

การพัฒนาวิธีการตรวจให้คะแนนความรู้บางส่วนและเปรียบเทียบคุณสมบัติทางจิตมิติ ของแบบสอบถูกผิดหลายตัวเลือก Development of Partial Knowledge Scoring Methods and Comparison of Psychometric Properties in Multiple True-False Tests

ภคจิรา บวรธรรมรัตน์¹ ญัตถุภรณ์ หลาวทอง²

Pakjira Bowornthammarat¹ Nuttaporn Lawthong²

¹นิสิตมหาบัณฑิตสาขาวิชาวิธีวิทยาการพัฒนานวัตกรรมทางการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา
คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

¹Graduate Student of Methodology for Innovation Development in Education Division, Educational
Research and Psychology Department, Faculty of Education, Chulalongkorn University

²อาจารย์ประจำภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

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

E-mail: 6380124427@student.chula.ac.th; โทรศัพท์มือถือ: 082-4801200

Received: 10 June 2022

Revised: 23 July 2022

Accepted: 25 July 2022

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อ 1. เพื่อพัฒนาวิธีการตรวจให้คะแนนความรู้บางส่วนของแบบสอบถูกผิดหลายตัวเลือก 2. เพื่อเปรียบเทียบคุณสมบัติทางจิตมิติของแบบสอบถูกผิดหลายตัวเลือกที่ตรวจให้คะแนนความรู้บางส่วนด้วยวิธีการนับ 2 (Count-2) วิธีการให้คะแนนบางส่วน 50 (PS₅₀) วิธีการเพิ่มคะแนนตัวเลือกที่เว้นไว้ (LO) และวิธีประยุกต์การเพิ่มคะแนนตัวเลือกที่เว้นไว้ (MLO) ตัวอย่างวิจัย ได้แก่ นักเรียนชั้นมัธยมศึกษาปีที่ 6 แผนการเรียนคณิต - วิทยาศาสตร์ จำนวน 1,178 คน เครื่องมือวิจัย ได้แก่ แบบสอบถูกผิดหลายตัวเลือก เรื่องเคมีอินทรีย์ จำนวน 20 ข้อ การวิเคราะห์ข้อมูลประกอบด้วย วิเคราะห์ค่าดัชนีความสอดคล้องของเนื้อหาและวัตถุประสงค์ การวิเคราะห์ค่าสถิติพื้นฐานของคะแนน การวิเคราะห์ความเที่ยง และการวิเคราะห์อำนาจจำแนก โดยใช้โปรแกรม SPSS

ผลการวิจัยพบว่า 1) ได้พัฒนาวิธีการตรวจให้คะแนน MLO ซึ่งพัฒนาจาก LO2) เมื่อตรวจให้คะแนนด้วยวิธีที่ต่างกัน พบว่า ทุกวิธีให้ค่าสัมประสิทธิ์ความเที่ยงที่ยอมรับได้คือมากกว่า 0.70 โดยวิธี Count-2 ให้ค่าสัมประสิทธิ์ความเที่ยงสูงที่สุด ส่วนค่าอำนาจจำแนกเฉลี่ยมีค่าสูงสุดเมื่อใช้วิธี

Count-2 โดยวิธี Count-2 และ MLO ข้อสอบทุกข้อมีค่าอำนาจจำแนกมากกว่า 0.2 แต่วิธี PS_{50} และ LO มีข้อสอบที่อำนาจจำแนกต่ำกว่า 0.2 จำนวน 1 ข้อ และ 3 ข้อ ตามลำดับ

คำสำคัญ: แบบสอบถูกผิดหลายตัวเลือก / วิธีการตรวจให้คะแนนความรู้บางส่วน / คุณสมบัติทางจิตมิติ

Abstract

The purposes of this research were to 1) develop partial knowledge scoring methods in Multiple True-False Tests (MTF); and 2) compare the psychometric properties of MTF items when partially scored using Count-2, Partial Scoring 50 (PS_{50}), Left Out (LO), and Modified Left Out (MLO) methods. The sample consisted of 1,178 grade 12 students enrolled in the Mathematics-Science program. The research instrument was a 20-item multiple true-false test on Organic Chemistry. Index of Item-Objective Congruence (IOC), fundamental score statistics, reliability, and discrimination power were analyzed using SPSS.

The findings were as follows: 1) The MLO method was developed based on the LO method. 2) When scoring test items using different methods, it was observed that all methods exhibited acceptable reliability coefficients exceeding 0.70. Among these, the Count-2 method demonstrated the highest reliability coefficients. Moreover, Count-2 also had the highest discrimination index. Both the Count-2 and MLO methods delivered a discrimination index exceeding 0.2. In contrast, the discrimination index fell below 0.2 for 1 item and 3 items when utilizing PS_{50} and LO, respectively. Regarding the mean discrimination power, both the Count-2 and MLO methods had test items with discrimination power values exceeding 0.2, while the PS_{50} and LO methods had one and three test items with discrimination power values below 0.2, respectively.

Keywords: Multiple True-False Test / Partial Knowledge Scoring / Psychometric properties

บทนำ

แบบสอบเป็นเครื่องมือวัดและประเมินผลสำหรับใช้ในการวัดผลสัมฤทธิ์ทางการเรียนของผู้เรียน คุณภาพของเครื่องมือวัดจึงเป็นปัจจัยที่จะช่วยสะท้อนความรู้ที่แท้จริงของผู้เรียน เนื่องจากการใช้เครื่องมือที่มีคุณภาพจะลดความคลาดเคลื่อนจากการวัดได้ ดังนั้น จึงต้องมีการตรวจสอบคุณสมบัติทางจิตมิติ หรือคุณสมบัติของการเป็นเครื่องมือวัดที่ดี โดยคุณสมบัติทางจิตมิตินี้อยู่หลายประเภท เช่น ความเที่ยง ซึ่งหมายถึงความคงที่ของผลการวัดระหว่างช่วงเวลาที่แตกต่างกัน หรือแบบวัดที่สมมูลกัน หรือความสอดคล้องภายในของการวัดลักษณะเดียวกัน (ศิริชัย กาญจนวาสี. 2556, หน้า 58) ความตรง ซึ่งหมายถึงความสามารถของเครื่องมือที่ใช้วัดได้ตรงตามสิ่งที่ต้องการวัดหรือจุดประสงค์ของการวัด ความยาก ซึ่งหมายถึงสัดส่วนของผู้ที่ตอบข้อสอบถูก และอำนาจจำแนก ซึ่งหมายถึงความสามารถในการจำแนกความแตกต่างระหว่างผู้สอบที่มีความสามารถต่างกัน (ศิริชัย กาญจนวาสี. 2556, หน้า 99)

แบบสอบที่ได้รับความนิยมและใช้อย่างแพร่หลาย ได้แก่ แบบสอบเลือกตอบหลายตัวเลือก (multiple choicetest) เนื่องจากแบบสอบประเภทนี้สามารถนำไปใช้ได้สะดวก มีความเป็นปรนัยสูง ใช้เวลาในการตรวจน้อย อีกทั้งยังสามารถวัดพฤติกรรมการเรียนรู้ได้หลากหลายและครอบคลุม (โชติกา ภาชีผล. 2559, หน้า 42) แต่แบบสอบประเภทนี้มีข้อจำกัดในเรื่องโอกาสในการเดาคำตอบของผู้สอบ เนื่องจากเป็นแบบสอบที่ให้ผู้สอบเลือกคำตอบที่ถูกต้องเพียงตัวเลือกเดียว ส่งผลให้ไม่สามารถวัดผลการเรียนรู้ที่แท้จริงของผู้เรียนได้ นักวิชาการ นักวัดและประเมินผลจึงได้พยายามแก้ไขข้อจำกัดดังกล่าวโดยการปรับปรุงและพัฒนารูปแบบข้อสอบวิธีการในการตอบสนองข้อสอบและตรวจให้คะแนน แบบสอบถูกผิดหลายตัวเลือกเป็นรูปแบบหนึ่งของแบบสอบเลือกตอบหลายตัวเลือกที่พัฒนาขึ้นโดยเป็นแบบสอบที่มีการตอบสนองหลายครั้ง ในแต่ละข้ออาจจะมีตัวเลือกถูก 1 ตัวเลือก มากกว่า 1 ตัวเลือก หรือไม่มีตัวเลือกถูกเลยก็ได้ ผู้สอบต้องพิจารณาทุกตัวเลือกว่าเป็นตัวเลือกที่ถูกหรือผิด การตอบสนองข้อสอบในลักษณะดังกล่าวสามารถช่วยลดการเดาคำตอบของผู้สอบได้ (อรพรรณ แก้วน้อย. 2557, หน้า 2) แบบสอบประเภทนี้สามารถตรวจให้คะแนนได้ทั้งแบบประเพณีนิยมและแบบการให้คะแนนความรู้บางส่วน ซึ่งเป็นวิธีการให้คะแนนที่ใกล้เคียงกับความรู้ของผู้สอบ โดยอาจมีการให้คะแนนมากกว่า 2 ค่า ในแต่ละข้อ เพื่อที่จะสามารถวัดระดับความรู้ของผู้สอบให้ได้ใกล้เคียงกับความรู้ที่แท้จริงของผู้สอบได้มากที่สุดและสามารถจำแนกผู้เรียนได้ (เอมอร จังศิริพรภรณ์. 2545, หน้า 12-19) จากการศึกษางานวิจัยและเอกสารที่เกี่ยวข้องพบวิธีการตรวจให้คะแนนความรู้บางส่วนของแบบสอบถูกผิดหลายตัวเลือกหลายวิธี โดยวิธีการตรวจให้คะแนนที่ให้คุณสมบัติทางจิตมิติที่ดี เช่น วิธีการนับ 2 (Count-2) เป็นวิธีการที่ผู้สอบจะได้รับคะแนนเมื่อตอบตรงกับเฉลยของตัวเลือกอย่างน้อย

1 ใน 2 ของจำนวนตัวเลือกทั้งหมด และจะได้คะแนนตามเงื่อนไขการคำนวณคะแนน ซึ่งเป็นวิธีที่มีความเที่ยงสูงกว่าวิธีอื่น (วัชรภรณ์ จิตมาศ. 2538 อ้างถึงใน อรพรรณ แก้วน้อย. 2557, หน้า 31; สุรัชย์ มีชาญ, 2545 อ้างถึงใน อรพรรณ แก้วน้อย. 2557, หน้า 32) วิธีการให้คะแนนบางส่วน 50 (Partial Scoring 50; PS_{50}) เป็นวิธีให้คะแนนซึ่งมีการให้คะแนนอยู่ 3 ค่า คือ 0 คะแนนเมื่อผู้สอบเลือกตอบได้ตรงกับเฉลยน้อยกว่าครึ่งหนึ่งของจำนวนตัวเลือกได้ 0.5 คะแนนเมื่อผู้สอบเลือกตอบได้ตรงกับเฉลยตั้งแต่ครึ่งหนึ่งของจำนวนตัวเลือกแต่ไม่ถูกต้องทั้งหมด และ 1 คะแนนเมื่อผู้สอบเลือกตอบได้ตรงกับเฉลยของตัวเลือกทั้งหมด ซึ่งเป็นวิธีที่ให้ความเที่ยง และอำนาจจำแนกสูงกว่าวิธีอื่น (Lahner.2018, p. 459-460) นอกจากนี้ การเว้นคำตอบสามารถสะท้อนความรู้ที่แท้จริงของผู้สอบได้ เนื่องจากการเว้นคำตอบของผู้สอบ อาจเกิดจากความตั้งใจไม่รู้คำตอบ หรือทำไม่ทัน (Grandy. 1987, p. 17-19; Custer et al. 2012 p. 1) โดยแบบสอบถูกผิดหลายตัวเลือกมีวิธีการตรวจให้คะแนนที่นำการเว้นคำตอบมาคิดคะแนน ได้แก่ วิธีการเพิ่มคะแนนให้กับตัวเลือกที่เว้นไว้ (Let omit; LO) เป็นวิธีตรวจให้คะแนนโดยเพิ่มคะแนนให้กับข้อที่ผู้สอบไม่ได้ตอบหรือเว้นไว้ คะแนนที่ได้ในแต่ละข้อจึงคำนวณมาจากจำนวนตัวเลือกทั้งหมด จำนวนตัวเลือกที่ตอบตรงกับเฉลย และจำนวนตัวเลือกที่เว้นไว้ (อรพรรณ แก้วน้อย. 2557, หน้า 30; Tsai & Suen. 1993, p. 401) แต่การตรวจให้คะแนนด้วยวิธีนี้มีข้อจำกัดในเรื่องของการให้คะแนนติดลบในแต่ละข้อ ซึ่งอาจมีผลต่อคะแนนรวมทั้งหมดและไม่สะท้อนความรู้ที่แท้จริงของผู้เรียน (เอมอร จังศิริพรภรณ์. 2545, หน้า 6-7)

จากองค์ความรู้ที่กล่าวมาและข้อจำกัดของวิธีการเพิ่มคะแนนให้กับตัวเลือกที่เว้นไว้ ผู้วิจัยจึงได้มีแนวคิดที่จะพัฒนาวิธีประยุกต์การเพิ่มคะแนนตัวเลือกที่เว้นไว้ (Modified Let Omit; MLO) และสนใจศึกษาคุณสมบัติทางจิตมิติตามทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory; CTT) ของแบบสอบถูกผิดหลายตัวเลือกที่ตรวจให้คะแนนด้วย MLO โดยเปรียบเทียบกับวิธีการตรวจให้คะแนนอื่น ได้แก่ Count-2, PS_{50} และ LO ผลการวิจัยจะทำให้ได้วิธี MLO สำหรับเป็นทางเลือกในการตรวจให้คะแนนความรู้บางส่วนแบบสอบถูกผิดหลายตัวเลือก อีกทั้งได้ทราบคุณภาพของแบบสอบถูกผิดหลายตัวเลือกเมื่อตรวจให้คะแนนด้วยวิธี Count-2, PS_{50} , LO และ MLO ซึ่งจะประโยชน์ต่อครูผู้สอน นักการศึกษา นักวัดและประเมินผล ในการเลือกวิธีการตรวจให้คะแนนแบบสอบถูกผิดหลายตัวเลือกไปใช้วัดความรู้บางส่วนของผู้สอบทั้งในระดับการสอบในชั้นเรียนและการสอบขนาดใหญ่

วัตถุประสงค์การวิจัย

1. เพื่อพัฒนาวิธีการตรวจให้คะแนนความรู้บางส่วนของแบบสอบถูกผิดหลายตัวเลือก
2. เพื่อเปรียบเทียบคุณสมบัติทางจิตมิติของแบบสอบถูกผิดหลายตัวเลือกที่มีวิธีการให้คะแนนความรู้บางส่วน 4 วิธีประกอบด้วยวิธีการนับ 2 (Count-2) วิธีการให้คะแนนบางส่วน 50 (PS_{50}) วิธีการเพิ่มคะแนนตัวเลือกที่เว้นไว้ (LO) และวิธีประยุกต์การเพิ่มคะแนนตัวเลือกที่เว้นไว้ (MLO)

ขอบเขตการวิจัย

1. ประชากรและกลุ่มตัวอย่าง

ประชากรการวิจัยคือนักเรียนระดับชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2564 แผนการเรียนคณิตศาสตร์ - วิทยาศาสตร์โรงเรียนมัธยมศึกษา สังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษากรุงเทพมหานครเขต 1 และเขต 2 จำนวน 120 โรงเรียน

ตัวอย่างการวิจัยคือนักเรียนระดับชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2564 แผนการเรียนคณิตศาสตร์-วิทยาศาสตร์โรงเรียนมัธยมศึกษา สังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษากรุงเทพมหานครเขต 1 และเขต 2 จำนวน 1,178 คนจาก 10 โรงเรียนได้มาจากการเลือกอย่างเจาะจง (Purposive sampling) คัดเลือกตัวอย่างวิจัยจากโรงเรียนโดยมีเกณฑ์คือ (1) เป็นโรงเรียนที่มีการจัดการเรียนการสอนเรื่องเคมีอินทรีย์หรือในช่วงภาคเรียนที่ 1 หรือก่อนการสอบกลางภาคของภาคเรียนที่ 2 ปีการศึกษา 2564 (2) โรงเรียนมีการจัดการเรียนการสอนโดยครูผู้สอนในรูปแบบการสอนสดผ่านสัญญาณอินเทอร์เน็ตและ/หรือการบันทึกวิดีโอและ(3) เป็นโรงเรียนที่นักเรียนมีความพร้อมในเรื่องของสัญญาณอินเทอร์เน็ตและอุปกรณ์อิเล็กทรอนิกส์ที่ใช้ในการสอบ

2. ตัวแปรที่ศึกษา

2.1 ตัวแปรอิสระ ได้แก่ วิธีการตรวจให้คะแนนความรู้บางส่วนของแบบสอบถูกผิดหลายตัวเลือกประกอบด้วย วิธีการนับ 2 (Count-2) วิธีการให้คะแนนบางส่วน 50 (PS_{50}) วิธีการเพิ่มคะแนนตัวเลือกที่เว้นไว้ (LO) และวิธีประยุกต์การเพิ่มคะแนนตัวเลือกที่เว้นไว้ (MLO)

2.2 ตัวแปรตาม ได้แก่ คุณสมบัติทางจิตมิติตามทฤษฎีการทดสอบแบบดั้งเดิมประกอบด้วย ความตรงด้านเนื้อหา ความเที่ยง และอำนาจจำแนก

วิธีดำเนินการวิจัย

เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยครั้งนี้ได้แก่แบบสอบถูกผิดหลายตัวเลือกวิชาเคมีเรื่องเคมีอินทรีย์จำนวน 20 ข้อโดยคัดเลือกมาจากข้อสอบจำนวน 40 ข้อที่ได้นำไปทดลองใช้ (try out) แล้วนำผล

การตอบมาตรวจให้คะแนนตอบผิดได้ 0 คะแนน ตอบถูกได้ 1 คะแนน และตรวจสอบคุณภาพข้อสอบ โดยแบบสอบมีค่าสัมประสิทธิ์ความเที่ยง 0.61 ค่าความยากอยู่ในช่วง 0.34 ถึง 0.68 และค่าอำนาจจำแนกอยู่ในช่วง -0.01 ถึง 0.43 ผู้วิจัยคัดเลือกข้อสอบที่มีคุณภาพเหมาะสมคือค่าความยากอยู่ในช่วง 0.20 ถึง 0.80 และค่าอำนาจจำแนกมากกว่า 0.2 มาสร้างเป็นแบบสอบ

ตัวอย่างข้อสอบถูกผิดหลายตัวเลือก

ข้อใดกล่าวถูกต้องเกี่ยวกับธาตุคาร์บอน

- ก. สัญลักษณ์ธาตุคือ C
- ข. เลขอะตอมเท่ากับ 6
- ค. จำนวนเวเลนซ์อิเล็กตรอนเท่ากับ 4
- ง. การจัดเรียงอิเล็กตรอนในระดับพลังงานหลักคือ 2 6

ตัวเลือก (เลขย)	ถูก	ผิด	สภาพของการตอบ
ก (ถูก)	X		ตรงกับเฉลย
ข (ผิด)	X		ไม่ตรงกับเฉลย
ค (ถูก)		X	ไม่ตรงกับเฉลย
ง (ผิด)			ไม่ตรงกับเฉลย* เว้นคำ**

* เมื่อตรวจให้คะแนนด้วยวิธี count 2 และ PS50 ** เมื่อตรวจให้คะแนนด้วยวิธี LO

ภาพที่ 1 ตัวอย่างกระดาษคำตอบวิธีการตอบและสภาพการตอบ

การเก็บรวบรวมข้อมูล

งานวิจัยครั้งนี้ทำการเก็บรวบรวมข้อมูลในช่วงภาคเรียนที่ 2 ปีการศึกษา 2564 ของโรงเรียนมัธยมศึกษาโดยเก็บรวบรวมข้อมูลผลการตอบแบบสอบถามถูกผิดหลายตัวเลือกผ่านระบบออนไลน์ด้วยแพลตฟอร์มกูเกิลฟอร์ม (Google Form) และควบคุมการสอบผ่าน Google Meet หรือ Zoom ใช้เวลาในการทำแบบสอบ 60 นาทีจากนั้นนำผลการตอบมาตรวจให้คะแนนด้วยวิธีการ Count-2, PS₅₀, LO และ MLO ซึ่งมีวิธีการและเงื่อนไขการตรวจให้คะแนนดังนี้

1. วิธีการนับ 2

วิธีการนับ 2 (count-2) ผู้สอบจะได้คะแนนความรู้บางส่วนเมื่อเลือกตอบตรงกับเฉลยของตัวเลือกตั้งแต่ 1 ใน 2 ของตัวเลือกทั้งหมดในข้อนั้น และได้ 1 คะแนน เมื่อเลือกตอบตรงกับเฉลยของตัวเลือกทั้งหมด โดยเงื่อนไขการให้คะแนนคือ

$$f(i) = \begin{cases} 0 & i = 0 \text{ หรือ } 1 \\ \frac{i-1}{n-1} & \text{กรณีอื่นๆ} \end{cases}$$

เมื่อ i คือ จำนวนตัวเลือกที่เลือกตอบตรงกับเฉลยของตัวเลือก
 n คือ จำนวนตัวเลือกทั้งหมด

ดังนั้น จะมีคะแนนที่เป็นไปได้คือ 0, 0.33, 0.67 และ 1 คะแนน

2. วิธีการให้คะแนนบางส่วน 50

วิธีการให้คะแนนบางส่วน 50 (Partial Scoring 50; PS_{50}) ผู้สอบจะได้คะแนนครึ่งหนึ่งเมื่อเลือกตอบตรงกับเฉลยของตัวเลือกอย่างน้อยร้อยละ 50 ของตัวเลือก และได้คะแนนเต็มเมื่อเลือกตอบตรงกับเฉลยของตัวเลือกทั้งหมด และมีคะแนนที่เป็นไปได้ 3 ค่า คือ 0, 0.5 และ 1 ทั้งนี้กรณีที่มีจำนวนตัวเลือก 4 ตัวเลือก และ 5 ตัวเลือก โดยมีเงื่อนไขการให้คะแนนคือ

$$f(i) = \begin{cases} 1, & i = n \\ 0.5, & \frac{n}{2} \leq i < n \\ 0, & \text{กรณีอื่นๆ} \end{cases}$$

เมื่อ i คือ จำนวนตัวเลือกที่เลือกตอบตรงกับเฉลยของตัวเลือก
 n คือ จำนวนตัวเลือกทั้งหมด

3. วิธีการให้คะแนนแบบการเพิ่มคะแนนตัวเลือกที่เว้นไว้

วิธีการให้คะแนนแบบการเพิ่มคะแนนตัวเลือกที่เว้นไว้ (Let omit; LO) (Tsai and Suen, 1993) เป็นวิธีที่ถือว่าข้อที่ผู้สอบเว้นไว้คือข้อที่ผู้สอบไม่มีความรู้หรือทำไม่ทัน มีเงื่อนไขการให้คะแนนคือ

$$f(i) = \frac{i - (n - i - j)}{n} \quad \text{เมื่อ } n \leq i, j < 0$$

เมื่อ i คือ จำนวนตัวเลือกที่เลือกตอบตรงกับเฉลยของตัวเลือก
 j คือ จำนวนตัวเลือกที่ผู้สอบเว้นไว้
 n คือ จำนวนตัวเลือกทั้งหมดของข้อสอบ

ดังนั้น จะมีคะแนนที่เป็นไปได้ คือ -1, -0.75, -0.50, -0.25, 0, 0.25, 0.50, 0.75 และ 1 คะแนน

ตารางที่ 1 รูปแบบการตอบและคะแนนเมื่อตรวจให้คะแนนด้วยวิธีการนับ 2 (Count-2) วิธีการให้คะแนนบางส่วน 50 (PS₅₀) และวิธีการเพิ่มคะแนนตัวเลือกที่เว้นไว้ (LO)

รูปแบบการตอบ	จำนวนตัวเลือก			คะแนน		
	ตอบตรงกับเฉลย	ตอบไม่ตรงกับเฉลย	เว้นคำตอบ	Count-2	PS ₅₀	LO
1	0	0	4	0	0	0
2	0	1	3	0	0	-0.25
3	0	2	2	0	0	-0.5
4	0	3	1	0	0	-0.75
5	0	4	0	0	0	-1
6	1	3	0	0	0	-0.5
7	1	2	1	0	0	-0.25
8	1	1	2	0	0	0
9	1	0	3	0	0	0.25
10	2	2	0	0.33	0.5	0
11	2	1	1	0.33	0.5	0.25
12	2	0	2	0.33	0.5	0.5
13	3	1	0	0.67	0.5	0.5
14	3	0	1	0.67	0.5	0.75
15	4	0	0	1	1	1

หมายเหตุ: เมื่อตรวจให้คะแนนด้วยวิธี Count-2 และ PS₅₀ การเว้นคำตอบถือเป็นการตอบไม่ตรงเฉลย

การวิเคราะห์ข้อมูล

งานวิจัยครั้งนี้ทำการวิเคราะห์ค่าสถิติและคุณสมบัติทางจิตมิติโดยใช้โปรแกรม SPSS ดังนี้

1. วิเคราะห์ค่าสถิติพื้นฐาน ได้แก่ คะแนนต่ำสุด (Min) คะแนนสูงสุด (Max) ค่าเฉลี่ย (Mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation)
2. วิเคราะห์ค่าดัชนีความสอดคล้องของเนื้อหาและวัตถุประสงค์ (Index Of item object Congruence: IOC)
3. วิเคราะห์ความเที่ยงแบบสอดคล้องภายในโดยคำนวณค่าสัมประสิทธิ์แอลฟาของครอนบาค (Cronbach's alpha) โดยค่าสัมประสิทธิ์ความเที่ยงที่ยอมรับได้คือมีค่าสูงกว่า 0.5 (ศิริชัย กาญจนวาสี. 2556, หน้า 97)
4. วิเคราะห์อำนาจจำแนกโดยการคำนวณค่าอำนาจจำแนก(Corrected Item Total Correlation)โดยมีเกณฑ์การแปลความหมาย ดังนี้ต่ำกว่า 0.20 หมายถึง ปรับปรุง, 0.20-0.29 หมายถึง พอใช้, 0.30-0.39 หมายถึง ดี และ 0.40 ขึ้นไป หมายถึง ดีมาก (McGehee & Ball. 2009, p.5)

ผลการวิจัย

ตอนที่ 1 การพัฒนาวิธีการตรวจให้คะแนนความรู้บางส่วนของแบบสอบถูกผิดหลายตัวเลือก

ผู้วิจัยพัฒนาวิธีประยุกต์การเพิ่มคะแนนตัวเลือกที่เว้นไว้ (Modified Let Omit; MLO) โดยพัฒนามาจากวิธี LO ซึ่งมีข้อจำกัดเนื่องจากมีการให้คะแนนติดลบ และกรณีที่ผู้สอบเว้นคำตอบทั้งหมดในแต่ละข้อจะให้คะแนนมากกว่าการตอบไม่ตรงเฉลยทั้งหมดในแต่ละข้อ ผู้วิจัยจึงปรับวิธีคิดคะแนนให้คะแนนรวมในแต่ละข้อไม่ติดลบโดยการนำสมการในการตรวจให้คะแนนแบบการเพิ่มคะแนนตัวเลือกที่เว้นไว้มาบวกด้วย 1 และปรับให้คะแนนรวมในแต่ละข้อมีค่าสูงสุดเต็ม 1 คะแนน โดยนำสมการที่บวกด้วย 1 ข้างต้นหารด้วย 2 และกำหนดเงื่อนไขในการคำนวณคะแนนเพิ่มเติมคือ ผู้สอบจะได้ 0 คะแนนเมื่อเว้นคำตอบทั้งหมดในข้อนั้นโดยมีเงื่อนไขการให้คะแนนดังนี้

$$f(i) = \begin{cases} \left[\frac{i - (n - i - j)}{n} \right] + 1 & , \text{เมื่อ } n \leq i, j < 0 \\ 0 & , j = n \end{cases}$$

เมื่อ i คือ จำนวนตัวเลือกที่เลือกตอบตรงกับเฉลยของตัวเลือก

j คือ จำนวนตัวเลือกที่ผู้สอบเว้นไว้

n คือ จำนวนตัวเลือกทั้งหมดของข้อสอบ

ดังนั้น จะมีค่าคะแนนที่เป็นไปได้คือ 0, 0.125, 0.250, 0.375, 0.500, 0.625, 0.750, 0.875 และ 1 คะแนน

ตารางที่ 2 รูปแบบการตอบและคะแนนเมื่อตรวจให้คะแนนด้วยวิธีประยุกต์การเพิ่มคะแนนตัวเลือกที่เว้นไว้

รูปแบบการตอบ	จำนวนตัวเลือกที่ตอบตรงกับเฉลย	จำนวนตัวเลือกที่ตอบไม่ตรงกับเฉลย	เว้นคำตอบ	คะแนน
1	0	0	4	0
2	0	1	3	0.375
3	0	2	2	0.25
4	0	3	1	0.125
5	0	4	0	0
6	1	3	0	0.25
7	1	2	1	0.375
8	1	1	2	0.5
9	1	0	3	0.625
10	2	2	0	0.5
11	2	1	1	0.625
12	2	0	2	0.75
13	3	1	0	0.75
14	3	0	1	0.875
15	4	0	0	1

เมื่อเปรียบเทียบคะแนนที่ได้จากการตรวจให้คะแนนด้วยวิธีต่างๆ พบว่า วิธีประยุกต์การเพิ่มคะแนนตัวเลือกที่เว้นไว้มีค่าคะแนนเฉลี่ยสูงสุด 12.05 คะแนน ส่วนเบี่ยงเบนมาตรฐาน 2.82 โดยมีคะแนนอยู่ในช่วง 1.125 ถึง 19 คะแนนและวิธีการเพิ่มคะแนนให้กับตัวเลือกที่เว้นไว้ เป็นวิธีการตรวจให้คะแนนที่มีค่าเฉลี่ยคะแนนต่ำที่สุด 4.91คะแนนส่วนเบี่ยงเบนมาตรฐาน 4.66 โดยมีคะแนนอยู่ในช่วง -5 ถึง 18 คะแนน ในขณะที่วิธีการนับ 2 และวิธีการให้คะแนนบางส่วน 50 มีค่าเฉลี่ยคะแนน ส่วนเบี่ยงเบนมาตรฐาน และช่วงคะแนนใกล้เคียงกัน ค่าแสดงดังตารางที่3

ตารางที่ 3 ค่าสถิติพื้นฐานของคะแนนจากการตรวจให้คะแนนด้วยวิธีที่ต่างกัน

วิธีการตรวจให้คะแนน	Min	Max	M	SD
Count-2	0.02	18.67	9.40	3.67
PS ₅₀	0	18.5	9.41	3.43
LO	-5	18	4.91	4.66
MLO	1.125	19	12.05	2.82

ตอนที่ 2 การเปรียบเทียบคุณสมบัติทางจิตมิติของแบบสอบถูกผิดหลายตัวเลือกเมื่อตรวจให้คะแนนความรู้บางส่วนด้วยวิธีที่ต่างกัน

งานวิจัยนี้ทำการวิเคราะห์คุณสมบัติทางจิตมิติตามทฤษฎีการทดสอบแบบดั้งเดิมของแบบสอบถูกผิดหลายตัวเลือกเมื่อตรวจให้คะแนนความรู้บางส่วนด้วยวิธีที่ต่างกัน 4 วิธี ได้แก่ Count-2, PS₅₀, LO และ MLO โดยวิเคราะห์คุณสมบัติทางจิตมิติ ได้แก่ ความสอดคล้องของเนื้อหาและวัตถุประสงค์ ความเที่ยง และอำนาจจำแนก

1. ผลการวิเคราะห์ความสอดคล้องของเนื้อหาและวัตถุประสงค์

แบบสอบถูกผิดหลายตัวเลือก เรื่องเคมีอินทรีย์ที่นำมาใช้เป็นเครื่องมือในการวิจัยครั้งนี้ได้ผ่านการตรวจสอบความสอดคล้องของเนื้อหาและวัตถุประสงค์โดยผู้เชี่ยวชาญจำนวน 5 ท่าน โดยผลการวิเคราะห์ค่าดัชนีความสอดคล้องของเนื้อหาและวัตถุประสงค์พบว่าข้อสอบจำนวน 20 ข้อ มีค่าดัชนีความสอดคล้องของเนื้อหาและวัตถุประสงค์มีค่า 0.4 จำนวน 1 ข้อ, 0.6 จำนวน 3 ข้อ, 0.8 จำนวน 3 ข้อ และ 1 จำนวน 13 ข้อ โดยข้อสอบที่นำมาใช้คือข้อสอบที่มีค่าดัชนีความสอดคล้องของเนื้อหาและวัตถุประสงค์มากกว่า 0.5 ดังนั้น ผู้วิจัยจึงได้ทำการปรับปรุงข้อสอบ จำนวน 1 ข้อที่มีค่าดัชนีความสอดคล้องของเนื้อหาและวัตถุประสงค์ต่ำกว่า 0.5 ตามคำแนะนำของผู้เชี่ยวชาญก่อนนำข้อสอบไปใช้

2. ผลการวิเคราะห์ความเที่ยง

ผลการวิเคราะห์ความเที่ยงแบบสอดคล้องภายในของแบบสอบถูกผิดหลายตัวเลือก เรื่องเคมีอินทรีย์ เมื่อตรวจให้คะแนนด้วยวิธีต่างๆ พบว่า วิธีการตรวจให้คะแนนแต่ละวิธีให้ค่าความเที่ยงใกล้เคียงกัน โดยวิธี Count-2 มีความเที่ยงสูงที่สุด เท่ากับ 0.739 รองลงมาคือวิธี PS₅₀ (0.737) วิธี MLO (.732) และวิธี LO (0.720) ซึ่งความเที่ยงของแบบสอบถูกผิดหลายตัวเลือก เมื่อตรวจให้คะแนนทั้ง 4 วิธี มีค่ามากกว่า 0.5 จึงอยู่ในเกณฑ์ที่ยอมรับได้ (ศิริชัย กาญจนวาสี. 2556, หน้า97) ดังตารางที่ 4

วิธีการตรวจให้คะแนน	สัมประสิทธิ์แอลฟาของครอนบาค
Count-2	0.739
PS ₅₀	0.737
LO	0.720
MLO	0.732

ผลการวิเคราะห์อำนาจจำแนก

ผลการวิเคราะห์อำนาจจำแนกของแบบสอบถูกผิดหลายตัวเลือก เรื่องเคมีอินทรีย์ เมื่อตรวจให้คะแนนด้วยวิธีต่าง ๆ ด้วยการหาค่า Corrected Item-Total Correlation พบว่า วิธี Count-2 ให้ค่าอำนาจจำแนกเฉลี่ยสูงที่สุดคือ 0.48 โดยมีข้อสอบที่มีอำนาจจำแนกดีมากที่สุด 15 ข้อ มีอำนาจจำแนกดีจำนวน 3 ข้อ และมีอำนาจจำแนกพอใช้จำนวน 2 ข้อ รองลงมาคือ PS₅₀ มีค่าอำนาจจำแนกเฉลี่ย 0.47 โดยมีข้อสอบที่มีอำนาจจำแนกดีมากที่สุด 15 ข้อ มีอำนาจจำแนกดีจำนวน 5 ข้อ และควรปรับปรุงอำนาจจำแนก จำนวน 1 ข้อ วิธี MLO ค่าอำนาจจำแนกเฉลี่ย 0.44 โดยมีข้อสอบที่มีอำนาจจำแนกดีมากที่สุด 16 ข้อ มีอำนาจจำแนกดี จำนวน 3 ข้อ และมีอำนาจจำแนกพอใช้ จำนวน 1 ข้อ และวิธี LO มีค่าเฉลี่ยอำนาจจำแนกต่ำที่สุดคือ 0.32 โดยมีข้อสอบที่มีอำนาจจำแนกดีมากที่สุด 12 ข้อ มีอำนาจจำแนกดี จำนวน 2 ข้อ มีอำนาจจำแนกพอใช้ จำนวน 3 ข้อ และควรปรับปรุงอำนาจจำแนก จำนวน 3 ข้อ ค่าอำนาจจำแนกแสดงดังตารางที่ 5

ตารางที่ 5 ค่าอำนาจจำแนก (Corrected Item-Total Correlation)

อภิปรายผลการวิจัย

ข้อที่	Corrected Item-Total Correlation			
	Count-2	PS ₅₀	LO	MLO
1	0.30	0.31	0.25	0.24
2	0.54	0.53	0.50	0.51
3	0.59	0.59	0.57	0.56
4	0.50	0.41	0.44	0.41
5	0.64	0.64	0.58	0.57
6	0.54	0.49	0.44	0.49
7	0.50	0.51	0.42	0.45
8	0.44	0.40	0.27	0.45
9	0.63	0.62	0.53	0.58
10	0.23	0.31	0.16	0.24
11	0.57	0.46	0.43	0.51
12	0.63	0.59	0.55	0.58
13	0.52	0.48	0.48	0.49
14	0.42	0.44	0.33	0.42
15	0.38	0.41	0.29	0.33
16	0.58	0.60	0.50	0.54
17	0.44	0.49	0.33	0.46
18	0.31	0.36	0.17	0.31
19	0.55	0.48	0.42	0.51
20	0.23	0.19	0.12	0.24
เฉลี่ย	0.48	0.47	0.39	0.44

การพัฒนาวิธีการตรวจให้คะแนนแบบสอบถูกผิดหลายตัวเลือกวิธี MLO ผู้วิจัยได้ทำการพัฒนาขึ้นโดยปรับจากวิธี LO โดยการปรับให้คะแนนแต่ละข้อไม่ติดลบ และปรับเงื่อนไขการให้คะแนนในกรณีที่ผู้สอบเว้นคำตอบทั้งหมดในแต่ละข้อ จากผลการวิเคราะห์ข้อมูลผู้วิจัยได้เสนอประเด็นในการอภิปรายผลการวิจัย ดังนี้

1. ค่าสถิติพื้นฐานของคะแนนจากการตรวจให้คะแนนด้วยวิธี Count-2, PS₅₀, LO และ MLO

วิธีการให้คะแนนที่พัฒนาขึ้นให้ค่าเฉลี่ยคะแนนสูงสุด โดยมีช่วงคะแนน 1.125 ถึง 19 คะแนน เนื่องจาก การตรวจให้คะแนนด้วยวิธี MLO มีค่าคะแนนที่เป็นไปได้ทั้งหมด 9 ค่า โดยในแต่ละข้อผู้สอบมีโอกาสได้ 0 คะแนนเพียงร้อยละ 13.33 ทำให้ผู้สอบมีโอกาสได้คะแนนรวมมากกว่า ในขณะที่วิธี Count-2 และวิธี PS₅₀ ในแต่ละข้อผู้สอบมีโอกาสได้ 0 คะแนน ร้อยละ 60 ส่วนวิธี LO เป็นวิธีที่มีการให้คะแนนแบบติดลบจึงทำให้คะแนนรวมน้อยกว่าการตรวจให้คะแนน

วิธีอื่น

2. อำนาจจำแนกของข้อสอบเมื่อตรวจให้คะแนนด้วยวิธี Count-2, PS_{50} , LO และ MLO แบบสอบถูกผิดหลายตัวเลือกที่พัฒนาขึ้นมีค่าดัชนีความสอดคล้องของเนื้อหาและวัตถุประสงค์อยู่ระหว่าง 0.60 – 1 เมื่อนำผลการตอบมาตรวจให้คะแนนและวิเคราะห์อำนาจจำแนกพบว่า วิธี Count-2 มีค่าอำนาจจำแนกเฉลี่ยสูงสุดรองลงมาคือ PS_{50} , MLO และ LO เนื่องจากเมื่อพิจารณารูปแบบการตอบและคะแนนที่เป็นไปได้เมื่อตรวจให้คะแนนด้วย LO และ MLO ในกรณีที่ผู้สอบมีจำนวนตัวเลือกที่ตอบตรงเฉลยน้อยกว่าและมีการเว้นคำตอบ แต่ได้คะแนนเท่ากับเมื่อตอบตรงเฉลยมากกว่าและมีการตอบไม่ตรงเฉลย เช่น ผู้สอบตอบตรงเฉลย 2 ตัวเลือกและเว้นคำตอบ 2 ตัวเลือก ได้คะแนนเท่ากับผู้สอบที่ตอบตรงเฉลย 3 ตัวเลือก และตอบไม่ตรงเฉลย 1 ตัวเลือก อาจส่งผลให้ไม่สามารถจำแนกผู้สอบได้ดีเท่าที่ควร (Siddiqui et al., 2016, p.117) และมีค่าต่ำกว่าวิธี Count-2 และ PS_{50} ซึ่งเป็นวิธีที่มีโอกาสในการได้คะแนนใกล้เคียงกัน ทำให้คะแนนที่ได้จากการตรวจให้คะแนนด้วยวิธีดังกล่าวอยู่ในช่วงใกล้เคียงกัน

3. ความเที่ยงของแบบสอบเมื่อตรวจให้คะแนนด้วยวิธี Count-2, PS_{50} , LO และ MLO ความเที่ยงมีค่าสูงที่สุดเมื่อตรวจให้คะแนนด้วยวิธี Count-2 สอดคล้องกับสุรชัย มีชาญ (2545) ซึ่งเปรียบเทียบคุณสมบัติทางจิตมิติของแบบสอบถูกผิดหลายตัวเลือกที่ตรวจให้คะแนนด้วยวิธีที่ต่างกันทั้งการให้คะแนนแบบประเพณีนิยมและการให้คะแนนความรู้บางส่วน พบว่าวิธีการนับ 2 มีค่าสัมประสิทธิ์ความเที่ยงสูงกว่าวิธี pass-3 และแบบประเพณีนิยมวิธี PS_{50} มีสัมประสิทธิ์ความเที่ยงรองลงมาและใกล้เคียงกับ Count- 2 เนื่องจากวิธีการตรวจให้คะแนนทั้ง 2 วิธีนี้ให้คะแนนการเว้นคำตอบเช่นเดียวกับการตอบไม่ตรงเฉลยแต่วิธี LO และวิธี MLO เมื่อผู้สอบมีการเว้นคำตอบและตอบไม่ตรงเฉลยจะมีการคำนวณคะแนนที่แตกต่างกัน ซึ่งในงานวิจัยนี้ถือว่าการเว้นคำตอบของผู้สอบเกิดจากการไม่ทราบคำตอบ แต่ในความเป็นจริงการเว้นคำตอบของผู้สอบอาจเกิดจากปัจจัยอื่น ซึ่งอาจส่งผลต่อคะแนนที่ได้จากวิธีการตรวจให้คะแนนดังกล่าว ส่วนสัมประสิทธิ์ความเที่ยงของวิธี MLO มีค่าสูงกว่าวิธีการเพิ่มคะแนนตัวเลือกที่เว้นไว้ LO เนื่องจากคะแนนที่ได้จากวิธี MLO อยู่ในช่วงบวกจึงมีสัมประสิทธิ์ความเที่ยงสูงกว่าวิธีการเพิ่มคะแนนตัวเลือกที่เว้นไว้ซึ่งมีช่วงคะแนนติดลบ (สุพจน์ เกิดสุวรรณ. 2545, หน้า158)

ข้อเสนอแนะเพื่อนำผลการวิจัยไปใช้

เมื่อพิจารณาคุณสมบัติทางจิตมิติของแบบสอบถูกผิดหลายตัวเลือกที่ตรวจให้คะแนนความรู้บางส่วนด้วยวิธีที่ต่างกัน พบว่า วิธี Count-2 และวิธี MLO เป็นวิธีการตรวจให้คะแนนที่มีความเที่ยง และอำนาจจำแนกเหมาะสม หากครูผู้สอนต้องการนำวิธีการตรวจให้คะแนนไปใช้ วิธีการคำนวณคะแนนเมื่อตรวจให้คะแนนด้วยวิธี Count-2 สามารถทำได้ง่ายและสะดวกกว่าในกรณี

ที่ครูผู้สอนต้องคำนวณโดยไม่ใช้โปรแกรม แต่ในกรณีที่ครูผู้สอนใช้โปรแกรมในการคำนวณคะแนน ครูผู้สอนสามารถเลือกใช้วิธีการตรวจให้คะแนนวิธีการใดก็ได้

ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

การวิจัยในครั้งนี้ทำการตรวจสอบคุณสมบัติทางจิตมิติตามทฤษฎีการทดสอบแบบดั้งเดิม ซึ่งผลการวิเคราะห์อาจเปลี่ยนแปลงไปตามกลุ่มตัวอย่างวิจัย ดังนั้น จึงควรทำการตรวจสอบคุณสมบัติทางจิตมิติตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) เพื่อให้การวิจัยมีความน่าเชื่อถือและมีครอบคลุมมากยิ่งขึ้น

เอกสารอ้างอิง

- โชติกา ภาชีผล. (2559). *การวัดและประเมินผลการเรียนรู้*. กรุงเทพฯ ฯ: สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี. (2556). *ทฤษฎีการทดสอบแบบดั้งเดิม* (พิมพ์ครั้งที่ 7). กรุงเทพฯ ฯ: สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.
- สุพจน์ เกิดสุวรรณ. (2545). *การพัฒนาการวัดความรู้บางส่วนของผู้ตอบแบบสอบถามเลือกตอบ* วิทยานิพนธ์ดุษฎีบัณฑิต, จุฬาลงกรณ์มหาวิทยาลัย.
- อรพรรณ แก้วน้อย. (2557). *การเปรียบเทียบคุณสมบัติทางจิตมิติของแบบสอบถูกผิดหลายตัวเลือกกับแบบสอบถูกผิดหลายตัวเลือกปรับปรุงใหม่ที่ใช้วิธีการตรวจให้คะแนนที่ต่างกัน*, วิทยานิพนธ์มหาบัณฑิต, จุฬาลงกรณ์มหาวิทยาลัย.
- เอมอร จังศิริพรกรณ์. (2545). *การเปรียบเทียบคุณภาพของแบบสอบเลือกตอบเมื่อตรวจด้วยวิธีการให้คะแนนความรู้บางส่วนกับวิธีประเพณีนิยม*. Chulalongkorn University Intellectual Repository (CUIR). <http://cuir.car.chula.ac.th/handle/123456789/2181>.
- Baghaei, P. & Dourakhshan, A. (2016). Properties of Single-Response and Double-Response Multiple-Choice Grammar Items. *International Journal of Language Testing*, 6(1), 33-42.
- Custer, M., Sharairi, S., & Swift, D. (2012, April, 12-16). *A Comparison of Scoring Options for Omitted and Not-Reached Items Through the Recovery of IRT Parameters When Utilizing the Rasch Model and Joint Maximum Likelihood Estimation*, Paper presentation. The Annual Meeting of the Nation Council on Measurement in Education, Vancouver, British Columbia.
- Grandy, J. (1987). *Characteristics of examinees who leave questions unanswered*

- on the GRE general test under right-only scoring.,GRE Board Professional Report. Princeton, NJ: Educational Testing Service.
- McGahee, T. & Ball, J. (2009). How to read and really use an item analysis. **Nurse Educator**, 34(4), 166-171
- Siddiqui, N.I., Bhavsar, V.H., Bhavsar, A.V., & Bose, S. (2016). Contemplation on marking scheme for Type X multiple choice questions, and an illustration of a practically applicable scheme. **Indian Journal of Pharmacology**, 48(2), 114-21.
- Tsai, F., Suen, H.K., (1993) A brief report on a comparison of six scoring methods for multiple true-false items. **Educational and Psychological Measurement**, 53, 399-404.