

# Reliability of Multiple True-False Test Using Different Scoring Methods

Surachai Meechan

Ed.D.(Testing and Measurement), Lecturer  
Department of Educational Evaluation and Research, Faculty of Education  
Prince of Songkla University

## Abstract

This research was intended to compare the reliability of a multiple true-false test using five scoring methods: MR, Pass-3, Count-2, Count-3 and MTF. Cutting scores of test were determined by Berk's method and test reliabilities were determined by Livingston's method.

Samples in this research were undergraduates at the Faculty of Education, Prince of Songkla University enrolled in the course 267-201 Educational Measurement and Evaluation in the second semester of the academic year 1999. One hundred and seventy-three subjects were derived from purposive sampling. Instrument for data collection was a multiple true-false test developed to measure the achievement in learning the course 267-201. The test was a 30 item-four-choiced-test on Fequency Distribution, Measure of Central Tendency and Measure of Dispersion. The results of this study were as follows:

1. The reliability of the multiple true-false test using different scoring methods were significantly different at .05 level; Pass-3 method provided lowest reliability when compared with other methods with the exception of MR .
2. The proportion of examinees that were classified as mastery by the multiple true-false test using different scoring methods were significantly different at .05 level. Pass-3 method provided the proportion of mastery learners lower than Count-2 method. There was no difference among other methods.
3. The Phi-correlation coefficient between the results of mastery and nonmastery classification by the multiple true-false test using different scoring methods were positively high and significant at .05 level. It could be concluded that the multiple true-false test using different scoring methods could mostly classify mastery and nonmastery examinees consistently.

**Keywords:** Criterion-referenced test, multiple true-false test, scoring method, test reliability

## ความเชื่อมั่นของแบบทดสอบถูกผิดหลายตัวเลือก ที่ใช้วิธีตรวจให้คะแนนต่างกัน

สุรัชย์ มีชาญ

กศ.ด.(การทดสอบและวัดผลการศึกษา), อาจารย์

ภาควิชาประเมินผลและวิจัยทางการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยสงขลานครินทร์

การวิจัยครั้งนี้มีวัตถุประสงค์หลักเพื่อเปรียบเทียบความเชื่อมั่นของแบบทดสอบถูกผิดหลายตัวเลือกที่ใช้วิธีตรวจให้คะแนนต่างกัน 5 วิธี คือ วิธี MR, วิธี Pass-3, วิธี Count-2, วิธี Count-3 และวิธี MTF โดยกำหนดคะแนนจุดตัดของแบบทดสอบ ตามวิธีของ Berk (1976) และประมาณค่าความเชื่อมั่นของแบบทดสอบ ตามวิธีของ Livingston (1972)

กลุ่มตัวอย่างที่ใช้ในการวิจัย เป็นนักศึกษาระดับปริญญาตรี สังกัดคณะศึกษาศาสตร์ ที่ลงทะเบียนเรียนรายวิชา 267-201 การวัดและการประเมินผลการศึกษา ภาคการศึกษาที่ 2 ปีการศึกษา 2542 จำนวน 173 คน ซึ่งได้มาโดยวิธีกำหนดตัวอย่างแบบเจาะจง (Purposive Sampling) เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบอิงเกณฑ์ที่ผู้วิจัยสร้างขึ้นเพื่อวัดผลสัมฤทธิ์ทางการเรียนในรายวิชา 267-201 เรื่อง การแจกแจงความถี่ การวัดแนวโน้มเข้าสู่ส่วนกลาง และการวัดการกระจาย ประกอบด้วยข้อสอบถูกผิดหลายตัวเลือกที่มี 4 ตัวเลือก จำนวน 30 ข้อ

ผลการวิจัยพบว่า

1. แบบทดสอบถูกผิดหลายตัวเลือกที่ใช้วิธีตรวจให้คะแนนต่างกัน มีความเชื่อมั่นแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยการตรวจให้คะแนนด้วยวิธี Pass-3 ทำให้แบบทดสอบถูกผิดหลายตัวเลือกมีความเชื่อมั่นต่ำกว่าวิธีอื่นๆ ยกเว้นเมื่อเปรียบเทียบกับวิธี MR ซึ่งไม่แตกต่างกัน
2. สัดส่วนของผู้สอบที่ได้รับการจำแนกเป็นผู้รอบรู้ โดยแบบทดสอบถูกผิดหลายตัวเลือกที่ใช้วิธีตรวจให้คะแนนต่างกัน มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยการตรวจให้คะแนนด้วยวิธี Pass-3 ทำให้มีสัดส่วนของผู้สอบที่ได้รับการจำแนกเป็นผู้รอบรู้ ต่ำกว่าวิธี Count-2 สำหรับวิธีอื่นๆ ไม่แตกต่างกัน
3. สัมประสิทธิ์สหสัมพันธ์แบบพี (Phi-Correlation Coefficient) ระหว่างผลการจำแนกผู้สอบเป็นผู้รอบรู้และไม่รอบรู้ของแบบทดสอบถูกผิดหลายตัวเลือกที่ใช้วิธีตรวจให้คะแนนต่างกัน มีค่าเป็นบวกและค่อนข้างสูง โดยมีนัยสำคัญทางสถิติที่ระดับ .05 ทุกค่า แสดงว่าแบบทดสอบถูกผิดหลายตัวเลือกที่ใช้วิธีตรวจให้คะแนนต่างกัน สามารถจำแนกผู้สอบเป็นผู้รอบรู้และไม่รอบรู้ได้อย่างสอดคล้องกันเป็นส่วนใหญ่

คำสำคัญ: ความเชื่อมั่นของแบบทดสอบ, แบบทดสอบถูกผิดหลายตัวเลือก, แบบทดสอบอิงเกณฑ์, วิธีตรวจให้คะแนน

### บทนำ

แบบทดสอบอิงเกณฑ์ (Criterion-Referenced Test) เป็นแบบทดสอบที่สร้างขึ้น โดยยึดหลักการเรียน เพื่อความรอบรู้ สอบเพื่อปรับปรุงการเรียนการสอนเป็น

หลัก เนื้อหาในแบบทดสอบจึงมักเป็นวงแคบ และเน้นความครอบคลุมเนื้อหาและจุดประสงค์อย่างมาก แบบทดสอบแบบนี้มุ่งสอบเพื่อพัฒนาความรู้ความสามารถ

นั่นคือ คนใดสอบไม่ผ่านเกณฑ์ที่กำหนดก็ร่ำเรียนใหม่ แล้วสอบซ่อมเสริมด้วยแบบทดสอบเดิม ถ้าเกิดการ เรียนรู้เพิ่มขึ้นก็จะสามารถสอบผ่านเกณฑ์ได้ (ลัวัน สายยศ และ อังคณา สายยศ, 2539, 20) ซึ่งโดยทั่วไป มักจะนิยามเกณฑ์หรือมาตรฐานว่า หมายถึง คะแนน จุดตัดของแบบทดสอบที่ใช้แบ่งผู้สอบออกเป็นผู้สอบได้-ผู้สอบตก หรือผู้มีสมรรถภาพ-ผู้ไม่มีสมรรถภาพ (บุญเชิด ภิญโญอนันตพงษ์, 2527, 110) คะแนนจุดตัดนั้นบ่งชี้ความสำคัญอย่างยิ่งต่อคุณภาพของแบบทดสอบอิงเกณฑ์ และเป็นปัญหาเบื้องต้นที่สำคัญยิ่ง เพราะมักจะไม่สามารถ กำหนดคะแนนจุดตัดได้อย่างเป็นปรนัย สิ่งนี้นับว่าเป็น จุดอ่อนของการทดสอบแบบอิงเกณฑ์อย่างมาก (บุญเชิด ภิญโญอนันตพงษ์, 2527, 108) การกำหนดคะแนน จุดตัดที่สูงหรือต่ำเกินไป จะทำให้เกิดความคลาดเคลื่อน ในการจำแนกความรอบรู้ของผู้เรียน นอกจากนั้นยังส่ง ผลกระทบต่อสภาพการเรียนการสอนอีกด้วย กล่าวคือ ถ้ากำหนดคะแนนจุดตัดไว้ต่ำเกินไป อาจจะทำให้การ เรียนการสอนด้อยกว่ามาตรฐานที่ต้องการ หรือถ้า กำหนดสูงเกินไป ก็อาจทำให้จำนวนผู้เรียนที่สอบไม่ ผ่านเกณฑ์มีมาก ทั้งๆ ที่เขาอาจมีความรู้พอที่จะเรียน ต่อไปได้ (สงบ ลักษณะ, 2523, 22)

สำหรับรูปแบบของแบบทดสอบที่นิยมใช้กันโดยทั่วไปนั้น มี 5 แบบ คือ แบบความเรียง (Essay) แบบ ถูกผิด (True-False) แบบเติมคำ (Completion) แบบ จับคู่ (Matching) และแบบเลือกตอบ (Multiple Choices) ข้อสอบเลือกตอบเป็นข้อสอบที่มีผู้นิยมใช้อย่างมาก เพราะ สามารถวัดได้ครอบคลุมจุดประสงค์และตรวจให้คะแนน ได้แน่นอน สามารถใช้แทนข้อสอบรูปแบบอื่นๆ ได้ดี แม้แต่ข้อสอบความเรียง (ลัวัน สายยศ และ อังคณา สายยศ, 2539, 85-93) อย่างไรก็ตามข้อสอบเลือกตอบ ก็มีข้อบกพร่องบางประการ เช่น 1) ปัญหาบางปัญหาที่มีคำตอบที่ถูกต้องมากกว่าหนึ่งคำตอบ แต่ข้อสอบเลือก ตอบไม่สะดวกที่จะชี้ชัดว่าผู้สอบทราบคำตอบที่ถูกต้อง ทั้งหมดหรือไม่ 2) การมีตัวเลือกที่ถูกเพียงตัวเลือกเดียว อาจทำให้ผู้สอบตอบได้ถูกต้องโดยไม่มีความรู้ในเรื่องนั้น อย่างสมบูรณ์ 3) การตรวจให้คะแนน 1 คะแนน สำหรับ ผู้ตอบได้ถูกต้อง และให้ 0 คะแนน สำหรับผู้ที่ตอบไม่ถูก นั้น ไม่สามารถแยกแยะผู้ที่มีความรู้ถูกต้องสมบูรณ์

ผู้ที่มีความรู้บางส่วนและผู้ที่ไม่มีความรู้ออกจากกันได้ (จักรกฤษณ์ สำราญใจ, 2531, 3-4) ซึ่งจุดอ่อนหรือข้อ บกพร่องเหล่านี้เป็นสิ่งที่นักวัดผลการศึกษาพยายาม ปรับปรุงแก้ไขตลอดมา ทั้งนี้เพื่อให้ผลการวัดมีความ ถูกต้องมากยิ่งขึ้น การเลือกใช้ข้อสอบถูกผิดหลายตัว เลือกเป็นแนวทางหนึ่งในการแก้ไขจุดอ่อนหรือข้อบกพร่อง ดังกล่าว

ข้อสอบถูกผิดหลายตัวเลือก (Multiple True-False Item) มีลักษณะคล้ายคลึงกับข้อสอบหลายตัวเลือก หรือที่นิยมเรียกกันว่า ข้อสอบเลือกตอบ (Multiple-Choice Item) เกือบทุกประการ โดยมีส่วนประกอบอยู่ 2 ส่วน ได้แก่ ส่วนที่เป็นข้อคำถามหรือตอนนำ และ ส่วนที่เป็นตัวเลือก สิ่งที่แตกต่างกันก็คือ ข้อสอบถูกผิด หลายตัวเลือกอาจมีจำนวนตัวเลือกที่เป็นคำตอบถูก อยู่ 1, 2, 3, ... หรือถูกทุกตัวเลือก หรือไม่มีตัวถูกเลยก็ได้ ขึ้นอยู่กับเนื้อหาที่ต้องการวัด ดังตัวอย่าง

ข้อใดเป็นการทำลายพันธรั้วสัตว์ ?

ก. ให้สัตว์ผสมพันธุ์กันนอกฤดู

ข. ตอนสัตว์ที่มีลักษณะดี

ค. ฆ่าสัตว์ที่เป็นโรคทิ้ง

ง. ขายสัตว์ที่พิการออกไป

จ. ให้แม่พันธุ์ผสมพันธุ์กับลูก

จากคำถามนี้ จะเห็นได้ว่ามีตัวเลือกที่ถูก จำนวน 2 ตัวเลือก คือ ข. และ จ. นอกนั้นเป็นตัวเลือกที่ผิด จึง เห็นได้ว่าจุดประสงค์ที่สำคัญของการใช้ข้อสอบถูกผิด หลายตัวเลือก ก็เพื่อให้ผลจากการสอบมีความแม่นยำ และวัดได้ลึก ครอบคลุมเนื้อหาในบางตอนที่มีความสำคัญ ที่ผู้สร้างไม่สามารถถามโดยใช้ข้อสอบแบบเลือกตอบที่มีคำตอบถูกเพียงคำตอบเดียว (วัชรารักษ์ จิตรมาศ, 2538, 9-10)

ในการตอบข้อสอบรูปแบบนี้ ผู้สอบจึงต้องตอบ ว่าตัวเลือกแต่ละตัว ถูก (True) หรือผิด (False) ภายใต้ เงื่อนไขของข้อคำถามหรือตอนนำที่กำหนดมาให้ ดังนั้น ข้อความหรือรายละเอียดในแต่ละตัวเลือก จึงต้องพยายาม แยกให้อิสระจากกันมากที่สุด ยกเว้นประเด็นของการ พิจารณาถึงความถูกต้องของตัวเลือกนั้น ซึ่งพิจารณาอยู่ ภายใต้กรอบของข้อคำถาม หรือตอนนำเดียวกัน

สำหรับการตรวจให้คะแนนข้อสอบถูกผิดหลายตัวเลือกนั้น มีผู้เสนอแนะไว้หลายวิธีภายใต้แนวคิดหรือพื้นฐานความเชื่อที่แตกต่างกัน เช่น วิธี MR (Multiple-Response Method) หรือวิธี MTF-A (Multiple True-false Items with Absolute Knowledge Scoring) หรืออาจเรียกว่า วิธี 0-1, วิธี MTF (Multiple True-False Method) และวิธีอื่นๆ อีกหลายวิธี โดยในช่วงเวลาที่ผ่านมามีผู้ศึกษาค้นคว้าเกี่ยวกับคุณภาพของแบบทดสอบถูกผิดหลายตัวเลือกที่ใช้วิธีตรวจให้คะแนนที่แตกต่างกัน จำนวนหลายคน เช่น Albanese และ Sabers (1988) ได้ศึกษาเกี่ยวกับความเชื่อมั่นของแบบทดสอบถูกผิดหลายตัวเลือก เมื่อตรวจให้คะแนนต่างกัน 4 วิธี Fu-Ju Tsai และ Hoi K. Suen (1993) ศึกษาเกี่ยวกับความเชื่อมั่น ความยากง่าย และอำนาจจำแนกของแบบทดสอบถูกผิดหลายตัวเลือกที่ตรวจให้คะแนนต่างกัน 6 วิธี วัชรารักษ์ จิตรมาศ (2538) ศึกษาเกี่ยวกับความยากง่าย อำนาจจำแนก และความเชื่อมั่นของแบบทดสอบถูกผิดหลายตัวเลือก ที่ตรวจให้คะแนนต่างกัน 5 วิธี สมพงษ์ อัครศุภฤกษ์ (2542) ทำการเปรียบเทียบความเชื่อมั่นของแบบทดสอบที่มีวิธีตรวจให้คะแนน และความยาวของแบบทดสอบต่างกัน พรทิพย์ เพชรภาค (2543) ศึกษาเปรียบเทียบค่าฟังก์ชันสารสนเทศของแบบทดสอบถูกผิดหลายตัวเลือกที่ตรวจให้คะแนนต่างกัน 3 วิธี และ พรรณเพ็ญ เมืองสง (2543) ศึกษาเปรียบเทียบค่าฟังก์ชันสารสนเทศของแบบทดสอบถูกผิดหลายตัวเลือก ที่มีจำนวนข้อ ตัวเลือก และวิธีตรวจให้คะแนนต่างกัน เป็นต้น ซึ่งโดยภาพรวมพบว่า วิธีตรวจให้คะแนนที่ต่างกัน ส่งผลให้แบบทดสอบถูกผิดหลายตัวเลือกมีคุณภาพแตกต่างกัน อย่างไรก็ตาม การศึกษาค้นคว้าทั้งหมดที่กล่าวมา ล้วนมุ่งเน้นไปที่การนำแบบทดสอบถูกผิดหลายตัวเลือกไปใช้ในการทดสอบแบบอิงกลุ่ม ขณะที่ระบบการเรียนการสอนของประเทศไทยเกือบทุกระดับชั้นเน้นการทดสอบและประเมินแบบอิงเกณฑ์ ดังนั้นผู้วิจัยจึงต้องการศึกษาเพิ่มเติมว่าถ้านำข้อสอบถูกผิดหลายตัวเลือกมาใช้ในแบบทดสอบอิงเกณฑ์และเลือกใช้วิธีตรวจให้คะแนนต่าง ๆ กัน สิ่งเหล่านี้จะมีผลอย่างไรต่อคะแนนจุดตัดของแบบทดสอบ และส่งผลต่อเนื่องไปถึงความเชื่อมั่นของแบบทดสอบดังกล่าวหรือไม่ อย่างไร ทั้งนี้เพื่อใช้เป็นแนวทาง

ในการพัฒนาวิธีการวัดและประเมินแบบอิงเกณฑ์ และการประยุกต์ใช้ข้อสอบถูกผิดหลายตัวเลือกให้สามารถวัดผลการเรียนรู้ได้อย่างมีประสิทธิภาพยิ่งขึ้น และเพื่อใช้เป็นข้อมูลพื้นฐานสำหรับการส่งเสริมให้มีการใช้แบบทดสอบถูกผิดหลายตัวเลือกในห้องเรียนทั่วไปให้มากยิ่งขึ้น

### วัตถุประสงค์ของการวิจัย

1. เพื่อเปรียบเทียบความเชื่อมั่นของแบบทดสอบถูกผิดหลายตัวเลือกที่ใช้วิธีตรวจให้คะแนนต่างกัน
2. เพื่อเปรียบเทียบสัดส่วนของผู้สอบที่ได้รับการจำแนกเป็นผู้รอบรู้ของแบบทดสอบถูกผิดหลายตัวเลือกที่ใช้วิธีตรวจให้คะแนนต่างกัน
3. เพื่อศึกษาความสอดคล้องในการจำแนกผู้สอบตามความรอบรู้และไม่รอบรู้ ของแบบทดสอบถูกผิดหลายตัวเลือกที่ใช้วิธีตรวจให้คะแนนต่างกัน

### ขอบเขตของการวิจัย

1. ประชากร
 

ประชากรที่ใช้ในการวิจัยครั้งนี้ เป็นนักศึกษาระดับปริญญาตรี คณะศึกษาศาสตร์ มหาวิทยาลัยสงขลานครินทร์ ที่ลงทะเบียนเรียน รายวิชา 267-201 การวัดและการประเมินผลการศึกษา ภาคการศึกษาที่ 2 ปีการศึกษา 2542 จำนวน 352 คน
2. ตัวแปร
  - 2.1 ตัวแปรอิสระ ได้แก่ วิธีตรวจให้คะแนน แบ่งออกเป็น 5 วิธี
    - วิธี MR
    - วิธี Pass-3
    - วิธี Count-2
    - วิธี Count-3
    - วิธี MTF
  - 2.2 ตัวแปรตาม ได้แก่
    1. ความเชื่อมั่นของแบบทดสอบ
    2. สัดส่วนของผู้สอบที่ได้รับการจำแนกเป็นผู้รอบรู้
    3. ความสอดคล้องในการจำแนกผู้สอบตามความรอบรู้และไม่รอบรู้

3. เนื้อหาวิชาที่ใช้ในแบบทดสอบ ได้แก่ รายวิชา 267-201 การวัดและการประเมินผลการศึกษา เรื่อง การแจกแจงความถี่ การวัดแนวโน้มเข้าสู่ส่วนกลาง และการวัดการกระจาย

4. ในการวิจัยครั้งนี้ผู้วิจัยทำการประมาณค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ ตามวิธีการของ Livingston ซึ่งเป็นการประมาณค่าความเชื่อมั่นในรูปของความไวต่อระดับของการรอบรู้หรือไม่รอบรู้โดยใช้ความต่อเนื่องของคะแนน (Livingston, 1972 อ้างถึงใน บุญเชิด ภิญาญอนันตพงษ์, 2527, 189) ดังนั้นคะแนนจุดตัดจึงหมายถึงคะแนนจุดตัดที่คำนวณได้อย่างแท้จริง ไม่มีการบิดเบือนแต่อย่างใด โดยคะแนนจุดตัดดังกล่าวคำนวณหาได้ตามวิธีการของ Berk (1976)

#### นิยามศัพท์เฉพาะ และแนวคิดที่เกี่ยวข้อง

1. ข้อสอบถูกผิดหลายตัวเลือก (Multiple True-False Item) หมายถึง ข้อสอบชนิดให้เลือกคำตอบ (Selected-Response Item) รูปแบบหนึ่ง ซึ่งข้อสอบนี้มีส่วนประกอบ 2 ส่วน ได้แก่ ส่วนที่เป็นข้อคำถาม หรือ ตอนนำ อาจเป็นคำ, วลี, ประโยคสั้นๆ หรือข้อความบรรยายรายละเอียดของปรากฏการณ์หรือเหตุการณ์ และส่วนที่เป็นตัวเลือก จะเป็นคำหรือข้อความที่เกี่ยวข้องกับข้อคำถามหรือสอดคล้องกันในเรื่อง ซึ่งอาจมีจำนวนตัวเลือกที่เป็นคำตอบถูก อยู่ 1, 2, 3, ... หรือถูกทุกตัวเลือก หรือไม่มีตัวถูกเลยก็ได้ ในการตอบข้อสอบรูปแบบนี้ ผู้สอบจะต้องระบุว่าตัวเลือกแต่ละตัว ถูก (true) หรือ ผิด (False) ภายใต้เงื่อนไขข้อคำถามหรือตอนนำได้กล่าวถึงไว้

2. แบบทดสอบอิงเกณฑ์ (Criterion-Referenced Test) ในที่นี้หมายถึง แบบทดสอบผลสัมฤทธิ์ทางการเรียน รายวิชา 267-201 การวัดและการประเมินผลการศึกษา เรื่อง การแจกแจงความถี่ การวัดแนวโน้มเข้าสู่ส่วนกลาง และการวัดการกระจาย ที่ผู้วิจัยสร้างขึ้น ซึ่งประกอบด้วยข้อสอบถูกผิดหลายตัวเลือก ที่มี 4 ตัวเลือก จำนวน 30 ข้อ

3. วิธี MR (Multiple-Response Method) หรือ วิธี MTF-A (Multiple True-False Item with Absolute

Knowledge) บางครั้งก็เรียกกันว่า วิธี 0-1 หมายถึง วิธีตรวจให้คะแนนที่ให้ความสำคัญกับความรู้ที่ถูกต้องสมบูรณ์ โดยผู้สอบจะต้องตอบตรงกับสภาพความเป็นจริง (ตรงกับคำตอบ) ทุกตัวเลือก จึงจะได้ 1 คะแนน แต่ถ้าไม่ตรงกับสภาพความเป็นจริง แม้เพียงตัวเลือกเดียวก็จะได้ 0 คะแนน (สุรชัย มีชาญ, 2537, 34) วิธีนี้จึงมีจุดมุ่งหมายเพื่อป้องกันผู้สอบมีความรู้ความสามารถในเรื่องนั้นๆ ถูกต้องสมบูรณ์หรือไม่ ถ้าได้ 0 คะแนนก็แสดงว่ายังมีบางประเด็นที่บกพร่องอยู่ ดังนั้นคะแนนที่เป็นไปได้ของข้อสอบที่ตรวจให้คะแนนด้วยวิธีนี้จึงมี 2 ค่า ได้แก่ 0 และ 1 คะแนน

4. วิธี Pass-3 หมายถึง วิธีตรวจให้คะแนนที่มีจุดมุ่งหมายเพื่อป้องกันผู้สอบมีความรู้ความสามารถเพียงพอตามเกณฑ์ขั้นต่ำที่คาดหวังไว้หรือไม่ ในกรณีนี้จึงกำหนดไว้ว่า ถ้าผู้สอบตอบได้ตรงกับสภาพความเป็นจริง ตั้งแต่ 3 ตัวเลือกขึ้นไป จะถือว่ามีความรอบรู้ในเรื่องนั้นๆ มากพอที่จะยอมรับได้ จึงพิจารณาให้ผ่านคือ ได้ 1 คะแนน แต่ถ้าตอบได้ตรงกับสภาพความเป็นจริงน้อยกว่า 3 ตัวเลือก ก็จะได้ 0 คะแนน (สุรชัย มีชาญ, 2537, 34-35; พรทิพย์ เพชรภาค, 2543, 22) ดังนั้นคะแนนที่เป็นไปได้ของข้อสอบที่ตรวจให้คะแนนด้วยวิธีนี้ จึงมี 2 ค่า ได้แก่ 0 และ 1 คะแนน

5. วิธี Count-2 (Count for 2 Correct Options Method) หมายถึง วิธีตรวจให้คะแนนตามแนวคิดของ Fu-Ju Tsai และ Hoi K. Suen (1993) ที่ยึดหลักการว่า ผู้สอบควรจะได้คะแนนบางส่วน (ในข้อสอบแต่ละข้อ) ก็ต่อเมื่อสามารถระบุว่า ตัวเลือกนั้น ถูก หรือ ผิด ได้ตรงกับสภาพจริงของตัวเลือกนั้น ตั้งแต่ 2 ตัวเลือกขึ้นไป มีเงื่อนไขดังนี้

$$f(i) = \begin{cases} 0 & \text{เมื่อ } i = 0 \text{ หรือ } 1 \\ \frac{i-1}{k-1} & \text{สำหรับกรณีอื่นๆ} \end{cases}$$

โดยที่  $i$  แทน จำนวนตัวเลือกที่ผู้สอบตอบได้ตรงกับสภาพจริงของตัวเลือกนั้นๆ

$k$  แทน จำนวนตัวเลือกทั้งหมดของข้อสอบแต่ละข้อ



ดังนั้นในการวิจัยครั้งนี้ คะแนนที่เป็นไปได้ของข้อสอบที่ตรวจให้คะแนนด้วยวิธีนี้ จึงมี 4 ค่า ได้แก่ 0,  $\frac{1}{3}$ ,  $\frac{2}{3}$  และ 1 คะแนน

6. วิธี Count-3 (Count for 3 Correct Options Method) หมายถึง วิธีตรวจให้คะแนนตามแนวคิดของ Fu-Ju Tsai และ Hoi K. Suen (1993) ที่ยึดหลักการว่าผู้สอบควรจะได้คะแนนบางส่วน (ในข้อสอบแต่ละข้อ) ก็ต่อเมื่อสามารถระบุคำตอบที่ถูกต้อง ถูก หรือผิด ได้ตรงกับสภาพจริงของตัวเลือกนั้น ตั้งแต่ 3 ตัวเลือกขึ้นไป มีเงื่อนไขดังนี้

$$f(i) = \begin{cases} 0 & \text{เมื่อ } i = 0, 1 \text{ หรือ } 2 \\ \frac{i-2}{k-2} & \text{สำหรับกรณีอื่นๆ} \end{cases}$$

โดยที่  $i$  แทน จำนวนตัวเลือกที่ผู้สอบตอบได้ตรงกับสภาพจริงของตัวเลือกนั้นๆ

$k$  แทน จำนวนตัวเลือกทั้งหมดของข้อสอบแต่ละข้อ

ดังนั้นในการวิจัยครั้งนี้ คะแนนที่เป็นไปได้ของข้อสอบที่ตรวจให้คะแนนด้วยวิธีนี้ จึงมี 3 ค่า ได้แก่ 0, .50 และ 1 คะแนน

7. วิธี MTF (Multiple True-False Method) หมายถึง วิธีตรวจให้คะแนนตามแนวคิดของ Albanese และคณะ (1979) ซึ่งเป็นการตรวจให้คะแนนรายตัวเลือก โดยจะได้ตัวเลือกละ 1 คะแนน เมื่อผู้สอบตอบได้ตรงกับสภาพจริงของตัวเลือกนั้นๆ และได้ 0 คะแนน เมื่อตอบไม่ตรงกับสภาพจริงของตัวเลือกนั้นๆ ดังนั้นคะแนนเต็มทั้งฉบับจึงเท่ากับจำนวนตัวเลือกทั้งหมดของแบบทดสอบนั้น ด้วยเหตุนี้จึงมีผู้เสนอแนะให้ตรวจให้คะแนนในแต่ละตัวเลือกเป็นจำนวนทศนิยมแทน โดยเมื่อคำนวณคะแนนรวมจากทุกตัวเลือกของข้อสอบแต่ละข้อ แล้วให้มีค่าเท่ากับ 1 คะแนน ทั้งนี้เพื่อให้คะแนนรวมทั้งฉบับมีคะแนนเต็มเท่ากับจำนวนข้อสอบทั้งหมดนั่นเอง เช่น ในกรณีที่ข้อสอบนั้นมี 4 ตัวเลือก ก็ให้ตัวเลือกละ .25 คะแนน เป็นต้น ในการวิจัยครั้งนี้ ผู้วิจัยเลือกใช้กรณีหลัง คือ ให้คะแนนเป็นรายตัวเลือก ตัวเลือกละ .25 คะแนน เพื่อจำกัดขอบเขตของการศึกษาไว้เฉพาะวิธีตรวจให้

คะแนนที่ให้คะแนนรายข้อ มีค่าตั้งแต่ 0-1 คะแนน เท่านั้น ดังนั้นในการวิจัยครั้งนี้ คะแนนที่เป็นไปได้ของข้อสอบที่ตรวจให้คะแนนด้วยวิธีนี้ จึงมี 5 ค่า ได้แก่ 0, .25, .50, .75 และ 1 คะแนน

8. คะแนนจุดตัด (Cutting Score) หมายถึง คะแนนที่ใช้เป็นเกณฑ์ในการตัดสินว่าผู้สอบเป็นผู้รอบรู้หรือไม่รอบรู้ จากการสอบด้วยแบบทดสอบอิงเกณฑ์ฉบับนั้น ในการวิจัยครั้งนี้กำหนดคะแนนจุดตัดตามวิธีของ Berk (1976) และผู้สอบที่จะได้รับการจำแนกเป็นผู้รอบรู้ก็คือ ผู้ที่สอบได้คะแนนไม่น้อยกว่าคะแนนจุดตัด

9. ค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ โดยทั่วไปหมายถึง ดัชนีที่บ่งชี้ว่าแบบทดสอบอิงเกณฑ์ฉบับนั้นมีความคงที่มากน้อยเพียงใดในการจำแนกผู้สอบเป็นผู้รอบรู้หรือไม่รอบรู้ในเรื่องที่สอบ แต่ในการวิจัยครั้งนี้ ผู้วิจัยประมาณค่าความเชื่อมั่นของแบบทดสอบตามวิธีของ Livingston ซึ่งเป็นการประมาณค่าความเชื่อมั่นในรูปของความไวต่อระดับของการรอบรู้หรือไม่รอบรู้ โดยใช้ความต่อเนื่องของคะแนน

10. สัดส่วนของผู้สอบที่ได้รับการจำแนกเป็นผู้รอบรู้ หมายถึง สัดส่วน (Proportion) ของจำนวนผู้สอบที่ได้รับการจำแนกเป็นผู้รอบรู้โดยแบบทดสอบอิงเกณฑ์ฉบับนั้น เมื่อเทียบกับจำนวนของผู้สอบทั้งหมด

11. ความสอดคล้องในการจำแนกผู้สอบตามความรู้และไม่รอบรู้ ในที่นี้หมายถึง ความสอดคล้องกันของผลการจำแนกผู้สอบใน 2 ครั้งใดๆ ซึ่งพิจารณาได้จากสัมประสิทธิ์สหสัมพันธ์แบบพี (Phi-Correlation:  $\Phi$ ) ของผลการจำแนกผู้สอบ ของแบบทดสอบถูกผิดหลายตัวเลือกที่ใช้วิธีตรวจให้คะแนน 2 วิธีที่ต่างกันนั้น

#### สมมติฐานของการวิจัย

1. ความเชื่อมั่นของแบบทดสอบถูกผิดหลายตัวเลือกที่ใช้วิธีตรวจให้คะแนนต่างกัน มีความแตกต่างกัน

2. สัดส่วนของผู้สอบที่ได้รับการจำแนกเป็นผู้รอบรู้ของแบบทดสอบถูกผิดหลายตัวเลือกที่ใช้วิธีตรวจให้คะแนนต่างกัน มีความแตกต่างกัน

3. ผลการจำแนกผู้สอบตามความรู้และไม่รอบรู้ ของแบบทดสอบถูกผิดหลายตัวเลือกที่ใช้วิธีตรวจให้คะแนนต่างกัน ไม่สอดคล้องกัน

### วิธีดำเนินการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์หลักเพื่อเปรียบเทียบความเชื่อมั่นแบบอิงเกณฑ์ของแบบทดสอบถูกผิดหลายตัวเลือกที่ใช้วิธีตรวจให้คะแนนต่างกัน โดยมีรายละเอียดและขั้นตอนดำเนินการดังต่อไปนี้

#### 1. ประชากรและกลุ่มตัวอย่าง

ประชากรที่ใช้ในการวิจัยครั้งนี้ เป็นนักศึกษา ระดับปริญญาตรี สังกัดคณะศึกษาศาสตร์ มหาวิทยาลัย สงขลานครินทร์ ที่ลงทะเบียนเรียนในรายวิชา 267-201 การวัดและการประเมินผลการศึกษา ภาคการศึกษาที่ 2 ปีการศึกษา 2542 จำนวน 352 คน และกลุ่มตัวอย่างที่ใช้ในการวิจัย จำนวน 173 คน ซึ่งได้มาโดยวิธีกำหนดตัวอย่างแบบเจาะจง (Purposive Sampling) เพื่อให้ครอบคลุมกลุ่มนักศึกษาที่เรียนกับอาจารย์ผู้สอนแต่ละคน และเรียนอยู่ในสาขาวิชาเอกต่างๆ กัน ทั้งสายวิทยาศาสตร์ และสายศิลปศาสตร์

#### 2. เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบอิงเกณฑ์ที่ผู้วิจัยสร้างขึ้น เพื่อวัดผลสัมฤทธิ์ทางการเรียน ในรายวิชา 267-201 การวัดและการประเมินผลการศึกษา เรื่อง การแจกแจงความถี่ การวัดแนวโน้มเข้าสู่ ส่วนกลาง และการวัดการกระจาย ประกอบด้วยข้อสอบ ถูกผิดหลายตัวเลือก ที่มี 4 ตัวเลือก จำนวน 30 ข้อ

#### 3. วิธีดำเนินการสร้างเครื่องมือ

ผู้วิจัยดำเนินการสร้างเครื่องมือที่ใช้ในการวิจัย ดังนี้

1. ศึกษารายละเอียดเกี่ยวกับเนื้อหาสาระ ในรายวิชา 267-201 การวัดและการประเมินผลการศึกษา ในส่วนที่เกี่ยวข้องกับสถิติพื้นฐานเพื่อการวัดและประเมินผลการศึกษา จากประมวลการสอนและเอกสารคำสอน ที่อาจารย์ผู้สอนทุกคนร่วมกันจัดทำไว้

2. คัดเลือกจุดประสงค์การเรียนรู้ที่สำคัญ ที่ จะนำมาสร้างแบบทดสอบ

3. เขียนข้อสอบถูกผิดหลายตัวเลือก ที่มี 4 ตัวเลือก เพื่อวัดในแต่ละจุดประสงค์การเรียนรู้ ทั้งนี้โดย ยึดถือนำหนักความสำคัญของจุดประสงค์ ตามกรอบที่ อาจารย์ผู้สอนร่วมกันกำหนดไว้ รวมข้อสอบที่เขียนขึ้น ทั้งหมด จำนวน 50 ข้อ

4. ให้อาจารย์ผู้สอนช่วยตรวจสอบความ สอดคล้องของข้อสอบกับจุดประสงค์การเรียนรู้ โดยให้ พิจารณาว่าข้อสอบแต่ละข้อวัดความสามารถของผู้เรียน ได้ตรงตามที่จุดประสงค์การเรียนรู้ นั้นๆ กำหนดไว้หรือไม่ ซึ่งในการวิจัยครั้งนี้ ผู้วิจัยกำหนดเกณฑ์ไว้ว่าข้อสอบที่ ผ่านการคัดเลือกต้องเป็นข้อสอบที่อาจารย์ผู้สอนเห็นว่า สามารถวัดความสามารถของผู้เรียนได้ตรงตามจุดประสงค์ การเรียนรู้ นั้นๆ อย่างแท้จริง

5. คัดเลือกข้อสอบเพื่อนำไปจัดฉบับและ ทดลองใช้ โดยนำแบบทดสอบดังกล่าวไปสอบกับนักศึกษา ที่ไม่ใช่กลุ่มตัวอย่างในการวิจัย ในที่นี้คือนักศึกษาที่เรียน รายวิชา 267-201 ในห้องเรียนอื่นๆ จำนวน 92 คน โดย การทดสอบ 2 ครั้ง ก่อนเรียนและหลังเรียน

6. นำกระดาษคำตอบของนักศึกษา มาตรวจ ให้คะแนน และทำการวิเคราะห์คุณภาพของข้อสอบเป็น รายข้อ เพื่อหาค่าอำนาจจำแนก โดยใช้ดัชนี PPD (Pre-to-Post Different Index) ของ Cox และ Vargas (1966 อ้างถึงใน ล้วน สายยศ และ อังคณา สายยศ, 2539, 197) เมื่อใช้วิธีตรวจให้คะแนนแบบสองค่า ได้แก่ วิธี MR และ วิธี Pass-3 และในกรณีที่ตรวจให้คะแนนแบบหลายค่า (วิธีอื่นๆ) จะทำการคำนวณค่าความยากง่ายของข้อสอบ ทั้งก่อนเรียนและหลังเรียน โดยใช้สูตรของ Whitney และ Sabers (1970 อ้างถึงใน ล้วน สายยศ และ อังคณา สายยศ, 2539, 199) ทั้งนี้การวิเคราะห์คุณภาพรายข้อ ดังกล่าว กระทำซ้ำ จำนวน 5 ครั้ง โดยยึดถือวิธีตรวจให้ คะแนนที่ต่างกัน 5 วิธี

7. คัดเลือกข้อสอบ เฉพาะข้อที่มีค่าอำนาจ จำแนกเป็นบวก หรือมีค่าความยากง่ายหลังเรียนสูงกว่า ก่อนเรียน เพื่อนำไปสร้างเป็นแบบทดสอบอิงเกณฑ์ฉบับ สมบูรณ์ โดยจะพิจารณาค่าอำนาจจำแนก ค่าความยาก ง่าย ทั้งก่อนเรียนและหลังเรียน และคำนึงถึงน้ำหนัก ความสำคัญของจุดประสงค์การเรียนรู้แต่ละข้อ รวม ข้อสอบที่คัดเลือกไว้ทั้งหมด จำนวน 30 ข้อ

8. นำข้อมูลผลการตอบข้อสอบ ทั้ง 30 ข้อ ที่ผ่านการคัดเลือก มาจัดเรียงใหม่ และคำนวณหาคะแนน จุดตัดของแบบทดสอบอิงเกณฑ์ฉบับสมบูรณ์ โดย กำหนดคะแนนจุดตัดตามวิธีการของ Berk (1976) ทั้งนี้ เพื่อนำไปใช้เป็นเกณฑ์ในการตัดสินความรอบรู้และไม่

รอบรู้ของผู้สอบในโอกาสต่อไป ทั้งนี้โดยคำนวณซ้ำ 5 ครั้ง ภายใต้งื่อนไขของการตรวจให้คะแนน 5 วิธี ดังกล่าวแล้ว

9. จัดทำแบบทดสอบอิงเกณฑ์ฉบับสมบูรณ์ เพื่อใช้เป็นเครื่องมือในการเก็บรวบรวมข้อมูลต่อไป

#### 4. การเก็บรวบรวมข้อมูล

ผู้วิจัยได้ดำเนินการตามขั้นตอนดังต่อไปนี้

1. ทำหนังสือถึงอาจารย์ผู้สอน รายวิชา 267-201 เพื่อขอความอนุเคราะห์ในการเก็บรวบรวมข้อมูล

2. นำแบบทดสอบที่ผู้วิจัยสร้างขึ้น ไปทดสอบกับนักศึกษาที่เป็นกลุ่มตัวอย่าง

3. ตรวจให้คะแนนผลการสอบ และประมาณค่าความเชื่อมั่นของแบบทดสอบ ตามวิธีการของ Livingston และคำนวณหาสัดส่วนของผู้สอบที่ได้รับการจำแนกเป็นผู้รอบรู้ ทั้งนี้โดยทำการคำนวณซ้ำ 5 ครั้ง ภายใต้งื่อนไขของคะแนนจุดตัดและการตรวจให้คะแนนแต่ละวิธี และคำนวณสัมประสิทธิ์สหสัมพันธ์แบบพี (Φ) ของผลการจำแนกผู้สอบใน 2 ครั้งใดๆ เมื่อแบบทดสอบถูกผิดหลายตัวเลือกใช้วิธีตรวจให้คะแนน 2 วิธีที่ต่างกัน

4. นำค่าความเชื่อมั่นของแบบทดสอบ, สัดส่วนของผู้สอบที่ได้รับการจำแนกเป็นผู้รอบรู้ และสัมประสิทธิ์สหสัมพันธ์แบบพี ที่คำนวณได้ไปใช้ในการทดสอบสมมติฐานของการวิจัยต่อไป

#### 5. การวิเคราะห์ข้อมูล

1. คำนวณค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบ

2. ประมาณค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์โดยใช้สูตรของ Livingston จากนั้นจึงแปลงค่าความเชื่อมั่นของแบบทดสอบให้อยู่ในรูปของคะแนนมาตรฐานของฟิชเชอร์ (Fisher's Z Transformation) และทดสอบความแตกต่างของค่าความเชื่อมั่นของแบบทดสอบ โดยการทดสอบไค-สแควร์ (Chi-square Test) หากพบว่าแบบทดสอบมีค่าความเชื่อมั่นแตกต่างกันอย่างมีนัยสำคัญทางสถิติ จะทดสอบความแตกต่างเป็นรายคู่โดยการทดสอบซี (Z-test)

3. คำนวณสัดส่วนของผู้สอบที่ได้รับการจำแนกเป็นผู้รอบรู้ และทดสอบความแตกต่างของสัดส่วนดังกล่าวโดยใช้การทดสอบซี (Z-proportion)

4. ศึกษาความสอดคล้องในการจำแนกผู้สอบตามความรอบรู้และไม่รอบรู้ ของแบบทดสอบที่ใช้วิธีตรวจให้คะแนนต่างกัน 2 วิธีใดๆ โดยคำนวณสัมประสิทธิ์สหสัมพันธ์แบบพี (Phi-Correlation Coefficient)

#### ผลการวิจัย

1. ค่าสถิติพื้นฐาน คะแนนจุดตัด และค่าความเชื่อมั่นของแบบทดสอบ

ผู้วิจัยได้นำคะแนนผลการสอบของนักศึกษาในกลุ่มตัวอย่างทั้ง 173 คน มาคำนวณหาค่าสถิติพื้นฐาน ได้แก่ ค่าเฉลี่ย ( $\bar{X}$ ) และส่วนเบี่ยงเบนมาตรฐาน (S.D.) และนำคะแนนจุดตัด (C) ของแบบทดสอบที่ได้กำหนดไว้มาใช้ในการประมาณค่าความเชื่อมั่นของแบบทดสอบ

ตารางที่ 1 ค่าสถิติพื้นฐาน คะแนนจุดตัด และค่าความเชื่อมั่นของแบบทดสอบ เมื่อใช้วิธีตรวจให้คะแนนแต่ละวิธี

วิธีตรวจให้คะแนน	คะแนนเต็ม	$\bar{X}$	S.D.	C	$r_{tt}^*$	$r_{cc}$
MR	30	7.8439	3.8359	6.00	.6764	.7371
Pass-3	30	16.2832	4.0871	16.00	.6360	.6377
Count-2	30	16.3622	3.1422	14.33	.7134	.7979
Count-3	30	12.0636	3.7477	10.00	.7130	.7798
MTF	30	19.5159	2.5382	18.50	.7191	.7579

\* ประมาณค่าความเชื่อมั่นตามวิธีการของ Kuder-Richardson สูตร 20 (K-R 20) และใช้สัมประสิทธิ์แอลฟา (Alpha Coefficient) ของ Cronbach เมื่อตรวจให้คะแนนแบบ 0-1 (วิธี MR และวิธี Pass-3) และแบบหลายค่า (วิธีอื่นๆ) ตามลำดับ



ฉบับดังกล่าว ทั้งแบบอิงกลุ่ม ( $r_{cc}$ ) และแบบอิงเกณฑ์ ( $r_{cc}$ ) ทั้งนี้โดยการคำนวณซ้ำ รวม 5 ครั้ง ตามวิธีตรวจให้คะแนนที่แตกต่างกัน 5 วิธี คือ วิธี MR, วิธี Pass-3, วิธี Count-2, วิธี Count-3 และวิธี MTF ปรากฏผลการคำนวณดังตารางที่ 1

จากตารางที่ 1 แสดงให้เห็นว่า คะแนนผลการสอบของผู้สอบทั้งหมดที่ได้จากแบบทดสอบที่ตรวจให้คะแนนต่างกัน 5 วิธี คือ วิธี MR, วิธี Pass-3, วิธี Count-2, วิธี Count-3 และวิธี MTF มีค่าเฉลี่ย (และส่วนเบี่ยงเบนมาตรฐาน) เท่ากับ 7.8439 (3.8359), 16.2832 (4.0871), 16.3622 (3.1422), 12.0636 (3.7477) และ 19.5159 (2.5382) ตามลำดับ เมื่อนำคะแนนผลการสอบทั้งหมดที่ได้จากการตรวจให้คะแนนในแต่ละวิธีไปใช้ในการประมาณค่าความเชื่อมั่นแบบอิงกลุ่ม ปรากฏว่าแบบทดสอบมีค่าความเชื่อมั่นเท่ากับ .6764, .6360, .7134, .7130 และ .7191 ตามลำดับ ซึ่งเมื่อนำค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของคะแนนรวมทั้งนำคะแนนจุดตัด และค่าความเชื่อมั่นแบบอิงกลุ่มไปใช้ประมาณค่าความเชื่อมั่นแบบอิงเกณฑ์ตามวิธีการของ Livingston แบบทดสอบจะมีค่าความเชื่อมั่นเท่ากับ .7371, .6377, .7979, .7798 และ .7579 ตามลำดับ

2. ผลการเปรียบเทียบความเชื่อมั่นของแบบทดสอบถูกผิดหลายตัวเลือกที่ใช้วิธีตรวจให้คะแนนต่างกัน เพื่อให้ทราบว่าคุณค่าความเชื่อมั่นของแบบทดสอบ เมื่อตรวจให้คะแนนด้วยวิธีที่ต่างกันมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติหรือไม่ ผู้วิจัยจึง

แปลงค่าความเชื่อมั่นแต่ละค่านี้นี้ให้เป็นคะแนนมาตรฐาน ฟิชเชอร์ (Fisher's Z Transformation) และนำไปทดสอบความแตกต่าง โดยการทดสอบไค-สแควร์ (Chi-square Test) ปรากฏผลดังตารางที่ 2

จากตารางที่ 2 แสดงให้เห็นว่า เมื่อใช้วิธีตรวจให้คะแนนต่างกัน แบบทดสอบถูกผิดหลายตัวเลือกจะมีความเชื่อมั่นแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

เพื่อให้ทราบว่าค่าความเชื่อมั่นคู่ใดบ้างที่มีความแตกต่างกัน จึงทดสอบความแตกต่างเป็นรายคู่ โดยการทดสอบซี (Z-test) ปรากฏผลดังตารางที่ 3

จากตารางที่ 3 แสดงให้เห็นว่า แบบทดสอบถูกผิดหลายตัวเลือกที่ตรวจให้คะแนนด้วยวิธี Pass-3 จะมีความเชื่อมั่นต่ำกว่า วิธี Count-2, วิธี Count-3 และวิธี MTF อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สำหรับคู่อื่นๆ ไม่แตกต่างกัน

3. ผลการเปรียบเทียบสัดส่วนของผู้สอบที่ได้รับการจำแนกเป็นผู้รอบรู้ ของแบบทดสอบถูกผิดหลายตัวเลือก ที่ใช้วิธีตรวจให้คะแนนต่างกัน

เนื่องจากมีความเป็นไปได้ที่แบบทดสอบอิงเกณฑ์ที่มีความเชื่อมั่นเท่ากันจะจำแนกผู้เรียนเป็นผู้รอบรู้และไม่รอบรู้แตกต่างกัน ทั้งในแง่ของจำนวนและความสอดคล้องของผลการจำแนก ดังนั้นผู้วิจัยจึงได้ตรวจนับจำนวน และคำนวณหาสัดส่วนของผู้ที่ได้รับการตัดสินว่าเป็นผู้รอบรู้และไม่รอบรู้จากการตอบแบบทดสอบถูกผิดหลายตัวเลือกที่ใช้วิธีตรวจให้คะแนนต่างกัน และ

ตารางที่ 2 ผลการทดสอบความแตกต่างของคุณค่าความเชื่อมั่นของแบบทดสอบถูกผิดหลายตัวเลือกที่ใช้วิธีตรวจให้คะแนนต่างกัน

วิธีตรวจให้คะแนน	$r_{cc}$	$Z_r$	$\chi^2$
MR	.7371	.9441	11.6946*
Pass-3	.6377	.7543	
Count-2	.7979	1.0928	
Count-3	.7798	1.0449	
MTF	.7579	.9913	

ตารางที่ 3 ผลการทดสอบความแตกต่างเป็นรายคู่ของค่าความเชื่อมั่นของแบบทดสอบ ถูกผิดหลายตัวเลือก ที่ใช้วิธีตรวจให้คะแนนต่างกัน

วิธีตรวจให้คะแนน	MR		Pass-3	Count-2	Count-3	MTF
	Z	.9441	.7543	1.0928	1.0449	.9913
MR	.9441	+	1.7499	-1.3709	-.9293	-.4352
Pass-3	.7543		+	-3.1208*	-2.6792*	-2.1850*
Count-2	1.092			+	.4416	.9358
Count-3	1.0449				+	.4942
MTF	.9913					+

ตารางที่ 4 จำนวน และสัดส่วนของผู้สอบที่ได้รับการจำแนกเป็นผู้รอบรู้และผู้ไม่รอบรู้ จากการทดสอบด้วยแบบทดสอบถูกผิดหลายตัวเลือกที่ใช้วิธีตรวจให้คะแนนต่างกัน

วิธีตรวจให้คะแนน	จำนวนผู้สอบทั้งหมด	จำนวนผู้รอบรู้	จำนวนผู้ไม่รอบรู้	สัดส่วนของผู้รอบรู้
MR	173	116	57	.6705
Pass-3	173	102	71	.5896
Count-2	173	132	41	.7630
Count-3	173	125	48	.7225
MTF	173	117	56	.6763

ทำการทดสอบความแตกต่างของสัดส่วนดังกล่าว เป็นรายคู่ โดยการทดสอบซี้ (Z-proportion) ปรากฏผลดังตารางที่ 4 และ 5

จากตารางที่ 4 แสดงให้เห็นว่า จากจำนวนผู้สอบทั้งหมด 173 คน ถ้าแบบทดสอบถูกผิดหลายตัวเลือกใช้วิธีตรวจให้คะแนนแต่ละวิธี คือ วิธี MR, วิธี Pass-3, วิธี Count-2, วิธี Count-3 และวิธี MTF จะมีผู้สอบที่ได้รับการจำแนกเป็นผู้รอบรู้ จำนวน 116, 102, 132, 125 และ 117 คน ตามลำดับ โดยคิดเป็นสัดส่วนของจำนวนผู้รอบรู้ต่อจำนวนผู้สอบทั้งหมดเท่ากับ .6705, .5896, .7630, .7225 และ .6763 ตามลำดับ

จากตารางที่ 5 แสดงให้เห็นว่า เฉพาะการตรวจให้คะแนนด้วยวิธี Pass-3 กับวิธี Count-2 เท่านั้นที่ทำให้สัดส่วนของผู้สอบที่ได้รับการจำแนกเป็นผู้รอบรู้ มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยเมื่อตรวจให้คะแนนด้วยวิธี Pass-3 จะมีสัดส่วนของ

ผู้สอบที่ได้รับการจำแนกเป็นผู้รอบรู้ต่ำกว่าเมื่อตรวจให้คะแนนด้วยวิธี Count-2 สำหรับคู่อื่นๆ ไม่แตกต่างกัน

4. ผลการศึกษาความสอดคล้องในการจำแนกผู้สอบตามความรอบรู้และไม่รอบรู้ ของแบบทดสอบถูกผิดหลายตัวเลือก ที่ใช้วิธีตรวจให้คะแนนต่างกัน

เพื่อให้เห็นถึงความแตกต่างหรือความสอดคล้องกันของผลการตรวจให้คะแนนแต่ละวิธีเพิ่มเติมยิ่งขึ้น ผู้วิจัยจึงได้ตรวจสอบความสอดคล้องในการจำแนกผู้สอบตามความรอบรู้และไม่รอบรู้ โดยการคำนวณสัมประสิทธิ์สหสัมพันธ์แบบฟี (Phi-Correlation Co-efficient) ของผลการจำแนก ระหว่างวิธีตรวจให้คะแนนต่างๆ กัน เป็นรายคู่ ปรากฏผลดังตารางที่ 6

จากตารางที่ 6 แสดงให้เห็นว่า สัมประสิทธิ์สหสัมพันธ์แบบฟี ( $\phi$ ) ของผลการจำแนกผู้สอบ ของแบบทดสอบถูกผิดหลายตัวเลือกที่ใช้วิธีตรวจให้คะแนนต่างกัน มีค่าเป็นบวกและค่อนข้างสูง โดยมีนัยสำคัญทาง

ตารางที่ 5 ผลการทดสอบความแตกต่างของสัดส่วน (P) ของผู้สอบที่ได้รับการจำแนกเป็นผู้รอบรู้ของแบบทดสอบถูกผิดหลายตัวเลือกที่ใช้วิธีตรวจให้คะแนนต่างกัน

วิธีตรวจให้คะแนน	MR Pass-3 Count-2 Count-3 MTF					
	สัดส่วน (P)	.6705	.5896	.7630	.7225	.6763
MR	.6705	+	.0809	-.0925	-.0520	-.0058
Pass-3	.5896		+	-.1734*	-.1329	-.0867
Count-2	.7630			+	.0405	.0867
Count-3	.7225				+	.0462
MTF	.6763					+

ตารางที่ 6 สัมประสิทธิ์สหสัมพันธ์แบบพี (Ø) ของผลการจำแนกผู้สอบเป็นผู้รอบรู้หรือไม่รอบรู้ของแบบทดสอบถูกผิดหลายตัวเลือกที่ใช้วิธีตรวจให้คะแนนต่างกัน

วิธีตรวจให้คะแนน	MR	Pass-3	Count-2	Count-3	MTF
MR	+	.540*	.650*	.692*	.619*
Pass-3		+	.668*	.743*	.779*
Count-2			+	.839*	.806*
Count-3				+	.940*
MTF					+

สถิติที่ระดับ .05 ทุกค่า แสดงว่าแบบทดสอบถูกผิดหลายตัวเลือกที่ใช้วิธีตรวจให้คะแนนต่างกัน สามารถจำแนกผู้สอบเป็นผู้รอบรู้หรือไม่รอบรู้ได้อย่างสอดคล้องกันเป็นส่วนใหญ่

#### อภิปรายผล

1. การเปรียบเทียบความเชื่อมั่นของแบบทดสอบถูกผิดหลายตัวเลือกที่ใช้วิธีตรวจให้คะแนนต่างกันพบว่า เมื่อใช้วิธีตรวจให้คะแนนต่างกัน แบบทดสอบถูกผิดหลายตัวเลือกจะมีความเชื่อมั่นแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยเมื่อตรวจให้คะแนนด้วยวิธี Pass-3 จะทำให้แบบทดสอบมีความเชื่อมั่นต่ำกว่าวิธี Count-2, วิธี Count-3 และวิธี MTF สำหรับคู่อื่นๆ ไม่แตกต่างกัน ส่วนใหญ่จึงไม่สอดคล้องกับสมมติฐานของการวิจัยที่ตั้งไว้ ทั้งนี้ผู้วิจัยคาดว่าสาเหตุสำคัญที่

ทำให้การตรวจให้คะแนนด้วยวิธี Pass-3 ส่งผลให้แบบทดสอบถูกผิดหลายตัวเลือกมีความเชื่อมั่นต่ำกว่าวิธีอื่นๆ (ยกเว้นเมื่อเปรียบเทียบกับวิธี MR) คงเป็นเพราะวิธี Pass-3 ยึดหลักการตรวจให้คะแนนแบบ 0-1 คือจะให้ 1 คะแนน เมื่อมีความรู้มากพอในระดับที่ยอมรับได้ ซึ่งในที่นี้คือตอบได้ถูกต้องตรงตามเฉลย จำนวน 3 ตัวเลือกขึ้นไป แต่ถ้าตอบได้ถูกต้องน้อยกว่า 3 ตัวเลือก จะไม่ให้คะแนน (ได้ 0 คะแนน) ดังนั้นผู้สอบที่ตอบถูกต้องตรงตามเฉลย 2 ตัวเลือก, 1 ตัวเลือก และตอบไม่ถูกต้องเลย ต่างก็ได้รับการตัดสินว่าไม่มีความรู้เหมือนกัน (ได้ 0 คะแนน) ในทำนองกลับกัน ผู้สอบที่ตอบได้ถูกต้องตรงตามเฉลย จำนวน 3 ตัวเลือก และ 4 ตัวเลือก ก็ได้รับการตัดสินว่าเป็นผู้มีระดับความรู้เท่ากัน (ได้ 1 คะแนน) ดังนั้นช่วงต่อเนืองระหว่างการได้รับการตัดสินว่าเป็นผู้มี ความรู้กับผู้ไม่มีความรู้ ก็คือการตอบได้ถูกต้อง จำนวน

2 และ 3 ตัวเลือก นั่นเอง จึงนับว่าเป็นจุดแบ่งที่สำคัญมาก เพราะหากข้อสอบแต่ละข้อมีความยากง่ายแตกต่างกันไป โอกาสที่ความสอดคล้องของคะแนนที่ได้จากการตอบข้อสอบแต่ละข้อก็น่าจะมีค่าต่ำด้วย โดยเฉพาะกับผู้สอบที่มีความสามารถปานกลาง เพราะเมื่อตรวจให้คะแนนแบบ 0-1 ด้วยวิธี Pass-3 นี้ หากผู้สอบที่มีความสามารถปานกลาง มีความคลาดเคลื่อนในการตอบแม้เพียงเล็กน้อย (ตอบได้ถูกต้องระหว่าง 2 กับ 3 ตัวเลือก) ก็จะได้รับผลการตัดสินระดับความรู้เปลี่ยนแปลงไปเป็นตรงกันข้ามทันที (ระหว่าง “0” กับ “1” คะแนน หรือระหว่าง “ไม่รอบรู้” กับ “รอบรู้”) อันส่งผลให้แบบทดสอบฉบับนั้นมีความเชื่อมั่นต่ำไปด้วย

สำหรับการตรวจให้คะแนนวิธี MR ซึ่งให้คะแนนแบบ 0-1 ก็พบแนวโน้มของข้อสรุปในทำนองเดียวกัน

ขณะที่การตรวจให้คะแนนด้วยวิธีอื่นนั้น ยึดหลักการให้คะแนนสำหรับความรู้บางส่วน โดยจะมีการให้คะแนนมากน้อยลดหลั่นกันไปตามจำนวนของตัวเลือกที่ผู้สอบคนนั้นตอบได้ถูกต้อง ซึ่งสิ่งเหล่านี้จะเป็นสาเหตุสำคัญที่ทำให้แบบทดสอบนั้นมีความสอดคล้องภายในสูงขึ้นไป เพราะความคลาดเคลื่อนในการตอบแต่ละครั้งคงจะส่งผลกระทบต่อคะแนนสอบในข้อนั้นน้อยกว่า การตรวจให้คะแนนแบบ 0-1 ดังกล่าวแล้ว ซึ่งแนวโน้มดังกล่าวนี้เป็นไปในทิศทางเดียวกับข้อมูลที่ปรากฏในงานวิจัยของ วัชรภรณ์ จิตรมาศ (2538) และ สมพงษ์ อิศวศุภฤกษ์ (2542)

อย่างไรก็ตามที่กล่าวมาทั้งหมดนั้นมุ่งเน้นไปที่ความสอดคล้องภายใน (Internal Consistency) ของแบบทดสอบเป็นสำคัญ ซึ่งถือเป็นเพียงปัจจัยหนึ่งเท่านั้นที่ส่งผลต่อความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ตามวิธีการของ Livingston ซึ่งเป็นความเชื่อมั่นในรูปของความไวต่อระดับของการรอบรู้หรือไม่รอบรู้โดยใช้ความต่อเนื่องของคะแนน เพราะอีกปัจจัยหนึ่งที่สำคัญและส่งผลอย่างมากต่อความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ก็คือระยะห่างระหว่างคะแนนจุดตัดกับคะแนนเฉลี่ยของแบบทดสอบนั้น โดยมีการศึกษาพบว่าเมื่อคะแนนจุดตัดมีค่าห่างจากคะแนนเฉลี่ยมากขึ้น ค่าความเชื่อมั่นของแบบทดสอบฉบับนั้นจะสูงขึ้นตามไปด้วย แต่จะมีค่าลด

ต่ำลงเมื่อคะแนนจุดตัดมีค่าใกล้เคียงกับคะแนนเฉลี่ย และจะมีค่าต่ำสุดเมื่อคะแนนจุดตัดและคะแนนเฉลี่ยมีค่าเท่ากัน (Berk, 1980) ซึ่งแสดงว่าความคลาดเคลื่อนในการจำแนกผู้สอบตามความรอบรู้และไม่รอบรู้จะสูงขึ้น ถ้าผู้สอบส่วนใหญ่สอบได้คะแนนใกล้เคียงกับคะแนนจุดตัด เพราะหากความคลาดเคลื่อนในการตอบเกิดขึ้นแม้เพียงเล็กน้อย ก็อาจจะทำให้ผลการจำแนกผู้สอบเป็นผู้รอบรู้หรือไม่รอบรู้เปลี่ยนแปลงไปได้ ซึ่งเมื่อพิจารณาค่าสถิติพื้นฐานของคะแนนผลการสอบ จะเห็นได้ว่าการตรวจให้คะแนนวิธี Pass-3 นั้น แม้ผู้สอบส่วนใหญ่จะได้คะแนนค่อนข้างสูง แต่คะแนนจุดตัดที่ใช้ในการตัดสินความรอบรู้ก็มีค่าค่อนข้างสูงด้วย โดยมีค่าใกล้เคียงกันอย่างมาก (ดูรายละเอียดในตารางที่ 1) ความเชื่อมั่นของแบบทดสอบในรูปของความไวต่อระดับของการรอบรู้หรือไม่รอบรู้ตามวิธีของ Livingston จึงมีค่าต่ำกว่าวิธีอื่นๆ

2. การเปรียบเทียบสัดส่วนของผู้สอบที่ได้รับการจำแนกเป็นผู้รอบรู้ของแบบทดสอบถูกผิดหลายตัวเลือกที่ใช้วิธีตรวจให้คะแนนต่างกันพบว่าสัดส่วนของผู้สอบที่ได้รับการจำแนกเป็นผู้รอบรู้ โดยแบบทดสอบถูกผิดหลายตัวเลือกที่ใช้วิธีตรวจให้คะแนนต่างกัน มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยวิธี Pass-3 จะทำให้มีสัดส่วนของผู้สอบที่ได้รับการจำแนกเป็นผู้รอบรู้ต่ำกว่าวิธี Count-2 สำหรับคู่อื่นๆ ไม่แตกต่างกัน ทั้งนี้คงเป็นผลสืบเนื่องมาจากสาเหตุต่างๆ ตามที่ได้กล่าวถึงแล้วในข้อ 1 กล่าวคือ เมื่อตรวจให้คะแนนด้วยวิธี Pass-3 คะแนนจุดตัดของแบบทดสอบจะมีค่าใกล้เคียงกับคะแนนเฉลี่ย โดยมีค่าต่ำกว่าคะแนนเฉลี่ยเล็กน้อย จึงส่งผลให้จำนวนผู้สอบที่ได้รับการจำแนกเป็นผู้รอบรู้ มีน้อย ขณะที่เมื่อตรวจให้คะแนนด้วยวิธีอื่นๆ คะแนนจุดตัดจะมีค่าต่ำกว่าคะแนนเฉลี่ยมากยิ่งขึ้น จำนวนผู้สอบที่ได้รับการจำแนกเป็นผู้รอบรู้ จึงมีเพิ่มขึ้น

3. การศึกษาความสอดคล้องในการจำแนกผู้สอบตามความรอบรู้และไม่รอบรู้ของแบบทดสอบถูกผิดหลายตัวเลือกที่ใช้วิธีตรวจให้คะแนนต่างกันพบว่า สัมประสิทธิ์สหสัมพันธ์แบบพี (Ø) ของผลการจำแนกผู้สอบใน 2 ครั้งใดๆ ที่ใช้วิธีตรวจให้คะแนนต่างกัน มีค่าเป็นบวก และค่อนข้างสูง โดยมีนัยสำคัญทางสถิติที่ระดับ .05 ทุกค่า ซึ่งแสดงว่าแบบทดสอบถูกผิดหลาย

ตัวเลือกที่ใช้วิธีตรวจให้คะแนนต่างกัน สามารถจำแนกผู้สอบเป็นผู้รอบรู้หรือไม่รอบรู้ได้อย่างสอดคล้องกันเป็นส่วนใหญ่ ซึ่งไม่สอดคล้องกับสมมติฐานของการวิจัยที่ตั้งไว้ ทั้งนี้อาจเป็นเพราะว่าแม้จะตรวจให้คะแนนด้วยวิธีต่างกัน แต่คะแนนรวมที่ผู้สอบแต่ละคนได้รับจากการตรวจแต่ละวิธี ยังคงมีความสอดคล้องหรือเป็นไปในทิศทางเดียวกัน กล่าวคือ ผู้ที่ได้คะแนนรวมสูงก็จะได้คะแนนรวมสูงในทุกวิธี และผู้ที่ได้คะแนนรวมต่ำก็จะต่ำเหมือนกันในทุกวิธี ซึ่งในกรณีนี้ผู้วิจัยได้ทดลองคำนวณสัมประสิทธิ์สหสัมพันธ์อย่างง่ายของคะแนนรวมที่ได้จากการตรวจให้คะแนนด้วย 2 วิธีที่ต่างกัน โดยใช้สูตรคำนวณสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (Pearson Product Moment Correlation Coefficient) พบว่ามีค่าตั้งแต่ .790 ถึง .991 แสดงว่ามีความสัมพันธ์กันสูงมาก ซึ่งแม้ว่าลักษณะความสัมพันธ์ดังกล่าวจะไม่มีอิทธิพลมากพอที่จะทำให้จำนวนผู้สอบที่ได้รับการจำแนกเป็นผู้รอบรู้ในแต่ละวิธีมีจำนวนใกล้เคียงกันก็ตาม (ตามผลการวิจัยที่ปรากฏในข้อ 2) แต่เมื่อเทียบกับจำนวนผู้สอบทั้งหมด ก็ยังพบว่ามีความสอดคล้องกันในการจำแนกผู้สอบโดยส่วนใหญ่

#### ข้อเสนอแนะ

##### 1. ข้อเสนอแนะในการนำผลการวิจัยไปใช้

เนื่องจากผลการวิจัยพบว่าแบบทดสอบถูกผิดหลายตัวเลือกที่ใช้วิธีตรวจให้คะแนนต่างกันมีความเชื่อมั่นแตกต่างกัน ผู้ใช้แบบทดสอบจึงต้องพิจารณาเพิ่มเติมว่าหากคำนึงถึงความเชื่อมั่นในแง่ของความไวต่อระดับของการรอบรู้หรือไม่รอบรู้ พบแนวโน้มว่าวิธีที่มีการตรวจให้คะแนนสำหรับความรู้อ่างส่วน ในลักษณะที่สลับกันไปตามจำนวนตัวเลือกที่ตอบได้ถูกต้อง ซึ่งในที่นี้หมายถึง วิธี Count-2, วิธี Count-3 และวิธี MTF เหมาะสมกว่าวิธีที่ให้คะแนนแบบ 0-1 (โดยเฉพาะวิธี Pass-3) อย่างไรก็ตามวิธีตรวจให้คะแนนแต่ละวิธีต่างก็มีแนวคิดพื้นฐานและความสะดวกในการนำไปใช้ที่ต่างกัน ผู้ใช้จึงควรพิจารณาความเหมาะสมหลาย ๆ ด้านประกอบกัน แต่ในเบื้องต้นนี้ ผู้วิจัยมีความเห็นว่าวิธี MTF น่าจะเป็นวิธีหนึ่งที่เหมาะสม เพราะนอกจากจะแปลความหมายคะแนนในลักษณะของการเปรียบเทียบความ

สามารถระหว่างบุคคลแล้ว ยังสามารถแปลความหมายของคะแนนดิบ (Raw Score) เพิ่มเติมอย่างง่ายได้อีกว่า ผู้สอบตอบข้อสอบผิดพลาดก็ตัวเลือกในแต่ละข้อหรือทั้งฉบับ จึงน่าจะสะดวกกว่าวิธีอื่น ๆ ในการคำนวณและแปลความหมายของคะแนนที่ได้

นอกจากนี้แม้โดยทางสถิติจะยอมรับว่าวิธีตรวจให้คะแนนที่ต่างกันเหล่านี้จะสามารถจำแนกผู้สอบเป็นผู้รอบรู้หรือไม่รอบรู้ได้อย่างสอดคล้องกันโดยส่วนใหญ่อีกตาม แต่เมื่อพิจารณาร่วมกับผลการวิจัยในข้อ 1 และ 2 จะเห็นได้ว่าวิธีตรวจให้คะแนนบางวิธี เช่น วิธี Pass-3 อาจจะต้องได้รับการพิจารณาอย่างรอบคอบในการนำไปปรับใช้ เพราะมีผู้ที่สอบไม่ผ่าน (ไม่รอบรู้) มากกว่าวิธีอื่นอย่างมาก ซึ่งในขั้นนี้อาจจะยังไม่ถือว่าเป็นสิ่งดีหรือไม่อย่างไร เพราะการวิจัยครั้งนี้ไม่สามารถกำหนดกลุ่มที่รู้ชัด (Known Group) หรือระบุสถานภาพจริง (True Status) ได้ว่าใครเป็นผู้รอบรู้-ไม่รอบรู้ จึงไม่สามารถยืนยันได้ว่าผลการจำแนกผู้สอบเป็นผู้รอบรู้และไม่รอบรู้ที่ได้จากการตรวจให้คะแนนโดยวิธีใดที่มีความถูกต้องตรงตามสภาพจริง ผลการวิจัยนี้จึงทำให้ทราบเพียงว่าการตรวจให้คะแนนโดยวิธีที่ต่างกันเหล่านี้ให้ผลที่แตกต่างกันอยู่บ้าง ซึ่งประเด็นดังกล่าวนี้ก็เป็นที่ประเด็นหนึ่งที่จะต้องทำการศึกษาวิจัยเพื่อหาข้อสรุปที่ชัดเจนยิ่งขึ้นต่อไป

##### 2. ข้อเสนอแนะในการวิจัยครั้งต่อไป

2.1 ควรมีการศึกษาเกี่ยวกับความตรงของแบบทดสอบ (Test Validity) โดยใช้ตัวบ่งชี้สถานภาพจริง (True Status) ที่ได้มาอย่างหลากหลายและน่าเชื่อถือเพื่อตรวจสอบถึงความตรงเชิงสภาพ (Concurrent Validity) ของแบบทดสอบถูกผิดหลายตัวเลือกที่ใช้วิธีตรวจให้คะแนนต่าง ๆ กัน

2.2 ควรมีการศึกษาเกี่ยวกับความเชื่อมั่นของแบบทดสอบถูกผิดหลายตัวเลือกภายใต้เงื่อนไขของการประมาณค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ที่ใช้หลักการที่ต่างกันออกไป เช่น ระดับความสอดคล้องของการตัดสินความรอบรู้หรือไม่รอบรู้จากการสอบซ้ำสองครั้ง หรือหลักการอื่น ๆ ที่มีผู้เสนอแนะไว้จำนวนมาก รวมไปถึงการใช้วิธีกำหนดคะแนนจุดตัดวิธีอื่น ๆ ทั้งนี้เพื่อให้ได้แนวคิดที่หลากหลายในการพิจารณาตัดสินใจ



เลือกใช้วิธีกำหนดคะแนนจุดตัด วิธีตรวจให้คะแนน และการตัดสินใจเลือกใช้สูตรหรือวิธีการในการประมาณค่าความเชื่อมั่นในโอกาสต่อไป

2.3 เนื่องจากจำนวนตัวเลือกและความยากง่ายของข้อสอบ (และตัวเลือก) มีผลโดยตรงต่อโอกาสในการตอบข้อสอบได้ถูกต้องและคะแนนที่ผู้สอบคนนั้นๆ จะได้รับ จึงควรมีการศึกษาเพิ่มเติมในประเด็นเหล่านี้เพิ่มเติม

#### เอกสารอ้างอิง

- จักรกฤษณ์ สํารานใจ. (2531). ประสิทธิภาพเชิงสัมพัทธ์ของข้อสอบเลือกตอบชนิดตัดสินทุกตัวเลือก เทียบกับข้อสอบเลือกตอบชนิดแบบจับในแบบทดสอบวัดผลสัมฤทธิ์. วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต, จุฬาลงกรณ์มหาวิทยาลัย.
- บุญเชิด ภิญโญอนันตพงษ์. (2527). การทดสอบแบบอิงเกณฑ์: แนวคิดและวิธีการ. กรุงเทพฯ: โอเดียนสโตร์.
- พรทิพย์ เพชรกาศ. (2543). การเปรียบเทียบค่าฟังก์ชันสารสนเทศของแบบทดสอบเลือกตอบแบบถูกผิดที่ตรวจให้คะแนนต่างวิธี. วิทยานิพนธ์ศึกษาศาสตร์มหาบัณฑิต, มหาวิทยาลัยสงขลานครินทร์.
- พรรณเพ็ญ เมืองสง. (2543). การเปรียบเทียบค่าฟังก์ชันสารสนเทศของแบบทดสอบถูกผิดแบบหลายตัวเลือกที่มีจำนวนข้อ, ตัวเลือก และวิธีการตรวจให้คะแนนต่างกัน. วิทยานิพนธ์ศึกษาศาสตร์มหาบัณฑิต, มหาวิทยาลัยสงขลานครินทร์.
- ล้วน สายยศ และอังคณา สายยศ. (2539). เทคนิคการวัดผลการเรียนรู้. กรุงเทพฯ: ชมรมเด็ก.
- วัชรภรณ์ จิตรมาศ. (2538). การเปรียบเทียบคุณภาพแบบทดสอบเลือกตอบชนิดถูกผิดหลายตัวเลือกที่มีจำนวนตัวเลือกและวิธีการตรวจให้คะแนนต่างกัน. ปริญญาการศึกษามหาบัณฑิต, มหาวิทยาลัยศรีนครินทร-วิโรฒ ประสานมิตร.

- สงบ ลักษณะ. (2523). การตรวจสอบความเที่ยงตรงของข้อสอบอิงเกณฑ์. วารสารการวัดผลการศึกษา, 1(3), 36-44.
- สงบ ลักษณะ. (2523). ปัญหามาตรฐานและเกณฑ์กับการวัดผลสัมฤทธิ์. วารสารการวิจัยทางการศึกษา, 10(1), 16-23.
- สมพงษ์ อัครศุภฤกษ์. (2542). การเปรียบเทียบความเชื่อมั่นของแบบทดสอบ ที่มีวิธีการตรวจและความยาวของแบบทดสอบต่างกัน. ปริญญาการศึกษามหาบัณฑิต, มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร.
- สุรัช มีชาญ. (2537). การให้คะแนนสำหรับความรู้บางส่วน (Partial Knowledge) โดยใช้ข้อสอบเลือกตอบถูกผิดทุกตัวเลือก (Multiple True-False Item) และข้อสอบเลือกตอบถูกผิดทุกขั้นตอน (Sequential True-False Item). ที่ระลึกเนื่องในโอกาสเกษียณอายุราชการ ดร.วัน สังข์สะอาด (น.31-40). กรุงเทพฯ: มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร.
- Albanese, M.A. & Sabers, D.L. (1988). Multiple True-False Item: A Study of Inter-Item Correlations, Scoring Alternatives, and Reliability Estimation. *Journal of Educational Measurement*, 25, 111-123.
- Berk, R.A. (1980). *Criterion-Referenced Testing: State of the Art*. Baltimore: John Hopkins University.
- Berk, R.A. (1976). Determination of Optimal Cutting scores in Criterion-Referenced Measurement. *Journal of Experimental Education*, 45, 4-9.
- Fu-Ju Tsai. & Hoi K. Suen. (1993). A Brief Report on a Comparison of Six Scoring Methods for Multiple True-False Items. *Educational and Psychological Measurement*, 53, 399-405.