



ประสิทธิภาพของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนรูปแบบผสม : การประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนความรู้บางส่วน และทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนความรู้บางส่วนแบบทั่วไป
The Efficiency of Mixed-Format Achievement Tests : An Application of Partial Credit Model and Generalized Partial Credit Model

รุ่งนภา แสนอานวยผล (Rungnapha Saen-amnuaiphon)*

ดร.ประกฤติยา ทักชีโน (Dr.Prakittiya Tuksino)**

ดร.ชนะศึก นิชานนท์ (Dr.Chanasuek Nichanong)***

บทคัดย่อ

การศึกษาค้นคว้าครั้งนี้ มีวัตถุประสงค์ คือ 1) เพื่อศึกษาประสิทธิภาพของแบบทดสอบรูปแบบผสม และ 2) เพื่อศึกษาปฏิสัมพันธ์ระหว่างโมเดลการตรวจให้คะแนน สัดส่วนของข้อสอบที่ตรวจให้คะแนนแบบสองค่าและมากกว่าสองค่า และความยาวของแบบทดสอบ และเปรียบเทียบประสิทธิภาพของแบบทดสอบรูปแบบผสม เงื่อนไขที่ทำการศึกษามี 18 เงื่อนไข ประกอบด้วยโมเดลการตรวจให้คะแนน 2 โมเดล คือโมเดลโลจิสติก 1 พารามิเตอร์กับโมเดลการตรวจให้คะแนนความรู้บางส่วน (PCM) และโมเดลโลจิสติก 3 พารามิเตอร์กับโมเดลการตรวจให้คะแนนความรู้บางส่วนแบบทั่วไป (GPCM) สัดส่วนข้อสอบที่ตรวจให้คะแนนแบบสองค่าและมากกว่าสองค่า 3 สัดส่วน คือ 20:80 50:50 และ 80:20 และความยาวของแบบทดสอบ 3 เงื่อนไข คือ 10 30 และ 50 ข้อ การประเมินประสิทธิภาพของแบบทดสอบรูปแบบผสมพิจารณาจากดัชนีความคลาดเคลื่อนมาตรฐานในการประมาณค่า ($SE(\theta)$) ดัชนีความลำเอียง (BIAS) พร้อมทั้งวิเคราะห์ความแปรปรวนแบบพหุจำแนก 3 ทาง (Three-way MANOVA) เพื่อเปรียบเทียบค่าเฉลี่ยของดัชนี $SE(\theta)$ และ BIAS ผลการวิจัยพบว่า 1. โมเดลโลจิสติก 1 พารามิเตอร์กับ PCM และโมเดลโลจิสติก 3 พารามิเตอร์กับ GPCM มีค่า $SE(\theta)$ และ BIAS ต่ำสุดที่สัดส่วนข้อสอบที่ตรวจให้คะแนนสองค่าและมากกว่าสองค่า คือ 20:80 และความยาวของแบบทดสอบ 50 ข้อ 2. มีปฏิสัมพันธ์ระหว่างโมเดลการตรวจให้คะแนน สัดส่วนของข้อสอบที่ตรวจให้คะแนนสองค่าและมากกว่าสองค่า และความยาวของแบบทดสอบ ที่ส่งผลต่อค่า $SE(\theta)$ และ BIAS อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนปฏิสัมพันธ์รายคู่ พบว่า ปฏิสัมพันธ์ระหว่างโมเดลการตรวจให้คะแนน กับสัดส่วนของข้อสอบที่ตรวจให้คะแนนสองค่าและมากกว่าสองค่า ระหว่างโมเดลการตรวจให้คะแนน กับความยาวของแบบทดสอบ และระหว่างสัดส่วนของข้อสอบที่ตรวจให้คะแนนสองค่าและมากกว่าสองค่ากับความยาวของแบบทดสอบ ส่งผลต่อค่า $SE(\theta)$ และ BIAS อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 นอกจากนี้พบว่าโมเดลการตรวจให้คะแนน สัดส่วนของข้อสอบที่ตรวจให้คะแนนสองค่าและมากกว่าสองค่า และความยาวของแบบทดสอบที่แตกต่างกันส่งผลต่อค่า $SE(\theta)$ และ BIAS ที่ต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

คำสำคัญ : แบบทดสอบรูปแบบผสม, โมเดลการตรวจให้คะแนนความรู้บางส่วน, โมเดลการตรวจให้คะแนนความรู้บางส่วนแบบทั่วไป

Keywords : Mixed-Format Tests, Partial Credit Model, Generalized Partial Credit Model

* นักศึกษาระดับปริญญาเอก สาขาวิชาการวัดและประเมินผลการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น

** อาจารย์ประจำสาขาวิชาการวัดและประเมินผลทางการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น

*** ผู้อำนวยการสถาบันวิจัยและพัฒนา มหาวิทยาลัยราชภัฏสวนดุสิต

Abstract

This research was the experimental research. There were 3 objectives: 1) to study the efficiency of mixed-format achievement tests and 2) to study the interaction of the models, proportion of the dichotomous and polytomous items, and the test length, and to compare the efficiency of mixed-format tests. There were 18 studied conditions including 2 models as: one-parameter logistic model and partial credit model (PCM), and three-parameters logistic model and generalized partial credit model (GPCM), the proportions of the dichotomous and polytomous items as: 20:80, 50:50, and 80:20, and the test length as: 10, 30, and 50 items. The evaluation of efficiency in mixed-format achievement tests, was considered by index of standard error of estimation (SE (θ)), index of BIAS, and effect size whereas the Three-way MANOVA to compare the mean of SE (θ) index, and BIAS to study interaction between the studied conditions.

The research findings found that:

1. One-parameter logistic model with PCM, and three-parameter logistic model with GPCM, had SE (θ), and BIAS in lowest level at the proportion of the dichotomous and polytomous items as: 20:80, and 50 items of the test length.
2. There was an interaction among model, proportion of the dichotomous and polytomous items, and test length affected on the SE (θ) and BIAS at .05 significant level. The interaction between components showed that there were interaction between model and proportion of the dichotomous and polytomous items, model and test length, and proportion of the dichotomous and polytomous items and test length which affected on the SE (θ) and BIAS at .05 significant level. Besides, it was found that model, proportion of the dichotomous and polytomous items, and different test length affected between the SE (θ) and BIAS values at .05 significant level.

ความเป็นมาและความสำคัญของปัญหา

การวัดทางการศึกษาและจิตวิทยาเป็นการวัดคุณลักษณะภายในมนุษย์ซึ่งเป็นสิ่งที่สังเกตไม่ได้โดยตรงแต่มีความสำคัญและจำเป็นต้องศึกษา เพราะว่าคุณลักษณะภายในส่งผลหรือมีอิทธิพลต่อการแสดงออกภายนอกของบุคคล จะทำให้เข้าใจถึงสาเหตุแห่งการเกิดพฤติกรรมของบุคคล (ศิริชัย กาญจนวาสี, 2545) การวัดผลทางการศึกษาที่ใช้เป็นส่วนใหญ่ คือ การวัดผลสัมฤทธิ์ทางการเรียนของผู้เรียนซึ่งเป็นการวัดความสำเร็จทางการเรียนวัดประสบการณ์ทางการเรียนที่นักเรียนได้รับจากการเรียนการสอนในโรงเรียน ได้แก่ ความรู้พื้นฐานหรือความพร้อมของนักเรียน วัดความก้าวหน้าในการเรียน วัดเพื่อวินิจฉัยการเรียนรู้ วัดเพื่อประเมินผลรวมเมื่อการเรียนการสอนสิ้นสุดลง ซึ่งจะเป็นประโยชน์ต่อการปรับปรุงและพัฒนาการสอนของครู ส่วนใหญ่มักจะใช้แบบทดสอบวัดผลสัมฤทธิ์ (Achievement Test) เป็นเครื่องมือ (เพ็ญศรี

ทิพย์สุวรรณกุล, 2544 ; พิชิต ฤทธิ์จำรูญ, 2548 ; ศิริชัย กาญจนวาสี, 2548)

ในปัจจุบันการประเมินผู้เรียนมีความพยายามที่จะวัดทักษะความสามารถในการคิดวิเคราะห์ คิดสังเคราะห์ คิดสร้างสรรค์ ของผู้เรียนมากกว่าการวัดแค่เพียงความรู้ ความจำ และความเข้าใจในเนื้อหาเท่านั้นแต่การเลือกใช้ข้อสอบชนิดใดชนิดหนึ่งจะมีข้อจำกัดในการวัดเพราะทั้งข้อสอบแบบเลือกตอบ และข้อสอบเขียนตอบ มีทั้งจุดแข็งและจุดอ่อน เนื่องจากการใช้แบบทดสอบเพียงรูปแบบใดรูปแบบหนึ่ง อาจจะไม่สามารถวัดความรู้ ความสามารถของผู้เรียนได้อย่างครอบคลุม (Kinsey, 2003; Kim Walker & McHale, 2007; Reshetar & Melican, 2010) การทดสอบจึงจำเป็นต้องใช้แบบทดสอบทั้ง 2 รูปแบบควบคู่กันไป เรียกว่า “แบบทดสอบรูปแบบผสม” (Mixed-format Tests) ซึ่งประกอบด้วย ข้อสอบแบบเลือกตอบ หรือข้อสอบที่มี



การตรวจให้คะแนนสองค่า และข้อสอบเขียนตอบ หรือข้อสอบที่ตรวจให้คะแนนมากกว่าสองค่า

จากเหตุผลที่กล่าวมาข้างต้นจะเห็นได้ว่า แบบทดสอบรูปแบบผสมสามารถช่วยลดข้อจำกัดของข้อสอบแบบเลือกตอบ และข้อสอบเขียนตอบ และใช้จุดเด่นของข้อสอบทั้งสองชนิด พร้อมทั้งได้มีการนำแบบทดสอบรูปแบบผสมมาใช้ในสถานการณ์การทดสอบทั้งในระดับประเทศ และระดับนานาชาติ แต่ทั้งนี้ปัญหาที่พบในการออกแบบการสร้างแบบทดสอบรูปแบบผสม คือ ยังไม่มีความชัดเจนในการเลือกใช้โมเดลว่าจะใช้โมเดลใดเป็นฐานในการพัฒนา และวิเคราะห์คุณภาพแบบทดสอบ จะกำหนดจำนวนข้อสอบเท่าใด และจะกำหนดสัดส่วนของข้อสอบที่ตรวจให้คะแนนสองค่า และมากกว่าสองค่าเท่าใด จึงจะช่วยให้การใช้แบบทดสอบรูปแบบผสมมีประสิทธิภาพสูง

วัตถุประสงค์ของการวิจัย

1. ศึกษาประสิทธิภาพของแบบทดสอบรูปแบบผสม ตามโมเดลการตรวจให้คะแนน สัดส่วนของข้อสอบที่ตรวจให้คะแนนแบบสองค่าและมากกว่าสองค่า และความยาวของแบบทดสอบ
2. ศึกษาปฏิสัมพันธ์ระหว่างโมเดลการตรวจให้คะแนน สัดส่วนของข้อสอบที่ตรวจให้คะแนนแบบสองค่า และมากกว่าสองค่า และความยาวของแบบทดสอบในการส่งผลต่อประสิทธิภาพของแบบทดสอบรูปแบบผสม และเปรียบเทียบประสิทธิภาพของแบบทดสอบรูปแบบผสม เมื่อโมเดลการตรวจให้คะแนน สัดส่วนของข้อสอบที่ตรวจให้คะแนนแบบสองค่าและมากกว่าสองค่า และความยาวของแบบทดสอบแตกต่างกัน

วิธีดำเนินการวิจัย

งานวิจัยครั้งนี้ใช้ระเบียบวิธีวิจัยเชิงทดลอง โดยมีขั้นตอนการดำเนินการวิจัย ดังรายละเอียดต่อไปนี้

ขั้นตอนการดำเนินการวิจัย

1. ศึกษาค้นคว้าเกี่ยวกับ ทฤษฎี หลักการเกี่ยวกับการวิเคราะห์ประสิทธิภาพของข้อสอบ แบบทดสอบรูปแบบผสม และโมเดลการตรวจให้คะแนนสองค่า และมากกว่าสองค่า จากงานวิจัยที่เกี่ยวข้อง เอกสาร ตำรา วารสารทั้งในและต่างประเทศ

2. ศึกษาเงื่อนไขการทดสอบที่เกิดขึ้นในการศึกษาวิจัย จากเอกสาร และงานวิจัย ที่เกี่ยวข้อง รวมถึงศึกษาการใช้โปรแกรมสำเร็จรูป WinGen เพื่อใช้ในการจำลองข้อมูลตามเงื่อนไขต่าง ๆ ที่ต้องการศึกษา และโปรแกรม MULTILOG เพื่อใช้ในการประมาณค่าพารามิเตอร์แบบทดสอบรูปแบบผสม พร้อมทั้งศึกษารายละเอียดการใช้แนวทางในการจำลองข้อมูลสำหรับ ใช้ในการวิจัย เพื่อให้ได้ค่าพารามิเตอร์ที่แท้จริงมาใช้ในการวิจัย

3. จำลองข้อมูลการตอบของผู้สอบตามแบบแผน และตัวแปรที่กำหนดด้วยโปรแกรม WinGen

4. ตรวจสอบความเป็นเอกมิติของข้อมูลเพื่อตรวจสอบข้อตกลงเบื้องต้นตามทฤษฎีการตอบสนองข้อสอบ ด้วยการวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory Factor Analysis) ด้วยโปรแกรม SPSS

5. ประมาณค่าพารามิเตอร์ของข้อสอบ และแบบทดสอบตามทฤษฎีการตอบสนองข้อสอบตามโมเดลการตรวจให้คะแนนแบบโลจิสติก 1 พารามิเตอร์กับ PCM และ โมเดลโลจิสติก 3 พารามิเตอร์กับ GPCM จากข้อมูลผลการตอบที่ได้จากการจำลองข้อมูล ด้วยโปรแกรม MULTILOG

6. เปรียบเทียบผลการวิเคราะห์ข้อมูลแต่ละเงื่อนไข จากค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า (SE()) ค่าความลำเอียง (BIAS) และขนาดอิทธิพล (Effect Size)

7. สรุปผลการวิเคราะห์ข้อมูล เพื่อเปรียบเทียบผลการวิเคราะห์ข้อมูลในแต่ละเงื่อนไขที่ทำการศึกษา

เงื่อนไขที่ใช้ในการจำลองข้อมูล

1. โมเดลการตรวจให้คะแนนที่ใช้ในการวิเคราะห์ข้อมูลครั้งนี้ คือโมเดลโลจิสติก 1 พารามิเตอร์กับ PCM และ โมเดลโลจิสติก 3 พารามิเตอร์กับ GPCM กำหนดรายการคำตอบเป็น 5 ลำดับขั้น คือ 0 1 2 3 และ 4
2. ความยาวของแบบทดสอบ 10 30 และ 50 ข้อ
3. สัดส่วนของข้อสอบที่ตรวจให้คะแนนสองค่าและมากกว่าสองค่าคือ 20:80 50:50 และ 80:20

การศึกษาการจำลองข้อมูล (Simulation Study)

1. จำลองข้อมูลพารามิเตอร์ผู้สอบ ตามเงื่อนไขที่กำหนด คือ เลือกจำนวนผู้สอบ 1,000 คน กำหนดให้ค่าเฉลี่ยพารามิเตอร์ผู้สอบเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1

2. จำลองข้อมูลพารามิเตอร์ข้อสอบ

3. ตรวจสอบความถูกต้องของข้อมูลที่ได้จากการจำลอง โดย พิจารณาค่าพารามิเตอร์ผู้สอบว่าต้องอยู่ในช่วง -3.0 ถึง 3.0 และค่าพารามิเตอร์ข้อสอบ คือ ค่าพารามิเตอร์ความยาก (bi) ต้องมีค่าอยู่ในช่วง -2.5 ถึง 2.5 ค่าพารามิเตอร์อำนาจจำแนก (ai) มีค่าอยู่ในช่วง 0.5 ถึง 2.5 และค่าพารามิเตอร์การเดา (ci) ที่ต้องมีค่าไม่เกิน 0.3 ถ้าค่าพารามิเตอร์ที่ได้ไม่เป็นไปตามที่กำหนด ผู้วิจัยก็จะดำเนินการจำลองข้อมูลใหม่ เพื่อให้ได้ค่าพารามิเตอร์ตามที่กำหนด

4. นำข้อมูลพารามิเตอร์ผู้สอบ และพารามิเตอร์ข้อสอบที่ได้จากการจำลองข้อมูลคำนวณหาความน่าจะเป็นในการตอบข้อสอบตามลักษณะของโมเดลการตรวจให้คะแนน

5. หาผลการตอบของผู้สอบ

6. นำข้อมูลการตอบที่ได้จากการจำลองข้อมูล ไปคำนวณค่าพารามิเตอร์ข้อสอบด้วยโปรแกรม MULTILOG

การวิเคราะห์ข้อมูล

1. การตรวจสอบความเป็นเอกมิติของข้อมูล ใช้การวิเคราะห์องค์ประกอบ (Factor analysis) โดยการคำนวณอัตราส่วนของค่าไอเกน (Eigen Ratio: ER) ขององค์ประกอบที่ 1 กับองค์ประกอบที่ 2

2. ความคลาดเคลื่อนมาตรฐานในการประมาณค่า (SE()) คำนวณได้จากสูตร

$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}}$$

3. ดัชนีความลำเอียง (BIAS)

$$MSE(\theta) = \frac{\sum_{i=1}^r (\theta_i - \hat{\theta}_i)^2}{r}$$

ผลการวิจัยและอภิปรายผล

1. ปฏิสัมพันธ์ระหว่างโมเดลการตรวจให้คะแนน สัดส่วนข้อสอบ ที่ตรวจให้คะแนนสองค่าและมากกว่าสองค่า และความยาวของแบบทดสอบ เมื่อพิจารณาจากค่า $SE(\theta = -3.0)$ $SE(\theta = 0.0)$ $SE(\theta = 3.0)$ และ BIAS

ผลการวิจัยพบว่า 1) มีปฏิสัมพันธ์ระหว่างโมเดลการตรวจให้คะแนน สัดส่วนข้อสอบที่ตรวจให้คะแนนสองค่าและมากกว่าสองค่า และความยาวของแบบทดสอบ ที่ส่งผลต่อประสิทธิภาพของแบบทดสอบรูปแบบผสม คือ $SE(\theta = -3.0)$ $SE(\theta = 0.0)$ $SE(\theta = 3.0)$ และ BIAS อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 2) มีปฏิสัมพันธ์ระหว่างโมเดลการตรวจให้คะแนน กับสัดส่วนข้อสอบที่ตรวจให้คะแนนสองค่าและมากกว่าสองค่า อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 3) มีปฏิสัมพันธ์ระหว่างโมเดลการตรวจให้คะแนน กับความยาวของแบบทดสอบ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 4) มีปฏิสัมพันธ์ระหว่างสัดส่วนข้อสอบที่ตรวจให้คะแนนสองค่าและมากกว่าสองค่า กับความยาวของแบบทดสอบ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งแสดงให้เห็นว่าในการพิจารณาเลือกใช้ หรือสร้างแบบทดสอบรูปแบบผสมให้มีประสิทธิภาพทั้งด้านความตรง (Validity) และความเที่ยง (Reliability) ผู้ใช้ หรือผู้สร้างควรพิจารณาเลือกทั้งโมเดลการตรวจให้คะแนน สัดส่วนข้อสอบที่ตรวจให้คะแนน สองค่าและมากกว่าสองค่า และความยาวของแบบทดสอบ ควบคู่กันทั้งสามเงื่อนไข ไม่ควรพิจารณาเพียงเงื่อนไขใดเงื่อนไขหนึ่ง เพราะทั้งสามเงื่อนไขได้ร่วมกันส่งผลต่อประสิทธิภาพของแบบทดสอบรูปแบบผสมทั้งสิ้น

2. เปรียบเทียบโมเดลการตรวจให้คะแนน สัดส่วนข้อสอบแบบตรวจให้คะแนนสองค่าและมากกว่าสองค่า และความยาวของแบบทดสอบ ที่ส่งผลต่อประสิทธิภาพของแบบทดสอบรูปแบบผสม

ผลการวิจัย พบว่า โมเดลการตรวจให้คะแนน สัดส่วนข้อสอบปรนัยหรือที่ตรวจให้คะแนนสองค่าและข้อสอบอัตนัยหรือที่ตรวจให้คะแนนมากกว่าสองค่า และความยาวของแบบทดสอบที่แตกต่างกัน ต่างก็มีอิทธิพลต่อประสิทธิภาพของแบบทดสอบรูปแบบผสม คือ $SE(\theta = -3.0)$ $SE(\theta = 0.0)$ $SE(\theta = 3.0)$ และ BIAS



อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ทั้งนี้เมื่อพิจารณาจากผลการวิจัยตามวัตถุประสงค์ข้อที่ 1 จะพบว่าทั้งสัดส่วนข้อสอบปรนัยและอัตนัย และความยาวของแบบทดสอบต่างเป็นปัจจัยที่ส่งผลต่อการประมาณค่าความตรงและความเที่ยงทั้งสิ้น (Nunnally, 1964; Mehren & Lehmann, 1973) สอดคล้องกับผลการศึกษาของ Kinsey (2003) ที่พบว่าทั้งความยาวของแบบทดสอบและสัดส่วนของข้อสอบมีอิทธิพลต่อค่าสารสนเทศของแบบทดสอบอย่างมีนัยสำคัญทางสถิติ ที่ระดับ .05

เมื่อพิจารณาโมเดลการตรวจให้คะแนน พบว่าแบบทดสอบรูปแบบผสมที่ใช้โมเดลโลจิสติก 1 พารามิเตอร์กับ PCM มีประสิทธิภาพดีกว่าโมเดลโลจิสติก 3 พารามิเตอร์กับ GPCM อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เนื่องจากเมื่อพิจารณาการประมาณค่าพารามิเตอร์ของข้อสอบปรนัยหรือที่ตรวจให้คะแนนสองค่า โดยใช้โมเดลโลจิสติก 1 พารามิเตอร์ พบว่า ค่าพารามิเตอร์ความยาก (bi) จะมีค่าแปรเปลี่ยนตามลักษณะของข้อสอบแต่ละข้อ แต่พารามิเตอร์อำนาจจำแนก (ai) คงที่ และพารามิเตอร์การเดา (ci) เป็น 0 ในขณะที่แบบทดสอบรูปแบบผสมที่ใช้โมเดลโลจิสติก 3 พารามิเตอร์กับ GPCM การประมาณค่าพารามิเตอร์ของข้อสอบที่ตรวจให้คะแนนสองค่า โดยใช้โมเดลโลจิสติก 3 พารามิเตอร์ ซึ่งพบว่าค่า พารามิเตอร์ความยาก (bi) พารามิเตอร์อำนาจจำแนก (ai) และค่าพารามิเตอร์การเดา (ci) จะแปรเปลี่ยนตามลักษณะของข้อสอบแต่ละข้อ จึงส่งผลให้ค่า SE() และค่า BIAS ของโมเดล PCM ต่ำกว่าโมเดล GPCM ยังพบว่า แบบทดสอบรูปแบบผสมที่ใช้โมเดล PCM มีสัดส่วนข้อสอบที่ตรวจให้คะแนนสองค่าและมากกว่าสองค่า คือ 20:80 มีความยาวของแบบทดสอบ 50 ข้อ มีค่า SE(θ) ต่ำสุดที่ระดับความสามารถ 0.0 ถึง -3.0 (กลุ่มปานกลางถึงต่ำ) ส่วนแบบทดสอบที่ใช้โมเดล GPCM มีสัดส่วนข้อสอบปรนัยหรือที่ตรวจให้คะแนนสองค่าและข้อสอบอัตนัยหรือที่ตรวจให้คะแนนมากกว่าสองค่า คือ 20:80 มีความยาวของแบบทดสอบ 50 ข้อ มีค่าความคลาดเคลื่อนมาตรฐานต่ำสุดที่ระดับความสามารถ 0.0 ถึง 3.0 (กลุ่มปานกลางถึงสูง) ทั้งนี้เนื่องจากผู้สอบที่มีระดับความสามารถต่ำมีโอกาสในการเดามากกว่าผู้สอบที่มีระดับความสามารถสูง ดังนั้นโมเดล PCM จึงให้ค่า SE(θ) ต่ำกว่า แต่สำหรับผู้สอบที่

มีระดับความสามารถสูงโอกาสการเดาจะน้อยกว่ากลุ่มต่ำ หรืออาจจะเข้าใกล้ 0 นอกจากนี้โมเดล 3PL ได้ทำการประมาณค่าอำนาจจำแนกด้วย จึงส่งผลให้ค่า SE(θ) ของโมเดล GPCM ต่ำกว่า โมเดล PCM ทั้งนี้สอดคล้องกับแนวคิดของ Hambleton and Swaminatan (1985) Embretson and Reise (2000) และ McIntire and Miller (2007) ที่ได้กล่าวถึงเกณฑ์ในการพิจารณาขนาดค่า SE() ที่กำหนดไว้ว่าถ้าค่าพารามิเตอร์อำนาจจำแนก (ai) สูง และค่าพารามิเตอร์การเดา (ci) ต่ำ ขนาดของค่า SE() จะลดลง นอกจากนี้ผลการศึกษาค่า BIAS ยังพบอีกว่าโมเดล GPCM จะมีค่า BIAS ต่ำกว่า PCM สำหรับแบบทดสอบที่มีสัดส่วนข้อสอบที่ตรวจให้คะแนนสองค่าและมากกว่าสองค่า คือ 20:80 มีความยาวของแบบทดสอบ 10 ข้อ ทั้งนี้เนื่องจากแบบทดสอบ มีสัดส่วนข้อสอบปรนัย ร้อยละ 20 จากข้อสอบ 10 ข้อ แสดงว่ามีข้อสอบปรนัยเพียง 2 ข้อ จาก 10 ข้อ ซึ่งเมื่อข้อสอบปรนัยมีจำนวนข้อสอบน้อยโอกาสในการเดาของผู้สอบจะมีน้อยกว่าข้อสอบที่มีสัดส่วนข้อสอบปรนัยและอัตนัย คือ 50:50 หรือ 80:20 จึงทำให้โมเดล GPCM มีค่า BIAS ต่ำกว่าโมเดล PCM

เมื่อพิจารณาสัดส่วนข้อสอบที่ตรวจให้คะแนนสองค่าและมากกว่าสองค่า พบว่าแบบทดสอบที่มีสัดส่วนข้อสอบที่ตรวจให้คะแนนสองค่าและมากกว่าสองค่า คือ 20:80 มีประสิทธิภาพสูงกว่า 50:50 และ 80:20 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และแบบทดสอบที่มีสัดส่วนข้อสอบที่ตรวจให้คะแนนสองค่าและมากกว่าสองค่า คือ 50:50 มีประสิทธิภาพสูงกว่า 80:20 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ผลการวิจัยแสดงให้เห็นว่าประสิทธิภาพของแบบทดสอบรูปแบบผสมจะเกี่ยวข้องกับจำนวนของข้อสอบที่ตรวจให้คะแนนมากกว่าสองค่าซึ่งเป็นข้อสอบอัตนัยชนิดตรวจให้คะแนนความรู้บางส่วน มากกว่าข้อสอบปรนัยชนิดเลือกตอบ ซึ่ง Kim Walker & McHale (2007) & Reshetar and Melican (2010) ได้กล่าวถึงจุดเด่นของข้อสอบอัตนัยที่สำคัญคือ ลดการเดาคำตอบถูกของผู้สอบ ในขณะที่ข้อสอบปรนัย เพิ่มโอกาสในการเดาคำตอบถูก (ci) ของผู้สอบอยู่ด้วย ซึ่งค่าโอกาสในการเดา (ci) จะส่งผลไปยังค่าสารสนเทศของข้อสอบ (Item Information) กล่าวคือเมื่อโอกาสในการเดาเข้าใกล้ 0

หรือเมื่อไม่มีโอกาสในการเดาค่าสารสนเทศของข้อสอบ จะสูงขึ้น และสารสนเทศของข้อสอบ (Item Information) เมื่อรวมกันทั้งฉบับจะได้ค่าสารสนเทศของแบบทดสอบ (TIF) ที่มีค่าสูงขึ้นด้วย พร้อมส่งผลให้ค่า $SE(\theta)$ ต่ำลงด้วย ดังนั้นการที่แบบทดสอบที่มีสัดส่วนข้อสอบ 20:80 มีประสิทธิภาพสูงกว่าแบบทดสอบที่มีสัดส่วน 50:50 และ 80:20 เนื่องมาจากการที่สัดส่วนข้อสอบอัตนัยหรือตรวจให้คะแนนมากกว่าสองค่ามีจำนวนเพิ่มขึ้น จะทำให้ข้อสอบแบบเลือกตอบมีจำนวนลดลง ซึ่งส่งผลให้ค่าโอกาสในการเดาลดลง เมื่อโอกาสในการเดาลดลงค่า ขนาดของ $SE(\theta)$ ก็จะลดลงด้วย (Hambleton & Swaminatan, 1985; Embretson & Reise, 2000; Mehren & Lehmann, 1973) นอกจากนี้ Cao (2008) และ Messick (1995) ยังได้เปรียบเทียบข้อสอบอัตนัยและข้อสอบปรนัยไว้ว่า ข้อสอบอัตนัยจะมีค่าความตรงเชิงเนื้อหา (Content Validity) สูงกว่าข้อสอบปรนัยเนื่องจากสามารถกำหนดเนื้อหาการวัดและทักษะได้ตรงตามจุดประสงค์ได้มากกว่า และในทำนองเดียวกันข้อสอบอัตนัยมักจะมีค่าความตรงเชิงโครงสร้างสูงกว่าข้อสอบปรนัยชนิดเลือกตอบ เพราะโดยธรรมชาติของข้อสอบปรนัยชนิดเลือกตอบจะถูกจำกัดการตอบในกรอบที่กำหนดให้เท่านั้น จึงเป็นการยากที่จะดึงความรู้ หรือความคิดที่แตกต่างกันหรือความคิดที่อยู่นอกกรอบออกมาได้ นอกจากนี้ข้อสอบอัตนัยยังสามารถวัดได้ตรงตามสภาพจริงหรือตรงตามความสามารถที่แท้จริงอีกด้วย ดังนั้นเมื่อสัดส่วนของข้อสอบอัตนัยหรือข้อสอบที่ตรวจให้คะแนนมากกว่าสองค่าเพิ่ม ค่า $SE(\theta)$ และค่า BIAS จึงลดลง

เมื่อพิจารณาความยาวของแบบทดสอบ พบว่าแบบทดสอบที่มีความยาว 50 ข้อ มีประสิทธิภาพสูงกว่า 30 ข้อ และ 10 ข้อ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และแบบทดสอบที่มี ความยาวของแบบทดสอบ 30 ข้อ มีประสิทธิภาพสูงกว่า 10 ข้อ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 จากผลการวิจัยที่แสดงให้เห็นว่า เมื่อความยาวของแบบทดสอบเพิ่มขึ้นค่า $SE(\theta)$ และค่า BIAS ก็จะลดลงด้วย เนื่องจากค่า $SE(\theta)$ เป็นค่าที่สะท้อนถึงค่าฟังก์ชันสารสนเทศของแบบทดสอบ (TIF) กล่าวคือเมื่อแบบทดสอบมีค่า $SE(\theta)$ ต่ำ แสดงว่าค่าฟังก์ชันสารสนเทศของแบบทดสอบ (TIF) สูง ทั้งนี้ค่าฟังก์ชันสารสนเทศของ

แบบทดสอบ (TIF) ยังสะท้อนถึงค่าความเที่ยง (Reliability) ตามทฤษฎีการทดสอบแบบดั้งเดิมอีกด้วย พร้อมกันนี้ค่า BIAS ยังแสดงถึงความสามารถของแบบทดสอบรูปแบบผลมในการประมาณค่าได้ใกล้เคียงกับค่าที่แท้จริง ซึ่งคำนวณมาจากผลต่างของระดับความสามารถที่แท้จริงกับระดับความสามารถที่ได้จากการประมาณค่า และจะสะท้อนถึงค่าความตรง (Validity) ของแบบทดสอบตามทฤษฎีการทดสอบแบบดั้งเดิม อีกด้วยเช่นกัน ทั้งนี้ Alen and Yen (1979) McIntire & Miller (2007) และ Mehren & Lehmann (1973) ได้กล่าวถึงปัจจัยที่ส่งผลต่อค่าสัมประสิทธิ์ความเที่ยง และความตรง (Validity) อีกประการหนึ่ง คือความยาวของแบบทดสอบ ซึ่งเมื่อพิจารณาจากแบบทดสอบที่วัดเนื้อหาเดียวกันจะพบว่าแบบทดสอบที่มีจำนวนข้อสอบมากกว่าจะทำให้ค่าสัมประสิทธิ์ความเที่ยงสูงกว่าแบบทดสอบที่มีจำนวนข้อสอบน้อย นอกจากนี้ถ้าข้อสอบมีจำนวนข้อน้อยเกินไปก็จะขาดความเป็นตัวแทนของสิ่งของที่มุ่งวัด จึงส่งผลให้ค่าความตรงลดลงได้ และสอดคล้องกับหลักการพิจารณาค่า $SE(\theta)$ ของการประมาณค่าตามทฤษฎีการตอบสนองข้อสอบคือแบบทดสอบที่มีจำนวนข้อสอบมากขึ้น ขนาดของ $SE(\theta)$ จะลดลง (Hambleton & Swaminatan, 1985; Embretson & Reise, 2000; ศิริชัย กาญจนวาสี, 2550) นอกจากนี้หลักการประมาณค่าความสามารถโดยใช้วิธีการประมาณค่าที่เป็นไปได้สูงสุด (Maximum Likelihood Estimate ; MML) พบว่าถ้าแบบทดสอบมีจำนวนข้อสอบยาวขึ้นการประมาณค่าที่เป็นไปได้สูงสุดของ จะมี การแจกแจงแบบปกติโดยมีค่าเฉลี่ยเท่ากับ และเป็นการประมาณค่า ที่ไม่ลำเอียง (Unbiased)

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลการวิจัยไปใช้

การเลือกใช้โมเดลการตรวจให้คะแนนพบว่าโมเดลโลจิสติก 1 พารามิเตอร์กับ PCM มีประสิทธิภาพของแบบทดสอบรูปแบบผลมสูง กว่าโมเดลโลจิสติก 3 พารามิเตอร์กับ GPCM ดังนั้นจากผลการวิจัยพบว่า การพิจารณาเลือกใช้โมเดลการตรวจให้คะแนนควรพิจารณาเลือกใช้โมเดล โลจิสติก 1 พารามิเตอร์กับ PCM แต่ทั้งนี้ถ้ากำหนดสัดส่วนข้อสอบแบบตรวจให้คะแนนสอง



ค่าและมากกว่า สองค่า คือ 20:80 หรือกำหนดให้มีสัดส่วนข้อสอบอัตนัยมากกว่าปรนัยจะพบว่า ทั้งสองโมเดลไม่แตกต่างกัน โดยจะเห็นว่าโมเดลโลจิสติก 3 พารามิเตอร์กับGPCM เป็นโมเดลที่เหมาะสมกับการทดสอบในสถานการณ์จริงมากที่สุดเนื่องจากการสอบด้วยข้อสอบปรนัยจะมีการเดาเข้ามาเกี่ยวข้อง และโมเดลโลจิสติก 3 พารามิเตอร์กับ GPCM ก็มีการประมาณค่าพารามิเตอร์การเดาด้วย ในขณะที่โมเดล โลจิสติก 1 พารามิเตอร์กับ PCM กำหนดให้ไม่มีการเดา ซึ่งไม่สอดคล้องกับสถานการณ์จริง ดังนั้นถ้ากำหนดสัดส่วนข้อสอบอัตนัยมากกว่าปรนัย ผู้สร้างแบบทดสอบรูปแบบผสมควรเลือกใช้โมเดล โลจิสติก 3 พารามิเตอร์กับ GPCM

การกำหนดสัดส่วนของข้อสอบที่ตรวจให้คะแนนสองค่าและมากกว่าสองค่า ในแบบทดสอบรูปแบบผสม ผู้วิจัยกำหนดสัดส่วนข้อสอบโดยอาศัยฐานแนวคิด 3 ประการ คือ กลุ่มที่ 1 กำหนดจากสัดส่วนข้อสอบปรนัยน้อยกว่าอัตนัย กลุ่มที่ 2 กำหนดจากสัดส่วนข้อสอบปรนัยเท่ากับอัตนัย และกลุ่มที่ 3 กำหนดจากสัดส่วนข้อสอบปรนัยมากกว่าอัตนัย และผลการวิจัยพบว่าการเลือกใช้แบบทดสอบที่มีสัดส่วนข้อสอบปรนัยน้อยกว่าอัตนัย จะให้ประสิทธิภาพของแบบทดสอบคือ ความเที่ยง และความตรง สูงที่สุด ดังนั้นจึงเป็นข้อเสนอแนะในการนำไปใช้คือ การกำหนดสัดส่วนข้อสอบที่ตรวจให้คะแนนสองค่าและมากกว่าสองค่า จะให้ประสิทธิภาพสูงที่สุดเมื่อกำหนดสัดส่วนข้อสอบปรนัยต่ออัตนัย คือ 20:80 หรือพิจารณาให้มีสัดส่วนข้อสอบปรนัยน้อยกว่าข้อสอบอัตนัย

การกำหนดความยาวของแบบทดสอบ จากการวิจัยครั้งนี้ กำหนด 3 กลุ่มคือกลุ่มแบบทดสอบฉบับสั้น คือ 10 ข้อ กลุ่มแบบทดสอบฉบับกลาง คือ 30 ข้อ และกลุ่มแบบทดสอบฉบับยาว คือ 50 ข้อ ผลการวิจัย แสดงให้เห็นว่าแบบทดสอบรูปแบบผสมไม่เหมาะสมกับแบบทดสอบฉบับสั้นเพราะจะทำให้ความเที่ยงและความตรงของแบบทดสอบมีค่าต่ำที่สุด ควรพิจารณากำหนดความยาวของแบบทดสอบไว้ที่ 30 และ 50 ข้อ หรืออยู่ระหว่าง 30-50 ข้อ เพราะความเที่ยงและความตรงของแบบทดสอบจะมีค่าใกล้เคียงกัน แต่ทั้งนี้ควรพิจารณาความสอดคล้องและครอบคลุมกับเนื้อหาที่ต้องการวัดอีกด้วย

2. ข้อเสนอแนะสำหรับการวิจัยต่อไป

1) เนื่องจากการศึกษาครั้งนี้เป็นการศึกษาประสิทธิภาพของแบบทดสอบรูปแบบผสมโดยใช้ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) ทั้งนี้ควรมีการศึกษาโดยใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (Generalizability Theory: GT) ด้วย ซึ่ง GT จะเป็นการวิเคราะห์ความน่าเชื่อถือ หรือความเที่ยง (Reliability) ในสถานการณ์หรือเงื่อนไขต่าง ๆ ของการวัดที่ทำให้สามารถอธิบายและควบคุมแหล่งความคลาดเคลื่อนของการวัดได้ตรงประเด็น เพื่อให้ผลการวัดมีความน่าเชื่อถือหรือมีความเที่ยง สูงถึงระดับที่ต้องการได้

2) การศึกษาครั้งนี้ผู้วิจัยกำหนดระดับคะแนนสำหรับข้อสอบที่ตรวจให้คะแนนมากกว่าสองค่าไว้ที่ 5 ระดับคะแนน คือ 0-4 จึงควรศึกษา ในเงื่อนไขของระดับคะแนนที่แตกต่างกันเพื่อเปรียบเทียบประสิทธิภาพของแบบทดสอบรูปแบบผสมต่อไป

3) การศึกษาครั้งนี้ผู้วิจัยได้ศึกษาเปรียบเทียบประสิทธิภาพของแบบทดสอบรูปแบบผสมที่มีวัดความสามารถ หรือคุณลักษณะเดียว ซึ่งในปัจจุบันเครื่องมือที่ใช้ในการประเมินผลผู้เรียนสามารถวัดหลากหลายคุณลักษณะมากขึ้น ดังนั้นควรมีการศึกษาสำหรับแบบสอบรูปแบบผสมที่ถูกพิจารณาในลักษณะที่แบบสอบมีความเป็นพหุมิติ ซึ่งลักษณะของความเป็นพหุมิติแสดงให้เห็นถึงการฝ่าฝืนข้อตกลงเบื้องต้นตามทฤษฎีการตอบสนองข้อสอบ โดยกำหนดระดับความสัมพันธ์ของความเป็นพหุมิตินี้หลายระดับแตกต่างกันไป เพื่อศึกษาผลของวิธีการเปรียบเทียบประสิทธิภาพแต่ละเงื่อนไขที่เหมาะสมกับสถานการณ์การทดสอบดังกล่าว

4) การศึกษาครั้งนี้ผู้วิจัยได้ศึกษาเงื่อนไขสัดส่วนข้อสอบปรนัย หรือที่ตรวจให้คะแนนสองค่าต่อข้อสอบอัตนัยหรือที่ตรวจให้คะแนนมากกว่าสองค่าโดยมิได้ศึกษาสัดส่วนของคะแนนที่ได้ ดังนั้นควรมีการศึกษาประสิทธิภาพของแบบทดสอบรูปแบบผสมโดยเปรียบเทียบจากเงื่อนไขสัดส่วนของคะแนนจากข้อสอบปรนัยและข้อสอบอัตนัยที่แตกต่างกัน

5) ควรทำการศึกษาประสิทธิภาพของแบบทดสอบรูปแบบผสมกับสถานการณ์การทดสอบจริง เพื่อเป็นขยายองค์ความรู้ เนื่องจากการวิจัยเกี่ยวกับ



การศึกษาประสิทธิภาพของแบบทดสอบรูปแบบผสม กับโมเดลที่ต้องการ และเป็นไปตามข้อตกลงเบื้องต้นของ
ที่ผ่านมามีส่วนใหญ่เป็นการจำลองข้อมูลเพื่อให้สอดคล้อง ทฤษฎีที่ศึกษา

เอกสารอ้างอิง

- ชวาล แพทย์กุล. (2518). **เทคนิคการวัดผล**. พิมพ์ครั้งที่ 6. กรุงเทพฯ: วัฒนาพานิช.
- พรทิพย์ ไชยใส. (2534). **การพัฒนาสูตรการให้คะแนนแบบสอบเลือกตอบสำหรับความรู้บางส่วนของผู้ตอบ : การประยุกต์ใช้วิธีการของอาร์โนลด์ และวิธีการของแฮมตัน**. วิทยานิพนธ์ครุศาสตรดุษฎีบัณฑิต สาขาวิชาการศึกษา และประเมินผลการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- พัชรี จันทรเพ็ง. (2550). **การเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบ แบบพหุมิติภายใต้การหมุนแกน โครงสร้างเชิงมิติและระดับความสัมพันธ์ที่แตกต่างกัน**. วิทยานิพนธ์ ปริญญาครุศาสตรดุษฎีบัณฑิต สาขาวิชาการศึกษา และประเมินผลการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- พิชิต ฤทธิ์จำรูญ. (2548). **หลักการวัดและประเมินผลการศึกษา**. พิมพ์ครั้งที่ 3. กรุงเทพฯ: เข้าส์ ออฟ เคอร์ มีสท์.
- เพ็ญศรี ทิพย์สุวรรณกุล. (2544). **เครื่องมือวัดผลสัมฤทธิ์ทางการเรียน**. เอกสารการสอนชุดวิชาการพัฒนาแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน. กรุงเทพฯ: สำนักพิมพ์มหาวิทยาลัยสุโขทัย ธรรมาธิราช.
- ศิริชัย กาญจนวาสี. (2545). **ทฤษฎีการทดสอบแนวใหม่**. พิมพ์ครั้งที่ 2. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- _____. (2550ก). **นโยบายการประเมินผลการเรียนรู้ตามพระราชบัญญัติการศึกษาแห่งชาติ**. ใน สุวิมล ว่องวานิช (บรรณาธิการ). **หนังสือชุดปฏิรูปการศึกษา: การประเมินผลการเรียนรู้แนวใหม่**. (หน้า 3-24). กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- อัญชลี ศรีกลชาญ. (2552). **คุณภาพของการปรับเทียบคะแนนสำหรับแบบสอบรูปแบบผสม: การประยุกต์ใช้ การปรับเทียบตามทฤษฎีการตอบสนองข้อสอบด้วยวิธีไค้คุณลักษณะและการปรับค่าพารามิเตอร์ พร้อมกัน**. วิทยานิพนธ์ปริญญาครุศาสตรดุษฎีบัณฑิต สาขาวิชาการศึกษา และประเมินผลการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- Andersen, E. B. (1997). The Rating scale model. In W. J. van der Linden, & R. K. Hambleton (Eds.). **Handbook of modern item response theory**. New York: Springer-Verlag.
- Hambleton, R. K & Swaminatan, H. (1985). **Item response theory: Principles and application**. 2nd ed. Boston: Kluwer Nijhoff Publishing.
- Hambleton, R. K. , Swaminatan, H. & Roger, H. J. (1991). **Fundamentals of item response theory**. C. A: SAGE Publications.
- Hambleton, R. K. (1996). Advances in assessment models, methods, and practices. Berliner D. C. and Calfee R. C. P. (Eds.). **Handbook of educational psychology**, (pp. 899 – 925). New York: Simon & Schuster Maemillan.
- Kim, S. H. & Lee, W. C. (2006). An extension of four IRT linking methods for mixed-format tests. **Journal of Educational Measurement**. 43(1). 53–70.



- Kim, S., Walker, M. E., & McHale, F. (2007). **Equating of mixed-format tests in large scale Assessments.** Paper Presented at the Annual Meeting of the American Educational Research Association (AERA) and The National Council on Measurement in Education (NCME). [n.p].
- Mehren, W. A. & Lehmann I. J. (1973). **A Measurement and evaluation in education and psychology.** New York: Holt. Rinehart and Winston.
- Reshetar, R. & Melican, G. J. (2010). **Design and evaluation of mixed-format large scale assessments for the advanced placement program (AP).** Paper presented at the Annual Meeting of the American Education Research Association in Denver, April, 2010. Denver: AERA.

