

บทที่ 5

สรุป อภิปรายผล และข้อเสนอแนะ

การวิจัยครั้งนี้มีวัตถุประสงค์ คือ 1) เพื่อเปรียบเทียบผลการประมาณค่าพารามิเตอร์ ความยากของข้อสอบ (δ_i) พารามิเตอร์ความสามารถของผู้สอบ (θ_j) สำหรับผู้สอบจำแนก ตามเพศ สถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน ระหว่างวิธี HGLM วิธี MIMIC และวิธี BAYESIAN 2) เพื่อเปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) สำหรับผู้สอบ จำแนกตามเพศ สถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน ระหว่างวิธี HGLM วิธี MIMIC และ วิธี BAYESIAN และ 3) เพื่อศึกษาลักษณะของข้อสอบที่เกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF) ที่ได้จากการวิเคราะห์การทำหน้าที่ต่างกัน โดยวิธี HGLM วิธี MIMIC และวิธี BAYESIAN ด้วยการวิเคราะห์ลักษณะและเนื้อหาของคำหรือข้อความที่ใช้ในการเขียนข้อสอบ

ประชากรที่ใช้ในการวิจัยครั้งนี้ เป็นนักเรียนชั้นประถมศึกษาปีที่ 3 ในสังกัดสำนักงาน คณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ ที่เข้าสอบวัดผลสัมฤทธิ์ทางการเรียน เพื่อประเมินคุณภาพการศึกษาระดับชาติปีการศึกษา 2553 ซึ่งเป็นนักเรียนที่เข้าสอบวิชา ภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์ จำนวนทั้งหมด 592,525 คน จำแนกตามเพศชาย จำนวน 307,306 คน เพศหญิง 285,219 คน

ในการกำหนดขนาดกลุ่มตัวอย่างสำหรับการวิจัย ผู้วิจัยได้ศึกษาการกำหนดขนาด กลุ่มตัวอย่างจากประชากรตามการวิเคราะห์คุณภาพตามทฤษฎีการทดสอบตอบสนองข้อสอบ (Item-Response Theory: IRT) และตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning: DIF) ซึ่งตามเงื่อนไขแล้วกลุ่มตัวอย่างไม่ต่ำกว่า 10 – 20 เท่าของจำนวนข้อคำถาม (นงลักษณ์ วิรัชชัย, 2542, หน้า 311; สุวิมล ติรภานันท์, 2553, หน้า 168; สุกมาศ อังศุโชติ สมถวิล วิจิตรวรรณ และรัชนิภา ภิญญภาณุวัฒน์, 2554, หน้า 31) ผู้วิจัยจึงเพิ่มกลุ่มตัวอย่าง ในการวิจัยครั้งนี้ ให้ได้อย่างน้อย 10 เท่า คือประมาณ 900 คน (ข้อคำถาม 90 ข้อ) ในการศึกษา ครั้งนี้ใช้กลุ่มตัวอย่าง 1,000 คน จำแนกเป็นเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน โดยมีขั้นตอนการสุ่มตัวอย่าง แบบหลายขั้นตอน (Multi – stage Random Sampling Technique) ดังนี้

1. จำแนกการแบ่งประชากร ในแต่ละจังหวัดออกเป็นประชากรย่อย (Subpopulation) โดยในแต่ละจังหวัดจะมีโรงเรียนเป็นสมาชิกอยู่หลายโรงเรียน ผู้วิจัยกำหนดให้แต่ละโรงเรียน ถือเป็น 1 กลุ่ม (Cluster) โดยที่ทุกโรงเรียนภายในชั้นภูมิเดียวกันหรือในเขตพื้นที่เดียวกัน

จะมีความแตกต่างกันมากที่สุดและระหว่างสถานศึกษาภายในชั้นภูมิเดียวกันหรือในเขตพื้นที่เดียวกันจะมีความคล้ายคลึงกันมากที่สุด การแบ่งลักษณะนี้ยึดวิธีการสุ่มแบบกลุ่ม (Cluster Sampling)

2. ดำเนินการสุ่มโดยใช้วิธีการสุ่มตัวอย่างแบบกลุ่ม (Cluster Random Sampling) ซึ่งกำหนดให้โรงเรียนเป็นหน่วยการสุ่ม (Sampling Unit) ผู้วิจัยจึงสุ่มโรงเรียนจากจังหวัด

3. จากโรงเรียน ที่สุ่มได้ในข้อที่ 2 เมื่อใช้สถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนเป็นตัวแปรจำแนกกลุ่ม (Classify Variable) พบว่า แต่ละโรงเรียนจะถูกแบ่งออกเป็น 2 ชั้นภูมิ (Stratum) ได้แก่ ในเขตกรุงเทพมหานครและปริมณฑล และนอกเขตกรุงเทพมหานครและปริมณฑล จากนั้นดำเนินการสุ่มโดยใช้วิธีการสุ่มตัวอย่างแบบชั้นภูมิ (Stratified Random Sampling)

4. ในแต่ละชั้นภูมิ (Stratum) ที่ได้จากแบบแผนการสุ่มในข้อที่ 3 นั้น เมื่อทำการแบ่งอีกครั้ง โดยใช้เพศเป็นตัวแปรจำแนกกลุ่ม เราจะได้ชั้นภูมีย่อย ๆ อีก 2 ชั้นภูมิ คือ เพศชายและเพศหญิง ผู้วิจัยดำเนินการสุ่มโดยใช้วิธีการสุ่มตัวอย่างแบบชั้นภูมิ (Stratified Random Sampling) ได้กลุ่มตัวอย่างจำนวน 1,000 คน

การวิเคราะห์ข้อมูล เริ่มต้นจากการสร้างไฟล์ข้อมูลสำหรับการวิเคราะห์ด้วยโปรแกรม HLM โปรแกรม Mplus และโปรแกรม WinBUGS เพื่อให้ได้ไฟล์ข้อมูลที่มีความเหมาะสมสำหรับนำเข้าสู่การวิเคราะห์ในแต่ละโปรแกรมซึ่งมีลักษณะของ Data Input ที่แตกต่างกัน แล้วดำเนินการวิเคราะห์ข้อมูลเพื่อตอบคำถามการวิจัยดังนี้

1. เปรียบเทียบผลการประมาณค่าพารามิเตอร์ความยากของข้อสอบ (δ_i) และพารามิเตอร์ความสามารถของผู้สอบ (θ_j) จำแนกตามเพศ และสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน ด้วยวิธี HGLM – 2L จากโปรแกรม HLM วิธี MIMIC จากโปรแกรม Mplus และวิธี BAYESIAN จากโปรแกรม WinBUGS แล้วศึกษาความสอดคล้องของผลการประมาณค่าพารามิเตอร์ที่คำนวณได้ด้วยการหาค่าสัมประสิทธิ์สหสัมพันธ์

2. เปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) สำหรับผู้สอบจำแนกตามเพศ และสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน ระหว่างวิธี HGLM – 2L วิธี MIMIC และวิธี BAYESIAN

3. ศึกษาลักษณะของข้อสอบที่เกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF) ที่ได้จากการวิเคราะห์การทำหน้าที่ต่างกัน โดยวิธี HGLM วิธี MIMIC และวิธี BAYESIAN ด้วยการวิเคราะห์ลักษณะและเนื้อหาของคำหรือข้อความที่ใช้ในการเขียนข้อสอบ

สรุปผลการวิจัย

จากการวิเคราะห์ค่าพารามิเตอร์ความยากของข้อสอบ (δ_i) และค่าพารามิเตอร์ความสามารถของผู้สอบ (θ_j) ครั้งนี้ ประกอบด้วยวิชาภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์ ซึ่งผู้วิจัยได้ทำการวิเคราะห์ด้วยวิธี HGLM – 2L ประยุกต์ใช้โปรแกรม HLM แล้วนำผลการวิเคราะห์มาเปรียบเทียบกับผลการประมาณค่าด้วยวิธี MIMIC ใช้โปรแกรม Mplus และด้วยวิธี BAYESIAN ใช้โปรแกรม WinBUGS14 ได้ดังนี้

1. ผลการเปรียบเทียบการประมาณค่าพารามิเตอร์ความยากของข้อสอบและพารามิเตอร์ความสามารถของผู้สอบ

1.1 เปรียบเทียบผลการประมาณค่าพารามิเตอร์ความยากของข้อสอบ

ค่าพารามิเตอร์ความยากของข้อสอบ (δ_i) สามารถคำนวณได้โดยนำค่าสัมประสิทธิ์ของอิทธิพลคงที่ (Fixed Effect) ของข้อสอบแต่ละข้อลบออกจากข้อสอบข้ออ้างอิงผลของการประมาณค่าพารามิเตอร์ความยากของข้อสอบ ทั้งวิชาภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์ ด้วยวิธี HGLM – 2L ซึ่งผลของการประมาณค่าพารามิเตอร์ความยากของข้อสอบวิชาภาษาไทย พบว่า วิธี HGLM – 2L สอดคล้องสัมพันธ์กับผลการประมาณค่าพารามิเตอร์ความยากของข้อสอบด้วยวิธี MIMIC และสอดคล้องสัมพันธ์กับผลการประมาณค่าพารามิเตอร์ความยากของข้อสอบด้วยวิธี BAYESIAN อย่างสมบูรณ์ (1.00) และมีนัยสำคัญทางสถิติที่ระดับ .01

ส่วนผลการประมาณค่าพารามิเตอร์ความยากของข้อสอบวิชาคณิตศาสตร์ ค่าสัมประสิทธิ์ สัมพันธ์ของค่าพารามิเตอร์ความยากทั้ง 3 วิธี พบว่า วิธี HGLM – 2L วิธี MIMIC และวิธี BAYESIAN มีความสัมพันธ์ในระดับสูงมาก เมื่อพิจารณาผลการประมาณค่าด้วยวิธี MIMIC เทียบกับผลการประมาณค่าด้วยวิธี BAYESIAN พบว่า δ_{MIMIC} มีความสัมพันธ์กับ δ_{BAYESIAN} อย่างสมบูรณ์ และมีนัยสำคัญทางสถิติที่ระดับ .01 รองลงมา คือ ผลการประมาณค่าด้วยวิธี HGLM – 2L เทียบกับผลการประมาณค่าด้วยวิธี BAYESIAN พบว่า $\delta_{\text{HGLM-2L}}$ สัมพันธ์กับ δ_{MIMIC} อยู่ในระดับสูงมาก เท่ากับ .999 และมีนัยสำคัญทางสถิติที่ระดับ .01 และผลการประมาณค่าด้วยวิธี HGLM – 2L เทียบกับผลการประมาณค่าด้วยวิธี BAYESIAN พบว่า $\delta_{\text{HGLM-2L}}$ สัมพันธ์กับ δ_{BAYESIAN} อยู่ในระดับสูงมาก เท่ากับ .999 และมีนัยสำคัญทางสถิติที่ระดับ .01

ผลการประมาณค่าพารามิเตอร์ความยากของข้อสอบวิชาวิทยาศาสตร์ พบว่า วิธี HGLM – 2L สอดคล้องสัมพันธ์กับผลการประมาณค่าพารามิเตอร์ความยากของข้อสอบ

ด้วยวิธี MIMIC และสอดคล้องสัมพันธ์กับผลการประมาณค่าพารามิเตอร์ความยากของข้อสอบ ด้วยวิธี BAYESIAN อย่างสมบูรณ์ (1.00) และมีนัยสำคัญทางสถิติที่ระดับ .01

1.2 การเปรียบเทียบผลการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ

การเปรียบเทียบผลการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ (θ_j)

ในการวิจัยครั้งนี้ ประกอบด้วยวิชาภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์ ซึ่งทั้ง 3 วิชา ใช้วิธีการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบได้แก่ วิธี HGLM – 2L แล้วนำผลการวิเคราะห์มาเปรียบเทียบกับผลการประมาณค่าด้วย วิธี MIMIC และวิธี BAYESIAN โดยประยุกต์ใช้โปรแกรม HLM โปรแกรม Mplus และโปรแกรม WinBUGS ตามลำดับ โดยโปรแกรม HLM ศึกษาผลการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบจาก Residual File ส่วนโปรแกรม Mplus และโปรแกรม WinBUGS ศึกษาผลการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบจากการเขียนคำสั่งโปรแกรมในโปรแกรม Mplus และ WinBUGS ตามลำดับ

ซึ่งผลการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าพารามิเตอร์ความสามารถของผู้สอบ วิชาภาษาไทย ระหว่างวิธี HGLM วิธี MIMIC และวิธี BAYESIAN พบว่า ผลการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ของค่าพารามิเตอร์ความสามารถของผู้สอบ (θ_j) วิชาภาษาไทยของทั้ง 3 วิธี ซึ่งได้แก่ โมเดล HGLM 2 ระดับ จากโปรแกรม HLM วิธี MIMIC และวิธี BAYESIAN มีความสัมพันธ์ในระดับสูงมาก เมื่อพิจารณาผลการประมาณค่าด้วยวิธี HGLM – 2L เทียบกับผลการประมาณค่าด้วยวิธี MIMIC พบว่า $\theta_{\text{HGLM-2L}}$ มีความสัมพันธ์กับ θ_{BAYESIAN} อย่างสมบูรณ์ และมีนัยสำคัญทางสถิติที่ระดับ .01 รองลงมา คือ ผลการประมาณค่าด้วยวิธี HGLM – 2L เทียบกับผลการประมาณค่าด้วยวิธี BAYESIAN พบว่า $\theta_{\text{HGLM-2L}}$ สัมพันธ์กับ θ_{MIMIC} ในระดับสูงมาก เท่ากับ .999 และมีนัยสำคัญทางสถิติที่ระดับ .01 และผลการประมาณค่าด้วยวิธี MIMIC เทียบกับผลการประมาณค่าด้วยวิธี BAYESIAN พบว่า θ_{MIMIC} สัมพันธ์กับ θ_{BAYESIAN} ในระดับสูงมากเช่นกัน เท่ากับ .999 และมีนัยสำคัญทางสถิติที่ระดับ .01

ส่วนผลการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าพารามิเตอร์ความสามารถของผู้สอบ วิชาคณิตศาสตร์ ระหว่างวิธี HGLM วิธี MIMIC และวิธี BAYESIAN พบว่า ผลการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ของค่าพารามิเตอร์ความสามารถของผู้สอบ (θ_j) วิชาคณิตศาสตร์ของทั้ง 3 วิธี ซึ่งได้แก่ โมเดล HGLM 2 ระดับ จากโปรแกรม HLM วิธี MIMIC และวิธี BAYESIAN มีความสัมพันธ์ในระดับสูงมาก เมื่อพิจารณาผลการประมาณค่าด้วยวิธี MIMIC

เทียบกับผลการประมาณค่าด้วยวิธี BAYESIAN พบว่า θ_{MIMIC} มีความสัมพันธ์กับ $\theta_{BAYESIAN}$ ในระดับสูงมากที่สุด เท่ากับ .999 และมีนัยสำคัญทางสถิติที่ระดับ .01 รองลงมาคือ ผลการประมาณค่าด้วยวิธี HGLM – 2L เทียบกับผลการประมาณค่าด้วยวิธี BAYESIAN พบว่า $\theta_{HGLM-2L}$ สัมพันธ์กับ $\theta_{BAYESIAN}$ ในระดับสูงมาก เท่ากับ .996 และมีนัยสำคัญทางสถิติที่ระดับ .01 และผลการประมาณค่าด้วยวิธี HGLM – 2L เทียบกับผลการประมาณค่าด้วยวิธี MIMIC พบว่า $\theta_{HGLM-2L}$ สัมพันธ์กับ θ_{MIMIC} ในระดับสูงมาก เท่ากับ .994 และมีนัยสำคัญทางสถิติที่ระดับ .01

และผลการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าพารามิเตอร์ความสามารถของผู้สอบ วิชาวิทยาศาสตร์ ระหว่างวิธี HGLM วิธี MIMIC และวิธี BAYESIAN พบว่า ผลการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ของค่าพารามิเตอร์ความสามารถของผู้สอบ (θ_j) วิชาวิทยาศาสตร์ของทั้ง 3 วิธี ซึ่งได้แก่ โมเดล HGLM 2 ระดับ จากโปรแกรม HLM วิธี MIMIC และวิธี BAYESIAN มีความสัมพันธ์ในระดับสูงมาก เมื่อพิจารณาผลการประมาณค่าด้วยวิธี MIMIC เทียบกับผลการประมาณค่าด้วยวิธี BAYESIAN พบว่า θ_{MIMIC} มีความสัมพันธ์กับ $\theta_{BAYESIAN}$ ในระดับสูงมากเท่ากับ .999 และมีนัยสำคัญทางสถิติที่ระดับ .01 รองลงมา คือ ผลการประมาณค่าด้วยวิธี HGLM – 2L เทียบกับผลการประมาณค่าด้วยวิธี BAYESIAN พบว่า $\theta_{HGLM-2L}$ สัมพันธ์กับ $\theta_{BAYESIAN}$ ในระดับสูงมาก เท่ากับ .986 และมีนัยสำคัญทางสถิติที่ระดับ .01 และผลการประมาณค่าด้วยวิธี HGLM – 2L เทียบกับผลการประมาณค่าด้วยวิธี MIMIC พบว่า $\theta_{HGLM-2L}$ สัมพันธ์กับ θ_{MIMIC} ในระดับสูงมาก เท่ากับ .981 และมีนัยสำคัญทางสถิติที่ระดับ .01

2. ผลการเปรียบเทียบการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) จำแนกตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน

การศึกษาการทำหน้าที่ต่างกันของข้อสอบในครั้งนี้ประกอบด้วยวิชาภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์ มีจำนวนข้อสอบวิชาละ 30 ข้อ รวมทั้งหมด 90 ข้อ ผู้วิจัยได้ใช้วิธี HGLM – 2L แล้วนำผลการวิเคราะห์มาเปรียบเทียบกับวิธี MIMIC และวิธี BAYESIAN โดยประยุกต์ใช้โปรแกรม HLM โปรแกรม Mplus และโปรแกรม WinBUGS14 ตามลำดับ โดยจำแนกตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน ซึ่งมีการวิเคราะห์ด้วยวิธีต่าง ๆ ดังนี้

2.1 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) จำแนกตามเพศ

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย จากการวิเคราะห์ข้อสอบจำนวน 30 ข้อ พบว่าข้อสอบข้อที่ 1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 18, 20,

ของข้อสอบ (DIF) ของวิธี HGLM – 2L และวิธี BAYESIAN ระหว่างผู้สอบเพศชายและเพศหญิง มีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .959 แสดงว่าค่าดัชนีการทำหน้าที่ต่างกันของข้อสอบทั้ง 2 วิธี มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และค่าดัชนีการทำหน้าที่ต่างกันของข้อสอบ (DIF) ของวิธี MIMIC และวิธี BAYESIAN ระหว่างผู้สอบเพศชายและเพศหญิง มีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .924 แสดงว่าค่าดัชนีการทำหน้าที่ต่างกันของข้อสอบทั้ง 2 วิธี มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

2.2 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย จากการวิเคราะห์ข้อสอบจำนวน 30 ข้อ พบว่าข้อสอบข้อที่ 4, 14, 15, 26 และข้อที่ 29 ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ ซึ่งผลการตรวจสอบค่าดัชนีการทำหน้าที่ต่างกันของข้อสอบ (DIF) ของวิธี HGLM – 2L และวิธี MIMIC ระหว่างผู้สอบในเขตกรุงเทพมหานครและปริมณฑลและนอกเขตกรุงเทพมหานครและปริมณฑล มีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .945 แสดงว่าค่าดัชนีการทำหน้าที่ต่างกันของข้อสอบทั้ง 2 วิธี มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนค่าดัชนีการทำหน้าที่ต่างกันของข้อสอบ (DIF) ของวิธี HGLM – 2L และวิธี BAYESIAN ระหว่างผู้สอบในเขตกรุงเทพมหานครและปริมณฑลและนอกเขตกรุงเทพมหานครและปริมณฑล มีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .975 แสดงว่าค่าดัชนีการทำหน้าที่ต่างกันของข้อสอบทั้ง 2 วิธี มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และค่าดัชนีการทำหน้าที่ต่างกันของข้อสอบ (DIF) ของวิธี MIMIC และวิธี BAYESIAN ระหว่างผู้สอบระหว่างผู้สอบในเขตกรุงเทพมหานครและปริมณฑลและนอกเขตกรุงเทพมหานครและปริมณฑลมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .913 แสดงว่าค่าดัชนีการทำหน้าที่ต่างกันของข้อสอบทั้ง 2 วิธี มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ จากการวิเคราะห์ข้อสอบจำนวน 30 ข้อ พบว่าข้อสอบข้อที่ 12 ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ ซึ่งผลการตรวจสอบค่าดัชนีการทำหน้าที่ต่างกันของข้อสอบ (DIF) ของวิธี HGLM – 2L และวิธี MIMIC ระหว่างผู้สอบในเขตกรุงเทพมหานครและปริมณฑลและนอกเขตกรุงเทพมหานครและปริมณฑล มีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .860 แสดงว่าค่าดัชนีการทำหน้าที่ต่างกันของข้อสอบทั้ง 2 วิธี มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนค่าดัชนีการทำหน้าที่ต่างกัน

ของข้อสอบ (DIF) ของวิธี HGLM – 2L และวิธี BAYESIAN ระหว่างผู้สอบในเขตกรุงเทพมหานคร และปริมณฑลและนอกเขตกรุงเทพมหานครและปริมณฑลมีค่าสัมประสิทธิ์สหสัมพันธ์ เท่ากับ .881 แสดงว่าค่าดัชนีการทำหน้าที่ต่างกันของข้อสอบทั้ง 2 วิธี มีความสัมพันธ์กัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และค่าดัชนีการทำหน้าที่ต่างกันของข้อสอบ (DIF) ของวิธี MIMIC และวิธี BAYESIAN ระหว่างผู้สอบระหว่างผู้สอบในเขตกรุงเทพมหานครและ ปริมณฑลและนอกเขตกรุงเทพมหานครและปริมณฑลมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .781 แสดงว่าค่าดัชนีการทำหน้าที่ต่างกันของข้อสอบทั้ง 2 วิธี มีความสัมพันธ์กันอย่างมีนัยสำคัญ ทางสถิติที่ระดับ .01

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาวิทยาศาสตร์ จากการวิเคราะห์ ข้อสอบจำนวน 30 ข้อ พบว่าพบว่าข้อสอบข้อที่ 1, 2, 11, 12, 15, 18, 26, 27, 29 และข้อที่ 30 ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ ซึ่งผลการตรวจสอบค่าดัชนีการทำหน้าที่ต่างกัน ของข้อสอบ (DIF) ของวิธี HGLM – 2L และวิธี MIMIC ระหว่างผู้สอบในเขตกรุงเทพมหานครและ ปริมณฑลและนอกเขตกรุงเทพมหานครและปริมณฑล มีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .541 แสดงว่าค่าดัชนีการทำหน้าที่ต่างกันของข้อสอบทั้ง 2 วิธี มีความสัมพันธ์กันอย่างมีนัยสำคัญ ทางสถิติที่ระดับ .01 ส่วนค่าดัชนีการทำหน้าที่ต่างกันของข้อสอบ (DIF) ของวิธี HGLM – 2L และ วิธี BAYESIAN ระหว่างผู้สอบในเขตกรุงเทพ มหานครและปริมณฑลและนอกเขตกรุงเทพมหานคร และปริมณฑลมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .902 แสดงว่าค่าดัชนีการทำหน้าที่ต่างกัน ของข้อสอบทั้ง 2 วิธี มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และค่าดัชนี การทำหน้าที่ต่างกันของข้อสอบ (DIF) ของวิธี MIMIC และวิธี BAYESIAN ระหว่างผู้สอบระหว่าง ผู้สอบในเขตกรุงเทพมหานครและปริมณฑลและนอกเขตกรุงเทพมหานครและปริมณฑล มีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .546 แสดงว่าค่าดัชนีการทำหน้าที่ต่างกันของข้อสอบทั้ง 2 วิธี มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

โดยสรุป ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธี HGLM – 2L วิธี MIMIC และวิธี BAYESIAN จำแนกตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน พบว่า ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย เมื่อจำแนกตามเพศ วิธี HGLM – 2L จะพบข้อสอบที่ทำหน้าที่ต่างกันมากที่สุดจำนวน 24 ข้อ ได้แก่ข้อที่ 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 รองลงมาคือ จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน วิธี BAYESIAN พบข้อสอบที่ทำหน้าที่ต่างกัน มากที่สุดจำนวน 5 ข้อ ได้แก่ข้อที่ 4, 14, 15, 26 และข้อที่ 29 ซึ่งเมื่อจำแนกตามเพศ ส่วนใหญ่

ตรวจพบการทำหน้าที่ต่างกันของข้อสอบวิธี HGLM – 2L และเมื่อจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน ส่วนใหญ่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบวิธี BAYESIAN

ส่วนผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ จำแนกตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน พบว่าเมื่อจำแนกตามเพศ วิธี HGLM จะพบข้อสอบที่ทำหน้าที่ต่างกันมากที่สุดจำนวน 21 ข้อ ได้แก่ข้อที่ 2, 3, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 20, 21, 23, 27, 28, 29 และข้อที่ 30 รองลงมาคือ จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน วิธี HGLM และวิธี BAYESIAN พบข้อสอบที่ทำหน้าที่ต่างกันมากที่สุดจำนวน 1 ข้อ ได้แก่ ข้อที่ 12 ซึ่งเมื่อจำแนกตามเพศ ส่วนใหญ่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบวิธี HGLM – 2L และเมื่อจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน ส่วนใหญ่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบวิธี HGLM – 2L และวิธี BAYESIAN

และผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาวิทยาศาสตร์ จำแนกตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน พบว่าเมื่อจำแนกตามเพศ วิธี HGLM จะพบข้อสอบที่ทำหน้าที่ต่างกันมากที่สุดจำนวน 13 ข้อ ได้แก่ข้อที่ 1, 6, 7, 13, 18, 21, 23, 24, 25, 26, 27, 28 และข้อที่ 30 รองลงมาคือ จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน วิธี HGLM พบข้อสอบที่ทำหน้าที่ต่างกันมากที่สุดจำนวน 10 ข้อ ได้แก่ข้อที่ 1, 2, 11, 12, 15, 18, 26, 27, 29 และข้อที่ 30 ซึ่งเมื่อจำแนกตามเพศ ส่วนใหญ่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบวิธี HGLM – 2L และเมื่อจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน ส่วนใหญ่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบวิธี HGLM – 2L

3. ผลการศึกษาลักษณะของข้อสอบที่เกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF)

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

ข้อสอบวิชาภาษาไทยที่เกิดการทำหน้าที่ต่างกันของข้อสอบ ได้แก่ข้อที่ 1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 17, 18, 20, 21, 23, 24, 25, 26, 28, 29 และข้อที่ 30 ซึ่งข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ จากทั้ง 3 วิธี คือ วิธี HGLM – 2L วิธี MIMIC และวิธี BAYESIAN ได้แก่ข้อที่ 24 และข้อที่ 30

ข้อสอบวิชาคณิตศาสตร์ที่เกิดการทำหน้าที่ต่างกันของข้อสอบ ได้แก่ข้อที่ 1, 2, 3, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 20, 21, 22, 23, 27, 28, 29 และข้อที่ 30 ซึ่งข้อสอบ

ที่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ จากทั้ง 2 วิธี คือ วิธี HGLM – 2L และวิธี BAYESIAN ได้แก่ข้อที่ 30

ข้อสอบวิชาวิทยาศาสตร์ที่เกิดการทำหน้าที่ต่างกันของข้อสอบ ได้แก่ข้อที่ 1, 6, 7, 13, 18, 19, 21, 23, 24, 25, 26, 27, 28 และข้อที่ 30 ซึ่งข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ จากทั้ง 3 วิธี คือ วิธี HGLM – 2L วิธี MIMIC และวิธี BAYESIAN ได้แก่ข้อที่ 6, 7, 13 และข้อที่ 24

จากลักษณะของข้อสอบที่เกิดการทำหน้าที่ต่างกันในภาพรวมทั้ง 3 วิชา โดยจำแนกตามเพศ ส่วนใหญ่ข้อสอบที่เกิดการทำหน้าที่ต่างกันของข้อสอบ จะมีคำศัพท์ที่เกี่ยวข้องกับเพศนั้น จึงทำให้ข้อสอบเข้าข้างเพศนั้น และอาจเป็นเพราะความสามารถที่แตกต่างระหว่างเพศชายและเพศหญิงที่มีลักษณะความสามารถ ความถนัด และความสนใจในเรื่องนั้น ๆ ต่างกัน

และผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ(DIF) ทั้ง 3 วิธี ได้แก่ วิธี HGLM – 2L วิธี MIMIC และวิธี BAYESIAN ด้วยการวิเคราะห์ลักษณะและเนื้อหาของคำหรือข้อความที่ใช้ในการเขียนข้อสอบ โดยจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน มีลักษณะของข้อสอบที่เกิดการทำหน้าที่ต่างกัน ดังนี้

ข้อสอบวิชาภาษาไทยที่เกิดการทำหน้าที่ต่างกันของข้อสอบ ได้แก่ข้อที่ 4, 14, 15, 26 และข้อที่ 29 ซึ่งข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ จากทั้ง 2 วิธี คือ

วิธี HGLM – 2L และวิธี BAYESIAN ได้แก่ข้อที่ 4 และข้อที่ 26

เมื่อพิจารณาข้อสอบวิชาคณิตศาสตร์ที่เกิดการทำหน้าที่ต่างกันของข้อสอบ ได้แก่ ข้อที่ 12 ซึ่งข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ จากทั้ง 2 วิธี คือ วิธี HGLM – 2L และวิธี BAYESIAN ได้แก่ข้อที่ 12

และเมื่อพิจารณาข้อสอบวิชาวิทยาศาสตร์ที่เกิดการทำหน้าที่ต่างกันของข้อสอบ ได้แก่ ข้อที่ 1, 2, 11, 12, 15, 18, 26, 27, 29 และข้อที่ 30 ซึ่งข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ จากทั้ง 2 วิธี คือ วิธี HGLM – 2L และวิธี BAYESIAN ได้แก่ข้อที่ 18 และข้อที่ 27

จากลักษณะของข้อสอบที่เกิดการทำหน้าที่ต่างกันในภาพรวม ทั้ง 3 วิชา โดยจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน ส่วนใหญ่สาเหตุที่ทำให้ข้อสอบเกิดการทำหน้าที่ต่างกัน อาจเป็นเพราะประสบการณ์ ความคุ้นเคยเกี่ยวกับเรื่องนั้น สภาพแวดล้อมและการฝึกปฏิบัติที่แตกต่างกันระหว่างนักเรียนในเขตกรุงเทพมหานครและปริมณฑลนักเรียนนอกเขต กรุงเทพมหานครและปริมณฑล

อภิปรายผลการวิจัย

การอภิปรายผลการวิจัยนี้ ผู้วิจัยได้นำเสนอ 3 ประเด็น ประเด็นแรกเป็นการเปรียบเทียบผลการประมาณค่า พารามิเตอร์ความยากของข้อสอบ (δ_i) และพารามิเตอร์ความสามารถของผู้สอบ (θ_j) ประเด็นที่สองเป็นการเปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) สำหรับผู้สอบจำแนกตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน และประเด็นที่สามอภิปรายการศึกษาลักษณะของข้อสอบที่เกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF) สำหรับผู้สอบจำแนกตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่ได้จากการวิเคราะห์การทำหน้าที่ต่างกัน โดยทั้ง 3 ประเด็น มีการวิเคราะห์ด้วยวิธี HGLM – 2L วิธี MIMIC และวิธี BAYESIAN จากข้อมูลที่มีผลการตอบให้คะแนนแบบสองค่า (Dichotomous) ซึ่งมีรายละเอียดดังนี้

1. การเปรียบเทียบผลการประมาณค่าพารามิเตอร์ความยากของข้อสอบ (δ_i) และพารามิเตอร์ความสามารถของผู้สอบ (θ_j)

1.1 ผลการประมาณค่าพารามิเตอร์ความยากของข้อสอบ (δ_i)

ผลการประมาณค่าพารามิเตอร์ความยากของข้อสอบ (δ_i) วิชาภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์ พบว่า วิธี HGLM – 2L วิธี MIMIC และวิธี BAYESIAN สามารถประมาณค่า พารามิเตอร์ความยากของข้อสอบได้ และวิธี MIMIC มีการประมาณค่าพารามิเตอร์ความยากของข้อสอบได้สูงกว่า วิธี HGLM – 2L และวิธี BAYESIAN ทั้งนี้ อาจเนื่องมาจากวิธีหลายตัวชี้วัดหลายสาเหตุในรูปแบบองค์ประกอบเชิงยืนยัน หรือวิธี MIMIC เป็นหลักการของ CFA กับตัวแปร (Muthén, 1988) แล้ววิธี MIMIC ยังสามารถนำไปใช้สำหรับการวิเคราะห์ DIF ได้ด้วย (Muthén et al., 1991) ซึ่งผลที่ได้ต้องมีค่าเป็นแบบ 2 ค่า (Dichotomous)

ค่าสัมประสิทธิ์สหสัมพันธ์ของการประมาณค่า พารามิเตอร์ความยากของข้อสอบ วิชาภาษาไทย ที่วิเคราะห์ด้วยวิธี HGLM – 2L วิธี MIMIC และวิธี BAYESIAN มีความสอดคล้องสัมพันธ์กันอย่างสมบูรณ์ และมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนค่าสัมประสิทธิ์สหสัมพันธ์ของการประมาณค่าพารามิเตอร์ความยากของข้อสอบวิชาคณิตศาสตร์ ที่วิเคราะห์ด้วยวิธี HGLM – 2L วิธี MIMIC และวิธี BAYESIAN มีความสัมพันธ์ในระดับสูงมาก และมีนัยสำคัญทางสถิติที่ระดับ .01 และค่าสัมประสิทธิ์สหสัมพันธ์ของการประมาณค่าพารามิเตอร์ความยากของข้อสอบวิชาวิทยาศาสตร์ ที่วิเคราะห์ด้วยวิธี HGLM – 2L วิธี MIMIC และวิธี BAYESIAN มีความสอดคล้องสัมพันธ์กันอย่างสมบูรณ์ และมีนัยสำคัญทางสถิติที่ระดับ .01 อาจเนื่องมาจากการประมาณค่าพารามิเตอร์ความยากของข้อสอบ (δ_i) คือ การสร้างโมเดลการวิเคราะห์

ที่มีตัวแปรตามเป็นผลการตอบที่ให้คะแนนแบบทวิภาค (0, 1) ซึ่งมีการกระจายแบบทวินาม (Binomial) และเป็นกรณีหนึ่งของการกระจายแบบเบอร์นูลลี (Bernoulli) ซึ่ง Raudenbush & Bryk (2002) ได้เสนอให้ใช้ฟังก์ชันการเชื่อมโยง (Link Function) แบบฟังก์ชันโลจิท (Logit Link Function) เพื่อทำให้การประมาณค่าระดับที่ 1 เชื่อมโยงไปสู่การวิเคราะห์ในระดับที่ 2 และระดับที่สูงขึ้นไปได้ โดยระดับการวิเคราะห์ที่ 1 จะเป็นระดับข้อสอบ ใช้โมเดลเชิงเส้นทั่วไป (GLM) ในการวิเคราะห์แบบทวนซ้ำ (Iterations) ก่อนเข้าสู่สมการวิเคราะห์ที่ 2 ดังนั้นในระดับการวิเคราะห์ที่ 1 จึงไม่สามารถใส่ตัวแปรทำนายเข้าสู่สมการได้ ส่วนการวิเคราะห์ระดับที่ 2 ก็จะเป็นการวิเคราะห์ด้วยโมเดลเชิงเส้นทั่วไประดับลดหลั่น (HGLM) (อิทธิฤทธิ์ พงษ์ปิยะรัตน์, 2551)

นอกจากนี้ อาจเนื่องจากการวิเคราะห์จากโปรแกรม HLM ด้วยวิธี HGLM – 2L โปรแกรม Mplus ด้วยวิธี MIMIC และโปรแกรม WinBUGS ด้วยวิธี BAYESIAN มีความสอดคล้องสัมพันธ์กันอย่างสมบูรณ์ และมีนัยสำคัญทางสถิติที่ระดับ .01 ทั้งนี้เนื่องจาก ผู้วิจัยได้วิเคราะห์วิธี HGLM – 2L แบบ 1-PL วิเคราะห์ด้วยโปรแกรม HLM (Kamata, 2001) ซึ่งวิธี MIMIC วิเคราะห์ด้วยโปรแกรม Mplus สามารถประมาณค่าพารามิเตอร์ได้ 2PL และวิธี BAYESIAN วิเคราะห์ด้วยโปรแกรม WinBUGS ก็สามารถประมาณค่าได้ 3PL จึงทำให้มีความสัมพันธ์กันอย่างสมบูรณ์ ซึ่งสอดคล้องกับงานวิจัยของ Kim (2003) ที่พบว่า ค่าพารามิเตอร์ของข้อสอบจากการประมาณค่าด้วยโมเดล HGLM – 2L มีความสัมพันธ์กับค่าพารามิเตอร์ของข้อสอบที่ประมาณค่าด้วยโมเดล Rasch อย่างสมบูรณ์ ($r = 1.00$)

1.2 พารามิเตอร์ความสามารถของผู้สอบ (θ_j)

ผลการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ วิชาภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์ พบว่า วิธี HGLM – 2L วิธี MIMIC และวิธี BAYESIAN สามารถประมาณค่าพารามิเตอร์ความสามารถของผู้สอบได้ ซึ่งวิธี MIMIC มีการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ วิชาภาษาไทย และวิชาคณิตศาสตร์ได้สูงกว่า วิธี HGLM – 2L และวิธี BAYESIAN ทั้งนี้เนื่องมาจากวิธี MIMIC หรือโมเดล MIMIC ใช้พิจารณาความสอดคล้องระหว่างข้อมูลกับโมเดล ว่ามีคุณลักษณะแฝงเพียง 1 คุณลักษณะ ดังนั้นจึงนำมาใช้เปรียบเทียบความเป็นเอกมิติของแบบสอบกับวิธีตรวจสอบแบบอื่น ๆ การวิเคราะห์โมเดลเป็นการตรวจสอบความสอดคล้องระหว่างข้อมูลกับโมเดลตามทฤษฎี อาศัยค่าไค – สแควร์ และค่าความสอดคล้องกับข้อมูล (GFI) ถ้าข้อมูลมีความสอดคล้องกับโมเดลมาก ค่าไค – สแควร์ จะมีค่าต่ำและค่า GFI

จะมีค่าสูง นั่นคือ ในขณะที่ข้อมูลเชิงประจักษ์ได้ให้แบบสอปที่มีความเป็นเอกมิตสูง และลดต่ำลง
 อย่างเป็นระบบ ค่าความสอดคล้องระหว่างข้อมูลกับโมเดลก็จะให้ค่าตัวเลขที่เปลี่ยนแปลง
 อย่างเป็นระบบคล้ายตามกัน (วรรณุช แหยมแสง, 2536) และอาจเนื่องจากวิธี BAYESIAN
 มีส่วนที่แตกต่างจากวิธี HGLM – 2L และวิธี MIMIC คือ การแจกแจงเริ่มต้น (Prior Distribution)
 การประมาณค่าพารามิเตอร์ความยากของข้อสอบด้วยวิธี BAYESIAN ที่มีการแจกแจงโพสทีเรีย
 (Posterior Distribution) ซึ่งประกอบด้วยผลคูณของฟังก์ชันไลค์ลิฮูด (Likelihood Function)
 กับการแจกแจงเริ่มต้น (Prior Distribution) มีค่าสูงสุด และ Swaminathan & Gifford (1985,
 pp. 358 – 359) ได้กล่าวว่า เมื่อกลุ่มตัวอย่างมีขนาดใหญ่ (มากกว่า 200 คน) และข้อสอบ
 มีจำนวนมาก การแจกแจงเริ่มต้นนี้ มีค่าน้อยมากจนทำให้การแจกแจงโพสทีเรียและฟังก์ชัน
 ไลค์ลิฮูดมีค่าใกล้เคียงกัน ซึ่งการวิจัยครั้งนี้ใช้กลุ่มตัวอย่าง 1, 000 คน และมีจำนวนข้อสอบ
 วิชาภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์ จำนวนวิชาละ 30 ข้อ รวมทั้งหมด 90 ข้อ ดังนั้น
 วิธีการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบด้วยวิธี BAYESIAN จึงมีวิธีการประมาณ
 ค่าพารามิเตอร์ความสามารถของผู้สอบ ต่างจากวิธีอื่น

ส่วนวิธี HGLM – 2L มีการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ
 วิชาวิทยาศาสตร์ได้สูงกว่าวิธี MIMIC และวิธี BAYESIAN ทั้งนี้เนื่องจากจากการประมาณ
 ค่าพารามิเตอร์ความยากของข้อสอบ (δ_i) ทั้ง 3 วิธี มีแนวคิดพื้นฐานทางสถิติที่ไม่เหมือนกัน
 โดยวิธี HGLM – 2L และวิธี MIMIC ใช้สถิติแบบนอนเบส์เซียน (Non Bayesian Statistics)
 ซึ่งแนวคิดนี้จะพิจารณาว่าเป็นค่าคงที่ ไม่ทราบค่าและต้องการประมาณค่า ในขณะที่วิธี
 BAYESIAN ใช้สถิติแบบเบส์เซียน (Bayesian Statistics) และมีวิธีการประมาณค่าพารามิเตอร์
 ความยากของข้อสอบ (δ_i) โดยมีการกำหนดการแจกแจงความน่าจะเป็นของค่าพารามิเตอร์
 ที่เรียกว่า การแจกแจงเริ่มต้น (Prior Distribution) และสอดคล้องกับผลการศึกษา
 ของอิทธิฤทธิ์ พงษ์ปิยะรัตน์ (2551) ได้ศึกษาการประมาณค่าพารามิเตอร์ความยากของข้อสอบ
 และพารามิเตอร์ความสามารถของผู้สอบ จากโมเดล HGLM ทั้ง 2 ระดับ และ 3 ระดับ
 ด้วยโปรแกรม HLM และการประมาณค่าพารามิเตอร์ทั้งสองจากโปรแกรม BILOG – MG
 ผลการวิเคราะห์ พบว่า โมเดล HGLM สามารถประมาณค่าพารามิเตอร์ความยากของข้อสอบได้
 และเมื่อให้ผลการวิเคราะห์ค่าพารามิเตอร์ความยากของข้อสอบจากโปรแกรม BILOG – MG
 เป็นเกณฑ์การพิจารณาแล้วพบว่า ผลค่าพารามิเตอร์ความยากของข้อสอบจากโมเดล
 HGLM – 2L และ HGLM – 3L นั้นมีความสัมพันธ์กับค่าพารามิเตอร์ความยากของข้อสอบ
 จากโปรแกรม BILOG – MG อย่างสมบูรณ์ (1.00)

จากการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบทั้ง 3 วิธี นี้ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าพารามิเตอร์ความสามารถของผู้สอบของวิชาภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์ ระหว่างวิธี HGLM วิธี MIMIC และวิธี BAYESIAN มีความสัมพันธ์ในระดับสูงมาก และมีนัยสำคัญทางสถิติที่ระดับ .01 ทั้งนี้อาจเนื่องมาจากความสัมพันธ์ของดัชนี ต่อไปนี้

$$RMSD^2 = MCSE^2 + BIAS^2 \text{ (BIAS: Different Between The Average Estimated Parameter, RMSD: Root Mean Square Deviation, MCSE: Monte Carlo Standard Error)}$$

ซึ่งเป็น การประเมินความเหมาะสมของการประมาณค่าพารามิเตอร์ โดยสร้างดัชนีเพื่อตรวจสอบและ ประเมินความสามารถในการประมาณค่า (GOR: Goodness of Recovery) ซึ่งการประมาณค่าพารามิเตอร์จะดำเนินการผ่านวิธีการคำนวณ คือ วิธี Sixth – order Laplace โดยใช้โปรแกรม HLM และวิธีการ MCMC โดยใช้โปรแกรม WinBUGS และวิธีการอื่น (Rijman et al., 2005) ซึ่งสอดคล้องกับการศึกษาของ Rijman et al. ผลการศึกษา พบว่า เทคนิคการประมาณค่าพารามิเตอร์ทั้งวิธี Sixth – order Laplace โดยใช้โปรแกรม HLM และวิธีการ MCMC โดยใช้โปรแกรม WinBUGS ให้ผลการประมาณค่าพารามิเตอร์ความยากข้อสอบ และการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบที่เท่าเทียมกันและมีความสัมพันธ์กับค่าพารามิเตอร์จริง ซึ่งสอดคล้องกับการศึกษาของวิชิตา บัวคง (2533) ที่ศึกษาการประมาณค่าพารามิเตอร์ความสามารถของผู้เข้าสอบ พบว่า แบบสอบที่ประมาณค่าความสามารถของผู้เข้าสอบด้วยวิธีของเบย์ มีความตรงเชิงทำนายสูงกว่าวิธีอื่น อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 และมีค่าฟังก์ชันสารสนเทศของแบบสอบสูงกว่าวิธีอื่นเช่นกัน แล้วถ้าความสามารถของกลุ่มตัวอย่างมีค่าเฉลี่ยอยู่ในระดับปานกลางถึงค่อนข้างต่ำ เมื่อนำค่าพารามิเตอร์ความสามารถของผู้เข้าสอบที่ประมาณได้ในแต่ละวิธีอื่น มาหาความสัมพันธ์กับคะแนนเกณฑ์ด้วยวิธีของเบย์ก็จะทำให้มีความตรงเชิงทำนายสูงที่สุด และสอดคล้องกับการศึกษาของนพดล มีชั้นช่วง (2544) ที่ศึกษาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าความยากที่ประมาณค่าด้วยวิธีแมกซิมัมไลค์ลิสต์ วิธีอีวีรอสติก และวิธีของเบย์ พบว่า ค่าสัมประสิทธิ์สหสัมพันธ์มีค่าตั้งแต่ .263 ถึง .446 วิธีแมกซิมัมไลค์ลิสต์ และวิธีของเบย์ มีนัยสำคัญทางสถิติที่ระดับ .05 และสอดคล้องกับการศึกษาของอิทธิฤทธิ์ พงษ์ปิยะรัตน์ (2551) ที่ได้ศึกษาการประมาณค่าพารามิเตอร์ความยากของข้อสอบที่วิเคราะห์ค่าพารามิเตอร์ความยากของข้อสอบด้วยโมเดล HGLM – 2L ซึ่งเมื่อเชื่อมโยงโมเดลการวิเคราะห์ค่าพารามิเตอร์ความยากของข้อสอบกับหลักการวิเคราะห์เศษเหลือที่ Goldstein (1997) ได้กล่าวไว้ว่าค่าส่วนที่เหลือที่เกิดจากการวิเคราะห์ถดถอยเมื่อตัวแปรตามเป็นผลสัมฤทธิ์ทางการเรียน ส่วนที่เหลือที่ได้จากการวิเคราะห์ เป็นผลสัมฤทธิ์

ที่ได้รับการจัดอิทธิพลแทรกซ้อนต่าง ๆ ออกไปจึงสามารถเป็นตัวบ่งชี้ความสามารถเฉพาะของผลสัมฤทธิ์ทางการเรียนที่เกิดจากการจัดการศึกษาได้ ซึ่งในการวิเคราะห์ข้อสอบพหุระดับนี้ค่าส่วนที่เหลือก็คือค่าอิทธิพลสุ่ม (Random Effect) ซึ่งเป็นค่าส่วนเบี่ยงเบนของโอกาสในการตอบข้อสอบถูกของผู้สอบคนที่ j จากค่าเฉลี่ยของโอกาสในการตอบข้อสอบถูกของนักเรียนในโรงเรียน จึงเป็นค่าที่แสดงถึงความสามารถเฉพาะของผู้สอบแต่ละคนในโรงเรียนที่ m ซึ่งค่าดังกล่าวมีการแจกแจงเป็นแบบปกติค่าเฉลี่ยเท่ากับศูนย์และค่าความแปรปรวนเป็น $\tau(U_{ojm} \sim N[0, \tau])$

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

2.1 เปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) สำหรับผู้สอบจำแนกตามเพศ ดังนี้

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์ พบข้อสอบที่เกิดการทำหน้าที่ต่างกันของข้อสอบ ส่วนใหญ่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบจากโปรแกรม HLM ด้วยวิธี HGLM-2L ที่เป็นเช่นนี้อาจเนื่องจากวิธี HGLM-2L ในระดับการวิเคราะห์ที่ 1 จะกำหนดให้ค่าพารามิเตอร์ความยากของข้อสอบที่มีประมาณค่าได้เป็นค่าคงที่ (Fixed Effect) กับผู้สอบแต่ละกลุ่ม เพื่อไม่ให้เกิดความไม่ผันแปรของค่าพารามิเตอร์ไปตามกลุ่มผู้สอบ ซึ่งเมื่อนำความแตกต่างของเพศไปทดสอบ ด้วยสถิติที่ก็สามารถใช้เป็นตัวบ่งชี้การทำหน้าที่ต่างกันของข้อสอบในแต่ละข้อได้ ซึ่งถ้าผลการทดสอบมีนัยสำคัญทางสถิติแสดงว่าข้อสอบข้อนั้นทำหน้าที่ต่างกันระหว่างกลุ่มเพศ เมื่อสัมประสิทธิ์ติดลบ กลุ่มอ้างอิงจะเสียเปรียบ แต่ถ้าค่าสัมประสิทธิ์เป็นบวกแสดงถึงกลุ่มอ้างอิงได้เปรียบจากการทำหน้าที่ต่างกันของข้อสอบ ในการวิเคราะห์ครั้งนี้ ผู้วิจัยได้กำหนดให้เพศชายเป็น 1 และเพศหญิงเป็น 0 และอาจเนื่องจากวิธี BAYESIAN มีส่วนที่แตกต่างจากวิธี HGLM-2L และวิธี MIMIC คือ การแจกแจงเริ่มต้น (Prior Distribution) การประมาณค่าพารามิเตอร์ความยากของข้อสอบด้วยวิธี BAYESIAN ที่มีการแจกแจงโพสทีเรีย (Posterior Distribution) ซึ่งประกอบด้วยผลคูณของฟังก์ชันไลค์ลิฮูด (Likelihood Function) กับการแจกแจงเริ่มต้น (Prior Distribution) มีค่าสูงสุด (Swaminathan & Gifford, 1985, pp. 358 – 359)

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์ เมื่อจำแนกตามเพศ ในการวิจัยครั้งนี้ วิธี HGLM-2L ตรวจสอบพบการทำหน้าที่ต่างกันของข้อสอบได้สูงกว่าวิธี MIMIC และวิธี BAYESIAN จึงสอดคล้องกับการศึกษา

ของ Fukahara & Kamata (2007) ได้ศึกษาการประเมินประสิทธิภาพการทำงานของวิธี MIMIC แบบละเมิดข้อตกลงเบื้องต้น โดยการทำชุดข้อสอบ พบว่าการทำหน้าที่ต่างกันของข้อสอบ (DIF) มีแนวโน้มที่จะการประมาณค่าได้ภายใต้ข้อสอบที่ไม่เป็นอิสระกัน และสอดคล้องกับการศึกษาของ Finch (2005) ที่ศึกษาการเปรียบเทียบประสิทธิภาพของโมเดล MIMIC กับการทดสอบโดยวิธี แมนเทลเฮนเซล (Mantel & Haenszel, 1959) และวิธี SIBTEST (Shealy & Stout, 1993) และวิธีการทดสอบ IRT likelihood ratio (Thissen et al., 1986) กับความคลาดเคลื่อนประเภทที่ 1 และอำนาจการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) แสดงให้เห็นว่าวิธี MIMIC มีค่าสูงขึ้นและความคลาดเคลื่อนประเภทที่ 1 มีค่าลดลง เมื่อจำนวนข้อสอบมีจำนวน 50 ข้อ นอกจากนี้วิธี MIMIC ยังสามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปได้

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย วิชาคณิตศาสตร์ และวิชาวิทยาศาสตร์ จากโปรแกรม HLM ด้วยวิธี HGLM – 2L สอดคล้องกับผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ได้จากโปรแกรม Mplus ด้วยวิธี MIMIC และสอดคล้องกับโปรแกรม WinBUGS ด้วยวิธี BAYESIAN มีความสัมพันธ์ในระดับสูงมาก และมีนัยสำคัญทางสถิติที่ระดับ .01 ทั้งนี้เนื่องจากวิธี HGLM – 2L วิธี MIMIC และวิธี BAYESIAN อยู่บนพื้นฐานของการวิเคราะห์คุณภาพของข้อสอบตามทฤษฎีการตอบสนองข้อสอบ (IRT) (1 – PL) Parameter Logistic Measurement Model (1 – PL) เช่นเดียวกัน จึงทำให้มีความสัมพันธ์ในระดับที่สูงมาก

2.2 เปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) สำหรับผู้สอบจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน ดังนี้

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย พบข้อสอบที่เกิดการทำหน้าที่ต่างกันของข้อสอบ ส่วนใหญ่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบจากโปรแกรม WinBUGS ด้วยวิธี BAYESIAN ทั้งนี้เนื่องจากการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบวิธี BAYESIAN มีค่าความคลาดเคลื่อนมาตรฐาน (SE) น้อยกว่าวิธี HGLM – 2L และวิธี MIMIC ซึ่งการประยุกต์ใช้โปรแกรม WinBUGS ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยใช้เทคนิค Markov Chain Monte Carlo (MCMC) มีการแจกแจงเริ่มต้น (Prior Distribution) ทำให้ประมวลผลออกมาใกล้เคียงกับความเป็นจริง และสอดคล้องกับผลการศึกษาของ Saengla Chaimongkol et al. (2007) ได้ศึกษาเรื่องตัวแบบอธิบายการทำหน้าที่ต่างกันของข้อสอบโดยใช้ WinBUGS1.4 เสนอตัวแบบพหุระดับการถดถอยโลจิสติกที่ใช้ในการตรวจสอบสาเหตุของการทำหน้าที่ต่างกันของข้อสอบ (DIF) โดยที่ตัวแบบที่เสนอจะพิจารณาโครงสร้าง

ของข้อมูลที่มีการซ้อนทับกัน 3 ระดับ ที่มีการรวมผลลัพธ์ที่ได้จากการวิเคราะห์การถดถอย
โลจิสติก เพื่อที่จะระบุลักษณะตัวแปรของข้อมูลระดับ 3 ที่สามารถใช้อธิบายสาเหตุการผันแปร
ของ DIF ได้ การศึกษาครั้งนี้จะใช้วิธีการจำลองข้อมูลในการตรวจสอบความถูกต้องและ
ความเหมาะสมของตัวแบบ โดยที่ค่าพารามิเตอร์ต่าง ๆ ในตัวแบบจะถูกประมาณโดยการใช้
หลักการของเบย์ที่ใช้โปรแกรม WinBUGS

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ พบข้อสอบ
ที่เกิดการทำหน้าที่ต่างกันของข้อสอบ ส่วนใหญ่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ
จากโปรแกรม HLM ด้วยวิธี HGLM – 2L เช่นเดียวกับการตรวจสอบจากโปรแกรม WinBUGS
ด้วยวิธี BAYESIAN ที่เป็นเช่นนี้อาจเนื่องจากวิธี HGLM – 2L เป็นวิธีที่มีความไวในการตรวจสอบ
การทำหน้าที่ต่างกันของข้อสอบเหมือนกันกับการตรวจสอบด้วยวิธี BAYESIAN และวิธี
HGLM – 2L ในการวิเคราะห์ระดับที่ 1 ระดับข้อสอบที่สอดแทรกในผู้สอบ (Between Item Within
Person) และระดับที่ 2 ระดับของบุคคล ที่สอดแทรกอยู่ในโรงเรียน (Between Person Within
School) ผลการวิเคราะห์ระดับนี้จะได้ค่าพารามิเตอร์ความสามารถของผู้สอบ ซึ่งวิชาคณิตศาสตร์
เป็นวิชาที่ต้องมีทักษะทางเป็นทางการคิดวิเคราะห์ที่ซับซ้อน ไม่สามารถแสดงออกได้โดยตรง
จำเป็นต้องอาศัยเครื่องมือวัดเพื่อให้แสดงถึงกระบวนการคิดทางสมองที่อยู่ในตัวบุคคล
อาจเป็นไปได้ว่าวิธี HGLM – 2L นี้มีความไวในการตรวจสอบวิชาคณิตศาสตร์ได้เหมือนกันกับ
วิธี BAYESIAN เพราะการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี BAYESIAN
โดยโปรแกรม WinBUGS เป็นการใช้เทคนิค Markov Chain Monte Carlo (MCMC)

มีการแจกแจงเริ่มต้น (Prior Distribution) ทำให้การประมวลผลออกมาใกล้เคียงกับความเป็นจริง

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาวิทยาศาสตร์ พบข้อสอบ
ที่เกิดการทำหน้าที่ต่างกันของข้อสอบ ส่วนใหญ่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบจาก
โปรแกรม HLM ด้วยวิธี HGLM – 2L ที่เป็นเช่นนี้อาจเนื่องจากวิธี HGLM – 2L มีความไวในการ
ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ดีกว่าวิธี MIMIC และวิธี BAYESIAN เมื่อจำแนกตาม
สถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน โดยการวิเคราะห์ข้อสอบพหุระดับสามารถดำเนินการ
วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ได้โดยอาศัยหลักการศึกษาค้นคว้าผลของตัวแปรภายนอกที่มี
ต่อค่าพารามิเตอร์ความสามารถของผู้สอบ โดยตัวแปรภายนอกสามารถจัดกระทำในลักษณะของ
ข้อมูลแบบ 2 ค่า (Dichotomous)

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาภาษาไทย วิชาคณิตศาสตร์
และวิชาวิทยาศาสตร์ จากโปรแกรม HLM ด้วยวิธี HGLM – 2L สอดคล้องกับผลการตรวจสอบ

การทำหน้าที่ต่างกันของข้อสอบที่ได้จากโปรแกรม Mplus ด้วยวิธี MIMIC และสอดคล้องกับโปรแกรม WinBUGS ด้วยวิธี BAYESIAN มีความสัมพันธ์ในระดับสูงมาก และมีนัยสำคัญทางสถิติที่ระดับ .01 ทั้งนี้อาจเนื่องมาจากทั้ง 3 วิธี คือ วิธี HGLM วิธี MIMIC และวิธี BAYESIAN ตั้งอยู่บนพื้นฐานทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) เหมือนกัน และสอดคล้องกับผลการ ศึกษาของ French & Miller (1996) ที่พบว่า เมื่อกลุ่มตัวอย่างมีขนาดเล็กลง อำนาจในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบลดลง และวิธีที่มีพื้นฐานมาจากการใช้ทฤษฎีการตอบสนองข้อสอบนั้น ยิ่งมีขนาดกลุ่มตัวอย่างใหญ่ขึ้นเพียงใด จะส่งผลให้การประมาณค่าพารามิเตอร์ของข้อสอบได้ดียิ่งขึ้น

โดยสรุป ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์ ระหว่างวิธี HGLM – 2L วิธี MIMIC และวิธี BAYESIAN จำแนกตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน พบว่า เมื่อจำแนกตามเพศจะพบข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด รองลงมาคือ จำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน ซึ่งสอดคล้องกับการศึกษาของเกสร หวังจิตร (2539) พบว่า ผลการระบุลักษณะของข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบระหว่างกลุ่มผู้สอบ เมื่อจำแนกตามตัวแปรเพศ ภูมิฐานะ ประสบการณ์ในการสอบ และสังกัดของสถานศึกษา มีความแตกต่างกัน โดยเมื่อจำแนกผู้สอบตามเพศ พบว่า มีข้อสอบที่มีการระบุว่ามี DIF มากที่สุด รองลงมาคือ ภูมิฐานะ สังกัดของสถานศึกษา ตามลำดับ ส่วนตัวแปรประสบการณ์ในการสอบมีข้อสอบที่พบ DIF น้อยที่สุดซึ่งผลการวิจัยนี้สอดคล้องกันทั้งวิชาภาษาไทยและภาษาอังกฤษ และลักษณะของข้อสอบที่พบ DIF เมื่อจำแนกผู้สอบตามภูมิฐานะ ในแบบสอบวิชาภาษาไทยและวิชาภาษาอังกฤษ พบว่า ส่วนใหญ่เข้าข้างกลุ่มผู้สอบในเขตกรุงเทพมหานครและปริมณฑลมากกว่าผู้สอบนอกเขตกรุงเทพมหานครและปริมณฑล และสอดคล้องกับผลการศึกษาของรักชนก ยี่สุนศรี (2544) ที่ศึกษาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ ระหว่างกลุ่มผู้สอบเมื่อจำแนกตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา พบว่าเมื่อจำแนกกลุ่มผู้สอบตามเพศจะมีข้อสอบที่ถูกระบุว่ามี DIF และถูกตัดออกมากกว่าการจำแนกกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษาของผู้สอบ ทั้งในแบบสอบวิชาภาษาอังกฤษและวิชาคณิตศาสตร์

3. ผลการศึกษาลักษณะของข้อสอบที่เกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF)

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

ข้อสอบวิชาภาษาไทยที่เกิดการทำหน้าที่ต่างกันของข้อสอบ ได้แก่ข้อที่ 1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 17, 18, 20, 21, 23, 24, 25, 26, 28, 29 และข้อที่ 30 ซึ่งข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ จากทั้ง 3 วิธี คือ วิธี HGLM – 2L วิธี MIMIC และวิธี BAYESIAN ได้แก่ข้อที่ 24 และข้อที่ 30

ข้อสอบวิชาคณิตศาสตร์ที่เกิดการทำหน้าที่ต่างกันของข้อสอบ ได้แก่ข้อที่ 1, 2, 3, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 20, 21, 22, 23, 27, 28, 29 และข้อที่ 30 ซึ่งข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ จากทั้ง 2 วิธี คือ วิธี HGLM – 2L และวิธี BAYESIAN ได้แก่ข้อที่ 30

ข้อสอบวิชาวิทยาศาสตร์ที่เกิดการทำหน้าที่ต่างกันของข้อสอบ ได้แก่ข้อที่ 1, 6, 7, 13, 18, 19, 21, 23, 24, 25, 26, 27, 28 และข้อที่ 30 ซึ่งข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ จากทั้ง 3 วิธี คือ วิธี HGLM – 2L วิธี MIMIC และวิธี BAYESIAN ได้แก่ข้อที่ 6, 7, 13 และข้อที่ 24

จากลักษณะของข้อสอบที่เกิดการทำหน้าที่ต่างกันในทุกภาพรวมทั้ง 3 วิชา โดยจำแนกตามเพศ ลักษณะของข้อสอบเป็นปัจจัยที่มีผลต่อความถูกต้องและความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (Narayanan & Swamitnathan, 1994) ข้อสอบ NT ปี 2553 มีลักษณะเป็นข้อสอบปรนัยแบบเลือกตอบ (Multiple Choice) ชนิด 4 ตัวเลือก ทั้งนี้เนื่องจากตัวเลือกสามารถตอบได้หลายคำตอบตามความเห็นของแต่ละบุคคล ข้อสอบบางข้อ ตัวเลือกไม่ได้เรียงจากข้อความสั้นไปหาข้อความยาว และข้อสอบที่เกิดการทำหน้าที่ต่างกันของข้อสอบจะมีคำศัพท์ที่เกี่ยวข้องกับเพศนั้น จึงทำให้ข้อสอบเข้าข้างเพศนั้น และทั้งนี้อาจเป็นเพราะความสามารถที่แตกต่างระหว่างเพศชายและเพศหญิงที่มีลักษณะความสามารถ ความถนัด และความสนใจในเรื่องนั้น ๆ ต่างกัน

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง พบว่า ลักษณะของข้อสอบคณิตศาสตร์ที่เพศชายจะทำได้ดีกว่าเพศหญิง ได้แก่ ข้อสอบเรขาคณิตและพีชคณิต ที่มีเหตุผลเชิงคณิตศาสตร์ งานที่มีลักษณะเป็นการแก้ปัญหา (Problem Solving) และงานที่ใช้ความคิดในขั้นสูง (Higher

Level Cognitive) ข้อสอบเกี่ยวกับอัตราส่วน (Ratio) สัดส่วน (Proportions) หรือเปอร์เซ็นต์ (Percent) และข้อสอบที่มีตาราง กราฟหรือภาพ ส่วนลักษณะของข้อสอบคณิตศาสตร์ที่เพศหญิงจะทำได้ดีกว่าเพศชาย ได้แก่ การคิดคำนวณ พีชคณิต ข้อสอบที่ใช้สัญลักษณ์และข้อสอบคณิตศาสตร์ที่เป็นนามธรรม (Abstract Mathematics) และงานที่มีลักษณะคล้ายคลึงกับในตำรา แต่ในสภาพจริงของการเรียนการสอนในประเทศไทย อาจมีปัญหาและตัวแปรที่เป็นปัจจัยให้ผลการศึกษาไม่เป็นไปตามหลักการที่อ้างอิงจากการศึกษาของต่างประเทศ เช่น พฤติกรรมการสอนของครู พฤติกรรมการเรียนของนักเรียน (อุทัยวรรณ สายพัฒนา, 2547) และอาจเนื่องจากความสามารถที่แตกต่างระหว่างเพศชายและเพศหญิงที่มีลักษณะความสามารถ ความถนัด และความสนใจในเรื่องนั้น ๆ ต่างกัน

Maier & Casselman (1970) กล่าวว่า เพศชายจะมีความสามารถในการแก้ปัญหาทางคณิตศาสตร์ และเพศหญิงจะมีความสามารถในทางภาษา และเป็นการถกเถียงกันมานานในการเรียนที่ต่างกันระหว่างเพศหญิงและเพศชาย เมื่อคนส่วนใหญ่เชื่อว่า เพศชายเรียนวิชาคณิตศาสตร์และวิทยาศาสตร์เก่งกว่าเพศหญิง ขณะที่เพศหญิงเรียนภาษาศาสตร์และสังคมศาสตร์เก่งกว่าเพศชาย ซึ่งในประเทศไทยก็มีผลที่สอดคล้องกันเมื่อเพศชายส่วนใหญ่เลือกเรียนสายวิทย์ – คณิต ขณะที่เพศหญิงเลือกเรียนศิลป์ – ภาษาที่ต่างกันไป ส่วนต่างประเทศก็มีความเข้าใจผิดในเรื่องนี้ไม่แพ้กัน เมื่อ ศาสตราจารย์ นิโคล เอลส์ เควสท์ จากมหาวิทยาลัย วิลลาโนวา รัฐฟิลาเดเฟีย ประเทศอเมริกา พบว่า คนส่วนใหญ่เกือบครึ่งล้านคน ใน 69 ประเทศ มีความเชื่อว่าเพศชายเรียนวิชาคณิตศาสตร์และวิทยาศาสตร์เก่งกว่าเพศหญิงจริง ๆ แต่เมื่อทีมนักวิจัยได้ศึกษากับนักเรียน จำนวน 493,495 คน โดยให้ทำข้อสอบนานาชาติ พบว่า ผลการสอบของนักเรียนเพศหญิงไม่ได้ด้อยไปกว่านักเรียนเพศชาย ซึ่งสอดคล้องกับนักวิจัยชาวอเมริกันอีกหลายท่าน และจากการศึกษาของ ศาสตราจารย์ นิโคล เอลส์ เควสท์ ยังพบว่า เพศหญิงสามารถเรียนรู้และนำวิชาคณิตศาสตร์และวิทยาศาสตร์ไปใช้ได้ไม่แพ้เพศชาย ยิ่งไปกว่านั้น เพศหญิงส่วนใหญ่ทำได้ดีกว่าเพศชาย แต่ไม่ได้หมายความว่า เพศหญิงจะเรียนในวิชาคณิตศาสตร์และวิทยาศาสตร์เก่งกว่าเพศชาย แต่สามารถสรุปได้ว่า เพศชายและหญิงมีความสามารถในวิชาคณิตศาสตร์และวิทยาศาสตร์พอ ๆ กัน

นอกจากนี้ข่าวคณิตศาสตร์ของวารสารการศึกษาวิทยาศาสตร์ คณิตศาสตร์และเทคโนโลยี ได้กล่าวถึง ความเชื่อยอดนิยม (ที่ผิด ๆ) เกี่ยวกับการเรียนคณิตศาสตร์ ที่ว่า “โดยธรรมชาติผู้ชายมีความถนัดในวิชาคณิตศาสตร์มากกว่าผู้หญิง” แต่ข่าวนี้ก็ไม่มีข้อมูลยืนยันในการศึกษาอย่างชัดเจน ควรศึกษาเรื่องนี้เพิ่มเติมต่อไป

และจากการศึกษาทฤษฎีคอนสตรัคติวิสต์ หรือทฤษฎีการสร้างความรู้ด้วยตนเอง ซึ่งเชื่อว่า ความรู้เป็นสิ่งที่มนุษย์แต่ละคนสร้างขึ้นด้วยตนเอง โดยใช้ทักษะทางปัญญาของแต่ละบุคคล การสร้างความรู้เกิดขึ้นเมื่อมนุษย์ได้รับข้อมูลใหม่ ๆ ข้อมูลจะผ่านกระบวนการซึมซาบเข้าไปปฏิสัมพันธ์กับความรู้ที่มีอยู่เดิม หากไม่สามารถสัมพันธ์กันได้ จะเกิดภาวะไม่สมดุล และได้แย่งภายในขึ้น มนุษย์จะพยายามปรับสภาวะให้อยู่ในภาวะสมดุลโดยใช้เหตุและผลภายใต้กระบวนการปรับโครงสร้างทางปัญญา จนเกิดเป็นความรู้ใหม่ที่มีความหมายสำหรับตนเอง ธรรมชาติของเด็ก จะมีความสนใจ อยากรู้ อยากเห็น ไม่หยุดนิ่ง เมื่อใดที่เผชิญกับสิ่งที่ไม่รู้ แต่มีความต้องการที่จะรู้ จะพยายามค้นหาคำตอบ เพื่ออธิบายสิ่งนั้นว่าคืออะไร เพราะอะไร หรือทำไม ซึ่งสังเกตได้จากเวลาที่เด็กมักมีคำถามแปลก ๆ ที่ตัวเองอยากรู้มาถาม จนเราไม่สามารถตอบได้ทัน หรือตอบได้หมดทุกคำถาม การฝึกให้เด็กคิดแบบวิทยาศาสตร์ จะช่วยให้เด็กค้นหาคำตอบในสิ่งที่สงสัยได้ด้วยตัวเอง อย่างมีเหตุผลและเชื่อถือได้

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

ข้อสอบวิชาภาษาไทยที่เกิดการทำหน้าที่ต่างกันของข้อสอบ ได้แก่ ข้อที่ 4, 14, 15, 26 และข้อที่ 29 ซึ่งข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ จากทั้ง 2 วิธี คือ วิธี HGLM – 2L และวิธี BAYESIAN ได้แก่ ข้อที่ 4 และข้อที่ 26

เมื่อพิจารณาข้อสอบวิชาคณิตศาสตร์ที่เกิดการทำหน้าที่ต่างกันของข้อสอบ ได้แก่ ข้อที่ 12 ซึ่งข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ จากทั้ง 2 วิธี คือ วิธี HGLM – 2L และวิธี BAYESIAN ได้แก่ ข้อที่ 12

เมื่อพิจารณาข้อสอบวิชาวิทยาศาสตร์ที่เกิดการทำหน้าที่ต่างกันของข้อสอบ ได้แก่ ข้อที่ 1, 2, 11, 12, 15, 18, 26, 27, 29 และข้อที่ 30 ซึ่งข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ จากทั้ง 2 วิธี คือ วิธี HGLM – 2L และวิธี BAYESIAN ได้แก่ ข้อที่ 18 และข้อที่ 27

จากลักษณะของข้อสอบที่เกิดการทำหน้าที่ต่างกันในทุกวิชาทั้ง 3 วิชา โดยจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน ทั้งนี้อาจเนื่องมาจากประสบการณ์ ความคุ้นเคยเกี่ยวกับเรื่องนั้น สภาพแวดล้อมและการฝึกปฏิบัติการที่แตกต่างกันระหว่างนักเรียนในเขตกรุงเทพมหานครและปริมณฑลนักเรียนนอกเขตกรุงเทพมหานครและปริมณฑล อาจเนื่องจากการไม่สนใจเรียนในบางวิชา เช่น จากคำกล่าว “ปัจจุบันวิชาภาษาไทยไม่เป็นที่สนใจของนักเรียนเท่าที่ควร

ดูได้จากผลการสอบ NT ระดับประเทศอยู่ในระดับต่ำ จึงจำเป็นต้องพัฒนากระบวนการเรียนวิชาภาษาไทยของนักเรียนให้มีนิสัยรักการเรียนและเรียนภาษาไทยให้ดีขึ้น" ซึ่งสอดคล้องกับการศึกษาของอุมารณ ภัทรวณิช และ ปัทมา อมรสิริสมบุรณ์ (2550) ซึ่งได้ศึกษาการขยายโอกาสทางการศึกษาเป็นนโยบายที่สำคัญของหลายประเทศ เนื่องจากทำให้ความไม่เท่าเทียมด้านเศรษฐกิจและสังคมลดลง มีหลักฐานที่สำคัญที่ทำให้เห็นความสัมพันธ์ระหว่างการพัฒนาทางการศึกษากับการพัฒนาทางเศรษฐกิจและสังคม (Hill & King, 1993) ในระดับมหภาค การเข้าถึงการศึกษา มีความสัมพันธ์กับการเพิ่มประสิทธิภาพของกำลังแรงงาน เพิ่มอายุขัยเฉลี่ยของประชากร และปรับปรุงด้านสุขภาพของประชากร รวมถึงลดภาวะเจริญพันธุ์ด้วย (Schultz, 1993) งานวิจัยของ Knodel & Jones (1996) ยืนยันว่าช่องว่างระหว่างเพศกำลังจะลดน้อยลง ไม่ว่าจะเป็นภูมิภาคใด และความแตกต่างของการศึกษาเกิดขึ้นในประเทศที่มีกลุ่มสถานภาพทางเศรษฐกิจและสังคมที่แตกต่างกันอย่างมากระหว่างเด็กยากจนหรือเด็กจากครอบครัวชนบททั้งเด็กหญิงและชาย ยังคงไม่ได้รับความเป็นธรรมด้านการได้รับการศึกษา เขาเหล่านั้นก็ต้องเผชิญกับความไม่เท่าเทียมกันในด้านตลาดแรงงาน และชีวิตความเป็นอยู่ สำหรับประเทศไทย พบว่า เนื่องจากการไม่ได้เรียนระดับมัธยมต้นของเด็กและเยาวชนมีเพียงร้อยละ 16 แต่การไม่ได้เรียนต่อระดับมัธยมปลายหรือเทียบเท่า มีถึงร้อยละ 33 (สำนักงานสถิติแห่งชาติ, 2543) ดังนั้น การให้ความสนใจกับการศึกษาที่สูงกว่าภาคบังคับ จึงเป็นสิ่งที่น่าสนใจ งานวิจัยของอุมารณ ภัทรวณิช และ ปัทมา อมรสิริสมบุรณ์ (2550) ได้ศึกษาการเรียนต่อระดับมัธยมปลายหรือเทียบเท่า ระหว่างปี พ.ศ. 2533 และปี พ.ศ. 2543 ซึ่งเห็นว่า ไม่มีความแตกต่างระหว่างเพศในการเรียนต่อระดับมัธยมปลายหรือเทียบเท่า ในปี พ.ศ. 2533 แต่ในปี พ.ศ. 2543 พบความแตกต่างระหว่างเพศ โดยเพศหญิงมีโอกาสได้เรียนต่อมากกว่าเพศชาย นอกจากนี้ ความแตกต่างระหว่างเขตเมืองและชนบท ยังเป็นปัจจัยสำคัญที่ส่งผลต่อการเรียนต่อระดับมัธยมปลายทั้งในปี พ.ศ. 2533 และ 2543 แม้ว่าความแตกต่างระหว่างเมืองและชนบทในการเรียนต่อระดับมัธยมปลายของปี พ.ศ. 2543 ได้ลดลงจากปี พ.ศ. 2533 ก็ตาม ประเด็นความไม่เท่าเทียมทางการศึกษา จึงมีคำถามวิจัยว่า ในช่วงเวลา หลังจาก พ.ศ. 2543 ถึงปัจจุบัน โอกาสการศึกษาระดับมัธยมปลาย (ซึ่งนอกเหนือจากการศึกษาภาคบังคับ และเป็นช่วงขั้นสุดท้ายของการศึกษาขั้นพื้นฐาน) ได้เพิ่มขึ้นหรือไม่ ความแตกต่างระหว่างเพศ และความแตกต่างระหว่างเมืองและชนบทในด้านโอกาสทางการศึกษาในระดับมัธยมปลายหรือเทียบเท่า ยังมีอยู่หรือไม่ รวมถึงมีปัจจัยอะไรบ้างที่เกี่ยวข้องกับโอกาสที่จะได้รับการศึกษา ส่วนความไม่เท่าเทียม

ด้านการศึกษา เมืองและชนบท มีความแตกต่างระหว่างเมือง/ ชนบทในการจ้างงาน ความยากจน ข้อมูลจากสำนักงานสถิติแห่งชาติ (2543) ชี้ว่า ประเทศไทยกำลังเพิ่มความเป็นเมืองมากขึ้น นั่นคือหนึ่งในสามของประชากรไทยอาศัยอยู่ในเขตเมือง ซึ่งเป็นการพิจารณาในระดับภาคจะยังคงพบความแตกต่างในเรื่องเมืองและชนบทอยู่ ตัวอย่างคือ ในปี พ.ศ. 2543 นั้น 1 ใน 3 (ร้อยละ 34) ของประชากรภาคกลางอาศัยในเขตเมือง เปรียบเทียบกับเพียงร้อยละ 17 ของประชากรภาคอีสาน ที่อาศัยในเขตเมือง หมายถึงประชากรภาคกลางที่อาศัยอยู่ในเขตชนบทมีน้อยกว่าประชากรในภาคอีสาน นอกจากนี้โอกาสการจ้างงานมีความเกี่ยวข้องกับเมืองและชนบทอย่างชัดเจน นั่นคือมีเพียงประมาณ 1 ใน 3 (ร้อยละ 37) ของประชากรภาคกลางทำงานในภาคเกษตรกรรม ขณะที่ภาคอื่น ๆ มีประชากรมากกว่าร้อยละ 70 อยู่ในภาคเกษตรกรรม ซึ่งก็ตีความว่า คนภาคกลางมีโอกาสดำรงงานในภาคเกษตรกรรมน้อยกว่าคนในภาคอื่น เนื่องจากว่าภาคกลางมีความเป็นเมืองมากกว่าภาคอื่น ๆ ข้อมูลจากสำนักงานสถิติแห่งชาติ ชี้ว่า อุปติการณ์ความยากจนในเขตชนบทมีเพิ่มขึ้นมากกว่าในเขตเมือง ในปีพ.ศ. 2547 เพียงร้อยละ 5 และ 7 ของประชากรที่อาศัยในชนบท ภาคกลางและภาคใต้เป็นคนยากจน ขณะที่ประชากรชนบทภาคอีสานและภาคเหนือมีคนยากจนถึงร้อยละ 20 และ ร้อยละ 17 ตามลำดับ แต่ความยากจนในเขตเมืองของทุกภาคยังคงมีอัตราที่ค่อนข้างต่ำ และเมื่อศึกษาความแตกต่างระหว่างเพศ สัดส่วนของเยาวชนหญิงอายุ 16 – 18 ปี ได้เปรียบในด้านการได้รับการศึกษาตามเกณฑ์สูงกว่าสัดส่วนเยาวชนชาย นั่นคือ ในเขตเมือง เยาวชนหญิงได้เรียนตามเกณฑ์มีถึงร้อยละ 47 ขณะที่เยาวชนชายมีเพียงร้อยละ 40 ส่วนในชนบท เยาวชนหญิงได้เรียนตามเกณฑ์ร้อยละ 33 เยาวชนชายได้เรียนตามเกณฑ์ร้อยละ 24 เห็นได้ว่าไม่ว่าจะเป็นเมืองหรือชนบท ผู้หญิงมีโอกาสในการได้รับการศึกษาในระดับมัธยมปลายหรือเทียบเท่าสูงกว่าผู้ชาย การค้นพบเช่นนี้เหมือนกับการศึกษาของอุมารณ ภัทรวาณิช และ ปัทมา อมรสิริสมบุญ (2550) ที่พบว่า เด็กผู้หญิงมีโอกาสได้เรียนต่อในระดับมัธยมศึกษาตอนปลายหรือเทียบเท่ามากกว่าเด็กผู้ชาย ด้วยเหตุผลของโอกาสการจ้างงานในภาคเกษตรกรรมลดลง เนื่องจากการปรับปรุงผลผลิตทางการเกษตรของไทยตั้งแต่ทศวรรษ 2520 โดยการมีใช้เทคโนโลยีเข้ามาใช้มากขึ้น ซึ่งเพิ่มโอกาสในการทำงานภาคการเกษตรให้กับผู้ชายมากกว่าผู้หญิง ประกอบกับในช่วงทศวรรษ 2530 การเติบโตของเศรษฐกิจในเมืองมีมากขึ้น ซึ่งสิ่งนี้น่าที่จะทำให้เกิดการเพิ่มโอกาสให้ผู้หญิงได้มีโอกาสเรียนต่อในระดับที่สูงขึ้น เพราะเป็นหนทางได้เข้าไปสู่ภาค อุตสาหกรรมและการบริการ ซึ่งกำลังเป็นภาคที่ทำให้เกิดการเติบโตทางเศรษฐกิจในเมือง นั่นเอง

ข้อเสนอแนะ

ข้อเสนอแนะในการนำไปใช้

1. การวิเคราะห์การประมาณค่าพารามิเตอร์ความยากของข้อสอบ พารามิเตอร์ความสามารถของผู้สอบ และการทำหน้าที่ต่างกันของข้อสอบ ควรเลือกการวิเคราะห์ด้วยวิธี MIMIC จากโปรแกรม Mplus เพราะมีกระบวนการในการวิเคราะห์ที่ง่าย ใช้งานสะดวก และไม่ซับซ้อน โดยการเขียนคำสั่งและวิเคราะห์ในขั้นตอนเดียว ซึ่งผลการวิเคราะห์ได้ค่าที่ใกล้เคียงและสอดคล้องกันกับวิธี HGLM – 2L และวิธี BAYESIAN

2. การวิเคราะห์ด้วยวิธี BAYESIAN จากโปรแกรม WinBUGS มีความคลาดเคลื่อนมาตรฐาน (SE) น้อยกว่าวิธี HGLM – 2L และวิธี MIMIC โดยโปรแกรม WinBUGS ใช้เทคนิค Markov Chain Monte Carlo (MCMC) มีการแจกแจงเริ่มต้น (Prior Distribution) ทำให้การประมวลผลออกมาใกล้เคียงกับความเป็นจริง

3. การศึกษาครั้งนี้ได้ข้อค้นพบว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี BAYESIAN โดยการเขียนคำสั่งในโปรแกรม WinBUGS และการใช้ข้อมูลจริง (Real Data) มีวิธีการเขียนคำสั่งที่ซับซ้อนค่อนข้างใช้เวลาในการศึกษา ควรเลือกวิเคราะห์ด้วย วิธี HGLM – 2L และวิธี MIMIC

4. ถ้าต้องการเพิ่มตัวแปรภายนอกเข้าสู่สมการ เพื่ออธิบายความผันแปรที่เกิดขึ้นเมื่อพบว่า ข้อสอบที่วิเคราะห์ข้อใดมีความผันแปรระหว่างโรงเรียน ตามอิทธิพลของตัวแปรความแตกต่างด้านเพศ และสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน สามารถใส่ตัวแปรพยากรณ์ในระดับโรงเรียนเข้าสู่สมการ เพื่ออธิบายความผันแปรที่เกิดขึ้นในข้อที่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบได้ โดยการวิเคราะห์ด้วยวิธี HGLM – 2L จากโปรแกรม HLM

5. ควรมีการปรับปรุง แก้ไขลักษณะของข้อสอบ ทั้งลักษณะของคำถามและตัวเลือก ให้มีความเป็นปรนัยมากขึ้น ในข้อที่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ เมื่อจำแนกเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน เพื่อหลีกเลี่ยงการนำไปสู่การทำหน้าที่ต่างกันของข้อสอบ

ข้อเสนอแนะสำหรับการทำวิจัยครั้งต่อไป

1. ควรศึกษาการประมาณค่าพารามิเตอร์ของข้อสอบ ด้วยวิธี BAYESIAN แบบ 3PL โดยการเขียนคำสั่งในสมการ จากการประยุกต์ใช้โปรแกรม WinBUGS

2. ควรศึกษาเปรียบเทียบค่าพารามิเตอร์ของข้อสอบแบบ 2 PL และพารามิเตอร์ความสามารถของผู้สอบด้วยวิธี MIMIC และวิธี BAYESIAN กับโปรแกรมอื่น ๆ เช่น TESTFACT, BILOG-MG, MULTILOG, XCALIBRE ที่อยู่บนพื้นฐานทฤษฎีการตอบสนองข้อสอบ (IRT) เช่นเดียวกัน

3. เนื่องจากข้อจำกัดของวิธี HGLM – 2L ในการวิเคราะห์ค่าพารามิเตอร์ข้อสอบ ได้เพียงค่าพารามิเตอร์ความยากของข้อสอบ จึงควรมีการเปรียบเทียบค่าพารามิเตอร์ของข้อสอบ แบบ 2 PL และพารามิเตอร์ความสามารถของผู้สอบด้วยวิธี MIMIC และวิธี BAYESIAN

4. ควรศึกษาเปรียบเทียบประสิทธิภาพของผลการตรวจสอบการทำหน้าที่ต่างกัน ของข้อสอบ ในแบบทดสอบที่มีการให้คะแนนแบบ 2 ค่า (Dichotomous) และการให้คะแนน แบบหลายค่า (Polytomous) โดยใช้วิธีการจำลองข้อมูล (Simulation Data)

5. ควรศึกษาการเปรียบเทียบประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกัน ของข้อสอบ ด้วยวิธี HGLM – 2L วิธี MIMIC และวิธี BAYESIAN โดยใช้ผลการตอบข้อสอบ i – net ของนักเรียน วิเคราะห์ตามตัวแปรอื่น ๆ เช่น การใช้ภาษาในชีวิตประจำวัน โดยกำหนดให้ ภาษาไทย เป็น 1 (Focal Group) และไม่ใช่ภาษาไทย เป็น 0 (Reference Group)

มหาวิทยาลัยบูรพา
Burapha University