

การเปรียบเทียบผลการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบสำหรับแบบสอบรูปแบบผสม : วิธีการประมาณค่าภาวะน่าจะเป็นสูงสุดชายขอบและวิธีการปรับเรียบเคอร์เนล

A Comparison of Parameter Estimation of Ability Parameters for Mixed Format Tests: The Marginal Maximum Likelihood and Kernel Smoothing Methods

นัยนา จันตะเสน และคณะ

บทคัดย่อ

การวิจัยครั้งนี้เป็นการวิจัยเชิงทดลอง มีวัตถุประสงค์เพื่อ ๑) เปรียบเทียบผลการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบสำหรับแบบสอบรูปแบบผสม ระหว่างวิธีการประมาณค่าภาวะน่าจะเป็นสูงสุดชายขอบ (MML) และวิธีการปรับเรียบเคอร์เนล (KS) ๒) วิเคราะห์สัดส่วนของขนาดตัวอย่าง ความยาวของแบบสอบ และเปอร์เซ็นต์ของข้อสอบที่ไม่สอดคล้องกับโมเดลการตอบสนองข้อสอบ ภายใต้เงื่อนไขการฝ่าฝืนข้อตกลงเบื้องต้นด้านการเลือกใช้โมเดลการตอบสนองข้อสอบ ในการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบระหว่างวิธี MML และวิธี KS ที่ส่งผลต่อค่าประมาณพารามิเตอร์ความสามารถของผู้สอบ ข้อมูลที่ใช้ในการศึกษาครั้งนี้ ใช้ข้อมูลจำลองภายใต้การเขียนคำสั่งการจำลองข้อมูลและประมวลผลด้วยโปรแกรม R จากเงื่อนไขในการศึกษา ๒๗ เงื่อนไข ดังนี้ [ขนาดตัวอย่าง ๓ เงื่อนไข ได้แก่ ๓๐๐ ๕๐๐ และ ๑,๐๐๐ แทนตัวอย่างขนาดเล็ก กลาง และใหญ่ตามลำดับ] $\times$ (ความยาวของแบบสอบ ๓ เงื่อนไข ได้แก่ ๑๐ ๒๐ และ ๔๐ ข้อ แทนความยาวของแบบสอบสั้น ปานกลางและยาวตามลำดับ) $\times$ (เปอร์เซ็นต์ของข้อสอบที่ไม่สอดคล้องกับโมเดลการตอบสนองข้อสอบภายใต้เงื่อนไขการฝ่าฝืนข้อตกลงเบื้องต้นด้านการเลือกใช้โมเดลการตอบสนองข้อสอบ ๓ เงื่อนไข ได้แก่ ๑๐ ๒๐ และ ๓๐ เปอร์เซ็นต์)

ผลการวิจัยในภาพรวมสรุปได้ว่า ๑) ผลการวิเคราะห์ เปรียบเทียบค่าประมาณพารามิเตอร์ความสามารถของผู้สอบ สำหรับแบบสอบรูปแบบผสม ระหว่างวิธี MML และวิธี KS กรณีข้อสอบที่สอดคล้องกับโมเดลการตอบสนองข้อสอบ การประมาณค่าความสามารถของผู้สอบโดยรวมด้วยวิธี MML มีค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) น้อยกว่าวิธี KS และกรณีข้อสอบที่ไม่สอดคล้องกับโมเดลการตอบสนองข้อสอบภายใต้เงื่อนไขการฝ่าฝืนข้อตกลงเบื้องต้นด้านการเลือกใช้โมเดลการตอบสนองข้อสอบ การประมาณค่าความสามารถของผู้สอบโดยรวมด้วยวิธี MML มีค่า AMSE มากกว่าวิธี KS ๒) ผลการวิเคราะห์สัดส่วนของขนาดตัวอย่าง ความยาวของแบบสอบ และเปอร์เซ็นต์

ของข้อสอบที่ไม่สอดคล้องกับโมเดลการตอบสนองข้อสอบ ด้วยการวิเคราะห์ความแปรปรวนสองทาง พบว่า ภายใต้เงื่อนไขการประมาณค่าความสามารถของผู้สอบด้วยวิธี KS สำหรับแบบสอบที่มีเปอร์เซ็นต์ของข้อสอบที่ไม่สอดคล้องกับโมเดลการตอบสนองข้อสอบ ๑๐ เปอร์เซ็นต์ มีปฏิสัมพันธ์ระหว่างขนาดตัวอย่างและจำนวนข้อสอบร่วมกันส่งผลต่อค่าดัชนี AMSE อย่างมีนัยสำคัญทางสถิติที่ระดับ .๐๕ ($F=3.398$, $p=.01$)

Abstract

This research was an experimental research aiming to: 1) Compare the results of parameter estimation of ability parameters for mixed format tests between the Marginal Maximum Likelihood (MML) and the Kernel Smoothing (KS) methods. 2) Analyze the proportion of the sample size, the test length and the percentage of mis-fitted items under the violation of assumptions on the use of item response models to estimate the ability of the examinees between the MML and the KS methods. In this study, the data was simulated and assessed with program R. The total data studied included 27 conditions. The data were generated with three sample sizes (300, 500 and 1,000 samples), each fewer than three test lengths (10, 20 and 40 items), and percentage of mis-fitted items (10%, 20% and 30%).

The research results were as follows: 1) The analysis results of compared estimated ability of the examinees for mixed format tests between the MML and the KS methods; it was found that the estimate of the overall ability parameter with the MML method had a lower Average Mean Squared error (AMSE) than the KS method for the fitting item response model. However, for mis-fitting items under the violation of assumptions on the use of item response models, the overall ability parameter with MML method had a higher AMSE than the KS method. 2) The analysis results of the proportion of the sample size and the test length and percentage of mis-fitted items under violation of the assumptions on the use of item response models with two way of variance (two way ANOVA) found that there is an interaction between sample size and number of items. 10 percent of mis-fitted items with the KS approach were effected significantly on the AMSE indices at p -value < 0.05 ($F=3.398$, $p=.01$).

คำสำคัญ: ทฤษฎีการตอบสนองข้อสอบแบบพารามेटริก/ทฤษฎีการตอบสนองข้อสอบแบบนึ่งพารามेटริก/วิธีการประมาณค่าภาวะน่าจะเป็นสูงสุดชายขอบ (MML)/วิธีการปรับเรียบเคอร์เนล (KS)

KEYWORD: PARAMETRIC ITEM RESPONSE THEORY/NONPARAMETRIC ITEM RESPONSE THEORY/MARGINAL MAXIMUM LIKELIHOOD (MML)/ KERNEL SMOOTHING (KS)

บทนำ

ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) เป็นทฤษฎีที่เกี่ยวข้องกับรูปแบบทางคณิตศาสตร์ ซึ่งเกี่ยวกับความสัมพันธ์ระหว่างระดับความสามารถของผู้สอบ (θ) และความน่าจะเป็นในการตอบสนองข้อสอบ $P(\theta)$ ในแต่ละข้อว่ามีความน่าจะเป็นในการตอบถูกได้มากน้อยเพียงใด อธิบายโดยใช้โค้งคุณลักษณะข้อสอบ (Item Characteristic Curve: ICC) ซึ่งมีลักษณะเป็นฟังก์ชันทางคณิตศาสตร์ที่เรียกว่า ฟังก์ชันโลจิสติก (logistic function) ซึ่งทำให้ค่าพารามิเตอร์ของข้อสอบที่ได้จากการวิเคราะห์คุณภาพของข้อสอบ ทั้งค่าความยาก (a) ค่าอำนาจจำแนก (b) และค่าโอกาสในการเดาของข้อสอบ (c) แต่ละข้อเป็นคุณลักษณะเฉพาะของข้อสอบที่มีอยู่ประจำ และคงที่ในตัวข้อสอบจึงทำให้การวิเคราะห์ข้อสอบด้วยทฤษฎีการตอบสนองข้อสอบนี้ ค่าพารามิเตอร์ของข้อสอบไม่ผันแปรไปตามกลุ่มผู้สอบ (ศิริชัย กาญจนวาสี, ๒๕๕๐) ซึ่งคุณลักษณะที่อธิบายไว้นี้ เป็นการอ้างอิงตามโมเดลการตอบสนองข้อสอบแบบพาราเมตริก (อิงพารามิเตอร์) เพราะต้องการฟังก์ชันการตอบสนองข้อสอบ (Item Response Function: IRF) ที่เป็นไปตามพารามิเตอร์ที่กำหนดไว้ อย่างไรก็ตามในข้อสอบที่สอดคล้องกับทฤษฎีการตอบสนองข้อสอบแบบพาราเมตริก (Parametric Item Response Theory: PIRT) ข้อสอบบางข้อที่ใช้ในการประเมินทางการศึกษาหรือทางจิตวิทยา การตอบสนองข้อสอบอาจไม่เป็นไปตามที่แสดงออกหรือตามที่กำหนดไว้ ดังนั้นการนำโมเดลการตอบสนองข้อสอบแบบพาราเมตริกไปใช้ต้องพิจารณาความสอดคล้องระหว่างรูปแบบที่เลือกไว้กับลักษณะของข้อมูลเป็นสำคัญ นอกจากนี้ผู้ใช้เป็น

ต้องเลือกรูปแบบที่เหมาะสมแล้วยังต้องเลือกวิธีการประมาณค่าพารามิเตอร์ที่มีประสิทธิภาพเหมาะสมกับสภาพการทดสอบของแต่ละประเภทเนื่องจากวิธีประมาณค่าแต่ละวิธีมีความแตกต่างกันหลาย ๆ ด้าน ไม่ว่าจะเป็นเรื่องของความสะดวกในการคำนวณ หรือขนาดของกลุ่มผู้เข้าสอบที่จำเป็นต้องมีเพื่อให้ค่าประมาณใกล้เคียงกับความสามารถที่แท้จริงของผู้สอบ โดยเฉพาะอย่างยิ่งความแตกต่างในด้านแนวคิดทางสถิติที่ทำให้ค่าพารามิเตอร์ที่ประมาณได้จากแต่ละวิธีแตกต่างกันเป็นผลให้ค่าฟังก์ชันสารสนเทศ ซึ่งเป็นดัชนีผสมของค่าพารามิเตอร์แตกต่างกันด้วย สำหรับการศึกษเกี่ยวกับการประเมินความสอดคล้องของโมเดลการตอบสนองข้อสอบแบบพาราเมตริก (PIRT) นั้น พารามิเตอร์ของรูปแบบการตอบสนองข้อสอบมักจะประมาณค่าด้วยวิธีการประมาณค่าภาวะน่าจะเป็นสูงสุดชายขอบ (MML) ส่วนการประเมินความสอดคล้องของโมเดลการตอบสนองข้อสอบตามทฤษฎีการตอบสนองข้อสอบแบบนันทพาราเมตริก (Nonparametric Item Response Theory: NIRT) นิยมใช้วิธีการปรับเรียบเคอร์เนล (Kernel smoothing) เนื่องจากการปรับเรียบเคอร์เนลได้แสดงให้เห็นถึงประสิทธิภาพที่ดีที่สามารถนำไปตรวจสอบข้อสอบที่สอดคล้องกับโมเดลการตอบสนองข้อสอบแบบพาราเมตริก (PIRT) ได้ ดังนั้นจึงมีบทความ และวรรณคดีที่เกี่ยวข้องเสนอแนะให้พัฒนาและขยายวิธีการปรับเรียบเคอร์เนลให้สามารถใช้งานได้ทั่วไปในสถานการณ์ต่างๆ ของการวัดให้มากขึ้น โดย (Chon, Lee & Dunbar 2010; Lee, Wollack & Douglas 2009) เสนอแนะให้สร้างแบบสอบรูปแบบผสม เพื่อให้สอดคล้องกับสถานการณ์จริงของปัญหาในการสร้างแบบสอบมากที่สุด เนื่องจาก

ในทศวรรษที่ ๒๐ และในทศวรรษหน้านั้นมีความแตกต่างกันอย่างมากจากการประเมินในสองทศวรรษที่ผ่านมา รูปแบบการประเมินจะมีความเป็นปรนัยมากขึ้น เช่น แบบสอบเลือกตอบจะถูกนำมาใช้ร่วมกับแบบสอบรูปแบบอื่นมากขึ้น แบบสอบรูปแบบผสมระหว่างข้อสอบ ที่มีรูปแบบต่างกันจะถูกนำไปใช้กับการสอบทั้งในระดับชั้นเรียนและการสอบระดับมหภาค (Hambleton, 1996)

ดังนั้นผู้วิจัยจึงสนใจศึกษาการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ สำหรับแบบสอบรูปแบบผสม ตามแนวทางของ Chon, Lee & Dunbar (2010) ได้แก่ แบบสอบรูปแบบผสมรูปแบบ 1PL/PCM ประกอบด้วยข้อสอบแบบตรวจให้คะแนน ๒ ค่า ที่มีโมเดลการตอบสนองข้อสอบชนิด ๑ พารามิเตอร์ และข้อสอบแบบตรวจให้คะแนนมากกว่า ๒ ค่า ที่มีโมเดลการตอบสนองข้อสอบแบบการตรวจให้คะแนนความรู้บางส่วน โดยมีสัดส่วนของข้อสอบที่ไม่สอดคล้องกับโมเดลการตอบสนองข้อสอบภายใต้เงื่อนไขการฝ่าฝืนข้อตกลงเบื้องต้นด้านการเลือกใช้โมเดลการตอบสนองข้อสอบ (Item Response Models) ร่วมอยู่ด้วย โดยศึกษาภายใต้เงื่อนไขสัดส่วนของขนาดตัวอย่าง ความยาวของแบบสอบ และเปอร์เซ็นต์ของข้อสอบที่ไม่สอดคล้องกับโมเดลการตอบสนองข้อสอบภายใต้เงื่อนไขการฝ่าฝืนข้อตกลงเบื้องต้นด้านการเลือกใช้โมเดลการตอบสนองข้อสอบ ซึ่งการเป็นตัวแทนของเงื่อนไขทั้งสามดังกล่าว นั้น สอดคล้องกับบทความและงานวิจัยของ (Chon, Lee & Dunbar 2010; Douglas & Cohen 2001; Lee (2007; Lee, Wollack & Douglas 2009; Liang & Wells 2009; Sueiro & Abad 2011 และ Wells & Bolt 2008) และเพื่อศึกษาสัดส่วนของเงื่อนไขทั้งสามที่ส่งผลต่อ

การประมาณค่าพารามิเตอร์ความสามารถของผู้สอบว่าควรเลือกที่สัดส่วนเท่าใด สำหรับการประมาณค่าด้วยวิธี MML หรือวิธี KS เพื่อให้ได้ค่าประมาณพารามิเตอร์ความสามารถของผู้สอบที่เหมาะสมใกล้เคียงกับความสามารถที่แท้จริงของผู้สอบมากที่สุด

วัตถุประสงค์ของการวิจัย

๑. เพื่อเปรียบเทียบผลการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบสำหรับแบบสอบรูปแบบผสม ระหว่างวิธีการประมาณค่าภาวะน่าจะเป็นสูงสุดชายขอบ และวิธีการปรับเรียบเคอร์เนล

๒. เพื่อวิเคราะห์สัดส่วนของขนาดตัวอย่าง ความยาวของแบบสอบ และเปอร์เซ็นต์ของข้อสอบที่ไม่สอดคล้องกับโมเดลการตอบสนองข้อสอบภายใต้เงื่อนไขการฝ่าฝืนข้อตกลงเบื้องต้นด้านการเลือกใช้โมเดลการตอบสนองข้อสอบ ในการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบระหว่างวิธีการประมาณค่าภาวะน่าจะเป็นสูงสุดชายขอบ และวิธีการปรับเรียบเคอร์เนล ที่ส่งผลต่อค่าประมาณความสามารถของผู้สอบ

ขอบเขตของการวิจัย

การศึกษาในครั้งนี้มีขอบเขตของการศึกษาดังนี้

๑. การศึกษาในครั้งนี้เป็นการศึกษาโดยการจำลองข้อมูล (simulation) ให้สอดคล้องกับโมเดลการตอบสนองข้อสอบตามแบบสอบรูปแบบผสมรูปแบบ 1PL/PCM ประกอบด้วยประชากรข้อสอบแบบตรวจให้คะแนน ๒ ค่า ที่มีโมเดลการตอบสนองข้อสอบชนิด ๑ พารามิเตอร์ และ

ประชากรข้อสอบแบบตรวจให้คะแนนมากกว่า ๒ ค่า ที่มีโมเดลการตอบสนองข้อสอบแบบการตรวจให้คะแนนความรู้บางส่วน (PCM) โดยมีสัดส่วนของข้อสอบที่ไม่สอดคล้องกับโมเดลการตอบสนองข้อสอบภายใต้เงื่อนไขการฝ่าฝืนข้อตกลงเบื้องต้นด้านการเลือกใช้โมเดลการตอบสนองข้อสอบ ได้แก่ ประชากรข้อสอบแบบตรวจให้คะแนน ๒ ค่า ที่มีโมเดลการตอบสนองข้อสอบชนิด ๔ พารามิเตอร์ และประชากรข้อสอบแบบตรวจให้คะแนนมากกว่า ๒ ค่า ที่มีโมเดลการตอบสนองข้อสอบแบบการตรวจให้คะแนนความรู้บางส่วนทั่วไป (GPCM) รวมอยู่ด้วย

๒. การประมาณค่าพารามิเตอร์ความสามารถของผู้สอบสำหรับแบบสอบรูปแบบผสมในครั้งนี้ ผู้วิจัยดำเนินการประมาณค่าความสามารถของผู้สอบด้วยวิธีการประมาณค่าภาวะน่าจะเป็นสูงสุดชายขอบ (แทนวิธีการประมาณค่าแบบพารามิเตอร์) และวิธีการปรับเรียบเคอร์เนล (แทนวิธีการประมาณค่าแบบนัยพารามิเตอร์)

๓. เงื่อนไขในการจำลองข้อมูล ประกอบด้วย (รูปแบบของแบบสอบรูปแบบผสม ๑ เงื่อนไข ได้แก่ แบบสอบรูปแบบผสมรูปแบบ 1PL/PCM) $\times$ (ขนาดตัวอย่าง ๓ เงื่อนไข ได้แก่ ๓๐๐ ๕๐๐ และ ๑,๐๐๐ แทนตัวอย่างขนาดเล็ก กลาง และใหญ่ตามลำดับ) $\times$ (ความยาวของแบบสอบ ๓ เงื่อนไข ได้แก่ ๑๐ ๒๐ และ ๔๐ ข้อ แทนความยาวของแบบสอบสั้น ปานกลาง และยาวตามลำดับ) $\times$ (เปอร์เซ็นต์ของข้อสอบที่ไม่สอดคล้องกับโมเดลการตอบสนองข้อสอบภายใต้เงื่อนไขการฝ่าฝืนข้อตกลงเบื้องต้นด้านการเลือกใช้โมเดลการตอบสนองข้อสอบ ๓ เงื่อนไข ได้แก่ ๑๐ ๒๐ และ ๓๐ เปอร์เซ็นต์) รวมเงื่อนไขข้อมูลที่ต้องวิเคราะห์ภายใต้แบบสอบรูปแบบผสม ($๑ \times ๓ \times ๓ \times ๓$

$= ๒๗$ เงื่อนไข) ๒๗ เงื่อนไข ซึ่งสอดคล้องกับการศึกษาของ (Chon, Lee & Dunbar, 2010; Douglas & Cohen, 2001; Lee 2007; Lee, Wollack & Douglas, 2009; Liang & Wells, 2009; Sueiro & Abad, 2011 และ Wells & Bolt, 2008)

๔. ตัวแปรในการวิจัยประกอบด้วย

๔.๑ ตัวแปรอิสระ ได้แก่ สัดส่วนของแบบสอบรูปแบบผสม 1PL/PCM เป็น ๔๐ : ๒๐ สำหรับแบบสอบรูปแบบตรวจให้คะแนน ๒ ค่า และแบบสอบรูปแบบตรวจให้คะแนนมากกว่า ๒ ค่า ตามลำดับ ภายใต้เงื่อนไขในการวิเคราะห์ข้อมูล ๓ ตัวแปร ประกอบด้วย จำนวนตัวอย่างในการวิเคราะห์ ๓ เงื่อนไข ได้แก่ ๓๐๐ ๕๐๐ และ ๑,๐๐๐ คน จำนวนข้อสอบที่ใช้ในการวิเคราะห์ข้อมูล ๓ เงื่อนไข ได้แก่ ๑๐ ๒๐ และ ๔๐ ข้อ และสัดส่วนเปอร์เซ็นต์ของข้อสอบที่ไม่สอดคล้องกับโมเดลการตอบสนองข้อสอบ ๓ เงื่อนไข ได้แก่ ๑๐ ๒๐ และ ๓๐ เปอร์เซ็นต์

๔.๒ ตัวแปรตาม ได้แก่ ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ย (Average Mean Square Error: AMSE)

วิธีการวิจัย

การศึกษาในครั้งนี้มี ๒ ขั้นตอนในการวิจัยดังต่อไปนี้

ขั้นตอนการจำลองข้อมูล

๑. ศึกษาเอกสาร และงานวิจัยที่เกี่ยวข้องทั้งในประเทศ และต่างประเทศที่เกี่ยวข้องกับแนวคิด วิธีการ จุดเด่น จุดด้อย ข้อเสนอแนะในการวิจัย ในทฤษฎีการตอบสนองข้อสอบทั้งแนวคิด

ทฤษฎีแบบพาราเมตริกและแบบนัยพาราเมตริก และศึกษาแนวคิด หลักการ วิธีการใช้งาน การเขียน คำสั่งในการประมวลผลสำหรับการวิเคราะห์ ข้อมูลด้วยโปรแกรม R

๒. สร้างข้อสอบรูปแบบผสมภายใต้เงื่อนไข การสร้างพารามิเตอร์ข้อสอบที่สอดคล้องกับโมเดล การตอบสนองข้อสอบและข้อสอบที่ไม่สอดคล้อง กับโมเดลการตอบสนองข้อสอบที่กำหนดตามแบบ สอบรูปแบบผสม ในครั้งนี้ประกอบด้วย

๒.๑ การสร้างพารามิเตอร์ข้อสอบแบบ ตรวจให้คะแนน ๒ ค่า

๒.๑.๑ สร้างพารามิเตอร์ข้อสอบที่ สอดคล้องกับข้อสอบชนิด ๑ พารามิเตอร์ ตาม แนวทางของ Orlando & Thissen (2000) ที่ กำหนดให้พารามิเตอร์อำนาจจำแนก (a) มีค่าคงที่ เท่ากับ ๑ และพารามิเตอร์ความยาก (b) แจกแจงแบบ normal มีค่าเฉลี่ยเท่ากับ ๐ ส่วนเบี่ยงเบน มาตรฐานเท่ากับ ๑ ($b \sim N(0, 1)$)

๒.๑.๒ สร้างพารามิเตอร์ข้อสอบที่ ไม่สอดคล้องกับโมเดลการตอบสนองข้อสอบ ที่มี โมเดลการตอบสนองข้อสอบชนิด ๔ พารามิเตอร์ เพื่อหาคุณลักษณะของการตอบสนองข้อสอบ ที่ไม่ใช่ด้านความรู้ความจำ โดยกำหนดให้ พารามิเตอร์อำนาจจำแนก (a) มีค่าเฉลี่ยเท่ากับ ๒ ส่วนเบี่ยงเบนมาตรฐานเท่ากับ ๒.๕ พารามิเตอร์ ความยาก (b) แจกแจงแบบ normal มีค่า เฉลี่ยเท่ากับ ๐ ส่วนเบี่ยงเบนมาตรฐานเท่ากับ ๑ พารามิเตอร์การเดาข้อสอบได้ถูก (c) มีค่าเฉลี่ย เท่ากับ ๐ ส่วนเบี่ยงเบนมาตรฐานเท่ากับ ๐.๒๕ และพารามิเตอร์ความรอบคอบหรือพารามิเตอร์ ขีดจำกัดบน (d) มีค่าเฉลี่ยเท่ากับ ๐.๗ ส่วน เบี่ยงเบนมาตรฐานเท่ากับ ๐

๒.๒ การสร้างพารามิเตอร์ข้อสอบแบบ ตรวจให้คะแนนมากกว่า ๒ ค่า

๒.๒.๑ สร้างพารามิเตอร์ข้อสอบที่ สอดคล้องกับโมเดลการตอบสนองข้อสอบที่มี โมเดลการตอบสนองข้อสอบแบบการตรวจให้ คะแนนความรู้บางส่วน (PCM) ตามแนวทางของ Chon, Lee & Dunbar (2010) ที่กำหนดให้ ค่าพารามิเตอร์ความชัน (a) แจกแจงแบบ log normal มีค่าเฉลี่ยเท่ากับ ๐ ส่วนเบี่ยงเบน มาตรฐานเท่ากับ ๑ ($a_j \sim LN(0,1)$) ค่าพารามิเตอร์ ระดับความยาก β_1 แจกแจงแบบ normal มีค่า เฉลี่ยเท่ากับ - ๑ ส่วนเบี่ยงเบนมาตรฐานเท่ากับ ๐.๕ ($d_{j1} \sim N(-1, 0.5)$) ค่าพารามิเตอร์ระดับความ ยาก β_2 แจกแจงแบบ normal มีค่าเฉลี่ยเท่ากับ ๐ ส่วนเบี่ยงเบนมาตรฐานเท่ากับ ๐.๕ ($d_{j2} \sim N(0, 0.5)$) และค่าพารามิเตอร์ระดับความยาก β_3 แจกแจงแบบ normal มีค่าเฉลี่ยเท่ากับ ๑ ส่วน เบี่ยงเบนมาตรฐานเท่ากับ ๐.๕ ($d_{j3} \sim N(1, 0.5)$) และพารามิเตอร์ความสามารถของผู้สอบแจกแจง แบบ normal มีค่าเฉลี่ยเท่ากับ ๐ ส่วนเบี่ยงเบน มาตรฐานเท่ากับ ๑ ($\theta \sim N(0, 1)$)

๒.๒.๒ สร้างพารามิเตอร์ข้อสอบที่ ไม่สอดคล้องกับโมเดลการตอบสนองข้อสอบ ที่ มีโมเดลการตอบสนองข้อสอบแบบการตรวจให้ คะแนนความรู้บางส่วนทั่วไป (GPCM) มีค่า พารามิเตอร์ความชัน (a) แจกแจงแบบ log normal มีค่าเฉลี่ยเท่ากับ ๐ ส่วนเบี่ยงเบน มาตรฐานเท่ากับ ๐.๒ ($a_j \sim LN(0,0.2)$) ค่า พารามิเตอร์ระดับความยาก β_1 แจกแจงแบบ normal มีค่าเฉลี่ยเท่ากับ - ๑ ส่วนเบี่ยงเบน มาตรฐานเท่ากับ ๐.๕ ($d_{j1} \sim N(-1, 0.5)$) ค่า พารามิเตอร์ระดับความยาก β_2 แจกแจงแบบ normal มีค่าเฉลี่ยเท่ากับ ๐ ส่วนเบี่ยงเบน

การเปรียบเทียบผลการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบสำหรับแบบสอบรูปแบบผสม :
วิธีการประมาณค่าภาวะน่าจะเป็นสูงสุดชายขอบและวิธีการปรับเรียบเคอร์เนล

มาตรฐานเท่ากับ ๐.๕ ($d_{j2} \sim N(0, 0.5)$) และค่าพารามิเตอร์ระดับความยาก β_3 แจกแจงแบบ normal มีค่าเฉลี่ยเท่ากับ ๑ ส่วนเบี่ยงเบนมาตรฐานเท่ากับ ๐.๕ ($d_{j3} \sim N(1, 0.5)$)

ขั้นตอนการวิเคราะห์ข้อมูล

๓. การประมาณค่าพารามิเตอร์ความสามารถของผู้สอบสำหรับแบบสอบรูปแบบผสมในครั้งนี้ ผู้วิจัยดำเนินการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบด้วยวิธี MML และวิธี KS ภายใต้เงื่อนไขข้อมูลแบบสอบรูปแบบผสม ๒๗ เงื่อนไข

๔. พิจารณาประสิทธิภาพของวิธี MML และวิธี KS โดยพิจารณาที่ดัชนี AMSE จากการทำซ้ำ ๕๐๐ ครั้ง โดยเปรียบเทียบค่าที่ได้ ถ้าเงื่อนไขที่ศึกษาใดมีค่า AMSE น้อยกว่าแสดงว่าวิธีประมาณค่าความสามารถของผู้สอบวิธีนั้นสามารถประมาณค่าความสามารถของผู้สอบได้ดีกว่า หรือมีความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบน้อย

๕. สรุปผลการวิเคราะห์ข้อมูลตามวัตถุประสงค์ของการวิจัย โดยเปรียบเทียบประสิทธิภาพของผลการประมาณค่าความสามารถของผู้สอบจากสถิติทั้งสองในแต่ละเงื่อนไขที่ทำการศึกษาจากค่า AMSE โดยใช้สถิติการทดสอบที่ (Independent sample t-test)

๖. ศึกษาสัดส่วนของขนาดตัวอย่าง ความยาวของแบบสอบ และเปอร์เซ็นต์ของข้อสอบที่ไม่สอดคล้องกับโมเดลการตอบสนองข้อสอบของสถิติทั้งสองด้วยการวิเคราะห์ความแปรปรวนสองทาง

ผลการวิจัย

ผลการวิจัยสามารถสรุปตามวัตถุประสงค์การวิจัยดังนี้

๑) ผลการเปรียบเทียบการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบสำหรับแบบสอบรูปแบบผสมรูปแบบ 1PL/PCM ระหว่างวิธี MML และวิธี KS โดยพิจารณาที่ค่า AMSE ดำเนินการเปรียบเทียบทั้งหมด ๒๗ เงื่อนไขพบว่า

แบบสอบที่มีเปอร์เซ็นต์ของข้อสอบที่ไม่สอดคล้องกับโมเดลการตอบสนองข้อสอบ ๑๐ เปอร์เซ็นต์ (10% misfit) นั้น สำหรับข้อสอบแบบตรวจให้คะแนน ๒ ค่า ที่มีโมเดลการตอบสนองข้อสอบชนิด ๑ พารามิเตอร์ การประมาณค่าความสามารถของผู้สอบด้วยวิธี MML มีค่า AMSE น้อยกว่าวิธี KS ทุกเงื่อนไข ส่วนโมเดลการตอบสนองข้อสอบชนิด ๔ พารามิเตอร์ การประมาณค่าความสามารถของผู้สอบด้วยวิธี MML มีค่า AMSE มากกว่าวิธี KS ทุกเงื่อนไข ส่วนเงื่อนไขข้อสอบแบบตรวจให้คะแนนมากกว่า 2 ค่า ที่มีโมเดลการตอบสนองข้อสอบแบบการตรวจให้คะแนนความรู้บางส่วน การประมาณค่าความสามารถของผู้สอบด้วยวิธี MML มีค่า AMSE น้อยกว่าวิธีการปรับเรียบ KS ทุกเงื่อนไข เมื่อพิจารณาโมเดลการตอบสนองข้อสอบแบบการตรวจให้คะแนนความรู้บางส่วนทั่วไป การประมาณค่าความสามารถของผู้สอบด้วยวิธี MML มีค่า AMSE น้อยกว่าวิธี KS ยกเว้นกรณีเงื่อนไขตัวอย่าง ๓๐๐ คน กับความยาวของแบบสอบ ๔๐ ข้อเท่านั้นที่วิธี MML มีค่า AMSE มากกว่าวิธี KS

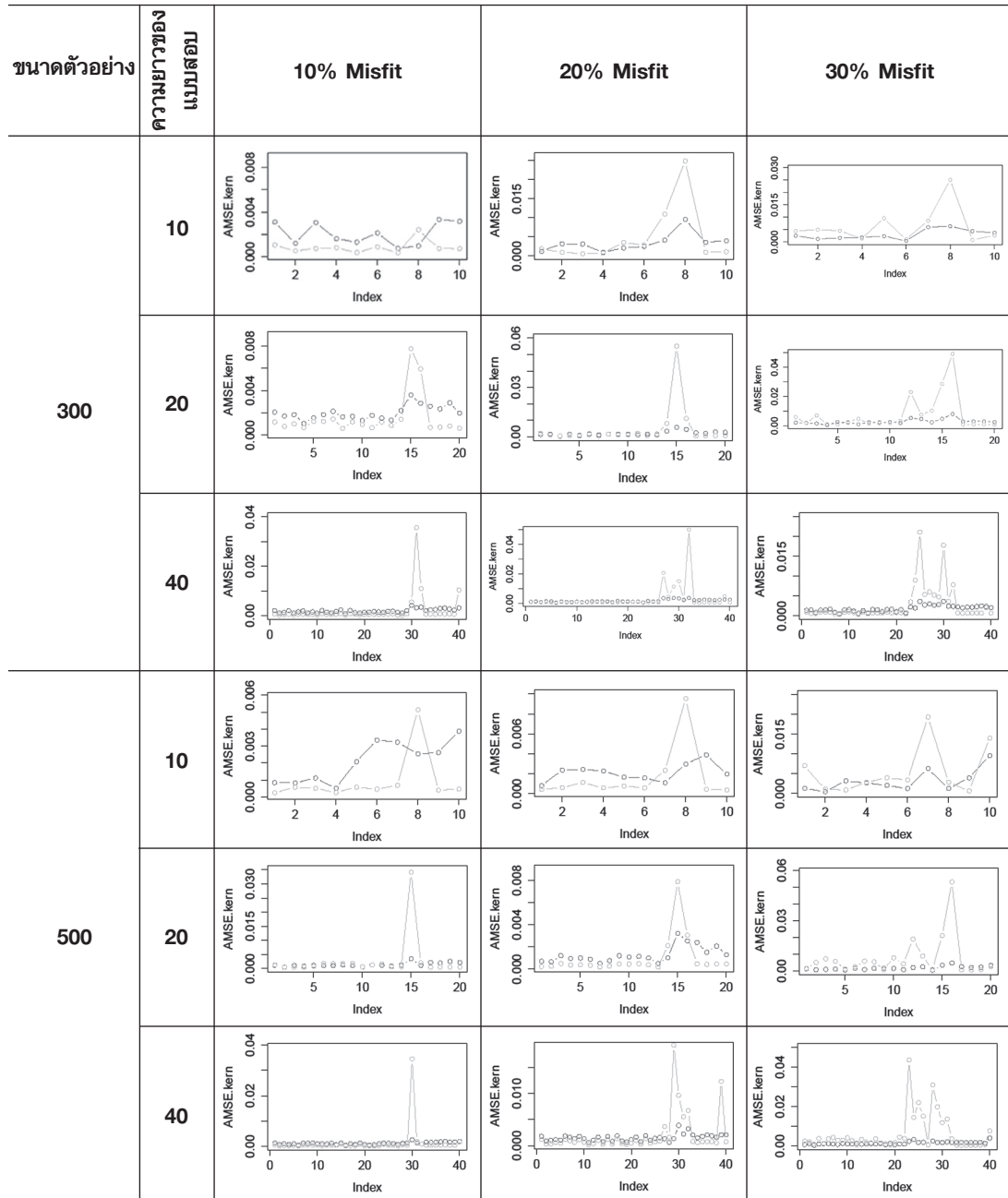
แบบสอปที่มีเปอร์เซ็นต์ของข้อสอปที่ไม่ สอดคล้องกับโมเดลการตอบสนองข้อสอป ๒๐ เปอร์เซ็นต์ (20% misfit) นั้น สำหรับข้อสอปแบบ ตรวจให้คะแนน ๒ ค่า ที่มีโมเดลการตอบสนอง ข้อสอปชนิด ๑ พารามิเตอร์ การประมาณค่าความ สามารถของผู้สอปด้วยวิธี MML มีค่า AMSE น้อยกว่าวิธี KS ทุกเงื่อนไข ส่วนโมเดลการตอบ สอนของข้อสอปชนิด ๔ พารามิเตอร์ การประมาณ ค่าความสามารถของผู้สอปด้วยวิธี MML มีค่า AMSE มากกว่าวิธี KS ทุกเงื่อนไข ส่วนเงื่อนไข ข้อสอปแบบตรวจให้คะแนนมากกว่า ๒ ค่า ที่มี โมเดลการตอบสนองข้อสอปแบบการตรวจให้ คะแนนความรู้บางส่วน การประมาณค่าความ สามารถของผู้สอปด้วยวิธี MML มีค่า AMSE น้อยกว่าวิธี KS ทุกเงื่อนไข เมื่อพิจารณาโมเดล การตอบสนองข้อสอปแบบการตรวจให้คะแนน ความรู้บางส่วนทั่วไป การประมาณค่าความ สามารถของผู้สอปด้วยวิธี MML มีค่า AMSE น้อยกว่าวิธี KS ยกเว้นกรณีเงื่อนไขขนาดตัวอย่าง ๕๐๐ คน กับความยาวของแบบสอป ๔๐ ข้อเท่านั้น ที่ข้อสอปบางข้อประมาณค่าความสามารถของ ผู้สอปด้วยวิธี MML มีค่า AMSE มากกว่าวิธี KS

แบบสอปที่มีเปอร์เซ็นต์ของข้อสอปที่ไม่ สอดคล้องกับโมเดลการตอบสนองข้อสอป ๓๐ เปอร์เซ็นต์ (30% misfit) นั้น สำหรับข้อสอปแบบ ตรวจให้คะแนน ๒ ค่า ที่มีโมเดลการตอบสนอง ข้อสอปชนิด ๑ พารามิเตอร์ ภายใต้เงื่อนไขขนาด ตัวอย่าง ๓๐๐ คน กับความยาวของแบบสอป ๑๐ และ ๒๐ ข้อ และขนาดตัวอย่าง ๕๐๐ คน กับ ความยาวของแบบสอป ๑๐ ข้อ การประมาณค่า ความสามารถของผู้สอปด้วยวิธี MML มีค่า

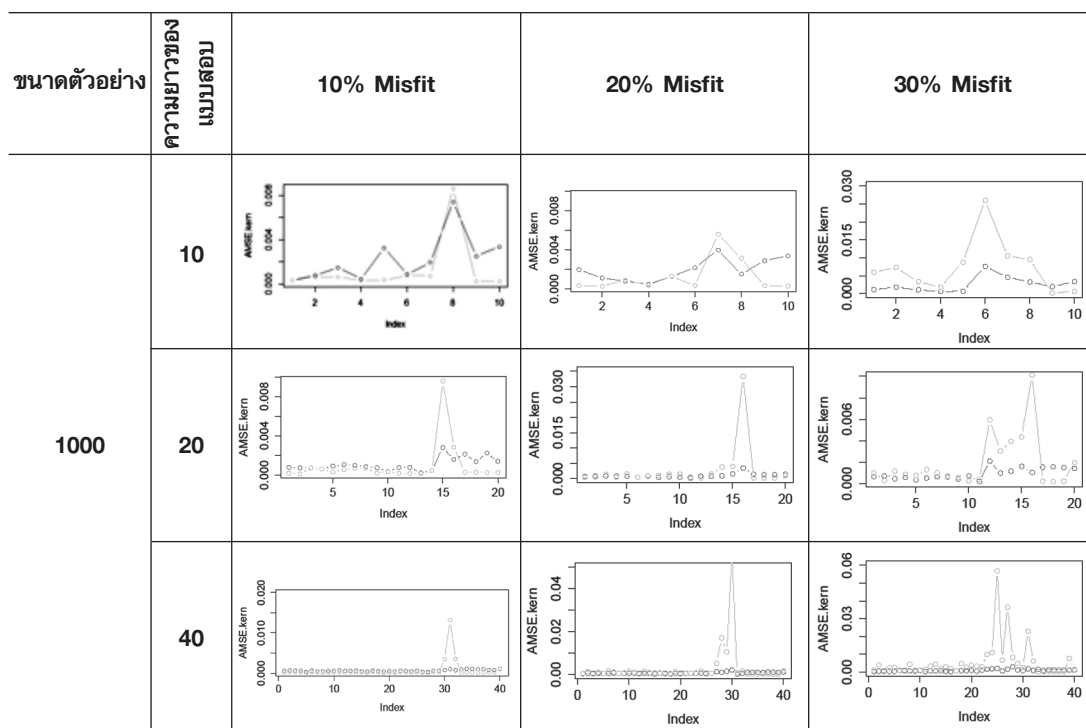
AMSE น้อยกว่าวิธี KS บางข้อ เงื่อนไขขนาด ตัวอย่าง ๓๐๐ คน กับความยาวของแบบสอป ๔๐ ข้อ การประมาณค่าความสามารถของผู้สอป ด้วยวิธี MML มีค่า AMSE น้อยกว่าวิธี KS ทุก ข้อ เงื่อนไขขนาดตัวอย่าง ๕๐๐ คน กับความยาว ของแบบสอป ๒๐ และ ๔๐ ข้อ และขนาดตัวอย่าง ๕๐๐ คน กับความยาวของแบบสอป ๑๐ ๒๐ และ ๔๐ ข้อ การประมาณค่าความสามารถของผู้สอป ด้วยวิธี MML มีค่า AMSE มากกว่าวิธี KS ทุกเงื่อนไข ส่วนโมเดลการตอบสนองข้อสอปชนิด ๔ พารามิเตอร์ การประมาณค่าความสามารถ ของผู้สอปด้วยวิธี MML มีค่า AMSE มากกว่า วิธี KS ทุกเงื่อนไข ส่วนเงื่อนไขข้อสอปแบบ ตรวจให้คะแนนมากกว่า ๒ ค่า ที่มีโมเดลการ ตอบสนองข้อสอปแบบการตรวจให้คะแนน ความรู้บางส่วน การประมาณค่าความสามารถ ของผู้สอปด้วยวิธี MML มีค่า AMSE น้อยกว่า วิธี KS ทุกเงื่อนไข เมื่อพิจารณาโมเดลการตอบ สอนของข้อสอปแบบการตรวจให้คะแนนความรู้ บางส่วนทั่วไป การประมาณค่าความสามารถ ของผู้สอปด้วยวิธี MML มีค่า AMSE น้อยกว่า วิธี KS ยกเว้นกรณีเงื่อนไขขนาดตัวอย่าง ๕๐๐ คน กับความยาวของแบบสอป ๑๐ และ ๔๐ ข้อ และขนาดตัวอย่าง ๑,๐๐๐ คน กับความยาวของ แบบสอป ๔๐ ข้อ การประมาณค่าความสามารถ ของผู้สอปด้วยวิธี MML มีค่า AMSE มากกว่าวิธี KS โดยมีรายละเอียดการเปรียบเทียบค่า AMSE ทั้งกรณีแบบสอปที่มีเปอร์เซ็นต์ของข้อสอปที่ไม่ สอดคล้องกับโมเดลการตอบสนองข้อสอป ๑๐ - ๓๐ เปอร์เซ็นต์ ของทั้งสองวิธีดังแสดงตาม แผนภาพที่ ๑

การเปรียบเทียบผลการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบสำหรับแบบสอบรูปแบบผสม :
 วิธีการประมาณค่าภาวะจะเป็นสูงสุดชายขอบและวิธีการปรับเรียบเคอร์เนล

แผนภาพที่ ๑ เปรียบเทียบค่า AMSE ระหว่างวิธี MML และวิธี KS ภายใต้ ๒๗ เงื่อนไขของแบบสอบรูปแบบผสมรูปแบบ 1PL/PCM



แผนภาพที่ ๑ เปรียบเทียบค่า AMSE ระหว่างวิธี MML และวิธี KS ภายใต้ ๒๗ เงื่อนไขของแบบสอบบรูปแบบผสมรูปแบบ 1PL/PCM (ต่อ)



—— วิธีการประมาณค่าภาวะน่าจะเป็นสูงสุดชายขอบ (MML)

—— วิธีการปรับเรียบเคอร์เนล (KS)

ผลการวิจัยพบว่า วิธีการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบทั้งสองวิธีดังกล่าวมีความแตกต่างกัน ดังนั้นผู้วิจัยจึงได้นำผลการวิเคราะห์ค่า AMSE ของแต่ละวิธีมาวิเคราะห์เปรียบเทียบประสิทธิภาพของการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบของทั้งสองวิธีในแต่ละเงื่อนไขที่ทำการศึกษา โดยใช้สถิติการทดสอบที (Independent sample t-test) ผลการวิเคราะห์พบว่า การประมาณค่าพารามิเตอร์ความสามารถของผู้สอบด้วยวิธี MML และวิธี KS ที่เงื่อนไขขนาดตัวอย่าง ๕๐๐ คน กับข้อสอบ ๒๐ ข้อและมีข้อสอบร่วม 30% misfit มีความ

แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .๐๕ และที่เงื่อนไขขนาดตัวอย่าง ๓๐๐ คน กับข้อสอบ ๑๐ ข้อและมีข้อสอบร่วม 10% misfit ขนาดตัวอย่าง ๕๐๐ คน กับข้อสอบ ๔๐ ข้อและมีข้อสอบร่วม 30% misfit และขนาดตัวอย่าง ๑,๐๐๐ คน กับข้อสอบ ๔๐ ข้อและมีข้อสอบร่วม 30% misfit มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .๐๑ แสดงว่าการวิเคราะห์คุณภาพของการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบจากค่า AMSE ของทั้งสองวิธีให้ผลที่แตกต่างกันในสี่เงื่อนไข โดยมีรายละเอียดในการวิเคราะห์ดังแสดงในตารางที่ ๑

การเปรียบเทียบผลการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบสำหรับแบบสอบรูปแบบผสม :
วิธีการประมาณค่าภาวะน่าจะเป็นสูงสุดชายขอบและวิธีการปรับเรียบเคอร์เนล

ตารางที่ ๑ ผลการวิเคราะห์ค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐานและการวิเคราะห์ความแตกต่างด้วยสถิติการทดสอบ
ที (t-test) ของค่า AMSE ระหว่างวิธี MML และวิธี KS จำแนกตามขนาดตัวอย่างความยาวของ
แบบสอบ และ % misfit

เปอร์เซ็นต์ข้อสอบ ที่ไม่สอดคล้อง	ขนาดตัวอย่าง	ความยาวของแบบสอบ	วิธี MML		วิธี KS		t-test	p-value
			M	SD	M	SD		
10	300	10	0.0009	0.0006	0.0021	0.0010	-3.237**	0.006
		20	0.0015	0.0019	0.0020	0.0006	-1.030	0.309
		40	0.0021	0.0059	0.0019	0.0008	0.166	0.869
	500	10	0.0009	0.0015	0.0021	0.0012	-1.932	0.069
		20	0.0027	0.0074	0.0013	0.0007	0.818	0.418
		40	0.0014	0.0054	0.0012	0.0005	0.198	0.843
	1,000	10	0.0013	0.0026	0.0022	0.0021	-0.908	0.376
		20	0.0010	0.0021	0.0011	0.0007	-0.268	0.791
		40	0.0007	0.0022	0.0007	0.0002	0.100	0.921
20	300	10	0.0048	0.0077	0.0034	0.0024	0.554	0.591
		20	0.0047	0.0122	0.0021	0.0013	0.972	0.343
		40	0.0034	0.0086	0.0019	0.0008	1.117	0.271
	500	10	0.0017	0.0028	0.0021	0.0009	-0.455	0.655
		20	0.0010	0.0018	0.0013	0.0007	-0.741	0.463
		40	0.0020	0.0038	0.0015	0.0007	0.868	0.391
	1,000	10	0.0013	0.0017	0.0020	0.0011	-1.013	0.325
		20	0.0028	0.0073	0.0010	0.0068	1.088	0.290
		40	0.0029	0.0092	0.0008	0.0004	1.470	0.150
30	300	10	0.0063	0.0072	0.0031	0.0020	1.372	0.187
		20	0.0079	0.0122	0.0030	0.0017	1.775	0.084
		40	0.0027	0.0045	0.0018	0.0008	1.214	0.232
	500	10	0.0056	0.0062	0.0032	0.0028	1.111	0.287
		20	0.0078	0.0122	0.0018	0.0011	2.189*	0.041
		40	0.0061	0.0092	0.0015	0.0007	3.164**	0.003
	1,000	10	0.0074	0.0075	0.0026	0.0022	1.938	0.068
		20	0.0019	0.0025	0.0009	0.0005	1.683	0.107
		40	0.0058	0.0107	0.0010	0.0005	2.803**	0.008

*p < .05, **p < .01

๒) ผลการวิเคราะห์สัดส่วนของขนาดตัวอย่าง ความยาวของแบบสอบถาม และเปอร์เซ็นต์ของข้อสอบที่ไม่สอดคล้องกับโมเดลการตอบสนองข้อสอบภายใต้เงื่อนไขการฝ่าฝืนข้อตกลงเบื้องต้นด้านการเลือกใช้โมเดลการตอบสนองข้อสอบ ในการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบระหว่างวิธี MML และวิธี KS ที่ส่งผลต่อค่าประมาณพารามิเตอร์ความสามารถของผู้สอบ มีผลการวิเคราะห์ ดังนี้

๒.๑ ผลการวิเคราะห์ข้อมูลเปรียบเทียบประสิทธิภาพของการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบด้วยวิธี MML ในแต่ละเงื่อนไขที่ทำการศึกษา ผู้วิจัยใช้การวิเคราะห์ความแปรปรวนสองทาง เพื่อทดสอบความแตกต่างของค่า AMSE จำแนกตามขนาดตัวอย่าง และความยาวของแบบสอบถาม เมื่อพิจารณาแบบสอบถามที่มี ๑๐ ๒๐ และ ๓๐ เปอร์เซนต์ของข้อสอบที่ไม่สอดคล้องกับโมเดลการตอบสนองข้อสอบรวมอยู่ด้วย พบว่าไม่มีปฏิสัมพันธ์ระหว่างขนาดตัวอย่างและความยาวของแบบสอบถามร่วมกันส่งผลต่อค่า AMSE ดังนั้นขนาดตัวอย่างและความยาวของแบบสอบถาม

ต่างแยกกันส่งผลต่อค่า AMSE

๒.๒ ผลการวิเคราะห์ข้อมูลเปรียบเทียบประสิทธิภาพของการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบด้วยวิธี KS ในแต่ละเงื่อนไขที่ทำการศึกษา ผู้วิจัยใช้การวิเคราะห์ความแปรปรวนสองทาง เพื่อทดสอบความแตกต่างของค่า AMSE จำแนกตามขนาดตัวอย่าง และความยาวของแบบสอบถาม เมื่อพิจารณาแบบสอบถามที่มี ๑๐ เปอร์เซนต์ของข้อสอบที่ไม่สอดคล้องกับโมเดลการตอบสนองข้อสอบรวมอยู่ด้วย พบว่า มีปฏิสัมพันธ์ระหว่างขนาดตัวอย่าง และความยาวของแบบสอบถามร่วมกันส่งผลต่อค่า AMSE อย่างมีนัยสำคัญทางสถิติที่ระดับ .๐๕ ($F=3.398, p=.01$) และแบบสอบถามที่มี ๒๐ และ ๓๐ เปอร์เซนต์ของข้อสอบที่ไม่สอดคล้องกับโมเดลการตอบสนองข้อสอบรวมอยู่ด้วย พบว่า ขนาดตัวอย่าง และความยาวของแบบสอบถาม มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .๐๑ ทั้งสองเงื่อนไข โดยมีรายละเอียดในการวิเคราะห์ดังแสดงในตารางที่ ๒

การเปรียบเทียบผลการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบสำหรับแบบสอบรูปแบบผสม :
วิธีการประมาณค่าถ่วงน้ำหนักจะเป็นสูงสุดขायชอบและวิธีการปรับเรียบเคอร์เนล

ตารางที่ ๒ ผลการวิเคราะห์ความแตกต่างของค่า AMSE ของวิธี KS ด้วยการวิเคราะห์ความแปรปรวนสองทาง
จำแนกตามจำนวนตัวอย่าง ความยาวของแบบสอบ และ % misfit

เปอร์เซ็นต์ของ ข้อสอบที่ไม่ สอดคล้องๆ	แหล่งความแปรปรวน	Levene's test (p-value)	df	MS	F	p-value
10	Corrected Model	11.760	8	.000007198	11.914	.000
	Intercept	(.000)	1	.000	676.181	.000
	จำนวนตัวอย่าง		2	.000005824	9.639**	.000
	ความยาวของแบบสอบ		2	.000008710	14.417**	.000
	จำนวนตัวอย่าง*ความยาวของ แบบสอบ		4	.000002053	3.398*	.010
	Error		201	.0000006042		
	Total		210			
	Corrected Total		209			
20	Corrected Model	5.734	8	.00001020	11.976	.000
	Intercept	(.000)	1	.000	575.523	.000
	จำนวนตัวอย่าง		2	.00002045	24.023**	.000
	ความยาวของแบบสอบ		2	.00001472	17.286**	.000
	จำนวนตัวอย่าง*ความยาวของ แบบสอบ		4	.000001024	1.203	.311
	Error		201	.0000008513		
	Total		210			
	Corrected Total		209			
30	Corrected Model	10.432	8	.00001401	9.812	.000
	Intercept	(.000)	1	.001	480.301	.000
	จำนวนตัวอย่าง		2	.00001569	10.987**	.000
	ความยาวของแบบสอบ		2	.00002809	19.667**	.000
	จำนวนตัวอย่าง*ความยาวของ แบบสอบ		4	.00000330	2.311	.059
	Error		201	.000001428		
	Total		210			
	Corrected Total		209			

*p < .05, **p < .01

จากตารางที่ ๒ ในการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบด้วยวิธี KS ที่สัดส่วน ๑๐ เปอร์เซนต์ของข้อสอบที่ไม่สอดคล้องกับโมเดลการตอบสนองข้อสอบรวมอยู่ด้วย พบว่ามีปฏิสัมพันธ์ระหว่างขนาดตัวอย่างและความยาวของแบบสอบร่วมกันส่งผลต่อค่า AMSE อย่างมีนัยสำคัญทางสถิติที่ระดับ .๐๕ ($F=3.398$, $p=.01$) ดังนั้นจึงไม่อาจกล่าวได้ว่าขนาดตัวอย่างใดส่งผลต่อค่า AMSE มากที่สุด หรือความยาวของแบบสอบใดส่งผลต่อค่า AMSE มากที่สุด ผู้วิจัยจึงดำเนินการวิเคราะห์ Sample Main Effect และกำหนดวิธีการเปรียบเทียบรายคู่ด้วยวิธี Bonferroni พบว่า

ผลการวิเคราะห์ความยาวของแบบสอบ ๑๐ ข้อ พบว่า กลุ่มผู้สอบที่ใช้ความยาวของแบบสอบ ๑๐ ข้อ จากขนาดตัวอย่างผู้สอบต่างกัน ๓ กลุ่ม ส่งผลต่อค่า AMSE ไม่แตกต่างกัน

ผลการวิเคราะห์ความยาวของแบบสอบ ๒๐ ข้อ พบว่า กลุ่มผู้สอบที่ใช้ความยาวของแบบสอบ ๒๐ ข้อ จากขนาดตัวอย่างผู้สอบต่างกัน ๓ กลุ่ม ส่งผลต่อค่า AMSE แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .๐๑ โดยความยาวของแบบสอบ ๒๐ ข้อ ที่ใช้กับขนาดตัวอย่างผู้สอบ ๓๐๐ คน ส่งผลต่อค่า AMSE สูงกว่าความยาวของแบบสอบ ๒๐ ข้อ ที่ใช้กับขนาดตัวอย่างผู้สอบ ๕๐๐ คน และ ๑,๐๐๐ คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .๐๑ ส่วนความยาวของแบบสอบ ๒๐ ข้อ ที่ใช้กับขนาดตัวอย่างผู้สอบ ๕๐๐ คน และ ๑,๐๐๐ คน ส่งผลต่อค่า AMSE ไม่แตกต่างกัน

ผลการวิเคราะห์ความยาวของแบบสอบ ๔๐ ข้อ พบว่า กลุ่มผู้สอบที่ใช้ความยาวของแบบสอบ ๔๐ ข้อ ที่มีขนาดตัวอย่างผู้สอบต่างกัน ๓ กลุ่ม ส่งผลต่อค่า AMSE แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .๐๑ โดยมีรายละเอียดในการวิเคราะห์ดังแสดงในตารางที่ ๓

ตารางที่ ๓ เปรียบเทียบรายคู่ค่า AMSE จากวิธี KS จำแนกตามความยาวของแบบสอบที่มีขนาดตัวอย่างผู้สอบต่างกัน ๓ กลุ่ม

ความยาวของแบบสอบ	ขนาดตัวอย่าง	ขนาดตัวอย่าง		
		300 คน	500 คน	1000 คน
10	300 คน		-.00003	-.00016
	500 คน		-	-.00013
	1000 คน			-
20	300 คน		.00065**	.00089**
	500 คน		-	.00024
	1000 คน			-
40	300 คน		.00071**	.00124**
	500 คน		-	.00053**
	1000 คน			-

**p < .01

การเปรียบเทียบผลการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบสำหรับแบบสอบรูปแบบผสม :
วิธีการประมาณค่าภาวะนั้นจะเป็นสูงสุดขายขอบและวิธีการปรับเรียบเคอร์เนล

เมื่อพิจารณาผลการวิเคราะห์ขนาดตัวอย่างผู้สอบ ๓๐๐ คน พบว่า ขนาดตัวอย่างผู้สอบ ๓๐๐ คน ที่ใช้ความยาวของแบบสอบต่างกัน ๓ กลุ่ม ส่งผลต่อค่า AMSE ไม่แตกต่างกัน

ผลการวิเคราะห์ขนาดตัวอย่างผู้สอบ ๕๐๐ คน พบว่า ขนาดตัวอย่างผู้สอบ ๕๐๐ คน ที่ใช้ความยาวของแบบสอบต่างกัน ๓ กลุ่ม ส่งผลต่อค่า AMSE แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .๐๑ โดยขนาดตัวอย่างผู้สอบ ๕๐๐ คน ที่ใช้ความยาวของแบบสอบ ๑๐ ข้อ ส่งผลต่อค่า AMSE สูงกว่าขนาดตัวอย่างผู้สอบ ๕๐๐ คน ที่ใช้ความยาวของแบบสอบ ๒๐ และ ๔๐ ข้อ อย่างมีนัยสำคัญทางสถิติที่ระดับ .๐๕ ส่วนขนาดตัวอย่างผู้สอบ ๕๐๐ คน ที่ใช้ความยาวของแบบสอบ ๒๐ และ ๔๐ ข้อ ส่งผลต่อค่า AMSE ไม่แตกต่างกัน

ผลการวิเคราะห์ขนาดตัวอย่างผู้สอบ ๑,๐๐๐ คน พบว่า ขนาดตัวอย่างผู้สอบ ๑,๐๐๐ คน ที่ใช้ความยาวของแบบสอบต่างกัน ๓ กลุ่ม ส่งผลต่อค่า AMSE แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .๐๑ โดยขนาดตัวอย่างผู้สอบ ๑,๐๐๐ คน ที่ใช้ความยาวของแบบสอบ ๑๐ ข้อ ส่งผลต่อค่า AMSE สูงกว่าขนาดตัวอย่างผู้สอบ ๑,๐๐๐ คน ที่ใช้ความยาวของแบบสอบ ๒๐ และ ๔๐ ข้อ อย่างมีนัยสำคัญทางสถิติที่ระดับ .๐๑ ส่วนขนาดตัวอย่างผู้สอบ ๑,๐๐๐ คน ที่ใช้ความยาวของแบบสอบ ๒๐ และ ๔๐ ข้อ ส่งผลต่อค่า AMSE ไม่แตกต่างกัน โดยมีรายละเอียดในการวิเคราะห์ดังแสดงในตารางที่ ๔

ตารางที่ ๔ เปรียบเทียบรายคู่ค่า AMSE จากวิธี KS จำแนกตามขนาดตัวอย่างผู้สอบที่ใช้ความยาวของแบบสอบต่างกัน ๓ กลุ่ม

ขนาดตัวอย่าง	ความยาวของแบบสอบ	ความยาวของแบบสอบ		
		10 ข้อ	20 ข้อ	40 ข้อ
300	10 ข้อ		.00007	.00013
	20 ข้อ		-	.00005
	40 ข้อ			-
500	10 ข้อ		.00076*	.0087**
	20 ข้อ		-	.00010
	40 ข้อ			-
1000	10 ข้อ		.00114*	.00154**
	20 ข้อ		-	.00040
	40 ข้อ			-

*p < .05, **p < .01

การอภิปรายผล

จากผลการวิจัยมีประเด็นหลักๆ ที่ควรนำมาอภิปรายดังนี้

๑) ผลการเปรียบเทียบการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบสำหรับแบบสอบรูปแบบผสม ระหว่างวิธี MML และวิธี KS ที่พิจารณาที่ค่า AMSE ดำเนินการเปรียบเทียบทั้งหมด ๒๗ เงื่อนไข พบว่า ข้อสอบที่สอดคล้องกับโมเดลการตอบสนองข้อสอบ ที่ประมาณค่าพารามิเตอร์ความสามารถของผู้สอบโดยรวม ด้วยวิธี MML มีค่า AMSE น้อยกว่าวิธี KS แสดงว่าวิธี MML สามารถประมาณค่าพารามิเตอร์ความสามารถของผู้สอบได้ดีกว่าวิธี KS ส่วนข้อสอบที่ไม่สอดคล้องกับโมเดลการตอบสนองข้อสอบ ภายใต้เงื่อนไขการฝ่าฝืนข้อตกลงเบื้องต้นด้านการเลือกใช้โมเดลการตอบสนองข้อสอบ การประมาณค่าพารามิเตอร์ความสามารถของผู้สอบโดยรวม ด้วยวิธี MML มีค่า AMSE มากกว่าวิธี KS แสดงว่าวิธี KS สามารถประมาณค่าความสามารถของผู้สอบได้ดีกว่าวิธี MML สอดคล้องกับงานวิจัยของ (Douglas & Cohen, 2001; Lee Wollack & Douglas, 2009; Orlando & Thissen, 2000 และ Sueiro & Abad, 2011) สำหรับข้อสอบแบบตรวจให้คะแนน ๒ ค่า และสอดคล้องกับงานวิจัยของ (Liang & Wells, 2009; Van der Ark, 2001) สำหรับข้อสอบแบบตรวจให้คะแนนมากกว่า ๒ ค่า ซึ่งการประเมินความสอดคล้องของโมเดลการตอบสนองข้อสอบแบบพาราเมตริก นั้น เป็นแนวทางหนึ่งที่ใช้จำเป็นต้องเลือกรูปแบบการตอบสนองข้อสอบที่เหมาะสมกับลักษณะของข้อมูลเป็นสำคัญ และต้องเลือกวิธีการประมาณค่าพารามิเตอร์ที่มีประสิทธิภาพเหมาะสมกับสภาพการทดสอบของแต่ละประเภท เพื่อให้

ค่าประมาณพารามิเตอร์ความสามารถของผู้สอบใกล้เคียงกับความสามารถที่แท้จริงของผู้สอบ ในการพิจารณาวิธีการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบข้ามโมเดลการตอบสนองข้อสอบทั้งรูปแบบการตอบสนองข้อสอบแบบพาราเมตริก และรูปแบบการตอบสนองข้อสอบแบบนัยพาราเมตริก เป็นแนวทางหนึ่งในการประเมินความสอดคล้องของข้อสอบกับโมเดลการตอบสนองข้อสอบแบบพาราเมตริก (PIRT) หากการประเมินความสอดคล้องของโมเดลการตอบสนองข้อสอบมีความเหมาะสมก็จะให้ข้อมูลสารสนเทศอย่างมากเกี่ยวกับประสิทธิภาพการทำงานของข้อสอบและผู้สอบ ซึ่งผลที่ได้จะเป็นประโยชน์ต่อการแสดงให้เห็นถึงความสัมพันธ์ที่แท้จริงระหว่างความสามารถของผู้สอบและความน่าจะเป็นของการตอบสนองข้อสอบแต่ละประเภท ซึ่งมีความสำคัญมากเนื่องจากในการประยุกต์ใช้งานทางด้านการวัดผลการศึกษาประเภทต่างๆ เช่น การปรับเหมาะของแบบสอบ การทำหน้าที่ต่างกันของแบบสอบ การทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ ต้องพึ่งพาโมเดลการตอบสนองข้อสอบที่สอดคล้องกับโมเดลการตอบสนองข้อสอบแบบพาราเมตริก (PIRT) เป็นสำคัญ

๒) ผลการศึกษาสัดส่วนของขนาดตัวอย่างความยาวของแบบสอบ และเปอร์เซ็นต์ของข้อสอบที่ไม่สอดคล้องกับโมเดลการตอบสนองข้อสอบ ภายใต้เงื่อนไขการฝ่าฝืนข้อตกลงเบื้องต้นด้านการเลือกใช้โมเดลการตอบสนองข้อสอบ ในการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบระหว่างวิธี MML และวิธี KS ที่ส่งผลต่อค่าประมาณพารามิเตอร์ความสามารถของผู้สอบพบว่า วิธีการประมาณค่าพารามิเตอร์ความ

ความสามารถของผู้สอบด้วยวิธี KS ที่สัดส่วน ๑๐ เปอร์เซ็นต์ของข้อสอบที่ไม่สอดคล้องกับโมเดล การตอบสนองข้อสอบรวมอยู่ด้วย พบว่ามี ปฏิสัมพันธ์ระหว่างขนาดตัวอย่าง และความยาว ของแบบสอบร่วมกันส่งผลต่อค่า AMSE อย่างมี นัยสำคัญทางสถิติที่ระดับ .๐๕ สอดคล้องกับผล การวิจัยของ Sueiro & Abad (2011) ที่กล่าวว่า ขนาดตัวอย่างเป็นปัจจัยที่มีอิทธิพลหลักในความ เหมาะสมของผลการประมาณค่าพารามิเตอร์ ความสามารถของผู้สอบ เนื่องจากตัวอย่างขนาด ใหญ่ย่อมให้ความแม่นยำมากขึ้นในการประมาณ ค่าพารามิเตอร์ความสามารถของผู้สอบ และมีความ สอดคล้องของโมเดลการตอบสนองข้อสอบ แบบพาราเมตริก ส่วนความยาวของแบบสอบ เป็นปัจจัยที่สองที่สำคัญที่สุด ขณะที่ลดความยาว ของแบบสอบ อำนาจในการทดสอบจะลดลง เนื่องจากความสั้นของแบบสอบจะทำให้การ ประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ โดยเฉพาะวิธี KS ลดลง ดังนั้นการเลือกพิจารณา สัดส่วนของขนาดตัวอย่าง ความยาวของแบบสอบ และเปอร์เซ็นต์ของข้อสอบที่ไม่สอดคล้องกับ โมเดลการตอบสนองข้อสอบ จึงเป็นปัจจัยหนึ่ง ที่สำคัญในการนำมาใช้พิจารณาร่วมในการเลือก ใช้สถิติในการประมาณค่าพารามิเตอร์ความ สามารถของผู้สอบเพื่อให้ผลการประมาณค่า พารามิเตอร์ความสามารถของผู้สอบมีความใกล้เคียง ความสามารถที่แท้จริงของผู้สอบมากที่สุด

ข้อเสนอแนะ

ข้อเสนอแนะในการนำผลการวิจัยไปใช้

๑) จากผลการวิจัย พบว่า วิธีการประมาณ ค่าพารามิเตอร์ความสามารถของผู้สอบที่สามารถ ประมาณค่าความสามารถของผู้สอบได้ใกล้เคียง

ความสามารถที่แท้จริงได้มากที่สุด ขึ้นอยู่กับเงื่อนไข ของความสอดคล้องของโมเดลการตอบสนอง ข้อสอบ ตามสถานการณ์การทดสอบ ซึ่งใน สถานการณ์การทดสอบที่เกิดขึ้นจริงข้อมูลการ สอบที่นำมาใช้ในการประมาณค่าอาจไม่สอดคล้อง กับโมเดลการตอบสนองข้อสอบที่ต้องการหรืออาจ มีการฝ่าฝืนข้อตกลงเบื้องต้นตามทฤษฎีการตอบ สอนข้อสอบแบบพาราเมตริก (PIRT) ดังนั้นผล การวิจัยจึงเป็นแนวทางในการพิจารณาเลือกใช้วิธี การประมาณค่าพารามิเตอร์ความสามารถของ ผู้สอบให้สอดคล้องกับสถานการณ์การทดสอบ ให้มากที่สุด

๒) การศึกษานี้เป็นการศึกษาจากการจำลอง ข้อมูลเท่านั้น ดังนั้นพารามิเตอร์ข้อสอบจึงเป็นไป ตามทฤษฎี สอดคล้องและไม่สอดคล้องกับโมเดล การตอบสนองข้อสอบทุกประการ ดังนั้นผลการ ประมาณค่าความสามารถของผู้สอบที่เกิดขึ้นจึง เป็นผลในเชิงทฤษฎี หากมีการนำไปประยุกต์ใช้ใน สถานการณ์จริง ควรต้องพิจารณาค่าพารามิเตอร์ ของข้อสอบ และมีการตรวจสอบข้อมูลว่าเป็นไป ตามข้อตกลงเบื้องต้นตามทฤษฎีการตอบสนอง ข้อสอบแบบพาราเมตริกหรือไม่ แล้วจึงพิจารณา วิธีการประมาณค่าที่เหมาะสมกับสถานการณ์ของ การทดสอบนั้นๆ

๓) การเลือกวิธีประมาณค่าความสามารถ ของผู้สอบ ควรพิจารณาสัดส่วนของขนาดตัวอย่าง ผู้สอบและความยาวของแบบสอบด้วย เนื่องจาก ข้อค้นพบครั้งนี้แสดงให้เห็นว่าขนาดตัวอย่างที่ ใหญ่ย่อมให้ความแม่นยำมากขึ้นในการประมาณ ค่าพารามิเตอร์ความสามารถของผู้สอบ และความ สอดคล้องของโมเดลการตอบสนองข้อสอบแบบ พาราเมตริก และการลดความยาวของแบบสอบ อำนาจในการทดสอบจะลดลงด้วยเช่นกัน ดังนั้น

ในการประมาณค่าความสามารถของผู้สอบควรพิจารณาคุณสมบัติของวิธีการประมาณค่าความสามารถของผู้สอบวิธีนั้นกับขนาดตัวอย่างและจำนวนข้อสอบร่วมกันด้วย จะช่วยให้การประมาณค่าความสามารถของผู้สอบมีประสิทธิภาพมากขึ้น

ข้อเสนอแนะในการศึกษาวิจัยครั้งต่อไป

๑) จากการศึกษาในครั้งนี้เป็นการศึกษาเปรียบเทียบการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบสำหรับแบบสอบรูปแบบผสมระหว่างวิธี MML และวิธี KS ร่วมกับการศึกษาสัดส่วนของขนาดตัวอย่าง ความยาวของแบบสอบ และเปอร์เซ็นต์ของข้อสอบที่ไม่สอดคล้องกับโมเดลการตอบสนองข้อสอบภายใต้เงื่อนไขการฝ่าฝืนข้อตกลงเบื้องต้นด้านการเลือกใช้โมเดลการตอบสนองข้อสอบ พบว่า ข้อสอบที่สอดคล้องกับโมเดลการตอบสนองข้อสอบโดยรวม วิธี MML สามารถประมาณค่าความสามารถของผู้สอบได้ดีกว่าวิธี KS สำหรับข้อสอบที่ไม่สอดคล้องกับโมเดลการตอบสนองข้อสอบโดยรวม วิธี KS สามารถประมาณค่าความสามารถของผู้สอบได้ดีกว่าวิธี MML ดังนั้นเพื่อให้การศึกษาค้นคว้าการประเมินความสอดคล้องของข้อสอบกับโมเดลการตอบสนองข้อสอบแบบพาราเมตริก (PIRT) กว้างขวาง และมีคุณค่ามากยิ่งขึ้น

การศึกษาในอนาคตจึงควรศึกษาเพิ่มเติม เพื่อเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าความสามารถของผู้สอบวิธีอื่นด้วย ภายใต้เงื่อนไขการฝ่าฝืนข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบแบบพาราเมตริก (PIRT) เช่น ด้านความเป็นเอกมิติ ด้านความเป็นอิสระระหว่างข้อสอบ เป็นต้น และควรพิจารณาสัดส่วนเปอร์เซ็นต์ของข้อสอบที่ไม่สอดคล้องกับโมเดลการตอบสนองข้อสอบมากกว่าร้อยละ ๓๐ เพื่อเปรียบเทียบประสิทธิภาพของการทำนายความสามารถของผู้สอบเพื่อหาตัวแทนทั้งในด้านสถิติและความสอดคล้องของข้อสอบด้วย ว่าตัวแปรทั้งสองจะสามารถร่วมกันส่งผลต่อประสิทธิภาพการประมาณค่าความสามารถของผู้สอบหรือไม่อย่างไร

๒) การศึกษาในครั้งนี้เป็นการศึกษาในเชิงทฤษฎี เพื่อบุกเบิกองค์ความรู้ใหม่ โดยไม่มีการศึกษาจากข้อมูลจริงแต่อย่างใด ดังนั้น หากการศึกษาค้นคว้าในอนาคตควรศึกษาเพิ่มเติมในส่วนของการประเมินความสอดคล้องข้อสอบกับโมเดลการตอบสนองข้อสอบแบบพาราเมตริก (PIRT) ด้วยข้อมูลจริงเปรียบเทียบกับการศึกษาในเชิงทฤษฎีจะเป็นการศึกษาที่มีประโยชน์และมีคุณค่าในทางปฏิบัติมากยิ่งขึ้น

เอกสารอ้างอิง

ภาษาไทย

ศิริชัย กาญจนวาสี. (๒๕๕๐). *ทฤษฎีการทดสอบแนวใหม่*. (พิมพ์ครั้งที่ 3). กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ภาษาอังกฤษ

- Chon, K. H., Lee, W. C. & Dunbar, S. (2010). A comparison of item fit statistics for mixed irt models. *Journal of Educational Measurement*, 47, 318 - 338.
- Douglas, J. & Cohen, A. (2001). Nonparametric item response function estimation for assessing parametric model fit. *Applied Psychological Measurement*, 25, 234 - 243.
- Hambleton, R. K. (1996). *advances in assessment models, methods, and practices*. Berliner D. C. & Calfee R. C. P. (Ed.), *Handbook of Educational Psychology*, 899 - 925. New York: Simon & Schuster Maemillan.
- Lee, Y.S. (2007). A comparison of methods for nonparametric estimation of item characteristic curves for binary items. *Applied Psychological Measurement*, 31, 121 - 134.
- Lee, Y. S., Wollack, J. A. & Douglas, J. (2009). On the use of nonparametric item characteristic curve estimation techniques for checking parametric model fit. *Educational and Psychological Measurement*, 69, 181 - 197.
- Liang, T. & Wells, C. (2009). A model fit statistic for generalized partial credit model. *Educational and Psychological Measurement*, 69, 913 - 928.
- Orlando, M. & Thissen, D. (2000). Likelihood-based item-fit indices item response theory models. *Applied Psychological Measurement*, 24, 50 - 64.
- Sueiro, M. & Abad, F. (2011). Assessing goodness of fit in item response theory with nonparametric models: A comparison of posterior probabilities and kernel smoothing approaches. *Educational and Psychological Measurement*, (X), 1-15.
- Van Der Ark, L. A. (2001). Relationships and properties of polytomous item response theory models. *Applied Psychological Measurement*, 25 (3), 273-282.
- Wells, C. S. & Bolt D. M. (2008). Investigation of a nonparametric procedure for assessing goodness-of-fit in item response theory. *Applied Measurement in Education*, 21, 22 - 40.

ผู้เขียน

นางนัยนา จันตะเสน ศึกษานิเทศก์ประจำ สำนักงานเขตพื้นที่การศึกษาประถมศึกษาพระนครศรีอยุธยา เขต ๒ สังกัดสำนักงานคณะกรรมการ การศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ
ศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี อาจารย์ที่ปรึกษา อาจารย์ประจำสาขาวิชาการวัดและประเมินผล การศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ผู้ช่วยศาสตราจารย์ ดร.กมลวรรณ ตังธนภานนท์ อาจารย์ที่ปรึกษาร่วม อาจารย์ประจำสาขาวิชา การวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
หมายเหตุ: การวิจัยครั้งนี้ได้รับการสนับสนุนทุนวิจัยจาก “ทุน ๙๐ ปี จุฬาลงกรณ์มหาวิทยาลัย” กองทุนรัชดาภิเษกสมโภช