

ประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้วิธีถดถอยโลจิสติก
สำหรับข้อสอบที่ตรวจให้คะแนนแบบทวิภาค: เปรียบเทียบขนาดอิทธิพลสองเกณฑ์**

Efficacy of Detection Dif in Logistic Regression
for Dichotomously Scored Items by Using Two Effect Size Criteria

ธเกียรติกมล ทองงอก* โชติกา ภาณุผล และศิริชัย กาญจนวาสี

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

บทคัดย่อ

วัตถุประสงค์ของการศึกษาเพื่อเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีรูปแบบการให้คะแนนแบบ ทวิภาค โดยการจำลองข้อมูลในวิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin & Gierl กับ Zumbo & Thomas ใช้ทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์ จำลองผลการตอบภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน ขนาดของการทำหน้าที่ต่างกัน จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับ ความยาวแบบสอบทั้งฉบับ ข้อมูลที่ศึกษา 24 เงื่อนไข ทุกเงื่อนไขจำลองข้อมูลซ้ำ 25 ครั้ง วิเคราะห์ข้อมูลทุกเงื่อนไขด้วยวิธีถดถอยโลจิสติกโดยใช้ระดับนัยสำคัญที่ .05 สรุปผลการวิจัยดังนี้

1. วิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin & Gierl มีอัตราความถูกต้องในการตรวจสอบสูงกว่าเกณฑ์ Zumbo & Thomas ภายใต้เกือบทุกเงื่อนไข
2. ข้อสอบที่ทำหน้าที่ต่างกันแบบอนุกรมมีอัตราความถูกต้องจากการวัดขนาดอิทธิพลทั้ง 2 เกณฑ์สูงกว่าแบบเอกรูป แบบสอบที่มีจำนวนข้อสอบทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 20 มีอัตราความถูกต้องของขนาดอิทธิพลทั้ง 2 เกณฑ์สูงกว่าในแบบสอบที่มีจำนวนข้อสอบ ที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10 เมื่อขนาดอิทธิพลของข้อสอบที่การทำหน้าที่ต่างกันเพิ่มขึ้นทำให้อัตราความถูกต้องของขนาดอิทธิพลทั้ง 2 เกณฑ์เพิ่มขึ้นภายใต้เกือบทุกเงื่อนไข

* ผู้ประสานงานหลัก (Corresponding Author)

**งานวิจัยเรื่องนี้ได้รับการสนับสนุนทุนวิจัยจาก “ทุน 90 ปี จุฬาลงกรณ์มหาวิทยาลัย”

กองทุนรัชดาภิเษกสมโภช

(THE 90th ANNIVERSARY OF CHULALONGNORN UNIVERSITY FUND (Ratchadaphiseksomphot Endowment Fund))

e-mail: thongngok@hotmail.com

ข้อเสนอแนะ : ภายใต้การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอย โลจิสติก ควรใช้ผลการทดสอบระดับนัยสำคัญร่วมกับผลของการวัดขนาดอิทธิพล

คำสำคัญ : การทำหน้าที่ต่างกันของข้อสอบ เกณฑ์การวัดขนาดอิทธิพล วิธีถดถอยโลจิสติก

Abstract

The objective of this study was to compare the correct identification and Type I error rate of DIF with dichotomously scored items by simulation data in a logistic regression procedure between effect size measures of Jodoin and Gierl's criteria and Zumbo and Thomas's criteria. In this study, data was simulated under the IRT theory of two-parameter item response, simulating dichotomous response under the condition of 4 varied factors. The total data studied were 24 conditions: DIF Type, amounts of DIF, items with DIF and Test length. The data were replicated 25 times for each condition. In each condition, the data were analyzed with effect size measures. Significance .05 was used in the analysis of all DIF. The research results were as follows:

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

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

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

Keywords : differential item functioning, effect size criteria, logistic regression

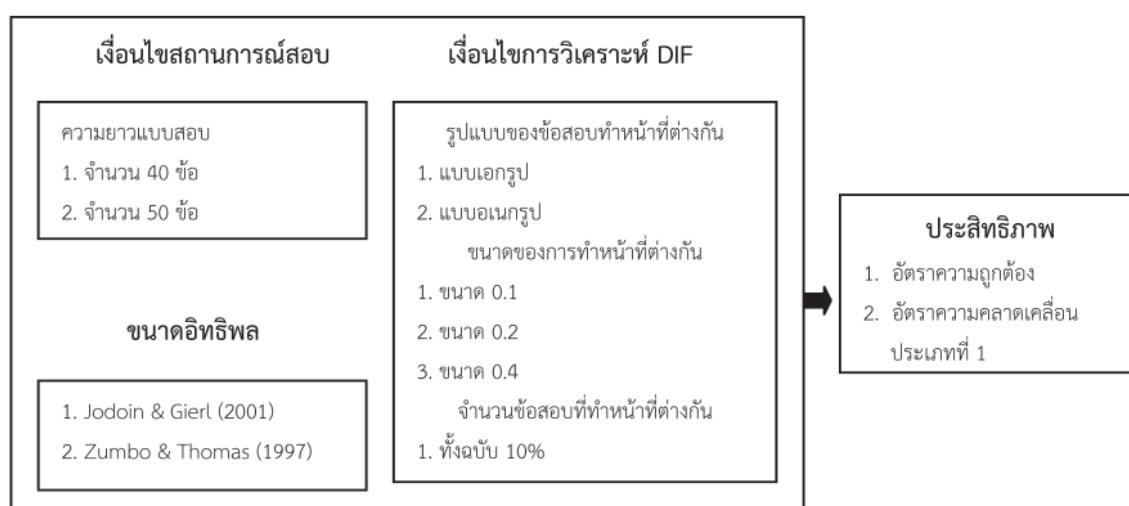
บทนำ

การสอบเพื่อสรุปบุคคลจากคะแนนที่ได้โดยอาศัยเครื่องมือต้องดำเนินการด้วยความยุติธรรม เนื่องจากความไม่ยุติธรรมถือเป็นภาวะคุกคามต่อคุณภาพด้านความตรงของเครื่องมือ (Zumbo, 2005) ถ้าเครื่องมือเอนเอียงเข้าข้างผู้สอบกลุ่มใดกลุ่มหนึ่งแสดงว่าเครื่องมือเกิดความลำเอียง (Shumacker, 2005) การทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning : DIF) หมายถึงข้อสอบที่ทำให้ผู้ตอบจากต่างกลุ่มที่มีความสามารถเท่ากันเกิดความน่าจะเป็นในการตอบได้ถูกต้องแตกต่างกันทำให้เกิดการได้เปรียบเสียเปรียบกันระหว่างกลุ่มผู้สอบ

วัตถุประสงค์

1. เพื่อเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ในวิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin & Gierl และการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo & Thomas ภายใต้เงื่อนไขเดียวกันของปัจจัยที่แปรเปลี่ยน 4 ปัจจัยและปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย
2. เพื่อเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ในวิธีถดถอยโลจิสติก ด้วยการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin & Gierl และการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo & Thomas ภายใต้เงื่อนไขต่างกันของปัจจัยที่แปรเปลี่ยน 4 ปัจจัยและปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย

กรอบแนวคิดในการวิจัย



แผนภาพที่ 1 กรอบแนวคิดในการวิจัย

เอกสารและงานวิจัยที่เกี่ยวข้อง

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก การศึกษาผลของตัวแปรต้นที่มีต่อตัวแปรตาม โดยความสัมพันธ์ระหว่างตัวแปรต้นกับโอกาสของการเกิดเหตุการณ์ของตัวแปรตามมีลักษณะเป็นฟังก์ชันโลจิสติกไม่ใช่ความสัมพันธ์เชิงเส้นตรง มีจุดเด่นคือโมเดลการถดถอยโลจิสติกสามารถทดสอบอิทธิพลของปฏิสัมพันธ์ระหว่างระดับความสามารถกับการเป็นสมาชิกของกลุ่มผู้สอบจึงมีประสิทธิภาพตรวจสอบ DIF แบบเอกรูปและอนเอกรูป (Li & Stout, 1996, Gómez-Benito; Hidalgo & Padilla, 2009) เมื่อแทนความน่าจะเป็นในการตอบข้อสอบถูกต้องของผู้สอบและ Z แทนผลรวมเชิงเส้นของตัวแปรทำนาย โมเดลของการวิเคราะห์ ดังนี้

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

โมเดลของสมการถดถอยโลจิสติกในการวิเคราะห์ DIF ดังสมการ

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

เมื่อ X แทนระดับความสามารถของผู้สอบเป็นคะแนนสอบ G แทนสมาชิกของกลุ่มผู้สอบคือกลุ่มอ้างอิงหรือกลุ่มเปรียบเทียบ XG แทนปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับกลุ่มผู้สอบ β_0 แทนค่าจุดตัดแกน (Intercept) β_1 แทนพารามิเตอร์ความสามารถที่แตกต่างรายข้อ β_2 แทนพารามิเตอร์ความสามารถที่แตกต่างระหว่างกลุ่มรายข้อและ β_3 แทนพารามิเตอร์ปฏิสัมพันธ์ระหว่างกลุ่มกับความสามารถ

ขนาดอิทธิพล (Effect Size) ของการทำหน้าที่ต่างกันหรือความเข้มของการทำหน้าที่ต่างกันของข้อสอบ (Magnitude of DIF) บอกระดับความสัมพันธ์ของตัวแปรต้นกับตัวแปรตาม มีความสำคัญ 2 ประการ (Kim, Chosen, Alagoz & Kim, 2007) ประการแรก ขนาดอิทธิพลช่วยตอบคำถามเกี่ยวกับความสัมพันธ์ระหว่างตัวแปรที่ศึกษา เมื่อกลุ่มตัวอย่างมีขนาดใหญ่จะมีความไวในการตรวจสอบสูงมาก ประการที่สอง Zumbo & Hubley (1998 b) เสนอว่าควรศึกษาผลการวิเคราะห์ระหว่างการทดสอบระดับนัยสำคัญกับการวัดขนาดอิทธิพล Zumbo & Thomas (1997) กำหนดดัชนีทดสอบด้วยสถิติไค-สแควร์ (χ^2) ที่มี df=2 ใช้อัตราส่วนไลค์ลิฮูด (Likelihood Ratio Test) ด้วยวิธีถดถอยโลจิสติกรวมกับการวัดขนาดอิทธิพลของการทำหน้าที่ต่างกันของข้อสอบ การคำนวณ R^2 สำหรับวัดขนาดอิทธิพลใน DIF ที่มี

รูปแบบการให้คะแนนแบบพหุวิภาคมีเพียง R^2 สำหรับข้อมูลที่เป็นลำดับชั้น (R-squared for ordinal) การคำนวณ R^2 ในรูปแบบการให้คะแนนแบบพหุวิภาค มีดังนี้ (1) R^2 ของ Nagelkerke เป็นกลุ่มนามบัญญัติคำนวณง่ายสามารถคำนวณโดยใช้โปรแกรมสำเร็จรูป SPSS (2) R^2 ของ WLS (Weighted-least-squares Squared) เป็นกลุ่มนามบัญญัติ เหมาะสำหรับกลุ่มตัวอย่างที่มีตัวแปรเป็นลำดับชั้นภายใน (3) R^2 สำหรับข้อมูลที่เป็นลำดับชั้น (R-squared for ordinal) เป็นกลุ่มจัดลำดับ เหมาะกับตัวแปรที่มีความเป็นคุณลักษณะแฝงที่มีความต่อเนื่องประเภทพฤติกรรมศาสตร์และสังคมศาสตร์

การศึกษานี้ใช้ R^2 ตัดสินตามเกณฑ์ Zumbo & Thomas มีความเข้ม 3 ระดับ คือ ขนาดเล็กเมื่อค่าความแตกต่าง $R^2 < .13$ ขนาดปานกลางเมื่อค่าความแตกต่าง $.13 < R^2 < .26$ และขนาดใหญ่เมื่อค่าความแตกต่าง $R^2 > .26$ (Zumbo, 1999) และเกณฑ์ Jodoin & Gierl ขนาดเล็กเมื่อค่าความแตกต่าง $R^2 < .035$ ขนาดปานกลางเมื่อค่าความแตกต่าง $.035 < R^2 < .07$ และขนาดใหญ่เมื่อค่าความแตกต่าง $R^2 > .07$

ระเบียบวิธีวิจัย

ตัวแปรที่ศึกษา

1. ตัวแปรอิสระ (1) เกณฑ์ขนาดอิทธิพล เกณฑ์ Jodoin & Gierl และ Zumbo & Thomas (2) รูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูปและแบบอนเอกรูป (3) ขนาดของการทำหน้าที่ต่างกัน ขนาด 0.1, 0.2 และ 0.4 (4) จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10 และ 20 (5) ความยาวของแบบสอบทั้งฉบับ 40 และ 50 ข้อ

2. ตัวแปรตาม (1) อัตราความถูกต้อง คำนวณจากจำนวนของข้อสอบที่ตรวจสอบได้ถูกต้องว่าทำหน้าที่ต่างกัน ต่อจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งหมดในแบบสอบ คำนวณเป็นค่าร้อยละ (2) อัตราความคลาดเคลื่อนประเภทที่ 1 เป็นการระบุผิดพลาดว่าข้อสอบทำหน้าที่ต่างกัน ทั้งที่ความเป็นจริงข้อสอบทำหน้าที่ไม่ต่างกัน คำนวณจากสัดส่วนของจำนวนข้อสอบที่ตรวจสอบผิดพลาดว่าทำหน้าที่ต่างกันทั้งที่ในความเป็นจริงข้อสอบไม่ได้ทำหน้าที่ต่างกันคำนวณเป็นค่าร้อยละ

การวิเคราะห์ข้อมูล

1. วิเคราะห์ค่าพารามิเตอร์ใช้โมเดลภายใต้ทฤษฎีการตอบสนองข้อสอบด้วยโปรแกรม MULTILOG - MG พัฒนาขึ้นโดย David Thissen ที่มีคุณสมบัติในการประมาณค่าพารามิเตอร์ของข้อสอบและผู้สอบ ความเหมาะสำหรับการตอบข้อสอบที่มีการให้คะแนนแบบพหุวิภาค

2. วิเคราะห์คุณภาพข้อสอบด้วยโปรแกรม SPSS

3. วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก ใช้การเขียนคำสั่งและประมวลผลด้วยโปรแกรม R

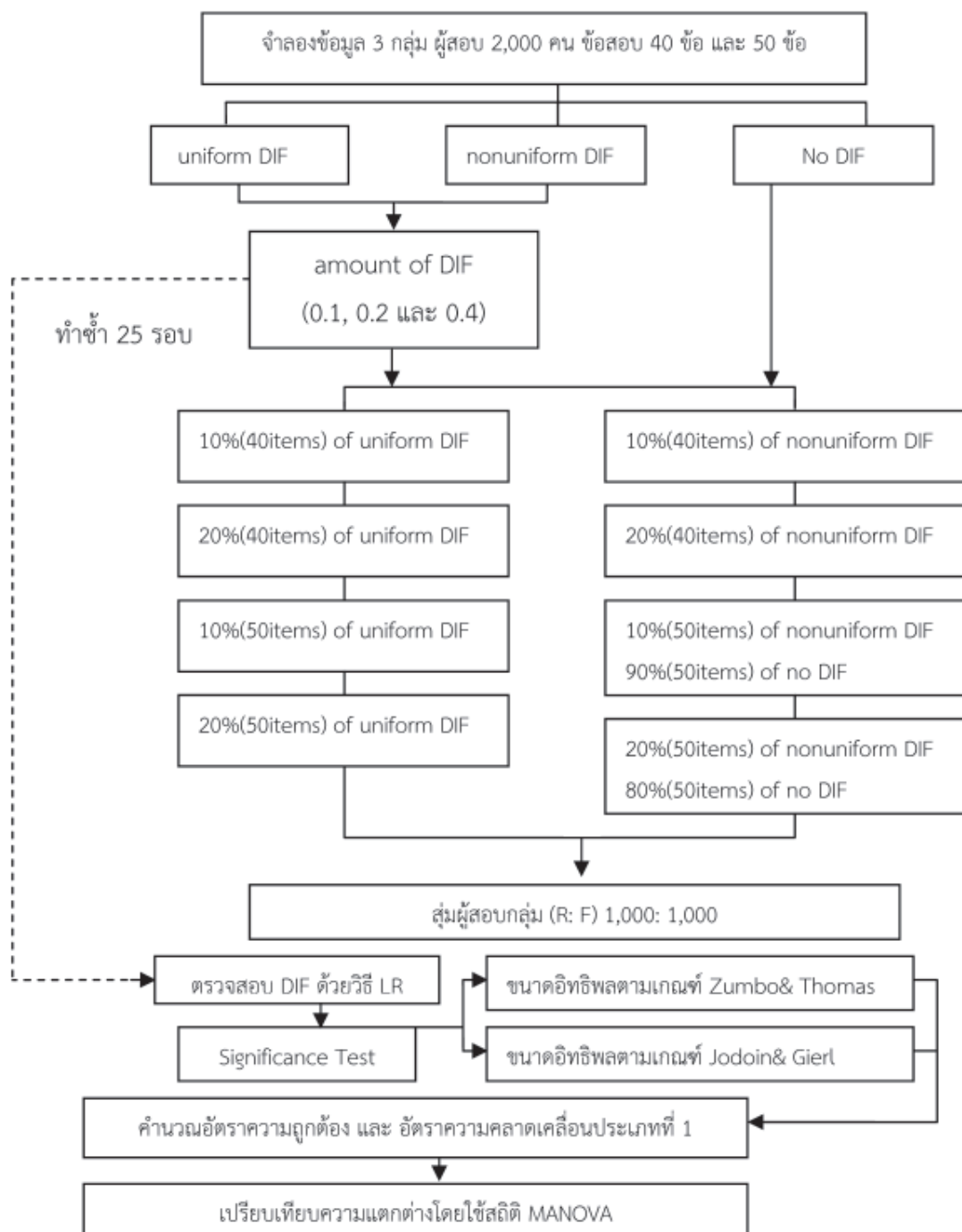
4. การเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ ตัวแปรอิสระ โดยสถิติการวิเคราะห์ความแปรปรวนหลายตัวแปร (MANOVA) ที่ระดับนัยสำคัญ .001 กำหนดการวิเคราะห์ให้มีปฏิสัมพันธ์ระหว่างตัวแปรอิสระไม่เกินอันดับที่สอง ถ้าผลการทดสอบมีนัยสำคัญทางสถิติจะทดสอบผลระหว่างกลุ่ม (Test of between-subjects effects) ของตัวแปรตามทีระดับนัยสำคัญ .001 ทดสอบผลย่อย (Simple effect) ภายใต้ตัวแปรที่ศึกษาระดับนัยสำคัญ .001 ทดสอบภายหลังด้วยวิธีของเชฟเฟ (Scheffé) โดยใช้ระดับนัยสำคัญระดับเดียวกับการทดสอบผลย่อย

ข้อมูลจำลอง (Simulation data) จำลองข้อมูลตามเงื่อนไขปัจจัยที่ศึกษาด้วยโปรแกรม WinGen (Windows software that generates IRT parameters and item responses) โดย Kyung T. Han (2007) จำลองข้อมูลเป็นกลุ่มข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป แบบอนเอกรูป และกลุ่มข้อสอบที่ไม่ทำหน้าที่ต่างกัน กำหนดขนาดกลุ่มตัวอย่าง สุ่มขนาดกลุ่มตัวอย่างกลุ่มอ้างอิงต่อกลุ่มเปรียบเทียบในสัดส่วน 1,000 : 1,000 ดำเนินการตรวจสอบความถูกต้องของข้อมูลว่าเป็นไปตามเงื่อนไขที่กำหนด ทุกเงื่อนไขทำซ้ำ 25 รอบ แสดงดังแผนภาพที่ 2

ผลการวิจัย

1. ค่าพารามิเตอร์ในทฤษฎีการตอบสนองข้อสอบคือพารามิเตอร์ของผู้สอบหรือคุณลักษณะของผู้สอบ (θ) ส่วนใหญ่มีค่าอยู่ในช่วง -3.0 ถึง +3.0

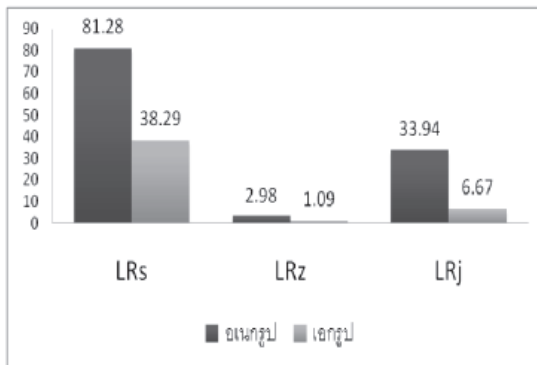
2. ประสิทธิภาพด้านอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 แสดงค่าเฉลี่ยของอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้วิธีถดถอยโลจิสติก ระหว่างการทดสอบระดับนัยสำคัญ (LR_S) ขนาดอิทธิพลตามเกณฑ์ Jodoin & Gierl (LR_J) และขนาดอิทธิพลตามเกณฑ์ Zumbo & Thomas (LR_Z) จำแนกตามปัจจัยที่ศึกษา ดังแผนภาพที่ 3-6



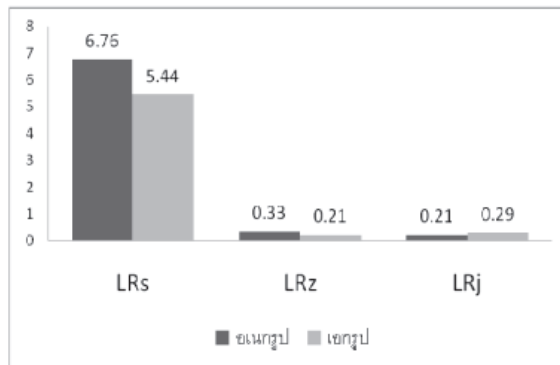
แผนภาพที่ 2 ขั้นตอนการจำลองข้อมูล

2.1) อัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน 2 เงื่อนไข

ก. ร้อยละของอัตราความถูกต้อง



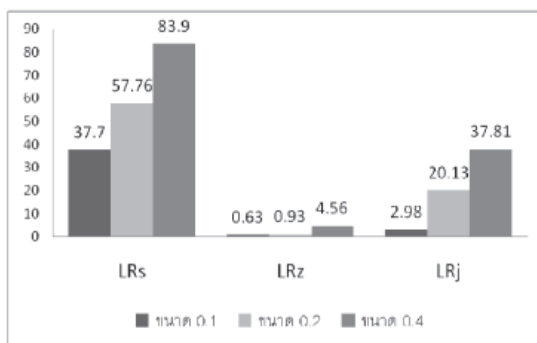
ข. ร้อยละของอัตราความคลาดเคลื่อนประเภทที่ 1



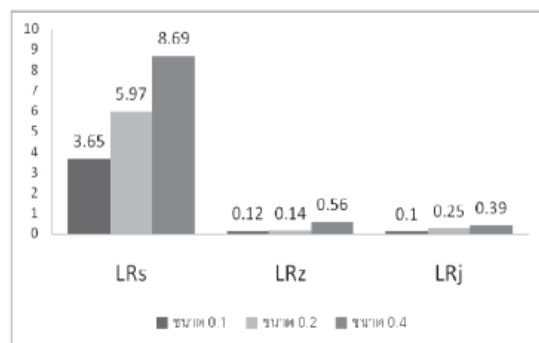
แผนภาพที่ 3 ภายใต้ปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน 2 เงื่อนไข

2.2) อัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ปัจจัยขนาดของการทำหน้าที่ต่างกัน 3 เงื่อนไข

ก. ร้อยละของอัตราความถูกต้อง



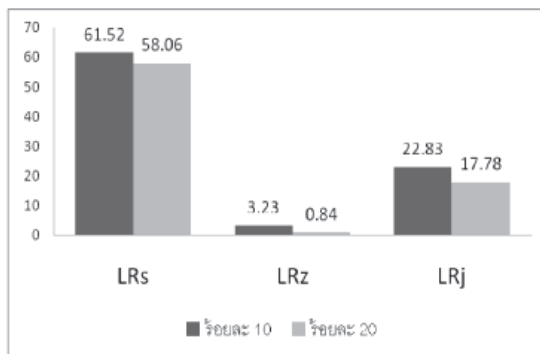
ข. ร้อยละของอัตราความคลาดเคลื่อนประเภทที่ 1



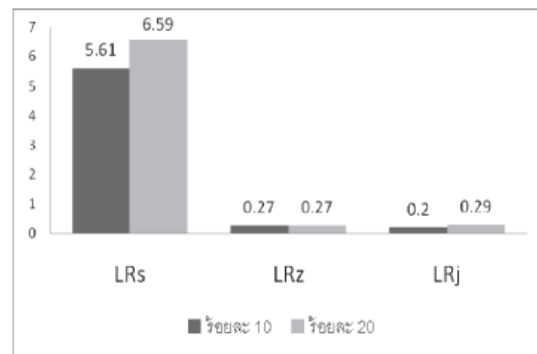
แผนภาพที่ 4 ภายใต้ปัจจัยขนาดของการทำหน้าที่ต่างกัน 3 เงื่อนไข

2.3) อัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 2 เงื่อนไข

ก. ร้อยละของอัตราความถูกต้อง



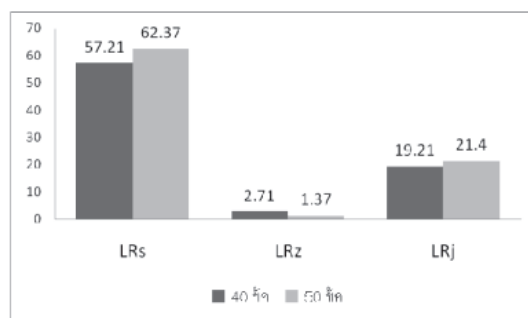
ข. ร้อยละของอัตราความคลาดเคลื่อนประเภทที่ 1



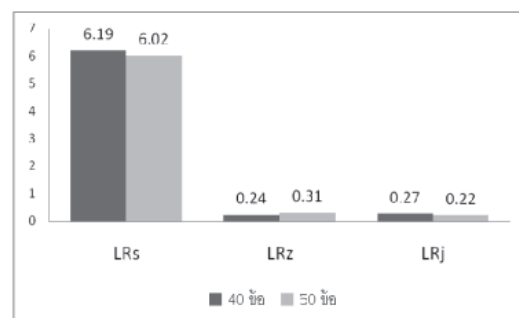
แผนภาพที่ 5 ภายใต้ปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 2 เงื่อนไข

2.4) อัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ภายใต้ปัจจัยความยาวของแบบสอบทั้งฉบับ 2 เงื่อนไข

ก. ร้อยละของอัตราความถูกต้อง



ข. ร้อยละของอัตราความคลาดเคลื่อนประเภทที่ 1



แผนภาพที่ 6 ภายใต้ปัจจัยความยาวของแบบสอบทั้งฉบับ 2 เงื่อนไข

สรุปผลการวิจัย

การเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ใช้สถิติการวิเคราะห์ความแปรปรวนพหุ (Multivariate analysis of variance; MANOVA) ที่ระดับนัยสำคัญ .001 ตัวแปรตามคือ อัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ตัวแปรอิสระคือวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยวิธีถดถอยโลจิสติกในสองเกณฑ์และปัจจัยที่แปรเปลี่ยน 4 ปัจจัย สรุปผลการเปรียบเทียบ แสดงดังตารางที่ 2 และ 3

ตารางที่ 2 สรุปผลการเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างการวัดขนาดอิทธิพลตามเกณฑ์สองเกณฑ์

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

ตารางที่ 2 (ต่อ)

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

จากตารางที่ 2 วิธีถดถอยโลจิสติกโดยการวัดขนาดอิทธิพลเกณฑ์ Jodoin & Gierl มีอัตราความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูงกว่าเกณฑ์ Zumbo & Thomas และภายใต้ปฏิสัมพันธ์ระหว่างปัจจัยที่แปรเปลี่ยนเกือบทุกเงื่อนไข

ตารางที่ 3 สรุปผลการเปรียบเทียบอัตราความถูกต้อง (CI) และอัตราความคลาดเคลื่อนประเภทที่ 1 (TE) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก ระหว่างการวัดชนิดอิทธิพลตามเกณฑ์ Jodoin and Gierl และเกณฑ์ Zumbo and Thomas ภายใต้เงื่อนไขต่างกัน

ปัจจัย		การวัดชนิดอิทธิพล 2 เกณฑ์			
		Zumbo and Thomas		Jodoin and Gierl	
		CI	TE	CI	TE
TYPE อนกรุป	เอกรุป	อนกรุปสูงกว่าเอกรุป	ไม่มีการเปรียบเทียบ	อนกรุปสูงกว่าเอกรุป	ไม่มีการเปรียบเทียบ
LENGTH 40 ข้อ	50 ข้อ	ไม่แตกต่างอย่างมีนัยสำคัญ	ไม่มีการเปรียบเทียบ	ไม่แตกต่างอย่างมีนัยสำคัญ	ไม่มีการเปรียบเทียบ
AMOUNT DIF0.1	DIF0.2 DIF0.4	0.2 สูงกว่า 0.1 0.4 สูงกว่า 0.1	ไม่มีการเปรียบเทียบ ไม่มีการเปรียบเทียบ	0.2 สูงกว่า 0.1 0.4 สูงกว่า 0.1	ไม่มีการเปรียบเทียบ ไม่มีการเปรียบเทียบ
TYPE * %DIF อนกรุป กับ 20%DIF เอกรุป กับ 10%DIF เอกรุป กับ 20%DIF	อนกรุป กับ 20%DIF เอกรุป กับ 10%DIF เอกรุป กับ 20%DIF	ไม่แตกต่างอย่างมีนัยสำคัญ ไม่แตกต่างอย่างมีนัยสำคัญ ไม่แตกต่างอย่างมีนัยสำคัญ	ไม่มีการเปรียบเทียบ ไม่มีการเปรียบเทียบ ไม่มีการเปรียบเทียบ	ไม่แตกต่างอย่างมีนัยสำคัญ ไม่แตกต่างอย่างมีนัยสำคัญ ไม่แตกต่างอย่างมีนัยสำคัญ	ไม่มีการเปรียบเทียบ ไม่มีการเปรียบเทียบ ไม่มีการเปรียบเทียบ
LENGTH * %DIF 40 ข้อ กับ 10% DIF 50 ข้อ กับ 10%DIF	40 ข้อ กับ 20%DIF 50 ข้อ กับ 10%DIF	ไม่มีการเปรียบเทียบ ไม่มีการเปรียบเทียบ	40 ข้อ กับ 10% DIF สูงกว่า 40 ข้อ กับ 20%DIF 40 ข้อ กับ 10% DIF สูงกว่า 50 ข้อ กับ 10%DIF	ไม่มีการเปรียบเทียบ ไม่มีการเปรียบเทียบ	40 ข้อ กับ 10% DIFต่ำกว่า 40 ข้อ กับ 20%DIF 40 ข้อ กับ 10% DIFต่ำกว่า 50 ข้อ กับ 10%DIF

ตารางที่ 3 (ต่อ)

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

ตารางที่ 3 (ต่อ)

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

จากตารางที่ 3 ข้อสอบที่ทำหน้าที่ต่างกันแบบอนุกรมมีอัตราความถูกต้องจากการวัดขนาดอิทธิพลทั้ง 2 เกณฑ์สูงกว่าแบบอนุกรม แบบสอบที่มีจำนวนข้อสอบทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 20 มีอัตราความถูกต้องจากการวัดขนาดอิทธิพลทั้ง 2 เกณฑ์สูงกว่าในแบบสอบที่มีจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10 และเมื่อขนาดอิทธิพลของข้อสอบที่ทำหน้าที่ต่างกันเพิ่มขึ้น มีผลทำให้อัตราความถูกต้องจากการวัดขนาดอิทธิพลทั้ง 2 เกณฑ์เพิ่มขึ้นภายใต้เกือบทุกเงื่อนไข

1. สรุปปัจจัยรูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน ร้อยละเฉลี่ย CI รูปแบบเอกรูปให้ค่าต่ำกว่าแบบอนุกรม เมื่อขนาดอิทธิพลของแบบเอกรูปมากกว่าร้อยละเฉลี่ย CI จะมีค่าสูงขึ้น ร้อยละเฉลี่ย CI แบบอนุกรมจะมีค่าสูงขึ้นเมื่อขนาดของการทำหน้าที่ต่างกันสูงขึ้นจาก 0.1 เป็น 0.4 เมื่อความยาวเพิ่มจาก 40 เป็น 50 ข้อ และแบบสอบที่มีข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10 ของทั้งฉบับ ร้อยละเฉลี่ย TE ของขนาดอิทธิพลเกณฑ์ Zumbo & Thomas ต่ำสุดในเงื่อนไขข้อสอบที่ทำหน้าที่ต่างกันแบบอนุกรม ขนาดอิทธิพลตามเกณฑ์ Jodoin & Gierl ต่ำสุดอยู่ในเงื่อนไขข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป

2. สรุปปัจจัยขนาดของการทำหน้าที่ต่างกัน ร้อยละเฉลี่ย CI เมื่อมีขนาดของการทำหน้าที่ต่างกันสูงขึ้นส่งผลให้อัตราความถูกต้องสูงขึ้นในทุกวิธีที่ศึกษา ค่า CI สูงสุดคือวิธีถดถอยโลจิสติกโดยการทดสอบระดับนัยสำคัญ รองลงมาคือขนาดอิทธิพลตามเกณฑ์ Jodoin & Gierl และ Zumbo & Thomas ตามลำดับ TE ที่คำนวณได้จากการตรวจสอบด้วยวิธีถดถอยโลจิสติกโดยการทดสอบระดับนัยสำคัญ ให้ค่าร้อยละสูงกว่าการวัดขนาดอิทธิพลทั้งสองเกณฑ์ในทุกระดับของขนาดของการทำหน้าที่ต่างกัน เมื่อพิจารณาขนาดอิทธิพลตามเกณฑ์ Zumbo & Thomas และ Jodoin & Gierl มี TE ต่ำที่สุดเมื่อขนาดของการทำหน้าที่ต่างกันเป็น 0.1 และสูงสุดเมื่อเป็น 0.4

3. สรุปปัจจัยจำนวนข้อสอบที่ทำหน้าที่ต่างกัน แบบสอบที่มีข้อสอบทำหน้าที่ต่างกันทั้งฉบับ ร้อยละ 10 ค่าร้อยละเฉลี่ยของ CI สูงร้อยละ 20 วิธีถดถอยโลจิสติกโดยการทดสอบระดับนัยสำคัญและการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo & Thomas และ Jodoin & Gierl ให้ผลสอดคล้องตรงกัน TE ที่คำนวณได้จากการตรวจสอบด้วยวิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญมีค่าสูงกว่าขนาดอิทธิพลจากการวัดขนาดอิทธิพลทั้งสองเกณฑ์ TE ตามเกณฑ์ Zumbo & Thomas มีค่าต่ำสุดในเงื่อนไขข้อสอบที่ทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 20 สำหรับขนาดอิทธิพลตามเกณฑ์ Jodoin & Gierl มีค่าต่ำสุดในเงื่อนไขข้อสอบทำหน้าที่ต่างกันทั้งฉบับคิดเป็นร้อยละ 10

4. สรุปปัจจัยความยาวของแบบสอบทั้งฉบับ มีนัยสำคัญทางสถิติในกรณีเดียวเท่านั้นเมื่อข้อสอบมีจำนวนข้อเพิ่มขึ้น ผลจากวิธีตรวจสอบภายใต้วิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญกับการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin & Gierl ให้ผลตรงกันคือมีร้อยละเฉลี่ยของ CI สูงขึ้น ร้อยละเฉลี่ยของ TE ในแบบสอบที่มีความยาว 40 ข้อ สูงกว่าความยาว 50 ข้อ

5. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบตามวิธีการตรวจสอบที่ศึกษา

5.1) การทดสอบระดับนัยสำคัญ ข้อสอบทำหน้าที่ต่างกันแบบอนุกรมมีอัตราความถูกต้องสูงกว่าแบบเอกรูป การทำหน้าที่ต่างกันขนาดใหญ่กว่าจะพบอัตราความถูกต้องสูงกว่าการทำหน้าที่ต่างกันที่มีขนาดเล็ก หากในแบบสอบมีจำนวนข้อที่ทำหน้าที่ต่างกันหลายข้อจะส่งผลให้อัตราความถูกต้องลดลง

เมื่อข้อสอบทำหน้าที่ต่างกันทั้งแบบเอกรูปและแบบอนกรูป ในการทำหน้าที่ต่างกันที่มีขนาดใหญ่จะพบ TE สูงกว่าขนาดเล็กในทุกเงื่อนไข

5.2) ขนาดอิทธิพลตามเกณฑ์ Zumbo & Thomas การตรวจสอบการทำหน้าที่ต่างกันด้วยวิธีถดถอยโลจิสติก ตามเกณฑ์ Zumbo & Thomas มีผลของอัตราความถูกต้องสอดคล้องกับผลที่ได้จากการตรวจสอบโดยการทดสอบระดับนัยสำคัญ

5.3) ขนาดอิทธิพลตามเกณฑ์ Jodoin & Gierl ข้อสอบทำหน้าที่ต่างกันแบบเอกรูปการตรวจสอบการทำหน้าที่ต่างกันด้วยวิธีถดถอยโลจิสติกตามเกณฑ์ Jodoin & Gierl จะไม่สามารถตรวจพบข้อสอบที่ทำหน้าที่ต่างกันกรณีการทำหน้าที่ต่างกันมีขนาดเล็ก เมื่อการทำหน้าที่ต่างกันมีขนาดใหญ่อัตราความถูกต้องจะสูงขึ้น เมื่อข้อสอบทำหน้าที่ต่างกันทั้งแบบเอกรูปและแบบอนกรูปการทำหน้าที่ต่างกันที่มีขนาดเล็กจะไม่เกิด TE ส่วนร้อยละเฉลี่ยสูงสุดของ TE ภายใต้เงื่อนไขรูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบอนกรูปที่มีความยาวของแบบสอบทั้งฉบับ 40 ข้อ

อภิปรายผลการวิจัย

1. ประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้วิธีถดถอยโลจิสติก ระหว่างการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin & Gierl และ Zumbo & Thomas ภายใต้สถานการณ์จำลองตามปัจจัยที่แปรเปลี่ยนและศึกษาปฏิสัมพันธ์สองทางระหว่างปัจจัยที่แปรเปลี่ยน 4 ปัจจัย การวัดขนาดอิทธิพลตามเกณฑ์ Jodoin & Gierl มีอัตราความถูกต้องสูงกว่าเกณฑ์ Zumbo & Thomas เกือบทุกเงื่อนไข ดังนี้ (1.1) รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน และขนาดของการทำหน้าที่ต่างกัน มีผลต่ออัตราความถูกต้องของการวัดขนาดอิทธิพลตามเกณฑ์ทั้งสองเกณฑ์ (1.2) ความยาวของแบบสอบทั้งฉบับ ไม่มีผลต่ออัตราความถูกต้องและ TE เนื่องจากกำหนดความยาวของแบบสอบใกล้เคียงกันมากระหว่าง 40 กับ 50 ข้อ จึงไม่มีผลของความแตกต่าง (1.3) ปฏิสัมพันธ์สองทางของปัจจัยที่แปรเปลี่ยนระหว่างรูปแบบของข้อสอบที่ทำหน้าที่ต่างกับขนาดของการทำหน้าที่ต่างกัน ปฏิสัมพันธ์สองทางระหว่างรูปแบบของข้อสอบที่ทำหน้าที่ต่างกับจำนวนข้อสอบที่ทำหน้าที่ต่างกัน และปฏิสัมพันธ์สองทางระหว่างจำนวนข้อสอบที่ทำหน้าที่ต่างกับขนาดของการทำหน้าที่ต่างกัน มีผลต่ออัตราความถูกต้องของการวัดขนาดอิทธิพลตามเกณฑ์ทั้งสองเกณฑ์

2. ประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของวิธีถดถอยโลจิสติก โดยการวัดขนาดอิทธิพลของการทำหน้าที่ต่างกัน ภายใต้ปัจจัยที่ศึกษา พบว่า (2.1) รูปแบบของข้อสอบที่ทำหน้าที่ต่างกันแบบอนกรูปมีอัตราความถูกต้องสูงกว่าแบบเอกรูป รูปแบบอนกรูปตามเกณฑ์ Jodoin & Gierl ให้ค่า TE ต่ำกว่าเกณฑ์ Zumbo & Thomas (2.2) ขนาดของการทำหน้าที่ต่างกัน มีผลต่อประสิทธิภาพการตรวจสอบระหว่างการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin & Gierl และเกณฑ์ Zumbo & Thomas (2.3) จำนวนข้อสอบที่ทำหน้าที่ต่างกันและความยาวของแบบสอบทั้งฉบับ มีผลต่อประสิทธิภาพการตรวจสอบขนาดอิทธิพลทั้งสองเกณฑ์ อย่างมีนัยสำคัญทางสถิติที่ระดับ .001

ข้อเสนอแนะจากงานวิจัย

ขนาดอิทธิพลตามเกณฑ์ Jodoin & Gierl มีประสิทธิภาพดีกว่าเกณฑ์ Zumbo & Thomas จึงควรเลือกใช้การวัดขนาดอิทธิพลตามเกณฑ์ Jodoin & Gierl เพื่อได้สารสนเทศเกี่ยวกับขนาดของการทำหน้าที่ต่างกันของข้อสอบที่เหมาะสม (ศิริรัตน์ สุคันธพฤกษ์, 2554) สำหรับด้านการได้ประโยชน์ของหน่วยงานที่จัดสอบ รูปแบบการสอบที่มีความยากสูงมีอำนาจจำแนกที่ดีการตัดสินใจโดยเกณฑ์ Zumbo & Thomas กลับเป็นผลดีต่อหน่วยงานเนื่องจากตรวจไม่พบข้อสอบที่ทำหน้าที่ต่างกัน ความเป็นจริงสิ่งที่เราต้องการคือการไม่เกิดการทำหน้าที่ต่างกันของข้อสอบแต่เมื่อตรวจพบก็เป็นหน้าที่ของนักการศึกษาที่จะดำเนินการบางอย่างเกี่ยวกับการปรับปรุงเพื่อให้ข้อสอบมีคุณภาพ นอกจากนี้ นักวิจัยควรพิจารณาเงื่อนไขขนาดของกลุ่มตัวอย่าง ความแตกต่างเกี่ยวกับการแจกแจงความสามารถของผู้สอบ ความยากของข้อสอบ และเงื่อนไขเกี่ยวกับความยาวของแบบสอบเพื่อให้เกิดความเหมาะสมและรัดกุม เมื่อต้องการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบต้องไม่เน้นเพียงเพื่อตัดสินใจ ตัดข้อสอบออกไปจากแบบสอบเท่านั้นแต่ควรหาสารสนเทศอื่นมาประกอบการตัดสินใจอย่างรัดกุม เนื่องจากการสอบแต่ละครั้งโดยเฉพาะการสอบระดับชาติเป็นเรื่องที่ต้องมีการลงทุนสูงในหลายด้านและใช้ทุนทรัพย์ค่อนข้างสูง การออกแบบบริหารจัดการการสอบเป็นเรื่องที่ต้องตั้งอยู่บนหลักของความคุ้มค่า เกิดประสิทธิผลและมีประสิทธิภาพสูงสุด ดังนั้นนักศึกษาต้องอาศัยหลักฐานที่มีน้ำหนักเพียงพอ วิธีการที่น่าสนใจคืออาศัยความชำนาญด้านวัดผลประเมินผลของผู้เชี่ยวชาญมาตัดสินการทำหน้าที่ต่างกันของข้อสอบรวมกับการใช้หลักฐานในทางสถิติ เพื่อตรวจสอบความถูกต้องของผลการตัดสินใจให้เกิดความน่าเชื่อถือและมีความยุติธรรมมากที่สุดในการตัดสินใจข้อสอบ

เอกสารอ้างอิง

- ศิริชัย กาญจนวาสี. (2550). *ทฤษฎีการทดสอบแนวใหม่ (MODERN TEST THEORIES)*. พิมพ์ครั้งที่ 3. กรุงเทพมหานคร. โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- ศิริรัตน์ สุคันธพฤกษ์. (2554). การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบวัดความวิตกกังวลในการสอบคณิตศาสตร์ : การเปรียบเทียบระหว่างไฮราซิคอลลีเนียร์โมเดล พาเชียลเครดิตโมเดล. วารสารวิจัย มสค สาขามนุษยศาสตร์และสังคมศาสตร์. 7(1), 113-128.
- Elizabeth Kristjansson, et al. (2005). "A comparison of four methods for detecting differential item functioning in ordered response items". *Journal of Educational and Psychological Measurement*, 65(6): 935-953.
- Gómez-Benito, J.Hidalgo, M.D. & Padilla, J.L.,(2009). Efficacy of Effect Size Measures in Logistic Regression An Application for Detecting DIF. *Methodology : European Journal of Research Methods for the Behavioral and Social Sciences*. Vol 5(1),18-25.
- Kim, S.H., Chosen, A.S. & Kim, S. (2007). DIF Detection and Effect Size Measures for Polytomously Scored Items. *Journal of Educational Measurement Summer*, Vol. 44, No. 2, pp. 93-116.
- Kyung T. H. (2007). User's Manual for WinGen : Windows Software that Generates IRT Model Parameters and Item Responses. Center for Educational Assessment Research Report No. 642. Amherst, MA : University of Massachusetts, Centerfor Educational Assessment.
- Li, H. & Stout, W. (1996). A new procedure for detecting crossing DIF. *Psychometrika*. 61(4): 647-677.
- Shumacker, R.E. (2005). <http://www.appliedmeasurementassociates.com/White%20Papers/TEST%20BIAS%20AND%20DIFFERENTIAL%20ITEM%20FUNCTIONING.pdf>. [วันที่ 11 มิถุนายน 2551]
- Zumbo, B.D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type item scores*. Ottawa, Canada : Directorate of Human Resources Research and Evaluation, Department of National Defense. Retrieved from <http://www.edu.ubc.ca/faculty/zumbo/DIF/index.html>

- Zumbo, B.D., & Hubley, A.M. (1998 b). A note on misconceptions concerning prospective and retrospective power. *Journal of the Royal Statistical Society, Series D : The Statistician*, 47, 385-388.
- Zumbo, B.D., & Thomas, D.R. (1997). *A measure of effect size for a model-based approach for studying DIF*. Prince George, Canada : University of Northern British Columbia, Edgeworht Laboratory for Quantitative Behavioral Science.

คณะผู้เขียน

ดร. ธเกียรติกมล ทองงอก

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
e-mail: thongngok@hotmail.com

รองศาสตราจารย์ ดร. โชติกา ภาษีผล

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
e-mail: aimornj@hotmail.com

ศาสตราจารย์ ดร. ศิริชัย กาญจนวาสี

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
e-mail: skanjanawasee@hotmail.com