

การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบที่มีเทสต์เลต ของแบบทดสอบระดับชาติ ของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้วิธี BAYESIAN

The Analysis Differential Item Functioning by BAYESIAN TESTLET Model of National Test Scores for Grade Sixth Students.

วาราสินี อยู่สุนทร¹ รศ.ดร.ไพรัตน์ วงษ์นาม² ผศ.ดร.สมพงษ์ ปั่นหุ่น³

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ 1) เพื่อเปรียบเทียบผลการประมาณค่าพารามิเตอร์ของข้อสอบ (อำนาจจำแนก ความยาก และโอกาสการเดา) และพารามิเตอร์ความสามารถผู้สอบด้วยวิธีของเบส์ และวิธีของเบส์แบบมีอิทธิพลเทสต์เลต 2) เพื่อเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธีของเบส์ (Bayes) และวิธีของเบส์แบบมีอิทธิพลเทสต์เลต (Bayes_y) จำแนกตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน 3) เพื่อศึกษาความสัมพันธ์ของค่าพารามิเตอร์ของข้อสอบ พารามิเตอร์ความสามารถผู้สอบ และขนาดของการทำหน้าที่ต่างกัน ระหว่างวิธีของเบส์และวิธีของเบส์แบบมีอิทธิพลเทสต์เลต

ข้อมูลที่ใช้ในการวิเคราะห์ครั้งนี้ เป็นคะแนนการสอบวัดผลสัมฤทธิ์ทางการเรียน เพื่อประเมินคุณภาพการศึกษาระดับชาติ ของนักเรียนชั้นประถมศึกษาปีที่ 6 ของสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) ในรายวิชาภาษาอังกฤษ ซึ่งได้มาจากการสุ่มแบบหลายขั้นตอน (Multi-stage Random Sampling Technique) จำนวน 1,000 คน จำแนกเป็นชายและหญิง ที่อยู่ในเขตกรุงเทพมหานครและปริมณฑลและนอกเขตกรุงเทพมหานครและปริมณฑล ในการวิเคราะห์ค่าพารามิเตอร์และการทำหน้าที่ต่างกันของข้อสอบ

ผลการวิจัยสรุปได้ ดังนี้

1. ผลการประมาณค่าพารามิเตอร์ระหว่างวิธีของเบส์ และวิธีของเบส์แบบมีอิทธิพลเทสต์เลต พบว่าค่าพารามิเตอร์อำนาจจำแนกที่ประมาณด้วยวิธีของเบส์แบบมีอิทธิพลเทสต์เลตประมาณค่าได้ดีกว่า ส่วนค่าพารามิเตอร์ความยาก พารามิเตอร์โอกาสการเดา และพารามิเตอร์ความสามารถของผู้สอบพบว่า ทั้งสองวิธีประมาณค่าได้ไม่แตกต่างกันมากนัก

2. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ พบว่าวิธีของเบส์และวิธีของเบส์แบบมีอิทธิพลเทสต์เลต ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ได้ไม่แตกต่างกันมากนัก

3. ผลการศึกษาความสัมพันธ์ผลการวิเคราะห์ระหว่างวิธีของเบส์และวิธีของเบส์แบบมีอิทธิพลเทสต์เลต ที่ระดับนัยสำคัญ .01 ค่าสัมประสิทธิ์สหสัมพันธ์ของค่าพารามิเตอร์ของข้อสอบ (อำนาจจำแนก ความยาก และโอกาสการเดา) พารามิเตอร์ความสามารถผู้สอบ ขนาดของการทำหน้าที่ต่างกันจำแนกตามเพศ และขนาดของการทำหน้าที่

¹ นิสิตหลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิจัย วัดผลและสถิติการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา อีเมลล์

wiw_warasinee@hotmail.com

² อาจารย์ ภาควิชาและจิตวิทยาประยุกต์ คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา

³ อาจารย์ ภาควิชาและจิตวิทยาประยุกต์ คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา

ต่างกันสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน มีความสัมพันธ์กันสูงมีค่า .579, .998, .912, .996, .978 และ .983 ตามลำดับ

คำสำคัญ: การทำหน้าที่ต่างกันของข้อสอบ, เทสต์เลท, วิธีของเบส์

Abstract

This study aimed to: 1) compare the item parameters and person parameter among Bayesian and Bayesian TESTLET MODEL; 2) compare the differential item functioning (DIF) on gender and school locations variables with which estimate by the Bayesian (Bayes) and Bayesian TESTLET MODEL (Bayes_γ); and 3) analyze the correlation among item parameters, person parameters, DIF on gender and school location which estimating by Bayesian and Bayesian TESTLET MODEL.

The national test scores of grade sixth students in 2016 academic year were drawn from National Institute of Educational Testing Service (Public Organization) by using the multi-stage random sampling technique. All one thousand cases of test scores were divided into two groups along gender (male vs female) and location (Bangkok and Metropolitan areas vs non Bangkok and Metropolitan areas) variables

The research results were as follows:

1. The research findings revealed that the discrimination parameter showed that the Bayesian Testlet Model (Bayesian_γ) was the best estimator; the difficulty parameter, the guessing parameter and the ability level showed that the Bayesian Testlet Model (Bayesian_γ), it is same.

2. The research findings revealed that the detection of Differential Item Functioning of the Bayesian (Bayes) and Bayesian TESTLET MODEL (Bayes_γ), it is the same.

3. Estimated parameter of the item parameters, person parameter, DIF on gender and school location of the Bayesian (Bayes) and Bayesian TESTLET MODEL (Bayes_γ) were significantly correlated in all aspects ($p < .01$), with high correlation (.579, .998, .912, .996, .978 and .983 respectively).

Keywords: Differential Item Functioning, TESTLET, BAYESIAN Method

ความสำคัญและที่มาของปัญหาวิจัย

การทดสอบในแต่ละครั้ง อาจเกิดความลำเอียงของแบบทดสอบซึ่งเกิดจากลักษณะของผู้สอบที่แตกต่างกัน เช่น เชื้อชาติ ศาสนา วัฒนธรรม สังคม เพศ ภาษา อายุ และประสบการณ์ เป็นต้น ปัจจัยข้างต้นอาจส่งผล ทำให้เกิดการได้เปรียบเสียเปรียบระหว่างผู้เข้าสอบแต่ละกลุ่ม ทำให้คะแนนที่ได้เอนเอียงไปทางกลุ่มใดกลุ่มหนึ่ง สาเหตุดังกล่าวอาจทำให้การประเมินคุณภาพไม่ตรงกับเป้าหมายที่ตั้งไว้ ซึ่งในการหาคุณภาพของแบบสอบที่วิเคราะห์ค่าพารามิเตอร์ของข้อสอบสิ่งที่จะต้องคำนึงถึงในการพัฒนาแบบทดสอบ คือ แบบทดสอบต้องมีความยุติธรรมกับผู้สอบ

ข้อสอบเกิดการทำหน้าที่ต่างกันของข้อสอบ อาจเป็นเพราะประสบการณ์ ความคุ้นเคยเกี่ยวกับเรื่องนั้น สภาพแวดล้อม และการฝึกปฏิบัติที่แตกต่างกันระหว่างนักเรียนในเขตกรุงเทพมหานครและปริมณฑลนักเรียนนอกเขต กรุงเทพมหานครและปริมณฑล อริสหา เตห์ลิม (2559) ศึกษาประสิทธิภาพการประมาณค่าพารามิเตอร์และการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีแมกซิมไลซ์ไลค์ลิตู๊ด (ML) วิธีของเบส์ (Bayes) และวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayesy) พบว่าวิธีของเบส์แบบมีอิทธิพลของทดสอบ (Bayesy) ประมาณค่าพารามิเตอร์ได้ดีที่สุดเมื่อการแจกแจงข้อมูลเป็นแบบปกติ นอกจากนี้วิธีของเบส์ (Bayes) และวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayesy) สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ดี

จากผลข้างต้นผู้วิจัยจึงสนใจที่จะศึกษาการประมาณค่าพารามิเตอร์ และวิเคราะห์ผลการทำหน้าที่ต่างกันของข้อสอบโดยวิธีของเบส์ทั้งในแบบที่คำนึงถึงอิทธิพลของทดสอบและไม่คำนึงถึงอิทธิพลของทดสอบ รวมไปถึงศึกษาความสัมพันธ์ระหว่างพารามิเตอร์ของข้อสอบกับขนาดของการทำหน้าที่ต่างกันของข้อสอบ จากแบบทดสอบที่มีทดสอบ (testlet) เพื่อเป็นการขยายความรู้ในเชิงทฤษฎีในการประมาณค่าพารามิเตอร์และการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ให้ตรงกับข้อตกลงเบื้องต้นของ สำหรับการวิเคราะห์ข้อสอบตามแนวทางทฤษฎีการตอบสนองข้อสอบ การศึกษาข้างต้นจึงเป็นการสะท้อนให้เห็นถึงคุณภาพของแบบทดสอบ เพื่อได้ข้อค้นพบที่เป็นประโยชน์ต่อผู้ที่นำวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบไปใช้ในการตรวจสอบคุณภาพของแบบทดสอบ ในการตัดสินใจว่าสมควรตัดข้อสอบที่พบว่ามีการทำหน้าที่ที่ต่างออกไปจากแบบทดสอบหรือควรปรับปรุงข้อสอบข้อนั้นเสียใหม่ เพื่อรักษาโครงสร้าง และคุณภาพของแบบทดสอบฉบับนั้นไว้ ซึ่งจะ使得แบบทดสอบมีคุณภาพที่ดียิ่งขึ้น

วัตถุประสงค์ของการวิจัย

1. เพื่อเปรียบเทียบผลการประมาณค่าพารามิเตอร์ของข้อสอบ (อำนาจจำแนก ความยาก และโอกาสการเดา) และพารามิเตอร์ความสามารถผู้สอบด้วยวิธีของเบส์ (Bayes) และวิธีของเบส์แบบมีอิทธิพลของทดสอบ (Bayesy)
2. เพื่อเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีของเบส์ (Bayes) และวิธีของเบส์แบบมีอิทธิพลของทดสอบ (Bayesy) จำแนกตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน
3. เพื่อศึกษาความสัมพันธ์ของค่าพารามิเตอร์ของข้อสอบ พารามิเตอร์ความสามารถผู้สอบ และขนาดของการทำหน้าที่ต่างกัน ระหว่างวิธีของเบส์ (Bayes) และวิธีของเบส์แบบมีอิทธิพลของทดสอบ (Bayesy)

วิธีดำเนินการวิจัย

ประชากรและกลุ่มตัวอย่าง

ประชากรการวิจัยครั้งนี้ เป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ในสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ ที่เข้าสอบเพื่อประเมินคุณภาพการศึกษาระดับชาติ ปีการศึกษา 2559 ซึ่งเป็นนักเรียนที่เข้าสอบวิชาภาษาอังกฤษ จำนวนทั้งหมด 716,771 คน การเลือกกลุ่มตัวอย่างผู้วิจัยได้กำหนดขนาดของกลุ่มตัวอย่างจากการประชากรตามการวิเคราะห์คุณภาพตามทฤษฎีทางทดสอบการตอบสนองของข้อสอบ (IRT) และตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ตามเงื่อนไขกลุ่มตัวอย่างไม่ต่ำกว่า 10-20 เท่าของจำนวนข้อคำถาม (นงลักษณ์ วิรัชชัย, 2542, หน้า 311) กลุ่มตัวอย่างในการวิจัยครั้งนี้ใช้ 1,000 คน จำแนกเป็นเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน โดยมีขั้นตอนการสุ่มตัวอย่างแบบหลายขั้นตอนเพื่อให้ได้กลุ่มตัวอย่าง 1,000 คน ที่แยกเป็น

เพศชาย 500 คน โดยแบ่งตามสถานที่ตั้งทางภูมิศาสตร์คือ ในเขตกรุงเทพฯและปริมณฑล 250 คน นอกเขตกรุงเทพฯและปริมณฑล 250 คน และ เพศหญิง 500 คน โดยแบ่งตามสถานที่ตั้งของโรงเรียนคือ ในเขตกรุงเทพฯ และปริมณฑล 250 คน กับนอกเขตกรุงเทพฯ และปริมณฑล 250 คน

เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยครั้งนี้ เป็นแบบวัดผลสัมฤทธิ์ทางการเรียนเพื่อประเมินคุณภาพการศึกษาระดับชาติ ของนักเรียนชั้นประถมศึกษาปีที่ 6 วิชาภาษาอังกฤษ ปีการศึกษา 2559 ปรนัยจำนวน 40 ข้อ สร้างโดยผู้เชี่ยวชาญจากสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน)

การเก็บข้อมูล

ผู้วิจัยได้รวบรวมข้อมูลผลการตอบของนักเรียนที่เป็นกลุ่มตัวอย่าง ซึ่งเป็นข้อมูลทุติยภูมิจากผลการตอบของนักเรียนชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2559 วิชาภาษาอังกฤษ จากสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) การดำเนินการแยกตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน แล้วนำข้อมูลมาวิเคราะห์ตามวัตถุประสงค์ของการวิจัย

การวิเคราะห์ข้อมูล

ผู้วิจัยแบ่งการวิเคราะห์ออกเป็น 4 ขั้นตอน คือ

1. คำนวณค่าสถิติพื้นฐาน ได้แก่ค่าเฉลี่ย (Mean) ค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ค่าความเบ้ (Skewness) ค่าความโด่ง (Kurtosis) เพื่อบรรยายลักษณะการแจกแจงของคะแนนการสอบวิชาภาษาอังกฤษโดยใช้โปรแกรม SPSS

2. วิเคราะห์โดยประยุกต์ใช้โปรแกรม WinBUGS14 จากข้อมูลผลการตอบข้อสอบมีการให้คะแนนแบบ 2 ค่า (Dichotomous) ในการวิเคราะห์การประมาณค่าพารามิเตอร์อำนาจจำแนกศึกษาจากค่า $a[i]$ ค่าพารามิเตอร์ความยากศึกษาจากค่า $b[i]$ ค่าพารามิเตอร์โอกาสการเดาศึกษาจากค่า $c[i]$ และพารามิเตอร์ความสามารถผู้สอบศึกษาจากค่า $z[i]$ ดังสมการโลจิสติก 3 PL ด้วยวิธีของเบส์ (Bayes) และวิธีของเบส์แบบมี อิทธิพลทดสอบ (Bayes_y) ต่อไปนี้

สมการวิธีของเบส์แบบมีอิทธิพลทดสอบ

$$p[j, i] = (c[i] + (1 - c[i])) * (1 / (1 + \exp(-a[i] * (z[j] - b[i] - \gamma[j, i]))))$$

ผู้วิจัยได้แสดงผลการวิเคราะห์ข้อมูลในช่วง 5001-2500 เนื่องจากความถี่เริ่มคงที่ตั้งแต่ 5001 ขึ้นไป โดยการเบริน (Burn) ข้อมูลก่อนหน้า 5001 (Fukuhara, 2009)

3. วิเคราะห์ โดยประยุกต์ใช้โปรแกรม WinBUGS14 ในกรณีนี้กำหนดให้ตัวแปรเพศชาย = 1 (Reference Group) และหญิง = 0 (Focal Group) และกำหนดให้ตัวแปรสถานที่ตั้งอยู่ในเขตกรุงเทพฯ และปริมณฑล = 1 (Reference Group) และนอกเขตกรุงเทพฯ และปริมณฑล = 0 (Focal Group) ซึ่งในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ศึกษาได้จากค่า $\beta_{1[i]}$ ดังสมการโลจิสติก 3 PL ด้วยวิธีของเบส์ (Bayes) และวิธีของเบส์แบบมี อิทธิพลทดสอบ (Bayes_y) ต่อไปนี้

สมการวิธีของเบส์ $p[j, i] = (c[i] + (1 - c[i])) * (1 / (1 + \exp(-a[i] * (z[j] - b[i] - (\beta_{1[i]} * \text{Group}[j]))))))$

สมการวิธีของเบส์แบบมี อิทธิพลทดสอบ

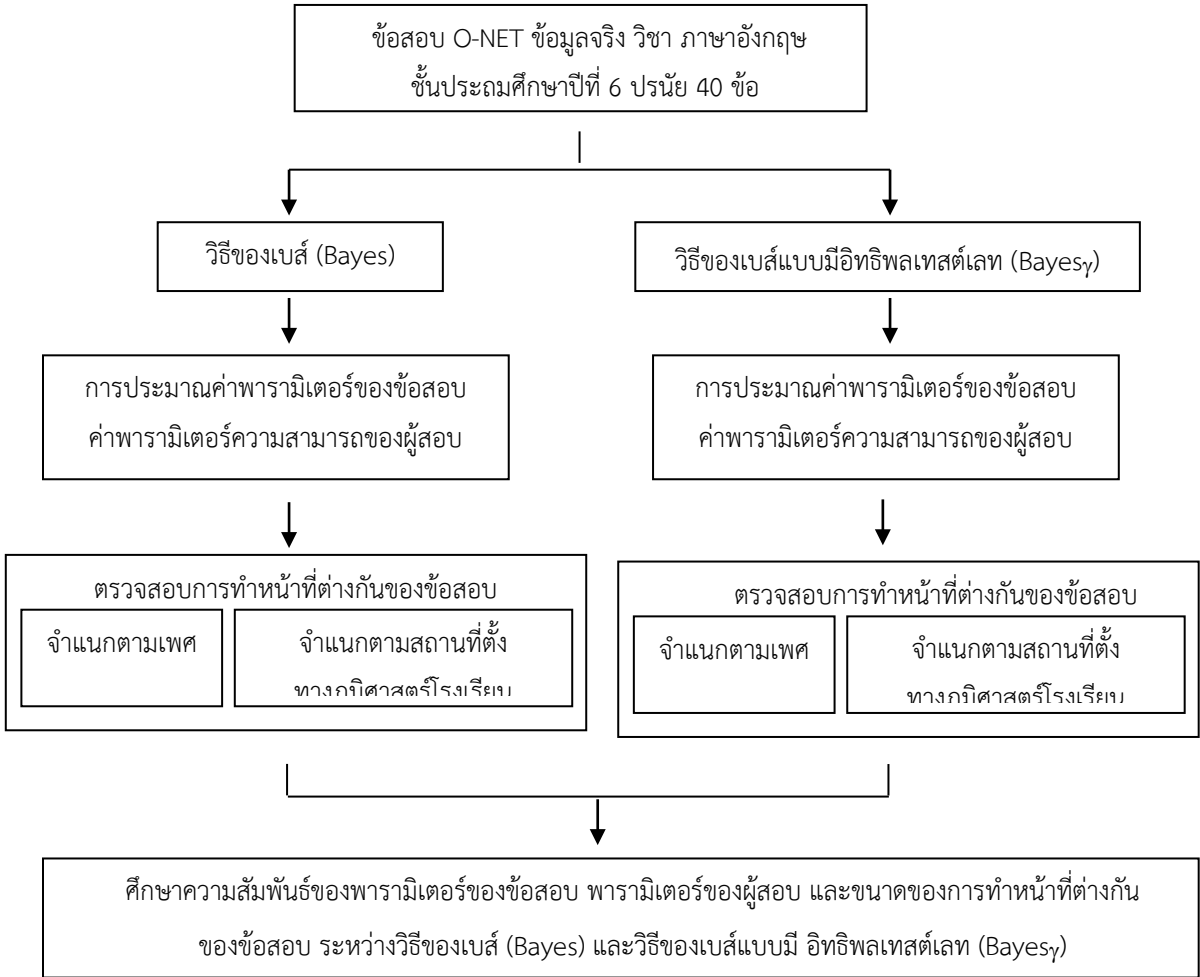
$$p[j, i] = (c[i] + (1 - c[i])) * (1 / (1 + \exp(-a[i] * (z[j] - b[i] - \gamma[j, i] - (\beta_{1[i]} * \text{Group}[j]))))))$$

ผู้วิจัยได้แสดงผลการวิเคราะห์ข้อมูลในช่วง 5001-2500 เนื่องจากความถี่เริ่มคงที่ตั้งแต่ 5001 ขึ้นไป โดยการเบริน (Burn) ข้อมูลก่อนหน้า 5001 (Fukuhara, 2009) และการพิจารณาถ้าค่า $\text{val}2.5\text{pc}$ และ $\text{val}97.5\text{pc}$ ไม่

กลุ่ม 0 และค่าสัมบูรณ์ของขนาด DIF มากกว่า .426 แสดงว่าข้อสอบข้อนี้ จะเกิดการทำหน้าที่ที่ต่างกันของข้อสอบอยู่ในระดับมาก (Vaughn, 2006)

4. วิเคราะห์ความสัมพันธ์ของค่าพารามิเตอร์ของข้อสอบ (ความยาก อำนาจจำแนกและโอกาสการเดา) พารามิเตอร์ความสามารถผู้สอบ และการทำหน้าที่ที่ต่างกัน ระหว่างวิธีของเบส์ (Bayes) และวิธีของเบส์แบบมี อิทธิพลทดสอบ (Bayes_y) โดยใช้โปรแกรม SPSS ที่ระดับนัยสำคัญ $p < 0.01$ โดยใช้เกณฑ์การแปลงขนาดความสัมพันธ์ของ โคเฮ็น (Cohen, 1988)

กรอบแนวคิดในการวิจัย



ภาพที่ 1 กรอบแนวคิดในการวิจัย

สรุปผลการวิจัย

จากการวิเคราะห์ผู้วิจัยได้วิเคราะห์ด้วยวิธีของเบส์ (Bayes) และวิธีของเบส์แบบมี อิทธิพลทดสอบ (Bayes_y) โดยใช้โปรแกรม WinBUGS14 และนำผลการประมาณค่ามาเปรียบเทียบกันได้ดังนี้

1. ผลการเปรียบเทียบผลการประมาณค่าพารามิเตอร์ของข้อสอบ และพารามิเตอร์ความสามารถของผู้สอบ ด้วยวิธีของเบส์ (Bayes) และวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayes_y)

1.1 เปรียบเทียบผลการประมาณค่าพารามิเตอร์อำนาจจำแนกของข้อสอบด้วยวิธีของเบส์ (Bayes) และวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayes_γ) ผลการประมาณค่าพารามิเตอร์อำนาจจำแนกพบว่า ค่าพารามิเตอร์อำนาจจำแนกที่ประมาณค่าด้วยวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayes_γ) ส่วนใหญ่จะมีค่าสูงกว่าเล็กน้อยเมื่อเทียบกับค่าพารามิเตอร์อำนาจจำแนกที่ประมาณค่าด้วยวิธีของเบส์ (Bayes)

1.2 เปรียบเทียบผลการประมาณค่าพารามิเตอร์ความยากของข้อสอบด้วยวิธีของเบส์ (Bayes) และวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayes_γ) ผลการประมาณค่าพารามิเตอร์ความยากพบว่า ค่าพารามิเตอร์ความยากที่ประมาณค่าด้วยวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayes_γ) ส่วนใหญ่จะมีค่าสูงกว่าเล็กน้อยเมื่อเทียบกับค่าพารามิเตอร์ความยากที่ประมาณค่าด้วยวิธีของเบส์ (Bayes)

1.3 เปรียบเทียบผลการประมาณค่าพารามิเตอร์โอกาสการเดาของข้อสอบด้วยวิธีของเบส์ (Bayes) และวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayes_γ) ผลการประมาณค่าพารามิเตอร์โอกาสการเดาของข้อสอบพบว่า ค่าพารามิเตอร์โอกาสการเดาที่ได้จากการประมาณค่าด้วยวิธีของเบส์ (Bayes) จะมีค่าสูงกว่าค่าพารามิเตอร์โอกาสการเดาที่ประมาณด้วยวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayes_γ)

1.4 เปรียบเทียบผลการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ ด้วยวิธีของเบส์ (Bayes) และวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayes_γ) ผลการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบพบว่า ค่าพารามิเตอร์ความสามารถของผู้สอบที่ประมาณค่าด้วยวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayes_γ) มีค่าใกล้เคียงกับค่าพารามิเตอร์ความยากที่ประมาณค่าด้วยวิธีของเบส์ (Bayes)

2. ผลเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีของเบส์ (Bayes) และวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayes_γ)

2.1 เปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ จำแนกตามเพศ

จากการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในการวิเคราะห์ครั้งนี้ผู้วิจัยได้กำหนดให้เพศชายเป็น 1 เป็นกลุ่มอ้างอิง และเพศหญิงเป็น 0 เป็นกลุ่มเปรียบเทียบ ซึ่งเมื่อนำความแตกต่างของเพศไปทดสอบด้วยสถิติที่ ก็สามารถใช้เป็นตัวบ่งชี้การทำหน้าที่ต่างกันของข้อสอบในแต่ละข้อได้ ซึ่งถ้าผลการทดสอบมีนัยสำคัญทางสถิติแสดงว่าข้อสอบข้อนั้นเกิดการทำหน้าที่ต่างกันระหว่างกลุ่มเพศ เมื่อสัมประสิทธิ์ติดลบกลุ่มอ้างอิงจะเสียเปรียบในการตอบ แต่ถ้าสัมประสิทธิ์เป็นบวกกลุ่มอ้างอิงจะได้เปรียบ ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่จำแนกตามเพศ จากข้อสอบวิชาภาษาอังกฤษจำนวน 40 ข้อ พบว่าเมื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีของเบส์ (Bayes) และวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayes_γ) ไม่พบข้อสอบตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

2.2 เปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ จำแนกตามสถานที่ตั้งภูมิศาสตร์ของโรงเรียน ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ จำแนกตามสถานที่ตั้งภูมิศาสตร์ของโรงเรียน ในการวิเคราะห์ครั้งนี้ผู้วิจัยได้กำหนดให้โรงเรียนในเขตกรุงเทพฯ และปริมณฑล เป็น 1 เป็นกลุ่มอ้างอิง และโรงเรียนนอกเขตกรุงเทพฯ และปริมณฑล 0 เป็นกลุ่มเปรียบเทียบ จากข้อสอบวิชาภาษาอังกฤษจำนวน 40 ข้อ พบว่าเมื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีของเบส์ (Bayes) มีข้อสอบข้อที่ 2 ที่มีการทำหน้าที่ต่างกัน ซึ่งตรวจพบข้อสอบที่มีการทำหน้าที่ต่างกันของข้อสอบตรงกันกับวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayes_γ) แต่ขนาดของการทำหน้าที่ต่างกันของข้อสอบที่วิเคราะห์ด้วยเบส์แบบมีอิทธิพลทดสอบ (Bayes_γ) มีค่าการทำหน้าที่ต่างกันของข้อสอบอยู่ในระดับมาก ซึ่งสูงกว่าวิธีของเบส์ (Bayes)

3.6 ความสัมพันธ์ของขนาดการทำหน้าที่ต่างกันของข้อสอบ จำแนกตามสถานที่ตั้งภูมิศาสตร์ของโรงเรียน ระหว่างวิธีของเบส์ (Bayes) และวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayes_y) พบว่าขนาดของการทำหน้าที่ต่างกันของข้อสอบ ที่วิเคราะห์ด้วยวิธีของเบส์ (Bayes) และวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayes_y) มีความสอดคล้องไปในแนวทางเดียวกัน ผู้วิจัยจึงต้องการศึกษาผลความสัมพันธ์ของขนาดของการทำหน้าที่ต่างกันของข้อสอบของทั้งสองวิธี ซึ่งค่าสัมประสิทธิ์สหสัมพันธ์ของขนาดของการทำหน้าที่ต่างกันของข้อสอบที่ได้จากการวิเคราะห์ด้วยค่าด้วยวิธีของเบส์ (Bayes) และวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayes_y) มีความสัมพันธ์สูง เท่ากับ .983 และมีนัยสำคัญทางสถิติที่ระดับ .01

อภิปรายผล

การอภิปรายผลในงานวิจัยนี้นำเสนอ 3 ประเด็นหลักตามวัตถุประสงค์ โดยมีรายละเอียด ดังนี้

1. ผลการประมาณค่าพารามิเตอร์ข้อสอบ (อำนาจจำแนก ความยาก และโอกาสการเดา) และพารามิเตอร์ความสามารถผู้สอบ เมื่อพิจารณาวิธีประมาณค่าพารามิเตอร์จากผลการประมาณค่าอำนาจจำแนก พบว่าค่าพารามิเตอร์อำนาจจำแนกที่ประมาณด้วยวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayes_y) ส่วนใหญ่จะมีค่าสูงกว่าค่าพารามิเตอร์อำนาจจำแนกที่ประมาณด้วยวิธีของเบส์ (Bayes) นั่นคือการประมาณค่าพารามิเตอร์อำนาจจำแนกด้วยวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayes_y) จะมีค่าอำนาจจำแนกที่สูงกว่า ซึ่งสอดคล้องกับงานวิจัยของ อริสพา เต หลิม (2559) ที่ศึกษาผลการประมาณค่าพารามิเตอร์ที่ได้จากการจำลองข้อมูลระหว่างวิธีของเบส์ (Bayes) และวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayes_y) พบว่าค่าพารามิเตอร์อำนาจจำแนกที่ประมาณด้วยวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayes_y) ประมาณค่าได้ดีกว่า เมื่อข้อมูลมีการแจกแจงปกติ ซึ่งในงานวิจัยครั้งนี้ข้อมูลที่ได้ผลจากการสอบระดับชาติ ที่มีการแจกแจงแบบปกติมาทำการวิเคราะห์ และพบว่าวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayes_y) ประมาณค่าพารามิเตอร์อำนาจจำแนกได้ดีที่สุด

เมื่อพิจารณาที่ผลการประมาณค่าพารามิเตอร์ความยากจะเห็นว่าค่าพารามิเตอร์ความยากที่ประมาณด้วยวิธีของเบส์ (Bayes) และวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayes_y) มีค่าความยากใกล้เคียงกัน ซึ่งค่าพารามิเตอร์ความยากที่ประมาณด้วยวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayes_y) จะมีค่าสูงกว่าเล็กน้อย

เมื่อพิจารณาที่ผลการประมาณค่าพารามิเตอร์ความสามารถผู้สอบ จะเห็นว่าค่าพารามิเตอร์ความสามารถผู้สอบที่ประมาณด้วยวิธีของเบส์ (Bayes) และวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayes_y) มีค่าใกล้เคียงกันมาก ซึ่งในการวิเคราะห์คุณภาพของแบบทดสอบนั้น จะให้ความสำคัญกับค่าพารามิเตอร์ของข้อสอบมากกว่าในการวิเคราะห์ค่าพารามิเตอร์ผู้สอบ

ในงานวิจัยครั้งนี้ที่ได้ผลจากการสอบระดับชาติ ที่มีการแจกแจงแบบปกติ มาวิเคราะห์ข้อสอบหากมองในภาพรวมในการประมาณค่าพารามิเตอร์ข้อสอบ และพารามิเตอร์ความสามารถผู้สอบ พบว่าวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayes_y) จะให้ผลที่ดีกว่า เมื่อพิจารณาจากผลของค่าอำนาจจำแนก ที่วิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayes_y) นั้นได้ค่าอำนาจจำแนกที่ดีกว่า ที่ประมาณด้วยวิธีของเบส์ (Bayes) ส่วนค่าพารามิเตอร์ความยาก ค่าพารามิเตอร์โอกาสการเดา และค่าพารามิเตอร์ความสามารถผู้สอบ ทั้งสองวิธีไม่แตกต่างกันมากนัก และเมื่อพิจารณาจะพบว่าผลการวิเคราะห์ด้วยวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayes_y) จะให้ค่าที่ดีกว่าเนื่องจากกำจัด

อิทธิพลของทดสอบที่เลวทำให้ข้อสอบมีความเป็นอิสระออกจากกัน ตามข้อตกลงเบื้องต้นของวิเคราะห์คุณภาพของข้อสอบตามทฤษฎีการตอบสนองของข้อสอบ (IRT)

2. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ พบว่าผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่จำแนกตามเพศและจำแนกตามสถานที่ตั้งภูมิศาสตร์ของโรงเรียน ด้วยวิธีวิธีของเบส์ (Bayes) และวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayes_y) มีจำนวนข้อที่พบการทำหน้าที่ต่างกันของข้อสอบไม่แตกต่างกัน เมื่อพิจารณาจำนวนข้อสอบที่ตรวจพบจะเห็นว่ามีการตรวจพบการทำหน้าที่ต่างกันของข้อสอบเมื่อจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนในข้อสอบข้อที่ 2 และข้อสอบที่มีการทำหน้าที่ต่างกันของข้อสอบตรงกับวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayes_y) มีขนาดของการทำหน้าที่ต่างกันของข้อสอบอยู่ในระดับมาก ซึ่งสูงกว่าวิธีของเบส์ (Bayes) แต่ไม่พบการทำหน้าที่ต่างกันของข้อสอบที่จำแนกตามเพศ ทั้งนี้อาจเนื่องมาจากข้อมูลที่ใช้ในการวิเคราะห์ครั้งนี้เป็นข้อมูลจริงและเป็นเพียงการวิเคราะห์จากวิชาภาษาอังกฤษวิชาเดียว จึงอาจทำให้ไม่เจอข้อมูลที่แตกต่างกันในการวิเคราะห์ทั้งสองวิธีในการวิจัยครั้งต่อไปหากต้องการเปรียบเทียบการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของทั้งสองวิธีให้ชัดเจนควรใช้ข้อมูลที่ได้จากการแบบทดสอบที่หลากหลายทั้งเนื้อหาวิชาและรูปแบบของคำถาม หรืออาจทำการวิเคราะห์จากการจำลองข้อมูลเพื่อเปรียบเทียบประสิทธิภาพของการตรวจการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีวิธีของเบส์ (Bayes) และวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayes_y) ต่อไป

3. ความสัมพันธ์ระหว่างค่าพารามิเตอร์ของข้อสอบ พารามิเตอร์ผู้สอบ และการทำหน้าที่ต่างกันของข้อสอบ ค่าสัมประสิทธิ์สหสัมพันธ์ของการประมาณค่าพารามิเตอร์อำนาจจำแนก ค่าพารามิเตอร์ความยาก ค่าพารามิเตอร์โอกาสการเดา ค่าพารามิเตอร์ความสามารถของผู้สอบ และขนาดการทำหน้าที่ต่างกันของข้อสอบที่วิเคราะห์ด้วยวิธีวิธีของเบส์ (Bayes) และวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayes_y) มีความสัมพันธ์กัน ที่นัยสำคัญที่ .01 ทั้งนี้เนื่องจากวิธีของเบส์ (Bayes) และวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayes_y) อยู่บนพื้นฐานของการวิเคราะห์คุณภาพของข้อสอบตามทฤษฎีการตอบสนองของข้อสอบ (IRT) และฟังก์ชันของสมการเช่นเดียวกัน จึงทำให้มีความสัมพันธ์ในระดับสูงมาก สอดคล้องกับหลักการของ Campbell และ Fiske (1959) ที่ว่าการโมเดลในการประมาณค่าพารามิเตอร์ต่างกันจากแบบทดสอบเดียวกัน ค่าพารามิเตอร์ที่ได้จึงมีความสัมพันธ์กัน

ข้อเสนอแนะ

ข้อเสนอแนะในการนำผลวิจัยไปใช้

1. การวิเคราะห์ค่าพารามิเตอร์ข้อสอบควรเลือกวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayes_y) เนื่องจากให้ค่าพารามิเตอร์อำนาจจำแนกสูงกว่า วิธีของเบส์ (Bayes)
2. การวิเคราะห์เพื่อตรวจการทำหน้าที่ต่างกันของข้อสอบควรเลือกวิธีของเบส์ (Bayes) เนื่องจากมีกระบวนการที่ไม่ซับซ้อน ซึ่งผลการวิเคราะห์ได้ค่าใกล้เคียงกับวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayes_y)

ข้อเสนอแนะในการวิจัยครั้งต่อไป

1. ควรศึกษาเปรียบเทียบประสิทธิภาพของการตรวจการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีวิธีของเบส์ (Bayes) และวิธีของเบส์แบบมีอิทธิพลทดสอบ (Bayes_y) โดยใช้ข้อมูลที่ได้จากการแบบทดสอบที่หลากหลายทั้งเนื้อหาวิชาและรูปแบบของคำถาม หรืออาจทำการวิเคราะห์จากการจำลองข้อมูลเพื่อเปรียบเทียบประสิทธิภาพของทั้งสองวิธีให้ชัดเจนยิ่งขึ้น

2. การศึกษาเปรียบเทียบค่าพารามิเตอร์ของข้อสอบแบบ 3 PL และพารามิเตอร์ความสารถผู้สอบ ด้วยวิธีของเบส์ (Bayes) และวิธีของเบส์แบบมี อิทธิพลทดสอบ (Bayes_y) กับโปรแกรมอื่นๆ

เอกสารอ้างอิง

- คมศักดิ์ ชื่นชม. (2539). การศึกษาผลการวิเคราะห์ความลำเอียงที่ใช้วิธีต่างกันของแบบทดสอบวัดจริยธรรมด้านความซื่อสัตย์. วิทยานิพนธ์การศึกษามหาบัณฑิต, สาขาวิชาการวัดผลการศึกษา, บัณฑิตวิทยาลัย, มหาวิทยาลัยศรีนครินทรวิโรฒ.
- จิตติมา วรรณศรี. (2539). การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซลกับวิธีซิบเทสต์ เมื่อความยาวแบบทดสอบ ขนาดของกลุ่มตัวอย่างอัตราส่วนของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบต่างกัน. วิทยานิพนธ์ ครุศาสตรบัณฑิต, สาขาการวัดและประเมินผลการศึกษา, คณะครุศาสตร์, จุฬาลงกรณ์มหาวิทยาลัย.
- ซัชชัย เผ่าพงษ์. (2527). การเปรียบเทียบความลำเอียงของข้อสอบจากแบบทดสอบวัดความถนัดทางการเรียนคณิตศาสตร์และภาษาระดับชั้นมัธยมตอนต้น. วิทยานิพนธ์ ศษ.ม. บัณฑิตวิทยาลัย, มหาวิทยาลัยขอนแก่น, ขอนแก่น.
- ทัศน์ีย์ พรหมนตรี. (2530). การวิเคราะห์ความลำเอียงของแบบทดสอบวิชาคณิตศาสตร์ตามโครงการตรวจสอบคุณภาพทางการศึกษา ชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526. วิทยานิพนธ์ ศษ.ม. ภาควิชาวิจัย การศึกษา, จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพฯ.
- นงลักษณ์ วิรัชชัย. (2542). โมเดลลิสรล: สถิติวิเคราะห์สำหรับการวิจัย (พิมพ์ครั้งที่ 3). กรุงเทพฯ: โรงพิมพ์ จุฬาลงกรณ์มหาวิทยาลัย.
- สุรศักดิ์ อมรรัตนศักดิ์. (2544). การศึกษาเปรียบเทียบผลของวิธีวิเคราะห์ความลำเอียงของข้อสอบที่ต่างกัน 4 วิธี. วิทยานิพนธ์ ค.ด., จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพฯ.
- สุพัฒนา หอมบุปผา. (2556). การเปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี HGLM วิธี MIMIC และวิธี BAYESIAN. วิทยานิพนธ์ปรัชญาดุษฎีบัณฑิต, สาขาวิชาวิจัย วัดผล และสถิติการศึกษา, คณะศึกษาศาสตร์, มหาวิทยาลัยบูรพา.
- อริสพา เตห์ลิม. (2559). การเปรียบเทียบประสิทธิภาพการประมาณค่าพารามิเตอร์และการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีแมซิมมไลค์ลิสต์ วิธีของเบส์แบบมีอิทธิพลทดสอบ. วิทยานิพนธ์ปรัชญาดุษฎีบัณฑิต, สาขาวิชาวิจัย วัดผล และสถิติการศึกษา, คณะศึกษาศาสตร์, มหาวิทยาลัยบูรพา.
- Cambell, D.T., &Fiske, D, W. (1959). Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix. Psychological Bulletin, 56, 81-105.
- Cohen, J. (1988). Statistical power analysis for the behavioral sciences. (2nd ed.). New Jersey: Lawrence Erlbaum.
- Fukuhara, H., &Kamata, A. (2011). A differential item functioning model for testlet-based items using a bi-factor multidimensional item response theory model: Abayesian approach. Applied Psychological Measurement, 35(8), 969-980.

- Holland, P. W., & Thayer, D. T. (1988). **Differential item performance and the Mantel - Haenszel procedure**. New Jersey Erlbaum.
- Vaughn, B. (2006). **A hierarchical generalized linear model of random differential item functioning for polytomous items: A bayesian multilevel approach**. Tallahassee: Florida State University.