่องหมาย (UA) วิธีผลรวมของกำลังสองแบบคิดเครื่องหมาย (SSOS) และวิธีผลรวมของกำลังสองแบบไม่คิดเครื่องหมาย (USOS) มีอัตราความถูกต้องลดลง

เมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้นมีผลทำให้อัตราความถูกต้องของวิธีการถดถอยโลจิสติกแบบจัดอันดับลดลงมีสาเหตุมาจาก สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนที่เพิ่มขึ้นมีผลทำให้ค่าประมาณความสามารถมีความเชื่อมั่นต่ำลง ซึ่งจะมีผลทำให้เกณฑ์การจับคู่

ความสามารถ (Matching criterion) ที่ใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบขาดความแม่นยำ ดังนั้นจึงมีผลทำให้อัตราความถูกต้องของวิธีการตรวจสอบลดลง (Narayanan; & Swaminathan. 1994: 318) ปัจจัยดังกล่าวไม่มีผลกระทบต่ออัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลิ-ชิปเทสท์เพราะว่า วิธีโพลิ-ชิปเทสท์ใช้คะแนนเกณฑ์การจับคู่คำนวณแยกจากข้อสอบที่ศึกษา โดยถือว่าเกณฑ์การจับคู่ประมาณค่ามาจากข้อสอบที่ทำหน้าที่ไม่เบี่ยงเบน ส่วนข้อสอบที่ศึกษาเป็นข้อสอบที่สงสัยว่าทำหน้าที่เบี่ยงเบน ดังนั้นเกณฑ์การจับคู่จึงมีความบริสุทธิ์ (Purification) นอกจากนี้ยังใช้โปรแกรม DIFPACK ที่มีประสิทธิภาพในการวิเคราะห์สามารถวิเคราะห์แบบทวนซ้ำ (Iterative algorithm) โดยนำข้อสอบที่ทำหน้าที่เบี่ยงเบนออกจากคะแนนเกณฑ์ จนกระทั่งได้เกณฑ์การจับคู่เป็นอิสระจากข้อสอบทำหน้าที่เบี่ยงเบน (DIF-free) (Clauser; & Mazor. 1998: 35) ดังนั้นจึงสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ที่สูงเกินปกติ ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับ ซึ่งตามปกติใช้ในการตรวจสอบข้อสอบที่วัดความสามารถมิติเดียวและให้คะแนนหลายค่า (Zumbo. 1999) ไม่มีกระบวนการวิเคราะห์ที่ทำให้เกณฑ์การจับคู่มีความบริสุทธิ์ อีกทั้งยังวิเคราะห์โดยใช้คะแนนรวมของแบบทดสอบแทนระดับความสามารถ ซึ่งแบบทดสอบที่วัดคุณลักษณะซับซ้อน (Complex traits) ไม่เหมาะสมที่จะใช้คะแนนรวมของแบบทดสอบแทนระดับความสามารถ เพราะจะทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้น (Mazor; Hambleton; & Clauser. 1998: 358) ด้วยเหตุผลดังกล่าวจึงทำให้วิธีการถดถอยโลจิสติกแบบจัดอันดับขาดความแม่นยำในการตรวจสอบ สำหรับวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติใช้ค่าประมาณความสามารถทั้งสองมิติเป็นเกณฑ์การจับคู่ ซึ่งเหมาะสมกับโมเดลที่ใช้วิเคราะห์ ถึงแม้ว่าสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้น ก็ไม่ส่งผลกระทบต่ออัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1

ดังนั้น ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนในช่วง 10% ถึง 30% ไม่มีผลต่อประสิทธิภาพของวิธีโพลิชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ

3. ปัจจัยความแตกต่างของการแจกแจงความสามารถ

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า พบว่า ปัจจัยความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบมีผลต่ออัตราความถูกต้องของวิธีโพลิ-ชิปเทสท์อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่อ

วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่ามีผลต่อทั้งสามวิธีอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ผลการทดสอบที่มีนัยสำคัญเป็นไปตามสมมติฐานที่กำหนดในข้อ 1-3 จะเห็นว่า ปัจจัยความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบมีผลต่ออัตราความถูกต้องของวิธีโพลี-ชิปเทสท์ กล่าวคือ เมื่อกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถเพิ่มขึ้นในช่วง 0.0SD ถึง 1.0SD มีผลทำให้วิธีโพลี-ชิปเทสท์มีอัตราความถูกต้องเพิ่มขึ้นในช่วง 5% ถึง 16% และการเพิ่มขึ้นของความแตกต่างดังกล่าว ส่งผลกระทบต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับค่อนข้างมาก โดยมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นประมาณ 20% แต่ส่งผลต่อวิธีโพลี-ชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติค่อนข้างน้อย ดังจะเห็นได้ว่าทั้งสองวิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นประมาณ 2% และ 5% ตามลำดับ

ผลการศึกษาวิธีโพลี-ชิปเทสท์ดังกล่าว สอดคล้องกับผลการศึกษาของ ชาง มาซซีโอ และ รุสโซ (Chang, Mazzeo, & Roussos. 1996) โบลท์ (Bolt. 2002) ที่พบว่า เมื่อการแจกแจงความสามารถระหว่างกลุ่มผู้สอบเพิ่มขึ้น 1.0SD มีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสท์เพิ่มขึ้นเล็กน้อย ส่วนผลการศึกษาวิธีการถดถอยโลจิสติกแบบจัดอันดับสอดคล้องกับผลการศึกษาของ นารายานานและสวามินาทาน (Narayanan; & Swaminathan. 1996) ที่พบว่า เมื่อกลุ่มผู้สอบมีการแจกแจงความสามารถเพิ่มขึ้น 1.0SD มีผลทำให้วิธีการถดถอยโลจิสติกมีอัตราความถูกต้องลดลง แต่อัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้น

ตามกรอบความคิดเชิงทฤษฎี เมื่อกลุ่มผู้สอบมีการแจกแจงความสามารถแตกต่างกัน มีผลกระทบต่อประสิทธิภาพของการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ทั้งนี้เพราะว่าความแตกต่างของผลการตอบข้อสอบ อันเนื่องมาจากความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มทั้งหมด เป็นผลกระทบของข้อสอบ (Item impact) มากกว่าการทำหน้าที่เบี่ยงเบนของข้อสอบ (Camilli. 1992: 131; Clauser; & Mazor. 1998: 31; Dorans; & Holland. 1993: 37) ดังนั้น ถ้าใช้เทคนิคการตรวจสอบภายใต้โมเดลแบบมิติเดียวจะทำให้ผลการตรวจสอบเกิดความผิดพลาด เพราะไม่สามารถแยกการทำหน้าที่เบี่ยงเบนของข้อสอบ ออกจากผลกระทบของข้อสอบ แต่วิธีโพลี-ชิปเทสท์ใช้เทคนิคการตรวจสอบภายใต้ทฤษฎีการตอบข้อสอบแบบหลายมิติสามารถคำนวณคะแนนเกณฑ์ที่ใช้ในการจับคู่แยกจากข้อสอบที่ศึกษา ผลการตอบข้อสอบถูกจึงเป็นความแตกต่างของความสามารถแทรกซ้อน (Nuisance ability) ซึ่งเป็นอิสระจากความสามารถ

เป้าหมาย (Target ability) ถึงแม้ว่าการแจกแจงความสามารถเป้าหมายระหว่างกลุ่มจะไม่เท่ากันก็สามารถใช้วิธีทางสถิติ คือ การปรับแก้ค่าการถดถอย (Regression correction) (Chang; Mazzeo; & Roussos. 1996: 338-339; Narayanan; Swaminathan. 1993: 327; Shealy; & Stout. 1993: 190-193) เพื่อลดความแตกต่างของค่าความสามารถเป้าหมายระหว่างกลุ่ม ทำให้สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ภายในขอบเขตที่กำหนด ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติถึงแม้ว่าไม่มีเทคนิคการปรับแก้ค่าการถดถอยดังกล่าว แต่ก็มิใช่ได้เปรียบตรงที่ใช้ค่าประมาณความสามารถทั้งสองมิติเป็นเกณฑ์การจับคู่ ทำให้เกณฑ์การจับคู่มีความแม่นยำ ซึ่งมีผลทำให้สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ที่สูงเกินปกติได้ดีกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับที่ใช้คะแนนรวมของแบบทดสอบ

ดังนั้น ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า เมื่อความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มเพิ่มขึ้นในช่วง 0.0SD ถึง 1.0SD มีผลทำให้อัตราความถูกต้องของวิธีโพลี-ชิปเทสท์เพิ่มขึ้น และสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ภายในขอบเขตที่กำหนด

4. ปัจจัยขนาดตัวอย่าง

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า พบว่า ปัจจัยขนาดตัวอย่างมีผลต่ออัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งสามวิธีอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ซึ่งเป็นไปตามสมมติฐานข้อ 1-3 จะเห็นว่า การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ เมื่อใช้ขนาดตัวอย่าง 250 คนต่อกลุ่มทั้งสามวิธีมีอัตราความถูกต้องต่ำสุด เมื่อเทียบกับขนาดตัวอย่างอื่นๆ โดยวิธีโพลี-ชิปเทสท์วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติสามารถตรวจสอบได้ถูกต้องประมาณ 27%, 50% และ 49% ตามลำดับ เมื่อขนาดตัวอย่างเพิ่มขึ้นมีผลทำให้อัตราความถูกต้องของทั้งสามวิธีเพิ่มขึ้นเกือบทุกเงื่อนไข ถ้าเพิ่มขนาดตัวอย่างจาก 250 เป็น 2,000 คนต่อกลุ่ม ทั้งสามวิธีมีอัตราความถูกต้องเพิ่มขึ้นมากที่สุด (ประมาณ 35%, 48% และ 51% ตามลำดับ) แต่วิธีโพลี-ชิปเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 ลดลงประมาณ 5% ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้น ประมาณ 21% และ 4% ตามลำดับ ถ้าเพิ่มขนาดตัวอย่างจาก 1,000 เป็น 2,000 คนต่อกลุ่ม ทั้งสามวิธีมีอัตราความถูกต้องเพิ่มขึ้นน้อยที่สุด (ประมาณ 4%,

6% และ 2% ตามลำดับ) ขนาดตัวอย่างในช่วง 500 ถึง 2,000 คนต่อกลุ่ม ไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสต์ ซึ่งค่อนข้างคงที่ (ประมาณ 3% ถึง 4%) ส่วนขนาดตัวอย่างในช่วง 250 ถึง 1,000 คนต่อกลุ่ม มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติค่อนข้างน้อย (ประมาณ 7% ถึง 9%) ซึ่งต่างจากวิธีการถดถอยโลจิสติกแบบจัดอันดับ ขนาดตัวอย่างในทุกเงื่อนไขไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ค่อนข้างมาก (ประมาณ 8% ถึง 29%)

ผลของวิธีโพลี-ชิปเทสต์สอดคล้องกับผลการศึกษาของ ชาง มาซซีโอ และรูสโซ (Chang, Mazzeo, & Roussos. 1996) ที่พบว่า เมื่อขนาดตัวอย่างเพิ่มขึ้นในทุกเงื่อนไข (500, 1,000 และ 2,000 คนต่อกลุ่ม) ไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสต์ โดยมีค่าค่อนข้างคงที่ทั้งสามระดับ (ประมาณ 7% ถึง 8%) สอดคล้องกับผลการศึกษาของ โบลท์ (Bolt. 2002) ที่พบว่า เมื่อขนาดตัวอย่างเพิ่มขึ้นจาก 300 คน เป็น 1,000 คนต่อกลุ่ม วิธีโพลี-ชิปเทสต์มีอัตราความคลาดเคลื่อนประเภทที่ 1 ค่อนข้างคงที่ สอดคล้องกับผลการศึกษาของ สเตาท์และคนอื่นๆ (Stout; et al. 1997) ที่พบว่า เมื่อขนาดตัวอย่างเพิ่มขึ้นในทุกเงื่อนไข (300, 500, 1,000, 1,500 3,000 คนต่อกลุ่ม) วิธีมัลติชิปมีอัตราความถูกต้องเพิ่มขึ้น แต่อัตราความคลาดเคลื่อนประเภทที่ 1 ลดลง และสอดคล้องกับผลการศึกษาของอุทยวรรณ สายพัฒนา (2547) ที่พบว่า ในเงื่อนไขขนาดตัวอย่าง 500 และ 1,000 คนต่อกลุ่ม วิธีโพลี-ชิปเทสต์มีอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่แตกต่างกัน ส่วนผลการศึกษาวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ สอดคล้องกับผลการศึกษาของ โรเจอร์สและสวามินาธาน (Rogers; & Swaminathan. 1993) เฟรนช์และมิลเลอร์ (French; & Miller. 1996) นารายานานและสวามินาธาน (Narayanan; & Swaminathan. 1996) ที่พบว่า เมื่อขนาดตัวอย่างเพิ่มขึ้นมีผลทำให้วิธีในกลุ่มที่ใช้โมเดลการถดถอยโลจิสติกมีอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้น

วิธีโพลี-ชิปเทสต์เป็นวิธีในกลุ่มทฤษฎีการตอบข้อสอบ (Item response theory; IRT) มีรูปแบบนันทพารามิเตอร์ (Nonparametric form) ไม่จำเป็นต้องใช้โมเดลประมาณค่าพารามิเตอร์ ดังนั้นจึงไม่มีข้อตกลงของโมเดลที่ใช้อธิบายความสัมพันธ์ระหว่างผลการตอบข้อสอบกับตัวแปรการจับคู่ (Ackerman. 1992: 76; Chang; Mazzeo; & Roussos. 1996: 334; Potenza; & Dorans. 1995: 24) ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติเป็นวิธีในกลุ่มทฤษฎีการทดสอบมาตรฐานเดิม (Classical test theory; CTT) มีรูปแบบพารามิเตอร์ (Parametric form) ต้องใช้โมเดลประมาณค่า โดยทั่วไปแล้วรูปแบบพารามิเตอร์มักใช้

ขนาดตัวอย่างมากกว่ารูปแบบนันทราเมทริก เพื่อให้การประมาณค่าพารามิเตอร์มีความคงที่ ดังนั้นวิธีโพลี-ชิปเทสท์จึงมีจุดเด่นที่ไม่จำเป็นต้องใช้ตัวอย่างขนาดใหญ่ อย่างไรก็ตาม ทั้งวิธีโพลี-ชิปเทสท์ และวิธีที่ใช้โมเดลการถดถอยโลจิสติกมีการทดสอบนัยสำคัญภายใต้การแจกแจงแบบเชิงเส้นกำกับ (Asymptotic distribution) ซึ่งจำเป็นต้องใช้ขนาดตัวอย่างที่เพียงพอ (Rogers; & Swaminathan. 1993: 106; Roussos; & Stout. 1996a: 215) จากการศึกษาของ นารายานานและสวามินาธาน (Narayanan; & Swaminathan. 1994: 315) พบว่า ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถมิติเดียวและให้คะแนนสองค่า เมื่อใช้ขนาดตัวอย่าง 300 คนต่อในแต่ละกลุ่มมีความถูกต้องเพียงพอ (Sufficient power) เมื่อวิเคราะห์ด้วยวิธีชิปเทสท์ ส่วนผลการศึกษาในครั้งนี้ได้ข้อค้นพบว่า ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า ขนาดตัวอย่าง 1,000 คนในแต่ละกลุ่มเป็นเงื่อนไขที่มีความถูกต้องเพียงพอ เมื่อวิเคราะห์ด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ซึ่งมีอัตราความถูกต้องเกิน 80% (Cohen. 1988: 5) แต่วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าประมาณสองเท่า

5. ปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า พบว่า ปฏิสัมพันธ์ระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน มีผลต่ออัตราความถูกต้องของวิธีโพลี-ชิปเทสท์และวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่ออัตราความถูกต้องของวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ในขณะเดียวกัน มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ผลดังกล่าวที่มีนัยสำคัญเป็นไปตามสมมติฐานข้อ 1-3 จะเห็นว่า ภายใต้ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน เมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้นมีผลทำให้วิธีโพลี-ชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องลดลง โดยส่งผลกระทบต่ออัตราความถูกต้องของวิธีการถดถอยโลจิสติกแบบจัดอันดับ

มากกว่าวิธีโพลี-ชิปเทสท์ เมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้น 10% (จาก 10% ถึง 20%) มีผลทำให้วิธีโพลี-ชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องลดลงประมาณ 5% และ 7% ตามลำดับ แต่ถ้าสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้น 20% (จาก 10% ถึง 30%) มีผลทำให้มีอัตราความถูกต้องลดลงประมาณ 9% และ 14% ตามลำดับ ทั้งยังทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับเพิ่มขึ้นด้วย ไม่ว่าเพิ่มในระดับ 10% หรือ 20% แต่ไม่มีผลต่อวิธีโพลี-ชิปเทสท์ ซึ่งค่อนข้างคงที่ในทุกเงื่อนไข (ประมาณ 2% ถึง 4%) ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ไม่มีผลทั้งอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 สำหรับภายใต้ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน เมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้นในช่วง 10% ถึง 30% ไม่มีผลต่ออัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งสามวิธี

ผลการศึกษาดังกล่าว สอดคล้องกับผลการศึกษาของ นารายานานและสวามินาทาน (Narayanan; & Swaminathan. 1993) ที่พบว่า ในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนรูปที่เป็นรูปแบบเดียวกัน เมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้น 10% (จาก 10% ถึง 20%) ส่งผลให้วิธีชิปเทสท์มีอัตราความถูกต้องลดลงเล็กน้อยประมาณ 1% ถึง 5% และสอดคล้องกับผลการศึกษาของ โรเจอร์สและสวามินาทาน (Rogers; & Swaminathan. 1993) ที่พบว่า ในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน เมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้น 15% (จาก 0% ถึง 15%) มีผลทำให้วิธีการถดถอยโลจิสติกมีอัตราความถูกต้องลดลง แต่ในข้อสอบที่ทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกันไม่มีผลต่ออัตราความถูกต้อง

ถึงแม้ว่าในแบบทดสอบมาตรฐานสามารถตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกันมากกว่าข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน แต่ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกันสามารถเกิดขึ้นในข้อมูลจริง (Narayanan; & Swaminathan. 1996; citing Hambleton; & Rogers. 1989; Linn; et al. 1981; Mellenbergh. 1982) ตลอดจนปริมาณของข้อสอบทำหน้าที่เบี่ยงเบนในแบบทดสอบ จำนวน 10% ถึง 30% มีโอกาสเกิดขึ้นในข้อมูลจริง ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนและสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนอาจมีผลกระทบต่อประสิทธิภาพของวิธีการตรวจสอบดังที่กล่าวมาข้างต้น ดังนั้นในการเลือกวิธีการตรวจสอบที่มีประสิทธิภาพสูงควรคำนึงถึงปัจจัยดังกล่าว สำหรับวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติน่าจะเป็นตัวเลือกหนึ่ง ในการนำไปใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถมิติเดียวและให้คะแนนหลายค่า เพราะไม่ได้รับผลกระทบจากปฏิสัมพันธ์ระหว่างรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน

6. ปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับ ปัจจัยความแตกต่างของการแจกแจงความสามารถ

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า พบว่า ปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับ ปัจจัยความแตกต่างของการแจกแจงความสามารถ มีผลต่ออัตราความถูกต้องของวิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีต่ออัตราความถูกต้องของวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ในขณะเดียวกัน มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ผลดังกล่าวที่มีนัยสำคัญเป็นไปตามสมมติฐานข้อ 1-3 จะเห็นว่า ภายใต้ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน เมื่อกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบเพิ่มขึ้น 1.0SD (จาก 0.0SD ถึง 1.0SD) มีผลทำให้วิธีโพลี-ชิปเทสต์ มีอัตราความถูกต้องเพิ่มขึ้นประมาณ 33% ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องลดลงประมาณ 10% และอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นประมาณ 22% ในขณะที่วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติไม่มีผลกระทบทั้งอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 สำหรับภายใต้ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน เมื่อกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถในช่วง 0.0SD ถึง 1.0SD ไม่ส่งผลกระทบต่อทั้งสามวิธี

ผลการศึกษาวิธีโพลี-ชิปเทสต์ไม่สอดคล้องกับผลการศึกษาของ ชาง มาซซีโอ และรูสโซ (Chang, Mazzeo, & Roussos. 1996) ที่พบว่า ภายใต้ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน เมื่อกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถเพิ่มขึ้น 1.0SD มีผลทำให้วิธีโพลี-ชิปเทสต์มีอัตราความถูกต้องลดลง ส่วนภายใต้ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน เมื่อกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถเพิ่มขึ้น 1.0SD ไม่มีผลต่ออัตราความถูกต้อง ส่วนผลการศึกษาวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ สอดคล้องกับผลการศึกษาของ คริสต์แจนส์สันและคนอื่นๆ (Krisjansson; et al. 2005) ที่พบว่า ในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน กลุ่มผู้สอบที่มีความแตกต่างของการแจกแจงความสามารถในช่วง 0.0SD ถึง 0.5SD ไม่มีผลต่ออัตราความถูกต้องของวิธีการถดถอยโลจิสติก (UCLOLR)

การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในข้อมูลจริง ควรคำนึงถึงรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน และความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบ ซึ่งอาจมีผลต่อประสิทธิภาพของการตรวจสอบดังที่กล่าวมาข้างต้น ค่าเฉลี่ยของความสามารถระหว่างกลุ่มผู้สอบแตกต่างกัน 1.0SD (Standard deviation; SD) ถือว่าเป็นค่าประมาณที่มีค่ามากสุดในการนำมาประยุกต์ใช้ตามสภาพที่เป็นจริง (Stout; et al. 1997: 203) จากผลการศึกษาในครั้งนี้จะเห็นว่า ปฏิสัมพันธ์ระหว่างรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน และความแตกต่างของการแจกแจงความสามารถไม่มีผลกระทบต่อประสิทธิภาพในการตรวจสอบด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ดังนั้นวิธีดังกล่าวน่าจะเป็นตัวเลือกหนึ่ง ในการนำไปใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถมิติเดียวและให้คะแนนหลายค่า

7. ผลการทดสอบปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยขนาดตัวอย่าง

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า พบว่า ปฏิสัมพันธ์สองทางระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยขนาดตัวอย่าง มีผลต่ออัตราความถูกต้องของทั้งสามวิธีอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 และมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ผลดังกล่าวที่มีนัยสำคัญเป็นไปตามสมมติฐานข้อ 1-3 จะเห็นว่า ภายใต้ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน หรือที่ไม่เป็นรูปแบบเดียวกัน เมื่อขนาดตัวอย่างเพิ่มขึ้นในช่วง 250 ถึง 2,000 คนในแต่ละกลุ่มมีผลทำให้วิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องเพิ่มขึ้นเกือบทุกเงื่อนไข และยังมีผลทำให้วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นในทุกเงื่อนไข

ผลการศึกษาดังกล่าว สอดคล้องกับผลการศึกษาของ ทองอยู่ สาระ (2543) วลีมาศ แซ่ฮั้ง (2543) เฟรนช์และมิลเลอร์ (French; & Miller. 1996) คริสต์แจนส์สันและคนอื่นๆ (Krisjansson; et al. 2005) นารายานานและสวามินาทาน (Narayanan; & Swaminathan. 1994, 1996) โรเจอร์สและสวามินาทาน (Rogers; & Swaminathan. 1993) วิทมอร์และชูแม็คเกอร์ (Whitmore; & Schumacker. 1999) ที่พบว่า ภายใต้ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน หรือข้อสอบ

ทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน เมื่อขนาดตัวอย่างเพิ่มขึ้นมีผลทำให้วิธีชิปเทสท์ และวิธีที่ใช้โมเดลการถดถอยโลจิสติกมีอัตราความถูกต้องเพิ่มขึ้น

เมื่อพิจารณาขนาดตัวอย่างที่ทำให้ความถูกต้องเพียงพอ (Sufficient power) ตามเกณฑ์ของโคเฮน (Cohen, 1988: 5) ที่กำหนดอัตราความถูกต้องมากกว่า 80% ที่ระดับนัยสำคัญ .05 มีข้อค้นพบว่า ในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน ขนาดตัวอย่าง 1,000 คนต่อกลุ่ม เป็นเงื่อนไขที่มีความเพียงพอสำหรับวิธีโพลี-ชิปเทสท์ ซึ่งสามารถตรวจสอบได้ถูกต้องประมาณ 96% และมีความผิดพลาดประมาณ 1% ส่วนในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ขนาดตัวอย่าง 1,000 คนในแต่ละกลุ่ม เป็นเงื่อนไขที่มีความเพียงพอสำหรับวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ซึ่งสามารถตรวจสอบได้ถูกต้องประมาณ 98% และมีความผิดพลาดประมาณ 9%

8. ผลการทดสอบปฏิสัมพันธ์สองทางระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถ

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า พบว่า ปฏิสัมพันธ์สองทางระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถ มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ผลดังกล่าวที่มีนัยสำคัญเป็นไปตามสมมติฐานข้อ 1-3 จะเห็นว่า ภายใต้ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10%, 20% และ 30% เมื่อกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถเพิ่มขึ้น 1.0SD (จาก 0.0SD ถึง 1.0SD) มีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับเพิ่มขึ้น ประมาณ 23%, 20% และ 18% ตามลำดับ แต่ไม่มีผลต่อวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ และวิธีโพลี-ชิปเทสท์

ผลการศึกษาวิธีโพลี-ชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ สอดคล้องกับผลการศึกษาของ นารายานานและสวามินาทาน (Narayanan; & Swaminathan, 1994, 1996) โรเจอร์สและสวามินาทาน (Rogers; & Swaminathan, 1993) ที่พบว่า ในแบบทดสอบที่มีสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนในช่วง 0% ถึง 20% เมื่อกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถเพิ่มขึ้น 1.0SD (จาก 0.0SD ถึง 1.0SD) ไม่มีผลต่ออัตราความถูกต้องของวิธีชิปเทสท์ และวิธีการถดถอยโลจิสติก

9. ผลการทดสอบปฏิสัมพันธ์สองทางระหว่างปัจจัยสัดส่วนของข้อสอบทำ หน้าที่เบี่ยงเบน กับปัจจัยขนาดตัวอย่าง

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า พบว่า ปฏิสัมพันธ์ระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยขนาดตัวอย่าง มีผลต่ออัตราความถูกต้องของวิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่อวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ในขณะเดียวกัน ปฏิสัมพันธ์ระหว่างปัจจัยดังกล่าว มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่อวิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ผลดังกล่าวที่มีนัยสำคัญทางสถิติเป็นไปตามสมมติฐานที่กำหนดในข้อ 1-3 จะเห็นว่า ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10%, 20% หรือ 30% เมื่อขนาดตัวอย่างเพิ่มขึ้นในช่วง 250 ถึง 2,000 คนต่อกลุ่ม มีผลต่อประสิทธิภาพของวิธีการถดถอยโลจิสติกแบบจัดอันดับมากที่สุด โดยทำให้อัตราความถูกต้องเพิ่มขึ้นในช่วง 4% ถึง 54% และมีอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นในช่วง 5% ถึง 26%

ผลการศึกษาวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติสอดคล้องกับผลการศึกษาของ นารายานานและสวามินาทาน (Narayanan; & Swaminathan. 1994, 1996) โรเจอร์สและสวามินาทาน (Rogers; & Swaminathan. 1993) ที่พบว่า ในแบบทดสอบที่มีสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนในช่วง 0% ถึง 20% เมื่อใช้ขนาดตัวอย่างของกลุ่มสนใจในช่วง 100 ถึง 500 คน และขนาดตัวอย่างของกลุ่มอ้างอิงในช่วง 250 ถึง 1,000 คน ไม่มีผลต่ออัตราความถูกต้องของวิธีชิปเทสต์ และวิธีการถดถอยโลจิสติก

10. ปฏิสัมพันธ์สองทางระหว่างปัจจัยความแตกต่างของการแจกแจงความ- สามารถ กับปัจจัยขนาดตัวอย่าง

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า พบว่า ปฏิสัมพันธ์ระหว่างปัจจัยความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบ กับปัจจัยขนาดตัวอย่าง มีผลต่ออัตราความถูกต้องของวิธีโพลี-ชิปเทสต์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 แต่ไม่มีผลต่อวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ในขณะเดียวกัน ปฏิสัมพันธ์ระหว่างปัจจัยดังกล่าว มีผล

ต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบทั้งสามอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ผลดังกล่าวที่มีนัยสำคัญทางสถิติเป็นไปตามสมมติฐานที่กำหนดในข้อ 1-3 จะเห็นว่า ภายใต้กลุ่มผู้สอบที่มีความแตกต่างของการแจกแจงความสามารถ 0.0SD, 0.5SD หรือ 1.0SD เมื่อขนาดตัวอย่างเพิ่มขึ้นในช่วง 250 ถึง 1,000 คนต่อกลุ่ม มีผลทำให้อัตราความถูกต้องของวิธีโพลี-ชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับเพิ่มขึ้นเกือบทุกเงื่อนไข แต่ถ้าเพิ่มเป็น 2,000 คนต่อกลุ่ม มีผลทำให้อัตราความถูกต้องเพิ่มขึ้นน้อยกว่าในช่วงดังกล่าว สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 มีผลต่อวิธีโพลี-ชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติน้อยมาก ซึ่งต่างจากวิธีการถดถอยโลจิสติกแบบจัดอันดับที่มีผลค่อนข้างมาก

ผลการศึกษาวิธีโพลี-ชิปเทสท์สอดคล้องกับผลการศึกษาของ โบลท์ (Bolt. 2002) ที่พบว่า ภายใต้กลุ่มผู้สอบที่มีความแตกต่างของการแจกแจงความสามารถ 0.0SD และ 1.0SD เมื่อขนาดตัวอย่างเพิ่มขึ้นจาก 300 คน เป็น 1,000 คนต่อกลุ่ม ทำให้อัตราความถูกต้องของวิธีโพลี-ชิปเทสท์เพิ่มขึ้น แต่ไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ส่วนผลการศึกษาวิธีการถดถอยโลจิสติกแบบจัดอันดับสอดคล้องกับผลการศึกษาของ นารายานานและสวามินาทาน (Narayanan; & Swaminathan. 1996) ที่พบว่า ภายใต้กลุ่มผู้สอบที่มีความแตกต่างของการแจกแจงความสามารถ 0.0SD และ 1.0SD เมื่อขนาดตัวอย่างเพิ่มขึ้น (กลุ่มอ้างอิง 500 และ 1,000 คน $\times$ กลุ่มสนใจ 200 และ 500 คน) มีผลทำให้วิธีการถดถอยโลจิสติกมีอัตราความถูกต้องเพิ่มขึ้นในทุกเงื่อนไข

จากผลการศึกษาครั้งนี้ได้ข้อค้นพบว่า เมื่อกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถ 0.0SD, 0.5SD และ 1.0SD ถ้าใช้ขนาดตัวอย่าง 1,000 คนต่อกลุ่มจะเป็นเงื่อนไขที่มีความถูกต้องเพียงพอ (Sufficient power) ตามเกณฑ์ของโคเฮน (Cohen. 1988: 5) สำหรับการตรวจสอบด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ซึ่งสามารถตรวจสอบถูกต้องประมาณ 99%, 98% และ 96% ตามลำดับ โดยมีโอกาสผิดพลาดประมาณ 7%, 8% และ 12% ตามลำดับ

11. ผลการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ

ผลการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภท 1 ระหว่างวิธีโพลี-ชิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้

คะแนนหลายค่า ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบ และขนาดตัวอย่าง รวมทั้งปฏิสัมพันธ์สองทางระหว่างปัจจัยดังกล่าว พบว่า ภายใต้เงื่อนไขส่วนมาก วิธีการถดถอยโลจิสติกแบบจัดอันดับและวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิตีมีอัตราความถูกต้องไม่แตกต่างกัน ทั้งสองวิธีมีอัตราความถูกต้องแตกต่างกับวิธีโพลี-ชิปเทสอย่างมีนัยสำคัญที่ระดับ .001 ส่วนผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ภายใต้เกือบทุกเงื่อนไข ทั้งสามวิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกันอย่างมีนัยสำคัญที่ระดับ .001 ผลการเปรียบเทียบดังกล่าวที่มีนัยสำคัญสอดคล้องกับสมมติฐานในข้อ 4

ถึงแม้ว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับจะมีอัตราความถูกต้องใกล้เคียงกับวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิตี แต่วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่า ประมาณ 8% เมื่อเปรียบเทียบกับเกณฑ์ที่คาดหวัง ระดับนัยสำคัญ .05 หรืออัตราความคลาดเคลื่อนประเภทที่ 1 ไม่เกิน 5% ซึ่งเป็นค่าที่ยอมรับได้ในทางสถิติ จะเห็นว่า วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าขอบเขตที่กำหนด (Nominal limits) ประมาณ 12% ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิตีมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าขอบเขตที่กำหนดประมาณ 4% แสดงว่า **ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิตีและให้คะแนนหลายค่า วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิตีมีประสิทธิภาพสูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ** สาเหตุที่ทำให้วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูง อาจเนื่องมาจากการวิจัยนี้ใช้ข้อมูลที่วัดความสามารถหลายมิตีและให้คะแนนหลายค่า วิธีการถดถอยโลจิสติกแบบจัดอันดับเป็นวิธีที่ซัมโบ (Zumbo, 1999) พัฒนาเพื่อใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถมิตีเดียวและให้คะแนนหลายค่า กระบวนการวิเคราะห์ที่ใช้คะแนนรวมของแบบทดสอบเป็นเกณฑ์การจับคู่ (Matching criterion) คะแนนรวมดังกล่าวถือว่าเป็นระดับความสามารถของผู้สอบ ซึ่งคำนวณมาจากค่าเฉลี่ยของระดับความสามารถในทุกมิตี เมื่อนำมาทดสอบความเหมาะสมกับโมเดล (Model data fit) ผลการทดสอบจึงไม่คงที่ หรือไม่เหมาะสมกับโมเดล แล้วสรุปว่าข้อสอบทำหน้าที่เบี่ยงเบน ซึ่งแท้ที่จริงแล้วข้อสอบทำหน้าที่ไม่เบี่ยงเบน คะแนนรวมของแบบทดสอบตามทฤษฎีการทดสอบแบบมาตรฐานเดิมจึงไม่น่าเพียงพอที่จะนำมาใช้เป็นเกณฑ์การจับคู่ความสามารถของผู้สอบ เมื่อ

ผู้สอบมาจากกลุ่มที่แตกต่างกันภายใต้การแจกแจงความสามารถหลายมิติ ผลดังกล่าวชี้ให้เห็นถึงความผิดพลาดของการเลือกใช้โมเดลสำหรับการวิเคราะห์ (Model misspecification) ดังนั้นจึงมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 สูงเกินปกติ ผลการศึกษาดังกล่าวสอดคล้องกับผลการศึกษาของแอ็คเคอร์แมน (Ackerman. 1992) ที่ใช้คะแนนรวมของแบบทดสอบที่วัดความสามารถหลายมิติเป็นเกณฑ์การจับคู่ ซึ่งมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าสูงเช่นเดียวกัน

ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ผู้วิจัยได้ปรับขยายมาจากโมเดลที่วัดความสามารถมิติเดียวของซัมโบ (Zumbo. 1999) โดยนำค่าประมาณความสามารถในทุกมิติ (θ_1 และ θ_2) เป็นเกณฑ์การจับคู่ เพื่อให้การเปรียบเทียบค่าความสามารถระหว่างกลุ่มมีความถูกต้องและแม่นยำ ดังนั้นผลการตรวจสอบจึงสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ภายในขอบเขตที่กำหนดได้ดีกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ ในขณะที่อัตราความถูกต้องของสองวิธีใกล้เคียงกัน ผลดังกล่าวสอดคล้องกับผลการศึกษาของ นพมาศ พิพัฒน์สุข (2541) เมเซอร์ แฮมเบิลตัน และคลอเซอร์ (Mazor; Hambleton; & Clauser. 1998) ที่พบว่า ในข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนสองค่า วิธีการถดถอยโลจิสติกเมื่อใช้คะแนนแบบทดสอบหลายชุดย่อย (Multiple subtest scores) เป็นเกณฑ์การจับคู่ความสามารถมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ลดลง เมื่อเทียบกับการใช้คะแนนรวมของแบบทดสอบเป็นเกณฑ์การจับคู่

สำหรับวิธีโพลี-ชิปเทสท์มีแนวคิดพื้นฐานมาจากทฤษฎีการตอบข้อสอบแบบหลายมิติ (Multidimensional IRT) ซึ่งช่าง มาซซีโอ และรูสโซ (Chang, Mazzeo, & Roussos. 1996) ปรับขยายมาจากวิธีชิปเทสท์ของชีลลีและสเตาท์ (Shealy; & Stout. 1993) เพื่อใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนหลายค่า ข้อจำกัดของวิธีนี้ก็คือไม่สามารถตรวจพบข้อสอบที่ทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) ดังเหตุผลที่อธิบายไว้ในข้อ 1 ผลของปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่เบี่ยงเบน จึงส่งผลให้อัตราความถูกต้องเกือบทุกเงื่อนไขมีค่าต่ำกว่าวิธีที่ใช้โมเดลการถดถอยโลจิสติก ส่วนข้อได้เปรียบของวิธีโพลี-ชิปเทสท์ที่เหนือกว่าอีกสองวิธีก็คือ สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ดี จะเห็นว่าเกือบทุกเงื่อนไขมีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ภายในขอบเขตที่กำหนด (Nominal limits) สำหรับปัจจัยที่แปรเปลี่ยนทั้ง 4 ปัจจัยมีผลน้อยมากต่ออัตราความคลาดเคลื่อนประเภทที่ 1 สาเหตุดังกล่าวอาจเนื่องมาจาก วิธีโพลี-ชิปเทสท์ใช้เทคนิคการตรวจสอบแบบหลายมิติ โดยแยกแบบทดสอบ

ออกเป็นสองชุดย่อย คือ แบบทดสอบที่ใช้ในการจับคู่ (Matching subtests) และแบบทดสอบที่ใช้ในการศึกษา (Studied subtests) ทำให้จำแนกความสามารถเป้าหมาย (Target ability) และความสามารถแทรกซ้อน (Nuisance ability) ได้อย่างชัดเจน ข้อสอบทำหน้าที่เบี่ยงเบนจึงเป็นความแตกต่างของผลการตอบข้อสอบระหว่างกลุ่มผู้สอบ ซึ่งคำนวณมาจากความแตกต่างของความสามารถแทรกซ้อน โดยเปรียบเทียบความสามารถเป้าหมายเดียวกัน แต่ถ้าความแตกต่างของผลการตอบข้อสอบไม่ได้เปรียบเทียบที่ความสามารถเป้าหมายเดียวกันจะเป็น **“ผลกระทบของข้อสอบ”** (Item impact) (Camilli. 1992: 131; Clauser; & Mazor. 1998: 31; Dorans; & Holland; 1993: 37; Zumbo. 1999: 12) ซึ่งมีผลทำให้ตัดสินข้อสอบทำหน้าที่เบี่ยงเบนผิดพลาด ดังนั้นการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบโดยใช้โมเดลแบบหลายมิติสามารถแยกผลกระทบของข้อสอบออกจากข้อสอบทำหน้าที่เบี่ยงเบน ซึ่งการตรวจสอบภายใต้มิติเดียวไม่สามารถทำได้ ส่วนเกณฑ์การจับคู่ค่อนข้างเป็นคุณลักษณะแฝงมากกว่าคะแนนที่ได้จากการสอบ ทำให้ผลการตรวจสอบมีความแม่นยำสูง ถึงแม้ว่ากลุ่มผู้สอบจะมีการแจกแจงของความสามารถเป้าหมายแตกต่างกัน ซึ่งมีผลทำให้เกิดเงื่อนไขผลกระทบของข้อสอบดังที่กล่าวมาข้างต้น ก็สามารถปรับแก้ความแตกต่างของการแจกแจงความสามารถเป้าหมายโดยใช้วิธีทางสถิติคือ การปรับแก้การถดถอย (Regression correction) นอกจากนี้ยังใช้การคำนวณทวนซ้ำเพื่อนำข้อสอบที่ทำหน้าที่เบี่ยงเบนออกจากเกณฑ์การจับคู่ จนกระทั่งได้คะแนนเกณฑ์ที่มีความบริสุทธิ์ (Purification) (Chang, Mazzeo, & Roussos. 1996: 351; Clauser; & Mazor. 1998: 35) ด้วยเหตุผลเหล่านี้จึงทำให้วิธีโพลี-ชิปเทสต์สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในขอบเขตที่กำหนดได้ดีกว่าวิธีอื่นๆ

ตามปกติแล้ว วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่มีประสิทธิภาพคาดหวังว่ามีอัตราความถูกต้องสูง และอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำ ถ้าตรวจสอบที่ระดับนัยสำคัญ .05 แล้วมีอัตราความถูกต้องตั้งแต่ 80% ขึ้นไป ถือว่ามีความถูกต้องเพียงพอ (Sufficient power) (Cohen. 1988: 5) ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ซึ่งเกิดขึ้นจริงถ้ามีค่าอยู่ภายในขอบเขตที่กำหนด (Nominal limits) ไม่เกิน 5% ถือว่ามีความทนทาน (Robustness) นั้นแสดงว่า จะต้องพิจารณาทั้งอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ควบคู่กันไป จากผลการศึกษาในครั้งนี้ วิธีโพลี-ชิปเทสต์มีอัตราความถูกต้องในภาพรวมค่อนข้างต่ำ แต่มีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ภายในขอบเขตที่กำหนดเกือบทุกเงื่อนไข ในขณะที่วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องในภาพรวมค่อนข้างสูง และมี

อัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าขอบเขตที่กำหนด ประมาณ 4% ดังนั้นจำเป็นต้องปรับแก้ (Adjustment) ระดับนัยสำคัญของวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติที่ใช้ในการตัดสินใจทำนายที่เบี่ยงเบนของข้อสอบให้มีค่าลดลง เพื่อให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าลดลงในระดับที่ยอมรับได้

ผลการปรับแก้ระดับนัยสำคัญได้ข้อค้นพบว่า ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า ด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ เมื่อกำหนดระดับนัยสำคัญ .023 แล้วใช้ขนาดตัวอย่าง 1,000 คนต่อกลุ่มจะทำให้ผลการตรวจสอบมีความถูกต้องเพียงพอ และมีความผิดพลาดอยู่ในขอบเขตที่กำหนด

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1.1 ผลการศึกษาค้นพบว่า การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำนายที่เบี่ยงเบน สัดส่วนของข้อสอบทำนายที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบ และขนาดตัวอย่าง โดยภาพรวมภายใต้ทุกเงื่อนไข วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีความเหมาะสมมากที่สุด เมื่อเทียบกับวิธีโพลีชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ถ้าตัดสินข้อสอบทำนายที่เบี่ยงเบนที่ระดับนัยสำคัญ .05 วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องประมาณ 81% และอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณ 9% ดังนั้นถ้านำวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติไปตรวจสอบในข้อมูลจริงจำเป็นต้องปรับระดับนัยสำคัญในการตัดสินใจสอบทำนายที่เบี่ยงเบนให้มีค่าลดลง เพื่อให้ผลการตรวจสอบมีความผิดพลาดไม่เกิน 5% ซึ่งอยู่ในระดับที่ยอมรับได้ ผลการปรับระดับนัยสำคัญพบว่า ที่ระดับนัยสำคัญ .02 ทำให้ผลการตรวจสอบมีอัตราความถูกต้องประมาณ 74% และอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณ 4%

1.2 ผลการศึกษาค้นพบว่า การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในแบบทดสอบที่วัดคุณลักษณะซับซ้อน (Complex traits) การใช้คะแนนรวมของแบบทดสอบเป็นเกณฑ์การจับคู่ความสามารถไม่เหมาะสม ซึ่งจะมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 สูงเกินไป ดังเช่น วิธีการถดถอยโลจิสติกแบบจัดอันดับ ดังนั้นควรใช้ระดับความสามารถใน ทุกมิติเป็นเกณฑ์การจับคู่ความสามารถ ซึ่งจะมีผลทำให้การตรวจสอบมีความถูกต้องและแม่นยำ ไม่เดล

การถดถอยโลจิสติกมีความยืดหยุ่นสูง สามารถนำไปปรับใช้ตามรูปแบบการให้คะแนนของข้อสอบ และจำนวนมิติที่ใช้วัดความสามารถ ในการศึกษาค้นคว้าครั้งนี้ใช้ค่าประมาณความสามารถสองมิติ ภายใต้ทฤษฎีการตอบข้อสอบแบบหลายมิติ (Multidimensional IRT) เป็นเกณฑ์การจับคู่ หากนำวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติไปประยุกต์ใช้ในข้อมูลจริง ไม่จำเป็นต้องใช้ค่าประมาณความสามารถดังกล่าว อาจใช้คะแนนรวมในแบบทดสอบชุดย่อยที่วัดความสามารถมิติเดียว (Unidimensional subtest scores) เป็นเกณฑ์การจับคู่ คะแนนรวมในแบบทดสอบชุดย่อยถือว่าเป็นระดับความสามารถในแต่ละมิติ

1.3 ผลการศึกษาค้นคว้าพบว่า ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนมีผลต่อประสิทธิภาพของวิธีโพลี-ชิปเทสท์ ซึ่งสามารถตรวจสอบได้เฉพาะข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF) ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีโมเดลที่เป็นข้อได้เปรียบสามารถตรวจพบข้อสอบทำหน้าที่เบี่ยงเบน ทั้งที่เป็นรูปแบบเดียวกัน (Uniform DIF) และที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) สำหรับวิธีการถดถอยโลจิสติกแบบจัดอันดับสามารถพบได้เช่นเดียวกัน แต่มีความผิดพลาดสูงกว่า ในข้อมูลจริงอาจเป็นไปได้ว่ามีข้อสอบทำหน้าที่เบี่ยงเบนทั้งสองรูปแบบ ดังนั้นวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติน่าจะมีความเหมาะสมในการนำไปใช้มากกว่า

1.4 ผลการศึกษาค้นคว้าพบว่า ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนในช่วง 10% ถึง 30% ไม่มีผลต่อประสิทธิภาพของวิธีโพลี-ชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ตามกรอบความคิดเชิงทฤษฎี เมื่อสัดส่วนข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้นมีผลทำให้คะแนนเกณฑ์การจับคู่มีความเชื่อมั่นต่ำ ซึ่งมีผลทำให้ความถูกต้องในการตรวจสอบลดลง และความผิดพลาดสูงขึ้น แต่วิธีโพลี-ชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติไม่ได้รับผลกระทบดังกล่าว ซึ่งต่างจากวิธีการถดถอยโลจิสติกแบบจัดอันดับ ในข้อมูลจริงอาจมีข้อสอบทำหน้าที่เบี่ยงเบนในช่วง 10% ถึง 30% ดังนั้นถ้าคำนึงถึงปัจจัยปริมาณสัดส่วนข้อสอบทำหน้าที่เบี่ยงเบน วิธีโพลี-ชิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติน่าจะเป็นตัวเลือกหนึ่งที่น่าไปใช้

1.5 ผลการศึกษาค้นคว้าพบว่า ปัจจัยความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบหนึ่งส่วนเบี่ยงเบนมาตรฐาน (1.0SD) มีผลต่อประสิทธิภาพของการตรวจสอบของทั้งสามวิธี เมื่อกลุ่มผู้สอบมีความแตกต่างของการแจกแจงความสามารถเพิ่มขึ้นมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 สูงขึ้น สำหรับวิธีโพลี-ชิปเทสท์มีอัตราความคลาดเคลื่อน

ประเภทที่ 1 อยู่ภายในขอบเขตที่กำหนดเกือบทุกเงื่อนไข ($\alpha = .05$) ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิตินี้มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าเกณฑ์ไม่มาก ซึ่งต่างจากวิธีการถดถอยโลจิสติกแบบจัดอันดับที่มีค่าสูงกว่าค่อนข้างมาก ความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบหนึ่งส่วนเบี่ยงเบนมาตรฐานถือว่าเป็นค่าระดับสูงที่สามารถเกิดขึ้นในข้อมูลจริง ดังนั้นการนำวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติไปใช้ตรวจสอบในข้อมูลจริงจำเป็นต้องปรับระดับนัยสำคัญในการตัดสินใจทำนายที่เบี่ยงเบนของข้อสอบให้มีค่าลดลง เพื่อให้ผลการตรวจสอบมีความผิดพลาดในระดับที่ยอมรับได้ ผลการปรับระดับนัยสำคัญพบว่า ภายใต้ความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มผู้สอบ 0.0SD, 0.5SD และ 1.0SD สามารถกำหนดระดับนัยสำคัญ .035, .03 และ .014 ตามลำดับ

1.6 ผลการศึกษาค้นพบว่า ขนาดตัวอย่าง 250 คนในแต่ละกลุ่ม (กลุ่มอ้างอิงและกลุ่มเปรียบเทียบ) ของสามวิธีไม่เหมาะสมที่จะนำมาตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า เพราะมีอัตราความถูกต้องค่อนข้างต่ำ สำหรับวิธีโพลี-ชิปเทสที่มีความถูกต้องไม่เพียงพอ (Insufficient power) ทุกขนาดตัวอย่าง เพราะมีอัตราความถูกต้องต่ำกว่า 80% ตามเกณฑ์ของโคเฮน (Cohen, 1998: 5) ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ และวิธีการถดถอยโลจิสติกแบบจัดอันดับมีความถูกต้องเพียงพอ เมื่อใช้ขนาดตัวอย่าง 1,000 หรือ 2,000 คนต่อกลุ่ม ทั้งสองวิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าขอบเขตที่กำหนด แต่วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ดังนั้นวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติน่าจะเป็นตัวเลือกที่ดีกว่า ถ้านำวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติไปตรวจสอบในข้อมูลจริงจำเป็นต้องปรับระดับนัยสำคัญในการตัดสินใจทำนายที่เบี่ยงเบนให้มีค่าลดลงเพื่อให้ผลการตรวจสอบมีความผิดพลาดในระดับที่ยอมรับได้ ผลการปรับระดับนัยสำคัญพบว่า เมื่อกำหนดระดับนัยสำคัญ .023 แล้วใช้ขนาดตัวอย่าง 1,000 คนต่อกลุ่ม ทำให้มีความถูกต้องเพียงพอ และมีความผิดพลาดอยู่ภายในขอบเขตที่กำหนด ซึ่งถือว่ามีความทนทาน (Robustness) ส่วนขนาดตัวอย่าง 2,000 คนต่อกลุ่มไม่ได้นำมาปรับระดับนัยสำคัญเพราะให้ผลความถูกต้องใกล้เคียงกับขนาดตัวอย่าง 1,000 คนต่อกลุ่ม จะเห็นว่าต้องใช้ตัวอย่างขนาดใหญ่ซึ่งเหมาะที่จะนำไปพัฒนาคุณภาพของแบบทดสอบมาตรฐาน (Standardized test) แบบทดสอบระดับชาติ (National test) หรือคลังข้อสอบ (Item pool)

2. ข้อเสนอแนะในการวิจัยครั้งต่อไป

2.1 จากข้อค้นพบในการวิจัยครั้งนี้ วิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ มีประสิทธิภาพในการตรวจสอบสูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับและวิธีโพลี-ชิปเทสท์ ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่างในการศึกษาภายใต้ปัจจัยดังกล่าวยังมีในบางประเด็นที่น่าสนใจ นอกจากนี้ยังมีปัจจัยอื่นๆ อีกที่คาดว่าจะมีผลต่ออัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ซึ่งสามารถศึกษาภายใต้การจำลองข้อมูล ดังนี้

2.1.1 ขนาดตัวอย่าง (Sample size) เป็นปัจจัยที่น่าสนใจเพิ่มเติม เนื่องจากในการศึกษานี้พบว่า การตรวจสอบด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ที่ระดับนัยสำคัญ .02 เมื่อใช้ขนาดตัวอย่าง 500 คนต่อกลุ่ม ทำให้มีอัตราความถูกต้อง ประมาณ 66.12% และมีอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณ 3.66% แต่เมื่อใช้ขนาดตัวอย่าง 1,000 คนต่อกลุ่ม ทำให้มีอัตราความถูกต้อง ประมาณ 94.94 และมีอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณ 4.51% แสดงว่าขนาดตัวอย่างที่เหมาะสมที่ทำให้อัตราความถูกต้องไม่ต่ำกว่า 80% และอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่เกิน 5% น่าจะอยู่ในช่วง 500 ถึง 1,000 คนต่อกลุ่ม ดังนั้นขนาดตัวอย่างในช่วงดังกล่าวเป็นประเด็นที่น่าสนใจ นอกจากนี้อาจศึกษาปัจจัยเกี่ยวกับอัตราส่วนของขนาดตัวอย่าง (Sample size ratio) ระหว่างกลุ่มอ้างอิงและกลุ่มสนใจ โดยให้กลุ่มสนใจมีขนาดตัวอย่างน้อยกว่าหรือเท่ากับกลุ่มอ้างอิง แล้วพิจารณาว่ามีผลต่ออัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 หรือไม่

2.1.2 ความแตกต่างของการแจกแจงความสามารถ (Ability distribution differences) เป็นปัจจัยที่น่าสนใจเพิ่มเติม เนื่องจากในการวิจัยครั้งนี้ ศึกษาเฉพาะความแตกต่างของค่าเฉลี่ยของค่าความสามารถในมิติที่หนึ่งและมิติที่สอง (θ_1 และ θ_2) ระหว่างกลุ่มอ้างอิงและกลุ่มสนใจเท่านั้น ซึ่งไม่ได้ศึกษาความแตกต่างของค่าความแปรปรวน และค่าสหสัมพันธ์ระหว่างค่าความสามารถในมิติที่หนึ่งและมิติที่สอง ปัจจัยดังกล่าวเป็นประเด็นที่น่าสนใจว่ามีผลกระทบต่อประสิทธิภาพในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หรือไม่

2.1.3 ขนาดของข้อสอบทำหน้าที่เบี่ยงเบน (Magnitude of DIF) เป็นปัจจัยที่ศึกษาเกี่ยวกับขนาดของข้อสอบทำหน้าที่เบี่ยงเบน มีลักษณะทำนองเดียวกับการวัดขนาดอิทธิพล (Effect size) ในทางสถิติ ซึ่งกำหนดจากพารามิเตอร์ของข้อสอบระหว่างกลุ่มผู้สอบ โดยสามารถ

กำหนดขนาดเล็กจนถึงขนาดใหญ่ แล้วพิจารณาอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 เพื่อนำมากำหนดเกณฑ์ที่เหมาะสมต่อไป

2.1.4 ทิศทางของข้อสอบทำหน้าที่เบี่ยงเบน (Direction of DIF) เป็นปัจจัยที่ศึกษาเกี่ยวกับทิศทางของข้อสอบทำหน้าที่เบี่ยงเบน เมื่อเข้าข้างผู้สอบกลุ่มใดกลุ่มหนึ่งแล้วมีผลต่ออัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 หรือไม่ ซึ่งสามารถกำหนดได้ 2 รูปแบบ คือ ข้อสอบทำหน้าที่เบี่ยงเบนแบบสมดุล (Balanced DIF) และข้อสอบทำหน้าที่เบี่ยงเบนแบบไม่สมดุล (Unbalanced DIF)

2.1.5 เกณฑ์การจับคู่ (Matching criterion) เป็นปัจจัยที่ศึกษาเกี่ยวกับการทำให้คะแนนเกณฑ์การจับคู่มีความบริสุทธิ์ โดยพยายามทำเกณฑ์ให้เป็นอิสระจากข้อสอบทำหน้าที่เบี่ยงเบน แล้วพิจารณาว่ามีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ลดลงหรือไม่ วิธีการทำให้เกณฑ์มีความบริสุทธิ์มักศึกษา 3 วิธี คือ วิธีหนึ่งขั้นตอน (One-stage) วิธีสองขั้นตอน (Two-stage) และวิธีคำนวณวนซ้ำ (Iterative)

2.1.6 ทิศทางของข้อสอบ (Item direction) เป็นปัจจัยที่ศึกษาเกี่ยวกับทิศทางของข้อสอบในแบบทดสอบที่วัดความสามารถหลายมิติ ซึ่งกำหนดจากพารามิเตอร์อำนาจจำแนกของข้อสอบแบบหลายมิติ (Multidimensional item discrimination; MDISC) โดยสามารถกำหนดให้ข้อสอบวัดความสามารถในแต่ละมิติแตกต่างกัน หรือเท่ากันก็ได้ แล้วพิจารณาว่ามีผลอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 หรือไม่

2.1.7 ความยาวของแบบทดสอบ (Test length) เป็นปัจจัยที่ต้องการทราบว่าเมื่อแบบทดสอบมีจำนวนข้อสอบเพิ่มขึ้น ซึ่งจะส่งผลทำให้คะแนนเกณฑ์การจับคู่มีความเชื่อมั่นสูงขึ้น แล้วจะมีผลทำให้อัตราความถูกต้องเพิ่มขึ้น และอัตราความคลาดเคลื่อนประเภทที่ 1 ลดลงหรือไม่

2.2 ควรทำการศึกษาลักษณะทำนองเดียวกับการศึกษาในครั้งนี้ แต่ใช้ข้อมูลจริง โดยอาจใช้ตัวแปรเพศ เชื้อชาติ ศาสนา ภูมิลำเนา อายุ หรือประสบการณ์ เป็นเกณฑ์ในการแบ่งกลุ่มผู้สอบ สำหรับแบบทดสอบที่วัดความสามารถหลายมิติที่ให้คะแนนหลายค่า อาจสร้างเป็นแบบทดสอบชุดย่อย (Subtest) เช่น แบบทดสอบวิชาคณิตศาสตร์ มีโครงสร้างของการวัดความรู้ ความเข้าใจ การคิดคำนวณ การแก้โจทย์ปัญหา เป็นต้น แล้วให้ผู้เชี่ยวชาญเป็นผู้ตรวจสอบว่าข้อสอบทำหน้าที่เบี่ยงเบนหรือไม่ เพื่อใช้เป็นเกณฑ์ในการตรวจสอบ ส่วนความสามารถของผู้สอบที่ใช้เป็นเกณฑ์การจับคู่สามารถคำนวณได้จากคะแนนรวมของแบบทดสอบชุดย่อยในแต่ละชุด แล้วพิจารณาผลการตรวจสอบว่าแตกต่าง หรือสอดคล้องกับผลการศึกษารั้งนี้

บรรณานุกรม

บรรณานุกรม

- ทองอยู่ สาระ. (2543). การเปรียบเทียบอำนาจการตรวจสอบและการจำแนกผิดพลาด
ในการตรวจสอบข้อสอบทำหน้าที่เบี่ยงเบนแบบสมำเสมอและแบบไม่สมำเสมอ
ระหว่างวิธีแมนเทิล-แฮนส์เซลกับวิธีการถดถอยโลจิสติก โดยใช้ความยาวของ
แบบทดสอบและขนาดกลุ่มตัวอย่างต่างกัน. ปรินญานิพนธ์ กศ.ม. (การวัดผล
การศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- นพมาศ พิพัฒน์สุข. (2541). การเปรียบเทียบประสิทธิภาพระหว่างวิธีแมนเทิล-แฮนส์เซลกับ
วิธีการถดถอยโลจิสติก ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ
เมื่อใช้เกณฑ์จับคู่เปรียบเทียบแตกต่างกันในแบบสอบพหุมิติ. วิทยานิพนธ์ ค.ม.
(การวัดและประเมินผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย
จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- วลีมาศ แซ่อึ้ง. (2543). การเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อน
ประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบหลายรูปแบบ ระหว่าง
วิธีชิปเทสต์ปรับปรุงใหม่ วิธีชิปเทสต์ วิธีแมนเทิล-แฮนส์เซล และวิธีการถดถอยโลจิสติก.
วิทยานิพนธ์ ค.ด. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย
จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- สิริรัตน์ วิชาสศิลป์. (2545). การเปรียบเทียบวิธีชิปเทสต์และดีเอฟไอทีในการตรวจสอบการทำ
หน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบ จากข้อมูลการตอบ
ข้อสอบที่ใช้ความสามารถหลายมิติ. ปรินญานิพนธ์ กศ.ด. (การทดสอบและ
วัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ.
ถ่ายเอกสาร.
- อุทัยวรรณ สายพัฒน์. (2547). การเปรียบเทียบประสิทธิภาพของผลการตรวจสอบการทำ
หน้าที่ต่างกันของข้อสอบ ในแบบทดสอบที่มีการให้คะแนนแบบหลายค่าระหว่าง
วิธี GMH และวิธี Poly-SIBTEST. ปรินญานิพนธ์ กศ.ด. (การทดสอบและ
วัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ.
ถ่ายเอกสาร.

- Ackerman, T. A. (1992, Spring). A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. *Journal of Educational Measurement*. 29(1): 67-91.
- Ackerman, T. A.; & Evans, J. A. (1994, December). The influence of conditioning scores in performing DIF analyses. *Applied Psychological Measurement*. 18(4): 329-342.
- American Educational Research Association; American Psychological Association; & National Council on Measurement in Education. (1999). *Standards for education and psychological testing*. Washington, DC: American Educational Research Association.
- Angoff, W. H. (1993). Perspective in differential item functioning methodology. In *Differential item functioning*. P. W. Holland; & H. Wainer. pp. 3-23. Hillsdale, NJ: Erlbaum.
- Baker, F. B. (1981, Spring). A criticism of Scheuneman's item bias technique. *Journal of Educational Measurement*. 18(1): 59-62.
- 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. (1992, June). A conceptual analysis of differential item functioning in terms of a multidimensional item response model. *Applied Psychological Measurement*. 16(2): 129-147.
- _____. (1993). The case against item bias detection techniques based on internal criteria: Do item bias procedures obscure test fairness issues. In *Differential item functioning*. P. W. Holland; & H. Wainer. pp. 397-417. Hillsdale, NJ: Erlbaum.
- Camilli, G.; & Shepard, L. A. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: Sage.

- Chang, H.; Mazzeo, J.; & Roussos, L. (1996, Fall). Detecting DIF for polytomously scored items: An adaptation of the SIBTEST procedure. *Journal of Educational Measurement*. 33(3): 333-353.
- Clauser, R. E.; & Mazor, K. M. (1998, Spring). Using statistical procedures to identify differentially functioning test items. *Educational Measurement: Issues and Practice*. 17(1): 31-44.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Cohen, A. S.; & Kim, S. H. (1993, March). 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.; & Baker, F. B. (1993, December). Detection of differential item functioning in the graded response model. *Applied Psychological Measurement*. 17(4): 335-350.
- Donoghue, J. R.; Holland, P. W.; & Thayer, D. T. (1993). A Monte Carlo study of factors that affect the Mantel-Haenszel and standardization measures of differential item functioning. In *Differential item functioning*. P. W. Holland; & H. Wainer. pp. 137-166. Hillsdale, NJ: Erlbaum.
- Dorans, N. J.; & Holland; P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In *Differential item functioning*. P. W. Holland; & H. Wainer. pp. 25-33. Hillsdale, NJ: Erlbaum.
- Dorans, N. J.; & Kulick, E. (1986, Winter). Demonstrating the unity of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement*. 23(4): 355-368.
- Flowers, C. P.; Oshima, T. C.; & Raju, N. S. (1999, December). A description and demonstration of the polytomous-DFIT framework. *Applied Psychological Measurement*. 23(4): 309-326.

- French, A. W.; & Miller, T. R. (1996, Fall). Logistic regression and its use in detecting differential item functioning in polytomous items. *Journal of Educational Measurement*. 33(3): 315-332.
- Gierl, M. J.; Gotzmann, A.; & Boughton, K. A. (2004). Performance of SIBTEST when the percentage of DIF items is large. *Applied Measurement in Education*. 17(3): 241-264.
- Hambleton, R. K.; Swaminathan, H.; & Rogers, H. J. (1990). *Fundamentals of item response theory*. Newbury Park, CA: Sage.
- Holland, W. P.; & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In *Test validity*. H. Wainer; & H. I. Braun. pp. 129-145. Hillsdale, NJ: Erlbaum.
- Hulin, C. L.; Drasgow, F.; & Parsons, C. K. (1983). *Item response theory: Applications to psychological measurement*. Homewood, IL: Dow Jones-Irwin.
- Jodoin, M. G.; & Gierl, M. J. (2001). Evaluating type I error and power rates using an effect size measure with the logistic regression procedure for DIF detection. *Applied Measurement in Education*. 14(4): 329-349.
- Kim, S.-H.; & Cohen, A. S. (1991, September). A comparison of two area measures for detecting differential item functioning. *Applied Psychological Measurement*. 15(3): 269-278.
- Krisjansson, E.; et al. (2005, December). A Comparison of four methods for detecting differential item functioning in ordered response items. *Educational and Psychological Measurement*. 65(6): 935-953.
- Li, H.; & Stout, W. (1996, December). 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). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.

- Mazor, K. M.; Hambleton, R. K.; & Clauser, R. E. (1998, December). Multidimensional DIF effects of matching on unidimensional subtest scores. *Applied Psychological Measurement*. 22(4): 357-367.
- Mellenbergh, G. J. (1982, Summer). Contingency table models for assessing item bias. *Journal of Educational Statistics*. 7(2): 105-107.
- Miller, M. D.; & Oshima, T. C. (1992, December). Effect of sample size, number of biased items, and magnitude of bias on a two-stage item bias estimation method. *Applied Psychological Measurement*. 16(4): 381-388.
- Miller, T. R.; & Spray, J. A. (1993, Summer). Logistic discriminant function analysis for DIF identification of polytomously scored items. *Journal of Educational Measurement*, 30(2): 107-122.
- Millsap, R. E.; & Everson, H. T. (1993, December). Methodology review: Statistical approaches for assessing measurement bias. *Applied Psychological Measurement*. 17(4): 297-334.
- Muraki, E. (1992, June). A generalized partial credit model: Application of an EM Algorithm. *Applied Psychological Measurement*. 16(2): 159-176.
- _____. (2000). *RESGEN: Item Response Generator*. Princeton, NJ: Educational Testing Service.
- Nandakumar, R. (1993, Winter). Simultaneous DIF amplification and cancellation: Shealy-Stout's test for DIF. *Journal of Educational Measurement*. 30(4): 293-311.
- Narayanan, P.; & Swaminathan, H. (1994, December). Performance of the Mantel-Haenszel and simultaneous item bias procedures for detecting differential item functioning. *Applied Psychological Measurement*. 18(4): 315-328.
- _____. (1996, September). Identification of items that show nonuniform DIF. *Applied Psychological Measurement*. 20(3): 257-274.
- Oshima, T. C.; & Miller, M. D. (1990, Fall). Multidimensional and IRT-based item invariance indexes: The effect of between-group variation in trait correlation. *Journal of Educational Measurement*. 27(3): 373-283.

- Oshima, T. C.; & Miller, M. D. (1992, September). Multidimensional and item bias in item response theory. *Applied Psychological Measurement*. 16(3): 237-248.
- Oshima, T. C.; Raju, N. S.; & Flowers, C. P. (1997, Fall). Development and demonstration of multidimensional IRT-based internal measures of differential functioning of items and tests. *Journal of Educational Measurement*. 34(3): 253-272.
- Potenza, M. T.; & Dorans, N. J. (1995, March). DIF assessment for polytomously scored items: A framework for classification and evaluation. *Applied Psychological Measurement*. 19(1): 23-37.
- Raju, N. S. (1990, June). Determining the significance of estimated signed and unsigned areas between two item response functions. *Applied Psychological Measurement*. 14(2): 197-207.
- Raju; N. S; & Ellis, B. B. (2002). Differential item functioning. In *Measurement and analyzing behavior in organizations: Advances in measurement and data analysis*. F. Drasgow; & N. Schmitt. pp. 156-188. San Francisco, CA: Jossey-Bass.
- Raju, N. S.; van der Linden, W.; & Fleer, P. (1995, December). An IRT-based internal measure of test bias with applications for differential item functioning. *Applied Psychological Measurement*. 19(4): 353-368.
- Reckase, M. D. (1997). A linear logistic multidimensional model for dichotomous item response data. In *Handbook of modern item response theory*. W. J. van der Linden; & R. K. Hambleton. pp. 271-286. New York: Springer.
- Rogers, H. J.; & Swaminathan, H. (1993, June). A comparison of logistic regression and Mantel-Haenszel procedures for detecting differential item functioning. *Applied Psychological Measurement*. 17(2): 105-116.
- Roussos, L.; & Stout, W. (1996a, Summer). Simulation studies of the effects of small sample size and studied item parameters on SIBTEST and Mantel-Haenszel Type I error performance. *Journal of Education Measurement*. 33(2): 215-230.

- Roussos, L.; & Stout, W. (1996b, December). 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.
- Scheuneman, J. D. (1979, Fall). A method of assessing bias in test items. *Journal of Educational Measurement*, 16(3): 143-152.
- _____. (1981, Spring). A response to Baker's criticism. *Journal of Education Measurement*. 18(1): 63-66.
- Shealy, R.; & Stout, W. (1993, June). A model-based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DTF as well as item bias/DIF. *Psychometrika*. 58(2): 159-194.
- Shepard, L. A.; Camilli, G.; & Williams, D. M. (1984, Summer). Accounting for statistical artifacts in item bias research. *Journal of Educational Statistics*. 9(2): 93-128.
- Stout, W.; et al. (1997, September). MULTISIB: A procedure to investigate DIF when a test is intentionally two-dimensional. *Applied Psychological Measurement*. 21(3): 195-213.
- Swaminathan, H.; & Rogers, H. J. (1990, Winter). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*. 27(4): 361-370.
- Thissen, D.; Steinberg, L.; & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In *Differential item functioning*. W. P. Holland; & H. Wainer. pp. 67-113. Hillsdale, NJ: Erlbaum.
- Zwick, R.; Donoghue, J. R.; & Grima, A. (1993, Fall). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement*. 30(3): 233-251.

- Zumbo, B. D. (1999). *A handbook of the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item score*. Ottawa, Ontario, Canada: Directorate of Human Resources Research and Evaluation, Department of National Defense.
- Whitmore, M.; & Schumacker, R. E. (1999, December). *A Comparison of logistic regression and analysis of variance differential item functioning detection methods*. *Educational and Psychological Measurement*. 59(6): 910-927.

ภาคผนวก

ภาคผนวก ก
ผลการเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง
และอัตราความคลาดเคลื่อนประเภทที่ 1
โดยใช้วิธีของเซฟเฟ (Scheffé)

ตาราง 26 การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบสามวิธี ภายใต้เงื่อนไขเดียวกันของปัจจัยรูปแบบของข้อสอบทำหน้าที่ย่อยแบบ

การเปรียบเทียบ	Poly-SIBTEST	OLR	MOLR	F	
				P	E_1
FDIF1					
1	1	-1	0	-	2,604.68***
2	1	0	-1	-	329.73***
3	0	1	-1	-	1,080.94***
FDIF2					
1	1	-1	0	5,093.03***	452.72***
2	1	0	-1	5,480.91***	71.94***
3	0	1	-1	7.12	163.72***

****p* < .001

ตาราง 28 การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขต่างกันของปัจจัยสัดส่วนของข้อสอบทำหน้าที่ย่อยแบบ

การเปรียบเทียบ	PDIF1	PDIF2	PDIF3	<i>F</i>	
				<i>P</i>	<i>E</i> ₁
1	1	-1	0	8.35	31.37***
2	1	0	-1	22.12***	212.89***
3	0	1	-1	3.29	80.81***

****p* < .001

ตาราง 29 การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภท
ที่ 1 ระหว่างวิธีการตรวจสอบสามวิธี ภายใต้เงื่อนไขเดียวกันของปัจจัยสัดส่วนของข้อสอบทำ
หน้าที่เพียงแบบ

การเปรียบเทียบ	Poly- SIBTEST	OLR	MOLR	<i>F</i>	
				<i>P</i>	<i>E</i> ₁
PDIF1					
1	1	-1	0	658.85***	417.26***
2	1	0	-1	633.59***	76.64***
3	0	1	-1	.25	136.25***
PDIF2					
1	1	-1	0	562.35***	817.55***
2	1	0	-1	657.15***	129.73***
3	0	1	-1	3.69	295.94***
PDIF2					
1	1	-1	0	596.63***	1,482.28***
2	1	0	-1	800.95***	146.31***
3	0	1	-1	15.02***	697.19***

***p < .001

ตาราง 31 การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธีการตรวจสอบ ภายใต้เงื่อนไขต่างกันของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เป็นแบบ

การเปรียบเทียบ	$d(\theta) 1$	$d(\theta) 2$	$d(\theta) 3$	F	
				P	E_1
Poly-SIBTEST					
1	1	-1	0	19.63***	.22
2	1	0	-1	169.90***	41.42***
3	0	1	-1	74.02***	35.59***
OLR					
1	1	-1	0	-	.48
2	1	0	-1	-	3,100.51***
3	0	1	-1	-	3,024.09***
MOLR					
1	1	-1	0	-	3.63
2	1	0	-1	-	214.29***
3	0	1	-1	-	162.13***

***p < .001

ตาราง 32 การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภท
ที่ 1 ระหว่างวิธีการตรวจสอบสามวิธี ภายใต้เงื่อนไขเดียวกันของปัจจัยสัดส่วนของข้อสอบทำ
หน้าที่เพียงแบบ

การเปรียบเทียบ	Poly- SIBTEST	OLR	MOLR	<i>F</i>	
				<i>P</i>	<i>E</i> ₁
<i>d(θ) 1</i>					
1	1	-1	0	999.75***	315.41***
2	1	0	-1	1,148.44***	88.19***
3	0	1	-1	5.15	70.04***
<i>d(θ) 2</i>					
1	1	-1	0	741.89***	323.31***
2	1	0	-1	834.15***	117.23***
3	0	1	-1	2.70	51.17***
<i>d(θ) 3</i>					
1	1	-1	0	241.69***	4,489.87***
2	1	0	-1	287.53***	309.55***
3	0	1	-1	1.99	2,441.59***

****p* < .001

ตาราง 34 การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภท
ที่ 1 ของแต่ละวิธีการตรวจสอบ ภายใต้เงื่อนไขต่างกันของปัจจัยขนาดตัวอย่าง

การเปรียบเทียบ	SIZE1	SIZE2	SIZE3	SIZE4	F	
					P	E ₁
Poly-SIBTEST						
1	1	-1	0	0	312.18***	58.46***
2	1	0	-1	0	797.36***	109.52***
3	1	0	0	-1	1007.59***	103.93***
4	0	1	-1	0	111.70***	7.95
5	0	1	0	-1	198.07***	6.50
6	0	0	1	-1	12.28	.07
OLR						
1	1	-1	0	0	487.59***	71.16***
2	1	0	-1	0	1,438.35***	483.59***
3	1	0	0	-1	1,853.15***	2,139.92***
4	0	1	-1	0	251.04***	183.74***
5	0	1	0	-1	439.61***	1,430.64***
6	0	0	1	-1	26.24***	588.97***
MOLR						
1	1	-1	0	0	629.16***	.43
2	1	0	-1	0	1,944.79***	13.68
3	1	0	0	-1	2,120.60***	94.06***
4	0	1	-1	0	361.64***	9.27
5	0	1	0	-1	439.61***	81.80***
6	0	0	1	-1	3.80	35.99***

***p < .001

ตาราง 35 การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภท
ที่ 1 ระหว่างวิธีการตรวจสอบสามวิธี ภายใต้เงื่อนไขเดียวกันของปัจจัยขนาดตัวอย่าง

การเปรียบเทียบ	Poly- SIBTEST	OLR	MOLR	F	
				P	E_1
SIZE1					
1	1	-1	0	414.36***	-
2	1	0	-1	369.54***	-
3	0	1	-1	1.28	-
SIZE2					
1	1	-1	0	613.48***	316.68***
2	1	0	-1	709.57***	64.11***
3	0	1	-1	3.49	95.82***
SIZE3					
1	1	-1	0	902.64***	1,167.59***
2	1	0	-1	1,231.01***	192.40***
3	0	1	-1	25.42***	412.05***
SIZE4					
1	1	-1	0	1,002.46***	3,383.52***
2	1	0	-1	1,124.32***	384.15***
3	0	1	-1	3.49	1,487.50***

*** $p < .001$

ตาราง 37 การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภท
ที่ 1 ของแต่ละวิธีการตรวจสอบ ภายใต้เงื่อนไขต่างกันของปฏิสัมพันธ์ระหว่างปัจจัยรูปแบบ
ของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน

การ เปรียบเทียบ	FDIF1 ×	FDIF1 ×	FDIF1 ×	FDIF2 ×	FDIF2 ×	FDIF2 ×	<i>F</i>	
	PDIF1	PDIF2	PDIF3	PDIF1	PDIF2	PDIF3	<i>P</i>	<i>E</i> ₁
Poly-SIBTEST								
1	1	-1	0	0	0	0	10.26	-
2	1	0	-1	0	0	0	39.80***	-
3	1	0	0	-1	0	0	1,838.38***	-
4	1	0	0	0	-1	0	1,703.31***	-
5	1	0	0	0	0	-1	1,801.30***	-
6	0	1	-1	0	0	0	9.64	-
7	0	1	0	-1	0	0	1,573.95***	-
8	0	1	0	0	-1	0	1,449.16***	-
9	0	1	0	0	0	-1	1,539.65***	-
10	0	0	1	-1	0	0	1,337.21***	-
11	0	0	1	0	-1	0	1,222.39***	-
12	0	0	1	0	0	-1	1,305.61***	-
13	0	0	0	1	-1	0	2.58	-
14	0	0	0	1	0	-1	.19	-
15	0	0	0	0	1	-1	1.37	-
OLR								
1	1	-1	0	0	0	0	23.94***	65.83***
2	1	0	-1	0	0	0	89.67***	456.53***
3	1	0	0	-1	0	0	9.56	1.62
4	1	0	0	0	-1	0	9.56	1.45
5	1	0	0	0	0	-1	2.51	1.75
6	0	1	-1	0	0	0	20.95***	175.64***
7	0	1	0	-1	0	0	3.25	88.09***

ตาราง 37 (ต่อ)

การ เปรียบเทียบ	FDIF1 × PDIF1	FDIF1 × PDIF2	FDIF1 × PDIF3	FDIF2 × PDIF1	FDIF2 × PDIF2	FDIF2 × PDIF3	<i>F</i>	
							<i>P</i>	<i>E</i> ₁
8	0	1	0	0	-1	0	3.25	86.80***
9	0	1	0	0	0	-1	10.95	89.06***
10	0	0	1	-1	0	0	40.69***	512.51***
11	0	0	1	0	-1	0	40.69***	509.40***
12	0	0	1	0	0	-1	62.18***	514.84***
13	0	0	0	1	-1	0	.00	.00
14	0	0	0	1	0	-1	2.27	.00
15	0	0	0	0	1	-1	2.27	.01

***p < .001

ตาราง 38 การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบสามวิธี ภายใต้เงื่อนไขเดียวกันของปฏิสัมพันธ์ระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถ

การเปรียบเทียบ	Poly- SIBTEST	OLR	MOLR	<i>F</i>		
				<i>P</i>	<i>E</i> ₁	
FDIF1 × PDIF1						
1	1	-1	0	-	333.88***	
2	1	0	-1	-	75.35***	
3	0	1	-1	-	92.00***	
FDIF1 × PDIF2						
1	1	-1	0	-	829.00***	
2	1	0	-1	-	130.67***	
3	0	1	-1	-	301.42***	
FDIF1 × PDIF3						
1	1	-1	0	1.76	1,866.00***	
2	1	0	-1	25.33***	143.96***	
3	0	1	-1	40.42***	973.37***	
FDIF2 × PDIF1						
1	1	-1	0	1,732.36***	133.83***	
2	1	0	-1	1,945.07***	16.74***	
3	0	1	-1	6.16	55.91***	
FDIF2 × PDIF2						
1	1	-1	0	1,601.32***	168.43***	
2	1	0	-1	1,762.83***	26.95***	
3	0	1	-1	3.88	60.63***	
FDIF2 × PDIF3						
1	1	-1	0	1,822.78***	170.22***	
2	1	0	-1	1,839.58***	32.37***	
3	0	1	-1	.04	54.13***	

****p* < .001

ตาราง 40 การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธีการตรวจสอบ ภายใต้เงื่อนไขต่างกันของปฏิสัมพันธ์ระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถ

การ เปรียบเทียบ	FDIF1 × $d(0)1$	FDIF1 × $d(0)2$	FDIF1 × $d(0)3$	FDIF2 × $d(0)1$	FDIF2 × $d(0)2$	FDIF2 × $d(0)3$	F	
							P	E_1
Poly-SIBTEST								
1	1	-1	0	0	0	0	.01	-
2	1	0	-1	0	0	0	1.39	-
3	1	0	0	-1	0	0	2,573.83***	-
4	1	0	0	0	-1	0	1,868.37***	-
5	1	0	0	0	0	-1	734.89***	-
6	0	1	-1	0	0	0	1.67	-
7	0	1	0	-1	0	0	2,585.51***	-
8	0	1	0	0	-1	0	1,878.32***	-
9	0	1	0	0	0	-1	741.13***	-
10	0	0	1	-1	0	0	2,455.66***	-
11	0	0	1	0	-1	0	1767.90***	-
12	0	0	1	0	0	-1	672.39***	-
13	0	0	0	1	-1	0	56.37***	-
14	0	0	0	1	0	-1	558.10***	-
15	0	0	0	0	1	-1	259.72***	-
OLR								
1	1	-1	0	0	0	0	.67	.07
2	1	0	-1	0	0	0	4.37	1343.91***
3	1	0	0	-1	0	0	34.80***	261.54***
4	1	0	0	0	-1	0	26.76***	222.66***
5	1	0	0	0	0	-1	1.98	810.68***
6	0	1	-1	0	0	0	1.62	1363.20***
7	0	1	0	-1	0	0	25.80***	253.13***
8	0	1	0	0	-1	0	18.96***	214.91***
9	0	1	0	0	0	-1	4.96	825.68***

ตาราง 40 (ต่อ)

การ เปรียบเทียบ	FDIF1 × $d(\theta)1$	FDIF1 × $d(\theta)2$	FDIF1 × $d(\theta)3$	FDIF2 × $d(\theta)1$	FDIF2 × $d(\theta)2$	FDIF2 × $d(\theta)3$	F	
							P	E_1
10	0	0	1	-1	0	0	14.50***	2,791.16***
11	0	0	1	0	-1	0	9.50	2,660.63***
12	0	0	1	0	0	-1	12.24	67.02***
13	0	0	0	1	-1	0	.53	1.56
14	0	0	0	1	0	-1	53.39***	1,993.14***
15	0	0	0	0	1	-1	43.32***	1,883.07***

***p < .001

ตาราง 41 การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบสามวิธี ภายใต้เงื่อนไขเดียวกันของปฏิสัมพันธ์ระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถ

การเปรียบเทียบ	Poly- SIBTEST	OLR	MOLR	<i>F</i>	
				<i>P</i>	<i>E</i> ₁
FDIF1 × <i>d</i> (θ) 1					
1	1	-1	0	1.21	637.37***
2	1	0	-1	14.72***	107.84***
3	0	1	-1	24.36***	220.86***
FDIF1 × <i>d</i> (θ) 2					
1	1	-1	0	-	568.09***
2	1	0	-1	-	128.43***
3	0	1	-1	-	156.30***
FDIF1 × <i>d</i> (θ) 3					
1	1	-1	0	-	3,226.68***
2	1	0	-1	-	254.41***
3	0	1	-1	-	1,669.02***
FDIF2 × <i>d</i> (θ) 1					
1	1	-1	0	3,083.86***	-
2	1	0	-1	2,971.58***	-
3	0	1	-1	1.04	-
FDIF2 × <i>d</i> (θ) 2					
1	1	-1	0	2,237.15***	-
2	1	0	-1	2,303.57***	-
3	0	1	-1	.49	-
FDIF2 × <i>d</i> (θ) 3					
1	1	-1	0	605.22***	1,683.88***
2	1	0	-1	881.39***	95.25***
3	0	1	-1	25.88***	978.15***

***p < .001

ตาราง 43 การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธีการตรวจสอบ ภายใต้เงื่อนไขต่างกันของปฏิสัมพันธ์ระหว่างปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยขนาดตัวอย่าง

การเปรียบเทียบ	FDIF1 x SIZE1	FDIF1 x SIZE2	FDIF1 x SIZE3	FDIF1 x SIZE4	FDIF2 x SIZE1	FDIF2 x SIZE2	FDIF2 x SIZE3	FDIF2 x SIZE4	F	
									P	E ₁
Poly-SIBTEST										
1	1	-1	0	0	0	0	0	0	1,138.85***	-
2	1	0	-1	0	0	0	0	0	3,085.46***	-
3	1	0	0	-1	0	0	0	0	3,479.35***	-
4	1	0	0	0	-1	0	0	0	381.64***	-
5	1	0	0	0	0	-1	0	0	256.97***	-
6	1	0	0	0	0	0	-1	0	241.71***	-
7	1	0	0	0	0	0	0	-1	134.48***	-
8	0	1	-1	0	0	0	0	0	475.24***	-
9	0	1	0	-1	0	0	0	0	637.01***	-
10	0	1	0	0	-1	0	0	0	2,839.04***	-
11	0	1	0	0	0	-1	0	0	2,477.77***	-
12	0	1	0	0	0	0	-1	0	2,429.90***	-
13	0	1	0	0	0	0	0	-1	2,056.01***	-
14	0	0	1	-1	0	0	0	0	11.83	-
15	0	0	1	0	-1	0	0	0	5,637.40***	-
16	0	0	1	0	0	-1	0	0	5,123.29***	-
17	0	0	1	0	0	0	-1	0	5,054.36***	-
18	0	0	1	0	0	0	0	-1	4,508.22***	-
19	0	0	0	1	-1	0	0	0	6,165.66***	-
20	0	0	0	1	0	-1	0	0	5,627.44***	-
21	0	0	0	1	0	0	-1	0	5,555.18***	-
22	0	0	0	1	0	0	0	-1	4,981.87***	-
23	0	0	0	0	1	-1	0	0	12.29	-
24	0	0	0	0	1	0	-1	0	15.91	-
25	0	0	0	0	1	0	0	-1	63.03***	-

ตาราง 43 (ต่อ)

การเปรียบเทียบ	FDIF1 X SIZE1	FDIF1 X SIZE2	FDIF1 X SIZE3	FDIF1 X SIZE4	FDIF2 X SIZE1	FDIF2 X SIZE2	FDIF2 X SIZE3	FDIF2 X SIZE4	<i>F</i>	
									<i>P</i>	<i>E</i> ₁
26	0	0	0	0	0	1	-1	0	.23	-
27	0	0	0	0	0	1	0	-1	19.66***	-
28	0	0	0	0	0	0	1	-1	15.61	-
OLR										
1	1	-1	0	0	0	0	0	0	693.20***	42.04***
2	1	0	-1	0	0	0	0	0	2,362.19***	442.97***
3	1	0	0	-1	0	0	0	0	2,870.42***	2,032.75***
4	1	0	0	0	-1	0	0	0	166.51***	7.16
5	1	0	0	0	0	-1	0	0	1,097.04***	10.73
6	1	0	0	0	0	0	-1	0	1,959.23***	75.60***
7	1	0	0	0	0	0	0	-1	2,509.94***	416.10***
8	0	1	-1	0	0	0	0	0	496.12***	212.09***
9	0	1	0	-1	0	0	0	0	742.43***	1,490.15***
10	0	1	0	0	-1	0	0	0	180.23***	83.91***
11	0	1	0	0	0	-1	0	0	46.14***	10.29
12	0	1	0	0	0	0	-1	0	321.65***	4.89
13	0	1	0	0	0	0	0	-1	565.04***	193.62***
14	0	0	1	-1	0	0	0	0	24.74***	577.88***
15	0	0	1	0	-1	0	0	0	1,274.39***	562.80***
16	0	0	1	0	0	-1	0	0	239.65***	315.83***
17	0	0	1	0	0	0	-1	0	18.83***	152.57***
18	0	0	1	0	0	0	0	-1	2.24	.42
19	0	0	0	1	-1	0	0	0	1,654.25***	2,281.27***
20	0	0	0	1	0	-1	0	0	418.39***	1,748.16***
21	0	0	0	1	0	0	-1	0	86.73***	1,324.33***
22	0	0	0	1	0	0	0	-1	12.09	609.48***
23	0	0	0	0	1	-1	0	0	408.76***	35.42***
24	0	0	0	0	1	0	-1	0	983.41***	129.31***
25	0	0	0	0	1	0	0	-1	1,383.51***	532.46***
26	0	0	0	0	0	1	-1	0	124.14***	29.37***
27	0	0	0	0	0	1	0	-1	288.24***	293.21***
28	0	0	0	0	0	0	1	-1	34.06***	136.98***

ตาราง 43 (ต่อ)

การเปรียบเทียบ	FDIF1 X SIZE1	FDIF1 X SIZE2	FDIF1 X SIZE3	FDIF1 X SIZE4	FDIF2 X SIZE1	FDIF2 X SIZE2	FDIF2 X SIZE3	FDIF2 X SIZE4	F	
									P	E ₁
MOLR										
1	1	-1	0	0	0	0	0	0	751.24***	-
2	1	0	-1	0	0	0	0	0	2,408.46***	-
3	1	0	0	-1	0	0	0	0	2,665.10***	-
4	1	0	0	0	-1	0	0	0	41.38***	-
5	1	0	0	0	0	-1	0	0	1,018.78***	-
6	1	0	0	0	0	0	-1	0	2,533.73***	-
7	1	0	0	0	0	0	0	-1	2,695.51***	-
8	0	1	-1	0	0	0	0	0	469.48***	-
9	0	1	0	-1	0	0	0	0	586.41***	-
10	0	1	0	0	-1	0	0	0	439.98***	-
11	0	1	0	0	0	-1	0	0	20.34***	-
12	0	1	0	0	0	0	-1	0	525.67***	-
13	0	1	0	0	0	0	0	-1	600.72***	-
14	0	0	1	-1	0	0	0	0	6.50	-
15	0	0	1	0	-1	0	0	0	1,818.44***	-
16	0	0	1	0	0	-1	0	0	294.39***	-
17	0	0	1	0	0	0	-1	0	1.59	-
18	0	0	1	0	0	0	0	-1	8.08	-
19	0	0	0	1	-1	0	0	0	2,042.29***	-
20	0	0	0	1	0	-1	0	0	388.34***	-
21	0	0	0	1	0	0	-1	0	1.66	-
22	0	0	0	1	0	0	0	-1	.09	-
23	0	0	0	0	1	-1	0	0	649.51***	-
24	0	0	0	0	1	0	-1	0	1,927.49***	-
25	0	0	0	0	1	0	0	-1	2,068.92***	-
26	0	0	0	0	0	1	-1	0	339.21***	-
27	0	0	0	0	0	1	0	-1	400.00***	-
28	0	0	0	0	0	0	1	-1	2.50	-

***p < .001

ตาราง 44 การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภท
ที่ 1 ระหว่างวิธีการตรวจสอบสามวิธี ภายใต้เงื่อนไขเดียวกันของปฏิสัมพันธ์ระหว่างปัจจัย
รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยขนาดตัวอย่าง

การเปรียบเทียบ	Poly- SIBTEST	OLR	MOLR	<i>F</i>	
				<i>P</i>	<i>E</i> ₁
FDIF1 × SIZE1					
1	1	-1	0	27.45***	22.92***
2	1	0	-1	52.94***	2.99
3	0	1	-1	4.15	9.36
FDIF1 × SIZE2					
1	1	-1	0	-	322.97***
2	1	0	-1	-	75.89***
3	0	1	-1	-	85.75***
FDIF1 × SIZE3					
1	1	-1	0	-	1,183.11***
2	1	0	-1	-	176.00***
3	0	1	-1	-	446.47***
FDIF1 × SIZE4					
1	1	-1	0	-	3,508.62***
2	1	0	-1	-	312.30***
3	0	1	-1	-	1,727.37***
FDIF2 × SIZE1					
1	1	-1	0	1,419.68***	-
2	1	0	-1	1,105.22***	-
3	0	1	-1	19.66***	-

ตาราง 44 (ต่อ)

การเปรียบเทียบ	Poly- SIBTEST	OLR	MOLR	F	
				P	E_1
FDIF2 × SIZE2					
1	1	-1	0	2,958.39***	68.27***
2	1	0	-1	3,049.78***	9.46
3	0	1	-1	.70	26.90***
FDIF2 × SIZE3					
1	1	-1	0	4,231.44***	255.77***
2	1	0	-1	5,352.29***	51.58***
3	0	1	-1	65.77***	77.63***
FDIF2 × SIZE4					
1	1	-1	0	4,480.27***	702.23***
2	1	0	-1	5,011.34***	126.30***
3	0	1	-1	14.87***	232.91***

***p < .001

ตาราง 46 การเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอย-
โลจิสติกแบบจัดอันดับ ภายใต้เงื่อนไขต่างกันของปฏิสัมพันธ์ระหว่างปัจจัยสัดส่วนของข้อสอบ
ทำหน้าที่เป็นเบง กับปัจจัยความแตกต่างของการแจกแจงความสามารถ

การเปรียบเทียบ	PDF1 X $d(\theta)1$	PDF1 X $d(\theta)2$	PDF1 X $d(\theta)3$	PDF2 X $d(\theta)1$	PDF2 X $d(\theta)2$	PDF2 X $d(\theta)3$	PDF3 X $d(\theta)1$	PDF3 X $d(\theta)2$	PDF3 X $d(\theta)3$	F
1	1	-1	0	0	0	0	0	0	0	4.65
2	1	0	-1	0	0	0	0	0	0	1,383.87***
3	1	0	0	-1	0	0	0	0	0	39.55***
4	1	0	0	0	-1	0	0	0	0	35.20***
5	1	0	0	0	0	-1	0	0	0	1,498.30***
6	1	0	0	0	0	0	-1	0	0	192.97***
7	1	0	0	0	0	0	0	-1	0	176.67***
8	1	0	0	0	0	0	0	0	-1	1,789.83***
9	0	1	-1	0	0	0	0	0	0	1,228.13***
10	0	1	0	-1	0	0	0	0	0	17.08
11	0	1	0	0	-1	0	0	0	0	14.26
12	0	1	0	0	0	-1	0	0	0	1,336.05***
13	0	1	0	0	0	0	-1	0	0	137.72***
14	0	1	0	0	0	0	0	-1	0	124.01***
15	0	1	0	0	0	0	0	0	-1	1,612.07***
16	0	0	1	-1	0	0	0	0	0	955.50***
17	0	0	1	0	-1	0	0	0	0	977.68***
18	0	0	1	0	0	-1	0	0	0	2.27
19	0	0	1	0	0	0	-1	0	0	543.31***
20	0	0	1	0	0	0	0	-1	0	571.63***
21	0	0	1	0	0	0	0	0	-1	26.07***
22	0	0	0	1	-1	0	0	0	0	.13
23	0	0	0	1	0	-1	0	0	0	1,050.97***
24	0	0	0	1	0	0	-1	0	0	57.79***

ตาราง 46 (ต่อ)

การเปรียบเทียบ	PDF1 X $d(\theta)1$	PDF1 X $d(\theta)2$	PDF1 X $d(\theta)3$	PDF2 X $d(\theta)1$	PDF2 X $d(\theta)2$	PDF2 X $d(\theta)3$	PDF3 X $d(\theta)1$	PDF3 X $d(\theta)2$	PDF3 X $d(\theta)3$	<i>F</i>
25	0	0	0	1	0	0	0	-1	0	49.03***
26	0	0	0	1	0	0	0	0	-1	1,297.24***
27	0	0	0	0	1	-1	0	0	0	1,074.22***
28	0	0	0	0	1	0	-1	0	0	63.34***
29	0	0	0	0	1	0	0	-1	0	54.16***
30	0	0	0	0	1	0	0	0	-1	1,323.05***
31	0	0	0	0	0	1	-1	0	0	615.86***
32	0	0	0	0	0	1	0	-1	0	645.99***
33	0	0	0	0	0	1	0	0	-1	12.95
34	0	0	0	0	0	0	1	-1	0	.36
35	0	0	0	0	0	0	1	0	-1	807.41***
36	0	0	0	0	0	0	0	1	-1	841.85***

***p < .001

ตาราง 47 การเปรียบเทียบค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบสามวิธี ภายใต้เงื่อนไขเดียวกันของปฏิสัมพันธ์ระหว่างปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยความแตกต่างของการแจกแจงความสามารถ

การเปรียบเทียบ	Poly-SIBTEST	OLR	MOLR	<i>F</i>
PDIF1 × $d(\theta)$ 1				
1	1	-1	0	-
2	1	0	-1	-
3	0	1	-1	-
PDIF1 × $d(\theta)$ 2				
1	1	-1	0	-
2	1	0	-1	-
3	0	1	-1	-
PDIF1 × $d(\theta)$ 3				
1	1	-1	0	1,241.80***
2	1	0	-1	68.61***
3	0	1	-1	726.63***
PDIF2 × $d(\theta)$ 1				
1	1	-1	0	103.62***
2	1	0	-1	45.25***
3	0	1	-1	11.92
PDIF2 × $d(\theta)$ 2				
1	1	-1	0	97.45***
2	1	0	-1	42.67***
3	0	1	-1	11.15

ตาราง 47 (ต่อ)

การเปรียบเทียบ	Poly-SIBTEST	OLR	MOLR	<i>F</i>
PDIF2 × <i>d(θ)</i> 3				
1	1	-1	0	1,523.51***
2	1	0	-1	104.61***
3	0	1	-1	829.67***
PDIF3 × <i>d(θ)</i> 1				
1	1	-1	0	349.29***
2	1	0	-1	28.96***
3	0	1	-1	177.10***
PDIF3 × <i>d(θ)</i> 2				
1	1	-1	0	293.55***
2	1	0	-1	50.64***
3	0	1	-1	100.35***
PDIF3 × <i>d(θ)</i> 3				
1	1	-1	0	1,911.14***
2	1	0	-1	156.19***
3	0	1	-1	974.64***

***p < .001

ตาราง 49 การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภท
ที่ 1 ของแต่ละวิธีการตรวจสอบ ภายใต้เงื่อนไขต่างกันของปฏิสัมพันธ์ระหว่างปัจจัยสัดส่วน
ของข้อสอบทำหน้าที่เบี่ยงเบน กับปัจจัยขนาดตัวอย่าง

การเปรียบเทียบ	PDIF1 X SIZE1	PDIF1 X SIZE2	PDIF1 X SIZE3	PDIF1 X SIZE4	PDIF2 X SIZE1	PDIF2 X SIZE2	PDIF2 X SIZE3	PDIF2 X SIZE4	PDIF3 X SIZE1	PDIF3 X SIZE2	PDIF3 X SIZE3	PDIF3 X SIZE4	F	
													P	E ₁
Poly-SIBTEST														
1	1	-1	0	0	0	0	0	0	0	0	0	0	112.47***	-
2	1	0	-1	0	0	0	0	0	0	0	0	0	222.27***	-
3	1	0	0	-1	0	0	0	0	0	0	0	0	235.53***	-
4	1	0	0	0	-1	0	0	0	0	0	0	0	8.87	-
5	1	0	0	0	0	-1	0	0	0	0	0	0	79.38***	-
6	1	0	0	0	0	0	-1	0	0	0	0	0	226.18***	-
7	1	0	0	0	0	0	0	-1	0	0	0	0	306.85***	-
8	1	0	0	0	0	0	0	0	-1	0	0	0	8.75	-
9	1	0	0	0	0	0	0	0	0	-1	0	0	28.26***	-
10	1	0	0	0	0	0	0	0	0	0	-1	0	176.53***	-
11	1	0	0	0	0	0	0	0	0	0	0	-1	271.38***	-
12	0	1	-1	0	0	0	0	0	0	0	0	0	18.52	-
13	0	1	0	-1	0	0	0	0	0	0	0	0	22.48	-
14	0	1	0	0	-1	0	0	0	0	0	0	0	184.52***	-
15	0	1	0	0	0	-1	0	0	0	0	0	0	2.87	-
16	0	1	0	0	0	0	-1	0	0	0	0	0	19.66	-
17	0	1	0	0	0	0	0	-1	0	0	0	0	47.77***	-
18	0	1	0	0	0	0	0	0	-1	0	0	0	183.95***	-
19	0	1	0	0	0	0	0	0	0	-1	0	0	27.98***	-
20	0	1	0	0	0	0	0	0	0	0	-1	0	7.19	-
21	0	1	0	0	0	0	0	0	0	0	0	-1	34.44***	-
22	0	0	1	-1	0	0	0	0	0	0	0	0	.19	-
23	0	0	1	0	-1	0	0	0	0	0	0	0	319.96***	-
24	0	0	1	0	0	-1	0	0	0	0	0	0	35.99***	-
25	0	0	1	0	0	0	-1	0	0	0	0	0	.02	-

ตาราง 49 (ต่อ)

การเปรียบเทียบ	PDIF1 X SIZE1	PDIF1 X SIZE2	PDIF1 X SIZE3	PDIF1 X SIZE4	PDIF2 X SIZE1	PDIF2 X SIZE2	PDIF2 X SIZE3	PDIF2 X SIZE4	PDIF3 X SIZE1	PDIF3 X SIZE2	PDIF3 X SIZE3	PDIF3 X SIZE4	<i>F</i>	
													<i>P</i>	<i>E</i> ₁
26	0	0	1	0	0	0	0	-1	0	0	0	0	6.80	-
27	0	0	1	0	0	0	0	0	-1	0	0	0	319.21***	-
28	0	0	1	0	0	0	0	0	0	-1	0	0	92.03***	-
29	0	0	1	0	0	0	0	0	0	0	-1	0	2.63	-
30	0	0	1	0	0	0	0	0	0	0	0	-1	2.45	-
31	0	0	0	1	-1	0	0	0	0	0	0	0	335.83***	-
32	0	0	0	1	0	-1	0	0	0	0	0	0	41.44***	-
33	0	0	0	1	0	0	-1	0	0	0	0	0	.09	-
34	0	0	0	1	0	0	0	-1	0	0	0	0	4.71	-
35	0	0	0	1	0	0	0	0	-1	0	0	0	335.06***	-
36	0	0	0	1	0	0	0	0	0	-1	0	0	100.63***	-
37	0	0	0	1	0	0	0	0	0	0	-1	0	4.25	-
38	0	0	0	1	0	0	0	0	0	0	0	-1	1.27	-
39	0	0	0	0	1	-1	0	0	0	0	0	0	141.33***	-
40	0	0	0	0	1	0	-1	0	0	0	0	0	324.64***	-
41	0	0	0	0	1	0	0	-1	0	0	0	0	420.07***	-
42	0	0	0	0	1	0	0	0	-1	0	0	0	.00	-
43	0	0	0	0	1	0	0	0	0	-1	0	0	68.79***	-
44	0	0	0	0	1	0	0	0	0	0	-1	0	264.55***	-
45	0	0	0	0	1	0	0	0	0	0	0	-1	378.39***	-
46	0	0	0	0	0	1	-1	0	0	0	0	0	37.57***	-
47	0	0	0	0	0	1	0	-1	0	0	0	0	74.08***	-
48	0	0	0	0	0	1	0	0	-1	0	0	0	140.84***	-
49	0	0	0	0	0	1	0	0	0	-1	0	0	12.92	-
50	0	0	0	0	0	1	0	0	0	0	-1	0	19.16	-
51	0	0	0	0	0	1	0	0	0	0	0	-1	57.21***	-
52	0	0	0	0	0	0	1	-1	0	0	0	0	6.14	-
53	0	0	0	0	0	0	1	0	-1	0	0	0	323.89***	-

ตาราง 49 (ต่อ)

การเปรียบเทียบ	PDIF1 X SIZE1	PDIF1 X SIZE2	PDIF1 X SIZE3	PDIF1 X SIZE4	PDIF2 X SIZE1	PDIF2 X SIZE2	PDIF2 X SIZE3	PDIF2 X SIZE4	PDIF3 X SIZE1	PDIF3 X SIZE2	PDIF3 X SIZE3	PDIF3 X SIZE4	<i>F</i>	
													<i>P</i>	<i>E</i> ₁
54	0	0	0	0	0	0	1	0	0	-1	0	0	94.55***	-
55	0	0	0	0	0	0	1	0	0	0	-1	0	3.07	-
56	0	0	0	0	0	0	1	0	0	0	0	-1	2.06	-
57	0	0	0	0	0	0	0	1	-1	0	0	0	419.22***	-
58	0	0	0	0	0	0	0	1	0	-1	0	0	148.87***	-
59	0	0	0	0	0	0	0	1	0	0	-1	0	17.90	-
60	0	0	0	0	0	0	0	1	0	0	0	-1	1.09	-
61	0	0	0	0	0	0	0	0	1	-1	0	0	68.45***	-
62	0	0	0	0	0	0	0	0	1	0	-1	0	263.87***	-
63	0	0	0	0	0	0	0	0	1	0	0	-1	377.58***	-
64	0	0	0	0	0	0	0	0	0	1	-1	0	63.53**	-
65	0	0	0	0	0	0	0	0	0	1	0	-1	124.50***	-
66	0	0	0	0	0	0	0	0	0	0	1	-1	10.16	-
OLR														
1	1	-1	0	0	0	0	0	0	0	0	0	0	129.80***	16.99
2	1	0	-1	0	0	0	0	0	0	0	0	0	362.53***	117.28***
3	1	0	0	-1	0	0	0	0	0	0	0	0	450.10***	450.93***
4	1	0	0	0	-1	0	0	0	0	0	0	0	22.63***	4.28
5	1	0	0	0	0	-1	0	0	0	0	0	0	98.23***	40.65***
6	1	0	0	0	0	0	-1	0	0	0	0	0	343.90***	166.00***
7	1	0	0	0	0	0	0	-1	0	0	0	0	427.37***	755.07***
8	1	0	0	0	0	0	0	0	-1	0	0	0	37.89***	7.03
9	1	0	0	0	0	0	0	0	0	-1	0	0	38.86***	82.69***
10	1	0	0	0	0	0	0	0	0	0	-1	0	307.94***	391.13***
11	1	0	0	0	0	0	0	0	0	0	0	-1	491.51***	1,408.20***
12	0	1	-1	0	0	0	0	0	0	0	0	0	58.48***	44.99***
13	0	1	0	-1	0	0	0	0	0	0	0	0	96.48***	292.85***
14	0	1	0	0	-1	0	0	0	0	0	0	0	260.83***	4.22
15	0	1	0	0	0	-1	0	0	0	0	0	0	2.19	5.08
16	0	1	0	0	0	0	-1	0	0	0	0	0	51.15***	76.77***

ตาราง 49 (ต่อ)

การเปรียบเทียบ	PDIF1 X SIZE1	PDIF1 X SIZE2	PDIF1 X SIZE3	PDIF1 X SIZE4	PDIF2 X SIZE1	PDIF2 X SIZE2	PDIF2 X SIZE3	PDIF2 X SIZE4	PDIF3 X SIZE1	PDIF3 X SIZE2	PDIF3 X SIZE3	PDIF3 X SIZE4	<i>F</i>	
													<i>P</i>	<i>E</i> ₁
17	0	1	0	0	0	0	0	-1	0	0	0	0	86.12***	545.50***
18	0	1	0	0	0	0	0	0	-1	0	0	0	307.94***	2.16
19	0	1	0	0	0	0	0	0	0	-1	0	0	26.62***	24.71***
20	0	1	0	0	0	0	0	0	0	0	-1	0	37.89***	245.07***
21	0	1	0	0	0	0	0	0	0	0	0	-1	116.15***	1,115.80***
22	0	0	1	-1	0	0	0	0	0	0	0	0	4.73	108.27***
23	0	0	1	0	-1	0	0	0	0	0	0	0	566.33***	76.77***
24	0	0	1	0	0	-1	0	0	0	0	0	0	83.34***	19.84
25	0	0	1	0	0	0	-1	0	0	0	0	0	.25	4.22
26	0	0	1	0	0	0	0	-1	0	0	0	0	2.67	277.18***
27	0	0	1	0	0	0	0	0	-1	0	0	0	634.82***	66.89***
28	0	0	1	0	0	0	0	0	0	-1	0	0	164.01***	3.02
29	0	0	1	0	0	0	0	0	0	0	-1	0	2.23	80.06***
30	0	0	1	0	0	0	0	0	0	0	0	-1	9.80	712.70***
31	0	0	0	1	-1	0	0	0	0	0	0	0	674.59***	367.39***
32	0	0	0	1	0	-1	0	0	0	0	0	0	127.78***	220.80***
33	0	0	0	1	0	0	-1	0	0	0	0	0	7.13	69.74***
34	0	0	0	1	0	0	0	-1	0	0	0	0	.29	38.98***
35	0	0	0	1	0	0	0	0	-1	0	0	0	749.17***	345.37***
36	0	0	0	1	0	0	0	0	0	-1	0	0	224.45***	147.43***
37	0	0	0	1	0	0	0	0	0	0	-1	0	13.45	2.13
38	0	0	0	1	0	0	0	0	0	0	0	-1	.91	265.39***
39	0	0	0	0	1	-1	0	0	0	0	0	0	215.17***	18.56
40	0	0	0	0	1	0	-1	0	0	0	0	0	542.99***	116.99***
41	0	0	0	0	1	0	0	-1	0	0	0	0	646.71***	645.70***
42	0	0	0	0	1	0	0	0	-1	0	0	0	1.95	.34
43	0	0	0	0	1	0	0	0	0	-1	0	0	120.81***	49.36***
44	0	0	0	0	1	0	0	0	0	0	-1	0	497.55***	313.62***
45	0	0	0	0	1	0	0	0	0	0	0	-1	725.09***	1,257.28***
46	0	0	0	0	0	1	-1	0	0	0	0	0	74.53***	42.36***
47	0	0	0	0	0	1	0	-1	0	0	0	0	115.81***	445.32***
48	0	0	0	0	0	1	0	0	-1	0	0	0	258.14***	13.87
49	0	0	0	0	0	1	0	0	0	-1	0	0	13.52	7.38

ตาราง 49 (ต่อ)

การเปรียบเทียบ	PDIF1 X SIZE1	PDIF1 X SIZE2	PDIF1 X SIZE3	PDIF1 X SIZE4	PDIF2 X SIZE1	PDIF2 X SIZE2	PDIF2 X SIZE3	PDIF2 X SIZE4	PDIF3 X SIZE1	PDIF3 X SIZE2	PDIF3 X SIZE3	PDIF3 X SIZE4	<i>F</i>	
													<i>P</i>	<i>E</i> ₁
50	0	0	0	0	0	1	0	0	0	0	-1	0	58.32***	179.59***
51	0	0	0	0	0	1	0	0	0	0	0	-1	150.28***	970.34***
52	0	0	0	0	0	0	1	-1	0	0	0	0	4.53	212.99*
53	0	0	0	0	0	0	1	0	-1	0	0	0	610.10***	104.72*
54	0	0	0	0	0	0	1	0	0	-1	0	0	151.56***	14.37
55	0	0	0	0	0	0	1	0	0	0	-1	0	.99	47.51***
56	0	0	0	0	0	0	1	0	0	0	0	-1	13.14	607.22***
57	0	0	0	0	0	0	0	1	-1	0	0	0	719.77***	616.40***
58	0	0	0	0	0	0	0	1	0	-1	0	0	208.49***	338.02***
59	0	0	0	0	0	0	0	1	0	0	-1	0	9.76	59.31***
60	0	0	0	0	0	0	0	1	0	0	0	-1	2.24	100.95***
61	0	0	0	0	0	0	0	0	1	-1	0	0	153.49***	41.50***
62	0	0	0	0	0	0	0	0	1	0	-1	0	561.87***	293.30***
63	0	0	0	0	0	0	0	0	1	0	0	-1	802.34***	1,216.26***
64	0	0	0	0	0	0	0	0	0	1	-1	0	128.02***	114.14***
65	0	0	0	0	0	0	0	0	0	1	0	-1	253.97***	808.43***
66	0	0	0	0	0	0	0	0	0	0	1	-1	21.36	315.03***

***p < .001

ตาราง 50 การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบสามวิธี ภายใต้เงื่อนไขเดียวกันของปฏิสัมพันธ์ระหว่างปัจจัย สัดส่วนของข้อสอบทำหน้าที่เปี่ยงเบน กับปัจจัยขนาดตัวอย่าง

การเปรียบเทียบ	Poly- SIBTEST	OLR	MOLR	<i>F</i>	
				<i>P</i>	<i>E</i> ₁
PDIF1 × SIZE1					
1	1	-1	0	181.55***	-
2	1	0	-1	120.92***	-
3	0	1	-1	6.14	-
PDIF1 × SIZE2					
1	1	-1	0	203.40***	52.19***
2	1	0	-1	142.95***	14.88***
3	0	1	-1	5.32	11.34
PDIF1 × SIZE3					
1	1	-1	0	309.96***	217.27***
2	1	0	-1	392.94***	49.17***
3	0	1	-1	4.92	59.72***
PDIF1 × SIZE4					
1	1	-1	0	374.14***	688.84***
2	1	0	-1	426.29***	126.95***
3	0	1	-1	1.70	224.36***
PDIF2 × SIZE1					
1	1	-1	0	136.78***	-
2	1	0	-1	123.92***	-
3	0	1	-1	.32	-
PDIF2 × SIZE2					
1	1	-1	0	209.55***	113.58***
2	1	0	-1	223.67***	28.11***
3	0	1	-1	.23	28.68***

ตาราง 50 (ต่อ)

การเปรียบเทียบ	Poly- SIBTEST	OLR	MOLR	<i>F</i>	
				<i>P</i>	<i>E</i> ₁
PDIF2 × SIZE3					
1	1	-1	0	288.31***	366.37***
2	1	0	-1	406.92***	65.38***
3	0	1	-1	10.19	122.21***
PDIF2 × SIZE4					
1	1	-1	0	276.56***	1,085.90***
2	1	0	-1	336.59***	131.77***
3	0	1	-1	2.95	461.12***
PDIF3 × SIZE1					
1	1	-1	0	105.61***	-
2	1	0	-1	128.73***	-
3	0	1	-1	1.14	-
PDIF3 × SIZE2					
1	1	-1	0	207.14***	182.45***
2	1	0	-1	379.41***	24.45***
3	0	1	-1	25.87***	73.32***
PDIF3 × SIZE3					
1	1	-1	0	314.57***	696.51***
2	1	0	-1	445.24***	87.58***
3	0	1	-1	11.32	290.13***
PDIF3 × SIZE4					
1	1	-1	0	367.51***	1,878.80***
2	1	0	-1	376.57***	139.80***
3	0	1	-1	.06	993.60***

***p < .001

ตาราง 52 การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภท
ที่ 1 ของแต่ละวิธีการตรวจสอบ ภายใต้เงื่อนไขต่างกันของปฏิสัมพันธ์ระหว่างปัจจัยความ-
แตกต่างของการแจกแจงความสามารถ กับปัจจัยขนาดตัวอย่าง

การเปรียบเทียบ	d(θ)1 × SIZE1	d(θ)1 × SIZE2	d(θ)1 × SIZE3	d(θ)1 × SIZE4	d(θ)2 × SIZE1	d(θ)2 × SIZE2	d(θ)2 × SIZE3	d(θ)2 × SIZE4	d(θ)3 × SIZE1	d(θ)3 × SIZE2	d(θ)3 × SIZE3	d(θ)3 × SIZE4	F	
													P	E ₁
Poly-SIBTEST														
1	1	-1	0	0	0	0	0	0	0	0	0	0	114.20***	5.79
2	1	0	-1	0	0	0	0	0	0	0	0	0	274.58***	14.75
3	1	0	0	-1	0	0	0	0	0	0	0	0	274.58***	9.64
4	1	0	0	0	-1	0	0	0	0	0	0	0	5.82	7.20
5	1	0	0	0	0	-1	0	0	0	0	0	0	193.51***	5.62
6	1	0	0	0	0	0	-1	0	0	0	0	0	323.84***	17.24
7	1	0	0	0	0	0	0	-1	0	0	0	0	429.04***	17.86
8	1	0	0	0	0	0	0	0	-1	0	0	0	53.42***	149.30***
9	1	0	0	0	0	0	0	0	0	-1	0	0	264.02***	.98
10	1	0	0	0	0	0	0	0	0	0	-1	0	619.92***	7.40
11	1	0	0	0	0	0	0	0	0	0	0	-1	803.69***	7.40
12	0	1	-1	0	0	0	0	0	0	0	0	0	34.62***	2.05
13	0	1	0	-1	0	0	0	0	0	0	0	0	34.62***	.49
14	0	1	0	0	-1	0	0	0	0	0	0	0	68.47***	25.90***
15	0	1	0	0	0	-1	0	0	0	0	0	0	10.40	.00
16	0	1	0	0	0	0	-1	0	0	0	0	0	53.42***	3.05
17	0	1	0	0	0	0	0	-1	0	0	0	0	100.54***	3.31
18	0	1	0	0	0	0	0	0	-1	0	0	0	11.41	213.91***
19	0	1	0	0	0	0	0	0	0	-1	0	0	30.94***	11.55
20	0	1	0	0	0	0	0	0	0	0	-1	0	201.97***	.10
21	0	1	0	0	0	0	0	0	0	0	0	-1	311.98***	.10
22	0	0	1	-1	0	0	0	0	0	0	0	0	.00	.54
23	0	0	1	0	-1	0	0	0	0	0	0	0	200.47***	42.55***
24	0	0	1	0	0	-1	0	0	0	0	0	0	7.07	2.16
25	0	0	1	0	0	0	-1	0	0	0	0	0	2.03	.10
26	0	0	1	0	0	0	0	-1	0	0	0	0	17.16	.15
27	0	0	1	0	0	0	0	0	-1	0	0	0	85.78***	257.89***
28	0	0	1	0	0	0	0	0	0	-1	0	0	.10	23.35***
29	0	0	1	0	0	0	0	0	0	0	-1	0	69.35***	1.26

ตาราง 52 (ต่อ)

การเปรียบเทียบ	$d(\theta)1 \times \text{SIZE1}$	$d(\theta)1 \times \text{SIZE2}$	$d(\theta)1 \times \text{SIZE3}$	$d(\theta)1 \times \text{SIZE4}$	$d(\theta)2 \times \text{SIZE1}$	$d(\theta)2 \times \text{SIZE2}$	$d(\theta)2 \times \text{SIZE3}$	$d(\theta)2 \times \text{SIZE4}$	$d(\theta)3 \times \text{SIZE1}$	$d(\theta)3 \times \text{SIZE2}$	$d(\theta)3 \times \text{SIZE3}$	$d(\theta)3 \times \text{SIZE4}$	F	
													P	E_1
30	0	0	1	0	0	0	0	0	0	0	0	-1	138.74***	1.26
31	0	0	0	1	-1	0	0	0	0	0	0	0	200.47***	33.50***
32	0	0	0	1	0	-1	0	0	0	0	0	0	7.07	.54
33	0	0	0	1	0	0	-1	0	0	0	0	0	2.03	1.10
34	0	0	0	1	0	0	0	-1	0	0	0	0	17.16	1.26
35	0	0	0	1	0	0	0	0	-1	0	0	0	85.78***	234.83***
36	0	0	0	1	0	0	0	0	0	-1	0	0	.10	16.79
37	0	0	0	1	0	0	0	0	0	0	-1	0	69.35***	.15
38	0	0	0	1	0	0	0	0	0	0	0	-1	138.74***	.15
39	0	0	0	0	1	-1	0	0	0	0	0	0	132.23***	25.53***
40	0	0	0	0	1	0	-1	0	0	0	0	0	242.85***	46.72***
41	0	0	0	0	1	0	0	-1	0	0	0	0	334.95***	47.73***
42	0	0	0	0	1	0	0	0	-1	0	0	0	23.98***	90.94***
43	0	0	0	0	1	0	0	0	0	-1	0	0	191.46***	2.86
44	0	0	0	0	1	0	0	0	0	0	-1	0	505.64***	29.18*
45	0	0	0	0	1	0	0	0	0	0	0	-1	672.77***	29.18*
46	0	0	0	0	0	1	-1	0	0	0	0	0	16.68	3.18
47	0	0	0	0	0	1	0	-1	0	0	0	0	46.27***	3.44
48	0	0	0	0	0	1	0	0	-1	0	0	0	43.58***	212.84***
49	0	0	0	0	0	1	0	0	0	-1	0	0	5.47	11.31
50	0	0	0	0	0	1	0	0	0	0	-1	0	120.72***	.12
51	0	0	0	0	0	1	0	0	0	0	0	-1	208.47***	.12
52	0	0	0	0	0	0	1	-1	0	0	0	0	7.39	.01
53	0	0	0	0	0	0	1	0	-1	0	0	0	114.20***	268.02***
54	0	0	0	0	0	0	1	0	0	-1	0	0	3.05	26.47***
55	0	0	0	0	0	0	1	0	0	0	-1	0	47.65***	2.05
56	0	0	0	0	0	0	1	0	0	0	0	-1	107.20***	2.05
57	0	0	0	0	0	0	0	1	-1	0	0	0	179.67***	270.43***
58	0	0	0	0	0	0	0	1	0	-1	0	0	19.93	27.23***
59	0	0	0	0	0	0	0	1	0	0	-1	0	17.51	2.27
60	0	0	0	0	0	0	0	1	0	0	0	-1	58.31***	2.27
61	0	0	0	0	0	0	0	0	1	-1	0	0	79.92***	126.04***

ตาราง 52 (ต่อ)

การเปรียบเทียบ	$d(\theta)1 \times \text{SIZE1}$	$d(\theta)1 \times \text{SIZE2}$	$d(\theta)1 \times \text{SIZE3}$	$d(\theta)1 \times \text{SIZE4}$	$d(\theta)2 \times \text{SIZE1}$	$d(\theta)2 \times \text{SIZE2}$	$d(\theta)2 \times \text{SIZE3}$	$d(\theta)2 \times \text{SIZE4}$	$d(\theta)3 \times \text{SIZE1}$	$d(\theta)3 \times \text{SIZE2}$	$d(\theta)3 \times \text{SIZE3}$	$d(\theta)3 \times \text{SIZE4}$	F	
													P	E_1
62	0	0	0	0	0	0	0	0	1	0	-1	0	309.38***	223.15***
63	0	0	0	0	0	0	0	0	1	0	0	-1	442.70***	223.15***
64	0	0	0	0	0	0	0	0	0	1	-1	0	74.81***	13.78
65	0	0	0	0	0	0	0	0	0	1	0	-1	146.43***	13.78
66	0	0	0	0	0	0	0	0	0	0	1	-1	11.91	.00
OLR														
1	1	-1	0	0	0	0	0	0	0	0	0	0	193.51***	25.72***
2	1	0	-1	0	0	0	0	0	0	0	0	0	633.12***	32.24***
3	1	0	0	-1	0	0	0	0	0	0	0	0	716.15***	285.13***
4	1	0	0	0	-1	0	0	0	0	0	0	0	.80	.35
5	1	0	0	0	0	-1	0	0	0	0	0	0	220.06***	5.36
6	1	0	0	0	0	0	-1	0	0	0	0	0	579.60***	63.30***
7	1	0	0	0	0	0	0	-1	0	0	0	0	684.59***	390.15***
8	1	0	0	0	0	0	0	0	-1	0	0	0	1.46	143.08***
9	1	0	0	0	0	0	0	0	0	-1	0	0	150.16***	605.31***
10	1	0	0	0	0	0	0	0	0	0	-1	0	387.26***	2,650.64***
11	1	0	0	0	0	0	0	0	0	0	0	-1	625.97***	7,700.94***
12	0	1	-1	0	0	0	0	0	0	0	0	0	126.59***	.37
13	0	1	0	-1	0	0	0	0	0	0	0	0	165.13***	139.58***
14	0	1	0	0	-1	0	0	0	0	0	0	0	169.49***	32.03***
15	0	1	0	0	0	-1	0	0	0	0	0	0	.85	7.60
16	0	1	0	0	0	0	-1	0	0	0	0	0	103.31***	8.32
17	0	1	0	0	0	0	0	-1	0	0	0	0	150.16***	215.53***
18	0	1	0	0	0	0	0	0	-1	0	0	0	161.35***	47.48**
19	0	1	0	0	0	0	0	0	0	-1	0	0	2.75	381.49***
20	0	1	0	0	0	0	0	0	0	0	-1	0	33.27***	2,154.18***
21	0	1	0	0	0	0	0	0	0	0	0	-1	123.40***	6,836.60***
22	0	0	1	-1	0	0	0	0	0	0	0	0	2.56	125.62***
23	0	0	1	0	-1	0	0	0	0	0	0	0	589.04***	39.26***
24	0	0	1	0	0	-1	0	0	0	0	0	0	106.66***	11.31
25	0	0	1	0	0	0	-1	0	0	0	0	0	1.18	5.19
26	0	0	1	0	0	0	0	-1	0	0	0	0	1.01	198.09***

ตาราง 52 (ต่อ)

การเปรียบเทียบ	$d(\theta)1 \times \text{SIZE1}$	$d(\theta)1 \times \text{SIZE2}$	$d(\theta)1 \times \text{SIZE3}$	$d(\theta)1 \times \text{SIZE4}$	$d(\theta)2 \times \text{SIZE1}$	$d(\theta)2 \times \text{SIZE2}$	$d(\theta)2 \times \text{SIZE3}$	$d(\theta)2 \times \text{SIZE4}$	$d(\theta)3 \times \text{SIZE1}$	$d(\theta)3 \times \text{SIZE2}$	$d(\theta)3 \times \text{SIZE3}$	$d(\theta)3 \times \text{SIZE4}$	F	
													P	E_1
27	0	0	1	0	0	0	0	0	-1	0	0	0	573.77***	39.49***
28	0	0	1	0	0	0	0	0	0	-1	0	0	166.62***	358.17***
29	0	0	1	0	0	0	0	0	0	0	-1	0	30.06***	2,098.26***
30	0	0	1	0	0	0	0	0	0	0	0	-1	.02	6736.69***
31	0	0	0	1	-1	0	0	0	0	0	0	0	669.21***	305.33***
32	0	0	0	1	0	-1	0	0	0	0	0	0	142.24***	212.31***
33	0	0	0	1	0	0	-1	0	0	0	0	0	7.22	79.74***
34	0	0	0	1	0	0	0	-1	0	0	0	0	.36	8.22
35	0	0	0	1	0	0	0	0	-1	0	0	0	652.93***	24.25***
36	0	0	0	1	0	0	0	0	0	-1	0	0	210.46***	59.55***
37	0	0	0	1	0	0	0	0	0	0	-1	0	50.16***	1,197.06***
38	0	0	0	1	0	0	0	0	0	0	0	-1	3.03	5,022.43***
39	0	0	0	0	1	-1	0	0	0	0	0	0	194.39***	8.43
40	0	0	0	0	1	0	-1	0	0	0	0	0	537.45***	73.00***
41	0	0	0	0	1	0	0	-1	0	0	0	0	638.71***	413.72***
42	0	0	0	0	1	0	0	0	-1	0	0	0	.10	157.49***
43	0	0	0	0	1	0	0	0	0	-1	0	0	129.09***	634.59***
44	0	0	0	0	1	0	0	0	0	0	-1	0	352.95***	2,711.53***
45	0	0	0	0	1	0	0	0	0	0	0	-1	582.14***	7,804.48***
46	0	0	0	0	0	1	-1	0	0	0	0	0	85.39***	31.82***
47	0	0	0	0	0	1	0	-1	0	0	0	0	128.38***	304.05***
48	0	0	0	0	0	1	0	0	-1	0	0	0	185.66***	93.05***
49	0	0	0	0	0	1	0	0	0	-1	0	0	6.66	496.75***
50	0	0	0	0	0	1	0	0	0	0	-1	0	23.47***	2,417.62***
51	0	0	0	0	0	1	0	0	0	0	0	-1	103.74***	7,299.97***
52	0	0	0	0	0	0	1	-1	0	0	0	0	4.37	139.15***
53	0	0	0	0	0	0	1	0	-1	0	0	0	522.87***	16.04
54	0	0	0	0	0	0	1	0	0	-1	0	0	139.74***	277.12***
55	0	0	0	0	0	0	1	0	0	0	-1	0	19.32	1,894.72***
56	0	0	0	0	0	0	1	0	0	0	0	-1	.89	6,367.88***
57	0	0	0	0	0	0	0	1	-1	0	0	0	622.81***	60.69***
58	0	0	0	0	0	0	0	1	0	-1	0	0	193.51***	23.53***

ตาราง 52 (ต่อ)

การเปรียบเทียบ	$d(\theta)1 \times \text{SIZE1}$	$d(\theta)1 \times \text{SIZE2}$	$d(\theta)1 \times \text{SIZE3}$	$d(\theta)1 \times \text{SIZE4}$	$d(\theta)2 \times \text{SIZE1}$	$d(\theta)2 \times \text{SIZE2}$	$d(\theta)2 \times \text{SIZE3}$	$d(\theta)2 \times \text{SIZE4}$	$d(\theta)3 \times \text{SIZE1}$	$d(\theta)3 \times \text{SIZE2}$	$d(\theta)3 \times \text{SIZE3}$	$d(\theta)3 \times \text{SIZE4}$	F	
													P	E_1
59	0	0	0	0	0	0	0	1	0	0	-1	0	42.06***	1,006.93***
60	0	0	0	0	0	0	0	1	0	0	0	-1	1.31	4,624.38***
61	0	0	0	0	0	0	0	0	1	-1	0	0	122.00***	159.81***
62	0	0	0	0	0	0	0	0	1	0	-1	0	341.16***	1,562.05***
63	0	0	0	0	0	0	0	0	1	0	0	-1	566.96***	5,744.64***
64	0	0	0	0	0	0	0	0	0	1	-1	0	55.13***	722.61***
65	0	0	0	0	0	0	0	0	0	1	0	-1	162.96***	3,988.17***
66	0	0	0	0	0	0	0	0	0	0	1	-1	28.52***	1,315.56***
MOLR														
1	1	-1	0	0	0	0	0	0	0	0	0	0	-	1.02
2	1	0	-1	0	0	0	0	0	0	0	0	0	-	.01
3	1	0	0	-1	0	0	0	0	0	0	0	0	-	2.86
4	1	0	0	0	-1	0	0	0	0	0	0	0	-	.02
5	1	0	0	0	0	-1	0	0	0	0	0	0	-	.00
6	1	0	0	0	0	0	-1	0	0	0	0	0	-	3.94
7	1	0	0	0	0	0	0	-1	0	0	0	0	-	13.78
8	1	0	0	0	0	0	0	0	-1	0	0	0	-	8.22
9	1	0	0	0	0	0	0	0	0	-1	0	0	-	31.20***
10	1	0	0	0	0	0	0	0	0	0	-1	0	-	99.54***
11	1	0	0	0	0	0	0	0	0	0	0	-1	-	455.86***
12	0	1	-1	0	0	0	0	0	0	0	0	0	-	1.26
13	0	1	0	-1	0	0	0	0	0	0	0	0	-	7.30
14	0	1	0	0	-1	0	0	0	0	0	0	0	-	1.34
15	0	1	0	0	0	-1	0	0	0	0	0	0	-	1.10
16	0	1	0	0	0	0	-1	0	0	0	0	0	-	8.97
17	0	1	0	0	0	0	0	-1	0	0	0	0	-	22.30
18	0	1	0	0	0	0	0	0	-1	0	0	0	-	15.03
19	0	1	0	0	0	0	0	0	0	-1	0	0	-	43.51***
20	0	1	0	0	0	0	0	0	0	0	-1	0	-	120.73***

ตาราง 52 (ต่อ)

การเปรียบเทียบ	$d(\theta)1 \times \text{SIZE1}$	$d(\theta)1 \times \text{SIZE2}$	$d(\theta)1 \times \text{SIZE3}$	$d(\theta)1 \times \text{SIZE4}$	$d(\theta)2 \times \text{SIZE1}$	$d(\theta)2 \times \text{SIZE2}$	$d(\theta)2 \times \text{SIZE3}$	$d(\theta)2 \times \text{SIZE4}$	$d(\theta)3 \times \text{SIZE1}$	$d(\theta)3 \times \text{SIZE2}$	$d(\theta)3 \times \text{SIZE3}$	$d(\theta)3 \times \text{SIZE4}$	F	
													P	E_1
21	0	1	0	0	0	0	0	0	0	0	0	-1	-	500.03***
22	0	0	1	-1	0	0	0	0	0	0	0	0	-	2.50
23	0	0	1	0	-1	0	0	0	0	0	0	0	-	.00
24	0	0	1	0	0	-1	0	0	0	0	0	0	-	.01
25	0	0	1	0	0	0	-1	0	0	0	0	0	-	3.51
26	0	0	1	0	0	0	0	-1	0	0	0	0	-	12.97
27	0	0	1	0	0	0	0	0	-1	0	0	0	-	7.60
28	0	0	1	0	0	0	0	0	0	-1	0	0	-	29.98***
29	0	0	1	0	0	0	0	0	0	0	-1	0	-	97.36***
30	0	0	1	0	0	0	0	0	0	0	0	-1	-	451.16***
31	0	0	0	1	-1	0	0	0	0	0	0	0	-	2.38
32	0	0	0	1	0	-1	0	0	0	0	0	0	-	2.73
33	0	0	0	1	0	0	-1	0	0	0	0	0	-	.09
34	0	0	0	1	0	0	0	-1	0	0	0	0	-	4.09
35	0	0	0	1	0	0	0	0	-1	0	0	0	-	1.38
36	0	0	0	1	0	0	0	0	0	-1	0	0	-	15.17
37	0	0	0	1	0	0	0	0	0	0	-1	0	-	68.67***
38	0	0	0	1	0	0	0	0	0	0	0	-1	-	386.53***
39	0	0	0	0	1	-1	0	0	0	0	0	0	-	.01
40	0	0	0	0	1	0	-1	0	0	0	0	0	-	3.38
41	0	0	0	0	1	0	0	-1	0	0	0	0	-	12.71
42	0	0	0	0	1	0	0	0	-1	0	0	0	-	7.40
43	0	0	0	0	1	0	0	0	0	-1	0	0	-	29.58***
44	0	0	0	0	1	0	0	0	0	0	-1	0	-	96.63***
45	0	0	0	0	1	0	0	0	0	0	0	-1	-	449.60***
46	0	0	0	0	0	1	-1	0	0	0	0	0	-	3.79
47	0	0	0	0	0	1	0	-1	0	0	0	0	-	13.50
48	0	0	0	0	0	1	0	0	-1	0	0	0	-	8.01

ตาราง 52 (ต่อ)

การเปรียบเทียบ	$d(\theta)1 \times \text{SIZE1}$	$d(\theta)1 \times \text{SIZE2}$	$d(\theta)1 \times \text{SIZE3}$	$d(\theta)1 \times \text{SIZE4}$	$d(\theta)2 \times \text{SIZE1}$	$d(\theta)2 \times \text{SIZE2}$	$d(\theta)2 \times \text{SIZE3}$	$d(\theta)2 \times \text{SIZE4}$	$d(\theta)3 \times \text{SIZE1}$	$d(\theta)3 \times \text{SIZE2}$	$d(\theta)3 \times \text{SIZE3}$	$d(\theta)3 \times \text{SIZE4}$	F	
													P	E_1
49	0	0	0	0	0	1	0	0	0	-1	0	0	-	30.79***
50	0	0	0	0	0	1	0	0	0	0	-1	0	-	98.81***
51	0	0	0	0	0	1	0	0	0	0	0	-1	-	454.29***
52	0	0	0	0	0	0	1	-1	0	0	0	0	-	2.98
53	0	0	0	0	0	0	1	0	-1	0	0	0	-	.78
54	0	0	0	0	0	0	1	0	0	-1	0	0	-	12.97
55	0	0	0	0	0	0	1	0	0	0	-1	0	-	63.88***
56	0	0	0	0	0	0	1	0	0	0	0	-1	-	375.06***
57	0	0	0	0	0	0	0	1	-1	0	0	0	-	.71
58	0	0	0	0	0	0	0	1	0	-1	0	0	-	3.51
59	0	0	0	0	0	0	0	1	0	0	-1	0	-	39.26***
60	0	0	0	0	0	0	0	1	0	0	0	-1	-	311.14***
61	0	0	0	0	0	0	0	0	1	-1	0	0	-	7.40
62	0	0	0	0	0	0	0	0	1	0	-1	0	-	50.56***
63	0	0	0	0	0	0	0	0	1	0	0	-1	-	341.67***
64	0	0	0	0	0	0	0	0	0	1	-1	0	-	19.28
65	0	0	0	0	0	0	0	0	0	1	0	-1	-	248.54***
66	0	0	0	0	0	0	0	0	0	0	1	-1	-	129.36***

***p < .001

ตาราง 53 การเปรียบเทียบค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภท
ที่ 1 ระหว่างวิธีการตรวจสอบสามวิธี ภายใต้เงื่อนไขเดียวกันของปฏิสัมพันธ์ระหว่างปัจจัย
ความแตกต่างของการแจกแจงความสามารถ กับปัจจัยขนาดตัวอย่าง

การเปรียบเทียบ	Poly- SIBTEST	OLR	MOLR	<i>F</i>	
				<i>P</i>	<i>E</i> ₁
<i>d</i> (θ) 1 × SIZE1					
1	1	-1	0	210.00***	-
2	1	0	-1	249.31***	-
3	0	1	-1	1.69	-
<i>d</i> (θ) 1 × SIZE2					
1	1	-1	0	313.85***	101.02***
2	1	0	-1	405.11***	26.09***
3	0	1	-1	5.82	24.43***
<i>d</i> (θ) 1 × SIZE3					
1	1	-1	0	532.81***	146.17***
2	1	0	-1	611.01***	58.71***
3	0	1	-1	2.68	19.61***
<i>d</i> (θ) 1 × SIZE4					
1	1	-1	0	609.19***	509.11***
2	1	0	-1	629.15***	72.37***
3	0	1	-1	.16	197.58***
<i>d</i> (θ) 2 × SIZE1					
1	1	-1	0	168.26***	-
2	1	0	-1	169.77***	-
3	0	1	-1	.00	-
<i>d</i> (θ) 2 × SIZE2					
1	1	-1	0	237.62***	52.68***
2	1	0	-1	273.19***	37.44***
3	0	1	-1	1.24	1.30

ตาราง 53 (ต่อ)

การเปรียบเทียบ	Poly-SIBTEST	OLR	MOLR	F	
				P	E_1
$d(\theta) 2 \times \text{SIZE3}$					
1	1	-1	0	423.16***	215.53***
2	1	0	-1	515.18***	96.99***
3	0	1	-1	4.52	23.35***
$d(\theta) 2 \times \text{SIZE4}$					
1	1	-1	0	397.71***	704.94***
2	1	0	-1	432.98***	135.70***
3	0	1	-1	.75	222.05***
$d(\theta) 3 \times \text{SIZE1}$					
1	1	-1	0	70.41***	5.36
2	1	0	-1	25.40***	31.82***
3	0	1	-1	11.23***	63.30***
$d(\theta) 3 \times \text{SIZE2}$					
1	1	-1	0	110.18***	685.56***
2	1	0	-1	105.46***	68.98***
3	0	1	-1	.05	319.62***
$d(\theta) 3 \times \text{SIZE3}$					
1	1	-1	0	85.97***	3,223.53***
2	1	0	-1	206.95***	269.23***
3	0	1	-1	26.15***	1,629.58***
$d(\theta) 3 \times \text{SIZE4}$					
1	1	-1	0	124.58***	8,657.70***
2	1	0	-1	173.92***	771.82***
3	0	1	-1	4.11	4,259.52***

***p < .001

ภาคผนวก ข
ตัวอย่างผลการวิเคราะห์ข้อมูล

1. ตัวอย่างผลการจำลองข้อมูลโดยใช้โปรแกรม RESGEN

1.1 แฟ้มข้อมูล .out

PC-REGEN V4.0 [DYNAMIC MEMORY ALLOCATION]

ITEM RESPONSE GENERATOR

*** ITEM RESPONSE GENERATOR ***

Manipulated Variables: Form of DIF=Nonuniform, Proportion of DIF
=10%, Ability Distribution=CaseA, Sample Size RG:FG=500:500

\$Comment SIMULATION DATAsample,
Compensatory Multidimensional Generalized Partial Credit
Model(CMGPCM)

\$GLOBAL Nblock=2,Ntest=2,Ngroup=2,Nsample=1,SEED=1000;

GLOBAL PARAMETERS
=====

NUMBER OF SUBTESTS	2
NUMBER OF BLOCK	2
NUMBER OF SUBJECT GROUPS	2
NUMBER OF SAMPLE UNITS	1
SEED FOR UNIFORM RANDOM NUMBER GENERATOR	1000
GENERATING METHOD: PROBABILITY-BASED (DEFAULT)	

\$SAVE dfname='sample.DAT';

FILE ASSIGNMENT AND DISPOSITION
=====

[OUTPUT FILES]

ITEM RESPONSE DATA FILE	sample.DAT
-------------------------	------------

BLOCK INFORMATION
=====

```
$LENGTH Nitm=(10,10),Model=(3,3),
      Ndimension=(2,2),Maxdim=2;
TOTAL NUMBER OF BLOCKS: 2
```

BLOCK	LENGTHS	MODEL	DIMENSION
1	10	POLYTOMOUS(MASTERS)	2
2	10	POLYTOMOUS(MASTERS)	2

```
TOTAL NUMBER OF ITEMS: 20
MAXIMUM LATENT TRAIT DIMENSION: 2
```

```
$block1 Bname=RG, Iname=(i01,i02,i03,i04,i05,i06,i07,i08,i09,i10),
```

```
INTERCEPT=(1.00,0.50,-0.50,-1.00,
              1.00,0.00,-0.50,-1.00,
              1.50,1.00,-1.00,-1.50,
              1.50,0.50,-1.00,-1.50,
              1.00,0.50,-0.50,-1.00,
              1.00,0.00,-0.50,-1.00,
              1.50,0.50,-0.50,-1.50,
              1.50,0.50,-0.50,-1.50,
              1.50,0.50,-0.50,-1.50,
              1.50,0.50,-0.50,-1.50),
```

```
SLOPE=(0.67,0.18,
        0.59,0.34,
        0.41,0.71,
        0.24,0.90,
        0.82,0.22,
        0.50,0.29,
        0.85,0.85,
        0.85,0.85,
        0.70,0.70,
        1.00,1.00),
```

```
LOGISTIC, SCALE=1.7, NPARM=2,
```

```
NCATEGORY=(5(0)10),
```

```
DIMENSION=2;
```

```

$block2 Bname=FG, Iname=(i01,i02,i03,i04,i05,i06,i07,i08,i09,i10),
      INTERCEPT=(1.00,0.50,-0.50,-1.00,
                    1.00,0.00,-0.50,-1.00,
                    1.50,1.00,-1.00,-1.50,
                    1.50,0.50,-1.00,-1.50,
                    1.00,0.50,-0.50,-1.00,
                    1.00,0.00,-0.50,-1.00,
                    1.50,0.50,-0.50,-1.50,
                    1.50,0.50,-0.50,-1.50,
                    1.50,0.50,-0.50,-1.50,
                    1.50,0.50,-0.50,-1.50),
      SLOPE=(0.67,0.18,
             0.59,0.34,
             0.41,0.71,
             0.24,0.90,
             0.82,0.22,
             0.50,0.29,
             0.70,0.70,
             1.00,1.00,
             0.85,0.85,
             0.85,0.85),
      LOGISTIC, SCALE=1.7, NPARAM=2,
      NCATEGORY=(5(0)10),
      DIMENSION=2;

```

```

1 BLOCK RG          DIMENSION: 2  PARAMETER: 2  SCALE CONSTANT: 1.7
(MASTERS)          LOGISTIC POLYTOMOUS ITEM RESPONSE MODEL

```

ITEM	INTERCEPT	SLOPE	THRESHOLD	ASYMPTOTE (CAT)
i01	1.000	0.670		0.000 (2)
	0.500	0.180		
	-0.500			
	-1.000			
i02	1.000	0.590		0.000 (2)
	0.000	0.340		
	-0.500			
	-1.000			

i03	1.500 1.000 -1.000 -1.500	0.410 0.710		0.000 (2)
i04	1.500 0.500 -1.000 -1.500	0.240 0.900		0.000 (2)
i05	1.000 0.500 -0.500 -1.000	0.820 0.220		0.000 (2)
i06	1.000 0.000 -0.500 -1.000	0.500 0.290		0.000 (2)
i07	1.500 0.500 -0.500 -1.500	0.850 0.850		0.000 (2)
i08	1.500 0.500 -0.500 -1.500	0.850 0.850		0.000 (2)
i09	1.500 0.500 -0.500 -1.500	0.700 0.700		0.000 (2)
i10	1.500 0.500 -0.500 -1.500	1.000 1.000		0.000 (2)

SCORING FUNCTION OF EACH DIMENSION

1	1.00	1.00
2	2.00	2.00
3	3.00	3.00
4	4.00	4.00
5	5.00	5.00

2 BLOCK FG DIMENSION: 2 PARAMETER: 2 SCALE CONSTANT: 1.7
 (MASTERS) LOGISTIC POLYTOMOUS ITEM RESPONSE MODEL

ITEM	INTERCEPT	SLOPE	THRESHOLD	ASYMPTOTE (CAT)
i01	1.000 0.500 -0.500 -1.000	0.670 0.180		0.000 (2)
i02	1.000 0.000 -0.500 -1.000	0.590 0.340		0.000 (2)

i03	1.500 1.000 -1.000 -1.500	0.410 0.710	0.000 (2)
i04	1.500 0.500 -1.000 -1.500	0.240 0.900	0.000 (2)
i05	1.000 0.500 -0.500 -1.000	0.820 0.220	0.000 (2)
i06	1.000 0.000 -0.500 -1.000	0.500 0.290	0.000 (2)
i07	1.500 0.500 -0.500 -1.500	0.700 0.700	0.000 (2)
i08	1.500 0.500 -0.500 -1.500	1.000 1.000	0.000 (2)
i09	1.500 0.500 -0.500 -1.500	0.850 0.850	0.000 (2)
i10	1.500 0.500 -0.500 -1.500	0.850 0.850	0.000 (2)

SCORING FUNCTION OF EACH DIMENSION

1	1.00	1.00
2	2.00	2.00
3	3.00	3.00
4	4.00	4.00
5	5.00	5.00

NOTE: TABLE '0' INDICATES THAT THE ITEM IS OBJECTIVELY SCORED.

SUBTEST INFORMATION

=====

\$test1 Tname=T1,SELECT=1,PACK;

\$test2 Tname=T2,SELECT=2,PACK;

LATENT TRAIT DISTRIBUTION

=====

```
$group1 Gname=RG, MEAN=(0.0,0.0), VARIANCE=(1.0,1.0),
        CORR=(1.00,0.50,1.00);
```

```
GROUP #:      1    NAME: RG          ABILITY DIMENSION:  2
```

```
MEAN VECTOR
```

```
  1      0.000
  2      0.000
```

```
COVARIANCE MATRIX
```

```
  1      1.000
  2      0.500      1.000
```

```
CORRELATION MATRIX
```

```
  1      1.000
  2      0.500      1.000
```

```
CHOLESKY FACTOR OF COVARIANCE MATRIX
```

```
  1      1.000
  2      0.500      0.866
```

```
$group2 Gname=FG, MEAN=(0.0,0.0), VARIANCE=(1.0,1.0),
        CORR=(1.00,0.50,1.00);
```

```
GROUP #:      2    NAME: FG          ABILITY DIMENSION:  2
```

```
MEAN VECTOR
```

```
  1      0.000
  2      0.000
```

```
COVARIANCE MATRIX
```

```
  1      1.000
  2      0.500      1.000
```

```
CORRELATION MATRIX
```

```
  1      1.000
  2      0.500      1.000
```

```
CHOLESKY FACTOR OF COVARIANCE MATRIX
```

```
  1      1.000
  2      0.500      0.866
```

```
SAMPLING INFORMATION
```

```
=====
```

```
$sample NSUBJECT=(500,500), TEST=(1,2), ALT;
```

GENERATING LATENT TRAIT DISTRIBUTION =====

SIMULATED ABILITY VALUES ARE CREATED AND STORED IN FILE SF.DAT

[THE FORMAT CARD]

PSU NAME	A8,2X
GROUP NAME	A8,2X
GROUP NUMBER	I5,2X
STIMULEE ID.	5A1,2X
WEIGHT	G12.6,2X
ABILITY	< >F10.5

SUBJECT:	1			
PSU 0001	RG	1	00001	1.00000
0.97726	-0.46867			

SUBJECT:	2			
PSU 0001	RG	1	00002	1.00000
1.70513	0.32029			

SUBJECT:	3			
PSU 0001	RG	1	00003	1.00000
0.36071	-0.37910			

PSU: PSU 0001	
GROUP: 1	
NAME: RG	
CASE: 500	
WGHT: 500.00	

MEAN VECTOR

1	-0.030
2	-0.026

COVARIANCE MATRIX

1	1.039	
2	0.493	0.936

CORRELATION MATRIX

1	1.000	
2	0.500	1.000

SUBJECT:	1			
PSU 0001	FG	2	00001	1.00000
-0.15381	0.54052			

SUBJECT:	2			
PSU 0001	FG	2	00002	1.00000
-0.31058	0.96761			

SUBJECT:	3			
PSU 0001	FG	2	00003	1.00000
-0.82951	0.95999			

PSU: PSU 0001
 GROUP: 2
 NAME: FG
 CASE: 500
 WGHT: 500.00

MEAN VECTOR

1 0.051
 2 0.027

COVARIANCE MATRIX

1 1.030
 2 0.491 0.932

CORRELATION MATRIX

1 1.000
 2 0.501 1.000

GENERATING ITEM RESPONSES =====

PSU: PSU 0001
 NUMBER OF VECTORS: 1000

--- ALTERNATIVE SUBTESTS ARE USED ---

1	GROUP: RG	CASE: 500		
1	TEST: T1	NUMBER: 250	PACKED FORM	
2	TEST: T2	NUMBER: 250	PACKED FORM	
2	GROUP: FG	CASE: 500		
1	TEST: T1	NUMBER: 250	PACKED FORM	
2	TEST: T2	NUMBER: 250	PACKED FORM	

THE NUMBER OF ALTERNATIVES, ANSWER KEY AND CATEGORIES

BLOCK:	1	RG												
ITEM	NAME	ALT	KEY	CAT	DIM	RESPONSE	FREQUENCY							
						NOT-P	OMIT	1	2	3	4	5		
1	i01	5	1	5	2									
			1	T1		0	0	63	109	151	115	62		
			2	T2		500	0	0	0	0	0	0		
2	i02	5	1	5	2									
			1	T1		0	0	93	137	101	84	85		
			2	T2		500	0	0	0	0	0	0		
3	i03	5	1	5	2									
			1	T1		0	0	50	71	243	80	56		
			2	T2		500	0	0	0	0	0	0		
4	i04	5	1	5	2									
			1	T1		0	0	58	120	193	78	51		
			2	T2		500	0	0	0	0	0	0		

BLOCK:		2		FG								
ITEM	NAME	ALT	KEY	CAT	DIM	RESPONSE		FREQUENCY				
						NOT-P	OMIT	1	2	3	4	5
1	i01	5	1	5	2							
			1	T1		500	0	0	0	0	0	0
			2	T2		0	0	70	102	160	107	61
2	i02	5	1	5	2							
			1	T1		500	0	0	0	0	0	0
			2	T2		0	0	80	154	113	82	71
3	i03	5	1	5	2							
			1	T1		500	0	0	0	0	0	0
			2	T2		0	0	49	80	239	94	38
4	i04	5	1	5	2							
			1	T1		500	0	0	0	0	0	0
			2	T2		0	0	62	129	189	70	50
5	i05	5	1	5	2							
			1	T1		500	0	0	0	0	0	0
			2	T2		0	0	83	85	150	110	72
6	i06	5	1	5	2							
			1	T1		500	0	0	0	0	0	0
			2	T2		0	0	74	161	121	70	74
7	i07	5	1	5	2							
			1	T1		500	0	0	0	0	0	0
			2	T2		0	0	76	127	130	97	70

8	i08	5	1	5	2	500	0	0	0	0	0	0
			1	T1		0	0	105	112	93	88	102
			2	T2								
9	i09	5	1	5	2	500	0	0	0	0	0	0
			1	T1		0	0	87	111	115	93	94
			2	T2								
10	i10	5	1	5	2	500	0	0	0	0	0	0
			1	T1		0	0	96	105	108	109	82
			2	T2								

Note: A correct item response of the dichotomous model
is specified by KEY.

GENERATED ITEM RESPONSES ARE STORED IN FILE: sample.DAT
[THE FORMAT CARD]

```

SIMULEE ID.      5A1,2X
PSU NAME         A8,2X
GROUP NAME       A8,2X
GROUP NUMBER     I5,2X
TEST NUMBER      I5,2X
CASE WEIGHT      G12.6,2X
ABILITY          < >F10.5
RESPONSES        < >(/,100A1)

```

```

13488 BYTES OF NUMERICAL WORKSPACE USED OF 409600 AVAILABLE
IN PHASE-1
5356 BYTES OF CHARACTER WORKSPACE USED OF 81920 AVAILABLE
IN PHASE-1

```

[RESGEN]

CURRENT DATE: 2-13-2006
CURRENT TIME: 14:02:38
NORMAL END

1.2 แฟ้มข้อมูล .dat

```

00001 PSU 0001 RG      1      1      1.00000      0.97726 -0.46867
2412324232
00002 PSU 0001 RG      1      2      1.00000      1.70513  0.32029
3241433343
00003 PSU 0001 RG      1      1      1.00000      0.36071 -0.37910
2123313222
00004 PSU 0001 RG      1      2      1.00000      2.72614  1.38766
5555525554
00005 PSU 0001 RG      1      1      1.00000      1.62758  0.06061
3233234332
00006 PSU 0001 RG      1      2      1.00000     -0.75884 -0.58834
2231222321
00007 PSU 0001 RG      1      1      1.00000      0.67292  0.79540
5343444445
00008 PSU 0001 RG      1      2      1.00000     -0.01528 -0.74875
3232221222
00009 PSU 0001 RG      1      1      1.00000      2.37671  1.71397
4555545555

```


2. ตัวอย่างผลการตรวจสอบ DIF ด้วยวิธี Poly-SIBTEST โดยใช้โปรแกรม DIFPACK

```

name of input parameter file = SIB.IN
number of items on test = 40
name of file for Ref. grp. scores = C:\Poly-SIBTEST\Rsampl.DAT
name of file for Focal grp. scores = C:\Poly-SIBTEST\Fsampl.DAT
minimum no. of examinees per matching score cell = 2
number of runs for this data set = 40
number of examinees in Reference Group = 500
number of examinees in Focal group = 500

```

Examinee Test Score Summary Statistics

```

Reference Group:           Mean = *****
                        Standard deviation = 47.41
Focal Group:              Mean = *****
                        Standard deviation = 46.99

```

Standardized Score Difference = 0.01

Item Statistics

= item number

m = mean score on item

r = point biserial (item score-test score correlation)

```

# :    1      2      3      4      5      6      7      8      9     10
m : 3.056 2.858 3.073 2.961 3.057 2.923 3.063 2.981 3.026 2.860
r : 0.733 0.798 0.784 0.803 0.817 0.737 0.877 0.863 0.747 0.745

# :   11     12     13     14     15     16     17     18     19     20
m : 3.078 3.003 3.039 2.968 3.109 2.903 3.065 3.003 3.121 2.982
r : 0.822 0.922 0.898 0.913 0.931 0.816 0.727 0.904 0.826 0.898

# :   21     22     23     24     25     26     27     28     29     30
m : 3.087 2.954 3.107 2.958 3.110 3.050 3.064 2.916 3.072 3.004
r : 0.801 0.889 0.925 0.835 0.913 0.916 0.890 0.853 0.881 0.919

# :   31     32     33     34     35     36     37     38     39     40
m : 3.089 3.040 3.078 2.903 3.106 3.064 3.118 3.105 3.078 3.111
r : 0.807 0.887 0.909 0.826 0.902 0.916 0.865 0.899 0.884 0.895

```

p-value notation:

R denotes p-value for test of DIF/DBF against Ref. group

F denotes p-value for test of DIF/DBF against Foc. group

E denotes p-value for test of DIF/DBF against either the
Ref. or Foc. group.

NOTES:

MS/SSD = Matching Subtest Standardized Score Difference.
Standardized difference in mean observed scores
between Reference group and Focal group on the
matching subtest.

p-elim = proportion of Reference (R) and Focal (F) groups
eliminated (not used) in SIBTEST calculations.

Positive Beta estimate indicates DIF/DBF favoring Ref. grp.
Negative Beta estimate indicates DIF/DBF favoring Foc. grp.

FLAG = error flag indicator. FLAG=0 indicates a normal successful completion of a SIBTEST run. All other values of FLAG come with short error messages.

SIBTEST-pooled weighting									F L A G
Run no.	Suspect Item	Subtest Numbers	Beta estimate	standard error	p-value	p-elim R F		MS SSD	
1	1		-0.036	0.062	0.564 E	.19	.25	0.01	0
2	2		0.056	0.063	0.371 E	.24	.23	0.01	0
3	3		0.025	0.048	0.606 E	.17	.17	0.01	0
4	4		0.005	0.049	0.911 E	.24	.21	0.01	0
5	5		-0.099	0.058	0.089 E	.23	.24	0.01	0
6	6		0.037	0.065	0.568 E	.26	.21	0.01	0
7	7		0.090	0.046	0.048 E	.22	.20	0.01	0
8	8		0.035	0.050	0.481 E	.21	.22	0.01	0
9	9		0.023	0.059	0.699 E	.22	.23	0.01	0
10	10		0.020	0.059	0.742 E	.16	.22	0.01	0
11	11		-0.006	0.048	0.909 E	.21	.18	0.01	0
12	12		0.024	0.043	0.579 E	.21	.20	0.01	0
13	13		-0.048	0.051	0.347 E	.19	.18	0.01	0
14	14		0.098	0.047	0.037 E	.24	.25	0.01	0
15	15		-0.041	0.040	0.298 E	.17	.20	0.01	0
16	16		0.001	0.050	0.985 E	.18	.17	0.01	0
17	17		0.039	0.061	0.518 E	.17	.22	0.01	0
18	18		0.024	0.050	0.627 E	.21	.23	0.01	0
19	19		0.024	0.053	0.653 E	.19	.18	0.01	0
20	20		-0.061	0.047	0.190 E	.19	.17	0.01	0
21	21		-0.086	0.059	0.144 E	.21	.21	0.01	0
22	22		-0.025	0.054	0.650 E	.17	.18	0.01	0
23	23		0.078	0.041	0.055 E	.24	.20	0.01	0
24	24		0.000	0.052	0.998 E	.19	.20	0.01	0
25	25		-0.052	0.046	0.258 E	.19	.20	0.01	0
26	26		0.078	0.044	0.077 E	.18	.22	0.01	0
27	27		-0.071	0.050	0.154 E	.18	.19	0.01	0
28	28		0.014	0.050	0.777 E	.23	.21	0.01	0
29	29		0.035	0.053	0.506 E	.24	.21	0.01	0
30	30		-0.005	0.045	0.911 E	.20	.20	0.01	0
31	31		-0.040	0.049	0.408 E	.16	.18	0.01	0
32	32		-0.046	0.048	0.337 E	.23	.17	0.01	0
33	33		-0.052	0.049	0.287 E	.21	.18	0.01	0
34	34		0.026	0.054	0.638 E	.21	.21	0.01	0
35	35		-0.013	0.047	0.790 E	.17	.22	0.01	0
36	36		0.015	0.039	0.706 E	.19	.22	0.01	0
37	37		-0.030	0.051	0.560 E	.23	.21	0.01	0
38	38		0.037	0.046	0.430 E	.21	.23	0.01	0
39	39		-0.045	0.043	0.295 E	.18	.20	0.01	0
40	40		-0.019	0.048	0.698 E	.20	.22	0.01	0

Program execution is completed.

Your output is stored on the file: C:\Poly-SIBTEST\sample

3. ตัวอย่างผลการตรวจสอบ DIF ด้วยวิธี OLR โดยใช้โปรแกรม SPSS

```

-> * SPSS SYNTAX written by: .
-> * Bruno D. Zumbo, PhD .
-> * Professor of Psychology and Mathematics, .
-> * University of Northern British Columbia .
-> * e-mail: zumbob@unbc.ca .

-> include file='ologit2.inc'.
22 set printback off.
23 ***** OLOGIT 2.0 (test version) *****
24
25 * By Prof. Dr. Steffen Kuehnel
26 Institut fuer Politikwissenschaft
27 Justus-Liebig-Universitat Giessen
28 Karl-Glockner-Str 21, Haus E
29 35394 Giessen
30 Germany.
31
32 * contrast subcommand added by
33 John Hendrickx <J.Hendrickx@maw.kun.nl>
34 Department of Sociology
35 University of Nijmegen
36 P.O. Box 9104
37 6500 HE Nijmegen
38 The Netherlands.
39
40 * This macro is available from
41 <http://www.socsci.kun.nl/maw/sociologie/resources/mlogist>.
42
43 * macro subroutine, called by class.

-> execute.

>Error # 105. Command name: EXECUTE
>This command is not valid before a working file has been defined.
>This command not executed.

->
-> GET FILE='sample.sav'.

-> EXECUTE.

-> compute item01= item01.

-> compute total= total.

-> compute g= g.

-> * Regression model with the conditioning variable, total score,
in alone.

-> * Step #1.

-> ologit var = item01 total
-> /output=all.
Run MATRIX procedure:

```

LOGISTIC REGRESSION with an ORDINAL DEPENDENT VARIABLE

(by Steffen M. KUEHNEL)

***** Information Section *****

Dependent variable is:
item01

Marginal distribution of dependent variable

Value	Frequ.	Percent	%>Value
1.00	127.00	12.70	87.30
2.00	185.00	18.50	68.80
3.00	340.00	34.00	34.80
4.00	201.00	20.10	14.70
5.00	147.00	14.70	.00

Effective sample size:
1000

Means and standard deviations of independent variables:

	Mean	Std.Dev.
total	121.1430	47.2216

***** Estimation Section *****

Running Iteration No.:
1

Running Iteration No.:
2

Running Iteration No.:
3

Running Iteration No.:
4

Running Iteration No.:
5

..... Optimal solution found.

***** OUTPUT SECTION *****

LR-test that all predictor weights are zero

-2 Log-Likelihood of Model with Constants only:
3090.757

-2 Log-Likelihood of full Model:
2347.083

LR-statistic		DF	Prob.	%-Reduct
Chisqu.				
743.674		1.000	.000	.241

Estimations, standard errors, and effects

	Coeff.=B	Std.Err.	B/Std.E.	Prob.	exp(B)
total	.042184	.001810	23.312152	.000000	1.043087
Const.1	-2.028908	.182915	-11.092069	.000000	.131479
Const.2	-3.759370	.201866	-18.623097	.000000	.023298
Const.3	-6.238911	.265712	-23.479974	.000000	.001952
Const.4	-7.995336	.306344	-26.099167	.000000	.000337

Results assuming a latent continuous variable

R-Square (%):
54.67

Standardized regression weights of the latent variable:
total .7394

----- END MATRIX -----

-> execute.

-> * Regression model adding uniform DIF to model.

-> * Step #2.

-> ologit var = item01 total g

-> /contrast g=indicator

-> /output=all.

Run MATRIX procedure:

LOGISTIC REGRESSION with an ORDINAL DEPENDENT VARIABLE

(by Steffen M. KUEHNEL)

Parameter coding for g using the indicator contrast

Value	Freq	g.1
1	500	1
2	500	0

***** Information Section *****

Dependent variable is:

item01

Marginal distribution of dependent variable

Value	Frequ.	Percent	%>Value
1.00	127.00	12.70	87.30
2.00	185.00	18.50	68.80
3.00	340.00	34.00	34.80
4.00	201.00	20.10	14.70
5.00	147.00	14.70	.00

Effective sample size:
1000

Means and standard deviations of independent variables:

	Mean	Std.Dev.
total	121.1430	47.2216
g.1	.5000	.5003

***** Estimation Section *****

Running Iteration No.:
1

Running Iteration No.:
2

Running Iteration No.:
3

Running Iteration No.:
4

Running Iteration No.:
5

..... Optimal solution found.

***** OUTPUT SECTION *****

LR-test that all predictor weights are zero

-2 Log-Likelihood of Model with Constants only:
3090.757

-2 Log-Likelihood of full Model:
2346.290

LR-statistic			
Chisqu.	DF	Prob.	%-Reduct
744.467	2.000	.000	.241

Estimations, standard errors, and effects

	Coeff.=B	Std.Err.	B/Std.E.	Prob.	exp(B)
total	.042209	.001810	23.316893	.000000	1.043112
g.1	-.105941	.118987	-.890354	.373276	.899478
Const.1	-1.976702	.191870	-10.302322	.000000	.138525
Const.2	-3.708700	.209467	-17.705401	.000000	.024509
Const.3	-6.189955	.271028	-22.838765	.000000	.002050
Const.4	-7.946381	.310914	-25.558137	.000000	.000354

Results assuming a latent continuous variable

R-Square (%):
54.71

Standardized regression weights of the latent variable:
total .7395
g.1 -.0197

----- END MATRIX -----

-> execute.

-> * Regression model adding non-uniform DIF to the model.

-> * Step #3.

```
-> ologit var = item01 total g total*g
-> /contrast g=indicator
-> /output=all.
Run MATRIX procedure:
```

LOGISTIC REGRESSION with an ORDINAL DEPENDENT VARIABLE

(by Steffen M. KUEHNEL)

Parameter coding for g using the indicator contrast

Value	Freq	g.1
1	500	1
2	500	0

Interaction term total*g

int1.1	total	g.1
--------	-------	-----

***** Information Section *****

Dependent variable is:
item01

Marginal distribution of dependent variable

Value	Frequ.	Percent	%>Value
1.00	127.00	12.70	87.30
2.00	185.00	18.50	68.80
3.00	340.00	34.00	34.80
4.00	201.00	20.10	14.70
5.00	147.00	14.70	.00

Effective sample size:
1000

Means and standard deviations of independent variables:

	Mean	Std.Dev.
total	121.1430	47.2216
g.1	.5000	.5003
int1.1	60.7010	69.3764

***** Estimation Section *****

Running Iteration No.:
1

Running Iteration No.:
2

Running Iteration No.:
3

Running Iteration No.:
4

Running Iteration No.:
5

..... Optimal solution found.

***** OUTPUT SECTION *****

LR-test that all predictor weights are zero

-2 Log-Likelihood of Model with Constants only:
3090.757

-2 Log-Likelihood of full Model:
2345.942

LR-statistic			
Chisqu.	DF	Prob.	%-Reduct
744.815	3.000	.000	.241

Estimations, standard errors, and effects

	Coeff.=B	Std.Err.	B/Std.E.	Prob.	exp(B)
total	.041463	.002204	18.813343	.000000	1.042335
g.1	-.286980	.329227	-.871679	.383383	.750527
int1.1	.001490	.002527	.589807	.555320	1.001491
Const.1	-1.885446	.246219	-7.657590	.000000	.151761
Const.2	-3.618215	.259129	-13.962977	.000000	.026831
Const.3	-6.099708	.310727	-19.630437	.000000	.002244
Const.4	-7.856439	.345785	-22.720582	.000000	.000387

Results assuming a latent continuous variable

R-Square (%):
54.73

Standardized regression weights of the latent variable:

total	.7263
g.1	-.0533
int1.1	.0384

----- END MATRIX -----

-> execute.

4. ตัวอย่างผลการตรวจสอบ DIF ด้วยวิธี MOLR โดยใช้โปรแกรม SPSS

```

-> * SPSS SYNTAX written by:.
-> * Bruno D. Zumbo, PhD.
-> * Professor of Psychology and Mathematics,.
-> * University of Northern British Columbia.
-> * SPSS SYNTAX modified by:.
-> * Arin Naumtanom.
-> * Student at Srinakharinwirote University.
-> include file='ologit2.inc'.
  24 set printback off.
  25 ***** OLOGIT 2.0 (test version)
*****
      **
  26
  27 * By Prof. Dr. Steffen Kuehnel
  28   Institut fuer Politikwissenschaft
  29   Justus-Liebig-Universitat Giessen
  30   Karl-Glockner-Str 21, Haus E
  31   35394 Giessen
  32   Germany.
  33
  34 * contrast subcommand added by
  35   John Hendrickx <J.Hendrickx@maw.kun.nl>
  36   Department of Sociology
  37   University of Nijmegen
  38   P.O. Box 9104
  39   6500 HE Nijmegen
  40   The Netherlands.
  41
  42 * This macro is available from
  43   <http://www.socsci.kun.nl/maw/sociologie/resources/mlogist>.
  44
  45 * macro subroutine, called by class.

-> execute.

>Error # 105.  Command name: EXECUTE
>This command is not valid before a working file has been defined.
>This command not executed.

-> GET FILE='sample.sav'.

-> EXECUTE.

-> compute item01= item01.

-> compute t1= t1.

-> compute t2= t2.

-> compute g= g.

-> * Multidimensional Polytomous Logistic Regression Model.
-> * Step #1.

```

```
-> ologit var = item01 t1 t2
-> /output=all.
Run MATRIX procedure:
```

LOGISTIC REGRESSION with an ORDINAL DEPENDENT VARIABLE
(by Steffen M. KUEHNEL)

***** Information Section *****

Dependent variable is:
item01

Marginal distribution of dependent variable

Value	Frequ.	Percent	%>Value
1.00	127.00	12.70	87.30
2.00	185.00	18.50	68.80
3.00	340.00	34.00	34.80
4.00	201.00	20.10	14.70
5.00	147.00	14.70	.00

Effective sample size:
1000

Means and standard deviations of independent variables:

	Mean	Std.Dev.
t1	.0129	1.0366
t2	.0703	1.0131

***** Estimation Section *****

Running Iteration No.:
1

Running Iteration No.:
2

Running Iteration No.:
3

Running Iteration No.:
4

Running Iteration No.:
5

..... Optimal solution found.

***** OUTPUT SECTION *****

LR-test that all predictor weights are zero

-2 Log-Likelihood of Model with Constants only:
3090.757

-2 Log-Likelihood of full Model:
2355.501
LR-statistic

Chisqu.	DF	Prob.	%-Reduct
735.255	2.000	.000	.238

Estimations, standard errors, and effects

	Coeff.=B	Std.Err.	B/Std.E.	Prob.	exp(B)
t1	-.062050	.069198	-.896693	.369883	.939836
t2	2.009425	.094346	21.298573	.000000	7.459025
Const.1	2.871541	.125954	22.798330	.000000	17.664219
Const.2	1.127525	.087145	12.938431	.000000	3.088005
Const.3	-1.190066	.088062	-13.514020	.000000	.304201
Const.4	-2.951950	.126132	-23.403564	.000000	.052238

Results assuming a latent continuous variable

R-Square (%):
54.93

Standardized regression weights of the latent variable:

t1 -.0238
t2 .7535

----- END MATRIX -----

-> execute.

-> * Regression model adding uniform DIF to model.

-> * Step #2.

-> ologit var = item01 t1 t2 g

-> /contrast g=indicator

-> /output=all.

Run MATRIX procedure:

LOGISTIC REGRESSION with an ORDINAL DEPENDENT VARIABLE

(by Steffen M. KUEHNEL)

Parameter coding for g using the indicator contrast

Value	Freq	g.1
1	500	1
2	500	0

***** Information Section *****

Dependent variable is:
item01

Marginal distribution of dependent variable

Value	Frequ.	Percent	%>Value
1.00	127.00	12.70	87.30
2.00	185.00	18.50	68.80
3.00	340.00	34.00	34.80
4.00	201.00	20.10	14.70
5.00	147.00	14.70	.00

Effective sample size:
1000

Means and standard deviations of independent variables:

	Mean	Std.Dev.
t1	.0129	1.0366
t2	.0703	1.0131
g.1	.5000	.5003

***** Estimation Section *****

Running Iteration No.:

1

Running Iteration No.:

2

Running Iteration No.:

3

Running Iteration No.:

4

Running Iteration No.:

5

..... Optimal solution found.

***** OUTPUT SECTION *****

LR-test that all predictor weights are zero

-2 Log-Likelihood of Model with Constants only:
3090.757

-2 Log-Likelihood of full Model:
2354.189

LR-statistic	Chisqu.	DF	Prob.	%-Reduct
	736.568	3.000	.000	.238

Estimations, standard errors, and effects

	Coeff.=B	Std.Err.	B/Std.E.	Prob.	exp(B)
t1	-.060558	.069251	-.874470	.381862	.941239
t2	2.011299	.094363	21.314519	.000000	7.473018
g.1	-.137649	.120174	-1.145419	.252036	.871404
Const.1	2.943973	.141417	20.817600	.000000	18.991144
Const.2	1.197292	.106648	11.226605	.000000	3.311138
Const.3	-1.123109	.105471	-10.648507	.000000	.325267
Const.4	-2.885650	.138407	-20.849083	.000000	.055819

Results assuming a latent continuous variable

R-Square (%):
55.00

Standardized regression weights of the latent variable:

t1 -.0232
t2 .7536
g.1 -.0255

----- END MATRIX -----

-> execute.

-> * Regression model adding non-uniform DIF to the model.

-> * Step #3.

-> ologit var = item01 t1 t2 g t1*g t2*g t1*t2*g

-> /contrast g=indicator

-> /output=all.

Run MATRIX procedure:

LOGISTIC REGRESSION with an ORDINAL DEPENDENT VARIABLE

(by Steffen M. KUEHNEL)

Parameter coding for g using the indicator contrast

Value Freq g.1

1	500	1
2	500	0

Interaction term t1*g
int1.1 t1 g.1

Interaction term t2*g
int2.1 t2 g.1

Interaction term t1*t2*g
int3.1 t1 t2 g.1

***** Information Section *****

Dependent variable is:
item01

Marginal distribution of dependent variable

Value	Frequ.	Percent	%>Value
1.00	127.00	12.70	87.30
2.00	185.00	18.50	68.80
3.00	340.00	34.00	34.80
4.00	201.00	20.10	14.70
5.00	147.00	14.70	.00

Effective sample size:
1000

Means and standard deviations of independent variables:

	Mean	Std.Dev.
t1	.0129	1.0366
t2	.0703	1.0131
g.1	.5000	.5003
int1.1	.0197	.7076
int2.1	.0422	.7220
int3.1	.2698	.8503

***** Estimation Section *****

Running Iteration No.:

1

Running Iteration No.:

2

Running Iteration No.:

3

Running Iteration No.:

4

Running Iteration No.:

5

..... Optimal solution found.

***** OUTPUT SECTION *****

LR-test that all predictor weights are zero

-2 Log-Likelihood of Model with Constants only:
3090.757

-2 Log-Likelihood of full Model:
2351.014

LR-statistic				
Chisqu.	DF	Prob.	%-Reduct	
739.743	6.000	.000	.239	

Estimations, standard errors, and effects

	Coeff.=B	Std.Err.	B/Std.E.	Prob.	exp(B)
t1	-.043490	.095154	-.457045	.647639	.957442
t2	1.977161	.120317	16.432870	.000000	7.222207
g.1	-.210713	.127049	-1.658521	.097212	.810006
int1.1	-.044751	.138760	-.322510	.747067	.956235
int2.1	.076859	.149981	.512458	.608331	1.079890
int3.1	.149056	.086836	1.716525	.086066	1.160738
Const.1	2.939523	.141126	20.829086	.000000	18.906835
Const.2	1.206724	.106834	11.295269	.000000	3.342516
Const.3	-1.115572	.105603	-10.563848	.000000	.327728
Const.4	-2.897527	.139409	-20.784286	.000000	.055159

Results assuming a latent continuous variable

R-Square (%):
55.28

Standardized regression weights of the latent variable:

t1	-.0166
t2	.7385

g.1	-.0389
int1.1	-.0117
int2.1	.0205
int3.1	.0467

----- END MATRIX -----

-> execute.

ประวัติย่อผู้วิจัย

ประวัติย่อผู้วิจัย

ชื่อ-ชื่อสกุล	นายอรินทร์ น่วมถนนอม
สถานที่เกิด	จังหวัดอ่างทอง
สถานที่อยู่	4/1330 ถ.นวมินทร์ แขวงคลองกุ่ม เขตบึงกุ่ม กรุงเทพฯ 10230
ตำแหน่งหน้าที่การงาน	คศ. 3
สถานที่ทำงาน	โรงเรียนบางกะปิ สังกัดสำนักงานเขตพื้นที่การศึกษากรุงเทพมหานครเขต 2
ประวัติการศึกษา	
พ.ศ. 2529	ค.บ. (คณิตศาสตร์) วิทยาลัยครูพระนครศรีอยุธยา
พ.ศ. 2537	กศ.ม. (การวัดผลการศึกษา) มหาวิทยาลัยศรีนครินทรวิโรฒ
พ.ศ. 2549	กศ.ด. (การทดสอบและวัดผลการศึกษา) มหาวิทยาลัยศรีนครินทรวิโรฒ