

ที่สุด และวิธี IRT-likelihood ratio ตรวจพบการทำหน้าที่ต่างกันของข้อสอบน้อยที่สุด เมื่อจำแนกตามเขตที่ตั้งทางภูมิศาสตร์ของโรงเรียน พบว่า วิธี Multiple group CFA ตรวจพบการทำหน้าที่ต่างกันของข้อสอบมากที่สุด และวิธี IRT-likelihood ratio ตรวจพบการทำหน้าที่ต่างกันของข้อสอบน้อยที่สุด ส่วนการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย เมื่อจำแนกตามเพศ และที่ตั้งทางภูมิศาสตร์ของโรงเรียน พบว่า วิธี Bayesian ตรวจพบการทำหน้าที่ต่างกันของข้อสอบมากที่สุด และวิธี Multiple group CFA ไม่พบการทำหน้าที่ต่างกันของข้อสอบเลย

คำสำคัญ : การทำหน้าที่ต่างกันของข้อสอบ วิธีอัตราส่วนความควรจะเป็น วิธีเบส์เซียน วิธีองค์ประกอบเชิงยืนยันกลุ่มพหุ

A Comparison of Differential Item functioning by IRT-likelihood ratio, Bayesian and Multiple Group CFA Method

Received	01/08/2560
Revised	06/10/2560
Accepted	20/11/2562

Pensri Thiamsuk Chade Sirisawat
Pairat Wongnam

Abstract

This study aimed: 1) to study the correlation between the discriminant item parameter (a), difficulty item parameter (b) and Ability person parameter (θ) using Maximum likelihood, Bayesian and Confirmatory factor analysis approaches; and 2) to compare the differential item functioning of the Mathematics and Thai language test items on gender and school locations variables which were estimated by IRT-likelihood ratio, Bayesian and Multiple group CFA approaches. The samples were 2,400 Mattayomsuksa 6 students who took the National Test in the academic year 2015 and were drawn from National Test database, Ministry of Education by using the multi-stage sampling. The samples were classified according to gender (male and female) and school geographical location (Bangkok and metropolitan areas and non-Bangkok and metropolitan areas). The research findings were that 1) the discrimination parameters of the Mathematics and Thai Language test scores revealed that the correlation of discriminant parameter estimated by Maximum likelihood, Bayesian and Confirmatory factor analysis methods were very highly positive correlated at 0.05 statistical significant level. For the Mathematics test, the correlation of difficulty parameters estimated by Maximum likelihood and Confirmatory factor analysis method were very highly positive correlated at 0.05 statistical significant level. Meanwhile, Bayesian method was positively correlated at a slightly moderate level with Maximum likelihood and Confirmatory factor analysis methods at 0.05

¹ Student of Doctor of Philosophy Program in Educational Research, Measurement and Statistics, Burapha University
e-mail: Pensrithiamsuk49@gmail.com

² Associate Professor, Faculty of Education, Burapha University

³ Assistant Professor, Faculty of Education, Burapha University

statistical significant level. The difficulty parameters of the Thai Language test estimated by all three methods were very highly positive correlated at 0.05 statistical significant level. The ability person parameter of the Mathematics and Thai Language test scores estimated by all three methods were very highly positive correlated at 0.05 statistical significant level; and 2) According to DIF detection, Bayesian method was very sensitive to detect DIF in Mathematics option to gender variables, and IRT- likelihood ratio was hardly to find DIF. Multiple group CFA method was very sensitive to detect DIF in Mathematics test option to location variables, and IRT-likelihood ratio was hardly to find DIF. Bayesian method was very sensitive to detect DIF in the Thai Language test option to gender and location variables and Multiple-group CFA method did not find DIF.

Keyword: Differential Item Functioning, IRT- likelihood ratio, Bayesian, Multiple-group CFA

บทนำ

ปัจจุบันนี้คุณภาพการศึกษาของไทยก็ยังคงเป็นปัญหา จากการประเมินผลนักเรียนนานาชาติ ตามโครงการ PISA จัดโดยกลุ่ม OECD พ.ศ.2558 พบว่า นักเรียนไทยได้คะแนนการรู้เรื่องการอ่าน 409 คะแนน จากคะแนนมาตรฐานที่ 493 คะแนน อยู่ในช่วงลำดับที่ 56-60 ซึ่งต่ำกว่าค่าเฉลี่ย OECD มากกว่าหนึ่งระดับ ประเทศในเอเชียที่ร่วมการประเมิน และมีคะแนนเฉลี่ยต่ำกว่าไทยมีเพียงอินโดนีเซียประเทศเดียวเท่านั้น และผลการทดสอบการศึกษาแห่งชาติขั้นพื้นฐาน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2554-2556 พบว่า คะแนนสอบไม่ผ่านเกณฑ์ร้อยละ 50 (ณัฐปฏิญ์ชาน์ พิชญุชมนชื่น, 2558, หน้า 78) และคะแนนภาพรวมของประเทศได้คะแนนอยู่ในระดับต่ำ จึงทำให้เกิดเสียงวิพากษ์วิจารณ์ว่าการสอบ O-NET เป็นการสอบแข่งขันของเด็กในเมืองมากกว่าเด็กในพื้นที่ต่างจังหวัด ปัญหาการสอบตกต่ำ ข้อสอบบางข้อมีความลำเอียงเชิงวัฒนธรรม บางคำถามเด็กฐานะเศรษฐิกิจดี หรือเด็กในเมืองตอบได้ แต่เด็กชนบทหรือต่างจังหวัดไม่ทราบคำตอบ และตอบข้อสอบไม่ได้ (ชินภัทร ภูมิรัตน, 2554)

ดังนั้นระบบการศึกษาของประเทศไทยจึงมีเป้าหมายที่จะพัฒนาผู้เรียนให้มีประสิทธิภาพทางการเรียนให้มากที่สุด เช่น กลุ่มสาระการเรียนรู้ภาษาไทยได้กำหนดเป้าหมายในการพัฒนาผู้เรียนให้มีความสามารถด้านทักษะการคิด ดังนั้น นักวิชาการหลายท่านจึงได้พัฒนาเทคนิคการสอน และทำการศึกษาปัจจัยที่ส่งผลต่อประสิทธิภาพทางการเรียนของนักเรียน เช่น งานวิจัยของ สุภภัทร์ ทัพชัย (2557, หน้า 63) ได้ศึกษาเปรียบเทียบการคิดอย่างมีวิจารณญาณระหว่างก่อนเรียนและหลังเรียนวิชาภาษาไทยโดยใช้เทคนิคการสอน PMI ของนักเรียนชั้นมัธยมศึกษาปีที่ 3 โรงเรียนบ้านเหล่าจันทนองท่อม จังหวัดร้อยเอ็ด ปีการศึกษา 2556 พบว่า นักเรียนมีคะแนนความสามารถในการคิดอย่างมีวิจารณญาณหลังใช้เทคนิค PMI สูงกว่าคะแนนก่อนเรียน งานวิจัยของ นवलแสง สุศิริ (2560, หน้า 76) พบว่า ปัจจัยที่ส่งผลต่อประสิทธิผลของทางวิชาการของนักเรียน ได้แก่ การจัดกระบวนการเรียนรู้ บรรยากาศทางวิชาการ ภาวะผู้นำทางวิชาการ และพฤติกรรมการสอนของครู

จากปัญหาผลการทดสอบการศึกษาแห่งชาติขั้นพื้นฐาน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2554- 2556 ที่มีข้อวิพากษ์วิจารณ์ว่าข้อสอบมีความลำเอียง ในต่างประเทศได้มีการศึกษาความลำเอียงของข้อสอบ หรือทำหน้าที่ต่างกันของข้อสอบ (DIF) เช่น Muthen และเพื่อนร่วมงานของเขา ได้ตั้งคำถามถึงแหล่งที่มาของการทำหน้าที่ต่างกันของข้อสอบ (Muthén, B., & Lehman, J., 1985, pp 133-142., 1988, pp 213-238.; Muthen, Kao, & Burstein, 1991, pp 1-22) และคิดวิธีการตรวจสอบ DIF โดยการใช้โครงสร้างโมเดลสมการโครงสร้างหลายตัวชี้วัดหลายสาเหตุ แต่ก็ยังไม่สามารถหาสาเหตุของการทำหน้าที่ต่างกันของข้อสอบได้อย่างชัดเจน นักวิชาการหลายท่านได้ทำการวิจัยกันมาอย่างต่อเนื่อง สำหรับงานวิจัยที่เกี่ยวข้องในประเทศไทยสืบค้นจาก ThaiLIS (TDC) ตั้งแต่ปี.ศ. 2539-2560 พบว่า มีงานวิจัยเกี่ยวกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ จำนวนทั้งสิ้น 26 เรื่อง งานวิจัยส่วนใหญ่นิยมใช้วิธี แมนเทล-แฮนส์เซล วิธี SIBTEST และโมเดลตอบสนองข้อสอบแบบ 3 พารามิเตอร์ ตามลำดับ จากเหตุผลดังกล่าวข้างต้น จึงทำให้ผู้วิจัยสนใจที่จะทำการศึกษากการทำหน้าที่ต่างกันของข้อสอบ โดยได้เลือก 3 วิธีการตรวจสอบที่น่าสนใจ

และเป็นวิธีการที่มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับโมเดลการตอบสนอง ข้อสอบที่ผ่านมา ได้แก่ วิธี IRT-likelihood ratio วิธี Bayesian และ วิธีการวิเคราะห์องค์ประกอบเชิงยืนยัน กลุ่มพหุ (Multiple group CFA : MCFA) ซึ่งแต่ละวิธีมีความแตกต่างกันดังนี้

วิธี IRT-likelihood ratio เป็นวิธีการทดสอบอัตราส่วนความควรจะเป็นในรูปของล็อกเชิงเส้น ประมาณค่าพารามิเตอร์ด้วยวิธีความเป็นไปได้สูงสุด (Maximum likelihood : ML) วิธี BAYESIAN ประมาณค่าโดยใช้ฟังก์ชันความน่าจะเป็นที่แสดงความสัมพันธ์ของข้อมูลที่หาค่าได้ กับเซตของพารามิเตอร์ที่ไม่ทราบค่า หรือเรียกการแจกแจงภายหลัง ส่วนวิธี Multiple group confirmatory Factor Analysis : Multiple group CFA เป็นการศึกษาความไม่แปรเปลี่ยนของโมเดลการวัด (MI) ที่สัมพันธ์กับพารามิเตอร์ของข้อสอบ ที่เป็นโมเดลการตอบสนองข้อสอบแบบ 2 PL ได้แก่ ค่าน้ำหนักองค์ประกอบ (factor loading) และค่าจุดตัด (Threshold) ของข้อสอบระหว่างประชากรต่างกลุ่ม

วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาความสัมพันธ์ระหว่างผลการประมาณค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ (a) ค่าพารามิเตอร์ความยากของข้อสอบ (b) ค่าพารามิเตอร์ความสามารถของผู้สอบ (θ) ด้วยวิธี Maximum likelihood วิธี Bayesian และวิธี Confirmatory factor analysis

2 เพื่อเปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบ (DIF) จำแนกตามเพศ และสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน ได้แก่ เขตกรุงเทพมหานคร และปริมณฑล และนอกเขตกรุงเทพมหานคร และปริมณฑล ด้วยวิธี IRT-likelihood ratio วิธี Bayesian และวิธี Multiple group CFA

วิธีดำเนินการวิจัย

1. ประชากร

ประชากรในการวิจัยครั้งนี้เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 6 โรงเรียนในสังกัดสำนักงานคณะกรรมการการศึกษาแห่งชาติขั้นพื้นฐาน กระทรวงศึกษาธิการ เป็นโรงเรียนสหศึกษา ขนาดใหญ่ พิเศษ ที่เข้ารับการทดสอบวัดผลสัมฤทธิ์ทางการเรียนเพื่อประเมินคุณภาพการศึกษาระดับชาติขั้นพื้นฐาน O-NET ประจำปีการศึกษา 2558 วิชาคณิตศาสตร์ และภาษาไทย จำนวน 358,512 คน เพศชาย จำนวน 304,751 คน และเพศหญิงจำนวน 285,412 คน กำหนดกลุ่มตัวอย่างตามการวิเคราะห์คุณภาพทฤษฎีการทดสอบตอบสนองข้อสอบ และการทำหน้าที่ต่างกันของข้อสอบโดยเงื่อนไขกำหนดให้กลุ่มตัวอย่างไม่ต่ำกว่า 10-20 เท่าของข้อคำถาม (นงลักษณ์ วิรัชชัย, 2542, หน้า 311 ; สุวิมล ติрганันท์, 2553, หน้า 168 ; สุภมาศ อังศุโชติ, สมถวิล วิจิตรวรรณา และรัชนีกุล ภิญโญภาณุวัฒน์, 2554, หน้า 31 อ้างถึงในสุพัฒน์นา หอมบุปผา, 2556, หน้า 109) ในงานวิจัยครั้งนี้ใช้กลุ่มตัวอย่าง 20 เท่าของข้อคำถาม โดยวิธีการสุ่มแบบหลายขั้นตอนได้กลุ่มตัวอย่าง 2,400 คน เพศชาย 1,200 คนและเพศหญิง 1,200 คน ที่อยู่ในเขตกรุงเทพมหานครและปริมณฑล และนอกเขตกรุงเทพมหานครและปริมณฑล

2. เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยครั้งนี้ เป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนเพื่อประเมินคุณภาพ การศึกษาระดับชาติขั้นพื้นฐาน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ประจำปีการศึกษา 2558 ได้แก่ วิชา คณิตศาสตร์ จำนวน 40 ข้อ และภาษาไทย จำนวน 80 ข้อ สร้างโดยสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) แบบทดสอบเป็นแบบปรนัย เลือกตอบ 5 ตัวเลือก

3. การเก็บรวบรวมข้อมูล

ผู้วิจัยรวบรวมข้อมูลของนักเรียนที่เป็นกลุ่มตัวอย่างซึ่งเป็นข้อมูลทุติยภูมิ ได้แก่ ผลการตอบแบบ ทดสอบวัดผลสัมฤทธิ์ทางการเรียน เพื่อประเมินคุณภาพการศึกษาระดับชาติขั้นพื้นฐาน O-NET ของนักเรียน ชั้นมัธยมศึกษาปีที่ 6 ประจำปีการศึกษา 2558 โรงเรียนสหศึกษา ขนาดใหญ่พิเศษ โดยผู้วิจัยได้ขอหนังสือจาก ภาควิชาวิจัยและจิตวิทยาประยุกต์ คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา เพื่อขอความอนุเคราะห์ในการเก็บ ข้อมูลจากสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน)

4. การวิเคราะห์ข้อมูล

4.1. คำนวณค่าสถิติพื้นฐาน ได้แก่ ค่าเฉลี่ยเลขคณิต (Mean) ค่าความเบ้ (SK) ค่าความโด่ง (KU) ค่าเบี่ยงเบนมาตรฐาน (Standard Deviation) เพื่อบรรยายการแจกแจงของคะแนนจากแบบทดสอบสอบวิชา คณิตศาสตร์ และภาษาไทย วิเคราะห์ความสัมพันธ์ด้วยค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (Pearson's Correlation) และการหาความสอดคล้องด้วยการสร้างตารางไขว้ (Cross tabulation table) โดยใช้โปรแกรม สำเร็จรูป

4.2. ประมาณค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ (a) ค่าพารามิเตอร์ความยาก (b) ของ ข้อสอบ และพารามิเตอร์ความสามารถของผู้สอบ (θ) ด้วยวิธี Maximum likelihood วิธี Bayesian และวิธี Confirmatory factor analysis และตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธี IRT-likelihood ratio วิธี Bayesian และวิธี Multiple group CFA จากคะแนนการตอบแบบทดสอบวิชาคณิตศาสตร์ และ ภาษาไทยที่มีการตรวจให้คะแนนแบบ 2 ค่า (Dichotomous Scores) โดยใช้โปรแกรม IRT-PRO โปรแกรม Open BUGS และโปรแกรม Mplus ตามลำดับ

สรุปและอภิปรายผลการวิจัย

สรุปผลการวิจัย

1. ผลการศึกษาความสัมพันธ์ระหว่างผลการประมาณค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ (a) ค่าพารามิเตอร์ความยากของข้อสอบ (b) และค่าพารามิเตอร์ความสามารถของผู้สอบ (θ) ด้วยวิธี Maximum likelihood วิธี Bayesian และวิธี Confirmatory factor analysis มีดังนี้

1.1 ผลการประมาณค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ

วิชาคณิตศาสตร์ จากการวิเคราะห์ทั้ง 3 วิธี ด้วยการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน พบว่าทั้งสามวิธีมีความสัมพันธ์ทางบวกในระดับสูงมาก ($r = 1.00$) และมีนัยสำคัญทางสถิติที่ระดับ 0.05

วิชาภาษาไทย จากการวิเคราะห์ทั้ง 3 วิธี ด้วยการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน พบว่า ทั้งสามวิธี มีความสัมพันธ์ทางบวกในระดับสูงมาก ($r = 1.00$) และมีนัยสำคัญทางสถิติที่ระดับ 0.05 เช่นกัน

1.2 ผลการประมาณค่าพารามิเตอร์ความยากของข้อสอบ

วิชาคณิตศาสตร์ จากการวิเคราะห์ทั้ง 3 วิธี ด้วยการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน พบว่า Maximum likelihood และวิธี Confirmatory factor analysis มีความสัมพันธ์ทางบวกในระดับสูงมาก ($r = 1.00$) และมีนัยสำคัญทางสถิติที่ระดับ 0.05 วิธี Bayesian กับวิธี Maximum likelihood และวิธี Bayesian กับวิธี Confirmatory factor analysis มีความสัมพันธ์ทางบวก ในระดับปานกลางค่อนข้างน้อย ($r = 0.43$) และมีนัยสำคัญทางสถิติที่ระดับ 0.05

วิชาภาษาไทย จากการวิเคราะห์ทั้ง 3 วิธี ด้วยการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน พบว่าทั้งสามวิธีมีความสัมพันธ์ทางบวกในระดับสูงมาก ($r = 1.00$) และมีนัยสำคัญทางสถิติที่ระดับ 0.05

1.3 ผลการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ

วิชาคณิตศาสตร์ จากการวิเคราะห์ทั้ง 3 วิธี ด้วยการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน พบว่า วิธี Bayesian และวิธี Confirmatory factor analysis มีความสัมพันธ์ทางบวกในระดับสูงมาก ($r = 1.00$) และมีนัยสำคัญทางสถิติที่ระดับ 0.05

วิชาภาษาไทย จากการวิเคราะห์ทั้ง 3 วิธี ด้วยการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน พบว่า วิธี Bayesian และวิธี Confirmatory factor analysis มีความสัมพันธ์ทางบวกในระดับสูงมาก ($r = 1.00$) และมีนัยสำคัญทางสถิติที่ระดับ 0.05 และวิธี Maximum likelihood กับวิธี Bayesian และวิธี Maximum likelihood กับวิธี Confirmatory factor analysis มีความสัมพันธ์ทางบวกในระดับสูง ($r = 0.99$) และมีนัยสำคัญทางสถิติที่ระดับ 0.05

2. เปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) จำแนกตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน

การศึกษาการทำหน้าที่ต่างกันของข้อสอบในครั้งนี้อยู่ ประกอบด้วยวิชาคณิตศาสตร์ จำนวน 40 ข้อ และภาษาไทย จำนวน 80 ข้อ รวมทั้งหมด 120 ข้อ ผู้วิจัยวิเคราะห์ข้อสอบด้วยวิธี IRT- likelihood ratio วิธี Bayesian และวิธี Multiple group CFA จำแนกตามเพศ และสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) และความสอดคล้องของวิธีการตรวจสอบด้วยการวิเคราะห์ตารางไขว้ (Cross tabulation table) ได้ผลดังนี้

2.1 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) จำแนกตามเพศ และสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน วิชาคณิตศาสตร์

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) วิชาคณิตศาสตร์ พบว่า วิธี Bayesian พบจำนวนข้อสอบที่มีการทำหน้าที่ต่างกันมากที่สุดจำนวน 12 ข้อ ได้แก่ ข้อ 6, 12, 13, 16, 17, 23, 25, 26, 30, 36, 37 และข้อ 39 คิดเป็นร้อยละ 30 รองลงมา คือ วิธี Multiple group CFA พบจำนวน 11 ข้อ ได้แก่ ข้อ 1, 4, 6, 9, 11, 12, 16, 17, 20, 23 และข้อ 24 คิดเป็นร้อยละ 27.50 และ วิธี IRT-likelihood ratio พบจำนวน 9 ข้อ ได้แก่ ข้อ 1, 9, 11, 12, 16, 17, 23, 24 และข้อ 39 คิดเป็นร้อยละ 22.50 และเมื่อตรวจสอบความสอดคล้องด้วยการสร้างตารางไขว้ (Cross tabulation table) ดังนี้

ตารางที่ 1 แสดงความสอดคล้องของการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ ด้วย วิธี IRT- likelihood ratio วิธี BAYSIAN และวิธี Multiple group CFA

วิธีตรวจสอบ DIF	จำนวนข้อที่พบ DIF	ข้อที่พบ DIF ตรงกัน	χ^2	Φ	df	p
จำแนกตามเพศ						
วิธี IRT-LR เปรียบเทียบ วิธี Bayesian	5(12.5%)	12,16,17,23,39	3.61	.30	1	.06
วิธี IRT-LR เปรียบเทียบ วิธี MGCF A	8 (20%)	1,9,11,12,16,17,23,24	21.95*	.74*	1	.00
วิธี Bayesian เปรียบเทียบ วิธี MGCF A	5(12.5%)	6,12,16,17,23	1.73	.21	1	.19
จำแนกตามสถานที่ตั้งทางภูมิศาสตร์						
วิธี IRT-LR เปรียบเทียบ วิธี Bayesian	2(5.0%)	10,40	.18	.08	1	.67
วิธี IRT-LR เปรียบเทียบ วิธี MGCF A	7(17.5%)	3,5,10,17,19,39,40	.22	.07	1	.64
วิธี Bayesian เปรียบเทียบ วิธี MGCF A	8(20.0%)	9,10,18,31,32,33,38,40	3.53	-.30	1	.60

* $p < .05$ IRT-LR= IRT-likelihood ratio, MGCF A =Multiple group CFA

ผลการวิจัย พบว่าเมื่อจำแนกตามเพศ วิธี IRT- likelihood ratio เปรียบเทียบ วิธี Multiple group CFA พบว่ามีความสอดคล้องกันอย่างมีนัยทางสถิติที่ระดับ 0.05 ส่วนวิธี IRT- likelihood ratio เปรียบเทียบกับ วิธี Bayesian และวิธี Bayesian เปรียบเทียบกับวิธี Multiple group CFA พบว่า ไม่มีความสอดคล้องกัน เมื่อจำแนกตามเขตที่ตั้งทางภูมิศาสตร์ของโรงเรียน พบว่า ผลการเปรียบเทียบทั้งสามวิธีไม่มีความสอดคล้องกัน

ผลตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) วิชาภาษาไทยจากการวิเคราะห์ข้อสอบจำนวน 80 ข้อ ด้วยวิธี IRT- likelihood ratio วิธี Bayesian และวิธี Multiple group CFA พบว่า วิธี Bayesian พบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน จำนวน 29 ข้อ ได้แก่ข้อ 5, 7, 9, 10, 11, 13, 22, 26, 28, 32, 34, 40, 44, 46, 47, 51, 58, 60, 62, 66, 67, 68, 69, 70, 71, 72, 77, 78 และข้อ 79 คิดเป็นร้อยละ 36.25 รองลง

มา คือวิธี IRT- likelihood ratio พบจำนวน 3 ข้อ ได้แก่ ข้อ 38, 60 และข้อ 72 คิดเป็นร้อยละ 3.75 และวิธี Multiple group CFA ไม่พบการทำหน้าที่ต่างกันของข้อสอบร้อยละ 0 และเมื่อตรวจสอบความสอดคล้องด้วยการสร้างตารางไขว้ (Cross tabulation table) ได้ผลดังนี้

ตารางที่ 2 แสดงความสอดคล้องของการทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทย ด้วย วิธี IRT- likelihood ratio วิธี BAYSIAN และวิธี Multiple group CFA

วิธีตรวจสอบ DIF	จำนวนข้อพบ DIF	ข้อที่พบ DIF ตรงกัน	χ^2	Φ	df	p
จำแนกตามเพศ						
วิธี IRT-LR เปรียบเทียบ วิธี Bayesian	2 (2.5%)	60,72	1.25	.13	1	.26
วิธี IRT-LR เปรียบเทียบ วิธี MGCFA	0(0%)	-	-	-	-	-
วิธี Bayesian เปรียบเทียบ วิธี MGCFA	0(0%)	-	-	-	-	-
จำแนกตามสถานที่ตั้งทางภูมิศาสตร์						
วิธี IRT-LR เปรียบเทียบ วิธี Bayesian	5(6.3%)	25,35,36,48,57	6.27*	.28*	1	.01
วิธี IRT-LR เปรียบเทียบ วิธี MGCFA	0(0%)	-	-	-	-	-
วิธี Bayesian เปรียบเทียบ วิธี MGCFA	0(0%)	-	-	-	-	-

* $p < .05$ IRT-LR= IRT-likelihood ratio, MGCFA =Multiple group CFA

ผลการวิจัยพบว่า เมื่อจำแนกตามเพศ ผลการเปรียบเทียบทั้งสามวิธี ไม่มีความสอดคล้องกัน และพบว่าวิธี Multiple group CFA ตรวจไม่พบข้อสอบที่ทำหน้าที่ต่างกัน เมื่อจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน พบว่า ผลการเปรียบเทียบระหว่าง IRT-LR กับวิธี BAYSIAN พบว่า มีความสอดคล้องกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05

อภิปรายผลการวิจัย

1. ผลการประมาณค่าอำนาจจำแนกของข้อสอบ ค่าพารามิเตอร์ความยากของข้อสอบและค่าพารามิเตอร์ความสามารถของผู้สอบ

ผลการประมาณค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ และค่าพารามิเตอร์ความสามารถของผู้สอบวิชาคณิตศาสตร์ และวิชาภาษาไทย พบว่าทั้งสามวิธีสามารถประมาณค่าพารามิเตอร์ ได้ใกล้เคียงกันมาก มีความสัมพันธ์ทางบวกในระดับสูง และมีนัยสำคัญทางสถิติที่ ระดับ 0.05 อาจเป็นเพราะทั้งสามวิธี วิเคราะห์ค่าพารามิเตอร์ อยู่บนพื้นฐานของการวิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบ (IRT) 2 PL parameter logistic measurement model ด้วยสมการเดียวกัน จึงทำให้ได้ผลการประมาณค่าพารามิเตอร์ ได้ใกล้เคียงกันมาก และมีความสัมพันธ์กันในระดับสูง

ส่วนการประมาณค่าพารามิเตอร์ความยากของข้อสอบ พบว่า วิธี Maximum likelihood และวิธี Confirmatory factor analysis ประมาณค่าพารามิเตอร์ได้ใกล้เคียงกันมาก และมีความสัมพันธ์ทางบวกในระดับสูงและมีนัยสำคัญทางสถิติที่ระดับ 0.05 ส่วนวิธี Bayesian สามารถประมาณค่าพารามิเตอร์ความยากของข้อสอบได้สูงกว่าวิธี Maximum likelihood และวิธี Confirmatory factor analysis และมีความสัมพันธ์ทางบวกในระดับปานกลางค่อนข้างน้อยกับวิธี Maximum likelihood และวิธี Confirmatory factor analysis ซึ่งสอดคล้องกับผลการศึกษาของ อริสสา เตหลิ้ม (2559, หน้า 138) พบว่าวิธี Bayesian สามารถประมาณค่าพารามิเตอร์ความยากของข้อสอบได้ดี เมื่อข้อมูลมีการแจกแจงความสามารถเป็นแบบเบ้ซ้าย และวิธี Maximum likelihood จะประมาณค่าพารามิเตอร์ความยากของข้อสอบได้ดี เมื่อข้อมูลมีการแจกแจงความสามารถเป็นแบบเบ้ขวา ซึ่งผลการแจกแจงความสามารถด้วยวิธี Bayesian พบว่ามีการแจกแจงแบบเบ้ซ้าย ($SK = -0.546$) ส่วนวิธี Maximum likelihood ($SK = -3.77$) และ Confirmatory factor analysis ($SK = -3.77$) มีการแจกแจงแบบเบ้ซ้าย ซึ่งอาจเป็นสาเหตุที่ทำให้วิธี Bayesian ประมาณค่าพารามิเตอร์ความยากได้สูงกว่าวิธีอื่นและสอดคล้องกับศาสตราจารย์ ฉ่ำสมัย (2548, หน้า 78) ได้ศึกษาเปรียบเทียบการอนุมานค่าเฉลี่ยประชากรระหว่างวิธีภาวะน่าจะเป็นและวิธีการแบบเบส์เซียน พบว่าในการอนุมานค่าเฉลี่ยกรณีประชากรมีการแจกแจงแบบปกติ และแบบไม่ปกติควรเลือกใช้วิธีเบส์เซียน ทุกขนาดตัวอย่าง เนื่องจากการประมาณค่าเฉลี่ยใกล้เคียงกับค่าเฉลี่ยที่แท้จริง การประมาณค่าแบบช่วงที่ระดับความเชื่อมั่น 95% และ 99% ได้ขนาดความกว้างช่วงเฉลี่ยแคบ และการทดสอบสมมติฐาน ที่ระดับนัยสำคัญ 0.01 และ 0.05 สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ดี และอำนาจการทดสอบสูงกว่าวิธีความควรจะเป็น

2. เปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบ (DIF) จำแนกตามเพศ และสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน ระหว่างวิธี IRT-likelihood ratio วิธี Bayesian และวิธี Multiple group CFA

2.1 เปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) วิชาคณิตศาสตร์ จำแนกตามเพศ พบว่าวิธี Bayesian พบการทำหน้าที่ต่างกันของข้อสอบมากที่สุด และพบว่าวิธี IRT-likelihood ratio เปรียบเทียบ วิธี Multiple group CFA มีความสอดคล้องกันอย่างมีนัยทางสถิติที่ระดับ 0.05 ส่วนวิธี IRT-likelihood ratio เปรียบเทียบ วิธี Bayesian และวิธี Bayesian เปรียบเทียบวิธี Multiple group CFA พบว่าไม่มีความสอดคล้องกัน เมื่อจำแนกตามสถานที่ตั้งทางภูมิศาสตร์ วิธี Multiple group CFA พบการทำหน้าที่ต่างกันของข้อสอบมากที่สุด และทั้งสามวิธีไม่มีความสอดคล้องกัน ทั้งนี้เนื่องจากการวิเคราะห์ด้วยทฤษฎีการตอบสนองข้อสอบ (IRT) โมเดลจะต้องมีความเป็นเอกมิติ (Unidimensional IRT model) และโมเดลสามารถวิเคราะห์คุณลักษณะของบุคคลได้ครั้งละหนึ่งองค์ประกอบเท่านั้น ดังนั้น ผู้วิจัยจึงได้ทำการตรวจสอบความเป็นเอกมิติของข้อมูลโดยการสกัดองค์ประกอบ (Extraction Factor Analysis) ด้วยการหมุนแกนแบบออร์โธกอนอล(Orthogonal) วิธีวาริแมกซ์ (Varimax) แล้วตรวจสอบโครงสร้างของข้อมูลเชิงประจักษ์ของวิชาคณิตศาสตร์ จำแนกตามเพศ และสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน โดยใช้เกณฑ์การตรวจแบบแนวตั้ง Morizot, Ainsworth และ Reise (2007, pp.407-423 อ้างถึงในชัยวิชิต เขียวชัยชนะ, 2552, หน้า 14) ได้

เสนอให้พิจารณาความเป็นเอกมิติ (Unidimensionality) จากค่าอัตราส่วนระหว่างค่า Eigen องค์ประกอบแรกต่อค่า Eigen องค์ประกอบที่สอง หากพบว่ามีค่ามากกว่าหรือเท่ากับ 3.00 จะบ่งบอกถึงความเป็นเอกมิติจากการวิเคราะห์ความเป็นเอกมิติของโครงสร้างคุณลักษณะแฝงวิชาคณิตศาสตร์ จำแนกตามเพศ และสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน ได้ผลดังนี้

กลุ่ม	Eigen value		Eigen 1: Eigen 2
	องค์ประกอบที่ 1	องค์ประกอบที่ 2	
เพศ			
เพศชาย	10.60	1.99	5.33
เพศหญิง	8.94	2.38	3.75
สถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน			
เขตกรุงเทพฯ	11.14	2.27	4.90
นอกเขตกรุงเทพฯ	7.24	2.13	3.39

จากการวิเคราะห์ พบว่า ข้อมูลเชิงประจักษ์ มีลักษณะเป็นเอกมิติ (Unidimensional) ซึ่งสอดคล้องกับโมเดลตอบสนองข้อสอบ (IRT) ซึ่งอาจเป็นสาเหตุทำให้วิธี Multiple group CFA สามารถตรวจสอบพบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ได้จำนวนมาก

2.2 เปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) วิชาภาษาไทย จำแนกตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน พบว่า วิธี Bayesian พบการทำหน้าที่ต่างกันของข้อสอบ (DIF) มากที่สุด และวิธี Multiple group CFA ไม่พบการทำหน้าที่ต่างกันของข้อสอบ ทั้งนี้อาจเป็นเพราะวิธี Bayesian สามารถประมาณค่าพารามิเตอร์จากการแจกแจงภายหลัง (Posterior distribution) โดยใช้เทคนิค Markov Chain Monte Carlo (MCMC) ทำการแจกแจงเริ่มต้นหรือการแจกแจงก่อนหน้า (Prior distribution) และทำซ้ำจำนวนมากจนทำให้ประมวลผลได้ใกล้เคียงความเป็นจริงมาก ซึ่งสอดคล้องกับเกียรติเทพ ตั้งสันติถาวร (2548, หน้า 117) พบว่าวิธีการของเบส์จะมีประสิทธิภาพดีเมื่อขนาดตัวอย่างที่ใช้มีขนาดเล็กถึงขนาดปานกลาง เมื่อขนาดตัวอย่างใหญ่ขึ้น ตัวประมาณค่าจะลู่เข้าสู่พารามิเตอร์จริงและสอดคล้องกับปรางทิพย์ รัชตะปิติ (2550, หน้า 243) พบว่าวิธีเบส์แบบการแจกแจงก่อนหน้าจะให้ข้อมูลมีประสิทธิภาพดีกว่าวิธีการวิเคราะห์กำลังสองน้อยสุดสองชั้นในทุกกรณีที่ศึกษา ตั้งแต่ขนาดตัวอย่างเล็ก ขนาดกลาง จนถึงขนาดใหญ่

จากการที่วิธี Multiple group CFA ตรวจสอบไม่พบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ผู้วิจัยได้ทำการตรวจสอบความเป็นเอกมิติของข้อมูลโดยทำการสกัดองค์ประกอบโดยการหมุนแกนแบบออร์โธกอนอล (Orthogonal) ด้วยวิธีวาริเมกซ์ (Varimax) แล้วพบว่าข้อมูลเชิงประจักษ์ มีแนวโน้มเป็นสององค์ประกอบ

(Two Factor) ดังนั้นจึงอาจเป็นสาเหตุให้การวิเคราะห์ด้วยวิธี Multiple group CFA ตรวจสอบไม่พบการทำหน้าที่ต่างกันของข้อสอบ (DIF) เหมือนกับวิธีอื่น เพราะข้อมูลเชิงประจักษ์ไม่มีความเป็นเอกมิติ ซึ่งได้ผลการวิเคราะห์ความเป็นเอกมิติของโครงสร้างคุณลักษณะแฝงวิชาภาษาไทย จำแนกตามเพศ และสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน ได้ผลดังนี้

กลุ่ม	Eigen value		Eigen 1: Eigen 2
	องค์ประกอบที่ 1	องค์ประกอบที่ 2	
เพศ			
เพศชาย	11.67	7.80	1.50
เพศหญิง	10.45	8.08	1.30
จำแนกตามที่ตั้งทางภูมิศาสตร์ของโรงเรียน			
เขตกรุงเทพฯ	12.27	7.95	1.54
นอกเขตกรุงเทพฯ	9.29	8.04	1.15

ข้อเสนอแนะจากการวิจัย

1. ผลการศึกษาค้นคว้าพบว่า การประมาณค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ ค่าพารามิเตอร์ความสามารถของผู้สอบ พบว่าทั้งสามวิธีให้ผลการประมาณค่าที่ใกล้เคียงกันมาก ดังนั้นควรเลือกวิธี Maximum likelihood ประยุกต์ใช้โปรแกรม IRT-PRO ในการวิเคราะห์ เนื่องจากเป็นโปรแกรมสำเร็จรูป มีวิธีการใช้งานง่าย ไม่ต้องเขียนคำสั่ง ใช้เวลาในการวิเคราะห์น้อย

2. ในการประมาณค่าพารามิเตอร์ความยากของข้อสอบ ควรมีการตรวจสอบลักษณะของการแจกแจงความสามารถของผู้สอบก่อนว่าเป็นการแจกแจงแบบใด เช่น การแจกแจงแบบปกติ แจกแจงแบบเบ้ซ้าย หรือเบ้ขวา ทั้งนี้เพื่อใช้เป็นข้อมูลในการเลือกวิธีการประมาณค่าความยากที่เหมาะสม เช่น วิธี Bayesian จะประมาณค่าพารามิเตอร์ความยากของข้อสอบได้ดีกว่าวิธี Maximum likelihood เมื่อข้อมูลมีการแจกแจงความสามารถเป็นแบบเบ้ซ้าย เป็นต้น

3. การศึกษาค้นคว้าครั้งนี้ผู้วิจัยไม่ได้ทำการตรวจสอบลักษณะของข้อสอบก่อนทำการวิเคราะห์ แต่ผลการวิจัย พบว่า วิธี Maximum likelihood วิธี Bayesian และวิธี Confirmatory factor analysis ประมาณค่าพารามิเตอร์ได้ใกล้เคียงกันทั้งสองวิชา และมีความสัมพันธ์กันในระดับสูงอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ทั้งนี้อาจเป็นเพราะอิทธิพลของข้อสอบทดสอบ (Testlet) มีไม่มากพอที่จะทำให้เกิดผลกระทบต่อการประมาณค่าพารามิเตอร์ จึงทำให้ผลการประมาณค่าดังกล่าวจากทั้งสามวิธีไม่แตกต่างกัน และเมื่อทำการตรวจสอบลักษณะของข้อสอบภายหลัง พบว่า วิชาภาษาไทยมีลักษณะข้อสอบที่มีอิทธิพลทดสอบอยู่ในแบบทดสอบตอนที่ 1 ตั้งแต่ข้อ 1-10 และตอนที่ 3 ข้อ 69 -70

