

ประสิทธิภาพของวิธีวิเคราะห์องค์ประกอบจำกัดและวิธีถดถอยโลจิสติก
ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเอกรูป
*Efficiency of Restricted Factor Analysis and Logistic Regression
Procedures in Detecting Uniform Differential Item Functioning*

อาภรณ์ นรสิงห์*

nga_ap@windowslive.com

กิตติมา พฤกษณ**

ศุภวรรณ พรหมเพรา***

พงศ์เทพ จิระโร****

บทคัดย่อ

การวิจัยครั้งนี้ มีวัตถุประสงค์เพื่อศึกษาอำนาจการตรวจสอบ และความสามารถในการควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเอกรูประหว่างวิธีการวิเคราะห์องค์ประกอบจำกัด (RFA) กับวิธีถดถอยโลจิสติก (LR) โดยศึกษาจากข้อมูลที่จำลองขึ้นภายใต้เงื่อนไข 32 เงื่อนไข คือ อัตราส่วนระหว่างกลุ่มอ้างอิงกับกลุ่มเปรียบเทียบ 4 อัตราส่วน (1:1, 0.9:1, 0.75:1, 0.5:1) ค่าความยากของข้อสอบ 2 ระดับ (ปานกลาง สูง) ค่าอำนาจจำแนกของข้อสอบ 2 ระดับ (ปานกลาง สูง) และขนาดการทำหน้าที่ต่างกันของข้อสอบ 2 ขนาด (มาก น้อย) ที่ระดับนัยสำคัญทางสถิติ 3 ระดับ ได้แก่ .10, .05 และ .01 ผลการวิจัยพบว่าวิธี RFA มีอำนาจการตรวจสอบสูงกว่า วิธี LR ในเกือบทุกกรณี ทั้งสองวิธีสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ดี ที่ระดับนัยสำคัญทางสถิติเป็น .05 และ .10 เมื่อข้อสอบมีค่าความยากอยู่ในระดับปานกลาง และในภาพรวมวิธี RFA มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปได้ดีกว่าวิธี LR โดยเฉพาะอย่างยิ่งเมื่อค่าความยากปานกลาง ค่าอำนาจจำแนกทั้ง 2 ระดับ และขนาดการทำหน้าที่ต่างกันของข้อสอบทั้ง 2 ขนาด

คำสำคัญ : การทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป วิธีถดถอยโลจิสติก วิธีวิเคราะห์องค์ประกอบจำกัด

*นิสิตระดับมหาบัณฑิต สาขาวิชาวิจัยและประเมินผลการศึกษา คณะครุศาสตร์ มหาวิทยาลัยราชภัฏกาญจนบุรี

**อาจารย์ ดร. คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏกาญจนบุรี

***ผู้ช่วยศาสตราจารย์ ดร. คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครศรีธรรมราช

****นางาตรี ดร. ภาควิชาวิจัยและจิตวิทยาประยุกต์ คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา

Abstract

The purpose of this research was to study power of the test and type I error rate controlling ability of restricted factor analysis (RFA) and logistic regression (LR) in detecting uniform differential item functioning (DIF). The data used in this study were simulated in 32 conditions namely 4 ratios between the reference and the focal group (1:1, 0.9:1, 0.75:1, 0.5:1), 2 levels of difficulty (moderate and high) , 2 levels of discrimination (moderate and high) and 2 sizes of DIF (large and small), with significance levels at .10, .05 and .01. The results revealed that, in almost cases RFA showed higher power of the test than LR. Both procedures could control the type I error rate at .05 and .10 significance levels in moderate difficulty. In perspective, the RFA showed more efficiency in detecting uniform DIF than LR especially at moderate difficulty, both levels of discrimination and both sizes of DIF.

Keywords : uniform differential item functioning, logistic regression, restricted factor analysis

บทนำ

การทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning : DIF) หมายถึง ข้อสอบทำหน้าที่ต่างกันภายใต้เงื่อนไขของผู้เข้าสอบที่มีความสามารถเท่ากัน แต่มาจากกลุ่มผู้เข้าสอบย่อยที่มีลักษณะต่างกัน มีความน่าจะเป็นในการตอบข้อสอบข้อนั้นไม่เท่ากัน (เสรี ชัดแจ้ง, 2540) โดยข้อสอบที่ทำหน้าที่ต่างกัน แบ่งออกเป็น 2 ประเภท คือข้อสอบทำหน้าที่ต่างกันแบบเอกรูป (Uniform DIF) และข้อสอบทำหน้าที่ต่างกันแบบอนเอกรูป (non-uniform DIF) (Mellenbergh, 1982, อ้างถึงใน ปิยะทิพย์ ดินวรร, 2549) ข้อสอบทำหน้าที่ต่างกันแบบเอกรูปเกิดขึ้นเมื่อไม่มีปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้เข้าสอบกับกลุ่มของผู้สอบ ส่วนข้อสอบทำหน้าที่ต่างกันแบบอนเอกรูปเกิดขึ้นเมื่อการมีปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้เข้าสอบกับกลุ่มของผู้สอบ

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ใช้การเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มผู้สอบย่อย 2 กลุ่ม ที่มีความสามารถระดับเดียวกัน โดยผู้เข้าสอบกลุ่มหนึ่งเป็นตัวแทนกลุ่มหลักในประชากรเรียก

ว่า “กลุ่มอ้างอิง” (Reference group: R) ซึ่งเป็นกลุ่มพื้นฐาน ส่วนอีกกลุ่มหนึ่งเป็นตัวแทนกลุ่มรองในประชากรเรียกว่า “กลุ่มเปรียบเทียบ” (Focal group: F) ซึ่งตามปกติแล้วเป็นกลุ่มผู้สอบที่สนใจจะศึกษาการทำหน้าที่ต่างกันของข้อสอบ (Angoff, 1993, อ้างถึงใน ปิยะทิพย์ ดินวรร, 2549) ถ้าข้อสอบทำหน้าที่ต่างกันแล้วโอกาสตอบข้อสอบถูกของผู้เข้าสอบทั้งสองกลุ่มจะไม่เท่ากัน โดยคาดว่าผู้เข้าสอบกลุ่มอ้างอิงจะได้เปรียบในการตอบข้อสอบ ส่วนผู้เข้าสอบกลุ่มเปรียบเทียบจะเสียเปรียบในการตอบข้อสอบ

ปัจจัยที่ส่งผลต่อประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เช่น อัตราส่วนของกลุ่มอ้างอิงต่อกลุ่มเปรียบเทียบก็เป็นปัจจัยหนึ่งที่ส่งผลต่ออัตราตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ จากการศึกษาของ จิตติมา วรรณศรี (2539) เปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยคงจำนวนผู้สอบกลุ่มอ้างอิงและลดจำนวนผู้สอบกลุ่มเปรียบเทียบแต่ละขนาดตัวอย่าง มีอัตราส่วนระหว่างผู้สอบกลุ่มอ้างอิงกับกลุ่มเปรียบเทียบต่างกัน ดังนี้ 1:1, 1:0.9, 1:0.75 และ 1:0.5 พบว่า ที่อัตราส่วน

1:1 เป็นอัตราส่วนที่ให้ประสิทธิภาพในการตรวจสอบดีมาก ในขณะที่ เคลาเซอร์ และคณะ (Clauser, Mazor & Hambleton, 1993) ได้เสนอแนะว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบหากเพิ่มผู้สอบกลุ่มอ้างอิงขึ้นและคงจำนวนผู้สอบกลุ่มเปรียบเทียบไว้เท่าเดิม จะทำให้อำนาจการตรวจสอบสูงขึ้นและไม่ทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 สูงขึ้น

นอกจากอัตราส่วนของกลุ่มอ้างอิงต่อกลุ่มเปรียบเทียบแล้ว ค่าความยากของข้อสอบและค่าอำนาจจำแนกของข้อสอบ ก็มีผลต่อประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เช่น จากการศึกษาของ โรเจอร์และสวามินาทาน (Rogers & Swaminathan, 1993) พบว่า วิธีถดถอยโลจิสติก ตรวจสอบข้อสอบทำหน้าที่ต่างกันได้ดีเฉพาะกรณีที่ข้อสอบมีค่าความยากปานกลาง และค่าอำนาจจำแนกสูง ซึ่งสอดคล้องกับผลการศึกษาของ นารายานันและสวามินาทาน (Narayanan & Swaminathan, 1996) พบว่า เมื่อข้อสอบมีค่าอำนาจจำแนกสูงขึ้นจะมีผลทำให้อำนาจการทดสอบของวิธีการถดถอยโลจิสติกมีค่าเพิ่มขึ้น ต่อมา นิคม กิรติวรังกูร (2542) พบว่า วิธีวิเคราะห์องค์ประกอบจำกัดมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูงเมื่อข้อสอบมีค่าความยากสูง ค่าความยากปานกลาง ค่าอำนาจจำแนกสูง และค่าอำนาจจำแนกปานกลาง

เมื่อพิจารณาถึงขนาดของข้อสอบที่ทำหน้าที่ต่างกัน จากการศึกษาของ นิคม กิรติวรังกูร (2542) กำหนดให้ขนาดของการทำหน้าที่ต่างกันของข้อสอบออกเป็น 2 ระดับ คือ ข้อสอบทำหน้าที่ต่างกันมาก (0.8SD) และน้อย (0.2SD) พบว่า ในกรณีที่ข้อสอบทำหน้าที่ต่างกันมากจะมีอำนาจการตรวจสอบสูงกว่า ข้อสอบที่ทำหน้าที่ต่างก็น้อย ซึ่งข้อค้นพบนี้สอดคล้องกับการศึกษาของ ออร์ต (Oort, 1998) ได้ศึกษาการทำหน้าที่ต่างกันของข้อสอบด้วยการจำลองข้อมูล โดยกำหนดขนาดของการทำหน้าที่ต่างกันของข้อสอบออกเป็น 3 ระดับ คือ

ข้อสอบทำหน้าที่ต่างกันมาก (0.8SD) ปานกลาง (0.5SD) และน้อย (0.2SD) พบว่าเมื่อข้อสอบทำหน้าที่ต่างกันปานกลาง และข้อสอบที่ทำหน้าที่ต่างกันมาก วิธี RFA ให้ผลการตรวจสอบได้ดีกว่าวิธี IRT

สำหรับความยาวของแบบสอบพบว่าไม่มีผลกระทบต่ออำนาจการตรวจสอบในการทำหน้าที่ต่างกันของข้อสอบ (Rogers & Swaminathan, 1993) และจากงานวิจัยของนารายานัน และสวามินาทาน (Narayanan & Swaminathan, 1996) พบว่า แบบสอบขนาด 40 ข้อ เป็นแบบสอบที่มีความยาวเพียงพอสำหรับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

จากการศึกษางานวิจัยที่เกี่ยวข้อง พบว่าในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปด้วยวิธีวิเคราะห์องค์ประกอบจำกัดและวิธีถดถอยโลจิสติก เป็นวิธีที่ประหยัดและสะดวกเพราะมีโปรแกรมคอมพิวเตอร์สำเร็จรูปซึ่งง่ายในการคำนวณ นอกจากนี้ยังพบว่าวิธีวิเคราะห์องค์ประกอบจำกัดให้ผลการตรวจสอบที่ดีกว่าวิธีการตอบสนองข้อสอบ แบบ 1 พารามิเตอร์ ในขนาดตัวอย่างเพียง 200 คน (Oort, 1998) และในปี พ.ศ. 2549 ปิยะทิพย์ ดินวรร ได้ศึกษาเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบพหุมิติ ในข้อมูลจริง พบว่าในภาพรวม วิธี LR มีประสิทธิภาพสูงกว่าวิธี RFA (ปิยะทิพย์ ดินวรร, 2549) และเท่าที่ศึกษามายังไม่พบผลการเปรียบเทียบประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูประหว่างวิธีการวิเคราะห์องค์ประกอบจำกัดกับวิธีถดถอยโลจิสติก ในข้อมูลจำลองว่าวิธีใดมีประสิทธิภาพดีกว่ากัน

จากเหตุผลที่กล่าวมาแล้วข้างต้นในการศึกษาครั้งนี้ผู้วิจัยจึงสนใจที่จะเปรียบเทียบประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูประหว่างวิธีวิเคราะห์องค์ประกอบจำกัด กับวิธีถดถอยโลจิสติกจากข้อมูลจำลองโดยใช้ค่าอัตราส่วนตามแนวคิดของจิตติมา วรรณศรี (2539) และคงจำนวนกลุ่มผู้สอบกลุ่มเปรียบเทียบตามคำแนะนำของเคลาเซอร์ และ

คณะ (Clauser, Mazor & Hambleton, 1993) ได้ อัตราส่วนระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบเป็น 1:1, 0.9:1, 0.75:1, 0.5:1 ค่าอำนาจจำแนกของข้อสอบ 2 ระดับ (ปานกลาง สูง) ค่าความยากของข้อสอบ 2 ระดับ (ปานกลาง สูง) และขนาดของข้อสอบที่ทำหน้าที่ต่างกัน 2 ระดับ (มาก น้อย) ศึกษาที่ระดับนัยสำคัญทางสถิติ 3 ระดับ ได้แก่ .10, .05 และ .01 กำหนดให้แบบสอบมีความยาว 40 ข้อ และขนาดตัวอย่างกลุ่มเปรียบเทียบเท่ากับ 200 และจากงานวิจัยที่กล่าวมาได้มีการศึกษาประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยพิจารณาจากอำนาจการตรวจสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 แต่ในการศึกษาครั้งนี้ผู้วิจัยศึกษา ประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยพิจารณาจากอำนาจการตรวจสอบและความสามารถในการควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ตามเกณฑ์ของ Bradley (Bradley, J.V., 1978) ผลที่ได้จากการศึกษาครั้งนี้ใช้เป็นแนวทางในการเลือกสถิติที่ใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบให้เหมาะสมกับลักษณะของข้อสอบเพื่อจะได้นำไปใช้ในการพัฒนาแบบสอบให้มีคุณภาพต่อไป

วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาอำนาจการตรวจสอบ (power of the test) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเอกรูประหว่างวิธีการวิเคราะห์องค์ประกอบจำกัดกับวิธีถดถอยโลจิสติก เมื่อ อัตราส่วนระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ค่าความยากของข้อสอบ ค่าอำนาจจำแนกของข้อสอบ และขนาดของข้อสอบที่ทำหน้าที่ต่างกัน มีค่าต่างกัน
2. เพื่อศึกษาความสามารถในการควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 (Type I error rate) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูประหว่างวิธีการวิเคราะห์องค์ประกอบจำกัดกับวิธี

ถดถอยโลจิสติก เมื่ออัตราส่วนระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ค่าความยากของข้อสอบ ค่าอำนาจจำแนกของข้อสอบ และขนาดของข้อสอบที่ทำหน้าที่ต่างกัน มีค่าต่างกัน

สมมติฐานของการวิจัย

วิธีวิเคราะห์องค์ประกอบจำกัดและวิธีถดถอยโลจิสติก จะมีอำนาจการตรวจสอบ และมีความสามารถในการควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้แตกต่างกัน ในเงื่อนไขต่อไปนี้

1. อัตราส่วนระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบเป็น 1:1, 0.9:1, 0.75:1 และ 0.5:1
2. ค่าความยากของข้อสอบ 2 ระดับ คือ ค่าความยากสูง และ ค่าความยากปานกลาง
3. ค่าอำนาจจำแนกของข้อสอบ 2 ระดับ คือ ค่าอำนาจจำแนกสูง และค่าอำนาจจำแนกปานกลาง
4. ขนาดของข้อสอบที่ทำหน้าที่ต่างกัน 2 ระดับ คือ ข้อสอบที่ทำหน้าที่ต่างกันมาก และข้อสอบที่ทำหน้าที่ต่างกันน้อย

วิธีดำเนินการวิจัย

1. การจำลองข้อมูล

ข้อมูลที่ใช้ในการวิจัยเป็นข้อมูลจำลองด้วยโปรแกรม SPSS ผู้วิจัยจำลองข้อมูลเช่นเดียวกับการศึกษาของ ออร์ต (Oort, 1998) โดย กำหนด ความยาวข้อสอบ 40 ข้อซึ่งมีรายละเอียดดังนี้

- 1.1. ค่าความยากของข้อสอบมี 2 ระดับ คือ ข้อสอบที่มีค่าความยากสูง (0.67 ถึง 2.50) และค่าความยากปานกลาง (-0.67 ถึง 0.67) (คณิต ไช่มุกด์, 2534)
- 1.2. ค่าอำนาจจำแนกของข้อสอบแบ่งออกเป็น 2 ระดับ คือ ค่าอำนาจจำแนกสูง (0.85 ถึง 2.50) ค่าอำนาจจำแนกปานกลาง (0.55 ถึง 0.85) (คณิต ไช่มุกด์, 2534)

1.3 ขนาดของข้อสอบที่ทำหน้าที่ต่างกัน แบ่งเป็น 2 ขนาด คือ ข้อสอบที่ทำหน้าที่ต่างกันมาก (0.80 SD) และข้อสอบที่ทำหน้าที่ต่างกันน้อย (0.30 SD) (นิคม กิรติวรังกูร, 2542)

2. การสุ่มข้อสอบ

เป็นการสุ่มข้อสอบเพื่อให้ได้แบบสอบ แต่ละแบบสอบมีค่าความยาก ค่าอำนาจจำแนกและขนาดของข้อสอบที่ทำหน้าที่ต่างกันตามเงื่อนไขที่กำหนด

3. สุ่มขนาดตัวอย่าง

ในการวิจัยครั้งนี้กำหนดขนาดตัวอย่างในกลุ่มเปรียบเทียบไว้ที่ 200 และกำหนดอัตราส่วนระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบเป็น 1:1, 0.9:1, 0.75:1, 0.5:1 ดังนั้น จึงมีเงื่อนไขเกี่ยวกับอัตราส่วนของขนาดตัวอย่าง ระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ 4 เงื่อนไข คือ 200:200, 180:200, 150:200 และ 100:200

การวิเคราะห์ข้อมูล

1. วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีถดถอยโลจิสติก ด้วยโปรแกรมสำเร็จรูป SPSS for Windows และวิธีวิเคราะห์องค์ประกอบจำกัด ด้วยโปรแกรม LISREL โดยในแต่ละเงื่อนไขวิเคราะห์ซ้ำทั้งหมด 10 รอบ

2. คำนวณหาค่าอำนาจการตรวจสอบ (power of the test) เฉลี่ย และอัตราความคลาดเคลื่อนประเภท

ที่ 1 (type I error rate) เฉลี่ย ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแต่ละวิธี

3. เปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเอกรูประหว่างวิธีวิเคราะห์องค์ประกอบจำกัดกับวิธีถดถอยโลจิสติก โดยพิจารณาจากค่าอำนาจการตรวจสอบเฉลี่ย และความสามารถในการควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 เฉลี่ยพิจารณาความสามารถในการควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ตามเกณฑ์ของ Bradley คือ $50\alpha \leq p \leq 1.5\alpha$ เมื่อ p คือ ค่าความคลาดเคลื่อนประเภทที่ 1 ที่คำนวณได้ มีค่า α เท่ากับ .01, .05 และ .10 (Bradley, J.V., 1978) นั่นคือ ถ้าอัตราความคลาดเคลื่อนของวิธีใดอยู่ในช่วงดังกล่าวจะถือว่าสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ที่ระดับนัยสำคัญนั้นๆ

สรุปผลการวิจัย

1. อำนาจการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีวิเคราะห์องค์ประกอบจำกัด (RFA) และวิธีถดถอยโลจิสติก (LR) พบว่า วิธี RFA มีอำนาจในการตรวจสอบสูงกว่าวิธี LR ในเกือบทุกกรณี ยกเว้นเมื่อค่าความยากและค่าอำนาจจำแนกอยู่ในระดับปานกลาง ขนาดการทำหน้าที่ของข้อสอบต่างกันน้อยที่ระดับนัยสำคัญ .05 และ .10 วิธี LR มีอำนาจในการตรวจสอบเทียบเท่าหรือสูงกว่าวิธี RFA ดังแสดงในตารางที่ 1

ตารางที่ 1 อำนาจการตรวจสอบเฉลี่ยในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี RFA กับวิธี LR เมื่ออัตราส่วนของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ค่าความยาก ค่าอำนาจจำแนก และขนาดของข้อสอบที่ทำหน้าที่ต่างกันมีค่าต่างกัน

อัตราส่วนค่าความยาก	ค่าอำนาจจำแนก	ขนาดที่ ทำหน้าที่ ต่างกัน	วิธีตรวจสอบ					
			LR			RFA		
			$\alpha = .01$	$\alpha = .05$	$\alpha = .10$	$\alpha = .01$	$\alpha = .05$	$\alpha = .10$
1:1	สูง	มาก	0	0.025	0.050	0.125	0.200	0.250
		น้อย	0	0	0	0.175	0.200	0.225
	ปานกลาง	มาก	0	0	0	0.150	0.150	0.150
		น้อย	0	0	0	0.050	0.050	0.050
	ปานกลาง	มาก	0	0.025	0.100	0.050	0.100	0.150
		น้อย	0	0.025	0.075	0.150	0.150	0.150
	ปานกลาง	มาก	0	0.025	0.025	0.075	0.150	0.200
		น้อย	0	0.025	0.150	0.025	0.025	0.050
0.9:1	สูง	มาก	0	0.050	0.050	0	0	0.025
		น้อย	0	0.025	0.025	0	0	0.050
	ปานกลาง	มาก	0	0	0	0.300	0.300	0.300
		น้อย	0	0	0	0.075	0.075	0.075
	ปานกลาง	มาก	0	0	0.075	0.225	0.225	0.250
		น้อย	0	0.025	0.125	0.025	0.075	0.250
	ปานกลาง	มาก	0	0	0	0.250	0.250	0.300
		น้อย	0	0.025	0.075	0	0.050	0.075
0.75:1	สูง	มาก	0	0	0.025	0	0	0.050
		น้อย	0	0	0.025	0	0.050	0.125
	ปานกลาง	มาก	0	0	0	0.150	0.150	0.250
		น้อย	0	0	0	0.075	0.100	0.166
	ปานกลาง	มาก	0	0	0.050	0.250	0.250	0.300
		น้อย	0	0	0.075	0	0.050	0.125
	ปานกลาง	มาก	0	0	0	0.200	0.225	0.275
		น้อย	0	0	0.100	0.050	0.075	0.100

อัตราส่วนค่าความยาก	ค่าอำนาจจำแนก	ขนาดที่ ทำหน้าที่ ต่างกัน	วิธีตรวจสอบ					
			LR			RFA		
			$\alpha = .01$	$\alpha = .05$	$\alpha = .10$	$\alpha = .01$	$\alpha = .05$	$\alpha = .10$
0.5:1	สูง	มาก	0	0.025	0.025	0	0	0.050
		น้อย	0	0	0	0	0	0
	ปานกลาง	มาก	0	0	0	0	0.200	0.200
		น้อย	0	0	0.025	0	0.100	0.100
	ปานกลาง	มาก	0	0	0.050	0.025	0.300	0.300
		น้อย	0	0	0.050	0	0.050	0.075
		มาก	0	0	0	0.050	0.075	0.075
		น้อย	0	0	0.025	0	0	0

2. อัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีวิเคราะห์องค์ประกอบจำกัด (RFA) และวิธีถดถอยโลจิสติก (LR) พบว่าวิธี LR มีอัตราความคลาดเคลื่อนประเภทที่ 1 น้อยกว่าวิธี RFA ในเกือบทุกกรณีแต่เมื่อพิจารณาความสามารถในการควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธี RFA สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ดีกว่าวิธี LR ในเกือบทุกกรณี โดยเฉพาะอย่างยิ่งเมื่อระดับนัยสำคัญทางสถิติเป็น .05 และ .10 ดังแสดงในตารางที่ 2

ตารางที่ 2 อัตราความคลาดเคลื่อนประเภทที่ 1 เฉลี่ย ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี RFA กับวิธี LR เมื่ออัตราส่วนของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ค่าความยาก ค่าอำนาจจำแนกและขนาดของข้อสอบที่ทำหน้าที่ต่างกันมีค่าต่างกัน

อัตราส่วน	ค่าความยาก	ค่าอำนาจจำแนก	ระดับนัยสำคัญ					
			.01		.05		.10	
			LR	RFA	LR	RFA	LR	RFA
1:1	สูง	สูง	0	0.119	0.019	0.169	0.041	0.238
		ปานกลาง	0	0.076	0	0.080	0.025	0.080a
	ปานกลาง	สูง	0	0.003	0.038b	0.025b	0.106a	0.038
		ปานกลาง	0	0.034	0.019	0.081	0.078a	0.134a
0.9:1	สูง	สูง	0	0	0.009	0.003	0.041	0.263
		ปานกลาง	0	0.117	0.006	0.121	0.025	0.125a
	ปานกลาง	สูง	0	0.003	0.028b	0.031b	0.072a	0.053a
		ปานกลาง	0	0.034	0.025b	0.078	0.056a	0.103a

อัตราส่วน	ค่าความยาก	ค่าอำนาจ จำแนก	ระดับนัยสำคัญ					
			.01		.05		.10	
			LR	RFA	LR	RFA	LR	RFA
0.75:1	สูง	สูง	0	0	0.016	0.003	0.025	0.028
		ปานกลาง	0	0.053	0.003	0.059b	0.016	0.066a
	ปานกลาง	สูง	0	0.003	0.022	0.022	0.053a	0.066a
		ปานกลาง	0	0.022	0.022	0.106	0.059a	0.147a
0.5:1	สูง	สูง	0	0	0.009	0	0.025	0.025
		ปานกลาง	0	0.019	0.025b	0.084	0.084a	0.100a
	ปานกลาง	สูง	0	0.006c	0.038b	0.038b	0.088a	0.050a
		ปานกลาง	0	0.006c	0.022	0.009	0.091a	0.034

a ควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ที่ระดับนัยสำคัญทางสถิติ .10

b ควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ที่ระดับนัยสำคัญทางสถิติ .05

c ควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ที่ระดับนัยสำคัญทางสถิติ .01

จากตารางที่ 2 เมื่อพิจารณาความสามารถในการควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ตามเกณฑ์ Bradley สรุปผลได้ดังนี้

2.1 ที่ความยากของข้อสอบสูง ค่าอำนาจจำแนกสูง ทั้งวิธี RFA และวิธี LR ไม่สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ ในทุกอัตราส่วน และทุกระดับนัยสำคัญทางสถิติ

2.2 ที่ความยากของข้อสอบสูง ค่าอำนาจจำแนกปานกลาง ระดับนัยสำคัญ .05 วิธี RFA สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ที่อัตราส่วนระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบเป็น 0.75:1 ส่วนวิธี LR สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้เมื่ออัตราส่วนระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบเป็น 0.5:1 และที่ระดับนัยสำคัญ .10 วิธี RFA สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ในทุกอัตราส่วน ในขณะที่ วิธี LR สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ที่อัตราส่วนระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบเป็น 0.5:1 เท่านั้น

2.3 ที่ความยากของข้อสอบปานกลาง ค่าอำนาจจำแนกสูง ระดับนัยสำคัญ .01 วิธี RFA สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ ที่อัตราส่วนระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบเป็น 0.5:1 ส่วนวิธี LR ไม่สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ ที่ทุกอัตราส่วน เมื่อพิจารณาที่ระดับนัยสำคัญ .05 ทั้ง 2 วิธีสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ที่อัตราส่วนระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบเป็น 1:1, 0.9:1 และ 0.5:1 และที่ระดับนัยสำคัญ .10 ทั้ง 2 วิธีสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ที่อัตราส่วนระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบเป็น 0.9:1, 0.75:1 และ 0.5:1 แต่เมื่ออัตราส่วนเป็น 1:1 วิธี LR สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ส่วนวิธี RFA ไม่สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้

2.4 เมื่อความยากของข้อสอบปานกลาง ค่าอำนาจจำแนกปานกลาง ระดับนัยสำคัญ .01 วิธี RFA สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ที่อัตราส่วนระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบเป็น

0.5:1 ส่วนวิธี LR ไม่สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้เลย และที่ระดับนัยสำคัญ .05 วิธี LR สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ที่อัตราส่วน 0.9:1 ส่วนวิธี RFA ไม่สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้เลย ส่วนที่ระดับนัยสำคัญ .10 ทั้ง 2 วิธีสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ ยกเว้นที่อัตราส่วน 0.5:1 วิธี

RFA ไม่สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้

3. ประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีวิเคราะห์องค์ประกอบจำกัด (RFA) และวิธีถดถอยโลจิสติก (LR) พิจารณาจากวิธีที่มีค่าอำนาจการตรวจสอบสูงสุด และสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ ดังแสดงในตารางที่ 3

ตารางที่ 3 วิธีที่มีอำนาจการตรวจสอบสูงและวิธีที่สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่ออัตราส่วนของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ค่าความยาก ค่าอำนาจจำแนก และขนาดของข้อสอบที่ทำหน้าที่ต่างกันมีค่าต่างกัน

อัตราส่วน	ค่าความยาก	ค่าอำนาจจำแนก	ขนาดที่ ทำหน้าที่ ต่างกัน	วิธีที่มีอำนาจการตรวจสอบสูง			วิธีที่สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้		
				$\alpha = .01$	$\alpha = .05$	$\alpha = .10$	$\alpha = .01$	$\alpha = .05$	$\alpha = .10$
1:1	สูง	สูง	มาก	RFA	RFA	RFA	-	-	-
			น้อย	RFA	RFA	RFA	-	-	-
		ปานกลาง	มาก	RFA	RFA	RFA	-	-	RFA
			น้อย	RFA	RFA	RFA	-	-	RFA
	ปานกลาง	สูง	มาก	RFA	RFA	RFA	-	LR, RFA	LR
			น้อย	RFA	RFA	RFA	-	LR, RFA	LR
		ปานกลาง	มาก	RFA	RFA	RFA	-	-	LR, RFA
			น้อย	RFA	LR, RFA	LR	-	-	LR, RFA
0.9:1	สูง	สูง	มาก	-	LR	LR	-	-	-
			น้อย	-	LR	RFA	-	-	-
		ปานกลาง	มาก	RFA	RFA	RFA	-	-	rFA
			น้อย	RFA	RFA	RFA	-	-	rFA
	ปานกลาง	สูง	มาก	RFA	RFA	RFA	-	LR, RFA	LR, RFA
			น้อย	RFA	RFA	RFA	-	LR, RFA	LR, RFA
		ปานกลาง	มาก	RFA	RFA	RFA	-	LR	LR, RFA
			น้อย	-	RFA	LR, RFA	-	LR	LR, RFA

อัตราส่วน	ค่าความยาก	ค่าอำนาจ จำแนก	ขนาดที่ ทำหน้าที่ ต่างกัน	วิธีที่มีอำนาจการตรวจสอบสูง			วิธีที่สามารถควบคุมอัตราความคลาด เคลื่อนประเภทที่ 1 ได้		
				$\alpha = .01$	$\alpha = .05$	$\alpha = .10$	$\alpha = .01$	$\alpha = .05$	$\alpha = .10$
0.75:1	สูง	สูง	มาก	-	-	RFA	-	-	-
			น้อย	-	RFA	RFA	-	-	-
		ปานกลาง	มาก	RFA	RFA	RFA	-	RFA	RFA
			น้อย	RFA	RFA	RFA	-	RFA	RFA
	ปานกลาง	สูง	มาก	RFA	RFA	RFA	-	-	LR, RFA
			น้อย	-	RFA	RFA	-	-	LR, RFA
		ปานกลาง	มาก	RFA	RFA	RFA	-	-	LR, RFA
			น้อย	RFA	RFA	LR,RFA	-	-	LR, RFA
	ต่ำ	สูง	มาก	-	LR	RFA	-	-	-
			น้อย	-	-	-	-	-	-
		ปานกลาง	มาก	-	RFA	RFA	-	LR	LR, RFA
			น้อย	-	RFA	RFA	-	LR	LR, RFA
0.5:1	สูง	สูง	มาก	RFA	RFA	RFA	RFA	LR, RFA	LR, RFA
			น้อย	-	RFA	RFA	-	-	-
		ปานกลาง	มาก	RFA	RFA	RFA	RFA	LR, RFA	LR, RFA
			น้อย	-	RFA	RFA	-	-	-
	ปานกลาง	สูง	มาก	RFA	RFA	RFA	RFA	LR, RFA	LR, RFA
			น้อย	-	RFA	RFA	-	-	-
		ปานกลาง	มาก	RFA	RFA	RFA	RFA	LR, RFA	LR, RFA
			น้อย	-	RFA	RFA	-	-	-

จากตารางที่ 3 เมื่อพิจารณาประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่าง RFA และวิธี LR สรุปได้ดังนี้

3.1 ที่ระดับนัยสำคัญ .01 วิธี RFA มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อข้อสอบมีความยากปานกลาง ค่าอำนาจจำแนกปานกลางและสูง ขนาดของข้อสอบทำหน้าที่ต่างกันมาก ที่อัตราส่วนระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบเป็น 0.5:1 ส่วนวิธี LR ไม่มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

3.2 ที่ระดับนัยสำคัญ .05 วิธี RFA จะมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อข้อสอบมีความยากปานกลาง ค่าอำนาจ

จำแนกสูง ทั้ง 2 ระดับ ของการทำหน้าที่ต่างกันของข้อสอบ ยกเว้นที่อัตราส่วนระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบเป็น 0.75:1 วิธี RFA จะมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อข้อสอบมีความยากสูง ค่าอำนาจจำแนกปานกลาง ขนาดของข้อสอบทำหน้าที่ต่างกันของข้อสอบทั้ง 2 ระดับ ส่วนวิธี LR ไม่มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

3.3 ที่ระดับนัยสำคัญ .10 วิธี RFA จะมีประสิทธิภาพ ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อข้อสอบมีความยากสูง ค่าอำนาจจำแนกปานกลาง ทั้ง 2 ระดับ ของข้อสอบที่ทำหน้าที่ต่างกัน ในทุกความต่างของอัตราส่วนระหว่างกลุ่มอ้างอิงและกลุ่ม

เปรียบเทียบ และที่อัตราส่วนระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบเป็น 0.9:1, 0.75:1 และ 0.5:1 ความยากของข้อสอบปานกลาง ค่าอำนาจจำแนกสูง ทั้ง 2 ระดับของข้อสอบที่ทำหน้าที่ต่างกัน และที่อัตราส่วนระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบเป็น 1:1 ค่าความยากปานกลาง ค่าอำนาจจำแนกปานกลาง ขนาดของข้อสอบที่ทำหน้าที่ต่างกันมาก ส่วนวิธี LR จะมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในทุกความต่างของอัตราส่วนระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบเมื่อข้อสอบมีความยากปานกลาง ค่าอำนาจจำแนกปานกลาง ขนาดของข้อสอบที่ทำหน้าที่ต่างกันน้อย เท่านั้น

3.4 ที่ค่าความยากสูง ค่าอำนาจจำแนกสูงทุกระดับนัยสำคัญ ทุกขนาดความต่างของข้อสอบที่ทำหน้าที่ต่างกัน และทุกความต่างของอัตราส่วนระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ทั้ง 2 วิธีไม่มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

อภิปรายผล

1. เมื่อพิจารณาอำนาจการตรวจสอบในการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป วิธี RFA จะมีอำนาจการตรวจสอบสูงกว่าวิธี LR ซึ่งข้อค้นพบนี้ สอดคล้องกับผลการศึกษาของ นิคม กิรติวรานุ (2542) ที่ได้ศึกษาผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวิเคราะห์องค์ประกอบจำกัด วิธีแมนเทิล-แฮนสเซล และวิธีการตอบสนองข้อสอบแบบสองพารามิเตอร์โดยใช้สถานการณ์จำลอง พบว่าในภาพรวมวิธีวิเคราะห์องค์ประกอบจำกัดมีอำนาจการตรวจสอบสูงกว่าวิธีอื่นๆ ที่ศึกษา

2. จากผลการศึกษาพบว่า ลักษณะของข้อสอบมีผลต่อประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ กล่าวคือ เมื่อข้อสอบมีค่าความยากปานกลาง ค่าอำนาจจำแนกปานกลางและสูง วิธี RFA

มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเอกรูปได้ดีกว่าวิธี LR สอดคล้องกับผลการศึกษาของศึกษาของนิคม กิรติวรานุ (2542) พบว่า วิธีวิเคราะห์องค์ประกอบจำกัดมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเมื่อข้อสอบมีความยากปานกลาง ค่าอำนาจจำแนกสูงและค่าอำนาจจำแนกปานกลาง แสดงว่าในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้ลักษณะข้อสอบที่มีค่าความยากปานกลางและอำนาจจำแนกปานกลางและสูง วิธี RFA สามารถระบุข้อสอบที่ทำหน้าที่ต่างกันได้ถูกต้องแม่นยำกว่าวิธี LR ในขณะที่วิธี LR จะมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เอกรูปเมื่อข้อสอบมีความยากปานกลาง ค่าอำนาจจำแนกปานกลาง ชัดแย้งกับงานวิจัยของโรเจอร์และสวามินาทาน (Rogers & Swaminathan, 1993) ที่พบว่า วิธี LR จะมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเอกรูป ได้ดี เมื่อข้อสอบมีความยากปานกลาง แต่ต้องมีค่าอำนาจจำแนกสูง ทั้งนี้ อาจเนื่องมาจากในศึกษาครั้งนี้ผู้วิจัยจำลองข้อมูลโดยกำหนดให้ค่าอำนาจจำแนกแบ่งออกเป็น 2 ระดับ คือค่าอำนาจจำแนกสูงมีค่าอยู่ระหว่าง 0.85 ถึง 2.50 ($\bar{\alpha} = 1.675$) และค่าอำนาจจำแนกปานกลาง มีค่าอยู่ระหว่าง 0.55 ถึง 0.85 ($\bar{\alpha} = 0.70$) (คณิต ไชมุฑ, 2534) ทำให้ค่าอำนาจจำแนกมีหลายค่า ส่วนโรเจอร์และสวามินาทาน (Rogers and Swaminathan, 1993) กำหนดให้ค่าอำนาจจำแนกสูงมีค่าเท่ากับ 1.60 ค่าอำนาจจำแนกปานกลางมีค่าเท่ากับ 1.00 และค่าอำนาจจำแนกต่ำมีค่าเท่ากับ 0.60 ซึ่งค่าอำนาจจำแนกจะมีค่าเดียว ผลจากการกำหนดค่าอำนาจจำแนกที่แตกต่างกันนี้อาจทำให้การศึกษาครั้งนี้พบว่า วิธี LR มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเอกรูปเมื่อข้อสอบมีความยากปานกลาง ค่าอำนาจจำแนกปานกลาง

3. จากผลการศึกษาพบว่า ขนาดของข้อสอบที่ทำหน้าที่ต่างกัน ไม่มีผลกระทบต่ออำนาจการตรวจสอบ

การทำหน้าที่ของข้อสอบ กล่าวคือ ทั้ง 2 วิธี สามารถตรวจพบจำนวนข้อสอบที่ทำหน้าที่ต่างกันได้ใกล้เคียงกันในทุกกรณีศึกษาซึ่งไม่สอดคล้องกับผลการศึกษาของออร์ต (Oort, 1998) ที่พบว่า ในกรณีที่ข้อสอบทำหน้าที่ต่างกันมาก จะตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ดีกว่า ในกรณีที่ข้อสอบทำหน้าที่ต่างกั น้อย

ข้อเสนอแนะ

ข้อเสนอแนะสำหรับการนำผลการวิจัยไปใช้

1. ถ้าพิจารณาอำนาจการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี RFA มีอำนาจในการตรวจสอบสูงกว่าวิธี LR ในเกือบทุกกรณี ดังนั้นวิธีที่เหมาะสมและควรนำไปใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปคือ วิธี RFA

2. ถ้าพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 วิธี LR มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าวิธี RFA ในเกือบทุกกรณี ดังนั้น วิธีที่เหมาะสมและควรนำไปใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปเมื่อพิจารณาเฉพาะกรณีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำ คือ วิธี LR

3. ถ้าพิจารณาจากวิธีที่มีอำนาจการตรวจสอบสูงสุด และสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ พบว่า วิธี RFA จะมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อข้อสอบมีค่าความยากปานกลาง ค่าอำนาจจำแนกปานกลางและสูง ดังนั้น วิธีที่เหมาะสมและควรนำไปใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปในกรณีนี้คือ วิธี RFA

ข้อเสนอแนะสำหรับการทำวิจัยครั้งต่อไป

1. ควรมีการศึกษาเพิ่มเติมเกี่ยวกับประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป โดยใช้ข้อมูลจริงเพื่อเปรียบเทียบกับการศึกษาในครั้งนี้

2. ในการศึกษาครั้งนี้ศึกษาเฉพาะกรณีขนาดกลุ่มเปรียบเทียบมีขนาดเล็กโดยกำหนดให้เป็น 200 คน ดังนั้น ควรมีการศึกษาเพิ่มเติมเกี่ยวกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในลักษณะคล้ายกับการวิจัยในครั้งนี้ โดยศึกษาในกรณีที่ขนาดตัวอย่างกลุ่มเปรียบเทียบมีขนาดอื่นๆ เช่น ขนาดกลางหรือขนาดใหญ่ เป็นต้น

เอกสารอ้างอิง

คณิต ไช่มุกด์. (2534). *การพัฒนาเกณฑ์ในการเลือกข้อสอบที่เหมาะสมกับความสามารถผู้สอบ*.

วิทยานิพนธ์ครุศาสตรดุษฎีบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย. กรุงเทพฯ.

จิตติมา วรรณศรี. (2539). *การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีแมนเทิล-แฮนสล์เชลกับวิธีชิปเทสต์ เมื่อความยาวของแบบทดสอบ และขนาด กลุ่มตัวอย่าง และอัตราส่วนของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบต่างกัน*. วิทยานิพนธ์ครุศาสตรมหาบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย. กรุงเทพฯ.

นิคม กิรติวางกูร. (2542). *การเปรียบเทียบประสิทธิภาพในการตรวจการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีการวิเคราะห์องค์ประกอบจำกัด แมนเทิล-แฮนสล์และการตอบสนองข้อสอบ*. วิทยานิพนธ์ครุศาสตรมหาบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย. กรุงเทพฯ.

- ปิยะทิพย์ ดินวร. (2549). *การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบพหุมิติ: การเปรียบเทียบประสิทธิภาพระหว่างวิธีวิเคราะห์องค์ประกอบจำกัดกับวิธีถดถอยโลจิสติก*. วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต, สาขาเทคโนโลยีการวัดทางการศึกษา, บัณฑิตวิทยาลัย, มหาวิทยาลัยบูรพา. ชลบุรี.
- เสรี ชัดเข้ม. (2540). *การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันแบบไม่สม่ำเสมอของข้อสอบระหว่างวิธีแมนเทิล-เฮนส์เซลล์แบบปกติ กับวิธีแมนเทิล – เฮนส์เซลล์แบบแบ่งกลุ่ม ความสามารถของผู้สอบและความยากของข้อสอบ*. วิทยานิพนธ์ครุศาสตรดุษฎีบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย. กรุงเทพฯ.
- Bradley, J. V. (1978). The robustness. *Br. J.math.statist. Psychol*, 31, 144-152. *Printed in Great Britain*.
- Clauser, B. E., Mazor, K. M., & Hambleton, R. K. (1993). The effects of purification of the matching criterion on the identification of DIF using the Mantel-Haenszel. *Applied Psychology Measurement*, 6(4), 269-279.
- Narayanan, P., & Swaminathan, H. (1994). Performance of the Mantel-Haenszel and simultaneous item bias procedures for detecting differential item function. *Applied Psychological Measurement*, (December 1994), 315-328.
- Narayanan, P., & Swaminathan, H. (1996). Identification of item that show non uniform DIF. *Applied Psychological Measurement*, (September 1996), 257-274
- Oort, F.J. (1998). Simulation study of item bias detection with restricted factor analysis. *Structural Equation Modeling*, 5(2), 107-124.
- Roger, H. J., & Swaminathan. (1993). A comparison of logistic regression and Mantel-Haenszel procedures for detecting difference item functioning. *Applied Psychological Measurement*, 17 (2) ,105 – 116.
- Swaminathan H., & Roger, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Education Measurement*, 27 (4), 361-370.