



การเปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี HGLM วิธี MIMIC และวิธี BAYESIAN

สุพัฒนา หอมบุปผา^{1*} ไพรัตน์ วงศ์นาม² และ สมพงษ์ ปันหุ่น³

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ 1) เปรียบเทียบผลการประมาณค่าพารามิเตอร์ความยากของข้อสอบ (δ_i) พารามิเตอร์ความสามารถของผู้สอบ (θ_j) สำหรับผู้สอบจำแนกตาม เพศ สถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน ระหว่าง วิธี HGLM วิธี MIMIC และ วิธี BAYESIAN 2) เปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) สำหรับผู้สอบ จำแนกตาม เพศ สถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน ระหว่าง วิธี HGLM วิธี MIMIC และ วิธี BAYESIAN 3) ศึกษาลักษณะของข้อสอบที่เกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF) ที่ได้จากการวิเคราะห์การทำหน้าที่ต่างกัน โดย วิธี HGLM วิธี MIMIC และ วิธี BAYESIAN ด้วยการวิเคราะห์ลักษณะและเนื้อหาของคำหรือข้อความที่ใช้ในการเขียนข้อสอบ

ข้อมูลที่ใช้ในการวิเคราะห์ครั้งนี้ เป็นคะแนนการสอบวัดผลสัมฤทธิ์ทางการเรียนเพื่อประเมินคุณภาพการศึกษา ระดับชาติ ของนักเรียนชั้นประถมศึกษาปีที่ 3 ในรายวิชาภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์ ของสำนักทดสอบทางการศึกษา กระทรวงศึกษาธิการ ปีการศึกษา 2553 ซึ่งได้มาโดยการสุ่มแบบหลายขั้นตอน (Multi-Stage Random Sampling Technique) จำนวน 1,000 คน จำแนกเป็นเพศชายและเพศหญิง ที่อยู่ในเขตกรุงเทพมหานครและปริมณฑล และนอกเขตกรุงเทพมหานครและปริมณฑล การวิเคราะห์ค่าพารามิเตอร์ความยากของข้อสอบ พารามิเตอร์ความสามารถของผู้สอบ และการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธี HGLM-2L วิธี MIMIC และ วิธี BAYESIAN โดยใช้โปรแกรมสำเร็จรูป 3 โปรแกรม ได้แก่ โปรแกรม HLM โปรแกรม Mplus และ โปรแกรม WinBUGS ตามลำดับ เมื่อพิจารณาการทำหน้าที่ต่างกันของข้อสอบ ผู้วิจัยทำการวิเคราะห์ที่เนื้อหา คำ ประโยค หรือข้อความ ที่ใช้ในการเขียนข้อสอบของแต่ละวิชา

ผลการวิจัยพบว่า ผลการวิเคราะห์ค่าพารามิเตอร์ความยากของข้อสอบ ค่าพารามิเตอร์ความสามารถของผู้สอบ และผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในรายวิชาภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์ ด้วยวิธี HGLM-2L วิธี MIMIC และ วิธี BAYESIAN พบร่วมกัน มีความสัมพันธ์กันในระดับสูงมาก อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 วิธีตรวจสอบที่พบการทำหน้าที่ต่างกันของข้อสอบมากที่สุด คือ วิธี HGLM-2L ส่วนวิธีที่ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบน้อยที่สุด คือ วิธี MIMIC ผลการศึกษาลักษณะของข้อสอบที่ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) วิชาภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์ ลักษณะของข้อสอบที่เกิดการทำหน้าที่ต่างกัน เมื่อจำแนกตามเพศ ส่วนใหญ่ ข้อสอบที่เกิดการทำหน้าที่ต่างกันของข้อสอบ จะมีคำศัพท์ที่เกี่ยวข้องกับเพศนั้น จึงทำให้ข้อสอบเข้าข้างเพศนั้น และอาจเป็นเพราะความสามารถที่แตกต่างระหว่างเพศชายและเพศหญิงที่มีลักษณะความสามารถ ความถนัด และความสามารถในเรื่องนันๆ ต่างกัน เมื่อจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน ส่วนใหญ่ส่วนใหญ่ที่ทำให้ข้อสอบเกิดการทำหน้าที่ต่างกันอาจเป็นเพราะประสบการณ์ ความคุ้นเคยเกี่ยวกับเรื่องนั้น สภาพแวดล้อมและการฝึกปฏิบัติที่แตกต่างกันระหว่างนักเรียนในเขตกรุงเทพมหานครและปริมณฑลนักเรียนนอกเขตกรุงเทพมหานครและปริมณฑล

คำสำคัญ : วิธี HGLM วิธี MIMIC วิธี BAYESIAN

^{1*} นักศึกษาปริญญาเอก สาขาวิชายัดผลและสถิติการศึกษา มหาวิทยาลัยบูรพา

² อาจารย์ประจำภาควิชาวิจัยและจิตวิทยาประยุกต์ คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา

³ อาจารย์ประจำภาควิชาวิจัยและจิตวิทยาประยุกต์ คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา

A COMPARISON OF DIFFERENTIAL ITEM FUNCTIONING BY HGLM, MIMIC AND BAYESIAN METHODS

Supattan Hombubpha^{1*} Pairat Wongnam² And Sompong Punhun³

Abstract

This study aimed to: 1) compare the item parameters (δ_i) and person parameters (θ_j) of grade three students' national test scores of Science, Mathematics and Thai Language on gender and locations variables among HGLM, MIMIC and Bayesian approaches; 2) compare the differential item functioning (DIF) of Science, Mathematics and Thai Language test items on gender and locations variables with which estimate by HGLM, MIMIC and Bayesian approaches; and 3) analyze the content of stems and their options of test items which were detected DIF by all of the detection approaches.

The national test scores of grade three students in 2010 academic year were drawn from the Bureau of Testing's database, Ministry of Education by using the multi-stage random sampling techniques. 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 analysis procedures to estimate the item parameters, person parameters and DIF by HGLM-2L, MIMIC, and BAYESIAN methods were used HLM, Mplus and WinBUGS program software, respectively. All parameter estimates and DIF detection results were compared in terms of congruence and correlation. The meaning of words, purposes of author, structure, grammar, and history were used to analyze test stems and options when met DIF.

The research findings revealed that the difficulty parameters of Thai Language and Science test scores which estimate by HGLM-2L, MIMIC, and BAYESIAN methods were perfectly correlate with .01 statistical significant level while in Mathematics those parameters were high correlation. The correlation among person parameters which estimating by HGLM, MIMIC, and BAYESIAN were very high with .01 statistical significant level in all subject test scores. According to DIF detection, the results of HGLM-2L, MIMIC, and BAYESIAN methods were found DIF in all subjects. The HGLM-2L method was very sensitive to detect DIF and the MIMIC method was hardly to find DIF. The consequences of DIF detection by all methods were high correlation in all subject test scores. Finally, the test items of Science, Mathematics, and Thai Language which found DIF were used familiar words, sentences, situations and experiences of examinees in their stems and options to gender and location variables.

Keywords : HGLM, MIMIC, BAYESIAN



บทนำ

การทดสอบเป็นการดำเนินการที่ตั้งอยู่บนพื้นฐานการวัดคุณลักษณะแห่งภายในตัวบุคคล (traits) โดยใช้ข้อสอบเป็นสิ่งร้าให้ผู้ทดสอบแสดงความสามารถอ่อนตบสนอง หากมีข้อมูลที่สามารถยืนยันได้ว่าข้อสอบที่สร้างขึ้นมีคุณสมบัติวัดได้ตรงตามสิ่งที่ต้องการวัด (Validity) และผลการวัดมีความคงเส้นคงวา (Reliability) ก็ย่อมมั่นใจได้ระดับหนึ่งว่าข้อสอบที่สร้างขึ้นมีคุณภาพเพียงใดนั้น ผู้พัฒนาข้อสอบต้องมีความรู้สึกแก่นแท้ของเนื้อหาวิชาที่จะวัด ประกอบกับความสามารถทักษะการเขียนข้อสอบ และต้องวางแผนการสร้างข้อสอบอย่างรอบคอบ ครอบคลุมเนื้อหาที่ต้องการวัด รวมทั้งมีการตรวจสอบคุณภาพของข้อสอบต้องนำเสนอข้อสอบที่สร้างขึ้มมาไปทดลองสอบกับกลุ่มตัวอย่าง แล้วนำผลการตอบของผู้สอบมาวิเคราะห์หาคุณภาพของข้อสอบเป็นรายข้อ ผลการวิเคราะห์คุณภาพข้อสอบเป็นรายข้อนี้ จะทำให้ทราบว่าข้อสอบแต่ละข้อสามารถทำหน้าที่ได้ตรงตามที่ผู้พัฒนาข้อสอบต้องการหรือไม่ เพื่อเป็นข้อมูลพื้นฐานสำหรับการจัดทำเป็นแบบสอบที่เหมาะสมต่อไป (ศิริชัย ภานุจนวงศ์, 2548; Murphy, R, Charles, และ Davidshofer, 2001)

สำหรับการวิเคราะห์คุณภาพของข้อสอบตามทฤษฎี IRT นั้น ไม่เดลขับชับข้อน้อยที่สุดคือ 1-Parameter Logistic Measurement Model (1 PL) ผลลัพธ์ที่ได้จากการประมาณค่าพารามิเตอร์คือค่าพารามิเตอร์ความสามารถของผู้สอบ (θ) และค่าพารามิเตอร์ของข้อสอบ (β_i) ซึ่งก็มีนักวัดผลตั้งคำถามต่อไปว่า ค่าพารามิเตอร์ความสามารถของผู้สอบมีความผันแปรระหว่างกลุ่มผู้สอบหรือไม่ และความผันแปรที่เกิดขึ้นมีสาเหตุมาจากตัวแปรอิสระใดบ้าง นักวัดผลจึงนำผลการตอบข้อสอบของผู้สอบมาประมาณค่าความสามารถของผู้สอบแต่ละคน ตามทฤษฎี IRT จากนั้นนำค่าพารามิเตอร์ความสามารถของผู้สอบที่ประมาณค่าได้ไปเป็นตัวแปรตามโดยมีตัวแปรคุณลักษณะของผู้สอบ (Person Characteristics) เป็นตัวแปรท่านายด้วยวิเคราะห์ผลโดยพหุ (Multiple Regression) เพื่อศึกษาว่าตัวแปรอิสระใดบ้างที่มีอิทธิพลต่อค่าพารามิเตอร์ความสามารถของผู้สอบ และสามารถอธิบายความผันแปรที่เกิดขึ้นได้อย่างไรบ้าง

จากการศึกษาโน้มเดลเชิงเส้นตรงทั่วไประดับลดหลั่น (Hierarchical Generalized Linear Model: HGLM) ค่าอิทธิพลของตัวแปรภายนอกต่อโอกาสในการตอบข้อสอบในการวิเคราะห์ ระดับที่ 2 (ระดับผู้สอบ) สามารถดำเนินการวิเคราะห์ได้จากโปรแกรม HLM ด้วยโน้มเดลเชิงเส้นตรงทั่วไประดับลดหลั่นได้ (Hierarchical Generalized Linear Model: HGLM) และทำการวิเคราะห์การประมาณค่าพารามิเตอร์ความยากของข้อสอบ (δ) ค่าพารามิเตอร์ความสามารถของผู้สอบ (θ) จากโปรแกรม HLM ซึ่งมีลักษณะเป็นพารามิเตอร์แบบสุ่ม (Random Parameter) การดำเนินการวิเคราะห์สามารถดำเนินการวิเคราะห์ในขั้นตอนเดียวตามโน้มเดล HGLM ด้วยโปรแกรมโน้มเดลเชิงเส้นตรงระดับลดหลั่น (HLM) ที่ผ่านมาส่วนใหญ่ก็จัดได้ดำเนินการวิเคราะห์ในลักษณะแยกส่วน ซึ่งในการวิจัยครั้งนี้ผลการวิเคราะห์ข้อสอบ นักวิจัยจะให้ค่าพารามิเตอร์ข้อสอบค่าพารามิเตอร์ผู้สอบแล้ว ยังจะทราบต่อไปได้ว่าตัวแปรคุณลักษณะของผู้สอบตัวแปรใด สามารถอธิบายความแปรปรวนในค่าความสามารถของผู้สอบได้ และจะนำไปสู่การศึกษาในรายละเอียดเชิงลึกของการพัฒนาการทดสอบ โดยประโยชน์จากการสนใจที่ได้จากการวิเคราะห์ที่น่าเชื่อถือ เพื่อวางแผนกำหนดนโยบายในการพัฒนาคุณภาพการศึกษาให้เกิดประสิทธิภาพต่อไป (อธิบุตร พงษ์ปิยะรัตน์, 2551) ซึ่งโปรแกรม HLM เป็นโปรแกรมที่มีประโยชน์ต่อการพัฒนาการทดสอบแล้ว โปรแกรม Mplus ก็มีความสำคัญต่อการพัฒนาการทดสอบเช่นเดียวกัน

วิธีมิมิก (MIMIC) ก็เป็นวิธีหนึ่งสำหรับการวิเคราะห์คุณภาพของข้อสอบตามทฤษฎี IRT ซึ่งวิธี MIMIC เป็นโน้มเดลลิสเรลที่มีตัวแปรแห่งเพียงตัวแปรเดียว โดยที่ตัวแปรแห่งนั้น ได้รับอิทธิพลจากตัวแปรภายนอก สังเกตได้หลายตัวแปร และส่งอิทธิพลไปยังตัวแปรภายนอก สังเกตได้หลายตัวแปร กล่าวอีกอย่างหนึ่งคือ เป็นโน้มเดลลิสเรลของคุณลักษณะแห่งที่มีหลายสาเหตุและวัดได้จากตัวบ่งชี้หลายตัว ลักษณะโน้มเดลจะเห็นว่าการวัดตัวแปรภายนอกสังเกตได้ต้องมีข้อตกลงข้างต้นว่าไม่มีความคลาดเคลื่อนในการวัด โน้มเดลมิมิกนี้เป็นประโยชน์มากใน



การตรวจสอบความเป็นเอกมิติ (Uni dimensionality) ใน การวิจัยสาหานการวัดผลการศึกษา สามารถวิเคราะห์ค่าพารามิเตอร์คุณลักษณะข้อสอบ และค่าความสามารถของผู้สอบไม่สามารถสังเกตโดยตรง จึงต้องประมาณจาก การตอบข้อสอบการประมาณค่าพารามิเตอร์ในทฤษฎี การตอบสนองข้อสอบ

นอกจากนี้ในการนำทฤษฎีการตอบสนองข้อสอบ มาใช้ ผู้วิจัยจำเป็นต้องเลือกวิธีที่เหมาะสมแล้ว วิธีประมาณค่าพารามิเตอร์ของข้อสอบ และความสามารถของผู้เข้าสอบก็เป็นอีกกระบวนการหนึ่งที่จำเป็นต้องเลือกใช้ให้เหมาะสมกับสภาพการวัดแต่ละครั้ง สำหรับทฤษฎีการตอบสนองข้อสอบนั้น วิธีการประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้เข้าสอบมีหนึ่งวิธีที่น่าสนใจ คือ วิธีของเบส์ (Bayesian Estimation) (Swaminathan and Gifford, 1985: 349-364)

การประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบด้วยวิธีแมกซิมัลไลค์ลิขิต เป็นส่วนมาก ทั้งๆ ที่วิธีแมกซิมัลไลค์ลิขิต มีข้อจำกัดที่ไม่สามารถประมาณค่าความสามารถของผู้เข้าสอบที่ได้คะแนนเต็มหรือศูนย์ได้ จำเป็นต้องมีการจำจัดผู้สอบเหล่านี้ออกไปจากการประมาณค่า ทำให้เหลือผู้เข้าสอบที่มีระดับความสามารถใกล้เคียงกัน ดังนั้นหากต้องการทำ การเปรียบเทียบความสามารถของนักเรียนเมื่อทำการสอบ ด้วยแบบทดสอบ 2 ฉบับใด ๆ จะมีแนวโน้มว่าความสามารถของนักเรียนจะถูกแบ่งกลุ่มไม่แตกต่างกันสักผลให้สรุปได้ว่าแบบทดสอบ 2 ฉบับนั้น มีคุณภาพไม่แตกต่างกัน ซึ่งข้อสรุปดังกล่าวอาจเป็นข้อค้นพบที่คลาดเคลื่อน อันเนื่องมาจากการความจำกัดของวิธีประมาณค่าแบบแมกซิมัลไลค์ลิขิตก็ได้ Swaminathan and Gifford (1985) กล่าวว่า การประมาณค่าพารามิเตอร์ด้วยวิธีของเบส์ไม่มีข้อจำกัดดังนี้ วิธีแมกซิมัลไลค์ลิขิต กล่าวคือสามารถประมาณค่าพารามิเตอร์ของผู้สอบที่ผู้เข้าสอบทุกคนตอบถูกหรือตอบผิดได้ และประมาณค่าความสามารถของผู้เข้าสอบที่ทำข้อสอบถูกหรือผิดทุกข้อได้ด้วย โดยเฉพาะในกลุ่มตัวอย่างขนาดเล็กประมาณ 200 - 300 คน วิธีของเบส์สามารถประมาณค่าสถิติของข้อสอบได้ใกล้เคียงกับ

ค่าพารามิเตอร์มากกว่าวิธีแมกซิมัลไลค์ลิขิต (สุนทร เทียนงาน, 2551)

วิธีเบส์เชียน มีความแตกต่างจากวิธีดังเดิมในการอนุมานรูปแบบความน่าจะเป็นของโมเดล สำหรับตัวแปรสังเกตและพารามิเตอร์ที่ไม่ทราบค่าเบส์เชียนจะอนุมานโดยการตรวจสอบเงื่อนไขของพารามิเตอร์ที่สอดคล้องกับข้อมูลที่สังเกตได้ จึงสามารถใช้งานได้ง่าย วิธีเบส์เชียนมีความยืดหยุ่นสูง สามารถแก้ปัญหาทั้งง่ายและซับซ้อนได้ดี มีการกำหนดการแจกแจงเริ่มต้นของค่าพารามิเตอร์ (prior distribution) ที่ใช้ในการกำหนดช่วงของค่าพารามิเตอร์ที่ต้องการประมาณค่า เป็นประโยชน์อย่างยิ่งสำหรับการวิเคราะห์ข้อมูล นอกจากนี้โมเดลที่มีความซับซ้อน ยังในปัจจุบันการพัฒนาเทคนิคการสุ่มตัวอย่างในการจำลองข้อมูล Markov Chain Monte Carlo (MCMC) มีโปรแกรมสนับสนุนการประมาณค่ามากmay เช่น โปรแกรม WinBUGS ซึ่งเป็นซอฟแวร์ที่ใช้งานได้ง่าย

จะเห็นได้ว่า การวิเคราะห์ข้อสอบ ด้วยทฤษฎีการตอบสนองข้อสอบ สามารถให้ห้างสารสนเทศที่เป็นค่าพารามิเตอร์ของข้อสอบเป็นรายข้อ (item parameter) พารามิเตอร์ของผู้สอบเป็นรายบุคคล (person parameter) รวมทั้งความสามารถในการสอบการทำหน้าที่ต่างกันของข้อสอบ ในการวิจัยนี้ จึงมุ่งศึกษาการวิเคราะห์คุณภาพของข้อสอบตามทฤษฎีการตอบสนองข้อสอบ (IRT) 1-parameter logistic measurement model (1-PL) ทั้งนี้ด้วยข้อจำกัดของวิธีการวิเคราะห์ด้วยวิธี HGLM สามารถวิเคราะห์ข้อสอบได้เพียง 1PL ส่วนวิธี MIMIC สามารถวิเคราะห์ได้ 2PL และวิธี BAYESIAN สามารถวิเคราะห์ได้ 3PL เพื่อให้สามารถเปรียบเทียบผลการประมาณค่าพารามิเตอร์ความยากของข้อสอบได้ ในการวิจัยครั้งนี้ ผู้วิจัยจึงศึกษาเพียง 1PL จากนั้นจึงตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยดำเนินการวิเคราะห์ด้วยวิธี HGLM ประยุกต์ใช้โปรแกรม HLM วิธี MIMIC ประยุกต์ใช้โปรแกรม Mplus และวิธี BAYESIAN ประยุกต์ใช้โปรแกรม WinBUGS ซึ่งโปรแกรมดังกล่าวสามารถวิเคราะห์สถิติขั้นสูงได้ดี และเป็นที่นิยมของนักสถิติและนักวัดผลในขณะนี้ โดยศึกษาจากการสอบวัดผลสัมฤทธิ์ทางการเรียนเพื่อประเมินคุณภาพการศึกษา ระดับชาติ (NT) ปีการศึกษา 2553 ขั้นประ楫ศึกษาปีที่



3. ได้แก่ วิชาภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์ เพื่อเป็นแนวทางสำหรับผู้ที่เกี่ยวข้องในการออกแบบข้อสอบระดับชาติ ในการนำไปปรับปรุงและพัฒนาข้อสอบต่อไป

วัตถุประสงค์ของการวิจัย

1. เพื่อเปรียบเทียบผลการประมาณค่าพารามิเตอร์ ข้อสอบ (δ_i) และพารามิเตอร์ความสามารถของผู้สอบ (θ_j) ระหว่างวิธี HGLM วิธี MIMIC และวิธี BAYESIAN

2. เพื่อเปรียบเทียบผลการวิเคราะห์การทำหน้าที่ ต่างกันของข้อสอบ (DIF) จำแนกตามเพศและสถานที่ตั้ง ทางภูมิศาสตร์ของโรงเรียน ระหว่างวิธี HGLM วิธี MIMIC และวิธี BAYESIAN

3. เพื่อศึกษาลักษณะของข้อสอบที่เกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF) ที่ได้จากการวิเคราะห์ การทำหน้าที่ต่างกัน โดยวิธี HGLM วิธี MIMIC และวิธี BAYESIAN ด้วยการวิเคราะห์ลักษณะและเนื้อหาของคำ หรือข้อความที่ใช้ในการเขียนข้อสอบ

วิธีดำเนินการวิจัย

ประชากรและกลุ่มตัวอย่าง

ประชากรในการวิจัยเป็นนักเรียนชั้นประถมศึกษาปีที่ 3 ที่เข้าสอบวัดผลสัมฤทธิ์ทางการเรียนเพื่อประเมินคุณภาพการศึกษาระดับชาติ ปีการศึกษา 2553 วิชาภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์ การเลือกกลุ่มตัวอย่างใช้วิธีการสุ่มแบบหลายขั้นตอน ได้นักเรียนจำนวน 1,000 คน จำแนก เป็น เพศชาย จำนวน 500 คน และเพศหญิง 500 คน ที่อยู่ในเขตกรุงเทพมหานคร และปริมณฑล และนอกเขตกรุงเทพมหานคร และปริมณฑล

เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยครั้งนี้ เป็นแบบสอบถามสัมฤทธิ์ทางการเรียนเพื่อประเมินคุณภาพการศึกษาระดับชาติ ของนักเรียนชั้นประถมศึกษาปีที่ 3 ได้แก่ วิชาภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์ สร้างโดยสำนักทดสอบทางการศึกษา คณะกรรมการออกแบบข้อสอบ ประกอบด้วย ครูผู้สอน ศึกษานิเทศก์ นักวิชาการ และผู้เชี่ยวชาญที่เกี่ยวข้อง การสร้างข้อสอบเป็นไปตาม

วัตถุประสงค์ที่กำหนดให้หลักสูตรชั้นประถมศึกษาปีที่ 3 ซึ่งประกอบด้วยข้อสอบดังต่อไปนี้

1. วิชาภาษาไทย จำนวน 30 ข้อ
2. คณิตศาสตร์ จำนวน 30 ข้อ
3. วิทยาศาสตร์ จำนวน 30 ข้อ

การเก็บรวบรวมข้อมูล

ผู้วิจัยรวบรวมข้อมูลผลการตอบของนักเรียนที่เป็นกลุ่มตัวอย่าง ซึ่งเป็นข้อมูลทุติยภูมิ ของนักเรียนที่สอบวัดผล สัมฤทธิ์ทางการเรียนเพื่อประเมินคุณภาพการศึกษาระดับชาติ ชั้นประถมศึกษาปีที่ 3 ปีการศึกษา 2553 จากสำนักทดสอบทางการศึกษา กระทรวงศึกษาธิการ แล้วนำข้อมูลมาวิเคราะห์ตามวัตถุประสงค์ของการวิจัย

การวิเคราะห์ข้อมูล

1. คำนวณค่าสถิติพื้นฐาน ได้แก่ ค่าเฉลี่ย (Mean) ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) ค่าความเบี้ยว (Skewness) ค่าความโด่ง (Kurtosis) เพื่อบรรยายลักษณะการแจกแจงของคะแนนการสอบวิชาภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์

2. การวิเคราะห์การประมาณค่าพารามิเตอร์ความยากของข้อสอบ (δ_i) ค่าพารามิเตอร์ความสามารถของผู้สอบ (θ_j) จากข้อมูลผลการตอบข้อสอบที่มีการให้คะแนนแบบ 2 ค่า (Dichotomous) ดังสมการโลジสติก 1 พารามิเตอร์ต่อไปนี้

$$P_i(\theta) = \frac{e^{(\theta-\delta_i)}}{1+e^{(\theta-\delta_i)}}$$

และการตรวจสอบการทำหน้าที่ต่างกันของ ข้อสอบ (DIF) ด้วยวิธี HGLM-2L ประยุกต์ใช้โปรแกรม HLM วิธี MIMIC ประยุกต์ใช้โปรแกรม Mplus และวิธี BAYESIAN ประยุกต์ใช้โปรแกรม WinBUGS

สรุปผลการวิจัย

1. ผลการเปรียบเทียบการประมาณค่าพารามิเตอร์ความยากของข้อสอบและพารามิเตอร์ความสามารถของผู้สอบ

1.1 เปรียบเทียบผลการประมาณค่าพารามิเตอร์ความยากของข้อสอบ

ผลของการประมาณค่าพารามิเตอร์ความยากของข้อสอบวิชาภาษาไทย พบว่า วิธี HGLM-2L สอดคล้องสัมพันธ์กับผลการประมาณค่าพารามิเตอร์ความยากของข้อสอบด้วยวิธี MIMIC และสอดคล้องสัมพันธ์กับผลการประมาณค่าพารามิเตอร์ความยากของข้อสอบด้วยวิธี BAYESIAN อย่างสัมบูรณ์ (1.00) และมีนัยสำคัญทางสถิติที่ระดับ .01

ส่วนผลการประมาณค่าพารามิเตอร์ความยากของข้อสอบ วิชาคณิตศาสตร์ ค่าสัมประสิทธิ์สัมพันธ์ของค่าพารามิเตอร์ความยากทั้ง 3 วิธี พบว่า วิธี HGLM-2L วิธี MIMIC และวิธี BAYESIAN มีความสัมพันธ์ในระดับสูงมาก และมีนัยสำคัญทางสถิติที่ระดับ .01

และผลการประมาณค่าพารามิเตอร์ความยากของข้อสอบวิชาวิทยาศาสตร์ พบร้า วิธี HGLM-2L สอดคล้องสัมพันธ์กับผลการประมาณค่าพารามิเตอร์ความยากของข้อสอบด้วยวิธี MIMIC และสอดคล้องสัมพันธ์กับผลการประมาณค่าพารามิเตอร์ความยากของข้อสอบด้วยวิธี BAYESIAN อย่างสัมบูรณ์ (1.00) และมีนัยสำคัญทางสถิติที่ระดับ .01

1.2 การเปรียบเทียบผลการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ

การเปรียบเทียบผลการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ (θ_j) ในการวิจัยครั้งนี้ ประกอบด้วยวิชาภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์ ซึ่งทั้ง 3 วิชา ใช้วิธีการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบได้แก่ วิธี HGLM-2L และนำผลการวิเคราะห์มาเปรียบเทียบกับการประมาณค่าด้วย วิธี MIMIC และวิธี BAYESIAN

ซึ่งผลการวิเคราะห์ค่าสัมประสิทธิ์สัมพันธ์ระหว่างค่าพารามิเตอร์ความสามารถของผู้สอบ วิชาภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์ ระหว่างวิธี HGLM-2L วิธี MIMIC และวิธี BAYESIAN พบร้า ผลการวิเคราะห์ค่าสัมประสิทธิ์สัมพันธ์ของค่าพารามิเตอร์ความสามารถของผู้สอบ (θ_j) ทั้ง 3 วิธี มีความสัมพันธ์ในระดับสูงมาก และมีนัยสำคัญทางสถิติที่ระดับ .01

2. ผลการเปรียบเทียบการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) จำแนกตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน

การศึกษาการทำหน้าที่ต่างกันของข้อสอบในครั้งนี้ประกอบด้วยวิชาภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์ มีจำนวนข้อสอบวิชาละ 30 ข้อ รวมทั้งหมด 90 ข้อ ผู้วิจัยได้ใช้วิธี HGLM-2L และนำผลการวิเคราะห์มาเปรียบเทียบกับวิธี MIMIC และวิธี BAYESIAN โดยประยุกต์ใช้โปรแกรม HLM โปรแกรม Mplus และโปรแกรม WinBUGS14 ตามลำดับ จำแนกตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน ซึ่งมีการวิเคราะห์ด้วยวิธีต่างๆ ดังนี้

2.1 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) จำแนกตามเพศ

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย จากการวิเคราะห์ข้อสอบจำนวน 30 ข้อ พบร้าข้อสอบข้อที่ 1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 18, 20, 21, 23, 24, 25, 26, 28, 29 และข้อที่ 30 ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ ซึ่งผลการตรวจสอบค่าดัชนีการทำหน้าที่ต่างกันของข้อสอบ (DIF) ของวิธี HGLM-2L วิธี MIMIC และวิธี BAYESIAN มีความสัมพันธ์กันสูงมาก อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

ส่วนผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ จากการวิเคราะห์ข้อสอบจำนวน 30 ข้อ พบร้าข้อสอบข้อที่ 1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 18, 20, 21, 23, 24, 25, 26, 28, 29 และข้อที่ 30 ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ ซึ่งผลการตรวจสอบค่าดัชนีการทำหน้าที่ต่างกันของข้อสอบ (DIF) ของวิธี HGLM-2L วิธี MIMIC และวิธี BAYESIAN มีความสัมพันธ์กันสูงมาก อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

และผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิทยาศาสตร์ จากการวิเคราะห์ข้อสอบจำนวน 30 ข้อ พบร้าข้อสอบข้อที่ 1, 6, 7, 13, 18, 21, 23, 24, 25, 26, 27, 28 และข้อที่ 30 ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ ซึ่งผลการตรวจสอบค่าดัชนีการทำหน้าที่ต่างกันของข้อสอบ (DIF) ของวิธี HGLM-2L วิธี MIMIC และวิธี BAYESIAN มีความสัมพันธ์กันสูงมาก อย่างมีนัยสำคัญทางสถิติที่ระดับ .01



ทำหน้าที่ต่างกันของข้อสอบ (DIF) ของวิธี HGLM-2L วิธี MIMIC และวิธี BAYESIAN มีความสัมพันธ์กันสูงมากอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

2.2 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย จากการวิเคราะห์ข้อสอบจำนวน 30 ข้อ พบว่าข้อสอบข้อที่ 4,14,15,26 และข้อที่ 29 ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ ซึ่งผลการตรวจสอบค่าดัชนีการทำหน้าที่ต่างกันของข้อสอบ (DIF) ของวิธี HGLM-2L วิธี MIMIC และวิธี BAYESIAN มีความสัมพันธ์กันสูงมากอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

ส่วนผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ จากการวิเคราะห์ข้อสอบจำนวน 30 ข้อ พบว่าข้อสอบข้อที่ 12 ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ ซึ่งผลการตรวจสอบค่าดัชนีการทำหน้าที่ต่างกันของข้อสอบ (DIF) ของวิธี HGLM-2L วิธี MIMIC และวิธี BAYESIAN มีความสัมพันธ์กันสูงมากอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

และผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิทยาศาสตร์ จากการวิเคราะห์ข้อสอบจำนวน 30 ข้อ พบว่าพบว่าข้อสอบข้อที่ 1, 2, 11, 12, 15, 18, 26, 27, 29 และข้อที่ 30 ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ ซึ่งผลการตรวจสอบค่าดัชนีการทำหน้าที่ต่างกันของข้อสอบ (DIF) ของวิธี HGLM-2L วิธี MIMIC และวิธี BAYESIAN มีความสัมพันธ์กันสูงมากอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

โดยสรุป ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย ระหว่างวิธี HGLM-2L วิธี MIMIC และวิธี BAYESIAN จำแนกตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน พบว่า เมื่อจำแนกตามเพศ วิธี HGLM-2L จะพบข้อสอบที่ทำหน้าที่ต่างกันมากรองลงมาคือ จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน วิธี BAYESIAN พบข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ จำแนกตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน พบว่า เมื่อจำแนกตามเพศ วิธี HGLM จะพบข้อสอบที่ทำหน้าที่ต่างกันมากที่สุดรองลงมาคือ จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน วิธี HGLM และวิธี BAYESIAN พบข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด

และการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิทยาศาสตร์ จำแนกตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน พบว่า เมื่อจำแนกตามเพศ วิธี HGLM จะพบข้อสอบที่ทำหน้าที่ต่างกันมากที่สุดรองลงมาคือ จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน วิธี HGLM พบข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด

3. ผลการศึกษาลักษณะของข้อสอบที่เกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF)

ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) หัว 3 วิธี ได้แก่ วิธี HGLM วิธี MIMIC และวิธี BAYESIAN ด้วยการวิเคราะห์ลักษณะและเนื้อหาของคำหรือข้อความที่ใช้ในการเขียนข้อสอบ โดยจำแนกตามเพศ ซึ่งแต่ละวิธีสามารถตรวจพบข้อสอบที่ทำหน้าที่ต่างกันแตกต่างกันไป ผู้วิจัยจึงขอกล่าวถึงลักษณะของข้อสอบที่เกิดการทำหน้าที่ต่างกัน ดังนี้

ข้อสอบวิชาภาษาไทยที่เกิดการทำหน้าที่ต่างกันของข้อสอบจากหัว 3 วิธี คือ วิธี HGLM-2L วิธี MIMIC และวิธี BAYESIAN ได้แก่ข้อที่ 24 และข้อที่ 30 ส่วนข้อสอบวิชาคณิตศาสตร์ที่เกิดการทำหน้าที่ต่างกันของข้อสอบจากหัว 2 วิธี คือ วิธี HGLM-2L และวิธี BAYESIAN ได้แก่ข้อที่ 30 และข้อสอบวิทยาศาสตร์ที่เกิดการทำหน้าที่ต่างกันของข้อสอบ จากหัว 3 วิธี คือ วิธี HGLM-2L วิธี MIMIC และวิธี BAYESIAN ได้แก่ข้อที่ 6,7,13 และข้อที่ 24

จากลักษณะของข้อสอบที่เกิดการทำหน้าที่ต่างกันในภาพรวมหัว 3 วิชา โดยจำแนกตามเพศ ส่วนใหญ่ข้อสอบที่เกิดการทำหน้าที่ต่างกันของข้อสอบ จะมีคำศัพท์ที่เกี่ยวข้องกับphenomenon จึงทำให้ข้อสอบเข้าข้างเพศนั้น และอาจเป็นเพราะความสามารถที่แตกต่างระหว่าง

เพศชาย และเพศหญิงที่มีลักษณะความสามารถ ความดันต์ และความสนใจในเรื่องนั้น ๆ ต่างกัน

ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) ทั้ง 3 วิธี ได้แก่ วิธี HGLM วิธี MIMIC และวิธี BAYESIAN ด้วยการวิเคราะห์ลักษณะและเนื้อหาของคำหรือข้อความที่ใช้ในการเขียนข้อสอบ โดยจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน มีลักษณะของข้อสอบที่เกิดการทำหน้าที่ต่างกัน ดังนี้

ข้อสอบวิชาภาษาไทยที่เกิดการทำหน้าที่ต่างกันของข้อสอบจากทั้ง 2 วิธี คือ วิธี HGLM-2L และวิธี BAYESIAN ได้แก่ข้อที่ 4 และข้อที่ 26 ส่วนข้อสอบวิชาคณิตศาสตร์ที่เกิดการทำหน้าที่ต่างกันของข้อสอบจากทั้ง 2 วิธี คือ วิธี HGLM-2L และวิธี BAYESIAN ได้แก่ข้อที่ 12 และเมื่อพิจารณาข้อสอบวิทยาศาสตร์ที่เกิดการทำหน้าที่ต่างกันของ จากทั้ง 2 วิธี คือ วิธี HGLM-2L และวิธี BAYESIAN ได้แก่ข้อที่ 18 และข้อที่ 27

จากลักษณะของข้อสอบที่เกิดการทำหน้าที่ต่างกันในภาพรวมทั้ง 3 วิชา โดยจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน ส่วนใหญ่ส่าเหตุที่ทำให้ข้อสอบเกิดการทำหน้าที่ต่างกันอาจเป็นเพราะประสบการณ์ ความคุ้นเคยเกี่ยวกับเรื่องนั้น สภาพแวดล้อมและการฝึกปฏิบัติที่แตกต่างกันระหว่างนักเรียนในเขตกรุงเทพฯ และปริมณฑล นักเรียนนอกเขตกรุงเทพฯ และปริมณฑล

อภิปรายผลการวิจัย

จากการวิจัยผู้วิจัยมีประเด็นสำคัญที่จะอภิปรายผลดังนี้

1. การเปรียบเทียบผลการประมาณค่าพารามิเตอร์ความยากของข้อสอบ (δ_i) และพารามิเตอร์ความสามารถของผู้สอบ (θ_j)

1.1 ผลการประมาณค่าพารามิเตอร์ความยากของข้อสอบ (δ_i) ผลการประมาณค่าพารามิเตอร์ความสามารถของข้อสอบ (δ_i) วิชาภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์ พบว่า วิธี HGLM-2L วิธี MIMIC และวิธี BAYESIAN สามารถประมาณค่าพารามิเตอร์ความยากของ

ข้อสอบได้ และวิธี MIMIC มีการประมาณค่าพารามิเตอร์ความยากของข้อสอบได้สูงกว่า วิธี HGLM-2L และวิธี BAYESIAN ทั้งนี้อาจเนื่องมาจากวิธีทั้งสองคือการวิเคราะห์ลักษณะของข้อสอบโดยใช้แบบจำลองเชิงเส้นยั่น หรือวิธี MIMIC เป็นหลักการของ CFA กับตัวแปร (Muthén, 1988) แล้ววิธี MIMIC ยังสามารถนำไปใช้สำหรับการวิเคราะห์ DIF ได้ด้วย (Muthén et al., 1991) ซึ่งผลที่ได้ต้องมีค่าเป็นแบบ 2 ค่า (dichotomous)

และค่าสัมประสิทธิ์สหสัมพันธ์ของการประมาณค่าพารามิเตอร์ความยากของข้อสอบวิชาภาษาไทย ที่วิเคราะห์ด้วยวิธี HGLM-2L วิธี MIMIC และวิธี BAYESIAN มีความสอดคล้องสัมพันธ์กันอย่างสัมบูรณ์ และมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนค่าสัมประสิทธิ์สหสัมพันธ์ของการประมาณค่าพารามิเตอร์ความยากของข้อสอบวิชาคณิตศาสตร์ ที่วิเคราะห์ด้วยวิธี HGLM-2L วิธี MIMIC และวิธี BAYESIAN มีความสัมพันธ์ในระดับสูงมาก และมีนัยสำคัญทางสถิติที่ระดับ .01 และค่าสัมประสิทธิ์สหสัมพันธ์ของการประมาณค่าพารามิเตอร์ความยากของข้อสอบวิทยาศาสตร์ ที่วิเคราะห์ด้วยวิธี HGLM-2L วิธี MIMIC และวิธี BAYESIAN มีความสอดคล้องสัมพันธ์กันอย่างสัมบูรณ์ และมีนัยสำคัญทางสถิติที่ระดับ .01 ทั้งนี้อาจเนื่องจากการวิเคราะห์จากโปรแกรม HLM ด้วยวิธี HGLM-2L โปรแกรม Mplus ด้วยวิธี MIMIC และโปรแกรม WinBUGS ด้วยวิธี BAYESIAN มีความสอดคล้องสัมพันธ์กันอย่างสัมบูรณ์ และมีนัยสำคัญทางสถิติที่ระดับ .01 ทั้งนี้เนื่องมาจากผู้วิจัยได้วิเคราะห์วิธี HGLM-2L แบบ 1-PL วิเคราะห์ด้วยโปรแกรม HLM (Kamata 2001) ซึ่งวิธี MIMIC วิเคราะห์ด้วยโปรแกรม Mplus สามารถประมาณค่าพารามิเตอร์ได้ 2 PL และวิธี BAYESIAN วิเคราะห์ด้วยโปรแกรม WinBUGS ที่สามารถประมาณค่าได้ 3 PL จึงทำให้มีความสัมพันธ์กันอย่างสัมบูรณ์ ซึ่งสอดคล้องกับงานวิจัยของ Kim, W. (2003) ที่พบว่าค่าพารามิเตอร์ของข้อสอบจาก การประมาณค่าด้วยโมเดล HGLM-2L มีความสัมพันธ์กับค่าพารามิเตอร์ของข้อสอบที่ประมาณค่าด้วยโมเดลรากช่องย่างสมบูรณ์ ($r=1.00$)

1.2 พารามิเตอร์ความสามารถของผู้สอบ (θ_j) ผลการประมาณค่าพารามิเตอร์ความสามารถของ



ผู้สอบ วิชาภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์ พบว่า วิธี HGLM-2L วิธี MIMIC และวิธี BAYESIAN สามารถประมาณค่า พารามิเตอร์ความสามารถของผู้สอบได้ ซึ่งวิธี MIMIC มีการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบได้สูงกว่า วิธี HGLM-2L และวิธี BAYESIAN ทั้งนี้เนื่องจากวิธี MIMIC หรือโมเดล MIMIC ใช้พิจารณาความสอดคล้องระหว่างข้อมูลกับโมเดล ว่ามีคุณลักษณะแห่งเพียง 1 คุณลักษณะ และจากวิธี BAYESIAN มีส่วนที่แตกต่างจากวิธี HGLM-2L และวิธี MIMIC คือ การแจกแจงเริ่มต้น (Prior Distribution) การประมาณค่าพารามิเตอร์ความยากของข้อสอบด้วยวิธี BAYESIAN ที่มีการแจกแจงโพสท์เรีย (Posterior Distribution) ซึ่งประกอบด้วยผลคุณของฟังก์ชันไอลิคหลัก (Likelihood Function) กับการแจกแจงเริ่มต้น(Prior Distribution) มีค่าสูงสุด และสนวนในทาง และกิฟฟอร์ด (Swaminathan and Gifford. 1985: 358-359) ได้กล่าวว่า เมื่อกลุ่มตัวอย่างมีขนาดใหญ่ (มากกว่า 200 คน) และ ข้อสอบมีจำนวนมาก การแจกแจงเริ่มต้นนี้ มีค่าน้อยมากจนทำให้การแจกแจงโพสท์เรียและฟังก์ชันไอลิคหลักมีค่าใกล้เคียงกัน ซึ่งการวิจัยครั้งนี้ใช้กลุ่มตัวอย่าง 1,000 คน และมีจำนวนข้อสอบวิชาภาษาไทย คณิตศาสตร์ และ วิทยาศาสตร์ จำนวนวิชาละ 30 ข้อ รวมทั้งหมด 90 ข้อ ดังนั้นวิธีการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบด้วยวิธี BAYESIAN จึงมีวิธีการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ ต่างจากวิธีอื่น

ส่วนวิธี HGLM-2L มีการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ วิชาวิทยาศาสตร์ได้สูงกว่าวิธี MIMIC และวิธี BAYESIAN ทั้งนี้อาจเนื่องมาจากการประมาณค่าพารามิเตอร์ความยากของข้อสอบ (δ_i) ทั้ง 3 วิธี มีแนวคิดพื้นฐานทางสถิติที่ไม่เหมือนกัน โดยวิธี HGLM-2L และวิธี MIMIC ใช้สถิติแบบอนเบสเซียน (Non Bayesian Statistics) ซึ่งแนวคิดนี้จะพิจารณาว่า เป็นค่าคงที่ ไม่ทราบค่าและต้องการประมาณค่า ในขณะที่วิธี BAYESIAN ใช้สถิติแบบเบสเซียน (Bayesian Statistics) และมีวิธีการประมาณค่าพารามิเตอร์ความยากของข้อสอบ (δ_i) โดยมีการกำหนดการแจกแจงความน่าจะเป็นของค่าพารามิเตอร์ที่เรียกว่า การแจกแจง

เริ่มต้น (Prior Distribution) และสอดคล้องกับผลการศึกษาของอิทธิฤทธิ์ พงษ์ปิยะรัตน์ (2551) ได้ศึกษา การประมาณค่าพารามิเตอร์ความยากของข้อสอบ และ พารามิเตอร์ความสามารถของผู้สอบ จากโมเดล HGLM ทั้ง 2 ระดับ และ 3 ระดับ ด้วยโปรแกรม HLM และการประมาณค่าพารามิเตอร์ทั้งสองจากโปรแกรม BILOG-MG จากผลการวิเคราะห์ พบว่า โมเดล HGLM สามารถประมาณค่าพารามิเตอร์ความยากของข้อสอบได้ และเมื่อให้ผลการวิเคราะห์ค่าพารามิเตอร์ความยากของข้อสอบ จากโปรแกรม BILOG-MG เป็นเกณฑ์การพิจารณาแล้ว พบว่า ผลค่าพารามิเตอร์ความยากของข้อสอบจากโมเดล HGLM-2L และ HGLM-3L นั้นมีความสัมพันธ์กับค่าพารามิเตอร์ความยากของข้อสอบจากโปรแกรม BILOG-MG อย่างสมบูรณ์ (1.00)

จากการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบทั้ง 3 วิธี นี้ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าพารามิเตอร์ความสามารถของผู้สอบของวิชาภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์ ระหว่างวิธี HGLM วิธี MIMIC และวิธี BAYESIAN มีความสัมพันธ์ในระดับสูงมาก และมีนัยสำคัญทางสถิติที่ระดับ .01 ซึ่งสอดคล้องกับการศึกษาของวิชuda บัววงศ์ (2533) ที่ศึกษาผลการประมาณค่า พารามิเตอร์ความสามารถของผู้เข้าสอบ พบว่า แบบสอบที่ประมาณค่าความสามารถของผู้เข้าสอบด้วยวิธีของเบส์ มีความตรงเชิงทำนายสูงกว่าวิธีอื่น อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 และมีค่าฟังก์ชันสารสนเทศของแบบสอบสูงกว่าวิธีอื่น เช่นกัน และถ้าความสามารถของกลุ่มตัวอย่างมีค่าเฉลี่ยอยู่ในระดับปานกลางถึงค่อนข้างต่ำ เมื่อนำค่าพารามิเตอร์ความสามารถของผู้เข้าสอบที่ประมาณได้ในแต่ละวิธีอื่น มาหาความสัมพันธ์กับคะแนนเกณฑ์ด้วยวิธีของเบส์ ก็จะทำให้มีความตรงเชิงทำนายสูงที่สุด

2. เปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) สำหรับผู้สอบจำแนกตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนระหว่างวิธี HGLM วิธี MIMIC และวิธี BAYESIAN

2.1 เปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) สำหรับผู้สอบ จำแนกตามเพศ ดังนี้

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์ พน ข้อสอบที่เกิดการทำหน้าที่ต่างกันของข้อสอบ ส่วนใหญ่ ตรวจพับการทำหน้าที่ต่างกันของข้อสอบจากโปรแกรม HLM ด้วยวิธี HGLM-2L ที่เป็นเช่นนี้อาจเนื่องจากวิธี HGLM-2L ในระดับการวิเคราะห์ที่ 1 จะกำหนดให้ค่าพารามิเตอร์ความยากของข้อสอบที่ประมาณค่าได้เป็นค่าคงที่ (fixed effect) กับผู้สอบแต่ละกลุ่ม เพื่อไม่ให้เกิดความไม่ผันแปรของค่าพารามิเตอร์ไปตามกลุ่มผู้สอบ ซึ่ง เมื่อนำความแตกต่างของเพศไปทดสอบ ด้วยสถิติที่ ก สามารถใช้เป็นตัวปัจจัยการทำหน้าที่ต่างกันของข้อสอบในแต่ละข้อได้ ซึ่งถ้าผลการทดสอบมีนัยสำคัญทางสถิติ แสดงว่าข้อสอบข้อนั้นทำหน้าที่ต่างกันระหว่างกลุ่มเพศ เมื่อสัมประสิทธิ์ติดลบ กคู่ม อ้างอิงจะเสียเปรียบ แต่ถ้าค่าสัมประสิทธิ์เป็นบวกแสดงถึงกลุ่ม อ้างอิงได้เปรียบจากการทำหน้าที่ต่างกันของข้อสอบ ในการวิเคราะห์ครั้งนี้ ผู้วิจัยได้กำหนดให้เพศชายเป็น 1 และเพศหญิงเป็น 0

เมื่อจากวิธี BAYESIAN มีส่วนที่แตกต่างจากวิธี HGLM-2L และวิธี MIMIC คือ การแจกแจงเริ่มต้น (Prior Distribution) การประมาณค่า พารามิเตอร์ความยากของข้อสอบด้วยวิธี BAYESIAN ที่มีการแจกแจงเพสทีเรีย (Posterior Distribution) ซึ่งประกอบด้วยผลคูณของฟังก์ชันลิคิลิคุต (Likelihood Function) กับการแจกแจงเริ่มต้น(Prior Distribution) มีค่าสูงสุด และสามารถหาได้โดยวิธี Gifford (Swaminathan and Gifford. 1985: 358-359)

ซึ่งการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์ เมื่อ จำแนกตามเพศ ในการวิจัยครั้งนี้ วิธี HGLM-2L ตรวจสอบพับการทำหน้าที่ต่างกันของข้อสอบได้สูงกว่าวิธี MIMIC และวิธี BAYESIAN จึงสอดคล้องกับการศึกษาของ Fukuhara and Kamata (2007) ได้ศึกษาการประเมินประสิทธิภาพการทำงานของวิธี MIMIC แบบ ละเอียดข้อตกลงเบื้องต้น โดยการทำชุดข้อสอบ พนว่าการ

ทำหน้าที่ต่างกันของข้อสอบ (DIF) มีแนวโน้มที่จะการประมาณค่าได้ภายใต้ข้อสอบที่ไม่เป็นอิสระกัน

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย วิชาคณิตศาสตร์ และวิชา วิทยาศาสตร์ จากโปรแกรม HLM ด้วยวิธี HGLM-2L สอดคล้องกับผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ได้จากโปรแกรม Mplus ด้วยวิธี MIMIC และ สอดคล้องกับโปรแกรม WinBUGS ด้วยวิธี BAYESIAN มี ความสัมพันธ์ในระดับสูงมาก และมีนัยสำคัญทางสถิติที่ ระดับ .01 ทั้งนี้อาจเนื่องจากวิธี HGLM-2L วิธี MIMIC และวิธี BAYESIAN อุยุบันพื้นฐานของการวิเคราะห์ คุณภาพของข้อสอบตามทฤษฎีการตอบสนองข้อสอบ (IRT) (1-PL) parameter logistic measurement model (1-PL) เช่นเดียวกัน จึงทำให้มีความสัมพันธ์ในระดับที่สูงมาก

2.2 เปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) สำหรับผู้สอบจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน ดังนี้

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย พนข้อสอบที่เกิดการทำหน้าที่ต่างกันของข้อสอบ ส่วนใหญ่ ตรวจพับการทำหน้าที่ต่างกันของข้อสอบจากโปรแกรม WinBUGS ด้วยวิธี BAYESIAN ทั้งนี้ อาจเนื่องจากการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ วิธี BAYESIAN มีค่าความคลาดเคลื่อนมาตรฐาน (S.E) น้อยกว่าวิธี HGLM-2L และวิธี MIMIC ซึ่งการประยุกต์ใช้ โปรแกรม WinBUGS ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยใช้เทคนิค Markov Chain Monte Carlo (MCMC) มีการแจกแจงเริ่มต้น (prior distribution) ทำให้ ประมาณผลลอกอกมาใกล้เคียงกับความเป็นจริง และสอดคล้อง กับผลการศึกษาของ Saengla Chaimongkol1, Fred W. Hufferand Akihito Kamata (2007) ได้ศึกษาเรื่องตัวแบบ อธิบายการทำหน้าที่ต่างกันของข้อสอบโดยใช้ WinBUGS1.4 เสนอตัวแบบพหุระดับการทดสอบโดยโลจิสติกที่ ใช้ในการตรวจสอบสาเหตุของการทำหน้าที่ต่างกันของข้อสอบ (DIF) โดยที่ตัวแบบที่เสนอจะพิจารณาโครงสร้าง ของข้อมูลที่มีการข้อนับกัน 3 ระดับ ที่มีการรวมผลลัพธ์ที่ ได้จากการวิเคราะห์การทดสอบโดยโลจิสติก เพื่อที่จะระบุ



ลักษณะตัวแปรของข้อมูลระดับ 3 ที่สามารถใช้อธิบายสาเหตุการผันแปรของ DIF ได้ การศึกษาครั้งนี้จะใช้วิธีการจำลองข้อมูลในการตรวจสอบความถูกต้องและความเหมาะสมของตัวแบบ โดยที่ค่า พารามิเตอร์ต่างๆในตัวแบบจะถูกประมาณโดยการใช้หลักการของเบส์ที่ใช้โปรแกรม WinBUGS

ส่วนผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ พบที่มีข้อสอบที่เกิดการทำหน้าที่ต่างกันของข้อสอบ ส่วนใหญ่ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจากโปรแกรม HLM ด้วยวิธี HGLM-2L เช่นเดียวกับการตรวจสอบจากโปรแกรม WinBUGS ด้วยวิธี BAYESIAN ที่เป็นเช่นนี้อาจเนื่องจากวิธี HGLM-2L เป็นวิธีที่มีความไวในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อเทียบกับการตรวจสอบด้วยวิธี BAYESIAN และวิธี HGLM-2L ในกรณีที่ระดับที่ 1 ระดับข้อสอบที่สอดแทรกในผู้สอบ (between item within person) และระดับที่ 2 ระดับของบุคคล ที่สอดแทรกอยู่ในโรงเรียน (between person within school) ผลการวิเคราะห์ระดับนี้จะได้ค่าพารามิเตอร์ความสามารถของผู้สอบ ซึ่งวิชาคณิตศาสตร์ เป็นวิชาที่ต้องมีทักษะทางเป็นการคิดวิเคราะห์ที่ซับซ้อน ไม่สามารถแสดงออกได้โดยตรง จำเป็นต้องอาศัยเครื่องมือวัดเพื่อให้แสดงถึงกระบวนการคิดทางสมองที่อยู่ในตัวบุคคล อาจเป็นไปได้วิธี HGLM-2L นี้มีความไวในการตรวจสอบวิชาคณิตศาสตร์ได้เหมือนกับวิธี BAYESIAN เพราการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี BAYESIAN โดยโปรแกรม WinBUGS เป็นการใช้เทคนิค Markov Chain Monte Carlo (MCMC) มีการแจกแจงเริ่มต้น (prior distribution) ทำให้การประมาณผลออกมากล้าเลียงกับความเป็นจริง

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ พบที่มีข้อสอบที่เกิดการทำหน้าที่ต่างกันของข้อสอบ ส่วนใหญ่ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจากโปรแกรม HLM ด้วยวิธี HGLM-2L ที่เป็นเช่นนี้อาจเนื่องจากวิธี HGLM-2L มีความไวในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ดีกว่าวิธี MIMIC และวิธี BAYESIAN เมื่อจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน โดยการวิเคราะห์ข้อสอบพหุ

ระดับสามารถดำเนินการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ได้โดยอาศัยหลักการศึกษาผลของตัวแปรภายนอกที่มีต่อค่าพารามิเตอร์ความสามารถของผู้สอบ โดยตัวแปรภายนอกสามารถจัดกระทำในลักษณะของข้อมูลแบบ 2 ค่า (dichotomous)

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาภาษาไทย วิชาคณิตศาสตร์ และวิชาวิทยาศาสตร์ จากโปรแกรม HLM ด้วยวิธี HGLM-2L สอดคล้องกับผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ได้จากโปรแกรม Mplus ด้วยวิธี MIMIC และสอดคล้องกับโปรแกรม WinBUGS ด้วยวิธี BAYESIAN มีความสัมพันธ์ในระดับสูงมาก และมีนัยสำคัญทางสถิติที่ระดับ .01 ทั้งนี้อาจเนื่องมาจากการทั้ง 3 วิธี คือ วิธี HGLM วิธี MIMIC และวิธี BAYESIAN ดังอยู่บนพื้นฐานทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) เมื่อมองกันและสอดคล้องกับผลการศึกษาของ French & Miller (1996) ที่พบว่าเมื่อกลุ่มตัวอย่างมีขนาดเล็กลง จำนวนในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบลดลง และวิธีที่มีพื้นฐานมาจากการใช้ทฤษฎีการตอบสนองข้อสอบนั้น ยังคงมีขนาดกลุ่มตัวอย่างใหญ่ขึ้นเพียงใด จะส่งผลให้การประมาณค่าพารามิเตอร์ของข้อสอบได้ดียิ่งขึ้น

โดยสรุป ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์ ระหว่างวิธี HGLM-2L วิธี MIMIC และวิธี BAYESIAN จำแนกตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน พบว่า เมื่อจำแนกตามเพศจะพบข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด รองลงมานี้คือ จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน

ซึ่งสอดคล้องกับการศึกษาของเกสร ห่วงจิตร (2539) พบว่า ผลการระบุลักษณะของข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบระหว่างกลุ่มผู้สอบ เมื่อจำแนกตามตัวแปรเพศ ภูมิลำเนา ประสบการณ์ในการสอบ และสังกัดของสถานศึกษา มีความแตกต่างกัน โดยเมื่อจำแนกผู้สอบตามเพศพบว่ามีข้อสอบที่มีการระบุว่าเกิด DIF มากที่สุด รองลงมานี้คือ ภูมิลำเนา สังกัดของสถานศึกษา ตามลำดับ



3. ผลการศึกษาลักษณะของข้อสอบที่เกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF)

ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) ทั้ง 3 วิธี ได้แก่ วิธี HGLM วิธี MIMIC และวิธี BAYESIAN ด้วยการวิเคราะห์ลักษณะและเนื้อหาของคำหรือข้อความที่ใช้ในการเขียนข้อสอบ โดยจำแนกตามเพศ ซึ่งแต่ละวิธีสามารถตรวจพบรหัสของข้อสอบที่ทำหน้าที่ต่างกันแตกต่างกันไป ผู้วิจัยจึงขอถือว่าลักษณะของข้อสอบที่เกิดการทำหน้าที่ต่างกัน ดังนี้

ข้อสอบวิชาภาษาไทยที่เกิดการทำหน้าที่ต่างกันของข้อสอบจากทั้ง 3 วิธี คือ วิธี HGLM-2L วิธี MIMIC และวิธี BAYESIAN ได้แก่ข้อที่ 24 และข้อที่ 30 ส่วนข้อสอบวิชาคณิตศาสตร์ที่เกิดการทำหน้าที่ต่างกันของข้อสอบจากทั้ง 2 วิธี คือ วิธี HGLM-2L และวิธี BAYESIAN ได้แก่ข้อที่ 30 และข้อสอบวิชาวิทยาศาสตร์ที่เกิดการทำหน้าที่ต่างกันของข้อสอบจากทั้ง 3 วิธี คือ วิธี HGLM-2L วิธี MIMIC และวิธี BAYESIAN ได้แก่ข้อที่ 6,7,13 และข้อที่ 24

จากลักษณะของข้อสอบที่เกิดการทำหน้าที่ต่างกันในภาพรวมทั้ง 3 วิชา โดยจำแนกตามเพศ ลักษณะของข้อสอบเป็นปัจจัยที่มีผลต่อความถูกต้องและความผิดพลาดในการตรวจ สอบการทำหน้าที่ต่างกันของข้อสอบ(Narayanan & Swamtnathan. 1994) ข้อสอบ NT ปี 2553 มีลักษณะเป็นข้อสอบปรนัยแบบเลือกตอบ (Multiple choice) ชนิด 4 ตัวเลือก ทั้งนี้อาจเนื่องจากตัวเลือกสามารถตอบได้หลายคำตอบตามคิดเห็นของแต่ละบุคคล ข้อสอบบางข้อ ตัวเลือกไม่ได้เรียงจากข้อความสันไปทางข้อความยawa และข้อสอบที่เกิดการทำหน้าที่ต่างกันของข้อสอบ จะมีคำศัพท์ที่เกี่ยวข้องกับเพศนั้น จึงทำให้ข้อสอบเข้าข้างเพศนั้น และทั้งนี้อาจเป็นเพราะความสามารถที่แตกต่างระหว่างเพศชายและเพศหญิงที่มีลักษณะความสามารถ ความถนัด และความสนใจในเรื่องนั้นๆต่างกัน

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องพบว่า ลักษณะของข้อสอบคณิตศาสตร์ที่เพศชายจะทำได้ดีกว่าเพศหญิง ได้แก่ ข้อสอบเรขาคณิตและพีชคณิต ที่มีเหตุผลเชิงคณิตศาสตร์ งานที่มีลักษณะเป็นการแก้ปัญหา (Problem Solving) และงานที่ใช้ความคิดในขั้นสูง (higher level cognitive) ข้อสอบเกี่ยวกับอัตราส่วน สัดส่วน (proportions) หรือเปอร์เซ็นต์ และข้อสอบที่มีตาราง กราฟ หรือภาพ ส่วนลักษณะของข้อสอบคณิตศาสตร์ที่เพศหญิงจะทำได้ดีกว่าเพศชาย ได้แก่ การคิดคำนวณ พีชคณิต ข้อสอบที่ใช้สัญลักษณ์และข้อสอบคณิตศาสตร์ที่เป็นนามธรรม (abstract mathematics) และงานที่มีลักษณะคล้ายคลึงกับใบคำ แต่ในสภาพจริงของเรียนการสอนในประเทศไทย อาจมีปัญหาและตัวแปรที่เป็นปัจจัยให้ผลการศึกษาไม่เป็นไปตามหลักการที่อ้างอิงจากการศึกษาของต่างประเทศ เช่น พฤติกรรมการสอนของครู พฤติกรรมการเรียนของนักเรียน (อุทัยวรรณ สายพัฒนา, 2547) และอาจเนื่องจากความสามารถที่แตกต่างระหว่างเพศชายและเพศหญิงที่มีลักษณะความสามารถ ความถนัด และความสนใจในเรื่องนั้นๆ ต่างกัน ส่วนไม่เอกสาร กล่าวว่า เพศชายจะมีความสามารถในการแก้ปัญหาทางคณิตศาสตร์ และเพศหญิงจะมีความสามารถในทางภาษา (Maier & Casselman; 1970)

ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) ทั้ง 3 วิธี ได้แก่ วิธี HGLM วิธี MIMIC และวิธี BAYESIAN ด้วยการวิเคราะห์ลักษณะและเนื้อหาของคำหรือข้อความที่ใช้ในการเขียนข้อสอบ โดยจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน มีลักษณะของข้อสอบที่เกิดการทำหน้าที่ต่างกัน ดังนี้

ข้อสอบวิชาภาษาไทยที่เกิดการทำหน้าที่ต่างกันของข้อสอบจากทั้ง 2 วิธี คือ วิธี HGLM-2L และวิธี BAYESIAN ได้แก่ข้อที่ 4 และข้อที่ 26 เมื่อพิจารณาข้อสอบวิชาคณิตศาสตร์ที่เกิดการทำหน้าที่ต่างกันของข้อสอบจากทั้ง 2 วิธี คือ วิธี HGLM-2L และวิธี BAYESIAN ได้แก่ข้อที่ 12 และเมื่อพิจารณาข้อสอบวิชาวิทยาศาสตร์ที่เกิดการทำหน้าที่ต่างกันของข้อสอบจากทั้ง 2 วิธี คือ วิธี HGLM-2L และวิธี BAYESIAN ได้แก่ข้อที่ 18 และข้อที่ 27

จากลักษณะของข้อสอบที่เกิดการทำหน้าที่ต่างกันในภาพรวมทั้ง 3 วิชา โดยจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน ทั้งนี้อาจเนื่องมาจากการสนับสนุน ความคุ้นเคยเกี่ยวกับเรื่องนั้น สภาพแวดล้อมและการฝึกปฏิบัติการที่แตกต่างกันระหว่างนักเรียนในเขตกรุงเทพมหานครและปริมณฑลนักเรียนนอกเขต



กรุงเทพมหานครและปริมณฑล ซึ่งสอดคล้องกับการศึกษาของอุมาภรณ์ ภัทรวาณิชย์ และคณะ (2550) ซึ่งได้ศึกษาการขยายโอกาสทางการศึกษาเป็นนโยบายที่สำคัญของหลายประเทศ เนื่องจากทำให้ความไม่เท่าเทียมด้านเศรษฐกิจและสังคมลดลง มีหลักฐานที่สำคัญที่ทำให้เห็นความสัมพันธ์ระหว่างการพัฒนาด้านการศึกษากับการพัฒนาทางเศรษฐกิจและสังคม (Hill and King 1993; Baker and Holsinger 1996) ในระดับมหาวิทยาลัย การศึกษา มีความสัมพันธ์กับการเพิ่มประสิทธิภาพของกลั่งแรงงาน เพิ่มอายุขัยเฉลี่ยของประชากร และปรับปรุงด้านสุขภาพของประชากร รวมถึงลดภาวะเริ่มพัฒนาด้วย (Schultz 1993; Raghupathy 1996; Axinn and Barber 2001) งานวิจัยของ Knodel and Jones (1996) ยืนยันว่าซึ่งว่าระหว่างเพศกลั่งจะลดน้อยลง ไม่ว่าจะในภูมิภาคใด และความแตกต่างของการศึกษาเกิดขึ้นในประเทศที่มีคุณภาพสถานภาพทางเศรษฐกิจและสังคมที่แตกต่างกันอย่างมาก ทราบเท่าที่เด็กยากจนหรือเด็กจากครอบครัวชนบททั้งเด็กหญิงและชาย ยังคงไม่ได้รับความเป็นธรรมด้านการได้รับการศึกษา

ข้อเสนอแนะในการวิจัย

1. ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1.1 การวิเคราะห์การประมาณค่าพารามิเตอร์ ความยากของข้อสอบ พารามิเตอร์ความสามารถของผู้สอบ และการทำหน้าที่ต่างกันของข้อสอบ ควรเลือกวิเคราะห์ด้วยวิธี MIMIC จากโปรแกรม Mplus เพราะมีกระบวนการในการวิเคราะห์ที่ง่าย ใช้งานสะดวก และไม่ซับซ้อน โดยการเขียนคำสั่งและวิเคราะห์ในขั้นตอนเดียว ซึ่งผลการวิเคราะห์ได้ค่าที่ใกล้เคียงและสอดคล้องกับวิธี HGLM-2L และวิธี BAYESIAN

1.2 การวิเคราะห์ด้วยวิธี BAYESIAN จากโปรแกรม WinBUGS มีความคล้ายคลึงกับวิธี HGLM-2L และวิธี MIMIC โดยโปรแกรม WinBUGS ใช้เทคนิค Markov Chain Monte Carlo (MCMC) มีการแจกแจงเริ่มต้น (prior distribution) ทำให้การประมาณผลออกมากใกล้เคียงกับความเป็นจริง

1.3 การศึกษาครั้งนี้ได้ข้อค้นพบว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี BAYESIAN โดยการเขียนคำสั่งในโปรแกรม WinBUGS และการใช้ข้อมูลจริง (real data) มีวิธีการเขียนคำสั่งที่ซับซ้อนค่อนข้างใช้เวลาในการศึกษา ควรเลือกวิเคราะห์ด้วย วิธี HGLM-2L และวิธี MIMIC

1.4 ถ้าต้องการเพิ่มตัวแปรภายนอกเข้าสู่สมการ เพื่ออธิบายความผันแปรที่เกิดขึ้น เมื่อพบว่าข้อสอบที่วิเคราะห์ขึ้นได้มีความผันแปรระหว่างโรงเรียน ตามอิทธิพลของตัวแปรความแตกต่างด้านเพศ และสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน สามารถใส่ตัวแปรพยากรณ์ในระดับโรงเรียนเข้าสู่สมการ เพื่ออธิบายความผันแปรที่เกิดขึ้นในข้อที่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบได้ โดยการวิเคราะห์ด้วยวิธี HGLM-2L จากโปรแกรม HLM

1.5 ควรมีการปรับปรุง แก้ไขลักษณะของข้อสอบ ทั้งลักษณะของคำถามและตัวเลือกให้มีความเป็นปรนัยมากขึ้น ในข้อที่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ เมื่อจำแนกเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน เพื่อหลีกเลี่ยงการนำไปสู่การทำหน้าที่ต่างกันของข้อสอบ

2. ข้อเสนอแนะสำหรับการทำวิจัยครั้งต่อไป

2.1 ควรศึกษาการประมาณค่าพารามิเตอร์ของข้อสอบ ด้วยวิธี BAYESIAN แบบ 3PL โดยการเขียนคำสั่งในสมการ จากการประยุกต์ใช้โปรแกรม WinBUGS

2.2 ควรศึกษาเบรย์บินค่าพารามิเตอร์ของข้อสอบแบบ 2 PL และพารามิเตอร์ความสามารถของผู้สอบด้วยวิธี MIMIC และวิธี BAYESIAN กับโปรแกรมอื่นๆ เช่น TESTFACT, BILOG-MG, MULTILOG, XCALIBRE ที่อยู่บนพื้นฐานทฤษฎีการตอบสนองข้อสอบ (IRT) เช่นเดียวกัน

2.3 เนื่องจากข้อจำกัดของวิธี HGLM-2L ใน การวิเคราะห์ค่าพารามิเตอร์ข้อสอบได้เพียงค่าพารามิเตอร์ความยากของข้อสอบ จึงควรมีการเบรย์บินค่าพารามิเตอร์ของข้อสอบ แบบ 2 PL และพารามิเตอร์ความสามารถของผู้สอบด้วยวิธี MIMIC และวิธี BAYESIAN

2.4 ควรศึกษาเปรียบเทียบประสิทธิภาพของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในแบบทดสอบที่มีการให้คะแนนแบบ 2 ค่า (dichotomous) และการให้คะแนนแบบหลายค่า (polytomous) โดยใช้วิธีการจำลองข้อมูล (simulation data)

2.5 ควรศึกษาเปรียบเทียบประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี

เอกสารอ้างอิง

- เกรศ ห่วงจิตร. (2539). การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบสำหรับแบบสอบถามคัดเลือกระดับบัณฑิตศึกษา วิชาภาษาไทยและภาษาอังกฤษด้วยวิธีแม่นเหล็ก. วิทยานิพนธ์ปริญญามหาบัณฑิต. ภาควิชาวิจัยการศึกษา จุฬาลงกรณ์มหาวิทยาลัย.
- วิชุดา บัวคง. (2533). การเปรียบเทียบประสิทธิผลของวิธีประมาณค่าพารามิเตอร์ของแบบจำลอง โลจิสติก 3 พารามิเตอร์ ระหว่างวิธีแมกซิมัลล์ลิลิอูด วิธีเชิร์ลิสติก และวิธีของเบส ในแบบสอบถามวัดผลสัมฤทธิ์และแบบสอบถามความถนัด. วิทยานิพนธ์ ค.ม. กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.
- ศิริชัย กานุจนาเส. (2548). (เล่ม ๑). ทฤษฎีการทดสอบแบบดั้งเดิม (CLASSICAL TEST THEORIES). พิมพ์ครั้งที่ ๕. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- สุนทร เทียนงาม. (2551). ผลของการไม่เป็นอิสระของข้อสอบที่มีต่อค่าความเที่ยงค่าพารามิเตอร์ของข้อสอบ ค่าความสามารถของผู้สอบ และค่าสารสนเทศของแบบสอบถามเมื่อมีเงื่อนไขการทดสอบที่แตกต่างกัน. วิทยานิพนธ์ปริญญาครุศาสตร์ดุษฎี บัณฑิตภาควิชาวิจัยและจิตวิทยาการศึกษา. บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- อิทธิฤทธิ์ พงษ์ปิยะรัตน์. (2551). การวิเคราะห์ข้อสอบและการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ : การวิเคราะห์พหุระดับ. วิทยานิพนธ์ปริญญาครุศาสตร์ดุษฎีบัณฑิต ภาควิชาวิจัยและจิตวิทยาการศึกษา. บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- อุมากรณ์ ภัทรภานิชย์ และปัทมา อุมรสถิรสมบูรณ์. (2550). ความไม่เท่าเทียมด้านการศึกษา: เมืองและชนบท. ใน ประชากรและสังคม 2550. วชิรชัย ทองไทย และสุรีย์พร พันพึ่ง. นครปฐม: สำนักพิมพ์ประชากรและสังคม.
- อุทัยวรรณ สายพัฒน์. (2547). การเปรียบเทียบประสิทธิภาพของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในแบบทดสอบที่มีการให้คะแนนแบบหลายค่าระหว่างวิธี GMH และวิธี Polytomous SIBTEST. ดุษฎีบัณฑิต กศ.ค. ดุษฎีนิพนธ์ กรุงเทพฯ: มหาวิทยาลัยศรีนครินทร์วิโรฒประสานมิตร.
- French, A. W., & Miller, T.R. (1996). Logistic Regression and Its Use in Detection Differential Item Functioning in Polytomous Items. *Journal of Education Measurement*, 33: 315-332
- Fukuhara,H., & Kamata, A (2007). DIF Detection in a Presence of Locally Dependent Items. Paper presented at the annual meeting of the Florida Educational Research Association,Tampa,FL.
- Hill, M. Anne and Elizabeth M. King. 1993. "Women's Education in Developing Countries: An Overview." Pp. 1-50 in Women's Education in Developing Countries, edited by Elizabeth M. King and Anne M. Hill. Baltimore: The Johns Hopkins University Press.

HGLM-2L วิธี MIMIC และวิธี BAYESIAN โดยใช้ผลการตอบข้อสอบ i-net ของนักเรียน วิเคราะห์ตามตัวแปรอื่นๆ เช่น การใช้ภาษาในชีวิตประจำวัน โดยกำหนดให้ภาษาไทย เป็น 1 (focal group) และไม่ใช่ภาษาไทย เป็น 0 (reference group)



- Kamata, A. (2001). Item Analysis by the Hierarchical Generalized Linear model. *Journal of Educational Measurement*, 38(1): 79-93.
- Kim, W. (2003). Development of a Differential Item Functioning (DIF) Procedure using the Hierarchical Generalized Linear Model: A Comparison Study with Logistic Regression Procedure. Doctoral Dissertation in Education Psychology, The Pennsylvania State University.
- Knodel, J. and Jones G.W. 1996. "Post-Cairo Population Policy: Does Promoting Girls' Schooling Miss the Mark?" *Population and Development Review*, 22(4), pp.683-702.
- Maier, N.R.F. and Casselman, G.G. (1970). Locating the difficulty in insight problems: Individual and sex differences *Psychological Rep*, 26, 103-77.
- Muthen, B. O. (1988). Some uses of structural equation modeling in validity studies: Extending IRT to external variables. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 213-238). Hillsdale, NJ: Lawrence Erlbaum.
- Muthen, B. O., Kao, C. F., & Burstein, L. (1991). Instructionally sensitive psychometrics: Application of a new IRT-based detection technique to mathematics achievement test items. *Journal of Educational Measurement*, 28(1), 1-22.
- Narayanan, P., and Swaminathan, H. Identification of items that show nonuniform DIF. *Applied Psychological Measurement*. 20(3):257-274.
- Saengla Chaimongkol., Fred W. Huffer & Akihito Kamata. An explanatory differential item functioning (DIF) model by the WinBUG 1.4. *Songklanakarin J. Sci. Technol.*, 2007, 29(2): 449-458.
- Swaminathan, H., & Gifford, J. (1985). Bayesian estimation in the two-parameter logistic model. *Psychometrika*, 50, 349-364
- Schultz, T. Paul, 1993. "Returns to Women's Education," Pp. 51-99 in *Women's Education in Developing Countries: Barriers, Benefits, and Policies*, edited by Elizabeth M. King, and Anne M. Hill. Published for the World Bank by John's Hopkins University Press: Baltimore.