

ประสิทธิภาพของวิธีการให้คะแนนความรู้บางส่วนด้วยวิธีการประยุกต์วิธีของคูมบ์ และวิธีตัดตัวลงเลือกตัวถูก

EFFICIENCY OF PARTIAL SCORING METHODS USING MODIFIED COOMB'S AND NUMBER RIGHT ELIMINATION METHODS

ณัฐภรณ์ เลขาวัฒนพงษ์¹, กิตติทัตน์ หวานฉ่ำ², และ สีวะโชติ ศรีสุทธียากร³
Natthaporn Lekhawattanapong¹, Kittitas Wancham², and Siwachoat Srisuttiyakorn³

^{1,2}นิสิตระดับปริญญาโทบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

³อาจารย์สาขาวิชาสถิติการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

บทคัดย่อ

งานวิจัยเรื่องประสิทธิภาพของวิธีการให้คะแนนความรู้บางส่วนด้วยวิธีการประยุกต์วิธีของคูมบ์และวิธีตัดตัวลงเลือกตัวถูก มีวัตถุประสงค์เพื่อ (1) เปรียบเทียบค่าประมาณความเที่ยงของแบบสอบที่ให้คะแนนความรู้บางส่วนระหว่างวิธีการประยุกต์วิธีของคูมบ์กับวิธีให้คะแนนแบบตัดตัวลงเลือกตัวถูก (2) เปรียบเทียบความสัมพันธ์ระหว่างค่าความสามารถจริงกับค่าประมาณความสามารถระหว่างวิธีการประยุกต์วิธีของคูมบ์และวิธีตัดตัวลงเลือกตัวถูก และ (3) เปรียบเทียบฟังก์ชันสารสนเทศของแบบสอบระหว่างวิธีการประยุกต์วิธีของคูมบ์กับวิธีตัดตัวลงเลือกตัวถูก ข้อมูลที่ใช้จำลองด้วยวิธีการมอนติคาร์โลภายใต้สถานการณ์ที่แบบสอบหลายตัวเลือก จำนวน 4 ตัวเลือก มีจำนวนข้อสอบ 20 ข้อ 30 ข้อ และ 40 ข้อ และมีจำนวนผู้สอบ 1,000 คน แต่ละสถานการณ์กระทำซ้ำจำนวน 10 รอบ ประสิทธิภาพของวิธีการให้คะแนนความรู้บางส่วนพิจารณาจากความเที่ยงของแบบสอบ ความสัมพันธ์ระหว่างความสามารถจริงกับความสามารถที่ได้จากการประมาณค่า และฟังก์ชันสารสนเทศของแบบสอบ วิเคราะห์ข้อมูลด้วยการวิเคราะห์ความแปรปรวนพหุนามแบบสองทาง (two-way MANOVA)

ผลการวิจัยพบว่า (1) แบบสอบที่ให้คะแนนด้วยวิธีตัดตัวลงเลือกตัวถูกมีค่าสัมประสิทธิ์แอลฟาครอนบาคสูงกว่าวิธีการประยุกต์วิธีของคูมบ์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยแบบสอบที่มีจำนวน 40 ข้อ มีความเที่ยงสูงกว่าแบบสอบที่มีจำนวน 30 ข้อ และ 20 ข้อ ตามลำดับ (2) ความสัมพันธ์ระหว่างความสามารถจริงกับความสามารถที่ได้จากการประมาณค่าจากวิธีตัดตัวลงเลือกตัวถูกมีค่าสูงกว่าความสัมพันธ์ที่ได้จากวิธีการประยุกต์วิธีของคูมบ์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยแบบสอบที่มีจำนวน 40 ข้อ จะมีค่าความสัมพันธ์สูงกว่าแบบสอบที่มีจำนวน 30 ข้อ และ 20 ข้อ ตามลำดับ และ (3) วิธีตัดตัวลงเลือกตัวถูกมีฟังก์ชันสารสนเทศของแบบสอบสูงกว่าวิธีการประยุกต์วิธีของคูมบ์ โดยเฉพาะในช่วงค่าความสามารถของผู้สอบระดับปานกลางและสูง ทั้งแบบสอบที่มีจำนวนข้อสอบ 20 ข้อ 30 ข้อ และ 40 ข้อ

คำสำคัญ: วิธีการให้คะแนนความรู้บางส่วน วิธีการประยุกต์วิธีของคูมบ์ วิธีตัดตัวลงเลือกตัวถูก

วิธีการมอนติคาร์โล

¹Corresponding author: natthaporncu@gmail.com

ABSTRACT

The purposes of research entitled efficiency of partial scoring methods using modified Coomb's and number right elimination methods were to (1) compare reliability coefficient of test using modified Coomb's scoring method and number right elimination method, (2) compare correlation between true ability and estimated ability of students were taken modified Coomb's scoring method and number right elimination method, and (3) compare test information between test using modified Coomb's scoring method and number right elimination method. The data was generated using a Monte Carlo simulation with multiple choice test with four choices had three different number of items; 20 items, 30 items and 40 items and had 1,000 test takers, each of which was repeated 10 times. Determined efficiency of partial scoring methods from reliability coefficient, correlation between true ability and estimated ability, and test information. Data analyzed by two-way MANOVA.

The results showed that (1) The number right elimination method is higher reliability coefficient than modified Coomb's scoring method significantly at .05 which 40 items were higher than the 30 items and 20 items respectively. (2) The number right elimination method is higher correlation between true ability and estimated ability of students than modified Coomb's scoring method significantly at .05 which 40 items were higher than the 30 items and 20 items respectively. (3) The number right elimination method is higher test information than modified Coomb's scoring method which 40 items were higher than the 30 items and 20 items respectively. Especially in the range of intermediate and high ability.

KEYWORDS: Partial Scoring Methods, Modified Coomb's Methods, Number Right Elimination Methods, Monte Carlo Simulation

ความเป็นมาและความสำคัญของปัญหา

การวัดและประเมินผลการศึกษาเป็นองค์ประกอบสำคัญในกระบวนการจัดการเรียนการสอน (Lau, Lau, Hong, and Usop, 2011) ซึ่งแผนการศึกษาแห่งชาติ พ.ศ. 2560-2579 ยุทธศาสตร์ที่ 3 การพัฒนาศักยภาพคนทุกช่วงวัย และการสร้างสังคมแห่งการเรียนรู้ ในเป้าหมายที่ 5 ระบุว่าให้มีแผนงานเร่งด่วนที่เป็นเป้าหมายในการพัฒนาปี พ.ศ. 2560-2561 นั่นคือ มีระบบและกลไกการทดสอบ การวัดและประเมินความรู้ ทักษะ และสมรรถนะของผู้เรียนทุกระดับการศึกษา และทุกกลุ่มเป้าหมายที่มีประสิทธิภาพ (สำนักงานเลขาธิการสภาการศึกษา กระทรวงศึกษาธิการ, 2560) นอกจากนี้ การวัดและประเมินผลการเรียนรู้ในศตวรรษที่ 21 ควรมีการประเมินเพื่อหาจุดแข็ง จุดอ่อน และวินิจฉัยปัญหาการเรียนรู้ของผู้เรียนเพื่อเป็นแนวทางในการปรับปรุงการเรียนการสอน และพัฒนากระบวนการเรียนรู้ของผู้เรียน (ศศิธร บัวทอง, 2560) ซึ่งการวัดและการประเมินผลการเรียนรู้ที่มีบทบาทมากในปัจจุบัน คือ การใช้แบบสอบหลายตัวเลือก (Multiple Choice) ในการ

วัดความรู้ความสามารถและผลสัมฤทธิ์ทางการเรียนของผู้เรียน (Chang, Lin, and Lin, 2007; รัตนา ไชยตรี, 2546) โดยแบบสอบหลายตัวเลือก (Multiple Choice) มีความสะดวกในการนำไปใช้หลายประการ เช่น วัดพฤติกรรมการเรียนรู้ได้หลากหลายและครอบคลุมเนื้อหา มีความเป็นปรนัยในการตรวจให้คะแนน ใช้เวลาในการตรวจน้อย สามารถนำผลการตอบไปวิเคราะห์ข้อมูลทางสถิติ เหมาะสำหรับการสอบที่มีผู้สอบจำนวนมาก เป็นต้น (โชติกา ภาษีผล ญัฏฐภรณ์ หลาวทอง และกมลวรรณ ตังธนากานนท์, 2558)

แบบสอบหลายตัวเลือกมีจุดอ่อนหลายประการ เช่น ข้อสอบออกยาก ใช้เวลาในการออกข้อสอบนาน ไม่สามารถวัดความคิดสร้างสรรค์ได้ และผู้เรียนสามารถเดาได้ ทำให้ความตรงลดลงเนื่องจากผู้เรียนมีโอกาสในการเดาและสามารถได้คะแนนสูงเนื่องจากความโชคดีในการเดา เพราะถ้ามี 4 ตัวเลือก ใน 1 ข้อ ผู้เรียนจะมีโอกาสในการตอบถูกอย่างน้อย 25% (Lau et al., 2011) ดังนั้น นักวิชาการด้านการวัดและประเมินผลการศึกษาจึงมีการทดสอบระหว่างแบบสอบหลายตัวเลือกและรูปแบบอื่นเพื่อแก้ปัญหาการเดาคำตอบของผู้เรียน เช่น แบบสอบตอบสั้น (Short-answer) และแบบสอบความเรียง (Essay) (Chang et al., 2007) โดยแบบสอบความเรียงเป็นรูปแบบที่สามารถวัดความรู้ความสามารถขั้นสูง ให้สิทธิเสรีภาพแก่ผู้เรียนในการตอบเพื่อแสดงความสามารถ ลดโอกาสในการเดา และเหมาะสำหรับการนำไปประเมินผลเพื่อแก้ปัญหา (โชติกา ภาษีผล ญัฏฐภรณ์ หลาวทอง และกมลวรรณ ตังธนากานนท์, 2558) จึงมีการศึกษาแบบสอบอัตนัยประยุกต์ (Modified essay question) จากงานวิจัยของ Palmer and Devitt (2007) และ Zafar-Khan and Aljarallah (2011) พบว่า แบบสอบอัตนัยประยุกต์สามารถประเมินความรู้ขั้นสูงได้จริงแต่ไม่สามารถประเมินความรู้พื้นฐานได้ ในขณะที่แบบสอบหลายตัวเลือกที่มีโครงสร้างที่ดีสามารถนำมาใช้แทนแบบสอบอัตนัยประยุกต์ได้ ดังนั้น แบบสอบอัตนัยจึงไม่เหมาะสำหรับการวัดและประเมินผลแบบอิงเกณฑ์ (Criterion-Referenced Evaluation) ที่ต้องใช้แบบสอบเพื่อการประเมินความรอบรู้ในเนื้อหาตามจุดประสงค์การเรียนรู้ (โชติกา ภาษีผล ญัฏฐภรณ์ หลาวทอง และกมลวรรณ ตังธนากานนท์, 2558) ซึ่งประกอบด้วยข้อสอบที่วัดความรู้ขั้นพื้นฐานและวัดความรู้ขั้นสูงเป็นไปตามจุดประสงค์การเรียนรู้ อย่างไรก็ตาม การวัดและประเมินผลของประเทศไทยเป็นไปตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน 2551 ซึ่งต้องประเมินตามตัวชี้วัดที่กำหนด (กระทรวงศึกษาธิการ, 2551) ในแต่ละตัวชี้วัดจะวัดความรู้ขั้นพื้นฐานและวัดความรู้ขั้นสูง ดังนั้น แบบสอบที่เหมาะสมในการวัดและประเมินผลการเรียนรู้ คือ แบบสอบหลายตัวเลือก (Multiple Choice) แต่แบบสอบหลายตัวเลือกในปัจจุบันมีการให้คะแนนแบบตอบถูกได้ 1 คะแนน และตอบผิดได้ 0 คะแนน ดังนั้น ผู้สอบจึงมีโอกาสในการเดาคำตอบได้ ซึ่งจะทำให้คะแนนที่ได้ไม่สะท้อนความสามารถที่แท้จริงของผู้เรียนและไม่สามารถให้สารสนเทศที่เพียงพอที่จำแนกผู้เรียนที่มีความรู้จริง (Full knowledge) มีความรู้บางส่วน (Partial knowledge) หรือไม่มีความรู้ (Absence of knowledge) (เอมอร์ จังศิริพรภรณ์, 2545b) จากข้อจำกัดของแบบสอบหลายตัวเลือกดังกล่าว จึงมีนักวิชาการได้ศึกษาวิธีการแก้ปัญหาการเดาคำตอบของผู้เรียนและสารสนเทศในการจำแนกผู้เรียน โดยใช้วิธีการให้คะแนนความรู้บางส่วน (Partial Knowledge) (โชติกา ภาษีผล, 2554) นอกจากจะแก้ปัญหาดังกล่าวแล้ว ยังทำให้ได้สารสนเทศในการวินิจฉัยความเข้าใจที่คลาดเคลื่อนหรือการไม่เข้าใจเนื้อหาของผู้เรียนได้ (Lau et al., 2011) ดังนั้น จึงเป็นที่มาของการศึกษาเกี่ยวกับวิธีการให้คะแนนความรู้บางส่วน (Partial Knowledge) ว่าวิธีใดมีประสิทธิภาพมากกว่ากัน

วิธีการให้คะแนนความรู้บางส่วนมีหลายวิธี ได้แก่ การให้คะแนนแก่ข้อสอบที่แตกต่างกัน (Differential Item Weighting) การให้น้ำหนักแก่ตัวเลือกที่แตกต่างกัน (Differential Option Weighting) การเปลี่ยนแปลงโครงสร้างของข้อสอบ (Changing the Item Structure) และการเปลี่ยนแปลงวิธีการตอบ (Changing the Response Method) การเปลี่ยนแปลงวิธีการตอบแบ่งเป็นวิธีการตอบและให้คะแนนแบบคูมพ์ (The

Coomb's response/scoring method) วิธีการประยุกต์วิธีของคูมบ์ วิธีการตอบและการให้คะแนนแบบอาร์โนลด์ (The Arnold Scoring) วิธีการตอบและการตรวจให้คะแนนแบบครอส (Cross Response/Scoring Method, CRS) วิธีการเลือกชุดย่อยของคำตอบ (The subset Selection Method, SST) วิธีประยุกต์การให้คะแนนความรู้บางส่วนของเดรสเซลและสมิท วิธีการทดสอบความน่าจะเป็น (Probability Testing) วิธีการแสดงความมั่นใจ (Confidence Marking) วิธีการจัดอันดับอย่างสมบูรณ์ (Complete Ordering) วิธีการจัดอันดับเป็นบางส่วน (Partial Ordering) และวิธีตัดตัวลงเลือกตัวถูก (The number right elimination testing: NRET) ผู้วิจัยสนใจศึกษาวิธีการประยุกต์การให้คะแนนของคูมบ์ และวิธีตัดตัวลงเลือกตัวถูก (The number right elimination testing: NRET) เนื่องจากวิธีประยุกต์วิธีของคูมบ์เป็นวิธีตัดตัวลงเลือกตัวถูก ซึ่งเป็นวิธีที่มีค่าอำนาจจำแนกเฉลี่ย ฟังก์ชันสารสนเทศของแบบสอบ ประสิทธิภาพสัมพัทธ์เฉลี่ยของแบบสอบ ความตรงเชิงโครงสร้างสูงกว่าวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท และวิธีประเพณีนิยม (สุพจน์ เกิดสุวรรณ, 2545; เอมอร์ จังศิริพรปรกรณ์, 2545ก) นอกจากนี้ยังพบว่า วิธีประยุกต์วิธีของคูมบ์มีคุณภาพดีกว่าวิธีประเพณีนิยม วิธีของคูมบ์ วิธีของอาร์โนลด์ วิธีประยุกต์ของอาร์โนลด์ วิธีของ เดรสเซลและสมิท และวิธีประยุกต์ของเดรสเซลและสมิท (สุพจน์ เกิดสุวรรณ, 2545) ส่วนวิธีตัดตัวลงเลือก ตัวถูก (The number right elimination testing: NRET) เป็นวิธีใหม่ที่ผสมระหว่างการตัดตัวลงและการให้คะแนนแบบประเพณีนิยมซึ่งมีความเที่ยงสูง และจากงานวิจัยพบว่า วิธีนี้เป็นวิธีที่มีโอกาสในการเดาน้อย สามารถวัดความรู้ที่คลาดเคลื่อนของนักเรียนและสามารถตรวจสอบความรู้บางส่วนได้ (Lau et al., 2011) นอกจากนี้ งานวิจัยของ Lau et al. (2014) ได้เปรียบเทียบความแกร่งของวิธีตัดตัวลงเลือกตัวถูก วิธีประเพณีนิยมและวิธีการตัดตัวลง พบว่า วิธีตัดตัวลงเลือกตัวถูกมีประสิทธิภาพในการประมาณค่าความสามารถ และความเที่ยงสูงกว่าวิธีประเพณีนิยม และวิธีการตัดตัวลง และมีความคลาดเคลื่อนมาตรฐานของการวัดต่ำที่สุด

นอกจากนี้ ความยาวของแบบสอบยังส่งผลต่อคุณภาพของแบบสอบ โดยมีผลกระทบต่อความแปรปรวนของคะแนนจริงและความแปรปรวนของคะแนนที่สังเกตได้ (ศิริชัย กาญจนวาสี, 2556) การเพิ่มความยาวของแบบสอบจะช่วยเพิ่มค่าสัมประสิทธิ์ความเที่ยงให้สูงขึ้น ซึ่งมีนักวิจัยจำนวนมากศึกษาเกี่ยวกับความยาวของแบบสอบที่เหมาะสม เช่น จากการสังเคราะห์งานวิจัยทั้งหมด 27 งานวิจัย พบว่าค่าเฉลี่ยของจำนวนข้อสอบในแบบสอบแต่ละฉบับ เท่ากับ 43 ข้อ (Rodriguez, 2005) ดังนั้น จึงมีการศึกษาเกี่ยวกับความยาวของแบบสอบที่ทำให้แบบสอบมีคุณภาพอยู่ในระดับที่ยอมรับได้ รวมถึงศึกษาความเที่ยงและฟังก์ชันสารสนเทศของแบบสอบที่ได้จากวิธีประยุกต์วิธีของคูมบ์ และวิธีตัดตัวลงเลือกตัวถูกว่าแตกต่างกันหรือไม่ อย่างไร เมื่อมีจำนวนข้อสอบ 20 ข้อ 30 ข้อ และ 40 ข้อ เพื่อนำไปสู่การเลือกใช้วิธีการให้คะแนนความรู้บางส่วนที่มีคุณภาพภายใต้เงื่อนไขจำนวนข้อสอบที่แตกต่างกัน

จากปัญหาที่กล่าวมาข้างต้นทำให้ผู้วิจัยสนใจศึกษาประสิทธิภาพของวิธีการตรวจให้คะแนนความรู้บางส่วน ได้แก่ วิธีการประยุกต์วิธีของคูมบ์ และวิธีตัดตัวลงเลือกตัวถูก เมื่อมีจำนวนข้อสอบต่างกัน เพื่อเป็นแนวทางในการวัดและประเมินผลการเรียนรู้ของผู้เรียนที่สะท้อนความสามารถที่แท้จริงของผู้เรียนอย่างแม่นยำ และให้สารสนเทศในการตัดสินใจเลือกลักษณะข้อสอบที่เหมาะสมกับการวัดผลการเรียนรู้ของผู้เรียน เช่น จำนวนข้อที่เหมาะสมกับวิธีการให้คะแนนความรู้บางส่วน

คำถามการวิจัย

1. ค่าประมาณความเที่ยงของแบบสอบที่ตรวจให้คะแนนด้วยวิธีการประยุกต์วิธีของคูมบ์และวิธีตัดตัวลงเลือกตัวถูกแตกต่างกันหรือไม่ อย่างไร เมื่อมีจำนวนข้อสอบ 20 ข้อ 30 ข้อ และ 40 ข้อ

2. ความสัมพันธ์ระหว่างความสามารถจริงกับความสามารถที่ได้จากการประมาณค่าจากวิธีการประยุกต์ของคัมภ์และวิธีตัดตัวลงเลือกตัวถูกแตกต่างกันหรือไม่ อย่างไร เมื่อมีจำนวนข้อสอบ 20 ข้อ 30 ข้อ และ 40 ข้อ

3. ฟังก์ชันสารสนเทศของแบบสอบที่ได้จากวิธีการประยุกต์ของคัมภ์และวิธีตัดตัวลงเลือกตัวถูกแตกต่างกันหรือไม่ อย่างไร เมื่อมีจำนวนข้อสอบ 20 ข้อ 30 ข้อ และ 40 ข้อ

วัตถุประสงค์การวิจัย

1. เพื่อเปรียบเทียบค่าประมาณความเที่ยงของแบบสอบที่ตรวจให้คะแนนด้วยวิธีการประยุกต์วิธีของคัมภ์และวิธีตัดตัวลงเลือกตัวถูก เมื่อมีจำนวนข้อสอบ 20 ข้อ 30 ข้อ และ 40 ข้อ

2. เพื่อเปรียบเทียบความสัมพันธ์ระหว่างความสามารถจริงกับความสามารถที่ประมาณค่าที่ได้จากวิธีการประยุกต์วิธีของคัมภ์และวิธีตัดตัวลงเลือกตัวถูก เมื่อมีจำนวนข้อสอบ 20 ข้อ 30 ข้อ และ 40 ข้อ

3. เพื่อเปรียบเทียบฟังก์ชันสารสนเทศของแบบสอบที่ได้จากวิธีการประยุกต์วิธีของคัมภ์และวิธีตัดตัวลงเลือกตัวถูก เมื่อมีจำนวนข้อสอบ 20 ข้อ 30 ข้อ และ 40 ข้อ

สมมติฐานการวิจัย

1. จากงานวิจัยของรัตน ไชยตรี (2546) และสุพจน์ เกิดสุวรรณ (2545) ได้ศึกษาความเที่ยงของแบบสอบ พบว่า วิธีการให้คะแนนความรู้บางส่วนแต่ละวิธีมีความเที่ยงแตกต่างกันอย่างมีนัยสำคัญทางสถิติ และงานวิจัยของ Lau, Hong, Lau, and Usop (2014) ได้เปรียบเทียบประสิทธิภาพของวิธีตัดตัวลงเลือกตัวถูก วิธีการให้คะแนนความรู้บางส่วนของคัมภ์ และวิธีแบบประเพณีนิยม พบว่า วิธีตัดตัวลงเลือกตัวถูกมีความเที่ยงสูงกว่าวิธีของคัมภ์ และวิธีแบบประเพณีนิยม ดังนั้นผู้วิจัยจึงตั้งสมมติฐานการวิจัย ดังนี้

วิธีตัดตัวลงเลือกตัวถูกน่าจะมีความเที่ยงสูงกว่าวิธีการประยุกต์วิธีของคัมภ์ทั้งแบบสอบที่มีจำนวนข้อสอบ 20 ข้อ 30 ข้อ และ 40 ข้อ

2. จากการศึกษางานวิจัยเกี่ยวกับวิธีการประยุกต์วิธีของคัมภ์ พบว่า ช่วงการให้คะแนนอยู่ระหว่าง 0 – 3 คะแนน (เอมอร์ จังศิริพรกรณ์, 2545) ส่วนวิธีตัดตัวลงเลือกตัวถูก มีช่วงการให้คะแนนอยู่ระหว่าง (-3) – 4 คะแนน ซึ่งวิธีนี้มีความสามารถในการจำแนกผู้เรียนตามระดับความสามารถในช่วงที่กว้างกว่า และมีความสามารถในการประมาณค่าความสามารถของผู้สอบได้ดีกว่าวิธีการให้คะแนนความรู้บางส่วนของคัมภ์และวิธีประเพณีนิยม (Lau, Hong, Lau, and Usop, 2014) ดังนั้น ผู้วิจัยจึงตั้งสมมติฐานการวิจัย ดังนี้

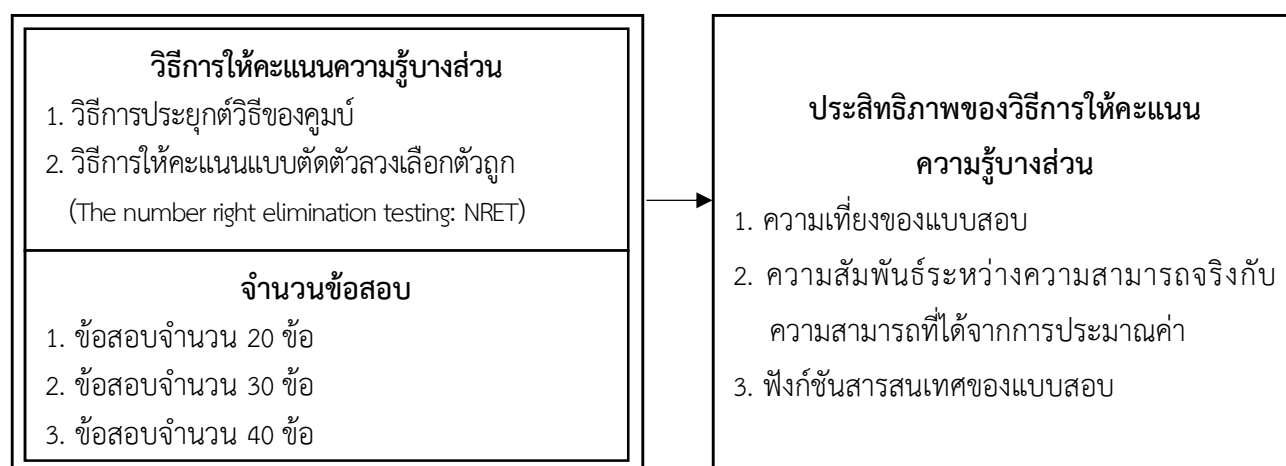
วิธีตัดตัวลงเลือกตัวถูก และวิธีการประยุกต์วิธีของคัมภ์น่าจะมีความสัมพันธ์ระหว่างความสามารถจริงกับความสามารถที่ได้จากการประมาณค่าแตกต่างกันทั้งแบบสอบที่มีจำนวนข้อสอบ 20 ข้อ 30 ข้อ และ 40 ข้อ

3. จากการศึกษางานวิจัยเกี่ยวกับฟังก์ชันสารสนเทศของแบบสอบที่มีการให้คะแนนความรู้บางส่วน การวิเคราะห์ฟังก์ชันสารสนเทศของแบบสอบตามความสามารถ พบว่า วิธีการให้คะแนนความรู้บางส่วนที่ต่างกันจะให้ฟังก์ชันสารสนเทศของแบบสอบแตกต่างกัน (สุพจน์ เกิดสุวรรณ, 2545; เอมอร์ จังศิริพรกรณ์, 2545a) เนื่องจากฟังก์ชันสารสนเทศของแบบสอบเป็นผลมาจากฟังก์ชันสารสนเทศของข้อสอบซึ่งเป็นดัชนีที่ประกอบด้วยพารามิเตอร์ความยาก อำนาจจำแนก และค่าความแปรปรวนของคะแนนรายข้อ (ศิริชัย กาญจนวาสี, 2555) ถ้าวิธีการให้คะแนนแตกต่างกันจะทำให้ความยากและอำนาจจำแนกแตกต่างกัน ซึ่งก็จะส่งผลต่อฟังก์ชันสารสนเทศของแบบสอบด้วยเช่นกัน นอกจากนี้ Lau, Lau, Hong, & Usop (2014) กล่าวว่า

วิธีตัดตัววงเลือกตัวถูก สามารถใช้แบบสอบที่มีความยาวข้อสอบสั้นได้ และยังคงมีประสิทธิภาพในการประมาณค่าความสามารถของนักเรียนและยังทำให้ความเที่ยงอยู่ในเกณฑ์ที่ยอมรับได้ ตัวอย่างเช่น ข้อสอบจำนวน 40 ข้อ สามารถลดจำนวนข้อสอบให้เหลือประมาณ 35-38 ข้อ ได้ เมื่อใช้วิธีตัดตัววงเลือกตัวถูก ดังนั้น ผู้วิจัยจึงตั้งสมมติฐานการวิจัย ดังนี้

วิธีตัดตัววงเลือกตัวถูก และวิธีการประยุกต์วิธีของคุมบ์น่าจะมีค่าฟังก์ชันสารสนเทศของแบบสอบแตกต่างกัน ทั้งแบบสอบที่มีจำนวนข้อสอบ 20 ข้อ 30 ข้อ และ 40 ข้อ

กรอบแนวคิดการวิจัย



ภาพที่ 1 กรอบแนวคิดการวิจัย

วิธีดำเนินการวิจัย

1. กลุ่มเป้าหมายที่ใช้ในการวิจัย

การวิจัยเรื่องนี้ใช้การจำลองข้อมูลด้วยโปรแกรม R ซึ่งจำลองผู้สอบจำนวน 1,000 คน

2. รูปแบบการวิจัย

การวิจัยนี้ใช้การจำลองข้อมูลโดยจำลองผู้สอบจำนวน 1,000 คน ทั้งนี้ผู้สอบแต่ละคนทำแบบสอบแบบหลายตัวเลือกจำนวน 4 ตัวเลือก โดยที่แบบสอบมีข้อสอบจำนวน 20 ข้อ 30 ข้อ และ 40 ข้อ และมีการให้คะแนนด้วยวิธีประยุกต์วิธีของคุมบ์ และวิธีตัดตัววงเลือกตัวถูก มีการกระทำซ้ำ 10 รอบ จากนั้น จำลองรูปแบบการตอบคำถามของผู้สอบแต่ละคนตามโมเดล Generalized Partial Credit Model (G-PCM) สำหรับแบบสอบที่มีจำนวนข้อสอบ 20 ข้อ 30 ข้อ และ 40 ข้อ

วิธีการให้คะแนนความรู้บางส่วน

วิธีการตรวจให้คะแนนความรู้บางส่วน หมายถึง การให้คะแนนที่ใกล้เคียงกับความสามารถที่แท้จริงของผู้เรียน ซึ่งในงานวิจัยนี้ศึกษาวิธีการตรวจให้คะแนนความรู้บางส่วน 2 วิธี ได้แก่ วิธีการประยุกต์วิธีของคุมบ์ และวิธีตัดตัววงเลือกตัวถูก

1. วิธีการประยุกต์วิธีของคุมบ์ เป็นวิธีที่พัฒนามาจากวิธีให้คะแนนแบบคุมบ์ดั้งเดิม นั่นคือ มีคำตอบถูกเพียงคำตอบเดียว โดยให้ผู้สอบตัดตัววงที่ไม่ใช่คำตอบถูกออกไปให้มากที่สุด และปรับไม่ให้คะแนนมีค่าติดลบ โดยวิธีนี้เป็นวิธีที่หลีกเลี่ยงการให้คะแนนแบบลงโทษ เนื่องจากคะแนนติดลบมีผลกระทบต่อคะแนน

ในข้ออื่นและคะแนนรวม ทำให้ไม่ได้สารสนเทศที่สะท้อนความรู้ที่แท้จริงของผู้เรียนและไม่สามารถจำแนกผู้เรียนได้ ซึ่งมีสูตรการให้คะแนน ดังนี้

$$X = nC \quad (\text{เอมอร์ จังศิริพรปกรณ์, 2545ก})$$

เมื่อ	X	คือ	คะแนนที่ได้จากการตอบข้อสอบ 1 ข้อ
	n	คือ	จำนวนตัวลวงที่ถูกตัดออก
	C	คือ	ค่าความถูกต้องของการตัดตัวลวง ซึ่งมี 2 ค่า ดังนี้ C = 1 เมื่อไม่มีตัวเลือกถูกอยู่ในตัวลวงที่ตัดออก C = 0 เมื่อมีตัวเลือกถูกอยู่ในตัวลวงที่ตัดออก

ตารางที่ 1 การให้คะแนนด้วยวิธีการประยุกต์วิธีของคัมป์ เมื่อข้อสอบมี 4 ตัวเลือก

รูปแบบการตอบ	จำนวนตัวเลือกที่ผู้สอบเลือก	ลักษณะการตอบ	จำนวนตัวเลือกที่เลือกถูกตรงกับคำตอบ	คะแนน
1	3	FFF	ตัดตัวลวงตรงกับจำนวนตัวลวงในข้อนั้น 3 ตัว	3
2	2	FF	ตัดตัวลวงตรงกับจำนวนตัวลวงในข้อนั้น 2 ตัว	2
3	1	F	ตัดตัวลวงตรงกับจำนวนตัวลวงในข้อนั้น 1 ตัว	1
4	3	TFF	ตัดตัวลวงตรงกับจำนวนตัวลวงในข้อนั้น 2 ตัว	0
5	2	TF	ตัดตัวลวงตรงกับจำนวนตัวลวงในข้อนั้น 1 ตัว	0
6	1	T	ตัดตัวลวงไม่ตรงกับจำนวนตัวลวงในข้อนั้น	0

หมายเหตุ F คือ ตัวเลือกที่เป็นตัวลวง, T คือ ตัวเลือกที่เป็นตัวถูก (ปรับปรุงจาก โชติกา ภาชีผล, 2554)

2. วิธีตัดตัวลวงเลือกตัวถูก (The number right elimination testing: NRET) เป็นวิธีที่ผสมผสานระหว่างวิธีการตัดตัวลวงของคัมป์ และวิธีการให้คะแนนแบบประเพณีนิยม โดยให้ผู้สอบตัดตัวลวงที่แน่ใจว่าไม่ใช่คำตอบออก เลือก 1 ตัวเลือกที่เป็นคำตอบที่ถูกต้อง โดยตัวเลือกที่ไม่แน่ใจให้ตอบ “ไม่แน่ใจ” ถ้าข้อสอบมี 4 ตัวเลือก จะมีช่วงคะแนนอยู่ระหว่าง (-3) – 4 คะแนน

การให้คะแนน (Lau, Hong, Lau, and Usop, 2014)

- | | |
|--|--------------------|
| 1) เลือกคำตอบที่ถูกต้อง | จะได้ 1 คะแนน |
| 2) ตัดตัวลวงถูกต้อง 1 ตัว | จะได้ 1 คะแนน |
| 3) ถ้าตัดตัวถูกออกโดยเข้าใจว่าเป็นตัวลวง | จะถูกหัก k-1 คะแนน |
| 4) ถ้าตอบ “ไม่แน่ใจ” | จะได้ 0 คะแนน |

การแปลความหมายคะแนน เมื่อข้อสอบมี 4 ตัวเลือก

- | | |
|---------------|---|
| 4 คะแนน | หมายถึง ผู้สอบมีความรู้สมบูรณ์ |
| 1, 2, 3 คะแนน | หมายถึง ผู้สอบมีความรู้บางส่วน |
| 0 คะแนน | หมายถึง ผู้สอบไม่มีความรู้ |
| -1, -2 คะแนน | หมายถึง ผู้สอบมีความรู้บางส่วนคลาดเคลื่อน |
| -3 คะแนน | หมายถึง ผู้สอบมีความรู้คลาดเคลื่อนทั้งหมด |

ตารางที่ 2 การให้คะแนนแบบตัดตัววงเลือกตัวถูก เมื่อข้อสอบมี 4 ตัวเลือก

รูปแบบการตอบ	ลักษณะการตอบและการตรวจให้คะแนน						คะแนนรวม
	จำนวนตัวลวงที่ตัด	คะแนนที่ได้	ตัวถูก	คะแนนที่ได้	จำนวนตัวเลือกที่ไม่แน่ใจ	คะแนนที่ได้	
1	3	3	เลือก	1	0	0	4
2	2	2	เลือก	1	1	0	3
3	1	1	เลือก	1	2	0	2
4	0	0	เลือก	1	3	0	1
5	3	3	ตัด	-3	0	0	0
6	2	2	ตัด	-3	1	0	-1
7	1	1	ตัด	-3	2	0	-2
8	0	0	ตัด	-3	3	0	-3
9	3	3	ไม่แน่ใจ	0	0	0	3
10	2	2	ไม่แน่ใจ	0	1	0	2
11	1	1	ไม่แน่ใจ	0	2	0	1
12	0	0	ไม่แน่ใจ	0	3	0	0

3. การวิเคราะห์ข้อมูล

การวิจัยเรื่องนี้วิเคราะห์ข้อมูลโดยใช้การจำลองข้อมูลด้วยโปรแกรม R ซึ่งมีขั้นตอนในการวิเคราะห์ข้อมูล ดังนี้

1. จำลองผู้สอบจำนวน 1,000 คน เนื่องจากการวิเคราะห์ข้อมูลโดยใช้ทฤษฎีการตอบสนองข้อสอบ (IRT) ในการวิเคราะห์พารามิเตอร์ต่างๆ ควรใช้ขนาดตัวอย่างวิจัยไม่ต่ำกว่า 500 คน (Liang and Wells, 2009; ศิริชัย กาญจนวาสี, 2555) เพราะถ้าตัวอย่างวิจัยน้อยจะไม่สามารถประมาณค่าพารามิเตอร์ได้ หรือจะประมาณค่าพารามิเตอร์โดยมีความคลาดเคลื่อนสูง และจะทำให้การวิเคราะห์ไม่น่าเชื่อถือ ดังนั้น ผู้วิจัยจึงจำลองผู้สอบจำนวน 1,000 คน เนื่องจากการวิเคราะห์ข้อมูลโดยใช้ทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่า 2 ค่า (Polytomous IRT Models) และเพื่อประมาณค่าพารามิเตอร์ได้อย่างแม่นยำ ซึ่งระดับความสามารถของผู้สอบมีการแจกแจงแบบโค้งปกติที่มีค่าเฉลี่ย เท่ากับ 0 และมีส่วนเบี่ยงเบนมาตรฐาน เท่ากับ 1 โดยใช้คำสั่ง `mnorm` ทั้งนี้ผู้สอบแต่ละคนทำแบบสอบแบบหลายตัวเลือกจำนวน 4 ตัวเลือก โดยที่แบบสอบมีข้อสอบจำนวน 20 ข้อ 30 ข้อ และ 40 ข้อ และมีการให้คะแนนด้วยวิธีประยุกต์วิธีของคุมบ์ และวิธีตัดตัววงเลือกตัวถูก และมีการกระทำซ้ำ 10 รอบ

2. จำลองรูปแบบการตอบคำถามของผู้สอบแต่ละคนตามโมเดล Generalized Partial Credit Model (G-PCM) สำหรับแบบสอบที่มีจำนวนข้อสอบ 20 ข้อ 30 ข้อ และ 40 ข้อ ซึ่งให้คะแนนด้วยวิธีประยุกต์วิธีของคุมบ์ และวิธีตัดตัววงเลือกตัวถูก โดยใช้คำสั่ง `simdata` ทั้งนี้ในการจำลองรูปแบบการตอบคำถามนั้น ผู้วิจัยจำลองให้ข้อสอบแต่ละข้อมีค่าพารามิเตอร์ระดับความยากของขั้นการตอบ (δ_{ij}) มีค่าสูงขึ้นตามขั้นการตอบ เนื่องจากขั้นการตอบที่สูงขึ้น ผู้สอบย่อมต้องมีความสามารถที่สูงขึ้นเพื่อที่จะได้คะแนนในรายการคำตอบที่

สูงขึ้น และค่าพารามิเตอร์ความชัน (α_i) มีค่าอยู่ระหว่าง 0.50 ถึง 2.50 ซึ่งมีการแจกแจงแบบโค้งปกติที่มีค่าเฉลี่ย เท่ากับ 0 และมีส่วนเบี่ยงเบนมาตรฐาน เท่ากับ 1

3. วิเคราะห์ผลการตอบคำถามของผู้สอบโดยใช้โมเดล Generalized Partial Credit Model (G-PCM) ซึ่งใช้คำสั่ง gpcm เพื่อนำผลการวิเคราะห์ข้อมูลไปประมาณค่าความสามารถของผู้สอบโดยใช้คำสั่ง factor.scores สำหรับแบบสอบที่มีจำนวนข้อสอบ 20 ข้อ 30 ข้อ และ 40 ข้อ ซึ่งให้คะแนนด้วยวิธีประยุกต์วิธีของคุมบ์และวิธีตัดตัวลงเลือกตัวถูก

4. วิเคราะห์ความเที่ยงแบบความสอดคล้องภายในวิธีสัมประสิทธิ์แอลฟาครอนบาคของแบบสอบที่มีจำนวนข้อสอบ 20 ข้อ 30 ข้อ และ 40 ข้อ ซึ่งให้คะแนนด้วยวิธีประยุกต์วิธีของคุมบ์ และวิธีตัดตัวลงเลือกตัวถูก โดยใช้คำสั่ง alpha

5. วิเคราะห์ความสัมพันธ์ระหว่างความสามารถจริงกับความสามารถที่ได้จากการประมาณค่าสำหรับแบบสอบที่มีจำนวนข้อสอบ 20 ข้อ 30 ข้อ และ 40 ข้อ ซึ่งให้คะแนนด้วยวิธีประยุกต์วิธีของคุมบ์ และวิธีตัดตัวลงเลือกตัวถูก โดยใช้คำสั่ง cor

6. วิเคราะห์ฟังก์ชันสารสนเทศของแบบสอบที่มีจำนวนข้อสอบ 20 ข้อ 30 ข้อ และ 40 ข้อ ซึ่งให้คะแนนด้วยวิธีประยุกต์วิธีของคุมบ์ และวิธีตัดตัวลงเลือกตัวถูก โดยใช้คำสั่ง information พร้อมทั้งสร้างโค้งสารสนเทศของแบบสอบเพื่อพิจารณาความแตกต่างระหว่างสารสนเทศของแบบสอบระหว่างวิธีการให้คะแนนความรู้บางส่วนทั้ง 2 วิธี ที่มีจำนวนข้อสอบต่างกัน

7. วิเคราะห์ความแตกต่างระหว่างความเที่ยงของแบบสอบและความสัมพันธ์ระหว่างความสามารถจริงกับความสามารถที่ได้จากการประมาณค่าเมื่อตรวจให้คะแนนด้วยวิธีประยุกต์วิธีของคุมบ์ และวิธีตัดตัวลงเลือกตัวถูก ซึ่งมีจำนวนข้อสอบ 20 ข้อ 30 ข้อ และ 40 ข้อ โดยใช้การวิเคราะห์ความแปรปรวนพหุนามแบบสองทาง (two-way MANOVA)

สรุปและอภิปรายผลการวิจัย

สรุปผลการวิจัย

1. การวิเคราะห์ความแตกต่างระหว่างความเที่ยงของแบบสอบและความสัมพันธ์ระหว่างความสามารถจริงกับความสามารถที่ได้จากการประมาณค่า

การวิเคราะห์ความแตกต่างระหว่างความเที่ยงของแบบสอบและความสัมพันธ์ระหว่างความสามารถจริงกับความสามารถที่ได้จากการประมาณค่าเมื่อตรวจให้คะแนนด้วยวิธีการประยุกต์วิธีของคุมบ์ และวิธีตัดตัวลงเลือกตัวถูก ซึ่งมีจำนวนข้อสอบ 20 ข้อ 30 ข้อ และ 40 ข้อ ผู้วิจัยได้ตรวจสอบข้อตกลงเบื้องต้น 3 ประการ ได้แก่ (1) การทดสอบความเท่ากันของเมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมของประชากรโดยใช้สถิติทดสอบ Box's test of equality of covariance matrices พบว่า เมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมระหว่างกลุ่มประชากรมีความแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 (Box's M = 14.93, df1 = 12, df2 = 15008.82, p-value = .34) (2) ทดสอบความเป็นเอกพันธ์ของความแปรปรวนระหว่างกลุ่มประชากรโดยใช้สถิติทดสอบ Levene's test of equality of error variances พบว่า ความแปรปรวนของความเที่ยงของแบบสอบ และความสัมพันธ์ระหว่างความสามารถจริงกับความสามารถที่ได้จากการประมาณค่าในประชากรแต่ละกลุ่มมีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 และ (3) ทดสอบความสัมพันธ์ระหว่างตัวแปรตาม พบว่า ความเที่ยงของแบบสอบและความสัมพันธ์ระหว่างความสามารถจริงกับความสามารถที่ได้จากการประมาณค่ามีความสัมพันธ์กันทางบวกในระดับสูง ($r = .90$,

p-value < .001) จากนั้นจึงวิเคราะห์ความแตกต่างระหว่างความเที่ยงของแบบสอบและความสัมพันธ์ระหว่างความสามารถจริงกับความสามารถที่ได้จากการประมาณค่าเมื่อมีวิธีการให้คะแนนความรู้บางส่วนและจำนวนข้อสอบแตกต่างกัน ผลการวิเคราะห์ข้อมูลสรุปได้ดังนี้

1) การตรวจให้คะแนนด้วยวิธีการประยุกต์วิธีของคุมป์ และวิธีตัดตัวลงเลือกตัวถูก และแบบสอบที่มีจำนวนข้อสอบ 20 ข้อ 30 ข้อ และ 40 ข้อ มีค่าความเที่ยงของแบบสอบแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

2) การตรวจให้คะแนนด้วยวิธีการประยุกต์วิธีของคุมป์ และวิธีตัดตัวลงเลือกตัวถูก และแบบสอบที่มีจำนวนข้อสอบ 20 ข้อ 30 ข้อ และ 40 ข้อ มีค่าความสัมพันธ์ระหว่างความสามารถจริงกับความสามารถที่ได้จากการประมาณค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

ตารางที่ 3 ผลการวิเคราะห์ความแปรปรวนพหุนามแบบสองทาง

แหล่งความแปรปรวน	Wilks' lambda	F	Hypothesis df	Error df	p
วิธีการให้คะแนน	.10	240.34	2.00	55.00	<.001
จำนวนข้อสอบ	.18	37.17	4.00	110.00	<.001

ตารางที่ 4 ผลการวิเคราะห์ความแปรปรวนพหุนามแบบสองทางเมื่อพิจารณาวิธีการให้คะแนนและจำนวนข้อสอบ

ตัวแปรตาม	แหล่งความแปรปรวน	SS	df	MS	F	p-value
ความเที่ยง	วิธีการให้คะแนน	0.01	1	0.01	301.29	<.001
	ความคลาดเคลื่อน	0.002	56	<.001		
	รวม	56.45	60			
ความสัมพันธ์	วิธีการให้คะแนน	0.09	1	0.09	254.34	<.001
	ความคลาดเคลื่อน	0.02	56	<.001		
	รวม	42.29	60			
ความเที่ยง	จำนวนข้อสอบ	0.01	2	0.002	68.99	<.001
	ความคลาดเคลื่อน	0.002	56	<.001		
	รวม	56.45	60			
ความสัมพันธ์	จำนวนข้อสอบ	0.05	2	0.02	72.44	<.001
	ความคลาดเคลื่อน	0.002	56	<.001		
	รวม	56.45	60			

จากการวิเคราะห์ภายหลังพหุคูณวิธี Bonferroni พบว่า วิธีตัดตัวลงเลือกตัวถูกมีค่าความเที่ยงของแบบสอบ และความสัมพันธ์ระหว่างความสามารถจริงกับความสามารถที่ได้จากการประมาณค่าสูงกว่าวิธีการประยุกต์วิธีของคุมป์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ทั้งนี้แบบสอบที่มีข้อสอบจำนวน 30 ข้อ มีค่าความเที่ยงของแบบสอบ และความสัมพันธ์ระหว่างความสามารถจริงกับความสามารถที่ได้จากการประมาณค่าสูงกว่าแบบสอบที่มีข้อสอบ 20 ข้อ ส่วนแบบสอบที่มีข้อสอบจำนวน 40 ข้อ มีค่าความเที่ยงของแบบสอบ และ

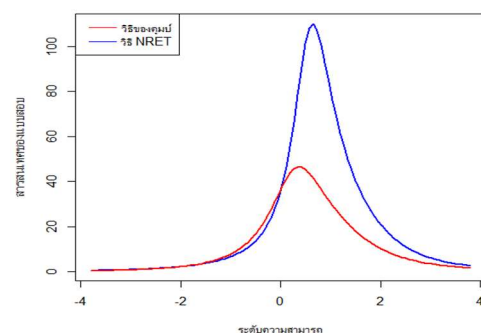
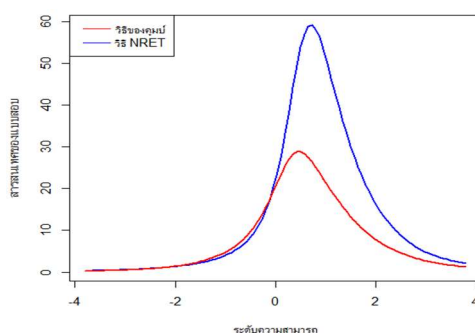
ความสัมพันธ์ระหว่างความสามารถจริงกับความสามารถที่ได้จากการประมาณค่าสูงกว่าแบบสอบที่มีข้อสอบ 30 ข้อ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

2. ผลการเปรียบเทียบฟังก์ชันสารสนเทศของแบบสอบที่ได้จากวิธีการประยุกต์วิธีของคัมป์และวิธีตัดตัวลงเลือกตัวถูก

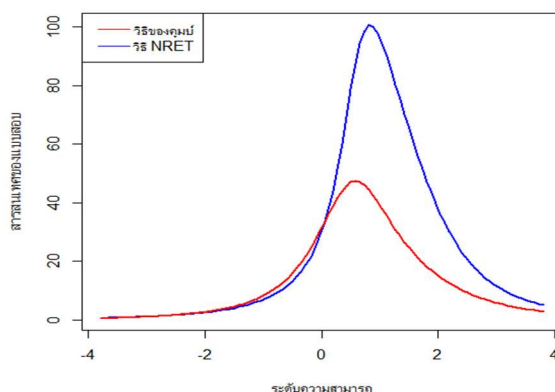
จากการวิเคราะห์ฟังก์ชันสารสนเทศของแบบสอบพบว่าฟังก์ชันสารสนเทศของแบบสอบที่ได้จากวิธีการประยุกต์วิธีของคัมป์และวิธีตัดตัวลงเลือกตัวถูก เมื่อแบบสอบมีจำนวนข้อสอบ 20 ข้อ 30 ข้อ และ 40 ข้อ มีค่ามากเมื่อผู้สอบมีความสามารถค่อนข้างสูง ($\theta = 0-2$) และจากการเปรียบเทียบฟังก์ชันสารสนเทศของแบบสอบที่ได้จากการให้คะแนนแบบความรู้บางส่วนทั้ง 2 วิธี พบว่าวิธีตัดตัวลงเลือกตัวถูกมีฟังก์ชันสารสนเทศของแบบสอบสูงกว่าวิธีการประยุกต์วิธีของคัมป์โดยมีสารสนเทศของแบบสอบแตกต่างกันมากเมื่อผู้สอบมีความสามารถระดับปานกลางและระดับสูงทั้งแบบสอบที่มีจำนวนข้อสอบ 20 ข้อ 30 ข้อ และ 40 ข้อ ดังภาพที่ 2-3 นอกจากนี้ ตารางที่ 5 ยังแสดงค่าฟังก์ชันสารสนเทศของแบบสอบแต่ละค่าความสามารถสำหรับวิธีการให้คะแนนความรู้บางส่วนทั้ง 2 วิธี เมื่อมีจำนวนข้อสอบต่างกัน

ตารางที่ 5 สารสนเทศของแบบสอบที่ได้จากวิธีการประยุกต์วิธีของคัมป์และวิธีตัดตัวลงเลือกตัวถูก เมื่อมีจำนวนข้อสอบต่างกัน

จำนวนข้อสอบ	วิธีการให้คะแนนความรู้บางส่วน	สารสนเทศของแบบสอบ								
		ระดับความสามารถ								
		-4 ถึง -3	-3 ถึง -2	-2 ถึง -1	-1 ถึง 0	0 ถึง 1	1 ถึง 2	2 ถึง 3	3 ถึง 4	รวม
20 ข้อ	วิธีประยุกต์ของคัมป์	0.34	0.9	2.74	10.72	25.92	13.78	4.83	1.74	60.97
	วิธี NRET	0.41	0.92	2.42	9.85	47.25	31.54	9.39	2.98	104.76
30 ข้อ	วิธีประยุกต์ของคัมป์	0.51	1.36	4.35	18.58	41.01	18.84	6.28	2.21	93.14
	วิธี NRET	0.64	1.46	3.95	16.05	83.11	43.76	11.97	3.81	164.75
40 ข้อ	วิธีประยุกต์ของคัมป์	0.75	1.81	4.98	17.26	42.56	25.44	9.71	3.9	106.41
	วิธี NRET	0.82	1.77	4.35	14.92	73.24	65.85	21.97	7.16	190.08



ภาพที่ 2 โคงสารสนเทศของแบบสอบที่ได้จากวิธีการประยุกต์วิธีของคัมป์ และวิธีตัดตัวลงเลือกตัวถูก เมื่อมีข้อสอบจำนวน 20 ข้อ (ซ้าย) และ 30 ข้อ (ขวา)



ภาพที่ 3 โค้งสารสนเทศของแบบสอบที่ได้จากวิธีการประยุกต์วิธีของคูมบ์ และวิธีตัดตัวลงเลือกตัวถูก เมื่อมีข้อสอบจำนวน 40 ข้อ

จากผลการวิเคราะห์ข้อมูลข้างต้นสรุปได้ว่า วิธีตัดตัวลงเลือกตัวถูกมีประสิทธิภาพมากกว่าวิธีการประยุกต์วิธีของคูมบ์ เนื่องจากมีค่าฟังก์ชันสารสนเทศของแบบสอบสูงกว่า รวมทั้งสามารถประมาณค่าความสามารถของผู้สอบได้แม่นยำมากกว่าวิธีการประยุกต์วิธีของคูมบ์ ทั้งแบบสอบที่มีจำนวนข้อสอบ 20 ข้อ, 30 ข้อ และ 40 ข้อ

อภิปรายผลการวิจัย

จากผลการวิเคราะห์ข้อมูลและข้อสรุปผลการวิจัย ผู้วิจัยนำเสนอประเด็นการอภิปรายที่น่าสนใจ ดังนี้

1. การเปรียบเทียบความเที่ยงของแบบสอบเมื่อให้คะแนนด้วยวิธีการประยุกต์วิธีของคูมบ์และวิธีตัดตัวลงเลือกตัวถูก

จากการวิเคราะห์ความเที่ยงแบบความสอดคล้องภายในวิธีสัมประสิทธิ์แอลฟาครอนบาคของแบบสอบที่มีข้อสอบจำนวน 20 ข้อ 30 ข้อ และ 40 ข้อ พบว่า ค่าความเที่ยงที่ได้มีค่าสูงกว่า .70 ซึ่งความเที่ยงอยู่ในเกณฑ์ที่ยอมรับได้ (ศิริชัย กาญจนวาสี, 2556) และเมื่อเปรียบเทียบความเที่ยงของแบบสอบเมื่อให้คะแนนด้วยวิธีการประยุกต์วิธีของคูมบ์และวิธีตัดตัวลงเลือกตัวถูก พบว่า วิธีตัดตัวลงเลือกตัวถูกมีความเที่ยงสูงกว่าวิธีการประยุกต์วิธีของคูมบ์เล็กน้อย เนื่องจากวิธีตัดตัวลงเลือกตัวถูกมีช่วงการให้คะแนนอยู่ระหว่าง -3 ถึง 4 คะแนน ส่วนวิธีการประยุกต์วิธีของคูมบ์มีช่วงการให้คะแนนอยู่ระหว่าง 0 ถึง 3 คะแนน ซึ่งช่วงการให้คะแนนน้อยจะทำให้ข้อมูลที่เป็นประโยชน์ขาดหายไปและความแม่นยำในการวัดต่ำลง (Ma, 2004) ในทางตรงข้ามกัน ถ้าช่วงการให้คะแนนเพิ่มขึ้นจะทำให้ความแปรปรวนเพิ่มขึ้น ดังนั้น ความเที่ยงจึงเพิ่มขึ้นด้วยเช่นกัน (Frey, 1989) เมื่อเปรียบเทียบความแตกต่างระหว่างความเที่ยงของทั้ง 2 วิธี พบว่า วิธีการให้คะแนนความรู้บางส่วนทั้ง 2 วิธี มีความเที่ยงไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ทั้งแบบสอบที่มีข้อสอบจำนวน 20 ข้อ 30 ข้อ และ 40 ข้อ ซึ่งสอดคล้องกับงานวิจัยของ Lau et al. (2011; 2014)

2. การเปรียบเทียบความสัมพันธ์ระหว่างความสามารถจริงกับความสามารถที่ได้จากการประมาณค่าที่ได้จากวิธีการประยุกต์วิธีของคูมบ์และวิธีตัดตัวลงเลือกตัวถูก

จากการวิเคราะห์ความสัมพันธ์ระหว่างความสามารถจริงกับความสามารถที่ได้จากการประมาณค่าที่ได้จากวิธีการประยุกต์วิธีของคูมบ์และวิธีตัดตัวลงเลือกตัวถูก พบว่า ค่าความสัมพันธ์ที่ได้มีค่าสูงกว่า .70 ซึ่งมีความสัมพันธ์ระดับสูง (Hinkle, Wiersma, and Jurs, 2003) ทั้งแบบสอบที่มีข้อสอบจำนวน 20 ข้อ 30 ข้อ

และ 40 ข้อ และเมื่อเปรียบเทียบความสัมพันธ์ระหว่างความสามารถจริงกับความสามารถที่ได้จากการประมาณค่าระหว่างวิธีการประยุกต์วิธีของคัมภ์และวิธีตัดตัวลงเลือกตัวถูก พบว่าความสัมพันธ์ระหว่างความสามารถจริงกับความสามารถที่ได้จากการประมาณค่าที่ได้จากวิธีตัดตัวลงเลือกตัวถูกมีค่าสูงกว่าความสัมพันธ์ที่ได้จากวิธีการประยุกต์วิธีของคัมภ์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เนื่องจากวิธีตัดตัวลงเลือกตัวถูกมีช่วงในการให้คะแนนกว้างทำให้สามารถจำแนกความสามารถที่แท้จริงของผู้เรียนได้อย่างแม่นยำกว่าวิธีการประยุกต์วิธีของคัมภ์ ซึ่งสอดคล้องกับงานวิจัยของ Lau, Hong, Lau, and Usop (2014) พบว่าวิธีตัดตัวลงเลือกตัวถูกมีประสิทธิภาพในการประมาณค่าความสามารถและให้ข้อมูลที่เป็นประโยชน์ในการจำแนกผู้สอบตามระดับความสามารถของแต่ละบุคคลสูงกว่าวิธีตัดตัวลงและวิธีแบบประเพณีนิยม ทั้งข้อสอบจำนวน 20 ข้อ 30 ข้อ และ 40 ข้อ

3. การเปรียบเทียบฟังก์ชันสารสนเทศของแบบสอบที่ได้จากวิธีการประยุกต์วิธีของคัมภ์และวิธีตัดตัวลงเลือกตัวถูก

จากการวิเคราะห์ฟังก์ชันสารสนเทศของแบบสอบพบว่าวิธีตัดตัวลงเลือกตัวถูกมีฟังก์ชันสารสนเทศของแบบสอบสูงกว่าวิธีการประยุกต์วิธีของคัมภ์ เนื่องจากวิธีตัดตัวลงเลือกตัวถูกมีความคลาดเคลื่อนมาตรฐานในการวัดต่ำ สามารถตรวจสอบความเข้าใจที่คลาดเคลื่อนได้และลดการเดาได้ (Lau, Hong, Lau, & Usop, 2014) จึงให้สารสนเทศของแบบสอบสูงกว่า โดยมีสารสนเทศของแบบสอบแตกต่างกันมากเมื่อผู้สอบมีความสามารถระดับปานกลางและระดับสูง ทั้งแบบสอบที่มีจำนวนข้อสอบ 20 ข้อ 30 ข้อ และ 40 ข้อ เนื่องจากผู้สอบที่มีความสามารถระดับต่ำ ส่วนใหญ่จะเดาคำตอบหรือตอบมั่ว และทำให้ไม่สามารถอธิบายความสัมพันธ์ระหว่างความสามารถที่มีอยู่ภายในบุคคล (Latent trait or ability) กับผลการตอบข้อสอบได้ นอกจากนี้ แบบสอบที่มีจำนวนข้อสอบ 30 ข้อ และ 40 ข้อ ที่ตรวจให้คะแนนด้วยวิธีตัดตัวลงเลือกตัวถูก มีฟังก์ชันสารสนเทศใกล้เคียงกัน ซึ่งสอดคล้องกับงานวิจัยของ Lau, Hong, Lau, and Usop (2014) พบว่า วิธีตัดตัวลงเลือกตัวถูกสามารถใช้กับแบบสอบที่มีจำนวนข้อสอบน้อยกว่าวิธีการตัดตัวลง แต่จะให้สารสนเทศและคุณสมบัติทางจิตมิติที่ใกล้เคียงกับข้อสอบที่ยาวกว่า

ข้อเสนอแนะ

ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1. จากผลการวิจัย พบว่า วิธีตัดตัวลงเลือกตัวถูกมีความเที่ยงของแบบสอบ ความสัมพันธ์ระหว่างความสามารถจริงกับความสามารถที่ได้จากการประมาณค่าที่ได้ และฟังก์ชันสารสนเทศของแบบสอบสูงกว่าวิธีการประยุกต์วิธีของคัมภ์ ดังนั้นครูผู้สอนควรใช้วิธีตัดตัวลงเลือกตัวถูกในการวัดผลการเรียนรู้ของผู้เรียน
2. จากผลการวิจัย พบว่า วิธีตัดตัวลงเลือกตัวถูกที่มีจำนวนข้อสอบ 30 ข้อให้สารสนเทศของแบบสอบใกล้เคียงกับจำนวน 40 ข้อ ดังนั้น การนำวิธีตัดตัวลงเลือกตัวถูกไปใช้จึงควรใช้ข้อสอบจำนวน 30 ข้อ เพื่อให้วัดผลการเรียนรู้ของผู้เรียนได้อย่างแม่นยำภายใต้ระยะเวลาและทรัพยากรที่จำกัด

ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

1. ควรเก็บข้อมูลกับตัวอย่างวิจัยจริงเพื่อยืนยันผลการวิจัยว่าสอดคล้องกับข้อมูลที่ได้จากการจำลองหรือไม่ โดยควรทำการศึกษาหลายๆ วิชา
2. ควรศึกษาตัวแปรอื่นๆ ร่วมกับวิธีการให้คะแนนความรู้บางส่วน เช่น จำนวนตัวเลือก จำนวนตัวถูก

เอกสารอ้างอิง

- กระทรวงศึกษาธิการ. (2551). **หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551**. กรุงเทพฯ: โรงพิมพ์ชุมนุมสหกรณ์การเกษตรแห่งประเทศไทย.
- โชติกา ภาชีผล. (2554). **การสร้างและพัฒนาเครื่องมือในการวัดและประเมินผลการศึกษา**. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- โชติกา ภาชีผล ญัฐภรณ์ หลาวทอง และกมลวรรณ ตั้งชนกานนท์. (2558). **การวัดและประเมินผลการเรียนรู้**. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- รัตนา ไชยตรี. (2546). **การเปรียบเทียบคุณภาพและอัตราความคลาดเคลื่อนของการกำหนดเกรดแบบอิงกลุ่ม และอิงเกณฑ์ เมื่อใช้แบบสอบถามเลือกตอบที่มีการตอบ และการตรวจให้คะแนนความรู้บางส่วนด้วยวิธีแตกต่างกัน**. [ม.ป.ท.: ม.ป.พ.]
- ศศิธร บัวทอง. (2560). **การวัดและประเมินทักษะการเรียนรู้ในศตวรรษที่ 21**. Veridian E-Journal, Silpakorn University, 10(2), 1856-1867.
- ศิริชัย กาญจนวาสี. (2555). **ทฤษฎีการทดสอบแนวใหม่ (Modern Test Theories)**. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- _____. (2556). **ทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory)**. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- สุพจน์ เกิดสุวรรณ. (2545). **การพัฒนาการวัดความรู้บางส่วนของผู้ตอบแบบสอบถามเลือกตอบ**. วิทยานิพนธ์ปริญญาครุศาสตรดุษฎีบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- สำนักงานเลขาธิการสภาการศึกษา กระทรวงศึกษาธิการ. (2560). **แผนการศึกษาแห่งชาติ พ.ศ. 2560 – 2579**. ค้นเมื่อ 6 ธันวาคม 2560, จาก http://www.sh.mahidol.ac.th/rdcenter/download/refer_APA.pdf
- เอมอร จังศิริพรพรรณ. (2545ก). **การเปรียบเทียบคุณภาพของแบบสอบถามเลือกตอบเมื่อตรวจด้วยวิธีการให้คะแนนความรู้บางส่วนกับวิธีประเมินนิยม**. *วารสารครุศาสตร์*, 34(3), 73-94.
- _____. (2545ข). **การให้คะแนนความรู้บางส่วนทำได้อย่างไร**. *วารสารวัดผลการศึกษา*, 23(67), 21-32.
- _____. (2546). **การเปรียบเทียบคุณภาพของแบบสอบถามเลือกตอบเมื่อตรวจด้วยวิธีการให้คะแนนความรู้บางส่วนกับวิธีประเมินนิยม**. *วารสารครุศาสตร์*, 31, 24-40.
- Chang, S. H. , Lin, P. C., & Lin, Z. C. (2007). Measures of partial knowledge and unexpected responses in multiple-choice tests. *Educational Technology & Society*, 10(4), 95-109.
- Frey, A. S. (1989). **A test validation study for capturing partial information in multiple-choice questions using polychotomous scoring**. EdD. Dissertation, University of San Francisco, California.
- Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). **Applied Statistics for the Behavioral Sciences**. 5 th ed. Boston, Mass.: Houghton Mifflin.



- Lau, P. N. K., Lau, S. H., Hong, K. S., & Usop, H. (2011). Guessing, Partial Knowledge, and Misconceptions in Multiple-Choice Tests. **Educational Technology & Society**, 14(4), 99-110.
- Lau, S. H., Hong, K. S., Lau, P. N. K., & Usop, H. (2014). Robustness Of Number Right Elimination Testing (Nret) Scoring Method For Multiple-Choice Items In Computeradaptive Assessment System (Caas). **Research and Practice in Technology Enhanced Learning**, 9(2), 283-300.
- Liang, T., & Wells, C. S. (2009). A model fit statistic for generalized partial credit model. **Educational and Psychological Measurement**, 69, 913-928.
- Ma, X. (2004). **An investigation of alternative approaches to scoring multiple response items on a certification examination**. Ph.D. Dissertation, University of Massachusetts, Amherst.
- Palmer, E. J., & Devitt, P. G. (2007). Assessment of higher order cognitive skills in undergraduate education: modified essay or multiple choice questions?. **BMC Medical Education**, 7(49), 1-7.
- Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: a meta-analysis of 80 years of research. **Educational Measurement: Issues and Practice**, 24(2), 3-13.
- Zafar-Khan, M., & Aljarallah, B. M. (2011). Evaluation of modified essay questions (meq) and multiple choice questions (mcq) as a tool for assessing the cognitive skills of undergraduate medical students. **International Journal of Health Sciences**, 5(1), 39-43.