

คุณภาพของแบบทดสอบแบบเลือกตอบภายใต้เงื่อนไขที่มีการวัดระดับพฤติกรรม การเรียนรู้ วิธีการตรวจให้คะแนนความรู้บางส่วน และจำนวนตัวถูกที่แตกต่างกัน*

ณัชชา มหปญญานนท์¹

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบคุณภาพของแบบทดสอบคู่ขนานที่วัดระดับพฤติกรรม
การเรียนรู้ วิธีการตรวจให้คะแนนความรู้บางส่วนและจำนวนตัวถูกต่างกัน ในประเด็นความเที่ยงตรงตามสภาพ
ค่าความยากและค่าอำนาจจำแนกของข้อสอบ สารสนเทศของข้อสอบ สารสนเทศเฉลี่ยของแบบทดสอบ และ
ประสิทธิภาพสัมพัทธ์ของแบบทดสอบ โดยเงื่อนไขระดับพฤติกรรมการเรียนรู้จำแนกเป็นพฤติกรรมการเรียนรู้
ขั้นความจำ และขั้นสูงกว่าความจำ เงื่อนไขวิธีการตรวจให้คะแนนความรู้บางส่วนจำแนกเป็นแบบเลือกชุดตัว
ถูก (SST) และแบบตัดตัวลง (ET) และเงื่อนไขจำนวนตัวถูกจำแนกเป็น ตัวถูกตัวเดียวและตัวถูกมากกว่า 1 ตัว
กลุ่มตัวอย่างในการวิจัย ได้แก่ นิสิตคณะศึกษาศาสตร์ มหาวิทยาลัยทักษิณที่ลงทะเบียนรายวิชาการวัดและ
ประเมินผลการศึกษา ปีการศึกษา 2555 จำนวน 174 คน เครื่องมือที่ใช้ในการวิจัยครั้งนี้ เป็นแบบทดสอบ
คู่ขนานที่ผู้วิจัยพัฒนาขึ้นจากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน วิชาการวัดและประเมินผลการศึกษา
จำนวน 8 ชุด ๆ ละ 10 ข้อ

ผลการวิจัย พบว่า

1. ความเที่ยงตรงตามสภาพของแบบทดสอบคู่ขนานทั้ง 8 ชุด ในภาพรวมไม่แตกต่างกันอย่างมี
นัยสำคัญทางสถิติที่ระดับ .05
2. ค่าอำนาจจำแนกเฉลี่ยของแบบทดสอบทั้ง 8 ชุด ในภาพรวมและภาพย่อยรายเงื่อนไข ไม่แตกต่าง
กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05
3. ค่าความยากเฉลี่ยของแบบทดสอบทั้ง 8 ชุด ในภาพรวมแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่
ระดับ .01
4. แบบทดสอบแบบเลือกชุดตัวถูกวัดความรู้ขั้นความจำที่มีคำตอบถูกมากกว่า 1 ตัว มีค่าฟังก์ชัน
สารสนเทศของข้อสอบ, แบบทดสอบ และประสิทธิภาพสัมพัทธ์สูงสุด

คำสำคัญ: แบบทดสอบแบบเลือกตอบ, ทฤษฎีการตอบสนองข้อสอบ, วิธีการตรวจให้คะแนนความรู้บางส่วน,
จำนวนตัวถูกถูก

* บทความวิจัย คณะศึกษาศาสตร์ มหาวิทยาลัยทักษิณ

¹ อาจารย์, คณะศึกษาศาสตร์ มหาวิทยาลัยทักษิณ



Quality of Multiple Choice Tests under the Different Conditions on Level of Learning, Partial Knowledge Scoring Methods and Number of Correct Answers*

Natcha Mahapoonyanont¹

Abstract

The purpose of this research was to compare the tests' quality between 8 parallel tests which lied on the different conditions on level of learning, partial knowledge scoring methods and number of correct answers in the following aspects: criterion related validity, item difficulty and discrimination parameters, item information, test average information and ratio of average information. The level of learning comprised remembering and higher than remembering level; partial knowledge scoring methods comprised the subset selection method (SST) and the elimination method (ET), and the number of correct answers comprised single and multiple correct answers. Research subjects were 174 bachelor students in Faculty of Education, Thaksin University who registered Educational measurement and evaluation subject from the second semester in 2012. Research instruments were 8 parallel tests of educational measurement and evaluation with 10 items each.

The results of this study were as follows:

1. The criterion related validity between 8 parallel tests was not significant different at the .05 level.
2. The mean item discrimination parameter between 8 parallel tests was not significant different at the .05 level.
3. The mean item difficulty parameter between 8 parallel tests was statistically significant at the .01 level.
4. The multiple correct answers on remembering level with scoring by the subset selection method (SSTRM) had the highest item information, test information and ratio of average information.

Keywords: Multiple Choice Tests, Item Response Theory, Partial Knowledge Scoring Methods, number of correct answers

* Research Article from Faculty of Education, Thaksin University, Thailand

¹ Lecturer, Faculty of Education, Thaksin University

ความสำคัญและปัญหาการวิจัย

คะแนนเป็นผลที่ได้จากการสอบและเป็นตัวชี้วัดสำคัญที่สามารถอธิบายความสามารถของบุคคล ซึ่งเป็นคุณลักษณะภายในที่ต้องอาศัยวิธีการวัดทางอ้อม โดยการใช้แบบทดสอบเป็นสิ่งเร้าแล้วผู้สอบตอบสนองสะท้อนความรู้ ความสามารถออกมาเป็นคะแนน แล้วจึงทำการสรุปอ้างอิงไปยังคุณลักษณะที่ต้องการวัดนั้นๆ อย่างไรก็ตาม คะแนนที่ได้จากการสอบเป็นข้อมูลเชิงปริมาณที่รวมเอาความคลาดเคลื่อนจากแหล่งต่างๆ ไม่ว่าจะเป็นความคลาดเคลื่อนจากตัวผู้สอบเอง จากข้อสอบ ลักษณะของผู้เข้าสอบ สถานการณ์ในการสอบ และความคลาดเคลื่อนจากแหล่งอื่นๆ ทั้งที่เป็นระบบและไม่เป็นระบบ (Lord and Novick, 1968)

ข้อสอบแบบเลือกตอบได้มีการนำมาใช้อย่างแพร่หลายทั้งในอดีตและปัจจุบัน เนื่องจากสามารถวัดความรู้ความสามารถและวัดผลสัมฤทธิ์ได้อย่างกว้างขวาง ประกอบกับมีความเป็นปรนัยในการตรวจให้คะแนน ใช้เวลาตรวจน้อย สามารถใช้เครื่องคอมพิวเตอร์ตรวจได้ ผลการวิเคราะห์ข้อสอบสามารถนำมาวินิจฉัยปัญหาการเรียนรู้ได้ แต่ยังมีจุดอ่อนที่สำคัญ คือมีการเปิดโอกาสให้มีการตอบถูกโดยการเดา ปัญหาการตอบแบบเดาสุ่มจึงเป็นแหล่งความคลาดเคลื่อนในการวัด นอกจากนี้ผลการวัดที่ได้ไม่สามารถให้สารสนเทศที่เพียงพอที่จะจำแนกผู้ตอบในระดับต่างๆ ของความรู้ระหว่างผู้รู้จริง (full knowledge) ผู้ที่มีความรู้บางส่วน (partial knowledge) และผู้ที่ไม่มีความรู้ (absence of knowledge) (เอมอร์ จังศิริพรปกรณ. 2548, น. 1)

ระบบการให้คะแนนความรู้บางส่วนได้ถูกนำมาใช้ เนื่องจากประโยชน์ของการจำแนกคำตอบที่มากกว่า 2 คำ ในแต่ละข้อจะทำให้ได้สารสนเทศเพื่อวินิจฉัยระดับความรู้ความเข้าใจของผู้เรียน เพราะหากคำตอบของข้อสอบ เป็น “ถูก” กับ “ผิด” นักเรียนทุกคนจะถูกจำแนกไว้ในประเภทเดียวกันทั้งหมดคือ รู้กับไม่รู้ และข้อมูลสารสนเทศจากแนวคิดคำนวณก็จะไม่ได้ถูกนำมาใช้ประโยชน์เพื่อเป็นข้อมูลสารสนเทศเกี่ยวกับความเข้าใจของนักเรียน หากเราให้ความสนใจและใส่ใจในกระบวนการที่จะแก้ปัญหาเพื่อตอบข้อสอบ จึงมีความเป็นไปได้ที่จะสามารถให้คะแนนตามระดับความเข้าใจของผู้เรียนที่แตกต่างกัน (Master, G. N. 1988, pp.179-282) อย่างไรก็ตาม จากการวิเคราะห์ข้อสอบและแบบทดสอบด้วยทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory) พบจุดอ่อนบางประการที่ไม่สามารถแก้ปัญหาในการเดารวมไปถึงความคลาดเคลื่อนในการวัด และมีได้ให้ความสนใจต่อสถานการณ์หรือเงื่อนไขในการวัดตลอดจนปฏิสัมพันธ์ที่อาจเกิดขึ้นระหว่างผู้สอบกับเงื่อนไขต่างๆ ซึ่งจะส่งผลต่อความน่าเชื่อถือของผลการวัด ปัจจุบันจึงได้มีนักทฤษฎีวัดผลอ้างอิงถึงความสัมพันธ์ระหว่างความสามารถที่แท้จริงกับลักษณะการตอบสนองข้อสอบ เพื่อทำนายลักษณะที่ซ่อนอยู่ในตัว เรียกว่า ทฤษฎีตอบสนองข้อสอบ (Item Response Theory: IRT) (ศิริชัย กาญจนวาสี, 2545, น. 45-46)

จากแนวคิดและข้อค้นพบต่าง ๆ ที่กล่าวมาเบื้องต้นทำให้ผู้วิจัยสนใจศึกษาและมีคำถามในการวิจัยครั้งนี้ คือ คุณภาพของแบบทดสอบแบบเลือกตอบที่มีจำนวนตัวถูกแตกต่างกัน ใช้วิธีการตรวจให้คะแนนความรู้บางส่วนต่างกัน ภายใต้เงื่อนไขของการวัดพฤติกรรมการเรียนรู้ที่ต่างกันมีความแตกต่างกันหรือไม่ เมื่อเงื่อนไขของจำนวนตัวถูกคือ ตัวเลือกตัวเดียวกับตัวถูกมากกว่า 1 ตัว เมื่อตรวจด้วยวิธีการตรวจให้คะแนนความรู้บางส่วนระหว่างวิธีการตรวจให้คะแนนแบบการตัดตัวลง (Elimination Scoring: ET) โดยใช้วิธีการตอบและให้คะแนนแบบคัมป์ (the Coombs Response/ Scoring method) และวิธีเลือกชุดตัวถูก (Subset Selection Method: SST) จะมีคุณภาพแตกต่างกันหรือไม่ นอกจากนี้ผู้วิจัยสนใจศึกษาเปรียบเทียบคุณภาพของแบบทดสอบที่มีจำนวนตัวถูกแตกต่างกัน ใช้วิธีการตรวจให้คะแนนความรู้บางส่วนต่างกัน ภายใต้เงื่อนไขของแบบทดสอบที่วัดพฤติกรรมการเรียนรู้ต่างกัน เนื่องจากการศึกษาผลการวิจัยของเอมอร์ จังศิริพรปกรณ์ (2548, น. 113) พบว่า แบบทดสอบที่สร้างขึ้นเพื่อวัดพฤติกรรมระดับที่สูงกว่าความรู้ความจำนั้น ไม่จำเป็นต้องสร้างแบบทดสอบที่มีตัวถูกมากกว่า 1 ตัว ในแต่ละข้อเพื่อป้องกันการเดา เพราะการสร้างแบบทดสอบที่มีตัวถูกตัวเดียวที่วัดพฤติกรรมที่สูงกว่าความรู้ความจำนั้น ไม่สามารถทำให้ผู้สอบทราบว่าตัวเลือกใดเป็นตัวเลือกจึงทำให้คุณภาพของแบบทดสอบที่มีจำนวนตัวถูกตัวเดียวและจำนวนตัวถูกมากกว่า 1 ตัว มีคุณภาพไม่แตกต่างกัน จึงต้องการศึกษาและเปรียบเทียบคุณภาพของแบบทดสอบภายใต้เงื่อนไขระดับพฤติกรรมการเรียนรู้ จำนวนตัวถูก และรูปแบบการตรวจให้คะแนนที่แตกต่างกันโดยอาศัยทฤษฎีการตอบสนองข้อสอบรายข้อ (Item response theory) ในการวิเคราะห์คุณภาพของแบบทดสอบด้านค่าความยาก อำนาจจำแนก สารสนเทศของข้อสอบและแบบทดสอบ

วัตถุประสงค์ของการวิจัย

1. เพื่อเปรียบเทียบความเที่ยงตรงตามสภาพของแบบทดสอบระหว่างแบบทดสอบคู่ขนานที่วัดระดับพฤติกรรมการเรียนรู้ วิธีการตรวจให้คะแนนความรู้บางส่วนและจำนวนตัวถูกต่างกัน
2. เพื่อเปรียบเทียบอำนาจจำแนกของข้อสอบระหว่างแบบทดสอบคู่ขนานที่วัดระดับพฤติกรรมการเรียนรู้ วิธีการตรวจให้คะแนนความรู้บางส่วนและจำนวนตัวถูกต่างกัน
3. เพื่อเปรียบเทียบความยากของข้อสอบระหว่างแบบทดสอบคู่ขนานที่วัดระดับพฤติกรรมการเรียนรู้ วิธีการตรวจให้คะแนนความรู้บางส่วนและจำนวนตัวถูกต่างกัน
4. เพื่อเปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบทดสอบ และอัตราส่วนสารสนเทศเฉลี่ยของข้อสอบเลือกตอบระหว่างแบบทดสอบคู่ขนาน ที่วัดระดับพฤติกรรมการเรียนรู้ วิธีการตรวจให้คะแนนความรู้บางส่วนและจำนวนตัวถูกต่างกัน

วิธีดำเนินการวิจัย

ประชากรและกลุ่มตัวอย่าง

ประชากรในการวิจัย ได้แก่ นิสิตคณะศึกษาศาสตร์ มหาวิทยาลัยทักษิณ ที่ลงทะเบียนรายวิชาการวัดและประเมินผลการศึกษา ปีการศึกษา 2555 กลุ่มตัวอย่างในการวิจัย ได้แก่ นิสิตคณะศึกษาศาสตร์ มหาวิทยาลัยทักษิณ ที่ลงทะเบียนรายวิชาการวัดและประเมินผลการศึกษา ปีการศึกษา 2555 ที่ได้มาโดยวิธีการสุ่มแบบแบ่งกลุ่ม (Cluster random sampling) จำนวน 4 ห้องเรียน รวมจำนวนนิสิต 174 คน

เครื่องมือการวิจัย

เครื่องมือที่ใช้ในการวิจัยครั้งนี้ เป็นแบบทดสอบคู่ขนานที่ผู้วิจัยพัฒนาขึ้นจากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน วิชาการวัดและประเมินผลการศึกษา จำนวน 8 ชุด โดยแบบทดสอบทั้ง 8 ฉบับมีรายละเอียด ดังนี้

ตารางที่ 1 รายละเอียดของแบบทดสอบทั้ง 8 ฉบับ

ฉบับที่	ระดับพฤติกรรมการเรียนรู้	วิธีการตอบ	จำนวนตัวถูก	จำนวนข้อ
1	ขั้นความจำ	แบบเลือกชุดตัวถูก	ตัวเดียว	10
2	ขั้นสูงกว่าความจำ	แบบเลือกชุดตัวถูก	ตัวเดียว	10
3	ขั้นความจำ	แบบเลือกชุดตัวถูก	มากกว่า 1 ตัว	10
4	ขั้นสูงกว่าความจำ	แบบเลือกชุดตัวถูก	มากกว่า 1 ตัว	10
5	ขั้นความจำ	แบบตัดตัวลง	ตัวเดียว	10
6	ขั้นสูงกว่าความจำ	แบบตัดตัวลง	ตัวเดียว	10
7	ขั้นความจำ	แบบตัดตัวลง	มากกว่า 1 ตัว	10
8	ขั้นสูงกว่าความจำ	แบบตัดตัวลง	มากกว่า 1 ตัว	10

วิธีการตรวจให้คะแนน

วิธีการตรวจที่ใช้ในการวิจัยครั้งนี้ มี 2 วิธี คือ

1. วิธีการตรวจให้คะแนนแบบเลือกชุดตัวถูก หมายถึง การตรวจให้คะแนนความรู้บางส่วนที่ เอมอร์ จังศิริพรปกรณ์ (2548) ประยุกต์การตรวจให้คะแนนจากแนวคิดวิธีการตอบและให้คะแนนของ เดรสเชลและสมิท โดยการพัฒนาสูตรการให้คะแนนที่ไม่มีค่าติดลบและมีค่าสูงสุด (k-1) เท่ากับ 3 ต่ำสุด เท่ากับ 0 ในแต่ละข้อ หลักการตรวจให้คะแนนรายข้อขึ้นอยู่กับจำนวนตัวเลือกที่ผู้ตอบเลือกและจำนวนตัวเลือกที่เลือกถูกต้องตรงกับคำตอบ โดยมีหลักการคือ ถ้าเลือกตัวถูกหรือชุดตัวถูกได้ครบตรงกับคำตอบพอดีถือว่ามีความรู้เต็มที่ (Full knowledge) ผู้ตอบจะได้คะแนนเต็ม ถ้าเลือกตัวถูกหรือชุดตัวถูกได้บางส่วน แต่ไม่ครบถือว่ามีความรู้บางส่วน (Partial knowledge) ผู้ตอบจะได้คะแนนบางส่วนตามจำนวนตัวเลือกที่เลือก ถ้าเลือกตัวถูกหรือชุดตัวถูกได้ แต่เลือกตัวลงเป็นตัวถูกด้วย ถือว่ามีความรู้ที่ผิดบางส่วน

(Partial misinformation) ผู้ตอบจะได้คะแนนบางส่วนตามจำนวนตัวถูกที่เลือกและจะถูกตัดคะแนนที่เข้าใจว่าตัวถูกเป็นตัวถูก ถ้าไม่สามารถเลือกตัวถูกหรือชุดตัวถูกได้เลย ถือว่ามีความรู้ที่ผิดเต็มที่ (Full misinformation) ผู้ตอบจะไม่ได้คะแนน ถ้าไม่ตอบถือว่าไม่มีความรู้ (Absence of knowledge) ผู้ตอบจะไม่ได้คะแนน

2. วิธีตรวจให้คะแนนแบบตัดตัวถูก หมายถึง วิธีการให้คะแนนความรู้บางส่วนที่เอมอร์ จังศิริพรปกรณ์ ได้ประยุกต์การตรวจให้คะแนนจากแนวคิดวิธีการตอบและให้คะแนนแบบคูมบี้ โดยพัฒนาสูตรการให้คะแนนที่ไม่มีค่าติดลบและมีค่าสูงสุด ($k-1$) เท่ากับ 3 ต่ำสุดเท่ากับ 0 ในแต่ละข้อ หลักการตรวจให้คะแนนรายข้อ ขึ้นอยู่กับจำนวนตัวเลือกที่ผู้ตอบเลือกและจำนวนตัวถูกที่ตัดได้ตรงกับคำตอบ โดยมีหลักการคือ ถ้าตัดตัวถูกได้ครบตรงกับคำตอบพอดี ถือว่ามีความรู้เต็มที่ (Full knowledge) ผู้ตอบจะได้คะแนนเต็ม ถ้าตัดตัวถูกได้บางส่วน แต่ไม่ครบถือว่ามีความรู้บางส่วน (Partial knowledge) ผู้ตอบจะได้คะแนนบางส่วนตามจำนวนตัวถูกที่ตัดถูก ถ้าตัดตัวถูกได้แต่ตัดตัวถูกออกไปด้วย ถือว่ามีความรู้ที่ผิดบางส่วน (Partial misinformation) ผู้ตอบจะได้คะแนนบางส่วนตามจำนวนตัวถูกที่ตัดถูก และจะถูกตัดคะแนนที่เข้าใจว่าตัวถูกเป็นตัวถูกถ้าตัดตัวถูกผิด โดยไปตัดตัวถูกถือว่ามีความรู้ที่ผิดเต็มที่ (Full misinformation) ผู้ตอบจะไม่ได้คะแนนถ้าไม่ตัดตัวถูกเลยถือว่าไม่มีความรู้ (Absence of knowledge) ผู้ตอบจะไม่ได้คะแนน

การเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูลในการวิจัยครั้งนี้ เก็บข้อมูลโดยให้นักศึกษาศึกษาสาส์ตร มหาวิทยาลัยทักษิณ ที่ลงทะเบียนเรียนวิชาการวัดและประเมินผลการศึกษา จำนวน 4 ห้องเรียน ทำแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาการวัดและประเมินผลการศึกษา ที่พัฒนาขึ้นเป็นแบบทดสอบคู่ขนาน จำนวน 8 ฉบับ โดยมีการเก็บข้อมูล 4 ครั้ง แต่ละครั้งห่างกัน 2 สัปดาห์ ซึ่งการเก็บข้อมูลเพื่อเปรียบเทียบวิธีการให้คะแนนครั้งนี้ เพื่อให้ผู้สอบทำอย่างเต็มความสามารถ และลดปัญหาการทำแบบไม่ตั้งใจจึงมีการให้คะแนนความก้าวหน้าถ้าผู้สอบทำคะแนนได้เท่าเดิมหรือดีขึ้น

การวิเคราะห์ข้อมูล

1. วิเคราะห์ความเที่ยงตรงตามสภาพ จากค่าสัมประสิทธิ์สหสัมพันธ์ (r) และแปลงค่าความเที่ยงตรงตามสภาพของแบบทดสอบทั้ง 8 ชุด จากค่าสัมประสิทธิ์สหสัมพันธ์ เป็นคะแนนพิชเชอร์ซี (Z_r) แล้วเปรียบเทียบความเป็นอิสระของความเที่ยงตรงตามสภาพในภาพรวม โดยใช้สถิติทดสอบไคสแควร์ (χ^2) และเปรียบเทียบความแตกต่างของค่าความเที่ยงตรงตามสภาพเป็นรายคู่โดยใช้สถิติทดสอบ Z

2. การวิเคราะห์คุณภาพของแบบทดสอบจากฟังก์ชันสารสนเทศของแบบทดสอบ อัตราส่วนสารสนเทศเฉลี่ย และฟังก์ชันสารสนเทศของข้อสอบโดยใช้การวิเคราะห์ตามทฤษฎีการตอบสนอง

ข้อสอบ (Item Response Theory) แบบพหุวิภาคที่ใช้โมเดลการให้คะแนนความรู้บางส่วน (Partial Credit Model) ซึ่งเป็นการให้คะแนนด้วยแนวทางการตรวจให้คะแนนที่ได้ประยุกต์การให้คะแนนของคูมพ์ และประยุกต์การให้คะแนนของเดรสเซลและสมิท และทำการประมาณค่าพารามิเตอร์ความยาก (b) อำนาจจำแนก (a) ค่าฟังก์ชันสารสนเทศของข้อสอบ และค่าฟังก์ชันสารสนเทศของแบบทดสอบ (TIF) โดยใช้โปรแกรม PARSCALE for WINDOWS

3. การเปรียบเทียบความแตกต่างของค่าเฉลี่ยของค่าอำนาจจำแนก ความยาก ค่าฟังก์ชันสารสนเทศของข้อสอบและค่าฟังก์ชันสารสนเทศของแบบทดสอบ (TIF) ใช้สถิติ F-test Repeated และเปรียบเทียบความแตกต่างรายคู่ด้วยค่าสถิติ F โดยใช้โปรแกรม SPSS for WINDOWS

สรุปผลการวิจัย

1. ผลการเปรียบเทียบความเที่ยงตรงตามสภาพของแบบทดสอบระหว่างแบบทดสอบคู่ขนานที่วัดระดับพฤติกรรมการเรียนรู้ วิธีการตรวจให้คะแนนความรู้บางส่วนและจำนวนตัวถูกต่างกัน

ตารางที่ 2 ความเที่ยงตรงตามสภาพจากคะแนนสอบที่วัดระดับพฤติกรรมการเรียนรู้ วิธีการตรวจให้คะแนนความรู้บางส่วนและจำนวนตัวถูกต่างกันในภาพรวม

ประเภทของแบบทดสอบ	N	ความเที่ยงตรงตามสภาพ		χ^2
		r	Z_r	
SSTRO	174	0.554**	0.623	54.655
ETRO	174	0.327**	0.339	
SSTRM	174	0.588**	0.675	
ETRM	174	0.092	0.092	
SSTHO	174	0.556**	0.627	
ETHO	174	0.538**	0.601	
SSTHM	174	0.097	0.097	
ETHM	174	0.208**	0.211	

** $p < .01$, $\chi^2_{173,.05} = 219.189$

จากตารางที่ 2 พบว่า ความเที่ยงตรงตามสภาพที่พิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์ (r) ระหว่างแบบทดสอบคู่ขนานทั้ง 8 ชุด กับคะแนนสอบปลายภาคเรียนที่ 2 ปีการศึกษา 2555 แต่ละชุดมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ยกเว้น แบบทดสอบวัดระดับพฤติกรรมการเรียนรู้ขึ้นความจำแบบตัดตัวลงที่มีคำตอบถูกมากกว่า 1 ตัว และแบบทดสอบวัดระดับพฤติกรรมการเรียนรู้ขึ้นสูงกว่าความจำแบบเลือกชุดตัวถูกที่มีคำตอบถูกมากกว่า 1 ตัว ที่ไม่มีความสัมพันธ์อย่างมีนัยสำคัญ

ทางสถิติ เมื่อเปรียบเทียบความแตกต่างของความเที่ยงตรงตามสภาพ โดยเปรียบเทียบคะแนนฟิชเชอร์ซี โดยใช้สถิติไคสแควร์ (χ^2) พบว่าไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

2. ผลการเปรียบเทียบอำนาจจำแนกของข้อสอบระหว่างแบบทดสอบคู่ขนานที่วัดระดับพฤติกรรมการเรียนรู้ วิธีการตรวจให้คะแนนความรู้บางส่วนและจำนวนตัวถูกต่างกัน

ตารางที่ 3 ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐานของค่าอำนาจจำแนก และค่าสถิติ F เปรียบเทียบค่าอำนาจจำแนกเฉลี่ยจากการวัดซ้ำ (repeated) ของแบบทดสอบทั้ง 8 ชุด จำแนกตามเงื่อนไขจำนวนตัวถูก วิธีการตรวจให้คะแนนความรู้บางส่วน และระดับพฤติกรรมการเรียนรู้

ประเภทของแบบทดสอบ	อำนาจจำแนก		F test	p-value
	Mean	S.D.		
ภาพรวมทั้ง 3 กลุ่ม			1.262	.284
กลุ่มที่ 1 เงื่อนไขจำนวนตัวถูก				
ตัวถูก 1 ตัว			2.489	.083
	SSTRO	0.261 0.130		
	ETRO	0.460 0.324		
	SSTHO	0.444 0.125		
	ETHO	0.363 0.085		
ตัวถูกมากกว่า 1 ตัว			0.340	.797
	SSTRM	0.502 0.257		
	ETRM	0.423 0.327		
	SSTHM	0.485 0.279		
	ETHM	0.417 0.227		
กลุ่มที่ 2 เงื่อนไขวิธีการตรวจ				
เลือกชุดตัวถูก			2.634	.070
	SSTRO	0.261 0.130		
	SSTRM	0.502 0.257		
	SSTHO	0.444 0.125		
	SSTHM	0.485 0.279		
ตัดตัวลวง			0.347	.792
	ETRO	0.460 0.324		
	ETRM	0.423 0.327		
	ETHO	0.363 0.085		
	ETHM	0.417 0.227		



ตารางที่ 3 (ต่อ)

ประเภทของแบบทดสอบ	อำนาจจำแนก		F test	p-value
	Mean	S.D.		
กลุ่มที่ 3 เจื่อนไ้ระดับพฤติกรรมการเรียนรู้				
วัดระดับพฤติกรรมการเรียนรู้ชั้นความจำ			2.629	.071
SSTRO	0.261	0.130		
ETRO	0.460	0.324		
SSTRM	0.502	0.257		
ETRM	0.423	0.327		
วัดระดับพฤติกรรมการเรียนรู้ชั้นสูงกว่าความจำ			0.737	.539
SSTHO	0.444	0.125		
ETHO	0.363	0.085		
SSTHM	0.485	0.279		
ETHM	0.417	0.227		

จากตารางที่ 3 พบว่า เมื่อทดสอบความแตกต่างของค่าอำนาจจำแนกเฉลี่ยด้วยสถิติ F พบว่าค่าอำนาจจำแนกเฉลี่ยของแบบทดสอบทั้ง 8 ชุด ในภาพรวมและภาพย่อยรายเงื่อนไข ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

3. ผลการเปรียบเทียบความยากของข้อสอบระหว่างแบบทดสอบคู่ขนานที่วัดระดับพฤติกรรมการเรียนรู้ วิธีการตรวจให้คะแนนความรู้บางส่วนและจำนวนตัวถูกต่างกัน

ตารางที่ 4 ค่าเฉลี่ยส่วนเบี่ยงเบนมาตรฐานของค่าความยาก และค่าสถิติ F เปรียบเทียบค่าความยากจากการวัดซ้ำ (repeated) ของแบบทดสอบ

ประเภทของแบบทดสอบ	ค่าความยาก		F test	p-value
	Mean	S.D.		
ภาพรวมทั้ง 3 กลุ่ม			9.442	.000
กลุ่มที่ 1 เจื่อนไ้จำนวนตัวถูก				
ตัวถูก 1 ตัว			8.161	.001
SSTRO	-0.584	1.173		
ETRO	-0.515	1.435		
SSTHO	-0.047	1.096		
ETHO	1.397	0.801		



ตารางที่ 4 (ต่อ)

ประเภทของแบบทดสอบ	ค่าความยาก		F test	p-value
	Mean	S.D.		
ตัวถูกมากกว่า 1 ตัว			14.283	.000
SSTRM	-0.592	0.876		
ETRM	-0.092	1.255		
SSTHM	1.677	0.979		
ETHM	1.554	1.018		
ตัวถูกมากกว่า 1 ตัว			14.283	.000
กลุ่มที่ 2 เจาะไข วิธีการตรวจ				
เลือกชุดตัวถูก			15.879	.000
SSTRO	-0.584	1.173		
SSTRM	-0.592	0.876		
SSTHO	-0.047	1.096		
SSTHM	1.677	0.979		
ตัดตัวลง			7.273	.001
ETRO	-0.515	1.435		
ETRM	-0.092	1.255		
ETHO	1.397	0.801		
ETHM	1.554	1.018		
กลุ่มที่ 3 เจาะไขระดับพฤติกรรมการเรียนรู้				
วัดระดับพฤติกรรมการเรียนรู้ขึ้นจากความจำ			0.408	.784
SSTRO	-0.584	1.173		
ETRO	-0.515	1.435		
SSTRM	-0.592	0.876		
ETRM	-0.092	1.255		
วัดระดับพฤติกรรมการเรียนรู้ขึ้นสูงกว่าความจำ			7.309	.001
SSTHO	-0.047	1.096		
ETHO	1.397	0.801		
SSTHM	1.677	0.979		
ETHM	1.554	1.018		

จากตารางที่ 4 เมื่อทดสอบความแตกต่างของค่าความยากเฉลี่ยด้วยสถิติ F พบว่า ค่าความยากเฉลี่ยของแบบทดสอบทั้ง 8 ชุด ในภาพรวมแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนในภาพย่อย

พบว่าเงื่อนไขการวัดระดับพฤติกรรมการเรียนรู้ต่างกัน ค่าความยากของแบบทดสอบในกลุ่มของการวัดระดับพฤติกรรมการเรียนรู้ขึ้นจากความจำจากแบบทดสอบ 4 ฉบับ ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ

การเปรียบเทียบค่าความยากเฉลี่ยของแบบทดสอบในกลุ่มเงื่อนไขจำนวนตัวถูก พบว่าแบบทดสอบแบบเลือกชุดตัวถูกวัดขึ้นสูงกว่าความจำที่มีคำตอบถูกมากกว่า 1 ตัว มีค่าความยากเฉลี่ยสูงกว่าแบบทดสอบแบบเลือกชุดตัวถูกวัดขึ้นสูงกว่าความจำที่มีคำตอบถูกตัวเดียว อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

ส่วนในกลุ่มเงื่อนไขวิธีการตรวจ พบว่า แบบทดสอบแบบตัดตัวลงวัดขึ้นสูงกว่าความจำที่มีคำตอบถูกตัวเดียว มีค่าความยากเฉลี่ยสูงกว่าแบบทดสอบแบบเลือกชุดตัวถูกวัดขึ้นสูงกว่าความจำที่มีคำตอบถูกตัวเดียวอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และในกลุ่มเงื่อนไขระดับพฤติกรรมการเรียนรู้ พบว่ามี 3 คู่ที่แตกต่างกันอย่างมีนัยสำคัญ ทางสถิติที่ระดับ .05 ได้แก่ แบบทดสอบแบบตัดตัวลงวัดขึ้นสูงกว่าความจำที่มีคำตอบถูกตัวเดียว มีค่าความยากเฉลี่ยสูงกว่าแบบทดสอบแบบตัดตัวลงวัดขึ้นความจำที่มีคำตอบถูกตัวเดียวและแบบทดสอบแบบเลือกชุดตัวถูกวัดขึ้นสูงกว่าความจำที่มีคำตอบถูกมากกว่า 1 ตัว มีค่าความยากเฉลี่ยสูงกว่าแบบทดสอบแบบเลือกชุดตัวถูกวัดขึ้นความจำที่มีคำตอบถูกมากกว่า 1 ตัวและแบบทดสอบแบบตัดตัวลงวัดขึ้นสูงกว่าความจำที่มีคำตอบถูกมากกว่า 1 ตัว มีค่าความยากเฉลี่ยสูงกว่าแบบทดสอบแบบตัดตัวลงวัดขึ้นความจำที่มีคำตอบถูกมากกว่า 1 ตัว อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ทุกคู่

ส่วนแบบทดสอบแบบเลือกชุดตัวถูกวัดขึ้นความจำที่มีคำตอบถูกตัวเดียว และแบบทดสอบแบบเลือกชุดตัวถูกวัดขึ้นสูงกว่าความจำที่มีคำตอบถูกตัวเดียวมีค่าความยากเฉลี่ยไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ

4. ผลการเปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบทดสอบ และอัตราส่วนสารสนเทศเฉลี่ยของข้อสอบเลือกตอบระหว่างแบบทดสอบคู่ขนาน ที่วัดระดับพฤติกรรมการเรียนรู้วิธีการตรวจให้คะแนนความรู้บางส่วนและจำนวนตัวถูกต่างกัน



ตารางที่ 5 ค่าเฉลี่ยส่วนเบี่ยงเบนมาตรฐานของฟังก์ชันสารสนเทศของข้อสอบและค่าสถิติ F เปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยจากการวัดซ้ำของแบบทดสอบทั้ง 3 เงื่อนไข

ประเภทของแบบทดสอบ	ฟังก์ชันสารสนเทศของข้อสอบ		F test	p-value
	Mean	S.D.		
ภาพรวมทั้ง 3 กลุ่ม			1.133	.345
กลุ่มที่ 1 เงื่อนไขจำนวนตัวถูก				
ตัวถูก 1 ตัว			1.447	.251
SSTRO	0.150	0.232		
ETRO	0.251	0.299		
SSTHO	0.243	0.188		
ETHO	0.217	0.290		
ตัวถูกมากกว่า 1 ตัว			1.101	.366
SSTRM	0.290	0.363		
ETRM	0.239	0.465		
SSTHM	0.258	0.283		
ETHM	0.187	0.149		
กลุ่มที่ 2 เงื่อนไขวิธีการตรวจ				
เลือกชุดตัวถูก			1.962	.144
SSTRO	0.150	0.232		
SSTRM	0.290	0.363		
SSTHO	0.243	0.188		
SSTHM	0.258	0.283		
ตัดตัวลวง			0.436	.729
ETRO	0.251	0.299		
ETRM	0.239	0.465		
ETHO	0.217	0.290		
ETHM	0.187	0.149		
กลุ่มที่ 3 เงื่อนไขระดับพฤติกรรมการเรียนรู้				
วัดระดับพฤติกรรมการเรียนรู้ขั้นความจำ			2.346	.095
SSTRO	0.150	0.232		
ETRO	0.251	0.299		
SSTRM	0.290	0.363		
ETRM	0.239	0.465		

ตารางที่ 5 (ต่อ)

ประเภทของแบบทดสอบ	ฟังก์ชันสารสนเทศของข้อสอบ		F test	p-value
	Mean	S.D.		
วัดระดับพฤติกรรมการเรียนรู้ขั้นสูงกว่าความจำ			0.669	.579
SSTHO	0.243	0.188		
ETHO	0.217	0.290		
SSTHM	0.258	0.283		
ETHM	0.187	0.149		

จากตาราง 5 พบว่า แบบทดสอบแบบเลือกชุดตัวถูกวัดขั้นความจำที่มีคำตอบถูกมากกว่า 1 ตัว มีค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงสุด (0.290) รองลงมาคือ แบบทดสอบแบบเลือกชุดตัวถูกวัดขั้นสูงกว่าความจำที่มีคำตอบถูกมากกว่า 1 ตัว ส่วนแบบทดสอบแบบเลือกชุดตัวถูกวัดขั้นความจำที่มีคำตอบถูกตัวเดียวมีค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยต่ำที่สุด (0.150)

เมื่อพิจารณาในแง่ของจำนวนตัวถูก พบว่า จำนวนจำนวนตัวถูกตัวเดียว แบบทดสอบแบบเลือกชุดตัวถูกวัดขั้นสูงกว่าความจำที่มีคำตอบถูกตัวเดียวมีค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงสุด (0.243) ส่วนจำนวนจำนวนตัวถูกมากกว่า 1 ตัว แบบทดสอบแบบเลือกชุดตัวถูกวัดขั้นความจำที่มีคำตอบถูกมากกว่า 1 ตัวมีค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงสุด (0.290)

เมื่อพิจารณาในแง่ของวิธีการตรวจให้คะแนนความรู้บางส่วน พบว่า วิธีการตรวจแบบเลือกชุดตัวถูกแบบทดสอบแบบเลือกชุดตัวถูกวัดขั้นความจำที่มีคำตอบถูกมากกว่า 1 ตัว มีค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงสุด (0.290) วิธีการตรวจแบบตัดตัวลง แบบทดสอบแบบตัดตัวลงวัดขั้นความจำที่มีคำตอบถูกตัวเดียวมีค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงสุด (0.251) และเมื่อพิจารณาในแง่ของระดับพฤติกรรมการเรียนรู้ที่วัด พบว่า แบบทดสอบที่วัดระดับพฤติกรรมการเรียนรู้ขั้นความจำ แบบทดสอบแบบเลือกชุดตัวถูกวัดขั้นความจำที่มีคำตอบถูกมากกว่า 1 ตัว มีค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงสุด (0.290) แบบทดสอบที่วัดระดับพฤติกรรมการเรียนรู้ขั้นสูงกว่าความจำแบบทดสอบแบบเลือกชุดตัวถูกวัดขั้นสูงกว่าความจำที่มีคำตอบถูกมากกว่า 1 ตัว มีค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงสุด (0.258) และเมื่อทดสอบความแตกต่างของค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยด้วยสถิติ F พบว่า ค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยทั้ง 8 ชุด ในภาพรวมและในภาพย่อย ไม่มีความแตกต่างอย่างมีนัยสำคัญทางสถิติ



ตารางที่ 6 ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของฟังก์ชันสารสนเทศของข้อสอบจากแบบทดสอบ
จำนวน 8 ชุด จำแนกตามความสามารถของผู้สอบ

ความสามารถ ของผู้สอบ (θ)	ฟังก์ชัน สารสนเทศ ของข้อสอบ		SSTRO	ETRO	SSTRM	ETRM	SSTHO	ETHO	SSTHM	ETHM
	Mean	S.D.								
-4	Mean		0.033	0.087	0.087	0.087	0.107	0.011	0.029	0.076
	S.D.		0.029	0.067	0.081	0.119	0.034	0.021	0.023	0.082
-3	Mean		0.076	0.146	0.186	0.202	0.158	0.062	0.064	0.133
	S.D.		0.054	0.095	0.173	0.330	0.055	0.156	0.047	0.119
-2	Mean		0.165	0.222	0.372	0.222	0.179	0.148	0.153	0.233
	S.D.		0.160	0.146	0.305	0.271	0.052	0.366	0.140	0.207
-1	Mean		0.368	0.353	0.480	0.233	0.180	0.161	0.253	0.245
	S.D.		0.322	0.344	0.309	0.087	0.035	0.192	0.246	0.200
0	Mean		0.346	0.378	0.625	0.696	0.218	0.282	0.344	0.219
	S.D.		0.372	0.378	0.633	1.431	0.080	0.266	0.316	0.140
1	Mean		0.121	0.387	0.425	0.322	0.393	0.431	0.450	0.182
	S.D.		0.075	0.459	0.413	0.406	0.219	0.297	0.412	0.093
2	Mean		0.052	0.292	0.188	0.211	0.498	0.423	0.440	0.173
	S.D.		0.043	0.376	0.140	0.229	0.311	0.381	0.379	0.087
3	Mean		0.031	0.163	0.068	0.089	0.286	0.236	0.310	0.176
	S.D.		0.031	0.172	0.049	0.047	0.194	0.283	0.187	0.148
4	Mean		0.017	0.068	0.029	0.047	0.087	0.060	0.172	0.076
	S.D.		0.020	0.069	0.023	0.030	0.062	0.062	0.100	0.082

จากตารางที่ 6 ผลการเปรียบเทียบค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยทั้ง 8 ชุด พบว่าแบบทดสอบที่มีค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงที่สุดที่ระดับความสามารถ 0 คือ แบบทดสอบแบบเลือกชุดตัวถูกวัดขึ้นความจำที่มีคำตอบถูกมากกว่า 1 ตัว (ค่าเฉลี่ยฟังก์ชันสารสนเทศ 0.696)

ที่ระดับความสามารถ 1 แบบทดสอบที่มีค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงที่สุด คือแบบทดสอบแบบเลือกชุดตัวถูกวัดขึ้นสูงกว่าความจำที่มีคำตอบถูกมากกว่า 1 ตัว (ค่าเฉลี่ยฟังก์ชันสารสนเทศ 0.450)

ที่ระดับความสามารถ 2 แบบทดสอบที่มีค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงที่สุด คือแบบทดสอบแบบเลือกชุดตัวถูกวัดขึ้นสูงกว่าความจำที่มีคำตอบถูกตัวเดียว (ค่าเฉลี่ยฟังก์ชันสารสนเทศ 0.498)

ที่ระดับความสามารถ 3 และ 4 แบบทดสอบที่มีค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงสุด คือแบบทดสอบแบบเลือกชุดตัวถูกวัดขั้นสูงกว่าความจำที่มีคำตอบถูกมากกว่า 1 ตัว (ค่าเฉลี่ยฟังก์ชันสารสนเทศ 0.310 และ 0.172 ตามลำดับ)

ที่ระดับความสามารถ -1 และ -2 แบบทดสอบที่มีค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงสุด คือ แบบทดสอบแบบเลือกชุดตัวถูกวัดขั้นความจำที่มีคำตอบถูกมากกว่า 1 ตัว (ค่าเฉลี่ยฟังก์ชันสารสนเทศ 0.480 และ 0.372 ตามลำดับ)

ที่ระดับความสามารถ -3 แบบทดสอบที่มีค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงสุด คือ แบบทดสอบแบบตัดตัวถูกวัดขั้นความจำที่มีคำตอบถูกมากกว่า 1 ตัว (ค่าเฉลี่ยฟังก์ชันสารสนเทศ 0.202)

ที่ระดับความสามารถ -4 แบบทดสอบที่มีค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงสุด คือ แบบทดสอบแบบเลือกชุดตัวถูกวัดขั้นสูงกว่าความจำที่มีคำตอบถูกตัวเดียว (ค่าเฉลี่ยฟังก์ชันสารสนเทศ 0.107)

แสดงว่า แบบทดสอบที่เหมาะสมกับผู้สอบที่มีระดับความสามารถสูงคือ แบบทดสอบแบบเลือกชุดตัวถูกวัดขั้นสูงกว่าความจำที่มีคำตอบถูกมากกว่า 1 ตัว และแบบทดสอบที่เหมาะสมกับผู้สอบที่มีระดับความสามารถต่ำ คือ แบบทดสอบแบบเลือกชุดตัวถูกวัดขั้นความจำที่มีคำตอบถูกมากกว่า 1 ตัว

ตารางที่ 7 ฟังก์ชันสารสนเทศของแบบทดสอบ (Test Information) และความคลาดเคลื่อนมาตรฐาน (SE (θ)) จำแนกตามระดับความสามารถของผู้สอบของแบบทดสอบจำนวน 8 ชุด

ความสามารถ ของผู้สอบ (θ)	ฟังก์ชัน สารสนเทศของ แบบทดสอบ	SSTRO	ETRO	SSTRM	ETRM	SSTHO	ETHO	SSTHM	ETHM
-4	Test Inf	0.326	0.868	0.872	0.874	1.072	0.105	0.285	0.764
	SE(θ)	1.752	1.073	1.071	1.070	0.966	3.086	1.872	1.144
-3	Test Inf	0.761	1.465	1.864	2.022	1.585	0.616	0.640	1.329
	SE(θ)	1.146	0.826	0.733	0.703	0.794	1.274	1.250	0.868
-2	Test Inf	1.655	2.218	3.212	2.217	1.790	1.481	1.243	1.875
	SE(θ)	0.777	0.671	0.558	0.672	0.747	0.822	0.897	0.730
-1	Test Inf	3.677	3.535	4.800	2.331	1.796	1.608	2.529	2.334
	SE(θ)	0.521	0.532	0.456	0.655	0.746	0.789	0.629	0.654
0	Test Inf	3.460	3.782	6.245	6.957	2.177	2.825	3.442	2.445
	SE(θ)	0.538	0.514	0.400	0.379	0.678	0.595	0.539	0.639
1	Test Inf	1.213	3.870	4.253	3.219	3.932	4.305	4.500	2.185
	SE(θ)	0.908	0.508	0.485	0.557	0.504	0.482	0.471	0.676

ตารางที่ 7 (ต่อ)

ความสามารถ ของผู้สอบ (θ)	ฟังก์ชัน สารสนเทศของ แบบทดสอบ	ฟังก์ชัน							
		SSTRO	ETRO	SSTRM	ETRM	SSTHO	ETHO	SSTHM	ETHM
2	Test Inf	0.525	2.924	1.438	1.692	4.771	4.225	4.401	1.821
	SE(θ)	1.381	0.585	0.834	0.769	0.458	0.486	0.477	0.741
3	Test Inf	0.315	1.635	0.683	0.891	2.855	2.363	3.096	1.729
	SE(θ)	1.783	0.782	1.210	1.059	0.592	0.651	0.568	0.761
4	Test Inf	0.171	0.682	0.288	0.465	0.865	0.598	1.720	1.756
	SE(θ)	2.421	1.211	1.863	1.466	1.075	1.293	0.763	0.755

จากตารางที่ 7 ผลการเปรียบเทียบค่าฟังก์ชันสารสนเทศของแบบทดสอบทั้ง 8 ชุด พบว่าแบบทดสอบที่มีค่าฟังก์ชันสารสนเทศของแบบทดสอบสูงที่สุดที่ระดับความสามารถ 0 คือ แบบทดสอบแบบตัดตัวลงวัดขึ้นความจำที่มีคำตอบถูกมากกว่า 1 ตัว (ค่าฟังก์ชันสารสนเทศ 6.957) มีรายละเอียดดังนี้

ที่ระดับความสามารถ 1 แบบทดสอบที่มีค่าฟังก์ชันสารสนเทศของแบบทดสอบสูงที่สุด คือแบบทดสอบแบบเลือกชุดตัวถูกวัดขึ้นสูงกว่าความจำที่มีคำตอบถูกมากกว่า 1 ตัว (ค่าฟังก์ชันสารสนเทศ 4.500)

ที่ระดับความสามารถ 2 แบบทดสอบที่มีค่าฟังก์ชันสารสนเทศของแบบทดสอบสูงที่สุด คือแบบทดสอบแบบเลือกชุดตัวถูกวัดขึ้นสูงกว่าความจำที่มีคำตอบถูกตัวเดียว (ค่าฟังก์ชันสารสนเทศ 4.771)

ที่ระดับความสามารถ 3 แบบทดสอบที่มีค่าฟังก์ชันสารสนเทศของแบบทดสอบสูงที่สุด คือแบบทดสอบแบบเลือกชุดตัวถูกวัดขึ้นสูงกว่าความจำที่มีคำตอบถูกมากกว่า 1 ตัว (ค่าฟังก์ชันสารสนเทศ 3.096)

ที่ระดับความสามารถ 4 แบบทดสอบที่มีค่าฟังก์ชันสารสนเทศของแบบทดสอบสูงที่สุด คือแบบทดสอบแบบเลือกตัดตัวลงวัดขึ้นสูงกว่าความจำที่มีคำตอบถูกมากกว่า 1 ตัว (ค่าฟังก์ชันสารสนเทศ 1.756)

ที่ระดับความสามารถ -1 และ -2 แบบทดสอบที่มีค่าฟังก์ชันสารสนเทศของแบบทดสอบสูงที่สุด คือแบบทดสอบแบบเลือกชุดตัวถูกวัดขึ้นความจำที่มีคำตอบถูกมากกว่า 1 ตัว (ค่าฟังก์ชันสารสนเทศ 4.800 และ 3.212 ตามลำดับ)

ที่ระดับความสามารถ -3 แบบทดสอบที่มีค่าฟังก์ชันสารสนเทศของแบบทดสอบสูงที่สุด คือแบบทดสอบแบบตัดตัวลงวัดขึ้นความจำที่มีคำตอบถูกมากกว่า 1 ตัว (ค่าฟังก์ชันสารสนเทศ 2.022)

ที่ระดับความสามารถ -4 แบบทดสอบที่มีค่าฟังก์ชันสารสนเทศของแบบทดสอบสูงที่สุด คือ แบบทดสอบแบบเลือกชุดตัวถูกวัดขึ้นสูงกว่าความจำที่มีคำตอบถูกต้องเดียว (ค่าฟังก์ชันสารสนเทศ 1.072)

จากผลการวิเคราะห์แสดงว่า แบบทดสอบที่เหมาะสมกับผู้สอบที่มีระดับความสามารถสูง ได้แก่ แบบทดสอบแบบเลือกชุดตัวถูกวัดขึ้นสูงกว่าความจำที่มีคำตอบถูกมากกว่า 1 ตัว แบบทดสอบแบบเลือกชุดตัวถูกวัดขึ้นสูงกว่าความจำที่มีคำตอบถูกต้องเดียว และแบบทดสอบแบบเลือกตัดตัวลงวัดขึ้นสูงกว่าความจำที่มีคำตอบถูกมากกว่า 1 ตัว และแบบทดสอบที่เหมาะสมกับผู้สอบที่มีระดับความสามารถต่ำ ได้แก่ แบบทดสอบแบบเลือกชุดตัวถูกวัดขึ้นความจำที่มีคำตอบถูกมากกว่า 1 ตัว แบบทดสอบแบบตัดตัวลงวัดขึ้นความจำที่มีคำตอบถูกมากกว่า 1 ตัว และแบบทดสอบแบบเลือกชุดตัวถูกวัดขึ้นสูงกว่าความจำที่มีคำตอบถูกต้องเดียว

เมื่อเปรียบเทียบค่าฟังก์ชันสารสนเทศของแบบทดสอบแต่ละฉบับตามความสามารถ ส่วนใหญ่ ให้ค่าฟังก์ชันสารสนเทศของแบบทดสอบสูงสุดและความคลาดเคลื่อนมาตรฐานต่ำสุดในช่วงความสามารถ 0 และ 1 ยกเว้นแบบทดสอบแบบเลือกชุดตัวถูกวัดขึ้นสูงกว่าความจำที่มีคำตอบถูกต้องเดียว แบบทดสอบแบบตัดตัวลงวัดขึ้นสูงกว่าความจำที่มีคำตอบถูกต้องเดียว แบบทดสอบแบบเลือกชุดตัวถูกวัดขึ้นสูงกว่าความจำที่มีคำตอบถูกมากกว่า 1 ตัว และแบบทดสอบแบบเลือกตัดตัวลงวัดขึ้นสูงกว่าความจำที่มีคำตอบถูกมากกว่า 1 ตัว ให้ค่าฟังก์ชันสารสนเทศของแบบทดสอบสูงสุดในช่วงความสามารถสูง ตารางที่ 8 อัตราส่วนสารสนเทศของแบบทดสอบจำแนกตามระดับพฤติกรรมการเรียนรู้ที่วัด วิธีการตรวจให้คะแนนความรู้บางส่วนและจำนวนตัวถูกต่างกัน

การเปรียบเทียบระหว่าง		อัตราส่วนสารสนเทศ
SSTRO	ETRO	0.597
SSTRO	SSTRM	0.517
SSTRO	ETRM	0.627
SSTRO	SSTHO	0.618
SSTRO	ETHO	0.690
SSTRO	SSTHM	0.580
SSTRO	ETHM	0.802
ETRO	SSTRM	0.866
ETRO	ETRM	1.050
ETRO	SSTHO	1.035
ETRO	ETHO	1.156
ETRO	SSTHM	0.972

ตารางที่ 8 (ต่อ)

การเปรียบเทียบระหว่าง		อัตราส่วนสารสนเทศ
ETRO	ETHM	1.342
SSTRM	ETRM	1.213
SSTRM	SSTHO	1.195
SSTRM	ETHO	1.335
SSTRM	SSTHM	1.122
SSTRM	ETHM	1.550
ETRM	SSTHO	0.985
ETRM	ETHO	1.101
ETRM	SSTHM	0.925
ETRM	ETHM	1.278
SSTHO	ETHO	1.117
SSTHO	SSTHM	0.939
SSTHO	ETHM	1.297
ETHO	SSTHM	0.841
ETHO	ETHM	1.161
SSTHM	ETHM	1.381

จากตารางที่ 8 พบว่าอัตราส่วนสารสนเทศทั้ง 28 ค่า มีค่าตั้งแต่ 0.517 – 1.550 โดยอัตราส่วนสารสนเทศระหว่าง แบบทดสอบแบบเลือกชุดตัวถูกวัดขึ้นความจำที่มีคำตอบถูกมากกว่า 1 ตัว (1.550) และ แบบทดสอบแบบตัดตัวลงวัดขึ้นสูงกว่าความจำที่มีคำตอบถูกมากกว่า 1 ตัว มีค่าสูงสุด รองลงมาคือ อัตราส่วนสารสนเทศระหว่างแบบทดสอบแบบเลือกชุดตัวถูกวัดขึ้นสูงกว่าความจำที่มีคำตอบถูกมากกว่า 1 ตัว และแบบทดสอบแบบตัดตัวลงวัดขึ้นสูงกว่าความจำที่มีคำตอบถูกมากกว่า 1 ตัว (1.381) ซึ่งแสดงให้เห็นว่า แบบทดสอบแบบเลือกชุดตัวถูกวัดขึ้นความจำที่มีคำตอบถูกมากกว่า 1 ตัว มีประสิทธิภาพสัมพัทธ์สูงสุด รองลงมาคือ แบบเลือกชุดตัวถูกวัดขึ้นสูงกว่าความจำที่มีคำตอบถูกมากกว่า 1 ตัว

อภิปรายผลการวิจัย

1. แบบทดสอบแบบตัดตัวลงวัดความรู้ขึ้นสูงกว่าความจำที่มีคำตอบถูกมากกว่า 1 ตัว มีรูปแบบการตอบกระจายมากกว่าแบบทดสอบอีก 7 ชุด อาจเนื่องมาจากผู้สอบไม่คุ้นเคยกับแบบทดสอบที่มีรูปแบบการตอบแบบตัดตัวลง อีกทั้งเป็นข้อสอบที่วัดระดับพฤติกรรมการเรียนรู้ขึ้นสูงกว่าความจำ

ซึ่งเป็นข้อสอบที่ค่อนข้างยากอยู่แล้ว ประกอบกับผู้สอบต้องตัดตัวลงที่ไม่ถูกออกไปทีละตัวจนตัดไม่ได้ ทำให้รูปแบบการตอบมีการกระจายค่อนข้างมาก สอดคล้องกับงานวิจัยของ เอมอร์ จังศิริพรปกรณ์ (2548) ที่พบว่า แบบทดสอบแบบตัดตัวลงมีรูปแบบการกระจายของคำตอบมากกว่าแบบเลือกชุดตัวถูก

2. แบบทดสอบทั้ง 8 ชุด ต่างมีความเที่ยงตรงตามสภาพ เมื่อใช้คะแนนสอบปลายภาคเป็นเกณฑ์ แสดงให้เห็นว่าแบบทดสอบทั้ง 8 ชุด นอกจากจะมีความเที่ยงตรงตามเนื้อหาโดยการตรวจสอบจากผู้ทรงคุณวุฒิแล้วยังมีความเที่ยงตรงตามสภาพด้วย แม้ว่าแบบทดสอบแบบเลือกชุดตัวถูกวัดความรู้ขั้นสูงกว่าความจำที่มีคำตอบถูกมากกว่า 1 ตัวจะมีค่าสัมประสิทธิ์สหสัมพันธ์กับคะแนนสอบปลายภาคอย่างไม่นัยสำคัญทางสถิติ แต่เมื่อทดสอบความแตกต่างของคะแนนฟิชเชอร์ซี แล้ว พบว่าไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ แสดงให้เห็นว่าแบบทดสอบทั้ง 8 ชุด มีความเที่ยงตรงตามสภาพ ทั้งนี้อาจเนื่องมาจากแบบทดสอบทั้ง 8 ชุด ผ่านกระบวนการตรวจสอบความเที่ยงตรงตามเนื้อหาโดยผู้ทรงคุณวุฒิ และสอดคล้องกับงานวิจัยของ เอมอร์ จังศิริพรปกรณ์ (2548) และงานวิจัยของ กาญจนา ศิริวัฒนพงษ์ (2520) ที่พบว่าแบบทดสอบที่มีจำนวนตัวถูกแตกต่างกันมีความเที่ยงตรงตามสภาพไม่แตกต่างกัน ในขณะเดียวกันวิธีการตรวจให้คะแนนความรู้บางส่วนต่างกันคือ แบบเลือกชุดตัวถูกและแบบตัดตัวลงก็ไม่มีผลต่อความเที่ยงตรงตามสภาพของแบบทดสอบอีกด้วย (เอมอร์ จังศิริพรปกรณ์, 2548; กาญจนา ศิริวัฒนพงษ์, 2520)

3. ค่าเฉลี่ยของค่าอำนาจจำแนกของแบบทดสอบคู่ขนานทั้ง 8 ชุด เมื่อวิเคราะห์โดยใช้ทฤษฎีการตอบสนองข้อสอบรายข้อ ในภาพรวมและในภาพย่อยรายเงื่อนไขไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ สอดคล้องกับงานวิจัยของ เอมอร์ จังศิริพรปกรณ์ (2548) ที่พบว่าแบบทดสอบที่มีจำนวนตัวถูกแตกต่างกันและวิธีการตรวจให้คะแนนความรู้บางส่วนต่างกันคือ แบบเลือกชุดตัวถูกและแบบตัดตัวลงก็ไม่มีผลต่อค่าอำนาจจำแนกของข้อสอบ รวมถึง ลำราญ มีแจ้ง (2525) พบว่าวิธีการตรวจให้คะแนนความรู้บางส่วนในแต่ละวิธีไม่มีผลต่อค่าอำนาจจำแนกของข้อสอบเมื่อวิเคราะห์ด้วยทฤษฎีการทดสอบแบบดั้งเดิม

4. ค่าความยากของแบบทดสอบคู่ขนานทั้ง 8 ชุด เมื่อวิเคราะห์โดยใช้ทฤษฎีการตอบสนองข้อสอบรายข้อพบว่าในภาพรวมแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 เมื่อพิจารณารายเงื่อนไขมีประเด็นที่น่าสนใจ ดังนี้

4.1 แบบทดสอบที่วัดระดับพฤติกรรมการเรียนรู้ที่ต่างกันมีผลต่อค่าความยากของแบบทดสอบ โดยแบบทดสอบที่วัดความรู้ขั้นสูงกว่าความจำมีค่าความยากสูงกว่าแบบทดสอบที่วัดความรู้ขั้นความจำ ซึ่งเป็นไปตามหลักอนุกรมวิธาน (Taxonomy) ซึ่งบลูมและคณะได้นำหลักการดังกล่าวมาใช้กับวัตถุประสงค์ทางการศึกษา ซึ่งไม่เพียงแต่จะช่วยให้ครูสามารถกำหนดวัตถุประสงค์ของการเรียนการสอนให้สอดคล้องกับเป้าหมายของการศึกษาตามวัตถุประสงค์ของหลักสูตรเท่านั้น แต่ยังช่วยให้ครูสามารถจัดลำดับวัตถุประสงค์ให้เหมาะสมกับการเรียนการสอน และช่วยให้บรรลุตรงตามเป้าหมายของ

การศึกษาอีกด้วย ทั้งนี้เนื่องจากการจำแนกวัตถุประสงค์ตามหลักอนุกรมวิธานนั้น มีคุณลักษณะที่เป็นการจัดลำดับหรือชั้นต่างๆ จากระดับที่ง่ายไปสู่ระดับที่ยากหรือซับซ้อนของการเรียนรู้ตนเอง (Bloom, B. S., 1979)

4.2 แบบทดสอบที่มีจำนวนตัวถูกต่างกันเมื่อใช้วิธีการตรวจเดียวกันและวัดความรู้ในระดับเดียวกัน ไม่มีผลต่อค่าความยากของแบบทดสอบ ยกเว้น แบบทดสอบแบบเลือกชุดตัวถูกที่วัดความรู้ขั้นสูงกว่าความจำที่มีจำนวนตัวถูกมากกว่า 1 ตัว จะมีค่าความยากมากกว่าแบบทดสอบที่มีจำนวนตัวถูกตัวเดียว ทั้งนี้อาจเนื่องมาจากแบบทดสอบที่วัดระดับพฤติกรรมการเรียนรู้ขั้นสูงกว่าความจำเป็นข้อสอบที่มีความยากมากอยู่แล้วแต่เมื่อมีจำนวนตัวถูกมากกว่า 1 ตัว ทำให้ระดับความยากของข้อสอบเพิ่มมากขึ้น แสดงให้เห็นว่าจำนวนตัวถูกที่มากกว่า 1 ตัว มีผลต่อค่าความยากของแบบทดสอบเมื่อวัดระดับพฤติกรรมการเรียนรู้ที่สูงกว่าความจำ สอดคล้องกับงานวิจัยของเอมอร์ จังศิริพรปกรณ์ (2548) ที่พบว่า แบบทดสอบที่มีจำนวนตัวถูกตัวเดียวมีความง่ายกว่าแบบทดสอบที่มีจำนวนตัวถูกมากกว่า 1 ตัว

4.3 แบบทดสอบที่มีจำนวนตัวถูกเหมือนกันและวัดความรู้ในระดับเดียวกันแต่ใช้วิธีการตรวจต่างกัน ไม่มีผลต่อค่าความยากของแบบทดสอบ ยกเว้น แบบทดสอบแบบตัดตัวลงที่วัดความรู้ขั้นสูงกว่าความจำที่มีจำนวนตัวถูกตัวเดียวจะยากมากกว่าแบบทดสอบแบบเลือกชุดตัวถูกวัดความรู้ขั้นสูงกว่าความจำที่มีจำนวนตัวถูกตัวเดียวอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทั้งนี้ผู้สอบต้องทำข้อสอบที่มีความยากสูงเพราะวัดความรู้ขั้นสูงกว่าความจำอีกทั้งต้องตัดตัวลงออกถึง 3 ตัว และผลการวิจัยนี้ไม่สอดคล้องกับ เอมอร์ จังศิริพรปกรณ์ (2548) และ สุพจน์ เกิดสุวรรณ (2545) ที่พบว่า แบบทดสอบแบบตัดตัวลงมีค่าความยากเฉลี่ยต่ำกว่าแบบทดสอบแบบเลือกชุดตัวถูก เหตุที่เป็นเช่นนี้อาจเนื่องมาจากงานวิจัยนี้มีเงื่อนไขของระดับพฤติกรรมการเรียนรู้เข้ามาเกี่ยวข้อง และผู้เรียนโดยทั่วไปไม่คุ้นเคยกับแบบทดสอบในลักษณะที่ต้องตัดตัวลงออกไปถึง 3 ตัว และจากผลการสังเกตในระหว่างที่นิสิตทำการสอบนิสิตให้ความเห็นว่า แบบทดสอบที่วัดความรู้ขั้นสูงกว่าความจำมีความยากมากอีกทั้งแบบทดสอบแบบตัดตัวลงมีความยากในการทำมากกว่าแบบทดสอบแบบเลือกชุดตัวถูก

5. แบบทดสอบแบบเลือกตัดตัวลงวัดความรู้ขั้นสูงกว่าความจำที่มีคำตอบถูกมากกว่า 1 ตัว เหมาะสำหรับผู้สอบที่มีความสามารถสูงเนื่องจากให้ค่าฟังก์ชันสารสนเทศของแบบทดสอบสูงสุดในช่วงความสามารถสูงสอดคล้องกับ เอมอร์ จังศิริพรปกรณ์ (2548, น. 112) ที่กล่าวว่า แบบทดสอบแบบเลือกตอบที่มีตัวถูกตัวเดียวเปิดโอกาสให้มีการเดาได้มากกว่าแบบทดสอบแบบเลือกตอบที่มีตัวถูกมากกว่า 1 ตัว เช่น ในกรณีที่ผู้สอบทราบว่าแบบทดสอบที่มีตัวถูกตัวเดียวและรู้ว่าตัวเลือกใดเป็นตัวถูกแต่ตัวเลือกที่เป็นตัวลงไม่ทราบว่าถูกหรือผิด เมื่อข้อสอบเป็นแบบที่มีตัวถูกตัวเดียวก็ทำให้สามารถเดาได้เช่นกัน และได้คะแนนเท่ากับผู้ที่รู้ว่าตัวเลือกใดถูกและตัวเลือกใดผิด

ข้อเสนอแนะ

ข้อเสนอแนะทั่วไป

1. จากผลการวิจัยพบว่า แบบทดสอบแบบเลือกชุดตัวถูกที่ใช้วัดความรู้ขั้นความจำที่มีคำตอบถูกมากกว่า 1 ตัว ให้ค่าฟังก์ชันสารสนเทศและมีประสิทธิภาพสัมพัทธ์สูงที่สุดจึงเป็นแนวทางที่มีความเป็นไปได้ในการนำไปใช้ในการวัดและประเมินผลระดับห้องเรียนสำหรับผู้สอน วิธีการตรวจให้คะแนนก็ไม่ยุ่งยากมากนัก ซึ่งสามารถแก้ปัญหาในเรื่องของการเดาคำตอบ และจำแนกคนที่รู้จริง รู้บางส่วน รู้ผิดบางส่วน และไม่รู้เลยออกจากกันได้ อีกทั้งยังเหมาะกับเนื้อหาวิชาโดยทั่วไป และกระบวนการสร้างและบริหารการสอบก็ใกล้เคียงกับแบบทดสอบที่ใช้กันโดยทั่วไป ส่วนพฤติกรรมการเรียนรู้ขั้นสูงกว่าความจำผู้สอนสามารถวัดความสามารถดังกล่าวโดยใช้วิธีการวัดประเมินแบบอื่นๆ เช่น การทำโครงงาน ชิ้นงาน การสาธิต แฟ้มสะสมผลงาน หรือการประเมินตามสภาพจริงในรูปแบบอื่นๆ

2. จากผลการวิจัยพบว่า แบบทดสอบแบบเลือกชุดตัวถูกวัดความรู้ขั้นสูงกว่าความจำที่มีคำตอบถูกมากกว่า 1 ตัว เป็นข้อสอบที่ยากเหมาะกับผู้สอบที่มีความสามารถสูง และแบบทดสอบแบบเลือกชุดตัวถูกวัดความรู้ขั้นสูงกว่าความจำที่มีคำตอบถูกตัวเดียวเป็นข้อสอบที่ง่ายเหมาะกับผู้สอบที่มีความสามารถต่ำ จึงเป็นทางเลือกให้นำไปใช้ในสถานการณ์ที่แตกต่างกัน เช่น การสอบคัดเลือก หรือการสอบเพื่อพัฒนาการหากเป็นการวัดประเมินเพื่อการพัฒนาโดยทั่วไปอาจเลือกใช้แบบทดสอบแบบเลือกชุดตัวถูกวัดความรู้ขั้นความจำหรือสูงกว่าความจำที่มีคำตอบถูกมากกว่า 1 ตัว เนื่องจากแบบทดสอบดังกล่าวให้ค่าฟังก์ชันสารสนเทศสูง

ข้อเสนอแนะในการวิจัยครั้งต่อไป

1. ควรมีการศึกษาคุณภาพของแบบทดสอบในรูปแบบต่างๆ เช่น วิธีการตรวจให้คะแนนความรู้บางส่วนแบบต่างๆ หรือจำนวนตัวถูกที่แตกต่างกัน วิธีการตอบแบบต่างๆ ที่มีคุณภาพและเหมาะสมกับนักเรียนในระดับประถมศึกษาเนื่องจากยังไม่พบงานวิจัยที่ศึกษากับกลุ่มตัวอย่างระดับประถมศึกษา เพื่อให้ได้องค์ความรู้ด้านการวัดผลการศึกษาอันจะนำไปใช้ในการพัฒนาการศึกษาของประเทศต่อไป

2. ควรมีการศึกษาเปรียบเทียบประสิทธิภาพของการทดสอบแบบปรับเหมาะ (Adaptive Testing) ที่ประยุกต์ใช้รูปแบบการให้คะแนนความรู้บางส่วนแบบต่างๆ

3. ควรมีการศึกษาปฏิสัมพันธ์ระหว่างเงื่อนไขต่างๆ ที่อาจมีผลต่อคุณภาพของแบบทดสอบแบบเลือกตอบ เพื่อเป็นสารสนเทศในการเลือกใช้รูปแบบการวัดและประเมินที่สอดคล้องกับสถานการณ์ต่างๆ ที่แตกต่างกัน

เอกสารอ้างอิง

- กาญจนา ศิริวัฒนพงษ์. (2520). การศึกษาเปรียบเทียบการตอบและการตรวจให้คะแนนแบบทดสอบปรนัย ชนิดเลือกตอบที่มีลักษณะต่างกัน. วิทยานิพนธ์การศึกษามหาบัณฑิต. มหาวิทยาลัย ศรีนครินทรวิโรฒประสานมิตร, กรุงเทพฯ.
- ศิริชัย กาญจนวาสี. (2545). ทฤษฎีการทดสอบแนวใหม่. กรุงเทพฯ ฯ: คณะครุศาสตร์ จุฬาลงกรณ์ มหาวิทยาลัย.
- สำราญ มีแจ้ง. (2525). ผลของการใช้คำสั่งและการให้คะแนนที่ต่างกันต่อค่าความเที่ยง ความตรงและ อำนาจจำแนกของแบบทดสอบชนิดเลือกตอบ. วิทยานิพนธ์ครุศาสตรมหาบัณฑิต. จุฬาลงกรณ์ มหาวิทยาลัย, กรุงเทพฯ.
- สุพจน์ เกิดสุวรรณ. (2545). การพัฒนาการวัดความรู้บางส่วนของผู้ตอบแบบสอบเลือกตอบ. วิทยานิพนธ์ ครุศาสตรดุษฎีบัณฑิต. จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพฯ.
- เอมอร จังศิริพรปกรณ์. (2548). การเปรียบเทียบคุณภาพของแบบสอบระหว่างแบบสอบเลือกตอบที่มี จำนวนตัวถูกเดียวกับตัวถูกมากกว่า 1 ตัว เมื่อตรวจด้วยวิธีการตรวจให้คะแนนความรู้บางส่วน. (อัคราณา). กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.
- Bloom, B. S., (1979). *Taxonomy of educational objectives: The classification of educational goals, handbook I cognitive domain*. USA: Longman.
- Lord, F. M., & Novick, M. R., (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Master, G. N., (1988). *The analysis of partial credit scoring: Applied measurement in education*, 1(4), 279-297.
- Wright, B. D., and Masters, G.N., (1982). *Rating scale analysis*. Chicago: MESA Press.
