

## A Comparison of the Quality of Multiple-Choice Tests Under Various Number-Right Options, Response Methods, and Scoring Procedures

Sirichai Kanjanawasee

### ABSTRACT

*The purpose of this research is to compare the quality of multiple-choice tests under various number-right options, response methods, and scoring procedures. Conditions of multiple-choice tests in this study were single/multiple right options; response methods of choosing a set of right options/ deleting distracters, and traditional/partial knowledge scoring procedures. The samples were 233 junior Education students enrolling Educational Measurement and Evaluation Course in academic year 2003. Data were collected 4 times though 4 various multiple-choice test conditions: the first and second time with each two multiple-choice test conditions; and the third and last time with midterm and final tests. Item and test analysis were conducted through CTT and Polytomous IRT procedures with PARSCALE for window. Multivariate repeated measure was analyzed via SPSS for Window. Major findings were as follows:*

*1. There was interaction effect between multiple-choice test conditions and scoring procedures. The different multiple-choice test conditions using traditional scoring procedure effected total observed scores of respondents while the different multiple-choice test conditions using partial knowledge scoring procedure had constant effect on total observed scores of respondents.*

*2. Multiple-choice test with single right option and response method of choosing a set of right options using partial knowledge scoring procedure had higher predictive validity than using traditional scoring procedure.*

*3. Multiple-choice test with single right option and response method of choosing a set of right options using partial knowledge scoring procedure provided maximum average item information function and test information function. Its relative efficiency was approximately 1.60 times higher than other multiple-choice test conditions.*

# การเปรียบเทียบคุณภาพของแบบสอบถามเลือกตอบที่มีจำนวน ตัวถูก วิธีการตอบและวิธีการตรวจให้คะแนนที่ต่างกัน

ศิริชัย กาญจนวาสี

## บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบคุณภาพของแบบสอบถามเลือกตอบที่มีจำนวนตัวถูก วิธีการตอบและวิธีการตรวจให้คะแนนที่ต่างกัน รูปแบบของแบบสอบถามเลือกตอบที่ใช้ในการวิจัยประกอบด้วยจำนวนตัวถูกเดียว/ตัวถูกหลายตัว วิธีการตอบแบบเลือกชุดตัวถูก/ตัดตัวลง และใช้วิธีการตรวจให้คะแนนแบบประเพณีนิยม/ให้คะแนนความรู้บางส่วน กลุ่มตัวอย่างที่ใช้ในการวิจัยได้แก่ นิสิตคณะครุศาสตร์ ชั้นปีที่ 3 จากทุกสาขาวิชาที่ลงทะเบียนเรียนวิชาการวัดและประเมินผลทางการศึกษา ปีการศึกษา 2546 จำนวน 233 คน ทำการเก็บรวบรวมข้อมูลจากกลุ่มตัวอย่างรวม 4 ครั้ง ด้วยแบบสอบถามเลือกตอบที่มีรูปแบบต่างกัน 4 ฉบับ ทำครั้งละ 2 ฉบับ อย่างสุ่ม จำนวน 2 ครั้ง แล้วทำการทดสอบกลางภาค และปลายภาคเรียน ข้อมูลที่ได้นำมาวิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิม และทฤษฎีการตอบสนองข้อสอบแบบพหุวิภาค โดยใช้โปรแกรม PARSCALE for Window และเปรียบเทียบความแตกต่างด้วยการวิเคราะห์ความแปรปรวนแบบวัดซ้ำด้วยสถิติทดสอบเอฟพหุตัวแปร โดยใช้โปรแกรม SPSS for Window ผลการวิจัยสรุปได้ดังนี้

1. มีปฏิสัมพันธ์ระหว่างรูปแบบของแบบสอบถามเลือกตอบกับวิธีการตรวจให้คะแนนโดยรูปแบบของแบบสอบถามเลือกตอบที่ต่างกันเมื่อใช้วิธีการตรวจให้คะแนนแบบประเพณีนิยม ส่งผลต่อคะแนนของผู้สอบ แต่รูปแบบของแบบสอบถามเลือกตอบที่ต่างกันเมื่อใช้วิธีการตรวจแบบให้คะแนนความรู้บางส่วน ส่งผลคงที่ต่อคะแนนของผู้สอบ
2. แบบสอบถามเลือกตอบที่มีตัวถูกเดียวและตอบแบบเลือกชุดตัวถูก เมื่อใช้วิธีการตรวจแบบให้คะแนนความรู้บางส่วน มีความตรงเชิงทำนายสูงกว่าวิธีการตรวจให้คะแนนแบบประเพณีนิยม
3. แบบสอบถามเลือกตอบที่มีตัวถูกเดียวและตอบแบบเลือกชุดตัวถูก เมื่อใช้วิธีการตรวจแบบให้คะแนนความรู้บางส่วน มีค่าฟังก์ชันสารสนเทศเฉลี่ยของข้อสอบและค่าฟังก์ชันสารสนเทศของแบบสอบถามรวมสูงสุด โดยมีประสิทธิภาพสัมพัทธ์สูงกว่าแบบสอบถามเลือกตอบรูปแบบอื่นโดยเฉลี่ย 1.60 เท่า

## ความเป็นมาและความสำคัญของปัญหา

แบบสอบถามตอบยังคงมีบทบาทสำคัญในการใช้เป็นเครื่องมือวัดผลการเรียนรู้ได้อย่างกว้างขวาง ตลอดจนใช้วัดความรู้ ความสามารถ และผลสัมฤทธิ์ของผู้สอบสำหรับการคัดเลือก วินิจฉัย และการพัฒนาบุคลากร ทั้งนี้เนื่องจากแบบสอบถามมีข้อดีหลายประการ ได้แก่ สามารถสร้างคำถามได้ครอบคลุมเนื้อหาและจุดมุ่งหมายของการเรียนรู้ วัดความสามารถทางสมองได้ตั้งแต่ขั้นต่ำจนถึงขั้นสูง ช่วยส่งเสริมทักษะการอ่าน การคิดและการเรียนรู้ สามารถนำมาใช้ทดสอบเพื่อวินิจฉัยปัญหาการเรียนรู้ ตรวจให้คะแนนง่าย ใช้เวลาตรวจน้อย สามารถตรวจโดยใช้คอมพิวเตอร์ การวิเคราะห์ข้อสอบและแบบสอบทำได้ง่ายและสะดวกโดยใช้โปรแกรมสำเร็จรูป สามารถนำมาหมุนเวียนใช้ใหม่ได้ นิยมพัฒนาเป็นแบบสอบมาตรฐานที่มีความเที่ยงและความตรงตามเนื้อหาสูง และมีประสิทธิภาพของการวัดได้ดีกว่าแบบสอบถามแบบอื่น ๆ (ศิริชัย กาญจนวาสี, 2544) อย่างไรก็ตามแบบสอบถามตอบก็มีข้อจำกัดอยู่หลายประการ ได้แก่ การสร้างข้อสอบให้มีคุณภาพจำเป็นต้องใช้ผู้ทรงคุณวุฒิที่มีความเชี่ยวชาญในเนื้อหาและมีทักษะสูงในการเขียนข้อสอบ การเลือกรูปแบบที่เหมาะสมในการเขียนข้อสอบยังขาดแนวทางปฏิบัติที่ชัดเจน เนื่องจากมีการศึกษาวิจัยในวงจำกัด จุดอ่อนที่สำคัญอีกประการหนึ่ง คือ เปิดโอกาสให้ผู้สอบตอบถูกโดยการเดา การศึกษาวิจัยถึงกลยุทธ์ในการออกแบบวิธีการตอบที่มีประสิทธิภาพในการลดโอกาสการเดาถูก ยังมีไม่มากนัก นอกจากนี้วิธีการตรวจให้คะแนนที่นิยมใช้กันคือ การตรวจแบบประเพณีนิยม (ตอบถูกได้ 1 คะแนน ตอบผิดได้ 0 คะแนน) นั้นให้สารสนเทศเกี่ยวกับความรู้ความสามารถของผู้เรียนน้อยมาก ไม่เพียงพอสำหรับจำแนกระดับความสามารถของผู้สอบระหว่างผู้ที่รู้จริง (Full knowledge) ผู้ที่มีความรู้บางส่วน (Partial knowledge) และผู้ที่ไม่มีความรู้ (Absence of knowledge) ดังนั้นปัญหาการใช้แบบสอบถามตอบที่ประสบอยู่ขณะนี้ คือ การเลือกใช้รูปแบบของแบบสอบถามตอบที่เหมาะสม การออกแบบวิธีการตอบที่ลดโอกาสของการเดาถูก และวิธีการตรวจให้คะแนนอย่างไรจึงจะทำให้เกิดความคลาดเคลื่อนของการวัดน้อยที่สุด และคะแนนผลการสอบสามารถให้สารสนเทศที่ถูกต้องตรงตามความเป็นจริง ตามระดับความรู้ความสามารถของผู้สอบมากที่สุด

ในการวัดผลการเรียนรู้ของผู้เรียนโดยใช้แบบสอบถามตอบ ผู้วิจัยมีแนวคิดเชิงระบบว่า ผู้สอบ (ซึ่งมีความรู้ความสามารถระดับหนึ่ง) เมื่อทำแบบสอบและมีปฏิสัมพันธ์กับข้อสอบแล้ว “รูปแบบของแบบสอบ กระบวนการคิดวิเคราะห์ของผู้สอบ วิธีการตอบข้อสอบ และวิธีการตรวจให้คะแนน ย่อมมีความสัมพันธ์เชื่อมโยงกันและส่งผลต่อคะแนนสอบที่ผู้สอบพึงได้รับ และคุณภาพของคะแนนผลการวัด” การศึกษาวิจัยคุณภาพของแบบสอบถามตอบจึงควรพิจารณาครอบคลุมปัจจัยที่เกี่ยวข้องกับกระบวนการทดสอบทั้งระบบดังกล่าว

จากการศึกษาทฤษฎีและงานวิจัยที่ผ่านมา มีการศึกษาวิจัยเกี่ยวกับแบบสอบเลือกตอบที่หลากหลาย ส่วนใหญ่ศึกษาเปรียบเทียบปัจจัยเฉพาะด้านที่ส่งผลต่อคุณภาพของแบบสอบเลือกตอบ ประกอบด้วยปัจจัยหลักดังต่อไปนี้

1. จำนวนตัวถูก ส่วนใหญ่ศึกษาแบบสอบเลือกตอบที่ข้อสอบมีจำนวนตัวถูก 1 ตัว (One answer option) เช่น การศึกษาของ Michael (1968), Hambleton และคณะ (1970) การศึกษาแบบสอบเลือกตอบที่ข้อสอบมีจำนวนตัวถูกมากกว่า 1 ตัว (Multiple answer options) ยังมีการศึกษาน้อยมาก เช่น การศึกษาของ Fisbie (1992) Kolstad และคณะ (1983), Frary (1989), กาญจนา ศิริวัฒน์พงษ์ (2520) เป็นต้น

2. วิธีตอบ มี 2 แนวทางสำคัญ ได้แก่ วิธีการเลือกชุดตัวถูก (Subset selection method) เช่น การศึกษาของ Dressel และ Schmid (1953), Jaradat และ Swagad (1986) และวิธีการตัดตัวถูก (Elimination method) เช่น การศึกษาของ Coombs และคณะ (1956), Arnold และ Arnold (1970), Tollefson และ Chung (1986), สำราญ มีแจ้ง (2525) เป็นต้น

3. วิธีการตรวจให้คะแนน มี 2 วิธีการสำคัญ ได้แก่ วิธีการตรวจให้คะแนนแบบประเพณีนิยม (Traditional Scoring) เช่น การศึกษาของ Davis และ Fifer (1959), Rippey (1970), Hopkins และคณะ (1973) เป็นต้น และวิธีการตรวจให้คะแนนแบบให้คะแนนความรู้บางส่วน (Partial Credit Scoring) เช่น การศึกษาของ Albert (1970), Smith (1987) Simon, Budescu และ Nevo (1997), พรทิพย์ ไชยใส (2533), สุพจน์ เกิดสุวรรณ (2545), เอมอร จังศิริพรภรณ์ (2546) เป็นต้น

จากรอบแนวคิดเชิงระบบเกี่ยวกับคุณภาพของแบบสอบเลือกตอบ และการศึกษาค้นคว้าผลการวิจัยที่ผ่านมา ผู้วิจัยจึงมีความสนใจศึกษาเปรียบเทียบคุณภาพของแบบสอบเลือกตอบที่มีจำนวนตัวถูก วิธีการตอบและวิธีการตรวจให้คะแนนที่ต่างกัน เพื่อค้นหารูปแบบของแบบสอบเลือกตอบทางเลือกที่มีคุณภาพสูงกว่ารูปแบบประเพณีนิยม โดยใช้กรอบการวิเคราะห์เปรียบเทียบคุณภาพของแบบสอบทั้งตามแนวทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory) และตามแนวทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) สำหรับการตรวจให้คะแนนแบบพหุวิภาค (Polytomous) ตรวจโดยใช้โมเดลการวิเคราะห์ความรู้บางส่วน (Generalized Partial Credit Model: GPCM)

## วัตถุประสงค์ของการวิจัย

การวิจัยนี้มีวัตถุประสงค์ทั่วไป เพื่อเปรียบเทียบคุณภาพของแบบสอบเลือกตอบที่มีจำนวนตัวถูก วิธีตอบ และวิธีการตรวจให้คะแนนที่ต่างกัน โดยมีวัตถุประสงค์เฉพาะดังนี้

1. เพื่อเปรียบเทียบค่าความยากและอำนาจจำแนกของแบบสอบเลือกตอบคู่ขนานที่มีจำนวนตัวถูกตัวเดียวและตัวถูกหลายตัว วิธีการตอบแบบเลือกชุดตัวถูกและตัดตัวลง และวิธีการตรวจให้คะแนนแบบประเพณีนิยม และให้คะแนนความรู้บางส่วน

2. เพื่อเปรียบเทียบค่าความเที่ยง ความตรงตามสภาพ และความตรงเชิงทำนายของแบบสอบเลือกตอบคู่ขนานที่มีจำนวนตัวถูกตัวเดียวและตัวถูกหลายตัว วิธีการตอบแบบเลือกชุดตัวถูกและตัดตัวลง และวิธีการตรวจให้คะแนนแบบประเพณีนิยมและให้คะแนนความรู้บางส่วน

3. เพื่อเปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบสอบ และอัตราส่วนสารสนเทศเฉลี่ยของแบบสอบเลือกตอบคู่ขนานที่มีจำนวนตัวถูกตัวเดียวและตัวถูกหลายตัว วิธีการตอบแบบเลือกชุดตัวถูกและตัดตัวลงเมื่อใช้วิธีการตรวจให้คะแนนแบบให้คะแนนความรู้บางส่วน

### สมมติฐานของการวิจัย

1. แบบสอบเลือกตอบที่มีจำนวนตัวถูก วิธีการตอบ และวิธีการตรวจที่ต่างกัน น่าจะมีระดับความยากเฉลี่ยของข้อสอบ และอำนาจจำแนกเฉลี่ยของข้อสอบแตกต่างกัน

2. แบบสอบเลือกตอบที่มีจำนวนตัวถูก และวิธีการตอบที่ต่างกัน เมื่อใช้วิธีการตรวจแบบให้คะแนนความรู้บางส่วน น่าจะมีค่าความเที่ยง ความตรงตามสภาพ และความตรงเชิงทำนายสูงกว่า เมื่อใช้วิธีการตรวจแบบประเพณีนิยม

3. แบบสอบเลือกตอบที่มีจำนวนตัวถูก และวิธีการตอบที่ต่างกัน เมื่อใช้วิธีการตรวจแบบให้คะแนนความรู้บางส่วน แบบสอบเลือกตอบที่มีตัวถูกเดียว และตอบแบบเลือกชุดตัวถูก น่าจะมีค่าฟังก์ชันสารสนเทศเฉลี่ยของข้อสอบ ค่าฟังก์ชันสารสนเทศของแบบสอบรวมสูงสุด และมีค่าความคลาดเคลื่อนมาตรฐานต่ำสุด

### ขอบเขตของการวิจัย

1. รูปแบบของแบบสอบเลือกตอบชนิด 4 ตัวเลือก ประกอบด้วยลักษณะที่ต่างกันของข้อสอบ 3 ด้าน ได้แก่ จำนวนคำตอบถูกของข้อสอบ วิธีการตอบข้อสอบ และวิธีการตรวจให้คะแนนข้อสอบ

#### 1.1 จำนวนตัวถูกของข้อสอบ

- 1) ตัวถูกเดียว: ข้อสอบทุกข้อมีคำตอบถูก 1 ตัวเลือก
- 2) ตัวถูกหลายตัว: ข้อสอบแต่ละข้อมีคำตอบถูก 1-3 ตัว

### 1.2 วิธีการตอบข้อสอบ

- 1) เลือกตัวเลือก: ผู้สอบเลือกตัวเลือก อาจเลือกชุดตัวเลือกจำนวน 1, 2, 3 หรือ 4 ตัว
- 2) ตัดตัวเลือก: ผู้สอบตัดตัวเลือก อาจตัดชุดตัวเลือกจำนวน 1, 2, 3 หรือ 4 ตัว

### 1.3 วิธีการตรวจให้คะแนนข้อสอบ

- 1) วิธีการตรวจแบบประเพณีนิยม: ผู้สอบตอบถูกได้คะแนนเต็ม แต่ถ้าตอบผิดได้ 0 คะแนน
- 2) วิธีการตรวจแบบให้คะแนนความรู้บางส่วน
  - วิธีตอบแบบเลือกชุดตัวเลือก: ตรวจให้คะแนนความรู้บางส่วน โดยประยุกต์วิธีการตรวจของ Dressel & Schmidt (1953)
  - วิธีตอบแบบตัดตัวเลือก: ตรวจให้คะแนนความรู้บางส่วนโดยประยุกต์วิธีการตรวจของ Coombs และคณะ (1956)

## 2. คุณภาพของข้อสอบและแบบสอบเลือกตอบ ตรวจสอบโดยใช้ตัวบ่งชี้ตามทฤษฎีการทดสอบแบบดั้งเดิม และทฤษฎีการตอบสนองข้อสอบ (Item Response Theory)

### 2.1 คุณภาพของข้อสอบและแบบสอบตามทฤษฎีการทดสอบแบบดั้งเดิม

- 1) ค่าความยากของข้อสอบ ( $p$ )
- 2) ค่าอำนาจจำแนกของข้อสอบ ( $r$ )
- 3) ค่าความเที่ยงของข้อสอบ (Reliability)
- 4) ค่าความตรงของแบบสอบ (Validity) โดยพิจารณาทั้งความตรงตามสภาพ (Concurrent Validity) และความตรงเชิงทำนาย (Predictive Validity)

### 2.2 คุณภาพของข้อสอบและแบบสอบตามทฤษฎีการตอบสนองข้อสอบ

- 1) ค่าความยากของข้อสอบ ( $\alpha$ )
- 2) ค่าอำนาจจำแนกของข้อสอบ ( $\beta$ )
- 3) ค่าฟังก์ชันสารสนเทศของข้อสอบ (Item Information Function: IIF)
- 4) ค่าฟังก์ชันสารสนเทศของแบบสอบ (Test Information Function: TIF)
- 5) อัตราส่วนสารสนเทศเฉลี่ยของแบบสอบ (Ratio of Average Information: RAI)

วิเคราะห์ตามทฤษฎีการตอบสนองแบบพหุวิภาค (Polytomous IRT) โดยใช้โมเดลการให้คะแนนบางส่วน (Partial Credit Model: PCM)

## ประชากรและกลุ่มตัวอย่าง

ประชากรของการวิจัย ได้แก่ นิสิตนักศึกษาระดับปริญญาบัณฑิต กลุ่มตัวอย่างในการวิจัยในครั้งนี้คือ นิสิตคณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ชั้นปีที่ 3 ทุกสาขาวิชาที่ลงทะเบียนเรียนวิชาการวัดและประเมินผลทางการศึกษา ประจำภาคต้น ปีการศึกษา 2546 จำนวน 250 คน ในการเก็บรวบรวมข้อมูลจากกลุ่มตัวอย่าง 4 ครั้ง มีจำนวนนิสิตที่เป็นกลุ่มตัวอย่างทำแบบสอบครบทั้ง 4 ครั้ง จำนวน 223 คน คิดเป็นร้อยละ 89.2

## เครื่องมือสำหรับการวิจัย

เครื่องมือที่ใช้ในการวิจัย ประกอบด้วย แบบสอบเลือกตอบสำหรับการทดลองและแบบสอบเลือกตอบสำหรับใช้เป็นเกณฑ์เปรียบเทียบ แบบสอบเลือกตอบสำหรับการทดลองมีลักษณะเป็นแบบสอบเลือกตอบคู่ขนาน 4 ฉบับ แต่ละฉบับเป็นแบบสอบเลือกตอบชนิด 4 ตัวเลือก จำนวน 25 ข้อ ซึ่งมีรูปแบบแตกต่างกัน 4 รูปแบบ ดังนี้ 1) แบบสอบเลือกตอบที่มีตัวถูกเดียวและตอบแบบเลือกชุด ตัวถูก (TEST A) 2) แบบสอบเลือกตอบที่มีตัวถูกเดียวและตอบแบบตัดตัววง (TEST B) 3) แบบสอบเลือกตอบที่มีตัวถูกหลายตัวและตอบแบบเลือกชุดตัวถูก (TEST C) และ 4) แบบสอบเลือกตอบที่มีตัวถูกหลายตัวและตอบแบบตัดตัววง (TEST D) สำหรับแบบสอบเลือกตอบที่ใช้เป็นเกณฑ์เปรียบเทียบ ได้แก่ แบบสอบกลางภาคและแบบสอบปลายภาค วิชาการวัดและประเมินผลทางการศึกษา ซึ่งมีความตรงตามเนื้อเรื่อง (Content validity) และมีความเที่ยง (Reliability) เท่ากับ 0.73 และ 0.82 ตามลำดับ

ผลการตอบแบบสอบเลือกตอบที่ใช้สำหรับการทดลองทั้ง 4 ฉบับ ได้รับการตรวจให้คะแนนแบบประเพณีนิยม และแบบให้คะแนนความรู้บางส่วน ในการตรวจแบบประเพณีนิยม ยึดหลักว่าถ้าผู้สอบเลือกตัวเลือกได้ตรงคำตอบจะได้คะแนนเต็มของข้อนั้น แต่ถ้าตอบผิดจะได้ 0 คะแนน ส่วนการตรวจแบบให้คะแนนความรู้บางส่วน สำหรับแบบสอบที่ใช้วิธีการตอบแบบเลือกชุดตัวถูก (TEST A และ TEST C) ทำการตรวจให้คะแนนความรู้บางส่วนโดยประยุกต์วิธีการของ Dressel และ Schmid (1953) สำหรับแบบสอบที่ใช้วิธีการตอบแบบตัดตัววง (TEST B และ TEST D) ทำการตรวจให้คะแนนความรู้บางส่วนโดยประยุกต์วิธีของ Coombs และคณะ (1953)

## การวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลสำหรับการวิจัยนี้ ใช้การวิเคราะห์ข้อสอบและแบบสอบตามทฤษฎีการทดสอบแบบดั้งเดิม เพื่อประมาณค่าความยาก ( $p$ ) อำนาจจำแนก ( $r$ ) ของข้อสอบ ค่าความเที่ยง ความตรงตามสภาพ และความตรงเชิงทำนายของแบบสอบ ตลอดจนใช้การวิเคราะห์ข้อสอบ

และแบบสอบถามตามทฤษฎีการตอบสนองข้อสอบแบบพหุวิภาค (Polytomous IRT) เพื่อประมาณค่าความยาก ( $\alpha$ ) อำนาจจำแนก ( $\beta$ ) ค่าฟังก์ชันสารสนเทศของข้อสอบ (IIF) ค่าฟังก์ชันสารสนเทศของแบบสอบ (TIF) และอัตราส่วนสารสนเทศเฉลี่ยของแบบสอบ (RAI) โดยใช้โปรแกรม PARSCALE (Muraki & Bock, 1993) สำหรับการเปรียบเทียบความแตกต่างของค่าเฉลี่ยความยาก อำนาจจำแนก คะแนนสอบ คะแนนความสามารถ และค่าฟังก์ชันสารสนเทศเฉลี่ยของข้อสอบ ใช้สถิติทดสอบเอฟแบบพหุตัวแปร (Multivariate F-Test) สำหรับการวิเคราะห์ความแปรปรวนแบบวัดซ้ำ (Repeated ANOVA) โดยใช้โปรแกรม SPSS for Window version 12.0

## สรุปผลการวิจัย

### ผลการวิจัยสามารถสรุปได้ดังนี้

1. แบบสอบถามเลือกตอบที่มีรูปแบบแตกต่างกันของจำนวนตัวถูกและวิธีการตอบ 4 รูปแบบ เมื่อตรวจให้คะแนนแบบประเพณีนิยมและทำการวิเคราะห์ข้อสอบตามทฤษฎีการทดสอบแบบดั้งเดิม พบว่า ข้อสอบมีค่าเฉลี่ยความยาก ( $p$ ) และค่าเฉลี่ยอำนาจจำแนก ( $r$ ) แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01

2. แบบสอบถามเลือกตอบที่มีรูปแบบแตกต่างกันของจำนวนตัวถูก และวิธีการตอบ 4 รูปแบบ เมื่อใช้วิธีการตรวจให้คะแนนแตกต่างกัน มีผลให้คะแนนที่ได้จากแบบสอบแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 โดยพบปฏิสัมพันธ์ระหว่างรูปแบบของแบบสอบกับวิธีการตรวจให้คะแนน แสดงว่า รูปแบบของแบบสอบถามเลือกตอบที่มีจำนวนตัวถูกและวิธีการตอบต่างกันมีผลต่อคะแนนสอบของผู้สอบกลุ่มเดียวกันตามวิธีการตรวจให้คะแนน ถ้าใช้วิธีการตรวจแบบให้คะแนนความรู้บางส่วน รูปแบบของแบบสอบจะไม่มีผลต่อคะแนนของผู้สอบ ผู้สอบจะได้คะแนนเฉลี่ยสูงกว่าวิธีการตรวจแบบประเพณีนิยม แต่ถ้าใช้วิธีการตรวจแบบประเพณีนิยม รูปแบบของแบบสอบจะมีผลต่อคะแนนของผู้สอบ

3. แบบสอบถามเลือกตอบที่มีรูปแบบแตกต่างกันของจำนวนตัวถูกและวิธีการตอบ 4 รูปแบบ เมื่อใช้วิธีการตรวจให้คะแนนแตกต่างกัน จะมีผลต่อความเที่ยง ความตรงตามสภาพและความตรงเชิงทำนายดังนี้

3.1 แบบสอบถามเลือกตอบที่มีตัวถูกเดียวและตอบแบบเลือกชุดตัวถูก เมื่อตรวจแบบให้คะแนนความรู้บางส่วน จะมีค่าความตรงเชิงทำนายสูงกว่าการตรวจแบบประเพณีนิยม

3.2 แบบสอบถามเลือกตอบที่มีตัวถูกเดียวและตอบแบบตัดตัวลง เมื่อตรวจแบบให้คะแนนความรู้บางส่วน จะมีค่าความเที่ยง ความตรงตามสภาพและความตรงเชิงทำนายไม่แตกต่างจากการตรวจแบบประเพณีนิยม



3.3 แบบสอบเลือกตอบที่มีตัวถูกหลายตัว ซึ่งตอบแบบเลือกชุดตัวถูกหรือตอบแบบตัดตัวลง จะมีค่าความเที่ยงสูงกว่าการตรวจแบบประเพณีนิยม

4. แบบสอบเลือกตอบที่มีรูปแบบแตกต่างกันของจำนวนตัวถูก และวิธีการตรวจ 4 รูปแบบ เมื่อตรวจให้คะแนนโดยวิธีให้คะแนนความรู้บางส่วน และทำการวิเคราะห์ข้อสอบและแบบสอบตามทฤษฎีการตอบสนองข้อสอบแบบพหุวิภาค (Polytomous IRT) พบว่า

4.1 ข้อสอบมีค่าเฉลี่ยความยาก ( $\alpha$ ) และค่าเฉลี่ยอำนาจจำแนก ( $\beta$ ) แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01

4.2 ข้อสอบมีค่าฟังก์ชันสารสนเทศเฉลี่ยของข้อสอบแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 โดยแบบสอบเลือกตอบที่มีตัวถูกเดี่ยวและตอบแบบเลือกชุดตัวถูกมีค่าฟังก์ชันสารสนเทศเฉลี่ยของข้อสอบสูงกว่าแบบสอบเลือกตอบที่มีตัวถูกหลายตัว ซึ่งตอบแบบเลือกชุดตัวถูกหรือตอบแบบตัดตัวลง

4.3 แบบสอบมีค่าฟังก์ชันสารสนเทศรวมแตกต่างกัน โดยแบบสอบเลือกตอบที่มีตัวถูกเดี่ยวและตอบแบบเลือกชุดตัวถูก มีค่าฟังก์ชันสารสนเทศของแบบสอบรวมสูงสุด และมีค่าความคลาดเคลื่อนมาตรฐานต่ำสุด

4.4 จากการเปรียบเทียบอัตราส่วนสารสนเทศเฉลี่ยระหว่างแบบสอบ พบว่า แบบสอบเลือกตอบที่มีตัวถูกเดี่ยวและตอบแบบเลือกชุดตัวถูกมีประสิทธิภาพสูงสุด โดยมีประสิทธิภาพสัมพัทธ์สูงกว่าแบบสอบเลือกตอบรูปแบบอื่นโดยเฉลี่ย 1.60 เท่า

## ข้อเสนอแนะ

1. จากการพบปฏิสัมพันธ์ระหว่างรูปแบบของแบบสอบกับวิธีการตรวจให้คะแนนที่มีผลต่อคุณภาพของคะแนนสอบ ดังนั้นการใช้แบบสอบเลือกตอบจึงควรออกแบบการสร้างข้อสอบให้มีลักษณะตัวถูกเดี่ยวที่ใช้วิธีตอบแบบเลือกชุดตัวถูก และทำการตรวจแบบให้คะแนนความรู้บางส่วน

2. ควรส่งเสริมการใช้แบบสอบเลือกตอบชนิดที่มีตัวถูกเดี่ยว ใช้วิธีตอบแบบเลือกชุดตัวถูกและตรวจแบบให้คะแนนความรู้บางส่วน โดยประยุกต์วิธีตรวจให้คะแนนของ Dressel และ Schmid จะทำให้ได้สารสนเทศสำคัญที่มีผลต่อคุณภาพของแบบสอบเลือกตอบสูงสุด อันจะเป็นประโยชน์สำหรับการวัดความสามารถที่แท้จริงของผู้สอบ ตลอดจนเป็นแนวปฏิบัติในการสร้างเครื่องมือที่มีคุณภาพสำหรับนำผลการวัดไปใช้เพื่อการตัดสินใจทางการศึกษาได้อย่างเที่ยงตรงดังนี้

ในกรณีที่ผู้สอบเลือกชุดคำตอบที่มีตัวถูกอยู่ด้วย จะได้คะแนนดังนี้

$$X = 2n_r - (k - 2)$$

เมื่อ  $X =$  คะแนนที่ได้

♦ การเปรียบเทียบคุณภาพของแบบสอบเลือกตอบที่มีจำนวนตัวถูก วิธีการตอบและวิธีการตรวจให้คะแนนที่ต่างกัน ♦

$n_R$  = จำนวนตัวเลือกที่ถูกตรงคำตอบ (กรณีนี้  $n_R = 1$ )

$k$  = จำนวนตัวเลือกที่ผู้ตอบเลือก

สำหรับกรณีที่ผู้สอบเลือกชุดคำตอบที่ไม่มีตัวถูกอยู่ด้วยจะได้ 0 คะแนน

## รายการอ้างอิง

- กาญจนา ศิริวัฒน์พงษ์. (2520). **การศึกษาเปรียบเทียบการตอบและการตรวจให้คะแนนแบบทดสอบปรนัยชนิดเลือกตอบที่มีลักษณะแตกต่างกัน**. วิทยานิพนธ์ปริญญาการศึกษา มหาวิทยาลัย บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒประสานมิตร.
- พรทิพย์ ไชยโส. (2533). **การพัฒนาสูตรการให้คะแนนแบบเลือกตอบสำหรับความรู้อย่างส่วนของผู้ตอบ: การประยุกต์ใช้วิธีการอาร์โนลด์ และวิธีการของแฮมตัน**. วิทยานิพนธ์ครุศาสตร์ ดุษฎีบัณฑิต บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี. (2544). **ทฤษฎีการทดสอบแบบดั้งเดิม**. (พิมพ์ครั้งที่ 5) กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี. (2545). **ทฤษฎีการทดสอบแนวใหม่**. (พิมพ์ครั้งที่ 2) กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- สุพจน์ เกิดสุวรรณ. (2545). **การพัฒนาการวัดความรู้บางส่วนของผู้ตอบแบบสอบเลือกตอบ**. วิทยานิพนธ์ครุศาสตร์ดุษฎีบัณฑิต บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- สำราญ มีแจ้ง. (2525). **ผลการใช้คำสั่งและการให้คะแนนที่ต่างกันต่อค่าความเที่ยง ความตรง และอำนาจจำแนกของแบบทดสอบชนิดเลือกตอบ**. วิทยานิพนธ์ครุศาสตร์มหาบัณฑิต บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- เอมอร จังศิริพรปกรณ์. (2546). **การเปรียบเทียบคุณภาพของแบบสอบเลือกตอบระหว่างวิธีการให้คะแนนความรู้บางส่วนกับวิธีประเพณีนิยม เอกสารอัดสำเนา**. ภาควิชาวิจัย การศึกษา จุฬาลงกรณ์มหาวิทยาลัย.
- Albert, W. G. (1970). Scoring for partial knowledge in mathematic testing: A study of modification and extension of multiple-choice item applied to the testing of achievement in mathematic. **Dessertation Abstracts International**, 31(4), 1619A-1620A.
- Arnold, J. C., and Arnold, P. L. (1970). On scoring multiple choice exams allowing for partial knowledge. **The Journal of Experimental Education**, 39, 8-13.

- Coombs, C. H., Milholland, J. E., and Womer, F. B. (1956). The assessment of partial knowledge. **Educational and Psychological Measurement**, 16, 13-37.
- Dressel, P. L., and Schmid, J. (1953). Some modifications of the multiple choice item. **Educational and Psychological Measurement**, 13, 574-595.
- Fisbie, D. A. (1992). The multiple true - false format: A status review. **Educational Measurement : Issues and Practice**, 11(4), 21-26.
- Frary, R. B. (1980). The effect of misinformation, partial information, and guessing on expected multiple-choice test item scores. **Applied Psychological Measurement**, 4, 79-90.
- Hambleton, R. K., Roberts, D. M., and Traub, R. E. (1970). A comparison of the reliability and validity of two methods for assessing partial knowledge on a multiple choice test. **Journal of Educational Measurement**, 7, 75-82.
- Hopkins, K. K., Hakstian, A. R., and Hopkins, B. R. (1973). Validity and reliability consequences of confidence weighting. **Educational and Psychological Measurement**, 33, 135-141.
- Jaradat, D., and Swagad, S. (1986). The subset selection technique for multiple choice tests: An empirical inquiry. **Journal of Educational Measurement**, 23, 369-376.
- Jaradat, D., and Tollefson, N. (1986). The subset selection technique for multiple choice tests: An empirical inquiry. **Journal of Educational Measurement**, 23, 369-376.
- Michael, J. C. (1968). The reliability of a multiple choice examination under various test-taking instructions. **Journal of Educational Measurement**, 5, 307-314.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM Algorithm. **Applied Psychological Measurement**, 16(2), 159-176.
- Muraki, E. (1993). Information functions of the generalized parital credit model. **Applied Psychological Measurement**, 17(4), 351-363.
- Muraki, E., and Bock, R. D. (1993). **PARSCALE: IRT based test scoring and item analysis for graded open-ended exercises and performance tasks**. Chicago: Scientific Software International.

◆ การเปรียบเทียบคุณภาพของแบบสอบเลือกตอบที่มีจำนวนตัวถูก วิธีการตอบและวิธีการตรวจให้คะแนนที่ต่างกัน ◆

- Rippey, R. M. (1970). A comparison of five different scoring functions for confidence tests. **Journal of Educational Measurement**, 7, 165–170
- Simon, A. B., Budescu, D. V., and Nevo, B. (1997). A comparative study of measures of partial knowledge in multiple-choice tests. **Applied Psychological Measurement**, 21, 65–88.
- Smith, R. M. (1987). Assessing partial knowledge in vocabulary. **Journal of Educational Measurement**, 24, 217–231.
- Tollefson, N., and Chung, J. M. (1986). **A comparison of two methods of assessing partial knowledge on multiple choice**. Kansas University, Lawrence.