Journal of Research Methodology, Volume 14, Number 2 (May - August 2001)

วารสารวิธีวิทยาการวิจัย ปีที่ 14 ฉบับที่ 2 (พฤษภาคม - สิงหาคม 2544)

A Comparison of the Power of the Test and Type I Error Rates in Detecting Nonuniform Differential Item Functioning Among the Modified SIBTEST, the SIBTEST, the Mantel-Haenszel and the Logistic Regression Methods*

Waleemass Saeaung

ABSTRACT

The purpose of this research was to compare the power of the test and type I error rates in detecting nonuniform differential item functioning (DIF) among the modified SIBTEST, the SIBTEST, the Mantel-Haenszel and the logistic regression methods. Data for the study were simulated under the three-parameter logistic model with fixed pseudo-guessing (c) values. Four factors were manipulated: (1) nine types of items with difficulty (b) and discrimination (a) values at high, medium and low levels, (2) two levels of test length, (3) three levels of the proportion of DIF items in the test and (4) six levels of sample size. Three hundred and twenty-four conditions were studied. Then the data of each condition have been calculated to become the power of the test and type I error rates of detecting nonuniform DIF.

The results of this research were:

1. The modified SIBTEST and the logistic regression methods were equally powerful in detecting nonuniform differential item functioning under most conditions. Both the modified SIBTEST and the logistic regression methods were more powerful than the SIBTEST and the Mantel-Haenszel methods in detecting nonuniform differential item functioning under most conditions.

2. The type I error rates for the modified SIBTEST, the SIBTEST, the Mantel-Haenszel and the logistic regression methods were within the criteria of the type I error rates at 10% level in detecting nonuniform differential item functioning under most conditions.

* Doctoral dissertation of Department of Educational Research, Chulalongkorn University under the advice of Assoc. Prof. Sirichai Kanjanawasee, Ph.D. and Assoc. Prof. Taweewat Pitayanon, Ph.D.

Journal of Research Methodology, Volume 14, Number 2 (May - August 2001)

วารสารวิธีวิทยาการวิจัย ปีที่ 14 ฉบับที่ 2 (พฤษภาคม - สิงหาคม 2544)

การเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติก*

วลีมาศ แซ่อึ้ง

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่าง วิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ข้อมูลที่ใช้ใน การศึกษาจำลองภายใต้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ ชนิดกำหนดค่าการเดา (c) คงที่แล้ว จัดกระทำข้อมูลตามปัจจัย 4 ตัว คือ (1) ลักษณะของข้อสอบที่มีค่าความยาก (b) และอำนาจจำแนก (a) ระดับต่ำ ปานกลาง และสูง จำนวน 9 ลักษณะ (2) ความยาวของแบบสอบ 2 ระดับ (3) สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบ 3 ระดับและ (4) ขนาดกลุ่มตัวอย่าง 6 ระดับ รวมข้อมูล ที่ศึกษาทั้งหมดจำนวน 324 เงื่อนไข แล้วนำข้อมูลของแต่ละเงื่อนไขมาคำนวณค่าอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป

ผลการวิจัยพบว่า

1. อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปของ วิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าเท่าเทียมกันภายใต้เกือบทุกเงื่อนไข และทั้งสองวิธีดังกล่าวมีอำนาจการทดสอบสูงกว่าวิธีซิปเทสท์ และวิธีแมนเทล-แฮนส์เซลภายใต้เกือบทุกเงื่อนไข

2. อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปของวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก มีค่าอยู่ภายในเกณฑ์ของอัตราความคลาดเคลื่อนประเภทที่ 1 ที่ระดับ 10% ภายใต้เกือบทุกเงื่อนไข

---

* อาจารย์ที่ปรึกษา รองศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี และรองศาสตราจารย์ ดร.ทวีวัฒน์ ปิตยานนท์ วิทยานิพนธ์ครุศาสตรดุษฎีบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา ปีการศึกษา 2543

ความเป็นมาและความสำคัญของปัญหา

"ความตรง" เป็นหัวใจสำคัญของแบบสอบ ในการสร้างและการตรวจสอบคุณภาพของแบบสอบจะต้องคำนึงถึงคุณภาพด้านความตรงเป็นสำคัญ ทั้งนี้เพราะว่าความตรงเป็นคุณสมบัติของแบบสอบที่แสดงถึงความสามารถในการวัดได้ถูกต้องแม่นยำ ถ้าผลการวัดได้ค่าที่ใกล้เคียงกับค่าคุณลักษณะที่แท้จริงเพียงใด ก็ถือว่าการวัดมีความตรงมากขึ้นเพียงนั้น นักวิธีวิทยาการวิจัยมักนิยมตรวจสอบความตรงของแบบสอบ 3 ประเภทหลัก คือ (1) ความตรงตามเนื้อหา (content validity) (2) ความตรงตามเกณฑ์สัมพันธ์ (criterion-related validity) และ (3) ความตรงตามภาวะสันนิษฐาน (construct validity) ส่วนการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (differential item functioning; DIF) ก็เป็นอีกประเภทหนึ่งที่ใช้ตรวจสอบคุณภาพด้านความตรงของแบบสอบ (Mazor, Clauser & Hambleton, 1992) โดยเป็นการตรวจสอบในประเด็นความอยุติธรรมของข้อสอบ (item unfairness)

ข้อสอบทำหน้าที่ต่างกันเมื่อ***ผู้สอบที่มีความสามารถระดับเดียวกันแต่อยู่ต่างกลุ่มกันมีโอกาสของการตอบข้อสอบได้ถูกต้องไม่เท่ากัน*** (Li & Stout, 1996) ซึ่งขนาดและทิศทางของข้อสอบที่ทำหน้าที่ต่างกันจะแปรเปลี่ยนไปตามระดับความสามารถที่แตกต่างกัน โดยข้อสอบที่ทำหน้าที่ต่างกันแบ่งออกเป็น 2 ประเภท (Mellenbergh, 1982) คือ ***ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป (uniform DIF) และข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป (nonuniform DIF)*** ข้อสอบที่ทำหน้าที่ต่างกันประเภทแรกเกิดขึ้นเมื่อไม่มีปฏิสัมพันธ์ (interaction) ระหว่างระดับความสามารถของผู้สอบและการเป็นสมาชิกของกลุ่ม (group membership) ส่วนข้อสอบที่ทำหน้าที่ต่างกันประเภทหลังเกิดขึ้นเมื่อมีปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบและการเป็นสมาชิกของกลุ่ม ซึ่งตามทฤษฎีการตอบสนองข้อสอบ (item response theory; IRT) สามารถพิจารณาปฏิสัมพันธ์ดังกล่าวได้จากความแตกต่างของค่าพารามิเตอร์อำนาจจำแนกของข้อสอบระหว่างผู้สอบกลุ่มย่อย 2 กลุ่ม กล่าวคือ ถ้าข้อสอบระหว่างผู้สอบกลุ่มย่อย 2 กลุ่มมีค่าอำนาจจำแนกเท่ากันแล้วโค้งลักษณะข้อสอบ (item characteristic curves; ICCs) ระหว่างผู้สอบดังกล่าวจะขนานกัน แสดงว่าข้อสอบทำหน้าที่ต่างกันแบบเอกรูป แต่ถ้าข้อสอบระหว่างผู้สอบ กลุ่มย่อย 2 กลุ่มมีค่าอำนาจจำแนกไม่เท่ากันแล้วโค้งลักษณะข้อสอบระหว่างผู้สอบดังกล่าวจะไม่ขนานกัน แสดงว่าข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป (Camilli & Shepard, 1994)

หลักการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจะเปรียบเทียบผลการตอบข้อสอบระหว่างผู้สอบกลุ่มย่อยสองกลุ่มที่มีความสามารถระดับเดียวกัน โดยที่ผู้สอบกลุ่มหนึ่งเป็นตัวแทนกลุ่มหลักในประชากรเรียกว่า ***"กลุ่มอ้างอิง" (reference group; R)*** ซึ่งเป็นกลุ่มพื้นฐาน ส่วนอีกกลุ่มหนึ่งเป็นตัวแทนกลุ่มรองในประชากรเรียกว่า ***"กลุ่มเปรียบเทียบ" (focal group; F)*** ซึ่งตาม

ปกติเป็นกลุ่มผู้สอบที่สนใจจะทำการศึกษาการทำหน้าที่ต่างกันของข้อสอบ (Angoff, 1993) ถ้าข้อสอบทำหน้าที่ต่างกันแล้วโอกาสในการตอบข้อสอบถูกของผู้สอบแต่ละกลุ่มจะไม่เท่ากัน โดยคาดว่าผู้สอบกลุ่มแรกจะได้เปรียบในการตอบข้อสอบ ส่วนผู้สอบกลุ่มหลังคาดว่าจะเสียเปรียบในการตอบข้อสอบ

วิธีการตรวจสอบ 3 วิธีต่อไปนี้ที่ผู้วิจัยสนใจนำมาศึกษา ได้แก่ วิธีซิปเทสท์ (Shealy & Stout, 1993) วิธีแมนเทล-แฮนส์เซล (Holland & Thayer, 1988) และวิธีการถดถอยโลจิสติก (Swaminathan & Rogers, 1990) โดยวิธีแรกเป็นวิธีนันพาราเมทริก (nonparametrice) ซึ่งถูกออกแบบมาเพื่อใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีทิศทางเดียวโดยเฉพาะ สามารถคำนวณได้ง่าย ไม่ยุ่งยากซับซ้อน การแปลผลไม่ยาก ทั้งยังใช้กลุ่มตัวอย่างขนาดเล็กทำให้เสียค่าใช้จ่ายไม่มาก นอกจากนี้ยังตัดสินข้อสอบที่ทำหน้าที่ต่างกันโดยใช้สถิติทดสอบนัยสำคัญ ส่วนวิธีที่สองเป็นวิธีนันพาราเมทริกเช่นเดียวกับวิธีแรก และมีคุณสมบัติคล้ายกับวิธีแรก ต่อมา Mazor และคณะ (1994) ได้นำมาปรับปรุงขั้นตอนการวิเคราะห์เพื่อนำมาตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป ปรากฏว่า สามารถตรวจสอบได้ดีกว่าวิธีแมนเทล-แฮนส์เซลแบบเดิม สำหรับวิธีสุดท้ายมีจุดเด่นตรงที่ใช้โมเดลการถดถอยโลจิสติกวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ โมเดลนี้มีเทอมที่สามารถคำนวณปฏิสัมพันธ์ระหว่างสมาชิกกลุ่มผู้สอบและระดับความสามารถ จึงทำให้สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ทั้งแบบมีทิศทางเดียวและแบบไม่มีทิศทาง ดังนั้นประเด็นปัญหาของการวิจัยในครั้งนี้ก็คือ ถ้านำวิธีซิปเทสท์มาปรับปรุงขั้นตอนการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปจะได้ผลเป็นอย่างไร โดยนำผลการตรวจสอบไปเปรียบเทียบกับวิธีซิปเทสท์แบบเดิม วิธีแมนเทล-แฮนส์เซลที่พัฒนาโดย Mazor และคณะ (1994) และวิธีการถดถอยโลจิสติก โดยศึกษาภายใต้เงื่อนไขของปัจจัยที่แตกต่างกันทางด้านลักษณะของข้อสอบ ความยาวของแบบสอบ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน และขนาดกลุ่มตัวอย่าง

วัตถุประสงค์ในการวิจัย

1. เพื่อเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยที่แตกต่างกันทางด้านลักษณะของข้อสอบ ความยาวของแบบสอบ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน และขนาดกลุ่มตัวอย่าง

2. เพื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล

และวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยที่แตกต่างกันทางด้านลักษณะของข้อสอบ ความยาวของแบบสอบ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน และขนาดกลุ่มตัวอย่าง

สมมติฐานการวิจัย

1. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปภายใต้เงื่อนไขของปัจจัยที่แตกต่างกันทางด้านลักษณะของข้อสอบ ความยาวของแบบสอบ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน และขนาดกลุ่มตัวอย่าง อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกน่าจะมีค่าสูงกว่าวิธีซิปเทสท์

2. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปภายใต้เงื่อนไขของปัจจัยที่แตกต่างกันทางด้านลักษณะของข้อสอบ ความยาวของแบบสอบ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน และขนาดกลุ่มตัวอย่าง อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกน่าจะมีค่าสูงกว่าวิธีซิปเทสท์ เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบ 4 วิธีดังกล่าวกับเกณฑ์ที่ระดับ 10% อัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้ง 4 วิธีน่าจะมีค่าอยู่ภายในเกณฑ์ที่กำหนด

ขอบเขตของการวิจัย

1. การวิจัยครั้งนี้ศึกษาในสถานการณ์จำลองภายใต้ทฤษฎีการตอบสนองข้อสอบ (IRT) โมเดลโลจิสติกแบบ 3 พารามิเตอร์ชนิดกำหนดค่า c คงที่ (3PLM-c)

2. ใช้วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 4 วิธี คือ วิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก

3. ตัวแปรที่ใช้ในการวิจัยประกอบด้วย

3.1 ตัวแปรอิสระมี 4 ตัว คือ ลักษณะของข้อสอบ ความยาวของแบบสอบ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันและขนาดกลุ่มตัวอย่างในแต่ละตัวแปรดังกล่าวยังแบ่งออกเป็นระดับต่างๆ ดังนี้

(1) ลักษณะของข้อสอบมี 9 ลักษณะ ได้แก่ ค่า a ต่ำกับ b ต่ำ, a ต่ำกับ b ปานกลาง, ค่า a ต่ำกับ b สูง, a ปานกลางกับ b ต่ำ, a ปานกลางกับ b ปานกลาง, a ปานกลางกับ b สูง, a สูงกับ b ต่ำ, a สูงกับ b ปานกลาง และ a สูงกับ b สูง

(2) ความยาวของแบบสอบมี 2 ระดับ ได้แก่ แบบสอบที่มีจำนวน 30 ข้อและ 60 ข้อ

(3) สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันมี 3 ระดับ ได้แก่ ข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบจำนวน 5 %, 10% และ 20%

(4) ขนาดกลุ่มตัวอย่างมี 6 ระดับ ได้แก่ จำนวนผู้สอบกลุ่มอ้างอิงต่อกลุ่มเปรียบเทียบเท่ากับ 250 คนต่อ 250 คน, 500 คนต่อ 250 คน, 500 คนต่อ 500 คน, 1000 คนต่อ 250 คน, 1000 คนต่อ 500 คน และ 1000 คนต่อ 1000 คน

3.2 ตัวแปรตามมี 2 ตัว คือ อำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1

4. ในการศึกษาครั้งนี้ผู้วิจัยใช้วิธีการวัดพื้นที่ชนิดไม่คิดเครื่องหมายของ Raju (1990) กรณีอเนกรูปภายใต้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ชนิดกำหนดค่า c คงที่ (3PLM-c) เป็นวิธีเกณฑ์

เพื่อใช้เป็นเกณฑ์สำหรับการเปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปซึ่งตรวจสอบด้วยวิธีที่ศึกษา 4 วิธี ได้แก่ วิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก

ประโยชน์ที่คาดว่าจะได้รับ

ผลที่ได้จากการวิจัยในครั้งนี้จะเป็นแนวทางในการปรับปรุงข้อสอบให้มีความยุติธรรมต่อกลุ่มผู้สอบ ซึ่งเป็นอีกทางเลือกหนึ่งในการพัฒนาแบบสอบให้มีคุณภาพ ทั้งยังเป็นแนวทางในการเลือกวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปที่มีประสิทธิภาพสูง ระหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกภายใต้เงื่อนไขของปัจจัยที่ศึกษา 4 ตัว ได้แก่ ลักษณะของข้อสอบ ความยาวของแบบสอบ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน และขนาดกลุ่มตัวอย่าง

วิธีดำเนินการวิจัย

1. จำลองข้อมูลเมทริกซ์คำตอบจำนวน 9 เมทริกซ์ตามลักษณะของข้อสอบ 9 ลักษณะคือ a ต่ำกับ b ต่ำ, a ต่ำกับ b ปานกลาง, a ต่ำกับ b สูง, a ปานกลางกับ b ต่ำ, a ปานกลางกับ b ปานกลาง, a ปานกลางกับ b สูง, a สูงกับ b ต่ำ, a สูงกับ b ปานกลาง และ a สูงกับ b สูง ในแต่ละเมทริกซ์มีจำนวนผู้สอบ 2,000 คน จำนวนข้อสอบ 90 ข้อ โดยใช้โปรแกรม IRTDATA version 1.0 ภายใต้ทฤษฎี IRT โมเดลโลจิสติกแบบ 3 พารามิเตอร์ชนิดกำหนดค่า c คงที่

2. จัดกระทำข้อมูลตามปัจจัยที่ศึกษา

2.1 ใช้โปรแกรม SPSS for windows version 7.52 สุ่มขนาดกลุ่มตัวอย่างในแต่ละเมทริกซ์คำตอบที่ได้จากการจำลองในข้อ 1 เพื่อให้ได้ขนาดผู้สอบกลุ่มอ้างอิงต่อกลุ่มเปรียบเทียบเท่ากับ 1,000 คนต่อ 1,000 คน

2.2 นำข้อมูลเมทริกซ์คำตอบที่แบ่งกลุ่มผู้สอบในข้อ 2.1 มาประมาณค่าพารามิเตอร์

ของข้อสอบภายใต้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ชนิดกำหนดค่า c คงที่ โดยใช้โปรแกรม BILOG version 3.04

2.3 นำค่าพารามิเตอร์ระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบในข้อ 2.2 มาปรับเทียบสเกลพารามิเตอร์ของข้อสอบให้อยู่บนสเกลเดียวกันด้วยวิธีโค้งลักษณะแบบสอบ (test characteristic curve; TCC) ภายใต้โมเดลโลจิสติกแบบ 2 พารามิเตอร์ โดยใช้โปรแกรม EQUATE version 2.0

2.4 นำข้อมูลที่ปรับเทียบสเกลพารามิเตอร์ของข้อสอบในข้อ 2.3 มาวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป โดยใช้วิธีการวัดพื้นที่ในช่วงเปิดชนิดไม่คิดเครื่องหมายของ Raju (1990) ภายใต้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ชนิดกำหนดค่า c คงที่ โดยใช้โปรแกรม IRTDIF version 1.0

2.5 สุ่มข้อสอบที่ทำหน้าที่ต่างกันและข้อสอบที่ทำหน้าที่ไม่ต่างกันในเมทริกซ์ข้อมูลซึ่งตรวจสอบแล้วในข้อ 2.4 เพื่อนำมาจัดกระทำเป็นแบบสอบ 6 ฉบับ

2.6 สุ่มขนาดกลุ่มตัวอย่างในแต่ละเมทริกซ์ข้อมูลจากข้อ 2.5 เพื่อให้ได้ขนาดกลุ่มผู้สอบจำนวน 6 ระดับ คือ จำนวนผู้สอบกลุ่มอ้างอิงต่อกลุ่มเปรียบเทียบเท่ากับ 250 คนต่อ 250 คน, 500 คนต่อ 250 คน, 500 คนต่อ 500 คน, 1000 คนต่อ 250 คน, 1000 คนต่อ 500 คน และ 1000 คนต่อ 1000 คน รวมข้อมูลที่จัดกระทำทั้งหมด 324 เงื่อนไข (9 x 2 x 3 x 6)

2.7 นำข้อมูลในข้อ 2.6 มาวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีซิปเทสท์ วิธีซิปเทสท์ปรับใหม่ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก โดยสองวิธีแรกใช้โปรแกรม SIBTEST วิธีต่อมาใช้โปรแกรม MHDIF version 1.0 และวิธีสุดท้ายใช้โปรแกรม SPSS/PC+ version 4.01

3. ทดสอบความแตกต่างของอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีการตรวจสอบ 4 วิธีโดยใช้สถิติการวิเคราะห์ความแปรปรวนแบบ 2 ทาง (two-way ANOVA) และแบบทางเดียว (one-way ANOVA) ส่วนการเปรียบเทียบความแตกต่างเป็นรายคู่ใช้วิธีการทดสอบของ Tukey และการทดสอบ t-test

สรุปผลการวิจัย

ผลการวิจัยสรุปเป็น 2 ประเด็นใหญ่ๆ ดังนี้ (รายละเอียดย่อยๆ สามารถศึกษาได้ในวิทยานิพนธ์ฉบับสมบูรณ์ของผู้วิจัย)

1. การเปรียบเทียบอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการ

ถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยที่แตกต่างกันทางด้านลักษณะของข้อสอบ ความยาวของแบบสอบ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน และขนาดกลุ่มตัวอย่าง ปรากฏว่า อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงใกล้เคียงกันภายใต้เกือบทุกเงื่อนไข ส่วนอำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีค่าต่ำใกล้เคียงกันภายใต้เกือบทุกเงื่อนไข ทั้งวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีอำนาจการทดสอบสูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ภายใต้เกือบทุกเงื่อนไข

2. การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซลและวิธีการถดถอยโลจิสติก ภายใต้เงื่อนไขของปัจจัยที่แตกต่างกันทางด้านลักษณะของข้อสอบ ความยาวของแบบสอบ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน และขนาดกลุ่มตัวอย่าง ปรากฏว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงใกล้เคียงกันภายใต้เกือบทุกเงื่อนไข ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ และวิธีแมนเทล-แฮนส์เซลมีค่าต่ำใกล้เคียงกันภายใต้เกือบทุกเงื่อนไข ทั้งวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ภายใต้เกือบทุกเงื่อนไข

เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบ 4 วิธี ภายใต้เงื่อนไขของปัจจัยที่ศึกษา 324 เงื่อนไขกับเกณฑ์ของอัตราความคลาดเคลื่อนประเภทที่ 1 ที่ระดับ 10% ปรากฏว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกมีค่าอยู่ภายในเกณฑ์ที่กำหนดดังกล่าว

อภิปรายผลการวิจัย

1. ผลการวิจัยที่สรุปในข้อ 1 เฉพาะวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ และวิธีการถดถอยโลจิสติกสอดคล้องกับสมมติฐานที่กำหนดไว้ในข้อ 1 แสดงว่า วิธีซิปเทสท์ปรับใหม่ที่ผู้วิจัยนำมาปรับปรุงขั้นตอนการวิเคราะห์สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปได้อย่างมีประสิทธิภาพ โดยมีความถูกต้องแม่นยำในการตรวจสอบสูงกว่าวิธีซิปเทสท์แบบเดิม ทั้งยังมีความถูกต้องแม่นยำในการตรวจสอบเทียบเท่ากับวิธีการถดถอยโลจิสติก ทั้งนี้อาจเนื่องมาจากการปรับปรุงขั้นตอนการวิเคราะห์ของวิธีซิปเทสท์ กล่าวคือ ข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปมี 2 ลักษณะ คือ ข้อสอบทำหน้าที่ต่างกันแบบมีทิศทางเดียว (unidirectional DIF) และข้อสอบทำหน้าที่ต่างกันแบบไม่มีทิศทาง (nondirectional DIF) (Li & Stout, 1993 cited in Narayanan & Swaminathan, 1996) ข้อสอบทำหน้าที่ต่างกันแบบมีทิศทางเดียวเกิดขึ้นเมื่อ

มีปฏิสัมพันธ์เป็นลำดับ (ordinal interaction) ระหว่างการเป็นสมาชิกของกลุ่มและระดับความสามารถ ซึ่งในทฤษฎีการตอบสนองข้อสอบ (item response theory) สามารถพิจารณาได้จากโค้งลักษณะข้อสอบ (item characteristic curves) ระหว่างกลุ่มผู้สอบ 2 กลุ่มตัดกันตรงปลายสุดของช่วงความสามารถต่ำหรือสูง ข้อสอบลักษณะดังกล่าวจะมีค่าความยากต่ำหรือสูงซึ่งไม่มีปัญหาเมื่อตรวจสอบด้วยวิธีซิปเทสท์ ส่วนข้อสอบทำหน้าที่ต่างกันแบบไม่มีทิศทางเกิดขึ้นเมื่อมีปฏิสัมพันธ์ ไม่เป็นลำดับ (disordinal interaction) ระหว่างการเป็นสมาชิกของกลุ่มและระดับความสามารถ ซึ่งทำให้โค้งลักษณะข้อสอบระหว่างกลุ่มผู้สอบ 2 กลุ่มตัดกันตรงจุดกึ่งกลางของช่วงความสามารถ ข้อสอบที่มีลักษณะนี้จะมีค่าความยากปานกลางซึ่งไม่สามารถตรวจสอบด้วยวิธีซิปเทสท์ ทั้งนี้เพราะว่าวิธีซิปเทสท์ใช้สถิติชนิดคิดเครื่องหมายคำนวณดัชนีข้อสอบที่ทำหน้าที่ต่างกัน เมื่อขนาดของข้อสอบที่ทำหน้าที่ต่างกันเปลี่ยนทิศทางตรงจุดกึ่งกลางของช่วงความสามารถจะทำให้ความแตกต่างที่มีเครื่องหมายลบของช่วงคะแนนส่วนหนึ่งหักล้างกับความแตกต่างที่มีเครื่องหมายบวกของช่วงคะแนนอีกส่วนหนึ่ง ข้อสอบที่มีลักษณะดังกล่าวจึงไม่สามารถตรวจสอบด้วยวิธีซิปเทสท์ ดังนั้นในการศึกษาครั้งนี้ผู้วิจัยจึงนำวิธีซิปเทสท์ของ Shealy และ Stout (1993) มาปรับปรุงขั้นตอนการวิเคราะห์โดยแบ่งกลุ่มผู้สอบออกเป็น 2 กลุ่มตามระดับคะแนนรวม คือ กลุ่มที่มีคะแนนสูงกับกลุ่มที่มีคะแนนต่ำ แล้ววิเคราะห์แยกกันในแต่ละกลุ่ม ผลการแบ่งกลุ่มผู้สอบจะทำให้โค้งลักษณะข้อสอบที่ตัดกันตรงจุดกึ่งกลางของช่วงความสามารถเปลี่ยนเป็นตัดกันตรงปลายสุดของช่วงความสามารถต่ำหรือสูง ดังนั้นจึงไม่มีผลต่อสถิติชนิดคิดเครื่องหมายของวิธีซิปเทสท์ปรับใหม่ นั่นคือ วิธีซิปเทสท์ปรับใหม่สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปที่ไม่มีทิศทางได้ สำหรับวิธีการถดถอยโลจิสติกจะไม่มีปัญหาในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปเนื่องจากเป็นวิธีที่ใช้โมเดลการถดถอยโลจิสติกวิเคราะห์ดัชนีการทำหน้าที่ต่างกันของข้อสอบ โมเดลดังกล่าวสามารถคำนวณปฏิสัมพันธ์ระหว่างการเป็นสมาชิกของกลุ่มและระดับความสามารถ ดังนั้นจึงทำให้สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปที่มีทิศทางเดียวและไม่มีทิศทางได้ (Swaminathan & Rogers, 1990) ส่วนวิธีแมนเทล-แฮนส์เซลไม่เป็นไปตามสมมติฐานที่กำหนดไว้ในข้อ 1 แสดงว่า วิธีแมนเทล-แฮนส์เซลแบบปรับปรุงขั้นตอนการวิเคราะห์ถึงแม้ว่าจะมีอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปสูงกว่าวิธีแมนเทล-แฮนส์เซลแบบเดิม (Mazor & others, 1994) แต่เมื่อนำมาเปรียบเทียบกับวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติก ปรากฏว่า มีอำนาจการทดสอบต่ำกว่าวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติก และเมื่อนำมาเปรียบเทียบกับวิธีซิปเทสท์ ปรากฏว่า มีอำนาจการทดสอบเท่าเทียมกัน

2. ผลการวิจัยที่สรุปในข้อ 2 เฉพาะวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ และวิธีการถดถอย

โลจิสติกสอดคล้องกับสมมติฐานที่กำหนดไว้ในข้อ 2 สาเหตุดังกล่าวอาจเนื่องมาจากในการศึกษาครั้งนี้ผู้วิจัยจำลองข้อมูลโดยกำหนดให้การแจกแจงค่าความสามารถของกลุ่มผู้สอบเป็นแบบปกติเหมือนกันในทุกเงื่อนไขของการจำลองข้อมูล ต่อจากนั้นจึงจัดกระทำขนาดกลุ่มตัวอย่างระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบโดยการสุ่ม ผลจากการสุ่มจะทำให้การแจกแจงค่าความสามารถระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบมีค่าเท่ากัน ซึ่งจะไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 แต่ถ้าการแจกแจงค่าความสามารถระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบมีค่าแตกต่างกันจะมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าเพิ่มขึ้น (Rogers & Swaminathan, 1993; Narayanan & Swaminathan, 1996) ในการวิเคราะห์ด้วยวิธีซิปเทสท์ผู้วิจัยใช้ข้อมูลที่ได้จากการสุ่มดังที่กล่าวมาข้างต้น ดังนั้นจึงไม่มีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ แต่การวิเคราะห์ด้วยวิธีซิปเทสท์ปรับใหม่ผู้วิจัยปรับปรุงขั้นตอนการวิเคราะห์โดยแบ่งกลุ่มผู้สอบออกเป็น 2 กลุ่ม ตามระดับคะแนนรวม คือ กลุ่มที่มีคะแนนสูงกับกลุ่มที่มีคะแนนต่ำ แล้ววิเคราะห์แยกกันในแต่ละกลุ่มผลจากการแบ่งกลุ่มดังกล่าวอาจมีผลทำให้การแจกแจงค่าความสามารถระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบมีค่าแตกต่างกัน ซึ่งจะมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่มีค่าเพิ่มขึ้น ดังนั้นวิธีซิปเทสท์ปรับใหม่จึงมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีซิปเทสท์ ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติก ปรากฏว่า มีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ เนื่องจากวิธีซิปเทสท์มีรูปแบบนันพาราเมทริก (nonparametric) เมื่อทดสอบกับสถิติ Z ซึ่งมีการแจกแจงแบบปกติไม่จำเป็นต้องใช้กลุ่มตัวอย่างขนาดใหญ่ แต่วิธีการถดถอยโลจิสติกมีรูปแบบพาราเมทริก (parametric) เมื่อทดสอบกับสถิติไค-สแควร์ซึ่งมีการแจกแจงแบบเชิงเส้นกำกับ (asymptotic) จะต้องใช้กลุ่มตัวอย่างขนาดใหญ่ ถ้าใช้กลุ่มตัวอย่างขนาดเล็กจะทำให้การทดสอบขาดความตรง (Swaminathan & Rogers, 1990) ซึ่งจะทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าเพิ่มขึ้น ในการศึกษาครั้งนี้ผู้วิจัยใช้ขนาดกลุ่มตัวอย่างหลายระดับ ตั้งแต่ขนาดเล็ก (N_R : N_F = 250 : 250) จนถึงขนาดใหญ่ (N_R : N_F = 1000 : 1000) ดังนั้นในกลุ่มตัวอย่างขนาดเล็กจะมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการถดถอยโลจิสติกมีค่าสูงกว่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ ส่วนวิธีแมนเทล-แฮนส์เซลไม่เป็นไปตามสมมติฐานที่กำหนดไว้ในข้อ 2 สาเหตุดังกล่าวอาจเนื่องมาจากในการศึกษาครั้งนี้ ผู้วิจัยใช้การวิเคราะห์แบบ 2 ขั้นตอน กล่าวคือ ในขั้นตอนแรกจะวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบโดยใช้คะแนนรวมของผู้สอบเป็นเกณฑ์ในการจับคู่ของกลุ่มผู้สอบ และในขั้นตอนที่สองจะวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบซ้ำอีกครั้งหนึ่ง แต่จะนำข้อสอบที่ตรวจพบว่าทำหน้าที่ต่างกันในขั้นตอนแรกออกก่อนที่จะคำนวณคะแนนรวมของผู้สอบ การวิเคราะห์ดังกล่าวจะทำให้เกณฑ์การจับคู่

ของกลุ่มผู้สอบมีความบริสุทธิ์ ซึ่งจะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าลดลง ดังนั้น วิธีแมนเทล-แฮนส์เซลจึงมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติก

เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบทั้ง 4 วิธี ภายใต้เงื่อนไขของปัจจัยที่ศึกษา 324 เงื่อนไขกับเกณฑ์ของอัตราความคลาดเคลื่อนประเภทที่ 1 ที่ระดับ 10% ปรากฏว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกมีค่าอยู่ภายในเกณฑ์ที่กำหนดดังกล่าว ผลการศึกษาดังกล่าวสอดคล้องกับสมมติฐานในข้อ 2 และสอดคล้องกับผลการศึกษาของ Narayanan และ Swaminathan (1994, 1996) ที่พบว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ แบบเอกรูปและแบบอเนกรูปอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซล วิธีซิปเทสท์ วิธีโคร-ซิป และวิธีการถดถอยโลจิสติกมีค่าเฉลี่ยต่ำกว่า 10%

ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1. จากผลการวิจัย พบว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปภายใต้ปัจจัยที่ศึกษา 4 ตัว คือ ลักษณะของข้อสอบ ความยาวของแบบสอบ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบ และขนาดกลุ่มตัวอย่าง อำนาจการทดสอบของวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าสูงใกล้เคียงกันภายใต้เกือบทุกเงื่อนไข ส่วนอำนาจการทดสอบของวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลมีค่าต่ำใกล้เคียงกันภายใต้เกือบทุกเงื่อนไข ทั้งวิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีอำนาจการทดสอบสูงกว่าวิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซลภายใต้เกือบทุกเงื่อนไข เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบทั้ง 4 วิธีภายใต้ปัจจัยที่ศึกษา 324 เงื่อนไขกับเกณฑ์ที่ระดับ 10% พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบทั้ง 4 วิธีมีค่าอยู่ภายในเกณฑ์ดังกล่าว นั่นคือ วิธีซิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีความถูกต้องแม่นยำในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปได้เท่าเทียมกัน และทั้ง 2 วิธีมีความถูกต้องแม่นยำสูงกว่า วิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซล โดยทั่วๆ ไปแล้วนักวิธีวิทยาการวิจัยได้ให้การยอมรับการตรวจสอบด้วยวิธีการถดถอยโลจิสติก ทั้งนี้เนื่องจากวิธีดังกล่าวใช้โมเดลการถดถอยโลจิสติกซึ่งสามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและแบบอเนกรูป แต่วิธีดังกล่าวยังมีข้อจำกัดหลายประการ ดังเช่น การประมาณค่าพารามิเตอร์ในโมเดลการถดถอยโลจิสติกมีความซับซ้อนเพราะต้องคำนวณทวนซ้ำหลายรอบ (iterative) ซึ่งทำให้การคำนวณต้องใช้เวลามาก ทั้งยังเสียค่าใช้จ่ายค่อนข้างสูง นอกจากนี้การใช้สถิติไค-สแควร์ที่มีการแจกแจงแบบเชิงเส้นกำกับ (asymptotic)

ทดสอบนัยสำคัญจะต้องใช้ขนาดกลุ่มตัวอย่างที่เพียงพอ เพื่อมิให้การทดสอบขาดความตรง ส่วนวิธีซิปเทสท์ปรับใหม่มีข้อได้เปรียบตรงที่สามารถคำนวณได้ง่ายไม่ซับซ้อน โดยประมาณค่าดัชนีการทำหน้าที่ต่างกันของข้อสอบจากสัดส่วนของการตอบข้อสอบระหว่างกลุ่มผู้สอบ ๒ กลุ่ม ไม่ต้องคำนวณทวนซ้ำหลายรอบ ซึ่งทำให้คำนวณได้เร็ว ทั้งยังไม่จำเป็นต้องใช้กลุ่มตัวอย่างที่มีขนาดใหญ่ ทำให้เสียค่าใช้จ่ายไม่มาก นอกจากนี้ยังตัดสินข้อสอบที่ทำหน้าที่ต่างกันโดยใช้สถิติ Z ทดสอบนัยสำคัญทางสถิติทำให้มีความถูกต้องสูง ดังนั้นวิธีซิปเทสท์ปรับใหม่จึงเป็นอีกทางเลือกหนึ่งที่สามารถนำมาใช้แทนวิธีการถดถอยโลจิสติกได้

๒. จากผลการวิจัย พบว่า ลักษณะของข้อสอบมีผลต่ออำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ ๑ ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีการถดถอยโลจิสติก กล่าวคือ เมื่อลักษณะของข้อสอบมีค่าความยากปานกลางจะมีผลทำให้อำนาจการทดสอบของวิธีการถดถอยโลจิสติกมีค่าเพิ่มมากขึ้น และเมื่อลักษณะของข้อสอบมีค่าอำนาจจำแนกปานกลางถึงสูงจะมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ ๑ ของวิธีดังกล่าวมีค่าลดลง แสดงว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปภายใต้ลักษณะของข้อสอบที่มีค่าความยากปานกลางและอำนาจจำแนกปานกลางถึงสูง วิธีการถดถอยโลจิสติกสามารถระบุข้อสอบที่ทำหน้าที่ต่างกันได้ถูกต้องแม่นยำ หรืออาจระบุผิดพลาดว่าข้อสอบทำหน้าที่ต่างกัน (ทั้งที่ความเป็นจริงข้อสอบทำหน้าที่ไม่ต่างกัน) ได้น้อยมาก จากผลการศึกษาดังกล่าว สามารถนำวิธีการถดถอยโลจิสติกไปตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปในแบบสอบวัดผลสัมฤทธิ์ทางการเรียน ทั้งนี้เนื่องจากในแบบสอบดังกล่าวมักจะมีข้อสอบที่มีค่าความยากปานกลางและอำนาจจำแนกสูง ดังนั้นถ้าตรวจสอบด้วยวิธีการถดถอยโลจิสติกจะส่งผลให้การตรวจสอบมีประสิทธิภาพสูง

๓. จากผลการวิจัย พบว่าความยาวของแบบสอบมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ ๑ ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่ กล่าวคือ อัตราความคลาดเคลื่อนประเภทที่ ๑ ในแบบสอบ ๖๐ ข้อ มีค่าต่ำกว่าอัตราความคลาดเคลื่อนประเภทที่ ๑ ในแบบสอบ ๓๐ ข้อ แสดงว่า เมื่อเพิ่มความยาวของแบบสอบจะมีผลทำให้อัตราความคลาดเคลื่อนประเภทที่ ๑ มีค่าลดลง นั่นคือ การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปในแบบสอบ ๖๐ ข้อ ควรใช้วิธีซิปเทสท์ปรับใหม่ ทั้งนี้จะทำให้ระบุผิดพลาดว่าข้อสอบทำหน้าที่ต่างกัน (ทั้งที่ความเป็นจริงข้อสอบทำหน้าที่ไม่ต่างกัน) ได้น้อยที่สุด

๔. จากผลการวิจัย พบว่า สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบไม่มีผลต่ออำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ ๑ ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ กล่าวคือ อำนาจการทดสอบและอัตราความคลาดเคลื่อน

ประเภทที่ 1 ของวิธีดังกล่าวภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 5%, 10% และ 20% มีค่าใกล้เคียงกัน แสดงว่า เมื่อตรวจสอบด้วยวิธีซิปเทสท์ภายใต้สัดส่วนของข้อสอบที่ทำหน้าที่ ต่างกัน 5%, 10% และ 20% จะสามารถระบุข้อสอบที่ทำหน้าที่ต่างกันได้ถูกต้องใกล้เคียงกัน หรืออาจระบุผิดพลาดว่าข้อสอบทำหน้าที่ต่างกัน (ทั้งที่ความเป็นจริงข้อสอบทำหน้าที่ไม่ต่างกัน) ได้ใกล้เคียงกัน จากข้อค้นพบดังกล่าวสามารถนำวิธีซิปเทสท์มาปรับปรุงขั้นตอนการวิเคราะห์โดยใช้วิธีการวิเคราะห์แบบ 2 ขั้นตอน กล่าวคือ ในขั้นตอนแรกจะคำนวณคะแนนรวมของผู้สอบจากข้อสอบทุกข้อเพื่อใช้เป็นเกณฑ์ในการจับคู่ของกลุ่มผู้สอบ แล้ววิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ในขั้นตอนที่สองจะนำข้อสอบที่ถูกระบุว่าทำหน้าที่ต่างกันออกจากการคำนวณคะแนนรวมของผู้สอบ แล้ววิเคราะห์การทำหน้าที่ต่างกันของข้อสอบซ้ำอีกครั้งหนึ่ง ผลการวิเคราะห์จะทำให้เกณฑ์ในการจับคู่ของกลุ่มผู้สอบมีความบริสุทธิ์ (purification) ซึ่งจะส่งผลให้การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์มีประสิทธิภาพมากยิ่งขึ้น

5. จากผลการวิจัย พบว่า ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 250 : 250, 500 : 250, 500 : 500, 1,000 : 250 และ 1,000 : 500 ไม่มีผลต่ออำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ และวิธีแมนเทล-แฮนส์เซล กล่าวคือ เมื่อใช้วิธีการตรวจสอบ 3 วิธีดังกล่าวภายใต้ขนาดกลุ่มตัวอย่าง $N_R : N_F$ = 250 : 250, 500 : 250, 500 : 500, 1,000 : 250 และ 1,000 : 500 จะสามารถระบุข้อสอบที่ทำหน้าที่ต่างกันได้ถูกต้องใกล้เคียงกัน หรืออาจระบุผิดพลาดว่าข้อสอบทำหน้าที่ต่างกัน (ทั้งที่ความเป็นจริงข้อสอบทำหน้าที่ไม่ต่างกัน) ได้ใกล้เคียงกัน ดังนั้นในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปด้วยวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์และวิธีแมนเทล-แฮนส์เซล เมื่อต้องการใช้ขนาดกลุ่มอ้างอิงหรือกลุ่มเปรียบเทียบต่ำกว่า 1,000 คน สามารถใช้เพียงกลุ่มละ 250 คน ก็สามารถตรวจสอบได้อย่างมีประสิทธิภาพ

ข้อเสนอแนะในการวิจัยต่อไป

1. ควรมีการศึกษาเพิ่มเติมเกี่ยวกับประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป โดยศึกษาเฉพาะกรณีของข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปที่มีทิศทางเดียว (unidirectional) และข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปที่ไม่มีทิศทาง (non-unidirectional) ตามกรอบของทฤษฎี IRT

2. ควรมีการศึกษาเพิ่มเติมเกี่ยวกับประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปโดยใช้ข้อมูลจริงเพื่อเปรียบเทียบกับการศึกษาในครั้งนี้ นอกจากนี้ควรศึกษาร่วมกับตัวแปรที่เกี่ยวกับลักษณะของผู้สอบ เช่น เพศ เชื้อชาติ ศาสนา ภูมิลำเนา เป็นต้น

3. ควรมีการศึกษาเพิ่มเติมเกี่ยวกับประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป ซึ่งมีลักษณะคล้ายกับการวิจัยในครั้งนี้ โดยศึกษาร่วมกับตัวแปรอื่นๆ ที่คาดว่าน่าจะมีผลต่ออำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ ๑ ของวิธีการตรวจสอบ เช่น ตัวแปรความแตกต่างของการแจกแจงค่าความสามารถ (ability distribution differences) ระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ตัวแปรขนาดอิทธิพลของข้อสอบที่ทำหน้าที่ต่างกัน (DIF effect size) ตัวแปรเกณฑ์ในการจับคู่ของกลุ่มผู้สอบ (matching criterion) และตัวแปรอัตราส่วน (ratio) ระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ เป็นต้น

4. ควรมีการศึกษาเพิ่มเติมเกี่ยวกับประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบพหุวิภาค (polytomous) เช่น แบบสอบวัดผลสัมฤทธิ์ทางการเรียนที่มีการให้คะแนนแบบบางส่วน มาตรวัดเจตคติ มาตรวัดบุคลิกภาพ การประเมินทักษะการปฏิบัติ และการประเมินตามสภาพที่แท้จริง เป็นต้น โดยใช้วิธีในกลุ่ม IRT ได้แก่ วิธีการทดสอบไค-สแควร์ของ Lord วิธีการทดสอบอัตราส่วนไลค์ลิฮู้ด วิธีการวัดพื้นที่ของ Raju วิธีการวัดพื้นที่ของ Kim และ Cohen วิธีซิปเทสท์ และวิธีซิปเทสท์ปรับใหม่ เป็นต้น หรือใช้วิธีในกลุ่ม non-IRT ได้แก่ วิธีการทำให้เป็นมาตรฐาน (standardization) วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก (logistic discriminate function analysis) และวิธีแมนเทล-แฮนส์เซล เป็นต้น

5. ควรมีการศึกษาเพิ่มเติมเกี่ยวกับประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเมื่อมีผู้สอบหลายกลุ่ม โดยใช้วิธีในกลุ่ม IRT และ non-IRT และควรศึกษาเกี่ยวกับการตรวจสอบการทำหน้าที่ต่างกันของกลุ่มข้อสอบ (differential bundle functioning; DBF)

เอกสารอ้างอิง

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

Camilli, G. & Shepard, L.A. (1994). **Methods for identifying biased test Items.** California: Sage Publications, Inc.

Holland, P.W. & Thayer, D.T. Differential item performance and the Mantel-Haenszel procedure. In Wainer, H. & Braun, H.I. (Eds.), (1998). **Test validity,** pp.129-145. Hillsdale, NJ: Lawrence Erlbaum, Associates, Inc.

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

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

Mazor, K.M., Clauser, B.E. & Hambleton, R.K. (1992). The effect of sample size on the functioning of the Mantel-Haenszel statistic. **Educational and Psychological Measurement** 52: 443-451.

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

Narayanan, P. & Swaminathan, H. (1996). Identification of items that show nonuniform DIF. **Applied Psychological Measurement** 20(3): 257-274.

Narayanan, P. & Swaminathan, H. (1994). Performance of the Mantel-Haenzel and Simultaneous item bias procedures for detecting differential item functioning. **Applied Psychological Measurement** 18(4): 315-328.

Raju, N.S. (1990). Determining the significance of estimated signed and unsigned areas between two item response functions. **Applied Psychological Measurement** 14(2): 197-207.

Rogers, H. J. & Swaminathan, H. (1993). A Comparison of logistic regression and Mantel-Haenszel procedures for detecting differential item functioning.

Applied Psychological Measurement 17(2): 105-116.

Shealy, R. & Stout, W.F.(1993). A model-based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DIF as well as item bias/DIF. **Psychometrika** 58(2): 159-194.

Swaminathan, H. & Rogers, H.J. (1990). Detecting differential item functioning using logistic Regression Procedures. **Journal of Educational Measurement** 27(4): 361-370.