

การศึกษาประสิทธิภาพของโมเดลการวิเคราะห์ถื่นตามติเชิงวัฒนธรรม
ในการวิเคราะห์การทำหน้าที่ต่างกันระหว่างผู้ประเมิน: การศึกษาจากสถานการณ์จำลอง
The Study of Cultural Consensus Model Efficiency
on Analysis of Differential Rater Functioning: A Simulation Study

ศิธา จุฑารัตน์^{1*} ศิริเดช สุชีวะ² และ สีวะโชติ ศรีสุทธิยากร³

Sitara Jutharat^{1*}, Siridej Sujiva² and Siwachoat Srisutiyakorn³

(Received: March 31, 2020 ; Revised: June 4, 2020 ; Accepted: June 8, 2020)

บทคัดย่อ

การศึกษาประสิทธิภาพของโมเดลการวิเคราะห์ถื่นตามติเชิงวัฒนธรรมในการวิเคราะห์การทำหน้าที่ต่างกันระหว่างผู้ประเมิน: การศึกษาจากสถานการณ์จำลอง มีวัตถุประสงค์เพื่อ 1) ตรวจสอบประสิทธิภาพของการประมาณค่าพารามิเตอร์ของโมเดลการวิเคราะห์ถื่นตามติเชิงวัฒนธรรม 2) ศึกษาปัจจัยที่ส่งผลต่อประสิทธิภาพในการประมาณค่าพารามิเตอร์ของโมเดล โดยศึกษาด้วยวิธีการจำลองสถานการณ์แบบมอนติ คาร์โล (MCMC)

ผลการศึกษาพบว่า 1) โมเดล MC-GCM สามารถประมาณค่าพารามิเตอร์ได้สอดคล้องกับค่าพารามิเตอร์ที่กำหนดอย่างมีประสิทธิภาพ โดยค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าพารามิเตอร์จากการจำลองข้อมูลกับค่าพารามิเตอร์ที่ได้จากการประมาณค่าของโมเดลมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติ 2) ผลการวิเคราะห์ปัจจัยที่ส่งผลต่อการประมาณค่าโดยการวิเคราะห์ความแปรปรวนพหุคูณ (MANOVA) ระหว่างจำนวนผู้ประเมิน (N) จำนวนรายการประเมิน (M) และการทำหน้าที่ต่างกันระหว่างผู้ประเมิน (T) ที่ส่งผลต่อประสิทธิภาพของการประมาณค่าของโมเดล พบว่า การทำหน้าที่ต่างกันระหว่างผู้ประเมินส่งผลต่อค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสองและค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าจริงและค่าที่ได้จากการประมาณค่าของโมเดลอย่างมีนัยสำคัญทางสถิติ

คำสำคัญ การวิเคราะห์ถื่นตามติเชิงวัฒนธรรม การทำหน้าที่ต่างกันระหว่างผู้ประเมิน

¹ นิสิตปริญญาเอก สาขาการวัดและประเมินผลการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

² รองศาสตราจารย์ ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

³ อาจารย์ ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

¹ Doctoral degree student in Educational Measurement and Evaluation Program, Faculty of Education, Chulalongkorn University

² Associate Professor, Department of Educational Research and Psychology, Faculty of Education, Chulalongkorn University

³ Lecturer, Department of Educational Research and Psychology, Faculty of Education, Chulalongkorn University

*Corresponding Author E-mail: sitara.jut@mahidol.edu

Abstract

The Study of Cultural Consensus Model Efficiency on Analysis of Differential Rater Functioning: A Simulation Study aimed to: 1) examine the efficiency of parameter estimating of the cultural consensus model, 2) study the factors affecting the model's efficiency on parameter estimating. The study was conducted in simulation scenario using the Markov chain Monte Carlo simulation method (MCMC).

The study found that: 1) the MC-GCM model could efficiently recover the model parameters. The Pearson correlation coefficient between the true value and the estimated value was statistically significant. 2) The result of MANOVA analysis between the number of raters (N), number of items (M) and the differential rater functioning (T) showed that the differential rater functioning had significant effect on MSE and correlation between the true value and the estimated value.

Keywords: cultural consensus analysis, differential rater functioning

บทนำ

การทำหน้าที่ต่างกันระหว่างผู้ประเมิน (Differential Rater Functioning: DRF) เป็นการศึกษาลักษณะของการประเมินที่แตกต่างกันระหว่างผู้ประเมินในแง่ของการกด/ปล่อยคะแนนเมื่อให้คะแนนผู้สอบที่มีคุณลักษณะต่างกัน เช่น ผู้ประเมินบางคนให้คะแนนผู้สอบเพศหญิงสูงหรือต่ำกว่าเพศชาย มโนทัศน์ของการทำหน้าที่ต่างกันระหว่างผู้ประเมินใกล้เคียงกับมโนทัศน์ของการทำหน้าที่ต่างกันของข้อสอบ ในขณะที่การทำหน้าที่ต่างกันระหว่างข้อสอบพิจารณาปฏิสัมพันธ์ระหว่างข้อสอบที่ส่งผลต่อผู้สอบต่างกลุ่ม การทำหน้าที่ต่างกันระหว่างผู้ประเมินจะพิจารณาปฏิสัมพันธ์ระหว่างการให้คะแนนของผู้ตรวจที่ส่งผลต่อผู้สอบต่างกลุ่มกัน

การตรวจสอบการทำหน้าที่ต่างกันระหว่างผู้ประเมินมีจุดประสงค์เพื่อตรวจสอบการประเมินและระบุความแตกต่างของการตอบสนองของผู้ประเมินที่มีต่อเนื้อหาหรือรายการประเมิน ซึ่งเป็นประโยชน์ต่อการวิเคราะห์ความคลาดเคลื่อนในการประเมิน เช่น การกดหรือปล่อยคะแนน วิธีการตรวจสอบ DRF ที่พบมากที่สุด คือ การวิเคราะห์ Many Facet Rasch Model (MFRM) (Schaefer, 2008; Muckle และ Karabatsos, 2009; Farrokhi, Esfandiani และ Schaefer, 2012; Myford และ Wolfe, 2009; Xun Yan, 2014; Engelhard Jr., Wind, Kobin และ Chajewski, 2013; Wesolowski, Wind และ Engelhard Jr., 2015) อย่างไรก็ตาม แม้ว่ากรวิเคราะห์ MFRM จะเป็นวิธีการที่ได้รับความนิยมในการศึกษาการทำหน้าที่ต่างกันระหว่างผู้ประเมินมากที่สุด แต่มีผู้เสนอว่าการวิเคราะห์ MFRM ควรใช้ตัวอย่างขนาดใหญ่ (Farrokhi และคณะ, 2012) รวมถึง MFRM ยังมีปัญหาเกี่ยวกับการประมาณค่าความคลาดเคลื่อนมาตรฐาน (Patz และคณะ, 2002) นอกจากนี้ ผลการศึกษาของ Muckle และ Karabatsos (2009) พบข้อจำกัดเกี่ยวกับข้อตกลงเบื้องต้นของการแจกแจงของโมเดลที่ต้องเป็นการแจกแจงปกติ ในกรณีที่การแจกแจงของประชากรไม่เป็นการแจกแจงปกติ

จะทำให้โมเดลมีความสอดคล้องกับข้อมูลต่ำ และการประมาณค่าประชากรของอิทธิพลสุ่มจะไม่แม่นยำ ซึ่งส่งผลเสียต่อการวิเคราะห์ข้อมูลเกี่ยวกับความเที่ยงของการประเมิน

เนื่องจากข้อจำกัดของการวิเคราะห์ข้อมูลดังกล่าว ผู้วิจัยได้ศึกษาทฤษฎีการวิเคราะห์วัฒนธรรม (Cultural Consensus Theory) (Romney, Weller และ Batchelder, 1986) ทฤษฎีดังกล่าวประกอบด้วยโมเดลการวิเคราะห์วัฒนธรรมซึ่งมีพารามิเตอร์หลักที่ให้สารสนเทศเช่นเดียวกับการวิเคราะห์ด้วยทฤษฎีการตอบสนองข้อสอบ ได้แก่ ระดับวัฒนธรรม (Z_k) ความสามารถของผู้ตอบ (D_i) และความลำเอียงในการตอบ (g_i) นอกจากนี้ ทฤษฎีการวิเคราะห์วัฒนธรรมยังประกอบด้วยโมเดลย่อยที่มีพารามิเตอร์ที่รองรับการวิเคราะห์ข้อมูลที่แตกต่างกัน ดังนั้น ผู้วิจัยจึงสนใจศึกษาการประยุกต์ใช้การวิเคราะห์วัฒนธรรมในการวิเคราะห์การทำหน้าที่ต่างกันระหว่างผู้ประเมิน โดยมีวัตถุประสงค์เพื่อตรวจสอบประสิทธิภาพของการวิเคราะห์ความเที่ยงระหว่างผู้ประเมินรวมถึงศึกษาปัจจัยที่ส่งผลต่อประสิทธิภาพในการประมาณค่าของโมเดล ทั้งนี้ จากการทบทวนเอกสารที่เกี่ยวข้อง สามารถให้สมมติฐานได้ว่า การวิเคราะห์วัฒนธรรมเป็นวิธีการรวบรวมและวิเคราะห์ข้อมูลทางสถิติที่สามารถลดข้อจำกัดของวิธีการทางสถิติที่ใช้ในการประมาณค่าความเที่ยงระหว่างผู้ประเมินที่มีอยู่ในปัจจุบันได้

Multicultural General Condorcet Model (MC-GCM) เป็นโมเดลการวิเคราะห์วัฒนธรรมที่ออกแบบให้รองรับการวิเคราะห์ข้อมูลที่มีคำตอบแฝงได้หลายคำตอบตามกลุ่มผู้ให้ข้อมูล ในกรณีการประเมินมีผู้ประเมินที่มีพฤติกรรมการประเมินที่แตกต่างกัน รวมถึงยอมให้มีความแตกต่างกันของความยากในการประเมิน โมเดล MC-GCM เป็นการขยายผลการศึกษาของ Batchelder และ Anders (2012) เกี่ยวกับการปรับปรุงการอนุมานทางสถิติแบบเบย์แบบลดหลั่น (Hierarchical Bayesian Inference) และได้นำมาใช้เป็นส่วนหนึ่งของโมเดลนี้ การเพิ่มพารามิเตอร์จำนวนกลุ่มวัฒนธรรมทำให้สามารถประเมินได้ว่าผู้ให้ข้อมูลมีลักษณะของการแบ่งกลุ่มย่อยภายใน (subgroup) หรือไม่ ในกรณีของการวิเคราะห์ความสอดคล้องของผู้ประเมิน MC-GCM จะช่วยให้ทราบได้ว่าผู้ประเมินมีการทำหน้าที่แตกต่างกันในการประเมินหรือไม่ และมีลักษณะการทำหน้าที่ต่างกันอย่างไร รวมถึงการประมาณค่าความน่าเชื่อถือ (D_{ik}) ของผู้ประเมินทั้งรายบุคคลและกลุ่ม

โมเดลการวิเคราะห์วัฒนธรรมประมาณค่าข้อมูลจากเมตริกซ์การประเมินของผู้ประเมิน ซึ่งเป็นเมตริกซ์ขนาด ($N \times M$) สมมติให้การประเมินทางการศึกษาโดยผู้ประเมินในสถานการณ์ k ทำการประเมินคุณภาพการจัดเรียนการสอนโดยมีระดับการประเมิน 0 คือ ไม่ผ่านเกณฑ์ และ 1 คือ ผ่านเกณฑ์ มีรายการประเมินจำนวน 25 รายการ ใช้ผู้ประเมินจำนวน 15 คน หากนำผลการประเมินของผู้เชี่ยวชาญแต่ละคนในการประเมินแต่ละรายการมาเขียนลงในเมตริกซ์การประเมิน จะเป็นดังนี้

ผู้ประเมิน (N)	รายการประเมิน (M)												
	1	2	3	4	5	6	7	8	...	22	23	24	25
1	0	0	0	1	0	0	0	1		1	0	1	0
2	0	0	0	0	0	0	0	0		0	0	0	1
3	0	0	1	1	0	0	0	0		0	0	1	1
4	0	0	0	0	0	1	0	0		1	0	1	1
5	1	0	1	1	1	1	0	1		1	0	1	1
...													
13	0	0	0	1	0	1	1	1		1	1	0	0
14	1	0	1	1	1	0	1	0		1	1	0	0
15	0	1	0	1	1	0	1	0		0	1	0	0

Response profile matrix ของผู้ประเมินในสถานการณ์ ก

จากเมตริกซ์จะเห็นว่าผู้ประเมินบางคนประเมินรายการเดียวกันต่างออกไปจากผู้ประเมินส่วนใหญ่ ซึ่งอาจมีสาเหตุมาจากคำถามประเมินไม่ชัดเจน ผู้ประเมินบางคนอาจมีความเชี่ยวชาญหรือประสบการณ์น้อยกว่าคนอื่น หรือผู้ประเมินมีความลำเอียงต่อการประเมิน เป็นต้น ด้วยเหตุนี้ การสรุปผลการประเมินครั้งนี้ อาจนำไปสู่ข้อสรุปที่มีความน่าเชื่อถือต่ำ หรือไม่สามารสรุปได้ว่าผลการประเมินที่แท้จริงควรเป็นอย่างไร

การประมาณค่าของโมเดล MC-GCM ใช้วิธีการประมาณค่าแบบเบย์แบบลดหลั่น (Bayesian hierarchical framework) ซึ่งอธิบายรายละเอียดใน Anders และ Batchelder (2012) ในบทความดังกล่าว ได้ศึกษาประสิทธิภาพของ MC-GCM ในการประมาณค่าฉันทมติเชิงวัฒนธรรมด้วยข้อมูลจำลองและข้อมูลจริง การศึกษาก็คือข้อมูลจำลองพบว่า MC-GCM ประมาณค่าความแตกต่างของกลุ่มฉันทมติได้ตรงตามที่กำหนด และโมเดลสามารถประมาณค่าข้อมูลที่มีความแตกต่างของกลุ่มวัฒนธรรมได้อย่างมีประสิทธิภาพ

จากคุณสมบัติของโมเดลการวิเคราะห์ฉันทมติเชิงวัฒนธรรมที่กล่าวมา ทำให้ผู้วิจัยสนใจศึกษาความเป็นไปได้ในการประยุกต์ใช้โมเดลในการวิเคราะห์ฉันทมติของผู้ประเมินในการประเมินทางการศึกษารวมถึงการระบุการทำหน้าที่ต่างกันระหว่างผู้ประเมิน อย่างไรก็ตาม เนื่องจากยังไม่มีการศึกษาเกี่ยวกับการประยุกต์ใช้โมเดลดังกล่าวในการประเมินทางการศึกษา ผู้วิจัยจึงทำการศึกษาโดยการจำลองสถานการณ์การประเมินที่มีจำนวนผู้ประเมิน รายการประเมิน และการทำหน้าที่ต่างกันระหว่างผู้ประเมินที่แตกต่างกัน เพื่อศึกษาประสิทธิภาพของการประมาณค่าของโมเดลการวิเคราะห์ฉันทมติเชิงวัฒนธรรมในสถานการณ์การประเมินที่แตกต่างกัน รวมถึงศึกษาปัจจัยที่ส่งผลต่อประสิทธิภาพของการประมาณค่าพารามิเตอร์ของโมเดล ได้แก่ จำนวนผู้ประเมิน จำนวนข้อของรายการประเมิน และการทำหน้าที่ต่างกันระหว่างผู้ประเมิน ว่าปัจจัยดังกล่าวส่งผลต่อความคลาดเคลื่อนของการประมาณค่าของโมเดลอย่างไรบ้าง

ความมุ่งหมายของการวิจัย

1. เพื่อตรวจสอบประสิทธิภาพของการประมาณค่าพารามิเตอร์ของโมเดลการวิเคราะห์ชั้นทามติเชิงวัฒนธรรม
2. เพื่อศึกษาปัจจัยที่ส่งผลต่อประสิทธิภาพในการประมาณค่าของโมเดล ด้วยวิธีการจำลองสถานการณ์แบบมอนติ คาร์โล

วิธีการดำเนินการวิจัย

1. การกำหนดสถานการณ์จำลองข้อมูล

การศึกษาค้นคว้าครั้งนี้มีวัตถุประสงค์เพื่อตรวจสอบประสิทธิภาพของการประมาณค่าพารามิเตอร์ของโมเดลการวิเคราะห์ชั้นทามติเชิงวัฒนธรรม MC-GCM ภายใต้การจำลองสถานการณ์การประเมิน 18 สถานการณ์ โดยกำหนดจำนวนผู้ประเมิน $N = (15, 30, 45)$ จำนวนรายการประเมิน $M = (25, 55, 85)$ การทำหน้าที่ต่างกันระหว่างผู้ประเมิน (T) โดยกำหนด $T = 1$ แทนกรณีที่รายการประเมินไม่มีการทำหน้าที่ต่างกันระหว่างผู้ประเมิน และ $T = 2$ แทนกรณีที่รายการประเมินมีการทำหน้าที่ต่างกันระหว่างผู้ประเมิน ผู้วิจัยจำลองข้อมูลผลการประเมินแบบ 2 ค่า (1 = ใช่, 0 = ไม่ใช่) แล้วตรวจสอบความถูกต้องของการประมาณค่าพารามิเตอร์ชั้นทามติเชิงวัฒนธรรม (Z_k) พารามิเตอร์ความสามารถของผู้ประเมิน (θ_i) พารามิเตอร์ความยากของรายการประเมิน (δ_k) และพารามิเตอร์ความลำเอียงในการประเมิน (g_i) ใน 2 กรณี คือ 1) รายการประเมินที่ไม่มีการทำหน้าที่ต่างกันระหว่างผู้ประเมินและ 2) รายการประเมินที่มีการทำหน้าที่ต่างกันระหว่างผู้ประเมิน โดยมีรายละเอียดการจำลองข้อมูล ดังนี้

- 1) กำหนดจำนวนผู้ประเมิน (N) ดังนี้ $N = (15, 30, 45)$
- 2) กำหนดจำนวนรายการประเมินความสอดคล้องในแนวเดียวกัน (M) ดังนี้ $M = (25, 55, 85)$
- 3) กำหนดเงื่อนไขการทำหน้าที่ต่างกันระหว่างผู้ประเมิน (T) โดย T_1 เป็นรายการประเมินที่มีความยากของคำถามประเมินเท่ากัน ($\delta_k = 0.5$) และ T_2 เป็นรายการประเมินที่มีความยากของคำถามประเมินไม่เท่ากัน ($\delta_k \neq 0.5$)
- 4) สร้างคำตอบชั้นทามติของผู้ประเมิน 2 กลุ่มขนาด (Z_{tk})_{n,m} โดยการสุ่มจากการแจกแจงความน่าจะเป็นแบบเบอร์นูลลี $Z_{tk} \sim Bernoulli(m, P_t)$ โดย $P_t \sim Uniform(0,1)$
- 5) สร้างค่าพารามิเตอร์กลุ่มวัฒนธรรม $e_i \sim Categorical(\lambda)$ โดย $\lambda \sim Dirichlet(L), L = (1)_{1 \times T}$
- 6) กำหนดค่าความสามารถของผู้ประเมิน (θ_i) โดยการสุ่มจากการแจกแจงความน่าจะเป็นแบบเบต้า $\theta_i \sim Beta(n, 2, 2)$
- 7) กำหนดค่าความยากของคำถามประเมิน (δ_k) โดยการสุ่มจากการแจกแจงความน่าจะเป็นแบบเบต้า $\delta_k \sim Beta(n, 2, 2)$
- 8) กำหนดค่าความลำเอียงในการประเมิน (g_i) โดยการสุ่มจากการแจกแจงความน่าจะเป็นแบบเบต้า $g_i \sim Beta(n, 2, 2)$

- 9) สร้างเมตริกซ์คำตอบของการประเมิน X_{ik} จากค่าพารามิเตอร์ในข้อ 2 ถึง 7 จากสมการ (1) เมื่อกำหนดให้ไม่มีการทำหน้าที่ต่างกันระหว่างผู้ประเมิน หรือสมการ (2) เมื่อกำหนดให้มีการทำหน้าที่ต่างกันระหว่างผู้ประเมิน

$$Pr(X_{ik}) = (D_i Z_{k,e_i}) + ((1 - D_i)g_i) \quad (1)$$

$$D_{ik} = \frac{\theta_i(1 - \delta_k)}{\theta_i(1 - \delta_k) + \delta_k(1 - \theta_i)} \quad (2)$$

2. การเก็บรวบรวมข้อมูล

จากการกำหนดสถานการณ์จำลองข้างต้น มีสถานการณ์จำลองข้อมูลทั้งหมด $3 \times 3 \times 2 = 18$ สถานการณ์จำลองที่ใช้ในการศึกษาประสิทธิภาพการประมาณค่าฉันทามติเชิงวัฒนธรรมของผู้ประเมินด้วยโมเดล MC-GCM

- 1) ประมาณค่าพารามิเตอร์จากเมตริกซ์ X_{ik} ได้จากการจำลองข้อมูล โดยใช้คำสั่ง `cctapply(data)` ในชุดคำสั่ง `CCTpack` ที่อยู่ในโปรแกรม R โดยกำหนดการประมาณค่าตามค่าตั้งต้นของโมเดล ซึ่งใช้จำนวนตัวอย่าง 10000 หน่วย จำนวนลูกโซ่ 3 ลูกโซ่ จากนั้น บันทึกค่าประมาณพารามิเตอร์ที่ได้จากการประมาณค่าของโมเดลลงในเมตริกซ์ที่สร้างไว้ จำนวน 6 เมตริกซ์ ดังนี้
 - (1) `theta.est` เป็นเมตริกซ์เก็บค่าประมาณความสามารถของผู้ประเมิน (θ_i) ขนาด $R \times N$
เมื่อ R คือ จำนวนรอบของการกระทำซ้ำการจำลองข้อมูล
 - (2) `g.est` เป็นเมตริกซ์เก็บค่าประมาณความลำเอียงในการประเมิน (g_i) ขนาด $R \times N$
 - (3) `del1.est` เป็นเมตริกซ์เก็บค่าประมาณความยากของรายการประเมิน (δ_{k1}) สำหรับผู้ประเมินกลุ่มที่ 1 ขนาด $R \times M$
 - (4) `del2.est` เป็นเมตริกซ์เก็บค่าประมาณความยากของรายการประเมิน (δ_{k2}) สำหรับผู้ประเมินกลุ่มที่ 2 ขนาด $R \times M$
 - (5) `Z1.est` เป็นเมตริกซ์เก็บค่าคำตอบฉันทามติของผู้ประเมินกลุ่มที่ 1 ($Z_{t_1,k}$) สำหรับผู้ประเมินกลุ่มที่ 1 ขนาด $R \times M$
 - (6) `Z2.est` เป็นเมตริกซ์เก็บค่าคำตอบฉันทามติของผู้ประเมินกลุ่มที่ 1 ($Z_{t_2,k}$) สำหรับผู้ประเมินกลุ่มที่ 2 ขนาด $R \times M$
- 2) ทำซ้ำขั้นตอนที่ 1 ถึง 8 จำนวน 100 ครั้ง

3. การวิเคราะห์ข้อมูล

ผู้วิจัยวิเคราะห์ผลการประมาณค่าพารามิเตอร์จากข้อมูลจำลองจำนวน 4 พารามิเตอร์ ได้แก่ ฉันทามติเชิงวัฒนธรรม (Z_k) ความสามารถของผู้ประเมิน (θ_i) ความยากของรายการประเมิน (δ_k) และความ

ลำเอียงในการประเมิน (g_i) โดยกำหนดเงื่อนไขการจำลองข้อมูลที่แตกต่างกัน 18 สถานการณ์สำหรับการประมาณค่าในหนึ่งโมเดล รวมทั้งสิ้น 54 สถานการณ์ ตัวแปรอิสระที่ศึกษา คือ จำนวนผู้ประเมิน (N) แบ่งเป็น 15 30 และ 45 คน จำนวนรายการประเมิน (M) แบ่งเป็น 25 55 และ 85 ข้อ และการทำหน้าที่ต่างกันระหว่างผู้ประเมิน (T) แบ่งเป็น คำถามที่ไม่มีการทำหน้าที่ต่างกันระหว่างผู้ประเมิน และคำถามที่มีการทำหน้าที่ต่างกันระหว่างผู้ประเมิน ตัวแปรตาม คือ ประสิทธิภาพของการประมาณค่าของโมเดล MC-GCM ได้แก่ ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) ค่าความลำเอียงในการประมาณค่า ($Bias$) และค่าสัมประสิทธิ์สหสัมพันธ์ (Pearson Correlation Coefficient) วิเคราะห์ข้อมูลโดยการวิเคราะห์ความแปรปรวนพหุคูณ เพื่อดูปัจจัยที่ส่งผลต่อการประมาณค่าพารามิเตอร์ของโมเดลการวิเคราะห์ฉันทามติเชิงวัฒนธรรม ได้แก่ จำนวนผู้ประเมิน (N) จำนวนรายการประเมิน (M) การทำหน้าที่ต่างกันระหว่างผู้ประเมิน (T)

ผลการวิจัย

1. ประสิทธิภาพการประมาณค่าของโมเดล MC-GCM

การศึกษาประสิทธิภาพในการประมาณค่าของโมเดล MC-GCM พบว่า โมเดลสามารถประมาณค่าพารามิเตอร์ได้ใกล้เคียงกับค่าจริงของพารามิเตอร์ที่กำหนดไว้ ในกรณีที่รายการประเมินมีความเป็นเอกพันธ์ของคำถามประเมิน ผลการเปรียบเทียบระหว่างค่าประมาณของพารามิเตอร์กับค่าที่กำหนด พบว่า ในสถานการณ์ที่ไม่มีการทำหน้าที่ต่างกันระหว่างผู้ประเมิน ($g_i = 0.5$, $\delta_{i,j} = 0.5$) โมเดล MC-GCM สามารถประมาณค่าพารามิเตอร์ได้ใกล้เคียงกับค่าจริงมากกว่าสถานการณ์ที่มีการทำหน้าที่ต่างกันระหว่างผู้ประเมิน ผลการประมาณค่าแสดงในรูป 1 – 4 โดย แถบสีเทาเป็นค่าพารามิเตอร์ที่กำหนดโดยการสุ่ม และแถบสีดำเป็นค่าที่ได้จากการประมาณค่าของโมเดล จากรูปจะเห็นว่าโมเดลสามารถประมาณค่าพารามิเตอร์ในสถานการณ์ที่ 11 ซึ่งเป็นสถานการณ์ที่ไม่มีการทำหน้าที่ต่างกันระหว่างผู้ประเมินได้ใกล้เคียงกับค่าจริงมากกว่าสถานการณ์ที่ 12 ซึ่งเป็นสถานการณ์ที่มีการทำหน้าที่ต่างกันระหว่างผู้ประเมิน โดยเฉพาะการประมาณค่าพารามิเตอร์ความยากในการประเมิน (δ) ($r_{g_{11}} = 0.932$, $r_{g_{12}} = 0.928$) และพารามิเตอร์ความสามารถของผู้ประเมิน (θ) ($r_{\theta_{11}} = 0.891$, $r_{\theta_{12}} = 0.775$)

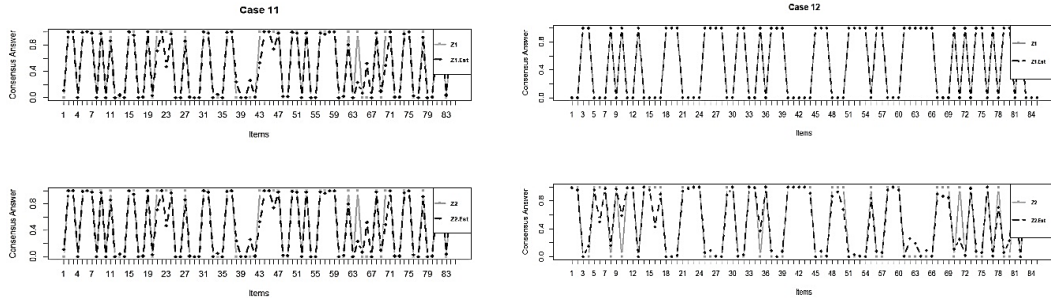
ผลการประมาณค่าพารามิเตอร์ฉันทามติเชิงวัฒนธรรม (Z_k) แสดงในภาพ 1 และตาราง 1 เปรียบเทียบผลการประมาณค่าในสถานการณ์การประเมินที่ไม่มีการทำหน้าที่แตกต่างกันระหว่างผู้ประเมิน (case 11) และสถานการณ์ที่มีการทำหน้าที่ต่างกันระหว่างผู้ประเมิน (case 12) ผลการตรวจสอบความถูกต้องแม่นยำของการประมาณค่าในสถานการณ์การประเมินที่ไม่มีการทำหน้าที่แตกต่างกันระหว่างผู้ประเมิน (case 11) พบว่า มีค่าความลำเอียงในการประมาณค่าฉันทามติเชิงวัฒนธรรมของผู้ประเมินกลุ่มที่ 1 ระหว่าง -0.024 ถึง 0.079 ค่าความลำเอียงในการประมาณค่าฉันทามติเชิงวัฒนธรรมของผู้ประเมินกลุ่มที่ 2 ระหว่าง -0.027 ถึง 0.053 มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองจากการประมาณค่าฉันทามติเชิงวัฒนธรรมของผู้ประเมินกลุ่มที่ 1 ระหว่าง 0.000 ถึง 0.056 ผู้ประเมินกลุ่มที่ 2 มีค่าระหว่าง 0.000 ถึง 0.167 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าจริงและค่าที่ได้จากการประมาณค่าฉันทามติเชิงวัฒนธรรมของผู้ประเมินกลุ่มที่ 1 มีค่าระหว่าง 0.889 ถึง 1.000 ผู้ประเมินกลุ่มที่ 2 มีค่าระหว่าง 0.695 ถึง 1.000 และมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติ

อย่างไรก็ตาม การประมาณค่าของโมเดล MC-GCM ในกรณีที่รายการประเมินมีการทำหน้าที่ต่างกันระหว่างผู้ประเมิน (case 12) จะมีความคลาดเคลื่อนสูงกว่ากรณีที่รายการประเมินไม่มีการทำหน้าที่ต่างกันระหว่างผู้ประเมิน โดยมีความลำเอียงในการประมาณค่าฉันทามติเชิงวัฒนธรรมของผู้ประเมินกลุ่มที่ 1 ระหว่าง -0.003 ถึง 0.015 ค่าความลำเอียงในการประมาณค่าฉันทามติเชิงวัฒนธรรมของผู้ประเมินกลุ่มที่ 2 ระหว่าง -0.006 ถึง 0.075 มีค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสองจากการประมาณค่าฉันทามติเชิงวัฒนธรรมของผู้ประเมินกลุ่มที่ 1 ระหว่าง 0.008 ถึง 0.059 ผู้ประเมินกลุ่มที่ 2 มีค่าระหว่าง 0.003 ถึง 0.062 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าจริงและค่าที่ได้จากการประมาณค่าฉันทามติเชิงวัฒนธรรมของผู้ประเมินกลุ่มที่ 1 มีค่าระหว่าง 0.868 ถึง 0.982 ผู้ประเมินกลุ่มที่ 2 มีค่าระหว่าง 0.766 ถึง 0.992

ตาราง 1 ผลการประมาณค่าพารามิเตอร์ฉันทามติเชิงวัฒนธรรมโดยโมเดล MC-GCM

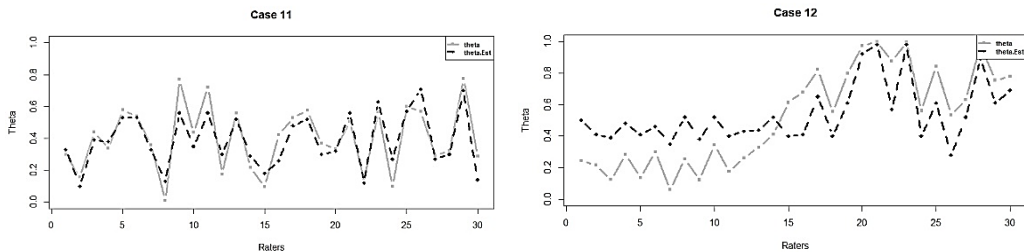
Case	N	M	Item type	Bias		MSE		Corr.	
				Z1	Z2	Z1	Z2	Z1	Z2
1		25	Homo.	0.002	0.053	0.056	0.051	0.889**	0.896**
2			Heter.	-0.103	0.009	0.059	0.003	0.872**	0.992**
3	15	55	Homo.	0.006	0.044	0.003	0.167	0.994**	0.695**
4			Heter.	-0.034	-0.019	0.034	0.017	0.913**	0.915**
5		85	Homo.	0.006	0.002	0.005	0.014	0.971**	0.990**
6			Heter.	0.009	-0.072	0.084	0.021	0.868**	0.766**
7		25	Homo.	-0.023	0	0	0.003	0.996**	0.999**
8			Heter.	-0.004	-0.037	0.011	0.041	0.977**	0.919**
9	30	55	Homo.	0.079	-0.062	0.021	0.008	0.958**	0.984**
10			Heter.	0.015	0.075	0.014	0.062	0.962**	0.919**
11		85	Homo.	-0.024	-0.027	0.022	0.024	0.953**	0.956**
12			Heter.	-0.011	-0.04	0.008	0.032	0.982**	0.936**
13		25	Homo.	0	0	0	0	1.000**	1.000**
14			Heter.	0.008	0	0.007	0	0.905**	0.949**
15	45	55	Homo.	0.001	-0.006	0	0.002	0.999**	0.996**
16			Heter.	-0.052	0.032	0.025	0.035	0.927**	0.949**
17		85	Homo.	0	0	0	0	0.999**	0.999**
18			Heter.	-0.003	0.018	0.021	0.023	0.957**	0.951**

**p-value \leq 0.01



ภาพ 1 การเปรียบเทียบผลการประมาณค่าพารามิเตอร์ Z_k กับค่าที่กำหนด

ผลการประมาณค่าพารามิเตอร์ความสามารถของผู้ประเมิน (θ_j) จากการจำลองข้อมูลผลการประเมินพบว่า ในกรณีที่รายการประเมินมีความเป็นเอกพันธ์ของคำถามประเมิน (ภาพ 2 ซ้าย) โมเดล MC-GCM สามารถประมาณค่าพารามิเตอร์ความสามารถของผู้ประเมินได้ใกล้เคียงค่าที่กำหนดไว้มากกว่ากรณีที่รายการประเมินเป็นวิวิธพันธ์ (ภาพ 2 ขวา) โดยมีค่าความลำเอียงในการประมาณค่าความสามารถของผู้ประเมินระหว่าง -0.088 ถึง 0.028 ค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสองจากการประมาณค่าความสามารถของผู้ประเมิน ระหว่าง 0.009 ถึง 0.028 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าจริงและค่าที่ได้จากการประมาณค่าความสามารถของผู้ประเมิน มีค่าระหว่าง 0.630 ถึง 0.940 และมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติ

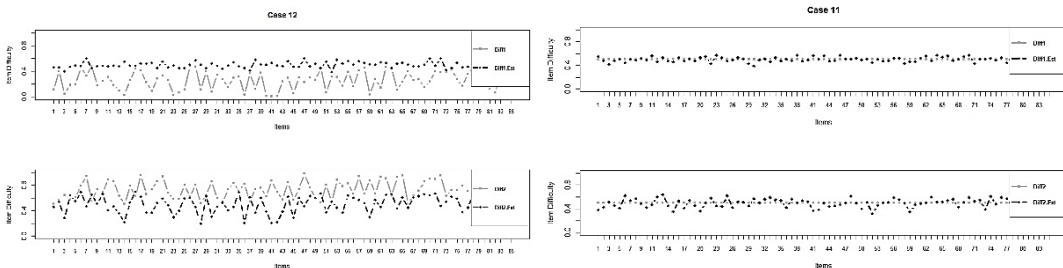


ภาพ 2 การเปรียบเทียบผลการประมาณค่าพารามิเตอร์ θ_j กับค่าที่กำหนด

กรณีที่รายการประเมินมีลักษณะเป็นวิวิธพันธ์ การประมาณค่าพารามิเตอร์ความสามารถของผู้ประเมินของโมเดล MC-GCM จะมีความคลาดเคลื่อนสูงกว่ากรณีที่รายการประเมินมีความเป็นเอกพันธ์ โดยมีค่าความลำเอียงในการประมาณค่าความสามารถของผู้ประเมินมีค่าระหว่าง -0.090 ถึง 0.079 ค่าความลำเอียงในการประมาณค่าฉันทมติเชิงวัฒนธรรมของผู้ประเมินกลุ่มที่ 2 ระหว่าง -0.027 ถึง 0.086 มีค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสองของการประมาณค่าความสามารถของผู้ประเมิน ระหว่าง 0.055 ถึง 0.091 และมีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าจริงและค่าที่ได้จากการประมาณค่าพารามิเตอร์ความสามารถของผู้ประเมินที่ต่ำกว่ารายการประเมินที่เป็นเอกพันธ์ คือมีค่าระหว่าง 0.510 ถึง 0.775 ซึ่งต่ำกว่าค่าสัมประสิทธิ์สหสัมพันธ์ของรายการประเมินที่มีความเป็นเอกพันธ์

ผลการตรวจสอบความถูกต้องของการประมาณค่าพารามิเตอร์ความยากของรายการประเมิน (δ_k) พบว่า ค่าพารามิเตอร์ความยากของรายการประเมิน (δ_k) มีค่าความลำเอียงในการประมาณค่าความยากของรายการประเมินของผู้ประเมินกลุ่มที่ 1 ระหว่าง -0.000 ถึง 0.003 ค่ากลุ่มที่ 2 ระหว่าง -0.002 ถึง 0.002

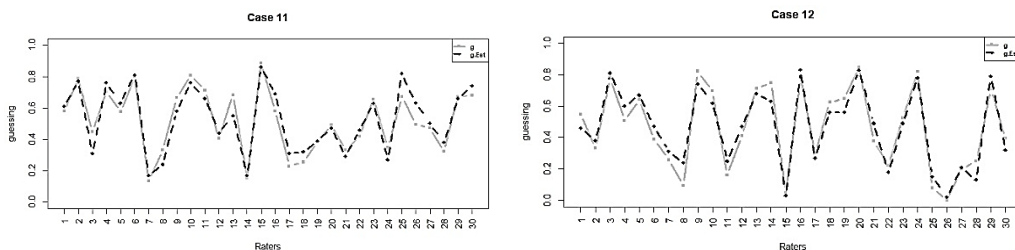
มีค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสองของการประมาณค่าความยากของรายการประเมินของผู้ประเมินกลุ่มที่ 1 ระหว่าง 0.000 ถึง 0.008 กลุ่มที่ 2 มีค่าระหว่าง 0.000 ถึง 0.005 อย่างไรก็ตาม การประมาณค่าของโมเดล MC-GCM ในกรณีที่รายการประเมินมีลักษณะเป็นวิวิธพันธ์ (ภาพ 3 ขวา) มีความคลาดเคลื่อนสูงกว่ากรณีที่รายการประเมินมีความเป็นเอกพันธ์ (ภาพ 3 ซ้าย) โดยมีค่าความลำเอียงในการประมาณค่าพารามิเตอร์ความยากของรายการประเมินของผู้ประเมินกลุ่มที่ 1 ระหว่าง 0.022 ถึง 0.026 ผู้ประเมินกลุ่มที่ 2 ระหว่าง -0.027 ถึง -0.022 มีค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสองจากการประมาณค่าความยากของรายการประเมินของผู้ประเมินกลุ่มที่ 1 ระหว่าง 0.049 ถึง 0.091 ผู้ประเมินกลุ่มที่ 2 มีค่าระหว่าง 0.060 ถึง 0.086 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าจริงและค่าที่ได้จากการประมาณค่าความยากของรายการประเมินของผู้ประเมินกลุ่มที่ 1 มีค่าระหว่าง 0.513 ถึง 0.783 ผู้ประเมินกลุ่มที่ 2 มีค่าระหว่าง 0.370 ถึง 0.697



ภาพ 3 การเปรียบเทียบผลการประมาณค่าพารามิเตอร์ (δ_k) กับค่าที่กำหนด

ผลการประมาณค่าพารามิเตอร์ความลำเอียงในการประเมิน (g_i) จากการจำลองข้อมูล พบว่าโมเดล MC-GCM สามารถประมาณค่าได้ใกล้เคียงกับค่าจริงของพารามิเตอร์ที่กำหนดไว้ ค่าประมาณพารามิเตอร์ความลำเอียงในการประเมิน (g_i) มีค่าความลำเอียงในการประมาณค่า ระหว่าง -0.058 ถึง 0.034 ค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสองจากการประมาณค่า ระหว่าง 0.005 ถึง 0.027 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าจริงและค่าที่ได้จากการประมาณค่าความสามารถของผู้ประเมิน มีค่าระหว่าง 0.663 ถึง 0.940 และมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติ

กรณีที่รายการประเมินมีการทำหน้าที่ต่างกันระหว่างผู้ประเมิน (รูป 4 ขวา) จะมีความคลาดเคลื่อนของการประมาณค่าสูงกว่ากรณีที่รายการประเมินไม่มีการทำหน้าที่ต่างกันระหว่างผู้ประเมิน (รูป 4 ซ้าย) โดยมีค่าความลำเอียงในการประมาณค่า ระหว่าง -0.038 ถึง 0.152 ค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสอง ระหว่าง 0.005 ถึง 0.057 ค่า และค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าจริงและค่าที่ได้จากการประมาณค่า มีค่าระหว่าง 0.702 ถึง 0.985



ภาพ 4 การเปรียบเทียบผลการประมาณค่าพารามิเตอร์ g_i กับค่าที่กำหนด

2. ผลการวิเคราะห์ความแปรปรวนพหุคูณ

ผลการวิเคราะห์ปัจจัยที่ส่งผลต่อการประมาณค่าโดยการวิเคราะห์ความแปรปรวนพหุคูณ (MANOVA) ระหว่างจำนวนผู้ประเมิน (N) จำนวนรายการประเมิน (M) และการทำหน้าที่ต่างกันระหว่างผู้ประเมินที่ส่งผลต่อประสิทธิภาพของการประมาณค่าของโมเดล จากตาราง 2 แสดงให้เห็นว่าปัจจัยที่ส่งผลต่อประสิทธิภาพในการประมาณค่าพารามิเตอร์ความสามารถของผู้ประเมิน (θ_k) และพารามิเตอร์ความยากของรายการประเมิน (δ_k) คือ ตัวแปรการทำหน้าที่ต่างกันระหว่างผู้ประเมิน (T) ซึ่งส่งผลต่อค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสอง (MSE) และค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าจริงและค่าที่ได้จากการประมาณค่าของโมเดล (Corr) อย่างมีนัยสำคัญทางสถิติ นอกจากนี้ จำนวนผู้ประเมิน (N) และจำนวนรายการประเมิน (M) ยังเป็นปัจจัยหนึ่งที่ส่งผลต่อการประมาณค่าพารามิเตอร์ g โดยจำนวนผู้ประเมินส่งผลต่อความลำเอียงในการประมาณค่า ในขณะที่จำนวนรายการประเมินส่งผลต่อค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสอง (MSE) และค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าจริงและค่าที่ได้จากการประมาณค่าของโมเดล (Corr) อย่างมีนัยสำคัญทางสถิติ

ตาราง 2 ผลการวิเคราะห์ปัจจัยที่ส่งผลต่อการประมาณค่าของโมเดล MC-GCM

Params	Source	Dependent Variables	SS	df	MS	F	Sig.
Theta	N	Bias	0.001	2.000	0.001	0.291	0.751
		MSE	0.000	2.000	0.000	0.258	0.776
		Corr	0.021	2.000	0.011	0.450	0.646
	M	Bias	0.008	2.000	0.004	2.133	0.153
		MSE	0.000	2.000	0.000	0.067	0.936
		Corr	0.010	2.000	0.005	0.194	0.825
	T	Bias	0.002	1.000	0.002	0.907	0.355
		MSE	0.011	1.000	0.011	79.736	0.000**
		Corr	0.243	1.000	0.243	29.037	0.000**
g	N	Bias	0.017	2.000	0.008	6.423	0.010**
		MSE	0.000	2.000	0.000	0.665	0.529
		Corr.	0.013	2.000	0.006	0.814	0.462
	M	Bias	0.002	2.000	0.001	0.392	0.682
		MSE	0.001	2.000	0.001	7.635	0.005**
		Corr.	0.076	2.000	0.038	10.190	0.002**
	T	Bias	0.003	1.000	0.003	1.415	0.252
		MSE	0.000	1.000	0.000	1.281	0.274
		Corr	0.008	1.000	0.008	1.052	0.320
Delta1	N	Bias	0.000	2.000	0.000	0.002	0.998
		MSE	0.000	2.000	0.000	0.063	0.940
		Corr	0.039	2	0.020	0.141	0.870
	M	Bias	0.000	2.000	0.000	0.008	0.992
		MSE	0.000	2.000	0.000	0.010	0.990
		Corr	0.002	2	0.001	0.008	0.993

ตาราง 2 (ต่อ)

Params	Source	Dependent Variables	SS	df	MS	F	Sig.	
Delta2	T	Bias	0.003	1.000	0.003	585.047	0.000**	
		MSE	0.027	1.000	0.027	415.577	0.000**	
		Corr	2.035	1	2.035	392.638	0.000**	
	N	M	Bias	0.000	2.000	0.000	0.007	0.993
			MSE	0.000	2.000	0.000	0.021	0.980
			Corr	0.025	2	0.013	0.109	0.898
		T	Bias	0.000	2.000	0.000	0.007	0.993
			MSE	0.000	2.000	0.000	0.030	0.971
			Corr	0.003	2	0.001	0.011	0.989
Z1	T	Bias	0.002	1.000	0.002	1180.568	0.000**	
		MSE	0.022	1.000	0.022	561.001	0.000**	
		Corr	1.665	1	1.665	305.268	0.000**	
	N	M	Bias	0.002	2.000	0.001	0.677	0.523
			MSE	0.004	2.000	0.002	4.259	0.034*
			Corr	0.010	2.000	0.005	3.171	0.071
		T	Bias	0.002	2.000	0.001	0.609	0.557
			MSE	0.000	2.000	0.000	0.140	0.870
			Corr	0.001	2.000	0.001	0.275	0.763
M	Bias	0.003	1.000	0.003	2.335	0.146		
	MSE	0.001	1.000	0.001	2.598	0.127		
	Corr	0.009	1.000	0.009	5.461	0.033*		
	Z2	T	Bias	0.002	2.000	0.001	0.458	0.647
			MSE	0.004	2.000	0.002	1.102	0.373
			Corr	0.032	2.000	0.016	2.287	0.157
N		M	Bias	0.003	2.000	0.002	0.833	0.466
			MSE	0.004	2.000	0.002	1.114	0.370
			Corr	0.007	2.000	0.004	0.526	0.608
		T	Bias	0.000	1.000	0.000	0.060	0.810
			MSE	0.000	1.000	0.000	0.057	0.815
			Corr	0.003	1.000	0.003	0.438	0.521

**p-value \leq 0.01, *p-value \leq 0.05

อภิปรายผล

การศึกษาประสิทธิภาพของโมเดลการวิเคราะห์ถดถอยพหุคูณเชิงพรรณนาในการวิเคราะห์การทำหน้าที่ต่างกันระหว่างผู้ประเมิน: การศึกษาจากสถานการณ์จำลอง มีวัตถุประสงค์เพื่อตรวจสอบประสิทธิภาพของการประมาณค่าพารามิเตอร์ของโมเดล รวมถึงศึกษาปัจจัยที่ส่งผลต่อประสิทธิภาพในการประมาณค่าของโมเดล โดยศึกษาด้วยวิธีการใช้สถานการณ์จำลองด้วยวิธี MCMC ผู้วิจัยกำหนดเงื่อนไขในการจำลองสถานการณ์ทั้งหมด 18 สถานการณ์ จากนั้นเปรียบเทียบผลการประมาณค่าระหว่างค่าจริงที่กำหนดกับค่าที่ได้จากการประมาณค่าของโมเดล MC-GCM นอกจากนี้ ผู้วิจัยได้วิเคราะห์ปัจจัยที่ส่งผลต่อผลการประมาณค่า คือ

ค่าความลำเอียงในการประมาณค่า (Bias) ค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสอง (MSE) และค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าจริงและค่าประมาณพารามิเตอร์ (Pearson correlation coefficient) โดยการวิเคราะห์ความแปรปรวนพหุคูณ

ผลการศึกษานี้แสดงให้เห็นว่าโมเดล MC-GCM สามารถประมาณค่าพารามิเตอร์ที่กำหนดขึ้นได้อย่างมีประสิทธิภาพ โดยเห็นได้จากค่าความลำเอียงในการประมาณค่า (Bias) ค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสอง (MSE) ที่มีค่าเข้าใกล้ 0 ($-0.103 \leq bias \leq 0.086$, $0 \leq mse \leq 0.167$) และค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าจริงและค่าประมาณพารามิเตอร์มีค่าใกล้ 1 ซึ่งสอดคล้องกับการศึกษาของ Anders และ Batchelder (2012) ว่าโมเดล MC-GCM สามารถประมาณค่าพารามิเตอร์ที่กำหนดได้อย่างมีประสิทธิภาพ อย่างไรก็ตาม การศึกษาจากการจำลองข้อมูล พบว่า โมเดล MC-GCM มีประสิทธิภาพในการประมาณค่าต่ำลงเมื่อกำหนดรายการประเมินให้มีการทำหน้าที่ต่างกันระหว่างผู้ประเมิน ทั้งนี้ ไม่พบประเด็นดังกล่าวในการศึกษาก่อนหน้า ซึ่งอาจเกิดจากความคลาดเคลื่อนในการสุ่มตัวอย่างในการจำลองข้อมูลหรือปัจจัยอื่นซึ่งควรมีการศึกษาเพิ่มเติมก่อนการนำโมเดลไปประยุกต์ใช้ในสถานการณ์จริงของการประเมินต่อไป ผลการวิเคราะห์ความแปรปรวนพหุคูณ แสดงให้เห็นว่าตัวแปรจำนวนผู้ประเมิน (N) จำนวนรายการประเมิน (M) และการทำหน้าที่ต่างกันระหว่างผู้ประเมิน (T) ส่งผลต่อประสิทธิภาพในการประมาณค่า ทั้งนี้ จากการศึกษาของ Romney, A. K., Weller, S. C., & Batchelder, W. H. (1986) และการศึกษาของ Batchelder W. H. และ Anders R. (2012) พบว่าหากเพิ่มจำนวน N และ M ขึ้น ผลการประมาณค่าจะมีความคลาดเคลื่อนน้อยลง

จากผลการศึกษาดังกล่าว แสดงว่าโมเดล MC-GCM สามารถประมาณค่าพารามิเตอร์ได้ใกล้เคียงกับค่าที่กำหนด แสดงให้เห็นว่าโมเดลการวิเคราะห์ฉันทมติเชิงวัฒนธรรมสามารถนำมาใช้ประยุกต์สำหรับการวิเคราะห์การประเมินที่มีการทำหน้าที่ต่างกันระหว่างผู้ประเมินได้อย่างมีประสิทธิภาพ

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลวิจัยไปใช้

ผลการศึกษาจากการจำลองสถานการณ์การประเมินที่มีการทำหน้าที่ต่างกันระหว่างผู้ประเมิน พบว่าโมเดลการวิเคราะห์ฉันทมติเชิงวัฒนธรรมสามารถประมาณค่าความสอดคล้องระหว่างผู้ประเมินได้อย่างมีประสิทธิภาพในกรณีที่ไม่มีการทำหน้าที่ต่างกันระหว่างผู้ประเมิน ซึ่งนักวิจัยและนักวัดผลสามารถประยุกต์ใช้โมเดลการวิเคราะห์ดังกล่าวในการวิเคราะห์ความสอดคล้องระหว่างผู้ประเมินได้ โดยสามารถศึกษาการใช้งานชุดคำสั่ง CCTpack ได้จากคู่มือที่มีอยู่ในโปรแกรม R

2. ข้อเสนอแนะในการวิจัยครั้งต่อไป

การศึกษานี้ได้ทำการศึกษาตัวแปรที่คาดว่าจะส่งผลต่อการประมาณค่าของโมเดลการวิเคราะห์ฉันทมติเชิงวัฒนธรรมจำนวน 3 ตัวแปร ได้แก่ จำนวนผู้ประเมิน จำนวนรายการประเมิน และการทำหน้าที่ต่างกันระหว่างผู้ประเมิน อย่างไรก็ตาม มีปัจจัยอื่นที่อาจจะส่งผลต่อการประมาณค่าที่ยังไม่มีการศึกษาเพิ่มเติม

อาทิ ระดับความสอดคล้องของการให้คะแนนการประเมินในเมตริกซ์การประเมิน หรือการประยุกต์ใช้โมเดลการวิเคราะห์ดังกล่าวในบริบทการประเมินในสถานการณ์จริง

เอกสารอ้างอิง

- Batchelder W. H., Anders R. (2012). Cultural Consensus Theory: Comparing Different Concepts of Cultural Truth. *Journal of Mathematical Psychology*, 56, 316-332.
- Engelhard, G. Jr., Wind, S. A., Jennifer, L. K., Chajewski, M. (2013). *Differential Item and Person Functioning in Large-Scale Writing Assessments within the Context of the SAT*. Research report. College Board.
- Farrokhi et al. (2012). A Many-Facet Rasch Measurement of Differential Rater Severity/Leniency in Three Types of Assessment. *JALT Journal*, 34(1), 79-102.
- Muckle, T. J., Karabatsos, G. (2009). Hierarchical Generalized Linear Models for the Analysis of Jude Rating. *Journal of Education Measurement*, 46(2), 198-219.
- Myford, C. M., & Wolfe, E. W. (2009). Monitoring Rater Performance over Time: A Framework for Detecting Differential Accuracy and Differential Scale Category Use. *Journal of Educational Measurement*. 46(4), 371-389.
- Patz, R. J., Junker, B. W., Johnson, M. S., and Mariano, L. T. (2002). The Hierarchical Rater Model for Rated Test Items and its Application to Large-Scale Educational Assessment Data. *Journal of Educational and Behavioral Statistics*, 27(4), 341-384.
- Romney, A. K., Weller, S. C., & Batchelder, W. H. (1986). Culture as consensus: A theory of culture and informant accuracy. *American anthropologist*, 88(2), 313-338.
- Schaefer, E. (2008). Rater Bias Patterns in an EFL Writing Assessment. *Language Testing*, 28(4), 465-493.
- Wesoloaki, B. C., Wind, S. A., & Engelhard, G. Jr. (2015). Rater Fairness in Music Performance Assessment: Evaluating Model-Data and Differential Rater Functioning. *Musicae Scientiae*, 19920, 147-170.
- Xun Yan. (2014). An Examination of rater performance on a local oral English proficiency test: A mixed-methods approach. *Language Testing*, 31(4), 501-527.