การเปรียบเทียบวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่า

ปริญญานิพนธ์

ของ

สุชาติ สิริมีนนันท์

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาการศึกษาดุษฎีบัณฑิต สาขาวิชาการทดสอบและวัดผลการศึกษา

พฤษภาคม 2554

การเปรียบเทียบวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่า

ปริญญานิพนธ์

ของ

สุชาติ สิริมีนนันท์

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาการศึกษาดุษฎีบัณฑิต สาขาวิชาการทดสอบและวัดผลการศึกษา

พฤษภาคม 2554

ลิขสิทธิ์เป็นของมหาวิทยาลัยศรีนครินทรวิโรฒ

การเปรียบเทียบวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่า

บทคัดย่อ

ของ

สุชาติ สิริมีนนันท์

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาการศึกษาดุษฎีบัณฑิต สาขาวิชาการทดสอบและวัดผลการศึกษา

พฤษภาคม 2554

สุชาติ สิริมีนนันท์. (2554). *การเปรียบเทียบวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่า.* ปริญญานิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. คณะกรรมการควบคุม: ผู้ช่วยศาสตราจารย์ ดร.สุรชัย มีชาญ, รองศาสตราจารย์ ดร.บุญเชิด ภิญโญอนันตพงษ์, อาจารย์ ดร.เสกสรรค์ ทองคำบรรจง.

การศึกษาครั้งนี้มีจุดมุ่งหมายเพื่อศึกษา และเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่1 จากการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่าด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ในการศึกษาครั้งนี้ใช้ข้อมูลจำลองโดยจำลองภายใต้โมเดลพาร์เชียลเครดิตทั่วไปแบบมิติเดียวของมูราคิ ซึ่งแต่ละข้อจะมีรายการตอบ 5 รายการ ให้คะแนนเป็น 1,2,3,4 หรือ 5 คะแนน ข้อมูลดังกล่าวจำลองภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน 2 รูปแบบ ความยาวของแบบทดสอบ 3 ขนาด สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 3 ขนาด และขนาดตัวอย่าง 3 ขนาด รวมข้อมูลที่ใช้ศึกษาทั้งหมด 54 เงื่อนไข ในแต่ละเงื่อนไขจำลองซ้ำ 50 ครั้ง

ผลการวิจัยสรุปได้ดังนี้

1. อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่าของวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก สูงกว่าวิธีโพลีโตมัสซิบเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับในทุกเงื่อนไข

2. อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก ต่ำกว่าวิธีโพลีโตมัสซิบเทสท์และวิธีการถดถอยโลจิสติกแบบจัดอันดับในทุกเงื่อนไข

3. โดยภาพรวมวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีประสิทธิภาพในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ สูงกว่าวิธีโพลีโตมัสซิบเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับ

A COMPARISON OF THE POLYTOMOUS SIBTEST, LOGISTIC DISCRIMINANT FUNCTION ANALYSIS AND ORDINAL LOGISTIC REGRESSION PROCEDURES ON DETECTING DIFFERENTIAL ITEM FUNCTIONING IN POLYTOMOUSLY SCORED ITEMS

AN ABSTRACT

BY

SUCHAT SIRIMINNUNTE

Presented in Partial Fullfillment of the Requirements for the

Doctor of Education Degree in Testing and Measurement

at Srinakharinwirot University

May 2011

Suchat Siriminnunte.(2011). *A Comparison of the Polytomous SIBTEST, Logistic Discriminant Function Analysis and Ordinal Logistic Regression Procedures on Detecting Differential Item Functioning in Polytomously Scored Items, Ed.D. (Testing and Measurement).* Bangkok: Graduate School, Srinakharinwirot University. Advisor Committee: Asst. Prof. Dr. Surachai Meechan, Assoc. Prof. Dr. Boonchird Pinyoanuntapong, Dr.Sakesan Thongkhambanchong.

The purpose of this research is to study and compare the power and type I error rates from the detection of differential item functioning (DIF) with Polytomous SIBTEST, Logistic discriminant function analysis and Ordinal logistic regression procedures. In this study, I used simulated data under the unidimensional generalized partial credit model according to Muraki. Type of all items were five response categories scoring as 1,2,3,4 or 5. This data was simulated under a variety of factors: 2 differing levels form of DIF, 3 differing levels of test length, 3 differing levels of proportion of DIF items, and 3 differing levels of sample sizes. A total of 54 conditions was studied. The data were replicated 50 times for each condition.

The major findings are as follow:

1. The Power rate of Logistic discriminant function analysis on procedure detecting differential item functioning in polytomously scored items is higher than Polytomous SIBTEST and Ordinal logistic regression procedures in every condition.

2. The Type I error rate of Logistic discriminant function analysis procedure on detecting differential item functioning in polytomously scored items is lower than Polytomous SIBTEST and Ordinal logistic regression procedure in every condition.

3. The Logistic discriminant function analysis is more efficient than Polytomous SIBTEST and Ordinal logistic regression procedures in detecting differential item functioning.

ปริญญานิพนธ์
เรื่อง

การเปรียบเทียบวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและ
วิธีการถดถอยโลจิสติกแบบจัดอันดับ ในการตรวจสอบการทำหน้าที่เบี่ยงเบน
ของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่า

ของ
สุชาติ สิริมีนนันท์

ได้รับอนุมัติจากบัณฑิตวิทยาลัยให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาการศึกษาดุษฎีบัณฑิต สาขาวิชาการทดสอบและวัดผลการศึกษา
ของมหาวิทยาลัยศรีนครินทรวิโรฒ

..... คณบดีบัณฑิตวิทยาลัย
(รองศาสตราจารย์ ดร.สมชาย สันติวัฒนกุล)
วันที่ ..... เดือน พฤษภาคม พ.ศ. 2554

คณะกรรมการควบคุมปริญญานิพนธ์

.....ประธาน
(ผู้ช่วยศาสตราจารย์ ดร.สุรชัย มีชาญ)

.....กรรมการ
(รองศาสตราจารย์ ดร.บุญเชิด ภิญโญอนันตพงษ์)

.....กรรมการ
(อาจารย์ ดร.เสกสรรค์ ทองคำบรรจง)

คณะกรรมการสอบปากเปล่า

.....ประธาน
(อาจารย์ ดร.สุวพร เซ็มเฮง)

.....กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.สุรชัย มีชาญ)

.....กรรมการ
(รองศาสตราจารย์ ดร.บุญเชิด ภิญโญอนันตพงษ์)

.....กรรมการ
(อาจารย์ ดร.เสกสรรค์ ทองคำบรรจง)

.....กรรมการ
(อาจารย์ ดร.อรินทร์ น่วมถนอม)

ประกาศคุณูปการ

ปริญญานิพนธ์ฉบับนี้สำเร็จสมบูรณ์ได้ด้วยความกรุณาช่วยเหลือ และเอาใจใส่เป็นอย่างดียิ่งจากผู้ช่วยศาสตราจารย์ ดร.สุรชัย มีชาญ ประธานกรรมการควบคุมปริญญานิพนธ์ รองศาสตราจารย์ ดร.บุญเชิด ภิญโญอนันตพงษ์ และอาจารย์ ดร.เสกสรรค์ ทองคำบรรจง กรรมการควบคุมปริญญานิพนธ์ ท่านทั้งสามได้เสียสละเวลาอันมีค่าในการให้คำปรึกษา แนะนำแก้ไขข้อบกพร่องต่าง ๆ พร้อมทั้งให้กำลังใจแก่ผู้วิจัยตลอดมา ผู้วิจัยรู้สึกซาบซึ้งและขอกราบขอบพระคุณเป็นอย่างสูงมา ณ โอกาสนี้

ขอกราบขอบพระคุณอาจารย์ ดร.สุวพร เซ็มเฮง และอาจารย์ ดร.อรินทร์ น่วมถนอม กรรมการที่แต่งตั้งเพิ่มเติมที่กรุณาให้คำแนะนำและข้อเสนอแนะที่เป็นประโยชน์ มีคุณค่าทำให้ปริญญานิพนธ์ฉบับนี้มีความสมบูรณ์ และมีคุณค่าสูงยิ่งขึ้น

ขอกราบขอบพระคุณคณะกรรมการสอบเค้าโครงปริญญานิพนธ์ทุกท่านที่กรุณาให้คำแนะนำและข้อเสนอแนะที่เป็นประโยชน์ทำให้ปริญญานิพนธ์ฉบับนี้สำเร็จลุล่วงมีความสมบูรณ์ยิ่งขึ้น

ขอกราบขอบพระคุณรองศาสตราจารย์ ดร.องอาจ นัยพัฒน์ อาจารย์ที่ปรึกษานิสิต ซึ่งให้การดูแลผู้วิจัยด้วยความห่วงใย และให้คำปรึกษาแนะนำอย่างดียิ่งตลอดระยะเวลาที่ศึกษาในมหาวิทยาลัย

ขอขอบพระคุณศาสตราจารย์ Bruno D. Zumbo แห่งมหาวิทยาลัยบริติช โคลัมเบีย ประเทศแคนาดาที่กรุณาให้ดาวน์โหลด SYNTAX ของโปรแกรม SPSS สำหรับใช้วิเคราะห์วิธีการถดถอยโลจิสติก และ Kyung T. Han และ Ronald K.Hambleton แห่งมหาวิทยาลัยแมสซาชูเซท สหรัฐอเมริกาที่กรุณาให้ดาวน์โหลดโปรแกรม WinGen2 สำหรับใช้ในการจำลองข้อมูล

ขอขอบพระคุณอาจารย์มิ่ง เทพครเมือง ที่กรุณาให้ความอนุเคราะห์โปรแกรมคอมพิวเตอร์ และช่วยเหลือในการวิเคราะห์ข้อมูลด้วยระบบคอมพิวเตอร์ ตลอดจนให้คำแนะนำต่าง ๆในการปรับปรุงและแก้ไขปริญญานิพนธ์เป็นอย่างดียิ่ง

ขอกราบขอบพระคุณ ดร.สุวัฒน์ วิวัฒนานนท์ อดีตผู้อำนวยการโรงเรียนบางกะปิ ขอกราบขอบพระคุณ ดร.ธงชาติ วงษ์สวรรค์ ผู้อำนวยการโรงเรียนบางกะปิ ที่สนับสนุนให้ผู้วิจัยมีโอกาสศึกษาต่ออย่างเต็มที่

สุดท้ายขอกราบขอบพระคุณบิดา มารดา และบุคคลในครอบครัวที่คอยห่วงใย ให้กำลังใจ และปรารถนาจะเห็นความสำเร็จของผู้วิจัยเป็นอย่างยิ่ง และขอกราบขอบพระคุณครูอาจารย์ทุกท่านทั้งในอดีตและปัจจุบันที่ได้อบรมสั่งสอนและประสิทธิ์ประสาทวิชาความรู้ จนผู้วิจัยประสบความสำเร็จในการศึกษาและหน้าที่การงานด้วยดี

สุชาติ สิริมีนนันท์

สารบัญ

บทที่	หน้า
1 บทนำ	1
ภูมิหลัง	1
ความมุ่งหมายของการวิจัย	9
ความสำคัญของการวิจัย	9
ขอบเขตของการวิจัย	10
นิยามศัพท์เฉพาะ	11
กรอบแนวคิดในการวิจัย	13
สมมติฐานในการวิจัย	21
2 เอกสารและงานวิจัยที่เกี่ยวข้อง	22
แนวคิดเกี่ยวกับการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ	22
โมเดลทางสถิติการทำหน้าที่เบี่ยงเบนของข้อสอบ	33
วิธีตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ	36
การวิเคราะห์อัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1	64
งานวิจัยที่เกี่ยวข้อง	66
3 วิธีดำเนินการวิจัย	79
การจัดกระทำตัวแปร	79
การจำลองข้อมูล	81
การวิเคราะห์ข้อมูล	84
4 ผลการวิเคราะห์ข้อมูล	89
สัญลักษณ์และความหมายที่ใช้ในการวิเคราะห์ข้อมูล	89
ตอนที่ 1 ค่าสถิติพื้นฐานของอัตราความถูกต้อง และอัตราความคลาดเคลื่อน ประเภทที่ 1 ของการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ	90

สารบัญ (ต่อ)

บทที่ **หน้า**

4 (ต่อ)

ตอนที่ 2 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่า..... 95

5 สรุป อภิปรายผล และข้อเสนอแนะ..... 137

สรุปผลการวิจัย..... 140

อภิปรายผลการวิจัย..... 153

ข้อเสนอแนะ..... 157

บรรณานุกรม..... 158

ภาคผนวก..... 167

ประวัติย่อผู้วิจัย..... 204

บัญชีตาราง

ตาราง **หน้า**

1 ค่าสถิติพื้นฐานของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบ DIF 3 วิธี ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัยหลัก..... 91

2 เปรียบเทียบอัตราความถูกต้อง และ อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบ DIF ของปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีโพลีโตมัสซิบเทสท์..... 95

3 เปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบ DIF ของปัจจัยความยาวของแบบทดสอบด้วยวิธีโพลีโตมัสซิบเทสท์..... 96

4 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ในการตรวจสอบ DIF ของปัจจัยความยาวของแบบทดสอบด้วยวิธีโพลีโตมัสซิบเทสท์..... 97

5 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1ในการตรวจสอบ DIF ของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีโพลีโตมัสซิบเทสท์..... 98

6 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ในการตรวจสอบ DIF ของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีโพลีโตมัสซิบเทสท์..... 99

7 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบ DIF ของปัจจัยขนาดตัวอย่างด้วยวิธีโพลีโตมัสซิบเทสท์..... 100

8 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ในการตรวจสอบ DIF ของปัจจัยขนาดตัวอย่างด้วยวิธีโพลีโตมัสซิบเทสท์..... 101

9 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบ DIF ของปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก..... 101

10 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบ DIF ของปัจจัยความยาวของแบบทดสอบด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก..... 102

11 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ในการตรวจสอบ DIF ของปัจจัยความยาวของแบบทดสอบด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก..... 103

12 เปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบ DIF ของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก..... 104

บัญชีตาราง (ต่อ)

ตาราง **หน้า**

13 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ในการตรวจสอบ DIF ของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก..... 105
14 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบ DIF ของปัจจัยขนาดตัวอย่างด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก 106
15 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ในการตรวจสอบ DIF ของปัจจัยขนาดตัวอย่างด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก..... 107
16 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบ DIF ของปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับ..... 108
17 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบ DIF ของปัจจัยความยาวของแบบทดสอบด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับ..... 109
18 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ในการตรวจสอบ DIF ของปัจจัยความยาวของแบบทดสอบด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับ..... 110
19 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบ DIF ของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับ..... 111
20 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ในการตรวจสอบ DIF ของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับ..... 112
21 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบ DIF ของปัจจัยขนาดตัวอย่างด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับ..... 113
22 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ในการตรวจสอบ DIF ของปัจจัยขนาดตัวอย่างด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับ..... 114
23 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน..... 115

บัญชีตาราง (ต่อ)

ตาราง **หน้า**

24 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่ป็นรูปแบบเดียวกัน 116

25 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน .. 117

26 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน.. 118

27 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขความยาวของแบบทดสอบ 20 ข้อ 119

28 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขความยาวของแบบทดสอบ 20 ข้อ..... 120

29 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขความยาวของแบบทดสอบ 40 ข้อ..... 121

30 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ ระหว่างวิธีตรวจสอบ DIF 3 วิธีภายใต้เงื่อนไขความยาวของแบบทดสอบ 40 ข้อ.... 122

31 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขความยาวของแบบทดสอบ 60 ข้อ..... 123

32 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขความยาวของแบบทดสอบ 60 ข้อ 124

33 เปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10%..... 125

34 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10%..... 126

35 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% 127

36 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20%..... 128

บัญชีตาราง (ต่อ)

ตาราง **หน้า**

37 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีตรวจสอบ DIF 3 วิธีภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30%..... 129

38 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30%..... 130

39 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขขนาดตัวอย่าง 250 คน 131

40 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขขนาดตัวอย่าง 250 คน..... 132

41 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขขนาดตัวอย่าง 500 คน 133

42 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขขนาดตัวอย่าง 500 คน..... 134

43 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขขนาดตัวอย่าง 1000 คน 135

44 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขขนาดตัวอย่าง 1000 คน. 136

บัญชีภาพประกอบ

ภาพประกอบ **หน้า**

1 กรอบแนวคิดในการวิจัย..... 20
2 ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF) 26
3 ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) 26
4 ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF)
ภายใต้โมเดลพาร์เชียลเครดิตทั่วไป (GPCM) 28
5 ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF)
ภายใต้โมเดลพาร์เชียลเครดิตทั่วไป (GPCM) 29
6 แผนผังของการจำลองข้อมูล..... 83

บทที่ 1
บทนำ

ภูมิหลัง

ในการวัดผลถ้าจะวัดได้ถูกต้องตรงกับสิ่งที่ต้องการวัด จะต้องมีเครื่องมือวัดที่ดีมีคุณภาพ ดัชนีบ่งชี้คุณภาพของเครื่องมือวัดที่สำคัญแบ่งเป็นสองกลุ่มคือ ดัชนีบ่งชี้คุณภาพรายข้อกับดัชนีบ่งชี้คุณภาพทั้งฉบับ ดัชนีบ่งชี้คุณภาพรายข้อประกอบด้วยความยากง่าย (Difficulty) และอำนาจจำแนก (Discrimination) ส่วนดัชนีบ่งชี้คุณภาพทั้งฉบับประกอบด้วยหลักฐานแสดงความเที่ยงตรง (Validity evidence) และหลักฐานแสดงความเชื่อมั่น (Reliability evidence) ความเที่ยงตรงถือเป็นลักษณะสำคัญที่สุดของเครื่องมือวัดที่มีคุณภาพ เครื่องมือวัดที่มีความเที่ยงตรงนั้นสามารถวัดในสิ่งที่ผู้วัดต้องการวัดได้อย่างถูกต้องแม่นยำ ไม่ว่าจะเป็นแบบทดสอบผลสัมฤทธิ์ แบบทดสอบความถนัด มาตรวัดเจตคติ จริยธรรม หรือบุคลิกภาพ นักวิจัยนิยมตรวจสอบความเที่ยงตรงของเครื่องมือวัดตามพยานหลักฐานซึ่งได้แก่ หลักฐานแสดงความเที่ยงตรงตามเนื้อหา (Content validity evidence) หลักฐานแสดงความเที่ยงตรงตามเกณฑ์ (Criterion validity evidence) และหลักฐานแสดงความเที่ยงตรงตามโครงสร้าง (Construct validity evidence) (บุญเชิด ภิญโญอนันตพงษ์. 2547: 171-173) นอกจากดัชนีบ่งชี้คุณภาพของเครื่องมือวัดดังกล่าวมาแล้ว เครื่องมือวัดที่ดียังประกอบด้วยลักษณะที่สำคัญซึ่งได้แก่ ความยุติธรรม (Fairness) ความเป็นปรนัย(Objectivity) ความมีประสิทธิภาพ (Efficiency) และความสามารถในการนำไปใช้ (Usability)

ความยุติธรรมของแบบทดสอบ (Test fairness) มีความสำคัญทั้งทางศีลธรรมสำหรับผู้สร้าง และผู้ใช้แบบทดสอบ สิ่งนี้เป็นหลักการขั้นพื้นฐาน ดังที่ทราบกันดีว่าการทดสอบอย่างเป็นทางการมีมากว่าสามพันปี ในสมัยราชวงศ์ชาง (1760-1122 ปีก่อนคริสต์ศักราช) และในสมัยราชวงศ์ฮั่น (202 ปีก่อนคริสต์ศักราช ถึงคริสต์ศักราช 220) กระดาษคำตอบของผู้เข้าสอบได้รับการเขียนซ้ำขึ้นมาใหม่โดยนักปราชญ์ราชบัณฑิต การเขียนใหม่นี้เพื่อให้แน่ใจถึงการไม่ปรากฏชื่อของผู้เข้าสอบก่อนที่ผู้สอบจะได้รับคะแนนจากการสอบ เป็นการขจัดความไม่เท่าเทียมกันอย่างหนึ่งในการได้รับคะแนนจากการตรวจให้คะแนน ขณะที่ปัจจุบันการให้คะแนนแบบทดสอบเลือกตอบ (Multiple choice test) ถือว่ามีความเป็นปรนัยในการตรวจ เกิดความยุติธรรมในการให้คะแนน แต่การทดสอบบางครั้ง เราจะพบความไม่ยุติธรรมของการตอบข้อสอบอันเนื่องมาจาก **ความลำเอียงของข้อสอบหรือแบบทดสอบ** (Item bias / Test bias) อันจะทำให้เกิดการได้เปรียบเสียเปรียบในการสอบระหว่างกลุ่มผู้สอบต่างกลุ่มที่มีความสามารถเท่ากันในคุณลักษณะที่ต้องการวัด ทางการศึกษาได้มีการศึกษาความลำเอียงข้อสอบมาอย่างต่อเนื่อง เริ่มต้นจากในปี ค.ศ. 1910 โดยอัลเฟรด บิเนต์ (Camilli; & Shepard. 1994: 5; citing Binet; & Simon. 1916/1973) ใช้แบบทดสอบทดสอบเด็กที่มีฐานะทางเศรษฐกิจและสังคมที่ต่างกัน เขาพบว่าข้อสอบในแบบทดสอบบางข้อวัดถึงผลต่าง ๆ ของผู้ที่ได้รับการฝึกฝนทางวัฒนธรรมที่บ้านหรือโรงเรียนมากกว่าความสามารถทางสมอง หลังจากศึกษา

โดยละเอียด บิเนต์ก็ได้ขจัดรายการของคะแนนนั้นๆ ออกไป เช่น ในเด็กอายุ 7 ขวบ ข้อสอบบางข้อที่เกี่ยวข้องกับนิ้วมือ การคัดลอกประโยคที่เขียนไว้ และการเรียกชื่อเหรียญ(เงิน) ก็ถูกตัดออกไป เพราะโดยปกติแล้วข้อสอบลักษณะนี้เกิดจากการฝึกฝนที่บ้าน (Home training) ความตั้งใจ (Attention) ภาษา (Language) นิสัยในการดูภาพ (Habit of looking at pictures) และแบบฝึกหัดเชิงวิชาการ (Scholastic exercise) ต่อมาในปี 1912 วิลเลียม สเทิร์น (Camilli; & Shepard. 1994: 5; citing William Stern. 1914) ได้ศึกษาความแตกต่างทางชนชั้นในประเทศเยอรมนีเช่นเดียวกับบิเนต์ โดยพยายามที่จะทำความเข้าใจเกี่ยวกับแหล่งที่ทำให้เกิดความแตกต่าง ซึ่งพบว่าความแตกต่างที่มีนัยสำคัญนั้นเป็นผลมาจากแบบทดสอบต่าง ๆ (Tests) เข้าข้างชนชั้นหนึ่งเหนือชนชั้นอื่นๆ นับแต่บิเนต์และสเทิร์นเป็นต้นมา นักวิจัยหลายคนได้แสดงให้เห็นถึงความสัมพันธ์ระหว่าง IQ กับชนชั้นทางสังคม พร้อมทั้งพยายามศึกษาว่าพันธุกรรมหรือสิ่งแวดล้อมเป็นสาเหตุของความแตกต่างที่สังเกตได้ในความสามารถหรือไม่ จนกระทั่งปี 1951 เอลส์และคณะ แสดงให้เห็นว่าความแตกต่างที่วัดได้ใน IQ ของนักเรียนขึ้นอยู่กับเนื้อหาเฉพาะของข้อสอบในแบบทดสอบซึ่งไม่ได้สะท้อนผลอย่างถูกต้องแม่นยำบนความสามารถพื้นฐานของนักเรียน ในทศวรรษ 1960 คณะสิทธิพลเมืองสหรัฐอเมริกาได้หันมาให้ความสนใจต่อบทบาทของแบบทดสอบอย่างจริงจังโดยศึกษาผลการสอบที่ปฏิเสธการเข้าสู่การศึกษาที่เท่าเทียมกันและโอกาสของการจ้างงาน ทั้งนี้เพราะสังคมอเมริกันประกอบด้วยประชากรกลุ่มย่อยๆ หลายเชื้อชาติ เช่น อเมริกันผิวขาว อเมริกันผิวดำ ชนพื้นเมือง และผิวสีอื่นๆ ที่มีภูมิหลังด้านวัฒนธรรมแตกต่างกัน ทำให้ผลการตอบข้อสอบจากแบบทดสอบที่วัดความสามารถด้านสติปัญญา (Cognitive) ที่มีการบรรจุข้อสอบที่อยู่นอกวัฒนธรรมของชนกลุ่มน้อย กลุ่มต่าง ๆ ที่มีความแตกต่างแตกต่างกัน ทั้ง ๆ ที่แต่ละกลุ่มมีความสามารถเท่าเทียมกัน ทำให้เกิดการได้เปรียบเสียเปรียบของกลุ่มต่างๆ ที่มีภูมิหลังด้านวัฒนธรรมต่างกัน ซึ่งชี้ให้เห็นว่าการนำแบบทดสอบที่มีความลำเอียงทางการวัดมาใช้จึงเป็นการไม่ยุติธรรมแก่ผู้สอบ และผลที่ได้จากการวัดก็ไม่สามารถสะท้อนถึงความสามารถที่แท้จริงของผู้สอบได้ ที่กล่าวมาจะพบว่ามีการศึกษาความลำเอียงของข้อสอบกันมาอย่างยาวนาน ระยะแรกมีการใช้ความหมายของความลำเอียงของข้อสอบปะปนกันระหว่างความหมายในทางสังคมและความหมายในทางสถิติ เพื่อให้เกิดความชัดเจนในเรื่องดังกล่าวจึงมีผู้เสนอให้ใช้คำว่า “**การทำหน้าที่เบี่ยงเบนของข้อสอบ**” หรือ “**การทำหน้าที่ต่างกันของข้อสอบ**” (Differential item functioning) เรียกสั้นๆว่า “**DIF**” (Holland; & Wainer. 1993: xiv; Angoff. 1993: 4) ซึ่งเน้นหลักฐานเชิงประจักษ์ที่ได้จากสารสนเทศทางสถิติโดยแยกออกจากการประเมินที่มีความหมายทางสังคมและมีความหมายไปในทางลบของคำว่า “ความลำเอียงของข้อสอบ” (Item bias) (Angoff. 1993: 4-5; Hambleton; Swaminathan; & Rogers.1991: 109)ดังนั้นคำว่า “การทำหน้าที่เบี่ยงเบนของข้อสอบ” (Differential item functioning: DIF) จึงเหมาะสมในเชิงวิชาการมากกว่าคำว่า “ความลำเอียงของข้อสอบ” (Item bias) คำๆ นี้จึงเป็นที่นิยมและใช้กันอย่างแพร่หลาย ซึ่งปัจจุบันนักวัดผลการศึกษาและนักวิจัยอื่นๆ ได้ให้ความสำคัญกับการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบหรือแบบทดสอบ ทั้งนี้เพราะภาค

ส่วนต่างๆ อาทิภาคการศึกษา ธุรกิจ อุตสาหกรรม การทหาร หรืออื่นๆ ได้ใช้การสอบด้วยแบบทดสอบในการตรวจสอบความรู้ความสามารถของบุคคล เพื่อเป็นพื้นฐานสำหรับการตัดสินใจคัดเลือกบุคคลเข้าศึกษาต่อ บรรจุงาน สนับสนุนการเลื่อนตำแหน่ง ออกใบอนุญาตหรือใบรับรอง การตัดสินใจเหล่านี้มีความสำคัญต่อบุคคล สังคม และนโยบายของรัฐอย่างเกี่ยวข้องกัน ดังนั้น แบบทดสอบที่นำมาใช้ควรเป็นแบบทดสอบที่เชื่อถือได้ วัดได้ตามคุณลักษณะที่ต้องการวัด เกิดความคลาดเคลื่อนในการวัดน้อยที่สุด มีความยากที่เหมาะสม รวมทั้งสามารถจำแนกความสามารถของผู้สอบได้ นอกจากนี้ควรเป็นแบบทดสอบที่มีความยุติธรรมระหว่างกลุ่มผู้สอบกลุ่มต่างๆที่มีลักษณะบางอย่างต่างกัน เช่น เพศ เชื้อชาติ เผ่าพันธุ์ ศาสนา วัฒนธรรม เศรษฐกิจ สังคม ภาษา ภูมิลำเนา อายุ ประสบการณ์ หรือภูมิหลังอื่นๆ ของผู้สอบ เพื่อไม่ให้เกิดการบิดเบือนในผลการสอบของผู้สอบกลุ่มเฉพาะเหล่านี้ ซึ่งจะช่วยให้ผู้มีหน้าที่เกี่ยวข้องกับการคัดเลือกได้นำไปใช้เป็นข้อมูลในการพิจารณาตัดสินผลการคัดเลือกให้มีความถูกต้องและเป็นธรรม ในทางตรงกันข้ามถ้าแบบทดสอบที่นำมาใช้ในการคัดเลือกขาดความยุติธรรม มีความเบี่ยงเบนเกิดขึ้น โดยข้อสอบหรือแบบทดสอบเข้าข้างผู้สอบกลุ่มหนึ่งมากกว่าอีกกลุ่มหนึ่ง จะส่งผลให้คะแนนสอบแตกต่างกันระหว่างกลุ่ม ทั้งที่มีความสามารถที่ต้องการวัดเท่ากัน ทำให้เกิดการได้เปรียบเสียเปรียบกัน การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบซึ่งถือเป็นส่วนหนึ่งของการพัฒนาแบบทดสอบจึงมีความจำเป็นสำหรับนักพัฒนาแบบทดสอบ เพื่อใช้เป็นหลักฐานแสดงความยุติธรรมของแบบทดสอบ (AERA/APA/NCME. 1999: 74)

แนวคิดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ จะทำการเปรียบเทียบผลการตอบข้อสอบระหว่างผู้สอบกลุ่มย่อยสองกลุ่มที่มีความสามารถระดับเดียวกัน โดยที่ผู้สอบกลุ่มหนึ่งเรียกว่า “**กลุ่มอ้างอิง**” (Reference group; **R**) ใช้เป็นกลุ่มสำหรับเปรียบเทียบกับอีกกลุ่มหนึ่งเรียกว่า “**กลุ่มสนใจ**” (Focal group; **F**) ถือเป็นกลุ่มเป้าหมายที่ต้องการศึกษา ข้อสอบที่ใช้ในการตรวจสอบเรียกว่า “**ข้อสอบที่ศึกษา**” (Studied item) ถ้าข้อสอบทำหน้าที่เบี่ยงเบนแล้ว ความน่าจะเป็นในการตอบข้อสอบถูกของผู้สอบแต่ละกลุ่มจะไม่เท่ากัน โดยคาดว่าผู้สอบกลุ่มอ้างอิงจะได้เปรียบในการตอบข้อสอบ ส่วนผู้สอบกลุ่มสนใจคาดว่าจะเสียเปรียบในการตอบข้อสอบ ส่วนวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบมีหลายวิธี ซึ่งแต่ละวิธีก็มีข้อดีข้อด้อยแตกต่างกันออกไปตามยุคตามสมัย แนวคิดในระยะแรกมุ่งเน้นการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบสองค่า (Dichotomously scored item) โดยใช้คะแนนสังเกต (Observed score) ภายใต้ทฤษฎีการทดสอบมาตรฐานเดิม (Classical test theory: CTT) เป็นเกณฑ์ในการจับคู่กลุ่มผู้สอบ ซึ่งในทางปฏิบัติโดยทั่วไปจะใช้คะแนนรวมของแบบทดสอบทั้งฉบับเป็นเกณฑ์การจับคู่กลุ่มผู้สอบ วิธีที่สำคัญในกลุ่มนี้ได้แก่ วิธีการวิเคราะห์ความแปรปรวน (Analysis of variance: ANOVA) วิธีการแปลงค่าความยากของข้อสอบ (Transformed item difficulty: TID) หรือวิธีการกำหนดจุดเดลต้า (Delta plot) (Angoff. 1993; citing Angoff. 1972; 5) และวิธีตารางการณ์จร (Contingency table method: CT) ซึ่งวิธีนี้ประกอบด้วยวิธีที่สำคัญ 3 วิธีคือ วิธีไค-สแควร์ (Chi –

square procedure) (Scheuneman. 1979) วิธีล็อก-ลิเนียร์ (Log-linear method) (Mellenbergh. 1982) และวิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel; MH) (Holland; & Thayer. 1988) นอกจากนี้ โดแรนส์และคูลิก (Dorans; & Kulick. 1986) ได้พัฒนาวิธีการทำให้เป็นมาตรฐาน (Standardization: STND) ต่อมามีการนำทฤษฎีการตอบข้อสอบ (Item response theory: IRT) มาประยุกต์ใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ทำให้การใช้เกณฑ์การจับคู่กลุ่มผู้สอบเปลี่ยนไป โดยจะใช้ค่าประมาณระดับความสามารถ (Unobserved score) ของผู้สอบ ซึ่งเป็นตัวแปรคุณลักษณะแฝงเป็นเกณฑ์การจับคู่กลุ่มผู้สอบแทนการใช้คะแนนรวมของแบบทดสอบ วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบแนวทางนี้จำแนกเป็น 2 กลุ่มใหญ่คือ กลุ่มที่ใช้การเปรียบเทียบค่าประมาณพารามิเตอร์ของข้อสอบระหว่างกลุ่มผู้สอบ วิธีที่สำคัญในกลุ่มนี้ได้แก่วิธีเปลี่ยนค่าความยาก (Difficulty shift) วิธีการทดสอบไค-สแควร์ของลอร์ด (Lord's chi-square test)(Lord. 1980) และวิธีการทดสอบอัตราส่วนความน่าจะเป็นในทฤษฎีการตอบข้อสอบ (IRT likelihood ratio test: IRT-LR) (Thissen; Steinberg; & Wainer. 1993) ส่วนอีกกลุ่มหนึ่งใช้การเปรียบเทียบค่าประมาณฟังก์ชันการตอบข้อสอบระหว่างกลุ่มผู้สอบ โดยการวัดพื้นที่ระหว่างฟังก์ชันการตอบข้อสอบ วิธีที่สำคัญในกลุ่มนี้ได้แก่ วิธีการวัดพื้นที่ชนิดไม่คิดเครื่องหมาย (Unsigned area) ในรูปรากของกำลังสองของความแตกต่างเฉลี่ย (Root mean squared difference) ระหว่าง IRFs (Linn; & Hamisch. 1981) วิธีการวัดพื้นที่ชนิดคิดเครื่องหมายและไม่คิดเครื่องหมายในรูปผลรวมของกำลังสอง (Sum of square: SOS) (Shepard; Camilli; & Williams. 1985) วิธีการวัดพื้นที่ชนิดคิดเครื่องหมายในช่วงเปิด (Exact signed area; ESA) และพื้นที่ชนิดไม่คิดเครื่องหมายในช่วงเปิด (Exact unsigned area; EUA) (Raju. 1990) และวิธีการวัดพื้นที่ชนิดคิดเครื่องหมายในช่วงปิด (Closed – interval signed area: CSA) และพื้นที่ชนิดไม่คิดเครื่องหมายในช่วงปิด (Closed – interval unsigned area; CUA) (Kim; & Cohen. 1991) นอกจากนี้ยังมีการปรับขยายวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบสำหรับข้อสอบที่วัดความสามารถหลายมิติเช่น วิธีมัลติซิบ (MULTISIB) (Stout; et al. 1997) และวิธีดีเอฟไอที (DFIT) (Oshima; Raju; & Flowers. 1997)

วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบตามวิธีที่ใช้คุณลักษณะแฝง (Unobserved score) โดยใช้ค่าประมาณระดับความสามารถของผู้สอบตามทฤษฎีการตอบข้อสอบเป็นเกณฑ์การจับคู่กลุ่มผู้สอบซึ่งทางการศึกษาถือว่าเป็นวิธีที่ยอมรับกันโดยทั่วไปว่ามีประสิทธิภาพสูงกว่าวิธีใช้คะแนนสังเกตตามทฤษฎีการทดสอบมาตรฐานเดิมเป็นเกณฑ์การจับคู่กลุ่มผู้สอบ วิธีที่สำคัญ ได้แก่ วิธีการทดสอบไค-สแควร์ของลอร์ด วิธีการทดสอบอัตราส่วนความน่าจะเป็นในทฤษฎีการตอบข้อสอบ วิธีการวัดพื้นที่ของราชู และ วิธีการวัดพื้นที่ของคิมและโคเฮ็น ทั้งนี้เนื่องจากการประมาณค่าพารามิเตอร์ข้อสอบตามทฤษฎีการตอบข้อสอบจะไม่แปรเปลี่ยนตามกลุ่มผู้สอบ ถึงแม้ว่าวิธีทั้งสี่จะมีประสิทธิภาพสูง แต่ก็มีข้อจำกัดตรงที่ข้อมูลจะต้องเป็นไปตามข้อตกลงที่เข้มงวด ต้องใช้กลุ่มตัวอย่างขนาดใหญ่ ใช้กระบวนการตรวจสอบหลายขั้นตอน การวิเคราะห์ข้อมูลค่อนข้างซับซ้อน มีการคำนวณทวนซ้ำหลายรอบ ทำให้เสียเวลาและค่าใช้จ่าย ตลอดจนการแปลผลค่อนข้าง

ยาก จากข้อจำกัดของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบตามวิธีที่ใช้คุณลักษณะแฝงดังกล่าว จึงทำให้นักการศึกษาและนักวิจัยอื่น ๆหันกลับมาให้ความสนใจพร้อมทั้งพัฒนาวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ใช้คะแนนสังเกตเป็นเกณฑ์การจับคู่กลุ่มผู้สอบ แต่ยังคงอาศัยแนวคิดและหลักการตรวจสอบการทำหน้าที่เบี่ยงเบนโดยใช้ค่าประมาณระดับความสามารถเป็นเกณฑ์การจับคู่กลุ่มผู้สอบ เพื่อให้สามารถตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบได้อย่างถูกต้องแม่นยำ ขั้นตอนการตรวจสอบไม่ยุ่งยากซับซ้อน และใช้ได้กับกลุ่มตัวอย่างขนาดเล็ก วิธีที่สำคัญ ได้แก่ วิธีการถดถอยโลจิสติก (Logistic regression method: LR) ซึ่งพัฒนาโดยสวามินาธานและโรเจอร์ส (Swaminathan; & Rogers. 1990) ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบจะใช้โมเดลการถดถอยโลจิสติกทำนายความน่าจะเป็นในการตอบข้อสอบถูก โมเดลดังกล่าวสามารถทดสอบอิทธิพลของปฏิสัมพันธ์ระหว่างระดับความสามารถกับการเป็นสมาชิกของกลุ่มผู้สอบ ดังนั้นจึงสามารถตรวจสอบข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกันและที่ไม่เป็นรูปแบบเดียวกัน จุดเด่นของวิธีนี้ คือใช้คะแนนรวม (คะแนนสังเกต) แทนระดับความสามารถ ทำให้คิดคำนวณง่าย สามารถใช้ได้กับกลุ่มตัวอย่างขนาดเล็ก และที่สำคัญคือสามารถตรวจสอบข้อสอบทำหน้าที่เบี่ยงเบนทั้งที่เป็นรูปแบบเดียวกันและที่ไม่เป็นรูปแบบเดียวกัน ส่วนข้อจำกัดของวิธีนี้ คือ ประสิทธิภาพในการตรวจสอบต่ำเมื่อข้อสอบมีค่าความยากและค่าอำนาจจำแนกสูง นอกจากนี้เชียลีและสเตาท์ (Shealy; & Stout.1993) ได้เสนอวิธีซิบเทสท์ (Simultaneous item bias test: SIBTEST) เพื่อใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบสองค่า หลักการตรวจสอบจะแบ่งแบบทดสอบออกเป็นสองชุดย่อย คือแบบทดสอบชุดย่อยที่สงสัยว่าทำหน้าที่เบี่ยงเบน (Suspect subtest) ประกอบด้วยข้อสอบที่ต้องการศึกษาหนึ่งข้อหรือหลายข้อ อีกชุดหนึ่งเป็นแบบทดสอบชุดย่อยที่มีความเที่ยงตรง (Valid subtest) ซึ่งนำไปใช้ในการจับคู่ความสามารถ ประกอบด้วยข้อสอบที่ไม่สงสัยว่าทำหน้าที่เบี่ยงเบนโดยแยกออกจากข้อสอบที่ได้ศึกษาข้างต้น ในการตรวจสอบจะเปรียบเทียบความน่าจะเป็นในการตอบข้อสอบถูกระหว่างกลุ่มผู้สอบที่มีคะแนนการจับคู่เท่ากันและใช้สถิติทดสอบนัยสำคัญ (Narayanan; & Swaminathan. 1994: 315) จุดเด่นของวิธีซิบเทสท์คือ สามารถคำนวณได้ง่าย ไม่ซับซ้อน เสียค่าใช้จ่ายไม่มาก ไม่จำเป็นต้องใช้กลุ่มตัวอย่างขนาดใหญ่ สามารถตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบได้คราวละหลายข้อพร้อมกัน

การปฏิรูปการศึกษาตามพระราชบัญญัติการศึกษาแห่งชาติ พุทธศักราช 2542 มีผลให้เกิดการเปลี่ยนแปลงทางการวัดผล โดยไม่ยึดอยู่กับการวัดแบบตรวจให้คะแนนแบบสองค่า (Dichotomously scored items) เพียงอย่างเดียว มีการนำวิธีการวัดผลแนวใหม่ที่มีการวัดประเมินผลที่หลากหลายให้คะแนนแบบหลายค่า (Polytomously scored items) เข้ามาใช้ควบคู่กับการวัดที่มีอยู่เดิมหรือใช้แทนการวัดแบบเดิม เพื่อให้การวัดผลการศึกษามีความเที่ยงตรง เหมาะสมและใช้การได้ดียิ่งขึ้น ทั้งนี้เพราะเชื่อว่าการวัดประเมินที่มีการให้คะแนนแบบหลายค่าให้รายละเอียดข้อมูลเกี่ยวกับการเรียนรู้ของนักเรียนมากกว่าการวัดด้วยข้อสอบแบบเลือกตอบ (Chang; Mazzeo;

& Roussos. 1996) นอกจากนี้การวัดประเมินดังกล่าวยังได้ให้วิธีการทดสอบที่มีความเสมอภาคมากกว่าวิธีที่ใช้รูปแบบการวัดแยกองค์ประกอบที่ไม่สอดคล้องกันทางโครงสร้างอีกด้วย(Zwick; Donoghue; & Grima. 1993: 233) ทำให้วิธีการวัดประเมินดังกล่าวได้รับความสนใจ พร้อมทั้งนำมาใช้ทางการศึกษาเพิ่มสูงขึ้น เช่น การตัดสินการเขียนแฟ้มสะสมงาน การปฏิบัติภาระงาน การพิสูจน์ทางคณิตศาสตร์ และการทดลองทางวิทยาศาสตร์ เป็นต้น การพัฒนาวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่า (Polytomous DIF) ซึ่งส่วนใหญ่พัฒนาโดยปรับขยายมาจากวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบสองค่า (Dichotomous DIF) ทั้งในกลุ่มวิธีที่ใช้คะแนนสังเกต (Observed score) ตามทฤษฎีการทดสอบมาตรฐานเดิม และในกลุ่มวิธีที่ใช้ตัวแปรแฝง (Unobserved score) ตามทฤษฎีการตอบข้อสอบ เช่น วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก (Logistic discriminant function analysis: LDFA) (Miller, & Spray. 1993) วิธีแมนเทล (Mantel) วิธีแมนเทล-แฮนส์เซลทั่วไป (Generalized Mantel-Haenszel: GMH) (Zwick; Donoghue; & Grima. 1993) วิธีการวัดพื้นที่ (Cohen; Kim; & Baker. 1993) วิธี HW1 และ HW3 (Welch; & Hoover. 1993) วิธีการทำให้เป็นมาตรฐานที่ให้คะแนนแบบหลายค่า (Polytomous STND) (Potenza; & Dorans. 1995) วิธีโพลีโตมัสซิบเทสท์ (Polytomuos SIBTEST: Poly-SIB)(Chang; Mazzeo; & Roussos. 1996) วิธีดีเอฟไอที (Differential functioning of items and test: DFIT) (Flower; Oshima; & Raju. 1999) วิธีการถดถอยโลจิสติกแบบจัดอันดับ (Ordinal logistic regression: OLR) (Zumbo. 1999) เป็นต้น

จากการศึกษางานวิจัยที่ผ่านมา (Swaminathan;& Rogers. 1990; Miller; & Spray. 1993; Spray; & Miller. 1994; Narayanan; & Sawaminathan. 1996; Roussos; Stout. 1996; Chang; Mazzeo; & Roussos. 1996; Oshima; Raju; & Flowers. 1997; Tian.1999; Ankenmann; Witt; & Dunbar. 1999; Kristjansson. 2001; Bolt. 2002; Gierl; Gotzmann; & Boughton. 2004; Su;& Wang. 2005; สิริรัตน์ วิภาสศิลป์ 2545; อุทัยวรรณ สายพัฒนะ. 2547; อรินทร์ น่วมถนอม. 2549) พบว่าวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบได้มีการคิดค้นและพัฒนามาอย่างต่อเนื่องเป็นลำดับ จากวิธีหนึ่งไปสู่อีกวิธีหนึ่ง จากวิธีที่ใช้คะแนนสังเกตตามทฤษฎีการทดสอบมาตรฐานเดิมไปสู่วิธีที่ใช้คะแนนความสามารถแฝงตามทฤษฎีการตอบข้อสอบ จากการตรวจในข้อสอบที่ให้คะแนนแบบสองค่าไปสู่ข้อสอบที่ให้คะแนนแบบหลายค่า จากการตรวจสอบในข้อสอบที่มุ่งวัดความสามารถเพียงมิติเดียวไปสู่ข้อสอบที่วัดความสามารถหลายมิติ ซึ่งมีประเด็นที่น่าสนใจ 3 ประเด็น ดังนี้

ประเด็นที่หนึ่ง การจัดการศึกษาในปัจจุบันต้องพัฒนาคนให้มีความรู้ ความคิด ทักษะ ความสามารถ คุณธรรม ความรับผิดชอบ และคุณลักษณะที่สามารถต่อสู้ แข่งขันกันเพื่อความสำเร็จ โดยมีการพัฒนาเพื่อให้เกิดคุณสมบัติต่างๆ อาทิ ความสามารถในการสื่อสารทั้งในการพูด การเขียน และการใช้เทคโนโลยีต่างๆ ความสามารถในความร่วมมือประสานประโยชน์และเจรจาแลกเปลี่ยน ความสามารถในการคิดอย่างมีวิจารณญาณและแก้ปัญหาที่ซับซ้อนอย่างมีประสิทธิภาพ ตลอดจน

ความสามารถในการเสริมสร้างประสบการณ์ของมนุษย์ในโลกกว้างที่ไร้พรมแดน เพื่อให้สามารถปรับตัวให้ทันกับโลกที่เปลี่ยนแปลงตลอดเวลา ทำให้การจัดการเรียนการสอน การวัดและประเมินผลจะต้องมีการเปลี่ยนแปลงอย่างมาก เพื่อนำไปสู่การพัฒนาผู้เรียนให้เกิดความสามารถดังกล่าว ในด้านการวัดและประเมินผลครูจะต้องปรับเปลี่ยนวิธีการวัดและประเมินผลจากแนวทางเดิมไปสู่วิธีการวัดและประเมินผลที่สามารถประเมินความก้าวหน้าในผลสัมฤทธิ์ทางการเรียนของนักเรียนที่มาจากการจัดกิจกรรมการเรียนการสอนที่เน้นนักเรียนเป็นศูนย์กลางและปฏิบัติจริง โดยจะใช้การวัดและประเมินผลแนวใหม่ เช่น วิธีการวัดประเมินการปฏิบัติ ที่ใช้แบบทดสอบที่มีการให้คะแนนแบบหลายค่าแทนการวัดแบบดั้งเดิมที่สอบวัดด้วยแบบทดสอบมาตรฐาน หรือแบบทดสอบที่ครูสร้างขึ้น ซึ่งมักเป็นแบบทดสอบที่ให้คะแนนแบบสองค่าที่ประกอบด้วยข้อสอบแบบเลือกตอบถูกผิด หรือเติมคำตอบสั้นๆในการตอบคำถาม ด้วยเหตุดังกล่าวทำให้การวัดผลปัจจุบันเริ่มใช้แบบทดสอบที่ประกอบด้วยข้อสอบที่มีการให้คะแนนแบบหลายค่าเพิ่มมากขึ้นเพราะเชื่อว่าการให้คะแนนข้อสอบแบบหลายค่าได้ให้สารสนเทศเกี่ยวกับแบบแผนการตอบและความเชื่อมั่นจากการตอบที่สูงกว่าการให้คะแนนข้อสอบแบบสองค่า (ศิริชัย กาญจนวาสี. 2545: 75) ขณะนี้การทดสอบทางการศึกษาและจิตวิทยาได้ให้ความสำคัญในประเด็นนี้สูงขึ้น ข้อสอบที่ให้คะแนนแบบหลายค่าจึงเป็นประเด็นหนึ่งที่นำมาศึกษาวิจัยในครั้งนี้

ประเด็นที่สอง ความเที่ยงตรงของการประเมินการทำหน้าที่เบี่ยงเบนของข้อสอบขึ้นอยู่กับคุณภาพของตัวแปรการจับคู่ นั่นคือวิธีต่างๆ ที่ใช้คุณลักษณะแฝงเป็นเกณฑ์การจับคู่เป็นสิ่งที่น่าสนใจ น่าใช้ในทางทฤษฎี เพราะเกิดจากความน่าเชื่อถือในคะแนนแบบทดสอบ สำหรับวิธีที่ใช้คะแนนสังเกตหรือคะแนนรวมเป็นเกณฑ์การจับคู่ ซวิกและคณะ (Zwick et al. 1993) ได้แสดงให้เห็นว่ามีคุณค่าทางทฤษฎีเช่นกัน ถ้าข้อมูลเป็นแบบมิติเดียว (Unidimensional) และเป็นไปตามโมเดลพาร์เชียลเครดิตหรือพาร์เชียลเครดิตทั่วไป ซึ่งคะแนนข้อสอบศึกษาถูกรวมไว้ในคะแนนแบบทดสอบรวม วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่าและใช้คะแนนรวมเป็นเกณฑ์การจับคู่กลุ่มผู้สอบที่น่าสนใจได้แก่ **วิธีโพลีโตมัสซิบเทสท์** (Polytomous SIBTEST) พัฒนาโดยชาง มาซซีโอ และรุสโซ (Chang; Mazzeo; & Roussos.1996) วิธีนี้ปรับขยายมาจากวิธีซิบเทสท์ (SIBTEST) (Shealy; & Stout.1993) เพื่อใช้ตรวจสอบการทำหน้าที่เบี่ยงเบน ของข้อสอบที่ให้คะแนนแบบหลายค่า วิธีนี้มีจุดเด่นคือ ใช้เทคนิคการตรวจสอบแบบหลายมิติจึงสามารถตรวจสอบได้ทั้งข้อสอบที่วัดความสามารถเพียงมิติเดียว และข้อสอบที่วัดความสามารถหลายมิติ การตรวจสอบจะแยกข้อสอบที่ใช้เป็นเกณฑ์การจับคู่ออกจากข้อสอบที่ใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนอย่างชัดเจน ทำให้เกณฑ์มีความเที่ยงตรงสูง คะแนนเกณฑ์การจับคู่ค่อนข้างเป็นคุณลักษณะแฝงมากกว่าคะแนนที่ได้จากการสอบ ทำให้มีความถูกต้องและแม่นยำ มีการคำนวณทวนซ้ำหลายรอบเพื่อคัดเลือกข้อสอบที่ทำหน้าที่เบี่ยงเบนออกจากคะแนนการจับคู่ทำให้เกณฑ์มีความบริสุทธิ์ มีการปรับแก้ค่าการถดถอยเพื่อลดความแตกต่างของค่าความสามารถเป้าหมายระหว่างกลุ่มผู้สอบ ทำให้สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ที่สูงเกิน

ปกติได้ **วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก** (Logistic discriminant function analysis) ซึ่งพัฒนาโดยมิลเลอร์และสเปรย์ (Miller; & Spray. 1993) ปรับขยายมาจากวิธีการถดถอยโลจิสติก (Logistic regression) (Swaminathan; & Rogers. 1990) เพื่อใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่า วิธีดังกล่าวอยู่บนโมเดลพื้นฐานที่ใช้สำหรับคัดแยกคนออกเป็นกลุ่มๆ ที่มีส่วนเหมือนกันบนพื้นฐานของการตอบข้อสอบและคะแนนแบบทดสอบรวม โมเดลดังกล่าวสามารถทดสอบอิทธิพลของคะแนนแบบทดสอบรวมและคะแนนข้อสอบทำให้มีประสิทธิภาพในการตรวจสอบข้อสอบทั้งข้อสอบที่ทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกันและข้อสอบที่ทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน และ**วิธีการถดถอยโลจิสติกแบบจัดอันดับ** (Oridinal logistic regression) พัฒนาโดยซัมโบ (Zumbo. 1999) วิธีนี้ปรับขยายมาจากวิธีการถดถอยโลจิสติก (Logistic regression) (Swaminathan; & Rogers. 1990) เช่นเดียวกัน เพื่อใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถเพียงมิติเดียวและให้คะแนนแบบหลายค่า ข้อดีของวิธีนี้คือใช้โมเดลการถดถอยโลจิสติกทำนายความน่าจะเป็นในการตอบข้อสอบถูกภายใต้คะแนนความสามารถแบบต่อเนื่องและการเป็นสมาชิกของกลุ่มผู้สอบ โมเดลดังกล่าวสามารถทดสอบอิทธิพลของปฏิสัมพันธ์ระหว่างระดับความสามารถกับกลุ่มผู้สอบ ทำให้มีประสิทธิภาพในการตรวจสอบข้อสอบทั้งข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน

ประเด็นที่สาม ปัจจุบันวิธีการวัดประเมินแนวใหม่โดยใช้แบบทดสอบที่ให้คะแนนแบบหลายค่ามีความสำคัญทางการศึกษาและจิตวิทยา เพื่อให้เกิดความยุติธรรมกับผู้สอบทุกคนในแต่ละกลุ่มจึงจำเป็นต้องมีการตรวจสอบคุณภาพแบบทดสอบด้านความเที่ยงตรงและคุณสมบัติของข้อสอบซึ่งรวมทั้งการทำหน้าที่เบี่ยงเบนของข้อสอบด้วย ในการเลือกใช้วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่าเพื่อให้ผลการตรวจสอบมีประสิทธิภาพและประสิทธิผล ควรคำนึงถึงองค์ประกอบหรือปัจจัยร่วมที่นำมาใช้ในการตรวจสอบ เช่น รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน (Form of DIF) ความยาวของแบบทดสอบ (Test length) สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน (Proportion of DIF) และขนาดตัวอย่าง (Sample size) เป็นต้น เพื่อเป็นสารสนเทศส่วนหนึ่งในการตัดสินใจเลือกใช้วิธีการตรวจสอบวิธีใดวิธีหนึ่ง ดังนั้นปัจจัยที่แปรเปลี่ยน 4 ปัจจัยดังกล่าวเป็นประเด็นที่สงสัยว่ามีผลต่อประสิทธิภาพในการตรวจสอบของวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับหรือไม่ ทั้งสามวิธีสามารถตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบได้ดีที่สุดในเงื่อนไขใด

จากประเด็นดังกล่าวผู้วิจัยมีความสนใจที่จะนำวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับมาใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่า ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ซึ่งประกอบด้วย รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความยาวของแบบทดสอบ สัดส่วนของข้อสอบที่ทำหน้าที่เบี่ยงเบน และขนาดตัวอย่าง เพื่อให้เป็นไปตามเงื่อนไขต่าง ๆ ผู้วิจัยจึงอาศัยการจำลองข้อมูล โดย

กำหนดให้การแจกแจงความสามารถของผู้สอบ และคุณลักษณะของข้อสอบเป็นแบบโค้งปกติ ผลจากการศึกษาครั้งนี้จะเป็นประโยชน์ต่อการทดสอบทางการศึกษาและจิตวิทยาต่อไป

ความมุ่งหมายของการวิจัย

การวิจัยครั้งนี้มีความมุ่งหมายสำคัญ เพื่อศึกษาวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่า วิธีใดมีประสิทธิภาพในการตรวจสอบมากที่สุด ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ซึ่งประกอบด้วยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความยาวของแบบทดสอบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และขนาดตัวอย่าง โดยพิจารณาจากอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 สำหรับความมุ่งหมายของการวิจัยนี้ แยกเป็นความมุ่งหมายเฉพาะดังนี้

1. เพื่อศึกษาอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่าด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ภายใต้เงื่อนไขต่างกันของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ซึ่งประกอบด้วยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความยาวของแบบทดสอบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และขนาดตัวอย่าง

2. เพื่อเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่า ด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ภายใต้เงื่อนไขต่างกันของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ซึ่งประกอบด้วยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความยาวของแบบทดสอบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และขนาดตัวอย่าง

3. เพื่อเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่า ระหว่างวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ภายใต้เงื่อนไขต่างกันของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ซึ่งประกอบด้วยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความยาวของแบบทดสอบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และขนาดตัวอย่าง

ความสำคัญของการวิจัย

ในการวิจัยครั้งนี้มีเป้าหมายในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่า ภายใต้เงื่อนไขรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความยาวของแบบทดสอบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และขนาดตัวอย่าง ด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ โดยพิจารณาจากอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ทำให้สามารถประเมินข้อดีและ

ข้อจำกัดของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบภายใต้ปัจจัยดังกล่าว นักพัฒนาแบบทดสอบสามารถเลือกใช้วิธีการตรวจสอบให้เหมาะสมกับลักษณะของข้อสอบและผู้สอบ เพื่อให้ได้สารสนเทศสูงสุดในการตรวจสอบ อีกทั้งยังเป็นแนวทางในการศึกษาการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่า ซึ่งจะทำให้มีการคิดค้นและพัฒนาวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่มีประสิทธิภาพและประสิทธิผลต่อไป

ขอบเขตของการวิจัย

ขอบเขตของการวิจัยมีดังนี้

1. ข้อมูลที่ศึกษาเป็นข้อมูลจำลอง โดยจำลองภายใต้โมเดลพาร์เชียลเครดิตทั่วไปแบบมิติเดียว (Unidimensional Generalized Partial Credit Model) ทำให้ได้ข้อมูลที่เป็นข้อสอบที่วัดเพียงมิติเดียว ข้อสอบแต่ละข้อมีรายการตอบ 5 รายการ โดยให้คะแนนเป็น 1,2,3,4 หรือ 5 ในการจำลองข้อมูลดังกล่าวจะจำลองผลการตอบข้อสอบภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน 2 รูปแบบ ความยาวของแบบทดสอบ 3 ขนาด สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 3 ขนาด และขนาดตัวอย่าง 3 ขนาด รวมข้อมูลทั้งหมดที่ต้องจัดกระทำเพื่อตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบจำนวน 54 เงื่อนไข (2×3×3×3) ในแต่ละเงื่อนไขจำลองข้อมูลซ้ำ 50 ครั้ง

2. ตัวแปรที่ศึกษาประกอบด้วย

2.1 ตัวแปรอิสระ มี 5 ตัวแปร ดังนี้

2.1.1 วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบมี 3 วิธี ดังนี้

1) วิธีโพลีโตมัสซิบเทสท์

2) วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก

3) วิธีการถดถอยโลจิสติกแบบจัดอันดับ

2.1.2 รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนมี 2 รูปแบบ ดังนี้

1) ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน

2) ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน

2.1.3 ความยาวของแบบทดสอบมี 3 ขนาด ดังนี้

1) จำนวน 20 ข้อ

2) จำนวน 40 ข้อ

3) จำนวน 60 ข้อ

2.1.4 สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนมี 3 ขนาด ดังนี้

1) จำนวน 10 %

2) จำนวน 20 %

3) จำนวน 30 %

2.1.5 ขนาดตัวอย่างประกอบด้วยกลุ่มอ้างอิงและกลุ่มสนใจที่มีจำนวนผู้สอบในแต่ละกลุ่มเท่ากันมี 3 ขนาด ดังนี้

1) จำนวน 250 คน

2) จำนวน 500 คน

3) จำนวน 1000 คน

2.2 ตัวแปรตาม มี 2 ตัวแปร ดังนี้

2.2.1 อัตราความถูกต้อง

2.2.2 อัตราความคลาดเคลื่อนประเภทที่ 1

นิยามศัพท์เฉพาะ

1. ข้อสอบที่ให้คะแนนแบบหลายค่า (Polytomously scored items) หมายถึง ข้อสอบที่มีรายการตอบหลายรายการ ในแต่ละรายการตอบจะมีการกำหนดขั้นของการให้คะแนนผลการตอบข้อสอบไว้เป็นจำนวนเต็ม ซึ่งในการวิจัยนี้ข้อสอบแต่ละข้อมีรายการตอบ 5 รายการ โดยให้คะแนนเป็น 1,2,3,4 หรือ 5

2. การทำหน้าที่เบี่ยงเบนของข้อสอบ (Differential item functioning: DIF) หมายถึง ความแตกต่างในการทำหน้าที่ของข้อสอบ หลังจากกลุ่มผู้สอบที่แตกต่างกันมีความสามารถ หรือคุณลักษณะที่ข้อสอบนั้นวัดเท่ากัน มีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องไม่เท่ากัน

3. กลุ่มอ้างอิง (Reference group: R) หมายถึง กลุ่มผู้สอบที่คาดว่าจะได้เปรียบในการตอบข้อสอบเมื่อข้อสอบทำหน้าที่เบี่ยงเบนโดยมีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องมากกว่ากลุ่มอื่น

4. กลุ่มสนใจ (Focal group: F) หมายถึง กลุ่มผู้สอบซึ่งเป็นเป้าหมายของการศึกษา ซึ่งคาดว่าจะเสียเปรียบในการตอบข้อสอบเมื่อข้อสอบทำหน้าที่เบี่ยงเบน โดยมีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องน้อยกว่ากลุ่มอ้างอิง

5. รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน (Form of DIF) หมายถึง ขนาด และทิศทางของข้อสอบทำหน้าที่เบี่ยงเบน ซึ่งแปรเปลี่ยนตามความแตกต่างของระดับความสามารถของผู้สอบ ในทฤษฎีการตอบข้อสอบที่วัดความสามารถมิติเดียวและให้คะแนนแบบหลายค่า สามารถพิจารณารูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนได้จากความแตกต่างของพารามิเตอร์อำนาจจำแนกของข้อสอบ และพารามิเตอร์ความยากของข้อสอบ ระหว่างกลุ่มอ้างอิงและกลุ่มสนใจ การวิจัยครั้งนี้ศึกษารูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน 2 รูปแบบดังนี้

5.1 ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF) หมายถึง ข้อสอบที่ทำให้ค่าความแตกต่างของความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจมีค่าคงที่ทุกระดับความสามารถ ในการวิจัยครั้งนี้กำหนดให้พารามิเตอร์ความยาก

ของข้อสอบระหว่างกลุ่มมีค่าแตกต่างกัน 0.25 ส่วนพารามิเตอร์อำนาจจำแนกข้อสอบระหว่างกลุ่มมีค่าเท่ากัน

5.2 ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) หมายถึง ข้อสอบที่ทำให้ความแตกต่างของความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องระหว่างผู้สอบกลุ่มอ้างอิง และกลุ่มสนใจมีค่าไม่คงที่ตลอดช่วงความสามารถ โดยที่ช่วงความสามารถหนึ่งกลุ่มสนใจมีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องสูงกว่ากลุ่มอ้างอิง ขณะที่อีกช่วงความสามารถหนึ่งกลุ่มอ้างอิงมีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องสูงกว่ากลุ่มสนใจ ในการวิจัยครั้งนี้กำหนดให้พารามิเตอร์อำนาจจำแนกข้อสอบระหว่างกลุ่มมีค่าแตกต่างกัน 1.0 ส่วนพารามิเตอร์ความยากของข้อสอบระหว่างกลุ่มมีค่าเท่ากัน

6. ความยาวของแบบทดสอบ (Test length) หมายถึง จำนวนของข้อสอบทั้งหมดในแบบทดสอบ ในการวิจัยนี้ ศึกษาความยาวของแบบทดสอบ 3 ขนาด คือ แบบทดสอบที่มีจำนวนข้อสอบ 20 ข้อ, 40 ข้อ และ 60 ข้อ

7. สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน (Proportion of DIF) หมายถึง จำนวนข้อสอบที่ทำหน้าที่เบี่ยงเบนต่อจำนวนข้อสอบทั้งหมดในแบบทดสอบ ในการวิจัยครั้งนี้ศึกษาสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 3 ขนาด คือ จำนวน 10%, 20% และ 30% ดังนั้น สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนจำนวน 10%, 20% และ 30% ในแบบทดสอบที่มีความยาว 20 ข้อ จะมีข้อสอบทำหน้าที่เบี่ยงเบนกับข้อสอบทำหน้าที่ไม่เบี่ยงเบน จำนวน 2 ข้อกับ 18 ข้อ, 4 ข้อกับ 16 ข้อ และ 6 ข้อกับ 14 ข้อ ตามลำดับ สำหรับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนจำนวน 10%, 20% และ 30% ในแบบทดสอบที่มีความยาว 40 ข้อ จะมีข้อสอบทำหน้าที่เบี่ยงเบนกับข้อสอบทำหน้าที่ไม่เบี่ยงเบนจำนวน 4 ข้อกับ 36 ข้อ, 8 ข้อกับ 32 ข้อ และ 12 ข้อกับ 28 ข้อ ตามลำดับ และสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนจำนวน 10%, 20% และ 30% ในแบบทดสอบที่มีความยาว 60 ข้อ จะมีข้อสอบทำหน้าที่เบี่ยงเบนกับข้อสอบทำหน้าที่ไม่เบี่ยงเบนจำนวน 6 ข้อกับ 54ข้อ , 12 ข้อกับ 48 ข้อ และ 18 ข้อกับ 42 ข้อ ตามลำดับ

8. ขนาดตัวอย่าง (Sample size) หมายถึง จำนวนผู้สอบกลุ่มอ้างอิง และกลุ่มสนใจที่ใช้ในการวิเคราะห์ข้อสอบที่ทำหน้าที่เบี่ยงเบน ในการวิจัยครั้งนี้ใช้จำนวนผู้สอบของกลุ่มอ้างอิง และกลุ่มสนใจเท่ากัน 3 ขนาด คือ 250 คน, 500 คน และ 1000 คน ดังนั้นจำนวนผู้สอบกลุ่มอ้างอิงต่อจำนวนผู้สอบกลุ่มสนใจ เป็น 250 คน : 250 คน, 500 คน : 500 คน และ 1000 คน : 1000 คน

9. วิธีโพลีโตมัสซิบเทสท์ (Polytomous SIBTEST: Poly-SIB) หมายถึงวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในกลุ่มวิธีที่ใช้คุณลักษณะแฝง ซึ่งเป็นไปตามทฤษฎีการตอบข้อสอบ รูปแบบนอนพาราเมตริก (Nonparametric form) ซึ่งพัฒนาโดยชาง มาซซีโอ และ รุสโซ (Chang; Mazzeo; & Roussos. 1996) เพื่อใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่า ในการศึกษาครั้งนี้นำมาตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัด

ความสามารถมิติเดียวและให้คะแนนแบบหลายค่า โดยใช้คะแนนรวมของแบบทดสอบซึ่งค่อนข้างเป็นคุณลักษณะแฝงเป็นเกณฑ์การจับคู่ความสามารถ

10. วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก (Logistic discriminant function analysis: LDFA) หมายถึงวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในกลุ่มวิธีที่ใช้คะแนนสังเกต ซึ่งเป็นไปตามทฤษฎีการทดสอบมาตรฐานเดิม รูปแบบพาราเมตริก (Parametric form) ซึ่งพัฒนาโดยมิลเลอร์และสเปรย์ (Miller; & Spray. 1993) เพื่อใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่า ในการศึกษาครั้งนี้นำมาตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถมิติเดียวและให้คะแนนแบบหลายค่า โดยใช้คะแนนรวมของแบบทดสอบเป็นเกณฑ์การจับคู่ความสามารถ

11. วิธีการถดถอยโลจิสติกแบบจัดอันดับ (Ordinal logistic regression: OLR) หมายถึงวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในกลุ่มวิธีที่ใช้คะแนนสังเกต ซึ่งเป็นไปตามทฤษฎีการทดสอบมาตรฐานเดิม รูปแบบพาราเมตริก (Parametric form) ซึ่งพัฒนาโดยซัมโบ (Zumbo. 1999) เพื่อใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่า ในการศึกษาครั้งนี้นำมาตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถมิติเดียวและให้คะแนนแบบหลายค่า โดยใช้คะแนนรวมของแบบทดสอบเป็นเกณฑ์การจับคู่ความสามารถ

12. อัตราความถูกต้อง (Power rate) หมายถึง สัดส่วนของจำนวนข้อสอบที่ตรวจสอบพบว่าทำหน้าที่เบี่ยงเบนได้อย่างถูกต้อง โดยคำนวณจากจำนวนข้อสอบที่ตรวจสอบได้ถูกต้องว่าทำหน้าที่เบี่ยงเบน ซึ่งตรวจสอบด้วยวิธีที่ศึกษา (วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ) ต่อจำนวนข้อสอบที่ทำหน้าที่เบี่ยงเบนทั้งหมดในแบบทดสอบ

13. อัตราความคลาดเคลื่อนประเภทที่ 1 (Type I error rate) หมายถึง สัดส่วนของจำนวนข้อสอบที่ตรวจสอบผิดพลาดว่าทำหน้าที่เบี่ยงเบนทั้งที่ความเป็นจริงข้อสอบทำหน้าที่ไม่เบี่ยงเบน โดยคำนวณจากจำนวนข้อสอบที่ตรวจสอบผิดพลาดว่าทำหน้าที่เบี่ยงเบน ซึ่งตรวจสอบด้วยวิธีที่ศึกษา (วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ) ต่อจำนวนข้อสอบที่ทำหน้าที่ไม่เบี่ยงเบนทั้งหมดในแบบทดสอบ

กรอบแนวคิดในการวิจัย

เนื่องจากการวัดผลการศึกษาปัจจุบันได้ให้ความสำคัญกับข้อสอบที่ให้คะแนนแบบหลายค่าสูงขึ้น คุณภาพของแบบทดสอบที่ประกอบด้วยข้อสอบดังกล่าวที่นำมาใช้จึงมีความสำคัญ ควรเป็นแบบทดสอบที่วัดได้ตรงตามสิ่งที่ต้องการวัด มีความเป็นมิติเดียว (Unidimensional test) คือวัดความสามารถอย่างใดอย่างหนึ่งเพียงความสามารถเดียวหรือคุณลักษณะเดียว แม้ว่าโดยทั่วไปแล้วจะเป็นการยากที่จะสร้างแบบทดสอบให้เป็นไปตามข้อตกลงดังกล่าว ในทางปฏิบัติเราสามารถเลี่ยงไปใช้ข้อตกลงที่ผ่อนคลาย (Stout. 1987: 589-618) โดยพิจารณามิติเด่น (Dominant dimension)

เพียงมิติเดียว คือให้แบบทดสอบนั้นวัดคุณลักษณะที่ต้องการวัดเป็นส่วนใหญ่ และอาจมีคุณลักษณะอื่นๆ ที่เกี่ยวข้องบ้างเป็นสิ่งที่หลีกเลี่ยงไม่ได้ ความเป็นมิติเดียวในลักษณะนี้เรียกว่า ความเป็นมิติเดียวที่สำคัญ (Essential unidimensional) และที่สำคัญคือ แบบทดสอบต้องมีความยุติธรรม ปราศจากความลำเอียง ซึ่งความยุติธรรมของแบบทดสอบนั้นจะต้องมีการแปลความหมายของคะแนนที่ได้จากแบบทดสอบของสมาชิกในกลุ่มได้ถูกต้อง และมีความเสมอภาค ซึ่งผู้สอบทุกคนมีโอกาสแสดงศักยภาพตามโครงสร้างแบบทดสอบที่วัด การที่จะทำให้แบบทดสอบมีความยุติธรรมนั้น นักพัฒนาแบบทดสอบจึงจำเป็นต้องตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบเพื่อเป็นหลักฐานพยานทางสถิติแล้วคัดเลือกข้อที่มีความลำเอียงทางการวัดออกจากแบบทดสอบ ซึ่งถือเป็นขั้นตอนการพัฒนาแบบทดสอบ

วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่าที่นักการศึกษาและนักวิจัยอื่น ๆ ให้ความสนใจในปัจจุบันที่สำคัญ 3 วิธี คือวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ซึ่งวิธีโพลีโตมัสซิบเทสท์ พัฒนาโดย ชาง มาซซีโอ และรุสโซ (Chang; Mazzeo; & Roussos. 1996) เป็นวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่า ในกลุ่มวิธีที่ใช้ทฤษฎีการตอบข้อสอบ รูปแบบนอนพาราเมตริก (Nonparametric form) ในการตรวจสอบจะใช้คะแนนรวมของแบบทดสอบเป็นเกณฑ์การจับคู่ความสามารถของกลุ่มผู้สอบ วิธีนี้สามารถคิดคำนวณง่ายไม่ยุ่งยากซับซ้อน เสียค่าใช้จ่ายไม่มาก การแปลผลไม่ยาก ทั้งยังใช้กลุ่มตัวอย่างขนาดเล็ก และใช้สถิติทดสอบนัยสำคัญ สำหรับวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก พัฒนาโดย มิลเลอร์และสเปรย์ (Miller; & Spray. 1993) เป็นวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่าในกลุ่มวิธีที่ใช้ทฤษฎีการทดสอบมาตรฐานเดิม รูปแบบพาราเมตริก (Parametric form) ในการตรวจสอบจะใช้คะแนนรวมของแบบทดสอบเป็นเกณฑ์การจับคู่กลุ่มผู้สอบ ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับ พัฒนาโดยซัมโบ (Zumbo. 1999) เป็นวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่าในกลุ่มวิธีที่ใช้ทฤษฎีการทดสอบมาตรฐานเดิม รูปแบบพาราเมตริก (Parametric form) เช่นเดียวกับวิธีการวิเคราะห์ฟังก์ชันการถดถอยโลจิสติก ในการตรวจสอบจะใช้คะแนนรวมของแบบทดสอบเป็นเกณฑ์การจับคู่เช่นเดียวกับวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก นักการศึกษาได้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่า เช่น มิลเลอร์และสเปรย์ (Miller; & Spray. 1993) และ เวลช์และมิลเลอร์ (Welch; & Miller. 1995) ได้ใช้วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก (LDFA) ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่มีการให้คะแนนแบบหลายค่าโดยใช้ข้อมูลจริงจากแบบทดสอบ ส่วนสเปรย์และมิลเลอร์ (Spray; & Miller. 1994) ได้นำวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก (LDFA) วิธีแมนเทล (Mantel) และวิธีแมน-เทลแฮนส์เซลทั่วไปมาใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) ขณะที่ชาง มาซซีโอและรุสโซ (Chang; Mazzeo; & Roussos. 1996) ได้เปรียบเทียบวิธีแมนเทล (Mantel) วิธี

ซิบเทสท์ปรับใหม่ (Modified SIBTEST) และวิธีความแตกต่างเฉลี่ยมาตรฐาน (Standardized mean difference) พบว่า ทั้งสามวิธีให้ผลเหมือนกัน นอกจากนี้เฟร็นช์และมิลเลอร์ (French; & Miller.1996) ได้ใช้วิธีการถดถอยโลจิสติกแบบมัลติโนเมียล (Multinomial logistic regression: MLR) ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ไม่เป็นรูปแบบเดียวกัน สำหรับคริสต์แจนเซ่น (Kristjansson. 2001) ได้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ โดยใช้วิธีการถดถอยโลจิสติกแบบจัดอันดับ (OLR) วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก (LDFA) วิธีแมนเทล (Mantel) และวิธีแมนเทล-แฮนส์เซลทั่วไป (GMH) โดยใช้ข้อมูลจำลอง

ในการศึกษาครั้งนี้ ผู้วิจัยได้นำวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับมาใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่มีการให้คะแนนแบบหลายค่า ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความยาวของแบบทดสอบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และขนาดตัวอย่าง มีรายละเอียดดังนี้

รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน

ปัจจัยแรกที่สนใจคือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน โดยปกตินักการศึกษาหรือนักวิจัยอื่น ๆ นิยมศึกษาในสองรูปแบบคือ ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF) และข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) จากการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบสองค่า (Swaminathan; & Rogers. 1990; Rogers; & Swaminathan. 1993; Narayanan; & Swaminathan. 1994, 1996; Hidalgo; & Lopez-Pina. 2004; จิติมา วรรณศรี. 2539; รัชรินทร์ มุคดา. 2540; นพมาศ พิพัฒนสุข. 2541; ทองอยู่ สาระ. 2543; วลีมาศ แซ่อึ้ง. 2543) ผลการศึกษาสรุปได้ว่าวิธีแมนเทล-แฮนส์เซล และวิธีซิบเทสท์ มีประสิทธิภาพในการตรวจสอบข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน ส่วนวิธีการถดถอยโลจิสติกมีประสิทธิภาพในการตรวจสอบข้อสอบทำหน้าที่เบี่ยงเบนทั้งที่เป็นรูปแบบเดียวกันและที่ไม่เป็นรูปแบบเดียวกัน ที่เป็นเช่นนี้สืบเนื่องมาจากวิธีแมนเทล-แฮนส์เซล และวิธีซิบเทสท์ ใช้สถิติแบบคิดเครื่องหมาย เมื่อข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกันในทฤษฎีการตอบข้อสอบ สามารถพิจารณาได้จากฟังก์ชันการตอบข้อสอบ (Item response functions: IRFs) ระหว่างกลุ่มตัดกันตรงช่วงกึ่งกลางของความสามารถ ทำให้ทิศทางของข้อสอบทำหน้าที่เบี่ยงเบน ณ ตำแหน่งนั้นมีผลทำให้ความแตกต่างของผลรวมฟังก์ชันการตอบข้อสอบแบบมาร์จินอล (Marginal IRFs) ที่เป็นบวกหักล้าง (Cancellation) กับผลรวมฟังก์ชันการตอบข้อสอบแบบมาร์จินอลที่เป็นลบ ดังนั้นวิธีแมนเทล-แฮนส์เซล และวิธีซิบเทสท์ จึงไม่มีความไวในการตรวจพบข้อสอบที่ทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ส่วนวิธีการถดถอยโลจิสติกสามารถใช้โมเดลทดสอบอิทธิพลของปฏิสัมพันธ์ระหว่างระดับความสามารถกับการเป็นสมาชิกของกลุ่มผู้สอบทำให้สามารถตรวจพบทั้งข้อสอบทำหน้าที่เบี่ยงเบนทั้งที่เป็นรูปแบบเดียวกันและไม่เป็นรูปแบบ

เดียวกัน สำหรับการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่า (Spray; & Miller. 1994; Zwick; Donoghue; & Grima. 1993; French; & Miller. 1996; Chang; Mazzeo; & Roussos. 1996; Kristjansson; et al. 2005; Su; & Wang. 2005; อุทัยวรรณ สายพัฒนะ. 2547; อรินทร์ น่วมถนอม. 2549) วิธีต่างๆ มักได้มาจากการขยายวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบสองค่า เช่นวิธีที่ได้จากการขยายวิธีแมนเทล-แฮนส์เซล ได้แก่ วิธีแมนเทล และวิธีแมนเทล-แฮนส์เซลทั่วไป และวิธีที่ได้จากการขยายของวิธีซิบเทสท์ ได้แก่ วิธีโพลีโตมัสซิบเทสท์ วิธีโคร-ซิบ อื่นๆ วิธีดังกล่าวข้างต้นมีประสิทธิภาพในการตรวจพบข้อสอบที่ทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน แต่จะไม่มีความไวในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ส่วนวิธีที่ได้จากการขยายของวิธีการถดถอยโลจิสติก ได้แก่ วิธีการถดถอยโลจิสติกแบบมัลติโนเมียล วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีประสิทธิภาพในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนทั้งที่เป็นรูปแบบเดียวกันและที่ไม่เป็นรูปแบบเดียวกัน จากการศึกษาของสเปรย์ และมิลเลอร์ (Spray; & Miller. 1994) พบว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก มีประสิทธิภาพในการตรวจสอบการทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกันสูงกว่า วิธีแมนเทล และวิธีแมนเทล-แฮนส์เซลทั่วไป สำหรับการศึกษาข้อมูลจำลองของความเบี่ยงเบนของข้อสอบที่มีการตรวจให้คะแนนหลายค่า ในกรณีการทำหน้าที่เบี่ยงเบนของข้อสอบที่มีรูปแบบเดียวกันมีการกำหนดขนาดของค่าความยากของกลุ่มอ้างอิงและกลุ่มสนใจแตกต่างกันเช่น .10, .25, .50, 1.0 ส่วนในกรณีข้อสอบทำหน้าที่เบี่ยงเบนไม่ที่เป็นรูปแบบเดียวกัน จะกำหนดค่าอำนาจจำแนกของกลุ่มอ้างอิงและกลุ่มสนใจแตกต่างกัน เช่น 0.5, 1.00 (Zwick; et al. 1993, 1997; Welch; & Hoover. 1993; Spray; & Miller. 1994; Chang; et al. 1996; French; & Miller. 1996; Tian. 1999; Ankenmann; et al. 1999) ในการศึกษาครั้งนี้ผู้วิจัยสนใจศึกษาทั้งข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และที่ไม่เป็นรูปแบบเดียวกัน เมื่อข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน จะกำหนดค่าความยากของกลุ่มอ้างอิงและกลุ่มสนใจแตกต่างกัน 0.25 และเมื่อข้อสอบทำหน้าที่เบี่ยงเบนไม่ที่เป็นรูปแบบเดียวกัน จะกำหนดค่าอำนาจจำแนกของกลุ่มอ้างอิงและกลุ่มสนใจแตกต่างกัน 1.0

ความยาวของแบบทดสอบ

ปัจจัยที่สองที่สนใจคือ ความยาวของแบบทดสอบ ทั้งนี้เพราะความยาวของแบบทดสอบจะมีผลกระทบต่อความถูกต้องในการจับคู่เปรียบเทียบระหว่างกลุ่มผู้สอบ เนื่องจากวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับใช้คะแนนรวมจากการสอบแทนคุณลักษณะภายในหรือระดับความสามารถของผู้สอบที่วัดได้ แบบทดสอบที่มีความยาวมากกว่าย่อมส่งผลให้มีความน่าเชื่อถือมากกว่า ซึ่งจะทำให้การจับคู่ระหว่างกลุ่มผู้สอบมีความถูกต้องมากขึ้น ผลการศึกษาของสวามินาธานและโรเจอร์ส (Swaminathan; & Rogers.1990) ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธี

แมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกที่มีความยาวของแบบทดสอบ 40 ข้อ, 60 ข้อ และ 80 ข้อ โดยใช้ข้อมูลจำลอง พบว่า เมื่อใช้แบบทดสอบที่ยาวมากขึ้น อัตราความถูกต้องของทั้งสองวิธีมีค่ามากขึ้น ยกเว้นกรณีข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกันของวิธีแมนเทล-แฮนส์เซล ต่อมาทั้งสองคนได้ศึกษาอีกครั้ง (Rogers; & Swaminathan. 1993) โดยใช้แบบทดสอบที่มีความยาว 40 และ 80 ข้อ ตรวจสอบด้วยวิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก พบว่า ความยาวของแบบทดสอบไม่มีผลต่ออัตราความถูกต้องของวิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ยกเว้นในกรณีข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกันของวิธีการถดถอยโลจิสติก ซึ่งให้ผลการศึกษาขัดแย้งกับครั้งแรก ส่วนผลการศึกษาของกาญจนา วัธนสุนทร (2537) ปรากฏว่า ความยาวของแบบทดสอบไม่มีผลกระทบต่ออัตราความถูกต้องของการตรวจสอบของวิธีแมนเทล- แฮนส์เซล และวิธีซิบเทสท์ ซึ่งสอดคล้องกับทองอยู่ สาระ (2543) ซึ่งได้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยแบบทดสอบที่มีความยาว 20 ข้อ, 40 ข้อ และ 60 ข้อ ด้วยวิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก พบว่า ความยาวของแบบทดสอบไม่มีผลต่ออัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ซึ่งขัดแย้งกับการศึกษาของจิติมา วรรณศรี (2539) พบว่า เมื่อใช้แบบทดสอบขนาด 60 ข้อ จะมีผลทำให้วิธีแมนเทล-แฮนส์เซล และวิธีซิบเทสท์มีประสิทธิภาพในการตรวจสอบดีที่สุด ส่วนผลการศึกษาของโคเฮ็นและคิม (Cohen; & Kim. 1993) ใช้ความยาวของแบบทดสอบ 20 ข้อ และ 60 ข้อ ในการตรวจสอบการทำหน้าที่ของข้อสอบด้วยวิธีการทดสอบไค-สแควร์ของลอร์ด และวิธีการวัดพื้นที่ชนิดคิดเครื่องหมายและชนิดไม่คิดเครื่องหมายของราชู พบว่าเมื่อเพิ่มความยาวของแบบทดสอบแล้วอัตราความคลาดเคลื่อนประเภทที่ 1 จะเพิ่มขึ้น ในขณะที่ผลการศึกษาของอุททาโรและมิลล์แซบ (Uttaro; & Millsap. 1994) พบว่าเมื่อใช้แบบทดสอบที่มีความยาวมากขึ้นแล้วอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซล จะลดลง ในการศึกษาครั้งนี้ผู้วิจัยสนใจศึกษาความยาวของแบบทดสอบ 3 ขนาด คือ 20 ข้อ, 40 ข้อ และ 60 ข้อ

สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน

ปัจจัยที่สามที่สนใจ คือ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน เนื่องจากการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบจำเป็นต้องเปรียบเทียบความสามารถของผู้สอบภายใต้คะแนนเกณฑ์การจับคู่ที่มีความเชื่อมั่น (Clauser; & Mazor. 1988: 37; Potenza; & Dorans. 1995: 32) ถ้าในแบบทดสอบมีข้อสอบทำหน้าที่เบี่ยงเบนปะปนอยู่ จะมีผลทำให้ค่าประมาณความสามารถมีความเชื่อมั่นต่ำลง เกณฑ์การจับคู่จึงขาดความแม่นยำ ซึ่งจะมีผลทำให้อัตราความถูกต้องลดลง หรืออัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าปกติ จากผลการศึกษาของ มอลเลอร์ (Maller. 2001) สังเกตเห็นว่าเศษหนึ่งส่วนสามโดยประมาณของข้อสอบใน Wechsler Intelligence Scale for Children-Third-Edition แสดง DIF โดยมีบางข้อเข้าข้างเด็กผู้ชาย และบางข้อเข้าข้างเด็กผู้หญิง ส่วนบัดเจลล์ ราชู และควอเท็ตไท (Budgell; Raju; & Quartetti. 1995) ได้ทบทวนการศึกษา DIF ในเครื่องมือการวัดประเมินที่แปล และพบว่าจำนวนของข้อสอบที่มี DIF อยู่ในช่วงจาก 1.5% ถึง

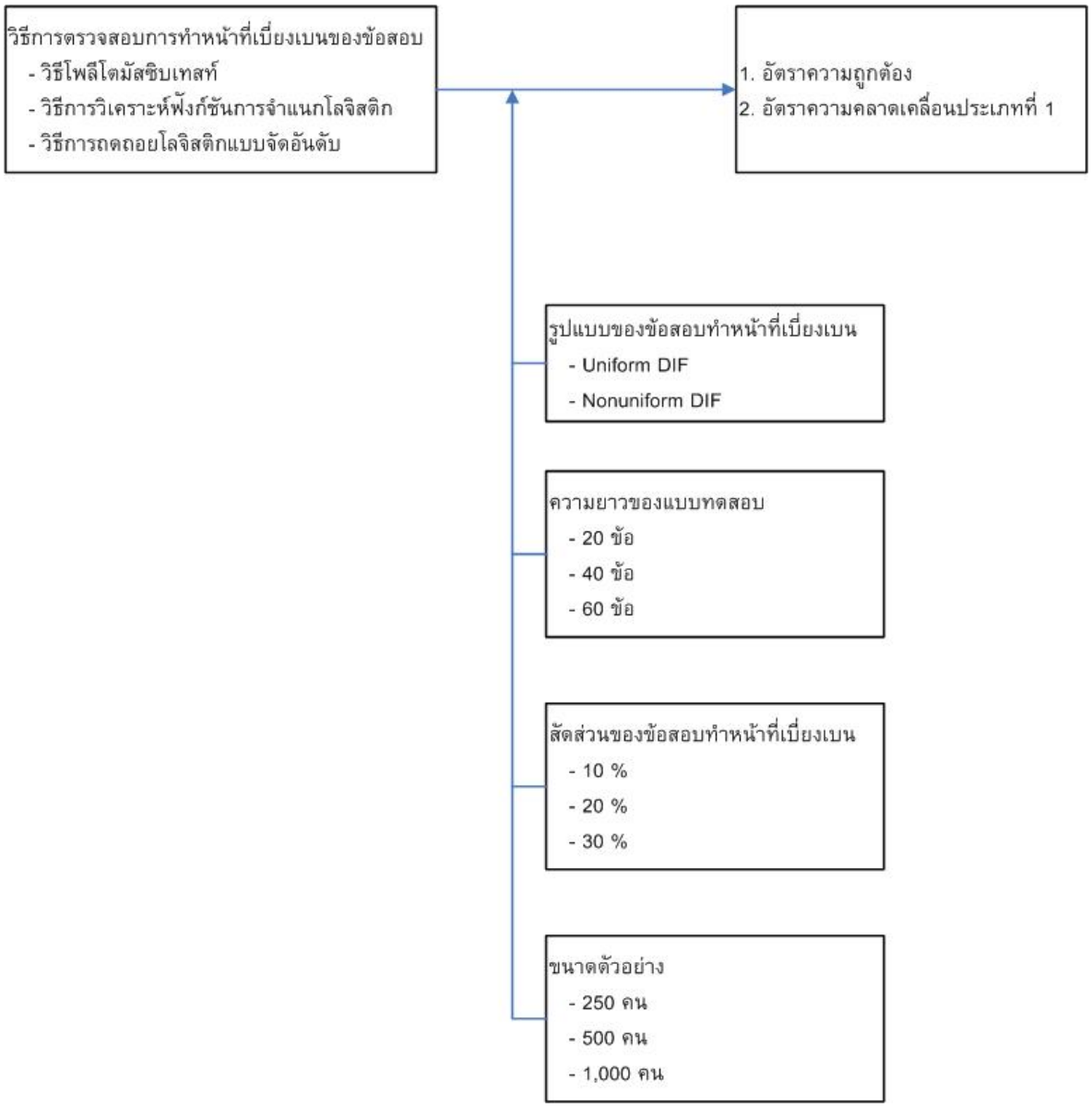
64% พร้อมกับมีการศึกษาหลายครั้งที่พบ DIF เกินกว่า 30% ของข้อสอบ ส่วนฮวง เชิร์ช และคาติกแบ็ก (Su; & Wang. 2005; citing Haung; Church; & Katigbak. 1977) ได้ทำการตรวจสอบถึงการวัดที่เท่าเทียมกันของข้อสอบที่วัดข้ามวัฒนธรรมของ NEO Personality Inventory ซึ่งเป็นการวัดของโมเดลบุคลิกภาพ 5 องค์ประกอบ ในบริบททางวัฒนธรรมที่ต่างกัน โดยใช้ภาษาอังกฤษเป็นเครื่องมือ เกือบจะ 40% ของข้อสอบ 180 ข้อ ถูกพบว่า แสดง DIF จะเห็นว่าแบบทดสอบที่ใช้กัน (Real test) อาจบรรจุจำนวนร้อยละของข้อสอบที่ทำหน้าที่เบี่ยงเบน (DIF)ระดับสูง และจากผลการศึกษาของ โรเจอร์ส และสวามินาธาน (Rogers; & Swaminathan. 1993) ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก พบว่าเมื่อใช้จำนวนสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 2 ระดับคือ 0% และ 15% จำนวนสัดส่วนของข้อสอบที่ทำหน้าที่เบี่ยงเบนดังกล่าวไม่มีผลต่ออัตราความถูกต้อง ยกเว้นในกรณีการทำหน้าที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกัน เมื่อจำนวนสัดส่วนของข้อสอบที่ทำหน้าที่เบี่ยงเบนในแบบทดสอบมีจำนวนเพิ่มขึ้น จะมีผลทำให้อัตราความถูกต้องของวิธีการถดถอยโลจิสติกมีค่าลดลง ส่วนนารายานานและสวามินาธาน (Narayanan; & Swaminathan. 1994) ได้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกันด้วยวิธีซิบเทสท์ และวิธีการถดถอยโลจิสติก โดยศึกษาปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 2 ระดับ คือ จำนวน 10% และ 20% พบว่าเมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนในแบบทดสอบมีจำนวนเพิ่มขึ้นจะมีผลทำให้อัตราความถูกต้องของวิธีซิบเทสท์และวิธีการถดถอยโลจิสติกมีค่าลดลง นอกจากนี้ยังส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งสองวิธีมีค่าลดลงด้วย ต่อมานารายานานและ สวามินาธาน (Narayanan; & Swaminathan. 1996) ได้ตรวจสอบอีกครั้งโดยตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ไม่เป็นรูปแบบเดียวกันด้วยวิธีแมนเทล-แฮนส์เซล วิธีโคร-ซิบ และวิธีการถดถอยโลจิสติก โดยศึกษาปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่เบี่ยงเบน 3 ระดับ คือ จำนวน 0%, 10% และ 20% พบว่า เมื่อสัดส่วนของข้อสอบที่ทำหน้าที่เบี่ยงเบนในแบบทดสอบมีจำนวนเพิ่มขึ้นจะมีผลทำให้อัตราความถูกต้องของวิธีการถดถอยโลจิสติกมีค่าลดลงแต่จะไม่มีผลต่ออัตราความถูกต้องของวิธีแมนเทล-แฮนส์เซล และวิธีโคร-ซิบ สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่าเมื่อจำนวนสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนในแบบทดสอบมีจำนวนเพิ่มขึ้นจะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งสามวิธีจะมีค่าเพิ่มขึ้นด้วย โดยวิธีแมนเทล-แฮนส์เซล มีค่าต่ำสุด ส่วนวลีมาศ แซ่อึ้ง (2543) ได้ศึกษาเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบแบบอเนกรูป (ที่ไม่เป็นรูปแบบเดียวกัน) ระหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ปัจจัยหนึ่งที่ใช้คือสัดส่วนของข้อสอบทำหน้าที่ต่างกัน 3 ขนาด (5%, 10% และ 20%) ผลการวิจัยพบว่า สัดส่วนของข้อสอบทำหน้าที่ต่างกันไม่มีผลต่ออำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่ และวีซิปเทสท์ แต่มีผลต่อวิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ในขณะที่อรินทร์ น่วมถนอม (2549) ได้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ พบว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการ

ถดถอยโลจิสติกแบบจัดอันดับหลายมิติ มีอัตราความถูกต้องใกล้เคียงกัน ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกันและที่ไม่เป็นรูปแบบเดียวกัน ทั้งสองวิธีมีอัตราความถูกต้องสูงกว่าวิธีโพลีโตมัสซิบเทสท์ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ไม่เป็นรูปแบบเดียวกัน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนในแบบทดสอบไม่มีผลต่อวิธีโพลีโตมัสซิบเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ แต่มีผลต่อวิธีการถดถอยโลจิสติกแบบจัดอันดับ ในการศึกษาครั้งนี้ ผู้วิจัยสนใจศึกษาสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 3 ขนาดคือ จำนวน 10%, 20% และ 30%

ขนาดตัวอย่าง

ปัจจัยสุดท้ายที่สนใจ คือ ขนาดตัวอย่าง ทั้งนี้เพราะการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ เป็นการเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มสนใจ ขนาดตัวอย่างถือว่ามีความสำคัญในการทดสอบนัยสำคัญทางสถิติ เพราะวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีรูปแบบพาราเมตริก (Parametric form) ซึ่งจำเป็นต้องใช้ข้อตกลงของโมเดลอธิบายความสัมพันธ์ระหว่างผลการตอบข้อสอบกับตัวแปรการจับคู่ ในขณะที่วิธีโพลีโตมัสซิบเทสท์มีรูปแบบนอนพาราเมตริก (Nonparametric form) ซึ่งไม่มีข้อตกลงดังกล่าว (Potenza; & Dorans. 1995: 24) โดยทั่วไปแล้ววิธีที่มีรูปแบบพาราเมตริกมักใช้ขนาดตัวอย่างมากกว่าวิธีที่มีรูปแบบนอนพาราเมตริก ถึงแม้ว่าวิธีในตระกูลโมเดลการถดถอยโลจิสติก (วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ) กับวิธีโพลีโตมัสซิบเทสท์มีรูปแบบทางสถิติแตกต่างกัน แต่วิธีดังกล่าวตั้งอยู่บนพื้นฐานของทฤษฎีการตอบข้อสอบและใช้สถิติการทดสอบภายใต้การแจกแจงแบบเชิงเส้นกำกับ (Asymptotic distribution) ซึ่งการแจกแจงดังกล่าวจำเป็นต้องใช้ขนาดตัวอย่างที่เพียงพอเพื่อให้ผลการวิเคราะห์มีความถูกต้อง จากผลการศึกษาของสวามินาธานและโรเจอร์ส (Swaminathan; & Rogers. 1990) พบว่าในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกันและที่ไม่เป็นรูปแบบเดียวกัน พบว่าเมื่อขนาดตัวอย่างเพิ่มขึ้นจะมีผลทำให้อัตราความถูกต้องของวิธีการถดถอยโลจิสติกมีค่าเพิ่มมากขึ้นเกือบทุกเงื่อนไข ส่วนอัตราความถูกต้องของวิธีแมนเทล-แฮนส์เซล จะมีค่าเพิ่มมากขึ้นเฉพาะกรณีข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกันเท่านั้น ขณะที่อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซล มีค่าลดลงเกือบทุกเงื่อนไข ต่อมาทั้งสองคนได้ศึกษาใหม่อีกครั้ง (Rogers; & Swaminathan. 1993) ผลปรากฏว่า ขนาดตัวอย่างมีผลต่ออัตราความถูกต้องของวิธีการถดถอยโลจิสติกและวิธีแมนเทล-แฮนส์เซล เมื่อขนาดตัวอย่างเพิ่มมากขึ้นจะมีผลทำให้อัตราความถูกต้องของทั้งสองวิธีเพิ่มมากขึ้น สำหรับทองอยู่ สาระ (2543) ได้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่มีรูปแบบเดียวกันและที่ไม่มีรูปแบบเดียวกันพบว่าขนาดตัวอย่าง และอัตราส่วนขนาดตัวอย่างของกลุ่มอ้างอิงและกลุ่มสนใจ (300:100, 300:200, 300:300, 500:100, 500:200, 500:300, 1000:100, 1000:200, 1000:300 และ 1000:1000) มีผลต่ออัตราความถูกต้อง

ของวิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก เกือบทุกเงื่อนไข กล่าวคือ เมื่อขนาดตัวอย่างเพิ่มขึ้นมีผลทำให้อัตราความถูกต้องเพิ่มขึ้น แต่จะไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งสองวิธีเกือบทุกเงื่อนไข ในขณะที่ สิริรัตน์ วิภาสศิลป์ (2545) พบว่า ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีซิบเทสท์ และวิธีดีเอฟไอที ขนาดตัวอย่าง 50, 100 และ 200 คน ไม่มีผลต่ออัตราความถูกต้องของวิธีซิบเทสท์และมีค่าต่ำกว่าเมื่อขนาดตัวอย่าง 500 และ 1000 คน สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 ก็จะมีค่าต่ำกว่าด้วย แต่การตรวจสอบด้วยวิธีดีเอฟไอที ทั้ง 5 ขนาด (50, 100, 200, 500 และ 1000 คน) ไม่มีผลต่ออัตราความถูกต้องของวิธีนี้ ในการศึกษาครั้งนี้ ผู้วิจัยสนใจศึกษาขนาดตัวอย่าง 3 ขนาด คือ 250, 500 และ 1000 คน โดยกลุ่มอ้างอิงกับกลุ่มสนใจมีขนาดตัวอย่างเท่ากัน คือ $N_R=N_F=250$, $N_R=N_F=500$ และ$N_R=N_F=1000$



ภาพประกอบ 1 กรอบแนวคิดในการวิจัย

สมมติฐานในการวิจัย

1. ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ประกอบด้วยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความยาวของแบบทดสอบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนและขนาดตัวอย่าง มีผลทำให้อัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่าด้วยวิธีโพลีโตมัสซิบเทสท์แตกต่างกัน

2. ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ประกอบด้วยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความยาวของแบบทดสอบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนและขนาดตัวอย่างมีผลทำให้อัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่าด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกแตกต่างกัน

3. ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ประกอบด้วยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความยาวของแบบทดสอบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนและขนาดตัวอย่างมีผลทำให้อัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่าด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับแตกต่างกัน

4. วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่าภายใต้เงื่อนไขปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ประกอบด้วยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความยาวของแบบทดสอบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และขนาดตัวอย่าง มีอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกัน

บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

ในการศึกษาครั้งนี้ ผู้วิจัยนำเสนอแนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้องเรียงลำดับดังนี้

1. แนวคิดเกี่ยวกับการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ
2. โมเดลการทำหน้าที่เบี่ยงเบนของข้อสอบทางสถิติ
3. วิธีตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ
4. งานวิจัยที่เกี่ยวข้อง

1. แนวคิดเกี่ยวกับการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ

การศึกษาเรื่องความยุติธรรม (Fairness) ของข้อสอบหรือแบบทดสอบในกรณีที่ข้อสอบทำให้ผู้สอบระหว่างกลุ่มย่อยเกิดการได้เปรียบเสียเปรียบกัน เดิมเรียกว่าความลำเอียงของข้อสอบ (Item bias) หรือความลำเอียงของแบบทดสอบ (Test bias) มีการศึกษาอย่างจริงจังในช่วงปลายทศวรรษ 1960 ในประเทศสหรัฐอเมริกา ซึ่งเป็นประเทศที่ประชากรมีความแตกต่างเกี่ยวกับเชื้อชาติ เผ่าพันธุ์ ศาสนา ภาษา วัฒนธรรม ประสบการณ์ สถานะทางเศรษฐกิจและสังคมหรือภูมิหลังอื่นๆ ในช่วงเวลาดังกล่าวมีการใช้แบบทดสอบเป็นเครื่องมือวัดที่สำคัญ มีการนำแบบทดสอบไอคิวมาใช้ในการคัดเลือกเพื่อศึกษาต่อ เพื่อบรรจุเข้าทำงาน เพื่อจัดบุคคลเข้าสู่ตำแหน่ง ในปี 1971 บริษัทกริกส์ วี. ดุ๊ก เพาเวอร์ (Griggs v. Duke Power) ชี้ว่าแบบทดสอบไอคิวที่ใช้ในการสอบคัดเลือกไม่เหมาะสม ผลการใช้แบบทดสอบไม่เป็นไปตามสัดส่วนระดับสติปัญญา และเกิดความลำเอียงต่อกลุ่มผู้สอบ บริษัทดังกล่าวจึงสรุปว่าการนำผลการทดสอบไอคิวมาใช้ในการคัดเลือกอาจไม่เหมาะสม ยกเว้นกรณีที่ผลของการปฏิบัติงานมีความสอดคล้องกับการทดสอบ ทำให้มีการนำเสนอวิธีต่างๆ เพื่อนำไปใช้ศึกษาความแตกต่างทางวัฒนธรรมในการทำข้อสอบของนักเรียนและบุคคลกลุ่มต่างๆ เพื่อระบุข้อสอบที่ไม่เหมาะสมหรือลำเอียงต่อผู้สอบกลุ่มย่อยและคัดข้อสอบที่ลำเอียงออกจากแบบทดสอบ จะเห็นว่าการนำแบบทดสอบที่มีความลำเอียงทางการวัดไปใช้จึงเป็นการไม่ยุติธรรมแก่ผู้สอบ และผลที่ได้จากการวัดก็ไม่สามารถสะท้อนถึงความสามารถที่แท้จริงของผู้สอบได้ ทั้งนี้เพราะแต่ละกลุ่มมีการได้เปรียบเสียเปรียบกันในการตอบข้อสอบข้อเดียวกัน ในการทดสอบถ้าจะให้เกิดความยุติธรรมกับผู้สอบ แบบทดสอบที่นำมาใช้ควรเป็นแบบทดสอบมาตรฐานที่มีความเที่ยงตรง ความเชื่อมั่น ความยากพอเหมาะ พร้อมทั้งมีประสิทธิภาพในการจำแนกระดับความสามารถผู้สอบได้ นอกจากนี้ยังต้องคำนึงถึงความยุติธรรมต่อผู้สอบ เพื่อไม่ให้เกิดการบิดเบือนในผลการสอบของผู้สอบ ซึ่งจะช่วยให้ผู้มีหน้าที่เกี่ยวข้องนำผลที่ได้ไปใช้ด้วยความถูกต้องและเป็นธรรม

แต่เดิมคำว่า **“ความลำเอียงของข้อสอบ”** เป็นคำที่ใช้ในการศึกษาหรือทดสอบความยุติธรรมของข้อสอบ คำดังกล่าวมีความหมายทั้งเชิงสถิติและเชิงสังคม ทำให้เกิดความสับสนในการนำไปใช้ เพื่อให้เกิดความชัดเจนในปัจจุบันนักวิจัยส่วนใหญ่ใช้คำว่า **“การทำหน้าที่เบี่ยงเบนของ**

ข้อสอบ" (Diferential item functioning) เรียกสั้นๆ ว่า **"DIF"** แทนคำว่า **"ความลำเอียงของข้อสอบ"** (Item bias) (Holland; & Wainer.1993:xiv; Angoff. 1993: 4) คำสองคำนี้ไม่เหมือนกัน แต่มีความเกี่ยวข้องกัน กล่าวคือการทำหน้าที่เบี่ยงเบนของข้อสอบเป็นคุณลักษณะทางสถิติของข้อสอบที่ตรวจได้ด้วยวิธีการวิเคราะห์ทางสถิติ ซึ่งเป็นส่วนประกอบหนึ่งของสิ่งที่แสดงถึงความลำเอียงของข้อสอบ วิธีการทางสถิติที่ใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบเป็นเงื่อนไขที่จำเป็น (Necessary condition) ในการตัดสินความลำเอียงของข้อสอบ แต่ไม่ใช่เงื่อนไขที่เพียงพอ (not sufficient condition) ในการตัดสินความลำเอียงของข้อสอบ (Clauser; & Hambleton. 1994: 88; Clauser; & Mazor. 1998: 31) เนื่องจากถ้าใช้วิธีการทางสถิติตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบเพียงอย่างเดียวแล้ว ผลการตรวจสอบพบว่าข้อสอบทำหน้าที่เบี่ยงเบนนั้นยังสรุปไม่ได้ว่าข้อสอบลำเอียงหรือไม่ ยังต้องใช้ผู้เชี่ยวชาญพิจารณาเนื้อหาสาระของข้อสอบและจุดมุ่งหมายในการวัดของข้อสอบที่เรียกว่า "วิธีการตัดสินข้อสอบ"(Judgmental method) (Camilli; & Shepard. 1994:135)

สำหรับความหมายของการทำหน้าที่เบี่ยงเบนของข้อสอบ ได้มีนักวิจัยหลายคนกำหนดนิยามไว้ดังนี้

การทำหน้าที่เบี่ยงเบนของข้อสอบ หมายถึง ผู้สอบที่มีความสามารถระดับเดียวกันแต่มาจากกลุ่มผู้สอบที่แตกต่างกัน มีความน่าจะเป็นในการตอบข้อสอบถูกไม่เท่ากัน (Hambleton; Swaminathan; & Rogers. 1990:110; Roussos; & Stout. 1996: 355; Stout; et al. 1997; 195; Mazor; Hambleton; & Clauser. 1998: 357; Bolt. 2002: 115; Bolt; & Gierl 2006: 313)

การทำหน้าที่เบี่ยงเบนของข้อสอบ หมายถึง การสังเกตได้ว่าข้อสอบแสดงคุณสมบัติทางสถิติที่แตกต่างกันในผลการตอบข้อสอบจากกลุ่มผู้สอบแตกต่างกัน ภายหลังการควบคุมความแตกต่างในความสามารถของกลุ่มผู้สอบ (Angoff. 1993 : 4-5)

การทำหน้าที่เบี่ยงเบนของข้อสอบ หมายถึง ความแตกต่างในการทำหน้าที่ของข้อสอบภายหลังการจับคู่ความสามารถของกลุ่มผู้สอบซึ่งเป็นความสามารถตามที่ข้อสอบต้องการวัด (Dorans; & Holland. 1993: 37)

การทำหน้าที่เบี่ยงเบนของข้อสอบ หมายถึง การเปรียบเทียบความสามารถระหว่างผู้สอบสองกลุ่มหรือมากกว่า ความน่าจะเป็นในการตอบข้อสอบถูกมีความเกี่ยวข้องกับการเป็นสมาชิกของกลุ่มผู้สอบ (Camilli. 1993: 397)

การทำหน้าที่เบี่ยงเบนของข้อสอบ หมายถึงฟังก์ชันการตอบข้อสอบซึ่งคำนวณจากสมาชิกของผู้สอบกลุ่มย่อยที่แตกต่างกันมีค่าไม่เท่ากัน (Narayanan; & Swaminathan. 1996: 257; Kim; Cohen; & Kim. 1994: 217)

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

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

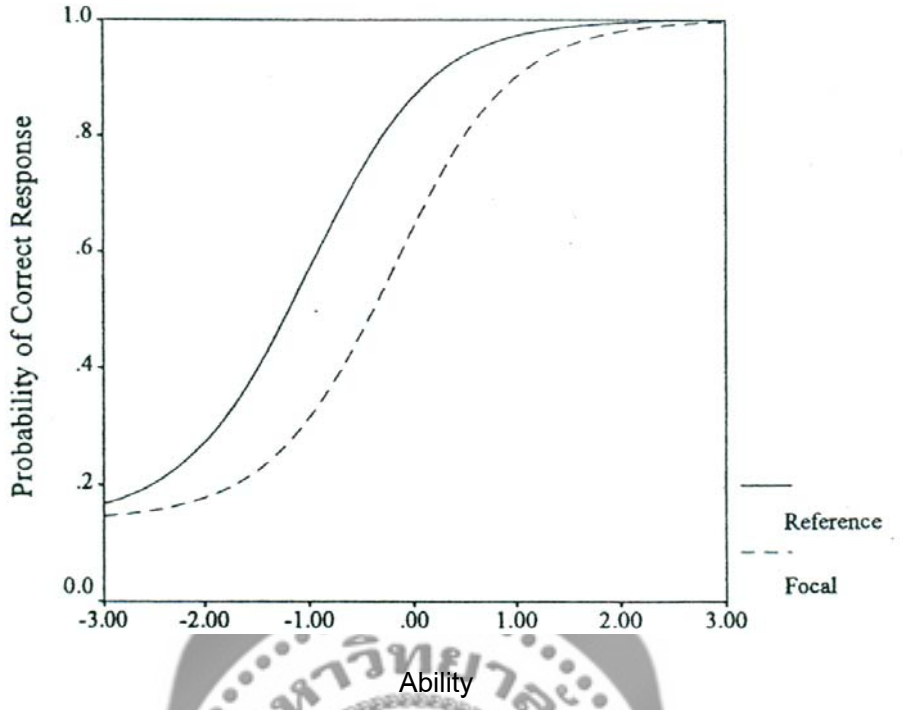
การทำหน้าที่เบี่ยงเบนของข้อสอบ หมายถึง ผู้สอบจากกลุ่มที่แตกต่างกันซึ่งมีระดับความสามารถเท่ากันตามที่ข้อสอบประเมิน มีความน่าจะเป็นในการตอบข้อสอบถูกแตกต่างกัน หรือมีคะแนนดิบที่คาดหวังจากการตอบข้อสอบแตกต่างกัน (Raju; & Ellis. 2002: 156)

การทำหน้าที่เบี่ยงเบนของข้อสอบ หมายถึง ผู้สอบจากกลุ่มที่แตกต่างกันมีความน่าจะเป็นในการตอบข้อสอบถูกแตกต่างกันภายหลังควบคุมความสามารถ หรือวางเงื่อนไขบนความสามารถ (Ackerman; Gierl; & Walker. 2003: 48; Gierl; Gotzmann; & Boughton. 2004: 242)

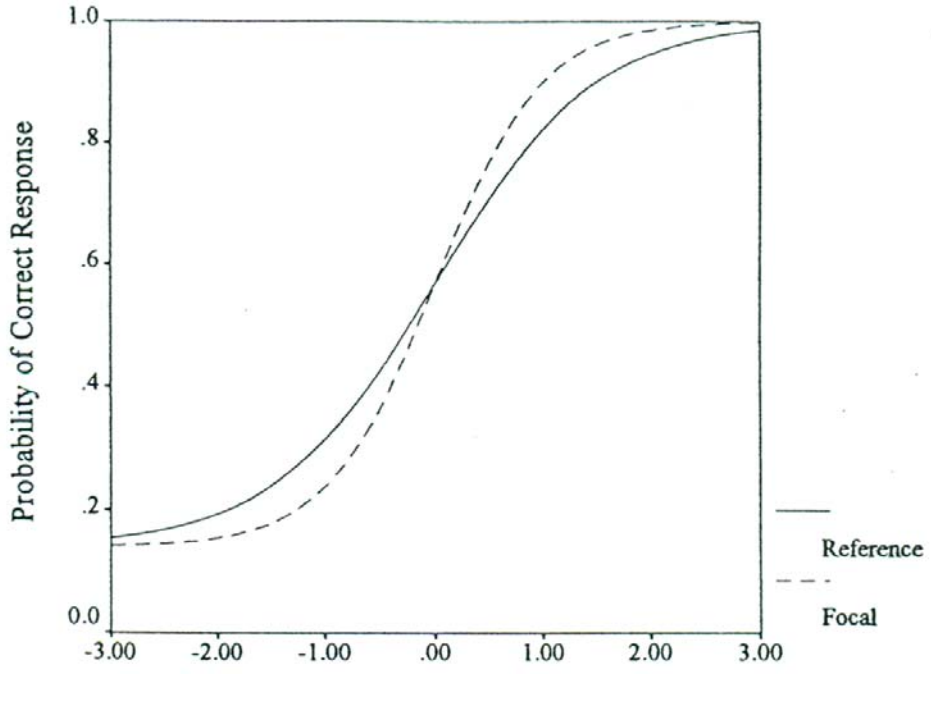
จากนิยามดังกล่าวสรุปได้ว่า **"การทำหน้าที่เบี่ยงเบนของข้อสอบ"** (Differential item functioning;DIF) หมายถึง ความแตกต่างในการทำหน้าที่ของข้อสอบ หลังจากกลุ่มผู้สอบที่แตกต่างกันมีความสามารถ หรือ คุณลักษณะที่ข้อสอบนั้นวัดเท่ากัน มีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องไม่เท่ากัน

ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ จะเปรียบเทียบผลการตอบข้อสอบระหว่างผู้สอบกลุ่มย่อยสองกลุ่มที่มีความสามารถระดับเดียวกัน โดยที่ผู้สอบกลุ่มหนึ่งเป็นตัวแทนกลุ่มหลัก (Majority group) ในประชากรเรียกว่า **"กลุ่มอ้างอิง"** (Reference group: **R**) ใช้เป็นกลุ่มสำหรับเปรียบเทียบกับอีกกลุ่มหนึ่งซึ่งเป็นตัวแทนกลุ่มรอง (Minority group) ในประชากรเรียกว่า **"กลุ่มสนใจ"** (Focal group: **F**) ถือเป็นกลุ่มเป้าหมายที่ต้องการศึกษา สำหรับเกณฑ์ที่ใช้ในการจำแนกผู้สอบเป็นกลุ่มสนใจและกลุ่มอ้างอิงมีหลายลักษณะ เช่น เพศ สีผิว เชื้อชาติ ภาษา วัฒนธรรม ภูมิลำเนา เป็นต้น (Holland; & Wainer.1993: xiii) ส่วนข้อสอบที่ใช้ในการตรวจสอบเรียกว่า **"ข้อสอบที่ศึกษา"** (Studied item) ถ้าข้อสอบทำหน้าที่เบี่ยงเบนแล้ว ความน่าจะเป็นในการตอบข้อสอบถูกของผู้สอบแต่ละกลุ่มจะไม่เท่ากัน โดยคาดว่ากลุ่มอ้างอิงเป็นกลุ่มที่ข้อสอบเข้าข้างทำให้ผู้สอบในกลุ่มนี้จะได้เปรียบในการตอบข้อสอบ ส่วนกลุ่มสนใจเป็นกลุ่มที่ข้อสอบไม่เข้าข้างทำให้ผู้สอบในกลุ่มนี้เสียเปรียบในการตอบข้อสอบ (Donoghue; Holland; & Thayer. 1993: 137-138; Roussos; Stout. 1996: 216) การทำหน้าที่เบี่ยงเบนของข้อสอบเป็นคุณลักษณะทางสถิติของข้อสอบที่แสดงให้เห็นความแตกต่างในการทำหน้าที่ของข้อสอบต่อกลุ่มอ้างอิงและกลุ่มสนใจที่นำมาเปรียบเทียบกัน โดยจะเปรียบเทียบผลการตอบข้อสอบที่ศึกษาระหว่างกลุ่มทั้งสอง หลังจากควบคุมความแตกต่างระดับความสามารถภายใต้ตัวแปรการจับคู่ (Matching variable) วิธีที่ใช้คะแนนสังเกตได้ (Observed score) ซึ่งเป็นไปตามทฤษฎีการทดสอบมาตรฐานเดิม (Classical test theory: CTT) มักใช้คะแนนรวมของแบบทดสอบเป็นเกณฑ์การจับคู่กลุ่มผู้สอบ ส่วนวิธีที่ใช้คุณลักษณะแฝง (Unobserved score) ซึ่งเป็นไปตามทฤษฎีการตอบข้อสอบ (Item response theory: IRT) จะใช้ค่าประมาณระดับความสามารถของผู้สอบเป็นเกณฑ์การจับคู่กลุ่มผู้สอบ ถ้าความแตกต่างของผลการตอบข้อสอบระหว่างกลุ่มผู้สอบเกิดขึ้นจากการแจกแจงความสามารถของผู้สอบ

ทั้งหมดซึ่งไม่ได้มีการควบคุมตัวแปรดังกล่าว จะเรียกว่า “ผลกระทบของข้อสอบ” (Item impact) (Ackerman. 1992: 69; Camilli. 1992: 131; Dorans; & Holland. 1993: 37; Donoghue; Holland; & Thayer. 1993:141) การเกิดผลกระทบของข้อสอบถือเป็นเรื่องปกติทั่ว ๆ ไป เนื่องจากบุคคลแต่ละคนมีคุณสมบัติแตกต่างกันออกไปตามหลักความแตกต่างระหว่างบุคคล เช่น กลุ่มผู้สอบเพศชายได้คะแนนเฉลี่ยในวิชาคณิตศาสตร์สูงกว่ากลุ่มผู้สอบเพศหญิง เป็นต้น ทิศทางของผลกระทบและทิศทางของข้อสอบที่ทำหน้าที่เบี่ยงเบนอาจจะไม่เป็นไปในทิศทางเดียวกัน ดังเช่น โดยภาพรวมแล้วข้อสอบวิชาคณิตศาสตร์อาจยากต่อกลุ่มผู้สอบเพศหญิง แต่เมื่อพิจารณาโดยการจับคู่กลุ่มผู้สอบตามความสามารถแล้วข้อสอบข้อนั้นอาจจะง่ายต่อกลุ่มผู้สอบเพศหญิงก็ได้ ดังเช่น ข้อสรุปของซิมป์สัน (Simpson’s paradox) (Dorans; & Holland. 1993: 37-38; citing Simpson. 1951) จากการวิเคราะห์การทำหน้าที่เบี่ยงเบนของข้อสอบ เมื่อพิจารณาขนาดและทิศทางของข้อสอบที่ทำหน้าที่เบี่ยงเบนพบว่า จะแปรเปลี่ยนตามระดับความสามารถของผู้สอบ สามารถแบ่งข้อสอบทำหน้าที่เบี่ยงเบนได้ 2 ประเภท คือ **“ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน”** (Uniform DIF) และ **“ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน”** (Nonuniform DIF) (Mellenbergh. 1982: 110) ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน เป็นข้อสอบที่ทำให้ผู้สอบกลุ่มหนึ่งมีโอกาสในการตอบข้อสอบถูกมากกว่าผู้สอบอีกกลุ่มหนึ่งอย่างสม่ำเสมอ อาจกล่าวได้ว่าความน่าจะเป็นในการตอบข้อสอบถูกของผู้สอบกลุ่มหนึ่งมากกว่าอีกกลุ่มหนึ่งอย่างคงที่ทุกระดับความสามารถ ข้อสอบที่ทำหน้าที่เบี่ยงเบนประเภทนี้เกิดขึ้นเมื่อไม่มีปฏิสัมพันธ์ (Interaction) ระหว่างความสามารถของผู้สอบกับการเป็นสมาชิกของกลุ่ม ภายใต้ทฤษฎีการตอบข้อสอบสามารถพิจารณาได้จากโค้งลักษณะข้อสอบ (Item characteristic curves: ICCs) หรือฟังก์ชันการตอบข้อสอบ (Item response functions: IRFs) ระหว่างกลุ่มผู้สอบเหมือนกันและขนานกัน ดังภาพประกอบ 2 ส่วนข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน เป็นข้อสอบที่ทำให้โอกาสในการตอบข้อสอบถูกของผู้สอบระหว่างกลุ่มแตกต่างกันอย่างไม่สม่ำเสมอในทุกระดับความสามารถ นั่นคือ ความแตกต่างของความน่าจะเป็นในการตอบข้อสอบถูกของผู้สอบสองกลุ่มไม่เหมือนกันทุกระดับความสามารถ เช่นที่ระดับความสามารถหนึ่ง ผู้สอบกลุ่มอ้างอิงมีความน่าจะเป็นในการตอบข้อสอบถูกมากกว่าผู้สอบกลุ่มสนใจ แต่ที่ระดับความสามารถอีกระดับหนึ่ง ผู้สอบกลุ่มสนใจมีความน่าจะเป็นในการตอบข้อสอบถูกมากกว่าผู้สอบกลุ่มอ้างอิง ข้อสอบทำหน้าที่เบี่ยงเบนประเภทนี้เกิดขึ้นเมื่อมีปฏิสัมพันธ์ระหว่างความสามารถของผู้สอบกับการเป็นสมาชิกของกลุ่ม นั่นแสดงว่าโค้งลักษณะข้อสอบ (ICCs) ระหว่างกลุ่มผู้สอบจะไม่ขนานกัน หรือมีฟังก์ชันการตอบข้อสอบ (IRFs) ต่างกันนั่นเอง ดังภาพประกอบ 3 (Li; & Stout. 1996: 647; Swaminathan; & Rogers. 1990: 361)



ภาพประกอบ 2 ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF)

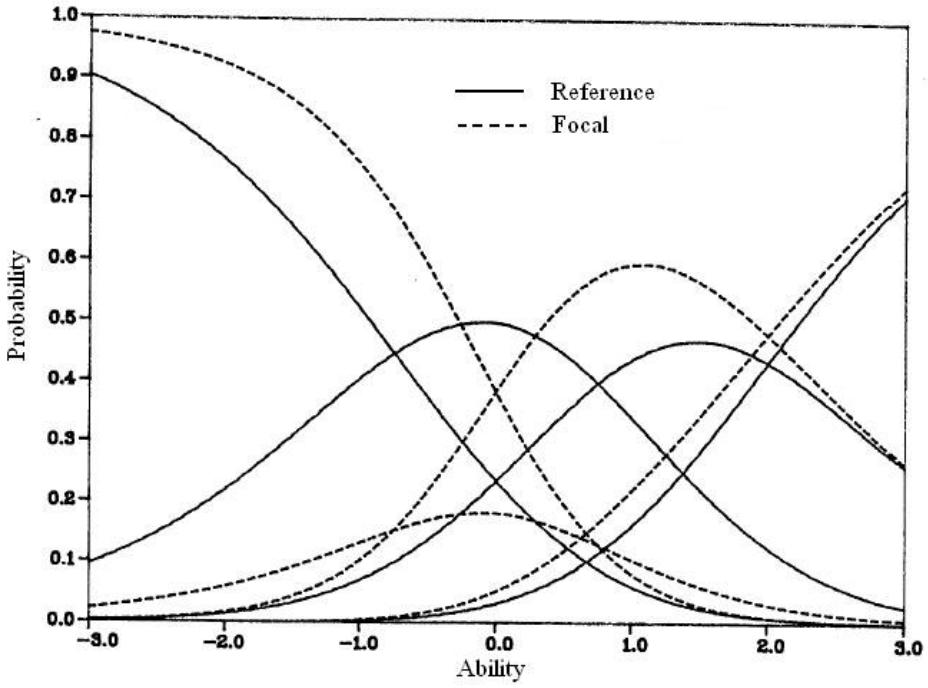


ภาพประกอบ 3 ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF)

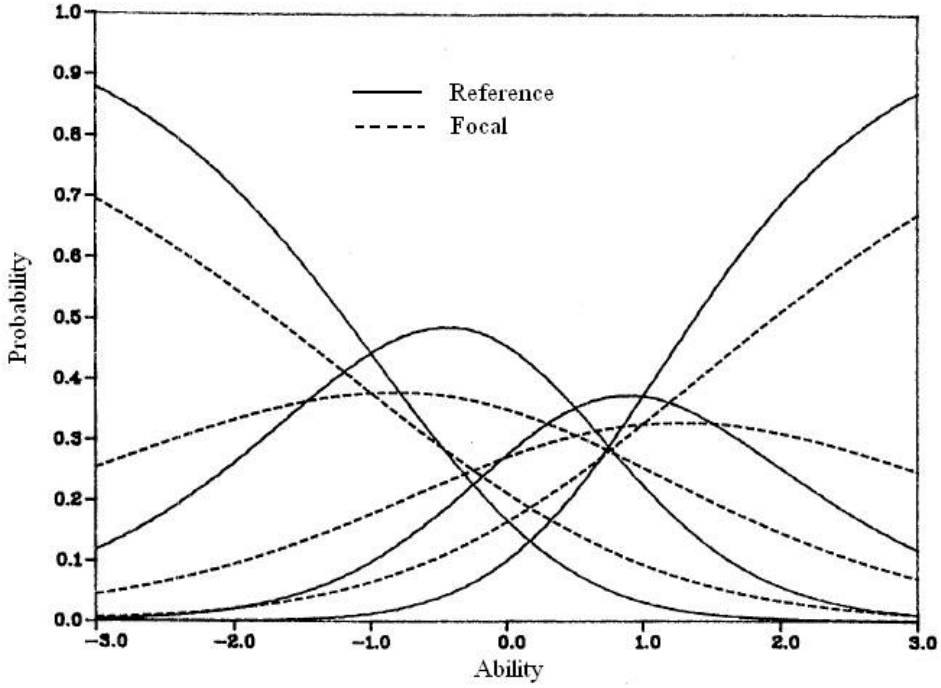
สามารถพิจารณา **“ปฏิสัมพันธ์”** ดังกล่าวได้จากความแตกต่างของค่าพารามิเตอร์อำนาจจำแนกข้อสอบ (Item discrimination) ระหว่างกลุ่มผู้สอบ กล่าวคือ ถ้าค่าพารามิเตอร์อำนาจจำแนกของข้อสอบระหว่างกลุ่มผู้สอบเท่ากันแล้วโค้งลักษณะข้อสอบ (Item characteristic curves: ICCs) ระหว่างผู้สอบกลุ่มย่อยจะขนานกัน แสดงว่าข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน แต่ถ้าค่าพารามิเตอร์อำนาจจำแนกของข้อสอบระหว่างกลุ่มผู้สอบไม่เท่ากัน จะทำให้โค้งลักษณะข้อสอบระหว่างกลุ่มไม่ขนานกัน แสดงว่าข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Swaminathan; & Rogers. 1990: 361; Li; & Stout. 1996: 647; Su; & Wang. 2005: 316) โค้งลักษณะข้อสอบดังกล่าวจะเป็นฟังก์ชันเพิ่มทิศทางเดียว (Monotonically inereasing function) ซึ่งแสดงถึงความสัมพันธ์ระหว่างผลการปฏิบัติข้อสอบกับระดับความสามารถ(คุณลักษณะ) ถ้าระดับของความสามารถเพิ่มขึ้นความน่าจะเป็นของการตอบข้อสอบได้ถูกต้องก็จะเพิ่มขึ้นด้วย นอกจากนี้ผลการศึกษาการทำหน้าที่เบี่ยงเบนของข้อสอบที่ไม่เป็นรูปแบบเดียวกันของ สวามินาธาน และ โรเจอร์ส (Swaminathan; & Rogers. 1990: 369) พบว่าข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกันมี 2 ลักษณะ คือข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกันโดยมี**ปฏิสัมพันธ์ไม่เป็นลำดับ** (Disordinal interaction) ซึ่งจะเกิดขึ้นเมื่อโค้งลักษณะข้อสอบตัดกันตรงจุดกึ่งกลางของช่วงความสามารถตามมาตรของทฤษฎีการตอบข้อสอบ (ความสามารถอยู่ในช่วง -3 ถึง +3) และข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกันโดยมี**ปฏิสัมพันธ์เป็นลำดับ** (Ordinal interaction) จะเกิดขึ้นเมื่อโค้งลักษณะข้อสอบตัดกันนอกช่วงความสามารถตามมาตรของทฤษฎีการตอบข้อสอบ กล่าวคือ อาจตัดกันตรงปลายสุดของช่วงความสามารถต่ำหรือสูง หรือโค้งลักษณะข้อสอบระหว่างกลุ่มซึ่งไม่ขนานกันแต่ไม่ตัดกัน สถานการณ์นี้อาจเกิดขึ้นเมื่อใช้โมเดลแบบ 3 พารามิเตอร์ (Narayanan; & Swaminathan. 1996: 258) ดังนั้นโค้งลักษณะข้อสอบที่ไม่ขนานกันของข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกันอาจจะตัดกันหรือไม่ตัดกันก็ได้ ส่วนเชียลี และ สเตาท์ (Shealy; & Stout. 1993) ได้ศึกษาการทำหน้าที่เบี่ยงเบนของข้อสอบและใช้คำว่า “ข้อสอบทำหน้าที่เบี่ยงเบนที่มีทิศทางเดียวกัน” (Unidirectional DIF) และ “ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่มีทิศทางเดียวกัน” (Nonunidirectional DIF) แทนคำว่า “ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน”(Uniform DIF) และ ”ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน” (Nonunidirectional DIF) ตามลำดับ ขณะที่ลิและสเตาท์ (Li; & Stout. 1993) ได้เรียกข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกันที่มีปฏิสัมพันธ์ไม่เป็นลำดับว่า **“ข้อสอบทำหน้าที่เบี่ยงเบนแบบไม่มีทิศทางเดียวกัน”** (Nondirectional DIF) และเรียกข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกันที่มีปฏิสัมพันธ์เป็นลำดับว่า **“ข้อสอบทำหน้าที่เบี่ยงเบนแบบมีทิศทางเดียวกัน”** (Unidirectional DIF) ซึ่งเมื่อพิจารณาถึงโค้งลักษณะข้อสอบเพื่อไม่ให้เกิดความสับสนและมีความชัดเจนมากขึ้น ต่อมา ลิ และ สเตาท์ (Li; & Stout. 1996: 648) จึงเรียกข้อสอบทำหน้าที่เบี่ยงเบนแบบไม่มีทิศทางเดียวกันว่า **“ข้อสอบทำหน้าที่เบี่ยงเบนแบบตัดกัน”** (Crossing DIF)

ส่วนข้อสอบทำหน้าที่เบี่ยงเบนแบบมีทิศทางเดียวกันใช้คำว่า **“ข้อสอบทำหน้าที่เบี่ยงเบนไม่ตัดกัน”** (Non-crossing DIF) อีกด้วย

สำหรับเงื่อนไขของ “ปฏิสัมพันธ์” ดังกล่าวเป็นรูปแบบของการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบสองค่า (Dichotomous) ส่วนข้อสอบที่ให้คะแนนหลายค่า (Polytomous) จะมีความสมบูรณ์มากกว่า ทั้งนี้ เพราะว่าไม่เพียงจะเกิดปฏิสัมพันธ์ระหว่างระดับความสามารถกับการเป็นสมาชิกของกลุ่มเท่านั้น แต่ยังมีตัวแปรที่สามคือระดับคะแนนของข้อสอบเข้ามาเกี่ยวข้องด้วย ดังนั้นการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่าจะเกิดขึ้นภายในรายการคะแนนทั้งหมด (Score categores) (French; & Miller. 1996: 315) รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่ให้คะแนนแบบหลายค่าจะมีลักษณะคล้ายกับรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่ให้คะแนนแบบสองค่า ซึ่งสามารถแสดงด้วยฟังก์ชันการตอบรายการข้อสอบ (Item-category response function: ICRFs) โดยใช้โมเดลพาร์เชียลเครดิตทั่วไป (Generalized partial credit model: GPCM) แสดงดังภาพประกอบ 4 และ 5



ภาพประกอบ 4 ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF) ภายใต้โมเดลพาร์เชียลเครดิตทั่วไป (GPCM)



ภาพประกอบ 5 ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) ภายใต้โมเดลพาร์เชียลเครดิตทั่วไป (GPCM)

ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบกลุ่มวิธีที่ใช้คุณลักษณะแฝงตามทฤษฎีการตอบข้อสอบภายใต้โมเดลแบบมิติเดียว ระยะเริ่มต้นมักใช้โมเดลโลจิสติกแบบ 1,2 หรือ 3 พารามิเตอร์ ซึ่งใช้กับข้อสอบที่มีการตรวจให้คะแนนแบบสองค่า (Dichotomous IRT model) โมเดลการตอบข้อสอบแบบ 3 พารามิเตอร์ (3 PLM) มีโค้งลักษณะข้อสอบที่เขียนด้วยฟังก์ชันโลจิส ดังสมการ

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

เมื่อ $P_i(\theta)$ แทนความน่าจะเป็นของผู้สอบที่มีความสามารถ θ จะตอบข้อที่ i ได้ถูกต้อง สำหรับ a_i แทนค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ (Item discrimination) b_i แทน ค่าพารามิเตอร์ความยากของข้อสอบ (Item difficulty) และ c_i แทนค่าพารามิเตอร์การเดา (Item guessing) ส่วน D เป็นค่าปรับมาตรมีค่าคงที่เท่ากับ 1 หรือ 1.7

ถ้ากำหนดให้ค่าพารามิเตอร์การเดา (c_i) มีค่าเป็นศูนย์ จะกลายเป็นโมเดลการตอบข้อสอบแบบ 2 พารามิเตอร์ (2 PLM) มีโค้งลักษณะข้อสอบที่เขียนด้วยฟังก์ชันโลจิส ดังสมการ

$$P_i(\theta) = \frac{\exp[Da_i(\theta - b_i)]}{1 + \exp[Da_i(\theta - b_i)]} \quad (2)$$

โมเดลการตอบข้อสอบแบบ 2 พารามิเตอร์ดังกล่าว ถ้ากำหนดให้ค่าพารามิเตอร์อำนาจจำแนก (a_i) มีค่าคงที่ทุกข้อในขณะที่พารามิเตอร์การเดามีค่าเป็นศูนย์ โมเดลนี้จะกลายเป็นโมเดลการตอบข้อสอบแบบ 1 พารามิเตอร์ (1 PLM) หรือโมเดลราส์ช (Rasch model) ซึ่งเป็นโมเดลที่อธิบายคุณลักษณะของข้อสอบด้วยพารามิเตอร์ความยาก (b_i) เพียงอย่างเดียวเท่านั้น โมเดลการตอบข้อสอบแบบ 1 พารามิเตอร์มีโค้งลักษณะข้อสอบที่เขียนด้วยฟังก์ชันโลจิส ดังสมการ

$$P_i(\theta) = \frac{\exp(\theta - b_i)}{1 + \exp(\theta - b_i)} \quad (3)$$

เมื่อนำโมเดลทฤษฎีการตอบข้อสอบที่ใช้กับข้อสอบที่มีการตรวจให้คะแนนแบบสองค่า (Dichotomous IRT model) มาวิเคราะห์ข้อคำถามที่ใช้มาตรประมาณค่าหรือแบบทดสอบวัดผลสัมฤทธิ์ที่มีการตรวจให้คะแนนแบบหลายค่า สารสนเทศเกี่ยวกับการตอบข้อสอบส่วนหนึ่งจะหายไป เช่นถ้าเป็นการวิเคราะห์ข้อสอบแบบเลือกตอบที่ให้คะแนนแบบสองค่าจะต้องกำหนดให้มีตัวถูกเพียงหนึ่งตัวและได้คะแนนเป็น 1 นอกนั้นเป็นตัวลวงได้คะแนนเป็น 0 ทั้งหมด แต่ถ้าเป็นข้อสอบหรือแบบวัดที่เป็นมาตรประมาณค่าอาจต้องกำหนดให้การตอบในรายการ (Category) ที่ 1,2 และ 3 ได้คะแนนเป็น 0 ส่วนการตอบในรายการที่ 4 และ 5 ได้คะแนนเป็น 1 เป็นต้น จากนั้นจึงจะทำการวิเคราะห์โดยใช้โมเดลการตอบข้อสอบที่ให้คะแนนแบบสองค่า ซึ่งทำให้สูญเสียสารสนเทศเกี่ยวกับแบบแผนการตอบส่วนหนึ่งไป ด้วยเหตุนี้ทำให้นักการศึกษาพยายามคิดค้นและพัฒนาโมเดลการตอบข้อสอบที่ให้คะแนนแบบหลายค่า (Polytomous IRT model) เพื่อแก้ปัญหาความไม่เหมาะสมระหว่างข้อมูลกับโมเดลการวิเคราะห์ โมเดลที่สำคัญได้แก่ โมเดลเกรดเรสพอนส์ (Graded Response Model: GRM) พัฒนาขึ้นในปี 1960 โดยซาเมจิมา (Sameijima) โมเดลเกรดเรสพอนส์แบบปรับขยาย (Modified Graded Response Model: M-GRM) พัฒนาขึ้นในปี 1960 โดยมูราคิ (Muraki.1990) โมเดลการตอบแบบโนมินอล (Nominal Response Model: NRM) พัฒนาโดยบ็อค (Bock. 1972) โมเดลมาตรประมาณค่า (Rating Scale Model: RSM) พัฒนาโดยแอนดริช (Andrich. 1978a, 1978b) โมเดลพาร์เชียลเครดิต (Partial Credit Model: PCM) พัฒนาโดยมาสเตอร์ (Master. 1982) และ โมเดลพาร์เชียลเครดิตทั่วไป (Generalized Partial Credit Model: GPCM) พัฒนาโดยมูราคิ (Muraki.1992)

โมเดลพาร์เชียลเครดิตทั่วไป (Generalized Partial Credit Model; GPCM)

มูรากิ (Muraki.1992) ได้พัฒนาโมเดลนี้จากโมเดลโมเดลพาร์เชียลเครดิต (Partial Credit Model: PCM) โดยมีพื้นฐานข้อตกลงเบื้องต้นที่ว่าในการตอบข้อสอบที่มีคำตอบหลายรายการ (Multicategory) ความน่าจะเป็นที่ผู้ตอบเลือกรายการที่ k มากกว่ารายการที่ k-1 จะอยู่ในกรอบของโมเดลการตอบที่ให้คะแนนแบบสองค่า ใน GPCM สำหรับแต่ละรายการที่อยู่ติดกัน (adjacent categories) ความน่าจะเป็นของการตอบรายการที่ k มากกว่ารายการที่ k-1 กำหนดโดยความน่าจะเป็นแบบมีเงื่อนไข (Conditional probability) ดังนี้

$$P_{jk|k-1,k}(\theta) = \frac{P_{jk}(\theta)}{P_{j,k-1}(\theta)+P_{jk}(\theta)} = \frac{\exp[Da_j(\theta-b_{jk})]}{1+\exp[Da_j(\theta-b_{jk})]} \quad (4)$$

เมื่อ $k = 2,3,...,m_j$

โดยที่ P_{jk} แทน ความน่าจะเป็นของการตอบรายการที่ k จากรายการที่เป็นไปได้ m_j ของข้อสอบ j

ถ้าให้ $C_{jk}n = P_{jk|k-1,k}(\theta)$ (5)

จะได้ $P_{jk}(\theta) = \frac{C_{jk}}{1-C_{jk}} P_{j,k-1}(\theta)$ (6)

มีข้อสังเกตว่า $\frac{C_{jk}}{1-C_{jk}}$ เป็นอัตราส่วนของความน่าจะเป็นตามเงื่อนไขสองความน่าจะเป็น สามารถแสดงในรูป $\exp[a_i(\theta - b_i)]$ และเรียกสมการ(6) ว่า ฟังก์ชันคุณลักษณะปฏิบัติการ (operating characteristic function; OCF)

ถ้า $P_{j1}(\theta) = \frac{1}{G}$ (7)

เมื่อ G เป็นองค์ประกอบที่ทำให้เป็นปกติ (normalizing factor) ความน่าจะเป็นต่อไปนี้ได้มาโดยการประยุกต์ใช้ OCF ในสมการ (6)

$$P_{j2}(\theta) = \frac{\exp[Da_j(\theta-b_{j2})}{G} \quad (8)$$

.

.

.

$$P_{jg}(\theta) = \frac{\exp[\sum_{v=2}^{g} Da_j(\theta-b_{jv})}{G} \quad (9)$$

.

.

.

และ

$$P_{jm_j}(\theta) = \frac{\exp[\sum_{v=2}^{m_j} Da_j(\theta - b_{jv})}{G} \quad (10)$$

เมื่อ g เป็นSubscript สำหรับการตอบรายการเฉพาะ k = g เพราะว่า

$$\sum_{k=1}^{m_j} P_{jk}(\theta) = 1 \quad (11)$$

และ

$$G = 1 + \sum_{c=2}^{m_j} \exp[\sum_{v=2}^{c} Da_j(\theta - b_{jv})] \quad (12)$$

นั่นคือ

$$P_{jk}(\theta) = \frac{\exp[\sum_{v=1}^{k} Da_j(\theta - b_{jv})]}{\sum_{c=1}^{m_j} \exp[\sum_{v=1}^{c} Da_j(\theta - b_{jv})]} \quad (13)$$

เมื่อ D แทน ค่าปรับมาตรมีค่าเท่ากับ 1.7

a_j แทน พารามิเตอร์อำนาจจำแนกหรือพารามิเตอร์ความชัน

b_{jv} แทน พารามิเตอร์ระดับความยากของขั้นการตอบที่ v (Item step difficulty)

นั่นคือ $P_{jk}(\theta) = \frac{\exp[\sum_{v=1}^{k} Da_j(\theta - b_j + d_v)]}{\sum_{c=1}^{m_j} \exp[\sum_{v=1}^{c} Da_j(\theta - b_j + d_v)]}$

โดยที่ $b_{jv} = b_j - d_v$

เมื่อ b_j แทน พารามิเตอร์ตำแหน่งข้อสอบ (Item location parameter)

d_v แทน ความยากสัมพัทธ์ของรายการคะแนน v

มาสเตอร์ (Master. 1982) เรียกพารามิเตอร์ลำดับขั้นข้อสอบ (Item step parameter); b_{jk} ว่าเป็นเหมือนพารามิเตอร์ความยากของข้อสอบ (Item difficulty parameter) ในแต่ละลำดับขั้น ซึ่งเป็นจุดบนสเกลความสามารถ θ ณ ที่ $P_{j,k-1}(\theta)$ กับ $P_{jk}(\theta)$ ตัดกัน ซึ่งเส้นโค้งทั้ง 2 เส้นถือเป็นฟังก์ชันการตอบรายการข้อสอบ (Item category response function: ICRFs) จะตัดกันเพียงจุดเดียว บนค่าใด ๆ ตลอดช่วงสเกล θ นั่นคือ

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

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

ถ้า $\theta < b_{jk}$, $P_{jk}(\theta) < P_{j,k-1}(\theta)$;

ภายใต้ข้อตกลงที่ว่า $a_j > 0$ และข้อสอบข้อที่ j นั้น b_{jk} ไม่จำเป็นต้องจัดเรียงลำดับ เพราะค่าพารามิเตอร์จะแสดงถึงขนาดที่สัมพันธ์กันของความน่าจะเป็นในการตอบ $P_{j,k-1}(\theta)$ และ $P_{jk}(\theta)$ ที่อยู่ลำดับติดกัน ส่วนพารามิเตอร์ a_j เป็นพารามิเตอร์ความชันของข้อสอบข้อที่ j โดยปกติค่า a_j มีพิสัยที่เป็นไปได้ในช่วง 0 ถึง ∞ ซึ่งค่าพารามิเตอร์ความชันจะบ่งชี้ระดับ (Degree) ที่การตอบแต่ละรายการแปรเปลี่ยนไปในแต่ละข้อเช่นเดียวกับระดับความสามารถ θ ที่แปรเปลี่ยนไป

2. โมเดลทางสถิติการทำหน้าที่เบี่ยงเบนของข้อสอบ

การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบสองค่าหรือหลายค่า ไม่ว่าจะใช้ทฤษฎีการทดสอบมาตรฐานเดิมหรือทฤษฎีการตอบข้อสอบซึ่งอยู่บนเงื่อนไขของการวัดเพียงมิติเดียว แบ่งออกได้เป็น 2 โมเดลดังนี้

2.1 โมเดลการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบสองค่า (Dichotomous DIF modeling)

โดยปกติแล้วในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบจะใช้การเปรียบเทียบผลการ ตอบข้อสอบของผู้สอบทั้งสองกลุ่ม กลุ่มที่สงสัยว่าจะเสียเปรียบในการตอบข้อสอบก็จะเป็นกลุ่มรอง (Minority group) ซึ่งเรียกว่า **“กลุ่มสนใจ”** (Focal group) ส่วนอีกกลุ่มที่มีผลการปฏิบัติตรงข้ามกับกลุ่มสนใจ ซึ่งได้เปรียบในการตอบข้อสอบจะเป็นกลุ่มหลัก (Majority group) ซึ่งเรียกว่า **“กลุ่มอ้างอิง”** (Reference group) สำหรับข้อสอบที่ใช้ภายใต้การตรวจสอบการทำหน้าที่เบี่ยงเบนจะเรียกว่า **“ข้อสอบที่ศึกษา”** (Studied item) วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบอาจใช้คะแนนสังเกต (Observed score) หรือไม่ก็ใช้ค่าประมาณตัวแปรแฝง (Latent variable) หรือประมาณค่าคะแนนจริง (Ture score) เป็นตัวแปรตามเงื่อนไขในการสรุปว่าข้อสอบที่ให้คะแนนแบบสองค่าว่าทำหน้าที่เบี่ยงเบนหรือไม่อาศัยข้อตกลงดังนี้

2.1.1 การไม่ปรากฏการทำหน้าที่เบี่ยงเบนของข้อสอบเมื่อใช้ตัวแปรแฝง (Latent-variable null DIF)

กำหนดให้ Y แทนคะแนน (Score) ของข้อสอบที่ต้องการศึกษา (Studied item) θ แทนตัวแปรแฝงที่ใช้ในการจับคู่ (Latent matching variable) จึงได้ว่า $E_R[Y|\theta]$ และ $E_F[Y|\theta]$ แทนการถดถอยของ Y บนตัวแปรแฝง θ ของกลุ่มอ้างอิงและกลุ่มสนใจตามลำดับ ข้อสอบหนึ่งข้อจะไม่แสดงการทำหน้าที่เบี่ยงเบน (Not exhibit DIF) ถ้าทุกค่าของ θ ที่ทำให้

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

ข้อสอบหนึ่งข้อแสดงการทำหน้าที่เบี่ยงเบน ถ้าการถดถอยของ Y บนตัวแปรแฝง θ ไม่ได้เหมือนกัน (Not identical) สำหรับกลุ่มต่าง ๆ ภายใต้การศึกษานั่นคือ มีบางค่าของ θ ที่ทำให้ $E_R[Y|\theta] \neq E_F[Y|\theta]$

2.1.2 การไม่ปรากฏการทำหน้าที่เบี่ยงเบนของข้อสอบเมื่อใช้คะแนนสังเกต (Observed -score null DIF)

กำหนดให้ Y แทนคะแนนของข้อสอบที่ต้องการศึกษา X แทนคะแนนสังเกตของแบบทดสอบที่ใช้ในการจับคู่ (Observed matching test score) ซึ่งรวมคะแนนของข้อสอบที่ต้องการศึกษาด้วย จึงได้ว่า $E_R[Y|X]$ และ $E_F[Y|X]$ แทนการถดถอยของ Y บนคะแนนสังเกต X ของกลุ่มอ้างอิงและกลุ่มสนใจตามลำดับ ข้อสอบหนึ่งข้อจะไม่แสดงการทำหน้าที่เบี่ยงเบน ถ้าทุกค่าของ X ที่ทำให้

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

ข้อสอบหนึ่งข้อแสดงการทำหน้าที่เบี่ยงเบน ถ้าคะแนนที่คาดหวัง (Expected scores) ไม่เหมือนกันสำหรับกลุ่มอ้างอิงและกลุ่มสนใจที่จับคู่บนคะแนน X นั่นคือ มีบางค่าของ X ที่ทำให้ $E_R[Y|X] \neq E_F[Y|X]$

ฮอลแลนด์และทาเยอร์ (Holland; & Thayer.1988) ได้แสดงให้เห็นว่าภายใต้เงื่อนไขโมเดลราส์ช (Rasch model) สมการ (1) และสมการ (2) เท่าเทียมกัน ในกรณีเช่นนี้การทดสอบสมมติฐานศูนย์ก็จะเป็นแบบเดียวกัน

2.1.3 การไม่ปรากฏการทำหน้าที่เบี่ยงเบนของข้อสอบเมื่อใช้ตัวแปรแฝงตามทฤษฎีการทดสอบมาตรฐานเดิม (Classical test theory latent-variable null DIF)

ตามทฤษฎีการทดสอบมาตรฐานเดิม ระบุว่า $X = T + E$ เมื่อ X แทนคะแนนสังเกต (Observed score) T แทนคะแนนจริง (True score) และ E แทนคะแนนความคลาดเคลื่อนในการวัด (Error score) ดังนั้น $E = X - T$ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบให้ถือว่าคะแนนความคลาดเคลื่อนในการวัดของกลุ่มอ้างอิงและกลุ่มสนใจมีค่าเท่ากับศูนย์พร้อมทั้งกำหนดให้ $E_R[Y|t]$ และ $E_F[Y|t]$ แทนการถดถอยของ Y บนคะแนนจริง t จากแบบทดสอบที่ใช้ในการจับคู่ของกลุ่มอ้างอิงและกลุ่มสนใจตามลำดับ ข้อสอบจะไม่แสดงการทำหน้าที่เบี่ยงเบน ถ้าทุกค่าของ t ที่ทำให้

$$E_R[Y|t] = E_F[Y|t] \quad (3)$$

ข้อสอบจะแสดงการทำหน้าที่เบี่ยงเบน ถ้าการถดถอยของ Y บนคะแนนจริง t จากแบบทดสอบที่ใช้ในการจับคู่ของกลุ่มอ้างอิงและกลุ่มสนใจไม่เหมือนกัน นั่นคือจะมีบางค่าของ t ที่ทำให้ $E_R[Y|t] \neq E_F[Y|t]$

2.2 โมเดลการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่า (Polytomous DIF modeling)

ชางและคณะ (Chang; Mazzeo; & Roussos. 1996) ได้ขยายวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่า ไว้ดังนี้

เมื่อให้ Y แทนคะแนนของข้อสอบที่ต้องการศึกษาโดยเป็นคะแนนของรายการแบบจัดอันดับ (Ordered categories) ซึ่งมี $m+1$ รายการ (ตัวอย่างเช่น $Y = k, 0 \le k \le m$)และ$P_{k,g}(\theta)$ แทนฟังก์ชันการตอบรายการข้อสอบ (Item category response function: ICRF) ที่ระดับคะแนน k ในกลุ่ม g คือเป็นความน่าจะเป็นของการได้คะแนน k สำหรับผู้สอบที่มีความสามารถ θ ซึ่งเลือกมาอย่างสุ่มจากกลุ่ม g ($g = R$ สำหรับกลุ่มอ้างอิง และ $g = F$ สำหรับกลุ่มสนใจ)นั่นเอง การถดถอยของคะแนนข้อสอบบนความสามารถθ จะกำหนดในรูปผลรวมของฟังก์ชันการตอบรายการข้อสอบแบบถ่วงน้ำหนักดังนี้

$$E_g\left[Y|\theta\right] = \sum_{k=1}^{m} kP_{k,g}(\theta) \qquad (4)$$

จากสมการ (4) ถ้า $m = 1$ จะเป็นการแสดงการถดถอยของ Y บนตัวแปรแฝงθ ของกลุ่ม g (R หรือ F) ที่มีการให้คะแนนข้อสอบแบบสองค่า ซึ่งการถดถอยนี้จะเป็นฟังก์ชันการตอบข้อสอบ (Item response function: IRF) ดังนั้นจึงถือได้ว่าข้อสอบที่มีการให้คะแนนแบบสองค่าเป็นกรณีเฉพาะ งข้อสอบที่มีการให้คะแนนแบบหลายค่า ที่มี $m = 1$

โมเดลการตอบข้อสอบที่ให้คะแนนแบบสองค่าโครงสร้างของข้อสอบได้จากฟังก์ชันการตอบข้อสอบ(IRF)ส่วนโมเดลการตอบข้อสอบที่ให้คะแนนแบบหลายค่าโครงสร้างของข้อสอบหนึ่งข้อโดยทั่วไปแล้วได้จากฟังก์ชันการตอบรายการข้อสอบ(ICRF)ทั้งหมดจำนวน m ฟังก์ชัน (all mICRFs) ความสอดคล้องที่เด่นชัดระหว่างฟังก์ชันการตอบข้อสอบ (IRF) และเซตของฟังก์ชันการตอบรายการข้อสอบ (Set of ICRFs) อาจจะปรากฏหรือไม่ก็ได้ ถ้าสมการ (1) ไม่ได้มีความหมายในรูปทั่วไป ดังนั้นการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่าตามทฤษฎีการตอบข้อสอบจะเปรียบเทียบความแตกต่างของ ICRFs ดังนี้

$$P_{kR}(\theta) = P_{kF}(\theta) \quad , \quad k = 1,2,...,m \qquad (5)$$

สำหรับทุกค่าθ เมื่อ $P_{kR}(\theta)$ และ $P_{kF}(\theta)$ แทน ICRFs ที่ระดับคะแนน k ของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจตามลำดับ ดังนั้นจึงนิยมให้สมการ (5) เป็นข้อตกลงของ **“การไม่ปรากฏการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่า เมื่อใช้ตัวแปรแฝง”**

ชางและมาซซีโอ (Chang; & Mazzeo.1994) ได้พิสูจน์ว่าสมการ (1) มีความหมายเหมือนกับ สมการ (5) เมื่อใช้กับโมเดลการตอบข้อสอบที่มีการให้คะแนนแบบหลายค่า เช่นโมเดลพาร์เชียลเครดิต (Partial credit model: PCM) (Master. 1982) โมเดลพาร์เชียลเครดิตทั่วไป (Generalized partial credit model: GPCM) (Muraki. 1992) และโมเดลเกรดเรสพอนส์ (Graded response model: GRM) (Samejima. 1969) เป็นต้น การทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้

คะแนนแบบหลายค่าจะประเมินผลเพียงความสอดคล้องกันของความแตกต่างในฟังก์ชันการตอบข้อสอบ(IRFs) ตลอด ความสามารถกลุ่มต่าง ๆ โดยไม่ต้องพิจารณาถึงความแตกต่างในฟังก์ชันการตอบรายการข้อสอบ (ICRFs) เพราะไม่มีสารสนเทศใดหายไปโดยการเปรียบเทียบเพียงฟังก์ชันการตอบข้อสอบ ดังนั้นการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่มีการให้คะแนนแบบหลายค่าตามทฤษฎีการตอบข้อสอบจะอาศัยข้อตกลง การไม่ปรากฏการทำหน้าที่เบี่ยงเบนของข้อสอบที่ทำให้คะแนนแบบสองค่าเมื่อใช้ตัวแปรแฝง จากสมการ (1) โดยตรง กล่าวคือ **"ข้อสอบหนึ่งข้อจะไม่แสดงการทำหน้าที่เบี่ยงเบน (Not exhibit DIF) ถ้าการถดถอยของคะแนนของข้อสอบที่ให้คะแนนแบบหลายค่าบนตัวแปรแฝง (ซึ่งขณะนี้คือผลรวมของ ICRFs แบบถ่วงน้ำหนัก) ของกลุ่มอ้างอิงและกลุ่มสนใจเหมือนกัน"** นั่นคือ $E_R[Y|\theta] = E_F[Y|\theta]$

สำหรับการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่า ซึ่งใช้คะแนนสังเกต (Observed score) ตามทฤษฎีการทดสอบมาตรฐานเดิม จะยังคงอาศัยข้อตกลง การไม่ปรากฏการทำหน้าที่เบี่ยงเบนของข้อสอบเมื่อใช้คะแนนสังเกตจากสมการ(2) โดยตรง กล่าวคือ **"ข้อสอบไม่แสดงการทำหน้าที่เบี่ยงเบน (Not exhibit DIF) ถ้าการถดถอยของคะแนนของข้อสอบที่ให้คะแนนแบบหลายค่า บนคะแนนสังเกตของแบบทดสอบที่ใช้ในการจับคู่เหมือนกันสำหรับกลุ่มผู้สอบภายใต้การศึกษา"** ในขณะที่การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่าซึ่งใช้คะแนนจริง (True score) ตามทฤษฎีการทดสอบมาตรฐานเดิม จะยังคงอาศัยข้อตกลงการไม่ปรากฏการทำหน้าที่เบี่ยงเบนของข้อสอบ เมื่อใช้ตัวแปรแฝงตามทฤษฎีการวัดมาตรฐานเดิม จากสมการ (3) โดยตรง กล่าวคือ **"ข้อสอบไม่แสดงการทำหน้าที่เบี่ยงเบน (Not exhibit DIF) ถ้าการถดถอยของคะแนนของข้อสอบที่ให้คะแนนแบบหลายค่าบนคะแนนจริงของแบบทดสอบที่ใช้ในการจับคู่เหมือนกันสำหรับกลุ่มผู้สอบภายใต้การศึกษา"**

จากที่กล่าวมาจะเห็นว่าการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่าเมื่อใช้คะแนนสังเกต หรือคะแนนจริงตามทฤษฎีการทดสอบมาตรฐานเดิมเป็นเกณฑ์ในการจับคู่กลุ่มอ้างอิงและกลุ่มสนใจจะยังคงอาศัยโมเดลการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบสองค่า เมื่อใช้คะแนนสังเกต หรือคะแนนจริงเป็นเกณฑ์การจับคู่กลุ่มผู้สอบในการตรวจสอบ ที่เป็นเช่นนี้เนื่องจากความไม่แปรเปลี่ยนทางโครงสร้างของข้อสอบ (Item structure invariance)

3. วิธีตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ

ในการศึกษาครั้งนี้ผู้วิจัยนำวิธีโพลีโตมัสซิบเทสท์(Polytomous SIBTEST: Poly-SIB) (Chang; Mazzeo; & Roussos. 1996)วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก(Logistic discriminant function analysis: LDFA) (Miller; & Spray. 1993) และวิธีการถดถอยโลจิสติกแบบจัดอันดับ (Ordinal logistic regression: OLR) (Zumbo. 1999) ไปตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่า ดังนั้นผู้วิจัยจะนำเสนอรายละเอียดของวิธีการตรวจสอบทั้ง

สามวิธีดังกล่าว โดยนำเสนอตั้งแต่วิธีที่พัฒนาเริ่มต้นคือวิธีซิบเทสท์ (SIBTEST) (Shealy; & Stout. 1993) และวิธีการถดถอยโลจิสติก (Logistic regression) (Swaminathan; & Rogers. 1990) ซึ่งใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในข้อมูลที่ให้คะแนนแบบสองค่า รวมวิธีการตรวจสอบทั้งหมดที่จะนำเสนอ 5 วิธี มีรายละเอียดดังนี้

3.1.วิธีซิบเทสท์ (SIBTEST)

เชียลีและสเตาท์ (Shealy; & Stout.1993) ได้เสนอวิธีซิบเทสท์ (Simultaneous item bias test: SIBTEST) เพื่อใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (Differential item functioning: DIF)และการทำหน้าที่เบี่ยงเบนของแบบทดสอบ (Differential test functioning: DTF) ในข้อสอบที่ให้คะแนนแบบสองค่า (Dichotomous) วิธีนี้สามารถใช้ตรวจสอบได้ทั้งแบบทดสอบที่วัดเพียงมิติเดียว (Unidimensional test) และแบบทดสอบที่วัดหลายมิติ (Multidimensional test) (Stout; Li; & Nandakumar. 1977) ในการศึกษาการทำหน้าที่เบี่ยงเบนของข้อสอบในแบบทดสอบที่วัดเพียงมิติเดียวด้วยวิธีซิบเทสท์ต้องอาศัยข้อตกลงที่ว่าข้อสอบในแบบทดสอบจะต้องมุ่งวัดคุณลักษณะหรือความสามารถแฝงเพียงลักษณะเดียวที่เรียกว่าความสามารถเป้าหมาย (Target ability: θ) ส่วนความสามารถแทรกซ้อน (Nuisance ability: η) ซึ่งเป็นความสามารถแฝงที่มีอิทธิพลต่อผลการตอบข้อสอบไม่ต้องการวัด หลักการตรวจสอบด้วยวิธีซิบเทสท์มีแนวคิดคล้ายกับวิธีการทำให้เป็นมาตรฐาน (Standardization: STND) (Dorans; & Kulick. 1986) โดยพัฒนามาจากโมเดลความลำเอียงของแบบทดสอบภายใต้ทฤษฎีการตอบข้อสอบแบบหลายมิติ (Multidimensional IRT) มีรูปแบบนอนพาราเมตริก (Nonparametric form) แต่ไม่ต้องใช้ฟังก์ชันการตอบข้อสอบประมาณค่าความสามารถ วิธีนี้จึงถูกจัดว่าเป็นวิธีนอนพาราเมตริกที่ใช้ตัวแปรแฝง (Latent-variable-nonparametric approach) (Potenza; & Dorans. 1995; Chang; Mazzeo; & Roussos. 1996: 334) วิธีซิบเทสท์ ได้รับการออกแบบมา เพื่อใช้สำหรับตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกัน (Uniform DIF) และการทำหน้าที่เบี่ยงเบนของข้อสอบแบบมีทิศทางเดียวกัน (Unidirectional DIF) ดังนั้นจึงมีข้อจำกัดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบแบบไม่มีทิศทางเดียวกัน (Nondirectional DIF) (Shealy; & Stout.1993:162; Li; & Stout. 1996: 650) ส่วนข้อได้เปรียบของวิธีซิบเทสท์คือสามารถคำนวณได้ง่ายไม่ซับซ้อน เสียค่าใช้จ่ายไม่มาก และไม่จำเป็นต้องใช้ตัวอย่างขนาดใหญ่ ทั้งยังใช้สถิติทดสอบนัยสำคัญ เพื่อตัดสินการทำหน้าที่เบี่ยงเบนของข้อสอบครั้งละหนึ่งข้อหรือมากกว่าหนึ่งข้อพร้อมกัน (Narayanan; & Swaminathan. 1996: 259; Nandakumar.1993: 294) นอกจากนี้ยังสามารถแสดงออกมาให้เห็นถึงการทำหน้าที่เบี่ยงเบนของแบบทดสอบ และผลกระทบที่เป็นความแตกต่างความสามารถระหว่างกลุ่มในความสามารถที่ตั้งใจว่าจะวัด รวมทั้งกลไกทางจิตวิทยาเกี่ยวกับสติปัญญาที่ใช้วิธี DIF ซึ่งสามารถแสดงให้เห็นได้อย่างชัดแจ้งโดยผ่านทางความสามารถเป้าหมาย (Target ability) ว่าต่างกับปัจจัยแทรกซ้อนต่างๆ (Nuisance determinants) (Nandakumar. 1993:294)

แนวคิดและหลักการ

ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบตามวิธีซิบเทสท์ เชียลี และสเตาท์ (Shealy; & Stout.1993:162-163) จะแบ่งกลุ่มผู้สอบออกเป็นสองกลุ่ม คือ กลุ่มอ้างอิง(R) และกลุ่มสนใจ (F) สุ่มตัวอย่างในแต่ละกลุ่มเพื่อทำแบบทดสอบที่มีจำนวน N ข้อ ในแบบทดสอบดังกล่าวจะมีข้อสอบส่วนหนึ่งที่ถูกตั้งข้อสงสัยว่าทำหน้าที่เบี่ยงเบน โดยเข้าข้างผู้สอบกลุ่มอ้างอิง ข้อสอบเช่นนี้เป็นเป้าหมายเพื่อศึกษาการทำหน้าที่เบี่ยงเบนของข้อสอบ ส่วนผลการตอบข้อสอบแทนด้วย $U=(U_1,U_2,...,U_N)$เมื่อ U_i แทนผลการตอบข้อสอบข้อที่ i ซึ่งให้คะแนนแบบสองค่า จะได้ว่า $U_i=1$ ถ้าตอบข้อสอบที่ i ถูกต้องและ $U_i=0$ ถ้าตอบข้อสอบข้อที่ i ผิด แบบแผน U ตามโมเดล IRT โดยทั่วไปจะต้องประกอบด้วย 2 องค์ประกอบคือ (1) พารามิเตอร์ความสามารถของผู้สอบจำนวน d มิติ และ (2) เซตของฟังก์ชันการตอบข้อสอบ (IRFs) ซึ่งฟังก์ชันการตอบข้อสอบในแต่ละข้อจะเป็นตัวกำหนดความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง ณ ที่นี้ $d=2$ เพราะมีพารามิเตอร์ความสามารถสองชนิดคือความสามารถเป้าหมายที่ต้องการวัด (θ) และความสามารถแทรกซ้อนที่ไม่ต้องการวัด (η) ส่วนเวกเตอร์ความสามารถของผู้สอบหนึ่งคนจากกลุ่มใดกลุ่มหนึ่งจะเป็น (θ,η) โดยที่ θ และ ηเป็นการวัดมิติเดียว (Unidimensional) และการแจกแจงของ (θ,η) เลือกมาจากกลุ่มผู้สอบรวมทั้งหมดโดยการสุ่ม แทนด้วย (Θ,η) ความสามารถทั้งสองมีข้อตกลงว่าข้อสอบส่วนมากของแบบทดสอบจะวัดความสามารถเป้าหมาย (θ) แต่อาจมีข้อสอบบางข้อที่ทำหน้าที่เบี่ยงเบนจะวัดทั้งความสามารถเป้าหมาย (θ) และความสามารถแทรกซ้อน (η) ซึ่งความสามารถแทรกซ้อนอาจมีเพียงความสามารถเดียวหรือมากกว่าหนึ่งความสามารถก็ได้ (Shealy; & Stout.1993: 162-163) ในทฤษฎีการตอบข้อสอบ (IRT) การตอบข้อสอบ i ที่ใช้ความสามารถเป้าหมายθ เพียงความสามารถเดียวจะแทนด้วยฟังก์ชันการตอบข้อสอบ $P_i(\theta)$ สามารถแสดงฟังก์ชันดังกล่าวโดยใช้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ได้ดังนี้

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

ส่วนการตอบข้อสอบ i ที่ต้องใช้ความสามารถเป้าหมายและความสามารถแทรกซ้อน η ซึ่งเป็นข้อสอบที่วัดความสามารถสองมิติ จะแทนด้วยฟังก์ชันการตอบข้อสอบ $P_i(\theta,\eta)$ สามารถแสดงฟังก์ชันดังกล่าวโดยใช้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ได้ดังนี้

$$P_i(\theta,\eta)=c_i+\frac{(1-c_i)}{1+\exp[-1.7\{a_{i\theta}(\theta-b_{i\theta})+a_{i\eta}(\eta-b_{i\eta})\}]} \quad , i=n+1,...,N \qquad (2)$$

ฟังก์ชันการตอบข้อสอบในสมการ(1)และ(2) เป็นโมเดลของทฤษฎีการตอบข้อสอบ ซึ่งมีข้อตกลงสามประการคือ **ประการแรก** ค่าของฟังก์ชันการตอบข้อสอบไม่แปรเปลี่ยนตามกลุ่มผู้สอบ (Group invariance)ซึ่งค่า $P_i(\theta)$ จะเหมือนกันสำหรับผู้สอบที่มีความสามารถθและค่า$P_i(\theta,\eta)$ จะเหมือนกันสำหรับผู้สอบที่มีความสามารถ (θ,η) โดยไม่คำนึงว่าเป็นสมาชิกผู้สอบกลุ่มใด **ประการที่สอง** การเพิ่มขึ้นของฟังก์ชันการตอบข้อสอบเป็นไปทิศทางเดียวไม่เปลี่ยนแปลง(Monotonicity)

กล่าวคือ ข้อสอบที่ใช้ความสามารถ θ เพียงความสามารถเดียว ค่า $P_i(\theta)$ จะเพิ่มขึ้นตามความสามารถ θ ส่วนข้อสอบที่ต้องใช้ความสามารถθ และ η ค่า $P_i(\theta,\eta)$ จะเพิ่มขึ้นตาม θ และ η และ **ประการที่สาม** ความเป็นอิสระในการตอบข้อสอบ (Local independence) ของU เพื่อก่อให้เกิด $P_i(\theta,\eta)$ กำหนดในรูปของความน่าจะเป็นดังนี้

$$P\left[U=(u_1,u_2,...,u_N)\middle|(\Theta=\theta,\eta=\eta)\right]=\prod_{i=1}^{N}P_i(\theta,\eta)^{u_i}\left(1-P_i(\theta,\eta)\right)^{1-u_i} \quad (3)$$

เชียลีและสเตาท์ (Shealy; & Stout. 1993:163-164; Li; & Stout. 1996: 651-652) ได้ใช้โมเดลทางคณิตศาสตร์ภายใต้ทฤษฎีการตอบข้อสอบที่วัดความสามารถหลายมิติ (Multidimensional IRT) อธิบายการทำหน้าที่เบี่ยงเบนของข้อสอบโดยใช้ฟังก์ชันการตอบข้อสอบแบบมาร์จินอล (Marginal IRFs) ของความสามารถเป้าหมายที่ต้องการวัดθ สำหรับกลุ่ม g (กลุ่มอ้างอิงหรือกลุ่มสนใจ) ดังนี้

$$M_{ig}(\theta)=E\left[P_i(\Theta,\eta)\middle|\Theta=\theta,G=g\right] \quad (4)$$

ถ้า $\eta|\Theta=\theta,G=g$ มีความหนาแน่นแบบมีเงื่อนไข เมื่อกำหนดความสามารถθ ของกลุ่ม g มีค่าคงที่ ดังนั้นความหนาแน่นแบบมีเงื่อนไขของ $\eta_g|\theta$ คือ $f_g(\eta|\theta)$ ซึ่งเป็นการแจกแจงแบบมีเงื่อนไขของกลุ่มผู้สอบสามารถเขียน (4) ใหม่ได้ดังนี้

$$M_{ig}(\theta)=\int_{-\alpha}^{\alpha}P_i(\theta,\eta)f_g(\eta|\theta)d\eta \quad (5)$$

จากสมการ (5) ถ้าการแจกแจงแบบมีเงื่อนไขของความสามารถ η มีค่าเท่ากันสำหรับผู้สอบสองกลุ่ม แล้วข้อสอบจะทำหน้าที่ไม่เบี่ยงเบน (No-DIF) เพราะว่ามีความสามารถθ เท่ากันจึงทำให้ความน่าจะเป็นในการตอบข้อสอบถูกเท่ากัน (Ackerman. 1992)

เราทราบมาแล้วว่าการทำหน้าที่เบี่ยงเบนของข้อสอบแบบมีทิศทางเดียวกัน(Unidirectional DIF) เกิดขึ้นเมื่อความน่าจะเป็นของการตอบข้อสอบถูกจากผู้สอบกลุ่มหนึ่งมากกว่าผู้สอบอีกกลุ่มหนึ่งตลอดช่วงความสามารถตามทฤษฎีการตอบข้อสอบสามารถแสดงได้ในรูปโค้งลักษณะข้อสอบของผู้สอบสองกลุ่มไม่ตัดกัน (Non-crossing ICCs) ข้อสอบใดเข้าข้างผู้สอบกลุ่มใดโดยพิจารณาเปรียบเทียบจากค่าฟังก์ชันการตอบข้อสอบแบบมาร์จินอล ระหว่างกลุ่มอ้างอิง(R) กับกลุ่มสนใจ (F)กล่าวคือ ถ้า $M_{iF}(\theta)<M_{iR}(\theta)$ ทุกค่าความสามารถ θ แสดงว่าข้อสอบข้อนั้นทำหน้าที่เบี่ยงเบนแบบมีทิศทางเดียวกัน โดยข้อสอบจะเข้าข้างผู้สอบกลุ่มอ้างอิงและถ้า $M_{iF}(\theta)>M_{iR}(\theta)$ ทุกค่าความสามารถ θ แสดงว่าข้อสอบข้อนั้นทำหน้าที่เบี่ยงเบนแบบมีทิศทางเดียวกัน โดยข้อสอบจะเข้าข้างผู้สอบกลุ่มสนใจ

กระบวนการตรวจสอบ

ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่มีทิศทางเดียวกันตามวิธีซิบเทสท์ (Shealy; & Stout. 1993) จะเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มสนใจ โดยใช้แบบทดสอบที่มีจำนวน N ข้อ แล้วแบ่งแบบทดสอบดังกล่าวออกเป็นสองชุดย่อย (Subtests) คือ แบบทดสอบชุดย่อยที่มีความเที่ยงตรง (Valid subtests) และแบบทดสอบชุดย่อยที่ต้องการศึกษา (Studied subtests)แบบทดสอบชุดแรกใช้ในการจับคู่เปรียบเทียบ (Matching subtests) ประกอบด้วยข้อสอบข้อที่ 1 ถึง n ซึ่งเป็นข้อสอบที่ไม่สงสัยว่าทำหน้าที่เบี่ยงเบนโดยวัดความสามารถเป้าหมาย θ เพียงอย่างเดียว ส่วนแบบทดสอบชุดหลัง เป็นส่วนที่เหลือจากชุดแรก ประกอบด้วยข้อสอบข้อที่ $n+1$ ถึง N ข้อสอบดังกล่าวสงสัยว่าทำหน้าที่เบี่ยงเบนโดยวัดทั้งความสามารถเป้าหมาย θ และความสามารถแทรกซ้อน η

ฟังก์ชันการตอบข้อสอบของแบบทดสอบที่ต้องการศึกษา (Studied subtest response function) สำหรับผู้สอบกลุ่มอ้างอิง และกลุ่มสนใจ กำหนดดังนี้

$$M_{SR}(\theta) = \sum_{i=n+1}^{N} M_{iR}(\theta) \quad (6)$$

$$M_{SF}(\theta) = \sum_{i=n+1}^{N} M_{iF}(\theta) \quad (7)$$

เมื่อ $M_{SR}(\theta)$และ $M_{SF}(\theta)$ แทนผลรวมฟังก์ชันการตอบข้อสอบแบบมาร์จินอล (Marginal IRFs) ของข้อสอบที่ต้องการศึกษา ณ ระดับความสามารถθ จากผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจตามลำดับ สำหรับปริมาณของการทำหน้าที่เบี่ยงเบนของข้อสอบ (Amount of DIF) สามารถคำนวณ จากความแตกต่างระหว่าง $M_{SR}(\theta)$ และ $M_{SF}(\theta)$ ดังนี้

$$B(\theta) = M_{SR}(\theta) - M_{SF}(\theta) \quad (8)$$

ขนาดของความแตกต่างดังกล่าวแสดงถึงปริมาณของการทำหน้าที่เบี่ยงเบนของข้อสอบแบบมีทิศทางเดียวกัน (การทำหน้าที่เบี่ยงเบนของข้อสอบแบบไม่ตัดกัน) จากแบบทดสอบชุดย่อยที่ต้องการศึกษาที่เข้าข้างกลุ่มอ้างอิง ณ ระดับความสามารถθ เชียลีและสเตาท์ (Shealy; & Stout. 1993:167; Li; & Stout.1996: 652) ได้คำนวณปริมาณเฉลี่ยของการทำหน้าที่เบี่ยงเบนของข้อสอบแบบมีทิศทางเดียวกัน ดังนี้

$$\beta_{uni} = \int_{\theta} [M_{SR}(\theta) - M_{SF}(\theta)] f_P(\theta) d\theta \quad (9)$$

หรือ $\beta_{uni} = \int_{\theta} B(\theta) f_p(\theta) d\theta$ (10)

เมื่อ β_{uni} แทนดัชนีการทำหน้าที่เบี่ยงเบนของข้อสอบที่มีทิศทางเดียวกัน และ $f_P(\theta)$ แทนฟังก์ชันความหนาแน่นความน่าจะเป็นของการแจกแจงความสามารถเป้าหมายจากผู้สอบทั้งสองกลุ่ม

การทดสอบสมมติฐาน

ในการทดสอบสมมติฐานของการทำหน้าที่เบี่ยงเบนของข้อสอบที่มีทิศทางเดียวกันนำดัชนี β_{uni} มากำหนดสมมติฐานศูนย์ (H_0) และสมมติฐานทางเลือก (H_1) ดังนี้

$$H_0 : \beta_{uni} = 0$$

$$H_1 : \beta_{uni} > 0 \qquad (11)$$

สมมติฐานทางเลือก (H_1) มีลักษณะทิศทางเดียวซึ่งใช้ทดสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่เข้าข้างกลุ่มอ้างอิง สำหรับค่าดัชนี β_{uni} คำนวณจากคะแนนของแบบทดสอบชุดย่อยที่ใช้ในการจับคู่เปรียบเทียบ (แบบทดสอบชุดย่อยที่มีความเที่ยงตรงนั่นเอง) และแบบทดสอบชุดย่อยที่ต้องการศึกษาดังนี้

$$X = \sum_{i=1}^{n} U_i \qquad (12)$$

$$Y = \sum_{i=n+1}^{n} U_i \qquad (13)$$

เมื่อ X แทน คะแนนรวมจากแบบทดสอบชุดย่อยที่มีความเที่ยงตรง

Y แทน คะแนนรวมจากแบบทดสอบชุดย่อยที่ต้องการศึกษา

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

คำนวณคะแนนเฉลี่ยจากผลการตอบข้อสอบในแบบทดสอบชุดย่อยที่ต้องการศึกษาของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจที่มีความสามารถระดับเดียวกันโดยที่ความสามารถระดับเดียวกันนั้นจะพิจารณาจากผู้สอบทั้งสองกลุ่มที่มีคะแนนจากผลการตอบแบบทดสอบชุดย่อยที่มีความเที่ยงตรงเท่ากันเพราะถือว่าเป็นผู้ที่มีความสามารถเป้าหมายที่ต้องการวัดเท่ากัน(โดยประมาณ)อาจกล่าวได้ว่า เป็นการพิจารณาจากคะแนนรวมที่เท่ากันของแบบทดสอบชุดย่อยที่มีความเที่ยงตรง $(X = k)$ ดังนี้

$$\overline{Y}_{Rk} - \overline{Y}_{Fk}, \quad k = 0,1,2,...,n \qquad (14)$$

เมื่อ $\overline{Y}_{Rk}$ และ $\overline{Y}_{Fk}$ แทนค่าเฉลี่ยของคะแนนรวม(Y)จากแบบทดสอบชุดย่อยที่ต้องการศึกษาที่ได้คะแนนรวม X = k (k= 0,1,2,...,n) จากผลการตอบแบบทดสอบชุดย่อยที่มีความเที่ยงตรงสำหรับผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจตามลำดับ ถ้า $\overline{Y}_{Rk} - \overline{Y}_{Fk}$ = 0 ทุกคะแนน k แสดงว่าข้อสอบที่สงสัยในแบบทดสอบชุดย่อย (ข้อสอบที่ต้องการศึกษา) ทำหน้าที่ไม่เบี่ยงเบน (No DIF) และถ้า $\overline{Y}_{Rk} - \overline{Y}_{Fk}$ > 0 ทุกคะแนน k แสดงว่าข้อสอบที่สงสัยในแบบทดสอบชุดย่อยทำหน้าที่เบี่ยงเบนแบบมีทิศทางเดียวกัน (Unidirectional DIF) โดยข้อสอบเข้าข้างกลุ่มอ้างอิง แต่ถ้า $\overline{Y}_{Rk} - \overline{Y}_{Fk}$ < 0 ทุกคะแนน k แสดงว่าข้อสอบที่สงสัยในแบบทดสอบชุดย่อยทำหน้าที่เบี่ยงเบนแบบมีทิศทางเดียวกัน โดยข้อสอบเข้าข้างกลุ่มสนใจ

ค่าความแตกต่างของผลการตอบข้อสอบสามารถใช้ประมาณค่า β_{uni} ได้ดังนี้

$$\hat{\beta}_{uni} = \sum_{k=0}^{n} \hat{P}_k \left(\overline{Y}_{Rk} - \overline{Y}_{Fk} \right) \quad (15)$$

โดยที่ $$\hat{P}_k = \frac{J_{Rk} + J_{Fk}}{\sum_{k=0}^{n} \left(J_{Rk} + J_{Fk} \right)} \quad (16)$$

เมื่อ $\hat{P}_k$ แทน สัดส่วนของผู้สอบทั้งหมด (กลุ่มอ้างอิงและกลุ่มสนใจ) ซึ่งตอบแบบทดสอบชุดย่อยที่มีความเที่ยงตรง แล้วได้คะแนนรวม $X = k$

J_{Rk} แทน จำนวนผู้สอบกลุ่มอ้างอิงซึ่งตอบแบบทดสอบชุดย่อยที่มีความเที่ยงตรงแล้วได้คะแนนรวม $X = k$

J_{Fk} แทน จำนวนผู้สอบกลุ่มสนใจซึ่งตอบแบบทดสอบชุดย่อยที่มีความเที่ยงตรงแล้วได้คะแนนรวม $X = k$

จากนั้นนำค่า $\hat{\beta}_{uni}$ ที่คำนวณได้มาทดสอบสมมุติฐานศูนย์ของการทำหน้าที่ไม่เบี่ยงเบน (No DIF) โดยใช้สถิติ B_{uni} ดังนี้

$$B_{uni} = \frac{\hat{\beta}_{uni}}{\hat{\sigma}(\hat{\beta}_{uni})} \quad (17)$$

โดยที่ $$\hat{\sigma}(\hat{\beta}_{uni}) = \left[\sum_{k=0}^{N} \hat{P}_k^2 \left(\frac{1}{J_{Rk}} \hat{\sigma}^2 (Y \mid k, R) + \frac{1}{J_{Fk}} \hat{\sigma}^2 (Y \mid k, F) \right) \right]^{\frac{1}{2}} \quad (18)$$

เมื่อ $\hat{\sigma}(\hat{\beta}_{uni})$ แทนค่าประมาณความคลาดเคลื่อนมาตรฐานของ $\hat{\beta}_{uni}$

$\hat{\sigma}^2(Y|k,R)$ แทนค่าประมาณความแปรปรวนของคะแนนจากแบบทดสอบชุดย่อยที่ต้องการศึกษาสำหรับผู้สอบกลุ่มอ้างอิงซึ่งมีคะแนนรวมเท่ากับ k

$\hat{\sigma}^2(Y|k,F)$ แทนค่าประมาณความแปรปรวนของคะแนนจากแบบทดสอบชุดย่อยที่ต้องการศึกษาสำหรับผู้สอบกลุ่มสนใจ ซึ่งมีคะแนนรวมเท่ากับ k

สถิติที่ใช้ในการทดสอบ B_{uni} มีการแจกแจงใกล้เคียงการแจกแจงแบบปกติมาตรฐาน $[N(0,1)]$ เมื่อข้อสอบทำหน้าที่ไม่เบี่ยงเบน (B_{uni}=0) และถ้าผลการทดสอบพบว่า $B_{uni} > Z_\alpha$ อย่างมีนัยสำคัญที่ระดับ α โดยที่ $P[N(0,1) > Z_\alpha] = \alpha$ แสดงว่าปฏิเสธ H_0 นั่นคือข้อสอบที่นำมาตรวจสอบทำหน้าที่เบี่ยงเบนที่มีทิศทางเดียวกัน (Unidirectional DIF)เมื่อ (B_{uni}) > 0 แสดงว่าข้อสอบ เข้าข้างกลุ่มอ้างอิง และเมื่อ (B_{uni}) < 0 แสดงว่าข้อสอบเข้าข้างกลุ่มสนใจ อย่างไรก็ตามสถิติการทดสอบ B_{uni} มีแนวโน้มที่จะแสดงความคลาดเคลื่อนประเภทที่ 1 (Type I error) กล่าวคือข้อสอบทำหน้าที่เบี่ยงเบน ทั้งที่ในความเป็นจริงแล้วข้อสอบทำหน้าที่ไม่เบี่ยงเบนที่เป็นเช่นนี้เนื่องจากเกิด ความแตกต่างของการแจกแจงความสามารถเป้าหมายระหว่างกลุ่มอ้างอิงและกลุ่มสนใจ คือถ้าการแจกแจงความสามารถเป้าหมายของผู้สอบกลุ่มอ้างอิงสูงกว่ากลุ่มสนใจจึงเกิด

เงื่อนไขที่เรียกว่า “ผลกระทบ” (Impact) ซึ่งจะทำให้สถิติ B_{uni} มีค่าเฟ้อ (Inflate) หรือมีค่าสูงกว่าปกติดังนั้นจึงมีความจำเป็นที่จะแก้ไขความแตกต่างของการแจกแจงความสามารถเป้าหมายโดยใช้การปรับแก้การถดถอย (Regression correction) เพื่อขจัดอิทธิพลค่าเฟ้อของผลกระทบ ซึ่งจะแปลงค่า $\overline{Y}_{Rk}$, $\overline{Y}_{Fk}$ เป็น $\overline{Y}^*_{Rk}$, $\overline{Y}^*_{Fk}$ ทีละคู่ (Shealy; & Stout. 1993: 169-170; 190-193) ดังนี้

$$\overline{Y}^*_{gk} = \overline{Y}_{gk} + \hat{M}_{gk}\left[\overline{V}(k) - \hat{V}_g(k)\right] \quad (19)$$

เมื่อ $\overline{Y}^*_{gk}$ แทน ค่าประมาณของคะแนนจริงจากแบบทดสอบชุดย่อยที่ต้องการศึกษาของผู้สอบกลุ่มย่อย k ในกลุ่ม g(R หรือ F)

$\overline{Y}_{gk}$ แทน ค่าเฉลี่ยของคะแนนสังเกตจากแบบทดสอบชุดย่อยที่ต้องการศึกษาของผู้สอบกลุ่ม g (R หรือ F) ซึ่งได้คะแนน $X = k$ จากแบบทดสอบชุดย่อยที่มีความเที่ยงตรง

$\hat{M}_{gk}$ แทน ค่าประมาณคร่าวๆของค่าเบี่ยงเบนของคะแนนจริงจากแบบทดสอบชุดย่อยที่ต้องการศึกษาที่มีต่อฟังก์ชันของคะแนนจริงจากแบบทดสอบชุดย่อยที่มีความเที่ยงตรงของผู้สอบกลุ่ม g (R หรือ F) ซึ่งได้คะแนน $X = k$

โดยที่ $$\hat{M}_{gk} = \frac{\overline{Y}_{g,k+1} - \overline{Y}_{g,k-1}}{\hat{V}_g(k+1) - \hat{V}_g(k-1)} \quad (20)$$

เมื่อ $\overline{Y}_{g,k+1}$ และ $\overline{Y}_{g,k-1}$ แทน ค่าเฉลี่ยของคะแนนสังเกตจากแบบทดสอบชุดย่อยที่ต้องการศึกษาของผู้สอบกลุ่มย่อย จำนวน $k+1$ และ $k-1$ กลุ่มตามลำดับ

$\hat{V}_g(k+1)$ และ $\hat{V}_g(k-1)$ แทนค่าประมาณของคะแนนจริงจากแบบทดสอบชุดย่อยที่ใช้ในการจับคู่เปรียบเทียบของผู้สอบกลุ่มย่อย จำนวน $k+1$ และ $k-1$ กลุ่มตามลำดับ

$\hat{V}_g(k)$ แทน ค่าประมาณคะแนนจริงของความสามารถเป้าหมาย สำหรับผู้สอบกลุ่ม g (R หรือ F) ซึ่งได้คะแนน $X = k$ จากแบบทดสอบชุดย่อยที่มีความเที่ยงตรง

โดยที่ $$\hat{V}_g(k) = \overline{X}_g + \left[1 - \frac{\hat{\sigma}^2(e|g)}{\hat{\sigma}^2(X|g)}\right](k - \overline{X}_g) \quad (21)$$

ซึ่ง $$\hat{\sigma}^2(e|g) = \sum_{i=1}^{n} \overline{U}_{ig}(1 - \overline{U}_{ig}) \quad (22)$$

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

เมื่อ X_{gi} แทน คะแนนจากแบบทดสอบชุดย่อยที่มีความเที่ยงตรงของผู้สอบคนที่ j ในกลุ่ม g

$\overline{X}_g$ แทน คะแนนเฉลี่ยจากแบบทดสอบชุดย่อยที่มีความเที่ยงตรงของผู้สอบคนที่ j ในกลุ่ม g

$\overline{U}_{ig}$ แทน สัดส่วนการตอบข้อสอบถูกของผู้สอบกลุ่ม g ซึ่งตอบข้อสอบข้อที่ i จากแบบทดสอบชุดย่อยที่มีความเที่ยงตรง

$\hat{V}_{(k)}$ แทน ค่าเฉลี่ยของค่าประมาณคะแนนจริงของความสามารถเป้าหมายสำหรับผู้สอบกลุ่ม g (R หรือ F) ซึ่งได้คะแนน $X = k$ จากแบบทดสอบชุดย่อยที่มีความเที่ยงตรง

โดยที่ $\hat{V}(k) = \frac{1}{2}\left[\hat{V}_R(k) + \hat{V}_F(k)\right]$ (24)

เมื่อ $\hat{V}_R(k)$ และ $\hat{V}_F(k)$ แทน ค่าประมาณคะแนนจริงของความสามารถเป้าหมายซึ่งได้คะแนน $X = k$ จากแบบทดสอบชุดย่อยที่มีความเที่ยงตรงสำหรับผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจตามลำดับ

หลังจากแก้ไขความแตกต่างของการแจกแจงความสามารถเป้าหมายด้วยวิธีการปรับแก้การถดถอยเพื่อขจัดอิทธิพลค่าเฟ้อของผลกระทบโดยการแปลงค่า $\overline{Y}_{Rk}$, $\overline{Y}_{Fk}$ ให้เป็น $\overline{Y}^*_{Rk}$, $\overline{Y}^*_{Fk}$ ทีละคู่ ดังนั้นค่า $\overline{Y}^*_{Rk} - \overline{Y}^*_{Fk}$ จึงเป็นความแตกต่างของผลการตอบข้อสอบในแบบทดสอบชุดย่อยที่ต้องการศึกษาระหว่างกลุ่มอ้างอิงและกลุ่มสนใจที่มีความสามารถระดับเดียวกันถ้า $\overline{Y}^*_{Rk} - \overline{Y}^*_{Fk} = 0$ ทุกคะแนน k แสดงว่าข้อสอบที่สงสัยในแบบทดสอบชุดย่อย (ข้อสอบที่ต้องการศึกษา) ทำหน้าที่ไม่เบี่ยงเบน (No-DIF) และถ้า $\overline{Y}^*_{Rk} - \overline{Y}^*_{Fk} > 0$ ทุกคะแนน k แสดงว่าข้อสอบที่สงสัยในแบบทดสอบชุดย่อยทำหน้าที่เบี่ยงเบนแบบมีทิศทางเดียวกัน (Uniderectional DIF) โดยข้อสอบเข้าข้างกลุ่มอ้างอิง แต่ถ้า $\overline{Y}^*_{Rk} - \overline{Y}^*_{Fk} < 0$ ทุกคะแนน k แสดงว่าข้อสอบที่สังสัยในแบบทดสอบชุดย่อยทำหน้าที่เบี่ยงเบนแบบมีทิศทางเดียวกัน โดยข้อสอบเข้าข้างกลุ่มสนใจ ค่าความแตกต่างของผลการตอบข้อสอบดังกล่าว สามารถนำมาใช้ประมาณค่า β_{uni} ดังนี้

$$\hat{\beta}_{uni} = \sum_{k=0}^{n} \hat{P}_k \left(\overline{Y}^*_{Rk} - \overline{Y}^*_{Fk}\right) \quad (25)$$

เมื่อ $\hat{P}_k$ แทนสัดส่วนของผู้สอบทั้งหมด (กลุ่มอ้างอิงและกลุ่มสนใจ) ซึ่งตอบแบบทดสอบชุดย่อยที่มีความเที่ยงตรง แล้วได้คะแนนรวม $X = k$ หลังจากนั้นจึงนำค่า $\hat{\beta}_{uni}$ ที่คำนวณได้มาทดสอบสมมติฐานศูนย์ของการทำหน้าที่ไม่เบี่ยงเบน (No-DIF) โดยใช้สถิติ B_{uni} ดังนี้

$$B_{uni} = \frac{\hat{\beta}_{uni}}{\hat{\sigma}(\hat{\beta}_{uni})} \quad (26)$$

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

รุสโซ และสเตาท์ (Roussos; & Stout. 1996: 220) ได้เสนอเกณฑ์เพื่อใช้จำแนกขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบไว้ดังนี้

- DIF ระดับ A (ขนาดเล็ก): ปฏิเสธสมมุติฐานศูนย์ และ $|B_{uni}| < 0.059$

- DIF ระดับ B (ขนาดปานกลาง):ปฏิเสธสมมุติฐานศูนย์ และ $0.059 \leq |B_{uni}| < 0.088$

- DIF ระดับ C (ขนาดใหญ่):ปฏิเสธสมมุติฐานศูนย์ และ $|B_{uni}| \geq 0.088$

3.2 วิธีโพลีโตมัสซิบเทสท์ (Polytomous SIBTEST)

ชาง มาซซีโอ และรุสโซ (Chang; Mazzeo; & Roussos.1996) ได้พัฒนาวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบสองค่า(dichotomous)โดยปรับขยายจากวิธีซิบเทสท์ (Shealy; & Stout.1993) เพื่อใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่า (Polytomous)วิธีที่พัฒนาขึ้นนี้เรียกว่า **"วิธีซิบเทสท์ที่ปรับขยาย"** (Modified SIBTEST procedure)(Chang; Mazzeo; & Roussos. 1996) หรือ **"วิธีโพลีโตมัสซิบเทสท์"** (Polytomous SIBTEST: Poly-SIB)

แนวคิดและหลักการ

ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่าโดยใช้วิธีโพลีโตมัสซิบเทสท์ (Polytomous SIBTEST) เป็นการปรับขยายวิธีซิบเทสท์ (SIBTEST) ที่ให้คะแนนแบบสองค่า วิธีโพลีโตมัสซิบเทสท์ได้รวมเอาการให้คะแนนข้อสอบแบบสองค่าเป็นกรณีเฉพาะ การตรวจสอบด้วยวิธีดังกล่าวข้อสอบที่ต้องการศึกษาสามารถเป็นข้อสอบที่ให้คะแนนแบบสองค่าหรือข้อสอบที่ให้คะแนนแบบหลายค่าก็ได้ ส่วนแบบทดสอบที่ใช้ในการจับคู่จะประกอบไปด้วยการผสมของข้อสอบที่ให้คะแนนแบบสองค่าและข้อสอบที่ให้คะแนนแบบหลายค่าก็ได้ นอกจากนี้การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่าโดยใช้วิธีโพลีโตมัสซิบเทสท์จะใช้กรอบแนวคิดเช่นเดียวกับวิธีซิบเทสท์ดั้งเดิม (Original SIBTEST) ที่ใช้ตรวจสอบในกรณีให้คะแนนแบบสองค่า

ถึงแม้ว่าจุดเริ่มต้นของวิธีซิบเทสท์เป็นการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบสองค่าโดยพัฒนาภายใต้โมเดลทฤษฎีการตอบข้อสอบ(litem response theory: IRT) แต่วิธีซิบเทสท์ก็สามารถทำให้เข้าใจและอธิบายได้ในมุมมองของทฤษฎีการทดสอบมาตรฐานเดิม (Classical test theory: CTT) หรือทฤษฎีคะแนนจริง (True score theory) ซึ่งตามข้อตกลงของทฤษฎีการทดสอบมาตรฐานเดิมเกี่ยวกับคะแนนสังเกต X กำหนดว่า $X = T + E$ ณ ที่นี้ T แทนคะแนนจริงของแบบทดสอบที่ใช้ในการจับคู่สำหรับผู้สอบที่เลือกมาอย่างสุ่ม ดังนั้นตัวแปร $E = X - T$ แทนความคลาดเคลื่อนในการวัด และสมมุติว่ามีค่าเฉลี่ยเป็นศูนย์ทั้งสองกลุ่มเมื่อให้ $f_g(t)$ แทนความหนาแน่นของคะแนนจริงของแบบทดสอบที่ใช้ในการจับคู่ของกลุ่ม g (R หรือ F) ซึ่งการถดถอยดังกล่าวสามารถเขียนในรูปสัญลักษณ์ได้ดังนี้ $E_g[Y|t]$ โดยที่ $E_g[Y|t] = E[Y|T = t, G = g]$

ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่า ตามวิธีโพลีโตมัสซิบเทสท์ ข้อสอบที่ต้องการศึกษาไม่แสดงการทำหน้าที่เบี่ยงเบน (Not exhibit DIF) จะเป็นไปตามข้อตกลง**การไม่ปรากฏการทำหน้าที่เบี่ยงเบนของข้อสอบที่ใช้ตัวแปรแฝงภายใต้ทฤษฎีการทดสอบมาตรฐานเดิม**(Classical test theory latent-variable null DIF) ซึ่งกำหนดว่า **"ถ้า** $E_R[Y|t]=E_F[Y|t]$ **สำหรับทุกค่าของคะแนนจริง t จากแบบทดสอบที่ใช้ในการจับคู่"** ถ้าข้อสอบจากแบบทดสอบที่ใช้ในการจับคู่ เกิดจากโมเดลตามทฤษฎีการตอบข้อสอบ เช่น โมเดลพาร์เชียลเครดิต (Partial credit model) โมเดลพาร์เชียลเครดิตทั่วไป (The generalized partial credit model) หรือโมเดลเกรดเรสพอนส์ (Graded response model) คะแนนจริงของแบบทดสอบที่ใช้ในการจับคู่ก็จะเป็นการเปลี่ยนแปลงแบบเพิ่มขึ้นทิศทางเดียวอย่างสมบูรณ์แบบ (Strictly monotonic) ของความสามารถ θ (Chang; Mazzeo; & Roussos. 1996: 337) ดังนั้นคะแนนจริงของแบบทดสอบที่ใช้ในการจับคู่จะเป็นการเปลี่ยนแปลงแบบหนึ่งต่อหนึ่งของตัวแปรความสามารถ θ และถ้าข้อสอบหนึ่งข้อสอดคล้องกับข้อตกลงดังกล่าวแล้วฟังก์ชันการตอบรายการข้อสอบ (ICRFs) ซึ่งได้ให้ $E_R[Y|t]$ ก็จะเหมือนกันกับฟังก์ชันการตอบรายการข้อสอบซึ่งได้ให้ $E_F[Y|t]$ ดังนั้นข้อตกลงดังกล่าวจึงสมมูลกับข้อตกลงการไม่ปรากฏการทำหน้าที่เบี่ยงเบนของข้อสอบเมื่อใช้ตัวแปรตามทฤษฎีการตอบข้อสอบ (Chang; Mazzeo; & Roussos. 1996: 337) กล่าวคือ $E_R[Y|\theta]=E_F[Y|\theta]$ เราจึงสามารถนำผลดังกล่าวไปอธิบายการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่า ด้วยวิธีโพลีโตมัสซิบเทสท์

กระบวนการตรวจสอบ

การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่าในรายการแบบจัดอันดับ (Ordered categories) จะกำหนดสัญลักษณ์เพื่อความเข้าใจตรงกันดังนี้

กำหนดให้

Y แทนคะแนนของข้อสอบที่ต้องการศึกษา ซึ่งมีรายการแบบจัดอันดับจำนวน $m+1$ รายการ (Y = 0,1,2,...,m)

$X_1, X_2, ..., X_n$ แทนคะแนนของข้อสอบที่ใช้ในการจับคู่เปรียบเทียบจำนวน n ข้อ

$m_1, m_2, ..., m_n$ แทนคะแนนมากสุดที่เป็นไปได้ของ $X_1, X_2, ..., X_n$ ตามลำดับ

X แทนคะแนนที่ใช้ในการจับคู่เปรียบเทียบ คำนวณจากสูตร

$$X = \sum_{j=1}^{n} X_j \qquad (1)$$

$X = 0, 1, 2, ..., n_H$ โดยที่ n_H เป็นคะแนนของข้อสอบที่ใช้ในการจับคู่เปรียบเทียบซึ่งมีค่ามากที่สุดที่เป็นไปได้ ซึ่งคำนวณมาจาก

$$n_H = \sum_{j=1}^{n} m_j \qquad (2)$$

$\overline{Y}_{gk}$ แทน คะแนนเฉลี่ยของข้อสอบที่ต้องการศึกษาสำหรับผู้เข้าสอบทั้งหมดในกลุ่ม g (R หรือ F) ซึ่งได้คะแนน $X = k$

เราทราบว่าการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบสองค่าโดยวิธีซิบเทสท์ (Shealy; & Stout.1993) จะอาศัยข้อตกลงเกี่ยวกับ**การไม่ปรากฏการทำหน้าที่เบี่ยงเบนของข้อสอบเมื่อใช้ตัวแปรแฝง คือว่า ข้อสอบหนึ่งข้อแสดงการทำหน้าที่เบี่ยงเบนถ้าคะแนนที่คาดหวังของกลุ่มอ้างอิงและกลุ่มสนใจไม่เท่ากันสำหรับทุกระดับของความสามารถ θ ที่ใช้ในการจับคู่** นั่นคือ **"การถดถอยของ Y บนตัวแปรแฝง θ ไม่ได้เหมือนกันทุกระดับความสามารถ"** กล่าวคือ **"มีบางค่าของความสามารถ θ ที่ทำให้** $E_R[Y|\theta] \neq E_F[Y|\theta]$" และปริมาณของการทำหน้าที่เบี่ยงเบนของข้อสอบ (B_0) สามารถวัดได้จาก $B_0 = E_R[Y|\theta] - E_F[Y|\theta]$ ส่วนดัชนีการทำหน้าที่เบี่ยงเบนของข้อสอบ (β_0) สามารถคำนวณได้จาก $\beta_0 = \int B_0(\theta) f_F(\theta) d\theta$ เมื่อ $f_F(\theta)$ แทนความหนาแน่นของความสามารถในกลุ่มสนใจ นอกจากนี้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่า คะแนนจริง t ของแบบทดสอบที่ใช้ในการจับคู่เป็นการแปลงค่าตัวแปรแฝง θ อย่างง่ายแบบหนึ่งต่อหนึ่ง ทำให้ปริมาณ (Local measure) ของการทำหน้าที่เบี่ยงเบนของข้อสอบ ณ คะแนนจริง t ที่ใช้ในการจับคู่จะเป็น $B(t) = E_R[Y|t] - E_F[Y|t]$ ดังนั้นดัชนีการทำหน้าที่เบี่ยงเบนของข้อสอบที่สอดคล้องกันกับความหนาแน่น $f_F(t)$ สามารถกำหนดจาก $\beta = \int B(t) f_F(t) dt$ ในกรณีทฤษฎีการตอบข้อสอบที่วัดมิติเดียว(Unidimension IRT) ค่า β ที่ได้จากสมการ $\beta = \int B(t) f_F(t) dt$ จะเหมือนกับค่า β ที่ได้จากสมการ $\beta = \int B_0(\theta) f_F(\theta) d\theta$ ซึ่งเป็นดัชนีการทำหน้าที่เบี่ยงเบนของข้อสอบกรณีข้อสอบให้คะแนนแบบสองค่า ปกติแล้วเราสามารถประมาณค่าการทำหน้าที่เบี่ยงเบนของข้อสอบโดยอาศัย

$$d_k = \overline{Y}_{Rk} - \overline{Y}_{Fk}, k = 0,...,n_H \quad (3)$$

เมื่อ $\overline{Y}_{Rk}$ และ $\overline{Y}_{Fk}$ แทนค่าเฉลี่ยของคะแนนข้อสอบที่ต้องการศึกษาของผู้สอบซึ่งได้คะแนน $X = k$ ของกลุ่มอ้างอิงและกลุ่มสนใจตามลำดับค่า $\overline{Y}_{Rk} - \overline{Y}_{Fk}$ เป็นความแตกต่างของผลการตอบข้อสอบที่ต้องการศึกษาระหว่างกลุ่มผู้สอบ ซึ่งมีคะแนนสังเกตจากแบบทดสอบที่ใช้ในการจับคู่เท่ากัน ถ้าสมมติว่าผู้สอบมีคะแนนสังเกตเท่ากับ(ใกล้เคียง)คะแนนจริงจากแบบทดสอบที่ใช้ในการจับคู่ ซึ่งในความเป็นจริงจะเป็นจริงได้เพียงถ้าแบบทดสอบที่ใช้ในการจับคู่มีความยาวมากพอเพื่อทำให้ความเชื่อมั่นสูงหรือกลุ่มผู้สอบที่ศึกษามีการแจกแจงความสามารถ θ เหมือนกัน ดังนั้นสมการ (3) จึงถือว่าเป็นค่าประมาณความแตกต่างของคะแนนสอบที่ระดับคะแนนจริงเท่ากัน ถ้าข้อสอบที่ศึกษาไม่มีคะแนนสังเกตที่ทำให้ข้อสอบทำหน้าที่เบี่ยงเบนแล้วคาดว่า $d_k \approx 0$ สำหรับการประมาณค่าดัชนีการทำหน้าที่เบี่ยงเบนของข้อสอบเป็นดังนี้

$$\hat{\beta} = \sum_{k=0}^{n_H} P_k d_k \quad (4)$$

โดยที่ $P_k = \frac{N_{Rk} + N_{Fk}}{N}$ (5)

เมื่อ P_k แทนสัดส่วนของผู้สอบทั้งหมด (กลุ่มอ้างอิง และกลุ่มสนใจ) ซึ่งตอบแบบทดสอบที่ใช้ในการจับคู่ $X_1, X_2, ..., X_n$ แล้วได้คะแนน $X = k$ ต่อจากนั้นจะนำค่า $\hat{\beta}$ มาทดสอบสมมุติฐานต่อไป

การทดสอบสมมุติฐาน

นำดัชนี $\hat{\beta}$ มาทดสอบสมมุติฐานศูนย์ โดยใช้สถิติ B ดังนี้

$$B = \frac{\hat{\beta}}{\hat{\sigma}(\hat{\beta})} \tag{6}$$

โดยที่ $\hat{\sigma}(\hat{\beta}) = \sqrt{\sum_{k=0}^{n_H} P_k^2 \left[\frac{\hat{\sigma}^2(Y|k,R)}{N_{Rk}} + \frac{\hat{\sigma}^2(Y|k,F)}{N_{Fk}} \right]}$ (7)

เมื่อ $\hat{\sigma}(\hat{\beta})$ แทนค่าประมาณความคลาดเคลื่อนมาตรฐานของ β และ $\hat{\sigma}^2(Y|k,g)$ แทนค่าประมาณความแปรปรวนของคะแนนจากแบบทดสอบที่ต้องการศึกษาในกลุ่ม g (R หรือ F) ซึ่งมีคะแนนรวมเท่ากับ k ในกรณีที่ $\beta = 0$ สถิติ B มีการแจกแจงใกล้เคียงปกติมาตรฐาน $[N(0,1)]$ ข้อสอบทำหน้าที่ไม่เบี่ยงเบน (No DIF) ถ้าผลการทดสอบพบว่า $|B| > Z_{1-\frac{\alpha}{2}}$ อย่างมีนัยสำคัญที่ระดับ α แสดงว่าปฏิเสธ H_0 นั่นคือข้อสอบทำหน้าที่เบี่ยงเบน (DIF) โดยเข้าข้างผู้สอบกลุ่มใดกลุ่มหนึ่ง

3.3 วิธีการถดถอยโลจิสติก (Logistic regression method)

สวามินาธานและโรเจอร์ส (Swaminathan; & Rogers. 1990) ได้พัฒนาโมเดลการถดถอยโล จิสติก (Logistic regression model)จากโมเดลล็อกลิเนียร์ (Log-linear model) และโมเดลโลจิท Logit model)ของเมลเลนเบิร์ก(Mellenbergh. 1982) วิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel: MH) ของฮอลแลนด์และทาเยอร์ (Holland; & Thayer. 1988) พร้อมทั้งเชื่อมแนวคิดของวิธีที่ใช้คุณลักษณะแฝงของทฤษฎีการตอบข้อสอบวิธีนี้ตั้งอยู่บนฐานของโมเดล (Model-based methods) ที่ให้ความสำคัญถึงธรรมชาติความต่อเนื่องของสเกลความสามารถ เพื่อใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนสองค่า (Dichotomous) หลักการตรวจสอบจะใช้โมเดลการถดถอยโลจิสติกทำนายความน่าจะเป็นในการตอบข้อสอบถูก โดยปกติใช้คะแนนแบบทดสอบรวมแทนระดับความสามารถซึ่งสมมุติว่าเป็นตัวแปรความสามารถแบบต่อเนื่อง โมเดลดังกล่าวมีพจน์ (Term) สำหรับทดสอบปฏิสัมพันธ์ระหว่างการเป็นสมาชิกของกลุ่มผู้สอบ (Group membership) กับระดับความสามารถ (Ability level) ดังนั้นจึงสามารถตรวจสอบได้ทั้งข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF) และข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) โมเดลการถดถอยโลจิสติกมีความยืดหยุ่นสามารถนำไปปรับขยายเพื่อใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่า (Polytomous) (Miller; &

Spray. 1993; French; & Miller. 1996; Zumbo. 1999) และผู้สอบหลายกลุ่ม (Agresti.1990; Millsap; & Everson. 1993: 305)

แนวคิดและหลักการ

ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบสองค่าจะใช้โมเดลการถดถอยโลจิสติก ทำนายความน่าจะเป็นในการตอบข้อสอบถูก 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 ความสามารถที่สังเกตได้ของผู้สอบแต่ละคน

β_0 แทน พารามิเตอร์ส่วนตัด (Intercept parameter)

β_1 แทน พารามิเตอร์ความชัน (Slope parameter)

สมการดังกล่าวเป็นโมเดลการถดถอยโลจิสติกแบบมาตรฐานสำหรับใช้ทำนายตัวแปรตามแบบสองค่าจากตัวแปรอิสระที่กำหนดให้ โมเดลการถดถอยโลจิสติกสามารถนำไปประยุกต์ใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ โดยระบุสมการแยกกลุ่มผู้สอบสองกลุ่มที่สนใจดังนี้

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

เมื่อ u_{ij} แทน ผลการตอบข้อสอบของผู้สอบคนที่ i ในกลุ่ม j

θ_{ij} แทน ระดับความสามารถของผู้สอบคนที่ i ในกลุ่ม j

β_{0j} แทน พารามิเตอร์ส่วนตัดในกลุ่ม j

β_{1j} แทน พารามิเตอร์ความชันในกลุ่ม j

จากนิยามที่ได้รับการยอมรับของการทำหน้าที่เบี่ยงเบนของข้อสอบ (Differential item functioning: DIF) กำหนดไว้ว่า“**ข้อสอบหนึ่งข้อแสดงการทำหน้าที่เบี่ยงเบนถ้าผู้สอบมีความสามารถเท่ากันแต่มาจากกลุ่มผู้สอบที่แตกต่างกัน มีความน่าจะเป็นในการตอบข้อสอบถูกไม่เท่ากัน”**(Swaminathan; & Rogers.1990: 362-363) จากนิยามดังกล่าวเมื่อพิจารณาโมเดลการถดถอยโลจิสติกสามารถสรุปได้ว่า ถ้า $\beta_{01}=\beta_{02}$ และ $\beta_{11}=\beta_{12}$ แล้วโค้งการถดถอยโลจิสติกของผู้สอบสองกลุ่มเหมือนกันแสดงว่าข้อสอบทำหน้าที่ไม่เบี่ยงเบน (No DIF) ถ้า $\beta_{11}=\beta_{12}$ แต่ $\beta_{01}\neq\beta_{02}$ แล้วโค้งการถดถอยโลจิสติกของผู้สอบสองกลุ่มจะขนานกันแต่ไม่ทับกัน แสดงว่าข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF) แต่ถ้า $\beta_{01}=\beta_{02}$ แต่ $\beta_{11}\neq\beta_{12}$ แล้วโค้งการถดถอยโลจิสติกระหว่างกลุ่มผู้สอบไม่ขนานกัน แสดงว่าข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF)

กระบวนการตรวจสอบ

ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบจะใช้โมเดลการถดถอยโลจิสติกทดสอบความแตกต่างของผลการตอบข้อสอบระหว่างกลุ่มผู้สอบที่ระดับความสามารถเดียวกัน โดยสามารถทดสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกันและที่ไม่เป็นรูปแบบเดียวกันได้พร้อมกัน โดยใช้โมเดลที่สมมูลกับโมเดลของสมการ (2) ดังนี้

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

$$\text{โดยที่ } Z_{ij} = \tau_0 + \tau_1\theta + \tau_2 g + \tau_3(\theta g) \quad (4)$$

เมื่อ $P(u_{ij}=1)$ แทน ความน่าจะเป็นแบบมีเงื่อนไขของการตอบถูกของผู้สอบที่มีความ สามารถ θ คนที่ i ในกลุ่ม j

Z_{ij} แทน ฟังก์ชั่นซึ่งกำหนดในรูปผลรวมเชิงเส้น (Linear combination) ของ θ และ g

จากโมเดลในสมการ (4) τ_0 เป็นพารามิเตอร์ส่วนตัด (Intercept parameter), τ_1 เป็นสัมประสิทธิ์ของความสามารถ (θ) τ_2 เป็นสัมประสิทธิ์สำหรับตัวแปรกลุ่ม (g) และ τ_3 เป็นพารามิเตอร์ปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับการเป็นสมาชิกของกลุ่มผู้สอบ (θg) ส่วนตัวแปร g ที่ใช้แทนสมาชิกของกลุ่มผู้สอบจะกำหนดรหัสเป็นตัวแปรดัมมี (Dummy -variable) ดังนี้

$$g = \begin{cases} 1 \text{ ถ้าผู้สอบเป็นสมาชิกในกลุ่มอ้างอิง} \\ 0 \text{ ถ้าผู้สอบเป็นสมาชิกในกลุ่มสนใจ} \end{cases} \quad (5)$$

จากโมเดลการถดถอยโลจิสติกในสมการ (4) พจน์ θg เป็นผลคูณของตัวแปรอิสระสองตัว คือ θ และ g ซึ่งกำหนดรหัสตัวแปรดังข้างต้น พารามิเตอร์ τ_2 สอดคล้องกับความแตกต่างกลุ่มจากผลการตอบข้อสอบ ส่วน τ_3 สอดคล้องกับปฏิสัมพันธ์ระหว่างการเป็นสมาชิกของกลุ่มและระดับความสามารถของโมเดลในสมการ (2) ดังนี้

$$\tau_2 = \beta_{01} - \beta_{02} \quad (6)$$

$$\tau_3 = \beta_{11} - \beta_{12} \quad (7)$$

ในการพิจารณาว่าข้อสอบทำหน้าที่เบี่ยงเบนหรือไม่ พิจารณาจากพารามิเตอร์ τ_2 และ τ_3 เทียบกับ 0 กล่าวคือ ถ้า $\tau_2 = 0$ และ $\tau_3 = 0$ แสดงว่าข้อสอบทำหน้าที่ไม่เบี่ยงเบน (No DIF) แต่ถ้า $\tau_2 \neq 0$ หรือ $\tau_3 \neq 0$ ข้อสอบจะทำหน้าที่เบี่ยงเบนโดยที่ ถ้า $\tau_2 \neq 0$ และ $\tau_3 = 0$ แสดงว่าข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และถ้า $\tau_3 \neq 0$ (ซึ่ง $\tau_2 = 0$ หรือ $\tau_2 \neq 0$ ก็ได้) แสดงว่าข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน

สำหรับสมการ (3) สามารถเขียนให้อยู่ในรูปลอการิทึมธรรมชาติของอัตราส่วนแต้มต่อ (Odds ratio) ได้ดังนี้ (Millsap; & Everson.1993: 305; Zumbo.1999: 23)

$$\ln\left[\frac{P}{(1-P)}\right] = \tau_0 + \tau_1\theta + \tau_2 g + \tau_3(\theta g) \quad (8)$$

การประมาณค่าพารามิเตอร์

ในการประมาณค่าของข้อสอบแต่ละข้อจะใช้การประมาณค่าความน่าจะเป็นสูงสุด (Maximum likelihood estimation: MLE) โดยใช้ฟังก์ชันการคำนวณดังนี้(Narayanan; & Swaminathan. 1996: 262; Swaminathan; & Rogers. 1990: 364-365)

$$L(u_{ij}|\theta) = \prod_{i=1}^{N}\prod_{j=1}^{n} P(u_{ij})^{u_{ij}}\left[1-P(u_{ij})\right]^{1-u_{ij}} \quad (9)$$

เมื่อ N แทน ขนาดตัวอย่าง (sample size)

n แทน ความยาวแบบทดสอบ

$u_{ij}=1$ และ $P(u_{ij})$ เป็นค่าที่คำนวณจาก (3)

ในการประมาณค่าพารามิเตอร์จะใช้วิธีการประมาณค่าความน่าจะเป็นสูงสุด (MLE) มีการแจกแจงปกติหลายตัวแปรแบบเชิงเส้นกำกับ (Asymptotically multivariate normal) โดยมีเวกเตอร์เฉลี่ย τ และพารามิเตอร์ความแปรปรวน-ความแปรปรวนร่วม $\sum$ ขณะที่ $\sum^{-1}$ เป็นเมตริกซ์สารสนเทศ (information matrix) ดังนี้

$$\sum^{-1} = -E\left[\frac{\partial^2}{\partial\tau_r\partial\tau_s}\right]; \quad r,s=0,...,3 \quad (10)$$

เมื่อ E แทนค่าคาดหวังการดำเนินการ และ lnL แทนลอการิทึมของความน่าจะเป็นที่กำหนดในสมการ (9) ดังนั้นการแจกแจงของค่าประมาณพารามิเตอร์ τ จะอยู่ในรูปดังนี้

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

โดยที่ $\hat{\tau}' = [\tau_0, \tau_1, \tau_2, \tau_3]$

สำหรับค่าความคาดเคลื่อนมาตรฐานแบบเชิงเส้นกำกับของค่าประมาณ $\tau_S(s=0,1,2,3)$ สามารถคำนวณจากรากที่สองของสมาชิกในแนวทแยงมุมสี่ค่าของเมตริกซ์ $\sum$ ดังนี้

$$SE(\hat{\tau}_S) = \sqrt{\sum^{ss}} \quad (12)$$

การทดสอบสมมุติฐาน

การทดสอบสมมุติฐานเพื่อตรวจหาข้อสอบที่ทำหน้าที่เบี่ยงเบนในแบบทดสอบ ต้องทดสอบสมมุติฐานเกี่ยวกับ τ บางค่าซึ่งสมมุติฐานที่สำคัญ คือ $H_0:\tau_2=0$ และ $H_0:\tau_3=0$ เนื่องจากการประมาณค่าของ τ_2 และ τ_3 เป็นแบบตัวแปรเดียวปกติ (Univariate normal) จึงสามารถใช้การทดสอบสมมุติฐานทีละค่า ในการควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 จะใช้การปรับแก้แบบบอนเฟอร์โรนี (Bonferroni adjustment) เช่นเดียวกับวิธีซิบเทสท์ อีกทางเลือกหนึ่งของการทดสอบสมมุติฐานของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF) และข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) จะทดสอบพร้อมกัน โดยกำหนดสมมติฐานศูนย์ $H_0:\tau_2=0$ 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 บ่งชี้ถึงปฏิสัมพันธ์ระหว่างระดับความสามารถกับการเป็นสมาชิกของกลุ่มผู้สอบ ถ้า $\tau_2\neq 0$ และ $\tau_3=0$ แสดงว่าข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน โดยเข้าข้างกลุ่มอ้างอิงเมื่อ $\tau_2>0$ หรือ เข้าข้างกลุ่มสนใจเมื่อ $\tau_2<0$ แต่ถ้า $\tau_3\neq 0$ ($\tau_2=0$ หรือ $\tau_2\neq 0$ ก็ได้) แสดงว่าข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) เมื่อ $\tau_3>0$ ข้อสอบจะเข้าข้างกลุ่มอ้างอิงที่ระดับความสามารถสูงกว่า และเข้าข้างกลุ่มสนใจที่ระดับความสามารถต่ำกว่า ในทางตรงกันข้าม เมื่อ $\tau_3<0$ ข้อสอบจะเข้าข้างกลุ่มสนใจที่ระดับความสามารถสูงกว่าและเข้าข้างกลุ่มอ้างอิงที่ระดับความสามารถต่ำกว่าสมมติฐานที่ใช้ทดสอบข้อสอบทำหน้าที่เบี่ยงเบนสองรูปแบบพร้อมกันเป็นดังนี้

$$H_0:\mathbf{C}\tau=\mathbf{0}$$
$$H_1:\mathbf{C}\tau\neq\mathbf{0} \qquad (13)$$

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

สำหรับสถิติที่ใช้ทดสอบสมมติฐาน คือ

$$\chi^2=\ \hat{\tau}'\mathrm{C}'\ \ (\mathrm{C}\sum\mathrm{C}')^{-1}\mathrm{C}\hat{\tau}' \qquad (15)$$

สถิติการทดสอบดังกล่าวมีการแจกแจงแบบไค-สแคว์ (χ^2) ที่ระดับชั้นของความเป็นอิสระเท่ากับ $2(\mathrm{df}=2)$ เมื่อค่าสถิติการทดสอบที่ได้จากสมการ(15) มีค่ามากกว่า $\chi^2_{\alpha:2}$ แสดงว่าปฏิเสธสมมติฐานของการทำหน้าที่ไม่เบี่ยงเบนของข้อสอบ (No DIF) นั่นคือข้อสอบทำหน้าที่เบี่ยงเบนนั่นเอง

ซัมโบและโทมัส (Hidalho; & Pina. 2004:907) ได้เสนอ ดัชนี $R^2\Delta$ เพื่อใช้จำแนกขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบโดยมีพื้นฐานจากค่าประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก (Weighted least squares) เกณฑ์สำหรับดัชนี $R^2\Delta$ มีดังนี้

- DIF ระดับ A (ขนาดเล็ก): $R^2\Delta < 0.13$
- DIF ระดับ B (ขนาดปานกลาง): $0.13 \leq R^2\Delta \leq 0.26$
- DIF ระดับ C (ขนาดใหญ่): $R^2\Delta \geq 0.26$

ความสัมพันธ์ระหว่างวิธีการถดถอยโลจิสติกและวิธีแมนเทล-แฮนส์เซล

สวามินาธาน และโรเจอร์ส (Swaminathan; & Rogers. 1990: 363-364) ได้แสดงให้เห็นถึงความเกี่ยวข้องระหว่างวิธีการถดถอยโลจิสติกและวิธีแมนเทล-แฮนส์เซล เมื่อกำหนดตัวแปรความ สามารถแบบไม่ต่อเนื่อง m ระดับ(แบบทดสอบมี $m-1$ ข้อ) โดยนิยามตัวแปร $X_k (k=1,...,m-1)$ ดังนี้

$$X_k = \begin{cases} 1 & \text{ถ้าผู้สอบที่เป็นสมาชิกที่มีความสามารถระดับ } k \\ 0 & \text{ถ้าผู้สอบที่เป็นสมาชิกมีความสามารถระดับอื่น ๆ} \end{cases} \quad (16)$$

ผู้สอบทุกคนที่มีความสามารถระดับ m ได้รับคะแนน 1(หรือ 0) จะกำหนดรหัส Z ในสมการ (3) ดังนี้

$$Z = \beta_0 + \sum_{k=1}^{m-1} \beta_k X_k + \tau g \quad (17)$$

จากสมการ(17) มีข้อสังเกตว่า ไม่มีเทอมปฏิสัมพันธ์ระหว่าง X_k และ g สูตรดังกล่าวจึงสามารถเขียนให้อยู่ในรูปลอการิทึมของอัตราส่วนของความน่าจะเป็นในการตอบข้อสอบถูกต่อความน่าจะเป็นในการตอบข้อสอบผิด ได้ดังนี้

$$\ln\left[\frac{P}{(1-P)}\right] = \beta_0 + \sum_{k=1}^{m-1} \beta_k X_k + \tau g \quad (18)$$

ในกรณีนี้ $\tau = \ln\alpha$ เมื่อ α เป็นอัตราส่วนแต้มต่อร่วม ซึ่งกำหนดโดยฮอลแลนด์และทาเยอร์ (Holland; & Thayer. 1988: 134) การทดสอบสมมุติฐาน $\tau = 0$ ของวิธีการถดถอยโลจิสติกจะสมมูล (Equivalent) กับการทดสอบ $\alpha = 1$ ของวิธีแมนเทล-แฮนส์เซล ดังนั้นจึงอาจกล่าวได้ว่าวิธีแมนเทล-แฮนส์เซลตั้งอยู่บนพื้นฐานโมเดลการถดถอยโลจิสติกเมื่อตัวแปรความสามารถ เป็นตัวแปรแบบไม่ต่อเนื่อง (Discrete variable) และไม่มีปฏิสัมพันธ์ระหว่างตัวแปรความสามารถและตัวแปรกลุ่ม เมื่อพิจารณาโมเดลในแบบแผนการทดลองพบว่า โมเดลการถดถอยโลจิสติกที่คำนวณโดยใช้สมการ (2) และ (3) สอดคล้องกับการวิเคราะห์ของโมเดลความแปรปรวนร่วม (Analysis of covariance model) ในขณะที่โมเดลแมนเทล-แฮนส์เซลที่คำนวณโดยใช้สมการ (17) สอดคล้องกับ

โมเดลแบบแผนบล็อคสุ่ม (Randomized block design: RBD) ทั้งนี้เพราะแบบแผนแบบบล็อคสุ่มมีข้อตกลงว่าไม่มีปฏิสัมพันธ์ระหว่างปัจจัย “บล็อค” กับปัจจัย “ทรีทเมนท์” จะเห็นว่าวิธีการถดถอยโลจิสติกและวิธีแมนเทล-แฮนส์เซลมีความคล้ายคลึงกัน

3.4 วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก (Logistic discriminant function analysis)

วิธีการถดถอยโลจิสติกสำหรับกรณีการให้คะแนนแบบสองค่าเป็นวิธีที่ใช้คะแนนสังเกตในรูปแบบพาราเมตริก (Parametric form) ของความสัมพันธ์ระหว่างคะแนนข้อสอบและตัวแปรที่ใช้ในการจับคู่ (Potenza; & Dorans. 1995) วิธีดังกล่าวตั้งอยู่บนโมเดลความน่าจะเป็นของการสังเกตการตอบข้อสอบที่มีคะแนนแบบสองค่า (U) ซึ่งเป็นฟังก์ชันของตัวแปรสำรวจ (Explanatory variable) สองตัว ซึ่งได้แก่ตัวแปรความสามารถที่สังเกตได้(คะแนนแบบทดสอบรวม) (X) และตัวแปรชี้วัดกลุ่ม (G) โมเดลการถดถอยโลจิสติกจะไม่มีสัญลักษณ์แสดงข้อสอบในโมเดล สามารถเขียนแสดงได้ดังนี้ (Miller; & Spray. 1993: 108)

$$P(U|X,G)=\frac{e^{(1-U)(-\beta_0-\beta_1X-\beta_2G-\beta_3(X*G))}}{1+e^{(1-U)(-\beta_0-\beta_1X-\beta_2G-\beta_3(X*G))}} \qquad (1)$$

เราไม่สามารถขยายโมเดลการถดถอยโลจิสติกซึ่งใช้ข้อมูลการให้คะแนนแบบสองค่าไปใช้กับข้อมูลการให้คะแนนแบบหลายค่าเพียงครั้งเดียวโดยตรง ในการขยายวิธีการถดถอยโลจิสติกไปใช้กับข้อสอบที่ให้คะแนนแบบหลายค่าข้อมูลต่างๆ จะมีการกำหนดรหัสให้ใหม่เข้าไปในกลุ่มของเซตของข้อมูลแบบสองค่า ซึ่งแต่ละค่าก็เหมาะสำหรับการวิเคราะห์การถดถอยแบบแยกกัน มีการขยายหลายครั้งของวิธีการถดถอยโลจิสติก เพื่อใช้กับข้อมูลที่มีการให้คะแนนแบบหลายค่า เฟร็นช์ และมิลเลอร์ (French; & Miller. 1996) ได้แนะนำการขยายออก 3 ครั้งของโมเดลโลจิสติกสำหรับโมเดลการตอบแบบมัลติโนเมียล (Multinomial response model) โดยใช้ข้อมูลการตอบแบบจัดอันดับ (Ordinal response data) แต่ละค่าก็จะเกี่ยวข้องกับเซตของความแตกต่างของการเปรียบเทียบรายคู่ (pairwise comparisons) ระหว่างรายการคะแนน (Score categories) หรือผลรวมของรายการคะแนน โมเดลเหล่านี้ได้แก่ โมเดลโลจิทอัตราส่วนแบบต่อเนื่อง (Continuation ratio logits) โมเดลโลจิทสะสม (Cumulative logits) และโมเดลโลจิทรายการที่อยู่ติดกัน (Adjacent categories logits) ในแต่ละวิธีเหล่านี้สมมุติว่ามีรายการตอบข้อสอบ (Item response categories) จำนวน J อันดับ (Ordered) ดังนั้นทั้งสามวิธีจะเกี่ยวข้องกับการประมวลผลข้อมูลแบบจัดอันดับเข้าไปในเซตของข้อมูลที่มีการตอบแบบสองค่าจำนวน J-1 เซต วิธีดังกล่าวตั้งอยู่บนพื้นฐานโลจิท (Logits) หรืออัตราส่วนของความน่าจะเป็นของการได้ค่าหนึ่ง ๆ ต่อความน่าจะเป็นของการได้คะแนนสูงหรือต่ำกว่าค่านั้น ๆ สมมุติว่าข้อสอบข้อหนึ่งมีรายการการตอบ 4 รายการ ก็จะมี 3 โลจิท และจะมีโค้ง

โลจิท 3 โค้ง ในแต่ละวิธีนั้นจะมีการคิดคำนวณการถดถอยที่แยกกันสำหรับแต่ละโลจิท เพื่อใช้ทดสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (ที่เป็นรูปแบบเดียวกัน , ไม่เป็นรูปแบบเดียวกัน) วิธีทั้งสามสามารถให้รายละเอียดเพิ่มเติมดังนี้

วิธีแรก โลจิทอัตราส่วนแบบต่อเนื่อง (Continuation ratio logits) วิธีนี้จะใช้จำนวนเล็กๆ ที่เพิ่มขึ้นของข้อมูลในแต่ละการถดถอยที่ต่อเนื่องกัน เพื่อที่จะทดสอบการมีอยู่ของข้อสอบที่ทำหน้าที่เบี่ยงเบน ในการถดถอยแรกความน่าจะเป็นของการได้คะแนน 0 ถูกเปรียบเทียบกับความน่าจะเป็นของการได้คะแนนมากกว่า 0 ในการถดถอยที่สองความน่าจะเป็นของการได้คะแนน 1 ถูกเปรียบเทียบกับความน่าจะเป็นของการได้คะแนนมากกว่า 1 และในการถดถอยที่สามความน่าจะเป็นของการได้คะแนน 2 ถูกเปรียบเทียบกับความน่าจะเป็นของการได้คะแนน 3 วิธีดังกล่าวสามารถแสดงง่ายๆ ดังนี้

เมื่อข้อมูลที่ใช้มีรายการตอบ 4 รายการในที่นี้ถ้าให้ k เป็นรายการการตอบ และ $k = 0,1,2,3$

ความน่าจะเป็นของการตอบรายการที่ $Y = k$ ถูกเปรียบเทียบกับความน่าจะเป็นของการตอบรายการที่ $Y > k$ ดังนั้นความน่าจะเป็นที่ $Y = 0$ จะถูกเปรียบเทียบกับความน่าจะเป็นที่ $Y = 1,2,3$ ความน่าจะเป็นที่ $Y = 1$ จะถูกเปรียบเทียบกับความน่าจะเป็นที่ $Y = 2,3$ ความน่าจะเป็นที่ $Y = 2$ จะถูกเปรียบเทียบกับความน่าจะเป็นที่ $Y = 3$

วิธีที่สอง โลจิทรายการที่อยู่ติดกัน (Adjacent categories logit) วิธีนี้ยึดหลักว่า ความน่าจะเป็นการตอบรายการใดถูกเปรียบเทียบกับความน่าจะเป็นการตอบรายการที่อยู่ติดกับมัน ในการถดถอยแรก ความน่าจะเป็นของการได้คะแนน 0 ถูกเปรียบเทียบกับความน่าจะเป็นของได้คะแนน 1 ในการถดถอยที่สองความน่าจะเป็นของการได้คะแนน 1 ถูกเปรียบเทียบกับความน่าจะเป็นของการได้คะแนน 2 และในการถดถอยที่สาม ความน่าจะเป็นของการได้คะแนน 2 ถูกเปรียบเทียบกับความน่าจะเป็นของการได้คะแนน 3 วิธีดังกล่าวสามารถแสดงง่าย ๆ ดังนี้

เมื่อข้อมูลที่ใช้มีรายการตอบ 4รายการ ในที่นี้ถ้าให้ k เป็นรายการการตอบ และ $k = 0,1,2,3$

ความน่าจะเป็นของการตอบรายการที่ $Y = k$ ถูกเปรียบเทียบกับความน่าจะเป็นของการตอบรายการที่ $Y = k+1$ ดังนั้นความน่าจะเป็นที่ $Y = 0$ จะถูกเปรียบเทียบกับความน่าจะเป็นที่ $Y = 1$ ความน่าจะเป็นที่ $Y = 1$ จะถูกเปรียบเทียบกับความน่าจะเป็นที่ $Y = 2$ ความน่าจะเป็นที่ $Y = 2$ จะถูกเปรียบเทียบกับความน่าจะเป็นที่ $Y = 3$ วิธีนี้พบว่ามีเพียงสองรายการถูกเปรียบเทียบในแต่ละการถดถอยทำให้สารสนเทศจำนวนมากหายไปจากการประมวลผลด้วยวิธีนี้

วิธีที่สาม โลจิทสะสม (Cumulative logits) วิธีนี้ยึดหลักว่า ความน่าจะเป็นของการได้คะแนนจากการตอบรายการใดหรือน้อยกว่าถูกเปรียบเทียบกับความน่าจะเป็นของการตอบรายการอื่นๆ ในการถดถอยแรกความน่าจะเป็นของการได้คะแนน 0 ถูกเปรียบเทียบกับความน่าจะเป็นของการได้คะแนนอื่นๆ ในการถดถอยที่สองความน่าจะเป็นของการได้คะแนน 1 หรือน้อยกว่าถูก

เปรียบเทียบกับความน่าจะเป็นของการได้คะแนนอื่นๆ และในการถดถอยที่สามความน่าจะเป็นของการได้คะแนน 2 หรือน้อยกว่าถูกเปรียบเทียบกับความน่าจะเป็นของการได้คะแนน 3 วิธีดังกล่าวสามารถแสดงง่ายๆ ดังนี้

เมื่อข้อมูลที่ใช้มีรายการตอบ 4 รายการในที่นี้ถ้าให้ k เป็นรายการการตอบ และ $k = 0,1,2,3$

ความน่าจะเป็นของการตอบรายการที่ $Y \leq k$ ถูกเปรียบเทียบกับความน่าจะเป็นของการตอบรายการที่ $Y > k$ ดังนั้นความน่าจะเป็นที่ $Y = 0$ จะถูกเปรียบเทียบกับความน่าจะเป็นที่ $Y = 1,2,3$ ความน่าจะเป็นที่ $Y = 0,1$ จะถูกเปรียบเทียบกับความน่าจะเป็นที่ $Y = 2,3$ และความน่าจะเป็นที่ $Y = 0,1,2$ จะถูกเปรียบเทียบกับความน่าจะเป็นที่ $Y = 3$

โมเดลโลจิทสะสมมีความสำคัญทั้งทางทฤษฎีและการประยุกต์ไปใช้มากกว่าสองโมเดลแรก เพราะไม่มีสารสนเทศใดสูญหายไปจากการให้รหัสและจากการศึกษาการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่าของเฟร็นช์และมิลเลอร์ (French; & Miller.1996)โดยใช้วิธีโลจิทอัตราส่วนแบบต่อเนื่อง และวิธีโลจิทสะสม พบว่าวิธีทั้งสองเป็นวิธีให้ผลค่อนข้างดี แต่การประมวลผลการถดถอยที่แยกกันสำหรับแต่ละโลจิทนั้น ทำให้แปรผลยากเพราะพารามิเตอร์หลายตัวต้องถูกทดสอบนัยสำคัญทางสถิติ นอกจากนี้นักวิจัยได้ปรับขยายโมเดลโลจิสติกโดยนำโมเดลการถดถอยโลจิสติกไปประยุกต์ใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่า (Polytomous) ด้วยวิธีอื่นๆ ดังเช่นมิลเลอร์และสเปรย์ (Miller; & Spray. 1993) ได้เสนอวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก (Logistic discriminant function analysis) ซึ่งเป็นวิธีที่ง่ายกว่าและใช้สะดวกกว่าวิธีที่ได้จากการขยายแบบให้คะแนนหลายค่าที่ใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ (Multinomial logistic regession: MLR) ส่วนซัมโบ (Zumbo. 1999) ได้ใช้โมเดลการถดถอยโลจิสติกแบบจัดอันดับ (Ordinal logistic regression model)โดยอาศัยวิธีโลจิทสะสม ทำให้ได้สารสนเทศครบถ้วน วิธีนี้โลจิททั้งหมดถูกทดสอบพร้อมกัน การตัดสินเกี่ยวกับการทำหน้าที่เบี่ยงเบนของข้อสอบตั้งอยู่บนการทดสอบนัยสำคัญทางสถิติ และการวัดขนาดของผล (Effect size)

แนวคิดและหลักการ

ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่าตามวิธีการถดถอยโลจิสติกแบบมัลติโนเมียล (Multinomial logistic regression: MLR) เกิดความยุ่งยากในการแปลผล มิลเลอร์และสเปรย์ (Miller; & Spray. 1993)จึงได้เสนอวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก (Logistic discriminant function analysis: LDFA) เพื่อใช้สำหรับตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่า วิธีดังกล่าวอยู่บนโมเดลพื้นฐานที่ใช้สำหรับคัดแยกคนออกเป็นกลุ่ม ๆ ที่มีส่วนเหมือนกันบนพื้นฐานของการตอบข้อสอบและคะแนนแบบทดสอบ วิธีนี้ไม่ต้องการสมมติฐานของความเป็นปกติ (Normality)คือตัวแปรอิสระของวิธีนี้ไม่จำเป็นต้องมี

การแจกแจงปกติ วิธีนี้ง่ายกว่าวิธีการถดถอยโลจิสติกแบบมัลติโนเมียล (MLR) เนื่องจากว่ามีเพียงแค่การถดถอยเดียวที่จำเป็นต้องใช้ในการทดสอบแต่ละโมเดล วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความเกี่ยวข้องสัมพันธ์อย่างใกล้ชิดกับวิธีการถดถอยโลจิสติกแบบมัลติโนเมียล และการวิเคราะห์ล็อก-ลิเนียร์ (Log-linear analysis) สำหรับความแตกต่างกันที่เห็นได้ชัดนั้นโมเดลการถดถอยในวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก ใช้เพื่อทำนายการเป็นสมาชิกกลุ่ม ส่วนโมเดลการถดถอยในวิธีการถดถอยโลจิสติกแบบมัลติโนเมียล จะใช้เพื่อทำนายคะแนนสอบสามารถเขียนโมเดลการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกได้ดังนี้

$$P(G|X,U)=\frac{e^{(1-G)(-\alpha_0-\alpha_1X-\alpha_2U-\alpha_3(X*U))}}{1+e^{(1-G)(-\alpha_0-\alpha_1X-\alpha_2U-\alpha_3(X*U))}} \qquad (2)$$

เมื่อ $\alpha_i (i=0,1,2,3)$ แทนสัมประสิทธิ์ฟังก์ชันการจำแนก

G แทน กลุ่มผู้สอบ

X แทน คะแนนรวมแบบทดสอบ

U แทน ผลการตอบข้อสอบ

P แทน ความน่าจะเป็นของการเป็นสมาชิกกลุ่ม

โมเดลนี้โดยพื้นฐานก็เหมือนกับโมเดลที่ให้ไว้ในสมการ (1) อย่างไรก็ตาม มีความแตกต่างกันบางประการกล่าวคือ **ประการแรก** การเป็นสมาชิกกลุ่ม (G) ในโมเดลจะแทนที่ของการตอบข้อสอบ (U) และตัวแปรกลุ่ม (G) จะเป็นตัวแปรชี้วัด (Indicator variable) กำหนดว่า $G=1$ สำหรับกลุ่มอ้างอิง และ $G=0$ สำหรับกลุ่มสนใจ **ประการที่สอง**การตอบข้อสอบ (U) ไม่จำเป็นต้องมีเพียง 2 รายการเท่านั้น แต่สามารถรับเอาไว้บนการตอบข้อสอบที่มีจำนวนของรายการใดก็ได้ **ประการที่สาม** ตัวแปรการตอบข้อสอบ (U) เป็นตัวทำนาย (Predictor) มากกว่าที่จะเป็นตัวแปรตาม และ**ประการที่สี่** สัมประสิทธิ์การถดถอยที่ให้ไว้โดย α_i เมื่อ $i=0,1,2,3$ เพื่อให้เห็นว่าแตกต่างจาก β ในโมเดลสมการ (1)

กระบวนการตรวจสอบ

ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ โดยใช้วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก (LDFA) ของมิลเลอร์และสเปรย์ (Miller; & Spray.1993) จะใช้กระบวนการตรวจสอบความเหมาะสมของข้อมูลกับโมเดล โดยจะทดสอบโมเดล 3 ขั้นตอนดังนี้

ขั้นตอนที่ 1 ทดสอบตัวแปรที่ใช้ในการจับคู่ความสามารถ (Matching variable) หรือตัวแปรสำรวจ (Explanatory variable) ในที่นี้คือคะแนนรวมแบบทดสอบ (X) ซึ่งเป็นระดับความสามารถของผู้สอบ ในขั้นตอนนี้ถือว่าอิทธิพลของตัวแปรคะแนนแบบทดสอบรวมมีค่าเท่ากันในแต่

ละกลุ่ม (G) ซึ่งกำหนดรหัสตัวแปรดัมมี (Dummy variable) โดยให้กลุ่มอ้างอิงแทนด้วย 1 และกลุ่มสนใจแทนด้วย 0 การนำตัวแปรเข้าไปในสมการถดถอยมีลักษณะดังนี้

$$P(G|X,U) = P(G,X) = \frac{e^{(1-G)(-\alpha_0-\alpha_1 X)}}{1+e^{(1-G)(-\alpha_0-\alpha_1 X)}} \quad (3)$$

สมการดังกล่าวจะทำนายถึงการเป็นสมาชิกกลุ่มจากคะแนนแบบทดสอบรวมเท่านั้น เมื่อเขียนสมการ (3) ในรูปโลจิทจะได้

$$\text{logit} = \ln\left[\frac{P(G=1)}{P(G=0)}\right] = \alpha_0 + \alpha_1 X \quad (4)$$

ขั้นตอนที่ 2 ทดสอบตัวแปรการตอบข้อสอบที่มีรายการคำตอบใดๆ (U) ซึ่งไม่จำเป็นต้องมีเพียง 2 รายการคำตอบ โดยเพิ่มตัวแปรดังกล่าวในโมเดลสมการ (3) ของขั้นตอนที่ 1 ดังนี้

$$P(G|X,U) = \frac{e^{(1-G)(-\alpha_0-\alpha_1 X-\alpha_2 U)}}{1+e^{(1-G)(-\alpha_0-\alpha_1 X-\alpha_2 U)}} \quad (5)$$

สมการดังกล่าวจะทำนายการเป็นสมาชิกกลุ่มจากคะแนนแบบทดสอบรวม และคะแนนข้อสอบ เมื่อเขียนสมการ (5) ในรูปโลจิท จะได้

$$\text{logit} = \ln\left[\frac{P(G=1)}{P(G=0)}\right] = \alpha_0 + \alpha_1 X + \alpha_2 U \quad (6)$$

ขั้นตอนที่ 3 ทดสอบตัวแปรปฏิสัมพันธ์ระหว่างคะแนนแบบทดสอบรวม (ระดับความสามารถ) กับการตอบข้อสอบ $(X * U)$ โดยเพิ่มตัวแปรดังกล่าวในโมเดลสมการ (5) ของขั้นตอนที่ 2 ดังนี้

$$P(G|X,U) = \frac{e^{(1-G)(-\alpha_0-\alpha_1 X-\alpha_2 U-\alpha_3 (X*U))}}{1+e^{(1-G)(-\alpha_0-\alpha_1 X-\alpha_2 U-\alpha_3 (X*U))}} \quad (7)$$

สมการดังกล่าวจะทำนายการเป็นสมาชิกกลุ่มจากคะแนนแบบทดสอบรวม,คะแนนข้อสอบ และปฏิสัมพันธ์ระหว่างคะแนนแบบทดสอบรวมกับคะแนนข้อสอบ เมื่อเขียนสมการ (7) ในรูปโลจิทจะได้

$$\text{logit} = \ln\left[\frac{P(G=1)}{P(G=0)}\right] = \alpha_0 + \alpha_1 X + \alpha_2 U + \alpha_3 (X * U) \quad (8)$$

การทดสอบสมมุติฐาน

การทดสอบการทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF) และที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) สามารถทดสอบได้พร้อมกัน โดยกำหนดสมมุติฐานศูนย์ (H_0) และสมมุติฐานทางเลือก (H_1) ดังนี้

$$H_0 : \alpha_2 = 0 \text{ และ } \alpha_3 = 0$$
$$H_1 : \alpha_2 \neq 0 \text{ หรือ } \alpha_3 \neq 0 \qquad (9)$$

การทดสอบนัยสำคัญทางสถิติของการทำหน้าที่เบี่ยงเบนของข้อสอบจะดำเนินการตามกระบวนการการทดสอบความเหมาะสมของข้อมูลกับโมเดล 3 ขั้นตอน โดยการนำตัวแปรเข้าทดสอบกับโมเดลตามลำดับขั้นตอนกล่าวคือ ขั้นตอนที่ 1 ใส่ตัวแปรสำรวจ (Explanatory variable) หรือตัวแปรคะแนนแบบทดสอบรวม (X) เข้าไปในสมการ ขั้นตอนที่ 2 ใส่ตัวแปรการตอบข้อสอบ (U) เข้าไปในสมการ และขั้นตอนที่ 3 ใส่ตัวแปรปฏิสัมพันธ์ระหว่างคะแนนแบบทดสอบรวมกับตัวแปรการตอบข้อสอบ $(X*U)$ เข้าไปในสมการ สถิติไค-สแควร์ของอัตราส่วนความน่าจะเป็น (G^2) ถูกคำนวณออกมาในแต่ละโมเดล ความแตกต่างในค่า G^2 ระหว่างโมเดลลำดับขั้น (Hierarchical model) มีการกระจายแบบเชิงเส้นกำกับ (Asymptotically) เป็นไค-สแควร์ ที่มีระดับชั้นความเป็นอิสระเท่ากับ 1 $(df = 1)$ สถิติการทดสอบดังกล่าวจะเป็น G^2_{diff} โดยที่

$$G^2_{diff} = -2(\ln L_{(i+1)} - \ln L_{(i)}) \qquad (10)$$

เมื่อ $L_{(i+1)}$ และ $L_{(i)}$ แทนความน่าจะเป็นของโมเดลลำดับขั้นที่อยู่ติดกัน ดังนั้นในการทดสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกัน และที่ไม่เป็นรูปแบบเดียวกัน ถ้าค่า G^2_{diff} ที่ได้จากความแตกต่างของ G^2 จากโมเดลในขั้นตอนที่ 3 และโมเดลในขั้นตอนที่ 1 สามารถเปรียบเทียบกับค่า G^2 ที่ระดับชั้นความเป็นอิสระเท่ากับ 2 ถ้าผลการทดสอบมีนัยสำคัญทางสถิติโดยปฏิเสธสมมติฐานศูนย์ นั่นแสดงว่าข้อสอบทำหน้าที่เบี่ยงเบนซึ่งเป็นผลมาจากอิทธิพลของการตอบข้อสอบ (คะแนนข้อสอบ) และ/หรืออิทธิพลของปฏิสัมพันธ์ระหว่างคะแนนแบบทดสอบรวมกับคะแนนข้อสอบ สำหรับการทดสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกัน ถ้าค่า G^2_{diff} ที่ได้จากความแตกต่างของ G^2 จากโมเดลในขั้นตอนที่ 2 และโมเดลในขั้นตอนที่ 1 สามารถเปรียบเทียบกับค่า G^2 ที่ระดับชั้นความเป็นอิสระเท่ากับ 1 $(df = 1)$ ถ้าผลการทดสอบมีนัยสำคัญทางสถิติแสดงว่าข้อสอบข้อนั้นทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน ในการทดสอบนัยสำคัญทางสถิติ ถ้าการทดสอบสำหรับการทำหน้าที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกันหรือที่ไม่เป็นรูปแบบเดียวกัน มีนัยสำคัญทางสถิติ หมายความว่ามีอย่างน้อย 1 รายการคำตอบ ในการที่จะเป็นสมาชิกกลุ่ม ซึ่งได้ให้คะแนนข้อสอบและคะแนนแบบทดสอบรวม แตกต่างจากรายการซึ่งถูกทำนายจากคะแนนรวมเพียงอย่างเดียว

3.5 วิธีการถดถอยโลจิสติกแบบจัดอันดับ (Ordinal logistic regression)

ซัมโบ (Zumbo.1999) ได้ขยายโมเดลการถดถอยโลจิสติกที่ใช้กับข้อสอบที่มีการให้คะแนนแบบสองค่าของสวามินาธานและโรเจอร์ส (Swaminathan; & Rogers.1990) และปรับปรุงโมเดลการถดถอยโลจิสติกที่ใช้กับข้อสอบที่มีการให้คะแนนแบบหลายค่าซึ่งได้แก่โมเดลโลจิทของอัตราส่วนแบบต่อเนื่อง (Continuation ratio logits) โมเดลโลจิทของรายการที่อยู่ติดกัน (Adjacent category logits) และโมเดลโลจิทสะสม (Cumulative logits) ของเฟร็นช์และมิลเลอร์ (French; & Miller.1996) โดยที่แต่ละโมเดลค่าการถดถอยคิดคำนวณแบบแยกกันสำหรับแต่ละโลจิท ทำให้ยากต่อการระบุการทำหน้าที่ต่างกันของข้อสอบ ซัมโบ (Zumbo. 1999) จึงได้นำโมเดลโลจิทสะสมมาพัฒนาใช้กับข้อสอบแบบจัดอันดับ โดยทำให้โลจิททั้งหมดสามารถทดสอบได้พร้อมกัน และเรียกวิธีที่พัฒนาขึ้นใหม่นี้ว่า **วิธีการถดถอยโลจิสติกแบบจัดอันดับ** (Ordinal logistic regression: OLR)

แนวคิดและหลักการ

โมเดลการถดถอยโลจิสติกมีพื้นฐานบนโมเดลทางสถิติของความน่าจะเป็นในการตอบข้อสอบหนึ่งข้อที่ให้คะแนนแบบสองค่าได้ถูกต้อง วิธีนี้จะใช้การตอบข้อสอบ (0 หรือ 1) เป็นตัวแปรตาม (Dependent variable) ที่มีตัวแปรกลุ่ม (ซึ่งกำหนดรหัสดัมมีเป็น 1 สำหรับกลุ่มอ้างอิงและ 0 สำหรับกลุ่มสนใจ ในที่นี้เป็นตัวแปร Group) คะแนนรวมสำหรับผู้เข้าสอบแต่ละคน (อาจเป็นความสามารถหรือคุณลักษณะ ในที่นี้เป็นตัวแปร TOT (Total test score) และปฏิสัมพันธ์ระหว่างคะแนนรวมกับกลุ่มผู้สอบ (TOT*GROUP) เป็นตัวแปรอิสระ (Independent variable) วิธีนี้ได้ให้การทดสอบการทำหน้าที่เบี่ยงเบนของข้อสอบโดยมีเงื่อนไขบนความสัมพันธ์ระหว่างการตอบข้อสอบและคะแนนรวม การทดสอบผลของกลุ่มผู้สอบเพื่อตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกัน (Uniform DIF) และการทดสอบปฏิสัมพันธ์ของคะแนนรวมและกลุ่มผู้สอบ เพื่อตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF)

สมการการถดถอยโลจิสติกดังกล่าวเป็นดังนี้

$$Y = b_0 + b_1 TOT + b_2 GROUP + b_3 TOT * GROUP \qquad (1)$$

เมื่อ Y เป็นลอการิทึมธรรมชาติของอัตราส่วนแต้มต่อ (Odds ratio) สมการ (1) เขียนใหม่ได้ดังนี้

$$\ln\left[\frac{P}{(1-P)}\right] = b_0 + b_1 tot + b_2 group + b_3 (tot * group) \qquad (2)$$

เมื่อ P แทนสัดส่วนของผู้สอบที่ตอบข้อสอบถูก

ขณะเดียวกันสามารถที่จะแสดงการถดถอยโลจิสติก (Logistic regression) ในรูปการถดถอยเชิงเส้น (Linear regression) ของตัวแปรตัวทำนาย (Predictor variable) บนตัวแปรสุ่มที่มีการแจกแจงแบบต่อเนื่องที่ไม่สามารถสังเกตได้ (Unobservable continuously distributed random variable) ดังนั้นสามารถเขียนแสดงสมการ (1) ได้ดังนี้

$$y^* = b_0 + b_1 TOT + b_2 GROUP + b_3 TOT * GROUP + \varepsilon_i \quad (3)$$

เมื่อ y^* แทน ตัวแปรตามแบบต่อเนื่องที่ไม่สามารถสังเกตได้

ε_i แทน ความคลาดเคลื่อน ซึ่งการแจกแจงของ ε_i มีค่าเฉลี่ยเป็น 0 และความแปรปรวนเป็น $\frac{\pi^2}{3}$

ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่า โดยใช้โมเดลการถดถอยโลจิสติกแบบจัดอันดับตามแนวคิดของซัมโบ จะประมาณค่าความน่าจะเป็นสะสม (Cumulative probability) ในรายการคำตอบที่ 1 ถึง J-1 เมื่อ J เป็นรายการคำตอบในมาตรแบบจัดอันดับ โดยที่ความน่าจะเป็นสะสมในรายการคำตอบสุดท้ายมีค่าเท่ากับ 1 ตัวอย่างเช่นข้อสอบมีรายการคำตอบ 3 รายการ โมเดลการถดถอยจะประมาณค่าความน่าจะเป็นสะสมในรายการที่ 1 ถึง 2 คือความน่าจะเป็นสะสมสำหรับการตอบในรายการที่ 1 หรือต่ำกว่า และความน่าจะเป็นสะสมสำหรับการตอบในรายการที่ 2 หรือต่ำกว่า โดยที่ความน่าจะเป็นสะสมในรายการที่ 3 ซึ่งเป็นรายการสุดท้ายมีค่าเท่ากับ 1 ดังนั้นสามารถเขียนสมการแสดงการถดถอยโลจิสติกแบบจัดอันดับในรูปลอการิทึมธรรมชาติของอัตราส่วนแต้มต่อ (Odds ratio) ดังนี้

$$\ln\left[\frac{P(Y \le 1)}{P(Y > 1)}\right] = \alpha_j + bX \quad (4)$$

โดยที่ $$\ln\left[\frac{P(Y \le j)}{P(Y > j)}\right] = \text{logit}[P(Y \le j)] \quad (5)$$

จะได้ว่า $$\text{logit}[P(Y \le j)] = \alpha_j + bX \quad (6)$$

เมื่อ $P(Y \le j)$ แทนความน่าจะเป็นสะสมสำหรับผู้ตอบในรายการที่ j หรือต่ำกว่า

α_j แทน พารามิเตอร์ส่วนตัด(Intercept parameter) ของความน่าจะเป็นสะสมในแต่ละรายการ

b แทน พารามิเตอร์ที่เกี่ยวข้องกับระดับความสามารถของผู้สอบ

X แทน ระดับความสามารถของผู้สอบ

โมเดลสมการ (6) เป็นสมการซึ่งไม่ปรากฏการทำหน้าที่เบี่ยงเบนของข้อสอบ (Null DIF) โมเดลนี้ทำนายคะแนนข้อสอบจากระดับความสามารถ (คะแนนรวม) ผู้สอบเท่านั้น เมื่อพิจารณาโมเดลในสมการ (6) ทำให้สามารถเขียนสมการที่เหมาะสมสำหรับการวิเคราะห์การทำหน้าที่เบี่ยงเบนของข้อสอบสำหรับสมการ (3) ดังนี้

$$\text{logit}[P(Y \le j)] = \alpha_j + b_1 tot + b_2 group + b_3 (tot * group) \quad (7)$$

หรือ $\text{logit}[P(Y \le j)] = \alpha_j + b_1 X + b_2 G + b_3 (X*G)$ (8)

เมื่อ $P(Y \le j)$ แทนความน่าจะเป็นสะสมสำหรับผู้ตอบในรายการที่ j หรือต่ำกว่า

α_j แทน พารามิเตอร์ส่วนตัดของความน่าจะเป็นสะสมในรายการที่ j

X แทน ระดับความสามารถของผู้สอบ

G แทน กลุ่มผู้สอบ

b_1 แทน พารามิเตอร์ที่เกี่ยวข้องกับระดับความสามารถของผู้สอบ (X)

b_2 แทน พารามิเตอร์ที่เกี่ยวข้องกับกลุ่มผู้สอบ (G)

b_3 แทน พารามิเตอร์ที่เกี่ยวข้องกับปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับกลุ่มผู้สอบ $(X*G)$

โมเดลสมการ (8) เป็นโมเดลเต็ม (Full model) ซึ่งประกอบด้วยการทำหน้าที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกัน (Uniform DIF) และการทำหน้าที่เบี่ยงเบนของข้อสอบที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) ในโมเดลเต็ม คะแนนข้อสอบถูกทำนายจากระดับความสามารถของผู้สอบ (คะแนนรวม) การเป็นสมาชิกกลุ่มและปฏิสัมพันธ์ของระดับความสามารถกับการเป็นสมาชิกกลุ่มผู้สอบ

โมเดลสำหรับการทำหน้าที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกัน (Uniform DIF) เป็นดังนี้

$$\text{logit}[P(Y \le j)] = \alpha_j + b_1 X + b_2 G \quad (9)$$

โมเดลสมการ (9) เป็นโมเดลสำหรับตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกัน โมเดลดังกล่าวคะแนนข้อสอบถูกทำนายจากระดับความสามารถของผู้สอบ (คะแนนรวม) และการเป็นสมาชิกกลุ่มผู้สอบ

กระบวนการตรวจสอบ

ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบโดยใช้วิธีการถดถอยโลจิสติกแบบจัดอันดับ (OLR) จะใช้กระบวนการตรวจสอบความเหมาะสมของข้อมูลกับโมเดลโดยจะทดสอบโมเดล 3 ขั้นตอนดังนี้

ขั้นตอนที่หนึ่ง ทดสอบตัวแปรที่ใช้ในการจับคู่ความสามารถ (Matching variable) หรือตัวแปรเงื่อนไข (Conditioning variable) ในที่นี้ คือ คะแนนซึ่งเป็นระดับความสามารถของผู้สอบ (X) ในขั้นตอนนี้ถือว่าอิทธิพลของตัวแปรคะแนนรวมมีค่าเท่ากันในทุกรายการตอบ และสมมติให้ค่าประมาณของพารามิเตอร์ความชันมีเพียงค่าเดียว การนำตัวแปรดังกล่าวเข้าไปในสมการถดถอยมีลักษณะดังนี้

$$\text{logit}[P(Y \le j)] = \alpha_j + b_1 X \quad (10)$$

ขั้นตอนที่สอง ทดสอบตัวแปรสมาชิกกลุ่มของผู้สอบ (G) ในที่นี้จะกำหนดรหัสเป็นตัวแปรดัมมี (Dummy variable) โดยให้กลุ่มอ้างอิงแทนด้วย 1 และกลุ่มสนใจแทนด้วย 0 แล้วเพิ่มตัวแปรดังกล่าวในสมการ (10) ของขั้นตอนที่หนึ่ง ดังนี้

$$\text{logit}[P(Y \le j)] = \alpha_j + b_1 X + b_2 G \qquad (11)$$

ขั้นตอนที่สาม ทดสอบตัวแปรปฏิสัมพันธ์ระหว่างระดับความสามารถ กับสมาชิกของกลุ่มผู้สอบ (X*G) โดยเพิ่มพจน์ทั้งสองดังกล่าวในโมเดลสมการ (11) ของขั้นตอนที่หนึ่งดังนี้

$$\text{logit}[P(Y \le j)] = \alpha_j + b_1 X + b_2 G + b_3 (X*G) \qquad (12)$$

การตรวจสอบสมมติฐาน

การทดสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกันและที่ไม่เป็นรูปแบบเดียวกัน สามารถทดสอบได้พร้อมกัน โดยกำหนดสมมติฐานศูนย์ (H_0) และสมมติฐานทางเลือก (H_1) ดังนี้

$$H_0 : b_2 = 0 \text{ และ } b_3 = 0$$
$$H_1 : b_2 \neq 0 \text{ หรือ } b_3 \neq 0$$

การทดสอบนัยสำคัญทางสถิติของการทำหน้าที่เบี่ยงเบนของข้อสอบจะดำเนินการตามกระบวนการการทดสอบความเหมาะสมของข้อมูลกับโมเดล 3 ขั้นตอนโดยการนำตัวแปรเข้าทดสอบกับโมเดลตามลำดับขั้นดังได้กล่าวมาแล้ว [ขั้นตอนที่ 1 ใส่ตัวแปรเงื่อนไข (คะแนนรวม; X) เข้าไปในสมการ ขั้นตอนที่ 2 ใส่ตัวแปรกลุ่ม (G) เข้าไปในสมการของขั้นตอนที่ 1 และ ขั้นตอนที่ 3 ใส่ตัวแปรปฏิสัมพันธ์ระหว่างระดับความสามารถ (คะแนนรวม) กับสมาชิกกลุ่มผู้สอบ (X*G) เข้าไปในสมการของขั้นตอนที่ 2]สารสนเทศดังกล่าวและการทดสอบไค-สแควร์สำหรับการถดถอยโลจิสติกหนึ่งการถดถอยนั้นสามารถคำนวณการทดสอบทางสถิติของการทำหน้าที่เบี่ยงเบนของข้อสอบซึ่งในการตัดสินว่าข้อสอบข้อใดทำหน้าที่เบี่ยงเบนหรือไม่ พิจารณาได้จากความแตกต่างระหว่างค่าไค-สแควร์ที่ได้รับมานั้นจากขั้นตอนที่ 3 และขั้นตอนที่ 1 ผลของค่าไค-สแควร์ที่ได้สามารถเปรียบเทียบกับฟังก์ชันการแจกแจงแบบไค-สแควร์ที่มีระดับชั้นความเป็นอิสระเท่ากับ 2 ($df = 2$) โดยมีค่า $p \le .01$ ซึ่งค่าของ df ดังกล่าวเกิดจากความแตกต่างของ df ของขั้นตอนทั้งสองที่เป็นเช่นนี้เป็นเพราะสถิติไค-สแควร์ของโมเดล ในขั้นตอนที่ 3 มีระดับชั้นความเป็นอิสระเท่ากับ 3 ($df = 3$) และสถิติไค-สแควร์ของโมเดลขั้นตอนที่ 1 มีระดับชั้นความเป็นอิสระเท่ากับ 1 ($df = 1$) ผลจากการทดสอบค่าไค-สแควร์ที่ระดับ $df = 2$ เป็นการทดสอบพร้อมกันของการทำหน้าที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกัน (Uniform DIF) และการทำหน้าที่เบี่ยงเบนของข้อสอบที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) ถ้าผลการทดสอบมีนัยสำคัญทางสถิติ โดยปฏิเสธสมมติฐานศูนย์ นั่นแสดงว่าข้อสอบทำหน้าที่เบี่ยงเบนซึ่งเป็นผลมาจากอิทธิพลของการเป็นสมาชิกของกลุ่ม

ผู้สอบ และ/หรืออิทธิพลของปฏิสัมพันธ์ระหว่างการเป็นสมาชิกของกลุ่มกับระดับความสามารถ (Zumbo.1999: 26)

สวามินาธาน และโรเจอร์ส (Swaminathan; & Rogers. 1990) มุ่งเน้นที่ความสนใจบนการทดสอบไค-สแควร์ที่มีระดับ $df = 2$ ดังนั้นขนาดของผล (Effect size) ที่ได้จะเป็น R^2 (มาจาก R^2 ของขั้นตอนที่ 3 ลบด้วย R^2 ของขั้นตอนที่ 1) ซึ่งเป็นผลมาจากทั้งอิทธิพลของการเป็นสมาชิกของกลุ่ม และ/หรืออิทธิพลของปฏิสัมพันธ์ระหว่างการเป็นสมาชิกของกลุ่มกับระดับความสามารถของผู้สอบพร้อมกัน ถือเป็นการวัดการทำหน้าที่เบี่ยงเบนของข้อสอบทั้งที่เป็นรูปแบบเดียวกัน และที่ไม่เป็นรูปแบบเดียวกันนอกจากนี้วิธีการดังกล่าวทำให้เราสามารถเปรียบเทียบค่า R^2 ขั้นตอนที่ 1 กับค่า R^2 ขั้นตอนที่ 2 ซึ่งเป็นผลมาจากอิทธิพลของกลุ่ม (ความแตกต่างกลุ่ม) และเปรียบเทียบค่า R^2 ขั้นตอนที่ 2 กับค่า R^2 ขั้นตอนที่ 3 ซึ่งเป็นผลมาจากอิทธิพลของปฏิสัมพันธ์กลุ่มและความสามารถซึ่งวิธีการดังกล่าวสามารถนำไปใช้ในการตัดสินว่าข้อสอบข้อใดทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน ข้อใดทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน โดยพิจารณาจากค่าของ R^2 ที่คิดคำนวณได้ระหว่างช่วงแรก(ขั้นตอนที่ 1 กับ 2) และช่วงหลัง (ขั้นตอนที่ 2 กับ 3) ถ้าค่าความแตกต่าง R^2 ในช่วงแรกมากกว่าค่าความแตกต่าง R^2 ในช่วงหลังสามารถสรุปได้ว่าข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน แต่ถ้าค่าความแตกต่าง R^2 ในช่วงแรกน้อยกว่าค่าความแตกต่าง R^2 ในช่วงหลัง สรุปได้ว่าข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Zumbo. 1999: 26,30-31)

4. การวิเคราะห์อัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1

ในการวิเคราะห์อัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ในสถานการณ์ที่ใช้ข้อมูลจริง เราดำเนินการโดยกำหนดวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่มีประสิทธิภาพขึ้นมาหนึ่งวิธี โดยถือว่าผลการวิเคราะห์การทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีดังกล่าวมีความถูกต้อง และจะใช้เป็นเกณฑ์ในการวิเคราะห์อัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ เช่น ถ้าผู้วิจัยต้องการวิเคราะห์อัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีซิบเทสท์ และวิธีการถดถอยโลจิสติก ถ้าให้วิธีการวัดพื้นที่ของราชูเป็นเกณฑ์ในการตรวจสอบดังนั้นผู้วิจัยต้องนำผลการวิเคราะห์การทำหน้าที่เบี่ยงเบนของข้อสอบวิธีซิบเทสท์ และวิธีการถดถอยโลจิสติกมาเทียบกับวิธีการวัดพื้นที่ของราชู โดยถือว่าผลการวิเคราะห์การทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีการวัดพื้นที่ของราชูเป็นเกณฑ์ที่ถูกต้อง กล่าวคือ ถ้าวิธีซิบเทสท์ และวิธีการถดถอยโลจิสติก ระบุข้อสอบที่ทำหน้าที่เบี่ยงเบนได้ตรงกับข้อสอบที่ถูกระบุด้วยวิธีการวัดพื้นที่ของราชู แสดงว่าวิธีดังกล่าวสามารถตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบได้ถูกต้อง และถ้าวิธีซิบเทสท์ และวิธีการถดถอยโลจิสติกระบุข้อสอบที่ทำหน้าที่เบี่ยงเบน ไม่ตรงกับข้อสอบที่ระบุด้วย

วิธีการวัดพื้นที่ของราชู แสดงว่าวิธีดังกล่าวตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบไม่ถูกต้อง สำหรับการศึกษาในสถานการณ์ที่ใช้ข้อมูลจำลอง เราสามารถกำหนดว่าจะให้ข้อสอบข้อใดบ้างทำหน้าที่เบี่ยงเบน จึงทำให้ผู้วิจัยทราบล่วงหน้าว่าข้อสอบข้อใดบ้างเป็นข้อสอบที่ลำเอียง จึงไม่จำเป็นต้องนำวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบวิธีใดวิธีหนึ่งมาเป็นเกณฑ์ในการตรวจสอบ ฉะนั้นในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีใดก็ตาม ถ้าวิธีดังกล่าวระบุข้อสอบที่ทำหน้าที่เบี่ยงเบนได้ตรงกับข้อสอบที่ทราบว่าทำหน้าที่เบี่ยงเบนอยู่ก่อนแล้ว (True positive: TP) แสดงว่าวิธีดังกล่าวตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบได้ถูกต้อง เราสามารถคำนวณหาอัตราความถูกต้อง (Power rate) ของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ได้จากสูตร

$$P = \frac{n_1}{N_1}$$

เมื่อ P แทน อัตราความถูกต้อง

n_1 แทน จำนวนข้อสอบที่ตรวจสอบได้ถูกต้องว่าทำหน้าที่เบี่ยงเบน

N_1 แทน จำนวนข้อสอบทำหน้าที่เบี่ยงเบนทั้งหมดในแบบทดสอบ

แต่ถ้าระบุข้อสอบที่ทำหน้าที่เบี่ยงเบนไม่ตรงกับข้อสอบที่ทราบว่าทำหน้าที่เบี่ยงเบนอยู่ก่อนแล้ว โดยระบุว่าข้อสอบทำหน้าที่เบี่ยงเบนทั้งที่ความเป็นจริงข้อสอบทำหน้าที่ไม่เบี่ยงเบน (False positive: FP) แสดงว่าวิธีดังกล่าวตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบไม่ถูกต้อง ทำให้เกิดความคลาดเคลื่อนประเภทที่ 1 (Type I error) สามารถคำนวณหาอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ได้จากสูตร

$$E = \frac{n_2}{N_2}$$

E แทน อัตราความคลาดเคลื่อนประเภทที่ 1

n_2 แทน จำนวนข้อสอบที่ระบุผิดพลาดว่าทำหน้าที่เบี่ยงเบน

N_2 แทน จำนวนข้อสอบทำหน้าที่ไม่เบี่ยงเบนทั้งหมดในแบบทดสอบ

5. งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีโพลีโตมัสซิบเทสท์ (Poly-SIB) วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก (LDFA) และวิธีการถดถอยโลจิสติกแบบจัดอันดับ (OLR) ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถมิติเดียวและให้คะแนนแบบหลายค่า

งานวิจัยต่างประเทศ

สวามินาธานและโรเจอร์ส (Swaminathan; & Rogers.1990: 361-370) ได้ศึกษาเปรียบเทียบวิธีการถดถอยโลจิสติกและวิธีแมนเทล-แฮนส์เซลในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบภายใต้เงื่อนไขของปัจจัย 3 ปัจจัย คือ ขนาดตัวอย่าง ความยาวแบบทดสอบ และธรรมชาติของการทำหน้าที่เบี่ยงเบนของข้อสอบ (การทำหน้าที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกัน และการทำหน้าที่เบี่ยงเบนของข้อสอบที่ไม่เป็นรูปแบบเดียวกัน) โดยจำลองข้อมูลภายใต้ขนาดกลุ่มตัวอย่าง 2 ขนาด (250 และ 500 คน) ความยาวแบบทดสอบ 3 ขนาด (40,60 และ 80 ข้อ) ซึ่งในแบบทดสอบแต่ละชุดประกอบด้วยสัดส่วนของข้อสอบที่ทำหน้าที่เบี่ยงเบนจำนวนร้อยละ 20 โดยครึ่งหนึ่งเป็นข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และอีกครึ่งหนึ่งเป็นข้อสอบที่ทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน สำหรับผลการตอบข้อสอบทั้งหมดจำลองโดยใช้โปรแกรม DATAGEN ของแฮมเบิลตัน และโรวิเนลลี (Swaminathan; & Rogers. 1990: 366) ซึ่งเป็นโมเดลการตอบข้อสอบแบบ 3 พารามิเตอร์ ในการจำลองข้อสอบที่ทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกันจะกำหนดให้พารามิเตอร์ของอำนาจจำแนกระหว่างผู้สอบสองกลุ่มมีค่าเท่ากัน ในขณะที่พารามิเตอร์ความยากจะมีค่าแปรเปลี่ยน สำหรับการจำลองข้อสอบที่ทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน จะกำหนดให้พารามิเตอร์ความยากระหว่างผู้สอบสองกลุ่มมีค่าเท่ากัน โดยกำหนดให้มีค่าเป็น 0 ส่วนพารามิเตอร์อำนาจจำแนกจะมีค่าแปรเปลี่ยน ในการควบคุมขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบ (amount of DIF) จะใช้พื้นที่ระหว่าง โค้งคุณลักษณะข้อสอบ (ICCs) ซึ่งคำนวณโดยใช้สูตรของราชู (Raju) ในการจำลองรูปแบบของการทำหน้าที่เบี่ยงเบน (Uniform DIF และ Nonuniform DIF) จะเลือกใช้พื้นที่เป็น .6 และ .8

ผลการศึกษาพบว่าการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกันด้วยวิธีการถดถอยโลจิสติกและวิธีแมนเทล-แฮนส์เซลมีอัตราความถูกต้องเท่ากัน ทั้งสองวิธีสามารถตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกันด้วยความถูกต้องประมาณ 75% ในกลุ่มตัวอย่างขนาด 250 คน และด้วยความถูกต้อง 100% ในกลุ่มตัวอย่างขนาด 500 คน ส่วนการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ไม่เป็นรูปแบบเดียวกัน วิธีการถดถอยโลจิสติกมีอัตราความถูกต้องสูงกว่าวิธีแมนเทล-แฮนส์เซล โดยที่วิธีการถดถอยโลจิสติกสามารถตรวจพบข้อสอบที่ทำหน้าที่เบี่ยงเบนมีความถูกต้องแม่นยำประมาณ 50% ในตัวอย่างขนาดเล็กและแบบทดสอบขนาดสั้น และมีความถูกต้องแม่นยำ 75% ในตัวอย่างขนาดใหญ่และแบบทดสอบขนาดยาว ส่วนวิธี

แมนเทล-แฮนส์เซลให้ผลการตรวจสอบพบข้อสอบทำหน้าที่เบี่ยงเบนภายใต้เงื่อนไขต่างๆ ค่อนข้างต่ำ สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีแมนเทล-แฮนส์เซลให้ผลค่อนข้างจะดีกว่าวิธีการถดถอยโลจิสติกที่ระดับนัยสำคัญ .01 กล่าวคือวิธีแมนเทล-แฮนส์เซลมีอัตราความคลาดเคลื่อนประเภทที่ 1 ประมาณร้อยละ 1 ภายใต้เงื่อนไขทั้งหมด ขณะที่วิธีการถดถอยโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างร้อยละ 1 ถึงร้อยละ 6

มิลเลอร์และสเปรย์ (Miller; & Spray. 1993: 107-122) ได้ปรับขยายวิธีการถดถอยโลจิสติก ที่ให้คะแนนแบบสองค่า (LR) และวิธีการทางโลจิท (Continuation ratio logits, Adjacent categories logits และ Cumulative logits) เพื่อใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่า วิธีดังกล่าวเรียกว่า “วิธีการวิเคราะห์ฟังก์ชันการจำแนก” (Logistic discriminant function analysis: LDFA) โดยนำไปเปรียบเทียบกับวิธีแมนเทล-แฮนส์เซล(Mantel-Haenszel) กรณีข้อสอบมีการให้คะแนนแบบสองค่า และวิธีแมนเทล (Mantel) กรณีข้อสอบมีการให้คะแนนแบบหลายค่า ศึกษาโดยใช้ข้อมูลจริงที่ได้จากแบบทดสอบการปฏิบัติทางคณิตศาสตร์ซึ่งพัฒนาโดย American College Testing [ACT] จำนวน 27 ข้อ ประกอบด้วยข้อสอบแบบเลือกตอบ (Multiple choice) จำนวน 12 ข้อ ข้อสอบการตอบแบบแจกแจงรายละเอียด (Grided-response items) จำนวน 9 ข้อ และแบบปลายเปิด(Open-ended) จำนวน 6 ข้อ โดยให้คะแนนข้อสอบแบบเลือกตอบ และการตอบแบบแจกแจงรายละเอียด เป็น 0-1 ส่วนสเกลคะแนนสำหรับข้อสอบแบบปลายเปิดมีความหลากหลาย มีช่วงจาก 0-3 ถึง 0-6 การตอบข้อสอบแบบปลายเปิดจะใช้ผู้ตรวจสองคนให้คะแนนตามเกณฑ์ที่วางไว้ (Scoring protocol) และจะมีผู้ตรวจอีกหนึ่งคนเป็นผู้ปรับแก้เพื่อให้เกิดความถูกต้อง ตัวอย่างที่ใช้จำนวน 1977 คน เป็นชาย 1006 คน และหญิง 971 คน วิธีการจำแนกนำมาใช้กับข้อสอบทั้ง 27 ข้อ โดยใช้คะแนนรวมเป็นตัวแปรตามเงื่อนไขหรือตัวแปรที่ใช้ในการจับคู่ คะแนนรวมเป็นผลรวมธรรมดาของผลที่ได้จากคะแนนข้อสอบ 27 ข้อนั้น โดยนำคะแนนของข้อสอบที่ให้คะแนนแบบสองค่าจำนวน 21 ข้อ รวมกับคะแนนของข้อสอบปลายเปิดที่ให้คะแนนแบบหลายค่า ดังนั้นช่วงกว้างของคะแนนจะมีค่าจาก 0 ถึง 46 ค่าประมาณความเชื่อมั่นของแบบทดสอบคำนวณโดยใช้สัมประสิทธิ์แอลฟาได้ .73 และ .78 สำหรับนักเรียนหญิงและนักเรียนชายตามลำดับ มีการประมาณค่าสัมประสิทธิ์การถดถอยเป็น 3 โมเดลได้แก่(1) โมเดลเต็ม (Full model) ซึ่งประกอบด้วย คะแนนตามเงื่อนไข(คะแนนรวม), คะแนนข้อสอบ และปฏิสัมพันธ์ระหว่างคะแนนรวมกับคะแนนข้อสอบ(2) โมเดลลด(Reduced model) มีเพียงคะแนนรวม และคะแนนข้อสอบ และ(3) โมเดลศูนย์ (Null model) จะมีเพียงคะแนนรวม ในแต่ละโมเดลจะให้ค่าสถิติไค-สแควร์อัตราส่วนความน่าจะเป็น (G^2) ความแตกต่างในค่า G^2 ระหว่างโมเดลลำดับขั้น (Hierachical model) มีการแจกแจงเชิงเส้นกำกับ (Asymptotically) ซึ่งเป็นตัวแปรสุ่มไค-สแควร์ ที่มีระดับชั้นความเป็นอิสระเท่ากับ 1 (df = 1)

ผลการศึกษาพบว่ากรณีข้อสอบที่ให้คะแนนแบบสองค่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก (LDFA) และวิธีแมนเทล-แฮนส์เซล (MH) เหมือนกัน ในการตรวจพบข้อสอบที่ทำหน้าที่

เบี่ยงเบน ที่เป็นรูปแบบเดียวกัน (Uniform DIF) ทั้งสองวิธีระบุข้อสอบแบบเลือกตอบจำนวน 7 ข้อ ว่าเป็นการแสดงการทำหน้าที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกันโดยที่ค่าไค-สแควร์ของวิธี LDFA จำนวน 6 ข้อ มี p< .001 และมี p< .005 รวมทั้งวิธี LDFA ระบุข้อสอบเลือกตอบ 2 ข้อ แสดงการทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) ที่ p< .005 สำหรับข้อสอบแบบปลายเปิดทั้งวิธี LDFA และวิธีแมนเทล (Mantel) ระบุข้อสอบ 2 ข้อเหมือนกันว่าข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน รวมทั้งวิธี LDFA ระบุว่าข้อสอบ 3 ข้อทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) เพื่อที่จะให้การแปลความได้ตรง และมีความหมายทางภาคปฏิบัติของDIF สำหรับข้อสอบที่มีนัยสำคัญทางสถิติ มิลเลอร์และสเปรย์ (Spray; & Miller. 1993) ได้กำหนดฟังก์ชันการจำแนกที่ประมาณได้จากภาพสำหรับโมเดลศูนย์และโมเดลเต็ม ที่มีช่วงความเชื่อมั่น 95% รอบๆ โมเดลเต็มที่ถูกประมาณค่า การตรวจบนภาพจะพบข้อสอบ DIF จากข้อสอบแบบเลือกตอบ 1 ข้อ และข้อสอบแบบปลายเปิด 1 ข้อ พร้อมทั้งสรุปว่าวิธีการตรวจพบ DIF ของข้อสอบที่มีการให้คะแนนแบบหลายค่า วิธีLDFAนั้นดีกว่าวิธีการถดถอยโลจิสติกที่ให้คะแนนแบบหลายค่า (Polytomous logistic regression) และวิธีแมนเทล (Mantel) ในแง่ที่ว่าง่ายและสะดวกกว่า นอกจากนี้มีประโยชน์สำหรับการระบุการทำหน้าที่เบี่ยงเบนของข้อสอบที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) ขณะที่วิธีแมนเทลทำไม่ได้

นารายานานและสวามินาธาน (Narayanan; & Swaminathan. 1996) ได้ศึกษาเปรียบเทียบวิธีแมนเทล-แฮนส์เซล(MH) วิธีโคร-ซิบ (CRO-SIB) (พัฒนามาจากวิธีซิบเทสท์ของเชียลี และ สเตาท์) และวิธีการถดถอยโลจิสติก (LR) ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ไม่เป็นรูปแบบ เดียวกัน (Nonuniform DIF) โดยจำลองข้อมูลภายใต้การจัดกระทำ 5 ปัจจัย จำนวน 384 เงื่อนไข คือขนาดตัวอย่าง 4 ขนาด (กลุ่มอ้างอิง 2 ขนาด จำนวน 500 คน และ 1000 คน กลุ่มสนใจ 2 ขนาด จำนวน 200 คน และ 500) สัดส่วนของข้อสอบที่ทำหน้าที่เบี่ยงเบน 3 ขนาด (0%, 10% และ 20%) ความแตกต่างของการแจกแจงค่าความสามารถ 2 ระดับ (แบบเท่ากัน และแบบไม่เท่ากัน) ขนาดอิทธิพลของข้อสอบทำหน้าที่เบี่ยงเบน 4 ขนาด (พื้นที่ระหว่าง IRFs มีค่าเท่ากับ .4, .6, .8 และ 1.0) ลักษณะของข้อสอบ 4 แบบ (ต่ำกับ สูง, ปานกลางกับ ต่ำ, ปานกลางกับ สูง และ สูงกับ ต่ำ) โดยกำหนดพารามิเตอร์ เหมือนกันสำหรับกลุ่มอ้างอิงและกลุ่มสนใจ คือ ต่ำ ปานกลาง และสูง ส่วนพารามิเตอร์ มีค่าแปรเปลี่ยนคือต่ำ (ในกลุ่มอ้างอิงและ ในกลุ่มสนใจ) และสูง (ในกลุ่มอ้างอิงและ ในกลุ่มสนใจ) สำหรับความยาวของแบบทดสอบใช้เพียง 40 ข้อ ข้อมูลที่ใช้ในการศึกษาจำลองตามโมเดลโลจิสติกแบบ 3 พารามิเตอร์โดยกำหนดค่าการเดาเท่ากับ ทุกข้อจากโปรแกรม DATAGEN ของ Hambleton และ Rovinelli (1973) ส่วนการวิเคราะห์การทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกใช้โปรแกรม DICHODIF ของRoger และ Swaminathan (1993) ส่วนวิธีโคร-ซิบใช้โปรแกรมCSIBTESTของ Li และStout (1994) แล้วใช้การวิเคราะห์ความแปรปรวน (ANOVA) แบบ 5 ทิศทางเพื่อทดสอบผลกระทบของปัจจัยที่ตรวจสอบด้วยวิธีทั้งสาม โดยทดสอบที่ระดับนัยสำคัญ .05 และ .01

ผลการศึกษาพบว่าในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ไม่เป็นรูปแบบเดียวกัน วิธีโคร-ซิบและวิธีการถดถอยโลจิสติกมีอัตราความถูกต้องเท่ากันภายใต้เกือบทุกเงื่อนไข ส่วนวิธีแมนเทล-แฮนส์เซลไม่มีอัตราความถูกต้องในการตรวจสอบดังกล่าว และจากการทดสอบปัจจัยที่มีผลต่อวิธีทั้งสามพบว่าขนาดตัวอย่าง ลักษณะข้อสอบและขนาดอิทธิพลของข้อสอบทำหน้าที่เบี่ยงเบนมีผลต่ออัตราความถูกต้องของทั้งสามวิธี กล่าวคือเมื่อขนาดตัวอย่างและขนาดอิทธิพลของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้นจะมีผลทำให้อัตราความถูกต้องของทั้งสามวิธีมีค่าเพิ่มขึ้นด้วย และเมื่อลักษณะของข้อสอบมีค่าอำนาจจำแนกเพิ่มขึ้นแล้วอัตราความถูกต้องของวิธีโคร-ซิบและวิธีการถดถอยโลจิสติกจะมีค่าเพิ่มขึ้น แต่เมื่อข้อสอบมีค่าความยากสูงหรือต่ำแล้วอัตราความถูกต้องของวิธีแมนเทล-แฮนส์เซลจะมีค่าสูงตามไปด้วย และเมื่อลักษณะข้อสอบมีค่าอำนาจจำแนกสูงและความยากต่ำแล้วอัตราความถูกต้องของทั้งสามวิธีมีค่าสูงสุด โดยที่ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนมีผลต่ออัตราความถูกต้องของวิธีการถดถอยโลจิสติก คือเมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้นแล้วอัตราความถูกต้องจะลดลง ส่วนปัจจัยการแจกแจงความสามารถพบว่ากลุ่มที่มีการแจกแจงความสามารถเท่ากันมีอัตราความถูกต้องของทั้งสามวิธีสูงกว่ากลุ่มที่มีการแจกแจงความสามารถไม่เท่ากัน สำหรับการตรวจสอบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่าวิธีแมนเทล-แฮนส์เซลมีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ภายในขอบเขตที่กำหนด แต่วิธีโคร-ซิบและวิธีการถดถอยโลจิสติกมีค่าสูงกว่าขอบเขตที่กำหนดในทุกเงื่อนไข โดยที่วิธีโคร-ซิบมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีการถดถอยโลจิสติกเพียงเล็กน้อย

รุสโซและสเตาท์ (Roussos; & Stout. 1996: 215-230) ได้ศึกษาผลของการใช้จำนวนตัวอย่าง ขนาดเล็ก และค่าพารามิเตอร์ของข้อสอบศึกษาที่มีต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิบเทสท์ (SIBTEST) และวิธีแมนเทล-แฮนส์เซล (MH) โดยจำลองข้อมูลศึกษา 2 การศึกษา

การศึกษาที่ 1 ศึกษาตัวอย่างขนาดเล็ก โดยใช้ขนาดตัวอย่าง 4 ขนาด (100คน 200คน 500คน และ 1000คน) ขนาดตัวอย่างในกลุ่มอ้างอิงเท่ากับกลุ่มสนใจ และความสามารถของกลุ่มอ้างอิง และกลุ่มสนใจมีการแจกแจงแบบปกติ โดยมีความแปรปรวนเท่ากับ 1.0 และความแตกต่างของการแจกแจง มีค่าเฉลี่ย 3 ระดับ แล้วจำลองข้อมูลภายใช้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ โดยประมาณค่าพารามิเตอร์จากข้อสอบจริงจำนวน 25 ข้อจาก ASVAB (Armed Services Vocational Aptitude Battery) ซึ่งมีอยู่ใน Mislevy และ Bock (1984) ในแบบ ทดสอบประกอบด้วยข้อสอบที่ทำหน้าที่ไม่เบี่ยงเบน 1 ข้อ ซึ่งมีค่าพารามิเตอร์ และ c=.25 โดยคัดเลือกมาจากข้อสอบที่มีค่าพารามิเตอร์ใกล้เคียงกับค่าเฉลี่ยในแบบทดสอบ และ สำหรับข้อมูลที่ศึกษาจำลอง 400 ครั้งในแต่ละเงื่อนไข แล้วใช้สถิติซิบเทสท์และแมนเทล-แฮนส์เซลทดสอบอัตราความคลาดเคลื่อนประเภทที่ 1 ที่ระดับนัยสำคัญ .05 ผลการศึกษาพบว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีซิบเทสท์และวิธีแมนเทล-แฮนส์เซลมีค่าไม่ต่างกันสำหรับกลุ่มตัวอย่างขนาดเล็ก อย่างไรก็ตามวิธีซิบเทสท์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีแมนเทล-แฮนส์เซลเล็กน้อย นอกจากนี้อัตราความ

คลาดเคลื่อนประเภทที่ 1 ของทั้งสองวิธีมีแนวโน้มว่าจะสูงขึ้นเมื่อขนาดตัวอย่างและความแตกต่างการแจกแจงความสามารถเพิ่มขึ้น

การศึกษาที่ 2 ศึกษาค่าพารามิเตอร์ของข้อสอบ โดยใช้ตัวอย่างขนาดปานกลางถึงใหญ่ 3 ขนาด (500 คน 1,000 คน และ 3,000 คน) ขนาดตัวอย่างในกลุ่มอ้างอิงเท่ากับกลุ่มสนใจ และความสามารถระหว่างกลุ่มดังกล่าวมีการแจกแจกปกติ โดยมีความแปรปรวนเท่ากับ 1.0 และความแตกต่างในค่าเฉลี่ยของการแจกแจง 2 ระดับ คือ 0.0 และ 1.0 สำหรับแบบทดสอบเหมือนการศึกษาที่ 1 โดยเลือกค่าพารามิเตอร์สำหรับนำมาวิเคราะห์ให้แตกต่างกัน คือเลือกค่าพารามิเตอร์อำนาจจำแนก 3 ระดับ คือ .4, 1.0 และ 2.5 พารามิเตอร์ความยาก 5 ระดับ คือ -1.5 , -0.5 ,0.0 , 0.5 และ 1.5 และพารามิเตอร์การเดา เมื่อ , พารามิเตอร์การเดา 3 ระดับคือ 0.20, 0.10 และ 0.05 เมื่อการศึกษาสำหรับกรณีนี้จะจำลองข้อมูล 100 ครั้งในแต่ละเงื่อนไขแล้วทดสอบด้วยสถิติซิบเทสท์และแมนเทล-แฮนส์เซลที่ระดับ.05ผลการศึกษาพบว่า เมื่อไม่มีความแตกต่างของการแจกแจงความสามารถระหว่างกลุ่มอ้างอิงและกลุ่มสนใจ แล้วอัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งสองวิธีอยู่ภายในขอบเขตที่กำหนดในทุกเงื่อนไข โดยที่วิธีแมนเทล-แฮนส์เซลดีกว่าเล็กน้อยเมื่อเพิ่มขนาดตัวอย่างแล้วอัตราความคลาดเคลื่อนของทั้งสองวิธีจะมีค่าลดลงโดยมีค่าลดลงถึง 0 เมื่อขนาดตัวอย่างเป็น 3,000 คน แต่ถ้าความแตกต่างของการแจกแจงความสามารถเป็น 1.0 แล้วอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธี ซิบเทสท์จะต่ำกว่าวิธีแมนเทล-แฮนส์เซลในทุกเงื่อนไขของการตรวจสอบ

ชาง มาซซีโอ และรุสโซ (Chang; Mazzeo; & Roussos.1996: 333-353) ได้ปรับขยายวิธีซิบเทสท์ของ เชียลี และสเตาท์ (Shealy; & Stout.1993.) เพื่อใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่าซึ่งเรียกว่าวิธีซิบเทสท์ที่ปรับขยาย (modified SIBTEST) โดยนำไปเปรียบเทียบกับวิธีแมนเทล (Mantel) และวิธีความแตกต่างเฉลี่ยมาตรฐาน (Standardized mean difference: SMD) โดยทำการศึกษา 2 การศึกษาภายใต้การจำลองข้อมูลคือการศึกษา 1 ใช้ข้อมูลของซวิค โดน็อคฮิว และกริมา (Zwick; Donoghue; & Grima.1993) ซึ่งได้ทำการศึกษาการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่าโดยใช้วิธีแมนเทล (Mantel) การศึกษาดังกล่าวใช้ข้อมูลจำลอง ภายใต้โมเดลพาร์เชียลเครดิต (Partial credit model: PCM) ของมาสเตอร์ (Master.1982)โดยออกแบบการจำลองข้อมูล 54 เงื่อนไขภายใต้ความแตกต่างของข้อสอบ (Different items) ความสามารถของกลุ่มผู้สอบ (Ability groups) และรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน (Type of DIF) แบบทดสอบที่ใช้ในการจับคู่ (Matching test) ประกอบด้วยข้อสอบ 24 ข้อซึ่งวิธีดังกล่าวจะรวมข้อสอบที่ต้องการศึกษา (Studied item) เข้าไว้ด้วย โดยที่ข้อสอบ 20 ข้อแรกให้คะแนนแบบสองค่า (Dichotomous) และอีก 4 ข้อให้คะแนนแบบหลายค่าชนิด 4 รายการตอบ (ให้คะแนน 0,1,2 หรือ 3) ข้อสอบที่ให้คะแนนแบบสองค่าจะจำลองข้อมูลโดยใช้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ (3PLM) กำหนดค่าการเดา ส่วนข้อสอบให้คะแนนแบบหลายค่าจำลองโดยใช้โมเดลพาร์เชียลเครดิต (PCM) ข้อสอบที่ต้องการศึกษา (Studied item) เป็นข้อสอบที่ให้คะแนนแบบ

หลายค่า (DIF & null DIF) จำนวน 3 ข้อ แต่ละข้อมี 4 รายการตอบแล้วใช้โมเดล PCM จำลองผลการตอบข้อสอบที่ทำหน้าที่เบี่ยงเบนระหว่างกลุ่มอ้างอิงและกลุ่มสนใจ โดยจำลองรูปแบบของข้อสอบ 4 รูปแบบ คือแบบคงที่ (Constant DIF) แบบเปลี่ยนขนาดที่ระดับความสามารถต่ำ (Low-shift DIF) แบบเปลี่ยนขนาดที่ระดับความสามารถสูง (High-shift DIF) และแบบสมดุล (Balanced DIF) ซึ่งสามรูปแบบแรกจัดเป็นข้อสอบที่ทำหน้าที่เบี่ยงเบนที่มีทิศทางเดียวกัน (Unidirectional) ส่วนรูปแบบสุดท้ายจัดเป็นข้อสอบที่ทำหน้าที่เบี่ยงเบนที่ไม่มีทิศทางเดียวกัน (Nonunidirectional) รูปแบบของข้อสอบดังกล่าวถูกกำหนดโดยพารามิเตอร์ความยากประจำขั้นของข้อสอบ (Threshold parameters) ระหว่างกลุ่มผู้สอบ โดยกลุ่มอ้างอิงมีค่าความยากประจำขั้นของข้อสอบเป็น (-0.91 , 0.98 , 0.21) , (-2.25 , -1.81 , 1.66) และ (-0.54 , -2.11 , 0.74) ตามลำดับ ส่วนพารามิเตอร์ความยาก (ประจำขั้น) ของข้อสอบของกลุ่มสนใจจำนวนโดยใช้สูตร เมื่อ i = 1 , 2 , 3 ในแต่ละรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนจะมีขนาดของข้อสอบทำหน้าที่เบี่ยงเบน 2 ขนาดคือ = .10 และ = .25 (สำหรับเงื่อนไข มีผลใน null-DIF) ทำให้ได้เงื่อนไขการทำหน้าที่เบี่ยงเบนของข้อสอบ 8 เงื่อนไข (รูปแบบของ DIF 4 แบบ ขนาดของ DIF 2 ขนาด) แต่ละเงื่อนไขบรรจุข้อสอบที่ต้องการศึกษา (Studied item) 3 ข้อ ดังนั้นข้อสอบทำหน้าที่เบี่ยงเบน 24 ข้อ (ข้อสอบ DIF 3 ข้อ รูปแบบของ DIF 4 แบบ ขนาดของ DIF 2 ขนาด) นั่นคือข้อสอบที่ใช้ในการศึกษาทั้งหมดมีทั้งหมด 27 ข้อ เป็นข้อสอบไม่ทำหน้าที่เบี่ยงเบน (No DIF) 3 ข้อ และข้อสอบทำหน้าที่เบี่ยงเบน (DIF) 24 ข้อ สำหรับผลการตอบข้อสอบดังกล่าวจำลองภายใต้ขนาดตัวอย่างขนาดกลุ่มละ 500 คน โดยมีการแจกแจงความสามารถ 2 ระดับ คือไม่มีความแตกต่างความสามารถระหว่างกลุ่ม และมีความแตกต่างความสามารถระหว่างกลุ่ม โดยกลุ่มอ้างอิงมีการแจกแจงแบบ เสมอ ขณะที่กลุ่มสนใจมีการแจกแจง 2 ระดับ คือ และ ดังนั้นจะต้องจำลองข้อมูลทั้งหมด 54 เงื่อนไข (27 $\times$ 2) ผลการศึกษาพบว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีความแตกต่างเฉลี่ยมาตรฐานมีค่าเท่ากับ 0.049 และ 0.046 ตามลำดับ ค่าดังกล่าวอยู่ภายในขอบเขตที่กำหนด ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิบเทสท์มีค่าเท่ากับ 0.063 ซึ่งถือว่ามีค่าสูงกว่าเกณฑ์ปกติ (เฟ้อ)เล็กน้อย ในกรณีการแจกแจงความสามารถแบบไม่เท่ากันจะมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่า หรือมีอัตราความถูกต้องต่ำกว่าการแจกแจงความสามารถแบบเท่ากันเพียงเล็กน้อย การตรวจที่ผิดพลาดของ DIFมักจะมีต่อกลุ่มอ้างอิงทั้งนี้เป็นเพราะการปรับแก้บนการถดถอยเกี่ยวกับความแตกต่างความสามารถกลุ่ม เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่าทั้งสามวิธีสามารถตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่มีรูปแบบคงที่ รูปแบบเปลี่ยนขนาดที่ระดับความสามารถต่ำ และรูปแบบเปลี่ยนขนาดที่ระดับความสามารถสูงซึ่งเป็นข้อสอบทำหน้าที่เบี่ยงเบนที่มีทิศทางเดียวกัน (Unidirectional DIF) ได้เป็นอย่างดีในทุกเงื่อนไข แต่ทั้งสามวิธีไม่สามารถตรวจพบการทำหน้าที่เบี่ยงเบนของข้อสอบที่มีรูปแบบสมดุล

ส่วนการศึกษา 2 ขยายจากการศึกษาที่ 1 โดยนำค่าพารามิเตอร์อำนาจจำแนกเข้าร่วมศึกษาด้วย โดยจำลองแบบทดสอบจับคู่ 24 ข้อ (25 ข้อสำหรับแมนเทลและวิธีความแตกต่างเฉลี่ย

มาตรฐาน) ในข้อสอบ 20 ข้อแรกให้คะแนนแบบสองค่า จำลองข้อมูลโดยใช้โมเดลแบบ 3 พารามิเตอร์ (3PLM) โดยมีค่าพารามิเตอร์ข้อสอบเหมือนกับในการศึกษาที่ 1 ส่วนข้อสอบอีก 4 ข้อ คือ ข้อที่ 21-24 เป็นข้อสอบที่ให้คะแนนแบบหลายค่า ซึ่งได้เพิ่มค่าพารามิเตอร์อำนาจจำแนกเข้ามาวิเคราะห์โดยนำมาจากผลการประมาณค่าพารามิเตอร์ที่เกิดขึ้นจริงของ 1992 NEAP (Nation Assessment of Educational Progress) แล้วจำลองข้อมูลโดยใช้โมเดลพาร์เชียลเครดิตทั่วไป (Generalized Partial Credit Model: GPCM) ของมูราคิ (Muraki. 1992) สำหรับข้อสอบที่ศึกษา (Studied items) มีค่าพารามิเตอร์ความยากประจำขั้น (threshold parameters) ของกลุ่มอ้างอิงเป็น (-1,0,1) เหมือนกันทุกข้อ ส่วนพารามิเตอร์อำนาจจำแนกจะมีค่าแตกต่างกัน 11 ระดับคือ 2.0, 1.5, 1.36, 1.12, 1.0, 0.869, 0.588, 0.33, 0.25, 0.23 และ 0.15 การจำลองข้อสอบดังกล่าวมี 2 เงื่อนไข คือ เงื่อนไขแรกเป็นข้อสอบไม่แสดงการทำหน้าที่เบี่ยงเบน (Null DIF) ซึ่งพารามิเตอร์ของข้อสอบที่ศึกษาระหว่างกลุ่มอ้างอิง และกลุ่มสนใจมีค่าเท่ากัน และเงื่อนไขที่สองเป็นข้อสอบที่ทำหน้าที่เบี่ยงเบน แบบคงที่ (Constant DIF) โดยกำหนดค่าความยาก(ประจำขั้น)แปรเปลี่ยนที่ขนาด 0.25 ดังนั้นค่าความยาก (ประจำขั้น)ของข้อสอบที่ศึกษาจึงเป็น (-0.75, 0.25, 1.25) ส่วนการแจกแจงความสามารถมีเพียงเงื่อนไขเดียวคือ กลุ่มอ้างอิงมีการแจกแจงความสามารถแบบ ส่วนกลุ่มสนใจมีการแจกแจงความสามารถแบบ ดังนั้นเงื่อนไขของการทดลองมีทั้งหมด 22 เงื่อนไข (Null DIF 11 ระดับ และ DIF 11 ระดับ) ด้วยขนาดตัวอย่างกลุ่มละ 500 คน โดยแต่ละเงื่อนไขจำลองข้อมูลซ้ำ 1000 ครั้ง ผลการศึกษาพบว่าเมื่อขนาดตัวอย่าง 500 คน(แต่ละกลุ่ม) และพารามิเตอร์อำนาจจำแนกของข้อสอบที่ศึกษามีค่าเพิ่มขึ้นแล้วอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทลและวิธีความแตกต่างเฉลี่ยมาตรฐานมีค่าสูงขึ้น โดยเพิ่มขึ้นถึง 40% ซึ่งต่างจากวิธีซิบเทสท์ ถึงแม้ว่าค่าอำนาจจำแนกจะสูงขึ้น แต่อัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นเล็กน้อย เมื่อเพิ่มขนาดตัวอย่างเป็น 1000 คน ต่อกลุ่ม อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทลและวิธีความแตกต่างเฉลี่ยมาตรฐานยิ่งเพิ่มมากขึ้นอย่างรวดเร็ว ส่วนของวิธีซิบเทสท์ค่อนข้างคงที่

เทียน (Tian. 1999:1-156) ได้ศึกษาเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ที่ได้จากการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบโดยใช้วิธีแมนเทล (Mantel) วิธีแมนเทล-แฮนส์เซลทั่วไป (GMH) วิธีการทำให้เป็นมาตรฐาน (STND) วิธีโพลีโตมัสซิบเทสท์ (Poly-SIBTEST) และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก (LDFA) จากแบบทดสอบที่มีขนาด 20 และ 40 ข้อ โดยที่เศษสี่ส่วนห้าของข้อสอบบนแต่ละแบบทดสอบเป็นข้อสอบที่ให้คะแนนแบบสองค่า (Dichotomous) และที่เหลือจะเป็นข้อสอบที่ให้คะแนนแบบหลายค่า (Polytomous) ที่มีรายการตอบ 4 รายการ (คือ 0,1,2,3) ดังนั้นแบบทดสอบที่มี 20 ข้อ จะประกอบด้วยข้อสอบที่ให้คะแนนแบบสองค่าจำนวน 16 ข้อ (ข้อ 1-16) และข้อสอบที่ให้คะแนนแบบหลายค่าที่มีรายการตอบ 4 รายการ จำนวน 4 ข้อ (ข้อ 17-20) ส่วนแบบทดสอบที่มี 40 ข้อ จะประกอบด้วยข้อสอบที่ให้คะแนนแบบสองค่า จำนวน 32 ข้อ(ข้อ 1-32) และข้อสอบที่ให้คะแนนแบบหลายค่าที่มีรายการตอบ 4 รายการ จำนวน 8 ข้อ (ข้อ 33-40) ข้อสอบสุดท้ายในแต่ละแบบทดสอบ

(ข้อ 20 ของแบบทดสอบแรก และข้อ 40 ของแบบทดสอบหลัง)เป็นข้อสอบที่ต้องการศึกษา (Studied item) ข้อมูลที่ใช้ในการศึกษาได้จากการจำลองภายใต้ปัจจัยที่แปรเปลี่ยน 6 ปัจจัยได้แก่ ความแตกต่างการแจกแจงความสามารถ 3 ระดับ (0,0.5 และ 1.0) ความยาวของแบบทดสอบ 2 ขนาด (20 และ 40) ขนาดตัวอย่าง 3 ขนาด (จำนวนรวมของขนาดตัวอย่างกลุ่มอ้างอิงและกลุ่มสนใจเป็น 600 คน,1200 คน และ 2400คน) และอัตราส่วนขนาดกลุ่มสนใจและกลุ่มอ้างอิงเป็น 2 ระดับ (1:1 และ 1:2) พารามิเตอร์อำนาจจำแนก 3 ระดับ (0.5,1.0 และ 1.5) และเงื่อนไข DIF 4 เงื่อนไข (Null DIF, Uniform DIF (Constant), Balanced uniform DIF, Nonuniform DIF) รวมข้อมูลที่จำลองทั้งหมด 432 เงื่อนไข (3×2×3×2×3×4) ในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน ความแตกต่างระหว่างพารามิเตอร์ b (ขนาดของ DIF) เป็น 0.25 และความแตกต่างในค่าอำนาจจำแนกข้อสอบสำหรับเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกันมีช่วงจาก 0.5 ถึง 1.5

ผลการศึกษาพบว่า ความแตกต่างการแจกแจงความสามารถเป็นองค์ประกอบหลักที่มีอิทธิพลต่อผลของวิธีทั้งหมด กล่าวคือเมื่อความแตกต่างความสามารถกลุ่มเพิ่มขึ้น อัตราความคลาดเคลื่อนประเภทที่ 1 และอัตราความถูกต้องจะเพิ่มขึ้นด้วย ซึ่งผลอันนี้จะจริงสำหรับค่าอำนาจจำแนกข้อสอบที่สูงขึ้นด้วย LDFA-nonuniform ดูเหมือนว่าจะมีความไวที่สุดต่อความแตกต่างการแจกแจงความสามารถ นั่นคือ มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงมากเกินกว่าที่จะยอมรับได้ภายใต้เงื่อนไขทั้งหมดที่มีความแตกต่างขนาดปานกลางและขนาดใหญ่ในการแจกแจงความสามารถ วิธี SIBTEST มีการควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ที่ค่อนข้างจะดีกว่าภายใต้เงื่อนไขเหล่านี้มากกว่าวิธีอื่น เมื่อกลุ่มสนใจ และกลุ่มอ้างอิงมีความสามารถเท่ากันวิธีทั้งหมดมีการควบคุมอัตราความคลาดเคลื่อนประเภทที่1 ที่ดี อย่างไรก็ตามวิธี SIBTEST มีอัตรา ความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีอื่นโดยเฉพาะแบบทดสอบที่มี 20 ข้อ จากการพิจารณาอัตราความถูกต้องของวิธีต่าง ๆ เมื่อการแจกแจงความสามารถเท่ากัน ทุกวิธีมีอัตราความถูกต้องสูงสำหรับการตรวจพบข้อสอบที่ทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF)

งานวิจัยภายในประเทศ

จิติมา วรรณศรี (2539) ได้เปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซลกับวิธีซิบเทสท์ โดยใช้ข้อมูลจำลองจากโปรแกรม IRTDATA โดยศึกษาภายใต้ปัจจัยที่แปรเปลี่ยนด้านความยาวของแบบทดสอบ 3 ขนาด คือ 30 ข้อ , 60 ข้อ และ 90 ข้อ ขนาดตัวอย่าง 3 ขนาด คือ 200 คน, 600 คน และ 1000 คน โดยแต่ละขนาดมีอัตราส่วนระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ (กลุ่มสนใจ) 4 อัตราส่วน คือ 1:1, 1:0.9, 1:0.75 และ 1:0.5 รวมเงื่อนไขที่ศึกษาทั้งหมด 36 เงื่อนไข

ผลการศึกษาพบว่า ภายใต้ขนาดตัวอย่างและอัตราส่วนขนาดตัวอย่าง วิธีแมนเทลแฮนส์เซล กับวิธีซิบเทสท์ มีประสิทธิภาพเท่าเทียมกันในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่

ทุกขนาดตัวอย่างและทุกอัตราส่วนขนาดตัวอย่าง ขณะที่ความยาวของแบบทดสอบเดียวกันสามารถตรวจพบข้อสอบที่ทำหน้าที่ต่างกันได้ถูกต้องร้อยละ 50 เมื่อขนาดตัวอย่าง 200 และ 600 คน และสามารถตรวจพบข้อสอบที่ทำหน้าที่ต่างกันได้ถูกต้องร้อยละ 100 เมื่อขนาดตัวอย่าง 1000 คน โดยส่วนมากวิธีซิบเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 มากกว่าวิธีแมนเทล-แฮนส์เซลเล็กน้อย ส่วนภายใต้ความยาวแบบทดสอบวิธีแมนเทล-แฮนส์เซลกับวิธีซิบเทสท์มีประสิทธิภาพเท่าเทียมกันที่ทุกระดับความยาวของแบบทดสอบ และถ้าใช้แบบทดสอบที่มีความยาว 60 ข้อ (ความยาวขนาดปานกลาง) จะมีประสิทธิภาพในการตรวจสอบดีที่สุด

วลีมาศ แซ่อึ้ง (2543) ได้เปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ไม่เป็นรูปแบบเดียวกัน (อเนกรูป) ระหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก โดยศึกษาภายใต้การจำลองข้อมูลด้วยการจัดกระทำปัจจัยที่แปรเปลี่ยน (ตัวแปรอิสระ) 4 ตัวคือ ลักษณะของข้อสอบ 9 ระดับ (aต่ำกับbต่ำ, aต่ำกับbปานกลาง, aต่ำกับbสูง, aปานกลางกับbต่ำ, aปานกลางกับ bปานกลาง , aปานกลางกับbสูง , aสูงกับbต่ำ , aสูงกับbปานกลาง และ aสูงกับbสูง) ความยาวของแบบทดสอบ 2 ขนาด (30 ข้อ และ 60 ข้อ) สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 3 ขนาด (5% 10% และ 20%) และขนาดตัวอย่าง 6 ขนาด (จำนวนผู้สอบกลุ่มอ้างอิงต่อจำนวนผู้สอบกลุ่มสนใจเท่ากับ 250 คนต่อ 250 คน , 500 คนต่อ 250 คน , 500 คนต่อ 500 คน ,1000 คน ต่อ 250 คน , 1000 คน ต่อ 500 คน, 1000 คน ต่อ 1000 คน) รวมข้อมูลที่จำลองทั้งหมด 324 เงื่อนไข (9×2×3×6) โดยใช้โปรแกรม IRTDATA ของ Johanson (1992) จำลองข้อมูลภายใต้ทฤษฎีการตอบข้อสอบ โมเดลโลจิสติกแบบสามพารามิเตอร์ชนิดกำหนดค่า c คงที่ (3 PLM-c) ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบวิธีซิปเทสท์และวิธีซิปเทสท์ปรับใหม่ใช้โปรแกรม SIBTEST ของ Stout และRoussos (1992) วิธีแมนเทล-แฮนส์เซล ใช้โปรแกรม MHDIF ของ Fidalgo (1995) และวิธีการถดถอยโลจิสติก ใช้โปรแกรม SPSS/PC สำหรับวิธีการวัดพื้นที่ชนิดไม่คิดเครื่องหมายของราชู ซึ่งใช้เป็นเกณฑ์ในการเปรียบเทียบ จะใช้โปรแกรม IRTDIF แล้วคำนวณอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธี ส่วนการเปรียบเทียบความแตกต่างใช้สถิติ ANOVA และ t-test

ผลการวิจัยพบว่าอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปของวิธีซิปเทสท์ปรับใหม่ และวิธีการถดถอยโลจิสติกมีค่าเท่าเทียมกันภายใต้เกือบทุกเงื่อนไข และทั้งสองวิธีดังกล่าวมีอำนาจการทดสอบสูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลภายใต้เกือบทุกเงื่อนไข ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 วิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล- แฮนส์เซล และวิธีการถดถอยโลจิสติก มีค่าอยู่ภายในเกณฑ์ของอัตราความคลาดเคลื่อนประเภทที่ 1 ที่ระดับ 10% ภายใต้เกือบทุกเงื่อนไข

ทองอยู่ สาระ(2543) ได้ศึกษาเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบแบบสม่ำเสมอและไม่สม่ำเสมอ

ระหว่างวิธี แมนเทล-แฮนส์เซลส์ และวิธีการถดถอยโลจิสติกโดยใช้ความยาวของแบบทดสอบและขนาดตัวอย่างต่างกัน กลุ่มตัวอย่างที่ใช้ในการศึกษาเป็นนักเรียนระดับมัธยมศึกษาปีที่ 3 จำนวน 3,242 คน เครื่องมือที่ใช้เป็นแบบทดสอบวัดความสามารถทางสมองตามแนวโครงสร้างของโอติส-เลนนอล ซึ่งเป็นแบบเลือกตอบชนิด 5 ตัวเลือก จำนวน 80 ข้อ วัดความสามารถทั่วไป 3 ด้าน คือ ความเข้าใจด้านภาษา เหตุผลด้านภาษา และเหตุผลด้านภาพ 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, 1000:100, 1000:200, 1000:300 และ 1000:1000 รวมเงื่อนไขทั้งหมด 30 เงื่อนไข

ผลการศึกษาพบว่า ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบแบบสม่ำเสมอและไม่สม่ำเสมอ ภายใต้ปัจจัยความยาวของแบบทดสอบวิธีแมนเทล-แฮนส์เซลส์ และวิธีการถดถอยโลจิสติกมีอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่แตกต่างกันเกือบทุกเงื่อนไข ปัจจัยความยาวของแบบทดสอบไม่มีผลต่ออัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีตรวจสอบทั้ง 2 วิธี ส่วนปัจจัยขนาดตัวอย่างมีผลต่ออัตราความถูกต้องของทั้ง 2 วิธี เกือบทุกเงื่อนไข กล่าวคือ เมื่อขนาดตัวอย่างเพิ่มขึ้นมีผลทำให้อัตราความถูกต้องเพิ่มขึ้น สำหรับขนาดตัวอย่างไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ในเกือบทุกเงื่อนไขที่ศึกษา

สิริรัตน์ วิภาสศิลป์ (2545) ได้เปรียบเทียบวิธีซิบเทสท์ และวิธีดีเอฟไอที ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบ จากข้อมูลการตอบข้อสอบที่ใช้ความสามารถหลายมิติภายใต้ปัจจัยที่ศึกษา 2 ปัจจัย คือ ความยาวของแบบทดสอบ 3 ขนาด (30 ข้อ , 40 ข้อ และ 50 ข้อ) และขนาดตัวอย่าง 5 ขนาด (50 คน , 100 คน, 200 คน ,500 คน และ 1000 คน) กลุ่มตัวอย่างที่ใช้ในการศึกษาได้มาจากการสุ่มแบบใส่คืนจากประชากรเทียม ซึ่งกำหนดจากนักเรียนชายและนักเรียนหญิงชั้นมัธยมศึกษาปีที่1 ในจังหวัดนนทบุรี โดยสุ่มแต่ละขนาด 50 ครั้ง เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบวิชาคณิตสาสตร์ ชั้นมัธยมศึกษาปีที่ 1 ที่ผู้วิจัยสร้างขึ้น ประกอบด้วยข้อสอบแบบเลือกตอบชนิด 5 ตัวเลือก จำนวน 50 ข้อ มีข้อสอบที่ผู้เชี่ยวชาญพิจารณาว่า เป็นข้อสอบที่แสดงการทำหน้าที่เบี่ยงเบนต่อเพศชายจำนวน 16 ข้อ จากนั้นจึงเก็บรวบรวมข้อมูล แล้วคัดเลือกข้อสอบตามสัดส่วนในตารางกำหนดข้อสอบเพื่อจัดเป็นแบบทดสอบที่มีความยาว 40 และ 30 ข้อ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบในเงื่อนไขต่าง ๆ ใช้โปรแกรม SIBTEST และ DFIT แล้วนำผลการตรวจสอบไปวิเคราะห์อัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบด้วยวิธีเดียวกัน และวิธีต่างกัน จากนั้นจึงเปรียบเทียบความแตกต่างโดยใช้สถิติ MANOVA และ Z-test

ผลการศึกษาในเงื่อนไขของข้อสอบ พบว่ากรณีปัจจัยความยาวแบบทดสอบ 30,40 และ 50 ข้อ และปัจจัยขนาดตัวอย่าง 50 ,100 และ 200 คน ไม่มีผลต่ออัตราความถูกต้องของวิธีซิบเทสท์ เมื่อขนาดตัวอย่าง 500 และ 1000 คน มีค่าสูงกว่าการตรวจสอบในขนาดตัวอย่าง 50,100 และ 200 คน ซึ่งอัตราความคลาดเคลื่อนประเภทที่ 1 ก็มีค่าสูงกว่าด้วยสำหรับการตรวจสอบด้วยวิธีดีเอฟไอที

พบว่าปัจจัยขนาดตัวอย่างทั้ง 5 ขนาดไม่มีผลต่ออัตราความถูกต้อง ส่วนกรณีปัจจัยความยาวแบบทดสอบทั้ง 3 ขนาด และปัจจัยขนาดตัวอย่างที่ต่างกันพบว่า อัตราความถูกต้องของวิธีซิบเทสท์น้อยกว่าวิธีดีเอฟไอทีในทุกเงื่อนไขและความสอดคล้องในการตรวจสอบด้วยวิธีทั้งสองมีค่าต่ำกว่าร้อยละ 10 ส่วนการศึกษาในเงื่อนไขของหมวดข้อสอบพบว่าอัตราความถูกต้องของวิธีซิบเทสท์มาก กว่าวิธีดีเอฟไอทีใน 2 เงื่อนไข คือความยาวของแบบทดสอบ 30 ข้อ และขนาดตัวอย่าง 1000 คน กับความยาวแบบทดสอบ 40 ข้อ และขนาดตัวอย่าง 500 คน ส่วนการศึกษาในเงื่อนไขแบบทดสอบพบว่าอัตราความถูกต้องของวิธีซิบเทสท์มากกว่าวิธีดีเอฟไอทีเมื่อความยาวของแบบทดสอบ 50 ข้อ และ ขนาดตัวอย่าง 100, 200 และ 1000 คน

อุทัยวรรณ สายพัฒนะ (2547) ได้เปรียบเทียบประสิทธิภาพของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนแบบหลายค่าระหว่างวิธีแมนเทล-แฮนส์เซลทั่วไป (GMH) และวิธีโพลีโตมัสซิบเทสท์ (Polytomous SIBTEST) โดยศึกษาภายใต้ปัจจัยที่แปรเปลี่ยน 2 ปัจจัย คือ ความยาวของแบบทดสอบ 3 ขนาด (20 ข้อ , 30 ข้อ และ 40 ข้อ) และขนาดตัวอย่าง 3 ขนาด (250 คน,500 คนและ 1000 คน) กลุ่มตัวอย่างในการศึกษาได้มาจากการสุ่มแบบใส่คืนจากประชากรเทียม ซึ่งกำหนดจากนักเรียนชายและนักเรียนหญิง ชั้นมัธยมศึกษาปีที่ 3 ในสังกัดกรมสามัญศึกษา จังหวัดพระนครศรีอยุธยา แต่ละขนาดตัวอย่างสุ่มจำนวน 10 ครั้ง เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบวิชาคณิตศาสตร์ เรื่องวงกลม ชั้นมัธยมศึกษาปีที่ 3 ที่ผู้วิจัยสร้างขึ้นประกอบ ด้วยข้อสอบเลือกตอบแบบถูกผิด ชนิด 5 ตัวเลือก จำนวน 40 ข้อ มีข้อสอบที่ผู้เชี่ยวชาญพิจารณาว่าเป็นข้อสอบที่ทำหน้าที่ต่างกันเอื้อต่อเพศชายจำนวน 8 ข้อ หลังจากเก็บรวบรวมข้อมูลแล้วคัดเลือกข้อสอบตามสัดส่วนในตารางกำหนดข้อสอบ จัดเป็นแบบทดสอบที่มีความยาว 30 และ 20 ข้อ ต่อจากนั้นจะตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้เงื่อนไขต่างๆโดยใช้วิธี GMH และวิธี Polytomous SIBTEST ส่วนวิธีพาเชียลเครดิตทั่วไป (GPCM) ใช้เป็นเกณฑ์การเปรียบเทียบกับ 2 วิธีดังกล่าว ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบวิธี GMH และวิธี GPCM จะใช้โปรแกรม MATLAB ส่วนวิธี Polytomous SIBTEST จะใช้โปรแกรม DIFPACK (ในส่วนโปรแกรมย่อย Polytomous SIBTEST)แล้วเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่1 ของวิธี GMH และ Polytomous SIBTEST โดยการวิเคราะห์ความแปรปรวนแบบหลายตัวแปร (MANOVA)

ผลการวิจัยพบว่าภายใต้ปัจจัยความยาวของแบบทดสอบและขนาดตัวอย่างส่งผลต่ออัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธี GMH ยกเว้นในเงื่อนไขความยาวแบบทดสอบ 30 และ 20 ข้อ ขนาดตัวอย่าง 500 และ 250 คน ส่วนการตรวจสอบด้วยวิธี Polytomous SIBTEST พบว่าปัจจัยความยาวของแบบทดสอบและขนาดตัวอย่างมีผลต่ออัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ยกเว้นในเงื่อนไขความยาวของแบบทดสอบ 20 ข้อ และขนาดตัวอย่าง 1000 และ 500 คน สำหรับการเปรียบเทียบประสิทธิภาพของสองวิธีพบว่าในทุกเงื่อนไขของความยาวของแบบทดสอบและขนาดตัวอย่างวิธี GMH และวิธี Polytomous

SIBTEST มีประสิทธิภาพในการตรวจสอบไม่แตกต่างกัน ยกเว้นในเงื่อนไขของขนาดตัวอย่าง 1000 คนที่มีความยาวแบบทดสอบ 20 ข้อ วิธี Polytomous SIBTEST มีประสิทธิภาพสูงกว่าวิธี GMH

อรินทร์ น่วมถนอม (2549) ได้เปรียบเทียบวิธีโพลี-ซิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับและวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า การศึกษาครั้งนี้ผู้วิจัยใช้โปรแกรม RESGEN ของ Muraki (2000) จำลองข้อมูลภายใต้โมเดลพาเชียลเครดิตทั่วไปแบบหลายมิติ จำลองผลการตอบจากแบบทดสอบที่วัดความสามารถสองมิติจำนวน 40 ข้อ ข้อสอบทุกข้อมีรายการตอบ 5 รายการ ให้คะแนนเป็น 1,2,3,4 หรือ 5 คะแนน ข้อมูลดังกล่าวจำลองภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน 2 รูปแบบ (ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน) สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 3 ขนาด (จำนวน 10% , จำนวน 20% และ จำนวน 30%) ความแตกต่างของการแจกแจงความสามารถ 3 ระดับ (แตกต่างกัน 0.0SD , แตกต่างกัน 0.5SD และ แตกต่างกัน 1.0 SD) และขนาดตัวอย่าง 4 ขนาด (250 คน , 500 คน, 1000 คน และ 2000 คน) รวมข้อมูลที่ใช้ศึกษาทั้งหมด 72 เงื่อนไข (2×3×3×4) ในแต่ละเงื่อนไขจำลองซ้ำ 50 ครั้งแล้ววิเคราะห์ข้อมูลในแต่ละเงื่อนไขด้วยวิธีโพลี-ซิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับและวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ในการวิเคราะห์การทำหน้าที่เบี่ยงเบนของข้อสอบวิธีโพลี-ซิปเทสท์ ใช้โปรแกรม DIFPACK ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับและวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติใช้โปรแกรม SPSS แล้วคำนวณอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของแต่ละวิธี ส่วนการเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ใช้การวิเคราะห์ความแปรปรวนหลายตัวแปร (MANOVA) ที่ระดับนัยสำคัญ .001

ผลการวิจัยพบว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับและวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องใกล้เคียงกัน ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกันและที่ไม่เป็นรูปแบบเดียวกันทั้งสองวิธีมีอัตราความถูกต้องสูงกว่าวิธีโพลี- ซิปเทสท์ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ไม่เป็นรูปแบบเดียวกัน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนในแบบทดสอบไม่มีผลต่อวิธีโพลี-ซิปเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ แต่มีผลต่อวิธีการถดถอยโลจิสติกแบบจัดอันดับ เมื่อความแตกต่างของการแจกแจงความสามารถเพิ่มขึ้นวิธีโพลี-ซิปเทสท์สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ที่สูงเกินปกติได้ดีกว่าวิธีอื่น นอกจากนี้เมื่อขนาดตัวอย่างเพิ่มขึ้นมีผลทำให้ทุกวิธีมีอัตราความถูกต้องเพิ่มขึ้นภายใต้เกือบทุกเงื่อนไข โดยภาพรวมวิธีโพลี-ซิปเทสท์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ภายในขอบเขตที่กำหนด ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับและวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าขอบเขตที่กำหนด

วริษษา ชะม้อย (2550) ได้เปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาค ระหว่างวิธีโครงสร้างความแปรปรวนร่วม และค่าเฉลี่ย (Mean and Covariance Structures : MACS) กับวิธีการวิเคราะห์การจำแนกแบบโลจิสติก (Logistic dis criminant function analysis : LDFA) โดยศึกษาภายใต้ปัจจัยที่แปรเปลี่ยน 2 ปัจจัย คือ ความยาวของแบบทดสอบ 2 ระดับ ได้แก่ แบบทดสอบที่ตรวจให้คะแนนหลายค่าที่มีจำนวน 30 ข้อ และ 40 ข้อ และขนาดกลุ่มตัวอย่าง มี 3 ระดับ ได้แก่ขนาด 200 คน , 500 คน และ 1000 คน รวมข้อมูลที่ศึกษาทั้งหมด 6 เงื่อนไข (2×3) กลุ่มตัวอย่างที่ใช้ในการศึกษาครั้งนี้เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ปีการศึกษา 2550 โรงเรียนมัธยมศึกษา สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานกรุงเทพมหานคร เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ 5 ตัวเลือก ที่มีคำตอบถูกหลายคำตอบ และให้ตัดสินทุกคำตอบ (Multiple True-false Answer)

ผลการวิจัยพบว่าอำนาจการทดสอบทั้ง 6 เงื่อนไขจากการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิธี LDFA มีอำนาจการทดสอบในช่วงตั้งแต่ร้อยละ 62.5 ถึงร้อยละ 100 และวิธี MACS มีค่าอำนาจการทดสอบอยู่ในช่วงตั้งแต่ร้อยละ 37.5 ถึงร้อยละ 100 และพบว่าทั้งวิธี LDFAและวิธี MACS จะมีอำนาจการทดสอบในเงื่อนไขความยาวของแบบทดสอบ 30 ข้อ สูงกว่าเงื่อนไขความยาวของแบบทดสอบ 40 ข้อทุกเงื่อนไข เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่าทั้ง 6 เงื่อนไขของการทดสอบ วิธี LDFA มีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในช่วงตั้งแต่ร้อยละ 0 ถึงร้อยละ 15.6 และวิธี MACS มีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในช่วงตั้งแต่ร้อยละ 8.3 ถึงร้อยละ 21.9 โดยวิธี LDFA เกือบทุกเงื่อนไขมีค่าอยู่ในเกณฑ์ที่กำหนด ส่วนวิธี MACS พบว่าส่วนใหญ่มีค่าเกินกว่าเกณฑ์ที่กำหนด และพบว่าทั้งวิธี LDFA และวิธี MACS จะมีอัตราความคลาดเคลื่อนประเภทที่ 1 ในเงื่อนไขความยาวของแบบทดสอบ 30 ข้อ ต่ำกว่าเงื่อนไขความยาวของแบบทดสอบ 40 ข้อทุกเงื่อนไข

บทที่ 3
วิธีดำเนินการวิจัย

ในการวิจัยครั้งนี้มีความมุ่งหมาย เพื่อศึกษา และเปรียบเทียบวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่า ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความยาวของแบบทดสอบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนและขนาดตัวอย่าง ในการดำเนินการวิจัยมีขั้นตอนดังนี้

1. การจัดกระทำตัวแปร
2. การจำลองข้อมูล
3. การวิเคราะห์ข้อมูล

การจัดกระทำตัวแปร

การศึกษาวิจัยครั้งนี้ ศึกษาในสถานการณ์จำลองโดยใช้ทฤษฎีการตอบข้อสอบแบบมิติเดียวโมเดลพาเชียลเครดิตทั่วไป (Unidimensional generalized partial credit model) จำลองข้อมูลภายใต้การจัดกระทำตัวแปรอิสระ 4 ตัวแปรคือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน 2 รูปแบบ ความยาวของแบบทดสอบ 3 ขนาด สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 3 ขนาด และขนาดตัวอย่าง 3 ขนาด โดยจำลองผลการตอบแบบทดสอบที่มีโครงสร้างแบบมิติเดียว ข้อสอบแต่ละข้อมีรายการตอบ 5 รายการ ให้คะแนนเป็น 1,2,3,4 และ 5 สำหรับการจัดกระทำตัวแปรอิสระทั้ง 4 ตัวแปรดังกล่าว ผู้วิจัยได้ออกแบบการวิจัยโดยมีรายละเอียด ดังนี้

1. รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน (Form of DIF)

รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนจะขึ้นอยู่กับขนาดและทิศทางของข้อสอบทำหน้าที่เบี่ยงเบน ซึ่งแปรเปลี่ยนตามความแตกต่างของระดับความสามารถของผู้สอบ ในการวิจัยครั้งนี้ศึกษารูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนสองรูปแบบ คือ ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF) และข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) ในการศึกษาครั้งนี้กำหนดความแตกต่างของพารามิเตอร์อำนาจจำแนกของข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มสนใจ เท่ากับ 1.0 และความแตกต่างพารามิเตอร์ความยากของข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มสนใจ เท่ากับ 0.25 การกำหนดค่าดังกล่าวขึ้นอยู่กับเงื่อนไขของการจำลองรูปแบบของการทำหน้าที่เบี่ยงเบน กล่าวคือการจำลองข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF) กำหนดให้พารามิเตอร์ความยากของข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มสนใจมีค่าแปรเปลี่ยน โดยผู้วิจัยทำการเพิ่มค่าความยากของกลุ่มสนใจ 0.25 เมื่อเทียบกับกลุ่มอ้างอิง ในขณะที่พารามิเตอร์อำนาจจำแนกของข้อสอบระหว่างกลุ่มมีค่าคงที่ สำหรับข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) กำหนดให้พารามิเตอร์อำนาจจำแนก

ของข้อสอบระหว่างกลุ่มอ้างอิง และกลุ่มสนใจมีค่าแปรเปลี่ยน โดยผู้วิจัยทำการลดค่าอำนาจจำแนกของกลุ่มสนใจ 1.0 เมื่อเทียบกับกลุ่มอ้างอิง ในขณะที่พารามิเตอร์ความยากของข้อสอบระหว่างกลุ่มมีค่าคงที่

2.ความยาวของแบบทดสอบ (Test length)

ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ โดยทั่วไปผู้วิจัยมักนำปัจจัยความยาวของแบบทดสอบมาร่วมตรวจสอบด้วย ความยาวของแบบทดสอบที่นำมาตรวจสอบมีตั้งแต่ขนาดสั้นเพียง 20 ข้อไปจนถึงขนาดยาว 80 ข้อ (Swaminathan; & Rogers. 1990; Kristjansson. 2001) ในการวิจัยครั้งนี้ ผู้วิจัยจัดกระทำความยาวของแบบทดสอบ 3 ขนาด คือ 20 ข้อ, 40 ข้อ และ 60 ข้อ โดยใช้การจำลองแบบทดสอบที่มีโครงสร้างแบบมิติเดียว และแต่ละกลุ่มผู้สอบมีความสามารถไม่ต่างกัน

3. สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน (Proportion of DIF)

แบบทดสอบมาตรฐานวัดผลสัมฤทธิ์ทางการเรียน ถ้ามีข้อสอบทำหน้าที่เบี่ยงเบน 10%-15% ของจำนวนข้อสอบทั้งหมดถือว่าแบบทดสอบฉบับนั้นผิดปกติ (Narayanan; & Swaminathan. 1994: 319; citing Clauser.1993) นักวิจัยส่วนใหญ่ศึกษาปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนในช่วง 0% ถึง 20% ซึ่งสอดคล้องกับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนที่สามารถเกิดขึ้นกับข้อมูลจริง ในการศึกษาครั้งนี้ศึกษาสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 3 ขนาด คือ 10%, 20% และ 30% ของแต่ละความยาวแบบทดสอบ ดังนั้นในการจำลองแบบทดสอบฉบับที่มี 20 ข้อ ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน จำนวน 10%, 20% และ 30% จะต้องจัดกระทำข้อสอบทำหน้าที่ไม่เบี่ยงเบนกับข้อสอบทำหน้าที่เบี่ยงเบนจำนวน 18 ข้อกับ 2 ข้อ, 16 ข้อกับ 4 ข้อ และ 14 ข้อกับ 6 ข้อตามลำดับ ในการจำลองแบบทดสอบฉบับที่มี 40 ข้อ ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนจำนวน 10%, 20% และ 30% จะต้องจัดกระทำข้อสอบทำหน้าที่ไม่เบี่ยงเบนกับข้อสอบทำหน้าที่เบี่ยงเบน จำนวน 36 ข้อกับ 4 ข้อ, 32 ข้อกับ 8 ข้อ และ 28 ข้อกับ 12 ข้อตามลำดับ และในการจำลองแบบทดสอบฉบับที่มี 60 ข้อ ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน จำนวน 10%, 20% และ 30% จะต้องจัดกระทำข้อสอบทำหน้าที่ไม่เบี่ยงเบนกับข้อสอบทำหน้าที่เบี่ยงเบน จำนวน 54 ข้อกับ 6 ข้อ, 48 ข้อกับ 12 ข้อ และ 42 ข้อกับ 18 ข้อตามลำดับ

4. ขนาดตัวอย่าง (Sample size)

การศึกษาปัจจัยขนาดตัวอย่างที่ใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ จะแบ่งผู้สอบออกเป็นสองกลุ่มคือกลุ่มสนใจและกลุ่มอ้างอิง สำหรับกลุ่มสนใจเป็นกลุ่มเป้าหมายที่ต้องการศึกษา ส่วนกลุ่มอ้างอิงเป็นกลุ่มใช้สำหรับเปรียบเทียบ ตามปกติจะกำหนดกลุ่มสนใจเท่ากับหรือน้อยกว่ากลุ่มอ้างอิง ในการศึกษาครั้งนี้กำหนดให้ขนาดของกลุ่มสนใจ (N_F) และกลุ่มอ้างอิง (N_R) เท่ากัน จากการศึกษางานวิจัยพบว่าวิธีที่ใช้โมเดลการถดถอยโลจิสติกจะใช้ตัวอย่างที่มีขนาดเล็กจนถึงขนาดใหญ่ในช่วง 100 ถึง 3000 คน (วลีมาศ แซ่อึ้ง. 2543; อุทัยวรรณ สายพัฒนะ. 2547;

อรินทร์ น่วมถนอม. 2549; Bolt. 2002; Chang; Mazzeo; & Roussos.1996; French; & Miller.1996; Krisjansson; et al. 2005; Mazor; Hambleton; & Clauser. 1998; Narayanan; & Swaminathan. 1994, 1996; Russos; & Stout. 1996) ในการวิจัยครั้งนี้ศึกษาขนาดตัวอย่าง 3 ขนาด คือ $N_R = N_F = 250$, $N_R = N_F = 500$ และ $N_R = N_F = 1000$

การจำลองข้อมูล

การศึกษาครั้งนี้ผู้วิจัยใช้ข้อมูลจำลอง โดยจำลองข้อสอบภายใต้โมเดลพาร์เชียลเครดิตทั่วไป (Generalized Partial Credit Model: GPCM) แบบมิติเดียวของมูราคิ (Muraki. 1992) ที่มีขนาดความยาว 20 ข้อ, 40 ข้อ และ 60 ข้อ ในข้อสอบแต่ละข้อจะวัดความสามารถเป้าหมายเพียงมิติเดียวและแต่ละข้อมีรายการตอบ 5 รายการ ให้คะแนนเป็น 1,2,3,4 หรือ 5 ผู้วิจัยกำหนดสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 3 ขนาด คือขนาด 10%, 20% และ 30% และรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน 2 รูปแบบ ได้แก่ ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกันและข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน สำหรับโมเดลพาเชียลเครดิตทั่วไป (GPCM) ของมูราคิ ปรับขยายมาจากโมเดลพาร์เชียลเครดิต (Partial Credit Model: PCM) ของมาสเตอร์ (Master. 1982) เป็นดังนี้

$$P_{jk}(\theta) = \frac{\exp[\sum_{v=1}^{k} Da_j(\theta - b_{jv})]}{\sum_{c=1}^{m_j} \exp[\sum_{v=1}^{c} Da_j(\theta - b_{jv})]}$$

เมื่อ $P_{jk}(\theta)$ แทน ความน่าจะเป็นของผู้สอบที่มีความสามารถ θ จะตอบรายการที่ k จากรายการที่เป็นไปได้ m_j ของข้อสอบ j

D แทน ค่าปรับมาตรมีค่าเท่ากับ 1.7

a_j แทน พารามิเตอร์อำนาจจำแนกหรือพารามิเตอร์ความชัน

b_{jv}แทน พารามิเตอร์ระดับความยากของขั้นการตอบที่v (item step difficulty) ของข้อสอบ j

ดังนั้น $$P_{jk}(\theta) = \frac{\exp[\sum_{v=1}^{k} Da_j(\theta - b_j + d_v)]}{\sum_{c=1}^{m_j} \exp[\sum_{v=1}^{c} Da_j(\theta - b_j + d_v)]}$$

โดยที่ $b_{jv} = b_j - d_v$

เมื่อ b_j แทน พารามิเตอร์ตำแหน่งข้อสอบ (item location parameter)

d_v แทน ความยากสัมพัทธ์ของรายการคะแนน v

พร้อมทั้งจำลองขนาดตัวอย่าง โดยผู้วิจัยได้จำลองขนาดตัวอย่างของกลุ่มอ้างอิงและกลุ่มสนใจมีจำนวนเท่ากัน 3 ขนาด คือ 250คน, 500คน และ 1000 คน

สำหรับการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ และประมาณค่าพารามิเตอร์ข้อสอบจากความสามารถผู้สอบ ของข้อสอบที่มีการตรวจให้คะแนนแบบหลายค่าจะใช้โปรแกรมวินเจน2 (WinGen2) ของฮานและแฮมเบิลตัน (Han; & Hambleton. 2007) จำลองผลการตอบข้อสอบ ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน 2 รูปแบบ (ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน) ความยาวของแบบทดสอบ 3 ขนาด (20 ข้อ, 40 ข้อ และ 60 ข้อ) สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 3 ขนาด (10%, 20% และ 30%) และขนาดตัวอย่าง 3 ขนาด (250 คน, 500 คน, 1000 คน) ซึ่งมีขั้นตอนหลัก 3 ขั้นตอน ดังนี้

ขั้นที่ 1 การสร้างข้อมูลผู้สอบ

ระบุจำนวนผู้สอบของขนาดตัวอย่าง เช่น 250 คน (ในการศึกษาครั้งนี้ใช้ขนาดตัวอย่าง 250คน, 500คน และ 1000 คน) แล้วเลือกชนิดของการแจกแจงคุณลักษณะหรือความสามารถของผู้สอบ โดยเลือกการแจกแจงแบบโค้งปกติ พร้อมทั้งระบุค่าเฉลี่ยของความสามารถผู้สอบ และส่วนเบี่ยงเบนมาตรฐานเป็น 0 และ 1 ตามลำดับ

ขั้นที่ 2 การสร้างข้อมูลข้อสอบ

ระบุจำนวนของข้อสอบในแบบทดสอบ เช่น 20 ข้อ (ในการศึกษาครั้งนี้ใช้แบบทดสอบที่มีความยาว 20 ข้อ, 40 ข้อ และ 60 ข้อ) จากนั้นระบุจำนวนรายการคำตอบโดยเลือก ‘5’ สำหรับข้อมูลที่มีการตรวจให้คะแนนแบบหลายค่าของโมเดลพาร์เชียลเครดิตทั่วไป (Generalized Partial Credit Model: GPCM) ในการประมาณค่าพารามิเตอร์ ต่อจากนั้นเลือกการแจกแจงของพารามิเตอร์ข้อสอบและระบุคุณลักษณะของการแจกแจง โดยเลือกพารามิเตอร์ a ที่มีการแจกแจงแบบโค้งปกติ มีค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 และเลือกพารามิเตอร์ b ที่มีการแจกแจงแบบโค้งปกติ มีค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 เช่นกัน

การศึกษาครั้งนี้ ผู้วิจัยสนใจศึกษารูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน 2 รูปแบบคือ ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และที่ไม่เป็นรูปแบบเดียวกัน รูปแบบแรกกำหนดให้พารามิเตอร์ความยากของข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มสนใจมีค่าแตกต่างกัน 0.25 นั่นคือ $b_{jF} = b_{jR} + .25$ ส่วนพารามิเตอร์อำนาจจำแนกของข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มสนใจมีค่าเท่ากัน นั่นคือ $a_{jF} = a_{jR}$ สำหรับรูปแบบหลังกำหนดให้อำนาจจำแนกข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มสนใจมีค่าต่างกัน 1.0 นั่นคือ $a_{jF} = a_{jR} -1.0$ ส่วนพารามิเตอร์ความยากของข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มสนใจมีค่าเท่ากัน นั่นคือ $b_{jF} = b_{jR}$ โดยมีสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 3 ขนาด คือ ขนาด 10%, 20% และ 30% ของแต่ละขนาดความยาวแบบทดสอบ

ขั้นที่ 3 การสร้างข้อมูลการตอบข้อสอบ

ในขั้นตอนนี้เป็นการนำเซตข้อมูลผู้ตอบที่สร้างขึ้นในขั้นที่ 1 และเซตข้อมูลพารามิเตอร์ข้อสอบที่สร้างขึ้นในขั้นที่ 2 มาจัดทำข้อมูลรายการคำตอบของกลุ่มตัวอย่างที่จะทำการศึกษา ในแต่ละเงื่อนไขจะทำข้อมูลซ้ำ 50 ชุด ซึ่งได้ข้อมูลที่เป็นไปตามเงื่อนไขทั้งสิ้น 2,700 ชุด

ในการศึกษาครั้งนี้ใช้โปรแกรม WinGen2 สำหรับแผนผังของการจำลองข้อมูลทั้ง 4 ปัจจัย แสดงในภาพประกอบ 6

รูปแบบของ DIF	ความยาวของแบบทดสอบ	สัดส่วนของ DIF	ขนาดตัวอย่าง
ที่เป็นรูปแบบเดียวกัน	20 ข้อ	10%	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1000$
		20%	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1000$
		30%	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1000$
	40 ข้อ	10%	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1000$
		20%	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1000$
		30%	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1000$
	60 ข้อ	10%	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1000$
		20%	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1000$
		30%	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1000$

ภาพประกอบ 6 แผนผังของการจำลองข้อมูล

รูปแบบของ DIF	ความยาวของแบบทดสอบ	สัดส่วนของ DIF	ขนาดตัวอย่าง
ที่ไม่เป็นรูปแบบเดียวกัน	20 ข้อ	10%	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1000$
		20%	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1000$
		30%	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1000$
	40 ข้อ	10%	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1000$
		20%	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1000$
		30%	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1000$
	60 ข้อ	10%	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1000$
		20%	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1000$
		30%	$N_R = N_F = 250$
			$N_R = N_F = 500$
			$N_R = N_F = 1000$

ภาพประกอบ 6 (ต่อ)

การวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลมีรายละเอียดดังนี้

1. การวิเคราะห์การทำหน้าที่เบี่ยงเบนของข้อสอบ

การวิเคราะห์การทำหน้าที่เบี่ยงเบนของข้อสอบ ด้วยวิธีโพลีโตมัสซิบเทสท์ (Poly-SIB) วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก (LDFA) และวิธีการถดถอยโลจิสติกแบบจัดอันดับ (OLR) จะใช้ผลการจำลองการตอบข้อสอบซึ่งเป็นคะแนนดิบมาวิเคราะห์ โดยวิธีแรกใช้โปรแกรม

DIFPACK ส่วนสองวิธีหลังใช้โปรแกรม MULTILOG และ SPSS ตามลำดับ ทั้งสามวิธีจะนำผลการตอบข้อสอบรายข้อมาวิเคราะห์และใช้คะแนนรวมของแบบทดสอบแทนระดับความสามารถ ดังนี้

1.1 การวิเคราะห์การทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีโพลีโตมัสซิบเทสท์

ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีโพลีโตมัสซิบเทสท์ ผู้วิจัยใช้โปรแกรม DIFPACK ซึ่งมีสูตรในการคำนวณ ดังนี้

$$\hat{\beta} = \sum_{k=0}^{n_H} P_k d_k$$

โดยที่ $P_k = \frac{N_{Rk} + N_{Fk}}{N}$

เมื่อ P_k แทนสัดส่วนของผู้สอบทั้งหมด (กลุ่มอ้างอิง และกลุ่มสนใจ) ซึ่งตอบแบบทดสอบที่ใช้ในการจับคู่ $X_1, X_2, ..., X_n$ แล้วได้คะแนน $X = k$

$$d_k = \overline{Y}_{Rk} - \overline{Y}_{Fk}, k = 0,...,n_H$$

เมื่อ $\overline{Y}_{Rk}$ และ $\overline{Y}_{Fk}$ แทนค่าเฉลี่ยของคะแนนข้อสอบที่ต้องการศึกษาของผู้สอบซึ่งได้คะแนน $X = k$ ของกลุ่มอ้างอิงและกลุ่มสนใจ

สมมติฐานของการทดสอบดัชนี DIF กำหนดดังนี้

$H_0 : \beta = 0$

$H_1 : \beta > 0$

นำดัชนี $\hat{\beta}$ มาทดสอบสมมุติฐานศูนย์ โดยใช้สถิติ B ดังนี้

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

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

เมื่อ $\hat{\sigma}(\hat{\beta})$ แทน ค่าประมาณความคลาดเคลื่อนมาตรฐานของ β

$\hat{\sigma}^2(Y|k,g)$ แทน ค่าประมาณความแปรปรวนของคะแนนจากแบบทดสอบที่ต้องการศึกษาในกลุ่ม g (R หรือ F) ซึ่งมีคะแนนรวมเท่ากับ k

ผู้วิจัยใช้เกณฑ์ตัดสินข้อสอบทำหน้าที่เบี่ยงเบนที่ระดับนัยสำคัญ .05 ถ้าผลการทดสอบพบว่า $|B| > Z_{1-\frac{\alpha}{2}}$ อย่างมีนัยสำคัญที่ระดับ .05 แสดงว่าปฏิเสธ H_0 นั่นคือข้อสอบที่นำมาตรวจทำหน้าที่เบี่ยงเบน (DIF) โดยเข้าข้างผู้สอบกลุ่มใดกลุ่มหนึ่ง

1.2 การวิเคราะห์การทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก

ในการวิเคราะห์ดัชนีการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก ผู้วิจัยใช้โปรแกรม MULTILOG ซึ่งมีสูตรในการคำนวณ ดังนี้

$$\text{logit} = \ln\left[\frac{P(G=1)}{P(G=0)}\right] = \alpha_0 + \alpha_1 X + \alpha_2 U + \alpha_3 (X * U)$$

เมื่อ $\alpha_i (i = 0,1,2,3)$ แทน สัมประสิทธิ์ฟังก์ชันการจำแนก

G แทน กลุ่มผู้สอบ

X แทน ระดับความสามารถของผู้สอบ

U แทน ผลการตอบข้อสอบ

P แทน ความน่าจะเป็นของการเป็นสมาชิกกลุ่ม

การทดสอบข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF) และที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) สามารถทดสอบได้พร้อมกัน โดยกำหนดสมมุติฐานศูนย์ (H_0) และสมมุติฐานทางเลือก (H_1) ดังนี้

$$H_0 : \alpha_2 = 0 \text{ และ } \alpha_3 = 0$$

$$H_1 : \alpha_2 \neq 0 \text{ หรือ } \alpha_3 \neq 0$$

การทดสอบนัยสำคัญทางสถิติของการทำหน้าที่เบี่ยงเบนของข้อสอบจะดำเนินการตามกระบวนการการทดสอบความเหมาะสมของข้อมูลกับโมเดล 3 ขั้นตอน โดยใช้สถิติ G^2 โดยที่

$$G^2_{\text{diff}} = -2(\ln L_{(i+1)} - \ln L_{(i)})$$

เมื่อ $L_{(i+1)}$ และ $L_{(i)}$ แทนความน่าจะเป็นของโมเดลลำดับขั้นที่อยู่ติดกัน

สถิติ G^2 ดังกล่าวมีการแจกแจงแบบเชิงเส้นกำกับ เป็นไค-สแควร์ที่ระดับชั้นความเป็นอิสระเท่ากับ 1 (df = 1) ในการศึกษาครั้งนี้ผู้วิจัยใช้เกณฑ์ตัดสินข้อสอบทำหน้าที่เบี่ยงเบนที่ระดับนัยสำคัญ .05

1.3 การวิเคราะห์การทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับ

ในการวิเคราะห์ดัชนีการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับ ผู้วิจัยใช้โปรแกรม SPSS ซึ่งมีสูตรในการคำนวณ ดังนี้

$$\text{logit} = \ln\left[\frac{P(Y \leq j)}{P(Y > j)}\right] = \alpha_j + \beta_1 X + \beta_2 G + \beta_3 (X * G)$$

เมื่อ $P(Y \leq j)$ แทน ความน่าจะเป็นสะสมสำหรับผู้ตอบในรายการที่ j หรือต่ำกว่า

α_j แทน พารามิเตอร์ส่วนตัดของความน่าจะเป็นสะสมในรายการที่ j

X แทน ระดับความสามารถของผู้สอบ

G แทน กลุ่มผู้สอบ

β_1 แทน พารามิเตอร์ที่เกี่ยวข้องกับระดับความสามารถของผู้สอบ

β_2 แทน พารามิเตอร์ที่เกี่ยวข้องกับกลุ่มผู้สอบ

β_3 แทน พารามิเตอร์ที่เกี่ยวข้องกับปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับกลุ่มผู้สอบ

การทดสอบการทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน (Uniform DIF) และที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) สามารถทดสอบได้พร้อมกัน โดยกำหนดสมมุติฐานศูนย์ (H_0) และสมมุติฐานทางเลือก (H_1) ดังนี้

$$H_0 : \beta_2 = 0 \text{ และ } \beta_3 = 0$$

$$H_1 : \beta_2 \neq 0 \text{ หรือ } \beta_3 \neq 0$$

การทดสอบสมมติฐานจะใช้การทดสอบความเหมาะสมของข้อมูลกับโมเดล 3 ขั้นตอน โดยใช้สถิติไค-สแควร์ที่ระดับชั้นความเป็นอิสระเท่ากับ 2 (df = 2) ในการศึกษาครั้งนี้ผู้วิจัยใช้เกณฑ์ตัดสินข้อสอบทำหน้าที่เบี่ยงเบนที่ระดับนัยสำคัญ .05

2. การคำนวณอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1

การคำนวณอัตราความถูกต้อง (Power rate) และอัตราความคลาดเคลื่อนประเภทที่ 1 (Type I error) ของวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติกแบบจัดอันดับ ใช้สูตรการคำนวณดังนี้

$$P = \frac{n_1}{N_1}$$

$$E = \frac{n_2}{N_2}$$

เมื่อ P แทน อัตราความถูกต้อง

E แทน อัตราความคลาดเคลื่อนประเภทที่ 1

n_1 แทน จำนวนข้อสอบที่ตรวจสอบได้ถูกต้องว่าทำหน้าที่เบี่ยงเบน

n_2 แทน จำนวนข้อสอบที่ตรวจสอบผิดพลาดว่าทำหน้าที่เบี่ยงเบน

N_1 แทน จำนวนข้อสอบทำหน้าที่เบี่ยงเบนในแบบทดสอบ

N_2 แทน จำนวนข้อสอบทำหน้าที่ไม่เบี่ยงเบนในแบบทดสอบ

3. การเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1

การเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความยาวของแบบทดสอบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และขนาดตัวอย่าง ซึ่งในการทดสอบปัจจัยที่ศึกษาซึ่งได้แก่รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน (ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน) ความยาวของแบบทดสอบ (20 ข้อ, 40 ข้อและ 60 ข้อ) สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน (10%, 20% และ 30%) และขนาดตัวอย่าง (250 คน, 500 คน และ 1000 คน) ว่ามีผลต่ออัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในแต่ละวิธีของทั้ง 3 วิธี ซึ่งได้แก่ วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับหรือไม่นั้น ผู้วิจัยใช้การทดสอบที (t-test) ทดสอบเงื่อนไขของปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน และการทดสอบเอฟ (F-test) จากการวิเคราะห์ความแปรปรวนทางเดียว (Oneway ANOVA) ทดสอบเงื่อนไขของปัจจัยความยาวของแบบทดสอบ เงื่อนไขของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และเงื่อนไขของปัจจัยขนาดตัวอย่าง และเมื่อมีนัยสำคัญทางสถิติแล้ว ผู้วิจัยทำการทดสอบความแตกต่างระหว่างค่าเฉลี่ยหลังการวิเคราะห์ความแปรปรวนทางเดียว (ทดสอบความแตกต่างรายคู่) โดยใช้วิธีการทดสอบของเชฟเฟ สำหรับการเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบระหว่างวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความยาวของแบบทดสอบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และขนาดตัวอย่าง ผู้วิจัยใช้การทดสอบเอฟ (F-test) จากการวิเคราะห์ความแปรปรวนทางเดียว (Oneway ANOVA) ทดสอบความแตกต่างของวิธีตรวจสอบทั้ง 3 วิธี ภายใต้เงื่อนไขของแต่ละปัจจัย ซึ่งได้แก่เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกันและข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกันของปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน เงื่อนไขความยาวของแบบทดสอบ 20 ข้อ, 40 ข้อและ 60 ข้อของปัจจัยความยาวของแบบทดสอบ เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10%, 20% และ 30% ของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และเงื่อนไขขนาดตัวอย่าง 250 คน, 500 คน และ 1000 คน ของปัจจัยขนาดตัวอย่าง และเมื่อมีนัยสำคัญทางสถิติแล้ว ผู้วิจัยทำการทดสอบความแตกต่างระหว่างค่าเฉลี่ยหลังการวิเคราะห์ความแปรปรวนทางเดียว (ทดสอบความแตกต่างรายคู่) โดยใช้วิธีการทดสอบของเชฟเฟ

บทที่ 4
ผลการวิเคราะห์ข้อมูล

การวิจัยครั้งนี้มีความมุ่งหมายเพื่อศึกษา และเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่า ของวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน 2 รูปแบบ ความยาวของแบบทดสอบ 3 ขนาด สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 3 ขนาด และขนาดตัวอย่าง 3 ขนาด ข้อมูลที่ศึกษาได้จากการจำลองด้วยโปรแกรม WinGen2 ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน รวมข้อมูลที่ใช้ศึกษาทั้งหมด 54 เงื่อนไข (2×3×3×3) และในแต่ละเงื่อนไขจำลองซ้ำ 50 ครั้ง

ผลการวิเคราะห์ข้อมูลในบทนี้ ผู้วิจัยได้นำเสนอเป็น 2 ตอนดังนี้

ตอนที่ 1 ค่าสถิติพื้นฐานของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ

ตอนที่ 2 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่า

สัญลักษณ์และความหมายที่ใช้ในการวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลและการแปลผลการวิเคราะห์ข้อมูลในการวิจัยครั้งนี้เพื่อให้การเสนอผลการวิเคราะห์ข้อมูลมีความสะดวกและเกิดความเข้าใจตรงกัน ผู้วิจัยจึงได้กำหนดสัญลักษณ์และความหมายที่ใช้ในการวิเคราะห์ข้อมูล ดังนี้

DIF	แทน การทำหน้าที่เบี่ยงเบนของข้อสอบ
Proc	แทน วิธีตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ
Poly-SIB	แทน การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีโพลีโตมัสซิบเทสท์
LDFA	แทน การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก
OLR	แทน การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับ
FDIF	แทน รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน
UDIF	แทน ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน
NDIF	แทน ข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน
TL	แทน ความยาวของแบบทดสอบ
TL20	แทน แบบทดสอบที่มีความยาว 20 ข้อ

TL40 แทน แบบทดสอบที่มีความยาว 40 ข้อ

TL60 แทน แบบทดสอบที่มีความยาว 60 ข้อ

PDIF แทน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน

PDIF10 แทน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10%

PDIF20 แทน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20%

PDIF30 แทน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30%

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

S250 แทน ขนาดตัวอย่าง 250 คน

S500 แทน ขนาดตัวอย่าง 500 คน

S1000 แทน ขนาดตัวอย่าง 1000 คน

P แทน อัตราความถูกต้อง

E แทน อัตราความคลาดเคลื่อนประเภทที่ 1

ตอนที่ 1 ค่าสถิติพื้นฐานของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ

ผู้วิจัยนำข้อมูลที่ได้จากการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความยาวของแบบทดสอบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และขนาดตัวอย่าง มาคำนวณอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ผลการคำนวณค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีทั้ง 3 วิธีภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย แสดงดังตาราง 1

ตาราง 1 ค่าสถิติพื้นฐานของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบ DIF 3 วิธี ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัยหลัก

ปัจจัย	วิธีการตรวจสอบ DIF					
	Poly-SIB		LDFA		OLR	
	P	E	P	E	P	E
รูปแบบของ DIF						
UDIF	.651	.402	.981	.014	.525	.154
	(.476)	(.490)	(.135)	(.116)	(.499)	(.361)
NDIF	.435	.320	.978	.017	.590	.211
	(.496)	(.467)	(.149)	(.131)	(.492)	(.408)
ความยาวของแบบทดสอบ						
20 ข้อ	.625	.190	.942	.047	.569	.371
	(.484)	(.392)	(.233)	(.211)	(.495)	(.483)
40 ข้อ	.699	.524	.974	.022	.942	.232
	(.459)	(.499)	(.157)	(.146)	(.233)	(.422)
60 ข้อ	.412	.310	.995	.001	.297	.087
	(.492)	(.462)	(.071)	(.034)	(.457)	(.282)
สัดส่วนของ DIF						
10%	.456	.309	.984	.013	.380	.075
	(.498)	(.462)	(.126)	(.111)	(.486)	(.263)
20%	.491	.361	.976	.020	.498	.199
	(.500)	(.480)	(.153)	(.139)	(.500)	(.399)
30%	.607	.429	.980	.015	.656	.302
	(.488)	(.495)	(.139)	(.121)	(.475)	(.459)
ขนาดตัวอย่าง						
$N_R = N_F = 250$	.462	.265	.969	.024	.472	.132
	(.499)	(.441)	(.172)	(.153)	(.499)	(.339)
$N_R = N_F = 500$	.563	.370	.979	.015	.569	.167
	(.496)	(.483)	(.142)	(.120)	(.495)	(.373)
$N_R = N_F = 1000$	.604	.448	.989	.008	.631	.248
	(.489)	(.497)	(.103)	(.090)	(.482)	(.432)
เฉลี่ย	.543	.361	.979	.016	.557	.182

ผลการวิเคราะห์ตามตาราง 1 สรุปได้ดังนี้

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ**ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน** พบว่า วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ภายใต้เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกันมีอัตราความถูกต้องคิดเป็นค่าเฉลี่ยประมาณ .651, .981 และ .525 โดยมีส่วนเบี่ยงเบนมาตรฐานเป็น .476, .135 และ .499 ตามลำดับ และภายใต้เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกันมีอัตราความถูกต้องคิดเป็นค่าเฉลี่ยประมาณ .435, .978 และ .590 โดยมีส่วนเบี่ยงเบนมาตรฐานเป็น .496, .149 และ .492 ตามลำดับ เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่าวิธีตรวจสอบทั้ง 3 วิธี ภายใต้เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกันมีอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นค่าเฉลี่ยประมาณ .402, .014 และ .154 โดยมีส่วนเบี่ยงเบนมาตรฐานเป็น .490, .116 และ .361 ตามลำดับ และภายใต้เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกันมีอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นค่าเฉลี่ยประมาณ .320, .017 และ .211 โดยมีส่วนเบี่ยงเบนมาตรฐานเป็น .467, .131 และ .408 ตามลำดับ โดยภาพรวมวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ภายใต้รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนมีอัตราความถูกต้องคิดเป็นค่าเฉลี่ยประมาณ .435 - .981 ซึ่งวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงสุดคิดเป็นค่าเฉลี่ยประมาณ .981ในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน วิธีโพลีโตมัสซิบเทสท์มีอัตราความถูกต้องต่ำสุดคิดเป็นค่าเฉลี่ยประมาณ .435 ในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ทั้ง 3 วิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นค่าเฉลี่ยประมาณ .014 - .402 ซึ่งวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุดคิดเป็นค่าเฉลี่ยประมาณ .014 ในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และวิธีโพลีโตมัสซิบเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุดคิดเป็นค่าเฉลี่ยประมาณ .402 ในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน **ภายใต้ปัจจัยความยาวของแบบทดสอบ** พบว่า วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ภายใต้เงื่อนไขความยาวของแบบทดสอบ 20 ข้อ มีอัตราความถูกต้องคิดเป็นค่าเฉลี่ยประมาณ .625, .942 และ .569 โดยมีส่วนเบี่ยงเบนมาตรฐานเป็น .484, .233 และ .495 ตามลำดับ ภายใต้เงื่อนไขความยาวของแบบทดสอบ 40 ข้อมีอัตราความถูกต้องคิดเป็นค่าเฉลี่ยประมาณ .699, .974 และ .942 โดยมีส่วนเบี่ยงเบนมาตรฐานเป็น .459, .157 และ .233 ตามลำดับ และภายใต้เงื่อนไขความยาวของแบบทดสอบ 60 ข้อมีอัตราความถูกต้องคิดเป็นค่าเฉลี่ยประมาณ .412, .995 และ .297 โดยมีส่วนเบี่ยงเบนมาตรฐานเป็น .492, .071 และ .457 ตามลำดับ เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีตรวจสอบทั้ง 3 วิธี ภายใต้เงื่อนไขความยาวของแบบทดสอบ 20 ข้อมีอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นค่าเฉลี่ยประมาณ .190, .047 และ .371 โดยมีส่วนเบี่ยงเบนมาตรฐานเป็น .392, .211 และ .483 ตามลำดับ ภายใต้เงื่อนไขความยาวของแบบทดสอบ 40 ข้อมี

อัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นค่าเฉลี่ยประมาณ .524, .022 และ .232 โดยมีส่วนเบี่ยงเบนมาตรฐานเป็น .499, .146 และ .422 ตามลำดับ และภายใต้เงื่อนไขความยาวของแบบทดสอบ 60 ข้อมีอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นค่าเฉลี่ยประมาณ .310, .001 และ .087 โดยมีส่วนเบี่ยงเบนมาตรฐานเป็น .462, .034 และ .282 ตามลำดับ โดยภาพรวมวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีอัตราความถูกต้องคิดเป็นค่าเฉลี่ยประมาณ .297 - .995 ซึ่งวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงสุดคิดเป็นค่าเฉลี่ยประมาณ .995 วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องต่ำสุดคิดเป็นค่าเฉลี่ยประมาณ .297 ในเงื่อนไขความยาวของแบบทดสอบ 60 ข้อ ทั้ง 3 วิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นค่าเฉลี่ยประมาณ .001 - .524 ซึ่ง วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุดคิดเป็นค่าเฉลี่ยประมาณ .001 ในเงื่อนไขความยาวของแบบทดสอบ 60 ข้อ และวิธีโพลีโตมัสซิบเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุดคิดเป็นค่าเฉลี่ยประมาณ .524 ในเงื่อนไขความยาวของแบบทดสอบ 40 ข้อ **ภายใต้ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน** พบว่า วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน10% มีอัตราความถูกต้องคิดเป็นค่าเฉลี่ยประมาณ .456, .984 และ .380 โดยมีส่วนเบี่ยงเบนมาตรฐานเป็น .498, .126 และ .486 ตามลำดับ ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% มีอัตราความถูกต้องคิดเป็นค่าเฉลี่ยประมาณ .491, .976 และ .498 โดยมีส่วนเบี่ยงเบนมาตรฐานเป็น.500, .153 และ .500 ตามลำดับ และภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% มีอัตราความถูกต้องคิดเป็นค่าเฉลี่ยประมาณ .607, .980 และ .656 โดยมีส่วนเบี่ยงเบนมาตรฐานเป็น .488, .139 และ .475 ตามลำดับ เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่าวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% มีอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นค่าเฉลี่ยประมาณ .309, .013 และ .075 โดยมีส่วนเบี่ยงเบนมาตรฐานเป็น .462, .111 และ .263 ตามลำดับ ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% มีอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นค่าเฉลี่ยประมาณ .361, .020 และ .199 โดยมีส่วนเบี่ยงเบนมาตรฐานเป็น .480, .139 และ .399 ตามลำดับ และภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% มีอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นค่าเฉลี่ยประมาณ .429, .015 และ .302 โดยมีส่วนเบี่ยงเบนมาตรฐานเป็น .495, .121 และ .459 ตามลำดับ โดยภาพรวมวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องคิดเป็นค่าเฉลี่ยประมาณ .380 - .984 ซึ่งวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงสุดคิดเป็นค่าเฉลี่ยประมาณ .984 ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% วิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องต่ำสุดคิดเป็นค่าเฉลี่ยประมาณ .380 ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% ทั้ง 3 วิธีมีอัตรา

ความคลาดเคลื่อนประเภทที่ 1 คิดเป็นค่าเฉลี่ยประมาณ .013 - .429 ซึ่งวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุดคิดเป็นค่าเฉลี่ยประมาณ .013 ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% และวิธีโพลีโตมัสซิบเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุดคิดเป็นค่าเฉลี่ยประมาณ .429 ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% **ภายใต้ปัจจัยขนาดตัวอย่าง** พบว่าวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ภายใต้เงื่อนไขขนาดตัวอย่าง 250 คน มีอัตราความถูกต้องคิดเป็นค่าเฉลี่ยประมาณ .462, .969 และ .472 โดยมีส่วนเบี่ยงเบนมาตรฐานเป็น .499, .172 และ .499 ตามลำดับ ภายใต้เงื่อนไขขนาดตัวอย่าง 500 คนมีอัตราความถูกต้องคิดเป็นค่าเฉลี่ยประมาณ .563, .979 และ .569 โดยมีส่วนเบี่ยงเบนมาตรฐานเป็น .496, .142 และ .495 ตามลำดับ และภายใต้เงื่อนไขขนาดตัวอย่าง 1000 คนมีอัตราความถูกต้องคิดเป็นค่าเฉลี่ยประมาณ .604, .989 และ .631 โดยมีส่วนเบี่ยงเบนมาตรฐานเป็น .489, .103 และ .482 ตามลำดับ เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่าวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับภายใต้เงื่อนไขขนาดตัวอย่าง 250 คนมีอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นค่าเฉลี่ยประมาณ .265, .024 และ .132 โดยมีส่วนเบี่ยงเบนมาตรฐานเป็น .441, .153 และ .339 ตามลำดับ ภายใต้เงื่อนไขขนาดตัวอย่าง 500 คนมีอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นค่าเฉลี่ยประมาณ .370, .015 และ .167 โดยมีส่วนเบี่ยงเบนมาตรฐานเป็น .483, .120 และ .373 ตามลำดับ และภายใต้เงื่อนไขขนาดตัวอย่าง 1000 คนมีอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นค่าเฉลี่ยประมาณ .448, .008 และ .248 โดยมีส่วนเบี่ยงเบนมาตรฐานเป็น .497, .090 และ .432 ตามลำดับ โดยภาพรวมวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องคิดเป็นค่าเฉลี่ยประมาณ .462 - .989 ซึ่งวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงสุดคิดเป็นค่าเฉลี่ยประมาณ .989 ในเงื่อนไขขนาดตัวอย่าง 1000 คน วิธีโพลีโตมัสซิบเทสท์มีอัตราความถูกต้องต่ำสุดคิดเป็นค่าเฉลี่ยประมาณ .462 ในเงื่อนไขขนาดตัวอย่าง 250 คน ทั้ง 3 วิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นค่าเฉลี่ยประมาณ .008 - .448 ซึ่งวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุดคิดเป็นค่าเฉลี่ยประมาณ .008 ในเงื่อนไขขนาดตัวอย่าง 1000 คน และวิธีโพลีโตมัสซิบเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุดคิดเป็นค่าเฉลี่ยประมาณ .448 ในเงื่อนไขขนาดตัวอย่าง 1000 คน

ตอนที่ 2 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่า

2.1 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ

การวิจัยตอนนี้ผู้วิจัยนำผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ปัจจัยความยาวของแบบทดสอบ ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และปัจจัยขนาดตัวอย่าง ซึ่งได้แก่อัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 มาวิเคราะห์ความแตกต่างเพื่อเปรียบเทียบผลภายใต้แต่ละวิธี โดยใช้การทดสอบที (t-test) ทดสอบเงื่อนไขของปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน และการทดสอบเอฟ (F-test) จากการวิเคราะห์ความแปรปรวนทางเดียว (Oneway ANOVA) ทดสอบเงื่อนไขของปัจจัยความยาวของแบบทดสอบ เงื่อนไขของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และเงื่อนไขของปัจจัยขนาดตัวอย่าง และเมื่อมีนัยสำคัญทางสถิติแล้วผู้วิจัยทำการทดสอบความแตกต่างระหว่างค่าเฉลี่ยหลังการวิเคราะห์ความแปรปรวนทางเดียว (ทดสอบความแตกต่างรายคู่) โดยใช้วิธีการทดสอบของเชฟเฟ่ ผลปรากฏดังตาราง 2-22

ตาราง 2 เปรียบเทียบอัตราความถูกต้อง และ อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบ DIF ของปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีโพลีโตมัสซิบเทสท์

		Mean	S.D.	t	df	p-value
P	UDIF	.651	.476	32.775	21598	.000
	NDIF	.435	.496			
E	UDIF	.402	.490	25.101	86398	.000
	NDIF	.320	.467			

ผลการวิเคราะห์ข้อมูลตามตาราง 2 พบว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีโพลีโตมัสซิบเทสท์แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีโพลีโตมัสซิบเทสท์มีความแตกต่างกัน กล่าวคือ การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของเงื่อนไขรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกันด้วยวิธีโพลีโตมัสซิบเทสท์ มีอัตราความถูกต้องในการตรวจพบสูงกว่าเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบ

เดียวกัน เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีโพลีโตมัสซิบเทสท์ มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีโพลีโตมัสซิบเทสท์มีความแตกต่างกัน กล่าวคือ การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกันด้วยวิธีโพลีโตมัสซิบเทสท์มีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนสูงกว่าเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน

ตาราง 3 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบ DIF ของปัจจัยความยาวของแบบทดสอบด้วยวิธีโพลีโตมัสซิบเทสท์

	อัตราความถูกต้อง	อัตราความคลาดเคลื่อนประเภทที่ 1
TL20	.625	.190
TL40	.699	.524
TL60	.412	.310
MS_b	192.467	647.648
df_b	2	2
MS_w	.230	.216
df_w	21597	86397
F-ratio	835.603	3002.318
p-value	.000	.000

ผลการวิเคราะห์ข้อมูลตามตาราง 3 พบว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยความยาวของแบบทดสอบ (20 ข้อ, 40 ข้อ และ 60 ข้อ) ด้วยวิธีโพลีโตมัสซิบเทสท์มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยความยาวของแบบทดสอบด้วยวิธีโพลีโตมัสซิบเทสท์มีความแตกต่างกัน และเมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบ พบว่า การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยความยาวของแบบทดสอบด้วยวิธีโพลีโตมัสซิบเทสท์มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยความยาวของแบบทดสอบด้วยวิธีโพลีโตมัสซิบเทสท์มีความแตกต่างกัน ดังนั้นผู้วิจัยจึงทำการทดสอบความแตกต่างรายคู่ด้วยวิธีการของเชฟเฟ่ ผลปรากฏดังตาราง 4

ตาราง 4 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ ในการตรวจสอบ DIF ของปัจจัยความยาวของแบบทดสอบด้วยวิธีโพลีโตมัสซิบเทสท์

	(I) TL	(J) TL	Mean difference	p-value
อัตราความถูกต้อง	20	40	-.074	.000
		60	.213	.000
	40	60	.287	.000
อัตราความคลาด	20	40	-.334	.000
เคลื่อนประเภทที่ 1		60	-.120	.000
	40	60	.214	.000

ผลการวิเคราะห์ข้อมูลตามตาราง 4 อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยความยาวของแบบทดสอบ (20 ข้อ, 40 ข้อ และ 60 ข้อ) ด้วยวิธีโพลีโตมัสซิบเทสท์ พบว่าที่ระดับความยาวของแบบทดสอบ 20 ข้อมีอัตราความถูกต้องในการตรวจสอบต่ำกว่าที่ระดับความยาวของแบบทดสอบ 40 ข้อ แต่สูงกว่าที่ระดับความยาวของแบบทดสอบ 60 ข้อ และที่ระดับความยาวของแบบทดสอบ 40 ข้อ มีอัตราความถูกต้องในการตรวจสอบสูงกว่าที่ระดับความยาวของแบบทดสอบ 60 ข้ออย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ที่ระดับความยาวของแบบทดสอบ 20 ข้อมีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าที่ระดับความยาวของแบบทดสอบ 40 ข้อแต่สูงกว่าที่ระดับความยาวของแบบทดสอบ 60 ข้อ และที่ระดับความยาวของแบบทดสอบ 40 ข้อ มีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนสูงกว่าที่ระดับความยาวของแบบทดสอบ 60 ข้อ เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ที่ระดับความยาวของแบบทดสอบ 20 ข้อ มีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบต่ำกว่าที่ระดับความยาวของแบบทดสอบ 40 ข้อ และ 60 ข้อ และที่ระดับความยาวของแบบทดสอบ 40 ข้อมีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบสูงกว่าที่ระดับความยาวของแบบทดสอบ 60 ข้ออย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ที่ระดับความยาวของแบบทดสอบ 20 ข้อ มีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าที่ระดับความยาวของแบบทดสอบ 40 ข้อ และ 60 ข้อ และที่ระดับความยาวของแบบทดสอบ 40 ข้อ มีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนสูงกว่าที่ระดับความยาวของแบบทดสอบ 60 ข้อ

ตาราง 5 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบ DIF ของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีโพลีโตมัสซิบเทสท์

	อัตราความถูกต้อง	อัตราความคลาดเคลื่อนประเภทที่ 1
PDIF10	.456	.309
PDIF20	.491	.361
PDIF30	.607	.429
MS_b	45.589	102.030
df_b	2	2
MS_w	.244	.228
df_w	21597	86397
F-ratio	186.888	446.822
p-value	.000	.000

ผลการวิเคราะห์ข้อมูลตามตาราง 5 พบว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน (10%, 20% และ 30%) ด้วยวิธีโพลีโตมัสซิบเทสท์ มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีโพลีโตมัสซิบเทสท์มีความแตกต่างกัน และเมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบ พบว่า การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีโพลีโตมัสซิบเทสท์มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีโพลีโตมัสซิบเทสท์มีความแตกต่างกัน ดังนั้นผู้วิจัยจึงทำการทดสอบความแตกต่างรายคู่ด้วยวิธีการของเชฟเฟ่ ผลปรากฏดังตาราง 6

ตาราง 6 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ ในการตรวจสอบ DIF ของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีโพลีโตมัสซิบเทสท์

	(I) PDIF	(J) PDIF	Mean difference	p-value
อัตราความถูกต้อง	10	20	-.035	.002
		30	-.151	.000
	20	30	-.116	.000
อัตราความคลาด	10	20	-.053	.000
เคลื่อนประเภทที่ 1		30	-.120	.000
	20	30	-.067	.000

ผลการวิเคราะห์ข้อมูลตามตาราง 6 อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน (10%, 20% และ 30%) ด้วยวิธีโพลีโตมัสซิบเทสท์ พบว่าที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% มีอัตราความถูกต้องในการตรวจสอบต่ำกว่าที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% และ 30% และที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% มีอัตราความถูกต้องในการตรวจสอบต่ำกว่าที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% วิธีโพลีโตมัสซิบเทสท์มีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% และ 30% และที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% วิธีโพลีโตมัสซิบเทสท์มีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่าที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% มีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบต่ำกว่าที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% และ 30% และที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% มีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบต่ำกว่าที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% มีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% และ 30% และที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% มีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30%

ตาราง 7 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบ DIF ของปัจจัยขนาดตัวอย่างด้วยวิธีโพลีโตมัสซิบเทสท์

	อัตราความถูกต้อง	อัตราความคลาดเคลื่อนประเภทที่ 1
S250	.462	.265
S500	.563	.370
S1000	.604	.448
MS_b	38.468	243.523
df_b	2	2
MS_w	.245	.225
df_w	21597	86397
F-ratio	157.272	1081.985
p-value	.000	.000

ผลการวิเคราะห์ข้อมูลตามตาราง 7 พบว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยขนาดตัวอย่าง (250 คน, 500 คน และ 1000 คน) ด้วยวิธีโพลีโตมัสซิบเทสท์ มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยขนาดตัวอย่างด้วยวิธีโพลีโตมัสซิบเทสท์มีความแตกต่างกัน และเมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบ พบว่า การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยขนาดตัวอย่างด้วยวิธีโพลีโตมัสซิบเทสท์มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยขนาดตัวอย่างด้วยวิธีโพลีโตมัสซิบเทสท์มีความแตกต่างกัน ดังนั้นผู้วิจัยจึงทำการทดสอบความแตกต่างรายคู่ด้วยวิธีการของเชฟเฟ่ ผลปรากฏดังตาราง 8

ตาราง 8 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ ในการตรวจสอบ DIF ของปัจจัยขนาดตัวอย่างด้วยวิธีโพลีโตมัสซิบเทสท์

	(I)S	(J)S	Mean difference	p-value
อัตราความถูกต้อง	250	500	-.101	.000
		1000	-.142	.000
	500	1000	-.041	.000
อัตราความคลาดเคลื่อนประเภทที่ 1	250	500	-.105	.000
		1000	-.183	.000
	500	1000	-.078	.000

ผลการวิเคราะห์ข้อมูลตามตาราง 8 อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยขนาดตัวอย่าง (250 คน, 500 คน และ 1000 คน) ด้วยวิธีโพลีโตมัสซิบเทสท์ พบว่าที่ขนาดตัวอย่าง 250 คน มีอัตราความถูกต้องในการตรวจสอบต่ำกว่าที่ขนาดตัวอย่าง 500 คน และ 1000 คน และที่ขนาดตัวอย่าง 500 คน มีอัตราความถูกต้องในการตรวจสอบต่ำกว่าที่ขนาดตัวอย่าง 1000 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ที่ขนาดตัวอย่าง 250 คน วิธีโพลีโตมัสซิบเทสท์มีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าที่ขนาดตัวอย่าง 500 คน และ 1000 คน และที่ขนาดตัวอย่าง 500 คน มีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าที่ขนาดตัวอย่าง 1000 คน เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ที่ขนาดตัวอย่าง 250 คน มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าที่ขนาดตัวอย่าง 500 คน และ 1000 คน และที่ขนาดตัวอย่าง 500 คน มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าที่ขนาดตัวอย่าง 1000 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ที่ขนาดตัวอย่าง 250 คน มีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าที่ขนาดตัวอย่าง 500 คน และ 1000 คน และที่ขนาดตัวอย่าง 500 คน มีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าที่ขนาดตัวอย่าง 1000 คน

ตาราง 9 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบ DIF ของปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก

		Mean	S.D.	t	df	p-value
P	UDIF	.981	.135	2.060	21598	.039
	NDIF	.978	.149			
E	UDIF	.014	.116	-4.561	86398	.000
	NDIF	.017	.131			

ผลการวิเคราะห์ข้อมูลตามตาราง 9 พบว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความแตกต่างกัน กล่าวคือ การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกันด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องในการตรวจพบสูงกว่าข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความแตกต่างกัน กล่าวคือ การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกันด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน

ตาราง 10 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบ DIF ของปัจจัยความยาวของแบบทดสอบด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก

	อัตราความถูกต้อง	อัตราความคลาดเคลื่อนประเภทที่ 1
TL20	.942	.047
TL40	.974	.022
TL60	.995	.001
MS_b	3.886	11.960
df_b	2	2
MS_w	.020	.015
df_w	21597	86397
F-ratio	196.041	794.141
p-value	.000	.000

ผลการวิเคราะห์ข้อมูลตามตาราง 10 พบว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยความยาวของแบบทดสอบ (20 ข้อ, 40 ข้อ และ 60 ข้อ) ด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความแตกต่างกัน และเมื่อพิจารณา

อัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบ พบว่าการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยความยาวของแบบทดสอบด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความแตกต่างกันอย่างมีนัย สำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยความยาวของแบบทดสอบด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความแตกต่างกัน ดังนั้นผู้วิจัยจึงทำการทดสอบความแตกต่างรายคู่ด้วยวิธีการของเชฟเฟ่ ผลปรากฏดังตาราง 11

ตาราง 11 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ ในการตรวจสอบ DIF ของปัจจัยความยาวของแบบทดสอบด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก

	(I) TL	(J) TL	Mean difference	p-value
อัตราความถูกต้อง	20	40	-.032	.000
		60	-.053	.000
	40	60	-.020	.000
อัตราความคลาด	20	40	.025	.000
เคลื่อนประเภทที่ 1		60	.045	.000
	40	60	.021	.000

ผลการวิเคราะห์ข้อมูลตามตาราง 11 อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยความยาวของแบบทดสอบ (20 ข้อ, 40 ข้อ และ 60 ข้อ) ด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก พบว่า ที่ระดับความยาวของแบบทดสอบ 20 ข้อมีอัตราความถูกต้องในการตรวจสอบต่ำกว่าที่ระดับความยาวของแบบทดสอบ 40 ข้อ และ 60 ข้อ และที่ระดับความยาวของแบบทดสอบ 40 ข้อ มีอัตราความถูกต้องในการตรวจสอบต่ำกว่าที่ระดับความยาวของแบบทดสอบ 60 ข้ออย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ที่ระดับความยาวของแบบทดสอบ 20 ข้อ มีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าที่ระดับความยาวของแบบทดสอบ 40 ข้อ และ 60 ข้อ และที่ระดับความยาวของแบบทดสอบ 40 ข้อ มีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าที่ระดับความยาวของแบบทดสอบ 60 ข้อ เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ที่ระดับความยาวของแบบทดสอบ 20 ข้อ มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าที่ระดับความยาวของแบบทดสอบ 40 ข้อ และ 60 ข้อ และที่ระดับความยาวของแบบทดสอบ 40 ข้อ มีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบสูงกว่าที่ระดับความยาวของแบบทดสอบ 60 ข้อ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ที่ระดับความยาวของแบบทดสอบ 20 ข้อ มีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนสูงกว่าที่ระดับความยาวของแบบทดสอบ 40 ข้อ และ 60 ข้อ และที่ระดับความยาวของแบบทดสอบ 40 ข้อมีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนสูงกว่าที่ระดับความยาวของแบบทดสอบ 60 ข้อ

ตาราง 12 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบ DIF ของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก

	อัตราความถูกต้อง	อัตราความคลาดเคลื่อนประเภทที่ 1
PDIF10	.984	.013
PDIF20	.976	.020
PDIF30	.980	.015
MS_b	.082	.404
df_b	2	2
MS_w	.020	.015
df_w	21597	86397
F-ratio	4.060	26.326
p-value	.017	.000

ผลการวิเคราะห์ข้อมูลตามตาราง 12 พบว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน (10%, 20% และ 30%) ด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าอัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความแตกต่างกัน และเมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบ พบว่า การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความแตกต่างกัน ดังนั้นผู้วิจัยจึงทำการทดสอบความแตกต่างรายคู่ด้วยวิธีการของเชฟเฟ่ ผลปรากฏดังตาราง 13

ตาราง 13 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ ในการตรวจสอบ DIF ของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก

	(I) PDIF	(J) PDIF	Mean difference	p-value
อัตราความถูกต้อง	10	20	.008	.024
		30	.004	.399
	20	30	-.004	.150
อัตราความคลาดเคลื่อนประเภทที่ 1	10	20	-.007	.000
		30	-.002	.079
	20	30	.005	.000

ผลการวิเคราะห์ข้อมูลตามตาราง 13 อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน (10%, 20% และ 30%) ด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก พบว่า ที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% มีอัตราความถูกต้องในการตรวจสอบสูงกว่าที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนคู่อื่น ๆ ไม่แตกต่างกัน กล่าวคือ ที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% มีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนสูงกว่าที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% ส่วนคู่อื่นๆ การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบไม่แตกต่างกัน เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% มีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบต่ำกว่าที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% และที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนคู่อื่น ๆ ไม่แตกต่างกัน กล่าวคือ ที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% มีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% และที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% มีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนสูงกว่าที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% ส่วนคู่อื่นๆ การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบไม่แตกต่างกัน

ตาราง 14 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบ DIF ของปัจจัยขนาดตัวอย่างด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก

	อัตราความถูกต้อง	อัตราความคลาดเคลื่อนประเภทที่ 1
S250	.969	.024
S500	.979	.015
S1000	.989	.008
MS_b	.710	1.824
df_b	2	2
MS_w	.020	.015
df_w	21597	86397
F-ratio	35.300	119.228
p-value	.000	.000

ผลการวิเคราะห์ข้อมูลตามตาราง 14 พบว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยขนาดตัวอย่าง (250 คน, 500 คน และ 1000 คน) ด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยขนาดตัวอย่างด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความแตกต่างกัน และเมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบ พบว่า การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยขนาดตัวอย่างด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยขนาดตัวอย่างด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความแตกต่างกัน ดังนั้นผู้วิจัยจึงทำการทดสอบความแตกต่างรายคู่ด้วยวิธีการของเชฟเฟ่ ผลปรากฏดังตาราง 15

ตาราง 15 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ในการตรวจสอบ DIF ของปัจจัยขนาดตัวอย่างด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก

	(I)S	(J)S	Mean difference	p-value
อัตราความถูกต้อง	250	500	-.010	.000
		1000	-.020	.000
	500	1000	-.010	.000
อัตราความคลาด	250	500	.009	.000
เคลื่อนประเภทที่ 1		1000	.016	.000
	500	1000	.007	.000

ผลการวิเคราะห์ข้อมูลตามตาราง 15 อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยขนาดตัวอย่าง (250 คน, 500 คน และ 1000 คน) ด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก พบว่า ที่ขนาดตัวอย่าง 250 คน มีอัตราความถูกต้องในการตรวจสอบต่ำกว่าที่ขนาดตัวอย่าง 500 คน และ 1000 คน และที่ขนาดตัวอย่าง 500 คน มีอัตราความถูกต้องในการตรวจสอบต่ำกว่าที่ขนาดตัวอย่าง 1000 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ที่ขนาดตัวอย่าง 250 คน มีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าที่ขนาดตัวอย่าง 500 คนและ 1000 คน และที่ขนาดตัวอย่าง 500 คน มีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าที่ขนาดตัวอย่าง 1000 คน เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ที่ขนาดตัวอย่าง 250 คน มีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบสูงกว่าที่ขนาดตัวอย่าง 500 คน และ 1000 คน และที่ขนาดตัวอย่าง 500 คน มีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบสูงกว่าที่ขนาดตัวอย่าง 1000 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ที่ขนาดตัวอย่าง 250 คน มีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนสูงกว่าที่ขนาดตัวอย่าง 500 คนและ 1000 คน และที่ขนาดตัวอย่าง 500 คน มีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนสูงว่าที่ขนาดตัวอย่าง 1000 คน.

ตาราง 16 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบ DIF ของปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับ

		Mean	S.D.	t	df	p-value
P	UDIF	.525	.499	-9.540	21598	.000
	NDIF	.590	.492			
E	UDIF	.154	.361	-21.932	86398	.000
	NDIF	.211	.408			

ผลการวิเคราะห์ข้อมูลตามตาราง 16 พบว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับมีความแตกต่างกัน กล่าวคือ การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกันด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องในการตรวจพบต่ำกว่าข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่าในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับมีความแตกต่างกัน กล่าวคือ การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกันด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับมีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน

ตาราง 17 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบ DIF ของปัจจัยความยาวของแบบทดสอบด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับ

	อัตราความถูกต้อง	อัตราความคลาดเคลื่อนประเภทที่ 1
TL20	.569	.371
TL40	.942	.232
TL60	.297	.087
MS_b	900.072	486.119
df_b	2	2
MS_w	.163	.138
df_w	21597	86397
F-ratio	5509.009	3524.549
p-value	.000	.000

ผลการวิเคราะห์ข้อมูลตามตาราง 17 พบว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยความยาวของแบบทดสอบ (20 ข้อ, 40 ข้อ และ60 ข้อ) ด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยความยาวของแบบทดสอบด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับมีความแตกต่างกัน และเมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบ พบว่า การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยความยาวของแบบทดสอบด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยความยาวของแบบทดสอบด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับมีความแตกต่างกัน ดังนั้นผู้วิจัยจึงทำการทดสอบความแตกต่างรายคู่ด้วยวิธีการของเชฟเฟ่ ผลปรากฏดังตาราง 18

ตาราง 18 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ ในการตรวจสอบ DIF ของปัจจัยความยาวของแบบทดสอบด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับ

	(I) TL	(J) TL	Mean difference	p-value
อัตราความถูกต้อง	20	40	-.373	.000
		60	.272	.000
	40	60	.645	.000
อัตราความคลาด	20	40	.139	.000
เคลื่อนประเภทที่ 1		60	.284	.000
	40	60	.144	.000

ผลการวิเคราะห์ข้อมูลตามตาราง 18 อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยความยาวของแบบทดสอบ (20 ข้อ, 40 ข้อ และ 60 ข้อ) ด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับ พบว่า ที่ระดับความยาวของแบบทดสอบ 20 ข้อมีอัตราความถูกต้องในการตรวจสอบต่ำกว่าที่ระดับความยาวของแบบทดสอบ 40 ข้อ แต่สูงกว่าที่ระดับความยาวของแบบทดสอบ 60 ข้อ และที่ระดับความยาวของแบบทดสอบ 40 ข้อมีอัตราความถูกต้องในการตรวจสอบสูงกว่าที่ระดับความยาวของแบบทดสอบ 60 ข้อ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ที่ระดับความยาวของแบบทดสอบ 20 ข้อ มีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าที่ระดับความยาวของแบบทดสอบ 40 ข้อ แต่สูงกว่าที่ระดับความยาวของแบบทดสอบ 60 ข้อ และที่ระดับความยาวของแบบทดสอบ 40 ข้อมีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนสูงกว่าที่ระดับความยาวของแบบทดสอบ 60 ข้อ เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ที่ระดับความยาวของแบบทดสอบ 20 ข้อ มีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบสูงกว่าที่ระดับความยาวของแบบทดสอบ 40 ข้อและ 60 ข้อ และที่ระดับความยาวของแบบทดสอบ 40 ข้อ มีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบสูงกว่าที่ระดับความยาวของแบบทดสอบ 60 ข้อ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ที่ระดับความยาวของแบบทดสอบ 20 ข้อ มีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนสูงกว่าที่ระดับความยาวของแบบทดสอบ 40 ข้อและ 60 ข้อ และที่ระดับความยาวของแบบทดสอบ 40 ข้อ มีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนสูงกว่าที่ระดับความยาวของแบบทดสอบ 60 ข้อ

ตาราง 19 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบ DIF ของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับ

	อัตราความถูกต้อง	อัตราความคลาดเคลื่อนประเภทที่ 1
PDIF10	.380	.075
PDIF20	.498	.199
PDIF30	.656	.302
MS_b	121.607	371.842
df_b	2	2
MS_w	.235	.141
df_w	21597	86397
F-ratio	516.442	2645.264
p-value	.000	.000

ผลการวิเคราะห์ข้อมูลตามตาราง 19 พบว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน (10%, 20% และ 30%) ด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับมีความแตกต่างกัน และเมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบ พบว่า การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับมีความแตกต่างกัน ดังนั้นผู้วิจัยจึงทำการทดสอบความแตกต่างรายคู่ด้วยวิธีการของเชฟเฟ่ ผลปรากฏดังตาราง 20

ตาราง 20 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ ในการตรวจสอบ DIF ของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับ

	(I) PDIF	(J) PDIF	Mean difference	p-value
อัตราความถูกต้อง	10	20	-.118	.000
		30	-.276	.000
	20	30	-.158	.000
อัตราความคลาด	10	20	-.124	.000
เคลื่อนประเภทที่ 1		30	-.227	.000
	20	30	-.103	.000

ผลการวิเคราะห์ข้อมูลตามตาราง 20 อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน (10%, 20% และ 30%) ด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับ พบว่า ที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% มีอัตราความถูกต้องในการตรวจสอบต่ำกว่าที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% และ 30% และที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% มีอัตราความถูกต้องในการตรวจสอบต่ำกว่าที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% มีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% และ 30% และที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% มีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% มีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบต่ำกว่าที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% และ 30% และที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% มีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบต่ำกว่าที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% มีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% และ 30% และที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% มีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าที่ระดับสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30%

ตาราง 21 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบ DIF ของปัจจัยขนาดตัวอย่างด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับ

	อัตราความถูกต้อง	อัตราความคลาดเคลื่อนประเภทที่ 1
S250	.472	.132
S500	.569	.167
S1000	.631	.248
MS_b	46.387	102.377
df_b	2	2
MS_w	.242	.147
df_w	21597	86397
F-ratio	191.336	697.358
p-value	.000	.000

ผลการวิเคราะห์ข้อมูลตามตาราง 21 พบว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยขนาดตัวอย่าง (250 คน, 500 คน และ 1000 คน) ด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยขนาดตัวอย่างด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับมีความแตกต่างกัน และเมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบ พบว่า การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยขนาดตัวอย่างด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยขนาดตัวอย่างด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับมีความแตกต่างกัน ดังนั้นผู้วิจัยจึงทำการทดสอบความแตกต่างรายคู่ด้วยวิธีการของเชฟเฟ่ ผลปรากฏดังตาราง 22

ตาราง 22 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ ในการตรวจสอบ DIF ของปัจจัยขนาดตัวอย่างด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับ

	(I)S	(J)S	Mean difference	p-value
อัตราความถูกต้อง	250	500	-.097	.000
		1000	-.159	.000
	500	1000	-.063	.000
อัตราความคลาด	250	500	-.035	.000
เคลื่อนประเภทที่ 1		1000	-.116	.000
	500	1000	-.081	.000

ผลการวิเคราะห์ข้อมูลตามตาราง 22 อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยขนาดตัวอย่าง (250 คน, 500 คน และ 1000 คน) ด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับ พบว่าที่ขนาดตัวอย่าง 250 คนมีอัตราความถูกต้องในการตรวจสอบต่ำกว่าที่ขนาดตัวอย่าง 500 คน และ 1000 คน และที่ขนาดตัวอย่าง 500 คน มีอัตราความถูกต้องในการตรวจสอบต่ำกว่าที่ขนาดตัวอย่าง 1000 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ที่ขนาดตัวอย่าง 250 คน มีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าที่ขนาดตัวอย่าง 500 คน และ 1000 คน และที่ขนาดตัวอย่าง 500 คน มีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าที่ขนาดตัวอย่าง 1000 คน เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่าที่ขนาดตัวอย่าง 250 คน มีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบต่ำกว่าที่ขนาดตัวอย่าง 500 คน และ 1000 คน และที่ขนาดตัวอย่าง 500 คน มีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบต่ำกว่าที่ขนาดตัวอย่าง 1000 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ที่ขนาดตัวอย่าง 250 คน มีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าที่ขนาดตัวอย่าง 500 คน และ 1000 คน และที่ขนาดตัวอย่าง 500 คน มีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าที่ขนาดตัวอย่าง 1000 คน

2.2 เปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบภายใต้ปัจจัยที่แปรเปลี่ยนของตัวแปรที่ศึกษา

การวิจัยตอนนี้ผู้วิจัยนำผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ปัจจัยความยาวของแบบทดสอบ ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และปัจจัยขนาดตัวอย่าง ซึ่งได้แก่อัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 มาวิเคราะห์ความแตกต่างเพื่อเปรียบเทียบระหว่างวิธีการตรวจสอบทั้ง 3 วิธี ภายใต้เงื่อนไขของปัจจัยแต่ละปัจจัยของตัวแปรที่ศึกษา โดยใช้การทดสอบเอฟ (F-test) จากการวิเคราะห์ความแปรปรวนทางเดียว (Oneway ANOVA) และเมื่อมีนัยสำคัญทางสถิติแล้ว ผู้วิจัยทำการทดสอบความแตกต่างรายคู่ โดยใช้วิธีการทดสอบของเชฟเฟ่ ผลปรากฏดังตาราง 23 – 44

ตาราง 23 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน

	อัตราความถูกต้อง	อัตราความคลาดเคลื่อนประเภทที่ 1
Poly-SIB	.651	.402
LDFA	.981	.014
OLR	.525	.154
MS_b	598.843	1670.387
df_b	2	2
MS_w	.165	.128
df_w	32397	129597
F-ratio	3631.988	13051.268
p-value	.000	.000

ผลการวิเคราะห์ตามตาราง 23 พบว่า ภายใต้เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน วิธีตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบทั้ง 3 วิธี ได้แก่วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีอัตราความถูกต้องในการตรวจสอบแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบระหว่างวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ภายใต้เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกันแตกต่างกัน และเมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบพบว่าทั้ง 3 วิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบระหว่างวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ภายใต้เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกันแตกต่างกัน เพื่อให้ทราบความแตกต่างระหว่างวิธี ผู้วิจัยจึงทำการทดสอบความแตกต่างรายคู่ด้วยวิธีการของเชฟเฟ่ ผลปรากฏดังตาราง 24

ตาราง 24 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน

	(I) Proc	(J) Proc	Mean difference	p-value
อัตราความถูกต้อง	Poly-SIB	LDFA	-.330	.000
		OLR	.126	.000
	LDFA	OLR	.456	.000
อัตราความคลาดเคลื่อนประเภทที่ 1	Poly-SIB	LDFA	.388	.000
		OLR	.248	.000
	LDFA	OLR	-.140	.000

ผลการวิเคราะห์ตามตาราง 24 พบว่า การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบภายใต้เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน วิธีโพลีโตมัสซิบเทสท์มีอัตราความถูกต้องต่ำกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก แต่สูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ภายใต้เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน วิธีโพลีโตมัสซิบเทสท์มีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก แต่สูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนสูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีโพลีโตมัสซิบเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ภายใต้เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน วิธีโพลีโตมัสซิบเทสท์มีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนสูงกว่าวิธีการวิเคราะห์ฟังก์ชัน การจำแนกโลจิสติก และการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ

ตาราง 25 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน

	อัตราความถูกต้อง	อัตราความคลาดเคลื่อนประเภทที่ 1
Poly-SIB	.435	.320
LDFA	.978	.017
OLR	.590	.211
MS_b	843.510	1015.455
df_b	2	2
MS_w	.170	.134
df_w	32397	129597
F-ratio	4963.199	7587.941
p-value	.000	.000

ผลการวิเคราะห์ตามตาราง 25 พบว่า ภายใต้เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน วิธีตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบทั้ง 3 วิธี ได้แก่ วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีอัตราความถูกต้องในการตรวจสอบแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบระหว่างวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ภายใต้เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกันแตกต่างกัน และเมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบ พบว่า ทั้ง 3 วิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบระหว่างวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับภายใต้เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกันแตกต่างกัน เพื่อให้ทราบความแตกต่างระหว่างวิธี ผู้วิจัยจึงทำการทดสอบความแตกต่างรายคู่ด้วยวิธีการของเชฟเฟ่ ผลปรากฏดังตาราง 26

ตาราง 26 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน

	(I) Proc	(J) Proc	Mean difference	p-value
อัตราความถูกต้อง	Poly-SIB	LDFA	-.543	.000
		OLR	-.155	.000
	LDFA	OLR	.388	.000
อัตราความคลาดเคลื่อนประเภทที่ 1	Poly-SIB	LDFA	.303	.000
		OLR	.109	.000
	LDFA	OLR	-.194	.000

ผลการวิเคราะห์ตามตาราง 26 พบว่า การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบภายใต้เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน วิธีโพลีโตมัสซิบเทสท์มีอัตราความถูกต้องต่ำกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ภายใต้เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน วิธีโพลีโตมัสซิบเทสท์มีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบน ต่ำกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบน สูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีโพลีโตมัสซิบเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ภายใต้เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน วิธีโพลีโตมัสซิบเทสท์มีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนสูงกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ

ตาราง 27 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขความยาวของแบบทดสอบ 20 ข้อ

	อัตราความถูกต้อง	อัตราความคลาดเคลื่อนประเภทที่ 1
Poly-SIB	.625	.190
LDFA	.942	.047
OLR	.569	.371
MS_b	145.664	380.003
df_b	2	2
MS_w	.178	.144
df_w	10797	43197
F-ratio	818.167	2641.549
p-value	.000	.000

ผลการวิเคราะห์ตามตาราง 27 พบว่า ภายใต้เงื่อนไขความยาวของแบบทดสอบ 20 ข้อ วิธีตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบทั้ง 3 วิธี ได้แก่วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีอัตราความถูกต้องในการตรวจสอบแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบระหว่างวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับภายใต้เงื่อนไขความยาวของแบบทดสอบ 20 ข้อ แตกต่างกัน และเมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบ พบว่า ทั้ง 3 วิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบระหว่างวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับภายใต้เงื่อนไขความยาวของแบบทดสอบ 20 ข้อ แตกต่างกัน เพื่อให้ทราบความแตกต่างระหว่างวิธี ผู้วิจัยจึงทำการทดสอบความแตกต่างรายคู่ด้วยวิธีการของเชฟเฟ่ ผลปรากฏดังตาราง 28

ตาราง 28 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขความยาวของแบบทดสอบ 20 ข้อ

	(I) Proc	(J) Proc	Mean difference	p-value
อัตราความถูกต้อง	Poly-SIB	LDFA	-.317	.000
		OLR	.056	.000
	LDFA	OLR	.373	.000
อัตราความคลาดเคลื่อนประเภทที่ 1	Poly-SIB	LDFA	.143	.000
		OLR	-.181	.000
	LDFA	OLR	-.324	.000

ผลการวิเคราะห์ตามตาราง 28 พบว่า การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบภายใต้เงื่อนไขความยาวของแบบทดสอบ 20 ข้อ วิธีโพลีโตมัสซิบเทสท์มีอัตราความถูกต้องต่ำกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก แต่สูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ภายใต้เงื่อนไขความยาวของแบบทดสอบ 20 ข้อ วิธีโพลีโตมัสซิบเทสท์มีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก แต่สูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนสูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีโพลีโตมัสซิบเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก แต่ต่ำกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความคลาดเคลื่อนประเภท ที่ 1 ต่ำกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ภายใต้เงื่อนไขความยาวของแบบทดสอบ 20 ข้อ วิธีโพลีโตมัสซิบเทสท์มีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนสูงกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก แต่ต่ำกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ

ตาราง 29 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขความยาวของแบบทดสอบ 40 ข้อ

	อัตราความถูกต้อง	อัตราความคลาดเคลื่อนประเภทที่ 1
Poly-SIB	.699	.524
LDFA	.974	.022
OLR	.942	.232
MS_b	163.243	1829.358
df_b	2	2
MS_w	.096	.150
df_w	21597	86397
F-ratio	1691.874	12235.888
p-value	.000	.000

ผลการวิเคราะห์ตามตาราง 29 พบว่า ภายใต้เงื่อนไขความยาวของแบบทดสอบ 40 ข้อ วิธีตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบทั้ง 3 วิธี ได้แก่ วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีอัตราความถูกต้องในการตรวจสอบแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบระหว่างวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับภายใต้เงื่อนไขความยาวของแบบทดสอบ 40 ข้อแตกต่างกัน และเมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบพบว่าทั้ง 3 วิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบระหว่างวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับภายใต้เงื่อนไขความยาวของแบบทดสอบ 40 ข้อแตกต่างกัน เพื่อให้ทราบความแตกต่างระหว่างวิธี ผู้วิจัยจึงทำการทดสอบความแตกต่างรายคู่ด้วยวิธีการของเชฟเฟ่ ผลปรากฏดังตาราง 30

ตาราง 30 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขความยาวของแบบทดสอบ 40 ข้อ

	(I) Proc	(J) Proc	Mean difference	p-value
อัตราความถูกต้อง	Poly-SIB	LDFA	-.275	.000
		OLR	-.243	.000
	LDFA	OLR	.032	.000
อัตราความคลาดเคลื่อนประเภทที่ 1	Poly-SIB	LDFA	.502	.000
		OLR	.292	.000
	LDFA	OLR	-.210	.000

ผลการวิเคราะห์ตามตาราง 30 พบว่า การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบภายใต้เงื่อนไขความยาวของแบบทดสอบ 40 ข้อ วิธีโพลีโตมัสซิบเทสท์มีอัตราความถูกต้องต่ำกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ภายใต้เงื่อนไขความยาวของแบบทดสอบ 40 ข้อ วิธีโพลีโตมัสซิบเทสท์มีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนสูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีโพลีโตมัสซิบเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ภายใต้เงื่อนไขความยาวของแบบทดสอบ 40 ข้อ วิธีโพลีโตมัสซิบเทสท์มีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนสูงกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ

ตาราง 31 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขความยาวของแบบทดสอบ 60 ข้อ

	อัตราความถูกต้อง	อัตราความคลาดเคลื่อนประเภทที่ 1
Poly-SIB	.412	.310
LDFA	.995	.001
OLR	.297	.087
MS_b	1512.517	1096.595
df_b	2	2
MS_w	.152	.098
df_w	32397	129597
F-ratio	9949.361	11169.645
p-value	.000	.000

ผลการวิเคราะห์ตามตาราง 31 พบว่า ภายใต้เงื่อนไขความยาวของแบบทดสอบ 60 ข้อ วิธีตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบทั้ง 3 วิธี ได้แก่ วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีอัตราความถูกต้องในการตรวจสอบแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบระหว่างวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับภายใต้เงื่อนไขความยาวของแบบทดสอบ 60 ข้อแตกต่างกัน และเมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบพบว่า ทั้ง 3 วิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบระหว่างวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับภายใต้เงื่อนไขความยาวของแบบทดสอบ 60 ข้อแตกต่างกัน เพื่อให้ทราบความแตกต่างระหว่างวิธี ผู้วิจัยจึงทำการทดสอบความแตกต่างรายคู่ด้วยวิธีการของเชฟเฟ่ ผลปรากฏดังตาราง 32

ตาราง 32 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขความยาวของแบบทดสอบ 60 ข้อ

	(I) Proc	(J) Proc	Mean difference	p-value
อัตราความถูกต้อง	Poly-SIB	LDFA	-.583	.000
		OLR	.115	.000
	LDFA	OLR	.698	.000
อัตราความคลาดเคลื่อนประเภทที่ 1	Poly-SIB	LDFA	.309	.000
		OLR	.223	.000
	LDFA	OLR	-.086	.000

ผลการวิเคราะห์ตามตาราง 32 พบว่า การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบภายใต้เงื่อนไขความยาวของแบบทดสอบ 60 ข้อวิธีโพลีโตมัสซิบเทสท์มีอัตราความถูกต้องต่ำกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก แต่สูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ภายใต้เงื่อนไขความยาวของแบบทดสอบ 60 ข้อ วิธีโพลีโตมัสซิบเทสท์มีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก แต่สูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนสูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีโพลีโตมัสซิบเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ภายใต้เงื่อนไขความยาวของแบบทดสอบ 60 ข้อ วิธีโพลีโตมัสซิบเทสท์มีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนสูงกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ

ตาราง 33 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10%

	อัตราความถูกต้อง	อัตราความคลาดเคลื่อนประเภทที่ 1
Poly-SIB	.456	.309
LDFA	.984	.013
OLR	.380	.075
MS_b	389.339	789.208
df_b	2	2
MS_w	.167	.098
df_w	10797	97197
F-ratio	2337.405	8027.387
p-value	.000	.000

ผลการวิเคราะห์ตามตาราง 33 พบว่า ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% วิธีตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบทั้ง 3 วิธี ได้แก่ วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีอัตราความถูกต้องในการตรวจสอบแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบระหว่างวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% แตกต่างกัน และเมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบ พบว่า ทั้ง 3 วิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบระหว่างวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% แตกต่างกัน เพื่อให้ทราบความแตกต่างระหว่างวิธี ผู้วิจัยจึงทำการทดสอบความแตกต่างรายคู่ด้วยวิธีการของเชฟเฟ่ ผลปรากฏดังตาราง 34

ตาราง 34 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10%

	(I) Proc	(J) Proc	Mean difference	p-value
อัตราความถูกต้อง	Poly-SIB	LDFA	-.528	.000
		OLR	.076	.000
	LDFA	OLR	.604	.000
อัตราความคลาดเคลื่อนประเภทที่ 1	Poly-SIB	LDFA	.296	.000
		OLR	.234	.000
	LDFA	OLR	-.062	.000

ผลการวิเคราะห์ตามตาราง 34 พบว่า การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% วิธีโพลีโตมัสซิบเทสท์มีอัตราความถูกต้องต่ำกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก แต่สูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% วิธีโพลีโตมัสซิบเทสท์มีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก แต่สูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนสูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีโพลีโตมัสซิบเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% วิธีโพลีโตมัสซิบเทสท์มีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนสูงกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ

ตาราง 35 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20%

	อัตราความถูกต้อง	อัตราความคลาดเคลื่อนประเภทที่ 1
Poly-SIB	.491	.361
LDFA	.976	.020
OLR	.498	.199
MS_b	555.931	840.159
df_b	2	2
MS_w	.174	.136
df_w	21597	86397
F-ratio	3186.193	6157.342
p-value	.000	.000

ผลการวิเคราะห์ตามตาราง 35 พบว่า ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% วิธีตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบทั้ง 3 วิธี ได้แก่ วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีอัตราความถูกต้องในการตรวจสอบแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบระหว่างวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% แตกต่างกัน และเมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบ พบว่า ทั้ง 3 วิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบระหว่างวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% แตกต่างกัน เพื่อให้ทราบความแตกต่างระหว่างวิธี ผู้วิจัยจึงทำการทดสอบความแตกต่างรายคู่ด้วยวิธีการของเชฟเฟ่ ผลปรากฏดังตาราง 36

ตาราง 36 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20%

	(I) Proc	(J) Proc	Mean difference	p-value
อัตราความถูกต้อง	Poly-SIB	LDFA	-.485	.000
		OLR	-.007	.608
	LDFA	OLR	.478	.000
อัตราความคลาดเคลื่อนประเภทที่ 1	Poly-SIB	LDFA	.341	.000
		OLR	.162	.000
	LDFA	OLR	-.179	.000

ผลการวิเคราะห์ตามตาราง 36 พบว่า การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% วิธีโพลีโตมัสซิบเทสท์มีอัตราความถูกต้องต่ำกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนคู่อื่นๆ ไม่แตกต่างกัน กล่าวคือ ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% วิธีโพลีโตมัสซิบเทสท์มีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนสูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ ส่วนคู่อื่นๆ การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบไม่แตกต่างกัน เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่าวิธีโพลีโตมัสซิบเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% วิธีโพลีโตมัสซิบเทสท์มีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนสูงกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ

ตาราง 37 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30%

	อัตราความถูกต้อง	อัตราความคลาดเคลื่อนประเภทที่ 1
Poly-SIB	.607	.429
LDFA	.980	.015
OLR	.656	.302
MS_b	443.820	1132.502
df_b	2	2
MS_w	.161	.157
df_w	32397	75597
F-ratio	2753.061	7223.070
p-value	.000	.000

ผลการวิเคราะห์ตามตาราง 37 พบว่า ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% วิธีตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบทั้ง 3 วิธี ได้แก่ วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีอัตราความถูกต้องในการตรวจสอบแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบระหว่างวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% แตกต่างกัน และเมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบ พบว่า ทั้ง 3 วิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบระหว่างวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% แตกต่างกัน เพื่อให้ทราบความแตกต่างระหว่างวิธี ผู้วิจัยจึงทำการทดสอบความแตกต่างรายคู่ด้วยวิธีการของเชฟเฟ่ ผลปรากฏดังตาราง 38

ตาราง 38 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30%

	(I) Proc	(J) Proc	Mean difference	p-value
อัตราความถูกต้อง	Poly-SIB	LDFA	-.373	.000
		OLR	-.049	.000
	LDFA	OLR	.324	.000
อัตราความคลาดเคลื่อนประเภทที่ 1	Poly-SIB	LDFA	.414	.000
		OLR	.126	.000
	LDFA	OLR	-.287	.000

ผลการวิเคราะห์ตามตาราง 38 พบว่า การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% วิธีโพลีโตมัสซิบเทสท์มีอัตราความถูกต้องต่ำกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% วิธีโพลีโตมัสซิบเทสท์มีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนสูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีโพลีโตมัสซิบเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% วิธีโพลีโตมัสซิบเทสท์มีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนสูงกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ

ตาราง 39 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขขนาดตัวอย่าง 250 คน

	อัตราความถูกต้อง	อัตราความคลาดเคลื่อนประเภทที่ 1
Poly-SIB	.462	.265
LDFA	.969	.024
OLR	.472	.132
MS_b	605.515	419.778
df_b	2	2
MS_w	.176	.111
df_w	21597	86397
F-ratio	3443.695	3783.094
p-value	.000	.000

ผลการวิเคราะห์ตามตาราง 39 พบว่า ภายใต้เงื่อนไขขนาดตัวอย่าง 250 คน วิธีตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบทั้ง 3 วิธี ได้แก่ วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีอัตราความถูกต้องในการตรวจสอบแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบระหว่างวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับภายใต้เงื่อนไขขนาดตัวอย่าง 250 คน แตกต่างกัน และเมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบ พบว่า ทั้ง 3 วิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบระหว่างวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับภายใต้เงื่อนไขขนาดตัวอย่าง 250 คน แตกต่างกัน เพื่อให้ทราบความแตกต่างระหว่างวิธี ผู้วิจัยจึงทำการทดสอบความแตกต่างรายคู่ด้วยวิธีการของเชฟเฟ่ ผลปรากฏดังตาราง 40

ตาราง 40 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขขนาดตัวอย่าง 250 คน

	(I) Proc	(J) Proc	Mean difference	p-value
อัตราความถูกต้อง	Poly-SIB	LDFA	-.507	.000
		OLR	-.010	.380
	LDFA	OLR	.497	.000
อัตราความคลาดเคลื่อนประเภทที่ 1	Poly-SIB	LDFA	.241	.000
		OLR	.133	.000
	LDFA	OLR	-.108	.000

ผลการวิเคราะห์ตามตาราง 40 พบว่า การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบภายใต้เงื่อนไขขนาดตัวอย่าง 250 คน วิธีโพลีโตมัสซิบเทสท์มีอัตราความถูกต้องต่ำกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนคู่อื่นๆ ไม่แตกต่างกัน กล่าวคือ ภายใต้เงื่อนไขขนาดตัวอย่าง 250 คน วิธีโพลีโตมัสซิบเทสท์มีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนสูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ ส่วนคู่อื่นๆ การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ไม่แตกต่างกัน เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีโพลีโตมัสซิบเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ภายใต้เงื่อนไขขนาดตัวอย่าง 250 คน วิธีโพลีโตมัสซิบเทสท์มีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนสูงกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ

ตาราง 41 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขขนาดตัวอย่าง 500 คน

	อัตราความถูกต้อง	อัตราความคลาดเคลื่อนประเภทที่ 1
Poly-SIB	.563	.370
LDFA	.979	.015
OLR	.569	.167
MS_b	410.223	915.188
df_b	2	2
MS_w	.170	.129
df_w	21597	86397
F-ratio	2406.164	7102.371
p-value	.000	.000

ผลการวิเคราะห์ตามตาราง 41 พบว่า ภายใต้เงื่อนไขขนาดตัวอย่าง 500 คน วิธีตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบทั้ง 3 วิธี ได้แก่ วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีอัตราความถูกต้องในการตรวจสอบแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบระหว่างวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับภายใต้เงื่อนไขขนาดตัวอย่าง 500 คน แตกต่างกัน และเมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบ พบว่า ทั้ง 3 วิธี มีอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบระหว่างวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับภายใต้เงื่อนไขขนาดตัวอย่าง 500 คน แตกต่างกัน เพื่อให้ทราบความแตกต่างระหว่างวิธี ผู้วิจัยจึงทำการทดสอบความแตกต่างรายคู่ด้วยวิธีการของเชฟเฟ่ ผลปรากฏดังตาราง 42

ตาราง 42 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขขนาดตัวอย่าง 500 คน

	(I) Proc	(J) Proc	Mean difference	p-value
อัตราความถูกต้อง	Poly-SIB	LDFA	-.416	.000
		OLR	-.006	.710
	LDFA	OLR	.411	.000
อัตราความคลาดเคลื่อนประเภทที่ 1	Poly-SIB	LDFA	.355	.000
		OLR	.203	.000
	LDFA	OLR	-.152	.000

ผลการวิเคราะห์ตามตาราง 42 พบว่า การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบภายใต้เงื่อนไขขนาดตัวอย่าง 500 คน วิธีโพลีโตมัสซิบเทสท์มีอัตราความถูกต้องต่ำกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนคู่อื่นๆ ไม่แตกต่างกัน กล่าวคือ ภายใต้เงื่อนไขขนาดตัวอย่าง 500 คน วิธีโพลีโตมัสซิบเทสท์มีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนสูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ ส่วนคู่อื่นๆ การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบไม่แตกต่างกัน เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีโพลีโตมัสซิบเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ภายใต้เงื่อนไขขนาดตัวอย่าง 500 คน วิธีโพลีโตมัสซิบเทสท์ มีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนสูงกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ

ตาราง 43 เปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขขนาดตัวอย่าง 1000 คน

	อัตราความถูกต้อง	อัตราความคลาดเคลื่อนประเภทที่ 1
Poly-SIB	.604	.448
LDFA	.989	.008
OLR	.631	.248
MS_b	332.338	1398.694
df_b	2	2
MS_w	.161	.147
df_w	21597	86397
F-ratio	2066.456	9492.073
p-value	.000	.000

ผลการวิเคราะห์ตามตาราง 43 พบว่า ภายใต้เงื่อนไขขนาดตัวอย่าง 1000 คน วิธีตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบทั้ง 3 วิธี ได้แก่ วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีอัตราความถูกต้องในการตรวจสอบแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบระหว่างวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับภายใต้เงื่อนไขขนาดตัวอย่าง 1000 คน แตกต่างกัน และเมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบ พบว่า ทั้ง 3 วิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบระหว่างวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับภายใต้เงื่อนไขขนาดตัวอย่าง 1000 คน แตกต่างกัน เพื่อให้ทราบความแตกต่างระหว่างวิธี ผู้วิจัยจึงทำการทดสอบความแตกต่างรายคู่ด้วยวิธีการของเชฟเฟ่ ผลปรากฏดังตาราง 44

ตาราง 44 เปรียบเทียบผลต่างค่าเฉลี่ยอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 รายคู่ระหว่างวิธีตรวจสอบ DIF 3 วิธี ภายใต้เงื่อนไขขนาดตัวอย่าง 1000 คน

	(I) Proc	(J) Proc	Mean difference	p-value
อัตราความถูกต้อง	Poly-SIB	LDFA	-.385	.000
		OLR	-.027	.000
	LDFA	OLR	.358	.000
อัตราความคลาดเคลื่อนประเภทที่ 1	Poly-SIB	LDFA	.440	.000
		OLR	.200	.000
	LDFA	OLR	-.240	.000

ผลการวิเคราะห์ตามตาราง 44 พบว่า การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบภายใต้เงื่อนไขขนาดตัวอย่าง 1000 คน วิธีโพลีโตมัสซิบเทสท์มีอัตราความถูกต้องต่ำกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ภายใต้เงื่อนไขขนาดตัวอย่าง 1000 คน วิธีโพลีโตมัสซิบเทสท์มีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนสูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่าวิธีโพลีโตมัสซิบเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ภายใต้เงื่อนไขขนาดตัวอย่าง 1000 คน วิธีโพลีโตมัสซิบเทสท์มีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนสูงกว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนต่ำกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ

บทที่ 5
สรุป อภิปรายผล และข้อเสนอแนะ

ในการวิจัยครั้งนี้มีความมุ่งหมายเพื่อศึกษาประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่า ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความยาวของแบบทดสอบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และขนาดตัวอย่าง โดยใช้ข้อมูลจำลองโดยมีความมุ่งหมายเฉพาะดังนี้

1. เพื่อศึกษาอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่าด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ภายใต้เงื่อนไขต่างกันของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ซึ่งประกอบด้วยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความยาวของแบบทดสอบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และขนาดตัวอย่าง

2. เพื่อเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่า ด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ภายใต้เงื่อนไขต่างกันของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ซึ่งประกอบด้วยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความยาวของแบบทดสอบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และขนาดตัวอย่าง

3. เพื่อเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่า ระหว่างวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ภายใต้เงื่อนไขต่างกันของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ซึ่งประกอบด้วยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความยาวของแบบทดสอบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และขนาดตัวอย่าง

สำหรับสมมติฐานของการวิจัย ผู้วิจัยตั้งสมมติฐานในการวิจัยดังนี้

1. ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ประกอบด้วยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความยาวของแบบทดสอบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และขนาดตัวอย่าง มีผลทำให้อัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่าด้วยวิธีโพลีโตมัสซิบเทสท์แตกต่างกัน

2. ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ประกอบด้วยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความยาวของแบบทดสอบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และขนาดตัวอย่างมีผลทำให้อัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่าด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกแตกต่างกัน

3. ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ประกอบด้วยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความยาวของแบบทดสอบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และขนาดตัวอย่างมีผลทำให้อัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่าด้วยวิธีการถดถอยโลจิสติกแบบจัดอันดับแตกต่างกัน

4. วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ให้คะแนนแบบหลายค่าภายใต้เงื่อนไขปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ประกอบด้วยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความยาวของแบบทดสอบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และขนาดตัวอย่าง มีอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกัน

วิธีดำเนินการวิจัย

1. จำลองข้อมูลภายใต้โมเดลพาร์เชียลเครดิตทั่วไป (Generalized Partial Credit Model: GPCM) ของมูรากิ (Muraki. 1992) ที่มีขนาดความยาว 20 ข้อ, 40 ข้อ และ 60 ข้อในข้อสอบแต่ละข้อจะวัดความสามารถเป้าหมายเพียงมิติเดียว และแต่ละข้อมีรายการตอบ 5 รายการ ให้คะแนนเป็น 1,2,3,4 หรือ 5 สำหรับการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ และประมาณค่าพารามิเตอร์ข้อสอบจากความสามารถผู้สอบของข้อสอบที่มีการตรวจให้คะแนนแบบหลายค่าจะใช้โปรแกรม WinGen2 ของฮานและแฮมเบิลตัน (Han; & Hambleton. 2007) จำลองผลการตอบข้อสอบภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ (1) รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน 2 รูปแบบ (ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน) (2) ความยาวของแบบทดสอบ 3 ขนาด (20 ข้อ, 40 ข้อ และ 60 ข้อ) (3) สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 3 ขนาด (10%, 20% และ 30%) (4) ขนาดตัวอย่าง 3 ขนาด (250 คน, 500 คน และ 1000 คน) ซึ่งมีขั้นตอนหลัก 3 ขั้นตอนดังนี้ **ขั้นแรก** เป็นการสร้างข้อมูลผู้สอบโดยระบุจำนวนผู้สอบของขนาดตัวอย่าง เช่น 250 คน จากนั้นเลือกชนิดของการแจกแจงคุณลักษณะหรือความสามารถของผู้สอบโดยเลือกการแจกแจงแบบโค้งปกติ พร้อมทั้งระบุค่าเฉลี่ยของความสามารถผู้สอบและส่วนเบี่ยงเบนมาตรฐานเป็น 0 และ 1 ตามลำดับ **ขั้นที่สอง**เป็นการสร้างข้อมูลข้อสอบ โดยระบุจำนวนของข้อสอบในแบบทดสอบ เช่น 20 ข้อพร้อมทั้งระบุจำนวนรายการคำตอบโดยเลือก ‘5’ สำหรับข้อมูลที่มีการตรวจให้คะแนนแบบหลายค่าของโมเดลพาร์เชียลเครดิตทั่วไป (GPCM) ในการประมาณค่าพารามิเตอร์ ต่อจากนั้นเลือกการแจกแจงของพารามิเตอร์ข้อสอบและระบุคุณลักษณะของการแจกแจง โดยเลือกพารามิเตอร์ a ที่มีการแจกแจงแบบโค้งปกติโดยมีค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ1 และเลือกพารามิเตอร์ b ที่มีการแจกแจงแบบโค้งปกติโดยมีค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 เช่นกัน

การศึกษาครั้งนี้ ผู้วิจัยสนใจศึกษารูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน 2 รูปแบบคือ ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และที่ไม่เป็นรูปแบบเดียวกัน รูปแบบแรก กำหนดให้พารามิเตอร์ความยากของข้อสอบระหว่างกลุ่มอ้างอิง และกลุ่มสนใจมีค่าแตกต่างกัน 0.25 นั่นคือ $b_{jF} = b_{jR} + .25$ ส่วนพารามิเตอร์อำนาจจำแนกของข้อสอบระหว่างกลุ่มอ้างอิง และกลุ่มสนใจมีค่าเท่ากัน นั่นคือ $a_{jF} = a_{jR}$ สำหรับรูปแบบหลังกำหนดให้อำนาจจำแนกข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มสนใจมีค่าต่างกัน 1.0 นั่นคือ $a_{jF} = a_{jR} -1.0$ ส่วนพารามิเตอร์ความยากของข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มสนใจมีค่าเท่ากัน นั่นคือ $b_{jF} = b_{jR}$ โดยมีสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 3 ขนาด คือ 10%, 20% และ 30% ของแต่ละขนาดความยาวของแบบทดสอบ **ขั้นสุดท้าย**เป็นการสร้างข้อมูลการตอบข้อสอบ ในขั้นนี้เป็นการนำเซตข้อมูลผู้ตอบที่สร้างขึ้นในขั้นแรกและเซตข้อมูลพารามิเตอร์ข้อสอบที่สร้างขึ้นในขั้นที่สอง มาจัดทำข้อมูลรายการคำตอบของกลุ่มตัวอย่างที่จะทำการศึกษา ในแต่ละเงื่อนไขจะทำข้อมูลซ้ำ 50 ชุดซึ่งได้ข้อมูลที่เป็นไปตามเงื่อนไขทั้งสิ้น 2,700 ชุด

2. นำข้อมูลที่ได้จากการจำลองทั้งหมดมาตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ 3 วิธี คือวิธีโพลีโตมัสซิบเทสท์ (Poly-SIB) วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก (LDFA) และวิธีการถดถอยโลจิสติกแบบจัดอันดับ (OLR) โดยวิธีแรกใช้โปรแกรม DIFPACK วิธีที่สองใช้โปรแกรม MULTILOG ส่วนวิธีสุดท้ายใช้โปรแกรม SPSS แล้วตัดสินข้อสอบทำหน้าที่เบี่ยงเบนที่ระดับนัยสำคัญ .05 จากนั้นนำผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบทั้งหมดมาคำนวณอัตราความถูกต้อง (Power rate: P) และอัตราความคลาดเคลื่อนประเภทที่ 1
(Type I error: E)

3. การเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความยาวของแบบทดสอบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และขนาดตัวอย่าง ผู้วิจัยใช้การทดสอบที (t-test) ทดสอบเงื่อนไขของปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน (ข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน) และการทดสอบเอฟ (F-test) จากการวิเคราะห์ความแปรปรวนทางเดียว (Oneway ANOVA) ทดสอบเงื่อนไขของปัจจัยความยาวของแบบทดสอบ (20 ข้อ, 40 ข้อและ 60 ข้อ) เงื่อนไขของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน (10%, 20% และ 30%) และเงื่อนไขของปัจจัยขนาดตัวอย่าง (250 คน, 500 คน และ 1000 คน) และเมื่อมีนัยสำคัญทางสถิติแล้ว ผู้วิจัยทำการทดสอบความแตกต่างรายคู่ โดยใช้วิธีการทดสอบของเชฟเฟ่สำหรับการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบระหว่างวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย

คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความยาวของแบบทดสอบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และขนาดตัวอย่าง ผู้วิจัยใช้การทดสอบเอฟ (F-test) จากการวิเคราะห์ความแปรปรวนทางเดียว (Oneway ANOVA) ทดสอบความแตกต่างของวิธีตรวจสอบทั้ง 3 วิธี ภายใต้เงื่อนไขของแต่ละปัจจัย ซึ่งได้แก่เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกันของปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน เงื่อนไขความยาวของแบบทดสอบ 20 ข้อ, 40 ข้อและ 60 ข้อของปัจจัยความยาวของแบบทดสอบ เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10%, 20% และ 30% ของปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และเงื่อนไขขนาดตัวอย่าง 250 คน, 500 คน และ 1000 คนของปัจจัยขนาดตัวอย่าง และเมื่อมีนัยสำคัญทางสถิติแล้ว ผู้วิจัยทำการทดสอบความแตกต่างรายคู่ โดยใช้วิธีการทดสอบของเชฟเฟ่

สรุปผลการวิจัย

การสรุปผลการวิจัยแบ่งออกเป็นสามส่วน คือ ค่าสถิติพื้นฐานของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ผลการทดสอบปัจจัยที่ศึกษาที่มีผลต่ออัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ และผลการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบภายใต้ปัจจัยที่ศึกษา ดังรายละเอียดต่อไปนี้

1. ค่าสถิติพื้นฐานของอัตราความถูกต้อง และ อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ

การสรุปผลการวิจัยในตอนนี้เป็นผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่าโดยใช้วิธีตรวจสอบ 3 วิธี คือ วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความยาวของแบบทดสอบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และขนาดตัวอย่าง โดยสรุปในรูปค่าเฉลี่ยของอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ซึ่งมีผลการตรวจสอบดังนี้

1.1 ผลการตรวจสอบภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน พบว่า วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงสุด คิดเป็นค่าเฉลี่ยประมาณ .981 ในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และวิธีโพลีโตมัสซิบเทสท์มีอัตราความถูกต้องต่ำสุด คิดเป็นค่าเฉลี่ยประมาณ .435 ในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่าวิธีการ

วิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด คิดเป็นค่าเฉลี่ยประมาณ .014 และวิธีโพลีโตมัสซิบเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด คิดเป็นค่าเฉลี่ยประมาณ .402 ในเงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน นอกจากนี้ยังพบว่า ภายใต้เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน และข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ทั้ง 3 วิธีมีอัตราความถูกต้อง คิดเป็นค่าเฉลี่ยประมาณ .719 และ .668 ตามลำดับ และมีอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นค่าเฉลี่ยประมาณ .190 และ .183 ตามลำดับ

1.2 ผลการตรวจสอบภายใต้ปัจจัยความยาวของแบบทดสอบ

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยความยาวของแบบทดสอบ พบว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงสุด คิดเป็นค่าเฉลี่ยประมาณ .995 และวิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องต่ำสุด คิดเป็นค่าเฉลี่ยประมาณ .297 ในเงื่อนไขความยาวของแบบทดสอบ 60 ข้อ เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด คิดเป็นค่าเฉลี่ยประมาณ .001 ในเงื่อนไขความยาวของแบบทดสอบ 60 ข้อ และวิธีโพลีโตมัสซิบเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด คิดเป็นค่าเฉลี่ยประมาณ .524 ในเงื่อนไขความยาวของแบบทดสอบ 40 ข้อ นอกจากนี้ยังพบว่าภายใต้เงื่อนไขความยาวของแบบทดสอบ 20 ข้อ, 40 ข้อ และ 60 ข้อ ทั้ง 3 วิธีมีอัตราความถูกต้อง คิดเป็นค่าเฉลี่ยประมาณ .712, .872 และ .568 ตามลำดับ และมีอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นค่าเฉลี่ยประมาณ .203, .259 และ .133 ตามลำดับ

1.3 ผลการตรวจสอบภายใต้ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน พบว่า วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงสุด คิดเป็นค่าเฉลี่ยประมาณ .984 และวิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องต่ำสุด คิดเป็นค่าเฉลี่ยประมาณ .380 ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด คิดเป็นค่าเฉลี่ยประมาณ .013 ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% และวิธีโพลีโตมัสซิบเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด คิดเป็นค่าเฉลี่ยประมาณ .429 ในเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% นอกจากนี้ยังพบว่าภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10%, 20% และ 30% ทั้ง 3 วิธีมีอัตราความถูกต้อง คิดเป็นค่าเฉลี่ยประมาณ .607, .655 และ .748 ตามลำดับ และมีอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นค่าเฉลี่ยประมาณ .132, .193 และ .249 ตามลำดับ

1.4 ผลการตรวจสอบภายใต้ปัจจัยขนาดตัวอย่าง

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยขนาดตัวอย่าง พบว่า วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงสุด คิดเป็นค่าเฉลี่ยประมาณ .989 ในเงื่อนไขขนาดตัวอย่าง 1000 คน และวิธีโพลีโตมัสซิบเทสท์มีอัตราความถูกต้องต่ำสุด คิดเป็นค่าเฉลี่ยประมาณ .462 ในเงื่อนไขขนาดตัวอย่าง 250 คน เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด คิดเป็นค่าเฉลี่ยประมาณ .008 ในเงื่อนไขขนาดตัวอย่าง 1000 คน และวิธีโพลีโตมัสซิบเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด คิดเป็นค่าเฉลี่ยประมาณ .448 ในเงื่อนไขขนาดตัวอย่าง 1000 คน นอกจากนี้ยังพบว่า ภายใต้เงื่อนไขขนาดตัวอย่าง 250 คน, 500 คน และ 1000 คน ทั้ง 3 วิธีมีอัตราความถูกต้อง คิดเป็นค่าเฉลี่ยประมาณ .634, .704 และ .741 ตามลำดับ และมีอัตราความคลาดเคลื่อนประเภทที่ 1 คิดเป็นค่าเฉลี่ยประมาณ .140, .184 และ .235 ตามลำดับ

2. ผลการทดสอบปัจจัยที่ศึกษาที่มีผลต่ออัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ

การสรุปผลการวิจัยในตอนนี้เป็นผลการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่าด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติกแบบจัดอันดับ ภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัยคือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความยาวของแบบทดสอบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และขนาดตัวอย่าง ผลการทดสอบปัจจัยดังกล่าว เป็นดังนี้

2.1 ผลการทดสอบปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน

ผลการทดสอบปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนภายใต้วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ พบว่า ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนมีผลต่ออัตราความถูกต้องของวิธีโพลีโตมัสซิบเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับ อย่างมีนัยสำคัญทางสถิติที่ระดับ.01 และมีผลต่ออัตราความถูกต้องของวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกอย่างมีนัยสำคัญทางสถิติที่ระดับ.05 กล่าวคือ การตรวจสอบในข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกันมีผลทำให้วิธีโพลีโตมัสซิบเทสท์ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงกว่าการตรวจสอบในข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ซึ่งตรงข้ามกับวิธีการถดถอยโลจิสติกแบบจัดอันดับที่พบว่าการตรวจสอบในข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกันมีอัตราความถูกต้องต่ำกว่าการตรวจสอบในข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ส่วนผลการทดสอบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่าปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลีโตมัส

ซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ อย่างมีนัยสำคัญทางสถิติที่ระดับ.01 กล่าวคือ การตรวจสอบในข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกันมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลีโตมัสซิบเทสท์สูงกว่าการตรวจสอบในข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ซึ่งตรงข้ามกับการตรวจสอบด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกที่พบว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ในข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกันมีค่าต่ำกว่าการตรวจสอบในข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับ พบว่าการตรวจสอบในข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกันอัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าต่ำกว่าการตรวจสอบในข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ซึ่งเป็นไปในทิศทางเดียวกันกับวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก

2.2 ผลการทดสอบปัจจัยความยาวของแบบทดสอบ

ผลการทดสอบปัจจัยความยาวของแบบทดสอบ ภายใต้วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ พบว่าปัจจัยความยาวของแบบทดสอบมีผลต่ออัตราความถูกต้องของวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ อย่างมีนัยสำคัญทางสถิติที่ระดับ.01 กล่าวคือ เมื่อความยาวของแบบทดสอบเพิ่มขึ้นจาก 20 ข้อ เป็น 40 ข้อ มีผลทำให้อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของแต่ละวิธีมีค่าเพิ่มขึ้น แต่เมื่อความยาวของแบบทดสอบเพิ่มขึ้นจาก 40 ข้อเป็น 60 ข้อ ยังคงมีผลทำให้อัตราความถูกต้องของวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีค่าเพิ่มสูงขึ้น ส่วนอัตราความถูกต้องของวิธีโพลีโตมัสซิบเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าลดลง ส่วนผลการทดสอบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่าปัจจัยความยาวของแบบทดสอบมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ อย่างมีนัยสำคัญทางสถิติที่ระดับ.01 กล่าวคือในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบเมื่อความยาวของแบบทดสอบเพิ่มขึ้นจาก 20 ข้อ เป็น 40 ข้อ มีผลทำให้การตรวจสอบด้วยวิธีโพลีโตมัสซิบเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้น ขณะที่การตรวจสอบด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 ลดลง แต่เมื่อความยาวของแบบ ทดสอบเพิ่มขึ้นจาก 40 ข้อเป็น 60 ข้อมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้ง 3 วิธี มีค่าลดลง

2.3 ผลการทดสอบปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน

ผลการทดสอบปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ภายใต้วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ พบว่า ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนมีผลต่ออัตราความถูกต้องของวิธีโพลีโตมัสซิบเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับ อย่างมีนัยสำคัญทางสถิติที่ระดับ.01 และมีผลต่ออัตราความถูกต้องของวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก

อย่างมีนัยสำคัญทางสถิติที่ระดับ.05 กล่าวคือ เมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้นจาก 10% ถึง 30% มีผลทำให้อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีโพลีโตมัสซิบเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเพิ่มขึ้น ซึ่งไม่สอดคล้องกับวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกที่พบว่าเมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้นจาก 10% เป็น 20% มีผลทำให้อัตราความถูกต้องในการตรวจสอบมีค่าลดลง แต่เมื่อขนาดสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเป็น 30% อัตราความถูกต้องจะมีค่าไม่แตกต่างจากที่ขนาดสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% และมีค่าใกล้เคียงกับที่ขนาดสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% ส่วนผลการทดสอบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่าปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ.01 กล่าวคือ เมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้นจาก 10% ถึง 30% มีผลทำให้วิธีโพลีโตมัสซิบเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้น ซึ่งไม่สอดคล้องกับวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกที่พบว่า เมื่อขนาดสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเป็น 10% และ 30% อัตราความคลาดเคลื่อนประเภทที่ 1 จากการตรวจสอบมีค่าใกล้เคียงกัน

2.4 ผลการทดสอบปัจจัยขนาดตัวอย่าง

ผลการทดสอบปัจจัยขนาดตัวอย่าง ภายใต้วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ พบว่า ปัจจัยขนาดตัวอย่างมีผลต่ออัตราความถูกต้องของวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับอย่างมีนัยสำคัญทางสถิติที่ระดับ.01 กล่าวคือ เมื่อขนาดตัวอย่างเพิ่มขึ้นมีผลทำให้วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเพิ่มขึ้น ส่วนผลการทดสอบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ปัจจัยขนาดตัวอย่างมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ อย่างมีนัยสำคัญทางสถิติที่ระดับ.01 กล่าวคือเมื่อขนาดตัวอย่างเพิ่มขึ้นมีผลทำให้วิธีโพลีโตมัสซิบเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้น ซึ่งไม่สอดคล้องกับวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกที่พบว่า เมื่อขนาดตัวอย่างเพิ่มขึ้นอัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าลดลง

3. ผลการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบภายใต้ปัจจัยที่ศึกษา

การสรุปผลการวิจัยในตอนนี้เป็นผลการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่าระหว่างวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก

และวิธีการถดถอยโลจิสติกแบบจัดอันดับภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัยคือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความยาวของแบบทดสอบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และขนาดตัวอย่าง ผลการเปรียบเทียบภายใต้ทุกเงื่อนไขสรุปได้ดังนี้

3.1 ผลการเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน

กรณีข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน จากผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ อัตราความถูกต้องในการตรวจสอบด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีค่าเฉลี่ยประมาณ .651, .981 และ .525 ตามลำดับ โดยผลการเปรียบเทียบอัตราความถูกต้อง พบว่า วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยที่วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงกว่าวิธีโพลีโตมัสซิบเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับ สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับภายใต้เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน มีค่าเฉลี่ยประมาณ .402, .014 และ .154 ตามลำดับ โดยผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยที่วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ส่วนวิธีโพลีโตมัสซิบเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด เมื่อพิจารณาอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีตรวจสอบทั้ง 3 วิธี ภายใต้เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน พบว่า วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงสุด และมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ส่วนวิธีโพลีโตมัสซิบเทสท์มีอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ

กรณีข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน จากผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ อัตราความถูกต้องในการตรวจสอบด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีค่าเฉลี่ยประมาณ .435, .978 และ .590 ตามลำดับ โดยผลการเปรียบเทียบอัตราความถูกต้อง พบว่า วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยที่วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงกว่าวิธีโพลีโตมัสซิบเทสท์และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ขณะที่วิธีโพลีโตมัสซิบเทสท์มีอัตราความถูกต้องต่ำสุด สำหรับอัตรา

ความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีค่าเฉลี่ยประมาณ .320, .017 และ .211 ตามลำดับ โดยผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยที่วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ส่วนวิธีโพลีโตมัสซิบเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด เมื่อพิจารณาอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีตรวจสอบทั้ง 3 วิธีภายใต้เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน พบว่า วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงสุดและมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องสูงกว่าวิธีโพลีโตมัสซิบเทสท์ แต่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่า

3.2 ผลการเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยความยาวของแบบทดสอบ

กรณีความยาวของแบบทดสอบ 20 ข้อ จากผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ อัตราความถูกต้องในการตรวจสอบด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีค่าเฉลี่ยประมาณ .625, .942 และ .569 ตามลำดับ โดยผลการเปรียบเทียบอัตราความถูกต้อง พบว่า วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยที่วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงกว่าวิธีโพลีโตมัสซิบเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับ สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีค่าเฉลี่ยประมาณ .190, .047 และ .371 ตามลำดับ โดยผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่าวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยที่วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด เมื่อพิจารณาอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีตรวจสอบทั้ง 3 วิธี ภายใต้เงื่อนไขความยาวของแบบทดสอบ 20 ข้อ พบว่า วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงสุด และมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ส่วนวิธีโพลีโตมัสซิบเทสท์มีอัตราความถูกต้องสูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ แต่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่า

กรณีความยาวของแบบทดสอบ 40 ข้อ จากผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ อัตราความถูกต้องในการตรวจสอบด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชัน

การจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีค่าเฉลี่ยประมาณ .699, .974 และ .942 ตามลำดับ โดยผลการเปรียบเทียบอัตราความถูกต้องพบว่าวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยที่วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงกว่าวิธีโพลีโตมัสซิบเทสท์และวิธีการถดถอยโลจิสติกแบบจัดอันดับ สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีค่าเฉลี่ยประมาณ .524, .022 และ.232 ตามลำดับ โดยผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยที่วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ส่วนวิธีโพลีโตมัสซิบเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด เมื่อพิจารณาอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีตรวจสอบทั้ง 3 วิธี ภายใต้เงื่อนไขความยาวของแบบทดสอบ 40 ข้อ พบว่า วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงสุด และมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องสูงกว่าวิธีโพลีโตมัสซิบเทสท์ แต่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่า

กรณีความยาวของแบบทดสอบ 60 ข้อ จากผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ อัตราความถูกต้องในการตรวจสอบด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีค่าเฉลี่ยประมาณ .412, .995 และ .297 ตามลำดับ โดยผลการเปรียบเทียบอัตราความถูกต้อง พบว่า วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยที่วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงกว่าวิธีโพลีโตมัสซิบเทสท์และวิธีการถดถอยโลจิสติกแบบจัดอันดับ สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีค่าเฉลี่ยประมาณ .310, .001 และ .087 ตามลำดับ โดยผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่าวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยที่วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ส่วนวิธีโพลีโตมัสซิบเทสท์มีอัตราความคลาดเคลื่อนประเภท ที่ 1 สูงสุด เมื่อพิจารณาอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีตรวจสอบทั้ง 3 วิธี ภายใต้เงื่อนไขความยาวของแบบทดสอบ 60 ข้อ พบว่า วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงสุด และมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด

ขณะที่วิธีโพลีโตมัสซิบเทสท์ มีอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ

3.3 ผลการเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน

กรณีสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% จากผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ อัตราความถูกต้องในการตรวจสอบด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีค่าเฉลี่ยประมาณ .456, .984 และ.380 ตามลำดับ โดยผลการเปรียบเทียบอัตราความถูกต้อง พบว่า วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยที่วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงกว่าวิธีโพลีโตมัสซิบเทสท์และวิธีการถดถอยโลจิสติกแบบจัดอันดับ สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีค่าเฉลี่ยประมาณ .309, .013 และ .075 ตามลำดับ โดยผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยที่วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ส่วนวิธีโพลีโตมัสซิบเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด เมื่อพิจารณาอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีตรวจสอบทั้ง 3 วิธี ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% พบว่า วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงสุด และมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ส่วนวิธีโพลีโตมัสซิบเทสท์มีอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีการถดถอยโลจิสติกแบบจัดอันดับ

กรณีสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% จากผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ อัตราความถูกต้องในการตรวจสอบด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีค่าเฉลี่ยประมาณ .491, .976 และ .498 ตามลำดับ โดยผลการเปรียบเทียบอัตราความถูกต้อง พบว่า วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยที่วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงกว่าวิธีโพลีโตมัสซิบเทสท์และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ขณะที่วิธีโพลีโตมัสซิบเทสท์และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีอัตราความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนใกล้เคียงกัน สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติก

แบบจัดอันดับ มีค่าเฉลี่ยประมาณ .361, .020 และ .199 ตามลำดับ โดยผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.01 โดยที่วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ส่วนวิธีโพลีโตมัสซิบเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด เมื่อพิจารณาอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีตรวจสอบทั้ง 3 วิธี ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 20% พบว่า วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงสุด และมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ส่วนวิธีโพลีโตมัสซิบเทสท์มีอัตราความถูกต้องใกล้เคียงกับวิธีการถดถอยโลจิสติกแบบจัดอันดับ แต่มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่า

กรณีสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% จากผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ อัตราความถูกต้องในการตรวจสอบด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีค่าเฉลี่ยประมาณ .607, .980 และ .656 ตามลำดับ โดยผลการเปรียบเทียบอัตราความถูกต้อง พบว่า วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยที่วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงกว่าวิธีโพลีโตมัสซิบเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับ สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีค่าเฉลี่ยประมาณ .429, .015 และ .302ตามลำดับ โดยผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยที่วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ส่วนวิธีโพลีโตมัสซิบเทสท์มีอัตราความคลาดเคลื่อนประเภท ที่ 1 สูงสุด เมื่อพิจารณาอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีตรวจสอบทั้ง 3 วิธี ภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% พบว่า วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงสุด และมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องสูงกว่าวิธีโพลีโตมัสซิบเทสท์ แต่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่า

3.4 ผลการเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้ปัจจัยขนาดตัวอย่าง

กรณีขนาดตัวอย่าง 250 คน จากผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ อัตราความถูกต้องในการตรวจสอบด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีค่าเฉลี่ยประมาณ .462, .969 และ .472

ตามลำดับ โดยผลการเปรียบเทียบอัตราความถูกต้อง พบว่า วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยที่วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงกว่าวิธีโพลีโตมัสซิบเทสท์และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ขณะที่วิธีโพลีโตมัสซิบเทสท์และวิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องใกล้เคียงกัน สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีค่าเฉลี่ยประมาณ .265, .024 และ .132 ตามลำดับ โดยผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยที่วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ส่วนวิธีโพลีโตมัสซิบเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด เมื่อพิจารณาอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีตรวจสอบทั้ง 3 วิธีภายใต้เงื่อนไขขนาดตัวอย่าง 250 คน พบว่า วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงสุด และมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ส่วนวิธีโพลีโตมัสซิบเทสท์มีอัตราความถูกต้องใกล้เคียงกับวิธีการถดถอยโลจิสติกแบบจัดอันดับ แต่มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่า

กรณีขนาดตัวอย่าง 500 คน จากผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ อัตราความถูกต้องในการตรวจสอบด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าเฉลี่ยประมาณ .563, .979 และ .569 ตามลำดับ โดยผลการเปรียบเทียบอัตราความถูกต้อง พบว่า วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยที่วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงกว่าวิธีโพลีโตมัสซิบเทสท์และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ขณะที่วิธีโพลีโตมัสซิบเทสท์และวิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องใกล้เคียงกัน สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีค่าเฉลี่ยประมาณ .370, .015 และ .167 ตามลำดับ โดยผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.01 โดยที่วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ส่วนวิธีโพลีโตมัสซิบเทสท์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด เมื่อพิจารณาอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีตรวจสอบ ทั้ง 3 วิธีภายใต้เงื่อนไขขนาดตัวอย่าง 500 คน พบว่า วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงสุด และมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ส่วนวิธี

โพลีโตมัสซิบเทสท์มีอัตราความถูกต้องใกล้เคียงกับวิธีการถดถอยโลจิสติกแบบจัดอันดับ แต่มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่า

กรณีขนาดตัวอย่าง 1000 คน จากผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ อัตราความถูกต้องในการตรวจสอบด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีค่าเฉลี่ยประมาณ .604, .989 และ .631 ตามลำดับ โดยผลการเปรียบเทียบอัตราความถูกต้อง พบว่า วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยที่วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงกว่าวิธีโพลีโตมัสซิบเทสท์และวิธีการถดถอยโลจิสติกแบบจัดอันดับ สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีค่าเฉลี่ยประมาณ .448, .008 และ .248 ตามลำดับ โดยผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยที่วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ส่วนวิธีโพลีโตมัสซิบเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงสุด เมื่อพิจารณาอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีตรวจสอบทั้ง 3 วิธีภายใต้เงื่อนไขขนาดตัวอย่าง 1000 คน พบว่า วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงสุด และมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำสุด ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความถูกต้องสูงกว่าวิธีโพลีโตมัสซิบเทสท์ แต่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่า

สรุปผลการวิจัยในภาพรวม

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่าของวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความยาวของแบบทดสอบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และขนาดตัวอย่าง สามารถสรุปโดยภาพรวมได้ดังนี้

1. ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีที่ใช้ตรวจสอบ

1.1 การตรวจสอบปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน

ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ พบว่าการตรวจสอบปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ให้ผลการตรวจสอบที่แตกต่างกัน ซึ่งการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบของปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน วิธี

โพลีโตมัสซิบเทสท์ และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกันสูงกว่าข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ซึ่งไม่สอดคล้องกับวิธีการถดถอยโลจิสติกแบบจัดอันดับ เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีโพลีโตมัสซิบเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกันสูงกว่าข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน ส่วนวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกันต่ำกว่าข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน

1.2 การตรวจสอบปัจจัยความยาวของแบบทดสอบ

การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในแบบทดสอบด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับของปัจจัยความยาวของแบบทดสอบ เมื่อความยาวของแบบทดสอบเพิ่มขึ้น อัตราความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีค่าเพิ่มขึ้น ส่วนวิธีโพลีโตมัสซิบเทสท์และวิธีการถดถอยโลจิสติกแบบจัดอันดับมีลักษณะคล้ายคลึงกัน เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า เมื่อความยาวของแบบทดสอบเพิ่มขึ้น อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติกแบบจัดอันดับมีค่าลดลง ขณะที่วิธีโพลีโตมัสซิบเทสท์มีค่าไม่แน่นอน

1.3 การตรวจสอบปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน

ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ พบว่า วิธีโพลีโตมัสซิบเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีอัตราความถูกต้องสูงขึ้นในทุกเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนที่เพิ่มขึ้น สำหรับเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 30% วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องไม่แตกต่างจากที่ขนาดสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% และ 20% เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า เมื่อขนาดสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้น วิธีโพลีโตมัสซิบเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงขึ้นในทุกเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนที่เพิ่มขึ้น ส่วนวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนของเงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 10% และ 30% มีค่าไม่แตกต่างกัน

1.4 การตรวจสอบปัจจัยขนาดตัวอย่าง

ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ พบว่า การตรวจสอบปัจจัยขนาดตัวอย่างด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับให้ผลการตรวจสอบที่คล้ายคลึงกัน กล่าวคือ อัตราความถูกต้องมีค่าเพิ่มขึ้นในทุกเงื่อนไขของปัจจัยขนาดตัวอย่างที่เพิ่มขึ้น เมื่อพิจารณาอัตราความคลาดเคลื่อน

ประเภทที่ 1 พบว่า เมื่อขนาดตัวอย่างเพิ่มขึ้น วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 ลดลง ซึ่งตรงข้ามกับวิธีโพลีโตมัสซิบเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับ

2. ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบภายใต้ปัจจัยที่ศึกษา

ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบระหว่างวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ภายใต้ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความยาวของแบบทดสอบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และขนาดตัวอย่าง ผลเป็นดังนี้

ผลการเปรียบเทียบอัตราความถูกต้องระหว่างวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่า ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความยาวของแบบทดสอบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และขนาดตัวอย่าง พบว่า การตรวจสอบภายใต้เงื่อนไขต่าง ๆ วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีอัตราความถูกต้องแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือทั้ง 3 วิธีมีความถูกต้องในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนแตกต่างกัน โดยที่วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องสูงกว่าวิธีโพลีโตมัสซิบเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับ

ส่วนผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ภายใต้เงื่อนไขต่าง ๆ วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ ทั้ง 3 วิธี มีความผิดพลาดในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนแตกต่างกัน โดยที่วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าวิธีโพลีโตมัสซิบเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับ

อภิปรายผลการวิจัย

1 ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่า พบว่าปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนมีผลต่ออัตราความถูกต้องของวิธีโพลีโตมัสซิบเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับ อย่างมีนัยสำคัญทางสถิติที่ระดับ.01 และมีผลต่ออัตราความถูกต้องของวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก อย่างมีนัยสำคัญทางสถิติที่ระดับ.05 ในขณะที่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งสามวิธีอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ผลดังกล่าวเป็นไปตามสมมติฐานข้อ 1-3 โดยผลการตรวจสอบการทำหน้าที่

เบี่ยงเบนของข้อสอบภายใต้ปัจจัยข้อสอบทำหน้าที่เบี่ยงเบนที่เป็นรูปแบบเดียวกัน ทั้งสามวิธีมีค่าอัตราความถูกต้องเกิน 50% ขณะที่ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบภายใต้เงื่อนไขข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน วิธีโพลีโตมัสซิบเทสท์มีค่าอัตราความถูกต้องเฉลี่ยประมาณ .435 คิดเป็น 43.5% ซึ่งต่ำที่สุด และต่ำกว่าการตรวจสอบด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับที่อัตราความถูกต้องมีค่าเฉลี่ยประมาณ .978 และ .590 คิดเป็น 97.8% และ 59.0% ตามลำดับ ทั้งนี้เป็นเพราะวิธีโพลีโตมัสซิบเทสท์ ไม่มีความไวในการตรวจสอบข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกัน เนื่องจากวิธีโพลีโตมัสซิบเทสท์พัฒนามาจากวิธีซิบเทสท์ซึ่งเป็นวิธีที่ใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่มีทิศทางเดียวโดยเฉพาะ ดังนั้นจึงมีข้อจำกัดในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ไม่มีทิศทางเดียวกันที่ใช้สถิติแบบคิดเครื่องหมาย (Li; & Stout. 1996; Shealy; & Stout. 1993)

2. ปัจจัยความยาวของแบบทดสอบ

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่า ความยาวของแบบทดสอบมีผลต่ออัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ผลดังกล่าวเป็นไปตามสมมติฐานข้อ 1-3 กล่าวคือ เมื่อเพิ่มความยาวแบบทดสอบจาก 20 ข้อ เป็น 40 ข้อ อัตราความถูกต้องในการตรวจสอบของวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับจะเพิ่มขึ้นเฉลี่ยประมาณ .074, .032 และ .373 คิดเป็น 7.4%, 3.2% และ 37.3% ตามลำดับ ซึ่งสอดคล้องกับสวามินาธานและโรเจอร์ส (Swaminathan; & Rogers. 1990) ที่ว่าแบบทดสอบที่ยาวและขนาดตัวอย่างที่มากขึ้น ทำให้โอกาสในการตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนสูงขึ้น แต่เมื่อความยาวแบบทดสอบเป็น 60 ข้อ พบว่าอัตราการตรวจสอบของวิธีโพลีโตมัสซิบเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับลดลง ซึ่งขัดแย้งกับคำกล่าวข้างต้นทั้งนี้อาจเนื่องมาจาก เมื่อความยาวของแบบทดสอบ 60 ข้อ ซึ่งเป็นแบบทดสอบที่ยาวมากจะทำให้ค่าความยากเฉลี่ยของข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มสนใจมีค่าใกล้เคียงกัน ทำให้ทั้งสองกลุ่มมีความสามารถใกล้เคียงกัน อาจมีผลทำให้ตรวจพบข้อสอบทำหน้าที่เบี่ยงเบนได้ถูกต้องลดลง

เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่าปัจจัยความยาวของแบบทดสอบมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 กล่าวคือ เมื่อความยาวของแบบทดสอบ 20 ข้อ, 40 ข้อ และ 60 ข้อ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก มีอัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าเฉลี่ยประมาณ .047, .022 และ .001 คิดเป็น 4.7%, 2.2% และ 0.1% ตามลำดับ ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับมีอัตราความคลาดเคลื่อนประเภทที่ 1 ของข้อสอบทำหน้าที่เบี่ยงเบนในแบบทดสอบที่มีความยาว 20 ข้อ, 40 ข้อ และ 60 ข้อ เฉลี่ยประมาณ .371, .232 และ .087

คิดเป็น 37.1%, 23.2% และ 8.7% ตามลำดับ ซึ่งสอดคล้องกับผลการศึกษาของอุททาโรและมิลล์แซบ (Uttaro; & Millsap. 1994) ที่ว่าแบบทดสอบที่ยาวขึ้นอัตราความคลาดเคลื่อนประเภทที่ 1 จะลดลง สำหรับวิธีโพลีโตมัสซิบเทสท์มีอัตราความคลาดเคลื่อนประเภทที่ 1 ในแบบทดสอบที่มีความยาว 20 ข้อ และ 40 ข้อ เฉลี่ยประมาณ .190 และ.524 คิดเป็น 19.0% และ 52.4% ตามลำดับ จะเห็นว่าเมื่อแบบทดสอบมีความยาวมากขึ้นอัตราความคลาดเคลื่อนประเภทที่ 1 จะเพิ่มขึ้น ซึ่งสอดคล้องกับผลการศึกษาของสวามินาธานและโรเจอร์ส (Swaminathan; & Rogers. 1990) และผลการศึกษาของ โคเฮ็นและคิม (Cohen; & Kim. 1993) ที่พบว่าเมื่อเพิ่มความยาวของแบบทดสอบแล้วอัตราความคลาดเคลื่อนประเภทที่ 1 จะเพิ่มขึ้น

3. ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่า พบว่า ปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนมีผลต่ออัตราความถูกต้องของวิธีโพลีโตมัสซิบเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และมีผลต่ออัตราความถูกต้องของวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ในขณะที่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งสามวิธี อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ผลดังกล่าวเป็นไปตามสมมติฐานข้อ 1-3 กล่าวคือ เมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้น จาก 10% เป็น 20% วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีอัตราความถูกต้องลดลง และมีอัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้น ซึ่งสอดคล้องกับผลการศึกษาของนารายานาน และสวามินาธาน (Narayanan; & Swaminathan. 1996) ขณะที่วิธีโพลีโตมัสซิบเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับ มีอัตราความถูกต้องและ อัตราความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้น ซึ่งไม่สอดคล้องกับผลการศึกษาของวลีมาศ แซ่อึ้ง (2543) และอรินทร์ น่วมถนอม (2549) ที่พบว่า สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนในช่วง 10% ถึง 20% ไม่มีผลต่อวิธีซิบเทสท์และผลการศึกษาของโอชิมาและมิลเลอร์ (Oshima; & Miller.1992) ที่พบว่าเมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนเพิ่มขึ้นแล้วอัตราความถูกต้องจะลดลง และการศึกษาของโคเฮ็นและคิม (Cohen; & Kim.1993) ซึ่งพบว่าเมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนในแบบทดสอบมีจำนวนลดลงแล้วอัตราความคลาดเคลื่อนประเภทที่ 1 จะมีค่าเพิ่มขึ้น

4. ปัจจัยขนาดตัวอย่าง

ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ พบว่า ขนาดตัวอย่างมีผลต่ออัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบด้วยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ผลดังกล่าวเป็นไปตามสมมติฐานข้อ 1-3 กล่าวคือเมื่อใช้ขนาดตัวอย่าง 250 คนต่อกลุ่ม ทั้งสามวิธีมีอัตราความถูกต้องต่ำสุดเมื่อเทียบกับขนาดตัวอย่างอื่นๆ โดยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับสามารถตรวจสอบได้ถูกต้องเฉลี่ยประมาณ .462, .969 และ .472 คิดเป็น 46.2%, 96.9%

และ 47.2% ตามลำดับ เมื่อขนาดตัวอย่างใหญ่ขึ้นเป็น 500 คนต่อกลุ่ม มีผลทำให้อัตราความถูกต้องของทั้งสามวิธีเพิ่มขึ้นทุกเงื่อนไข โดยวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับสามารถตรวจสอบได้ถูกต้องเฉลี่ยประมาณ .563, .979 และ .569 คิดเป็น 56.3%, 97.9% และ 56.9% ตามลำดับ ในขณะที่เมื่อขนาดตัวอย่างเป็น 1000 คนต่อกลุ่ม จะยังมีผลทำให้อัตราความถูกต้องของทั้งสามวิธีเพิ่มขึ้นเช่นเดิม โดยที่วิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับสามารถตรวจสอบได้ถูกต้องเฉลี่ยประมาณ .604, .989 และ .631 คิดเป็น 60.4%, 98.9% และ 63.1% ตามลำดับ ซึ่งสอดคล้องกับผลการศึกษาของโรเจอร์ส และสวามินาธาน (Rogers; & Swaminathan.1993) และผลการศึกษาของนารายานานและสวามินาธาน(Narayanan; & Swaminathan. 1996) ที่พบว่าเมื่อขนาดตัวอย่างเพิ่มขึ้นจะมีผลทำให้อัตราความถูกต้องเพิ่มขึ้น

สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ดี ไม่เกินร้อยละ 10 ในทุกเงื่อนไข และมีขนาดลดลงเมื่อขนาดตัวอย่างมีค่าเพิ่มขึ้น ขณะที่วิธีโพลีโตมัสซิบเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับมีแนวโน้มว่า เมื่อขนาดตัวอย่างเพิ่มขึ้นอัตราความคลาดเคลื่อนจะเพิ่มสูงขึ้นด้วย ซึ่งขัดแย้งกับการศึกษาของนารายานานและสวามินาธาน (Narayanan; & Swaminathan. 1996; ทองอยู่ สาระ 2543) ที่พบว่าขนาดกลุ่มตัวอย่างไม่มีผลกระทบต่ออัตราความคลาดเคลื่อนประเภทที่ 1 แต่สอดคล้องกับผลการศึกษาของอุทัยวรรณ สายพัฒนะ (2547) ที่พบว่าความผิดพลาดของการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ตรวจให้คะแนนแบบหลายค่าด้วยวิธีโพลีโตมัสซิบเทสท์เมื่อกลุ่มตัวอย่างมีขนาดใหญ่ขึ้นความผิดพลาดในการตรวจสอบจะมีค่ามากขึ้นด้วย

5. ผลการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ

ผลการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีโพลีโตมัสซิบเทสท์ วิธีวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่า ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความยาวของแบบทดสอบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และขนาดตัวอย่าง พบว่าภายใต้แต่ละเงื่อนไขทั้งสามวิธีมีอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ผลการเปรียบเทียบดังกล่าวมีนัยสำคัญสอดคล้องกับสมมติฐานข้อ 4 โดยวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีประสิทธิภาพในการตรวจสอบสูงสุด เพราะแต่ละเงื่อนไขมีอัตราความถูกต้องตั้งแต่ 80% ขึ้นไป ซึ่งถือว่ามีความถูกต้องเพียงพอ (Cohen. 1988) และมีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในขอบเขตที่กำหนด (Nominal limits) ไม่เกิน 5%

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1.1 ผลการศึกษาครั้งนี้พบว่า การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่าภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน ความยาวของแบบทดสอบ สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน และขนาดตัวอย่าง โดยภาพรวมภายใต้ทุกเงื่อนไข วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความเหมาะสมมากที่สุด เมื่อเทียบกับวิธีโพลีโตมัสซิบเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ทั้งนี้เพราะวิธีดังกล่าวให้ผลการตรวจสอบการทำหน้าที่เบี่ยงเบนมีอัตราความถูกต้องสูง และมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำ อยู่ในเกณฑ์ที่กำหนด

1.2 ผู้ปฏิบัติควรระลึกว่าการทำหน้าที่เบี่ยงเบนของข้อสอบไม่ใช่ความลำเอียงของข้อสอบ การทำหน้าที่เบี่ยงเบนของข้อสอบเป็นสารสนเทศทางสถิติ เป็นดัชนีตัวหนึ่งในการประเมินความลำเอียงของข้อสอบเท่านั้น ดังนั้นการตีความผลการวิเคราะห์ทางสถิติจึงควรกระทำด้วยความระมัดระวัง การที่ข้อสอบข้อใดข้อหนึ่งทำหน้าที่เบี่ยงเบนไม่ได้หมายความว่าข้อสอบข้อนั้นไม่ดีเสมอไป การนำดัชนีตัวนี้ไปใช้เพื่อคัดเลือกข้อสอบออกจากแบบทดสอบหรือคลังข้อสอบ คงต้องพิจารณาวัตถุประสงค์ในการวัดของแบบทดสอบและเนื้อหาสาระของข้อสอบ

2. ข้อเสนอแนะในการวิจัยครั้งต่อไป

2.1 ควรทำการศึกษาซ้ำด้วยวิธีการเดิม เพื่อเปรียบเทียบวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ภายใต้เงื่อนไขหรือปัจจัยอื่น ๆ เช่น ตรวจสอบข้อมูลที่วัดความสามารถหลายมิติ ศึกษาที่ระดับนัยสำคัญต่างๆ หรือศึกษาภายใต้ปัจจัยความแตกต่างของการแจกแจงความสามารถ ขนาดตัวอย่างระหว่างกลุ่มอ้างอิงและกลุ่มสนใจที่ไม่ใช่อัตราส่วน 1:1 ทิศทางของข้อสอบทำหน้าที่เบี่ยงเบน อื่น ๆ ซึ่งจะช่วยขยายขอบเขตทางการวัดผลให้กว้างขวางขึ้น

2.2 ควรศึกษาเปรียบเทียบผลการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบลักษณะทำนองเดียวกันกับการศึกษาครั้งนี้ แต่ใช้ข้อมูลจริง โดยอาจใช้ตัวแปร เพศ เชื้อชาติ ศาสนา ภูมิลำเนา อายุ หรือประสบการณ์ เป็นเกณฑ์ในการแบ่งกลุ่มผู้สอบ

2.3 ควรได้มีการศึกษาปฏิสัมพันธ์ระหว่างตัวแปรต่าง ๆ เช่นความยาวของแบบทดสอบกับรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน หรือตัวแปรอื่น ๆ เพื่อศึกษาว่ามีผลต่ออัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบหรือไม่

2.4 ควรมีการศึกษาเพิ่มเติมเกี่ยวกับประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ เมื่อมีผู้สอบหลายกลุ่ม และควรศึกษาเกี่ยวกับการตรวจสอบการทำหน้าที่เบี่ยงเบนของกลุ่มข้อสอบ (differential bundle functioning: DBF)

บรรณานุกรม

บรรณานุกรม

กาญจนา วัธนสุนทร. (2537). *การพัฒนาเกณฑ์ตัดสินข้อสอบลำเอียงทางเพศ.* วิทยานิพนธ์ ค.ด. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.

จิติมา วรรณศรี. (2539). *การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซลกับวิธีซิปเทสท์ เมื่อความยาวของแบบทดสอบ ขนาดกลุ่มตัวอย่างและอัตราส่วนของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบต่างกัน.* วิทยานิพนธ์ ค.ม. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.

ทองอยู่ สาระ. (2543). *การเปรียบเทียบอำนาจการตรวจสอบและการจำแนกผิดพลาดในการตรวจสอบข้อสอบทำหน้าที่เบี่ยงเบนแบบสม่ำเสมอและแบบไม่สม่ำเสมอระหว่างวิธีแมนเทล-แฮนส์เซลกับวิธีการถดถอยโลจิสติก โดยใช้ความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่างต่างกัน.* ปริญญานิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ: บัญฑิตวิทยาลัยมหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

นพมาศ พิพัฒนสุข. (2541). *การเปรียบเทียบประสิทธิภาพระหว่างวิธีแมนเทล-แฮนส์เซลกับวิธีการถดถอยโลจิสติก ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบเมื่อใช้เกณฑ์จับคู่เปรียบเทียบแตกต่างกันในแบบสอบพหุมิติ.* วิทยานิพนธ์ ค.ม. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.

บุญเชิด ภิญโญอนันตพงษ์. (2547). *การวัดประเมินการเรียนรู้* (การวัดประเมินแนวใหม่). กรุงเทพฯ: คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.

รัชรินทร์ มุคดา. (2540). *การเปรียบเทียบประสิทธิภาพระหว่างวิธีแมนเทล-แฮนส์เซลกับวิธีถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป ในกรณีการจัดกลุ่มความสามารถ ค่าความยากของข้อสอบ และค่าอำนาจจำแนกของข้อสอบต่างกัน.* วิทยานิพนธ์ปริญญามหาบัณฑิต. วิทยานิพนธ์ ค.ด. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.

วริษษา ชะม้อย. (2550). การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาค ระหว่างวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ยกับวิธีการวิเคราะห์ฟังชันก์เชิงจำแนกแบบโลจิสติก. วิทยานิพนธ์ ค.ม.(การวัดและประเมินผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.

วลีมาศ แซ่อึ้ง. (2543). *การเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบหลายรูปแบบ ระหว่างวิธีซิปเทสท์ปรับใหม่*

วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก. วิทยานิพนธ์ ค.ด. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.

สิริรัตน์ วิภาสศิลป์. (2545). *การเปรียบเทียบวิธีซิปเทสท์และดีเอฟไอทีในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบ จากข้อมูลการตอบข้อสอบที่ใช้วัดความสามารถหลายมิติ.* ปริญญานิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

ศิริชัย กาญจนวาสี. (2545). *ทฤษฎีการทดสอบแนวใหม่.* โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

อรินทร์ น่วมถนอม.(2549). *การเปรียบเทียบวิธีโพลี-ซิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับและวิธีการจัดอันดับหลายมิติ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติ และให้คะแนนหลายค่า.* ปริญญานิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา. กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

อุทัยวรรณ สายพัฒนะ. (2547). *การเปรียบเทียบประสิทธิภาพของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในแบบทดสอบที่มีการให้คะแนนแบบหลายค่าระหว่างวิธี GMH และวิธี Poly-SIBTEST.* ปริญญานิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา). 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; Gierl, M.J.; & Walker, C.M. (2003, Fall). Using multidimensional item response theory to evaluate educational and psychological tests. *Educational Measurement : Issues and practice.* 22(3) :37-51.

Agresti,A. (1990). *Categorical data analysis*. New York: Wiley.

American Educational Research Association; American Psychological Association; & National Council on Measurement in Educational. (1999). *Standards for education and psychological testing.* Washington, DC: American Educational Research Association.

Andrich, D. (1978). Application of a psychometric model to ordered categories which are scored with successive integers. *Applied Psychological Measurement*, 2, 581-594.

_____. (1978). A rating formulation for ordered response categories. *Psychometrika,* 43, 561-573.

Angoff, W.H. (1993). Perspectives on differential item functioning methodology.*In differential item functioning*. P.W.Holland; & H. Wainer. pp.3-23.Hillsdale, NJ: Erlbaum.

Ankenmann, R.D., Witt, E.A., Dunbar, S.B. (1999). An investigation of the power of the likelihood ratio goodness-of-fit statistic in detecting differential item functioning. *Journal of Educational Measureme.* 36, 277-300.

Bock, R.D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika,* 37,29-51.

Bolt, D.M. (2002). A monte carlo comparison of parametric and nonparametric polytomous DIF detection methods. *Applied Measurement in Education*. 15(2):113-141.

Bolt, D.M.; & Gierl, M.J. (2006, Winter). Testing features of graphical DIF : Application of a regression correctionto three nonparametric statistical tests. *Journal of Educational Measurement.* 43(4) : 313-333.

Budgell, G.R.; Raju, N.S.; & Quartetti,D.A.(1995). Analysis of differential item functioning in translated assessment instruments. *Applied Psychological Measurement*. 19(4) : 309-321.

Camilli, G. (1992, June). A conceptual analysis of differential 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.* California: Sage Publications, Inc.,.

Chang, H.; Mazzeo,J.; & Roussos, L. (1996, Fall). Detecting DIF for polytomously scored items: An adaptation of the SIBTEST procedure. *Journal of Educational Measuremen*t. 33(3): 333-353.

Clauser, B. E; & Hambleton, R.K. (1994, Spring). Differential item functioning. *Journal of Educational Measuremen*t. 31(1): 88-92.

Clauser, B. 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.

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. 35-66. Hillsdale, NJ : Erlbaum.

Dorans, N.J.; & Kulick, E. (1986, Winter). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement.* 23 (4) : 355-368.

French, A. W.; & 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.

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.

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.

Han, K.T.; & Hambleton, R. K. (2007). User 's Manual WinGen2. Amherst, MA: University of Massachusetts, Center for Educational Assessment.

Hidalgo, M.D.;& Pina,J.A. (2004). Differential item functioning detection and effect size : A comparison between logistic regression and Mantel-Haenszel procedures. *Educational and Psychological Measurement.* 64 (6) : 903-915.

Hulin, C. L.;Drasgow, F.; & Komocar, J. (1982). Applications of item response theory to analysis of attitude scale translations. *Journal of Applied Psychology*. 67: 818-825.

Holland, P. W.; & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In Test validity. H. Wainer, & H. I. Braun. 129-145. Hillsdale, NJ: Erlbaum.

Holland, P.W.; & Wainer,H. (1993). Differential item functioning. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Jodoin, M. G.; & Gierl, M. J. (2001). Evaluating type I error and power rates using an effect size measure with the logistic regression procedure for DIF detection. *Applied Measurement in Education*. 14(4): 329-349.

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.

_____. (1992,Spring). Effects of linking methods on detection of DIF. *Journal of Educational Measurement. 29* (1): 51-66.

Kim, S.-H.; Cohen, A.S.; & Kim, H.-O. (1994,September). An investigation of Lord's procedure for the detection of differential item functioning. *Applied Psychological Measurement*.18 (3): 217-228.

Kristjansson, E. (2001). Detecting DIF in polytomous item : An empirical comparison of the Ordinal logistic regression, Logistic discriminant function analysis, Mantel, and Generalized Mantel- Haenszel procedures. Ottawa, Canada: Unpublished doctorial dissertation, Faculty of Education, University of Ottawa.

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-H.; & Stout, W. (1996, December). A new procedure for detection of crossing DIF. *Psychometrika.* 61(4) : 647-677.

Linn, R. L.; & Hamisch, D. L. (1981). Interactions between item content and group membership on achievement test items. *Journal of Educational Measurement.* 18: 109-118.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.

Maller, S.J. (2001). Differential item functioning in the WISC-III : Item parameter for boys and girls in the national standardization sample. *Educational and Psychological Measurement.* 61(5) : 793-817.

Master,G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika*.47 : 149-174.

Mazor, K. M.; Clauser, B. E.; & Hambleton, R. K. (1994,Summer). Identification of nonuniform differential item functioning using a variation of the Mantel-Haenszel procedure. *Educational and Psychological Measurement*. 54 (2): 284- 291.

Mazor, K. M.; Hambleton, R. K.; & Clauser, R. E. (1998, December). Multidimensional DIF analyses: The effects of matching on unidimensional subtest scores. *Applied Psychological Measurement.* 22(4) : 357-367.

Mellenbergh, G. J. (1982).Contingency table models for assessing item bias. *Journal of Educational statistic*s. 7(2) : 105-118.

Miller, T. R.; & Spray, J. A. (1993). 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. (1990). Fitting a polytomous item response model to Likert-type data. *Applied Psychological Measuremen.* 14(1), 59-71.

Muraki, E. (1992, June). A generalized partial credit model : Application of an EM algorithm. *Applied Psychological Measurement*. 16(2) : 159-176.

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. (1992, September). Multidimensionality 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: Advance in measurement and data analysis*. F. Drasgow; and N. Schmitt. pp. 156-188. San Francisco: CA: Jossey-Bass.

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.A.; & Stout, W.F. (1996, 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 Educational Measurement* .33(2) : 215-230.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph No. 17,* 34 (4, Pt. 2).

Scheuneman, J. (1979,Fall). A method of assessing bias in test items. *Journal of Educational Measurement* .16 (3): 143-152.

Shealy, R.; & Stout, W.F. (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. (1985). Validity of approximation techniques for detecting item bias. *Journal of Educational Measurement*. 22(2) : 77-105.

Spray, J.; & Miller, T.(1994). Identifying nonuniform DIF in polytomously scored test items. Iowa City,bIowa : American College Testing Report Series 94-1.

Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika.* 52 :589-617.

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.

Su,Y.H.; & Wang, W.C. (2005). Efficiency of the Mantel, Generalized Mantel-Haenszel, and Logistic discriminant function analysis methods in detecting differential item functioning for polytomous items. *Applied Measurement in Education* .18 (4): 313-350.

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 P.W., Holland, and H. Wainer (Eds.), *Differential item functioning*, pp. 67-113. Hillsdale, NJ: Erlbaum.

Tian,F. (1999). Detecting DIF in polytomous item responses. Ottawa, Canada :Unpublished doctorial dissertation, Faculty of Education, University of Ottawa.

Uttaro, T.; & Millsap, R.E. (1994). Factors influencing the Mantel-Haenszel procedure in the detection of differential item functioning. *Applied Psychological Measurement*. 18(1) :15-25.

Welch, C.J.; & Hoover, H.D. (1993). Procedures for extending item bias techniques to polytomously scored items. *Applied Measurement in Education.* 6 : 1-19.

Welch, C.J.; & Miller, T.R. (1995). Assessing differential item functioning in direct writing assessments: problems and an example. *Journal of Educational Measurement. 32* : 163-178.

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.

Zwick, R.; et al. (1997). Descriptive and inferential procedures for assessing differential item functioning in polytomous items. *Applied Measurement in Education.*10 : 321-344.

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.

ภาคผนวก

ตัวอย่างคำสั่งที่วิเคราะห์ด้วยโปรแกรม DIFPACK

name of input parameter file = SIB.IN

number of items on test = 20

name of file for Ref. grp. scores = d:\SD\500i20\R\R_1.txt

name of file for Focal grp. scores = d:\SD\500i20\N10\F_1.txt

minimum no. of examinees per matching score cell = 2

number of runs for this data set = 20

number of examinees in Reference Group = 500

number of examinees in Focal group = 500

Examinee Test Score Summary Statistics

Reference Group: Mean = 61.03

Standard deviation = 5.33

Focal Group: Mean = 60.19

Standard deviation = 5.58

Standardized Score Difference = 0.15

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:	2.727	3.176	3.107	3.339	2.536	3.003	3.068	3.125	3.700	2.894
r:	-.072	0.329	0.245	0.296	0.335	-.040	0.229	0.231	-.050	0.020

#:	11	12	13	14	15	16	17	18	19	20
m:	2.983	2.924	3.920	3.244	3.174	2.756	2.671	2.704	3.020	2.536
r:	0.291	0.275	-.037	0.037	0.359	0.338	0.331	0.350	0.358	0.332

p-value notation:

R denotes p-value for test of DIF/DBF against Ref. group

F denotes p-value for test of DIF/DBF against Foc. group

E denotes p-value for test of DIF/DBF against either the Ref. or Foc. group.

NOTES:

MS/SSD = Matching Subtest Standardized Score Difference.
Standardized difference in mean observed scores between Reference group and Focal group on the matching subtest.

p-elim = proportion of Reference (R) and Focal (F) groups eliminated (not used) in SIBTEST calculations.

Positive Beta estimate indicates DIF/DBF favoring Ref. grp.
Negative Beta estimate indicates DIF/DBF favoring Foc. grp.

FLAG = error flag indicator. FLAG=0 indicates a normal successful completion of a SIBTEST run. All other values of FLAG come with short error messages.

SIBTEST-pooled weighting

Run no.	Suspect Subtest Item Numbers	Beta estimate	standard error	p-value	p-elim R	p-elim F	MS SSD	FLAG
1	1	-0.177	0.108	0.101 E	.02	.02	0.13	0
2	2	0.078	0.096	0.413 E	.01	.01	0.15	0
3	3	-0.047	0.082	0.565 E	.01	.02	0.17	0
4	4	0.004	0.086	0.958 E	.01	.03	0.16	0
5	5	0.050	0.074	0.500 E	.01	.02	0.16	0
6	6	-0.162	0.105	0.125 E	.02	.02	0.14	0
7	7	-0.019	0.087	0.826 E	.03	.02	0.17	0
8	8	-0.203	0.092	0.028 E	.01	.03	0.19	0
9	9	-0.270	0.106	0.011 E	.01	.02	0.15	0
10	10	-0.091	0.106	0.388 E	.01	.02	0.14	0
11	11	-0.048	0.094	0.610 E	.02	.04	0.17	0
12	12	0.144	0.088	0.102 E	.01	.02	0.13	0
13	13	0.081	0.091	0.371 E	.02	.02	0.13	0

14	14		-0.025	0.100	0.806	E	.01	.02	0.13	0
15	15		0.129	0.081	0.112	E	.01	.02	0.15	0
16	16		0.027	0.077	0.728	E	.01	.02	0.16	0
17	17		0.077	0.097	0.428	E	.02	.04	0.16	0
18	18		0.000	0.083	0.997	E	.01	.02	0.17	0
19	19		0.201	0.078	0.010	E	.01	.02	0.13	0
20	20		0.296	0.087	0.001	E	.01	.02	0.11	0

Program execution is completed.

Your output is stored on the file: d:\win1\Poly\500\I20\N10\01.TXT

ตัวอย่างคำสั่งที่วิเคราะห์ด้วยโปรแกรม MULTILOG

```
MULTILOG for Windows 7.00.2327.2
Created on: 13 March 2011, 15:50:03
>PROBLEM RANDOM,
         INDIVIDUAL,
         DATA = 'D:\LDFA\1000\I20\N10\N1000I20_0119.txt',
         NITEMS = 19,
         NGROUPS = 2,
         NEXAMINEES = 2000;
>TEST ALL,
      GRADED,
      NC = (5(0)19);
>EQUAL AJ,
 ITEM=(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18),
 WITH=(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18);
>END ;
5
12345
1111111111111111111
2222222222222222222
3333333333333333333
4444444444444444444
5555555555555555555
(I1,19A1)
```

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

MULTILOG for Windows 7.00.2327.2

Created on: 13 March 2011, 15:50:03

>PROBLEM RANDOM,

INDIVIDUAL,

DATA = 'D:\LDFA\1000\I20\N10\N1000I20_0119.txt',

NITEMS = 19,

NGROUPS = 2,

NEXAMINEES = 2000;

DATA FILE NAME IS

D:\LDFA\1000\I20\N10\N1000I20_0119.TXT

TYPE OF INPUT:

INDIVIDUAL RESPONSE VECTORS

>TEST ALL,

GRADED,

NC = (5(0)19);

>EQUAL AJ,

ITEM=(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19),

WITH=(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19);

NUMBER OF CODES 5

12345

VECTOR OF CATEGORIES FOR CODE=1

1111111111111111111

VECTOR OF CATEGORIES FOR CODE=2

2222222222222222222

VECTOR OF CATEGORIES FOR CODE=3

3333333333333333333

VECTOR OF CATEGORIES FOR CODE=4

4444444444444444444

VECTOR OF CATEGORIES FOR CODE=5

5555555555555555555

(I1,19A1)

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

MULTILOG for Windows 7.00.2327.2

Created on: 13 March 2011, 15:50:03

DATA PARAMETERS:

NUMBER OF LINES IN THE DATA FILE: 2000

NUMBER OF CATEGORICAL-RESPONSE ITEMS: 19

NUMBER OF CONTINUOUS-RESPONSE ITEMS, AND/OR GROUPS: 2

TOTAL NUMBER OF "ITEMS" (INCLUDING GROUPS): 21

NUMBER OF CHARACTERS IN ID FIELDS: 0

MAXIMUM NUMBER OF RESPONSE-CODES FOR ANY ITEM: 5

THE MISSING VALUE CODE FOR CONTINUOUS DATA: 9.0000

THE DATA WILL BE STORED IN MEMORY

ESTIMATION PARAMETERS:

THE ITEMS WILL BE CALIBRATED--

BY MARGINAL MAXIMUM LIKELIHOOD ESTIMATION

MAXIMUM NUMBER OF EM CYCLES PERMITTED: 25

NUMBER OF PARAMETER-SEGMENTS USED IS: 20

NUMBER OF FREE PARAMETERS IS: 96

MAXIMUM NUMBER OF M-STEP ITERATIONS IS 4 TIMES

THE NUMBER OF PARAMETERS IN THE SEGMENT

THE M-STEP CONVERGENCE CRITERION IS: 0.000100

THE EM-CYCLE CONVERGENCE CRITERION IS: 0.001000

THE RK CONTROL PARAMETER (FOR THE M-STEPS) IS: 0.9000

THE RM CONTROL PARAMETER (FOR THE M-STEPS) IS: 1.0000

THE MAXIMUM ACCELERATION PERMITTED IS: 0.0000

THETA-GROUP LOCATIONS WILL REMAIN UNCHANGED

QUADRATURE POINTS FOR MML,

AT THETA:

-4.500

-4.000

-3.500

-3.000

-2.500

-2.000

-1.500

-1.000

-0.500

0.000

0.500

1.000

1.500

2.000

2.500

3.000

3.500

4.000

4.500

MULTILOG for Windows 7.00.2327.2

READING DATA...

KEY-

CODE CATEGORY

1 1111111111111111111

2 2222222222222222222

3 3333333333333333333

4 4444444444444444444

5 5555555555555555555

FORMAT FOR DATA-

(I1,19A1)

FIRST OBSERVATION AS READ-

ITEMS 1335325311225443414

NORML 0.000 9.000

FINISHED CYCLE 25

MAXIMUM INTERCYCLE PARAMETER CHANGE= 0.12323 P(47)

ITEM SUMMARY

MULTILOG for Windows 7.00.2327.2

ITEM 1: 5 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 1 0.10 (0.08)

B(1) 2 -2.66 (2.03)

B(2) 3 -1.01 (1.01)

B(3) 4 0.88 (0.87)

B(4) 5 5.74 (4.08)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.003 0.003 0.003 0.003 0.003 0.003 0.003 0.003

-1.4 - 0.0 0.003 0.003 0.003 0.003 0.003 0.003 0.003 0.003

0.2 - 1.6 0.003 0.003 0.003 0.003 0.003 0.003 0.003 0.003

1.8 - 3.0 0.003 0.003 0.003 0.003 0.003 0.003 0.003

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2 3 4 5

OBS. FREQ. 917 90 100 241 652

OBS. PROP. 0.4585 0.0450 0.0500 0.1205 0.3260

GROUP 1:

EXP. PROP. 0.4306 0.0423 0.0487 0.1215 0.3569

GROUP 2:

EXP. PROP. 0.4316 0.0423 0.0487 0.1214 0.3560

ITEM 2: 5 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 6 1.43 (0.08)

B(1) 7 -1.31 (0.08)

B(2) 8 -0.47 (0.06)

B(3) 9 0.21 (0.05)

B(4) 10 1.04 (0.07)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.154 0.194 0.241 0.295 0.353 0.412 0.469 0.520

-1.4 - 0.0 0.562 0.594 0.616 0.631 0.640 0.646 0.648 0.647

0.2 - 1.6 0.644 0.638 0.627 0.610 0.584 0.549 0.503 0.450

1.8 - 3.0 0.392 0.332 0.275 0.224 0.179 0.141 0.110

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2 3 4 5

OBS. FREQ. 352 367 359 416 506

OBS. PROP. 0.1760 0.1835 0.1795 0.2080 0.2530

GROUP 1:

EXP. PROP. 0.1939 0.1779 0.1712 0.1965 0.2606

GROUP 2:

EXP. PROP. 0.2010 0.1807 0.1717 0.1944 0.2523

ITEM 3: 5 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 11 0.89 (0.07)

B(1) 12 -2.25 (0.20)

B(2) 13 -0.76 (0.10)

B(3) 14 0.51 (0.08)

B(4) 15 1.96 (0.16)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.183 0.196 0.206 0.216 0.224 0.230 0.235 0.239

-1.4 - 0.0 0.242 0.244 0.246 0.247 0.248 0.249 0.249 0.249

0.2 - 1.6 0.249 0.248 0.247 0.246 0.244 0.241 0.238 0.234

1.8 - 3.0 0.228 0.221 0.212 0.202 0.191 0.178 0.164

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2 3 4 5

OBS. FREQ. 271 420 487 461 361

OBS. PROP. 0.1355 0.2100 0.2435 0.2305 0.1805

GROUP 1:

EXP. PROP. 0.1453 0.2077 0.2353 0.2258 0.1858

GROUP 2:

EXP. PROP. 0.1494 0.2108 0.2356 0.2232 0.1810

ITEM 4: 5 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 16 0.96 (0.07)

B(1) 17 -2.15 (0.17)

B(2) 18 -0.95 (0.10)

B(3) 19 0.14 (0.07)

B(4) 20 1.30 (0.11)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.204 0.221 0.236 0.250 0.261 0.271 0.278 0.284

-1.4 - 0.0 0.288 0.291 0.293 0.294 0.295 0.295 0.295 0.294

0.2 - 1.6 0.293 0.291 0.287 0.283 0.277 0.269 0.259 0.247

1.8 - 3.0 0.232 0.216 0.199 0.181 0.163 0.145 0.127

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2 3 4 5

OBS. FREQ. 259 336 433 450 522

OBS. PROP. 0.1295 0.1680 0.2165 0.2250 0.2610

GROUP 1:

EXP. PROP. 0.1407 0.1692 0.2101 0.2153 0.2647

GROUP 2:

EXP. PROP. 0.1449 0.1721 0.2111 0.2136 0.2582

ITEM 5: 5 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 21 1.47 (0.08)

B(1) 22 -0.91 (0.07)

B(2) 23 0.01 (0.05)

B(3) 24 1.27 (0.07)

B(4) 25 2.34 (0.13)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.091 0.119 0.153 0.195 0.245 0.302 0.364 0.429

-1.4 - 0.0 0.491 0.545 0.590 0.622 0.642 0.654 0.658 0.658

0.2 - 1.6 0.654 0.650 0.648 0.648 0.650 0.653 0.654 0.652

1.8 - 3.0 0.647 0.637 0.619 0.591 0.550 0.499 0.439

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2 3 4 5

OBS. FREQ. 502 478 619 277 124

OBS. PROP. 0.2510 0.2390 0.3095 0.1385 0.0620

GROUP 1:

EXP. PROP. 0.2679 0.2248 0.2957 0.1406 0.0710

GROUP 2:

EXP. PROP. 0.2767 0.2267 0.2925 0.1364 0.0677

ITEM 6: 5 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 26 0.12 (0.06)

B(1) 27 -5.87 (3.90)

B(2) 28 -1.79 (1.43)

B(3) 29 1.06 (0.81)

B(4) 30 3.88 (2.57)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.004 0.004 0.004 0.004 0.004 0.004 0.004 0.004

-1.4 - 0.0 0.004 0.004 0.004 0.004 0.004 0.004 0.004 0.004

0.2 - 1.6 0.004 0.004 0.004 0.004 0.004 0.004 0.004 0.004

1.8 - 3.0 0.004 0.004 0.004 0.004 0.004 0.004 0.004

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2 3 4 5

OBS. FREQ. 615 246 184 173 782

OBS. PROP. 0.3075 0.1230 0.0920 0.0865 0.3910

GROUP 1:

EXP. PROP. 0.3356 0.1117 0.0821 0.0796 0.3910

GROUP 2:

EXP. PROP. 0.3366 0.1119 0.0821 0.0795 0.3898

ITEM 7: 5 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 31 0.27 (0.07)

B(1) 32 -5.07 (1.18)

B(2) 33 -1.35 (0.32)

B(3) 34 1.67 (0.53)

B(4) 35 5.30 (1.17)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.022 0.022 0.022 0.022 0.023 0.023 0.023 0.023

-1.4 - 0.0 0.023 0.023 0.023 0.023 0.023 0.023 0.023 0.023

0.2 - 1.6 0.023 0.023 0.023 0.023 0.023 0.023 0.023 0.023

1.8 - 3.0 0.023 0.023 0.023 0.023 0.022 0.022 0.022

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2 3 4 5

OBS. FREQ. 406 413 396 391 394

OBS. PROP. 0.2030 0.2065 0.1980 0.1955 0.1970

GROUP 1:

EXP. PROP. 0.2064 0.2033 0.1957 0.1946 0.2001

GROUP 2:

EXP. PROP. 0.2082 0.2041 0.1956 0.1938 0.1983

ITEM 8: 5 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 36 0.23 (0.06)

B(1) 37 -6.08 (1.57)

B(2) 38 -1.67 (0.46)

B(3) 39 1.39 (0.44)

B(4) 40 5.60 (1.38)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.017 0.017 0.017 0.017 0.017 0.017 0.017 0.017

-1.4 - 0.0 0.017 0.017 0.017 0.017 0.017 0.017 0.017 0.017

0.2 - 1.6 0.017 0.017 0.017 0.017 0.017 0.017 0.017 0.017

1.8 - 3.0 0.017 0.017 0.017 0.017 0.017 0.017 0.017

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2 3 4 5

OBS. FREQ. 382 423 354 415 426

OBS. PROP. 0.1910 0.2115 0.1770 0.2075 0.2130

GROUP 1:

EXP. PROP. 0.1971 0.2064 0.1730 0.2050 0.2184

GROUP 2:

EXP. PROP. 0.1986 0.2072 0.1731 0.2044 0.2169

ITEM 9: 5 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 41 0.21 (0.06)

B(1) 42 -4.33 (1.44)

B(2) 43 -2.49 (0.89)

B(3) 44 -1.46 (0.59)

B(4) 45 -0.52 (0.39)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.013 0.013 0.013 0.013 0.013 0.013 0.013 0.013

-1.4 - 0.0 0.013 0.013 0.013 0.013 0.013 0.013 0.013 0.013

0.2 - 1.6 0.012 0.012 0.012 0.012 0.012 0.012 0.012 0.012

1.8 - 3.0 0.011 0.011 0.011 0.011 0.011 0.010 0.010

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2 3 4 5

OBS. FREQ. 408 161 109 107 1215

OBS. PROP. 0.2040 0.0805 0.0545 0.0535 0.6075

GROUP 1:

EXP. PROP. 0.2846 0.0848 0.0522 0.0487 0.5297

GROUP 2:

EXP. PROP. 0.2864 0.0851 0.0523 0.0488 0.5275

ITEM 10: 5 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 46 0.10 (0.07)

B(1) 47 -5.78 (3.37)

B(2) 48 -1.47 (0.98)

B(3) 49 1.21 (0.89)

B(4) 50 6.22 (3.52)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.003 0.003 0.003 0.003 0.003 0.003 0.003 0.003

-1.4 - 0.0 0.003 0.003 0.003 0.003 0.003 0.003 0.003 0.003

0.2 - 1.6 0.003 0.003 0.003 0.003 0.003 0.003 0.003 0.003

1.8 - 3.0 0.003 0.003 0.003 0.003 0.003 0.003 0.003

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2 3 4 5

OBS. FREQ. 700 229 146 256 669

OBS. PROP. 0.3500 0.1145 0.0730 0.1280 0.3345

GROUP 1:

EXP. PROP. 0.3582 0.1040 0.0669 0.1208 0.3500

GROUP 2:

EXP. PROP. 0.3592 0.1041 0.0669 0.1207 0.3491

ITEM 11: 5 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 51 0.25 (0.08)

B(1) 52 -5.65 (1.43)

B(2) 53 -1.76 (0.49)

B(3) 54 1.37 (0.41)

B(4) 55 5.27 (1.29)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.020 0.020 0.020 0.020 0.020 0.020 0.020 0.020

-1.4 - 0.0 0.020 0.020 0.020 0.020 0.020 0.020 0.020 0.020

0.2 - 1.6 0.020 0.020 0.020 0.020 0.020 0.020 0.020 0.020

1.8 - 3.0 0.020 0.020 0.020 0.020 0.020 0.020 0.019

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2 3 4 5

OBS. FREQ. 384 396 388 409 423

OBS. PROP. 0.1920 0.1980 0.1940 0.2045 0.2115

GROUP 1:

EXP. PROP. 0.1969 0.1940 0.1905 0.2033 0.2153

GROUP 2:

EXP. PROP. 0.1985 0.1948 0.1906 0.2025 0.2136

ITEM 12: 5 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 56 0.68 (0.07)

B(1) 57 -2.20 (0.25)

B(2) 58 -0.50 (0.11)

B(3) 59 1.03 (0.14)

B(4) 60 2.57 (0.27)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.112 0.117 0.122 0.126 0.130 0.133 0.136 0.138

-1.4 - 0.0 0.140 0.141 0.143 0.144 0.144 0.145 0.145 0.145

0.2 - 1.6 0.145 0.145 0.145 0.145 0.145 0.144 0.143 0.142

1.8 - 3.0 0.140 0.138 0.136 0.133 0.130 0.126 0.122

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2 3 4 5

OBS. FREQ. 386 442 471 363 338

OBS. PROP. 0.1930 0.2210 0.2355 0.1815 0.1690

GROUP 1:

EXP. PROP. 0.1999 0.2177 0.2302 0.1798 0.1724

GROUP 2:

EXP. PROP. 0.2040 0.2196 0.2298 0.1777 0.1688

ITEM 13: 5 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 61 0.25 (0.07)

B(1) 62 -6.66 (1.99)

B(2) 63 -4.39 (1.35)

B(3) 64 -3.01 (0.95)

B(4) 65 -0.91 (0.40)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.019 0.019 0.019 0.018 0.018 0.018 0.018 0.018

-1.4 - 0.0 0.018 0.018 0.018 0.017 0.017 0.017 0.017 0.017

0.2 - 1.6 0.016 0.016 0.016 0.016 0.015 0.015 0.015 0.015

1.8 - 3.0 0.014 0.014 0.014 0.013 0.013 0.013 0.013

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2 3 4 5

OBS. FREQ. 242 173 143 258 1184

OBS. PROP. 0.1210 0.0865 0.0715 0.1290 0.5920

GROUP 1:

EXP. PROP. 0.1633 0.0904 0.0689 0.1201 0.5573

GROUP 2:

EXP. PROP. 0.1647 0.0909 0.0691 0.1204 0.5549

ITEM 14: 5 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 66 0.18 (0.06)

B(1) 67 -6.02 (2.93)

B(2) 68 -2.19 (1.24)

B(3) 69 0.21 (0.94)

B(4) 70 3.32 (1.65)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.010 0.010 0.010 0.010 0.010 0.010 0.010 0.010

-1.4 - 0.0 0.010 0.010 0.010 0.010 0.010 0.010 0.010 0.010

0.2 - 1.6 0.010 0.010 0.009 0.009 0.009 0.009 0.009 0.009

1.8 - 3.0 0.009 0.009 0.009 0.009 0.009 0.009 0.009

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2 3 4 5

OBS. FREQ. 446 309 227 287 731

OBS. PROP. 0.2230 0.1545 0.1135 0.1435 0.3655

GROUP 1:

EXP. PROP. 0.2581 0.1460 0.1033 0.1316 0.3610

GROUP 2:

EXP. PROP. 0.2594 0.1463 0.1034 0.1315 0.3594

ITEM 15: 5 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 71 1.58 (0.08)

B(1) 72 -1.74 (0.10)

B(2) 73 -0.51 (0.05)

B(3) 74 0.40 (0.04)

B(4) 75 1.17 (0.06)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.263 0.332 0.407 0.485 0.559 0.623 0.670 0.700

-1.4 - 0.0 0.717 0.727 0.735 0.744 0.754 0.762 0.768 0.772

0.2 - 1.6 0.775 0.775 0.770 0.758 0.735 0.698 0.645 0.578

1.8 - 3.0 0.501 0.420 0.343 0.273 0.213 0.164 0.124

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2 3 4 5

OBS. FREQ. 194 488 520 375 423

OBS. PROP. 0.0970 0.2440 0.2600 0.1875 0.2115

GROUP 1:

EXP. PROP. 0.1165 0.2368 0.2437 0.1801 0.2229

GROUP 2:

EXP. PROP. 0.1217 0.2418 0.2441 0.1775 0.2149

ITEM 16: 5 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 76 1.56 (0.08)

B(1) 77 -1.15 (0.07)

B(2) 78 -0.17 (0.05)

B(3) 79 0.96 (0.05)

B(4) 80 1.87 (0.10)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.123 0.161 0.209 0.267 0.334 0.408 0.485 0.558

-1.4 - 0.0 0.620 0.669 0.702 0.723 0.734 0.739 0.740 0.739

0.2 - 1.6 0.737 0.736 0.738 0.741 0.743 0.742 0.735 0.720

1.8 - 3.0 0.694 0.653 0.597 0.529 0.454 0.377 0.306

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2 3 4 5

OBS. FREQ. 378 489 623 313 197

OBS. PROP. 0.1890 0.2445 0.3115 0.1565 0.0985

GROUP 1:

EXP. PROP. 0.2109 0.2323 0.2893 0.1561 0.1115

GROUP 2:

EXP. PROP. 0.2187 0.2353 0.2873 0.1521 0.1066

ITEM 17: 5 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 81 1.97 (0.10)

B(1) 82 -0.57 (0.04)

B(2) 83 0.18 (0.04)

B(3) 84 0.74 (0.04)

B(4) 85 1.32 (0.06)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.032 0.047 0.069 0.101 0.146 0.208 0.292 0.401

-1.4 - 0.0 0.535 0.685 0.838 0.973 1.075 1.141 1.179 1.202

0.2 - 1.6 1.216 1.224 1.225 1.217 1.194 1.147 1.067 0.950

1.8 - 3.0 0.805 0.649 0.501 0.373 0.270 0.191 0.133

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2 3 4 5

OBS. FREQ. 608 480 339 269 304

OBS. PROP. 0.3040 0.2400 0.1695 0.1345 0.1520

GROUP 1:

EXP. PROP. 0.3220 0.2190 0.1596 0.1312 0.1682

GROUP 2:

EXP. PROP. 0.3330 0.2201 0.1580 0.1281 0.1607

ITEM 18: 5 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 86 1.99 (0.09)

B(1) 87 -1.02 (0.05)

B(2) 88 0.20 (0.04)

B(3) 89 0.89 (0.04)

B(4) 90 1.52 (0.07)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.074 0.108 0.157 0.224 0.315 0.432 0.573 0.727

-1.4 - 0.0 0.875 0.992 1.062 1.087 1.087 1.089 1.110 1.147

0.2 - 1.6 1.184 1.212 1.229 1.235 1.231 1.210 1.165 1.084

1.8 - 3.0 0.964 0.814 0.653 0.501 0.371 0.266 0.188

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2 3 4 5

OBS. FREQ. 374 731 405 255 235

OBS. PROP. 0.1870 0.3655 0.2025 0.1275 0.1175

GROUP 1:

EXP. PROP. 0.2120 0.3361 0.1904 0.1278 0.1335

GROUP 2:

EXP. PROP. 0.2210 0.3394 0.1881 0.1244 0.1272

ITEM 19: 5 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 91 2.13 (0.10)

B(1) 92 -1.09 (0.05)

B(2) 93 -0.39 (0.04)

B(3) 94 0.46 (0.03)

B(4) 95 1.61 (0.07)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.075 0.113 0.168 0.248 0.357 0.501 0.678 0.872

-1.4 - 0.0 1.059 1.211 1.310 1.362 1.382 1.383 1.372 1.360

0.2 - 1.6 1.350 1.334 1.303 1.264 1.239 1.237 1.241 1.214

1.8 - 3.0 1.125 0.975 0.790 0.604 0.441 0.311 0.214

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2 3 4 5

OBS. FREQ. 325 371 568 544 192

OBS. PROP. 0.1625 0.1855 0.2840 0.2720 0.0960

GROUP 1:

EXP. PROP. 0.1905 0.1769 0.2592 0.2606 0.1128

GROUP 2:

EXP. PROP. 0.1991 0.1803 0.2591 0.2544 0.1070

ITEM 20: GRP1, N[MU: 0.04 SIGMA: 1.00]

P(#);(S.E.): 96; (0.05) 98; (0.00)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000

-1.4 - 0.0 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000

0.2 - 1.6 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000

1.8 - 3.0 1.000 1.000 1.000 1.000 1.000 1.000 1.000

ITEM 21: GRP2, N[MU: 0.00 SIGMA: 1.00]

P(#);(S.E.): 100; (0.00) 101; (0.00)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000

-1.4 - 0.0 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000

0.2 - 1.6 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000

1.8 - 3.0 1.000 1.000 1.000 1.000 1.000 1.000 1.000

TOTAL TEST INFORMATION

FOR GROUP 1:

@THETA: INFORMATION:

-3.0 - -1.6 2.422 2.718 3.080 3.518 4.036 4.632 5.291 5.978

-1.4 - 0.0 6.639 7.210 7.646 7.937 8.112 8.212 8.274 8.322

0.2 - 1.6 8.355 8.361 8.327 8.254 8.148 7.996 7.755 7.375

1.8 - 3.0 6.833 6.161 5.430 4.714 4.062 3.498 3.026

@THETA: POSTERIOR STANDARD DEVIATION:

-3.0 - -1.6 0.643 0.607 0.570 0.533 0.498 0.465 0.435 0.409

-1.4 - 0.0 0.388 0.372 0.362 0.355 0.351 0.349 0.348 0.347

0.2 - 1.6 0.346 0.346 0.347 0.348 0.350 0.354 0.359 0.368

1.8 - 3.0 0.383 0.403 0.429 0.461 0.496 0.535 0.575

MARGINAL RELIABILITY: 0.8678

TOTAL TEST INFORMATION

FOR GROUP 2:

@THETA: INFORMATION:

-3.0 - -1.6 2.422 2.718 3.080 3.518 4.036 4.632 5.291 5.978

-1.4 - 0.0 6.639 7.210 7.646 7.937 8.112 8.212 8.274 8.322

0.2 - 1.6 8.355 8.361 8.327 8.254 8.148 7.996 7.755 7.375

1.8 - 3.0 6.833 6.161 5.430 4.714 4.062 3.498 3.026

@THETA: POSTERIOR STANDARD DEVIATION:

-3.0 - -1.6 0.643 0.607 0.570 0.533 0.498 0.465 0.435 0.409

-1.4 - 0.0 0.388 0.372 0.362 0.355 0.351 0.349 0.348 0.347

0.2 - 1.6 0.346 0.346 0.347 0.348 0.350 0.354 0.359 0.368

1.8 - 3.0 0.383 0.403 0.429 0.461 0.496 0.535 0.575

MARGINAL RELIABILITY: 0.8675

GROUP 1

NEGATIVE TWICE THE LOGLIKELIHOOD= 41649.5

(CHI-SQUARE FOR SEVERAL TIMES MORE EXAMINEES THAN CELLS)

GROUP 2

NEGATIVE TWICE THE LOGLIKELIHOOD= 41601.7

(CHI-SQUARE FOR SEVERAL TIMES MORE EXAMINEES THAN CELLS)

TOTAL, NEGATIVE TWICE THE LOGLIKELIHOOD, ALL GROUPS= 83251.2

NORMAL PROGRAM TERMINATION

START DATE: 03-19-2011

START TIME: 10:46:04

END TIME: 10:46:05

ตัวอย่างคำสั่งที่วิเคราะห์ด้วยโปรแกรม SPSS

```
GET FILE='F:\SD\250i20\N10\01.sav'.
PLUM
  item01  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total
  /PRINT = FIT SUMMARY.
PLUM
  item01  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g
  /PRINT = FIT SUMMARY.
PLUM
  item01  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g  g*total
  /PRINT = FIT SUMMARY.
PLUM
  item02  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total
  /PRINT = FIT SUMMARY.
PLUM
  item02  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
```

```
  /LOCATION = total g
  /PRINT = FIT SUMMARY.
PLUM
  item02  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g  g*total
  /PRINT = FIT SUMMARY.
PLUM
  item03  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total
  /PRINT = FIT SUMMARY.
PLUM
  item03  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g
  /PRINT = FIT SUMMARY.
PLUM
  item03  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g  g*total
  /PRINT = FIT SUMMARY.
PLUM
  item04  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
```

```
  /LINK = LOGIT
  /LOCATION = total
  /PRINT = FIT SUMMARY.
PLUM
  item04  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g
  /PRINT = FIT SUMMARY.
PLUM
  item04  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g  g*total
  /PRINT = FIT SUMMARY.
PLUM
  item05  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total
  /PRINT = FIT SUMMARY.
PLUM
  item05  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g
  /PRINT = FIT SUMMARY.
PLUM
  item05  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
```

```
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g  g*total
  /PRINT = FIT SUMMARY.
PLUM
  item06  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total
  /PRINT = FIT SUMMARY.
PLUM
  item06  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g
  /PRINT = FIT SUMMARY.
PLUM
  item06  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g  g*total
  /PRINT = FIT SUMMARY.
PLUM
  item07  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total
  /PRINT = FIT SUMMARY.
PLUM
  item07  WITH total g
```

```
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g
  /PRINT = FIT SUMMARY.
PLUM
  item07  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g  g*total
  /PRINT = FIT SUMMARY.
PLUM
  item08  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total
  /PRINT = FIT SUMMARY.
PLUM
  item08  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g
  /PRINT = FIT SUMMARY.
PLUM
  item08  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g  g*total
  /PRINT = FIT SUMMARY.
PLUM
```

```
  item09  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total
  /PRINT = FIT SUMMARY.
PLUM
  item09  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g
  /PRINT = FIT SUMMARY.
PLUM
  item09  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g  g*total
  /PRINT = FIT SUMMARY.
PLUM
  item10  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total
  /PRINT = FIT SUMMARY.
PLUM
  item10  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g
  /PRINT = FIT SUMMARY.
```

```
PLUM
  item10  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g  g*total
  /PRINT = FIT SUMMARY.
PLUM
  item11  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total
  /PRINT = FIT SUMMARY.
PLUM
  item11  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g
  /PRINT = FIT SUMMARY.
PLUM
  item11  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g  g*total
  /PRINT = FIT SUMMARY.
PLUM
  item12 WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total
```

```
  /PRINT = FIT SUMMARY.
PLUM
  item12  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g
  /PRINT = FIT SUMMARY.
PLUM
  item12  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g  g*total
  /PRINT = FIT SUMMARY.
PLUM
  item13  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total
  /PRINT = FIT SUMMARY.
PLUM
  item13  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g
  /PRINT = FIT SUMMARY.
PLUM
  item13  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
```

```
  /LOCATION = total g  g*total
  /PRINT = FIT SUMMARY.
PLUM
  item14  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total
  /PRINT = FIT SUMMARY.
PLUM
  item14  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g
  /PRINT = FIT SUMMARY.
PLUM
  item14  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g  g*total
  /PRINT = FIT SUMMARY.
PLUM
  item15  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total
  /PRINT = FIT SUMMARY.
PLUM
  item15  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
```

```
  /LINK = LOGIT
  /LOCATION = total g
  /PRINT = FIT SUMMARY.
PLUM
  item15  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g  g*total
  /PRINT = FIT SUMMARY.
PLUM
  item16  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total
  /PRINT = FIT SUMMARY.
PLUM
  item16  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g
  /PRINT = FIT SUMMARY.
PLUM
  item16  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g  g*total
  /PRINT = FIT SUMMARY.
PLUM
  item17  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
```

```
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total
  /PRINT = FIT SUMMARY.
PLUM
  item17  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g
  /PRINT = FIT SUMMARY.
PLUM
  item17  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g  g*total
  /PRINT = FIT SUMMARY.
PLUM
  item18  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total
  /PRINT = FIT SUMMARY.
PLUM
  item18  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g
  /PRINT = FIT SUMMARY.
PLUM
  item18  WITH total g
```

```
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g  g*total
  /PRINT = FIT SUMMARY.
PLUM
  item19  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total
  /PRINT = FIT SUMMARY.
PLUM
  item19  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g
  /PRINT = FIT SUMMARY.
PLUM
  item19  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g  g*total
  /PRINT = FIT SUMMARY.
PLUM
  item20  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total
  /PRINT = FIT SUMMARY.
PLUM
```

```
  item20  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g
  /PRINT = FIT SUMMARY.
PLUM
  item20  WITH total g
  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /LINK = LOGIT
  /LOCATION = total g  g*total
  /PRINT = FIT SUMMARY.
```

ประวัติย่อผู้วิจัย

ประวัติย่อผู้วิจัย

ชื่อ ชื่อสกุล	นายสุชาติ สิริมีนนันท์
วัน เดือน ปีเกิด	20 สิงหาคม 2501
สถานที่เกิด	จังหวัดพระนคร
สถานที่อยู่ปัจจุบัน	105/675 หมู่ 1 ถ.นวมินทร์ แขวงคลองกุ่ม เขตบึงกุ่ม กรุงเทพฯ
ตำแหน่งหน้าที่การงาน	ครู คศ. 3
สถานที่ทำงานปัจจุบัน	โรงเรียนบางกะปิ สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 2

ประวัติการศึกษา

พ.ศ. 2523	กศ.บ. (คณิตศาสตร์) จากมหาวิทยาลัยศรีนครินทรวิโรฒ บางเขน
พ.ศ. 2542	กศ.ม. (การวัดผลการศึกษา) จากมหาวิทยาลัยศรีนครินทรวิโรฒ
พ.ศ. 2554	กศ.ด. (การทดสอบและวัดผลการศึกษา) จากมหาวิทยาลัยศรีนครินทรวิโรฒ