การเปรียบเทียบคุณภาพแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ ตามทฤษฎีการตอบสนองข้อสอบ

ณพานันท์ ยมจินดา

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิจัย วัดผลและสถิติการศึกษา
คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา
มิถุนายน 2563
ลิขสิทธิ์เป็นของมหาวิทยาลัยบูรพา

คณะกรรมการควบคุมวิทยานิพนธ์และคณะกรรมการสอบวิทยานิพนธ์ ได้พิจารณาวิทยานิพนธ์ของ ณพานันท์ ยมจินดา ฉบับนี้แล้ว เห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิจัย วัดผลและสถิติการศึกษา ของมหาวิทยาลัยบูรพาได้

คณะกรรมการควบคุมวิทยานิพนธ์

.....อาจารย์ที่ปรึกษาหลัก
(ผู้ช่วยศาสตราจารย์ ดร.สุรีพร อนุศาสนนันท์)

..... อาจารย์ที่ปรึกษาร่วม
(ผู้ช่วยศาสตราจารย์ ดร.สมพงษ์ ปั้นหุ่น)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธาน
(ผู้ช่วยศาสตราจารย์ ดร.สุพัฒนา หอมบุปผา)

.....กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.สุรีพร อนุศาสนนันท์)

..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.สมพงษ์ ปั้นหุ่น)

.....กรรมการ
(รองศาสตราจารย์ ดร.เวชฤทธิ์ อังกนะภัทรขจร)

คณะศึกษาศาสตร์อนุมัติให้รับวิทยานิพนธ์ฉบับนี้ เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิจัย วัดผลและสถิติการศึกษา ของมหาวิทยาลัยบูรพา

..... คณบดีคณะศึกษาศาสตร์
(รองศาสตราจารย์ ดร.สฎายุ ธีระวณิชตระกูล)

วันที่ 25 เดือน มิถุนายน พ.ศ. 2563

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้ สำเร็จลุล่วงไปได้ด้วยดี เนื่องจากได้รับความกรุณาเป็นอย่างสูงจาก ผู้ช่วยศาสตราจารย์ ดร.สุรีพร อนุศาสนนันท์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ได้ให้คำปรึกษา แนะนำแนวทาง รวมถึง ผู้ช่วยศาสตราจารย์ ดร.สมพงษ์ ปั้นหุ่น อาจารย์ที่ปรึกษาร่วม ที่ได้กรุณาช่วยเหลือ ให้คำแนะนำปรึกษา ทั้งยังแก้ไขข้อบกพร่องต่าง ๆ ด้วยความเอาใจใส่อย่างยิ่งด้วยดีเสมอมา

ขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.สุพัฒนา หอมบุปผา ที่ได้กรุณาเป็นประธานในการสอบปากเปล่าวิทยานิพนธ์ รองศาสตราจารย์ ดร.เวชฤทธิ์ อังกนะภัทรขจร ที่ได้กรุณาเป็น กรรมการในการสอบปากเปล่าวิทยานิพนธ์ และที่สำคัญผู้วิจัยขอกราบขอพระคุณคณาจารย์ในภาควิชา วิจัย วัดผลและสถิติการศึกษา ทุกท่านที่ได้ให้คำปรึกษาและได้อบรมสั่งสอนให้ความรู้ ซึ่งเป็นประโยชน์ต่อวิทยานิพนธ์ฉบับนี้จนสำเร็จลุล่วง

ขอขอบพระคุณ รองศาสตราจารย์ ดร.ไพรัตน์ วงษ์นาม ที่ให้คำแนะนำและให้คำปรึกษาเกี่ยวกับทฤษฎีการตอบสนองข้อสอง (IRT) และ ดร.อาวีพร ปานทอง ที่ได้กรุณาให้ความรู้ คำแนะนำและให้คำปรึกษาเกี่ยวการวิเคราะห์ข้อมูลด้วยโปรแกรม IRTPRO

ขอขอบพระคุณ ดร.ภัทรภร ชัยประเสริฐ อาจารย์ธนายุต จันทราเขต อาจารย์พนา ชุมเกษียร อาจารย์ปวีณา พันทอง และอาจารย์อรรถการณ์ ทองแดงเจือ ที่กรุณาตรวจสอบคุณภาพเครื่องมือที่ใช้ในการวิจัย รวมทั้งให้ข้อเสนอแนะในการปรับปรุงเครื่องมือที่ใช้ในการเก็บข้อมูล

ขอขอบพระคุณ ท่านผู้บริหาร ครูและนักเรียนกลุ่มตัวอย่างของโรงเรียนต่าง ๆ ในเขตพื้นที่การศึกษามัธยมศึกษา เขต 18 ทั้งในจังหวัดชลบุรี และระยอง ที่ให้ความร่วมมือในการเก็บข้อมูลในการวิจัยเป็นอย่างยิ่ง

ขอขอบพระคุณ พี่ ๆ เพื่อน ๆ น้อง ๆ ในสาขาวิจัย วัดผลและสถิติการศึกษารวมทั้งที่ทำงาน ทุกคนที่คอยเป็นกำลังใจและช่วยเหลือผู้วิจัยเสมอมา

ท้ายนี้ ขอขอบพระคุณคุณพ่อ นาวาเอก (พิเศษ) ชัยวัฒน์ ยมจินดา และคุณแม่ ดวงกมล ยมจินดา บิดา-มารดา ผู้เป็นที่รัก และให้การสนับสนุนด้านการศึกษารวมถึงครอบครัวที่ให้กำลังใจ ห่วงใย เป็นแรงบันดาลใจ ขอบคุณนายไกรสิทธิ์ พิริยะอรุณโรจน์ ที่คอยช่วยเหลือ และให้กำลังใจ เพื่อให้ผู้วิจัยทำวิทยานิพนธ์เล่มนี้สำเร็จลุล่วงไปได้ด้วยดี คุณค่าและประโยชน์จากวิทยานิพนธ์ ผู้วิจัยขอมอบแด่ทุกท่านที่ทำให้ผู้วิจัยประสบความสำเร็จในวันนี้

ณพานันท์ ยมจินดา

58910153: สาขาวิชา: วิจัย วัดผลและสถิติการศึกษา: วท.ม. (วิจัย วัดผลและสถิติการศึกษา)
คำสำคัญ: โมเดลโลจิสติก/ จีพีซีเอ็ม/ ทวิภาค/ พหุวิภาค/ ให้คะแนนแบบพิจารณาความมั่นใจ

ณพานันท์ ยมจินดา: การเปรียบเทียบคุณภาพแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ ตามทฤษฎีการตอบสนองข้อสอบ (COMPARISON THE QUALITY OF MULTIPLE CHOICE TEST BETWEEN CONVENTIONAL METHOD AND CONFIDENTIAL MAKING METHOD USING ITEM RESPONSE THEORY) คณะกรรมการควบคุมวิทยานิพนธ์: สุรีพร อนุศาสนนันท์, ค.ด., สมพงษ์ ปั้นหุ่น, ค.ด. 221 หน้า. พ.ศ. 2563.

งานวิจัยครั้งนี้ผู้วิจัยมีวัตถุประสงค์ 1) เพื่อหาคุณภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ ในด้านฟังก์ชันสารสนเทศ ความยากและอำนาจจำแนก ความตรงตามสภาพ 2. เพื่อเปรียบเทียบคุณภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ ในด้านดังต่อไปนี้ ฟังก์ชันสารสนเทศ และความตรงตามสภาพ กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 4 ภาคเรียนที่ 2 ปีการศึกษา 2560 ของโรงเรียนในสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 18 จำนวน 658 คน เครื่องมือที่ใช้ในการวิจัยคือ แบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชา เคมี ชั้นมัธยมศึกษาปีที่ 4 เรื่อง ปริมาณสารสัมพันธ์ จำนวน 30 ข้อ นำมาวิเคราะห์ค่าความยาก, ค่าอำนาจจำแนก และฟังก์ชันสารสนเทศของแบบทดสอบเป็นรายข้อ โดยวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (Item response theory) แบบทวิภาค ที่ใช้โมเดลโลจิสติก (Logistic) 2 พารามิเตอร์ และโมเดลการให้คะแนนความรู้บางส่วน (Genralized partial credit model: GPCM) เปรียบเทียบค่าฟังก์ชันสารสนเทศจากค่าประสิทธิภาพสัมพัทธ์ RE (θ) ระหว่างแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ และค่าความตรงตามสภาพ

ผลการวิจัยพบว่า

1. ค่าความยากและค่าอำนาจจำแนกของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาเป็นรายข้อ พบว่า มีค่าความยากระหว่าง -1.86-4.22 ค่าอำนาจจำแนกระหว่าง 0.13-1.87 และแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ พบว่า มีค่าความยากระหว่าง -0.18-4.6 และมีค่าอำนาจจำแนกอยู่ระหว่าง 0.14-0.76

2. ค่าประสิทธิภาพสัมพัทธ์ RE (θ) พบว่า ระดับความสามารถ (θ = -2.8 ถึง 0.0) แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดามีค่าประสิทธิภาพสัมพัทธ์สูงกว่าแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ และระดับความสามารถ (θ = 0.4 ถึง 2.8) แบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจมีค่าประสิทธิภาพสัมพัทธ์สูงกว่าแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา

3. ค่าความตรงตามสภาพ พบว่า แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา และแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ กับ ระดับผลการเรียน วิชาเคมี มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ค่าความตรงตามสภาพขึ้นอยู่กับวิธีการให้คะแนนในแต่ละวิธี และแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาความตรงตามสภาพสูงกว่า แบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ

58910153: MAJOR: EDUCATIONAL RESEARCH, MEASUREMENT AND STATISTICS; M.Sc. (EDUCATIONAL RESEARCH, RESEARCH, MEASUREMENT, AND STATISTICS)

KEYWORDS: LOGISTIC MODELS/ GPCM/ DICHOTOMOUS/ POLYTOMOUS/ CONFIDENCE MARKING

NAPANUN YOMJINDA: COMPARISON THE QUALITY OF MULTIPLE CHOICE TEST BETWEEN CONVENTIONAL METHOD AND CONFIDENTIAL MAKING METHOD USING ITEM RESPONSE THEORY. ADVISORY COMMITTEE: SUREEPORN ANUSASANANT, Ph.D., SOMPONG PANHOON, Ph.D. 221 P. 2020.

The objectives of study were; 1) to find out the qualities of multiple choice test between conventional method and confidential making method in terms of test information, difficulty and discrimination index, concurrent validity, and 2) to compare the qualities of multiple choice test between conventional method and confidential making method of both test information and concurrent validity. The sample group was 658 grade 10 students in academic year 2017 of the schools under the Office of Secondary Education Service Area 18. The research instrument include was chemistry achievement test on a Stoichiometry, totaling 30 items. The data were analyzed by using IRTPRO in term of difficulty, discrimination index and item information through item response theory method. The model of dichotomously scoring used logistic model 2 parameters for and GPCM for polytomous scoring, the comparison of relative efficiency on test information and concurrent validity.

The research findings were as follows;

1. The difficulty and discrimination conventional method using dichotomous logistic 2 parameters are ranged from -1.86-4.22,discrimination are ranged from 0.13-1.87 and confidential making method difficulty are ranged from -0.18-4.6 and the discrimination are ranged from 0.14-0.76.

2. The relative efficiency RE (θ) at ability (θ = -2.8 to 0.0) of conventional method was higher than the confidential making method. The relative efficiency RE (θ) of confidential making method ability (θ = 0.4 to 2.8) was higher than conventional method.

3. The relationship between grade and test score of chemistry achievement test of both conventional method and confidential making method were significantly related at .01 and the concurrent validity of conventional method is higher than confidential making method.

สารบัญ

หน้า

บทคัดย่อภาษาไทย..... ง

บทคัดย่อภาษาอังกฤษ..... จ

สารบัญ..... ฉ

สารบัญตาราง..... ซ

สารบัญภาพ..... ฎ

บทที่

1 บทนำ..... 1

ความเป็นมาและความสำคัญของปัญหา..... 1

คำถามในการวิจัย..... 5

วัตถุประสงค์..... 5

สมมติฐานการวิจัย..... 5

กรอบแนวคิดในการวิจัย..... 6

ขอบเขตของการวิจัย..... 7

นิยามศัพท์เฉพาะ..... 7

ประโยชน์ที่ได้รับจากการวิจัย..... 10

2 เอกสารและงานวิจัยที่เกี่ยวข้อง..... 11

ตอนที่ 1 การศึกษาเกี่ยวกับแบบทดสอบเลือกตอบและวิธีการตรวจให้คะแนน..... 12

ตอนที่ 2 ทฤษฎีการตอบสนองข้อสอบ..... 29

ตอนที่ 3 การตรวจสอบคุณภาพของแบบทดสอบ..... 44

ตอนที่ 4 การสร้างแบบทดสอบโดยใช้การเรียนรู้ของบลูม..... 52

ตอนที่ 5 คะแนนจุดตัด..... 54

3 วิธีดำเนินการวิจัย..... 76

ตอนที่ 1 ประชากรและกลุ่มตัวอย่าง..... 76

ตอนที่ 2 การสร้างเครื่องมือที่ใช้ในการวิจัยและขั้นตอนการสร้างเครื่องมือ..... 78

ตอนที่ 3 การเก็บรวบรวมข้อมูล..... 93

ตอนที่ 4 การวิเคราะห์ข้อมูล..... 94

ตอนที่ 5 สถิติที่ใช้ในการวิเคราะห์ข้อมูล..... 95

สารบัญ (ต่อ)

บทที่	หน้า
4 ผลการวิเคราะห์ข้อมูล	100
สัญลักษณ์และอักษรย่อที่ใช้ในการนำเสนอผลการวิจัย	100
ตอนที่ 1 ผลการวิเคราะห์คุณภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจในการทดลองใช้ครั้งที่ 1 และครั้งที่ 2 ด้วยทฤษฎีการวัดแบบดั้งเดิม	101
ตอนที่ 2 ผลการวิเคราะห์คุณภาพของแบบทดสอบเลือกตอบที่ให้ คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจในการนำไปใช้จริง ด้วยทฤษฎีการตอบสนองข้อสอบ ในด้านฟังก์ชันสารสนเทศ ความยากและอำนาจจำแนก ความตรงตามสภาพ	115
ตอนที่ 3 ผลเปรียบเทียบประสิทธิภาพสัมพัทธ์และความตรงตามสภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ	129
5 สรุป อภิปรายผล และข้อเสนอแนะ	136
สรุปผลการวิจัย	137
อภิปรายผล	140
ข้อเสนอแนะ	146
บรรณานุกรม	148
ภาคผนวก	159
ภาคผนวก ก	160
ภาคผนวก ข	162
ภาคผนวก ค	179
ภาคผนวก ง	185
ภาคผนวก จ	194
ภาคผนวก ฉ	207
ประวัติย่อของผู้วิจัย	221

สารบัญตาราง

ตารางที่ หน้า

2-1 การเลือกตอบอย่างเสรี (Free choice test) โดยในแต่ละข้อผู้ตอบจะเลือกตอบกี่ตัวเลือกก็ได้และจะได้คะแนนตามเงื่อนไขถ้าข้อใดเว้นว่างไว้จะไม่ได้คะแนน 21

2-2 วิธีการตรวจให้คะแนนของกิบบอนส์และคณะ 26

2-3 การคำนวณค่าฟังก์ชันสารสนเทศ ค่าสารสนเทศสูงสุดของข้อสอบและตำแหน่ง θ ที่มีสารสนเทศของข้อสอบสูงสุด สำหรับโมเดลการตอบสนองข้อสอบโลจิสติก 1, 2, 3 พารามิเตอร์ 39

2-4 เกณฑ์การพิจารณาค่าอำนาจจำแนก 45

2-5 ตัวอย่างการคำนวณคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟ 58

2-6 การพิจารณาความสำเร็จที่คาดหวังในข้อสอบโดยวิธีของ อีเบล 60

2-7 ตัวอย่างแบบทดสอบฉบับหนึ่งมี 50 ข้อเมื่อให้ผู้เชี่ยวชาญ 5 คนพิจารณา แยกแยะลักษณะข้อสอบ 60

2-8 จำนวนคนที่ผ่านและไม่ผ่านเกณฑ์ที่กำหนดขึ้นในแต่ละจุดของคะแนนเกณฑ์เกณฑ์ภายนอก 63

2-9 คะแนนพยากรณ์ที่จะแบ่งการเรียนรู้ 64

3-1 รายชื่อโรงเรียนและจำนวนนักเรียนที่ใช้เป็นกลุ่มตัวอย่าง จำแนกตามขนาดโรงเรียนและจุดประสงค์การทดสอบ 77

3-2 โครงสร้างเนื้อหาแบบทดสอบประกอบด้วย รายละเอียดของเนื้อหา ผลการเรียนรู้และทักษะที่มุ่งวัด 83

3-3 การวิเคราะห์แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน วิชา เคมี เรื่อง ปริมาณสารสัมพันธ์ 85

3-4 การวิเคราะห์หน่วยการเรียนรู้ ผลการเรียนรู้ และจำนวนข้อสอบ ตามพฤติกรรมที่ต้องการวัด 87

3-5 เกณฑ์การพิจารณาค่าความยาก 96

3-6 เกณฑ์การพิจารณาค่าอำนาจจำแนก 97

4-1 ความสอดคล้องในแต่ละข้อของแบบทดสอบที่มีตัวเลือกตามวิธีการตรวจให้คะแนนทั้งสองวิธี 102

สารบัญตาราง (ต่อ)

ตารางที่ หน้า

4-2 ค่าความยาก (p) ค่าอำนาจจำแนก (r) ทดลองใช้ครั้งที่ 1 และครั้งที่ 2 ของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ..... 104

4-3 ปรับปรุงแบบทดสอบหลังจากการทดลองใช้ ครั้งที่ 1..... 108

4-4 ผลการพิจารณาคะแนนจุดตัดของแบบทดสอบตามวิธีการให้คะแนนของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ..... 113

4-5 คุณภาพแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจทั้งฉบับจากการทดลองใช้ครั้งที่ 1 และครั้งที่ 2 ทั้งฉบับ..... 115

4-6 ผลการวิเคราะห์ค่าสถิติพื้นฐานของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ..... 116

4-7 ผลการวิเคราะห์องค์ประกอบเพื่อตรวจสอบความเป็นมิติเดียว ค่าสถิติ Kaiser-Meyer-Olkin (KMO) และค่าสถิติ Bartlett's test of sphericity ของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ..... 117

4-8 จำนวนองค์ประกอบ ค่าไอเกน ค่าร้อยละความแปรปรวน และค่าร้อยละของความแปรปรวนสะสมในแต่ละองค์ประกอบของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ..... 119

4-9 ผลการวิเคราะห์ค่าฟังก์ชันสารสนเทศจำแนกรายข้อ จากแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ จำนวน 30 ข้อ..... 120

4-10 ผลการวิเคราะห์ค่าความยากและค่าอำนาจจำแนกของของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาโดยวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบแบบทวิภาค ใช้โมเดลโลจิสติก 2 พารามิเตอร์..... 124

4-11 ผลการวิเคราะห์ค่าความยากและค่าอำนาจจำแนกของของแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ โดยวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบที่ใช้โมเดลการให้คะแนนความรู้บางส่วน..... 126

4-12 ผลการวิเคราะห์ความยากและอำนาจจำแนกเฉลี่ยของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ..... 127

สารบัญตาราง (ต่อ)

ตารางที่ หน้า

4-13 ผลการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ (r) ระหว่างคะแนนจากแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาเคมี ภาคเรียนที่ 2 ชั้นมัธยมศึกษาปีที่ 4 ที่ได้จากวิธีการตอบ และการให้คะแนนในแต่ละวิธี กับระดับผลการเรียนวิชาเคมี ภาคเรียนที่ 1 ชั้นมัธยมศึกษาปีที่ 4..... 128

4-14 ผลการเปรียบเทียบฟังก์ชันสารสนเทศแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ จำแนกตามระดับความสามารถของผู้สอบ..... 129

4-15 ผลการเปรียบเทียบค่าต่ำสุด ค่าสูงสุด และค่าเฉลี่ยฟังก์ชันสารสนเทศแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ..... 130

4-16 ผลการเปรียบเทียบค่าความคลาดเคลื่อนมาตรฐาน SE (θ) ของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ..... 130

4-17 ผลการเปรียบเทียบค่าประสิทธิภาพสัมพัทธ์ RE (θ) ระหว่างค่าฟังก์ชันสารสนเทศระหว่างแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ..... 132

4-18 ผลการเปรียบเทียบค่าประสิทธิภาพสัมพัทธ์ RE (θ) ระหว่างค่าฟังก์ชันสารสนเทศระหว่างให้แบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจและแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา..... 133

4-19 การแปลงค่าสัมประสิทธิ์สหสัมพันธ์ (r) ระหว่างคะแนนจากแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาเคมี ภาคเรียนที่ 2 ชั้นมัธยมศึกษาปีที่ 4 ที่ได้จากวิธีการตอบ และการให้คะแนนในแต่ละวิธี กับระดับผลการเรียนวิชาเคมี ภาคเรียนที่ 1 ชั้นมัธยมศึกษาปีที่ 4 จำนวน 658 คน เป็นคะแนนฟิชเชอร์ซี (Z_r) เมื่อนำมาเปรียบเทียบความแตกต่างกันโดยใช้สถิติซี (Z) 135

สารบัญภาพ

ภาพที่ หน้า

1-1 กรอบแนวคิดในการวิจัย..... 6

2-1 โค้งลักษณะข้อสอบแบบ 1 พารามิเตอร์..... 33

2-2 โค้งลักษณะข้อสอบแบบ 2 พารามิเตอร์..... 35

2-3 โค้งลักษณะข้อสอบแบบ 3 พารามิเตอร์..... 36

2-4 โค้งสารสนเทศของข้อสอบ..... 41

2-5 บทบาทของการกำหนดคะแนนจุดตัด ในกระบวนการวัดและประเมินผลการเรียนรู้..... 55

3-1 ขั้นตอนในการสร้างแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชา เคมี..... 81

4-1 ค่าไอแกนของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา..... 118

4-2 ค่าไอแกนของแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ..... 118

4-3 เปรียบเทียบฟังก์ชันสารสนเทศแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและ
ให้คะแนนแบบพิจารณาความมั่นใจจำแนกตามระดับความสามารถของผู้สอบ..... 129

4-4 เปรียบเทียบค่าความคลาดเคลื่อนมาตรฐาน SE (θ) ของแบบทดสอบเลือกตอบที่ให้
คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ..... 131

4-5 ค่าประสิทธิ์ภาพสัมพัทธ์ระหว่างแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและ
ให้คะแนนแบบพิจารณาความมั่นใจ..... 134

บทที่ 1
บทนำ

ความเป็นมาและความสำคัญของปัญหา

การวัดผลการศึกษาจำเป็นต้องอาศัยเครื่องมือและเทคนิคต่าง ๆ ซึ่งได้แก่ แบบทดสอบ การสังเกต การจัดอันดับคุณภาพ การสัมภาษณ์ ฯลฯ เครื่องมือที่ถูกสร้างขึ้นมาสำหรับใช้วัดนั้น มีลักษณะที่แตกต่างกันออกไปตามวัตถุประสงค์ที่จะสร้าง เนื้อหาที่จะวัด ตลอดจนลักษณะของผู้ถูกวัด เครื่องมือที่ได้รับความนิยมมากที่สุดคือแบบทดสอบ ทั้งนี้อาจเนื่องมาจากกระบวนการต่าง ๆ และวิธีที่ใช้ครอบคลุมเนื้อหาที่จะวัด และวัตถุประสงค์ที่จะวัดโจทก์สามารถวัดได้ทุกพฤติกรรม แบบทดสอบที่ใช้วัดจึงจำเป็นต้องมีคุณภาพที่เชื่อถือได้ว่าคะแนนที่วัดได้นั้นเป็นคะแนนที่แท้จริงหรือมีความคาดเครื่องในการวัดน้อยที่สุด ซึ่งส่วนมากนิยมใช้แบบทดสอบปรนัยชนิดเลือกตอบเพราะสามารถเขียนคำถามได้ครอบคลุมเนื้อหามีความเป็นปรนัยในการตรวจให้คะแนนและมีประสิทธิภาพ (เอมอร จังศิริพรปกรณ์, 2544, หน้า 21) และในบรรดาแบบทดสอบปรนัยด้วยกันแล้ว ชนิดที่นิยมใช้กันอย่างแพร่หลาย ได้แก่ แบบสอบเลือกตอบ (Multiple choice)

แบบสอบเลือกตอบ (Multiple choice) เป็นที่นิยมใช้กันอย่างกว้างขวาง เพราะสามารถใช้วัดผลการเรียนรู้ ความเข้าใจ การนำไปใช้ และผลการเรียนรู้ขั้นสูงได้ สามารถสร้างให้วัดได้ครอบคลุมเนื้อเรื่องตามโครงสร้างอย่างมีประสิทธิภาพ และนำไปพัฒนาเป็นแบบสอบมาตรฐานได้ อีกทั้งมีความเป็นปรนัยในการตรวจให้คะแนน และสะดวกในการดำเนินการสอบ (ศิริชัย กาญจนวาสี, 2548) แต่แบบสอบแบบหลายตัวเลือกที่ให้นักเรียนเลือกตอบเฉพาะตัวเลือกถูกเพียงตัวเลือกเดียวและมีการให้คะแนนแบบ 2 ค่า (Dichotomously scored item) คือ ตอบถูกให้ 1 คะแนน ตอบผิดให้ 0 คะแนนนั้น นักวัดผลกลุ่มหนึ่งเกิดข้อสงสัยว่านักเรียนที่ตอบถูกนั้นเป็นผู้มีความรู้ในเรื่องนั้นจริงหรือเกิดจากการเดา และนักเรียนที่ตอบผิดนั้นเป็นเพราะไม่มีความรู้ในเรื่องนั้นเลย หรือเกิดจากปัญหาอื่น ๆ จึงได้พยายาม พัฒนาแบบทดสอบที่สามารถบ่งบอกความรู้จริงของผู้เรียนให้มากที่สุดเท่าที่จะทำได้ (อุทัยวรรณ สายพัฒนะ, 2547) การให้เลือกตัวเลือกเพียงตัวเดียวนั้น ยังไม่สามารถนำข้อมูลจากการตอบไปพิจารณาจำแนก ผู้มีความรู้สมบูรณ์ รู้บางส่วนและไม่รู้เลย ออกจากกันได้ เพราะเหตุที่ว่าผู้เลือกตัวเลือกถูกแล้วได้ 1 คะแนน อาจจะเป็นคนที่มีความรู้จริงในเรื่องนั้น รู้เพียงแค่บางส่วนแต่เผอิญเลือกตัวเลือกถูก หรือ ไม่รู้เลยแต่เดาถูกก็ได้ และในทำนองเดียวกัน ผู้ที่ได้คะแนน 0 คะแนนนั้นก็ไม่อาจจำแนกได้ว่าใครได้ 0 เพราะมีความรู้ผิด ๆ ไม่รู้เลยหรือรู้เพียงบางส่วน (Coombs, Milholland & Womer, 1956) เพื่อจะให้ได้ข้อมูลจาก

การตอบเพิ่มขึ้น เพียงพอที่จะจำแนกบุคคลตามความรู้ของเขา และลดการตอบถูกด้วยการเดา ในแบบทดสอบ แบบทดสอบเลือกตอบมีข้อดีคือสามารถเขียนคำถามได้รัดกุม ชัดเจน ครอบคลุมเนื้อหาสามารถวัดสมรรถภาพ ด้านต่าง ๆ ตามลำดับขั้นของความรู้ นอกจากนี้ยังประหยัดเวลาและแรงงานในการตรวจให้คะแนนที่ยุติธรรม แต่อย่างไรก็ตามแบบทดสอบเลือกตอบก็ยังมีปัญหาอยู่บ้าง คือ ไม่สามารถบอกได้ว่าผู้ตอบแบบทดสอบนั้นตอบด้วยความรู้ที่แท้จริงหรือมีความรู้เพียงบางส่วนหรือไม่มีความรู้เลยโจทก์จึงไม่เป็นการยุติธรรมที่บุคคลทั้งสามลักษณะดังกล่าวจะได้คะแนนเท่ากันหรือในกรณีที่ผู้ตอบตอบผิด อาจเนื่องมาจากผู้ตอบไม่มีความรู้เลย มีความรู้เพียงบางส่วนแต่ตัวลวงมีประสิทธิภาพ แต่ผู้ที่ตอบผิดทุกคนก็ได้ 0 คะแนน จากข้อคิดนี้ทำให้นักวัดผลหลายท่านพยายามหาวิธีที่จะวัดส่วนของความรู้ (Partial knowledge) ของผู้เรียนเพื่อจำแนกความรู้ของผู้เรียนได้ถูกต้องยิ่งขึ้น ซึ่งมีอยู่ด้วยกันหลายวิธี

Gibbons., Olkin and Sobel (1979, p. 261) เสนอวิธีการตอบและตรวจให้คะแนน โดยให้ผู้สอบเลือกตอบตัวเลือกที่ถูกได้ตั้งแต่ 1 ถึง K-1 ตัวตามความมั่นใจเมื่อ K คือจำนวนตัวเลือกทั้งหมดในแต่ละข้อกระทง ถ้าผู้สอบมั่นใจว่าตัวเลือกใดถูกก็ให้ตอบเพียงตัวเลือกเดียว ถ้าไม่มั่นใจหลายตัวตัวเลือกก็ให้ตอบทุกตัวเลือกที่ไม่มั่นใจ ส่วนวิธีการตรวจให้คะแนนยึดหลักว่าคะแนนคาดหวังที่ได้จากการเดาจะต้องเป็นศูนย์ เช่นเดียวกับ สำราญ มีแจ้ง (2527, หน้า 7) เสนอวิธีการตอบและตรวจ ให้คะแนนแบบสอบตัวเลือกที่มี 5 ตัวเลือก โดยให้ผู้สอบตอบได้ตั้งแต่ 1 ถึง 3 ตัวเลือกตามความมั่นใจ ถ้ามั่นใจในตัวเลือกใดก็ตอบตัวเลือกนั้น ถ้าไม่มั่นใจ 2 หรือ 3 ตัวเลือกก็ให้ตอบทั้ง 2 หรือ 3 ตัวเลือก การให้คะแนนถ้าตอบตัวเลือกเดียวแล้วถูกจะให้คะแนน 4 คะแนน ถ้าตอบ 2 ตัวเลือกแล้วถูกจะให้คะแนน 2 คะแนน ถ้าตอบ 3 ตัวเลือกแล้วถูกจะให้คะแนน 1 คะแนน ทุกกรณีถ้าผิดจะให้ 0 คะแนน (Gritten & Johnson 1941, pp. 432-430 cited in Echternacht, 1972, pp. 217-234) นำวิธีการให้คะแนนความมั่นใจมาใช้กับแบบสอบเลือกตอบและแบ่งระดับความมั่นใจเป็น 5 ระดับ เขาพบว่า ค่าความเที่ยงของแบบสอบเลือกตอบที่ให้คะแนนด้วยวิธีการให้คะแนนความมั่นใจมีค่าสูงขึ้นเช่นเดียวกับ (Pugh & Brunza, 1975, pp. 73-78) นำวิธีการให้คะแนนความมั่นใจมาใช้กับแบบสอบเลือกตอบเปรียบเทียบการให้คะแนนวิธีธรรมดา ผลการศึกษาพบว่าค่าความเที่ยงของแบบสอบเลือกตอบซึ่งคำนวณโดยวิธีวิเคราะห์ความแปรปรวนของแบบสอบที่มีวิธีต่อโดยบอกความมั่นใจสูงกว่าค่าความเที่ยงของแบบสอบที่มีวิธีตอบวิธีธรรมดาอย่างมีนัยสำคัญ (Abu-Sayf, 1975, pp. 160-162) ศึกษาวิธีการตอบโดยบอกระดับความมั่นใจกับแบบสอบเลือกตอบและแบ่งระดับความมั่นใจเป็น 3 ระดับ คือ มั่นใจ ไม่ค่อยมั่นใจ เดา เขาพบว่า จากคำตอบทั้งหมดที่ผู้ตอบตอบด้วยความมั่นใจมีเปอร์เซ็นต์ของผู้ตอบผิดน้อยที่สุดและผู้ที่ต่อด้วยการเดานั้นมีโอกาสถูกน้อยมาก จากการศึกษาคุณภาพของแบบทดสอบส่วนใหญ่จะวิเคราะห์ตามแนวทฤษฎีการวัด

แบบดั้งเดิม ซึ่งมีข้อจำกัดในเรื่องของแหล่งความคลาดเคลื่อนในการวัด และค่าพารามิเตอร์ของแบบทดสอบที่ได้ แปรเปลี่ยนไปตามกลุ่มตัวอย่างที่ใช้ ปัจจุบัน นักวัดผลการศึกษาได้พัฒนาทฤษฎีการตอบสนองข้อสอบ (Item response theory: IRT) เพื่อแก้ปัญหาการวิเคราะห์ข้อสอบตามแนวทฤษฎีการวัดแบบดั้งเดิม ซึ่งสามารถนำไปวิเคราะห์กับแบบทดสอบที่ตรวจให้คะแนนแบบทวิภาค และแบบทดสอบที่ตรวจให้คะแนนแบบพหุวิภาค โดยมีงานวิจัยที่ศึกษาในประเทศไทย (เพ็ญศรี สว่างเนตร, 2520, หน้า 41-42) นำวิธีการให้คะแนนความมั่นใจมาศึกษาเปรียบเทียบกับวิธีธรรมดา วิธีลดคะแนนข้อที่ตอบผิด วิธีเพิ่มคะแนนข้อที่เว้น ผลการศึกษาพบว่า ค่าความเที่ยงของแบบสอบที่ให้คะแนนวิธีบอกความมั่นใจมีค่าสูงกว่าวิธีการให้คะแนนวิธีอื่น ๆ อย่างมีนัยสำคัญ เช่นเดียวกับ เกริกชัย ฮวบเจริญ (2525, หน้า 36-37) นำวิธีการให้คะแนนความมั่นใจมาศึกษาเปรียบเทียบกับวิธีธรรมดาและวิธีตอบที่มีรูปแบบของกระดาษคำตอบชนิดพิเศษซึ่งเมื่อผู้ตอบตอบแล้วจะทราบทันทีว่าตนทำผิดหรือถูกในข้อกระทงนั้น ๆ ผลการศึกษาพบว่า วิธีการตอบโดยบอกระดับความมั่นใจมีค่าความเที่ยงสูงกว่าวิธีธรรมดาอย่างมีนัยสำคัญ สำหรับวิธีการให้คะแนนที่มีรูปแบบของกระดาษคำตอบชนิดพิเศษมีค่าความเที่ยงไม่แตกต่างกับวิธีการให้คะแนนโดยบอกระดับความมั่นใจอย่างมีนัยสำคัญ รุจิรา ขาวสะอาด (2543) ศึกษาผลการให้คะแนนน้ำหนักตามระดับความมั่นใจที่มีต่อค่าความเที่ยงและค่าความตรงของแบบสอบประเภทเลือกตอบที่ใช้การเขียนตัวเลือกต่างกัน ซึ่งพบว่า การเลือกใช้วิธีการให้น้ำหนักคะแนน ตามระดับความมั่นใจในการตอบควรกำหนดระดับความมั่นใจ 2 หรือ 3 ระดับ เพราะสะดวกต่อการคำนวณคะแนนและใช้เวลาในการสอบไม่มากเกินไป จากการศึกษางานวิจัยที่เกี่ยวข้องดังกล่าวข้างต้นมาจากแนวคิดเพื่อวัดส่วนของความรู้ของผู้สอบทั้งสิ้น นอกจากนี้ยังมีวิธีการวัดส่วนของความรู้ของผู้สอบอีกวิธีหนึ่ง นั่นคือ วิธีการกำหนดให้มีวิธีการตอบที่แตกต่างไปจากวิธีตอบแบบธรรมดา (Conventional) ที่เป็นอยู่ในปัจจุบัน กล่าวคือ นอกจากผู้เรียนจะต้องเลือกตัวเลือกที่คิดว่าเป็นคำตอบที่ถูกต้องแล้ว ผู้เรียนยังต้องตัดสินใจอีกว่าตัวเลือกที่เป็นคำตอบถูกนั้นผู้เรียนมีระดับความมั่นใจในการตอบ (Confidence weight scoring) มากน้อยเพียงใด กระดาษคำตอบนอกจากจะมีคำตอบให้ตอบแต่ละตัวเลือกแล้วต้องเพิ่มช่องให้ผู้ตอบบอกความมั่นใจในการตอบด้วย เช่นให้คะแนนเป็น 4, 3, 2, 1 เมื่อตอบถูกและบอกว่ามั่นใจที่สุด มั่นใจ ไม่ค่อยมั่นใจและไม่มั่นใจเลยตามลำดับ ถ้าผู้เรียนตอบผิดก็จะได้คะแนนเป็น -4, -3, -2, -1 เมื่อบอกว่ามั่นใจที่สุด มั่นใจ ไม่ค่อยมั่นใจและไม่มั่นใจเลยตามลำดับ หรือจะให้คะแนนเป็นอย่างอื่นที่แตกต่างจากนี้ไปก็ได้แต่หลักการคือข้อที่ตอบถูกด้วยความมั่นใจสูงจะกำหนดคะแนนให้มากกว่าข้อ ที่ตอบถูกแต่ไม่มั่นใจและกำหนดคะแนนให้ติดลบสำหรับข้อที่ตอบผิด

จะเห็นได้ว่า จากการศึกษางานวิจัยที่เกี่ยวข้อง จะศึกษาเกี่ยวกับค่าความเที่ยง ความตรง ค่าความยาก และค่าอำนาจจำแนก ซึ่งเป็นคุณลักษณะที่สำคัญของ คุณภาพของแบบทดสอบ ซึ่งเมื่อพิจารณาถึงทฤษฎีการวัดที่ใช้เป็นแนวทางในการศึกษาเกี่ยวกับ คุณภาพของแบบทดสอบเหล่านี้ ได้อาศัยแนวทฤษฎีการทดสอบแบบคลาสิคอล (Classical test theory: CTT) ซึ่งในปัจจุบัน การทดสอบตามแนวนี้ได้ถูกมองว่า มีจุดอ่อนหลายประการ โดยเฉพาะเรื่องค่าพารามิเตอร์ที่แปรเปลี่ยนไปตามกลุ่มตัวอย่าง ในช่วงปี ค.ศ. 1970 และ ค.ศ. 1980 ทฤษฎีการตอบสนองข้อสอบ (Item response theory) ได้รับความสนใจจากนักวัดผลทางจิตวิทยาเป็นอย่างมาก เนื่องจากสามารถแก้จุดอ่อนของทฤษฎีคลาสิคอล (Classical test theory: CTT) ได้โดยทฤษฎีการตอบสนองข้อสอบนี้เชื่อว่า พฤติกรรมที่บุคคลตอบสนองต่อข้อสอบนั้นถูกกำหนดโดยลักษณะบางอย่างภายในตัวบุคคล (Latent trait) ซึ่งไม่สามารถที่จะสังเกตได้จึงพยายามที่จะอธิบายความสัมพันธ์ระหว่างลักษณะภายในตัวบุคคล (Trait) กับพฤติกรรมการตอบสนองต่อข้อสอบ (Performance) โดยอาศัยแบบจำลองทางคณิตศาสตร์เป็นตัวเชื่อมที่สำคัญ จุดเด่นของทฤษฎีนี้ก็คือ ค่าพารามิเตอร์ของข้อสอบซึ่งเป็นตัวกำหนดลักษณะของข้อสอบ คือ ค่าอำนาจจำแนก (a) ค่าความยาก (b) และค่าการเดา (c) ไม่ขึ้นอยู่กับลักษณะของกลุ่มตัวอย่างผู้ตอบข้อสอบ และการประมาณค่าความสามารถของบุคคล ไม่ขึ้นอยู่กับกลุ่มตัวอย่างของข้อสอบ นั่นคือ แม้ผู้สอบจะตอบข้อสอบที่ไม่ใช่ชุดเดียวกัน ก็สามารถประมาณค่าความสามารถเปรียบเทียบกันได้ (พัชรี มีวรรณ์, 2540, หน้า 7) โดยมีดัชนีที่ใช้วัดความถูกต้องแม่นยำ (Precision) ในการประมาณค่าความสามารถที่ต้องการวัด ได้แก่ ค่าฟังก์ชันสารสนเทศ (Information function) (Birnbaum, 1968, p. 419)

จากแนวคิดวิธีการตอบและการตรวจให้คะแนนดังกล่าว ผู้วิจัยสนใจเปรียบเทียบคุณภาพแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจตามทฤษฎีการตอบสนองข้อสอบโดยงานวิจัยนี้ผู้วิจัยจึงใช้วิธีการวิเคราะห์โดยใช้ทฤษฎีการตอบสนองข้อสอบ (IRT) เนื่องจากทฤษฎีการตอบสนองข้อสอบให้แนวคิดที่ชัดเจนเกี่ยวกับการประมาณค่าความสามารถของผู้ตอบแบบทดสอบ โดยใช้ผลการตอบสนองข้อสอบนั้นอ้างอิงเกี่ยวกับความสามารถที่มีอยู่ในตัวบุคคล ในการประมาณค่าความสามารถที่ต้องการวัด ได้แก่ ค่าฟังก์ชั่นสารสนเทศ (Information function) (Birnbaum, 1968, p. 419) ค่าฟังก์ชันสารสนเทศนี้มีค่ามากหรือน้อยขึ้นอยู่กับค่าพารามิเตอร์ของแบบทดสอบ ซึ่งจะเป็นตัวกำหนดลักษณะของแบบทดสอบ เนื่องจากทฤษฎีนี้สามารถแก้จุดอ่อนของทฤษฎีการทดสอบดั้งเดิมได้ ผลการวิจัยนี้จะทำให้ทราบว่าฟังก์ชันสารสนเทศแบบทดสอบชนิดใดมีประสิทธิภาพกว่ากัน เพื่อเป็นข้อมูลในการตัดสินใจใช้แบบทดสอบที่มีวิธีการตอบและตรวจให้คะแนนทั้ง 2 วิธี เพื่อใช้ในการจัดการเรียนการสอนในรายวิชา เคมี ต่อไป และสามารถนำไปประยุกต์เพื่อแก้ปัญหาการวัดผลทาง

การศึกษาและจิตวิทยาได้ สามารถนำไปปรับแก้การเดาของผู้ตอบเป็นรายข้อ ซึ่งส่งผลต่อ คุณภาพของแบบทดสอบและทำให้ค่าคะแนนความคาดเคลื่อนน้อยลง และการพัฒนาวิธีการตอบและตรวจให้คะแนนแบบทดสอบเลือกตอบ สามารถนำมาใช้วัดหรือประเมินการเรียนของผู้เรียนต่อไป

คำถามในการวิจัย

1. แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ มีคุณภาพของแบบทดสอบเป็นอย่างไร

2. แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ แบบใดมีคุณภาพดีกว่ากัน

วัตถุประสงค์

1. เพื่อหาคุณภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ ในด้านฟังก์ชันสารสนเทศ ความยากและอำนาจจำแนก ความตรงตามสภาพ

2. เพื่อเปรียบเทียบคุณภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจในด้านดังต่อไปนี้

2.1 ฟังก์ชันสารสนเทศ

2.2 ความตรงตามสภาพ

สมมติฐานการวิจัย

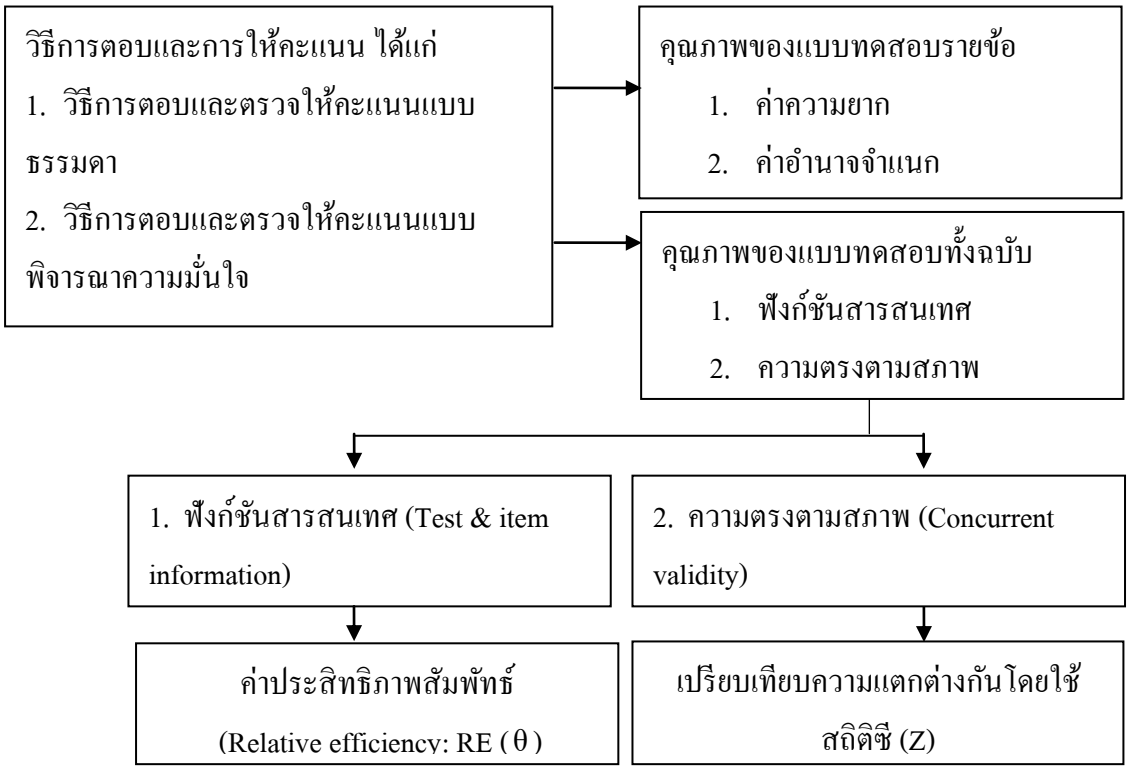
จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ พบว่า แบบทดสอบที่มีการให้คะแนนความรู้บางส่วนมีค่าฟังก์ชันสารสนเทศสูงกว่าการตรวจให้คะแนนแบบประเพณีนิยม (รัตนา ไชยตรี, 2546 อ้างถึงใน Semijima, 1996; Thissen, 1976; Muraki, 1993; ธนวัฒน์ แสนสุข, 2539) ศึกษาการใช้จีอาร์เอ็ม จีพีซีเอ็ม และโมเดลโลจิสติกในการเปรียบเทียบฟังก์ชันสารสนเทศของแบบวัดที่มีการให้คะแนนต่างกัน พบว่า การตรวจให้คะแนนแบบพหุวิภาคที่วิเคราะห์ตาม GRM ให้ค่าสารสนเทศเฉลี่ยสูงกว่าการตรวจให้คะแนนแบบทวิวิภาคเมื่อวิเคราะห์ตามโมเดลโลจิสติก 1, 2 และ 3 พารามิเตอร์ แต่การตรวจให้คะแนนแบบพหุวิภาคที่วิเคราะห์ตาม GPCM พบว่า ให้ค่าสารสนเทศเฉลี่ยต่ำกว่าการตรวจให้คะแนนแบบทวิวิภาคที่วิเคราะห์ตามโมเดลโลจิสติก 1, 2 และ3 พารามิเตอร์ เอมอร จังพรศิริพรปกรณ์ (2545, หน้า ง) ได้ศึกษาเปรียบเทียบคุณภาพของแบบทดสอบในด้าน

ฟังก์ชันสารสนเทศของข้อสอบและแบบสอบ โดยใช้ทฤษฎีการตอบสนองข้อสอบเมื่อตรวจให้คะแนนแบบประเพณีนิยมกับวิธีให้คะแนนความรู้บางส่วน พบว่า ค่าฟังก์ชันสารสนเทศของข้อสอบและแบบสอบและอัตราส่วนสารสนเทศเฉลี่ย วิธีประยุกต์การให้คะแนนของคูมบ์สสูงสุด รองลงมาคือ วิธีประยุกต์การให้คะแนนของเดรสเซลและชมิดท์ และแบบประเพณีนิยม ตามลำดับ รัตนา ไชยตรี (2546) ศึกษาความตรงตามสภาพตามวิธีประยุกต์การให้คะแนนของคูมบ์ส วิธีการให้คะแนนเดรสเซลและชมิดท์ และวิธีตอบโดยบอกระดับความมั่นใจ แล้วนำมาเปรียบเทียบความแตกต่างโดยใช้คะแนนฟิชเชอร์ซี เปรียบเทียบในภาพรวมโดยใช้สถิติทดสอบไคว์สแควร์ พบว่า ค่าความตรงตามสภาพขึ้นอยู่กับการตรวจให้คะแนนในแต่ละวิธี ดังนั้นการวิจัยครั้งนี้จึงตั้งสมมติฐาน

1. ฟังก์ชันสารสนเทศของแบบทดสอบเลือกตอบให้คะแนนแบบพิจารณาความมั่นใจมีประสิทธิภาพสัมพัทธ์สูงกว่าแบบทดสอบเลือกตอบให้คะแนนแบบธรรมดา

2. ความตรงตามสภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจมีความแตกต่างกัน

กรอบแนวคิดในการวิจัย



ภาพที่ 1-1 กรอบแนวคิดในการวิจัย

ขอบเขตของการวิจัย

1. ประชากรที่ใช้ในการวิจัยครั้งนี้ เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 4 ภาคเรียนที่ 2 ปีการศึกษา 2560 ของโรงเรียนในสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 18 ซึ่งมีทั้งหมด 11,060 คน จากโรงเรียนทั้งหมด 50 โรงเรียน

2. ตัวแปรที่ใช้ในการวิจัย คือ

2.1 ตัวแปรอิสระ ได้แก่ วิธีการตอบและการตรวจให้คะแนนจำแนกเป็น

2.1.1 วิธีการตอบและตรวจให้คะแนนแบบธรรมดา

2.1.2 วิธีการตอบและตรวจให้คะแนนแบบพิจารณาความมั่นใจ

2.2 ตัวแปรตาม ได้แก่ คุณภาพของแบบทดสอบจำแนกเป็น

2.2.1 ฟังก์ชันสารสนเทศ

2.2.2 ความยากและอำนาจจำแนก

2.2.3 ความตรงตามสภาพ

3. เครื่องมือที่ใช้ในการวิจัย เป็นแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชา เคมี กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 4 ภาคเรียนที่ 2 ปีการศึกษา 2560 ซึ่งเป็นแบบสอบชนิดเลือกตอบ 4 ตัวเลือก ที่ผู้วิจัยสร้างขึ้น จำนวน 1 ฉบับ จำนวน 30 ข้อ ซึ่งมีวิธีการตอบและตรวจให้คะแนน 2 วิธี แบ่งเป็นแบบทดสอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ

4. ขอบเขตด้านเนื้อหาวิชาเคมี ช่วงชั้นที่ 4 ชั้นมัธยมศึกษาปีที่ 4 ประกอบด้วย สาระทักษะการคิดคำนวณเกี่ยวกับคำนวณเกี่ยวกับมวลอะตอม มวลโมเลกุล โมล ความสัมพันธ์ระหว่างมวลสารกับจำนวนโมล ความสัมพันธ์ระหว่างปริมาตรต่อโมลของแก๊ส ความสัมพันธ์ระหว่างจำนวนอนุภาคกับจำนวนโมล ความเข้มข้นของสารละลาย การเตรียมสารละลาย และสมบัติบางประการของสารละลาย มวลสารในปฏิกิริยาเคมี ปริมาณของแก๊สในปฏิกิริยาเคมี ได้แก่ กฎของเกย์-ลุสแซก กฎของอาโวกาโดร การคำนวณหาสูตรเอมพิริกัล สูตรโมเลกุล การคำนวณหามวลเป็นร้อยละจากสูตร สมการเคมี ความสัมพันธ์ระหว่างปริมาณสารของสารในสมการเคมี ชนิดเลือกตอบ 4 ตัวเลือก จำนวน 1 ฉบับ 30 ข้อ

นิยามศัพท์เฉพาะ

แบบทดสอบ (Test) หมายถึง ชุดของคำถาม หรือ สถานการณ์ที่ใช้วัดความรู้ความสามารถ และทักษะวิชาวิทยาศาสตร์ของนักเรียน ซึ่งแต่ละข้อคำถามนี้นักเรียนตอบสนองออกมาให้สามารถสังเกต หรือนับจำนวนได้

แบบทดสอบเลือกตอบแบบธรรมดา (Multiple choice) หมายถึง แบบทดสอบที่แต่ละข้อคำถามประกอบด้วย ตอนนำ หรือ ตัวคำถาม (Stem or lead or problem) กับ ตัวเลือก (Choice or option or alternative or response) ในแต่ละข้อคำถาม มีตัวเลือกมากกว่า 2 ตัวเลือก สำหรับการศึกษาครั้งนี้มี 4 ตัวเลือก และมีตัวเลือกถูกเพียงตัวเลือกเดียว เพื่อให้ผู้สอบตอบสนองออกมาสังเกตได้และนับจำนวนได้ โดยมีวิธีการตรวจให้คะแนนแบบธรรมดา คือวิธีการให้ผู้สอบเลือกคำตอบที่ถูกที่สุดเพียงข้อเดียว โดยมีการตรวจให้คะแนน 1 คะแนน เมื่อผู้สอบเลือกตัวเลือกที่ถูกและตรวจให้คะแนน 0 คะแนน เมื่อผู้สอบเลือกตอบตัวเลือกผิดหรือไม่เลือกหรือเลือกตัวเลือกมากกว่า 1 ตัวเลือก การวิจัยครั้งนี้เป็นแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชา เคมี กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 4 ภาคเรียนที่ 2 ปีการศึกษา 2560 แบบทดสอบชนิดเลือกตอบ 4 ตัวเลือก จำนวน 30 ข้อ ที่มีรูปแบบการตอบและตรวจให้คะแนนแบบธรรมดา

แบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ (Confidential making) หมายถึง แบบทดสอบที่มีคำชี้แจงให้นักเรียนทำเครื่องหมายแก่ตัวเลือกที่ถูกที่สุดเพียงตัวเลือกเดียว พร้อมทั้งบอกความมั่นใจในการตอบตัวเลือกนั้นด้วยว่ามาก ปานกลาง หรือน้อย โดยมีวิธีการตรวจให้คะแนนความมั่นใจ คือการคาดคะเนของผู้สอบว่าคำตอบที่เลือกนั้นมีโอกาสเป็นคำตอบที่ถูกต้อง โดยระบุความมั่นใจในการตอบว่า มาก ปานกลาง หรือน้อยอย่างใดอย่างหนึ่ง ซึ่งคะแนนตามระดับความมั่นใจ 3 ระดับ การกำหนดคะแนนรายข้อที่ให้คะแนนแต่ละข้อเป็นไปตามเกณฑ์ ระดับความมั่นใจมากจะได้ 3 ปานกลางจะได้ 2 น้อยจะได้ 1 โดยข้อที่ไม่ตอบและตอบผิดจะได้ 0 คะแนน (รุจิรา ขาวสะอาด, 2543) การวิจัยครั้งนี้เป็นแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชา เคมี กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 4 ภาคเรียนที่ 2 ปีการศึกษา 2560 แบบทดสอบชนิดเลือกตอบ 4 ตัวเลือก จำนวน 30 ข้อ ที่มีรูปแบบการตอบและตรวจให้คะแนนแบบพิจารณาความมั่นใจ

ทฤษฎีการตอบสนองข้อสอบ (Item response theory) หมายถึง ทฤษฎีที่อธิบายถึงความสัมพันธ์ระหว่างความสามารถที่แท้จริง ที่มีอยู่ภายในตัวผู้สอบกับการตอบสนองข้อสอบ ซึ่ง ความสามารถที่แท้จริง คือ ลักษณะที่วัดได้ด้วยแบบทดสอบ

ฟังก์ชันสารสนเทศของแบบทดสอบ (Test information) หมายถึง ค่าความแม่นยำในการประมาณค่าความสามารถจริงของผู้สอบจากการตอบแบบทดสอบนั้น คำนวณได้จากผลรวมของค่าฟังก์ชันสารสนเทศของข้อสอบทุกข้อในแบบทดสอบทั้งฉบับถ้าสูงที่ระดับใดแสดงว่ามีความแม่นยำสูง การวิจัยครั้งนี้ซึ่งประมาณค่าได้จาก การวิเคราะห์ตามทฤษฎี IRT จะทำให้ค่าฟังก์ชันสารสนเทศเหมาะสมที่จะใช้เป็นดัชนีบ่งบอกคุณภาพของข้อสอบและแบบทดสอบแทนการหาค่าความเที่ยงและความคลาดเคลื่อนมาตรฐานของการวัดตามทฤษฎีแบบสอบดั้งเดิม

ฟังก์ชันสารสนเทศของข้อสอบ (Item information) หมายถึง ค่าความแม่นยำในการประมาณค่าความสามารถของผู้สอบจากการตอบข้อสอบในข้อนั้น ๆ คำนวณได้จากอัตราส่วน ระหว่างค่าอนุพันธ์ของการตอบได้คะแนนในระดับที่มีความรู้เต็มที่กับผลคูณของการตอบในระดับของความรู้บางส่วน

คุณภาพของแบบทดสอบ หมายถึง กระบวนการตรวจสอบคำตอบของผู้สอบในแต่ละข้อเพื่อพิจารณาข้อสอบแต่ละข้อมีระดับค่าความยาก ค่าอำนาจจำแนก ความเที่ยงและความตรง

. ค่าพารามิเตอร์ของข้อสอบ หมายถึง ค่าที่แสดงถึงคุณลักษณะของข้อสอบแต่ละข้อประกอบ ด้วย ค่าอำนาจจำแนก (a) ค่าความยาก (b) การวิจัยครั้งนี้มีการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (Item response theory) แบบทวิภาค ที่ใช้โมเดลโลจิสติก (Logistic) 2 พารามิเตอร์ และแบบพหุวิภาค ที่ใช้โมเดลการให้คะแนนความรู้บางส่วน (Generalized partial creadit model: GPCM) ซึ่งวิเคราะห์ด้วยโปรแกรมสำเร็จรูป IRTPRO

ค่าอำนาจจำแนกของข้อสอบ (a) หมายถึง ค่าที่แสดงถึงความสามารถของข้อสอบที่จะจำแนกผู้สอบระหว่างผู้ที่มีระดับความสามารถต่ำกว่าค่าความยากของข้อสอบและผู้ที่มีระดับความสามารถสูงกว่าค่าความยากของข้อสอบ ซึ่งหาได้จากค่าที่เป็นสัดส่วนโดยตรงกับความชันของโค้ง ลักษณะข้อสอบ ณ จุดเปลี่ยนโค้ง โดยที่ค่าอำนาจจำแนกในทางทฤษฎีมีค่าระหว่าง ($-\infty$ ถึง $+\infty$) ควรมีค่าเป็น + ตามปกติมีค่าไม่เกิน +2.50 (ศิริชัย กาญจนวาสี, 2545, หน้า 48)

ค่าความยากของข้อสอบ (b) หมายถึง ค่าที่แสดงถึงระดับความยากง่ายของข้อสอบ ซึ่งมี

ค่าเท่ากับระดับความสามารถ ณ จุดเปลี่ยนโค้งของโค้งลักษณะข้อสอบ มีค่าที่เป็นไปได้ตั้งแต่ -∞ ถึง + ∞ แต่ในทางปฏิบัติมักจะเลือกใช้ค่าที่อยู่ระหว่าง -2.50 ถึง +2.50, ค่า -2.50 แสดงว่าข้อสอบง่ายมาก และค่า +2.50 แสดงว่าข้อสอบยากมาก (ศิริชัย กาญจนวาสี, 2545, หน้า 48)

ประสิทธิภาพสัมพัทธ์ (Relative efficiency) หมายถึง อัตราส่วนระหว่างค่าฟังก์ชันสารสนเทศระหว่างแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ ที่วัดคุณลักษณะเดียวกัน

ความตรงตามสภาพ (Concurrent validity) หมายถึง ค่าสหสัมพันธ์ระหว่างคะแนนจากแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาเคมี ภาคเรียนที่ 2 ชั้นมัธยมศึกษาปีที่ 4 ที่ได้จากวิธีการตอบ และการให้คะแนนในแต่ละวิธี กับระดับผลการเรียนวิชาเคมี ภาคเรียนที่ 1 ชั้นมัธยมศึกษาปีที่ 4

ประโยชน์ที่ได้รับจากการวิจัย

1. ได้แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ ที่ได้มาตรฐานไปใช้ในห้องเรียน

2. เป็นข้อมูลในการตัดสินใจของครูเพื่อเลือกใช้แบบทดสอบที่เหมาะสม โดยเฉพาะอย่างยิ่งเพื่อแก้ปัญหาแบบทดสอบที่มีการเดา

3. ทำให้ทราบถึงความสามารถที่แท้จริงของผู้สอบและความสามารถในการประมาณค่าพารามิเตอร์เกี่ยวกับแบบทดสอบ รวมทั้งเป็นการเพิ่มสารสนเทศเกี่ยวกับการตอบแบบทดสอบแบบเลือกตอบให้มากขึ้น และนำมาซึ่งการพัฒนาแนวคิดเกี่ยวกับการหาค่าคะแนนความสามารถที่แท้จริงจากการใช้วิธีการตอบและการตรวจให้คะแนนแบบต่าง ๆ

บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

ในงานวิจัยครั้งนี้ผู้วิจัยมีวัตถุประสงค์ 1) เพื่อหาคุณภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ ในด้านฟังก์ชันสารสนเทศ ความยากและอำนาจจำแนก ความตรงตามสภาพ 2) เพื่อเปรียบเทียบคุณภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ ในด้านดังต่อไปนี้ 2.1) ฟังก์ชันสารสนเทศ 2.2) ความตรงตามสภาพ ผู้วิจัยนำเสนอผลการศึกษาเอกสารที่เกี่ยวข้องกับแนวคิด ดังนี้

ตอนที่ 1 การศึกษาเกี่ยวกับแบบทดสอบเลือกตอบและวิธีการตรวจให้คะแนน

1. แบบทดสอบเลือกตอบ (Test)
2. วิธีการตรวจให้คะแนน

ตอนที่ 2 ทฤษฎีการตอบสนองข้อสอบ

1. แนวคิดพื้นฐานทฤษฎีการตอบสนองข้อสอบ
2. หลักการของทฤษฎีตอบสนองข้อสอบ
3. ข้อตกลงเบื้องต้นของทฤษฎีตอบสนองข้อสอบ
4. โมเดลการตอบสนองข้อสอบ (Item response model)
   4.1 โมเดลการตอบสนองข้อสอบแบบทวิวิภาค (Dichotomous item response model)
   4.2 โมเดลการตอบสนองข้อสอบแบบพหุวิภาค (Polytomous item response model)
5. ฟังก์ชันสารสนเทศของข้อสอบและแบบทดสอบ (Item and test information)

ตอนที่ 3 การตรวจสอบคุณภาพของแบบทดสอบ

1. อำนาจจำแนก (Discrimination index)
2. ความยาก (Difficulty)
3. ความเที่ยง (Reliability)
4. ความตรง (Validity)

ตอนที่ 4 การสร้างแบบทดสอบโดยใช้ทฤษฎีการวัดและประเมินผลเรียนรู้ของบลูม (Bloom’s taxonomy)

1. ทฤษฎีการวัดและประเมินผลเรียนรู้ของบลูม (Bloom’s taxonomy)
2. การนำไปใช้ออกแบบจัดการเรียนรู้ และการประเมินผล
3. การสร้างแบบทดสอบ
4. การประยุกต์ IRT สำหรับการสร้างแบบทดสอบ

ตอนที่ 5 คะแนนจุดตัด (Cut score)

1. ความหมายของคะแนนจุดตัด
2. แนวคิดเรื่องการกำหนดคะแนนจุดตัด
3. วิธีการกำหนดคะแนนจุดตัด (Standard-setting method)

ตอนที่ 6 งานวิจัยที่เกี่ยวข้อง

1. งานวิจัยที่เกี่ยวข้องกับวิธีการตรวจให้คะแนน
2. งานวิจัยที่เกี่ยวกับการหาคุณภาพข้อสอบ

ตอนที่ 1 การศึกษาเกี่ยวกับแบบทดสอบเลือกตอบและวิธีการตรวจให้คะแนน

แบบทดสอบเลือกตอบ

สุรีพร อนุศาสนนันท์ (2558, หน้า 4) กล่าวว่า แบบสอบ คือชุดคำถามที่สร้างอย่างเป็นระบบ ที่มีแบบแผนเพื่อให้วัดพฤติกรรมของนักเรียน

พัชรี มีวรรณ์ (2540) กล่าวว่า แบบทดสอบเลือกตอบเป็นที่นิยมมากทั้งในแบบสอบมาตรฐาน และแบบสอบวัดผลสัมฤทธิ์ที่ใช้ในการทดสอบตามสถานศึกษาต่าง ๆ

แบบทดสอบเลือกตอบเป็นแบบทดสอบที่นิยมใช้กันโดยทั่วไปในทุกวงการ เพราะแบบทดสอบเลือกตอบสามารถวัดพฤติกรรมต่าง ๆ ได้หลายด้านและหลายระดับ สามารถใช้ได้กับเนื้อหาวิชาเกือบทุกลักษณะ การตรวจให้คะแนนได้ผลคงที่ ประหยัดเวลาและแรงงาน นอกจากนั้นยังสามารถวิเคราะห์ได้ว่า ข้อใดดีหรือไม่ดี (ชวาล แพรัตกุล, 2518, หน้า 164) ในการวัดผลสัมฤทธิ์นั้นสิ่งที่สำคัญประการหนึ่ง คือ ต้องการวัดผลการเรียนรู้ในระดับความรู้ ความจำ ความเข้าใจ การนำไปใช้ การวิเคราะห์ และการประเมินค่า (Gronlund, 1976, p. 270 อ้างถึงใน พิมพ์อุษา เจริญยศ, 2538, หน้า 5) จะเห็นได้ว่าแบบทดสอบเลือกตอบสามารถวัดระดับความรู้เหล่านี้ได้ดี จึงถูกนำไปใช้ในการสอบทุกระดับชั้น ตั้งแต่ประถมศึกษาจนถึงระดับมหาวิทยาลัยและได้มีการพัฒนาวิธีเขียนข้อคำถามของแบบทดสอบเลือกตอบ ให้มีคำถามและตัวเลือกแตกต่างกันออกไปเพื่อจะสามารถเลือกใช้ให้เหมาะสมกับเนื้อหาเรื่องราว หรือจุดประสงค์ของเนื้อหาวิชา หรือจุดประสงค์ของเนื้อหาวิชา ตลอดจนใช้ในการวัดผลการเรียนรู้ในระดับชั้นต่าง ๆ ได้อย่างมีประสิทธิภาพ

จากการศึกษาเอกสารและงานวิจัยของนักวัดผลและนักการศึกษาหลาย ๆ ท่าน พอจะสรุปได้ว่าแบบทดสอบเลือกตอบโดยทั่วไปประกอบด้วยส่วนที่สำคัญ 2 ส่วน คือ

1. ส่วนที่เป็นตัวปัญหาหรือคำถามหรือตอนนำ (Problem or question or stem)

2. ส่วนที่เป็นตอนตอบหรือตัวเลือก (Alternative or choice) ซึ่งตัวเลือกที่ถูก เรียกว่า ตัวถูก (Correct choice) และตัวเลือกที่ไม่ถูกต้อง เรียกว่า ตัวลวง หรือตัวผิด (Foils, distractors)

การจำแนกรูปแบบของแบบทดสอบเลือกตอบ

ชวาล แพรัตกุล (2520, หน้า 58) ได้จำแนกเป็นรูปแบบใหญ่ ๆ ได้ 3 แบบ คือ

1. แบบคำถามเดี่ยว (Single question) ลักษณะของข้อสอบแต่ละข้อจะมีคำถามเดียวโดด ๆ

2. แบบตัวเลือกคงที่ (Constant choice) ลักษณะคำถามประเภทนี้จะรวมเนื้อหา บางเรื่อง บางตอนที่มีความเป็นเอกพันธ์ร่วมกันอย่างใดอย่างหนึ่ง ไว้เป็นตัวเลือกตอบ แล้วเขียนคำถามเป็นชุด ๆ นักเรียนต้องใช้ความสามารถหลาย ๆ ด้านมาผสมผสานกันจึงจะสามารถตอบได้ถูกต้อง

3. แบบกำหนดสถานการณ์ (Situation) เป็นแบบที่ใช้วิธีการกำหนด ข้อความ ภาพ ตารางให้พิจารณา แล้วตั้งคำถามเกี่ยวกับข้อความ ภาพ หรือตารางที่กำหนดให้นั้น

การสร้างแบบทดสอบเลือกตอบ

ในการสร้างแบบทดสอบเลือกตอบ มีหลักการสร้างดังนี้ (Gronlund, 1981, pp. 40-53)

1. แต่ละข้อคำถามที่สร้าง ควรวัดการเรียนรู้ที่สำคัญเพียงเรื่องเดียว

2. คำถามในตอนนำของแต่ละข้อคำถามต้องชัดเจน

3. คำถามในตอนนำต้องไม่ซับซ้อน และใช้ภาษาที่ง่าย

4. คำถามที่ใช้ในตอนนำของแต่ละข้อคำถาม ต้องใช้คำที่มีความหมาย

5. ควรเขียนคำถามในตอนนำให้อยู่ในรูปเชิงบวกให้มากที่สุดเท่าที่จะเป็นไปได้

6. หากใช้คำที่มีความหมายในเชิงนิเสธ ให้เขียนหรือขีดเส้นใต้ให้เห็นชัดเพื่อเน้นคำถาม

7. สร้างตัวเลือกถูกให้เป็นคำตอบที่ถูกต้องที่สุด

8. สร้างตัวเลือกทั้งหมดให้สมเหตุสมผลกับตอนนำของข้อคำถาม

9. หลีกเลี่ยงถ้อยคำที่ทำให้สามารถเลือกคำตอบที่ถูกหรือตัดตัวลวงออกได้ เช่น ใช้คำที่คล้ายกับตอนนำให้เป็นคำตอบที่ถูก ยกคำหรือวลีที่เป็นคำตอบถูกมาทั้งหมด ตัวถูกเป็นคำที่กล่าวบ่อย ๆ จนขึ้นหู ตัวลวงเห็นเด่นชัดว่าผิด (มีคำว่าบ่อย ๆ เสมอ ทั้งหมด ไม่เพียงแต่) หรือคำที่มีความหมายคล้าย ๆ กัน

10. ตัวลวงต้องมีเหตุผลในการตอบ อาจทำได้โดย

10.1 ใช้จากความเข้าใจผิดโดยทั่ว ๆ ไป หรือจากคำตอบที่ผิดของนักเรียน

10.2 จากสถานการณ์การเลือกตอบด้านภาษาจากนักเรียน

10.3 ใช้คำที่ให้ความรู้สึกที่ดีในตัวลวงเหมือนกับที่ใช้ในคำตอบที่ถูก เช่น คำว่าสำคัญ บางครั้ง เป็นต้น

10.4 ใช้ตัวลวงที่ไม่เกี่ยวข้องแต่เป็นวลีที่เป็นแบบอย่างหรือมีเหตุผล หรือมีถ้อยคำที่ สัมพันธ์กับตอนนำของข้อคำถาม แต่ไม่มากเกินไปที่จะชี้ให้เห็นว่าไม่มีประสิทธิภาพ

10.5 ทำตัวลวงให้สอดคล้องกัน แต่ต้องระวังให้มีค่าอำนาจจำแนก และมีนัยสำคัญ

11. ให้คำตอบที่ถูกยาวพอ ๆ กับตัวลวง

12. หลีกเลี่ยงการใช้ตัวเลือก “ที่กล่าวมาข้างต้นถูกทั้งหมด” “ที่กล่าวมาข้างต้นผิดทั้งหมด” หากจำเป็นต้องใช้ อาจใช้คำว่า “คำตอบที่ถูกมิได้ให้ไว้”

13. เรียงตำแหน่งคำตอบถูกแบบสุ่ม

14. ต้องมั่นใจแต่ละข้อคำถามมีความเป็นอิสระจากข้อคำถามอื่น

15. ใช้รูปแบบข้อคำถามที่มีประสิทธิภาพ ได้แก่ มีตัวเลือกที่ง่ายต่อการตอบ ทั้งการอ่าน การเปรียบเทียบ และการตรวจให้คะแนน

ข้อดีของข้อสอบแบบเลือกตอบ

สุรีพร อนุศาสนนันท์ (2558, หน้า 67) ได้กล่าวถึงข้อดีของแบบทดสอบเลือกตอบ ดังนี้ สามารถวัดเนื้อหาได้ครอบคลุม ตรวจให้คะแนนง่ายและรวดเร็ว การตรวจให้คะแนนมีความยุติธรรม มีความเป็นปรนัย สามารถสร้างให้วัดสมรรถภาพสมองด้านต่าง ๆ ตามลำดับขั้นของความรู้ได้

ล้วน สายยศ และอังคณา สายยศ (2539, หน้า 108-109) ได้กล่าวถึงข้อดีของแบบทดสอบเลือกตอบไว้ดังนี้

1. วัดได้ครอบคลุมเนื้อหา ข้อคำถามของข้อสอบเลือกตอบเป็นการถามสั้น ๆ และเจาะจงส่วนใดส่วนหนึ่ง ดังนั้นสามารถถามรายละเอียดได้มากมาย คลุมเนื้อหาที่ผู้เรียนได้เรียนมาตลอด

2. วัดได้คลุมพฤติกรรม ข้อสอบแบบเลือกตอบสามารถเขียนเพื่อสอบวัดพฤติกรรมขั้นต่ำไปยังพฤติกรรมขั้นสูงได้ นั่นคือเราสามารถแปลงการสอบวัดพฤติกรรมที่ง่ายหรือที่ซับซ้อนมาวัดโดยข้อสอบแบบเลือกตอบได้อย่างดี แม้ว่าบางอย่างจะวัดโดยตรงไม่ได้ เราสามารถพลิกแพลงให้วัดสิ่งนั้นได้โดยทดแทนกัน ความจำ ความคิด การสร้างสรรค์ ก็สามารถใช้ข้อสอบเลือกตอบตรวจสอบได้

3. มีความเป็นปรนัยสูง นั่นคือข้อสอบเลือกตอบสามารถตรวจให้คะแนนตรงกัน ข้อคำถามเข้าใจตรงกัน การแปลคะแนนก็ทำให้ตรงกันได้ง่าย

4. ประหยัดเวลาทำงาน ข้อสอบแบบเลือกตอบสอบเสร็จตรวจให้คะแนนได้ทันที สามารถเจาะคีย์ตรวจมือได้รวดเร็วทันใจ ในสมัยนี้การตรวจได้พัฒนาขึ้นมาเป็นการตรวจด้วยเครื่อง ข้อสอบเลือกตอบใช้ได้ผลดีที่สุด และรวดเร็วมาก ประหยัดเวลา และแรงงาน

5. สามารถวิเคราะห์ได้ ข้อสอบเลือกตอบสอบเสร็จแล้วสามารถนำมาปรับปรุงแก้ไขจนเป็นที่เชื่อถือได้

6. ควบคุมความยากของแต่ละข้อได้ โดยอาศัยการเขียนตัวเลือกดี ๆ หรือเปลี่ยนแปลงตัวเลือกในรูปแบบต่าง ๆ เช่น ตัวเลือกมาก ๆ หรือตัวเลือกที่เป็นลักษณะเอกพันธุ์ (Homogenous) เป็นต้น

7. ตัวเลือกในข้อสอบแบบเลือกตอบใช้ประโยชน์ในการสอบเพื่อวินิจฉัย (Diagnostic purposes) ได้โดยใช้ข้อมูลจากการเลือกตอบตัวเลือกจากนักเรียนมาใช้พิจารณา นอกจากนั้นยังส่งเสริมความสามารถในการอ่านด้วยอย่างดี

8. ข้อสอบเลือกตอบที่ดีมีโอกาสเดาได้น้อย โอกาสของการเดาขึ้นอยู่กับจำนวนตัวเลือกกับจำนวนข้อสอบมากน้อยเท่าใด

9. ข้อสอบเลือกตอบมีโอกาสให้ความยุติธรรมสูง เพราะออกได้คลุมตัวอย่างของความรู้และพฤติกรรม ตลอดจนความแม่นยำในการตรวจสอบคะแนน

10. เป็นการส่งเสริมปรัชญาของการตัดสินใจ ทั้งนี้เพราะชีวิตคนในโลกนี้เต็มไปด้วยการเลือก เช่น เลือกเรียน เลือกอาหาร เลือกซื้อของ เลือกอาชีพ และกระทั่งการเลือกคู่เพื่อแต่งงาน ฯลฯ จึงเห็นว่าเมื่อชีวิตทุกขณะเต็มไปด้วยการเลือก ทำไมสังคมไม่สอนให้คนเลือกเป็น ทำไมการเรียนการสอนจึงไม่พัฒนาชีวิตที่จะเลือกสิ่งใดมีคุณค่ากว่า เหมาะสมกว่า การสร้างแบบทดสอบเลือกตอบที่ดี จึงเป็นการฝึกคนให้เลือกเป็น ตัดสินใจเป็น เสริมสร้างพัฒนาความคิดในการตัดสินใจดีขึ้น

ข้อเสียของแบบทดสอบเลือกตอบ

สุรีพร อนุศาสนนันท์ (2558, หน้า 67) ได้กล่าวถึงจุดบกพร่องของแบบทดสอบเลือกตอบ ดังนี้ เสียค่าใช้จ่ายในการสร้างและพัฒนาสูง สามารถเดาคำตอบได้ถูก

แม้ว่าแบบทดสอบเลือกตอบจะมีข้อดีในแง่สามารถวัดความรู้ความสามารถในระดับสูง ถามได้ครอบคลุมเนื้อหาสาระ ตรวจคะแนนได้ง่าย รวดเร็วและมีความเป็นปรนัยสูง แต่ยังมีจุดอ่อนที่ทำให้ผลการวัดโดยใช้แบบทดสอบชนิดนี้มีความถูกต้องน้อยกว่าที่ควรจะเป็นอยู่หลายประการ จุดอ่อนที่ควรได้รับการพิจารณา 3 ประการ คือ

1. การให้มีตัวเลือกที่เป็นคำตอบที่ถูกต้องเพียงตัวเดียวอาจไม่เหมาะสม ในกรณีที่ข้อความรู้ที่นำมาสร้างเป็นปัญหานั้นมีคำตอบได้หลายคำตอบและเป็นคำตอบที่ถูกต้องได้เท่า ๆ กัน การตัดสินว่าคำตอบใดเป็นคำตอบที่ถูกต้องที่สุดหรือดีที่สุดเพียงคำตอบเดียวบางที่ไม่อาจทำได้ การเขียนข้อสอบให้มีคำตอบที่ถูกต้องเพียงคำตอบเดียวจึงเป็นเรื่องยากมาก ถ้าจะแก้ปัญหาโดย

การแยกเอาคำตอบที่ถูกต้องคำตอบใดคำตอบหนึ่งมาเป็นตัวเลือกที่ถูกปนกับตัวเลือกที่ผิดอื่น ๆ ก็ไม่อาจวัดความรู้ที่สมบูรณ์ครบถ้วนในเรื่องนั้นได้ (จักรกฤษณ์ สำราญใจ, 2530, หน้า 3)

2. การมีคำตอบที่ถูกต้องเพียงคำตอบเดียวอาจจะเป็นการช่วยให้ผู้ตอบหาคำตอบที่ถูกต้องได้ง่ายขึ้น ดังผลการวิจัยของ Kolstad (1985, pp. 27-31 อ้างถึงใน จักรกฤษณ์ สำราญใจ, 2530, หน้า 3-4) ได้ทดลองให้ผู้ตอบข้อสอบ 2 กลุ่มตอบข้อสอบเลือกตอบที่ถูกต้องเพียงคำตอบเดียวโดยกลุ่มหนึ่งทราบว่าเป็นข้อสอบเลือกตอบที่มีคำตอบที่ถูกต้องเพียงคำตอบเดียว อีกกลุ่มหนึ่งก็ให้ตอบข้อสอบชุดเดียวกัน แต่ให้ตอบรับตัวเลือกที่ถูก และตอบปฏิเสธตัวเลือกที่ผิดโดยไม่ทราบว่าตัวเลือกที่เป็นคำตอบที่ถูกกี่ตัว ผลปรากฏว่า กลุ่มที่ไม่ทราบว่าข้อสอบมีคำตอบที่ถูกเพียงคำตอบเดียวยอมรับคำตอบที่ผิดมากกว่ากลุ่มที่ทราบ

3. การให้คะแนน 1 สำหรับผู้เลือกคำตอบได้ถูก และให้ 0 สำหรับผู้ที่เลือกคำตอบไม่ถูกนั้น ไม่สามารถแยกแยะผู้ที่มีความรู้ถูกต้องสมบูรณ์ ผู้ที่มีความรู้บางส่วนกับผู้ที่มีความรู้ไม่ถูกต้องออกจากกันได้เลย (Coombs et al., 1956, pp. 13-14 อ้างถึงใน จักรกฤษณ์ สำราญใจ, 2530, หน้า 4)

แนวทางปรับแก้ข้อบกพร่องของแบบทดสอบเลือกตอบ

ถึงแม้แบบทดสอบเลือกตอบจะมีจุดอ่อนหลายข้อดังกล่าว แต่แบบทดสอบปรนัยรูปแบบนี้ยังมีข้อได้เปรียบอยู่หลายประการที่มีคุณค่าพอที่จะยังคงใช้เป็นเครื่องมือในการวัดผลอยู่ นักวัดผลหลายท่านจึงได้พยายามหาทางแก้ไขจุดอ่อนเหล่านั้นเพื่อให้การใช้แบบทดสอบเลือกตอบมีประสิทธิภาพมากที่สุด วิธีการแก้ไขจุดอ่อนเหล่านี้พอสรุปได้เป็น 3 แนวทาง คือ

1. ปรับวิธีการตรวจให้คะแนน วิธีการนี้รูปแบบและวิธีการตอบยังคงเหมือนกับแบบทดสอบเลือกตอบทุกประการ แต่จะมีการเปลี่ยนแปลงวิธีการตรวจให้คะแนน (จักรกฤษณ์ สำราญใจ, 2530, หน้า 5)

2. ปรับวิธีการตอบและวิธีการตรวจให้คะแนน วิธีการนี้ผู้ตอบจะได้รับคำชี้แจงให้ตอบแบบทดสอบด้วยวิธีการที่แตกต่างจากการตอบแบบทดสอบเลือกตอบแต่แบบทดสอบจะยังคงมีรูปแบบเหมือนเดิม เช่น การให้ตอบปฏิเสธตัวเลือกที่ผิด การให้ตอบรับตัวเลือกที่ถูกและตอบปฏิเสธตัวที่ผิด การให้เลือกคำตอบได้หลายตัวเลือก การให้ผู้ตอบบอกระดับความมั่นใจใน การตอบ (อรวรรณ ตัณฑ์เจริญรัตน์, 2517, หน้า 23-26; กาญจนา ศิริวัฒนาพงษ์, 2520, หน้า 67-79; จักรกฤษณ์ สำราญใจ, 2530, หน้า 5)

3. ปรับรูปแบบข้อสอบ วิธีการตอบและวิธีการตรวจให้คะแนน วิธีการนี้รูปแบบของแบบสอบจะถูกปรับให้มีตัวเลือกที่เป็นคำตอบที่ถูกได้หลายตัวเลือก เช่น การใช้แบบทดสอบเลือกตอบชนิดเชิงซ้อน (Double or complex) แบบทดสอบเลือกตอบแบบถูกผิด (Multiple true-

false) เป็นต้น (จินดา โตอนันต์, 2526, หน้า 59-65; ปราณี ร่วมทอง, 2528, หน้า 71-79; จักรกฤษณ์ สำราญใจ, 2530, หน้า 5)

ประเภทของแบบทดสอบ

ประเภทของแบบทดสอบสามารถแบ่งประเภทออกได้หลายลักษณะขึ้นอยู่กับเกณฑ์ที่จะใช้

1. แบ่งตามสมรรถภาพที่จะวัด แบ่งเป็น 3 ประเภท

1.1 แบบทดสอบวัดผลสัมฤทธิ์ (Achievement test) หมายถึง แบบทดสอบที่วัดสมรรถภาพสมองด้านต่าง ๆ ที่นักเรียนได้รับการเรียนรู้ผ่านมาแล้วว่ามีอยู่เท่าใด แบบทดสอบแบ่งออกเป็น 2 ชนิด

1.1.1 แบบทดสอบที่ครูสร้างขึ้น หมายถึง แบบทดสอบที่มุ่งวัดผลสัมฤทธิ์ของผู้เรียนเฉพาะกลุ่มที่ครูสอน

1.1.2 แบบทดสอบมาตรฐาน หมายถึง แบบทดสอบที่มุ่งวัดผลสัมฤทธิ์ มีจุดมุ่งหมายเพื่อเปรียบเทียบคุณภาพต่าง ๆ ของนักเรียนที่ต่างกลุ่มกัน เช่น แบบทดสอบมาตรฐานระดับชาติ

1.2 แบบทดสอบวัดความถนัด (Aptitude test) หมายถึง แบบทดสอบที่มุ่งวัดสมรรถภาพสมองของผู้เรียน

1.2.1 แบบทดสอบวัดความถนัดทางการเรียน หมายถึง แบบทดสอบที่มุ่งวัดความถนัดทางด้านวิชาการต่าง ๆ เช่น ด้านภาษา

1.2.2 แบบทดสอบวัดความถนัดเฉพาะ หมายถึง แบบทดสอบที่มุ่งวัดความถนัดเฉพาะที่เกี่ยวกับงานอาชีพต่าง ๆ หรือความสามารถพิเศษ เช่น ความสามารถทางดนตรี

1.3 แบบทดสอบวัดบุคลิกภาพทางสังคม หมายถึง แบบทดสอบที่ใช้วัดบุคลิกภาพและการปรับตัวให้เข้ากับ สังคม ซึ่งเป็นเรื่องที่วัด ได้ยาก ผลที่ได้ไม่คงที่แน่นอน

1.3.1 แบบทดสอบวัดเจตคติที่มีต่อบุคคล สิ่งของ เรื่องราว

1.3.2 แบบทดสอบวัดความสนใจที่มีต่ออาชีพ การศึกษา

1.3.3 แบบทดสอบวัดการปรับตัว เช่น การปรับตัวเข้ากับเพื่อน ๆ

2. แบ่งตามลักษณะการตอบ

2.1 แบบทดสอบภาคปฏิบัติ หมายถึง แบบทดสอบที่ให้นักเรียนลงมือปฏิบัติจริง

2.2 แบบทดสอบข้อเขียน หมายถึง แบบทดสอบที่ใช้การเขียนตอบ

2.3 แบบทดสอบปากเปล่า หมายถึง แบบทดสอบที่ใช้การพูดโต้ตอบแทนการเขียน

3. แบ่งตามเวลาที่กำหนดให้ตอบ

3.1 แบบทดสอบที่จำกัดเวลาการตอบ หมายถึง แบบทดสอบที่ใช้เวลาน้อย

3.2 แบบทดสอบที่ไม่จำกัดเวลาการตอบ หมายถึง แบบทดสอบที่ใช้เวลาตอบมาก

4. แบ่งตามจำนวนผู้เข้าสอบ

4.1 แบบทดสอบรายบุคคล หมายถึง การสอบทีละคนมักเป็นการสอบภาคปฏิบัติ

4.2 แบบทดสอบเป็นชั้นหรือเป็นหมู่ หมายถึง การสอบทีละหลาย ๆ คน

5. แบ่งตามสิ่งเร้าของการถาม

5.1 แบบทดสอบทางภาษา หมายถึง แบบทดสอบที่ต้องอาศัยภาษาของสังคมนั้น ๆ เป็นหลัก ใช้กับผู้ที่อ่านออกเขียนได้

5.2 แบบทดสอบที่ไม่ใช้ภาษา หมายถึง แบบทดสอบที่ใช้รูปภาพ สัญลักษณ์หรือตัวเลข

6. แบ่งตามลักษณะของการใช้ประโยชน์

6.1 แบบทดสอบย่อย หมายถึง แบบทดสอบประจำบท หรือหน่วยการเรียน

6.2 แบบทดสอบรวม หมายถึง แบบทดสอบสรุปเนื้อหาที่เรียนผ่านมาตลอดภาคเรียน

7. แบ่งตามเนื้อหาของข้อสอบในฉบับ

7.1 แบบทดสอบอัตนัย หมายถึง แบบทดสอบที่มีเฉพาะคำถามนักเรียนต้องคิดหาคำตอบเอง

7.2 แบบทดสอบปรนัย หมายถึง แบบทดสอบที่มีทั้งคำถามและคำตอบเฉพาะคงที่

ในการวิจัยในครั้งนี้ผู้วิจัยสนใจแบบทดสอบเลือกตอบที่มีคำตอบที่ถูกต้องที่สุดเพียงคำตอบเดียว เพราะ เป็นแบบทดสอบที่นิยมใช้กันโดยทั่วไปในทุกวงการ สามารถใช้ได้กับเนื้อหาวิชาเกือบทุกลักษณะ การตรวจให้คะแนนได้ผลคงที่ ประหยัดเวลาและแรงงานและยังสามารถวัดได้ว่าผู้ตอบมีความรู้ความสามารถที่ถูกต้องสมบูรณ์ในเรื่องนั้นหรือไม่

วิธีการตรวจให้คะแนน

วิธีการตรวจให้คะแนนแบบ 0-1 มีข้อบกพร่อง หลายประการ (อนันต์ ศรีโสภา, 2532, หน้า 11) เช่น เป็นการยั่วยุหรือส่งเสริมให้นักเรียนที่ไม่มีความรู้ ความสามารถเดามากขึ้น แม้ว่าจะมีสูตรแก้การเดาก็มิได้ทำให้คะแนนหลังจากแก้การเดาแล้ว แทนความรู้ความสามารถที่แท้จริงของนักเรียนได้ การที่นักเรียนทำถูกได้ 1 คะแนน และทำผิดหรือไม่ทำเลยได้ 0 คะแนน ความจริงนักเรียนบางคนอาจจะมีความรู้ความสามารถไม่สมบูรณ์แต่ก็ไม่ถึงกับไม่รู้เลย คือ มีความรู้บางส่วน วิธีการแบบนี้ไม่สามารถวัดความรู้ความสามารถบางส่วนเหล่านั้นออกมาได้ อนึ่งข้อสอบแต่ละข้อระดับความยากไม่เท่ากัน คือ บางข้อยากมาก บางข้อยากน้อยแต่ก็ให้น้ำหนักของคะแนนเท่ากัน

หมดทุกข้อ คือ นักเรียนตอบถูกไม่ว่าข้อนั้นจะยากหรือง่ายก็ได้ 1 คะแนนเท่ากัน คะแนนรวมที่ได้จึงไม่สอดคล้องกับความรู้ความสามารถที่แท้จริงของนักเรียน จากข้อบกพร่องของวิธีการตรวจให้คะแนนแบบ 0-1 ดังกล่าว จึงได้มีนักวัดผลการศึกษาหลายท่านพยายามคิดหาวิธีเพื่อพัฒนาวิธีการตรวจให้คะแนนข้อสอบเลือกตอบให้มีประสิทธิภาพยิ่งขึ้น และเพื่อแสดงถึงตำแหน่งความรู้จริงของผู้ตอบช่วยลดความคลาดเคลื่อนจากการเดาสุ่มของ ผู้ตอบ ทำให้คะแนนมีความตรง และความเที่ยงมากขึ้น (เอมอร จังศิริพรปกรณ์, 2546, หน้า 12) แบ่งวิธีการให้คะแนนความรู้บางส่วนเป็นประเภทใหญ่ ๆ 4 วิธี ดังนี้

1. การให้น้ำหนักแก่ข้อสอบที่แตกต่างกัน (Differential item weighting)
2. การให้น้ำหนักแก่ตัวเลือกที่แตกต่างกัน (Differential option weighting)
3. การเปลี่ยนแปลงโครงสร้างของข้อสอบ (Changing the item structure)
4. การเปลี่ยนแปลงวิธีการตอบ (Changing the response method)

โดยแต่ละวิธีมีรายละเอียด ดังนี้

การให้น้ำหนักแก่ข้อสอบที่แตกต่างกัน (Differential item weighting)

หลักการของวิธีนี้ คือ ข้อสอบที่มีคุณภาพจะได้น้ำหนักมากกว่าข้อสอบที่ไม่มีคุณภาพ โดยใช้ผลที่ได้จากการวิเคราะห์ข้อสอบเป็นเกณฑ์ คือ ค่าความยาก ความสามารถความเที่ยง ความแปรปรวนหรือการพิจาณาของผู้เชี่ยวชาญ Wang and Stanley (1970, pp. 663-705) ให้ข้อสรุปว่าการให้น้ำหนักแก่ข้อสอบโดยใช้ทฤษฎีการวัดแบบดั้งเดิมทำให้น้ำหนักของคะแนนแต่ละข้อมีค่าไม่คงที่เปลี่ยนแปลงไปตามกลุ่มตัวอย่าง จึงควรให้คะแนนตามทฤษฎีการตอบสนองข้อสอบ การกำหนดน้ำหนักคะแนนที่เหมาะสมรายข้อนี้ยังวางอยู่บนพื้นฐานของวิธีการตอบและการให้คะแนนเฉพาะข้อที่ตอบถูกแบบเดิม ไม่ได้มีการพิจารณาให้คะแนนความรู้บางส่วนแก่ผู้ตอบ

การให้น้ำหนักแก่ตัวเลือกที่แตกต่างกัน (Differential option weighting)

หลักการของวิธีนี้ คือ ความรู้บางส่วนสามารถถูกวัดได้จากการให้น้ำหนักคะแนนแก่ตัวเลือกแบ่งวิธีการได้ 2 วิธี วิธีการแรกเป็นการให้น้ำหนักแก่ตัวเลือกโดยใช้วิจารณญาณของผู้เชี่ยวชาญหรือจากทฤษฎีของโครงสร้างความรู้ วิธีที่สองเป็นการให้น้ำหนักจากข้อมูลเชิงประจักษ์ที่ผ่านมาในอดีตหรือปัจจุบันโดยอาศัยสิ่งที่น่าสนใจของตัวเลือก เช่น มาตรฐานเฉลี่ยของแต่ละตัวเลือกที่ผู้สอบเลือกความสัมพันธ์ระหว่างตัวเลือกที่ผู้สอบเลือกกับคะแนนรวมทั้งหมด

1. วิธีการตรวจให้คะแนนของ Davis and Fifer (1959, pp. 160-161 อ้างถึงใน อนันต์ ศรีโสภา, 2532, หน้า 13) ได้ทดลองเกี่ยวกับ ค่าความเที่ยงและค่าความตรง ของการตรวจให้คะแนนแบบทดสอบความถนัดและแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน โดยให้คะแนนทุกตัวเลือก กลุ่มตัวอย่างที่ใช้เป็นนักเรียนศิษย์การบินที่ Lakeland Air Force Base จำนวน 65 คน วิธีการตรวจ

ให้คะแนนมี 2 วิธี คือ วิธีแบบธรรมดา หรือแบบ 0-1 และวิธีการให้น้ำหนักทุกตัวเลือกโดยให้ผู้เชี่ยวชาญเป็นผู้กำหนดน้ำหนักตัวเลือกแต่ละตัว พบว่า วิธีการตรวจให้คะแนนของเดวิสและฟิฟเฟอร์ เป็นวิธีหนึ่งที่จะช่วยให้คะแนนที่ได้จากแบบทดสอบมีความเที่ยงสูงขึ้น แต่มีข้อยุ่งยากในการปฏิบัติมาก

2. วิธีการตรวจให้คะแนนแฟรรีย์ (The proposed new method of Frary) (Frary, 1989) วิธีนี้เสนอโดยแฟรรีย์เขาต้องการจะปรับวิธีการให้คะแนนของครอสใหม่จากช่วงคะแนน 1 ถึง 7 เป็น -1 ถึง 1 คะแนน เพราะวิธีของครอสในกรณีที่ไม่มีการเลือกตัวเลือกใดเลยผู้ตอบจะได้คะแนนเท่ากับจำนวนตัวเลือกในข้อสอบ แฟรรีย์ เสนอว่า เขาควรได้ 0 คะแนน และใช้ระบบการแปลงคะแนนเชิงเส้นจากคะแนนของครอสใหม่ และได้วิเคราะห์ผลของการให้คะแนนแก่ความรู้บางส่วน การลงโทษผู้รู้ผิดและการเดาจากคะแนนเฉลี่ยรายข้อของวิธีการให้คะแนนแบบต่าง ๆ ให้ข้อเสนอว่า วิธีแบบแฟรรีย์ เป็นวิธีการให้คะแนนที่เหมาะกับการให้คะแนนความรู้บางส่วน เช่นเดียวกับวิธีการให้คะแนนของคูมส์และอาร์โนลด์ที่จะช่วยให้ความตรง ของการวัดสูงขึ้น เอมอร จังศิริพรปกรณ์ (2546) พบว่า การให้น้ำหนักคะแนนแก่ตัวเลือกมีประโยชน์บ้างในแง่ที่ช่วยเพิ่มค่าความตรง ภายใน การเปรียบเทียบความเที่ยงและความตรง ระหว่างแบบการให้น้ำหนักแก่ตัวเลือกกับแบบธรรมดา พบว่า มีค่าไม่คงที่ของความแตกต่างอย่างมีนัยสำคัญวิธีการเหล่านี้จึงไม่เป็นที่นิยม เนื่องจากการพัฒนาเพื่อให้ได้มาซึ่งน้ำหนักใช้ทรัพยากรมาก การคำนวณการให้คะแนนมีวิธีการยุ่งยากและยากในการอธิบายและตัดสินคะแนนแก่ผู้สอบ

การเปลี่ยนแปลงโครงสร้างของข้อสอบ (Changing the item structure)

วิธีนี้มีรูปแบบของโครงสร้างข้อสอบและ/ หรือคำอธิบายที่แตกต่างจากแบบเลือกตอบธรรมดา ดังนี้

1. ข้อสอบที่มี 2 ตัวเลือก คือ ถูกและผิด ผู้ตอบจะต้องเลือกคำตอบอย่างใดอย่างหนึ่ง คะแนนที่ได้ได้จากผลรวมของจำนวนคำตอบที่ถูกต้อง เนื่องจากการตอบแต่ละข้อใช้เวลาน้อย จึงต้องเพิ่มจำนวนข้อให้มากขึ้นมากกว่าข้อสอบแบบเลือกตอบธรรมดา

2. ข้อสอบที่มีตัวเลือกถูกหลาย ๆ ตัวเลือก โดยต้องมีคำชี้แจงอธิบายให้แก่ผู้สอบ ซึ่งอาจระบุจำนวนตัวเลือกที่ถูกหรือไม่ระบุก็ได้ คะแนนที่ได้เท่ากับผลรวมของการระบุตัวเลือกถูกได้ถูกต้อง ในบางกรณีมีการหักคะแนนการตอบที่ตอบตัวเลือกผิด โดยมีจุดประสงค์เพื่อลดการเดาและประเมินความรู้บางส่วน เช่น 1) วิธีการให้คะแนนของเดรชเซลและสมิท วิธีการให้คะแนนวิธีนี้ (Dressel & Schmidt, 1953, pp. 574-595) ได้ศึกษาวิธีการตอบ และวิธีการตรวจให้คะแนนที่ต่างกันเมื่อปรับปรุงอำนาจจำแนกของแบบทดสอบเลือกตอบโดยมีวิธีการตอบ 4 วิธี ดังนี้

2.1 เปิดโอกาสให้ผู้เลือกตอบอย่างเสรี (Free choice test) เป็นแบบทดสอบที่มี 5 ตัวเลือก โดยในแต่ละข้อผู้ตอบจะเลือกตอบกี่ตัวเลือกก็ได้และจะได้คะแนนตามเงื่อนไขถ้าข้อใดเว้นว่างไว้จะไม่ได้คะแนน

ตารางที่ 2-1 การเลือกตอบอย่างเสรี (Free choice test) โดยในแต่ละข้อผู้ตอบจะเลือกตอบกี่ตัวเลือกก็ได้และจะได้คะแนนตามเงื่อนไขถ้าข้อใดเว้นว่างไว้จะไม่ได้คะแนน

จำนวนคำตอบที่ทำเครื่องหมาย	**คะแนน**	
	ผู้ตอบทำเครื่องหมายถูกต้อง	**ผู้ตอบทำเครื่องหมายไม่ถูกต้อง**
1	4	-1
2	3	-2
3	2	-3
4	1	-4
5	0	-5

2.2 แบบบอกระดับความมั่นใจในการตอบ (A degree of certainly test) เป็นแบบที่ผู้เข้าสอบจะต้องแสดงระดับความมั่นใจในการตอบซึ่งมี 4 ระดับ แต่คำถามแต่ละข้อให้คำตอบได้เพียงคำตอบเดียว

2.3 แบบมีคำตอบถูกมากกว่าหนึ่ง (A multiple-answer test) แบบทดสอบชนิดนี้จะมีคำตอบถูกอยู่หลายคำตอบผู้เข้าสอบจะต้องทำเครื่องหมายทุกข้อที่ถูก

2.4 แบบมีคำตอบถูกสองคำตอบ (A two multiple-answer test) ผู้เข้าสอบจะต้องเลือกตอบเพียงสองคำตอบเท่านั้น

ผลการศึกษาพบว่า การตอบข้อสอบของนักเรียนเก่ง (พิจารณาจากคะแนนสอบ) จะแตกต่างจากนักเรียนปานกลางและอ่อนอย่างมีนัยสำคัญทางสถิติเมื่อใช้แบบทดสอบเลือกตอบชนิดตอบเสรี คือ กลุ่มเก่งจะทำเครื่องหมายคำตอบน้อยกว่ากลุ่มอ่อน ทั้งในข้อสอบที่มีระดับความยากสูง ปานกลาง และต่ำในทางตรงกันข้ามแบบทดสอบชนิดทดสอบความมั่นใจไม่ได้จำแนกนักเรียนเก่ง ปานกลาง อ่อน ได้ดีกว่าวิธีแรกเพราะเมื่อพิจารณาการตอบข้อสอบแล้วพบว่า ผู้เข้าสอบตอบข้อสอบที่มีความยากปานกลางและยากมากด้วยระดับความมั่นใจอย่างเดียวกัน

3. ข้อสอบที่มีกลุ่มของข้อสอบและกลุ่มของคำตอบที่เรียกว่าข้อสอบแบบจับคู่ โดยผู้สอบต้องเลือกคำตอบที่ถูกคู่กับคำถามที่กำหนดให้ ถ้าจำนวนคำถามเท่ากับจำนวนคำตอบ เรียกว่า

การจับคู่อย่างง่าย (Simple matching) ถ้าจำนวนคำตอบมากกว่าจำนวนคำถาม เรียกว่า การจับคู่แบบพหุ (Multiple matching) (Gulliksen, 1986) ความแตกต่างของสองแบบ คือ การจับคู่แบบพหุช่วยลดโอกาสในการเดามากกว่าแต่ข้อจำกัดที่สำคัญของวิธีการนี้ คือ ความยากในการสร้างข้อสอบให้มีความเหมาะสม พบว่า การจับคู่แบบพหุให้ค่าความตรง และค่าความเที่ยงมากกว่าแบบทดสอบเลือกตอบธรรมดา ความตรง และค่าความเที่ยงเพิ่มขึ้นเมื่อสัดส่วนของจำนวนตัวเลือกเพิ่มขึ้นและต้องการใช้เวลาในการทำข้อสอบมากกว่าปกติ

การเปลี่ยนแปลงวิธีการตอบ (Changing the response method)

วิธีการนี้แตกต่างจากวิธีการที่กล่าวมา คือ ใช้น้ำหนักที่ให้ผู้สอบเองซึ่งสะท้อนถึงความรู้ที่มีอยู่ในตัวผู้สอบ การประเมินตนเองนี้เป็นวิธีที่ดีที่สุดในการได้มาซึ่งสารสนเทศเกี่ยวกับความรู้ของผู้สอบในแต่ละข้อสอบจะให้สารสนเทศอย่างสมบูรณ์เกี่ยวกับน้ำหนัก วิธีการให้คะแนนและการควบคุมระดับความมั่นใจในการให้ ซึ่งจะช่วยลดการเดาและความคลาดเคลื่อนในการวัด วิธีนี้ใช้รูปแบบข้อสอบเหมือนแบบทดสอบชนิดเลือกตอบทั่วไปเพียงแต่คำสั่ง วิธีการตอบและการให้คะแนนแตกต่างจากแบบทดสอบชนิดเลือกตอบธรรมดาแต่ละวิธีการไม่ยุ่งยาก แบ่งได้ 6 วิธี ดังนี้ (เอมอร จังศิริพรปกรณ์, 2546, หน้า 14-19)

1. วิธีการแก้การเดา (Correction for guessing)

วิธีการนี้เป็นวิธีการที่ผู้สอบทำข้อสอบเหมือนกับแบบทดสอบชนิดเลือกตอบธรรมดาแต่ต่างกันที่การให้คะแนน วิธีการแก้การเดา มีวิธีที่น่าสนใจ ดังนี้

1.1 วิธีการแก้การเดาโดยการลงโทษ เป็นวิธีที่มีข้อตกลงเบื้องต้นว่าข้อที่ผู้สอบตอบผิดเป็นผลมาจากการเดา ดังนั้นข้อที่ผู้สอบตอบผิดแทนที่จะได้ 0 แต่จะให้คะแนนติดลบ -1/ (K-1) เมื่อ K คือ จำนวนตัวเลือกของแบบทดสอบ

1.1.1 วิธีตรวจให้คะแนนของลอร์ด (Lord, 1975, pp. 7-11) โดยใช้สูตรการให้คะแนนเพื่อแก้การเดาคำตอบ คือ X_c = R-W/ (K-1)

เมื่อ X_c คือ คะแนนที่ได้หลังจากการแก้การเดา

R คือ จำนวนข้อสอบที่ผู้สอบตอบถูก

W คือ จำนวนข้อสอบที่สอบตอบผิด

วิธีแก้การเดาสุ่มเหมาะกับแบบทดสอบความเร็วที่ผู้ตอบไม่ทัน แต่ถ้าเป็นทดสอบที่ไม่จำกัดเวลา ผู้ตอบอาจไม่ได้ตอบอย่างเดาสุ่มเพียงอย่างเดียวผู้ตอบอาจมีความรู้บางส่วนในการตัดตัวลวงบางตัวออกได้ ดังนั้นการใช้สูตรแก้การเดาอาจไม่เหมาะสมเพราะควรที่จะพิจารณาถึงการตอบโดยมีความรู้บางส่วนของผู้ตอบด้วย

1.1.2 วิธีการตรวจให้คะแนนของเดวิส (Davis, 1966, p. 79) วิธีต้องใช้สูตรการตรวจให้คะแนนเพื่อแก้การเดาของนักเรียน ดังนี้

$$X_c = R - \frac{W}{C-1}$$

เมื่อ X_c แทน คะแนนที่แก้การเดา

R แทน จำนวนข้อที่ตอบถูก

W แทน จำนวนข้อที่ตอบผิด

C แทน จำนวนตัวเลือกทั้งหมดในข้อสอบแต่ละข้อ

จากสูตรดังกล่าว เป็นการแก้การเดาโดยการหักสัดส่วนของจำนวนข้อสอบที่ตอบผิดออกจากจำนวนข้อที่ตอบถูก จากวิธีการตรวจให้คะแนนแบบนี้เป็นวิธีการที่ไม่ยุติธรรม เพราะถือว่าคำตอบที่ผู้ตอบผิดทุกข้อถือเป็นผลเนื่องมาจากการเดา โดยในความเป็นจริงผู้ตอบที่ตอบข้อสอบข้อนั้นผิด บางครั้งไม่ได้เกิดจากการเดาแต่เป็นเพราะได้ความรู้ที่ผิดหรือข้อสอบข้อนั้นมีตัวลวงที่สมเหตุสมผลจนทำให้ผู้สอบเข้าใจผิดไปได้

1.2 วิธีการให้รางวัลแก่ข้อที่ไม่ตอบ

วิธีการตรวจให้คะแนนของ Ebel (1972, p. 178) วิธีตรวจให้คะแนนวิธีนี้ เสนอขึ้นเพื่อแก้การเดาหรือการตอบถูกโดยบังเอิญของผู้ตอบ โดยไม่หักคะแนนข้อที่ผิดแต่จะเพิ่มคะแนนให้กับข้อที่ไม่ตอบ โดยถือว่าข้อที่ผู้ตอบเว้นไว้นั้นเป็นข้อที่ผู้ตอบไม่รู้หรือทำไม่ทัน แต่ถ้าผู้ตอบตอบโดยการเดาหรือทำเครื่องหมายลงในกระดาษคำตอบโดยที่ไม่ได้อ่านข้อสอบ สูตรการคะแนนมีดังนี้

$$X_c = R + \frac{0}{C}$$

เมื่อ X_c แทน คะแนนที่แก้การเดา

R แทน จำนวนข้อที่ตอบถูก

0 แทน จำนวนข้อที่เว้นไว้

C แทน จำนวนตัวเลือกทั้งหมดในข้อสอบแต่ละข้อ

2. วิธีการให้คะแนนจากการตัดตัวลวง (Elimination scoring)

วิธีการนี้เป็นวิธีการให้ผู้สอบตัดตัวลวงออกไปได้บ้างอย่างถูกต้องแม้จะไม่ทราบว่าคำตอบที่ถูกต้องคืออะไร มีวิธีการและงานที่สนับสนุน ดังนี้

2.1 วิธีการตอบและให้คะแนนแบบคูมบ์ (The Coombs response/ Scoring method) (Coombs et al., 1956) วิธีการนี้จำแนกระดับความรู้ของผู้สอบออกเป็น 5 ประเภท ดังนี้

2.1.1 มีความรู้เต็มที่ (Full knowledge) คือ ผู้ที่สามารตัดตัวลวงทั้งหมดออกได้

2.1.2 มีความรู้บางส่วน (Partial knowledge) คือ ผู้สามารถตัดตัวลวงบางตัวออกได้แต่ไม่ทั้งหมด

2.1.3 มีความรู้ผิดบางส่วน (Partial misinformation) คือ ผู้ที่สามารถตัดตัวลวงบางตัวได้บ้างและตัดตัวถูกออกด้วย

2.1.4 มีความรู้ผิดเต็มที่ (Full misinformation) คือ ผู้ที่ตัดตัวถูกออกเพียงตัวเดียว

2.1.5 ไม่มีความรู้ (Absence knowledge) คือ ผู้ที่ไม่ตอบหรือตัดตัวเลือกทุกตัวออกหมด

วิธีการตอบและการให้คะแนนของคูมบ์นี้ใช้รูปแบบของแบบทดสอบเลือกตอบที่มีคำตอบถูกเพียงคำตอบเดียวโดยให้ผู้สอบตัดตัวลวงที่ไม่ใช่คำตอบที่แน่ใจออกไปมากที่สุด การให้คะแนนผู้ตอบจะได้รับ 1 คะแนน ในแต่ละตัวลวงที่ตัดออกอย่างถูกต้องและในกรณีที่ผู้สอบมีความรู้ที่ผิดไปตัดตัวถูกออกโดยเข้าใจว่าเป็นตัวลวงจะถูกตัดคะแนน K-1 โดยที่ K คือ จำนวนตัวเลือก ดังนั้นช่วงคะแนนในแต่ละข้ออยู่ระหว่าง -(K-1) ถึง (K-1)

2.2 วิธีการตอบและการให้คะแนนแบบอาร์โนลด์ (The Arnold scoring/ Response method) (Arnold & Arnold, 1970, pp. 8-13) เป็นการให้คะแนนแบบทดสอบเลือกตอบซึ่งยอมให้คะแนนความรู้บางส่วนและควบคุมคะแนนที่เพิ่มขึ้นโดยโอกาสการเดาเป็น 0 ซึ่งเป็นการพัฒนามาจากทฤษฎีเกมเบื้องต้น (Game theory) โดยให้ผู้ตอบตัดตัวเลือกที่แน่ใจว่าเป็นตัวลวงหรือคำตอบที่ผิดออกไปให้มากที่สุดเท่าที่เขารู้ สามารถพัฒนาสูตรการให้คะแนน ดังนี้

$$C_d = (p)(d/k-d)$$

โดยที่ C_d คือ คะแนนเมื่อสมารถตัดตัวลวงบางตัวออกไปได้และคะแนนที่เพิ่มขึ้นจากการเดาเป็น 0

D คือ จำนวนตัวลวงที่ตัดออกได้ถูกต้อง

K คือ คะแนนลงโทษในกรณีที่ผู้สอบจัดตัวเลือกที่ถูกเข้าไว้เป็นตัวลวงในกรณีของ 4 ตัวเลือก คะแนนที่ได้เป็นผลจากการตอบของผู้สอบ ดังนี้

คะแนน C_0 ถ้าผู้เข้าสอบไม่สามารถตัดตัวลวงใด ๆ ได้เลย

คะแนน C_1 ถ้าผู้เข้าสอบไม่สามารถตัดตัวลวงออกได้ 1 ตัว

คะแนน C_2 ถ้าผู้เข้าสอบไม่สามารถตัดตัวลวงออกได้ 2 ตัว

คะแนน C_3 ถ้าผู้เข้าสอบไม่สามารถตัดตัวลวงออกได้ 3 ตัว

คะแนน -p ถ้าผู้สอบตัดตัวลวงถูกออกเพราะเข้าใจว่าเป็นตัวลวง

2.3 วิธีการตอบและให้คะแนนแบบครอส (The Cross response/ method: CRS) วิธีการนี้การตอบให้ผู้ตอบเลือกชุดย่อยของคำตอบที่เป็นตัวลวงออก เช่นเดียวกับวิธีการของคูมบ์และอาร์โนลด์ แต่การให้คะแนนแตกต่างกัน คือวิธีการให้คะแนนของครอสจะให้คะแนนรายข้อเฉลี่ยต่ำสุด เมื่อผู้ตอบตัดตัวเลือกที่เป็นคำตอบโดยเข้าใจว่าเป็นตัวลวงและตัดตัวลวงอีก K-2 ตัวโดยเข้าใจว่าเป็นชุดของตัวเลือกที่ผิด การให้คะแนนแบบครอสผู้สอบจะได้ 2 คะแนน ในแต่ละตัวลวงที่เขาตัดออกได้อย่างถูกต้องและจะได้ 1 คะแนน สำหรับแต่ละตัวเลือกที่ไม่ได้ตัดตัวลวง เมื่อคำตอบถูกยังไม่ได้ถูกตัด แต่เมื่อใดที่ผู้ตอบตัดตัวเลือกที่ถูกด้วย Misinformation จะได้คะแนนเฉพาะตัวเลือกที่ไม่ได้ตัดตัวเลือกละ 1 คะแนน ในกรณีที่ข้อสอบมี 4 ตัวเลือก คะแนนที่ยอมให้เกิดขึ้นได้มีค่าตั้งแต่ 1-7 คะแนน และคะแนนต่ำสุดจะเกิดจากการที่ผู้ตอบตัดตัวเลือกที่ถูกต้องหลังจากตัดตัวลวงมาแล้ว 2 ตัว (เอมอร จังศิริพรปกรณ์, 2546) เมื่อเปรียบเทียบการให้คะแนนแบบครอสกับแบบอื่น ๆ พบว่า คะแนนแบบครอสให้ค่าความตรง ต่ำกว่าวิธีการของคูมบ์และอาร์โนลด์

2.4 วิธีการเลือกชุดย่อยของคำตอบ (The subset selection method: SST)

2.4.1 เดรชเชลและสมิธ (Dressel & Schmidt, 1953, pp. 574-595) เป็นวิธีที่ตรงกันข้ามกับวิธีการให้คะแนนแบบตัดตัวลวง (ET) วิธีนี้ให้ผู้สอบเลือกชุดย่อยของคำตอบที่เขาเห็นว่าน่าจะมีตัวเลือกที่เป็นคำตอบถูกรวมอยู่ด้วย การให้คะแนนขึ้นอยู่กับจำนวนตัวเลือกที่มีอยู่ในชุดย่อยของคำตอบ ยิ่งน้อยเท่าใดคะแนนจะได้เพิ่มขึ้น ในกรณีที่ตัวเลือกที่ถูกต้องรวมอยู่ในชุดย่อยของคำตอบนั้นด้วย แต่ถ้าชุดย่อยนั้นไม่มีคำตอบที่ถูกต้องอยู่ด้วยคะแนนจะถูกหักออกไปเท่ากับขนาดของจำนวนตัวเลือกที่ผู้ตอบเลือก ในกรณีที่ผู้ตอบไม่เลือกตัวเลือกใดเลยจะได้ 0 คะแนน

2.4.2 วิธีการตรวจให้คะแนนของกิบบอนส์และคณะ (Gibbon, Olkin & Sobel, 1997, pp. 259-270) เสนอวิธีตรวจให้คะแนนที่เรียกว่า เทคนิคการเลือกกลุ่มย่อย (Subset selection technique: SST) ในแบบทดสอบเลือกตอบที่มีตัวเลือกถูกเพียงตัวเลือกเดียว โดยผู้ตอบสามารถเลือกตัวเลือกได้ตั้งแต่ 1 ตัว จนกระทั่งถึง 1-k ตัวเลือกที่ผู้ตอบคาดว่าจะมีคำตอบที่ถูกหรือตัวเลือก

ที่ถูกรวมอยู่ด้วย เมื่อ k คือ จำนวนตัวเลือกทั้งหมดของข้อสอบแต่ละข้อ การจะได้คะแนนของผู้ตอบขึ้นอยู่กับจำนวนตัวเลือกที่ผู้ตอบเลือก ยิ่งจำนวนตัวเลือกเพิ่มขึ้นค่าของคะแนนจะลดลง ส่วนการให้คะแนนยึดหลักว่าคะแนนที่คาดหวังที่ได้จากการเดาสุ่มของผู้ตอบในแต่ละข้อเป็น 0 เช่น กรณีที่แบบทดสอบเลือกตอบมี 5 ตัวเลือก และผู้ตอบอย่างสุ่มมา 2 ตัวเลือก ความน่าจะเป็นที่ตัวเลือกทั้ง 2 จะถูกคือ 2ใน 5 และความน่าจะเป็นที่จะตอบผิดคือ 3 ใน 5 ดังนั้น คะแนนที่คาดหวังคือ E (s) = 3 (2/ 5)-2 (3/ 5) = 0 ในกรณีนี้ถ้าตอบถูก หมายถึง มีคำตอบที่ถูกรวมอยู่จะได้ 3 คะแนน แต่ถ้าตอบผิด คือไม่มีคำตอบที่ถูกรวมอยู่ด้วย จะได้ -2 คะแนน โดยรูปแบบการให้คะแนนสามารถแสดงให้เห็นดังตารางต่อไปนี้

ตารางที่ 2-2 วิธีการตรวจให้คะแนนของกิบบอนส์และคณะ

จำนวนตัวเลือกที่ตอบ	คะแนนเมื่อมีตัวเลือกที่ถูก	คะแนนเมื่อไม่มีตัวเลือกที่ถูก
0	-	0
1	4	-1
2	3	-2
3	2	-3
4	1	-4
5	0	-5

จะเห็นว่า ในกรณีที่ผู้ตอบเลือกตอบแล้วไม่มีตัวเลือกที่ถูกรวมอยู่ด้วย คะแนนที่ได้จะมีค่าเป็นลบวิธีการนี้ เป็นวิธีลดค่าการเดาลงและยังสามารถวัดส่วนของความรู้ย่อย (Partial information) ของผู้ตอบในเรื่องนั้น ๆ

ทั้ง ET และ SST มีแนวโน้มในการเพิ่มคุณภาพของการจำแนกระดับความรู้ของผู้สอบและลดการเดา แต่อย่างไรก็ตามวิธีการยังมีความยุ่งยากและต้องใช้เวลาทำนานกว่าแบบทดสอบเลือกตอบธรรมดาซึ่งยังเป็นข้อจำกัดของการให้คะแนนและต้องมีการพัฒนาต่อไป

3. วิธีการของการทดสอบความน่าจะเป็น (Probability testing)

วิธีนี้เป็นวิธีที่มีความยืดหยุ่นและเปิดโอกาสให้ผู้สอบแสดงความรู้บางส่วนมากที่สุด โดยรายงานความน่าจะเป็นในตัวเลือกที่ผู้สอบคิดว่าเป็นคำตอบที่ถูกโดยมีตำแหน่งคะแนนที่เป็นไปได้ 101 ตำแหน่งช่วง (0-100) โดยแบ่งระดับความรู้เป็น 5 ประเภท ดังนี้ (Wallsten, Budescu & Zwick, 1993, pp. 176-190)

3.1 มีความรู้เต็มที่ (Full knowledge) คือ ให้ความน่าจะเป็น = 1 แก่คำตอบที่ถูกต้อง

3.2 มีความรู้บางส่วน (Partial knowledge) คือ ให้ความน่าจะเป็นมากกว่า 0 แต่ไม่ถึง 1 ($0 < p < 1$) แก่คำตอบที่ถูกต้อง

3.3 มีความรู้ผิดบางส่วน (Partial misinformation) คือ ให้ความน่าจะเป็นมากกว่า 0 แต่ไม่ถึง 1 ($0 < p < 1$) แก่คำตอบที่ผิด

3.4 มีความรู้ผิดเต็มที่ (Full misinformation) คือ ให้ความน่าจะเป็นเท่ากับ 0 แก่คำตอบที่ถูกต้อง

3.5 ไม่มีความรู้ (Absence knowledge) คือ ให้ความน่าจะเป็นแก่ทุกตัวเลือกเท่ากัน หรือไม่ตอบ

Michael (1968) พบว่า วิธีการให้คะแนนความน่าจะเป็น มีความตรง สูงกว่าวิธีเลือกตอบธรรมดา

Hambleton, Robert and Taub (1970) รายงานว่า ความเที่ยงสูงกว่าวิธีเลือกตอบธรรมดา

Koehler (1971) พบว่า มีความไม่แตกต่างกันของความตรง และความเที่ยงระหว่าง NC และ PT ซึ่งยังไม่พบข้อสรุปที่ชัดเจน

4. วิธีการแสดงความมั่นใจ (Confidence marking)

4.1 Dressel and Schmidt (1953, pp. 574-595) โดยที่ผู้สอบถูกถามให้แสดงความมั่นใจในคำตอบที่คิดว่าถูกโดยใช้ C-point scale (โดยทั่วไป $3 \leq c \leq 5$) หรืออาจเป็นการบอกความมั่นใจเป็นภาษา คือ ไม่แน่ใจ ค่อนข้างแน่ใจ แน่ใจมาก ซึ่งแบ่งระดับความรู้ออกเป็น 5 ระดับ

4.1.1 Full knowledge คือ การเลือกคำตอบได้ถูกต้องด้วยความมั่นใจระดับสูงที่สุด

4.1.2 Partial knowledge คือ การเลือกคำตอบได้ถูกต้องด้วยความมั่นใจระดับต่ำ

4.1.3 Partial misinformation คือ การเลือกคำตอบผิด ด้วยความมั่นใจระดับต่ำ

4.1.4 Full misinformation คือ การเลือกคำตอบผิด ด้วยความมั่นใจระดับสูงที่สุด

4.1.5 Absence knowledge คือ การไม่เลือกคำตอบเว้นว่างไว้

ผลการวิจัยพบว่า วิธีการนี้ให้ความเที่ยงของแบบทดสอบสูงกว่าการให้คะแนนแบบ 0,1 แต่วิธีการนี้ต้องการเวลาตอบและการให้คะแนนมากกว่าเดิม

4.2 Echternacth (1972, p. 217) วิธีการตรวจให้คะแนนโดยการบอกระดับความมั่นใจใช้กับแบบทดสอบถูกผิด ต่อมาก็เป็นแบบทดสอบเลือกตอบ วิธีการให้คะแนนแบบนี้จะศึกษาจากความรู้สึกของผู้ตอบที่มีต่อการทำข้อสอบว่าในการเลือกตอบแต่ละข้อนั้นผู้ตอบมีความมั่นใจมากน้อยเพียงใดว่าคำตอบที่เลือกจะเป็นคำตอบที่ถูกซึ่งแม้จะมีผู้ใดกำหนดน้ำหนักคะแนนต่างกัน

ออกไปแต่หลักการร่วมกันคือข้อที่ตอบถูกด้วยความมั่นใจจะกำหนดน้ำหนักคะแนนให้มากกว่าข้อที่ถูกแต่ไม่มั่นใจและการกำหนดน้ำหนักคะแนนนี้จะปรับคะแนนให้ติดลบสำหรับข้อที่ตอบผิด

5. วิธีการจัดอันดับอย่างสมบูรณ์ (Complete ordering)

เป็นวิธีเฉพาะของ การทดสอบความน่าจะเป็น (PT) คือ แทนที่จะแสดงให้เห็นความน่าจะเป็นในแต่ละตัวเลือกผู้สอบจะต้องจัดอันดับของตัวเลือก ซึ่งทำให้ง่ายต่อการนำไปใช้ทำให้สามารถจำแนกระดับความรู้ออกเป็น 3 ระดับ ดังนี้

5.1 Full knowledge คือ การจัดอันดับให้คำตอบถูกอยู่ในตำแหน่งที่สูงที่สุด

5.2 Partial knowledge คือการจัดอันดับให้คำตอบถูกอยู่ในตำแหน่งกลาง

5.3 Absence of knowledge คือ การจัดอันดับให้คำตอบถูกอยู่ในตำแหน่งต่ำที่สุด

6. วิธีการจัดอันดับเป็นบางส่วน (Partial ordering)

เป็นวิธีที่ผสมผสานระหว่าง วิธีการให้คะแนนแบบตัดตัวลวง (ET) และวิธีการจัดอันดับอย่างสมบูรณ์ (CO) โดยให้ผู้สอบจัดอันดับตัวเลือกที่ไม่สามารถตัดออกได้ วิธีนี้ เสนอโดย De Finetti (1965)

6.1 Full knowledge คือ การจัดอันดับแก่คำตอบถูกในตำแหน่งสูงที่สุด

6.2 Partial knowledge คือ การจัดอันดับแก่หลายตัวเลือก โดยคำตอบถูก ถูกจัดอันดับในตำแหน่งสูงที่สุด

6.3 Partial misinformation คือ การจัดอันดับแก่คำตอบถูกในตำแหน่งที่ต่ำกว่าตัวลวง

6.4 Full misinformation คือ การจัดอันดับตำแหน่งตัวลวง โดยตัดคำตอบถูกออก

6.5 Absence knowledge คือ การจัดอันดับแก่ตัวถูกในตำแหน่งต่ำที่สุดหรือไม่ตอบ

ในการวิจัยในครั้งนี้ผู้วิจัยสนใจวิธีการตรวจให้คะแนนแบบธรรมดาและวิธีการให้คะแนนแบบพิจารณาความมั่นใจ โดยจะเปรียบเทียบคุณภาพของแบบทดสอบที่มีวิธีการตรวจให้คะแนนทั้ง 2 วิธี ด้วยทฤษฎีการตอบสนองข้อสอบ (Item response theory) เนื่องจากทฤษฎีการวัดแบบดั้งเดิม มีข้อจำกัดในเรื่องความคลาดเคลื่อนในการวัดและค่าสถิติที่ได้แปรเปลี่ยนไปตามกลุ่มตัวอย่างที่ใช้ ในปัจจุบันนักวัดผลได้พัฒนาทฤษฎีการตอบสนองข้อสอบ (IRT) เพื่อแก้ปัญหาการวิเคราะห์ตามแนวทฤษฎีแบบดั้งเดิม การวิจัยด้านการวัดผลการศึกษาเกี่ยวกับทฤษฎีการตอบสนองรายข้อนอกจากจะเป็นการพัฒนาโมเดลและตรวจสอบโมเดลแล้วยังมีการวิจัยเกี่ยวกับการสร้างมาตรวัดคะแนนความสามารถ (Ability scores) ของผู้สอบ มีการพัฒนาคะแนนความสามารถในรูปฟังก์ชันของพารามิเตอร์ความสามารถของผู้สอบรูปต่าง ๆ โดยมีการกำหนดน้ำหนักคะแนนแบบต่าง ๆ และมีการวิจัยเกี่ยวกับการพัฒนาวิธีประมาณค่าพารามิเตอร์ความสามารถแบบต่าง ๆ ด้วย

ตอนที่ 2 ทฤษฎีการตอบสนองข้อสอบ

แนวคิดพื้นฐานทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบ (Item response theory: IRT) เป็นทฤษฎีการวัดที่อธิบายความสัมพันธ์ระหว่างคุณลักษณะภายในหรือความสามารถที่มีอยู่ในตัวบุคคลกับพฤติกรรมการตอบสนองข้อสอบของบุคคลนั้นว่ามีโอกาสตอบข้อสอบถูกมาน้อยเพียงไร เป็นทฤษฎีที่ได้แก้ไขข้อจำกัดหลายประการของทฤษฎีการวัดแบบดั้งเดิม (Classical test theory) มีพัฒนาการเริ่มต้นที่ Hambleton and Swaminatan (1985) บิเนทและไซมอน (Binet and Simon) เป็นนักจิตวิทยาได้สร้างกราฟที่แสดงความสัมพันธ์ระหว่างทักษะการคิดของเด็กกับอายุซึ่งเป็นที่มาของโค้งคุณลักษณะในระยะเวลาต่อมา ต่อมาในปี ค.ศ. 1936 ริชาร์ดสัน (Richarson) ได้แสดงความสัมพันธ์ระหว่างทฤษฎีตอบสนองข้อสอบกับทฤษฎีการทดสอบแบบดั้งเดิมและหาวิธีในการประมาณค่าพารามิเตอร์ตามทฤษฎีตอบสนองข้อสอบ ค.ศ. 1943 ลอว์ลี (Lawley) ได้เสนอบทความเกี่ยวกับปัญหาการสร้างและการเลือกข้อสอบโดยเสนอแนวคิดในรูปของโมเดล Normal ogive model ในปี ค.ศ. 1950 ราส์ช (Rasch) นักคณิตศาสตร์ชาวเดนมาร์กได้เสนอโมเดลราส์ช (Rasch model) แบบ 1 พารามิเตอร์โดยมีแนวคิดว่าความยากของข้อสอบเป็นสิ่งเดียวที่มีอิทธิพลต่อการตอบสนองข้อสอบ ต่อมาปี ค.ศ. 1952 ลอร์ด (Lord) ได้ขยายแนวคิดทฤษฎีตอบสนองข้อสอบโดยเสนอ Normal ogive function แบบ 2 พารามิเตอร์ที่มีการเพิ่มพารามิเตอร์ อำนาจจำแนก ในปี ค.ศ. 1968 เบิร์นบอม (Birbaum) ได้พัฒนาโมเดลโลจิสติก (Logistic model) ที่เป็นฟังก์ชันที่สามารถคำนวณ ได้ง่ายขึ้น ในปี ค.ศ. 1974 ลอร์ด (Lord) ได้เสนอแนวคิดโมเดลโลจิสติกแบบ 3 พารามิเตอร์ โดยเพิ่มพารามิเตอร์ในโอกาสการเดาข้อสอบในการประมาณค่าพารามิเตอร์ของผู้สอบด้วยโมเดลการตอบสนองข้อสอบในระยะเวลาแรก ๆ ใช้กับแบบทดสอบที่ให้คะแนนรายข้อเป็น 2 ระดับ หรือการให้คะแนนแบบ 0-1 เรียกว่า โมเดลการตอบสนองข้อสอบแบบทวิภาค (Dichotomous item response model) ต่อมาได้มีผู้พัฒนาเพื่อใช้กับแบบทดสอบและแบบวัดเจตคติที่ตรวจให้คะแนนเป็นแบบมาตรประมาณค่าหรือแบบพหุวิภาค (Polytomous) โมเดลในแนวนี้เรียกว่า (Polytomous item response model) (Murki, 1992; 1993) โดยมีความเชื่อที่สำคัญ ดังนี้

ทฤษฎีการตอบสนองข้อสอบ (Item response theory: IRT) นี้มีพื้นฐานความเชื่อว่าพฤติกรรมการตอบสนองต่อข้อสอบของผู้สอบซึ่งเป็นสิ่งที่สังเกตได้โดยตรงว่าถูกหรือผิดจะถูกกำหนดโดยคุณลักษณะภายในหรือความสามารถที่มีอยู่ในตัวบุคคล ซึ่งเป็นสิ่งที่ไม่สามารถสังเกตได้โดยตรง ทฤษฎีนี้อธิบายความสัมพันธ์ของคุณลักษณะภายใน หรือความสามารถที่มีในตัวบุคคล (Latent trait or ability) กับพฤติกรรมการตอบสนองข้อสอบ โดยใช้โค้งลักษณะข้อสอบ (Item characteristic curve: ICC) ซึ่งมีการกำหนดลักษณะของข้อสอบด้วยพารามิเตอร์ความยาก (b)

อำนาจจำแนก (a) และโอกาสการเดา (c) ทฤษฎีตอบสนองข้อสอบ (IRT) จึงอยู่บนพื้นฐานความคิดที่สำคัญ 2 ประการ คือ

1. ผลการตอบข้อสอบหรือข้อคำถามของผู้ตอบ สามารถอธิบายได้ด้วยความสามารถที่มีอยู่ภายในของผู้ตอบ

2. ความสัมพันธ์ระหว่างผลการตอบข้อสอบกับความสามารถที่มีอยู่ภายในสามารถอธิบายได้ด้วยฟังก์ชันลักษณะข้อสอบหรือโค้งลักษณะข้อสอบ (ICC) มีลักษณะเป็นฟังก์ชันทางคณิตศาสตร์ เรียกว่า ฟังก์ชันโลจิสติก (Logistic function) หรือใกล้เคียงฟังก์ชันปกติสะสม (Normal ogive function) ซึ่งเรียกว่า โมเดลการตอบสนองข้อสอบซึ่งมีหลายรูปแบบ เช่น โมเดล 1, 2, 3 พารามิเตอร์

สำหรับการเลือกโมเดลของทฤษฎีการตอบสนองข้อสอบมาใช้นั้น จะต้องเลือกให้เหมาะสมกับสถานการณ์ต่าง ๆ ตามข้อตกลงเบื้องต้นของแต่ละโมเดล โดยทฤษฎีนี้มีจุดเด่นหลายประการ ดังนี้ (Hambleton & Swaminathan, 1985 อ้างถึงใน กนกวรรณ รัตนธน, 2544)

1. ค่าพารามิเตอร์ของข้อสอบเป็นอิสระจากกลุ่มผู้สอบที่ใช้ในการประมาณค่า คือ ค่าพารามิเตอร์ของข้อสอบแต่ละข้อ ได้แก่ ค่าความยาก (b) ค่าอำนาจจำแนก (a) และค่าการเดา (c) เป็นค่าคงที่ไม่แปรเปลี่ยนไปตามกลุ่มของผู้สอบ ถึงแม้ว่าผู้สอบต่างกลุ่มกัน ค่าพารามิเตอร์ของข้อสอบจะมีค่าคงเดิมเสมอและในทำนองเดียวกันค่าพารามิเตอร์ความสามารถของผู้สอบเป็นอิสระจากข้อสอบที่ใช้ในการประมาณค่าจะไม่แปรเปลี่ยนไปตามค่าพารามิเตอร์ของข้อสอบ ไม่ว่าผู้สอบจะทำข้อสอบข้อใดหรือได้ทำข้อสอบเมื่อใดก็ตาม

2. การเปรียบเทียบความสามารถของผู้สอบ จะไม่ขึ้นอยู่กับคำถามในแบบสอบ ถึงแม้ว่าจะใช้ข้อคำถามต่างกัน ก็สามารถนำค่าความสามารถมาเปรียบเทียบกันได้ ทั้งนี้เนื่องจากค่าความสามารถที่ประมาณค่ามาได้นั้น เป็นคะแนนโลจิท (Logit) ซึ่งอยู่ในมาตรวัดเดียวกัน

3. การรายงานคุณภาพของข้อสอบในรูปค่าอินฟอร์เมชัน (Information) สามารถรายงานได้ทั้งเป็นรายข้อและทั้งฉบับ ค่าอินฟอร์เมชัน เป็นตัวบ่งชี้ถึงความถูกต้องแม่นยำ (Accuracy) ในการประมาณค่าความสามารถ สามารถนำมาใช้แทนค่าความเที่ยงและความคลาดเคลื่อนมาตรฐานในการวัดได้

หลักการของทฤษฎีตอบสนองข้อสอบ

ทฤษฎีการตอบข้อสอบ (Item response model) เชื่อเกี่ยวกับค่าพารามิเตอร์ของข้อสอบ (Item parameter) เช่น ค่าความยาก (b) ค่าอำนาจจำแนก (a) หรือค่าการเดา (c) ของข้อสอบแต่ละข้อว่าเป็นคุณลักษณะที่มีอยู่ประจำและคงที่พอสมควรในตัวข้อสอบนั้น ค่าพารามิเตอร์เหล่านี้ไม่ควรเปลี่ยนไปตามกลุ่มตัวอย่าง (Sample-free) และในทำนองเดียวกันค่าพารามิเตอร์ของผู้ตอบ

(Parameter) หรือคุณลักษณะ (Trait) หรือความสามารถ (Ability) ที่แท้จริงของผู้ตอบก็เป็นคุณลักษณะที่มีอยู่ในตัวผู้ตอบนั้นจริง จึงไม่ควรแปรเปลี่ยนไปตามชุดของข้อสอบที่เลือกใช้ แต่เนื่องจากความสามารถของผู้ตอบเป็นคุณลักษณะแฝง (Latent trait) ซึ่งไม่สามารถที่จะวัดหรือสังเกตได้โดยตรง จึงจำเป็นต้องใช้การพยากรณ์หรืออธิบายคุณลักษณะดังกล่าว โดยอาศัยผลที่ได้จากการตอบแบบทดสอบ ซึ่งเป็นสิ่งที่สามารถสังเกตหรือวัดได้ (Hambleton & Swaminathan, 1985, p. 9) นักวัดผลพยายามหาความสัมพันธ์ระหว่างผลที่ได้จากการตอบแบบทดสอบ หรือคะแนนกับปริมาณความสามารถของผู้ตอบแต่ละคน เพื่อเขียนมาเป็นโมเดลทางคณิตศาสตร์ ความสัมพันธ์ระหว่างปริมาณของการทำข้อสอบหรือคะแนนของผู้สอบ (Test performance) กับปริมาณความสามารถของผู้สอบ (Ability) สามารถเขียนในรูปของความสัมพันธ์ทั่วไป (Lord, 1982 อ้างถึงใน ตรึงใจ พูนผลอำนวย, 2534)

$$P = f(U_i \mid \theta_1, \theta_2, \theta_3, \ldots, \theta_k; \beta_j)$$

เมื่อ P แทน ผลการสอบ

f แทน ฟังก์ชัน

U_i แทน ผลการตอบข้อสอบ ตอบถูก $U_i = 1$ ตอบผิด $U_i = 0$

$\theta_1, \theta_2, \theta_3, \ldots, \theta_k$ แทน ความสามารถ (Ability หรือ Trait) ที่ 1, 2, 3, ..., k

β_j แทน ค่าพารามิเตอร์ของข้อสอบที่ j

ข้อตกลงเบื้องต้นของทฤษฎีตอบสนองข้อสอบ

1. ความเป็นมิติเดียว (Unidimension) หมายถึง ข้อสอบแต่ละข้อในแบบทดสอบจะต้องวัดความสามารถหรือคุณลักษณะเดียวกัน (Unidimensionality) หรือข้อสอบเหล่านี้มีความเป็นเอกพันธ์ (Homogeneous) การกำหนดเช่นนี้ก็เพื่อให้รูปแบบของทฤษฎีนี้มีความหมายซับซ้อนน้อยลง และง่ายต่อการแปลความหมายของคะแนนที่ได้จากแบบทดสอบ วิธีการตรวจสอบว่าแบบทดสอบนั้นวัดมิติเดียวกันหรือไม่นั้นทำได้โดยการวิเคราะห์องค์ประกอบ (Factor analysis) แล้วสังเกตค่าไอเกน (Eigen value) ค่าสูงสุดว่าแตกต่างจากค่าอื่น ๆ อย่างชัดเจนหรือไม่ ดังที่ Hambleton and Swaminathan (1985, pp. 16-17) กล่าวไว้ว่า ข้อตกลงนี้ไม่เข้มงวดนัก ขอให้มีลักษณะเด่นที่จะวัดองค์ประกอบใดองค์ประกอบหนึ่งก็ใช้ได้

2. ความเป็นอิสระในการตอบข้อสอบ (Local independence) หมายถึง โอกาสในการตอบข้อสอบแต่ละข้อได้ถูกต้องเป็นอิสระจากกัน นั่นคือ การตอบข้อสอบข้อใดข้อหนึ่งได้ถูกหรือผิดจะไม่มีผลต่อการตอบข้ออื่น ๆ

3. โค้งลักษณะข้อสอบ (Item characteristic curve) เป็นฟังก์ชันทางคณิตศาสตร์ที่แสดงถึงความสัมพันธ์ระหว่างโอกาสในการตอบข้อสอบข้อนั้นถูกต้องกับระดับความสามารถที่วัดได้โดยชุดของข้อสอบหรือแบบทดสอบนั้น (Hambleton & Swaminathan, 1985, p. 9) จะเห็นว่าโอกาสผู้สอบตอบข้อสอบถูกจะขึ้นอยู่กัลป์โค้งลักษณะของข้อสอบในแต่ละข้อมีคุณสมบัติไม่ แปรเปลี่ยนไปตามกลุ่มตัวอย่างหรือตัวผู้สอบ ดังนั้นโอกาสการตอบข้อสอบถูกจึงไม่แปรเปลี่ยนด้วยผู้วิจัยอธิบายคุณลักษณะเบื้องต้นของแบบทดสอบตามทฤษฎีการตอบข้อสอบ (Item response model) ด้วยการแสดงความเป็นมิติเดียว (Unidimension) โดยการวิเคราะห์องค์ประกอบ (Factor analysis) แล้วสังเกตค่าไอเกน (Eigen value)

โมเดลการตอบสนองข้อสอบ (Item response model)

ทฤษฎีการตอบสนองข้อสอบเป็นทฤษฎีที่อธิบายถึงความสัมพันธ์ระหว่างปริมาณของการกระทำหรือคะแนนของผู้สอบกับปริมาณความสามารถของผู้สอบ ซึ่งแสดงได้ดังสมการ (สำเริง บุญเรืองรัตน์, 2529)

$$P = f(\theta)$$

เมื่อ P แทน ผลการสอบ (Performance)

f แทน ฟังก์ชัน (Function)

θ แทน ความสามารถ (Ability หรือ Trait)

จากความสัมพันธ์ในสมการดังกล่าว เป็นการแสดงความสัมพันธ์ทั่ว ๆ ไปยังไม่เฉพาะเจาะจงจึงมีผู้สนใจศึกษาและพัฒนาเพื่อให้ได้สมการที่สามารถอธิบายลักษณะเฉพาะของข้อสอบแต่ละข้อที่สามารถอธิบายความสัมพันธ์ของสองตัวแปรนี้ได้เด่นชัดขึ้นคุณลักษณะของโมเดลการตอบสนองข้อสอบเป็นระบบความสัมพันธ์ระหว่างโอกาสตอบข้อสอบถูก (P_i) กับความสามารถที่มีอยู่ภายในผู้ตอบ (θ) ในรูปของโค้งลักษณะข้อสอบ (ICC) ซึ่งมีลักษณะเป็นฟังก์ชันโลจิสติก (Logistic function) หรือฟังก์ชันปกติสะสม (Normal ogive function) บางครั้งเรียกว่าโมเดลโลจิสติกหรือโมเดลปกติสะสมใช้ฟังก์ชันปกติสะสมแสดงความสัมพันธ์ระหว่างผลการตอบข้อสอบกับความสามารถของผู้สอบส่วนโมเดลโลจิสติกใช้ฟังก์ชัน โลจิสติกแสดงความสัมพันธ์ระหว่างผลการตอบกับความสามารถ ซึ่งฟังก์ชันทั้งสองให้ผลลัพธ์ของการประมาณค่าใกล้เคียงกันมากแต่ฟังก์ชันโลจิสติกมีลักษณะของสูตรทางคณิตศาสตร์ และวิธีคำนวณง่ายสะดวกกว่า นอกจากนี้โมเดลโลจิสติกยังมีความทนทานต่อความคลาดเคลื่อนที่เกิดขึ้นกับผู้สอบที่มีความสามารถสูงจะตอบข้อสอบได้ดีกว่า ทำให้โมเดลโลจิสติกเป็นที่นิยมกันมากในการนำไปใช้จริง

โมเดลการตอบสนองข้อสอบแบบทวิวิภาค (Dichotomous item response model)

โค้งลักษณะข้อสอบแสดงความสัมพันธ์ระหว่างระดับความสามารถกับโอกาสการตอบข้อสอบได้ถูกต้อง ทำให้เกิดโมเดลของโค้งลักษณะของข้อสอบขึ้นหลายโมเดล แต่รูปแบบที่นิยมใช้กันแพร่หลาย คือ รูปแบบโลจิสติก (Logistic model)

1. โมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ หรือราส์ชโมเดล (One-parameter model, Rasch model) มีข้อตกลงว่าข้อสอบทุกข้อมีค่าอำนาจจำแนกเท่ากันดังสมการ

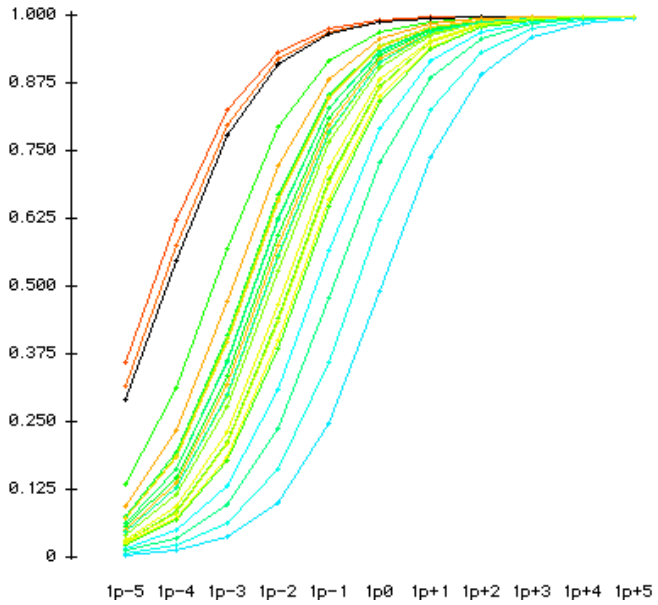
$$P_i(\theta) = \frac{e^{D(\theta-\beta_i)}}{1-e^{D(\theta-\beta_i)}} \quad (i = 1,2,\ldots,n)$$

เมื่อ $P_i(\theta)$ = ความน่าจะเป็นที่ผู้ตอบซึ่งมีความสามารถ θ จะตอบข้อสอบที่ i ได้ถูกต้อง

β_i = ค่าพารามิเตอร์ความยากของข้อสอบที่ i ซึ่งเป็นค่าที่แสดงตำแหน่งของ ICC ณ จุด θ ที่มีโอกาสตอบข้อสอบถูก 0.50

D = ค่าคงที่ของ Scale factor เพื่อปรับค่าฟังก์ชันโลจิสติกให้ใกล้เคียงกับค่าที่ได้จากฟังก์ชันรูปตัวเอสปกติ มีค่าประมาณ 1.7

e = 2.718



ภาพที่ 2-1 โค้งลักษณะข้อสอบแบบ 1 พารามิเตอร์ (Behrens, 1997)

โค้งลักษณะข้อสอบ (ICC) สำหรับโมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ มีลักษณะที่สำคัญดังนี้

b_i = มีค่าแปรเปลี่ยนตามลักษณะของข้อสอบแต่ละข้อ

a_i = มีค่าคงที่

$c_i = 0$

2. โมเดลการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ (Two-parameter model) มีโค้งลักษณะข้อสอบที่เขียนด้วยฟังก์ชันโลจิสติก ดังสมการ (Hambleton, Swaminathan & Roger, 1991, p. 14)

$$P_i(\theta) = \frac{e^{Da_i(\theta-\beta_i)}}{1+e^{Da_i(\theta-\beta_i)}} \quad (i = 1, 2, \ldots, n)$$

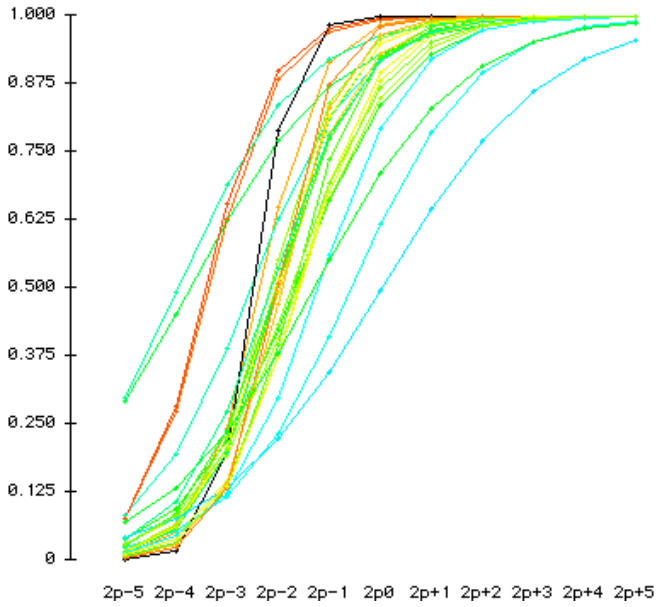
เมื่อ $P_i(\theta)$ = ความน่าจะเป็นที่ผู้ตอบซึ่งมีความสามารถ θ จะตอบข้อสอบที่ i ได้ถูกต้อง

b_i = ค่าพารามิเตอร์ความยากของข้อสอบที่ i ซึ่งเป็นค่าที่แสดงตำแหน่งของ ICC ณ จุด θ ที่มีโอกาสตอบข้อสอบถูก 0.50

a_i = ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบที่ i ซึ่งเป็นค่าความชันของ ICC ณ ตำแหน่ง b_i

e = 2.718

D = 1.70



ภาพที่ 2-2 โค้งลักษณะข้อสอบแบบ 2 พารามิเตอร์(Behrens, 1997)

โค้งลักษณะข้อสอบ (ICC) สำหรับโมเดลการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ มีลักษณะที่สำคัญดังนี้

b_i = มีค่าแปรเปลี่ยนตามลักษณะของข้อสอบแต่ละข้อ

a_i = มีค่าแปรเปลี่ยนตามลักษณะของข้อสอบแต่ละข้อ

$c_i = 0$

3. โมเดลการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ (Three-parameter model) มีโค้งลักษณะข้อสอบที่เขียนด้วยฟังก์ชันโลจิสติก ดังสมการ (Hambleton et al., 1991, p. 18)

$$P_i(\theta) = C_i(1-C_i)\frac{e^{Da_i(\theta-\beta_i)}}{1+e^{Da_i(\theta-\beta_i)}} \quad (i = 1, 2, \ldots, n)$$

เมื่อ $P_i(\theta)$ = ความน่าจะเป็นที่ผู้ตอบซึ่งมีความสามารถ θ จะตอบข้อสอบที่ i ได้ถูกต้อง

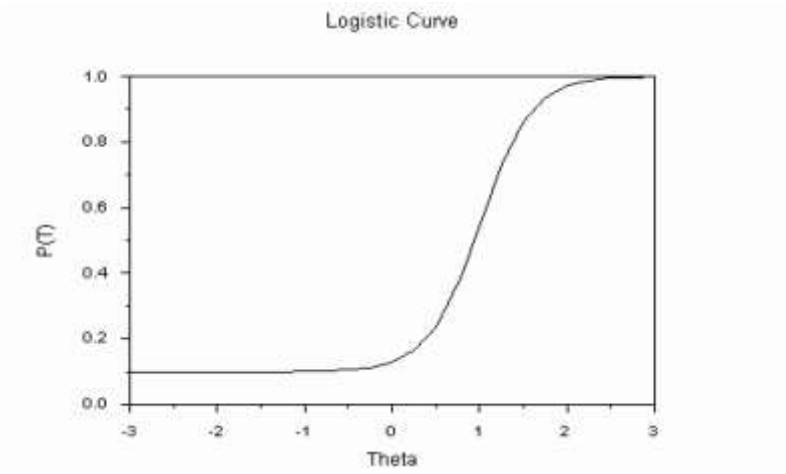
b_i = ค่าพารามิเตอร์ความยากของข้อสอบที่ i ซึ่งเป็นค่าที่แสดงตำแหน่งของ ICC ณ จุด θ ที่มีโอกาสตอบข้อสอบถูก $\frac{1+C_i}{2}$

a_i = ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบที่ i ซึ่งเป็นค่าความชันของ ICC ณ ตำแหน่ง b_i

c_i = ค่าพารามิเตอร์โอกาสเดาข้อสอบได้ถูก

e = 2.718

D = 1.70



ภาพที่ 2-3 โค้งลักษณะข้อสอบแบบ 3 พารามิเตอร์ (Brannick, 2007)

โค้งลักษณะข้อสอบ (ICC) สำหรับโมเดลการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ มีลักษณะที่สำคัญดังนี้

b_i = มีค่าแปรเปลี่ยนตามลักษณะของข้อสอบแต่ละข้อ

a_i = มีค่าแปรเปลี่ยนตามลักษณะของข้อสอบแต่ละข้อ

c_i = มีค่าแปรเปลี่ยนตามลักษณะของข้อสอบแต่ละข้อ

โมเดลการตอบสนองข้อสอบแบบพหุวิภาค (Plytomous item response model)

ทฤษฎีการตอบสนองของข้อสอบ (IRT) เสนอฟังก์ชั่นสำหรับอธิบายความสัมพันธ์ระหว่างความน่าจะเป็นของการตอบข้อสอบได้ถูกต้อง (หรือการเลือกรายการคำตอบ) กับความสามารถของผู้ตอบ (หรือคุณลักษณะภายในของผู้ตอบ) และคุณลักษณะของข้อสอบ โดยใช้โมเดล 2 ประเภทตามลักษณะการตรวจให้คะแนนคำตอบ ได้แก่ โมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนน 2 ค่า (Binary or dichotomous IRT models) และโมเดลการตอบสนองของข้อสอบแบบตรวจให้คะแนนมากกว่า 2 ค่า (Polytomous IRT models)

การตอบสนองข้อสอบแบบตรวจให้คะแนน 2 ค่า (Dichotomous IRT models) มีข้อจำกัดที่สำคัญคือ การตรวจให้คะแนนรายข้อต้องเป็นแบบ 2 ค่า (Binary) เช่น การตรวจให้คะแนนแบบ 0 หรือ 1 ถูกหรือผิ ด เห็นด้วยหรือไม่เห็นด้วย เป็นต้น ในขณะที่เครื่องมือทาง

การศึกษาและจิตวิทยาหลายประเภท แต่ละข้อคำถามมักมีหลายรายการคำตอบที่กำหนดลำดับหรือน้ำหนักคะแนนต่างกันเพราะน่าจะให้สารสนเทศและความเที่ยงจากการตอบที่สูงกว่าการตรวจให้คะแนนแบบ 2 ค่า เช่น แบบวัดเจตคติ แบบวัดบุคลิกภาพ แบบวัดความสนใจในอาชีพเป็นต้น จึงได้มีการพัฒนาโมเดลการตอบสนองข้อสอบสำหรับใช้กับการตรวจให้คะแนนรายข้อมากว่า 2 ค่า ซึ่งปัจจุบันมีรูปแบบของเครื่องมือที่นิยมใช้กัน เช่น มาตรประมาณค่า (Rating scale) การตรวจให้คะแนนความรู้บางส่วน (Partial credit) การตรวจให้คะแนนตามลำดับขั้นตอนของรายการหลายคำตอบ (Ordered-response categories) เป็นต้น

โมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่า 2 ค่า (Polytomous IRT models) เป็นโมเดลความสัมพันธ์ที่ไม่เชิงเส้นตรงระหว่างความสามารถของผู้ตอบกับโอกาสของการเลือกตอบแต่ละรายการคำตอบที่กำหนดให้ ผู้พัฒนาโมเดลการตอบสนองข้อสอบสำหรับการตรวจให้คะแนนรายข้อมากกว่า 2 ค่าไว้หลากหลายโมเดล ได้แก่

1. Graded response model (GRM)
2. Modified graded-response model (M-GRM)
3. Partial credit model (PCM)
4. Generalized partial credit model (G-PCM)
5. Rating scale model (RSM)
6. Nominal response model (NRM)

Generalized partial credit model

Muraki (1992; 1993) ได้พัฒนา Generalized partial credit model (G-PCM) ทำให้โมเดล PCM มีลักษณะเป็นโมเดลทั่วไป โดยยอมให้ข้อคำถามแต่ละข้อสามารถมีค่าพารามิเตอร์ความชันแตกต่างกันได้ G-PCM มีลักษณะเป็นโมเดลทั่วไปของ PCM และใช้หลักการคำนวณความน่าจะเป็นของการตอบแต่ละระดับขั้นการตอบโดยตรงแบบขั้นตอนเดียว (Direct IRT method) โมเดลใน G-PCM ลักษณะคำถามในแต่ละข้อ (i) อธิบายด้วยค่าพารามิเตอร์ระดับความยากของขั้นการตอบ (δ_{ij}) และค่าพารามิเตอร์ความชันของข้อคำถาม (α_i) ฟังก์ชันทางคณิตศาสตร์ของโมเดลนี้ จึงเขียนได้โดยแทนค่าพารามิเตอร์ความชันลงในลงในสูตรของโมเดล PCM ดังนี้

$$P_{ix}(\theta)=\frac{\exp\left[\sum_{j=0}^{x}\alpha_i\left(\theta-\delta_{ij}\right)\right]}{\sum_{r=0}^{m_i}\left[\exp\sum_{j=0}^{r}\alpha_i\left(\theta-\delta_{ij}\right)\right]}$$

เมื่อ $\sum_{j=0}^{0}\alpha_i\left(\theta-\delta_{ij}\right)=0$

$P_{ix}(\theta)$ = ความน่าจะเป็นที่ผู้ตอบซึ่งมีคุณลักษณะ θ จะตอบข้อ i ด้วยการเลือกหรือสามารถทำรายการคำตอบขั้นที่ x จากจำนวน m_i ขั้น (Step)

δ_{ij} = ค่าพารามิเตอร์ระดับความยากของขั้นการตอบที่ j ในข้อ (Item step difficulty) เมื่อ j = 1, 2, …, m_i

α_i = ค่าพารามิเตอร์ความชันของข้อคำถามที่ i

ค่าระดับความยากของขั้นการตอบ (δ_{ij}) ซึ่งเป็นค่าสเกล θ ตรงตำแหน่งที่ตัดกันของโค้งรายการคำตอบ มีความหมายเหมือนกับโมเดล PCM ส่วนค่าความชัน (α_i) มีความหมายต่างจากโมเดลการตอบสนองข้อสอบแบบมากกว่า 2 ค่า ค่าอำนาจจำแนกของข้อคำถามขึ้นอยู่กับผลร่วมกันระหว่างพารามิเตอร์ความชัน และการกระจายของพารามิเตอร์ Threshold ของรายการคำตอบ (สำหรับโมเดล GRM และ M-GRM) หรือตำแหน่งจุดตัดระหว่างรายการคำตอบ (สำหรับโมเดล PCM และG-PCM) ในโมเดล G-PCM ค่าพารามิเตอร์แสดงความชัน แสดงถึงระดับความผันแปรของรายการคำตอบระหว่างข้อเมื่อ θ ของผู้ตอบเปลี่ยนไป ถ้า α_i มีค่ามากกว่า 1.00 แสดงว่าโค้งรายการคำตอบของข้อนั้นมีความสูงชันมากกว่าโค้งรายการคำตอบในโมเดล PCM

ในการวิจัยในครั้งนี้ ผู้วิจัยใช้แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ โดยวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (Item response theory) แบบทวิภาค ที่ใช้โมเดลโลจิสติก (Logistic) 2 พารามิเตอร์ และโมเดลการให้คะแนนความรู้บางส่วน (Genralized partial credit model: GPCM)

ฟังก์ชันสารสนเทศของข้อสอบและแบบทดสอบ (Item and test information)

1. ฟังก์ชันสารสนเทศของข้อสอบ (Item information) ฟังก์ชันสารสนเทศของข้อสอบเป็นดัชนีผสมที่สร้างจากดัชนีคุณลักษณะของข้อสอบหลายลักษณะประกอบด้วยค่าพารามิเตอร์ความยากค่าพารามิเตอร์อำนาจจำแนกและค่าความแปรปรวนของคะแนนรายข้อเพื่อใช้บ่งชี้คุณภาพของข้อสอบ (Birnbaum, 1968 อ้างถึงใน ศิริชัย กาญจนวาสี, 2545, หน้า 54)

$$I_i(\theta) = \frac{[P_i'(\theta)]^2}{P_i(\theta)Q_i(\theta)} \qquad , i = 1, 2, \ldots, k$$

เมื่อ $I_i(\theta)$ = ค่าฟังก์ชันสารสนเทศหรือค่าสารสนเทศที่ได้รับจากข้อสอบที่ i สำหรับตอบที่มีความสามารถ θ

$P'_i(\theta) = P'_i$ = ความชันของฟังก์ชันการตอบสนองข้อสอบที่ i ณ ตำแหน่งความสามารถ θ

$P_i(\theta) = P_i =$ ความน่าจะเป็นที่ผู้ตอบซึ่งมีความสามารถ θ จะตอบข้อสอบที่ i ได้ถูกต้อง

$Q_i = Q_i = 1\text{-}P_i(\theta)$

สูตรคำนวณฟังก์ชันสารสนเทศของข้อสอบ สำหรับโมเดลการตอบสนองข้อสอบโลจิสแบบ 1-พารามิเตอร์ 2-พารามิเตอร์ และ 3-พารามิเตอร์ ตลอดจนค่าความชันของฟังก์ชันการตอบสนอง ณ ตำแหน่ง θ (P'_i) ค่าสารสนเทศสูงสุดของข้อสอบ [$I_i(\theta)_{max}$] และตำแหน่งของ θ ที่มีสารสนเทศของข้อสอบสูงสุด (θ_{max}) ดังตาราง

ตารางที่ 2-3 การคำนวณค่าฟังก์ชันสารสนเทศ ค่าสารสนเทศสูงสุดของข้อสอบและตำแหน่ง θ ที่มีสารสนเทศของข้อสอบสูงสุด สำหรับโมเดลการตอบสนองข้อสอบโลจิสติก 1, 2, 3 พารามิเตอร์ (ศิริชัย กาญจนวาสี, 2545, หน้า 55)

ค่าประมาณ	1-parameter	2-parameter	3-parameter
$I_i(\theta)$	$D^2P_iQ_i$	$D^2a_i^2P_iQ_i$	$D^2a_i^2Q_i(P_i-c_i)^2/(1-c_i)^2$
P_i'	$D\,P_iQ_i$	$D\,a_i^2P_iQ_i$	$D\,a_iQ_i(P_i-c_i)\,/(1-c_i)$
$I_i(\theta)_{max}$	$\frac{1}{4}D^2$	$\frac{1}{4}D^2a_i^2$	$D^2a_i^2\frac{\left[1-20c_i-8c_i^2+(1+8c_i)^{3/2}\right]}{8(1-c_i^2)}$
θ_{max}	b_i	b_i	$b_i+\frac{1}{Da_i}\left[In1+\frac{(1+c_i)^{1/2}}{2}\right]$

จากสูตรของฟังก์ชันสารสนเทศของข้อสอบ เมื่อพิจารณาเทียบกับค่าพารามิเตอร์ a, b และ c ของข้อสอบ พอสรุปเป็นแนวทางโดยทั่วไปได้ ดังนี้

ค่าสารสนเทศของข้อสอบจะสูงขึ้น สำหรับผู้สอบมีความสามารถ θ ใกล้กับค่าพารามิเตอร์ b ของข้อสอบ และค่าสารสนเทศของข้อสอบจะลดลงสำหรับผู้สอบที่มีความสามารถ θ ไกลจากค่าพารามิเตอร์ b ของข้อสอบ

ค่าสารสนเทศของข้อสอบโดยทั่วไปจะมีค่าสูงขึ้น ถ้าค่าพารามิเตอร์ a ของข้อสอบมีค่ามากขึ้น

ค่าสารสนเทศของข้อสอบมีค่าสูงขึ้น เมื่อค่าพารามิเตอร์ c ของข้อสอบเข้าใกล้ 0

$I_i(\theta)$ จะมีค่าสูงสุด ณ ตำแหน่ง θ_{max} ถ้า $C_i = 0$, $\theta_{max} = b$ แต่ถ้า $C_i > 0$, $\theta_{max} > b$

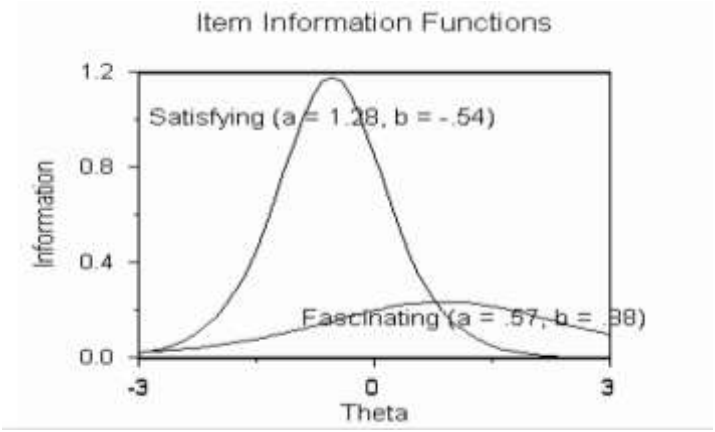
2. ฟังก์ชันสารสนเทศของแบบทดสอบ (Test Information) การวิเคราะห์ตาม IRT จะใช้แบบแผนการตอบสนองแบบทดสอบเป็นรายข้อในการประมาณค่าความสามารถของผู้สอบ ดังนั้นการประเมินคุณภาพของแบบทดสอบ จึงสามารถพิจารณาจากความถูกต้องแม่นยำใน การประมาณความสามารถของผู้ตอบโดยใช้ดัชนีที่เรียกว่า สารสนเทศของแบบทดสอบ (Test information: $I(\theta)$) ซึ่งเป็นค่าฟังก์ชันสารสนเทศของแบบทดสอบอันเกิดจากผลรวมเชิงพีชคณิตของค่าฟังก์ชันสารสนเทศของข้อสอบแต่ละข้อรวมเข้าด้วยกันทั้งฉบับ ณ ตำแหน่ง θ เดียวกันดังสูตร

$$I(\theta) = \sum_{i=1}^{k} I_i(\theta), \; i = 1, 2, \ldots, k$$

เมื่อ $I(\theta)$ = ค่าฟังก์ชันสารสนเทศหรือค่าที่ได้รับจากแบบทดสอบสำหรับผู้ตอบที่มีความสามารถ θ

ค่าฟังก์ชันสารสนเทศของข้อสอบแต่ละข้อ จึงมีส่วนอย่างเป็นอิสระจากกันต่อค่าสารสนเทศของแบบทดสอบ ลักษณะเช่นนี้ไม่ได้เกิดกับทฤษฎีการทดสอบแบบดั้งเดิม เนื่องจากค่าความยากและค่าอำนาจจำแนกของข้อสอบแต่ละข้อต่างก็ส่งผลต่อค่าความเที่ยงของแบบทดสอบทั้งฉบับแต่ไม่สามารถคำนวณค่าของแต่ละข้อได้อย่างเป็นอิสระจากกัน ดังนั้นคะแนนที่ได้จึงขึ้นอยู่กับลักษณะเฉพาะของกลุ่มข้อสอบ และแบบทดสอบเฉพาะฉบับที่เลือกมาใช้

เนื่องจากค่าสารสนเทศมีความสัมพันธ์ผกผันกับความคลาดเคลื่อนมาตรฐานของการประมาณค่าดังนั้นถ้าค่าสารสนเทศของแบบทดสอบมีค่าสูงในช่วง θ ใดก็จะมีความถูกต้องแม่นยำสูงในการประมาณค่าความสามารถของผู้ตอบในช่วง θ นั้น ๆ โดยมีความคลาดเคลื่อนมาตรฐานของการประมาณค่าต่ำ



ภาพที่ 2-4 โค้งสารสนเทศของข้อสอบ (Brannick, 2007)

3. ประสิทธิภาพสัมพัทธ์ (Relative efficiency)

3.1 การเปรียบเทียบค่าฟังก์ชันสารสนเทศระหว่างแบบทดสอบที่วัดคุณลักษณะเดียวกันสามารถทำได้โดยคำนวณประสิทธิภาพสัมพัทธ์ (Relative efficiency: RE (θ)) ซึ่งเป็นอัตราส่วนระหว่างค่าฟังก์ชันสารสนเทศของแบบทดสอบต่างฉบับ ณ ตำแหน่ง θ เดียวกัน ดังนี้

$$RE(\theta) = \frac{I_A(\theta)}{I_B(\theta)}$$

เมื่อ $I_A(\theta)$ = ค่าฟังก์ชันสารสนเทศของแบบทดสอบฉบับ A ณ ตำแหน่งบนสเกล ความสามารถร่วมกันระดับ θ

$I_B(\theta)$ = ค่าฟังก์ชันสารสนเทศของแบบทดสอบฉบับ B ณ ตำแหน่งบนสเกล ความสามารถร่วมกันระดับ θ

การแปลความหมายของค่าดัชนีประสิทธิภาพสัมพัทธ์ สามารถนำไปช่วยคัดเลือกแบบทดสอบได้ดังนี้

$RE(\theta)$ = 1 แสดงว่าแบบทดสอบทั้งสองฉบับมีประสิทธิภาพเท่ากันสำหรับผู้ตอบที่มีระดับความสามารถ θ

$RE(\theta) >$ 1 แสดงว่าแบบทดสอบฉบับ A มีประสิทธิภาพสูงกว่าแบบทดสอบฉบับ B สำหรับผู้ตอบที่มีระดับความสามารถ θ

$RE(\theta) <$ 1 แสดงว่าแบบทดสอบฉบับ A มีประสิทธิภาพต่ำกว่าแบบทดสอบฉบับ B สำหรับผู้ตอบที่มีระดับความสามารถ θ

สมมติว่า $I_A(\theta) = 25.0$ และ $I_B(\theta) = 20.0$

$$\therefore RE(\theta) = \frac{25}{20} = 1.25$$

แสดงว่า แบบฉบับ A มีประสิทธิภาพสูงกว่าแบบทดสอบฉบับ B สำหรับใช้สอบกับผู้สอบที่มีระดับความสามารถ θ หรือ อาจกล่าวได้ว่าแบบทดสอบฉบับ A ทำหน้าที่เสมือนเป็นแบบทดสอบที่มีความยาวกว่าแบบทดสอบฉบับ B 25% หรือแบบทดสอบฉบับ สามารถตัดข้อออกได้อีก 20% ก็ยังคงมีความแม่นยำในการประมาณค่าความสามารถ ณ ตำแหน่ง θ ได้ดีเท่ากับแบบทดสอบฉบับ B

3.2 การเปรียบเทียบค่าฟังก์ชันสารสนเทศเฉลี่ยระหว่างแบบทดสอบที่วัดคุณลักษณะเดียวกัน สามารถทำได้โดยคำนวณประสิทธิภาพสัมพัทธ์เฉลี่ย (RAI) ซึ่งเป็นอัตราส่วนระหว่างค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบทดสอบต่างฉบับ ณ ทุกตำแหน่ง θ ดังนี้

$$RAI(\theta; X, Y) = \frac{AI(\theta, X)}{AI(\theta, Y)}$$

เมื่อ $AI(\theta, X)$ = ค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบทดสอบฉบับ X ณ ทุกตำแหน่ง θ

$AI(\theta, Y)$ = ค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบทดสอบฉบับ Yณ ทุกตำแหน่ง θ

4. การประมาณค่าสารสนเทศของข้อสอบและแบบทดสอบ การประมาณค่าสารสนเทศของข้อสอบ (Item information function: IIF) หรือ $I_j(\theta)$ และฟังก์ชันสารสนเทศของแบบทดสอบ (Test information function: TIF) หรือ $I(\theta)$ โมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนน 2 ค่า สามารถนำแนวคิดมาประยุกต์ใช้กับโมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่า 2 ค่า ได้เช่นเดียวกัน กล่าวคือ เมื่อพิจารณาข้อสอบหรือข้อคำถามแต่ละข้อ เราสามารถแปลงโค้งรายการคำตอบให้เป็นโค้งสารสนเทศของข้อสอบหรือข้อคำถามได้ จากโค้งสารสนเทศของข้อสอบ เมื่อนำมารวมกัน (ณ ตำแหน่ง θ เดียวกัน) ทำให้ได้โค้งสารสนเทศของแบบทดสอบ สามารถเขียนในรูปของฟังก์ชันสารสนเทศของข้อสอบและแบบทดสอบได้ (De Ayala, Dodd & Koch, 1992) ดังสมการ

ฟังก์ชันสารสนเทศของข้อสอบ (Item information function: IIF) หรือ $I_j(\theta)$

$$\text{IIF หรือ } I(\theta)=\sum_{i=1}^{k} I_i(\theta)$$

เมื่อ $I_i(\theta)$ คือ ฟังก์ชันสารสนเทศของข้อสอบ ตั้งแต่ข้อ 1 ถึง ข้อ i

5. ความคลาดเคลื่อนมาตรฐานของการประมาณค่า (Standard error of estimation: SE (θ)) เป็นค่าส่วนเบี่ยงเบนมาตรฐานของการแจกแจงความน่าจะเป็นของค่าประมาณความสามารถที่แท้จริง (θ) ซึ่งเป็นค่าสัดส่วนผกผันกับความถูกต้องแม่นยำของการประมาณค่าความสามารถหรือค่าฟังก์ชันสารสนเทศของแบบทดสอบ ดังสูตร (ศิริชัย กาญจนวาสี, 2545, หน้า 59) ดังนี้

$$SE(\theta)=\frac{1}{\sqrt{I(\theta)}}$$

เมื่อ SE (θ) แทน ความคลาดเคลื่อนมาตรฐานของการประมาณค่าสำหรับผู้ตอบที่มีความสามารถ θ

I (θ) แทน สารสนเทศของแบบทดสอบสำหรับผู้ตอบที่มีความสามารถ θ

การสอบทุกครั้งย่อมมีความคลาดเคลื่อนในการประมาณค่าความสามารถ (θ) ความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถ มีค่าเท่ากับส่วนกลับของรากที่สองของค่าฟังก์ชันสารสนเทศของแบบทดสอบ ดังนั้น ค่าฟังก์ชันสารสนเทศของแบบทดสอบมีค่าสูง ความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถจะต่ำ นั่นคือ ถ้าค่าฟังก์ชันสารสนเทศของแบบทดสอบมีค่าสูง ณ ผู้สอบระดับความสามารถ θ ใด ความคลาดเคลื่อนมาตรฐานของการประมาณค่าของผู้สอบ ณ ระดับ θ จะต่ำ หรือมีความแม่นยำในการประมาณค่าความสามารถของผู้สอบ

การวิจัยด้านการวัดผลการศึกษาเกี่ยวทฤษฎีการตอบสนองรายข้อ นอกจากจะเป็นการพัฒนาโมเดลและตรวจสอบโมเดลแล้ว ยังมีการวิจัยเกี่ยวกับการสร้างมาตรวัดคะแนนความสามารถ (Ability score) ของผู้สอบมีการพัฒนาคะแนนความสามารถในรูปฟังก์ชันของพารามิเตอร์ความสามารถของผู้สอบรูปต่าง ๆ โดยมีการกำหนดน้ำหนักคะแนนแบบต่าง ๆ และมีการวิจัยเกี่ยวกับการพัฒนาวิธีการประมาณค่าพารามิเตอร์ความสามารถต่าง ๆ ด้วยทฤษฎี IRT ที่ใช้ในการวิเคราะห์ข้อสอบที่มีการตรวจให้คะแนนหลายค่า เช่น โมเดลเกรดเรสพอนส์ โมเดลเจเนอรอลไลซ์ พาเชียล เครดิต โมเดลพาเชียล เครดิต เป็นโมเดลที่รู้จักกันแพร่หลายและให้

ผลการวิเคราะห์ที่คล้ายคลึงกันเป็นส่วนใหญ่ทั้งในแบบทดสอบที่ข้อสอบตรวจให้คะแนนหลายค่า ผู้วิจัยมีการประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบ (IRT) ใช้โมเดลการให้คะแนนความรู้บางส่วน (Generalizeed partial creadit model: GPCM) โดยแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา และให้คะแนนแบบพิจารณาความมั่นใจ วิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (Item response theory) แบบทวิภาค ที่ใช้โมเดลโลจิสติก (Logistic) 2 พารามิเตอร์ และโมเดลการให้คะแนนความรู้บางส่วน (Genralized partial credit model: GPCM) ด้วยโปรแกรมสำเร็จรูป IRTPRO Student ในการประมาณค่าพารามิเตอร์ความยาก ค่าพารามิเตอร์อำนาจจำแนก และ ฟังก์ชันสารสนเทศ

ตอนที่ 3 การตรวจสอบคุณภาพของแบบทดสอบ

อำนาจจำแนก (Discrimination index)

ศิริชัย กาญจนวาสี (2556) กล่าวว่า อำนาจจำแนก หมายถึง ความสามารถข้องข้อสอบในการจำแนกความแตกต่างระหว่างผู้สอบที่มีผลสัมฤทธิ์ต่างกัน

ล้วน สายยศ และอังคณา สายยศ (2543, หน้า 197) กล่าวว่า ค่าอำนาจจำแนกข้อสอบอิงเกณฑ์นั้นจะเป็นค่าอำนาจจำแนกระหว่างกลุ่มที่ยังไม่ได้รับการเรียนรู้หรือกลุ่มที่ยังไม่รู้ (Nonmaster) กับกลุ่มที่ได้รับการเรียนรู้แล้วหรือที่รู้แล้ว (Master) ข้อสอบอิงเกณฑ์ไม่เน้นที่ค่าอำนาจจำแนก เนื่องจากแบบทดสอบอิงเกณฑ์จะใช้ในการวัดผลที่ใช้การเรียนการสอนแบบมีระบบ (Systematic instruction) เช่น การเรียนแบบรอบรู้ นักเรียนทุกคนจะเรียนรู้หมด นั่นคือ มีคะแนนเต็มหรือใกล้เต็มทุกคน เมื่อหาค่าอำนาจจำแนกจะมีค่าเป็น 0 หรือใกล้ 0 ทันที ดังนั้น ดัชนีค่าอำนาจจำแนกของข้อสอบอิงเกณฑ์ ควรมีค่าเท่ากับหรือมากกว่า 0

การคำนวณค่าอำนาจจำแนกของข้อสอบ มีอยู่ 2 วิธี

วิธีที่ 1 คำนวณจากผลการทดสอบซ้ำกัน 2 ครั้ง โดยการเอาแบบทดสอบฉบับเดิมไปสอบกับนักเรียนแล้วคำนวณจากสัดส่วนของนักเรียนที่ทำข้อนั้นถูกก่อนได้รับการสอน (Preinstruction) กับหลังได้รับการสอน (Postinstruction) ซึ่ง Cox and Vargas (1996; Tucker & Vargas, 1971; Vargas, 1969 อ้างถึงใน ล้วน สายยศ และอังคณา สายยศ, 2543) ได้เสนอสูตรดังนี้

$$PPDI = \frac{R_{pos}}{n} - \frac{R_{pre}}{n}$$

เมื่อ PPDI แทน ดัชนีค่าอำนาจจำแนกของข้อสอบอิงเกณฑ์ (Pre-to-post-different index)

R_{pos} แทน จำนวนนักเรียนที่ทำข้อสอบถูกหลังได้รับการสอน

R_{pre} แทน จำนวนนักเรียนที่ทำข้อสอบถูกก่อนได้รับการสอน

n แทน จำนวนนักเรียนที่ทำข้อสอบทั้งหมด

วิธีที่ 2 คำนวณจากผลการทดสอบเพียงครั้งเดียว วิธีนี้เอาแบบทดสอบไปสอบกับนักเรียนหลังจากนักเรียนรู้แล้ว แล้วนำข้อมูลมาคำนวณจากสัดส่วนของนักเรียนที่ทำข้อสอบถูกของกลุ่มที่สอบผ่านเกณฑ์กับกลุ่มที่สอบยังไม่ผ่านเกณฑ์ คำนวณโดยใช้สูตรของเบรนแนน (Brennan)

$$B = \frac{U}{n_1} - \frac{L}{n_2}$$

เมื่อ B แทน ดัชนีค่าอำนาจจำแนกของแบบทดสอบอิงเกณฑ์

U แทน จำนวนนักเรียนที่ทำข้อสอบถูกของกลุ่มที่สอบผ่านเกณฑ์

L แทน จำนวนนักเรียนที่ทำข้อสอบถูกของกลุ่มที่สอบไม่ผ่านเกณฑ์

n_1 แทน จำนวนักเรียนที่สอบผ่านเกณฑ์

n_2 แทน จำนวนักเรียนที่สอบไม่ผ่านเกณฑ์

ตารางที่ 2-4 เกณฑ์การพิจารณาค่าอำนาจจำแนก (Brennan et al., 1974, pp. 244-261)

ดัชนีค่า B	**ความหมาย**
+1.00	จำแนกผู้รอบรู้/ ไม่รอบรู้/ ได้ถูกต้องทุกคน
0.50-0.99	จำแนกผู้รอบรู้/ ไม่รอบรู้/ ได้ถูกต้องเป็นส่วนใหญ่
0.20-0.49	จำแนกผู้รอบรู้/ ไม่รอบรู้/ ได้ถูกต้องบางส่วน
0.00-0.19	จำแนกผู้รอบรู้/ ไม่รอบรู้/ ได้ถูกต้องน้อยมาก
ติดลบ	จำแนกผู้รอบรู้/ ไม่รอบรู้/ ตรงข้ามกับความตรง

ในการวิจัยครั้งนี้ ผู้วิจัยหาค่าอำนาจจำแนกของแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาเคมี กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 4 ภาคเรียนที่ 2 ปีการศึกษา 2560 โดยใช้ดัชนีบี (B-index) ตามวิธีของเบรนแนน (Brennan) (ล้วน สายยศ และอังคณา สายยศ, 2539, หน้า 198)

ความยาก (Difficulty)

ศิริชัย กาญจนวาสี (2556) กล่าวว่า ความยากง่ายของข้อสอบ หมายถึงสัดส่วนของจำนวนคนที่ตอบข้อสอบข้อนั้นถูก

ณัฏฐภรณ์ หลาวทอง (2551) กล่าวว่า ความยาก หมายถึง สัดส่วนของจำนวนคนที่ตอบข้อสอบนั้นได้ถูกต้องต่อจำนวนผู้เข้าสอบทั้งหมด

ความยาก เป็นคุณสมบัติของข้อสอบที่บอกให้ทราบว่าข้อสอบข้อนั้นมีคนตอบถูกมากหรือน้อย ถ้ามีคนตอบถูกมากข้อสอบนั้นก็ง่ายและถ้ามีคนตอบถูกน้อยข้อสอบนั้นก็ยาก ถ้ามีคนตอบ ผิดบ้างหรือมีคนตอบถูกปานกลาง ข้อสอบนั้นก็มีความยากปานกลาง ข้อสอบที่ดีควรมีความยาก พอเหมาะควรมีคนตอบถูกไม่ต่ำกว่า 20 คนและไม่เกิน 80 คน จากผู้สอบ 100 คน ค่าความยากหาได้ โดยการนำจำนวนคนที่ตอบถูกหารด้วยจำนวนคนที่ตอบทั้งหมด (ล้วน สายยศ และอังคณา สายยศ, 2538, หน้า 192-220; พิชิต ฤทธิ์จรูญ, 2544, หน้า 142-154)

กล่าวโดยสรุป ความยากของข้อสอบ หมายถึง จำนวนร้อยละหรือสัดส่วนของคนที่ตอบถูกในข้อนั้นเมื่อเปรียบเทียบกับจานวนคนทั้งหมด วิเคราะห์โดยสูตร ดังนี้

$$P = \frac{R}{N}$$

เมื่อ p แทน ค่าความยาก

R แทน จำนวนนักเรียนที่ตอบถูก

N แทน จำนวนนักเรียนทั้งหมด

หรือ

$$P = \frac{P_H + P_L}{2n}$$

เมื่อ P แทน ค่าความยาก

P_H แทน จำนวนคนที่ตอบถูกในกลุ่มสูง

P_L แทน จำนวนคนที่ตอบถูกในกลุ่มต่ำ

n แทน จำนวนคนในกลุ่มสูงหรือกลุ่มต่ำ

ค่าความยากมีค่าตั้งแต่ 0.00 ถึง 1.00 โดยทั่วไปข้อสอบที่มีความยากพอเหมาะควรมีค่าตั้งแต่ 0.20-0.80 ซึ่งมีรายละเอียด ดังนี้

$0.80 < P \leq 1.00$ แสดงว่า เป็นข้อสอบง่ายมาก ควรตัดทิ้งหรือปรับปรุง

$0.60 < P \leq 0.80$ แสดงว่า เป็นข้อสอบค่อนข้างง่าย (ดี)

$0.40 < P \leq 0.60$ แสดงว่า เป็นข้อสอบง่ายปานกลาง (ดีมาก)

$0.20 < P \leq 0.40$ แสดงว่า เป็นข้อสอบค่อนข้างยาก (ดี)

$0.00 < P \leq 0.20$ แสดงว่า เป็นข้อสอบยากมาก ควรตัดทิ้งหรือปรับปรุง

ถ้าข้อสอบข้อใดมีผู้ตอบถูกหมด แสดงว่า ข้อนั้นง่ายมาก มีค่า P = 1.00 แต่ถ้า ข้อสอบข้อใดมีผู้ตอบผิดหมด แสดงว่า ข้อนั้นยากมาก มีค่า P = 0.00

ความเที่ยง (Reliability)

Ebel (1965) กล่าวว่า ความเที่ยง หมายถึง ความคงเส้นคงวาของคะแนนจากแบบทดสอบที่มุ่งวัดในสิ่งที่ต้องการวัด

Gronlund (1976) กล่าวว่า ความเที่ยงหมายถึง ความคงที่ของผลการวัดจากเครื่องมือชนิดเดียวกันที่ทำการวัดซ้ำ

ศิริชัย กาญจนวาสี (2556) กล่าวว่า ความเที่ยง หมายถึงความคงที่หรือความคงเส้นคงวาของผลที่ได้จากการวัดซ้ำ หรือค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนจากแบบทดสอบคู่ขนาน 2 ชุด ซึ่งสอบโดยผู้สอบกลุ่มเดียวกัน

นักการศึกษาหลายท่านได้ให้ความหมายของความเที่ยงว่า ความเที่ยง เป็นความคงที่หรือความคงเส้นคงวาของแบบวัดหรือแบบทดสอบที่สามารถวัดบุคคลเดียวกัน ในเรื่องเดียวกัน ต่างเวลาและต่างโอกาสกัน ก็ให้ผลเหมือเดิมไม่ว่าจะวัดหรือทดสอบกี่ครั้งก็ตาม (ล้วน สายยศ และอังคณา สายยศ, 2543, หน้า 209) ได้กล่าวถึง ความเที่ยงของแบบทดสอบอิงเกณฑ์จากการทดสอบเพียงครั้งเดียว การหาค่า ความเที่ยงที่ได้จากการทดสอบเพียงครั้งเดียวมีวิธีหาหลายวิธีด้วยกัน ดังนี้

วิธีที่ 1 ความเที่ยงจากสูตรของ Livingston (1972; ล้วน สายยศ และอังคณา สายยศ, 2543, หน้า 236)

$$r_{cc} = \frac{\sigma^2(KR.20)+(\mu-KC)^2}{\sigma^2+(\mu-KC)^2}$$

เมื่อ r_{cc} แทน ความเที่ยงของแบบทดสอบอิงเกณฑ์

σ^2 แทน คะแนนความแปรปรวนของแบบทดสอบอิงเกณฑ์

K แทน จำนวนข้อของแบบทดสอบ

C แทน สัดส่วนของเกณฑ์ที่ผ่าน

μ แทน คะแนนเฉลี่ยของแบบทดสอบอิงเกณฑ์

KR-20 แทน ความเที่ยงของแบบทดสอบที่หาจากสูตร KR-20

วิธีที่ 2 ความเที่ยงจากสูตรไบโนเมียล (Binomial formula) ของ Lovett (1978) (ล้วน สายยศ และอังคณา สายยศ, 2539, หน้า 238)

$$r_{cc} = \frac{K\Sigma X_i - \Sigma X_i^2}{\{(K-1)+\Sigma(X_i - C)^2\}}$$

เมื่อ r_{cc} แทน ความเที่ยงของแบบทดสอบอิงเกณฑ์

K แทน จำนวนข้อสอบ

$\sum Xi$ แทน ผลรวมของคะแนนทั้งหมด

ΣX_i^2 แทน ผลรวมทั้งหมดของคะแนนแต่ละตัวยกกำลังสอง

X_i แทน คะแนนของนักเรียนแต่ละคน

C แทน คะแนนจุดตัดของแบบทดสอบอิงเกณฑ์

วิธีที่ 3 ความเที่ยงโดยการวิเคราะห์ความแปรปรวน Lovett (1978) ได้เสนอสูตรที่ใช้แนวความคิดของ Hoyt (1941) หาความเที่ยง ดังนี้ (ล้วน สายยศ และอังคณา สายยศ, 2539, หน้า 239)

$$r_{cc} = 1 - \frac{MS_E}{MS_p}$$

เมื่อ r_{cc} แทน ความเที่ยงตามวิธีของฮอยท์

MS_E แทน คะแนนความแปรปรวนของความคลาดเคลื่อน (Error*)*

MS_p แทน คะแนนความแปรปรวนระหว่างคน (Between people*)*

วิธีที่ *4* ความเที่ยงโดยวิธีแบ่งครึ่งแบบทดสอบ วิธีนี้จะหาความเที่ยงโดยใช้สูตรสเปียร์แมนบราวน์ (Spearman-Brown) แล้วใช้สูตรปรับของ Angoff (1953) ดังนี้ (ล้วน สายยศ และอังคณา สายยศ, 2539, หน้า 243)

$$r_{cc} = \frac{r_{12}\sigma^2}{(\sigma_1 + r_{12}\sigma_2)(\sigma_2 + r_{12}\sigma_1)}$$

เมื่อ r_{cc} แทน ความเที่ยงของแบบทดสอบ

σ^2 แทน คะแนนความแปรปรวนของแบบทดสอบทั้งฉบับ

σ_1 แทน ความเบี่ยงเบนของข้อสอบข้อคู่

σ_2 แทน ความเบี่ยงเบนของข้อสอบข้อคี่

R_{12} แทน สหสัมพันธ์ระหว่างข้อสอบข้อคู่กับข้อคี่

วิธีที่ 5 การหาค่าความเที่ยงโดยสูตรของ Harris (1972) ดังนี้ (ล้วน สายยศ และอังคณา สายยศ, 2543, หน้า 243)

$$r_{cc} = \frac{SS_b}{SS_b + SS_w}$$

เมื่อ r_{cc} แทน ความเที่ยงของแบบทดสอบ

SS_b แทน ผลรวมกาลังสองระหว่างกลุ่ม

SS_w แทน ผลรวมกาลังสองภายในกลุ่ม

4. ความเที่ยงที่คำนึงจุดประสงค์เชิงพฤติกรรมของแบบทดสอบอิงเกณฑ์การหาความเที่ยงของแบบทดสอบอิงเกณฑ์ที่คำนึงถึงจุดประสงค์เชิงพฤติกรรมนั้น Raju (1982) ได้เสนอสูตรคอนเจนเนอริค (Congeneric) ดังนี้ (ล้วน สายยศ และอังคณา สายยศ, 2543, หน้า 244)

$$r_{cc} = \frac{\sigma^2 + (\mu - C)^2 - \Sigma[\sigma_i^2 + (\mu - C)^2]}{[\sigma^2 + (\mu - C)^2[1 - \Sigma\lambda_i^2)}$$

เมื่อ σ^2 แทน ความแปรปรวนของทั้งฉบับ

μ แทน คะแนนเฉลี่ยทั้งฉบับ

λ_i แทน $\frac{K_i}{\Sigma K_i}$ เมื่อ K_i จำนวนข้อสอบในแต่ละจุดประสงค์เชิงพฤติกรรม

C แทน คะแนนจุดตัดของแต่ละจุดประสงค์

ในการวิจัยครั้งนี้ ผู้วิจัยหาหาค่าความเที่ยงของแบบทดสอบของแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชา เคมี กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 4 ภาคเรียนที่ 2 ปีการศึกษา 2560ในการวิจัยครั้งนี้ ผู้วิจัยใช้วิธีการคำนวณหาค่าความเที่ยงของแบบทดสอบอิงเกณฑ์จากโดยสูตรของ Livingston (1972)

ความตรง (Validity)

Waugh and Gronlund (2013) กล่าวว่า ความตรง หมายถึง คุณภาพของเครื่องมือที่ทำให้ผลการประเมินเป็นตัวแทนที่ดีและมีความสัมพันธ์กับเนื้อหาที่ประเมิน

ศิริชัย กาญจนวาสี (2556) กล่าวว่า ความตรง หมายถึง ความถูกต้องแม่นยำของเครื่องมือในการวัดสิ่งที่ต้องการจะวัด

นักการศึกษาหลายท่านได้ให้ความหมายของความตรง ไว้สอดคล้องกันว่า ความตรงเป็นคุณสมบัติของแบบวัดหรือแบบทดสอบที่สามารถวัดหรือประเมินผลได้ตรงตามลักษณะหรือจุดประสงค์ที่ต้องการได้อย่างถูกต้องและแม่นยำ (Bloom, 1967, p. 468; Gronlund. 1976, p. 65; Anastasi. 1988, p. 139; อนันต์ ศรีโสภา.2525, p. 43; ล้วน สายยศ และอังคณา สายยศ, 2543, หน้า 246; บุญเชิด ภิญโญอนันตพงษ์, 2545, หน้า 89)

คุณภาพของเครื่องมือด้านความตรง (Validity) หมายถึง การวัดได้ในสิ่งที่ต้องการวัดวัดได้ตรงตามวัตถุประสงค์ซึ่งความตรง แบ่งได้ 4 ประเภท (ล้วน สายยศ และอังคนา สายยศ, 2538) มีหลายประเภท ดังนี้

ความตรงตามเนื้อหา (Content validity)

ความตรงตามเนื้อหา (Content validity) หมายถึงการที่เครื่องมือวัดมีข้อคำถามตรงกับสิ่งที่วัด ตรงตามเนื้อหาของสิ่งที่ต้องการศึกษา โดยมีวิธีการหาดังนี้ คือตรวจสอบความสอดคล้องของข้อคำถามกับจุดประสงค์ที่ต้องการจะวัดจากผู้เชี่ยวชาญและตัวผู้วิจัยเอง ซึ่งขั้นตอนการหาค่าความตรง ด้านเนื้อหามีรายละเอียดดังนี้

1. ผู้วิจัยทำการหาคุณภาพเองโดยเอาเกณฑ์เอกสารที่เป็นเนื้อหามาพิจารณาว่าตรงตามวัตถุประสงค์หรือไม่ตรงประเด็นในเรื่องที่เราต้องการวัดหรือไม่หรือพิจารณาจากตารางวิเคราะห์หลักสูตร

2. ผู้เชี่ยวชาญทำการหาคุณภาพ มีวิธีการดังนี้

2.1 ให้ผู้เชี่ยวชาญการวิจัยอย่างน้อย 3 คน ช่วยประเมินเป็นรายบุคคลว่าข้อความของเครื่องมือแต่ละข้อนั้น สามารถวัดได้ตรงกับเนื้อหาในเรื่องที่เราวิจัยหรือไม่ โดยให้คะแนนตามเกณฑ์ ดังนี้

2.1.1 ข้อความสามารถวัดได้ตามเนื้อหาหรือประเด็นให้ +1

2.1.2 ข้อความที่ไม่แน่ใจว่าวัดได้ตามเนื้อหาหรือประเด็นให้ 0

2.1.3 ข้อความไม่สามารถวัดได้ตามเนื้อหาหรือประเด็นให้ -1

2.2 นำคะแนนของผู้เชี่ยวชาญทุกคนที่ประเมิน มากรอกลงในแบบวิเคราะห์ความสอดคล้องของข้อความกับวัตถุประสงค์หรือเกณฑ์ เพื่อหาค่าเฉลี่ยสำหรับข้อความแต่ละข้อ

ความตรงตามโครงสร้าง (Construct validity)

ความตรงตามโครงสร้าง (Construct validity) หมายถึง การที่เครื่องมือนั้นสามารถวัดได้ตรงตามลักษณะหรือทฤษฎีที่วางไว้ ความตรง เชิงโครงสร้างสามารถหาได้หลายวิธี ดังนี้

1. หาค่าสหสัมพันธ์ของคะแนนจากเครื่องมือของงานวิจัยกับคะแนนจากเครื่องมือต่างชุดกันที่วัดคุณลักษณะเดียวกัน ซึ่งมีการหาแล้วว่าสามารถวัดคุณลักษณะที่จะวัดนั้นได้จริง โดยหาค่าสัมประสิทธิ์สหสัมพันธ์ การหาค่าความเที่ยงจากเครื่องมือที่เป็นคู่ขนานกัน

2. ใช้กลุ่มที่รู้ชัดอยู่แล้ว (Known-group technique) เป็นวิธีเปรียบเทียบคะแนนเฉลี่ยระหว่างกลุ่มที่มีลักษณะที่ต้องการจะวัด กับกลุ่มที่ไม่มีคุณลักษณะของสิ่งที่ต้องการจะวัด ถ้ากลุ่มที่มีคุณลักษณะที่ต้องการจะวัดได้คะแนนเฉลี่ยสูงกว่ากลุ่มที่ไม่มี คุณลักษณะที่ต้องการจะวัด แสดงว่าเครื่องมือนั้นมีความ เที่ยงตรงเชิงโครงสร้าง

3. ใช้วิธีวิเคราะห์องค์ประกอบ (Factor analysis) โดยการหาค่าสหสัมพันธ์ภายใน (Inter correlation) ของข้อคำถามแต่ละข้อว่าสามารถวัดองค์ประกอบร่วมเดียวกันหรือไม่ ถ้าเครื่องมือนั้นมี 1 องค์ประกอบ แสดงว่ามีความตรง เชิงโครงสร้าง

4. ให้ผู้เชี่ยวชาญวิเคราะห์หาค่าความสอดคล้องระหว่างข้อความหรือข้อคำถามกับเกณฑ์หรือทฤษฎี ซึ่งเหมือนกับการหาค่าความตรง เชิงเนื้อหา (IOC)

ความตรงตามสภาพ (Concurrent validity)

ความตรงตามสภาพ (Concurrent validity) หมายถึง เครื่องมือที่มีความตรงตามสภาพ คือ เครื่องมือนั้นสามารถให้ผลการวัดสอดคล้องกับสภาพความเป็นจริงของคุณลักษณะนั้นในขณะนั้น เช่น คนที่เก่งภาษาอังกฤษมากที่สุดในชั้น ก็ควรจะสอบข้อสอบภาษาอังกฤษ ได้เป็นที่หนึ่ง เป็นต้น ในการหาความตรงตามสภาพ นี้ต้องมีข้อมูลแทนสภาพของผู้ถูกสอบแล้ว จึงนำข้อมูลเหล่านั้นมาหาความสัมพันธ์กับข้อมูลที่ได้จากเครื่องมือที่ต้องการหาความตรงตามสภาพ โดยการหาค่าสัมประสิทธิ์สหสัมพันธ์แบบ Pearson หรืออาจใช้สูตรการหาค่าสัมประสิทธิ์สหสัมพันธ์แบบอื่น ๆ ความตรงตามสภาพ นี้มีความเกี่ยวข้องกับค่าอำนาจจำแนก (Discrimination) ของเครื่องมือนั้น ๆ เครื่องมือที่มีค่าอำนาจจำแนกสูงจะมีความตรงเชิงสภาพสูงควบคู่กันไป คุณสมบัติของแบบทดสอบที่สามารถทำหน้าที่วัดสภาพความสามารถที่เป็นจริงของผู้สอบ สัมประสิทธิ์สหสัมพันธ์ในทางบวกที่สูงแสดงถึงคะแนนจากแบบทดสอบสามารถใช้เป็นตัวบ่งชี้ที่ดี ของสถานภาพของลักษณะที่มุ่งวัดนั้น (ศิริชัย กาญจนวาสี, 2544, หน้า 83) เช่น บุคคลใดที่ทำข้อสอบมนุษยสัมพันธ์ได้คะแนนสูง ก็น่าจะมีพฤติกรรมของผู้มีมนุษยสัมพันธ์ดีแต่หากว่าความเป็นจริงแล้วเขาไม่มีพฤติกรรมด้านมนุษยสัมพันธ์ ก็แสดงว่าเครื่องมือนั้นไม่มีความตรง ตามสภาพวิธีการหา ใช้วิธีการหาค่าสถิติมักใช้ค่าสหสัมพันธ์ หาได้ทั้งรายข้อและทั้งฉบับ เป็นความเที่ยงตรงของคะแนนที่ได้จากเครื่องมือวัด

ในการวิจัยนี้ ผู้วิจัย วิเคราะห์ค่าความตรงตามสภาพ ของวิธีการตอบและการตรวจให้คะแนนโดยคำนวณค่าสหสัมพันธ์ระหว่างคะแนนที่ได้จากแบบสอบที่มีวิธีการตอบและตรวจให้คะแนน ทั้ง 2 วิธี วิเคราะห์ค่าความตรงตามสภาพ จากค่าสัมประสิทธิ์สหสัมพันธ์ (r) ระหว่างคะแนนจากแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาเคมี ภาคเรียนที่ 2 ชั้นมัธยมศึกษาปีที่ 4 ที่ได้จากวิธีการตอบ และการให้คะแนนในแต่ละวิธี กับระดับผลการเรียนวิชาเคมี ภาคเรียนที่ 1 ชั้นมัธยมศึกษาปีที่ 4 แล้วแปลงค่าความตรงตามสภาพจากค่าสัมประสิทธิ์สหสัมพันธ์ เป็นคะแนนฟิชเชอร์ซี (Z_r) โดยใช้สูตร Glass and Stanley (1970, p. 265)

ความตรงตามการทำนาย (Predictive validity)

ความตรงตามการทำนาย (Predictive validity) หมายถึง เครื่องมือนั้นสามารถคาดคะเนพยากรณ์ สภาพหรือการกระทำในอนาคตได้ โดยพิจารณาได้จาก

1. ความสัมพันธ์ระหว่างคะแนนจากการทำเครื่องมือวิจัยกับเกณฑ์ในอนาคต โดยนำเอาผลการทดสอบไปพยากรณ์ผลความสำเร็จในอนาคต โดยหาค่าสัมประสิทธิ์ สหสัมพันธ์ เช่น การคัดเลือกบุคคลเข้าทำงานในสาขาอาชีพต่าง ๆ หากเขามีคะแนนสูงก็แสดงว่า เขาน่าจะทำงานที่มีประสิทธิภาพ

2. เมื่อได้ผลจากในขั้นที่ 1 แล้วก็ให้คำนวณหาค่าการพยากรณ์ โดยเอาคะแนนของตัวแปรต้น หรือตัวแปรอิสระ ซึ่งถือว่าเป็นตัวพยากรณ์ (Predictor) และตัวแปรตามซึ่งถือว่าเป็นเกณฑ์ (Criteria) ทั้งความตรงตามสภาพและความตรง ตามการทำนาย มีหลักการเหมือนกันคือใช้เกณฑ์ภายนอกเพื่อหาความตรง ส่วนความแตกต่างระหว่าง 2 ประเภทนี้คือ เรื่องของเวลา

ตอนที่ 4 การสร้างแบบทดสอบโดยใช้ทฤษฎีการเรียนรู้ของบลูม (Bloom's taxonomy)

พฤติกรรมที่ใช้ในการวัดในแบบทดสอบ

ลักษณะเด่นประการหนึ่งของแบบทดสอบเลือกตอบก็คือสามารถวัดพฤติกรรมการเรียนรู้ด้านต่าง ๆ ได้หลากหลาย ซึ่งมีผู้เสนอแนวคิดเกี่ยวกับพฤติกรรมการเรียนรู้ด้านต่าง ๆ ไว้ดังนี้

Bloom et al.. (1956 อ้างถึงใน ล้วน สายยศ และอังคณา สายยศ, 2539, หน้า 41-44) ได้แบ่งพฤติกรรมการเรียนรู้ออกเป็น 6 พฤติกรรมย่อย ๆ ดังนี้

1. ความรู้ (Knowledge) เป็นความสามารถในการระลึกนึกออกในสิ่งที่ได้เรียนรู้มาแล้ว หรือก็คือความจำนั่นเอง

2. ความเข้าใจ (Comprehension) หมายถึง ความสามารถในการแปลความ ตีความ และขยายความจากสื่อความหมายต่าง ๆ ที่ได้พบเห็น

3. การนำไปใช้ (Application) เป็นความสามารถในการนำหลักวิชาไปใช้แก้ปัญหาในสถานการณ์ได้

4. การวิเคราะห์ (Analysis) เป็นความสามารถในการแยกแยะเพื่อหาส่วนย่อย ๆ ของเหตุการณ์ เรื่องราวหรือเนื้อหาต่าง ๆ ว่าประกอบด้วยอะไร มีความสำคัญอย่างไร อะไรเป็นเหตุเป็นผล และที่เป็นอย่างนั้นอาศัยหลักการอะไร

5. การสังเคราะห์ (Synthesis) เป็นความสามารถในการผสานส่วนย่อยต่าง ๆ เข้าด้วยกันเพื่อให้เป็นสิ่งใหม่อีกรูปแบบหนึ่ง มีลักษณะโครงสร้างหรือหน้าที่ใหม่แปลกแตกต่างไปจากเดิม

6. การประเมินค่า (Evaluation) เป็นความสามารถในการพิจารณาตัดสินลงสรุปเกี่ยวกับคุณค่าของความคิดทุกชนิด เพื่อเปรียบเทียบเกณฑ์ หรือมาตรฐานที่กำหนดให้

การประยุกต์ IRT สำหรับการสร้างแบบทดสอบ

ในการสร้างแบบทดสอบต่าง ๆ เช่น แบบทดสอบผลสัมฤทธิ์ แบบวัดความถนัด แบบวัดความสามารถ เป็นต้น ค่าพารามิเตอร์ของข้อสอบมีลักษณะไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบการคัดเลือกข้อสอบแต่ละข้อจึงสามารถทำได้อย่างอิสระ รวมทั้งค่าพารามิเตอร์ความยากและความสามารถของผู้สอบได้รับการประมาณค่าอยู่บนสเกลเดียวกัน จึงทำให้เลือกข้อสอบแต่ละข้อให้ทำหน้าที่ดีที่สุด ณ บริเวณใดบริเวณหนึ่งบนสเกลของความสามารถได้ เช่น การกำหนดจุดตัดของผู้รอบรู้/ ไม่รอบรู้ ณ ตำแหน่ง θ ที่ต้องการ เป็นต้น ข้อได้เปรียบของ IRT คือ สามารถเลือกข้อสอบเป็นรายข้อบนพื้นฐานของปริมาณสารสนเทศที่จะได้รับ สำหรับผู้สอบที่มี θ ต่าง ๆ กัน สารสนเทศของข้อสอบสะท้อนความถูกต้องแม่นยำของการประมาณค่า เมื่อนำมารวมกันจะเป็นสารสนเทศของแบบทดสอบจึงสามารถใช้เป็นหลักประกันว่าจะได้แบบทดสอบตามเป้าหมายที่สนองต่อการนำไปใช้ให้ผลแม่นยำตามที่ต้องการ (ศิริชัย กาญจนวาสี, 2550, หน้า 80)

การสร้างแบบทดสอบทั่วไป ในกรณีที่มีคลังข้อสอบขนาดใหญ่ ซึ่งเก็บข้อสอบที่มีคุณภาพสูงไว้จำนวนมาก การคัดเลือกข้อสอบโดยพิจารณาจากสารสนเทศ เพื่อนำมาสร้างเป็นแบบทดสอบตามเป้าหมายที่ต้องการ มีขั้นตอนการคัดเลือก ดังนี้ (Lord, 1977 อ้างถึงใน ศิริชัย กาญจนวาสี, 2550, หน้า 80-81)

1. กำหนดรูปลักษณ์ของฟังก์ชันสารสนเทศเป้าหมายที่ต้องการ (Target information function)

2. เลือกข้อสอบจากคลังข้อสอบให้ได้ข้อสอบที่มีสารสนเทศของข้อสอบ [$I_j(\theta)$] ที่จะช่วยเติมเต็มให้เกิดรูปลักษณ์ตามฟังก์ชันสารสนเทศเป้าหมาย

3. หลังจากเลือกข้อสอบเพิ่มเข้ามาทีละข้อ คำนวณค่าฟังก์ชันสารสนเทศของแบบทดสอบ [$I(\theta)$]

4. เลือกข้อสอบข้อต่อไปตาม 2 และคำนวณ I (θ) จนได้รูปลักษณ์บรรลุตามฟังก์ชันสารสนเทศตามเป้าหมาย

ผู้วิจัยได้สร้างแบบสอบวัดผลสัมฤทธิ์ทางการเรียนในครั้งนี้ โดยเลือกใช้ทฤษฎีการเรียนรู้ของบลูม (Bloom's taxonomy) เป็นแนวทางในการพัฒนาแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชา เคมี นักเรียนชั้นมัธยมศึกษาปีที่ 4

ตอนที่ 5 คะแนนจุดตัด (Cut score)

ความหมายของคะแนนจุดตัด

นักการศึกษาและนักวัดผล เรียก คะแนนจุดตัด (Cut scores) ในชื่อต่าง ๆ กัน เช่น เกณฑ์ (Criteria) มาตรฐาน (Standard) คะแนนผ่าน (Passing score) ระดับความรอบรู้ (Mastery level) หรือความสามารถต่ำสุด (Minimal competence) โดยให้ความหมายไว้ต่าง ๆ กัน ดังนี้

Berk (1980) คะแนนจุดตัด คือ จุดบนสเกลคะแนนของแบบทดสอบที่แบ่งเป็น 2 กลุ่ม คือ กลุ่มรอบรู้ และกลุ่มไม่รอบรู้

อังคณา สายยศ (2525, หน้า 70) ให้ความหมายของคะแนนจุดตัดว่า หมายถึง คะแนนที่น้อยที่สุดที่นักเรียนจะต้องทำได้ในการที่จะได้รับการตัดสินให้เป็นผู้รอบรู้

ดำรง ศิริเจริญ (2529, หน้า 139) กล่าวถึง คะแนนจุดตัดว่าหมายถึงคะแนนที่น้อยที่สุดที่นักเรียนจะต้องทำได้ในการที่จะได้รับการตัดสินให้ผู้รอบรู้ คะแนนเกณฑ์อาจจะอยู่ในรูปของจำนวนข้อที่ผู้เรียนทำถูกในแต่ละจุดประสงค์

Chohen, Kane and Crooks (1999) กล่าวว่า คะแนนจุดตัดเป็นขอบเขตระหว่างรอยต่อของแต่ละกลุ่มซึ่งก็คือ จุดบนสเกลคะแนนนั้นเอง ตัวอย่างเช่น มีคะแนนจุดตัด 1 จุด ที่แบ่งออกเป็น 2 กลุ่ม คือ กลุ่มผ่านและกลุ่มตก ซึ่งคะแนนจุดตัดนั้นเป็นการกำหนดขึ้นในทางปฏิบัติสำหรับกำหนดกลุ่มที่แตกต่างกัน

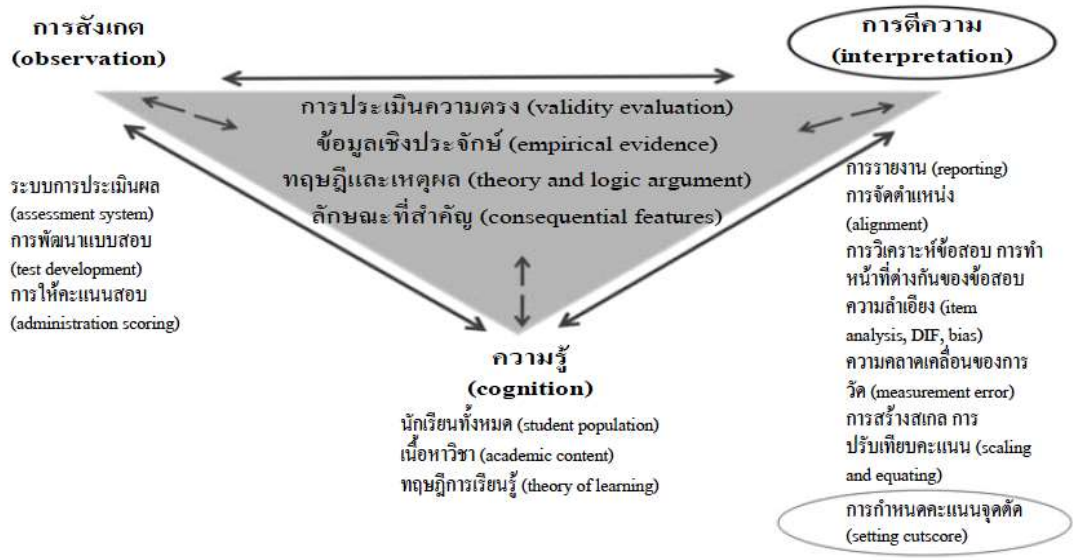
สรุปได้ว่า คะแนนจุดตัด คือ คะแนนที่เป็นเกณฑ์ต่ำสุดที่ใช้ตัดสินให้ผู้สอบเป็นผู้รอบรู้หรือไม่รอบรู้

แนวคิดเรื่องการกำหนดคะแนนจุดตัด

การกำหนดคะแนนจุดตัด (Cuts cores) มีวิวัฒนาการมากว่า 60 ปี ในปี ค.ศ. 1970 วิธีการกำหนดคะแนนจุดตัด (Cuts cores) ถูกใช้อย่างแพร่หลายสืบเนื่องมาจากความนิยมการทดสอบแบบอิงเกณฑ์ (Criterion referenced testing) การกำหนดคะแนนจุดตัด (Cuts cores) เป็นการดำเนินการเพื่อกำหนดคะแนนที่มีความเหมาะสมและน่าเชื่อถือสำหรับใช้เป็นเกณฑ์หรือขีดจำกัด (Threshold)

ในการแบ่งคะแนนเป็น 2 กลุ่มหรือมากกว่าสองกลุ่มที่สะท้อนถึงระดับความสามารถที่ต่างกัน (Berk, 1980; Cizek & Bunch, 2007; Zieky, 2008; Towles-Reeves, 2008)

ในปี ค.ศ. 2001 คะแนนจุดตัด (Cuts cores) ได้ถูกใช้อย่างกว้างขวางมากในสหรัฐอเมริกาเนื่องจากรัฐบาลกลางของสหรัฐอเมริกาออกกฎหมายปฏิรูปการศึกษาที่ชื่อว่า “No child left behind” ที่มีผลบังคับทั้งโรงเรียนประถมศึกษาและมัธยมศึกษาทั่วสหรัฐอเมริกา โดยมีสาระสำคัญว่านักเรียน เกรด 3-8 จะต้องมีพัฒนาการทางการเรียนด้านการอ่านและการคำนวณ และมีการจัดการทดสอบเพื่อวัดผลของพัฒนาการทางการเรียนเป็นรายปี เรียกว่า “Adequate yearly progress (AYP)”โดยได้กำหนดคะแนนจุดตัด (Cuts cores) เพื่อใช้ตัดสินระดับความสามารถทางด้านการอ่านและการคำนวณของนักเรียน (Zieky & Perie, 2006)



ภาพที่ 2-5 บทบาทของการกำหนดคะแนนจุดตัด (Cuts cores) ในกระบวนการวัดและประเมินผลการเรียนรู้ (Marion & Pellegrino, 2006)

จากภาพที่ 2-5 การกำหนดคะแนนจุดตัด (Cuts cores) มีบทบาทเพื่อการตีความหมายคะแนนสอบในกระบวนการวัดและประเมินผลการเรียนรู้ โดยใช้เป็นขีดจำกัดในการแบ่งคะแนนสอบเป็นกลุ่ม ๆ ซึ่งคะแนนแต่ละกลุ่มก็จะสามารถอธิบายความหมายตามเป้าหมายของการกำหนดคะแนนจุดตัด (Cuts cores) ผู้ที่จะนำคะแนนจุดตัด (Cuts cores) มาใช้ควรทำความเข้าใจเบื้องต้นเกี่ยวกับคะแนนจุดตัด (Cuts cores) ก่อนที่จะนำไปใช้ในการตัดสินใจ เพื่อให้สามารถใช้ได้อย่างถูกต้องและเหมาะสม ซึ่งผู้วิจัยได้รวบรวมไว้ดังนี้ (Kane, 1994; Norcini, 2003; Zieky, 2008)

1. คะแนนจุดตัด (Cuts cores) มีบทบาทเพื่อการตัดสินใจซึ่งการตัดสินโดยการใช้คะแนนจุดตัด (Cuts cores) จะทำให้เกิดลักษณะของการตัดสินใจ 4 ลักษณะ ดังนี้

1.1 การตัดสินว่าเป็นผู้ไม่ผ่านเกณฑ์ แต่ในสภาพความเป็นจริงแล้วเป็นผู้ที่มีความรู้พอที่จะผ่านเกณฑ์ได้ ซึ่งถือว่าเป็นการตัดสินใจคลาดเคลื่อนแบบไม่ยอมรับ (Error of rejection) หรือเรียกว่า ความผิดพลาดแบบลบ (False negative)

1.2 การตัดสินว่าเป็นผู้ผ่านเกณฑ์ แต่ในสภาพความเป็นจริงแล้วเป็นผู้ที่ไม่มีความรู้พอที่จะผ่านเกณฑ์ ซึ่งถือว่าเป็นการตัดสินใจคลาดเคลื่อนแบบยอมรับ (Error of acceptance) หรือเรียกว่า ความผิดพลาดแบบบวก (False positive)

1.3 การตัดสินว่าผ่านเกณฑ์และในสภาพความเป็นจริงก็เป็นผู้ที่มีความรู้พอที่จะผ่านเกณฑ์ ถือว่าเป็นการตัดสินที่ถูกต้องไม่มีความคลาดเคลื่อนเกิดขึ้น

1.4 การตัดสินว่าไม่ผ่านเกณฑ์และในสภาพความเป็นจริงก็เป็นผู้ที่ไม่มีความรู้พอที่จะผ่านเกณฑ์ ถือว่าเป็นการตัดสินที่ถูกต้องไม่มีความคลาดเคลื่อนเกิดขึ้น

2. คะแนนจุดตัด (Cuts cores) เป็นคะแนนซึ่งขึ้นอยู่กับการตัดสินของผู้เชี่ยวชาญ กล่าวคือคะแนนจุดตัด (Cuts cores) เป็นคะแนนที่ได้จากการตัดสินของผู้ตัดสินหรือผู้เชี่ยวชาญ ซึ่งจะตัดสินโดยอาศัยความรู้และประสบการณ์ในขณะนั้นเท่านั้น ดังนั้นการเลือกผู้เชี่ยวชาญจึงเป็นกระบวนการที่สำคัญมาก โดยต้องเลือกผู้ที่มีความเชี่ยวชาญในเรื่องนั้น ๆ และต้องมีเอกสารประกอบการตัดสินใจที่เพียงพอเพื่อให้คะแนนจุดตัดที่ได้มีความแม่นยำ

3. ไม่มีคะแนนจุดตัด (Cuts cores) จริง กล่าวคือคะแนนจุดตัด (Cuts cores) ที่กำหนดขึ้นเป็นเพียงคะแนนที่ได้จากการตัดสินของผู้เชี่ยวชาญกลุ่มนั้น ๆ เท่านั้น และใช้กับการตัดสินใจหนึ่ง ๆ เฉพาะเท่านั้น

วิธีการกำหนดคะแนนจุดตัด (Standard-setting method)

การกำหนดคะแนนจุดตัดนั้น Hambleton and Eignor (1979 อ้างถึงใน ล้วน สายยศ และอังคณา สายยศ, 2543, หน้า 268) ได้แบ่งวิธีหาคะแนนจุดตัดออกเป็น 3 วิธี คือ การกำหนดคะแนนจุดตัดโดยวิธีการพิจารณา (Judgmental methods) การกำหนดคะแนนจุดตัดโดยวิธีเชิงประจักษ์ (Empirical methods) และการกำหนดคะแนนจุดตัดแบบผสม (Combination methos) มีรายละเอียดดังนี้

การกำหนดคะแนนจุดตัดโดยวิธีการพิจารณา

วิธีนี้เป็นการกำหนดคะแนนจุดตัดโดยให้ผู้เชี่ยวชาญเป็นผู้พิจารณาตัดสินจากเนื้อหาของข้อสอบแต่ละข้อ แล้วคำนวณหาค่าคะแนนจุดตัด ซึ่งมีผู้เสนอหาคะแนนจุดตัดหลายวิธี ดังวิธีของ Nedelsky วิธีของ Angoff and Hambleton ดังนี้

1. วิธีของ Nedelsky เทคนิคนี้อาศัยการพิจารณาว่านักเรียนที่มีความสามารถต่ำสุดต้องมีคะแนนสอบผ่านขั้นต่ำเป็นเท่าไร โดยอาศัยความน่าจะเป็นของการเดา ตอบถูกจากแบบทดสอบเลือกตอบหลายตัวเลือก คะแนนจุดตัดจะกำหนดจากคะแนนสอบผ่านขั้นต่ำของนักเรียนที่มีความสามารถขั้นต่ำสุด ซึ่งคำนวณจากสูตร ดังนี้

$$D = M_{FD} + S_{FD}$$

เมื่อ D แทน คะแนนสอบผ่านขั้นต่ำ

M_{FD} แทน ค่าเฉลี่ยคะแนนคาตอบถูก หาจากผลรวมความน่าจะเป็นในการตอบถูก

S_{FD} แทน ค่าส่วนเบี่ยงเบนมาตรฐานของคะแนนการเดาตอบถูก

K แทน ค่าคงที่ ถ้าตัวเลือกผิดเด่นชัดจะใช้ค่า 0.5 ถ้าเลือกแตกต่างกันเล็กน้อยจะใช้ค่า 1.0 ถ้าเลือกคล้ายคลึงกันมากจะใช้ค่า 1.5

2. วิธีของแองกอฟ (Angolf method) ถูกคิดค้นโดย Willoam H. Angoff ในปี ค.ศ. 1970 เพื่อใช้ในการกำหนดคะแนนจุดตัดสำหรับข้อสอบแบบเลือกตอบ ซึ่งวิธีแองกอฟถูกนำไปใช้ในการกำหนดคะแนนจุดตัดอย่างกว้างขวางสำหรับการทดสอบที่มีความสำคัญ (Angoff, 1971, cited in Kane, 1994) และยังได้รับความนิยมอย่างต่อเนื่องจนถึงปัจจุบัน (Zieky et al., 2008) ทั้งนี้เนื่องจากว่าวิธีแองกอฟเป็นวิธีที่ง่ายต่อการปฏิบัติและเป็นการกำหนดคะแนนจุดตัดโดยการพิจารณาจากข้อสอบจึงสามารถดำเนินการได้ทั้งก่อนและหลังที่มีการจัดสอบแล้ว โดยให้ผู้เชี่ยวชาญพิจารณาความน่าจะเป็นที่ผู้มีความสามารถคาบเส้นจะสามารถตอบข้อสอบแต่ละข้อได้ถูกต้อง ซึ่งผู้ที่มีความสามารถคาบเส้น หมายถึง คนที่มีระดับความรู้ขั้นต่ำในระดับความสามารถนั้น ๆ เมื่อพิจารณาครบทุกข้อแล้วก็นำค่าความน่าจะเป็นที่ผู้มีความสามารถคาบเส้นจะสามารถตอบข้อสอบได้ถูกต้อง ซึ่งกำหนดโดยผู้เชี่ยวชาญแต่ละท่านมาหาค่าเฉลี่ยก็จะได้ความน่าจะเป็นที่ผู้มีความสามารถคาบเส้นจะสามารถตอบข้อสอบได้ถูกต้องเป็นรายข้อ จากนั้นก็รวมค่าความน่าจะเป็นที่ผู้มีความสามารถคาบเส้นจะสามารถตอบข้อสอบได้ถูกต้องจากทุกข้อในแบบสอบเพื่อกำหนดเป็นคะแนนจุดตัด

ขั้นตอน

การกำหนดคะแนนจุดตัดโดยวิธีแองกอฟมีขั้นตอนการดำเนินการดังต่อไปนี้

2.1 ผู้เชี่ยวชาญกำหนดระดับการผ่านขั้นต่ำ ซึ่งหมายถึงความรู้ขั้นต่ำในเรื่องที่จำเป็นต้องมี เพื่อนิยามผู้มีความสามารถคาบเส้น

2.2 ผู้เชี่ยวชาญพิจารณาข้อสอบทีละข้อเพื่อพิจารณาและกำหนดความน่าจะเป็นที่ผู้มีความสามารถคาบเส้นจะสามารถตอบข้อสอบข้อนั้นได้ถูกต้อง

ตัวอย่างการคำนวณคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟ ดังแสดงในตารางต่อไปนี้

ตารางที่ 2-5 ตัวอย่างการคำนวณคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟ (Ziek, 2008)

ข้อสอบ	ผู้ตัดสินคนที่ 1	ผู้ตัดสินคนที่ 2
1	1.00	0.85
2	0.65	0.50
3	0.80	0.75
4	0.45	0.50
5	0.30	0.40
รวม	3.20	3.00

คะแนนจุดตัด = (3.2 + 3.0)/ 2 = 3.1 นั่นคือ ผู้ที่จะผ่านได้ ต้องทำข้อสอบให้ถูก 3 ข้อ จาก 5 ข้อต่อมา Livingston and Zieky (1982) ได้ปรับปรุงวิธีแองกอฟใหม่ มีชื่อว่า Modified Angoff method สำหรับสาเหตุที่ต้องปรับปรุงวิธีแองกอฟก็เนื่องจากว่าต้องการเพิ่มความเที่ยงของผู้ตัดสินในการประมาณโอกาสในการตอบข้อสอบถูกโดยการเพิ่มความสอดคล้องในการตัดสินของผู้ตัดสิน (Intra-judge consistency) และความสอดคล้องในการตัดสินระหว่าง ผู้ตัดสิน (Inter-judge consistency) ซึ่งการเพิ่มความเที่ยงของผู้ตัดสินก็จะทำให้ค่าความผิดพลาดของคะแนนจุดตัดที่ได้ลดลง (Hurtz & Auerbach, 2003)

สำหรับวิธีการกำหนดจุดตัดทำโดยการกำหนดให้ผู้ตัดสินพิจารณาโอกาสการตอบถูกของผู้ที่มีความสามารถขั้นต่ำจำนวน 2 รอบ โดยรอบแรก ผู้ตัดสินกำหนดโอกาสการตอบถูกของผู้สอบที่มีความสามารถขั้นต่ำในข้อสอบข้อแรก เมื่อกำหนดเสร็จก็นำผลการตอบของผู้ตัดสินแต่ละคนมาเขียนบนกระดาน หรือกระดาษแผ่นใหญ่เพื่อให้ผู้ตัดสินทุกคนได้เห็น ถ้าผลการตอบของผู้ตัดสินแต่ละคนมาเขียนบนกระดาน หรือกระดาษแผ่นใหญ่เพื่อให้ผู้ตัดสินทุกคนได้เห็น ถ้าผลการตอบของผู้ตัดสินคล้ายกันต้องมีค่าห่างกันไม่เกิน 10% ก็ให้ตัดสินพิจารณาข้ออื่นถัดไป สำหรับในรอบที่สอง จะมีการอภิปรายร่วมกัน กรณีที่ผลการตัดสินของแต่ละคนไม่คล้ายกัน โดย

ให้ผู้ตัดสินที่ให้เปอร์เซ็นต์สูงสุด และต่ำสุดอธิบายเหตุผลที่ให้เปอร์เซ็นต์ต่ำสุด และสูงสุดจากนั้นเปิดโอกาสให้ผู้ตัดสินสามารถปรับเปลี่ยนผลการตัดสินของแต่ละคนได้ จากนั้นนำผลการตัดสินของแต่ละคนมาหาค่าเฉลี่ยก็จะได้คะแนนจุดตัด

วิธีแองกอฟได้รับการปรับปรุงอีกหลายครั้ง ซึ่ง Reckase (2000 cited in Brandon, 2004) กล่าวว่า ยังไม่มีข้อยุติเกี่ยวกับนิยามของกระบวนการแองกอฟ แต่สามารถสรุปขั้นตอนกว้าง ๆ ดังนี้ เลือกผู้ตัดสิน ฝึกผู้ตัดสิน กำหนดนิยามและอธิบายระดับการปฏิบัติที่ผู้สอบควรพึงมี หลังจากนั้นผู้ตัดสินจะได้รับข้อมูล เช่น ความยาก (P-value) ซึ่งจะนำมาใช้ประกอบการอภิปรายร่วมกันในแต่ละรอบ โดยปกติประมาณ 2-3 รอบ จากนั้นจึงคำนวณหาคะแนนจุดตัด

ต่อมาวิธีแองกอฟได้ถูกพัฒนาให้ง่ายต่อการปฏิบัติการขึ้นโดย Impara and Plake (1997) ซึ่งใช้ชื่อว่า วิธีแองกอฟแบบใช่/ ไม่ใช่ เพื่อให้สามารถกำหนดคะแนนจุดตัดได้ในกรณีที่ข้อสอบแต่ละข้อให้คะแนนไม่เท่ากัน การพิจาณากำหนดคะแนนจุดตัดจะต่างขากวิธีแองกอฟตรงที่วิธีแองกอฟแบบใช่/ ไม่ใช่ จะให้ผู้เชี่ยวชาญพิจารณาว่าผู้มีความสามารถคาบเส้นจะสามารถตอบข้อสอบได้ถูกต้องใช่หรือไม่ ซึ่ง Impara and Plake (1997) ได้ทำการทดลองเคลื่อนเทียบคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟและวิธีแองกอฟแบบใช่/ ไม่ใช่

การกำหนดคะแนนจุดตัดโดยวิธีแองกอฟและวิธีแองกอฟแบบใช่/ ไม่ใช่ จะดำเนินการกำหนดคะแนนจุดตัด 3 รอบ เนื่องจากเป็นจำนวนรอบที่นักวิชาการส่วนใหญ่นิยม (Busch & Jaeger, 1990; Fehrmann, Woehr & Arthur, 1991; Hambleton & Plake, 1995; Hess, Singer, Trinh, Nikkhoy & Bernstein, 2007) แต่ละรอบผู้เชี่ยวชาญสามารถเปลี่ยนแปลงคะแนนจุดตัด โดยดำเนินการดังต่อไปนี้

รอบที่ 1 ผู้เชี่ยวชาญแต่ละท่านพิจารณาข้อสอบทีละข้อและตัดสินว่าผู้มีความสามารถคาบเส้นของแต่ละระดับความสามารถจะสามารถตอบข้อสอบแต่ละข้อได้ถูกต้องใช่หรือไม่ หากผู้เชี่ยวชาญพิจารณาว่าผู้มีความสามารถคาบเส้นของระดับความสามารถที่กำลังพิจารณาอยู่สามารถตอบข้อสอบข้อนั้นได้ถูกต้องก็ให้ใส่เลข 1 ลงในช่องผลการพิจารณาของข้อนั้น แต่หากผู้เชี่ยวชาญพิจารณาว่าผู้ที่มีความสามารถคาบเส้นของระดับความสามารถที่กำลังพิจารณาอยู่ไม่สามารถตอบข้อสอบข้อนั้นได้ถูกต้องก็ให้ใส่เลข 0 ลงในช่องผลการพิจารณาของข้อนั้น โดยผู้เชี่ยวชาญจะต้องพิจารณาทีละระดับความสามารถจนครบทุกระดับ

รอบที่ 2 ผู้เชี่ยวชาญร่วมกันอภิปรายกลุ่มย่อยกับคะแนนจุดตัดที่กำหนดในรอบที่ 1 และกำหนดคะแนนจุดตัดอีกครั้งในรอบนี้ โดยใช้หลักการเดียวกันกับรอบที่ 1

รอบที่ 3 ผู้เชี่ยวชาญร่วมกันอภิปรายกลุ่มย่อยกับคะแนนจุดตัดที่กำหนดในรอบที่ 2 และกำหนดคะแนนจุดตัดอีกครั้งในรอบนี้ โดยใช้หลักการเดียวกันกับรอบที่ 1

3. วิธีของ Ebel วิธีนี้เป็นการใช้การพิจารณาจากลักษณะความยากง่ายและความเกี่ยวข้องในเนื้อหาของแบบทดสอบอิงเกณฑ์เป็นหลักในการพิจารณาความสำเร็จที่คาดหวังไว้ในข้อสอบซึ่งอีเบล ได้กำหนดไว้ ดังนี้

ตารางที่ 2-6 การพิจารณาความสำเร็จที่คาดหวังในข้อสอบโดยวิธีของ อีเบล

ลักษณะข้อสอบ	ระดับความยากง่ายของแบบทดสอบ		
	ง่าย	ปานกลาง	ยาก
ความจำเป็น	100%	-	-
ความสำคัญ	90%	70%	-
การยอมรับ	80%	60%	40%
ยังเป็นปัญหา	70%	50%	30%

จากตารางที่ 2-5 นำแบบทดสอบอิงเกณฑ์แต่ละข้อมาแจกแจงลักษณะ ของสิ่งที่เกี่ยวข้องในเนื้อหาแล้วคำนวณเป็นคะแนนจุดตัดหรือคะแนนการสอบผ่านของ นักเรียนดังตัวอย่าง

ตัวอย่างแบบทดสอบฉบับหนึ่งมี 50 ข้อเมื่อให้ผู้เชี่ยวชาญ 5 คนพิจารณา แยกแยะลักษณะข้อสอบ ซึ่งจะกลายเป็นมีจำนวนข้อทั้งหมด 250 ข้อ (50× 5) แล้วนำไปคำนวณคะแนนจุดตัด ดังนี้

ตารางที่ 2-7 ตัวอย่างแบบทดสอบฉบับหนึ่งมี 50 ข้อเมื่อให้ผู้เชี่ยวชาญ 5 คนพิจารณา แยกแยะลักษณะข้อสอบ

ลักษณะข้อสอบ	จำนวนข้อ	ความสำเร็จที่คาดหวังไว้	จำนวนข้อ × ความสำเร็จที่คาดหวังไว้
ความจำเป็น	47	100%	4,700
ความสำคัญ			
ง่าย	53	90%	4,770
ปานกลาง		70%	5,390
การยอมรับ			
ง่าย	12	80%	960
ปานกลาง	24	60%	1,440

ตารางที่ 2-7 (ต่อ)

ลักษณะข้อสอบ	จำนวนข้อ	ความสำเร็จที่คาดหวังไว้	จำนวนข้อ × ความสำเร็จที่คาดหวังไว้
ยาก	26	40%	1,040
ยังมีปัญหา			
ง่าย	2	70%	140
ปานกลาง	5	50%	250
ยาก	4	30%	120
รวม	250		

จากตารางที่ 2-7 ช่องลักษณะข้อสอบจะแยกแยะมาจากตารางที่ใช้เป็นหลักในการพิจารณาความสำเร็จที่คาดหวังไว้ในตารางข้างต้น ซึ่งแยกเป็นข้อสอบจำเป็นข้อสอบที่มีความสำคัญ โดยจำแนกย่อยเป็นข้อสอบง่ายปานกลางข้อสอบที่ยอมรับที่ใช้ในการเรียน โดยจำแนกย่อยเป็นข้อสอบง่ายปานกลางและยากข้อสอบที่ยังมีปัญหาว่าจำเป็นต้องเรียนหรือไม่ โดยจำแนกเป็นข้อสอบง่ายปานกลางและยาก เช่นกัน ส่วนช่องจำนวนข้อสอบนั้นเป็นตัวเลขที่ได้จากผู้เชี่ยวชาญแต่ละคนพิจารณา ข้อสอบว่ามีลักษณะใดจำนวนกี่ข้อรวมผู้เชี่ยวชาญ 5 คน แล้วจะมีจำนวนข้อสอบกี่ข้อดังเช่น ลักษณะข้อสอบความจำเป็นผู้เชี่ยวชาญพิจารณาจากข้อสอบ 50 ข้อ ว่าเป็นข้อสอบที่มีความจำเป็นต่อการเรียนรวมทั้ง 5 คน พิจารณาแล้วมี 47 ข้อ เป็นต้น เมื่อรวมทุกลักษณะและจากจำนวนข้อสอบ 50 ข้อ ก็จะมีข้อสอบรวมทั้งสิ้น 250 ข้อ จากช่องความสำเร็จที่คาดหวังไว้เป็นเปอร์เซ็นต์ที่คาดหวังไว้ว่านักเรียนควรจะทำได้จำแนกตามลักษณะข้อสอบจากตารางของอีเบลข้างต้น สำหรับช่องสุดท้ายนั้นจะเป็นผลมาจากการเอาช่องจำนวนข้อคูณกับช่อง ความสำเร็จที่คาดหวังไว้แล้วรวมตัวเลขของช่องนี้ซึ่งมีค่าเท่ากับ 18,810 จากนั้นจึงคำนวณหา คะแนนจุดตัดจากสูตร

$$\text{คะแนนจุดตัด} = \frac{\text{ผลรวมทั้งหมดของผลคูณระหว่างจำนวนข้อกับความสำเร็จที่คาดหวังไว้}}{\text{ผลรวมจำนวนข้อของผู้เชี่ยวชาญทั้งหมด}}$$

$$\text{แทนค่าคะแนนจุดตัด} = \frac{18,800}{250} = 75.24$$

นั่นคือ แบบทดสอบ 50 ข้อ นี้มีจุดตัดที่ 75%

ดังนั้น จึงหมายความว่า ถ้าข้อสอบมี 100 ข้อต้องทำถูกอย่างน้อย 75 ข้อ

ถ้าข้อสอบมี 50 ข้อ ต้องทำถูกอย่างน้อย $\frac{75\times50}{100}=37.5$ ข้อ

แสดงว่าคะแนนจุดตัดของแบบทดสอบฉบับนี้เท่ากับ 37.5 คะแนน หรือเท่ากับ 38 คะแนน (กรณีทำถูกได้ 1 คะแนน ทำผิดได้ 0 คะแนนในแต่ละข้อ)

4. วิธีของ Hambleton เป็นวิธีกำหนดคะแนนจุดตัดหรือเกณฑ์เพื่อใช้ในการแปลผลการปฏิบัติของผู้เรียนว่าได้เรียนรู้ หรือมีความสามารถตามจุดมุ่งหมายหรือไม่ โดยทั่วไปจะใช้ระดับ 80 ถึง 90 เปอร์เซ็นต์ของข้อสอบทั้งหมดเป็นเกณฑ์พิจารณาถ้าผู้เรียนทาข้อสอบ ได้ถูกต้องถึงระดับนี้แล้วก็จะถือว่าผู้ได้เรียนรู้แล้ว สำหรับวิชาที่เกี่ยวกับพฤติกรรมในการสร้างสรรค์หรือการแก้ปัญหาใหม่ ๆ อาจจะต้องใช้วิธีที่ผู้สอนและผู้เรียนร่วมกันกำหนดเกณฑ์

การกำหนดคะแนนจุดตัดโดยวิธีเชิงประจักษ์

การกำหนดจุดตัดวิธีนี้อาศัยผลการสอบมาใช้ประกอบการพิจารณาตัดสินคะแนนจุดตัดที่เหมาะสมซึ่งมีอยู่ด้วยกันหลายวิธี บางวิธีจะใช้การนิยามความรอบรู้ด้วยคะแนนสอบหรือคะแนนดิบ เช่น วิธีของ (Livingston) วิธีทฤษฎีการตัดสินใจ (Decision-theoretic approach) ของแกลสส์ (Glass), วิธีการของ Berk, วิธีของ Huynh, วิธีของ Krie-wall วิธีหาความคลาดเคลื่อนที่เกิดจากการเดาตอบและการสุ่มข้อสอบ (Errors dus to guessing and item sampling) ในที่นี้จะกล่าวถึงวิธีทฤษฎีการตัดสินใจของ Glass และวิธีการของ Berk ดังต่อไปนี้

1. วิธีทฤษฎีการตัดสินใจ วิธีนี้เป็นการกำหนดคะแนนจุดตัดโดย (Glass, 1978, pp. 237-257) เป็นวิธีการที่ แบ่งนักเรียนออกเป็นสองกลุ่ม โดยอาศัยเกณฑ์ภายนอก ซึ่งอาจจะเป็นผลการเรียนโดยปกติของนักเรียนหรือผลสำเร็จในการทำงานแล้วแบ่งเป็นกลุ่มผู้ผ่านเกณฑ์ภายนอก (Pass) และกลุ่มไม่ ผ่านเกณฑ์ภายนอก (Fail) ในแต่ละกลุ่มเมื่อทำแบบทดสอบอิงเกณฑ์ที่ต้องการหาคะแนนจุดตัดนั้นแล้ว มีจำนวนคนที่ผ่านและไม่ผ่านเกณฑ์ที่กำหนดขึ้นในแต่ละจุดของคะแนนเกณฑ์เท่าไร ดังนี้

ตารางที่ 2-8 จำนวนคนที่ผ่านและไม่ผ่านเกณฑ์ที่กำหนดขึ้นในแต่ละจุดของคะแนนเกณฑ์เกณฑ์ภายนอก

		ผ่าน	ไม่ผ่าน
คะแนนเกณฑ์	ไม่ผ่าน	P_A	P_B
ที่กำหนดในแบบทดสอบอิงเกณฑ์	ผ่าน	P_C	P_D

จากตารางที่กำหนดให้

P_A หมายถึง สัดส่วนนักเรียนที่สอบไม่ผ่านเกณฑ์แบบทดสอบอิงเกณฑ์แต่ผ่านเกณฑ์ภายนอก (False negative)

P_D หมายถึง สัดส่วนนักเรียนที่สอบผ่านเกณฑ์แบบทดสอบอิงเกณฑ์แต่ไม่ผ่านเกณฑ์ภายนอก (False positive)

P_B หมายถึง สัดส่วนนักเรียนที่สอบไม่ผ่านทั้งเกณฑ์แบบทดสอบอิงเกณฑ์และเกณฑ์ภายนอก

P_C หมายถึง สัดส่วนนักเรียนที่สอบผ่านทั้งเกณฑ์แบบทดสอบอิงเกณฑ์ และเกณฑ์ภายนอก

สำหรับเกณฑ์ภายนอกที่กำหนดนั้นจะมีค่าไม่เปลี่ยนแปลงแต่คะแนนของ แบบทดสอบอิงเกณฑ์นั้นจะแปรผันไปตามคะแนนแต่ละค่าของแบบทดสอบ ซึ่งจะทำให้ค่า P_A, P_B, P_C และ P_D แปรผันตามไปด้วย และค่าคะแนนจุดตัดของแบบทดสอบอิงเกณฑ์คือค่า ของฟังก์ชันของคะแนนเกณฑ์ f (Cx) ที่มีค่าน้อยที่สุดจากสูตรดังนี้

$$f(C_X) = \frac{P_A + P_D}{P_B + P_C}$$

ในการคำนวณคะแนนจุดตัดด้วยสมการดังกล่าวต้องยอมรับว่าโอกาสที่จะจำแนกผู้สอบผิดทางลบ (False negative: α) กับจำแนกผู้สอบผิดทางบวก (False positive: β) มีค่าเท่ากัน ถ้าพิสูจน์ได้ว่าโอกาสที่จำแนกผิดทางลบและทางบวกมีค่าไม่เท่ากันแล้ว จะต้องคำนวณคะแนนจุดตัดจากค่าฟังก์ชันที่ปรับแก้แล้วในสูตรดังนี้

$$f(C_X) = \frac{aP_A + \beta P_D}{P_B + P_C}$$

โดยกำหนดให้ค่าโอกาสที่จำแนกผิดทางลบ คือ α และโอกาสที่จำแนกผิดทางบวก คือ β มีค่าอยู่ระหว่าง 0 ถึง 1 และจะมีค่าเท่าไรนั้นขึ้นอยู่กับผู้ประเมินผลการสอบ จะต้องคำนึงถึงความสำคัญสองประการนี้ คือ

1. นักเรียนสอบผ่านเกณฑ์ แบบทดสอบอิงเกณฑ์ แต่สอบไม่ผ่านเกณฑ์ภายนอกหรือสอบตกหรือเรียนไม่สำเร็จควรให้ความสำคัญเท่าไรเป็นตัวกำหนด α

2. นักเรียนที่สอบไม่ผ่านเกณฑ์ของแบบทดสอบ แต่สามารถสอบผ่านเกณฑ์ภายนอกหรือสามารถเรียนสำเร็จควรให้ความสำคัญเท่าไร เป็นตัวกำหนด β โดยทั่วไปแล้วในทางปฏิบัติการคำนวณหาคะแนนจุดตัดโดยวิธีทฤษฎีการตัดสินใจนี้ มักจะกำหนดให้ค่าจำแนกผิดทางลบ (α) กับการจำแนกผิดทางบวก (β) มีค่าเท่ากัน

2. วิธีการของ Berk Berk ได้หาคะแนนจุดตัด โดยประยุกต์มาจากวิธีการเพิ่มคะแนนเกณฑ์อื่น ๆ ซึ่ง Berk กล่าวว่า การกำหนดเกณฑ์คือการกำหนดจุดตัดของคะแนน ที่แบ่งผู้เรียนออกเป็น สองพวก คือ พวกที่ได้รับการสอนให้เป็นพวกที่รอบรู้ (Master) พวกที่ไม่ได้รับการสอนเป็นพวกไม่รอบรู้ (Non-master) หลังจากให้นักเรียนทั้งสองกลุ่มทาแบบทดสอบแล้วพิจารณาการกระทำของคะแนนสองกลุ่มจะคาบเกี่ยวกัน จุดที่ฟังก์ชันทั้งสองตัดกัน คือ คะแนนพยากรณ์ที่จะแบ่งการเรียนรู้เป็นสี่พวกดังนี้

ตารางที่ 2-9 คะแนนพยากรณ์ที่จะแบ่งการเรียนรู้

		การจำแนกเกณฑ์	
คะแนนพยากรณ์	รอบรู้	รอบรู้ (TM)	รอบรู้ไม่จริง (FM)
	ไม่รอบรู้	ไม่รอบรู้ไม่จริง (FM)	ไม่รอบรู้จริง (TM)

คะแนนจุดตัดนี้เป็นคะแนนพยากรณ์ นามาหาค่าคะแนนเกณฑ์ที่ให้ค่าความน่าจะเป็นในการตัดสินอย่างถูกต้อง คือ ค่า P (TM) +P (TN) สูงสุด หรือให้ค่าความน่าจะเป็นในการตัดสินผิด คือ ค่า P (FM) + P (FN) ต่ำสุด ณ จุดคะแนนนั้นจะเป็นคะแนนจุดตัดที่เหมาะสมที่สุด

ในการหาจะเลื่อนค่าคะแนนพยากรณ์ ไปเรื่อย ๆ จุดคะแนนหนึ่งที่มีค่า P (TM) + P (TN) สูงสุดและค่า P (FM) + P (FN) ต่ำสุดเป็นคะแนนจุดตัดที่เหมาะสม

เมื่อ

$$P(TM) = \frac{TM}{M + N}$$

$$P(TN) = \frac{TN}{M + N}$$

$$P(FM) = \frac{FM}{M + N}$$

$$P(FN) = \frac{FN}{M + N}$$

N แทน จำนวนนักเรียนทั้งหมดที่ยังไม่ได้เรียน

M แทน จำนวนนักเรียนทั้งหมดที่เรียนแล้ว

คะแนนจุดตัดแต่ละคะแนนที่หาออกมาได้สามารถตรวจสอบความเที่ยงตรง โดยใช้สัมประสิทธิ์ความเที่ยงตรงของเกณฑ์ เพื่อเลือกค่าสัมประสิทธิ์ที่สูงที่สุดของความน่าจะเป็นในการตัดสินถูกของแต่ละคะแนนจุดตัดมาเป็นคะแนนเกณฑ์ สูตรการหาความเที่ยงตรงของเกณฑ์ได้จากสูตร ดังนี้

$$\phi_{vc} = \frac{P(TM) - BR(SR)}{\sqrt{BR(1-BR)SR(1-SR)}}$$

เมื่อ ϕ_{vc} แทน ค่าสัมประสิทธิ์ความเที่ยงตรงของคะแนนเกณฑ์

BR แทน ค่าความน่าจะเป็นของผู้รอบรู้ในประชากร = P (FN) +P (TM)

SR แทน ค่าความน่าจะเป็นของการพยากรณ์ผู้รอบรู้ในประชากร

= P (FM) +P (TM)

การกำหนดคะแนนจุดตัดโดยวิธีแบบผสม (Combination methods)

วิธีนี้เป็นการกำหนดคะแนนจุดตัดที่มีทั้งวิธีพิจารณาดุลยพินิจและเชิงประจักษ์ (Judgment-empirical) ซึ่งมีอยู่หลายวิธีด้วยกันเช่นวิธีกลุ่มตรงข้าม (Contrasting groups) ของไซกีและลิวิงตัน (Zieky and Novick) เป็นต้น

วิธีกำหนดคะแนนจุดตัดนั้นมีหลายวิธี อยู่ที่ดุลพินิจของผู้วิจัยว่ามีความสะดวก และความถูกต้องในการเก็บข้อมูลมากน้อยเพียงใด ก็ใช้วิธีนั้นหาคะแนนจุดตัดของแบบทดสอบในการวิจัย สำหรับการวิจัยครั้งนี้ผู้วิจัยเลือกวิธีของ Angoff มาใช้กำหนดคะแนนจุดตัด

ตอนที่ 6 งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องกับวิธีการตรวจให้คะแนน

สำราญ มีแจ้ง (2525, หน้า 55-56) ได้เสนอวิธีการตอบและการตรวจให้คะแนนแบบสอบชนิดเลือกตอบที่มี 5 ตัวเลือก โดยให้นักเรียนตอบได้ตั้งแต่ 1 ถึง 3 ตัวเลือกตามความมั่นใจ ถ้ามั่นใจตัวเลือกใดก็ตอบตัวเลือกนั้น ถ้าไม่มั่นใจก็ให้ตอบ 2-3 ตัวเลือก ถ้าตอบตัวเลือกเดียวแล้วถูกจะได้ 4 คะแนน ถ้าตอบ 2 ตัวเลือก แล้วมีตัวถูกจะได้ 2 คะแนน ถ้าตอบ 3 ตัวเลือก แล้วมีตัวถูกจะได้ 1 คะแนน ทุกกรณีถ้ามีการเลือกตัวเลือกที่ผิดจะได้ 0 คะแนน โดยได้เปรียบเทียบค่าความเที่ยง ความตรง และค่าอำนาจจำแนกของแบบสอบชนิดเลือกตอบที่มีการให้คะแนนแบบ 0-1 ตามวิธีของคูมบ์ ตามวิธีของอนันต์ ศรีโสภา และตามวิธีของสำราญ มีแจ้ง โดยใช้แบบสอบคณิตศาสตร์ 5 ตัวเลือก จำนวน 60 ข้อ กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 4 โรงเรียนสุรศักดิ์มนตรี และโรงเรียนวัดน้อยใน จำนวน 4 กลุ่ม ๆ ละ 60 คน และกลุ่มได้รับแบบสอบที่มีการตรวจให้คะแนนกลุ่มละ 1 วิธี ผลการวิจัยพบว่า ค่าความเที่ยงและความตรงของวิธีการให้คะแนนตามวิธีของสำราญ มีแจ้ง สูงกว่าการให้คะแนนแบบ 0-1 ส่วนค่าอำนาจจำแนกของแบบสอบที่มีการให้คะแนนทั้ง 4 วิธีไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ

ทวี ทองคำ (2526, หน้า ง-จ) ได้เปรียบเทียบค่าความเที่ยง ความตรงและอำนาจจำแนกของแบบสอบชนิดเลือกตอบที่ใช้คำสั่งและวิธีการให้คะแนนที่แตกต่างกัน 3 วิธี ได้แก่ วิธีประเพณีนิยม (0-1) วิธีของสำราญ มีแจ้ง และวิธีของกิบบอนส์และคณะ โดยศึกษากลุ่มตัวอย่างที่เป็นนักเรียนชั้นประกาศนียบัตรวิชาชีพปีที่ 1 ของวิทยาลัยเทคนิคยะลา จำนวน 180 คน ผลการวิจัยพบว่า ค่าความเที่ยง ความตรงตามสภาพและค่าอำนาจจำแนกของแบบสอบในทุกคำสั่งและวิธีการให้คะแนนไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ

รุจิรา ขาวสะอาด (2543, หน้า ง) ได้ศึกษาผลของการให้น้ำหนักตามระดับความมั่นใจที่มีต่อค่าความเที่ยง และค่าความตรงของแบบสอบเลือกตอบตัวเลือกธรรมดาและตัวเลือกซ้อนเมื่อให้

น้ำหนักตามระดับความมั่นใจ 2, 3, 4, 5 ระดับ และศึกษาปฏิสัมพันธ์ระหว่างระดับความมั่นใจในการตอบและรูปแบบการเขียนตัวเลือกที่มีต่อคะแนนเฉลี่ย ค่าความเที่ยง และค่าความตรงของแบบสอบ ของรูปแบบข้อสอบ 8 รูปแบบ โดยทำการศึกษากับนักเรียนชั้นมัธยมศึกษาปีที่ 2 จำนวน 8 กลุ่ม พบว่า แบบสอบเลือกตอบตัวเลือกธรรมดาและตัวเลือกซ้อนที่มีการระบุความมั่นใจ 3 ระดับ มีค่าความเที่ยงสูงสุด และแบบสอบเลือกตอบตัวเลือกธรรมดาที่มีการระบุความมั่นใจ 5 ระดับ มีค่าความตรงสูงสุด และแบบสอบเลือกตอบตัวเลือกซ้อนที่มีการระบุความมั่นใจ 2 ระดับ มีค่าความตรงสูงสุด

ศิริชัย กาญจนวาสี (2546, หน้า 68) ได้มีการเปรียบเทียบคุณภาพของแบบสอบเลือกตอบที่มีจำนวนตัวถูก วิธีการตอบและวิธีการตรวจให้คะแนนที่ต่างกัน รูปแบบของแบบสอบเลือกตอบที่ใช้ในการวิจัยประกอบด้วย จำนวนตัวถูกเดี่ยว ตัวถูกหลายตัว วิธีการตอบแบบเลือกชุดตัวถูกตัดตัวลวงและใช้วิธีการตรวจให้คะแนนแบบประเพณีนิยม ให้คะแนนความรู้บางส่วน ทำการศึกษากับนิสิตคณะครุศาสตร์ ชั้นปีที่ 3 จำนวน 233 คน ทำการเก็บรวบรวมข้อมูล 4 ครั้ง ด้วยแบบสอบเลือกตอบที่มีรูปแบบต่างกัน 4 ฉบับ ทำครั้งละ 2 ฉบับอย่างสุ่มจำนวน 2 ครั้ง แล้วทำการทดสอบกลางภาค และปลายภาคเรียน ผลการวิจัยพบว่า แบบสอบเลือกตอบที่มีตัวถูกเดี่ยวและแบบเลือกชุดตัวถูก เมื่อใช้วิธีตรวจแบบให้ความรู้บางส่วน จะมีความตรงเชิงทำนายสูงกว่าวิธีตรวจให้คะแนนแบบประเพณีนิยม และแบบสอบเลือกตอบที่มีตัวถูกเดี่ยวและตอบแบบเลือกชุดตัวถูก เมื่อใช้วิธีตรวจ แบบให้ความรู้บางส่วน มีค่าฟังก์ชันสารสนเทศเฉลี่ยของข้อสอบและค่าฟังก์ชันสารสนเทศของแบบสอบรวมสูงสุด โดยมีประสิทธิภาพสัมพัทธ์สูงกว่าแบบสอบเลือกตอบรูปแบบอื่น 1.6 เท่า

สุพจน์ เกิดสุวรรณ (2545, หน้า ง) ได้พัฒนาวิธีการวัดความรู้บางส่วน จากการตรวจและให้คะแนนวิธีของคูมบ์ส วิธีของเดรสเซลและชมิดท์ และวิธีของอาโนลด์ ซึ่งมีการพัฒนาสูตรจากสูตรเดิมที่มีการให้คะแนนลงโทษติดลบเมื่อผู้สอบตอบผิด สูตรที่พัฒนาขึ้นได้ทำการเปรียบเทียบคุณภาพกับสูตรเดิมและการให้คะแนนแบบประเพณีนิยมโดยใช้ทฤษฎีการตอบสนองข้อสอบ โดยทำการศึกษาในกลุ่มนักเรียนชั้นมัธยมศึกษาปีที่ 3 จังหวัดสระบุรี จำนวน 946 คน เครื่องมือที่ใช้เป็นแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ขั้นพื้นฐานระดับมัธยมศึกษาตอนต้น ผลการวิจัยพบว่า คุณภาพของวิธีการตอบและการตรวจให้คะแนน เมื่อพิจารณาตามเกณฑ์โดยรวมทุกด้าน พบว่า วิธีที่พัฒนาจากวิธีของคูมบ์ส มีคุณภาพดีกว่าวิธีอื่น ๆ รองลงมาคือ วิธีของอาร์โนลด์ และวิธีที่พัฒนาจากเดรสเซลและชมิดท์ ตามลำดับ

ฉวีวรรณ บุญมั่ง (2540) ได้ศึกษาการเปรียบเทียบความสัมพันธ์ระหว่างคะแนนความสามารถจากการวิเคราะห์โดยทฤษฎีการตอบสนองข้อสอบกับคะแนนที่ได้จากการตอบและตรวจให้คะแนน 4 วิธี กลุ่มตัวอย่างเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 โรงเรียนสังกัดสำนักงาน

การประถมศึกษาอำเภอบรบือ จังหวัดมหาสารคาม ภาคเรียนที่ 1 ปีการศึกษา 2540 จำนวน 496 คน พบว่า คะแนนความสามารถมีความสัมพันธ์กับคะแนนที่ได้จากวิธี 0-1 สูงสุดเท่ากับ 0.511 สัมพันธ์กับคะแนนที่ได้จากวิธีของคูมบ์ ต่ำสุดเท่ากับ 0.3896 และสัมพันธ์กับคะแนนที่ได้จากวิธีของกิบบอนส์และคณะ และวิธีของอนันต์ ศรีโสภา ใกล้เคียงกันค่าสัมประสิทธิ์สหสัมพันธ์ทุกค่าเป็นบวกและมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 เมื่อเปรียบเทียบความแตกต่างของค่าสัมประสิทธิ์สหสัมพันธ์พบว่า ไม่แตกต่างกัน

ผลการเปรียบเทียบค่าจากการวิเคราะห์โดยทฤษฎีการตอบสนองข้อสอบกับสัดส่วนการเดาที่ได้จากการตอบและตรวจให้คะแนน 4 วิธี พบว่า ค่าการเดาแตกต่างกับสัดส่วนการเดาที่ได้จากวิธีของ 0-1 วิธีของคูมบ์ วิธีของกิบบอนส์และคณะ และวิธีของอนันต์ ศรีโสภา และสัดส่วนการเดาที่ได้จากวิธีของคูมบ์แตกต่างกับวิธีของ 0-1 และวิธีของกิบบอนส์และคณะอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนสัดส่วนการเดาที่ได้จากการตอบและตรวจให้คะแนนคู่อื่น ๆ ไม่แตกต่างกัน ผลการเปรียบเทียบค่าการเดาจากการวิเคราะห์โดยทฤษฎีการตอบสนองข้อสอบกับสัดส่วนการเดาที่ได้จากการตอบและตรวจให้คะแนนวิธีของคูมบ์ วิธีของกิบบอนส์และคณะและวิธีของอนันต์ ศรีโสภา พบว่า ค่าการเดาแตกต่างกับสัดส่วนการเดาที่ได้จากวิธีของคูมบ์ และวิธีของกิบบอนส์และคณะ สัดส่วนการเดาที่ได้จากวิธีของคูมบ์แตกต่างกับวิธีของกิบบอนส์และคณะและวิธีของอนันต์ ศรีโสภา อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนค่าการเดากับสัดส่วนการเดาที่ได้จากวิธีของอนันต์ ศรีโสภา

พินิจ อุไรรักษ์ (2533) ศึกษาผลของวิธีการให้คะแนนที่มีต่อคะแนนสอบ ค่าความเชื่อมั่นและค่าความเที่ยงตรงจากวิธีการให้คะแนน 4 วิธี คือ วิธีให้น้ำหนักคะแนนรายข้อเท่ากัน วิธีการให้น้ำหนักคะแนนรายข้อต่างกันตามระดับความมั่นใจในการตอบ วิธีการให้คะแนนรายข้อต่างกันตามค่าพารามิเตอร์ตามสูตรของลอร์ด และวิธีการให้คะแนนรายข้อต่างกันตามค่าความสามารถของผู้สอบ กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2534 จำนวน 1,011 คน เครื่องมือที่ใช้เป็นแบบวัดผลสัมฤทธิ์วิทยาศาสตร์ ผลการวิจัยพบว่า

1. ความสัมพันธ์ของคะแนนสอบระหว่างวิธีให้คะแนนทั้ง 4 วิธี มีค่าตั้งแต่ .941-.983
2. ค่าความเชื่อมั่นของแบบทดสอบพบว่า วิธีการให้น้ำหนักคะแนนรายข้อเท่ากัน วิธีการให้น้ำหนักคะแนนรายข้อต่างตามระดับความมั่นใจในการตอบ วิธีการให้น้ำหนักคะแนนรายข้อต่างกันตามค่าพารามิเตอร์ตามสูตรของลอร์ด และวิธีการให้คะแนนรายข้อต่างกันตามค่าความสามารถของผู้สอบ มีค่าเท่ากับ .075

3. วิธีการให้น้ำหนักคะแนนรายข้อต่างตามระดับความมั่นใจในการตอบวิธีการให้น้ำหนักคะแนนรายข้อต่างกันตามค่าพารามิเตอร์ตามสูตรของลอร์ด และวิธีการให้คะแนนรายข้อต่างกันตามค่าความสามารถของผู้สอบ สูงกว่าวิธีการให้น้ำหนักคะแนนรายข้อเท่ากัน

Dressel and Schmidt (1953) ได้ศึกษาวิธีการตอบและวิธีการตรวจให้คะแนนที่ต่างกันเพื่อปรับปรุงอำนาจจำแนกของแบบทดสอบเลือกตอบ โดยมีวิธีการตอบ 4 วิธี คือ

1. แบบเลือกตอบเสรี เป็นแบบทดสอบเลือกตอบที่มี 5 ตัวเลือก ผู้สอบจะเลือกกี่ตัวเลือกก็ได้

2. แบบบอกระดับความมั่นใจในการตอบ เป็นแบบที่ผู้เข้าสอบจะต้องแสดงระดับความมั่นใจในการตอบ ซึ่งมี 4 ระดับ แต่คำถามแต่ละข้อให้ตอบเพียงคำตอบเดียว

3. แบบคำตอบถูกมากกว่าหนึ่ง แบบทดสอบชนิดนี้จะมีคำตอบถูกอยู่หลายคำตอบ ผู้เข้าสอบจะต้องทำเครื่องหมายทุกข้อที่ถูก

4. แบบมีคำตอบถูกสองคำตอบ ผู้เข้าสอบจะเลือกตอบได้เพียงสองคำตอบเท่านั้น

ผลการศึกษาพบว่า การข้อสอบของนักเรียนเก่ง (พิจารณาจากคะแนนสอบ) จะแตกต่างกับนักเรียนปานกลางและอ่อน อย่างมีนัยสำคัญทางสถิติ เมื่อใช้แบบทดสอบเลือกตอบชนิดตอบเสรีคือกลุ่มเก่งจะทำเครื่องหมายน้อยกว่ากลุ่มอ่อนทั้งในข้อสอบที่มีระดับความยากสูง ปานกลาง และต่ำ ในทางตรงกันข้ามแบบทดสอบชนิดทดสอบความมั่นใจไม่ได้จำแนกนักเรียนเก่ง ปานกลาง อ่อนได้ดีกว่าวิธีแรก เพราะเมื่อพิจารณาการตอบข้อสอบแล้วพบว่า ผู้เข้าสอบตอบข้อสอบที่มีความยาก ปานกลาง และยากมากด้วยระดับความมั่นใจเดียวกัน เดรชเซลและชมิด ได้สรุปว่าองค์ประกอบของความมั่นใจที่วัดโดยการให้คำตอบแบบเสรีนั้นแตกต่างกับที่วัดโดยผู้เข้าสอบบอกระดับความมั่นใจ

Smith (1987) ได้ศึกษาเปรียบเทียบความตรงสภาพและความเชื่อมั่นโดยใช้สัมประสิทธิ์แอลฟาระหว่างแบบทดสอบที่มีการให้คะแนนแบบทวิภาค (Dichotomous, 0-1) และแบบทดสอบที่มีการให้คะแนนแบบพหุวิภาค (Polytomous) โดยเก็บข้อมูลจากนักศึกษาที่มาสอบความถนัดทางภาษา จำนวน 1,399 คน ให้ทำแบบทดสอบทางภาษาที่ใช้โดยทั่วไปเพื่อนำไปเป็นเกณฑ์ ในการหาความตรงตามสภาพกับแบบทดสอบภาษาที่ใช้ในการศึกษาครั้งนี้ มี 2 ชุด ชุดแรกมีการให้คะแนนแบบทวิภาค (Dichotomous) และชุดที่ 2 ให้คะแนนแบบพหุวิภาค (Polytomous) โดยที่แต่ละตัวเลือกมีคะแนนแต่ละข้อไม่เท่ากันโดยมีหลักการให้คะแนนความรู้บางส่วนจากทฤษฎีโครงสร้างความรู้ และมีการแบ่งกลุ่มตัวอย่างออกเป็นสองกลุ่ม กลุ่มที่ 1 จำนวน 400 คน ทดสอบตามปกติ กลุ่มที่ 2 จำนวน 999 คน ทดสอบความตรงข้ามกลุ่ม โดยแบ่งเป็นกลุ่มที่มีความสามารถสูงและกลุ่มที่มีความสามารถต่ำ ผลการวิจัยพบว่า การให้คะแนนแบบพหุวิภาค (Polytomous)

ค่าความตรงสภาพและความเชื่อมั่นสูงกว่าการให้คะแนนแบบทวิภาค (Dichotomous) อย่างมีนัยสำคัญทางสถิติทั้งกลุ่มที่มีความสามารถสูงและกลุ่มที่มีความสามารถต่ำ

งานวิจัยที่เกี่ยวกับการหาคุณภาพแบบทดสอบ

เบญจพร ยนตจักรวิธี (2539) เปรียบเทียบการวิเคราะห์ข้อสอบระหว่างทฤษฎีการทดสอบแบบดั้งเดิมกับทฤษฎีการตอบสนองข้อสอบ (โมเดลราส์ช) ในด้านของสัดส่วนจำนวนข้อที่ได้รับการคัดเลือก ความสอดคล้องในการคัดเลือกข้อสอบจากทั้งสองทฤษฎี ความสัมพันธ์ระหว่างค่าความยากและคะแนนความสามารถของผู้สอบ กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 จำนวน 500 คน เครื่องมือที่ใช้เป็นแบบทดสอบแบบเลือกตอบ 4 ตัวเลือก จำนวน 33 ข้อ ผลการวิจัยพบว่า จำนวนข้อสอบที่ได้รับการคัดเลือกจากการวิเคราะห์ทั้งสองทฤษฎี มี 32 ข้อ และ 18 ข้อ ตามลำดับ แต่การคัดเลือกข้อสอบจากทั้งสองทฤษฎีไม่มีความสอดคล้องกัน เมื่อเปรียบเทียบสัดส่วน พบว่า จากการวิเคราะห์ด้วยทฤษฎีการทดสอบแบบดั้งเดิมมีมากกว่าโมเดลราส์ช อย่างมีนัยสำคัญที่ระดับ .01 ค่าความยากของข้อสอบและค่าความสามารถของผู้สอบที่ได้จากการวิเคราะห์ทั้งสองทฤษฎีความสัมพันธ์กันสูงมาก r = .99 และ r = .93 ตามลำดับ อย่างมีนัยสำคัญที่ระดับ .01

อวยพร เรืองตระกูล (2544) การพัฒนาและวิเคราะห์คุณภาพของวิธีการวัดคะแนนพัฒนาการ ตามทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบ พัฒนาและวิเคราะห์เปรียบเทียบคุณภาพของวิธีการวัดคะแนนพัฒนาการ 9 วิธี ประกอบด้วย วิธีการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิม 5 วิธี ได้แก่ 1) วิธีการวัดคะแนนพัฒนาการจากความแตกต่างระหว่างคะแนนดิบ 2) วิธีการวัดคะแนนพัฒนาการจากคะแนนมาตรฐาน 3) วิธีการวัดคะแนนพัฒนาการจากลอการิทึมของคะแนนดิบ 4) วิธีการวัดคะแนนพัฒนาการสัมพันธ์ และ 5) วิธีการวัดคะแนนพัฒนาการส่วนที่เหลือเทียบกับศักยภาพ วิธีการวัดคะแนนพัฒนาการตามทฤษฎีการตอบสนองข้อสอบ 2 วิธี ได้แก่ 1) วิธีการวัดคะแนนพัฒนาการจากความแตกต่างของความสามารถที่แท้จริง และ 2) วิธีการวัดคะแนนพัฒนาการโดยใช้โมเดลราส์ชพหุมิติ สำหรับการเรียนรู้และการเปลี่ยนแปลง และวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา 2 วิธี ได้แก่ 1) วิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับศักยภาพการพัฒนา และ 2) วิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์ เมื่อเทียบกับความสามารถที่แท้จริงก่อนเรียน กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 2 กลุ่ม ๆ ละ 698 คน และ 637 คน จากโรงเรียนสังกัดกรมสามัญศึกษาเขตกรุงเทพมหานคร 12 โรงเรียน เครื่องมือที่ใช้ในการวิจัยครั้งนี้เป็นแบบทดสอบวิชาคณิตศาสตร์ระดับชั้นมัธยมศึกษาปีที่ 3 จำนวน 3 ฉบับ ที่คู่ขนานกัน ข้อมูลที่ใช้ในการวิจัยมี 3 ชุด สองชุดแรกเป็นข้อมูลปฐมภูมิ ที่ได้จากการเก็บรวบรวมข้อมูลด้วยแบบทดสอบคณิตศาสตร์ที่วัดซ้ำ 3 ครั้ง

ด้วยแบบทดสอบฉบับเดิม และแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน ส่วนข้อมูลชุดที่สาม เป็นข้อมูลทุติยภูมิ ที่ได้จากคะแนนสอบวิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 2 ที่มีการวัดซ้ำ 5 ครั้ง มีการตรวจให้คะแนนแบบทวิวิภาค (Dichotomous) และตรวจให้คะแนนความรู้บางส่วน (Partial credit) การวิเคราะห์ข้อมูลเป็นการประมาณค่าคะแนนพัฒนาการ และการวิเคราะห์เปรียบเทียบคุณภาพของวิธีการวัดคะแนนพัฒนาการ ในด้านความตรงตามเกณฑ์สัมพันธ์และค่าความคลาดเคลื่อน ด้วยการวิเคราะห์โมเดลเชิงเส้นตรงระดับลดหลั่น และสถิติทดสอบ Hotelling ผลการวิจัยสรุปได้ดังนี้

1. เมื่อเปรียบเทียบผลการประมาณค่าคะแนนพัฒนาการระหว่าง 3 กลุ่มวิธี พบว่า กลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีตอบสนองข้อสอบ ที่ผู้วิจัยพัฒนามีคุณภาพไม่แตกต่างกัน

2. เมื่อเปรียบเทียบระหว่างวิธีภายใน กลุ่มทฤษฎีการทดสอบแบบดั้งเดิม พบว่า วิธีการวัดคะแนนพัฒนาการจากความแตกต่างระหว่างคะแนนดิบ และวิธีการวัดคะแนนพัฒนาการสัมพัทธ์ที่คุณภาพสูงกว่าวิธีอื่นสำหรับข้อมูล 2 ชุดแรก และวิธีการวัดคะแนนพัฒนาการส่วนที่เหลือเทียบกับศักยภาพของผู้สอบ มีคุณภาพสูงกว่าวิธีอื่นสำหรับข้อมูลชุดที่ 3

3. เมื่อเปรียบเทียบระหว่างวิธีภายในกลุ่มทฤษฎีการตอบสนองข้อสอบ พบว่า วิธีการวัดคะแนนพัฒนาการโดยใช้โมเดลราส์ชพหุมิติสำหรับการเรียนรู้และการเปลี่ยนแปลงมีคุณภาพสูงกว่าวิธีการวัดคะแนนพัฒนาการจากความแตกต่างของความสามารถที่แท้จริง

4. เมื่อเปรียบเทียบระหว่างวิธีภายในกลุ่มทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา พบว่า วิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับความสามารถที่แท้จริงก่อนเรียนมีคุณภาพสูงกว่าวิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับศักยภาพการพัฒนา สำหรับข้อมูลที่มีการตรวจให้คะแนนแบบทวิวิภาค

เอมอร จังศิริพรปกรณ์ (2548, หน้า ง) ได้ทำการเปรียบเทียบคุณภาพของแบบสอบในด้านความตรงตามสภาพ อำนาจจำแนก ความยาก ฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบสอบ และอัตราส่วนสารสนเทศเฉลี่ย ทำการศึกษาโดยใช้ทฤษฎีการตอบสนองข้อสอบระหว่างแบบสอบเลือกตอบที่มีจำนวนตัวถูกตัวเดียวกับตัวถูกมากกว่า 1 ตัว และมีการตรวจให้คะแนนความรู้บางส่วนตามวิธีการตรวจให้คะแนนแบบเลือกชุดตัวถูก และแบบตัดตัวลวง โดยทำการศึกษากับ นิสิตคณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ชั้นปีที่ 3 จำนวน 223 คน เครื่องมือที่ใช้เป็นแบบสอบวัดผลสัมฤทธิ์ทางการเรียน วิชาการวัดและประเมินผลการศึกษา จำนวน 4 ฉบับ คือ แบบสอบที่มีตัวถูกตัวเดียว 2 ฉบับ และมีจำนวนตัวถูกมากกว่า 1 ตัว 2 ฉบับ มีวิธีการตรวจให้คะแนน 2 วิธีข้างต้น

ผลการวิจัยสรุปว่า ค่าความตรงตามสภาพระหว่างแบบสอบไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ค่าอำนาจจำแนก พบว่า แบบสอบเลือกชุดตัวถูกที่มีคำตอบถูกตัวเดียวมีค่าอำนาจจำแนกเฉลี่ยสูงสุด ส่วนแบบสอบชนิดตัดตัวลวงที่มีคำตอบถูกตัวเดียวมีค่าความยากเฉลี่ยต่ำสุด ผลการเปรียบเทียบค่าฟังก์ชันสารสนเทศของแบบสอบ พบว่า แบบสอบเลือกชุดตัวถูกที่มีคำตอบถูกตัวเดียวมีค่าสูงที่สุด

เอมอร จังศิริพรปกรณ์ (2546) ได้ศึกษาการเปรียบเทียบคุณภาพของแบบสอบเลือกตอบเมื่อตรวจด้วยวิธีการให้คะแนนความรู้บางส่วน กับวิธีประเพณีนิยมกลุ่มตัวอย่างที่ใช้ในการศึกษาเป็นนิสิตคณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยชั้นปีที่ 3 ที่ลงทะเบียนเรียนวิชาการประเมินผลการเรียนการสอน ในภาคต้น ปีการศึกษา 2544 จำนวน 297 คน

ผลการวิจัยพบว่า

1. ผลการเปรียบเทียบความตรงตามสภาพ ระหว่างวิธีการตรวจให้คะแนนที่แตกต่างกันจากการศึกษาความตรงตามสภาพตามวิธีประยุกต์การให้คะแนนของคูมบ์ วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิทและวิธีประเพณีนิยม แล้วนำมาเปรียบเทียบความแตกต่างโดยใช้คะแนนฟิชเชอร์ซี เปรียบเทียบในภาพรวมโดยใช้สถิติทดสอบไคว์สแควร์ พบว่า ค่าความตรงตามสภาพไม่ขึ้นอยู่กับวิธีการตรวจให้คะแนนในแต่ละวิธี และเมื่อนำมาเปรียบเทียบความแตกต่างกันเป็นรายคู่ ด้วยสถิติทดสอบซี (Z) พบว่า ไม่มีความแตกต่างกันของความตรงสภาพในแต่ละคู่อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

2. ผลการเปรียบเทียบความยากและอำนาจจำแนกระหว่างวิธีการตรวจให้คะแนนที่แตกต่างกันผลการเปรียบเทียบความแตกต่างของค่าเฉลี่ยของค่าอำนาจจำแนกและค่าเฉลี่ยของค่าความยาก ด้วยสถิติ F พบว่า ค่าเฉลี่ยของค่าอำนาจจำแนกตามวิธีการตรวจให้คะแนนมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ($F = 85.369$, P-value $< .000$) และเมื่อนำมาเปรียบเทียบความแตกต่างรายคู่ด้วยวิธีของ Scheffe พบว่า วิธีประยุกต์การให้คะแนนของคูมบ์มีค่าอำนาจจำแนกเฉลี่ยสูงกว่าวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิทอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ในขณะเดียวกันวิธีประเพณีนิยมมีค่าอำนาจจำแนกเฉลี่ยสูงกว่าวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิทอย่างมีนัยสำคัญทางสถิติ เช่นเดียวกัน ส่วนค่าเฉลี่ยของความยากตามวิธีการตรวจให้คะแนนมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ($F = 20.592$, P-value $< .000$) และเมื่อนำมาเปรียบเทียบความแตกต่างรายคู่ด้วยวิธีของ Scheffe พบว่า วิธีประยุกต์การให้คะแนนของคูมบ์ มีค่าความยากเฉลี่ยต่ำกว่าวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิทและวิธีประเพณีนิยมอย่างมีนัยสำคัญทางสถิติ ในขณะเดียวกันวิธีประเพณีนิยม

มีค่าความยากเฉลี่ยไม่แตกต่างจากวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิทอย่างมีนัยสำคัญทางสถิติ

3. ผลการเปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบสอบและอัตราส่วนสารสนเทศ ระหว่างวิธีการตรวจให้คะแนนที่แตกต่างกัน

วิธีประยุกต์การให้คะแนนของคูมบ์มีค่าฟังก์ชันสารสนเทศของข้อสอบโดยเฉลี่ยสูงกว่าวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิทและวิธีประเพณีนิยมอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท มีค่าฟังก์ชันสารสนเทศของข้อสอบโดยเฉลี่ยสูงกว่าวิธีประเพณีนิยมอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เมื่อเปรียบเทียบค่าฟังก์ชันสารสนเทศของข้อสอบจำแนกตามค่าความสามารถของผู้สอบตามวิธีการตรวจให้คะแนน พบว่า วิธีประยุกต์การให้คะแนนของคูมบ์ ให้ค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงที่สุดในระดับความสามารถ -1 และมีค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยเท่ากับ 0.673 ซึ่งสูงกว่าวิธีอื่น ๆ ในขณะที่วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท ให้ค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงที่สุดในระดับความสามารถ 0 มีค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยเท่ากับ 0.395 ส่วนวิธีประเพณีนิยมให้ค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงที่สุดในระดับความสามารถ 0 มีค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยเท่ากับ 0.187 วิธีประยุกต์การให้คะแนนของคูมบ์ ให้ค่าฟังก์ชันสารสนเทศของแบบสอบ สูงกว่าวิธีอื่น ๆ รองลงมาคือ วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท ส่วนวิธีประเพณีนิยมให้ค่าฟังก์ชันสารสนเทศของแบบสอบต่ำที่สุด ส่วนอัตราส่วนสารสนเทศเฉลี่ยระหว่างวิธีประยุกต์การให้คะแนนของคูมบ์กับวิธีประเพณีนิยมมีค่าสูงสุด รองลงมาคือ อัตราส่วนสารสนเทศเฉลี่ยระหว่างวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิทกับวิธีประเพณีนิยม ซึ่งแสดงให้เห็นว่า วิธีประยุกต์การให้คะแนนของคูมบ์มีประสิทธิภาพสัมพัทธ์สูงที่สุด รองลงมา คือ วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท ส่วนวิธีประเพณีนิยมมีประสิทธิภาพสัมพัทธ์ต่ำที่สุด

ปรมินทร์ อริเดช (2547) ได้ศึกษาการเปรียบเทียบฟังก์ชันสารสนเทศในการใช้โมเดลโลจิสติก จีอาร์เอ็มและจีพีซีอีมของมาตรวัดเจตคติแบบลิเคิร์ทและมาตรวัดแบบตัวเลือกบังคับตอบที่มีวิธีการให้คะแนนแบบสองค่าและแบบหลายค่า กลุ่มตัวอย่างเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 จังหวัดเชียงราย ผลการวิจัยพบว่า ในมาตรวัดเจตคติแบบลิเคิร์ทการตรวจให้คะแนนแบบสองค่าที่วิเคราะห์ด้วยโมเดลโลจิสติกให้ค่าเฉลี่ยฟังก์ชันสารสนเทศสูงกว่าการตรวจให้คะแนนแบบหลายค่าที่วิเคราะห์ตาม GRM GCPM และ GPCM ที่มีการยุบรวมลำดับขั้นคะแนนใหม่ แต่ถ้าพิจารณาโค้งฟังก์ชันสารสนเทศจะพบว่า การวิเคราะห์ด้วย GPCM ที่มีการลำดับขั้นคะแนนใหม่จะให้ค่าฟังก์ชันสูงสุดและพิจารณาจากมาตรวัดแบบลิเคิร์ทแบบซึ่งมีค่าตรงกลางและไม่มีค่าตรง

กลาง การตรวจให้คะแนนแบบไม่มีค่าตรงกลางจะให้จะให้ค่าเฉลี่ยฟังก์ชันสารสนเทศสูงกว่าการตรวจให้คะแนนแบบมีค่าตรงกลาง ในมาตรวัดความมีระเบียบวินัยแบบตัวเลือกบังคับตอบการให้คะแนนแบบหลายค่าที่วิเคราะห์ด้วย GRM ให้ค่าเฉลี่ยฟังก์ชันสารสนเทศสูงสุด รองลงมาคือ GPCM ที่มีการยุบรวมลำดับขั้นคะแนนใหม่ GPCM ซึ่งสูงกว่าการให้คะแนนแบบสองค่าที่วิเคราะห์ด้วยโมเดลโลจิสติก 3, 2 และ 1 พารามิเตอร์ ตามลำดับ เมื่อพิจารณามาตรวัดความมีระเบียบวินัยแบบตัวเลือกบังคับตอบที่มีการให้คะแนนแบบสองค่า ที่มีวิธีกำหนดน้ำหนักคะแนนแตกต่างกัน 3 แบบ พบว่า การวิเคราะห์ด้วยโมเดลโลจิสติก ด้วยวิธีการให้คะแนนแบบที่ 3 ให้คะแนนเฉลี่ยฟังก์ชันสารสนเทศสูงสุดรองลงมา คือ 2 และ 1 พารามิเตอร์

ธนวัฒน์ แสนสุข (2539) ได้ศึกษาการใช้ จีอาร์เอ็ม จีพีซีเอ็มและโมเดลโลจิสติกในการเปรียบเทียบฟังก์ชันสารสนเทศของแบบวัดที่มีวิธีการให้คะแนนต่างกัน โดยที่แบบวัดมีการตรวจให้คะแนนแบบทวิวิภาค (0, 1) และแบบพหุวิภาค (1, 2, 3, 4) ทำการตรวจสอบความสมบูรณ์ของข้อมูลโดยวิเคราะห์ ด้วยโปรแกรม BIGSTEPS และวิเคราะห์หาค่าฟังก์ชันสารสนเทศด้วยโปรแกรม MULTILOG และ PARSCALE ผลการวิจัย พบว่า การตรวจให้คะแนนแบบพหุวิภาคเมือวิเคราะห์ตาม GRM ให้ค่าฟังก์ชันสารสนเทศสูงกว่าการให้คะแนนแบบทวิวิภาค สำหรับการตรวจให้คะแนนแบบพหุวิภาควิเคราะห์ตาม GPCM และแบบทวิวิภาคยังไม่สามารถสรุปได้ว่าวิธีใดให้ค่าฟังก์ชันสารสนเทศสูงกว่ากัน

Jaradat and Tollefson (1988) ได้ศึกษาคุณภาพของแบบสอบโดยใช้ทฤษฎีการวัดแบบดั้งเดิม ซึ่งเปรียบเทียบค่าความตรง ค่าความเที่ยง และการให้เกรดระหว่างวิธีการให้คะแนน 4 วิธีของข้อสอบแบบเลือกตอบ ได้แก่ วิธีประเพณีนิยม (FR) วิธีตัดตัวลวง (ET) วิธีเลือกชุดตัวถูก (SS) และวิธีการแก้การเดา (GC) โดยศึกษาในกลุ่มนักศึกษา 54 คน ในวิชาการวัดผลการศึกษา ข้อสอบเลือกตอบ มี 4 ตัวเลือก ใช้คะแนนสอบกลางภาค (MT) คะแนน จากวิธีประเพณีนิยม (FR) และคะแนนโครงงานในชั้นเรียนเป็นเกณฑ์ในการศึกษาความตรง ผลการวิจัยพบว่า ค่าความตรงและค่าความเที่ยงของ 4 วิธี ไม่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ส่วนการให้เกรด พบว่าวิธีตัดตัวลวงและวิธีเลือกชุดตัวถูก ทำให้ผู้สอบได้เกรด Aและ B สูงกว่าวิธีแก้การเดา

จากเอกสารและงานวิจัยที่เกี่ยวข้องกับวิธีการตอบและตรวจให้คะแนนของแบบทดสอบ พบว่า มีนักวิจัยได้พัฒนาและเปรียบเทียบคุณภาพของวิธีการให้ความรู้บางส่วนไว้หลายวิธี ซึ่งบางวิธีสามารถนำไปใช้ในการวัดและประเมินความรู้ของผู้เรียนได้จริงโดยไม่ยุ่งยากซับซ้อนในทางปฏิบัติ แต่ยังไม่ทราบแน่ชัดว่าวิธีการตอบและตรวจให้คะแนนแบบทดสอบเลือกตอบแบบธรรมดาและแบบทดสอบเลือกตอบแบบให้คะแนนความมั่นใจ แบบใดจะทำให้ได้แบบทดสอบที่มีคุณภาพ ประหยัดเวลา และลดความยุ่งยากในการสร้างแบบทดสอบ ผู้วิจัยจึงสนใจที่จะศึกษา

คุณภาพของแบบทดสอบเลือกตอบเมื่อตรวจด้วยวิธีการให้คะแนนความรู้บางส่วนว่าวิธีตรวจให้คะแนนวิธีใดให้ค่าความเที่ยงตรงตามสภาพและค่าฟังชันสารสนเทศของแบบทดสอบและข้อสอบสูงที่สุด เมื่อนำวิธีการให้คะแนนความรู้บางส่วนที่พัฒนาขึ้นและมีการตรวจสอบคุณภาพโดยทฤษฎีการตอบสนองข้อสอบ (IRT) ซึ่งเป็นประเด็นที่ผู้วิจัยจะได้ทำการศึกษาวิจัยต่อไป

บทที่ 3
วิธีดำเนินการวิจัย

การดำเนินการวิจัยครั้งนี้มีวัตถุประสงค์ 1) เพื่อหาคุณภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ ในด้านฟังก์ชันสารสนเทศ ความยากและอำนาจจำแนก ความตรงตามสภาพ 2) เพื่อเปรียบเทียบคุณภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ ในด้านดังต่อไปนี้ 2.1 ฟังก์ชันสารสนเทศ 2.2 ความตรงตามสภาพในการวิจัยครั้งนี้มีขั้นตอนการดำเนินการวิจัย ดังนี้

ตอนที่ 1 ประชากรและกลุ่มตัวอย่าง

ตอนที่ 2 เครื่องมือที่ใช้ในการวิจัย

ตอนที่ 3 การเก็บรวบรวมข้อมูล

ตอนที่ 4 การวิเคราะห์ข้อมูล

ตอนที่ 5 สถิติที่ใช้ในการวิเคราะห์ข้อมูล

ตอนที่ 1 ประชากรและกลุ่มตัวอย่าง

ประชากร

ประชากรที่ใช้ในการวิจัยครั้งนี้ เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 4 ภาคเรียนที่ 2 ปีการศึกษา 2560 ของโรงเรียนในสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 18 ซึ่งมีทั้งหมด 11,060 คน จากโรงเรียนทั้งหมด 50 โรงเรียน

กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 4 ภาคเรียนที่ 2 ปีการศึกษา 2560 ของโรงเรียนในสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 18 โดยมีประชากร 11,060 คน ผู้วิจัยใช้โปรแกรม G*Power 3.1 เลือกวิธีการวิเคราะห์ทางสถิติ t-test เลือกวิธีการวิเคราะห์อำนาจการทดสอบด้วย Correlation: Point biserial model ค่าขนาดอิทธิพล (Effect size) $|p| = 0.15$ ความน่าจะเป็นของความคลาดเคลื่อนในการทดสอบประเภทที่หนึ่ง $\alpha = 0.05$ อำนาจการทดสอบ $(1-\beta) = 0.95$ ได้ขนาดกลุ่มตัวอย่าง 567 คน สำหรับการวิจัยครั้งนี้ผู้วิจัยใช้กลุ่มตัวอย่าง 658 คนใช้เทคนิคการสุ่มตัวอย่างแบบหลายขั้นตอน (Multi-stage sampling) การสุ่มกลุ่มตัวอย่างมีขั้นตอน ดังนี้

สำรวจข้อมูลหน่วยสมาชิกของประชากรชั้นนักเรียนชั้นมัธยมศึกษาปีที่ 4 ปีการศึกษา 2560 ของโรงเรียนในสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 18 จัดทำกรอบการสุ่ม (Sampling frame) โดยอาศัยลักษณะการแบ่งเขตการปกครองของจังหวัดชลบุรีและจังหวัดระยอง จังหวัดชลบุรีมีจำนวน 34 โรงเรียน จำนวน 554 ห้องเรียน จำนวน 7,202 คน และจังหวัดระยอง มีจำนวน 19 โรงเรียน จำนวน 333 ห้องเรียน จำนวน 3,858 คน รวมทั้งสิ้น 11,060 คน และโดยใช้เทคนิคการสุ่มตัวอย่างแบบหลายขั้นตอน (Multi-stage sampling) ผู้วิจัยสุ่มโรงเรียนที่แบ่งตามขนาดโดยใช้วิธีการสุ่มแบบแบ่งชั้น (Stratified random sampling) ใช้โรงเรียนเป็นหน่วยการสุ่ม มีขนาดของโรงเรียนเป็นชั้น (Strata) ได้โรงเรียน 9 โรงเรียน สุ่มห้องเรียนได้ 21 ห้องเรียน สุ่มนักเรียนในแต่ละโรงเรียนในระดับชั้นมัธยมศึกษาปีที่ 4 ภาคเรียนที่ 2 โดยใช้วิธีการสุ่มกลุ่มตัวอย่างแบบยกกลุ่ม (Cluster random sampling) โดยจำแนกตามขนาดโรงเรียนดังตารางที่ 3-1

ตารางที่ 3-1 รายชื่อโรงเรียนและจำนวนนักเรียนที่ใช้เป็นกลุ่มตัวอย่าง จำแนกตามขนาดโรงเรียน และจุดประสงค์การทดสอบ

จุดประสงค์การทดสอบ	ขนาดโรงเรียน	ชื่อโรงเรียน	ประชากร		กลุ่มตัวอย่าง		รวม
			ห้อง	คน	ห้อง	คน	
เพื่อทดลองใช้ครั้งที่ 1	ใหญ่พิเศษ	โพธิสัมพันธ์พิทยาคาร	12	647	2	50	50
เพื่อทดลองใช้ครั้งที่ 2	ใหญ่พิเศษ	ชลราษฎรอำรุง	16	723	2	112	210
	ใหญ่	วังจันทร์วิทยา	6	263	1	40	
	กลาง	เกาะโพธิ์ถ้วยงาม	3	103	1	33	
	เล็ก	ผินแจ่มวิชาสอน	2	22	1	25	
นำไปใช้จริง	ใหญ่พิเศษ	สิงห์สมุทร	16	707	8	376	658
	ใหญ่	สัตหีบวิทยาคม	7	259	3	142	
	กลาง	พานทอง	4	150	2	100	
	เล็ก	ทุ่งเหียงพิทยาคม	2	50	1	40	

ตอนที่ 2 การสร้างเครื่องมือที่ใช้ในการวิจัยและขั้นตอนการสร้างเครื่องมือ

เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยครั้งนี้ คือ แบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชา เคมี กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 4 ภาคเรียน 2 ปีการศึกษา 2560 เรื่อง ปริมาณสารสัมพันธ์ ซึ่งเป็นแบบทดสอบชนิดเลือกตอบ 4 ตัวเลือก ที่ผู้วิจัยสร้างขึ้น 1 ฉบับ จำนวน 30 ข้อ

แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา (Multiple choice) หมายถึง

แบบทดสอบที่แต่ละข้อคำถามประกอบด้วย ตอนนำ หรือ ตัวคำถาม (Stem or lead or problem) กับ ตัวเลือก (Choice or option or alternative or response) ในแต่ละข้อคำถาม มีตัวเลือกมากกว่า 2 ตัวเลือก สำหรับการศึกษาครั้งนี้มี 4 ตัวเลือก และมีตัวเลือกถูกเพียงตัวเลือกเดียว เพื่อให้ผู้สอบตอบสนองออกมา สังเกตได้และนับจำนวนได้ โดยมีวิธีการตรวจให้คะแนนแบบธรรมดา คือ วิธีการให้ผู้สอบเลือกคำตอบที่ถูกที่สุดเพียงข้อเดียว โดยมีการตรวจให้คะแนน 1 คะแนน เมื่อผู้สอบเลือกตัวเลือกที่ถูก และตรวจให้คะแนน 0 คะแนน เมื่อผู้สอบเลือกตอบตัวเลือกผิดหรือไม่เลือกหรือเลือกตัวเลือกมากกว่า 1 ตัวเลือก

เกณฑ์การให้คะแนนแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา ดังนี้

ผู้สอบเลือกตัวเลือก “ถูก” ได้คะแนนเท่ากับ 1

ผู้สอบเลือกตัวเลือก “ผิด” ได้คะแนนเท่ากับ 0

ตัวอย่างแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา

คำชี้แจง วิธีการตอบและให้คะแนนแบบทดสอบเลือกตอบแบบธรรมดา ให้นักเรียนทำแบบสอบโดยเลือกตอบเพียง 1 ตัวเลือก โดยมีการตรวจให้คะแนน 1 คะแนน เมื่อนักเรียนเลือกตัวเลือกที่ถูก และตรวจให้คะแนน 0 คะแนน เมื่อนักเรียนเลือกตอบตัวเลือกผิดหรือไม่เลือกหรือเลือกตัวเลือกมากกว่า 1 ตัวเลือก

ข้อสอบข้อที่ (0) คำตอบที่ถูกต้อง คือ ข้อ ข จะได้คะแนน 1 คะแนน

ข้อ	ตัวเลือก				คะแนน
	ก	ข	ค	ง	
0		✘			1

ข้อสอบข้อที่ (0) คำตอบที่ถูกต้อง คือ ข้อ ข ถ้าผู้สอบตอบข้อ ก จะได้คะแนน 0 คะแนน

ข้อ	ตัวเลือก				คะแนน
	ก	ข	ค	ง	
0	✘				0

แบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ (Confidential making) หมายถึง แบบทดสอบที่มีคำชี้แจงให้นักเรียนทำเครื่องหมายแก่ตัวเลือกที่ถูกที่สุดเพียงตัวเลือกเดียว พร้อมทั้งบอกความมั่นใจในการตอบตัวเลือกนั้นด้วยว่ามาก ปานกลาง หรือน้อย โดยมีวิธีการตรวจให้คะแนนความมั่นใจ คือการคาดคะเนของผู้สอบว่าคำตอบที่เลือกนั้นมีโอกาสเป็นคำตอบที่ถูกต้อง โดยระบุความมั่นใจในการตอบว่า มาก ปานกลาง หรือน้อยอย่างใดอย่างหนึ่ง ซึ่งคะแนนตามระดับความมั่นใจ 3 ระดับ การกำหนดคะแนนรายข้อที่ให้คะแนนแต่ละข้อเป็นไปตามเกณฑ์ระดับความมั่นใจ มาก ปานกลาง น้อย โดยข้อที่ไม่ตอบและตอบผิดจะได้ 0 คะแนน

เกณฑ์การให้คะแนนแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ

ผู้สอบเลือกตัวเลือก “ ถูก ” ในระดับความมั่นใจ “มาก” ได้คะแนนเท่ากับ 3
ผู้สอบเลือกตัวเลือก “ ถูก ” ในระดับความมั่นใจ “ ปานกลาง” ได้คะแนนเท่ากับ 2
ผู้สอบเลือกตัวเลือก “ ถูก ” ในระดับความมั่นใจ “น้อย” ได้คะแนนเท่ากับ 1
ผู้สอบเลือกตัวเลือก “ ผิด ” ในระดับความมั่นใจ “มาก” ได้คะแนนเท่ากับ 0
ผู้สอบเลือกตัวเลือก “ ผิด ” ในระดับความมั่นใจ “ปานกลาง” ได้คะแนนเท่ากับ 0
ผู้สอบเลือกตัวเลือก “ ผิด ” ในระดับความมั่นใจ “น้อย” ได้คะแนนเท่ากับ 0

ตัวอย่างแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ

คำชี้แจง วิธีการตอบและให้คะแนนแบบพิจารณาความมั่นใจให้นักเรียนทำแบบสอบโดยเลือกตอบเพียง 1 ตัวเลือกโดยบอกระดับความมั่นใจของคำตอบที่ตอบไป ซึ่งกระดาษคำตอบที่จะให้นักเรียนทำข้อสอบในการตอบวิธีนี้เป็นกระดาษคำตอบที่มีสองสองส่วน คือ ส่วนแรกเป็นส่วนที่ให้ผู้ตอบทำเครื่องหมาย ✘ ลงในช่องของตัวเลือกที่ต้องการตอบ ส่วนที่สองเป็นส่วนที่ให้ทำเครื่องหมาย ✘ เพื่อระบุระดับความมั่นใจ ตัวอย่างเช่น

ข้อสอบข้อที่ (0) คำตอบที่ถูกต้อง คือ ข้อ ข ผู้สอบตอบข้อ ข และระบุระดับความมั่นใจมากจะได้คะแนน 3 คะแนน

ข้อ	ตัวเลือก				ระดับความมั่นใจในการตอบ			คะแนน
	ก	ข	ค	ง	มาก	ปานกลาง	น้อย	
0		✗			✗			3

ข้อสอบข้อที่ (0) คำตอบที่ถูกต้อง คือ ข้อ ข ถ้าผู้สอบตอบข้อ ข และระบุระดับความมั่นใจปานกลางจะได้คะแนน 2 คะแนน

ข้อ	ตัวเลือก				ระดับความมั่นใจในการตอบ			คะแนน
	ก	ข	ค	ง	มาก	ปานกลาง	น้อย	
0		✗				✗		2

ข้อสอบข้อที่ (0) คำตอบที่ถูกต้อง คือ ข้อ ข ถ้าผู้สอบตอบข้อ ข และระบุระดับความมั่นใจน้อย จะได้คะแนน 1 คะแนน

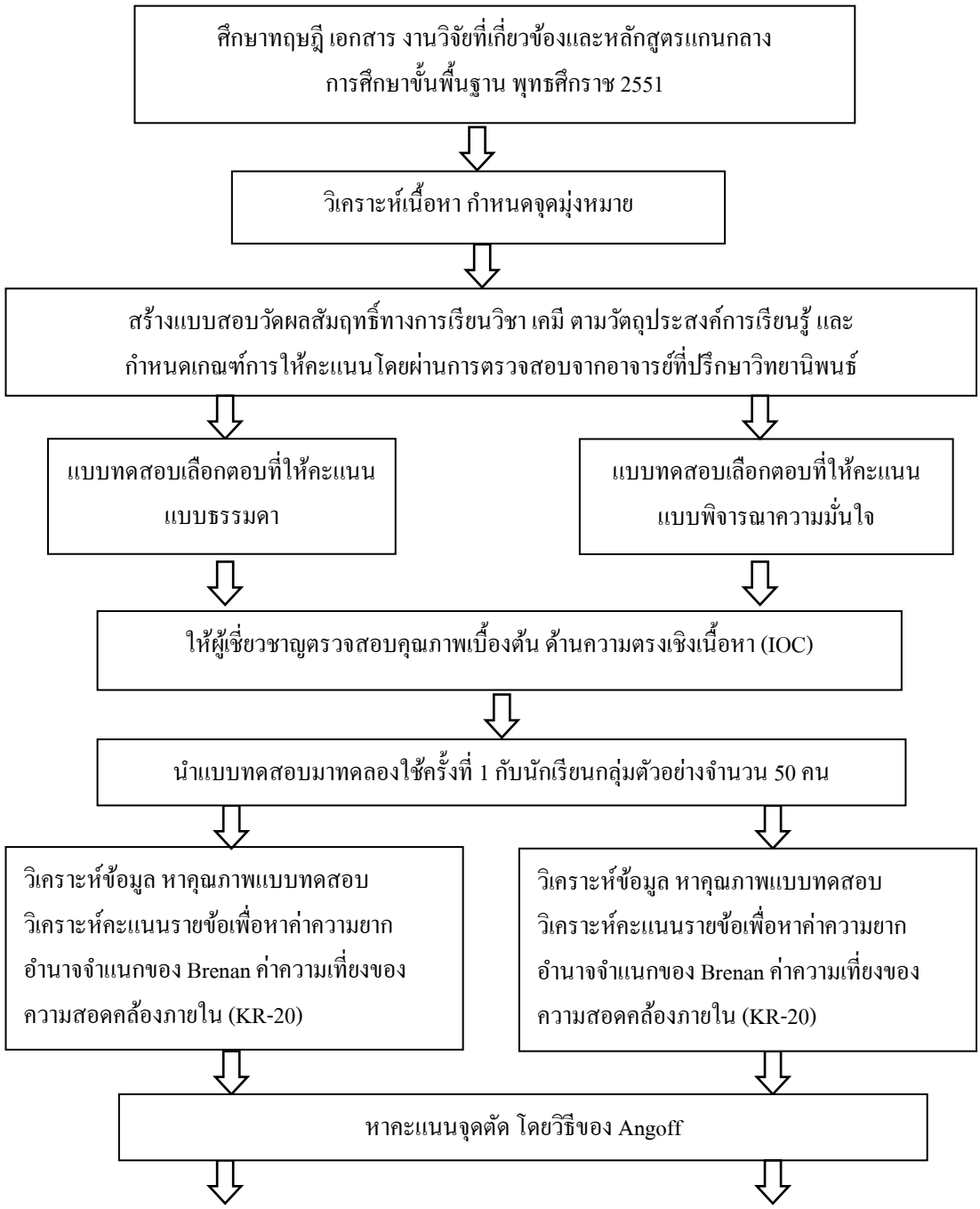
ข้อ	ตัวเลือก				ระดับความมั่นใจในการตอบ			คะแนน
	ก	ข	ค	ง	มาก	ปานกลาง	น้อย	
0		✗					✗	1

ข้อสอบข้อที่ (0) คำตอบที่ถูกต้อง คือ ข้อ ข ถ้าผู้สอบตอบผิดคือไม่ได้เลือกข้อ ข จะได้คะแนน 0 คะแนน ไม่ว่าจะระบุระดับความมั่นใจในการตอบระดับใด

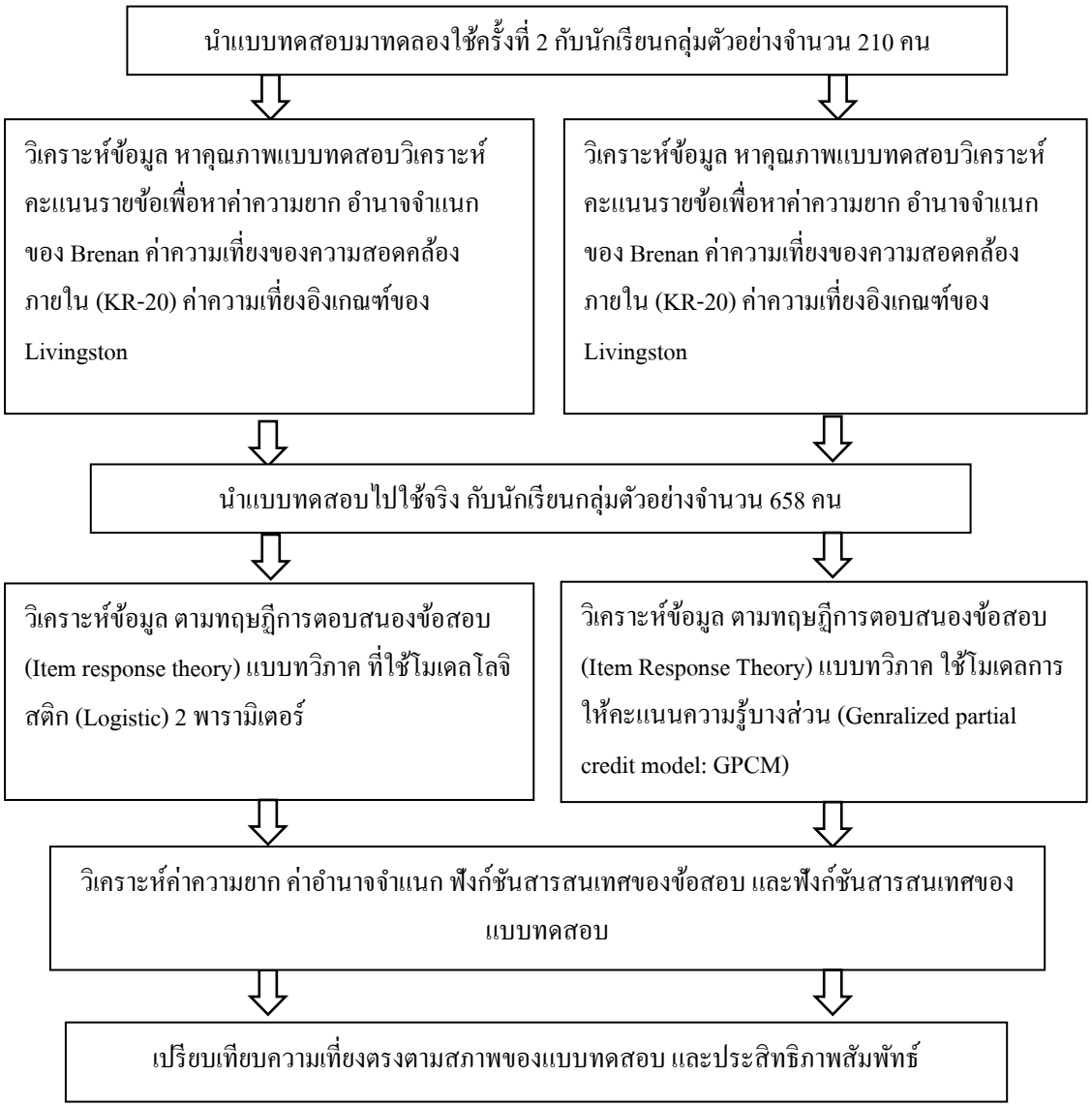
ข้อ	ตัวเลือก				ระดับความมั่นใจในการตอบ			คะแนน
	ก	ข	ค	ง	มาก	ปานกลาง	น้อย	
0	✗				✗			0

ขั้นตอนการสร้างเครื่องมือ

การดำเนินการสร้างเครื่องมือวัดแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชา เคมี กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 4 ภาคเรียนที่ 2 ปีการศึกษา 2560 เรื่อง ปริมาณสารสัมพันธ์ จำนวน 1 ฉบับ มีขั้นตอนการสร้าง ดังนี้



ภาพที่ 3-1 ขั้นตอนในการสร้างแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชา เคมี



ภาพที่ 3-1 (ต่อ)

ในการสร้างเครื่องมือวัดแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชา เคมี อาศัยเอกสารหลักสูตรและการวิเคราะห์หลักสูตร เพื่อที่จะระบุจุดมุ่งหมายในการชี้แนวทางการเลือกตัวแทนของเนื้อหาที่จะออกข้อสอบว่าออกข้อสอบอย่างไรจึงจะเป็นตัวแทนที่ดีของเนื้อหา วิธีการนี้จะเป็นหลักประกันได้ว่าแบบทดสอบนั้นจะมีความเที่ยงตรงด้านเนื้อหา (Content validity) มีการดำเนินการสร้างแบบทดสอบ ดังนี้

1. ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการสร้างแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชา เคมีเพื่อเป็นแนวทางการสร้างแบบทดสอบ

2. ศึกษาหลักสูตรคู่มือครูและแบบเรียนวิชาเคมี ช่วงชั้นที่ 4 ชั้นมัธยมศึกษาปีที่ 4 เพื่อวิเคราะห์เนื้อหา ผลการเรียนรู้ สาระทักษะการคิดคำนวณเกี่ยวกับคำนวณเกี่ยวกับมวลอะตอม มวลโมเลกุล โมล ความสัมพันธ์ระหว่างมวลสารกับจำนวนโมล ความสัมพันธ์ระหว่างปริมาตรต่อโมลของแก๊ส ความสัมพันธ์ระหว่างจำนวนอนุภาคกับจำนวนโมล ความเข้มข้นของสารละลาย การเตรียมสารละลาย และสมบัติบางประการของสารละลาย มวลสารในปฏิกิริยาเคมี ปริมาณของแก๊สในปฏิกิริยาเคมี ได้แก่กฎของเกย์-ลุสแซก กฎของอาโวกาโดร การคำนวณหาสูตรเอมพิริกัล สูตรโมเลกุล การคำนวณหามวลเป็นร้อยละจากสูตร สมการเคมี ความสัมพันธ์ระหว่างปริมาณสารของสารในสมการเคมี

3. กำหนดตารางโครงสร้างโดยมีรายละเอียด คือ รายละเอียดของเนื้อหา ผลการเรียนรู้และทักษะที่มุ่งวัด กำหนดน้ำหนักเป็นร้อยละตามเนื้อหา ผลการเรียนรู้ จำแนกตามทักษะและระบุจำนวนข้อตามเนื้อหาและทักษะกระบวนการทางเคมี ที่กำหนดนั้น ๆ ซึ่งในการเขียนข้อสอบครั้งแรกผู้วิจัยได้กำหนดให้ในแต่ละผลการเรียนรู้มีข้อสอบจำนวนมากกว่าข้อสอบที่ต้องการใช้จริง (30 ข้อ) ทำให้มีข้อสอบในครั้งแรกจำนวน 40 ข้อ

ตารางที่ 3-2 โครงสร้างเนื้อหาแบบทดสอบประกอบด้วย รายละเอียดของเนื้อหา ผลการเรียนรู้และทักษะที่มุ่งวัด

หน่วยที่	ผลการเรียนรู้	น้ำหนักเฉลี่ย	จำนวนข้อ	เลขที่ข้อ
	ปริมาณสัมพันธ์			
1	1. มวลอะตอม และมวลโมเลกุล			
	1.1 อธิบายความหมายของมวลอะตอม มวลโมเลกุล คำนวณหามวลอะตอมของธาตุ	22	9	1-9
	1.2 มวลโมเลกุลของสารมวลของธาตุ 1 อะตอม มวลอะตอมเฉลี่ยของธาตุ และมวลของสาร 1 โมเลกุล			
2	2. โมล	25	10	10-19
	2.1 อธิบาย และคำนวณความสัมพันธ์ระหว่างโมล มวล จำนวนอนุภาคและปริมาตรของแก๊ส			

ตารางที่ 3-2 (ต่อ)

หน่วยที่	ผลการเรียนรู้	น้ำหนักเฉลี่ย	จำนวนข้อ	เลขที่ข้อ
3	3. สารละลายและความเข้มข้น 3.1 อธิบายสมบัติบางประการของสารละลาย ทำการทดลอง เตรียมสารละลาย และคำนวณความเข้มข้นของสารละลายในหน่วยร้อยละ โมลาริตี โมแลล ส่วนในล้านส่วน และเศษส่วนโมล 3.2 อธิบายและทำการทดลองเกี่ยวกับกฎทรงมวลของสารและกฎสัดส่วนคงที่ 3.3 สืบค้นข้อมูล อธิบาย และทำการทดลองความสัมพันธ์ของปริมาตรแก๊สในสมการเคมี ใช้กฎของเกย์ลุสแซก กฎของอาโวกาโดร คำนวณหาปริมาตรของแก๊สในปฏิกิริยาเคมี	38	15	20-34
4	4. การคำนวณเกี่ยวกับสูตรเคมี 4.1 อธิบายและคำนวณเกี่ยวกับสูตร และสมการเคมี 4.2 ใช้ความสัมพันธ์ระหว่างจำนวนโมล กับมวลหรือปริมาตรของสาร คำนวณหาค่าโมล มวลหรือปริมาตรของสารสมการเคมี และหาค่าร้อยละของผลที่ได้จาการทดลองได้	15	6	35-40
	รวม	100	40	1-40

4. กำหนดตารางโครงสร้างการวิเคราะห์แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชา เคมี ชั้นมัธยมศึกษาปีที่ 4 ภาคเรียนที่ 2 ปีการศึกษา 2560 เรื่อง ปริมาณสารสัมพันธ์ ตามผลการเรียนรู้ และกำหนดจำนวนข้อสอบที่ใช้จริงจากจำนวนข้อสอบที่ออกทั้งหมดจากน้ำหนักเป็นร้อยละตามเนื้อหา ผลการเรียนรู้ โดยจำแนกตามพฤติกรรมที่ต้องการวัด

5. กำหนดตารางโครงสร้างการวิเคราะห์หน่วยการเรียนรู้ วัตถุประสงค์การเรียนรู้ และจำนวนข้อสอบ จากน้ำหนักเป็นร้อยละตามเนื้อหา ผลการเรียนรู้ โดยจำแนกตามพฤติกรรมที่ต้องการวัด และให้ผู้เชี่ยวชาญที่มีความชำนาญในเนื้อหาวิชา เคมี ตรวจสอบผลการเรียนรู้ให้ตรงตามพฤติกรรมที่ต้องการวัด เช่น ความรู้/ ความจำ ความเข้าใจ การนำไปใช้ การวิเคราะห์

ตารางที่ 3-3 การวิเคราะห์แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน วิชา เคมี เรื่อง ปริมาณสารสัมพันธ์

หน่วยการเรียนรู้	ผลการเรียนรู้	จำนวนข้อสอบที่ออกทั้งหมด	จำนวนข้อสอบที่ใช้จริง
1. มวลอะตอม และมวลโมเลกุล	1. สามารถอธิบายความหมายของมวลอะตอม มวลโมเลกุล คำนวณหามวลอะตอมของธาตุได้	2	2
	2. สามารถอธิบายความหมายของมวลโมเลกุลของสารมวลของธาตุ 1 อะตอม มวลอะตอมเฉลี่ยของธาตุ และมวลของสาร 1 โมเลกุลได้	7	5
2. โมล	1. สามารถอธิบายความสัมพันธ์ระหว่างโมล มวล จำนวนอนุภาคและปริมาตรของแก๊สได้	2	2
	2. สามารถคำนวณความสัมพันธ์ระหว่างโมล มวล จำนวนอนุภาคและปริมาตรของแก๊สได้	8	5
3. สารละลายและความเข้มข้น	1. สามารถอธิบายสมบัติบางประการของสารละลาย ทำการทดลอง เตรียมสารละลายได้	3	2
	2. สามารถคำนวณความเข้มข้นของสารละลายในหน่วยร้อยละ โมลาริตี โมแลล ส่วนในล้านส่วน ส่วนในพันล้านส่วนและเศษส่วนโมลได้	7	5
	3. สามารถอธิบายและทำการทดลองเกี่ยวกับกฎทรงมวลของสารและกฎสัดส่วนคงที่ได้	2	1
	4. สามารถอธิบายและทำการทดลองความสัมพันธ์ของปริมาตรแก๊สในสมการเคมี ใช้กฎของเกย์ลุสแซก กฎของอาโวกาโดรได้	1	1
	5. สามารถคำนวณหาปริมาตรของแก๊สในปฏิกิริยาเคมีได้	2	1

ตารางที่ 3-3 (ต่อ)

หน่วยการเรียนรู้	ผลการเรียนรู้	จำนวนข้อสอบที่ออกทั้งหมด	จำนวนข้อสอบที่ใช้จริง
4. การคำนวณเกี่ยวกับสูตรเคมี	1. สามารถอธิบายเกี่ยวกับสูตร และสมการเคมีได้	1	1
	2. สามารถคำนวณเกี่ยวกับสูตร และสมการเคมีได้	3	3
	3. สามารถใช้ความสัมพันธ์ระหว่างจำนวนโมล กับมวลหรือปริมาตรของสาร คำนวณหาค่าโมล มวลหรือปริมาตรของสาร สมการเคมี และหาค่าร้อยละของผลที่ได้จากการทดลองได้	2	2
	รวม	40	30

ตารางที่ 3-4 การวิเคราะห์หน่วยการเรียนรู้ ผลการเรียนรู้ และจำนวนข้อสอบ ตามพฤติกรรมที่ต้องการวัด

หน่วยการเรียนรู้	ผลการเรียนรู้	พฤติกรรมที่ต้องการวัด				รวม
		ความรู้/ความจำ	ความเข้าใจ	การนำไปใช้	การวิเคราะห์	
1. มวลอะตอม และ มวลโมเลกุล	1. สามารถอธิบายความหมายของมวลอะตอม มวลโมเลกุล คำนวณหามวลอะตอมของธาตุได้	(2)				(2)
	2. สามารถอธิบายความหมายของมวลโมเลกุลของสารมวลของธาตุ 1 อะตอม มวลอะตอมเฉลี่ยของธาตุ และมวลของสาร 1 โมเลกุลได้	(7)				(7)
2. โมล	3. สามารถอธิบายความสัมพันธ์ระหว่างโมล มวล จำนวนอนุภาคและปริมาตรของแก๊สได้		(2)			(2)
	4. สามารถคำนวณความสัมพันธ์ระหว่างโมล มวล จำนวนอนุภาคและปริมาตรของแก๊สได้			(8)		(8)
3. สารละลายและความเข้มข้น	5. สามารถอธิบายสมบัติบางประการของสารละลาย ทำการทดลอง เตรียมสารละลายได้	(1)		(1)	(1)	(3)
	6. สามารถคำนวณความเข้มข้นของสารละลายในหน่วยร้อยละโมลาริตี โมแลล ส่วนในล้านส่วน ส่วนในพันล้านส่วนและเศษส่วนโมลได้			(7)		(7)

ตารางที่ 3-4 (ต่อ)

หน่วยการเรียนรู้	ผลการเรียนรู้	พฤติกรรมที่ต้องการวัด			รวม
		ความรู้/ความจำ	ความเข้าใจ การนำไปใช้	การวิเคราะห์	
	7. สามารถอธิบายและทำการทดลองเกี่ยวกับกฎทรงมวลของสารและกฎสัดส่วนคงที่ได้		(1)	(1)	(2)
	8. สามารถอธิบายและทำการทดลองความสัมพันธ์ของปริมาตรแก๊สในสมการเคมี ใช้กฎของเกย์ลุสแซก กฎของอาโวกาโดรได้			(1)	(1)
	9. สามารถคำนวณหาปริมาตรของแก๊สในปฏิกิริยาเคมีได้		(2)		(2)
4. การคำนวณเกี่ยวกับสูตรเคมี	10. สามารถอธิบายเกี่ยวกับสูตร และสมการเคมีได้	(1)			(1)
	11. สามารถคำนวณเกี่ยวกับสูตร และสมการเคมีได้		(3)		(3)
	12. สามารถใช้ความสัมพันธ์ระหว่างจำนวนโมล กับมวลหรือปริมาตรของสาร คำนวณหาค่าโมล มวลหรือปริมาตรของสารสมการเคมี และหาค่าร้อยละของผลที่ได้จากการทดลองได้		(1)	(1)	(2)

6. สร้างข้อสอบตามเนื้อหาได้แบบทดสอบวัดผลสัมฤทธิ์วิชาเคมี ตามผลการเรียนรู้ สาระทักษะการคิดคำนวณเกี่ยวกับคำนวณเกี่ยวกับมวลอะตอม มวลโมเลกุล โมล ความสัมพันธ์ระหว่างมวลสารกับจำนวนโมล ความสัมพันธ์ระหว่างปริมาตรต่อโมลของแก๊ส ความสัมพันธ์ระหว่างจำนวนอนุภาคกับจำนวนโมล ความเข้มข้นของสารละลาย การเตรียมสารละลาย และสมบัติบางประการของสารละลาย มวลสารในปฏิกิริยาเคมี ปริมาณของแก๊สในปฏิกิริยาเคมี ได้แก่ กฎของเกย์-ลุสแซก กฎของอาโวกาโดร การคำนวณหาสูตรเอมพิริกัล สูตรโมเลกุล การคำนวณหามวลเป็นร้อยละจากสูตร สมการเคมี ความสัมพันธ์ระหว่างปริมาณสารของสารในสมการเคมี จำนวน 1 ฉบับ ชนิดเลือกตอบ 4 ตัวเลือก จำนวน 40 ข้อ โดยมีวิธีการตอบและตรวจให้คะแนน 2 วิธี ซึ่งแบ่งเป็น 2 ฉบับ

ฉบับที่ 1 เป็นแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา

ฉบับที่ 2 เป็นแบบทดสอบเลือกตอบที่ให้คะแนนแบบแบบพิจารณาความมั่นใจ

7. นำเครื่องมือที่สร้างขึ้นให้ผู้เชี่ยวชาญ จำนวน 5 คน ตรวจสอบคุณภาพเบื้องต้นด้านความตรงเชิงเนื้อหา (Content validity) ของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ โดยการนำแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาเคมี จำนวน 40 ข้อ ที่ผู้วิจัยสร้างขึ้นเพื่อให้ผู้เชี่ยวชาญประเมินว่าข้อสอบแต่ละข้อนั้น สามารถวัดได้ตรงกับเนื้อหาหรือไม่ และกำหนดค่าน้ำหนักคะแนนของตัวเลือกตามวิธีการตรวจให้คะแนนทั้ง 2 วิธี นำแบบทดสอบหาค่าดัชนีความสอดคล้องเป็นรายข้อ (Index of congruency: IOC) พบว่ามีค่าอยู่ระหว่าง 0.6-1.0 ตัดข้อสอบที่ค่าดัชนีความสอดคล้องต่ำกว่า 0.6 ทำให้ได้แบบทดสอบทั้งสิ้น จำนวน 30 ข้อ

8. นำแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชา เคมี 1 ฉบับ จำนวน 30 ข้อ ที่มีวิธีการตอบและตรวจให้คะแนนทั้ง 2 วิธีนำไปทดลองใช้เพื่อวิเคราะห์คุณภาพโดยใช้ทฤษฎีการวัดแบบดั้งเดิม (Classical test theory: CTT) ทดลองใช้ครั้งที่ 1 กับนักเรียนชั้นมัธยมศึกษาปีที่ 4 ที่กำลังศึกษาอยู่ในภาคเรียนที่ 2 ปีการศึกษา 2560 ของโรงเรียนในสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาเขต 18 นักเรียนกลุ่มตัวอย่างจำนวน 50 คน โดยนำแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาเคมี 1 ฉบับ จำนวน 30 ข้อ นำไปทดสอบและซักซ้อมความเข้าใจคำชี้แจงในการทำแบบทดสอบแต่ละวิธี โดยกำหนดให้เว้นระยะเวลาในการทำแบบทดสอบที่มีวิธีการตอบและตรวจให้คะแนนแต่ละวิธีอย่างน้อย 7 วัน และใช้การจับเวลา เพื่อหาเวลาที่ใช้ในการทำแบบทดสอบโดยเฉลี่ย ใน 1 คาบเรียน ได้เวลาโดยเฉลี่ย 51 นาที โดยผู้วิจัยได้เลือก โรงเรียนโพธิสัมพันธ์พิทยาคาร ซึ่งเป็นโรงเรียนที่มีสภาพการเรียนการสอนเช่นเดียวกับกลุ่มตัวอย่าง โดยการทดลองใช้ครั้งนี้มีวัตถุประสงค์เพื่อ

1) วิเคราะห์คะแนนรายข้อเพื่อหาค่าความยากและค่าอำนาจจำแนกของแบบทดสอบ 2) หาค่าความเที่ยงของความสอดคล้องภายใน

9. คัดเลือกข้อสอบที่มีคุณภาพให้ได้จำนวนข้อตามที่กำหนดไว้ในแต่ละจุดประสงค์จากแบบทดสอบ โดยเลือกข้อสอบที่มีค่าความยาก และค่าอำนาจจำแนกรายข้อตามเกณฑ์ที่กำหนด คือ มีค่าความยากตั้งแต่ 0.2-0.8 ค่าอำนาจำแนกตั้งแต่ 0.2-1.00 (ล้วน สายยศ และอังคณา สายยศ, 2536, หน้า 179) ได้ข้อสอบที่มีคุณสมบัติตามเกณฑ์ทั้งหมดและปรับปรุงแบบทดสอบที่มีค่าความยากและค่าอำนาจจำแนกไม่ผ่านเกณฑ์ ให้ได้แบบทดสอบเลือกตอบ 1 ฉบับ จำนวน 30 ข้อ เพื่อได้แบบทดสอบที่นำไปใช้ทดสอบจริง คัดเลือกข้อสอบตามผลการเรียนรู้ ต้องอาศัยข้อสอบที่มีค่าความยากค่อนข้างสูงเพื่อให้สามารถจำแนกกลุ่มผู้สอบที่มีความรอบรู้ และไม่รอบรู้ในแต่ละผลการเรียนรู้ออกจากกันได้อย่างชัดเจน

10. ผู้วิจัยนำแบบทดสอบที่คัดเลือกและปรับปรุงแก้ไขแล้วจากการสอบครั้งที่ 1 มาปรับปรุงคุณภาพและนำไปทดลองใช้ครั้งที่ 2 กับเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 4 ภาคเรียนที่ 2 ปีการศึกษา 2560 ของโรงเรียนในสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 18 จำนวน 5 ห้องเรียน นักเรียกลุ่มตัวอย่าง 210 คน ทดสอบและซักซ้อมความเข้าใจคำชี้แจงในการทำแบบทดสอบแต่ละวิธี โดยกำหนดให้เว้นระยะเวลาในการทำแบบทดสอบที่มีวิธีการตอบและตรวจให้คะแนนแต่ละวิธีอย่างน้อย 7 วัน และจับเวลาในการทำแบบทดสอบซึ่งนักเรียนสามารถทำแบบทดสอบให้เสร็จได้ภายใน เวลาเฉลี่ย 51 นาที

11. กำหนดคะแนนจุดตัดของแบบทดสอบอิงเกณฑ์ ในการวิจัยครั้งนี้ผู้วิจัยเลือกใช้การกำหนดคะแนนจุดตัดตามวิธีของแองกอฟ วิธีนี้จะอาศัยหลักการของความน่าจะเป็นที่นักเรียนซึ่งมีสมรรถภาพขั้นต่ำสุดที่จะยอมรับได้ตอบข้อสอบถูก โดยนำไปให้ผู้เชี่ยวชาญเนื้อหาวัดผลสัมฤทธิ์หรือครูผู้สอนพิจารณา ค่าความน่าจะเป็นในการตอบข้อสอบถูก จะนำมาใช้เป็นคะแนนจุดตัด วิธีการหาคะแนนจุดตัดมีลำดับขั้นดังนี้

11.1 นำข้อสอบทั้งหมดไปให้ครูผู้สอนกลุ่มหนึ่งพิจารณาเนื้อหาข้อสอบและความยาก ซึ่งผู้วิจัยยึดหลักเกณฑ์ในการเลือกผู้เชี่ยวชาญคือต้องเป็นผู้ที่มีความเชี่ยวชาญในเนื้อหาที่จะทำการกำหนดคะแนนจุดตัดมาก (Impara & Plake, 1997) ผู้วิจัยจึงกำหนดให้ผู้เชี่ยวชาญต้องเป็นครูผู้สอนในรายวิชา เคมี และมีประสบการณ์ไม่ต่ำกว่า 2 ปี จำนวน 5 คน

11.2 ผู้วิจัยสรุปขั้นตอนการหาคะแนนจุดตัด (Cut scores) ด้วยวิธีแองกอฟที่ได้รับการปรับปรุง (Busch & Jaaeger, 1990 อ้างถึงใน สุรีพร อนุศาสนนันท์, 2550; Fehrmann et al., 1991; Hambleton & Plake, 1995; Hess et al., 2007) ให้ครูผู้สอนและแจกคำอธิบายลักษณะ

ความสามารถของนักเรียนในแต่ละระดับ เพื่อให้ผู้ตัดสินเข้าใจตรงกัน และแจกค่าความยากง่ายของข้อสอบแต่ละข้อ

11.3 ดำเนินการกำหนดเปอร์เซ็นต์การตอบข้อสอบแต่ละข้อถูก จากผู้สอบที่มีความสามารถแต่ละระดับ มี 3 ขั้นตอน และนำค่าความน่าจะเป็นในการตอบข้อสอบถูกที่ครูผู้สอนแต่ละคนพิจารณาไว้มาหาค่าเฉลี่ยของความน่าจะเป็น ดังนี้

11.3.1 ขั้นที่ 1 ดำเนินการตัดสินเป็นรายบุคคล ดังนี้

ผู้ตัดสินแต่ละคนพิจารณาข้อสอบทีละข้อตั้งแต่ข้อสอบข้อที่ 1 เพื่อกำหนดคะแนนจุดตัด โดยตัดสินว่า

11.3.1.1 กลุ่มผู้ที่มีความสามารถระดับดีมาก (A) มีโอกาสตอบข้อข้อนั้นถูกกี่เปอร์เซ็นต์ หมายความว่า ผู้ตัดสินคาดว่า นักเรียนที่มีความสามารถระดับดีมาก (A) ประมาณกี่คนจากทั้งหมด 100 คนสามารถตอบข้อสอบข้อนั้นถูก จากนั้นกรอกตัวเลขลงในแบบประมาณเปอร์เซ็นต์การตอบข้อสอบแต่ละข้อถูกที่มีความสามารถระดับดีมาก (A) ในช่องเปอร์เซ็นต์การตอบถูก รอบที่ 1 (รายบุคคล) จากนั้นพิจารณาข้อสอบข้อถัดไปเรื่อย ๆ จนครบจำนวนข้อสอบทั้งหมด

11.3.1.2 กลุ่มผู้ที่มีความสามารถระดับดี (B) มีโอกาสตอบข้อข้อนั้นถูกกี่เปอร์เซ็นต์ หมายความว่า ผู้ตัดสินคาดว่า นักเรียนที่มีความสามารถระดับดี (B) ประมาณกี่คนจากทั้งหมด 100 คนสามารถตอบข้อสอบข้อนั้นถูก จากนั้นกรอกตัวเลขลงในแบบประมาณเปอร์เซ็นต์การตอบข้อสอบแต่ละข้อถูกที่มีความสามารถระดับดี (B) ในช่องเปอร์เซ็นต์การตอบถูก รอบที่ 1 (รายบุคคล) จากนั้นพิจารณาข้อสอบข้อถัดไปเรื่อย ๆ จนครบจำนวนข้อสอบทั้งหมด

11.3.1.3 กลุ่มผู้ที่มีความสามารถระดับพอใช้ (C) มีโอกาสตอบข้อข้อนั้นถูกกี่เปอร์เซ็นต์ หมายความว่า ผู้ตัดสินคาดว่า นักเรียนที่มีความสามารถระดับพอใช้ (C) ประมาณกี่คนจากทั้งหมด 100 คนสามารถตอบข้อสอบข้อนั้นถูก จากนั้นกรอกตัวเลขลงในแบบประมาณเปอร์เซ็นต์การตอบข้อสอบแต่ละข้อถูกที่มีความสามารถระดับพอใช้ (C) ในช่องเปอร์เซ็นต์การตอบถูก รอบที่ 1 (รายบุคคล) จากนั้นพิจารณาข้อสอบข้อถัดไปเรื่อย ๆ จนครบจำนวนข้อสอบทั้งหมด

11.3.1.4 กลุ่มผู้ที่มีความสามารถระดับอ่อน (D) มีโอกาสตอบข้อข้อนั้นถูกกี่เปอร์เซ็นต์ หมายความว่า ผู้ตัดสินคาดว่า นักเรียนที่มีความสามารถระดับอ่อน (D) ประมาณกี่คนจากทั้งหมด 100 คนสามารถตอบข้อสอบข้อนั้นถูก จากนั้นกรอกตัวเลขลงในแบบประมาณเปอร์เซ็นต์การตอบข้อสอบแต่ละข้อถูกที่มีความสามารถระดับอ่อน (D) ในช่องเปอร์เซ็นต์การตอบถูก รอบที่ 1 (รายบุคคล) จากนั้นพิจารณาข้อสอบข้อถัดไปเรื่อย ๆ จนครบจำนวนข้อสอบทั้งหมด

11.3.2 ขั้นที่ 2 ดำเนินการแบ่งผู้ตัดสินเป็นกลุ่มย่อย 2 กลุ่ม ดังนี้

11.3.2.1 ในแต่ละกลุ่มย่อย ผู้ตัดสินร่วมกันอภิปรายผลการตัดสินของละคน โดยเฉพาะผู้ตัดสินที่มีผลตัดสินต่ำสุด และสูงสุดให้อธิบายแสดงความคิดเห็นถึงเหตุผลที่ให้ผลการตัดสินต่ำสุด และสูงสุด และระหว่างการอภิปรายผู้อำนวยความสะดวกจะแสดงรูปการแจกแจงผลการตัดสินของผู้ตัดสินแต่ละคน

11.3.2.2 เมื่อผู้ตัดสินอภิปรายเสร็จ ก็เปิดโอกาสให้กำหนดคะแนนจุดตัดใหม่อีกครั้งหนึ่ง

11.3.2.3 นำผลการตัดสินของผู้ตัดสินทุกคนจากแต่ละกลุ่ม โดยผู้ตัดสินทุกคนจะมีคะแนนจุดตัด 4 ระดับ คือ 1) ระดับดีมาก (A) 2) ระดับดี (B) 3) ระดับพอใช้ (C) 4) ระดับอ่อน (D) นำคะแนนจุดตัดของผู้ตัดสินแต่ละคนมาหาค่าเฉลี่ยในแต่ละระดับ จะได้ค่าเฉลี่ยของคะแนนจุดตัด 4 จุด หรือมีมาตรฐาน 5 ระดับ

11.3.3 ขั้นที่ 3 ดำเนินการรวมผู้ตัดสินทั้งสองกลุ่มย่อยเป็นกลุ่มใหญ่เพียงกลุ่มเดียว ดำเนินการดังนี้

11.3.3.1 นำผู้ตัดสินทั้งสองกลุ่มย่อยเป็นกลุ่มใหญ่เพียงกลุ่มเดียวกัน ผู้อำนวยความสะดวกแสดงค่าเฉลี่ยคะแนนจุดตัดของแต่ละกลุ่มให้ผู้ตัดสินพิจารณา เปิดโอกาสให้ผู้ตัดสินอภิปรายร่วมกัน

11.3.3.2 เปิดโอกาสให้ผู้ตัดสินทำการตัดสินคะแนนจุดตัดทั้ง 4 ระดับครั้งอย่างอิสระจากกัน

11.3.3.3 นำผลการตัดสินของผู้ตัดสินแต่ละคนมาหาค่าเฉลี่ยในแต่ละระดับ

11.3.3.4 เมื่อได้คะแนนจุดตัด 4 ระดับ ให้ผู้ตัดสินร่วมกันเขียนบรรยายระดับการปฏิบัติการของนักเรียนในประเด็นความสามารถของนักเรียน ณ ระดับคะแนนจุดตัดของระดับนั้น รวม 5 ระดับ

12. นำแบบทดสอบนำมาวิเคราะห์หาคุณภาพแบบทดสอบเพื่อหาค่าความยากและค่าอำนาจจำแนกของแบบทดสอบรายข้อ และค่าความเที่ยงอิงเกณฑ์วิธีของ Livingston และปรับปรุงแบบทดสอบเพื่อนำไปใช้จริง

13. ผู้วิจัยนำแบบทดสอบที่คัดเลือกและปรับปรุงแก้ไขแล้วจากการสอบครั้งที่ 2 นำไปใช้จริงกับกลุ่มตัวอย่าง จำนวนนักเรียนทั้งสิ้น 658 คน

ตอนที่ 3 เก็บรวบรวมข้อมูล

ขั้นตอนการเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูลแบ่งออกเป็น 2 ส่วน คือ ส่วนที่ 1 ดำเนินการสอบกับนักเรียน และส่วนที่ 2 ดำเนินการเก็บข้อมูลกับกลุ่มผู้ตัดสิน มีขั้นตอนการดำเนินการเก็บรวบรวมข้อมูลในแต่ละระยะ ดังนี้

การเก็บรวบรวมข้อมูลส่วนที่ 1 (ดำเนินการเก็บข้อมูลกับนักเรียน)

1. ติดต่อขอหนังสือจากมหาวิทยาลัยบูรพา เพื่อขออนุเคราะห์ในการเก็บรวบรวมข้อมูลต่อ ผู้อำนวยการสำนักงานศึกษาธิการจังหวัดชลบุรี

2. ติดต่อขออนุญาตผู้บริหารโรงเรียนที่ทดลองใช้ครั้งที่ 1 เป็นนักเรียนกลุ่มตัวอย่าง จำนวน 50 คน

3. ติดต่อประสานงานกับคุณครูประจำชั้น และคุณครูผู้สอนนักเรียนที่ใช้เป็นกลุ่มตัวอย่าง เพื่อนัดหมายและกำหนดเวลาในการสอบ

4. เตรียมเครื่องมือให้เพียงพอกับจำนวนนักเรียนที่เข้าสอบในแต่ละครั้ง และวางแผนดำเนินการสอบ

5. ผู้คุมสอบอธิบายให้นักเรียนกลุ่มตัวอย่างเข้าใจวัตถุประสงค์ของแบบทดสอบและวิธีการตอบก่อนเริ่มต้นลงมือทำ โดยกำหนดให้เว้นระยะเวลาในการทำแบบทดสอบที่มีวิธีการตอบและตรวจให้คะแนนแต่ละวิธีอย่างน้อย 7 วัน

6. นำแบบทดสอบมาตรวจให้คะแนนตามที่ผู้วิจัยทำการตรวจให้คะแนนในแต่ละวิธีที่กำหนดไว้ด้วยตนเอง

7. นำแบบทดสอบไปทดสอบกับกลุ่มตัวอย่างครั้งที่ 2 ซึ่งเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 4 จำนวน 210 คน โดยกำหนดให้เว้นระยะเวลาในการทำแบบทดสอบที่มีวิธีการตอบและตรวจให้คะแนนแต่ละวิธีอย่างน้อย 7 วัน

8. นำแบบทดสอบมาตรวจให้คะแนนตามที่ผู้วิจัยทำการตรวจให้คะแนนในแต่ละวิธีที่กำหนดไว้ด้วยตนเอง

9. นำแบบทดสอบไปทดสอบกับกลุ่มตัวอย่าง ซึ่งเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 4 ภาคเรียนที่ 2 ปีการศึกษา 2560 จำนวน 658 คน โดยกำหนดให้เว้นระยะเวลาในการทำแบบทดสอบที่มีวิธีการตอบและตรวจให้คะแนนแต่ละวิธีอย่างน้อย 7 วัน

การเก็บข้อมูลส่วนที่ 2 (ดำเนินการเก็บข้อมูลกับผู้ตัดสิน)

1. ผู้วิจัยจัดทำคู่มือการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟที่ได้รับการปรับปรุง เพื่อให้ผู้ตัดสินที่ประกอบด้วยครูที่สอนรายวิชาเคมี จำนวน 5 คน

2. ผู้วิจัยส่งหนังสือขอเชิญครูผู้สอน เพื่อเป็นผู้ตัดสินในการกำหนดคะแนนจุดตัดในแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชา เคมี สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 4 เพื่อให้ศึกษาและนัด วัน เวลา สถานที่ เพื่อเข้าร่วมประชุมกลุ่ม กำหนดมาตรฐานร่วมกัน

3. ผู้วิจัยจัดเตรียม แบบฟอร์มสำหรับกำหนดคะแนนจุดตัด เพื่อใช้เป็นข้อมูลสารสนเทศสำหรับผู้ตัดสินใช้ในการกำหนดคะแนนจุดตัด

ตอนที่ 4 การวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูล

ขั้นตอนที่ 1 การวิเคราะห์ข้อมูลเบื้องต้น เพื่อหาค่าสถิติพื้นฐานของคะแนนจากการตอบแบบสอบที่มีวิธีการตอบและตรวจให้คะแนน ทั้ง 2 วิธี โดยคำนวณค่าเฉลี่ย ($\overline{X}$) ส่วนเบี่ยงเบนมาตรฐาน (SD) ความคลาดเคลื่อนมาตรฐานของแบบทดสอบ (SEM) ค่าสูงสุด และค่าต่ำสุด

ขั้นตอนที่ 2 การตรวจสอบความเป็นมิติเดียวของข้อสอบ วิเคราะห์ตัวประกอบสำคัญ (Principal componant) สำหรับการตรวจสอบพิจารณาค่าไอเกน (Eigen value) ถ้าหากค่าไอเกนของตัวประกอบที่ 1 ของแบบทดสอบมีค่าสูงกว่า ค่าไอเกนของตัวประกอบที่ 2 อย่างมาก และค่าไอเกนของตัวประกอบที่ 2 มีค่าสูงกว่าค่าไอเกนของตัวประกอบ ถัดไปเพียงเล็กน้อย จึงจะกล่าวได้ว่าแบบทดสอบทั้งฉบับนี้ มีคุณสมบัติเพียงมิติเดียว (Lord, 1980, p. 21) โดยนำแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจหาคุณภาพของข้อสอบแต่ละข้อ โดยมีขั้นตอนดังนี้วิเคราะห์องค์ประกอบ เนื่องจากตามข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ นำคะแนนผลการสอบของนักเรียนในแต่ละฉบับ มาวิเคราะห์ ด้วยโปรแกรม SPSS for Windows โดยวิเคราะห์ตัวประกอบสำคัญ (Principal componant) หมุนแกนด้วยวิธี Direct oblimin oblique สำหรับการตรวจสอบพิจารณาค่าไอเกน (Eigen value) ถ้าหากค่าไอเกนของตัวประกอบที่ 1 ของแบบทดสอบมีค่าสูงกว่า ค่าไอเกนของตัวประกอบที่ 2 อย่างมาก และค่าไอเกนของตัวประกอบที่ 2 มีค่าสูงกว่าค่าไอเกนของตัวประกอบ ถัดไปเพียงเล็กน้อย จึงจะกล่าวได้ว่า แบบทดสอบแต่ละฉบับ มีคุณสมบัติเพียงมิติเดียว (Lord, 1980, p. 21) ผลการวิเคราะห์ พอจะเชื่อได้ว่า แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจมีความเป็นมิติเดียว

ขั้นตอนที่ 3 วิเคราะห์ค่าความตรงตามตามสภาพ โดยวิเคราะห์จากคะแนนที่ได้จากวิธีการตอบและตรวจให้คะแนนทั้ง 2 วิธี ดังนี้

1. วิเคราะห์ค่าความตรงตามสภาพ จากค่าสัมประสิทธิ์สหสัมพันธ์ (r) ระหว่างคะแนนจากแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาเคมี ภาคเรียนที่ 2 ชั้นมัธยมศึกษาปีที่ 4 ที่ได้จากวิธี

การตอบ และการให้คะแนนในแต่ละวิธี กับระดับผลการเรียนวิชาเคมี ภาคเรียนที่ 1 ชั้นมัธยมศึกษาปีที่ 4 แล้วแปลงค่าความตรงตามสภาพจากค่าสัมประสิทธิ์สหสัมพันธ์ เป็นคะแนนฟิชเชอร์ซี (Z_r) โดยใช้สูตร Glass and Stanley (1970, p. 265) และเปรียบเทียบความแตกต่างของค่าความตรงตามสภาพ โดยใช้สถิติทดสอบ Z (Glass & Stanley, 1970 อ้างถึงใน เอมอร จังศิริพรปกรณ์, 2548)

ขั้นตอนที่ 4 การตรวจสอบคุณภาพของแบบทดสอบ

การตรวจสอบคุณภาพของแบบทดสอบ จากฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบสอบและค่าประสิทธิภาพสัมพัทธ์โดยใช้การวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (Item response theory) แบบทวิภาค ที่ใช้โมเดลโลจิสติก (Logistic) และแบบพหุวิภาค ที่ใช้โมเดลการให้คะแนนความรู้บางส่วน (Genralized partial credit model: GPCM)

1. การวิเคราะห์แบบทวิวิภาค ที่ใช้โมเดลโลจิสติก (Logistic) ซึ่งเป็นการให้คะแนนแบบทดสอบเลือกตอบแบบธรรมดา ประมาณค่าพารามิเตอร์ โดยการนำคะแนนผลการตอบแบบทดสอบแต่ละฉบับมาวิเคราะห์รายข้อตามรูปแบบโลจิสติก 2 พารามิเตอร์ โดยวิธีความเป็นไปได้สูงสุด (Maximum likelihood) หาค่าความยาก (b) ค่าอำนาจจำแนก (a) ค่าฟังก์ชันสารสนเทศของข้อสอบ (IIF) และค่าฟังก์ชันสารสนเทศของแบบทดสอบ (TIF) โดยใช้โปรแกรมสำเร็จรูป IRTPRO student

2. การวิเคราะห์แบบพหุวิภาค ที่ใช้โมเดลการให้คะแนนความรู้บางส่วน (Genralized partial credit model: GPCM) ซึ่งเป็นการให้คะแนนแบบทดสอบเลือกตอบแบบให้คะแนนความมั่นใจ ประมาณค่าพารามิเตอร์ความยาก (b) ค่าพารามิเตอร์อำนาจจำแนก (a) ค่าฟังก์ชันสารสนเทศของข้อสอบ (IIF) และค่าฟังก์ชันสารสนเทศของแบบทดสอบ (TIF) โดยใช้โปรแกรมสำเร็จรูป IRTPRO student

ตอนที่ 5 สถิติที่ใช้ในการวิเคราะห์ข้อมูล

1. ความเที่ยงตรงด้านเนื้อหา (Content validity) เพื่อให้ผู้เชี่ยวชาญประเมินว่าข้อสอบแต่ละข้อนั้น สามารถวัดได้ตรงกับเนื้อหาหรือไม่ และกำหนดค่าน้ำหนักคะแนนของตัวเลือกตามวิธีการตรวจให้คะแนนทั้งสองวิธี การหาค่าดัชนีความสอดคล้องเป็นรายข้อ (Index of congruency: IOC) โดยใช้สูตร (บุญเชิด ภิญโญอนันตพงษ์, 2545, หน้า 179)

$$IOC = \frac{\sum R}{N}$$

IOC แทน ดัชนีความสอดคล้อง

$\sum R$ แทน ผลรวมของคะแนนความคิดเห็นของผู้เชี่ยวชาญแต่ละคน

N แทน จำนวนผู้เชี่ยวชาญ

คะแนนค่าดัชนีความสอดคล้องระหว่างเนื้อหากับข้อสอบรายข้อของผู้เชี่ยวชาญ ดังนี้

+ 1 หมายถึง แน่ใจว่าสอดคล้องกับเนื้อหาและวิธีการให้คะแนน

0 หมายถึง ไม่แน่ใจสอดคล้องกับเนื้อหาและวิธีการให้คะแนน

-1 หมายถึง แน่ใจว่าไม่สอดคล้องกับเนื้อหาและวิธีการให้คะแนน

2. ค่าความยากโดยพิจารณาค่าความยากจากค่าเฉลี่ยของคะแนนของข้อนั้น ๆ ตามวิธีการให้คะแนนทั้งสองวิธี จากค่าความยากของแบบทดสอบ (Difficulty) (ล้วน สายยศ และอังคณา สายยศ, 2543, หน้า 196)

$$p = \frac{R}{N}$$

เมื่อ p แทน ดัชนีค่าความยาก

R แทน จำนวนนักเรียนที่ทำข้อสอบถูก

N แทน จำนวนนักเรียนที่ทำข้อสอบทั้งหมด

ตารางที่ 3-5 เกณฑ์การพิจารณาค่าความยาก (สมนึก ภัททิยธนี, 2553, หน้า 212)

ดัชนีค่า p	**ความหมาย**
มากกว่า 0.80	ง่ายมาก (ปรับปรุงหรือตัดทิ้ง)
0.60-0.80	ค่อนข้างง่าย
0.40-0.59	ปานกลาง
0.20-0.39	ค่อนข้างยาก
ต่ำกว่า 0.20	ยากมาก (ปรับปรุงหรือตัดทิ้ง)

3. ค่าอำนาจจำแนกของข้อสอบ โดยใช้ดัชนีบี (B-index) ตามวิธีของเบรนแนน (Brennan) (ล้วน สายยศ และอังคณา สายยศ, 2539, หน้า 198) จากสูตร

$$B = \frac{U}{N_1} - \frac{L}{N_2}$$

เมื่อ B แทน ค่าอำนาจจำแนกของข้อสอบ

U แทน จำนวนคนทำข้อสอบข้อนั้นถูกของกลุ่มที่ผ่านเกณฑ์

L แทน จำนวนคนทำข้อสอบข้อนั้นถูกของกลุ่มที่ไม่ผ่านเกณฑ์

N_1 แทน จำนวนคนที่สอบผ่านเกณฑ์

N_2 แทน จำนวนคนที่สอบไม่ผ่านเกณฑ์

ตารางที่ 3-6 เกณฑ์การพิจารณาค่าอำนาจจำแนก (Brennan et al., 1974, pp. 244-261)

ดัชนีค่า B	**ความหมาย**
+1.00	จำแนกผู้รอบรู้/ ไม่รอบรู้/ ได้ถูกต้องทุกคน
0.50-0.99	จำแนกผู้รอบรู้/ ไม่รอบรู้/ ได้ถูกต้องเป็นส่วนใหญ่
0.20-0.49	จำแนกผู้รอบรู้/ ไม่รอบรู้/ ได้ถูกต้องบางส่วน
0.00-0.19	จำแนกผู้รอบรู้/ ไม่รอบรู้/ ได้ถูกต้องน้อยมาก
ติดลบ	จำแนกผู้รอบรู้/ ไม่รอบรู้/ ตรงข้ามกับความตรง

4. ค่าความเที่ยง (Reliability) โดยการวิเคราะห์หาค่าความเที่ยงแบบอิงเกณฑ์ ที่เลือกใช้วิธีของ Livingston จากสูตร

$$r_{cc} = \frac{r_{tt}\delta_0^2 + (\overline{X} - c)^2}{\delta_0^2 + (\overline{X} - c)^2}$$

r_{cc} หมายถึง ความเชื่อมั่นของ Criterion referenced test

r_{tt} หมายถึง ความเชื่อมั่นของแบบทดสอบ

δ_0^2 หมายถึง ความแปรปรวนของคะแนนการสอบ

$\overline{X}$ หมายถึง ค่าเฉลี่ยของคะแนน

C หมายถึง คะแนนเกณฑ์ (Criterion score)

5. ค่าฟังก์ชันสารสนเทศของข้อสอบ (De Ayala et al., 1992) จากสูตร

$$\text{IIF หรือ } I_i(\theta) = \sum_{x=0}^{m}\left[\frac{P'_{ix}(\theta)^2}{P_{ix}(\theta)}\right]$$

เมื่อ $I_i(\theta)$ แทน สารสนเทศของข้อสอบข้อที่ i ที่ระดับความสามารถ θ

$P'_{ix}(\theta)$ แทน ความน่าจะเป็นหรือโอกาสสำหรับคนที่มีความสามารถ θ จะตอบข้อ i ได้คะแนน 1 มากกาว่าคะแนน x

$P_{ix}(\theta)$ แทน ความน่าจะเป็นหรือโอกาสของคนที่ทีความสามารถ θ จะตอบข้อสอบ i ได้คะแนน x

6. ค่าฟังก์ชันสารสนเทศของแบบทดสอบจาก IIF สามารถนำมาคำนวณ TIF

$$I(\theta) = \sum_{i=1}^{k} I_i(\theta)$$

เมื่อ $I_i(\theta)$ แทน ฟังก์ชันสารสนเทศของข้อสอบ ตั้งแต่ข้อ 1 ถึง ข้อ i

7. ค่าความคลาดเคลื่อนมาตรฐานของแบบทดสอบจาก TIF สามารถนำมาคำนวณ SE (θ)

$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}}$$

เมื่อ SE (θ) แทน ความคลาดเคลื่อนมาตรฐานของการประมาณค่าสำหรับผู้ตอบที่มีความสามารถ θ

$I(\theta)$ แทน สารสนเทศของแบบทดสอบสำหรับผู้ตอบที่มีความสามารถ θ

8. ค่าประสิทธิภาพสัมพัทธ์ (Relative efficiency: RE (θ)) จากสูตร

$$RE(\theta) = \frac{I_A(\theta)}{I_B(\theta)}$$

เมื่อ $I_A(\theta)$ แทน ค่าฟังก์ชันสารสนเทศของแบบทดสอบฉบับ A ณ ตำแหน่งบนสเกลความสามารถร่วมกันระดับ θ

$I_B(\theta)$ แทน ค่าฟังก์ชันสารสนเทศของแบบทดสอบฉบับ B ณ ตำแหน่งบนสเกลความสามารถร่วมกันระดับ θ

การแปลความหมายของค่าดัชนีประสิทธิภาพสัมพัทธ์ สามารถนำไปช่วยคัดเลือกแบบทดสอบได้ดังนี้

RE (θ) = 1 แสดงว่าแบบทดสอบทั้งสองฉบับมีประสิทธิภาพเท่ากันสำหรับผู้ตอบที่มีระดับความสามารถ θ

RE (θ) > 1 แสดงว่าแบบทดสอบฉบับ A มีประสิทธิภาพสูงกว่าแบบทดสอบฉบับ B สำหรับ ผู้ตอบที่มีระดับความสามารถ θ

RE (θ) < 1 แสดงว่าแบบทดสอบฉบับ A มีประสิทธิภาพต่ำกว่าแบบทดสอบฉบับ B สำหรับผู้ตอบที่มีระดับความสามารถ θ

9. ความตรงตามสภาพจากค่าสัมประสิทธิ์สหสัมพันธ์ เป็นคะแนนฟิชเชอร์ซี (Z_r) โดยใช้สูตร (Glass & Stanley, 1970, p. 265)

$$Z_r = \log_e \sqrt{(1+r)/(1-r)}$$

เมื่อ Z_r แทน คะแนนฟิชเชอร์ซี

r แทน ค่าสัมประสิทธิ์สหสัมพันธ์

10. สถิติทดสอบ Z มีสูตร คือ (Glass & Stanley, 1970 อ้างถึงใน เอมอร จังศิริพรปกรณ์, 2548)

$$Z = \frac{Z_{r1} - Z_{r2}}{\sqrt{\frac{1}{(n_1 - 3)}} + \sqrt{\frac{1}{(n_2 - 3)}}}$$

เมื่อ Z แทน สถิติทดสอบซี (Z-test)

Z_{r1}, Z_{r2} แทน คะแนนฟิชเชอร์ซี

n_1, n_2 แทน จำนวนข้อมูลของกลุ่มตัวอย่าง

บทที่ 4
ผลการวิเคราะห์ข้อมูล

การวิจัยครั้งนี้ มีวัตถุประสงค์ เพื่อหาคุณภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและแบบพิจารณาความมั่นใจ ในด้านฟังก์ชันสารสนเทศ ความยากและอำนาจจำแนก ความตรงตามสภาพและเปรียบเทียบคุณภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและแบบพิจารณาความมั่นใจในด้านดังต่อไปนี้ ความตรงตามสภาพ การนำเสนอผลการวิเคราะห์ข้อมูลแยกนำเสนอออกเป็น 3 ตอน ดังนี้

ตอนที่ 1 ผลการวิเคราะห์คุณภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ ในการทดลองใช้ครั้งที่ 1 และครั้งที่ 2 ด้วยทฤษฎีการวัดแบบดั้งเดิม (CTT)

ตอนที่ 2 ผลการวิเคราะห์คุณภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ ในการนำไปใช้จริง ด้วยทฤษฎีการตอบสนองข้อสอบ (IRT) ในด้านฟังก์ชันสารสนเทศ ความยากและอำนาจจำแนก ความตรงตามสภาพ

ตอนที่ 3 ผลเปรียบเทียบฟังก์ชันสารสนเทศและความตรงตามสภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ

สัญลักษณ์และอักษรย่อที่ในการนำเสนอผลการวิจัย

ผู้วิจัยจึงกำหนดสัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ เพื่อให้เข้าใจในการนำเสนอและเพื่อให้เกิดความเข้าใจตรงกันในการแปลความหมายของผลการวิเคราะห์ข้อมูล ดังนี้

p	แทน ค่าความยากของแบบทดสอบ (ทฤษฎีการวัดแบบดั้งเดิม CTT)
r	แทน ค่าอำนาจำแนกของแบบทดสอบ (ทฤษฎีการวัดแบบดั้งเดิม CTT)
r_{KR-20}	แทน ค่าความเที่ยงของความสอดคล้องภายใน
r_{cc}	แทน ค่าความเที่ยงแบบอิงเกณฑ์ของ Livingston
a	แทน ค่าอำนาจำแนกของแบบทดสอบ (ทฤษฎีการตอบสนองข้อสอบ IRT)
b	แทน ค่าความยากของแบบทดสอบ (ทฤษฎีการตอบสนองข้อสอบ IRT)
$\overline{X}$	แทน ค่าเฉลี่ย
Min	แทน คะแนนต่ำสุด
Max	แทน คะแนนสูงสุด

SD แทน ส่วนเบี่ยงเบนมาตรฐาน

SEM แทน ความคลาดเคลื่อนมาตรฐานของแบบทดสอบ

N แทน จำนวนนักเรียน

θ แทน ระดับความสามารถ

$I(\theta)$ แทน ฟังก์ชันสารสนเทศ

$SE(\theta)$ แทน ความคลาดเคลื่อนมาตรฐานของการประมาณค่าสำหรับผู้ตอบที่มีระดับความสามารถ θ

$RE(\theta)$ แทน ค่าประสิทธิภาพสัมพัทธ์สำหรับผู้ตอบที่มีระดับความสามารถ θ

Df แทน ชั้นแห่งความอิสระ (Degree of freedom)

TIF แทน ค่าฟังก์ชันสารสนเทศของแบบทดสอบ

Z_r แทน คะแนนฟิชเชอร์ซี

Z แทน สถิติซี

ตอนที่ 1 ผลการวิเคราะห์คุณภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ ในการทดลองใช้ครั้งที่ 1 และครั้งที่ 2 ด้วยทฤษฎีการวัดแบบดั้งเดิม (CTT)

ผู้วิจัยได้นำเสนอผลการวิเคราะห์ข้อมูลเบื้องต้น ได้แก่ ผลการตรวจสอบคุณภาพแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและแบบพิจารณาความมั่นใจรายข้อ และทั้งฉบับ

ผลการตรวจสอบคุณภาพแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและแบบพิจารณาความมั่นใจจำแนกเป็นรายข้อในการทดลองใช้ครั้งที่ 1

ผู้วิจัยได้นำเครื่องมือที่สร้างขึ้นให้ผู้เชี่ยวชาญ จำนวน 5 คน ตรวจสอบคุณภาพเบื้องต้นด้านความตรงเชิงเนื้อหา (Content validity) ของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและแบบพิจารณาความมั่นใจ โดยการนำแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาเคมี จำนวน 40 ข้อ ที่ผู้วิจัยสร้างขึ้นเพื่อให้ผู้เชี่ยวชาญประเมินว่าข้อสอบแต่ละข้อนั้น สามารถวัดได้ตรงกับเนื้อหาหรือไม่ และกำหนดค่าน้ำหนักคะแนนของตัวเลือกตามวิธีการตรวจให้คะแนนทั้ง 2 วิธี การหาค่าดัชนีความสอดคล้องเป็นรายข้อ (Index of congruency: IOC) ดังตารางที่ 4-1

ตารางที่ 4-1 ความสอดคล้องในแต่ละข้อของแบบทดสอบที่มีตัวเลือกตามวิธีการตรวจให้คะแนนทั้งสองวิธี

ข้อที่	ค่าความสอดคล้อง					ค่า IOC	แปลความหมาย
	คนที่ 1	คนที่ 2	คนที่ 3	คนที่ 4	คนที่ 5		
1.	1.0	1.0	1.0	0.0	1.0	0.8	สอดคล้อง
2.	1.0	0.0	1.0	1.0	0.0	0.6	สอดคล้อง
3.	1.0	1.0	1.0	0.0	0.0	0.6	สอดคล้อง
4.	0.0	0.0	0.0	1.0	0.0	0.2	ไม่สอดคล้อง
5.	0.0	1.0	1.0	0.0	1.0	0.6	สอดคล้อง
6.	1.0	1.0	1.0	0.0	1.0	0.8	สอดคล้อง
7.	-1.0	1.0	1.0	-1.0	0.0	0.0	ไม่สอดคล้อง
8.	1.0	0.0	0.0	0.0	1.0	0.4	ไม่สอดคล้อง
9.	1.0	0.0	1.0	1.0	1.0	0.8	สอดคล้อง
10.	1.0	1.0	1.0	1.0	1.0	1.0	สอดคล้อง
11.	1.0	1.0	0.0	1.0	1.0	0.8	สอดคล้อง
12.	1.0	1.0	-1.0	0.0	1.0	0.4	ไม่สอดคล้อง
13.	0.0	1.0	1.0	1.0	1.0	0.8	สอดคล้อง
14.	1.0	1.0	1.0	1.0	0.0	0.8	สอดคล้อง
15.	1.0	1.0	1.0	1.0	1.0	1.0	สอดคล้อง
16.	0.0	0.0	1.0	1.0	0.0	0.4	ไม่สอดคล้อง
17.	-1.0	1.0	1.0	0.0	1.0	0.4	ไม่สอดคล้อง
18.	1.0	1.0	0.0	1.0	0.0	0.6	สอดคล้อง
19.	1.0	1.0	1.0	-1.0	1.0	0.6	สอดคล้อง
20.	1.0	1.0	1.0	1.0	1.0	1.0	สอดคล้อง
21.	1.0	1.0	1.0	0.0	1.0	0.8	สอดคล้อง
22.	1.0	1.0	1.0	1.0	1.0	1.0	สอดคล้อง
23.	1.0	0.0	1.0	1.0	1.0	0.8	สอดคล้อง
24.	1.0	1.0	1.0	1.0	1.0	1.0	สอดคล้อง
25.	1.0	1.0	1.0	1.0	1.0	1.0	สอดคล้อง

ตารางที่ 4-1 (ต่อ)

ข้อที่	ค่าความสอดคล้อง					ค่า IOC	แปลความหมาย
	คนที่ 1	คนที่ 2	คนที่ 3	คนที่ 4	คนที่ 5		
26.	1.0	1.0	1.0	1.0	0.0	0.8	สอดคล้อง
27.	-1.0	1.0	1.0	0.0	0.0	0.2	ไม่สอดคล้อง
28.	1.0	1.0	1.0	1.0	1.0	1.0	สอดคล้อง
29.	0.0	1.0	1.0	1.0	1.0	0.8	สอดคล้อง
30.	0.0	0.0	0.0	1.0	0.0	0.2	ไม่สอดคล้อง
31.	1.0	0.0	1.0	1.0	1.0	0.8	สอดคล้อง
32.	1.0	1.0	1.0	1.0	1.0	1.0	สอดคล้อง
33.	1.0	1.0	1.0	1.0	1.0	1.0	สอดคล้อง
34.	1.0	1.0	1.0	1.0	1.0	1.0	สอดคล้อง
35.	0.0	1.0	1.0	1.0	1.0	0.8	สอดคล้อง
36.	1.0	1.0	0.0	1.0	1.0	0.8	สอดคล้อง
37.	1.0	-1.0	1.0	1.0	1.0	0.6	สอดคล้อง
38.	0.0	1.0	-1.0	0.0	1.0	0.2	ไม่สอดคล้อง
39.	1.0	-1.0	1.0	1.0	0.0	0.4	ไม่สอดคล้อง
40.	1.0	1.0	1.0	1.0	0.0	0.8	สอดคล้อง

จากตารางที่ 4-1 พบว่า ค่าดัชนีวามสอดคล้องระหว่างแบบทดสอบกับจุดประสงค์การเรียนรู้ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 4 มีค่าเฉลี่ยของคะแนนการพิจารณาจากผู้เชี่ยวชาญ 5 คนมีค่าอยู่ระหว่าง 0.3-1.0 ซึ่งผู้วิจัยคัดเลือกข้อคำถามที่มีค่าตั้งแต่ 0.6 ขึ้นไป และตัดข้อคำถามที่ค่าดัชนีวามสอดคล้อง ต่ำกว่า 0.6 จำนวน 10 ข้อ ได้แก่ข้อที่ 4, 7, 8, 12, 16, 17, 27, 30, 38 และ 39 แสดงว่าแบบทดสอบที่สร้างขึ้นจำนวน มีความตรงเชิงเนื้อหา 30 ข้อ

จากนั้น ผู้วิจัยนำแบบทดสอบมาทดลองใช้ครั้งที่ 1 กับนักเรียนกลุ่มตัวอย่างจำนวน 50 คน และทดลองใช้ครั้งที่ 2 กับกลุ่มตัวอย่างจำนวน 210 คน ซึ่งเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 4 โรงเรียนในสังกัดเขตพื้นที่การศึกษามัธยมศึกษา เขต 18 เพื่อวิเคราะห์หาคุณภาพแบบทดสอบ โดยพิจารณาค่าความยากและค่าอำนาจจำแนกของแบบทดสอบ โดยใช้ดัชนีบี (B-index) ตามวิธีของเบรนแนน (Brennan) ดังตารางที่ 4-2

ตารางที่ 4-2 ค่าความยาก (p) ค่าอำนาจจำแนก (r) ทดลองใช้ครั้งที่ 1 และครั้งที่ 2 ของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ

ข้อ	ครั้งที่ 1					ครั้งที่ 2				
	แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา		แบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ		การพิจารณา	แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา		แบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ		การพิจารณา
	p	r	p	r		p	r	p	r	
1.	0.50	-0.36**	0.40	0.11**	ปรับปรุง	0.53	0.27	0.54	0.22	ผ่านเกณฑ์
2.	0.78	0.41	0.72	0.17**	ปรับปรุง	0.70	0.25	0.76	0.38	ผ่านเกณฑ์
3.	0.15*	0.13**	0.20	0.26	ปรับปรุง	0.37	0.33	0.27	0.22	ผ่านเกณฑ์
4.	0.66	0.55	0.40	0.25	คัดไว้	0.73	0.45	0.52	0.37	ผ่านเกณฑ์
5.	0.56	0.73	0.37	0.41	คัดไว้	0.57	0.64	0.49	0.65	ผ่านเกณฑ์
6.	0.69	0.50	0.36	0.43	คัดไว้	0.57	0.68	0.48	0.71	ผ่านเกณฑ์
7.	0.59	0.71	0.48	0.45	คัดไว้	0.63	0.45	0.53	0.57	ผ่านเกณฑ์
8.	0.27	0.21	0.12*	0.02**	ปรับปรุง	0.57	0.33	0.25	0.24	ผ่านเกณฑ์
9.	0.66	0.77	0.46	0.64	คัดไว้	0.60	0.51	0.65	0.54	ผ่านเกณฑ์

ตารางที่ 4-2 (ต่อ)

ข้อ	ครั้งที่ 1					ครั้งที่ 2				
	แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา		แบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ		การพิจารณา	แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา		แบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ		การพิจารณา
	p	r	p	r		p	r	p	r	
10.	0.51	0.73	0.29	0.30	คัดไว้	0.63	0.42	0.39	0.56	ผ่านเกณฑ์
11.	0.58	0.77	0.34	0.45	คัดไว้	0.60	0.73	0.47	0.66	ผ่านเกณฑ์
12.	0.64	0.71	0.42	0.68	คัดไว้	0.60	0.70	0.55	0.73	ผ่านเกณฑ์
13.	0.35	0.02**	0.14*	-.04**	ปรับปรุง	0.67	0.27	0.45	0.23	ผ่านเกณฑ์
14.	0.49	0.64	0.30	0.25	คัดไว้	0.60	0.38	0.41	0.48	ผ่านเกณฑ์
15.	0.72	0.63	0.48	0.45	คัดไว้	0.67	0.50	0.62	0.55	ผ่านเกณฑ์
16.	0.63	0.68	0.50	0.34	คัดไว้	0.70	0.45	0.57	0.51	ผ่านเกณฑ์
17.	0.54	0.70	0.39	0.52	คัดไว้	0.60	0.59	0.44	0.51	ผ่านเกณฑ์
18.	0.09	0.04	0.35	0.10**	ปรับปรุง	0.20	0.27	0.37	0.27	ผ่านเกณฑ์
19.	0.36	0.66	0.24	0.21	คัดไว้	0.43	0.37	0.25	0.55	ผ่านเกณฑ์

ตารางที่ 4-2 (ต่อ)

ข้อ	ครั้งที่ 1					ครั้งที่ 2				
	แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา		แบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ		การพิจารณา	แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา		แบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ		การพิจารณา
	p	r	p	r		p	r	p	r	
20.	0.65	0.57	0.38	0.3	คัดไว้	0.70	0.56	0.52	0.59	ผ่านเกณฑ์
21.	0.47	0.30	0.24	0.21	คัดไว้	0.60	0.49	0.39	0.37	ผ่านเกณฑ์
22.	0.37	0.30	0.30	0.38	คัดไว้	0.47	0.28	0.31	0.40	ผ่านเกณฑ์
23.	0.19*	-0.05**	0.08*	0.01**	ปรับปรุง	0.40	0.27	0.37	0.27	ผ่านเกณฑ์
24.	0.50	0.55	0.31	0.32	คัดไว้	0.57	0.49	0.35	0.66	ผ่านเกณฑ์
25.	0.69	0.66	0.39	0.52	คัดไว้	0.57	0.64	0.49	0.69	ผ่านเกณฑ์
26.	0.56	0.75	0.37	0.50	คัดไว้	0.53	0.61	0.42	0.77	ผ่านเกณฑ์
27.	0.20	0.02**	0.30	0.17**	ปรับปรุง	0.33	0.40	0.36	0.41	ผ่านเกณฑ์
28.	0.45	0.71	0.30	0.11**	ปรับปรุง	0.50	0.27	0.43	0.35	ผ่านเกณฑ์
29.	0.50	0.70	0.37	0.27	คัดไว้	0.43	0.42	0.42	0.49	ผ่านเกณฑ์
30.	0.50	0.70	0.41	0.19**	ปรับปรุง	0.50	0.33	0.49	0.41	ผ่านเกณฑ์

จากตารางที่ 4-2 พบว่า ในทดลองใช้ครั้งที่ 1 ผู้วิจัยได้คัดเลือกแบบทดสอบที่มีค่าความยากตั้งแต่ 0.20 ถึง 0.78 และคัดเลือกแบบทดสอบที่มีค่าอำนาจจำแนก (Brennan) ตั้งแต่ 0.21 ถึง 0.77 ข้อที่มีความยากต่ำกว่า 0.20 และค่าอำนาจจำแนก (Brennan) ต่ำกว่า 0.20 นำมาปรับปรุงข้อสอบจำนวน 10 ข้อ ได้แก่ข้อ1, 2, 3, 8, 13, 18, 23, 27, 28 และ 30 เพื่อนำไปทดลองใช้ครั้งที่ 2 แล้วหาคุณภาพแบบทดสอบเพื่อให้ได้แบบทดสอบไปใช้จริง จากการหาคุณภาพแบบทดสอบโดยวิเคราะห์หาค่าความยากและค่าอำนาจจำแนก (Brennan) จากการนำไปทดลองใช้ครั้งที่ 2 พบว่าแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ มีค่าความยากอยู่ระหว่าง 0.20 ถึง 0.73 และ 0.25 ถึง 0.76 ตามลำดับ ข้อสอบผ่านเกณฑ์ทุกข้อ และค่าอำนาจจำแนก (Brennan) มีค่าอยู่ระหว่าง 0.25 ถึง 0.73 และ 0.24 ถึง 0.38 ตามลำดับ ข้อสอบผ่านเกณฑ์ทุกข้อ ได้แบบทดสอบจำนวน 30 ข้อ เพื่อนำไปใช้จริง

ตารางที่ 4-3 ปรับปรุงแบบทดสอบหลังจากการทดลองใช้ ครั้งที่ 1

ข้อ	ผลการเรียนรู้	แบบทดสอบก่อนปรับปรุง	แบบทดสอบหลังปรับปรุง
1	ผลการเรียนรู้ที่ 1 สามารถอธิบายความหมายของมวลอะตอม มวลโมเลกุล คำนวณหามวลอะตอมของธาตุได้	ธาตุ Ne มีมวลอะตอมเท่ากับ 20 Ne 1 อะตอมจะมีมวลกี่กรัม ก. 1.66×10^{-24} ข. 1.66×10^{-23} ค. 3.32×10^{-24} ง. 3.32×10^{-23}	ธาตุ A มีมวลอะตอม 20 หมายความว่าอย่างไร ก. ธาตุ A 1 อะตอมหนัก 20เท่าของ 6.02×10^{23} กรัม ข. ธาตุ A 1 อะตอมหนัก 20เท่าของ ไฮโดรเจน 1 อะตอม ค. ธาตุ A 1 อะตอมหนัก 20 กรัม ง. ธาตุ A 1 อะตอมมีมวล $20 \times 1.66 \times 10^{-24}$ กรัม
2	ผลการเรียนรู้ที่ 2 สามารถอธิบายความหมายของมวลโมเลกุลของสารมวลของธาตุ 1 อะตอม มวลอะตอมเฉลี่ยของธาตุ และมวลของสาร 1 โมเลกุลได้	สารประกอบ X_2 มีมวลโมเลกุล 4 หนักเท่าใด ก. 3.32×10^{-24} g ข. 6.64×10^{-24}g ค. 76.36×10^{-24}g ง. 146.08×10^{-24} g	สารประกอบ X มีมวลโมเลกุล 44 ถ้ามี 2 โมเลกุลจะหนักเท่าใด ค. $2 \times 1.66 \times 10^{-24}$ g ง. $44 \times 1.66 \times 10^{-24}$g ค. $46 \times 1.66 \times 10^{-24}$g ง. $88 \times 1.66 \times 10^{-24}$ g

ตารางที่ 4-3 (ต่อ)

ข้อ	ผลการเรียนรู้	แบบทดสอบก่อนปรับปรุง	แบบทดสอบหลังปรับปรุง
3		มวลแคลเซียม 1 อะตอม คือข้อใด ก. มวลที่แท้จริง ข. มวลเปรียบเทียบ ค. มวลอะตอมของธาตุนั้น ง. $\frac{\text{มวลของธาตุ 1 อะตอม(g)}}{\frac{1}{12}\text{มวลของคาร์บอน}-12\ 1\text{อะตอม(g)}}$	ธาตุแคลเซียมมวลอะตอม 40 ข้อใดถูกต้อง (ความเข้าใจ) ก. ธาตุแคลเซียม 1 อะตอม หนักเป็น 40 เท่า ของ $\frac{1}{12}$ มวลของคาร์บอน 12, 12 อะตอม ข. ธาตุแคลเซียม 1 อะตอม หนักเป็น 40 เท่า ของ $\frac{1}{16}$ มวลของออกซิเจน, 1 อะตอม ค. ธาตุแคลเซียม 1 อะตอม หนัก $40 \times 1.66 \times 10^{-24}$ กิโลกรัม ง. ธาตุแคลเซียม 1 อะตอม หนัก 6.64×10^{-23} กรัม
8	ผลการเรียนรู้ที่ 4 สามารถคำนวณความสัมพันธ์ระหว่างโมล มวล จำนวนอนุภาคและปริมาตรของก๊าซได้	ก๊าซ A 3 อะตอม จะมีปริมาตรเท่าใดที่ STP ก. 7.47 dm^3 ข. 11.20 dm^3 ค. 14.94 dm^3 ง. 22.40 dm^3	ก๊าซ A มีสูตรโมเลกุลเป็น A_3 ถ้าก๊าซนี้มีอยู่ 6.02×10^{23} อะตอม จะมีปริมาตรเท่าใดที่ STP ก. 7.47 dm^3 ข. 11.20 dm^3 ค. 14.94 dm^3 ง. 22.40 dm^3

ตารางที่ 4-3 (ต่อ)

ข้อ	ผลการเรียนรู้	แบบทดสอบก่อนปรับปรุง	แบบทดสอบหลังปรับปรุง
13	ผลการเรียนรู้ที่ 5 สามารถอธิบายสมบัติบางประการของสารละลาย ทำการทดลองเตรียมสารละลายได้	สารในกลุ่มใดต่อไปนี้แตกต่างจากสารกลุ่มอื่น ก. เหล็ก ทองแดง สังกะสี ข. อากาศ ก๊าซหุงต้ม น้ำเกลือ ค. เหรียญบาท เหล็กกล้า ทอง ง. น้ำอัดลม แอลกอฮอล์ล้างแผล ทิงเจอร์ไอโอดีน	สารในกลุ่มใดต่อไปนี้แตกต่างจากสารกลุ่มอื่น ก. เหล็ก ทองแดง สังกะสี ข. อากาศ ก๊าซหุงต้ม น้ำเกลือ ค. เหรียญบาท เหล็กกล้า ทองสัมฤทธิ์ ง. น้ำอัดลม แอลกอฮอล์ล้างแผล ทิงเจอร์ไอโอดีน
18	ผลการเรียนรู้ที่ 6 สามารถคำนวณความเข้มข้นของสารละลายในหน่วยร้อยละ โมลาริตี โมแลล ส่วนในล้านส่วน ส่วนในพันล้านส่วนและเศษส่วนโมลได้	ละลาย NaOH จำนวน 20 กรัม ลงในน้ำได้สารละลาย 500 cm^3 แบ่งสารละลายจำนวนนี้มา 100 cm^3 แล้วทำให้มีปริมาตร 250 cm^3 และสารละลายจำนวนนี้เมื่อนำมาเคี่ยวจนเหลือปริมาตร 100 cm^3 จะมีความเข้มข้นเท่าใด (Na = 23, H = 1, O = 16) ก. 0.5 mol/ L ข. 0.1 mol/ L ค. 0.2 mol/ L ง. 1.0 mol/ L	ละลาย NaOH จำนวน 20 กรัม ลงในน้ำได้สารละลาย 500 cm^3 แบ่งสารละลายจำนวนนี้มา 100 cm^3 แล้วเติมน้ำให้มีปริมาตร 250 cm^3 และสารละลายจำนวนนี้เมื่อนำมาเคี่ยวจนเหลือปริมาตร 100 cm^3 จะมีความเข้มข้นเท่าใด (Na = 23, H = 1, O = 16) ก. 0.5 mol/ L ข. 0.1 mol/ L ค. 0.2 mol/ L ง. 1.0 mol/ L

ตารางที่ 4-3 (ต่อ)

ข้อ	ผลการเรียนรู้	แบบทดสอบก่อนปรับปรุง	แบบทดสอบหลังปรับปรุง
23	ผลการเรียนรู้ที่ 9 สามารถคำนวณหาปริมาตรของก๊าซในปฏิกิริยาเคมีได้	นำ H_2 4 dm^3 และ O_2 3 dm^3 มาทำปฏิกิริยากันที่อุณหภูมิ 100°C เมื่อปฏิกิริยาเกิดอย่างสมบูรณ์จะมีก๊าซในระบบกี่ dm^3 ก. 4 ข. 5 ค. 6 ง. 8	นำ H_2 4 dm^3 และ O_2 3 dm^3 มาทำปฏิกิริยากันที่อุณหภูมิ 100°C ดังสมการ $2H_2 (g) + O_2 (g) \longrightarrow 2H_2O (g)$ เมื่อปฏิกิริยาเกิดอย่างสมบูรณ์จะมีก๊าซในระบบกี่ dm^3 ก. 4 ข. 5 ค. 6 ง. 8
27	ผลการเรียนรู้ที่ 11 สามารถคำนวณเกี่ยวกับสูตร และสมการเคมีได้	โลหะ M มีมวลอะตอม 200 เมื่อนำโลหะ M มาเผากับกำมะถัน ปรากฏว่า M หนัก 16.00 กรัม ให้สารประกอบหนัก 19.84 กรัม สูตรเอมพิริคัลของสารประกอบนี้คือข้อใด ก. M_2S_3 ข. M_3S_2 ค. MS ง. M_2S	โลหะ M มีมวลอะตอม 200 เมื่อนำมาเผากับกำมะถันที่มีมากเกินพอปรากฏว่า M หนัก 16.00 กรัม ให้สารประกอบหนัก 19.84 กรัม สูตรเอมพิริคัลของสารประกอบนี้คือข้อใด ก. M_2S_3 ข. M_3S_2 ค. MS ง. M_2S

ตารางที่ 4-3 (ต่อ)

ข้อ	ผลการเรียนรู้	แบบทดสอบก่อนปรับปรุง	แบบทดสอบหลังปรับปรุง
28	ผลการเรียนรู้ที่ 12 สามารถใช้ความสัมพันธ์ระหว่างจำนวนโมล กับมวลหรือปริมาตรของสาร คำนวณหาค่าโมล มวลหรือปริมาตรของสาร สมการเคมี และหาค่าร้อยละของผลที่ได้จากการทดลองได้	ถ้านำเบนซีน (C_6H_6) 15.6 กรัม มาทำปฏิกิริยากับกรดไนตริก (HNO_3) จำนวนมากเกินพอ พบว่าเกิดไนโตรเบนซีน ($C_6H_5NO_2$) 18.0 กรัม จงหาผลได้ร้อยละ ก. 50.0 ข. 62.5 ค. 73.2 ง. 78.3	ถ้านำเบนซีน (C_6H_6) 15.6 กรัม มาทำปฏิกิริยากับกรดไนตริก (HNO_3) จำนวนมากเกินพอ พบว่าเกิดไนโตรเบนซีน ($C_6H_5NO_2$) 18.0 กรัม ดังสมการ $C_6H_6\ (l) + HNO_3\ (aq) \longrightarrow C_6H_5NO_2\ (l) + H_2O\ (l)$ จงหาผลได้ร้อยละ ก. 50.0 ข. 62.5 ค. 73.2 ง. 78.3
30		ในการเผา $KClO_3$ จะเกิดปฏิกิริยาดังนี้ $2KClO_3 \longrightarrow 2KCl + 3O_2$ ถ้าเผา $KClO_3$ จำนวน 12.26 กรัม จะได้ก๊าซ O_2 กี่ลิตร (K = 39, Cl = 35.5, O = 16) ก. 1.12 ข. 2.24 ค. 3.36 ง. 4.48	ในการเผา $KClO_3$ จะเกิดปฏิกิริยาดังนี้ $2KClO_3 \longrightarrow 2KCl + 3O_2$ ถ้าเผา $KClO_3$ จำนวน 12.26 กรัม จะได้ก๊าซ O_2 กี่ลิตรที่ (STP) (K = 39, Cl = 35.5, O = 16) ก. 1.12 ข. 2.24 ค. 3.36 ง. 4.48

หลังการตรวจสอบคุณภาพของแบบทดสอบแบบทดสอบตามวิธีการให้คะแนนของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ ผู้วิจัยนำแบบทดสอบไปหาคะแนนจุดตัด โดยวิธีของ Angoff คะแนนจุดตัดของแบบทดสอบตามวิธีการให้คะแนนของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ เป็นวิธีที่กำหนดคะแนนจุดตัดของผู้เชี่ยวชาญในรายวิชาเคมี จำนวน 5 คน ซึ่งก็เป็นชุดเดียวกับที่พิจารณาหาค่าความตรงตามเชิงเนื้อหา (Content validity) ดังแสดงในตารางที่ 4-4

ตารางที่ 4-4 ผลการพิจารณาคะแนนจุดตัดของแบบทดสอบตามวิธีการให้คะแนนของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ

ข้อที่	ผู้เชี่ยวชาญคนที่				
	1	2	3	4	5
1	0.65	0.45	0.55	0.60	0.60
2	0.60	0.80	0.78	0.50	0.55
3	0.40	0.70	0.60	0.60	0.70
4	0.50	0.55	0.60	0.70	0.50
5	0.40	0.60	0.45	0.60	0.50
6	0.75	0.50	0.50	0.50	0.60
7	0.70	0.60	0.50	0.60	0.60
8	0.80	0.50	0.75	0.60	0.45
9	0.80	0.50	0.60	0.40	0.55
10	0.40	0.40	0.60	0.40	0.65
11	0.60	0.40	0.55	0.50	0.50
12	0.75	0.40	0.60	0.50	0.60
13	0.40	0.70	0.55	0.50	0.55
14	0.60	0.55	0.60	0.40	0.70
15	0.50	0.50	0.55	0.40	0.55
16	0.50	0.60	0.60	0.40	0.50
17	0.60	0.50	0.65	0.70	0.55
18	0.40	0.70	0.60	0.65	0.45

ตารางที่ 4-4 (ต่อ)

ข้อที่	ผู้เชี่ยวชาญคนที่				
	1	2	3	4	5
19	0.50	0.60	0.40	0.50	0.45
20	0.70	0.60	0.50	0.60	0.50
21	0.50	0.50	0.60	0.50	0.40
22	0.50	0.50	0.65	0.70	0.60
23	0.50	0.50	0.70	0.70	0.55
24	0.75	0.60	0.55	0.60	0.70
25	0.50	0.75	0.45	0.50	0.70
26	0.60	0.60	0.70	0.50	0.65
27	0.60	0.40	0.55	0.50	0.55
28	0.40	0.50	0.80	0.75	0.50
29	0.40	0.50	0.65	0.60	0.55
30	0.50	0.50	0.40	0.70	0.50
รวม	16.80	16.50	17.58	16.70	16.75
เฉลี่ย	16.87				

จากตารางที่ 4-4 พบว่า ผลการพิจารณาคะแนนจุดตัดของแบบทดสอบตามวิธีการให้คะแนนของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ โดยวิธี Angoff ทั้งฉบับ มีคะแนนจุดตัดเท่ากับ 16.87

ผลการตรวจสอบคุณภาพแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจทั้งฉบับ

ตารางที่ 4-5 คุณภาพแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจทั้งฉบับจากการทดลองใช้ครั้งที่ 1 และครั้งที่ 2 ทั้งฉบับ

วิธีการให้คะแนน	ครั้งที่ 1				ครั้งที่ 2				
	N	SD	SEM	r_{KR-20}	N	SD	SEM	r_{KR-20}	r_{cc}
แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา	50	4.79	2.30	0.77	210	5.01	1.94	0.85	0.97
แบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ	50	5.42	3.06	0.68	210	5.22	2.09	0.84	0.84

จากตารางที่ 4-5 แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ การนำไปทดลองใช้ครั้งที่ 1 พบว่า มีส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าเท่ากับ 4.79 และ 5.42 ตามลำดับ ความคลาดเคลื่อนของแบบทดสอบ (SEM) มีค่าเท่ากับ 2.30 และ 3.06 ตามลำดับ ความเที่ยงของความสอดคล้องภายใน (r_{KR-20}) มีค่าเท่ากับ 0.77 และ 0.68 ตามลำดับ จากการนำไปทดลองใช้ครั้งที่ 2 พบว่า มีส่วนเบี่ยงเบนมาตราฐาน (SD) มีค่าเท่ากับ 5.01 และ 5.22 ตามลำดับ ความคลาดเคลื่อนของแบบทดสอบ (SEM) มีค่าเท่ากับ 1.94 และ 2.09 ตามลำดับ ความเที่ยงของความสอดคล้องภายใน (r_{KR-20}) มีค่าเท่ากับ 0.85 และ 0.84 ตามลำดับ ค่าความเที่ยงแบบอิงเกณฑ์ (r_{cc}) มีค่าเท่ากับ 0.97 และ 0.84 ตามลำดับ

ตอนที่ 2 ผลการวิเคราะห์คุณภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ ในการนำไปใช้จริง ด้วยทฤษฎีการตอบสนองข้อสอบ (IRT) ในด้านฟังก์ชันสารสนเทศ ความยากและอำนาจจำแนก ความตรงตามสภาพ

ผู้วิจัยได้นำแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ นำแบบทดสอบใช้จริง ที่ได้รับการพัฒนาจนมีคุณภาพแล้วไปใช้จริง ซึ่งได้ผลการวิเคราะห์ดังต่อไปนี้

ผลการวิเคราะห์ผลการวิเคราะห์ค่าสถิติพื้นฐานของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ จากการนำแบบทดสอบใช้จริง

ตารางที่ 4-6 ผลการวิเคราะห์ค่าสถิติพื้นฐานของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ

วิธีการให้คะแนน	ข้อ	คะแนนเต็ม	$\overline{X}$	SD	Max	Min	SEM
แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา	30	30	13.56	5.78	29	3	2.28
แบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ	30	90	28.24	13.58	84	6	5.75

จากตารางที่ 4-6 ผลการวิเคราะห์ค่าสถิติพื้นฐานของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ พบว่า ค่าเฉลี่ย ($\overline{X}$) เท่ากับ 13.39 และ 28.24 ตามลำดับ ส่วนเบี่ยงเบนมาตรฐาน (SD) เท่ากับ 5.76 และ 13.58 ตามลำดับ ความคลาดเคลื่อนของแบบทดสอบ (SEM) มีค่าเท่ากับ 2.28 และ 5.75 ตามลำดับ แล้วนำแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ ไปหาคุณภาพแบบทดสอบที่วิเคราะห์ด้วยทฤษฎีการตอบสนองข้อสอบ (IRT)

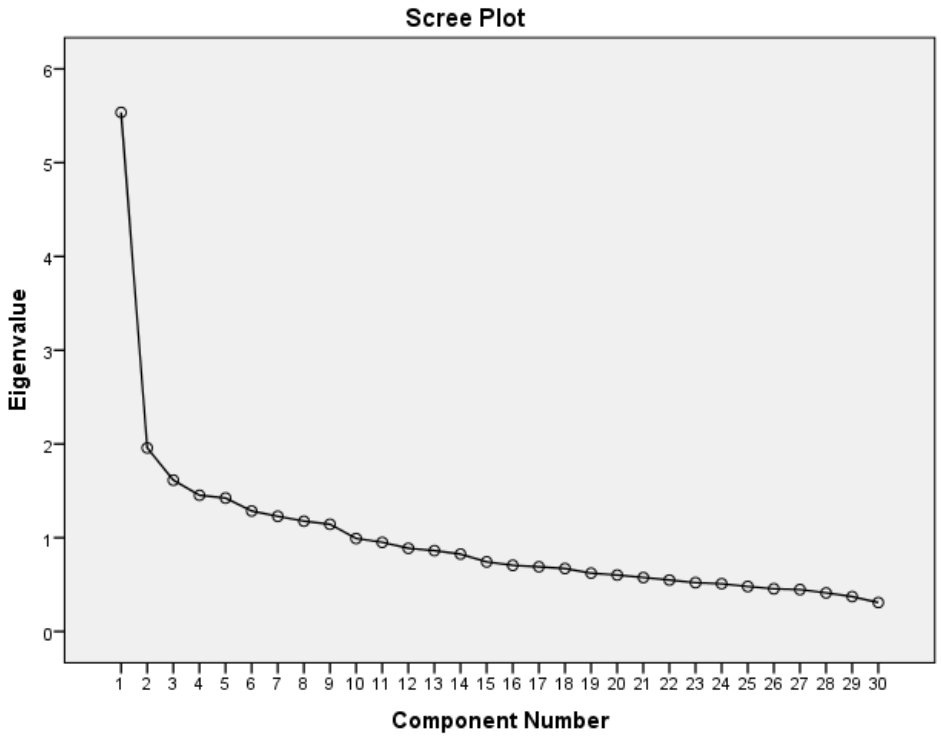
จากนั้นนำแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและแบบพิจารณาความมั่นใจนำมาวิเคราะห์องค์ประกอบ เพื่อตรวจสอบความเป็นมิติเดียว (Unidimensional) ตามข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ ซึ่งได้ผลดังตารางที่ 4-7

ตารางที่ 4-7 ผลการวิเคราะห์องค์ประกอบเพื่อตรวจสอบความเป็นมิติเดียว (Unidimensional) ค่าสถิติ Kaiser-Meyer-Olkin (KMO) และค่าสถิติ Bartlett's test of sphericity ของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ

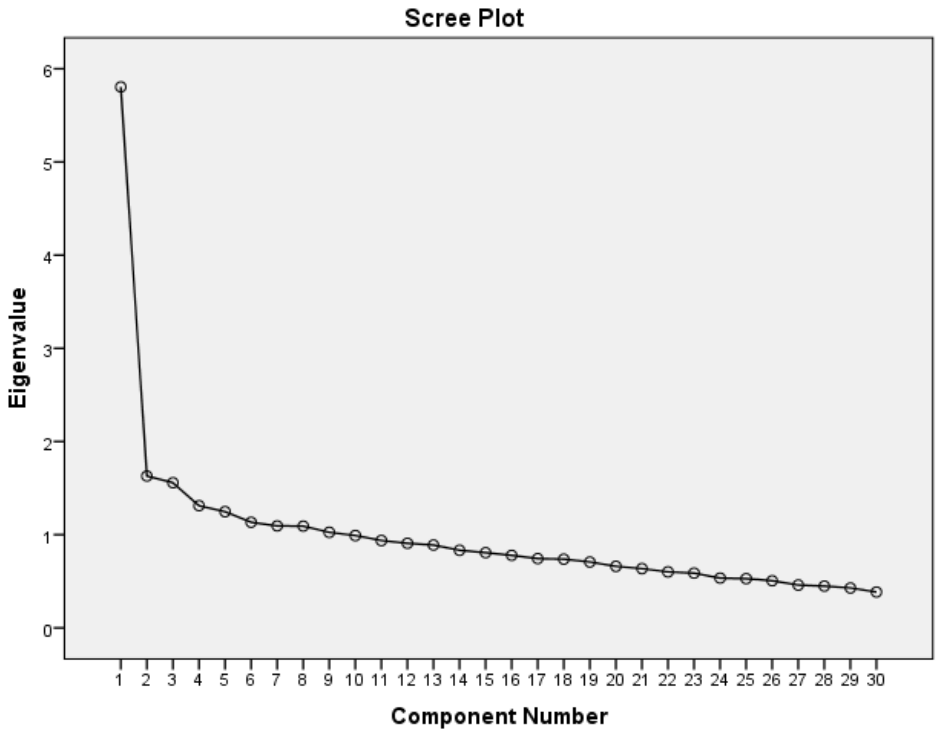
วิธีการให้คะแนน	Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy	Bartlett's test of sphericity		
		Chi-square (χ^2)	df	P-value
แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา	.807	4154.283*	435	.000
แบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ	.856	3625.837*	435	.000

*p < .05

จากตารางที่ 4-7 ในขั้นตอนแรกในการวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory factor analysis) ค่าสถิติ Kaiser-Meyer-Olkin (KMO) และค่าสถิติ Bartlett's test of sphericity ค่า KMO ไม่ควรน้อยกว่า 0.5 (Field, 2000) ผลการวิเคราะห์พบว่า ข้อมูลที่นำมาวิเคราะห์มีความสัมพันธ์กันดี และเมื่อพิจารณาค่า Bartlett's test of sphericity พบว่า ไคสแคว์มีค่า 4154.283 และ 3625.837 ที่ระดับนัยสำคัญ .05 แสดงว่าเมทริกซ์สหสัมพันธ์แตกต่างจากเมทริกซ์เอกลักษณ์ นั่นคือ ตัวแปรที่นำมาวิเคราะห์องค์ประกอบมีความสัมพันธ์กันจริง



ภาพที่ 4-1 ค่าไอแกน (Eigan value) ของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา



ภาพที่ 4-2 ค่าไอแกน (Eigan value) ของแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ

ตารางที่ 4-8 จำนวนองค์ประกอบ ค่าไอเกน ค่าร้อยละความแปรปรวน และค่าร้อยละของความแปรปรวนสะสมในแต่ละองค์ประกอบของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ

องค์ประกอบ		ค่าไอเกน		ค่าร้อยละความแปรปรวน		ค่าร้อยละของความแปรปรวนสะสม	
ธรรมดา	มั่นใจ	ธรรมดา	มั่นใจ	ธรรมดา	มั่นใจ	ธรรมดา	มั่นใจ
1	1	18.453	19.350	18.453	19.350	18.453	19.350
2	2	6.527	5.431	6.527	5.431	24.980	24.781
3	3	5.381	5.192	5.381	5.192	30.361	29.973
4	4	4.842	4.371	4.842	4.371	35.203	34.344
5	5	4.744	4.156	4.744	4.156	39.947	38.500
6	6	4.283	3.778	4.283	3.778	44.230	42.278
7	7	4.094	3.647	4.094	3.647	48.324	45.925
8	8	3.923	3.638	3.923	3.638	52.247	49.563
9	9	3.811	3.418	3.811	3.418	56.059	52.982

จากตารางที่ 4-8 ผู้วิจัยได้ดำเนินการสกัดองค์ประกอบของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ ด้วยวิธีการวิเคราะห์องค์ประกอบหลักและหมุนแกนด้วยวิธี Direct Oblimin oblique ได้องค์ประกอบทั้งหมด 9 องค์ประกอบ มีค่าไอเกนระหว่าง 18.453-3.811 และ 19.350-3.418 และมีค่าความแปรปรวนสะสม 56.059 และ 52.982 ตามลำดับ ค่าไอแกนขององค์ประกอบที่ 1 มีค่าเท่ากับ 18.453 และ 19.350 สูงกว่าค่าไอแกนขององค์ประกอบที่ 2 และค่าไอแกนขององค์ประกอบที่เหลือมีค่าใกล้เคียงกัน ซึ่งแตกต่างกันอย่างชัดเจน ถือได้ว่าแบบทดสอบนั้นมีความเป็นมิติเดียว (สุพัฒน์ สุกมลสันต์, 2539) สะท้อนให้เห็นว่าคะแนนของผลการตอบแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจมีความเป็นมิติเดียวค่อนข้างชัดเจน สามารถนำไปประมาณค่าพารามิเตอร์และวิเคราะห์ตามหลักการของทฤษฎีการตอบสนองข้อสอบได้

ผลการวิเคราะห์ค่าฟังก์ชันสารสนเทศแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ

ตารางที่ 4-9 ผลการวิเคราะห์ค่าฟังก์ชันสารสนเทศจำแนกรายข้อ จากแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ จำนวน 30 ข้อ

ข้อ	(θ)	I (θ)														
		-2.8	**-2.4**	**-2.0**	**-1.6**	**-1.2**	**-0.8**	**-0.4**	**0.0**	**0.4**	**0.8**	**1.2**	**1.6**	**2.0**	**2.4**	**2.8**
1	ธรรมดา	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	ความมั่นใจ	0.02	0.02	0.02	0.03	0.03	0.03	0.03	**0.03**	0.03	0.03	0.03	0.03	0.03	0.03	0.03
2	ธรรมดา	0.10	0.11	0.12	0.12	0.11	0.10	0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.02
	ความมั่นใจ	0.11	0.12	0.14	0.15	0.16	0.16	**0.16**	0.15	0.13	0.12	0.10	0.09	0.08	0.07	0.06
3	ธรรมดา	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	ความมั่นใจ	0.01	0.01	0.02	0.02	0.02	0.02	0.02	**0.03**	0.03	0.03	0.03	0.04	0.04	0.04	0.04
4	ธรรมดา	0.03	0.03	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.03	0.03	0.03
	ความมั่นใจ	0.03	0.04	0.04	0.05	0.06	0.06	0.07	0.07	0.07	**0.08**	0.08	0.08	0.07	0.07	0.07
5	ธรรมดา	0.04	0.07	0.11	0.16	0.23	0.31	0.37	0.39	0.37	0.31	0.23	0.16	0.11	0.07	0.04
	ความมั่นใจ	0.04	0.06	0.09	0.12	0.18	0.24	0.31	0.38	**0.42**	0.41	0.37	0.31	0.25	0.19	0.15
6	ธรรมดา	0.02	0.04	0.08	0.17	0.31	0.53	0.76	**0.85**	0.74	0.51	0.29	0.16	0.08	0.04	0.02
	ความมั่นใจ	0.03	0.05	0.08	0.13	0.21	0.31	0.44	0.55	0.61	0.59	0.51	0.41	0.31	0.24	0.19
7	ธรรมดา	0.06	0.09	0.12	0.17	0.22	0.26	**0.29**	0.29	0.26	0.22	0.18	0.13	0.09	0.06	0.04
	ความมั่นใจ	0.05	0.07	0.10	0.13	0.16	0.20	0.23	0.26	0.27	0.26	0.24	0.21	0.18	0.15	0.13

ตารางที่ 4-9 (ต่อ)

ข้อ	(θ)	I (θ)														
		-2.8	**-2.4**	**-2.0**	**-1.6**	**-1.2**	**-0.8**	**-0.4**	**0.0**	**0.4**	**0.8**	**1.2**	**1.6**	**2.0**	**2.4**	**2.8**
8	ธรรมดา	0.02	0.02	0.03	0.03	0.04	0.05	0.06	0.06	0.07	0.08	0.08	0.09	0.09	0.09	0.08
	ความมั่นใจ	0.01	0.02	0.02	0.03	0.03	0.04	0.04	0.05	0.06	0.07	0.08	0.08	0.09	0.09	**0.10**
9	ธรรมดา	0.10	0.15	0.21	0.28	0.34	**0.38**	0.37	0.32	0.25	0.18	0.12	0.08	0.05	0.03	0.02
	ความมั่นใจ	0.08	0.11	0.15	0.20	0.25	0.30	0.33	0.34	0.33	0.29	0.24	0.19	0.15	0.11	0.09
10	ธรรมดา	0.03	0.04	0.07	0.11	0.16	0.22	0.29	0.34	0.37	0.35	0.30	0.23	0.17	0.11	0.08
	ความมั่นใจ	0.03	0.04	0.05	0.08	0.11	0.16	0.23	0.30	0.38	**0.43**	0.43	0.39	0.33	0.26	0.19
11	ธรรมดา	0.04	0.06	0.10	0.16	0.24	0.33	0.42	0.46	0.44	0.37	0.27	0.18	0.12	0.07	0.04
	ความมั่นใจ	0.03	0.05	0.08	0.13	0.19	0.29	0.40	0.50	**0.55**	0.53	0.46	0.36	0.27	0.20	0.15
12	ธรรมดา	0.04	0.08	0.14	0.23	0.37	0.52	**0.62**	0.61	0.49	0.34	0.21	0.12	0.07	0.04	0.02
	ความมั่นใจ	0.05	0.07	0.11	0.17	0.24	0.34	0.44	0.52	0.53	0.48	0.39	0.30	0.22	0.16	0.11
13	ธรรมดา	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	ความมั่นใจ	0.02	0.02	0.02	0.02	**0.03**	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
14	ธรรมดา	0.04	0.05	0.06	0.08	0.10	0.11	0.13	**0.14**	0.14	0.14	0.13	0.12	0.11	0.09	0.07
	ความมั่นใจ	0.03	0.04	0.04	0.05	0.06	0.07	0.08	0.09	0.09	0.10	0.10	0.10	0.10	0.10	0.09
15	ธรรมดา	0.09	0.13	0.19	0.26	0.32	0.36	**0.37**	0.33	0.27	0.20	0.14	0.09	0.06	0.04	0.02
	ความมั่นใจ	0.07	0.09	0.12	0.15	0.18	0.21	0.23	0.24	0.24	0.23	0.21	0.19	0.16	0.13	0.11

ตารางที่ 4-9 (ต่อ)

ข้อ	(θ)	I (θ)														
		-2.8	**-2.4**	**-2.0**	**-1.6**	**-1.2**	**-0.8**	**-0.4**	**0.0**	**0.4**	**0.8**	**1.2**	**1.6**	**2.0**	**2.4**	**2.8**
16	ธรรมดา	0.07	0.10	0.13	0.16	0.18	0.20	0.21	0.21	0.19	0.16	0.13	0.10	0.08	0.06	0.04
	ความมั่นใจ	0.06	0.08	0.11	0.14	0.18	0.23	0.26	0.28	**0.29**	0.27	0.24	0.20	0.16	0.13	0.10
17	ธรรมดา	0.04	0.06	0.07	0.08	0.10	0.11	0.13	0.13	0.13	0.13	0.12	0.11	0.09	0.08	0.07
	ความมั่นใจ	0.04	0.05	0.07	0.09	0.11	0.14	0.17	0.20	0.22	**0.24**	0.23	0.22	0.20	0.17	0.14
18	ธรรมดา	0.02	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
	ความมั่นใจ	0.02	0.02	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.03	0.03	**0.04**	0.04	0.04	0.04
19	ธรรมดา	0.01	0.02	0.03	0.05	0.08	0.13	0.21	0.31	0.42	0.50	0.50	0.43	0.32	0.22	0.14
	ความมั่นใจ	0.01	0.02	0.03	0.04	0.06	0.10	0.16	0.25	0.37	0.50	0.59	**0.60**	0.52	0.40	0.29
20	ธรรมดา	0.06	0.08	0.12	0.17	0.23	0.28	0.31	**0.32**	0.30	0.25	0.19	0.14	0.10	0.07	0.04
	ความมั่นใจ	0.05	0.06	0.07	0.09	0.11	0.13	0.15	0.17	0.18	0.18	0.18	0.17	0.15	0.14	0.12
21	ธรรมดา	0.03	0.04	0.05	0.06	0.06	0.07	0.08	0.09	0.09	0.09	0.09	0.08	0.08	0.07	0.06
	ความมั่นใจ	0.03	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.10	**0.11**	0.11	0.11	0.11	0.10
22	ธรรมดา	0.03	0.03	0.04	0.05	0.07	0.08	0.09	0.11	0.12	0.13	0.13	0.13	0.12	0.11	0.10
	ความมั่นใจ	0.02	0.03	0.03	0.04	0.06	0.07	0.09	0.11	0.13	0.15	0.17	**0.18**	0.18	0.18	0.16
23	ธรรมดา	0.02	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
	ความมั่นใจ	0.02	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.03	**0.04**	0.04	0.04	0.04	0.04	0.04

ตารางที่ 4-9 (ต่อ)

ข้อ	(θ)	I (θ)														
		-2.8	**-2.4**	**-2.0**	**-1.6**	**-1.2**	**-0.8**	**-0.4**	**0.0**	**0.4**	**0.8**	**1.2**	**1.6**	**2.0**	**2.4**	**2.8**
24	ธรรมดา	0.01	0.02	0.04	0.08	0.15	0.26	0.41	0.57	0.67	**0.63**	0.49	0.32	0.19	0.11	0.06
	ความมั่นใจ	0.02	0.03	0.05	0.07	0.11	0.18	0.27	0.39	0.50	0.57	0.56	0.48	0.37	0.27	0.19
25	ธรรมดา	0.03	0.06	0.10	0.18	0.29	0.44	0.58	**0.63**	0.56	0.42	0.28	0.17	0.09	0.05	0.03
	ความมั่นใจ	0.04	0.06	0.09	0.13	0.20	0.29	0.39	0.47	0.52	0.49	0.42	0.33	0.25	0.18	0.13
26	ธรรมดา	0.01	0.03	0.06	0.11	0.22	0.41	0.65	**0.85**	0.85	0.65	0.41	0.22	0.11	0.06	0.03
	ความมั่นใจ	0.02	0.03	0.06	0.09	0.16	0.26	0.43	0.63	0.80	0.84	0.71	0.52	0.35	0.22	0.15
27	ธรรมดา	0.03	0.04	0.05	0.07	0.08	0.10	0.11	0.13	0.13	0.14	0.13	0.13	0.11	0.10	0.08
	ความมั่นใจ	0.03	0.03	0.04	0.05	0.06	0.08	0.09	0.11	0.12	0.14	**0.15**	0.15	0.15	0.14	0.13
28	ธรรมดา	0.04	0.05	0.05	0.06	0.07	0.08	0.08	0.09	0.09	0.09	0.09	0.08	0.07	0.07	0.06
	ความมั่นใจ	0.03	0.04	0.05	0.06	0.07	0.08	0.10	0.11	0.12	0.13	**0.14**	0.14	0.13	0.13	0.12
29	ธรรมดา	0.04	0.06	0.08	0.10	0.13	0.17	0.20	0.22	0.22	0.22	0.19	0.16	0.13	0.10	0.08
	ความมั่นใจ	0.03	0.04	0.06	0.07	0.10	0.13	0.16	0.20	0.24	0.26	**0.27**	0.27	0.24	0.21	0.17
30	ธรรมดา	0.05	0.06	0.07	0.08	0.09	0.10	0.11	0.11	0.11	0.10	0.10	0.09	0.08	0.06	0.05
	ความมั่นใจ	0.04	0.05	0.07	0.08	0.10	0.12	0.15	0.16	0.18	**0.19**	0.18	0.17	0.16	0.14	0.12

จากตารางที่ 4-9 ผลการวิเคราะห์ค่าฟังก์ชันสารสนเทศจำแนกรายข้อ พบว่า แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาจะให้ค่าฟังก์ชันสารสนเทศจำแนกรายข้อสูงกว่าแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ ซึ่งมีข้อ 6, 7, 9, 12, 14, 15, 20, 24, 25 และ 26 แบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจจะให้ค่าฟังก์ชันสารสนเทศจำแนกรายข้อสูงกว่า แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา ซึ่งมีข้อ 1, 2, 3, 4, 5, 8, 10, 11, 13, 16, 17, 18, 19, 21, 22, 23, 27, 28, 29 และ 30

ผลการวิเคราะห์ความยากและอำนาจจำแนกของแบบทดสอบ

ผลการวิเคราะห์ค่าความยากและค่าอำนาจจำแนกของของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจเป็นรายข้อ โดยวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (Item response theory) แบบทวิภาค ที่ใช้โมเดลโลจิสติก (Logistic) 2 พารามิเตอร์ และโมเดลการให้คะแนนความรู้บางส่วน (Genralized partial credit model: GPCM)

ตารางที่ 4-10 ผลการวิเคราะห์ค่าความยากและค่าอำนาจจำแนกของของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาโดยวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (Item response theory) แบบทวิภาค ที่ใช้โมเดลโลจิสติก (Logistic) 2 พารามิเตอร์

ข้อ	แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา	
	a	b
1	0.13	-1.35
2	0.68	-1.86
3	0.24	4.22
4	0.40	-0.18
5	1.25	-0.00
6	1.85	-0.02
7	1.08	-0.17
8	0.60	1.98
9	1.23	-0.68
10	1.21	0.44
11	1.36	0.07
12	1.58	-0.24

ตารางที่ 4-10 (ต่อ)

ข้อ	แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา	
	a	b
13	0.18	1.06
14	0.76	0.52
15	1.22	-0.57
16	0.92	-0.38
17	0.73	0.33
18	0.35	1.57
19	1.43	1.02
20	1.13	-0.12
21	0.60	0.78
22	0.72	1.22
23	0.35	1.54
24	1.64	0.50
25	1.59	-0.02
26	1.87	0.20
27	0.74	0.82
28	0.60	0.51
29	0.95	0.39
30	0.67	0.06

หมายเหตุ: b คือค่าความยาก และ a คือ ค่าอำนาจำแนก

จากตารางที่ 4-10 ผลการวิเคราะห์ค่าความยากและค่าอำนาจจำแนกของแบบทดสอบเลือกตอบที่ให้ให้คะแนนแบบธรรมดา พบว่า มีค่าความยากระหว่าง -1.86-4.22 ค่าอำนาจจำแนกระหว่าง 0.13-1.87 ซึ่งมีค่าอำนาจจำแนกของข้อสอบผ่านเกณฑ์ทุกข้อ ค่าความยากอยู่ระหว่าง -2.50 ถึง + 2.50 ค่าความยากที่อยู่ใกล้ -2.50 แสดงว่าเป็นแบบทดสอบที่ง่าย ค่าความยากที่อยู่ใกล้ +2.50 แสดงว่าเป็นแบบทดสอบที่ยาก และมีค่าอำนาจจำแนกอยู่ระหว่าง +0.50 ถึง +2.50
(ศิริชัย กาญจนวาสี, 2545, หน้า 48)

ตารางที่ 4-11 ผลการวิเคราะห์ค่าความยากและค่าอำนาจจำแนกของของแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ โดยวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (Item response theory) ที่ใช้โมเดลการให้คะแนนความรู้บางส่วน (Genralized partial credit model: GPCM)

ข้อ	แบบทดสอบเลือกตอบที่ให้ให้คะแนนแบบพิจารณาความมั่นใจ				
	a	**b**	**d_2**	**d_3**	**d_4**
1	0.14	2.10	-15.13	16.57	-1.44
2	0.36	-0.18	-2.63	4.18	-1.55
3	0.16	4.6	-13.19	12.49	0.70
4	0.23	1.66	-3.85	3.89	-0.03
5	0.56	0.95	-1.71	2.36	-0.65
6	0.73	1.01	-0.91	2.09	-1.18
7	0.47	1.13	-1.99	3.44	-1.45
8	0.23	3.14	-8.07	5.39	2.68
9	0.50	0.19	-1.54	1.71	-0.16
10	0.55	1.27	-1.87	1.83	0.05
11	0.64	0.86	-2.11	2.70	-0.59
12	0.61	0.51	-1.62	1.78	-0.16
13	0.15	2.68	-11.37	10.11	1.26
14	0.26	1.97	-5.76	5.29	0.47
15	0.45	0.71	-0.93	2.13	-1.20
16	0.45	0.60	-2.47	2.75	-0.28
17	0.40	1.32	-3.00	3.21	-0.21
18	0.17	4.36	-5.41	6.76	-1.36
19	0.63	1.68	-2.56	2.62	-0.06
20	0.38	1.36	-1.89	3.02	-1.13
21	0.26	1.98	-5.35	4.31	1.04
22	0.34	2.22	-3.98	3.50	0.48
23	0.18	4.18	-5.37	6.92	-1.55

ตารางที่ 4-11 (ต่อ)

ข้อ	แบบทดสอบเลือกตอบที่ให้ให้คะแนนแบบพิจารณาความมั่นใจ				
	a	b	d_2	d_3	d_4
24	0.62	1.22	-2.73	2.88	-0.16
25	0.60	0.76	-2.09	2.45	-0.36
26	0.76	0.83	-1.41	1.36	0.06
27	0.32	2.08	-4.02	3.84	0.17
28	0.32	1.97	-3.00	3.61	-0.61
29	0.43	1.41	-1.80	1.35	0.45
30	0.36	1.30	-2.41	2.69	-0.28

หมายเหตุ: a คือ ค่าอำนาจำแนก, b คือค่าความยาก และ d_2, d_3, d_4 คือ ค่า Threshold

จากตารางที่ 4-11 การวิเคราะห์ค่าความยากและอำนาจจำแนกของแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจมี พบว่า มีค่าความยากระหว่าง -0.18-4.6 และมีค่าอำนาจจำแนกอยู่ระหว่าง 0.14-0.76 ค่าอำนาจจำแนกผ่านเกณฑ์ทุกข้อ โดยที่ค่าความยากในทางทฤษฎีมีค่าระหว่าง ($-\infty$ถึง $+\infty$) แต่ในทางปฏิบัตินิยมใช้ข้อสอบที่ค่าความยากอยู่ระหว่าง -2.50 ถึง + 2.50 และค่าอำนาจจำแนกในทางทฤษฎีมีค่าระหว่าง ($-\infty$ถึง $+\infty$) ควรมีค่าเป็น + ตามปกติมีค่าไม่เกิน +2.50 (ศิริชัย กาญจนวาสี, 2545, หน้า 48)

ผลการวิเคราะห์ความยากและอำนาจจำแนกของแบบทดสอบทั้งฉบับ

ตารางที่ 4-12 ผลการวิเคราะห์ความยากและอำนาจจำแนกเฉลี่ยของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ

วิธีการให้คะแนน	จำนวนข้อสอบ	อำนาจจำแนก (a)				ความยาก (b)			
		Mean	SD	Max	Min	Mean	SD	Max	Min
แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา	30	0.93	0.50	1.87	0.13	0.58	1.20	4.22	-1.35
แบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ	30	0.41	0.18	0.76	0.14	1.66	1.17	4.60	-0.18

จากตารางที่ 4-12 ผลการวิเคราะห์ความยากและอำนาจจำแนกเฉลี่ย พบว่า อำนาจจำแนกเฉลี่ยของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดามีค่าสูงกว่าแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ ซึ่งมีอำนาจจำแนกเฉลี่ยเท่ากับ 0.93 และ 0.41 ตามลำดับ ความยากเฉลี่ยของแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจค่าสูงกว่าแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา ซึ่งมีความยากเฉลี่ยเท่ากับ 1.66 และ 0.58 ตามลำดับ

ผลการวิเคราะห์ความตรงตามสภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ

ผลการวิเคราะห์ค่าความตรงตามสภาพ ของวิธีการตอบและการตรวจให้คะแนน โดยคำนวณค่าสหสัมพันธ์ระหว่างคะแนนที่ได้จากแบบสอบที่มีวิธีการตอบและตรวจให้คะแนนทั้ง 2 วิธี วิเคราะห์ค่าความตรงตามสภาพ จากค่าสัมประสิทธิ์สหสัมพันธ์ (r) ระหว่างคะแนนจากแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาเคมี ภาคเรียนที่ 2 ชั้นมัธยมศึกษาปีที่ 4 ที่ได้จากวิธีการตอบ และการให้คะแนนในแต่ละวิธี กับระดับผลการเรียนวิชาเคมี ภาคเรียนที่ 1 ชั้นมัธยมศึกษาปีที่ 4 แล้วแปลงค่าความตรงตามสภาพจากค่าสัมประสิทธิ์สหสัมพันธ์ เป็นคะแนนฟิชเชอร์ซี (Z_r) โดยใช้สูตร Glass and Stanley (1970, p. 265)

ตารางที่ 4-13 ผลการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ (r) ระหว่างคะแนนจากแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาเคมี ภาคเรียนที่ 2 ชั้นมัธยมศึกษาปีที่ 4 ที่ได้จากวิธีการตอบ และการให้คะแนนในแต่ละวิธี กับระดับผลการเรียนวิชาเคมี ภาคเรียนที่ 1 ชั้นมัธยมศึกษาปีที่ 4

	คะแนนจากแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา	**คะแนนจากแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ**
ระดับผลการเรียนวิชาเคมี ภาคเรียนที่ 1	.984**	.581**

**Correlation is significant at the 0.01 level (2-tailed)

จากตารางที่ 4-13 ผลการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ (r) พบว่า ความสัมพันธ์ระหว่างระดับผลการเรียนวิชาเคมี ภาคเรียนที่ 1 กับคะแนนจากแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดามีค่าความสัมพันธ์อยู่ในระดับสูงซึ่งมีค่าเท่ากับ .984 และความสัมพันธ์ระหว่างระดับผลการเรียนวิชาเคมี ภาคเรียนที่ 1 กับคะแนนจากแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณา

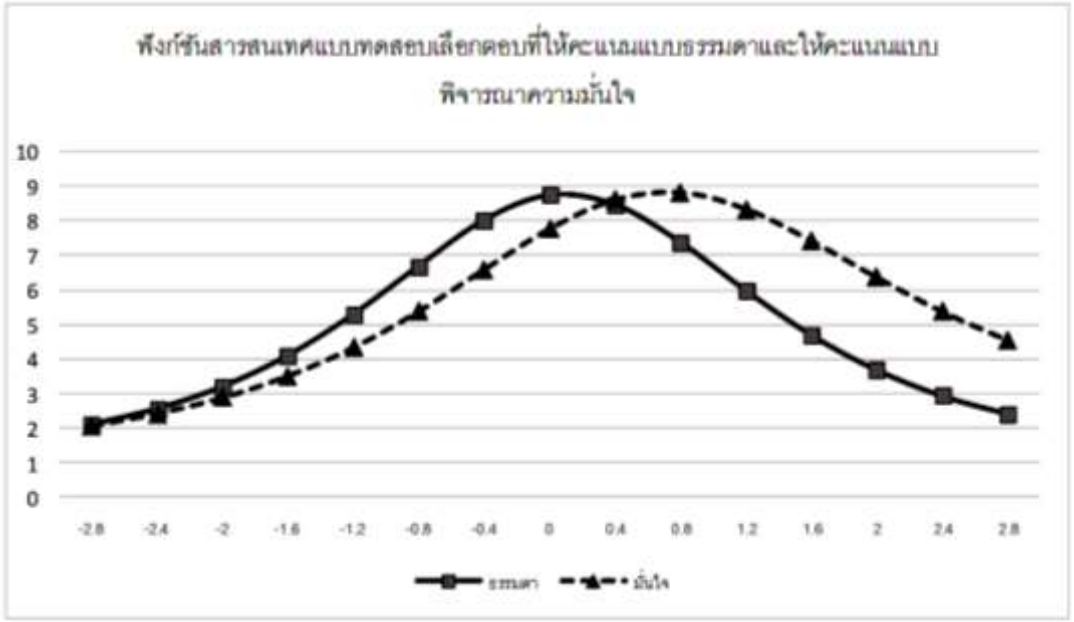
ความมั่นใจมีค่าความสัมพันธ์อยู่ในระดับปานกลางซึ่งมีค่าเท่ากับ .581 อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

ตอนที่ 3 ผลเปรียบเทียบประสิทธิภาพสัมพัทธ์และความตรงตามสภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ

ผลการเปรียบเทียบฟังก์ชันสารสนเทศแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ จำแนกตามระดับความสามารถของผู้สอบ

ตารางที่ 4-14 ผลการเปรียบเทียบฟังก์ชันสารสนเทศแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ จำแนกตามระดับความสามารถของผู้สอบ

	ฟังก์ชันสารสนเทศแบบทดสอบ (TIF)														
(θ)	**-2.8**	**-2.4**	**-2.0**	**-1.6**	**-1.2**	**-0.8**	**-0.4**	**0.0**	**0.4**	**0.8**	**1.2**	**1.6**	**2.0**	**2.4**	**2.8**
ธรรมดา	2.12	2.57	3.21	4.11	5.30	6.70	8.03	**8.75**	8.47	7.38	5.98	4.69	3.68	2.93	2.41
ความมั่นใจ	2.07	2.42	2.88	3.51	4.33	5.38	6.59	7.77	8.61	**8.82**	8.34	7.42	6.36	5.37	4.54



ภาพที่ 4-3 เปรียบเทียบฟังก์ชันสารสนเทศแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจจำแนกตามระดับความสามารถของผู้สอบ

จากตารางที่ 4-14 และ ภาพที่ 4-3 ผลการเปรียบเทียบฟังก์ชันสารสนเทศแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจจำแนกตามระดับความสามารถของผู้สอบ พบว่า แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา มีค่าฟังก์ชันสารสนเทศแบบทดสอบสูงเมื่อผู้สอบมีความสามารถระดับ (θ = 0.0) และแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ มีค่าฟังก์ชันสารสนเทศแบบทดสอบสูงเมื่อผู้สอบมีความสามารถระดับ (θ = 0.8)

ผลการเปรียบเทียบค่าต่ำสุด ค่าสูงสุด และค่าเฉลี่ยฟังก์ชันสารสนเทศแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ

ตารางที่ 4-15 ผลการเปรียบเทียบค่าต่ำสุด ค่าสูงสุด และค่าเฉลี่ยฟังก์ชันสารสนเทศแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ

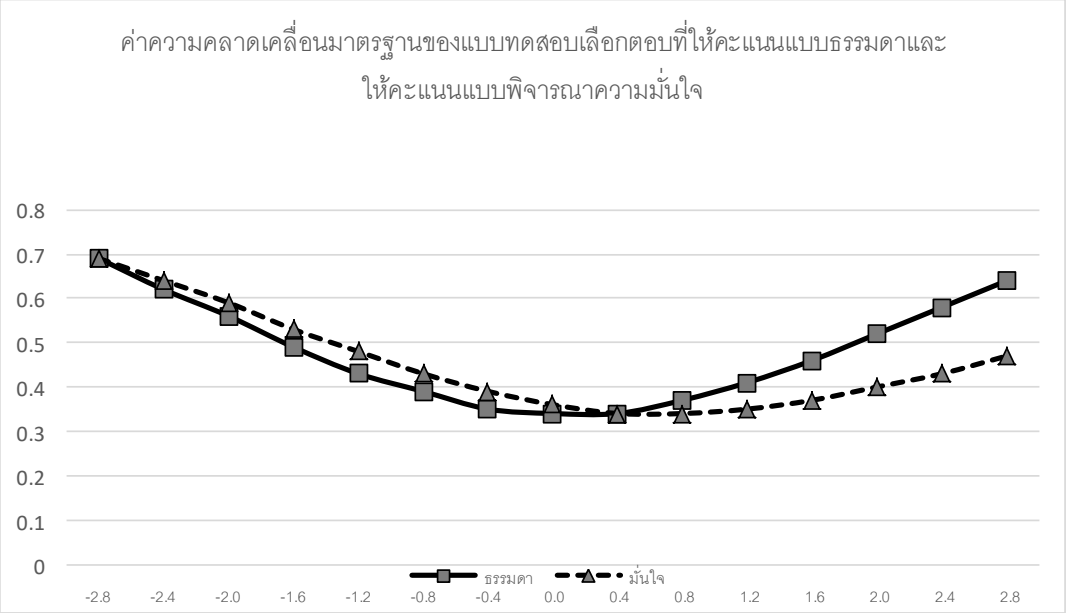
วิธีการให้คะแนน	**ฟังก์ชันสารสนเทศแบบทดสอบ (TIF)**		
	ต่ำสุด	**สูงสุด**	**เฉลี่ย**
ธรรมดา	2.12	8.75	5.09
ความมั่นใจ	2.07	8.82	5.63

จากตารางที่ 4-15 พบว่า แบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจจะมีค่าเฉลี่ยฟังก์ชันสารสนเทศแบบทดสอบสูงกว่าแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา

ผลการเปรียบเทียบค่าความคลาดเคลื่อนมาตรฐานของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ

ตารางที่ 4-16 ผลการเปรียบเทียบค่าความคลาดเคลื่อนมาตรฐาน SE (θ) ของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ

	ค่าความคลาดเคลื่อนมาตรฐาน SE (θ)														
(θ)	-2.8	**-2.4**	**-2.0**	**-1.6**	**-1.2**	**-0.8**	**-0.4**	**0.0**	**0.4**	**0.8**	**1.2**	**1.6**	**2.0**	**2.4**	**2.8**
ธรรมดา	0.69	0.62	0.56	0.49	0.43	0.39	0.35	**0.34**	0.34	0.37	0.41	0.46	0.52	0.58	0.64
ความมั่นใจ	0.69	0.64	0.59	0.53	0.48	0.43	0.39	0.36	**0.34**	0.34	0.35	0.37	0.40	0.43	0.47



ภาพที่ 4-4 เปรียบเทียบค่าความคลาดเคลื่อนมาตรฐาน SE (θ) ของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ

จากตารางที่ 4-16 และภาพที่ 4-4 ผลการเปรียบเทียบค่าความคลาดเคลื่อนมาตรฐาน SE (θ) ของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ พบว่า แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดามีค่าความคลาดเคลื่อนมาตรฐาน SE (θ) ของแบบทดสอบระหว่าง 0.34-0.69 ค่าความคลาดเคลื่อนมาตรฐาน SE (θ) ต่ำสุด ระดับความสามารถที่ (θ = 0.0) และแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจมีค่าความคลาดเคลื่อนมาตรฐาน SE (θ) ของแบบทดสอบระหว่าง 0.34-0.69 ค่าความคลาดเคลื่อนมาตรฐาน SE (θ) ต่ำสุด ระดับความสามารถที่ (θ = 0.4)

ผลการเปรียบเทียบค่าประสิทธิภาพสัมพัทธ์ RE (θ) ของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ

ตารางที่ 4-17 ผลการเปรียบเทียบค่าประสิทธิภาพสัมพัทธ์ RE (θ) ระหว่างค่าฟังก์ชันสารสนเทศระหว่างแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ

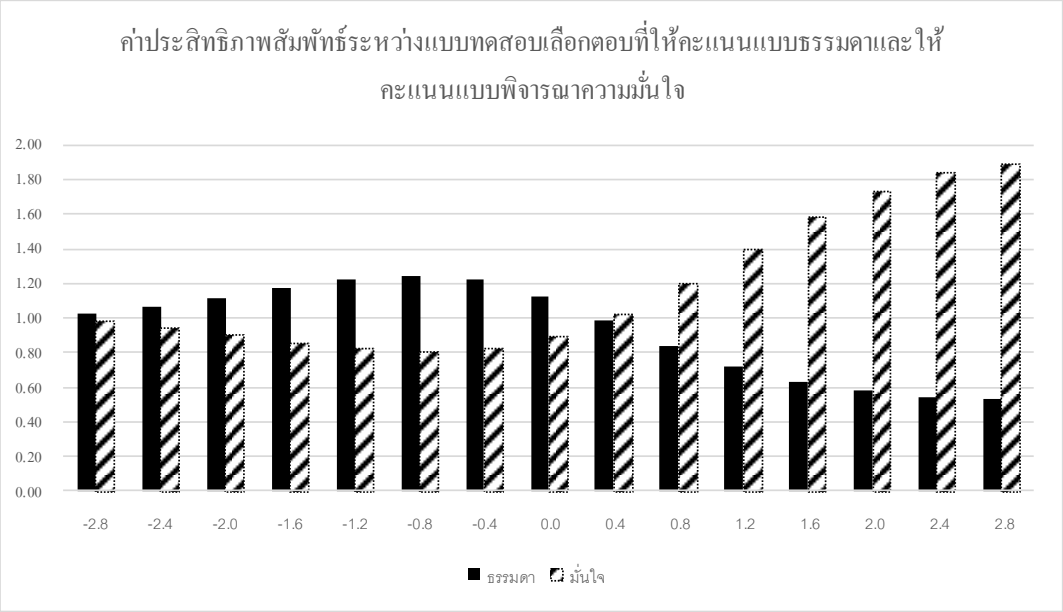
ระดับความสามารถ (θ)	วิธีการให้คะแนน		ค่าประสิทธิภาพสัมพัทธ์ RE (θ)
	แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา (TIF)	แบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ (TIF)	
-2.8	2.12	2.07	1.02
-2.4	2.57	2.42	1.06
-2.0	3.21	2.88	1.11
-1.6	4.11	3.51	1.17
-1.2	5.30	4.33	1.22
-0.8	6.70	5.38	1.25
-0.4	8.03	6.59	1.22
0.0	8.75	7.77	1.13
0.4	8.47	8.61	0.98
0.8	7.38	8.82	0.84
1.2	5.98	8.34	0.72
1.6	4.69	7.42	0.63
2.0	3.68	6.36	0.58
2.4	2.93	5.37	0.55
2.8	2.41	4.54	0.53

จากตารางที่ 4-17 และ ภาพที่ 4-5 ผลการเปรียบเทียบค่าประสิทธิภาพสัมพัทธ์ RE (θ) ระหว่างค่าฟังก์ชันสารสนเทศระหว่างแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและ

ให้คะแนนแบบพิจารณาความมั่นใจ พบว่า แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา มีค่าประสิทธิภาพสัมพัทธ์ RE (θ) สูงกว่า แบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ ที่มีความสามารถระดับ (θ = -2.8 ถึง 0.0)

ตารางที่ 4-18 ผลการเปรียบเทียบค่าประสิทธิภาพสัมพัทธ์ RE (θ) ระหว่างค่าฟังก์ชันสารสนเทศระหว่างให้แบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจและแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา

ระดับความสามารถ (θ)	วิธีการให้คะแนน		ค่าประสิทธิภาพสัมพัทธ์ RE (θ)
	แบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ (TIF)	แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา (TIF)	
-2.8	2.12	2.07	0.98
-2.4	2.57	2.42	0.94
-2.0	3.21	2.88	0.90
-1.6	4.11	3.51	0.85
-1.2	5.30	4.33	0.82
-0.8	6.70	5.38	0.80
-0.4	8.03	6.59	0.82
0.0	8.75	7.77	0.89
0.4	8.47	8.61	1.02
0.8	7.38	8.82	1.20
1.2	5.98	8.34	1.39
1.6	4.69	7.42	1.58
2.0	3.68	6.36	1.73
2.4	2.93	5.37	1.83
2.8	2.41	4.54	1.88



ภาพที่ 4-5 ค่าประสิทธิ์ภาพสัมพัทธ์ระหว่างแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ

จากตารางที่ 4-18 และ ภาพที่ 4-5 ผลการเปรียบเทียบค่าประสิทธิภาพสัมพัทธ์ RE (θ) ระหว่างค่าฟังก์ชันสารสนเทศระหว่างแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ พบว่า แบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจมีค่าประสิทธิภาพสัมพัทธ์ RE (θ) สูงกว่า แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา ที่มีความสามารถระดับ (θ = 0.4 ถึง 2.8)

ผลการเปรียบเทียบความตรงตามสภาพระหว่างแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ

ตารางที่ 4-19 การแปลงค่าสัมประสิทธิ์สหสัมพันธ์ (r) ระหว่างคะแนนจากแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาเคมี ภาคเรียนที่ 2 ชั้นมัธยมศึกษาปีที่ 4 ที่ได้จากวิธีการตอบ และการให้คะแนนในแต่ละวิธี กับระดับผลการเรียนวิชาเคมี ภาคเรียนที่ 1 ชั้นมัธยมศึกษาปีที่ 4 จำนวน 658 คน เป็นคะแนนฟิชเชอร์ซี (Z_r) เมื่อนำมาเปรียบเทียบความแตกต่างกันโดยใช้สถิติซี (Z)

วิธีการให้คะแนน	ค่าสัมประสิทธิ์สหสัมพันธ์ (r)	แปลงเป็นคะแนนฟิชเชอร์ซี (Z_r)	Z
คะแนนจากแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา	.984**	2.443	
คะแนนจากแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ	.581**	0.662	32.23

**Correlation is significant at the 0.01 level (2-tailed)

จากตารางที่ 4-19 พบว่า จากค่าความตรงตามสภาพที่พิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์ (r) ระหว่างคะแนนจากแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาเคมี ภาคเรียนที่ 2 ชั้นมัธยมศึกษาปีที่ 4 ที่ได้จากวิธีการตอบ และการให้คะแนนในแต่ละวิธี พบว่า แต่ละรูปแบบการตอบและตรวจให้คะแนนมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่าแบบทดสอบตามวิธีการให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจมีความตรงตามสภาพ และเมื่อนำมาเปรียบเทียบความแตกต่างของค่าความตรงตามสภาพโดยแปลงค่าสหสัมพันธ์เป็นคะแนนฟิชเชอร์ซี (Z_r) ซึ่งมีการแจกแจงใกล้เคียงกับโค้งปกติและมีค่าความคลาดเคลื่อนมาตรฐานขึ้นอยู่กับกลุ่มตัวอย่าง พบว่า ค่าความตรงตามสภาพขึ้นอยู่กับวิธีการให้คะแนนในแต่ละวิธี ทดสอบความแตกต่างระหว่างค่าสหสัมพันธ์ 2 ค่าที่มาจากกลุ่มตัวอย่างด้วย สถิติซี (Z) พบว่า ความตรงตามสภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาสูงกว่า แบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ

บทที่ 5
สรุป อภิปรายผล และข้อเสนอแนะ

ในงานวิจัยครั้งนี้ผู้วิจัยมีวัตถุประสงค์ 1) เพื่อหาคุณภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ ในด้านฟังก์ชันสารสนเทศ ความยากและอำนาจจำแนก ความตรงตามสภาพ 2) เพื่อเปรียบเทียบคุณภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ ในด้านดังต่อไปนี้ 2.1) ฟังก์ชันสารสนเทศ 2.2) ความตรงตามสภาพ โดยมีสมมติฐานของการวิจัย คือ 1) ฟังก์ชันสารสนเทศของแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ มีประสิทธิภาพสัมพัทธ์สูงกว่าแบบธรรมดา 2) ความตรงตามสภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจแตกต่างกัน กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 4 ภาคเรียนที่ 2 ปีการศึกษา 2560 ของโรงเรียนในสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 18 ซึ่งมีทั้งหมด 11,060 คน ผู้วิจัยใช้โปรแกรม G*Power 3.1 ได้ขนาดกลุ่มตัวอย่าง 567 คน สำหรับการวิจัยครั้งนี้ผู้วิจัยใช้กลุ่มตัวอย่าง 658 คน โดยมีวิธีการสุ่มแบบหลายขั้นตอน (Multi-stage sampling)

เครื่องมือที่ใช้ในการวิจัยครั้งนี้ คือ แบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชา เคมี กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 4 ภาคเรียนที่ 2 ปีการศึกษา 2560 เรื่อง ปริมาณสารสัมพันธ์ แบบสอบชนิดเลือกตอบ 4 ตัวเลือก จำนวน 1 ฉบับ จำนวน 30 ข้อ ซึ่งมีวิธีการตอบและตรวจให้คะแนน 2 วิธี แบ่งเป็นแบบทดสอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ

การเก็บรวบรวมข้อมูล โดยผู้วิจัยได้นำเครื่องมือที่สร้างขึ้นให้ผู้เชี่ยวชาญ จำนวน 5 คน ตรวจสอบคุณภาพเบื้องต้น ด้านความตรงเชิงเนื้อหา (Content validity) ได้แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชา เคมี จำนวน 30 ข้อ ผู้วิจัยนำแบบทดสอบมาทดลองใช้ครั้งที่ 1 กลุ่มตัวอย่างจำนวน 50 คน จากนั้นทำการวิเคราะห์หาคุณภาพแบบทดสอบเพื่อปรับปรุงคุณภาพของแบบทดสอบ ผู้วิจัยนำแบบทดสอบไปหาคะแนนจุดตัด โดยวิธีของ Angoff เป็นวิธีที่กำหนดคะแนนจุดตัดของผู้เชี่ยวชาญในรายวิชา เคมี จำนวน 5 คน ผู้วิจัยนำแบบทดสอบที่ได้รับการปรับปรุง ไปทดสอบครั้งที่ 2 กลุ่มตัวอย่างจำนวน 210 คน จากนั้นนำแบบทดสอบที่ผ่านการทดลองใช้ทั้ง 2 ครั้ง จำนวน 30 ข้อ นำไปใช้จริงกับกลุ่มตัวอย่าง ซึ่งเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 4 ภาคเรียนที่ 2 ปีการศึกษา 2560 จำนวน 658 คน

การวิเคราะห์ข้อมูลดังนี้ ความเที่ยงตรงด้านเนื้อหา (Content validity) จาก การหาค่าดัชนีความสอดคล้องเป็นรายข้อ (Index of congruency: IOC) ค่าความยากและค่าอำนาจจำแนกของข้อสอบ โดยใช้ดัชนีบี (B-index) ตามวิธีของเบรนแนน (Brennan) ค่าความเที่ยงของความสอดคล้องภายใน (KR-20) ค่าความเที่ยงแบบอิงเกณฑ์ วิธีของ Livingston ค่าคะแนนจุดตัดโดยวิธีของ Angoff ค่าสถิติพื้นฐานของแบบทดสอบ วิเคราะห์ตัวประกอบสำคัญ (Principal componant) สำหรับการตรวจสอบพิจารณาค่าไอเกน (Eigen value) ค่าพารามิเตอร์ความยาก (b) ค่าพารามิเตอร์อำนาจจำแนก (a) ค่าฟังก์ชันสารสนเทศของข้อสอบ (IIF) ค่าฟังก์ชันสารสนเทศของแบบทดสอบ (TIF) ค่าความคลาดเคลื่อนมาตรฐานของแบบทดสอบ SE (θ) ค่าประสิทธิภาพสัมพัทธ์ (Relative efficiency: RE (θ)) ค่าความตรงตามสภาพ (Concurrent validity)

สรุปผลการวิจัย

ผู้วิจัยนำเสนอผลการวิจัยออกเป็น 2 ข้อ ดังนี้

1. ผลการวิเคราะห์คุณภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ ในการนำไปใช้จริง ด้วยทฤษฎีการตอบสนองข้อสอบ (IRT) ในด้านฟังก์ชันสารสนเทศ ความยากและอำนาจจำแนก ความตรงตามสภาพ

1.1 จากผลการวิเคราะห์ค่าสถิติพื้นฐานของแบบทดสอบ พบว่า แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ มีค่าเฉลี่ย ($\overline{X}$) เท่ากับ 13.39 และ 28.24 ตามลำดับ แบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจมีส่วนเบี่ยงเบนมาตรฐาน (SD) และความคลาดเคลื่อนของแบบทดสอบ (SEM) มากกว่าแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา จากการที่ผู้วิจัยได้ดำเนินการสกัดองค์ประกอบของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจด้วยวิธีการวิเคราะห์องค์ประกอบหลักและหมุนแกนด้วยวิธี Direct Oblimin oblique ค่าไอแกนขององค์ประกอบที่ 1 มีค่า สูงกว่าค่าไอแกนขององค์ประกอบที่ 2 และค่าไอแกนขององค์ประกอบที่เหลือมีค่าใกล้เคียงกัน ซึ่งแตกต่างกันอย่างชัดเจน ถือได้ว่าแบบทดสอบนั้นมีความเป็นมิติเดียว (Lord, 1980, p. 21) สะท้อนให้เห็นว่าคะแนนของผลการตอบแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจมีความเป็นมิติเดียวค่อนข้างชัดเจน สามารถนำไปประมาณค่าพารามิเตอร์และวิเคราะห์ตามหลักการของทฤษฎีการตอบสนองข้อสอบได้ Lord (1980) ได้เสนอแนะว่า การนำแนวคิดของทฤษฎีการตอบสนองข้อสอบไปใช้ให้สอดคล้องกับความสามารถผู้เรียนที่ต้องการวัด ควรให้ความสำคัญกับข้อตกลงเบื้องต้นที่กำหนดไว้ ซึ่งหนึ่งในข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ คือแบบทดสอบต้องมีมิติเดียว

1.2 ผลการวิเคราะห์ค่าฟังก์ชันสารสนเทศแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ

จากผลการวิเคราะห์ค่าฟังก์ชันสารสนเทศจำแนกรายข้อ พบว่า แบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจจะให้ค่าฟังก์ชันสารสนเทศจำแนกรายข้อ สูงกว่าแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาที่ระดับความสามารถปานกลางถึงสูง (θ = 0.4 ถึง 2.8)

1.3 ผลการวิเคราะห์ความยากและอำนาจจำแนกของแบบทดสอบ

ผลการวิเคราะห์ค่าความยากและค่าอำนาจจำแนกของของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจเป็นรายข้อ โดยวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) แบบทวิภาค ที่ใช้โมเดลโลจิสติก (Logistic) 2 พารามิเตอร์ และโมเดลการให้คะแนนความรู้บางส่วน (Genralized partial credit model: GPCM)

จากผลการวิเคราะห์ค่าความยากและค่าอำนาจจำแนกของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา พบว่า มีค่าความยากระหว่าง -1.86-4.22 ค่าอำนาจจำแนกระหว่าง 0.13-1.87 ซึ่งมีค่าอำนาจจำแนกของข้อสอบผ่านเกณฑ์ทุกข้อ และจากผลการวิเคราะห์ค่าความยากและอำนาจจำแนกของแบบทดสอบเลือกตอบที่ให้ ให้คะแนนแบบพิจารณาความมั่นใจมี พบว่า มีค่าความยากระหว่าง -0.18-4.6 และมีค่าอำนาจจำแนกอยู่ระหว่าง 0.14-0.76 ค่าอำนาจจำแนกผ่านเกณฑ์ทุกข้อ โดยที่ค่าอำนาจจำแนกในทางทฤษฎีมีค่าระหว่าง ($-\infty$ถึง $+\infty$) ควรมีค่าเป็น + ตามปกติมีค่าไม่เกิน +2.50 และค่าความยากในทางทฤษฎีมีค่าระหว่าง ($-\infty$ถึง $+\infty$) แต่ในทางปฏิบัตินิยมใช้ข้อสอบที่ค่าความยากอยู่ระหว่าง -2.50 ถึง +2.50 ค่าความยากที่อยู่ใกล้ -2.50 แสดงว่าเป็นแบบทดสอบที่ง่าย ค่าความยากที่อยู่ใกล้ +2.50 แสดงว่าเป็นแบบทดสอบที่ยาก (ศิริชัย กาญจนวาสี, 2545, หน้า 48) เช่นเดียวกับ ณัฐภรณ์ เลขะวัฒนพงษ์ (2560) ได้ศึกษาการเปรียบเทียบค่าความยากและค่าอำนาจจำแนก จากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาเคมี โดยวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (Item response theory) โมเดลการให้คะแนนความรู้บางส่วน (Genralized partial credit model: GPCM) ด้วยโปรแกรม IRTPRO พบว่า วิธีการให้คะแนนความรู้บางส่วนที่ต่างกันส่งผลต่อความยากของข้อสอบที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และวิธีการให้คะแนนความรู้บางส่วนและจำนวนตัวเลือกที่ต่างกันส่งผลต่อค่าอำนาจำแนกของข้อสอบที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

จากผลการวิเคราะห์ความยากและอำนาจจำแนกเฉลี่ย พบว่า อำนาจจำแนกเฉลี่ยของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดามีค่าสูงกว่าแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ ซึ่งมีอำนาจจำแนกเฉลี่ยเท่ากับ 0.93 และ 0.41 ตามลำดับ ความยากเฉลี่ย

ของแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจค่าสูงกว่าแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดามี ซึ่งมีความยากเฉลี่ยเท่ากับ 1.66 และ 0.58 ตามลำดับ

1.4 ผลการวิเคราะห์ความตรงตามสภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ

จากผลการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ (r) พบว่า ความสัมพันธ์ระหว่างระดับผลการเรียนวิชาเคมี ภาคเรียนที่ 1 กับคะแนนจากแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดามีค่าความสัมพันธ์อยู่ในระดับสูงซึ่งมีค่าเท่ากับ .984 และความสัมพันธ์ระหว่างระดับผลการเรียนวิชาเคมี ภาคเรียนที่ 1 กับคะแนนจากแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจมีค่าความสัมพันธ์อยู่ในระดับปานกลางซึ่งมีค่าเท่ากับ .581 อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยพิจารณาจากเกณฑ์ (กัญวลัญช์ จิตรดี, 2559, หน้า 145 อ้างถึงใน ศราวุธ รัตนะ, 2562, หน้า 147) ค่าสัมประสิทธิ์สหสัมพันธ์เข้าใกล้ 1 (ประมาณ 0.70 ถึง 0.90) ถือว่ามีความสัมพันธ์กันอยู่ในระดับสูง (ถ้าสูงกว่า 0.90 ก็ถือว่ามีความสัมพันธ์กันอยู่ในระดับสูงมาก)

จากค่าความตรงตามสภาพที่พิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์ (r) ระหว่างคะแนนจากแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาเคมี ภาคเรียนที่ 2 ชั้นมัธยมศึกษาปีที่ 4 ที่ได้จากวิธีการตอบ และการให้คะแนนในแต่ละวิธี พบว่า แต่ละรูปแบบการตอบและตรวจให้คะแนนมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่าแบบทดสอบตามวิธีการให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ มีความตรงตามสภาพ และเมื่อนำมาเปรียบเทียบความแตกต่างของค่าความตรงตามสภาพ โดยใช้คะแนนฟิชเชอร์ซี (Z_r) พบว่าค่าความตรงตามสภาพขึ้นอยู่กับวิธีการให้คะแนนในแต่ละวิธี โดยใช้สถิติซี (Z) พบว่า ความตรงตามสภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา สูงกว่า แบบทดสอบตามวิธีการให้คะแนนแบบพิจารณาความมั่นใจ

2. ผลเปรียบเทียบฟังก์ชันสารสนเทศและความตรงตามสภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ

2.1 ผลการเปรียบเทียบฟังก์ชันสารสนเทศแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ จำแนกตามระดับความสามารถของผู้สอบ

จากผลการเปรียบเทียบฟังก์ชันสารสนเทศแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจจำแนกตามระดับความสามารถของผู้สอบ พบว่า แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา มีค่าฟังก์ชันสารสนเทศแบบทดสอบสูงเมื่อผู้สอบมีความสามารถระดับปานกลาง ($\theta = 0.0$) และแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ มีค่าฟังก์ชันสารสนเทศแบบทดสอบสูงเมื่อผู้สอบมีความสามารถระดับ

ค่อนข้างสูง (θ = 0.8) และแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจจะมีค่าเฉลี่ยฟังก์ชันสารสนเทศแบบทดสอบสูงกว่าแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา

จากผลการเปรียบเทียบค่าประสิทธิภาพสัมพัทธ์ RE (θ) ระหว่างค่าฟังก์ชันสารสนเทศระหว่างแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ พบว่า แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดามีค่าประสิทธิภาพสัมพัทธ์ RE (θ) สูงกว่า แบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ ที่มีความสามารถระดับต่ำถึงปานกลาง (θ = -2.8 ถึง 0.0) และแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจมีค่าประสิทธิภาพสัมพัทธ์ RE (θ) สูงกว่า แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา ที่มีความสามารถระดับปานกลางถึงสูง (θ = 0.4 ถึง 2.8)

2.2 ผลการเปรียบเทียบความตรงตามสภาพ

จากผลการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ (r) ระหว่างคะแนนจากแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาเคมี ภาคเรียนที่ 2 ชั้นมัธยมศึกษาปีที่ 4 ที่ได้จากวิธีการตอบ และการให้คะแนนในแต่ละวิธี กับระดับผลการเรียนวิชาเคมี ภาคเรียนที่ 1 ชั้นมัธยมศึกษาปีที่ 4 จำนวน 658 คน เป็นคะแนนฟิชเชอร์ซี (Z_r) เมื่อนำมาเปรียบเทียบความแตกต่างกันโดยใช้สถิติซี (Z) พบว่า ค่าความตรงตามสภาพที่พิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์ (r) ระหว่างคะแนนจากแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาเคมี ภาคเรียนที่ 2 ชั้นมัธยมศึกษาปีที่ 4 ที่ได้จากวิธีการตอบและการให้คะแนนทั้ง 2 วิธี พบว่า แต่ละรูปแบบการตอบและตรวจให้คะแนนมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่าแบบทดสอบตามวิธีการให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ มีความตรงตามสภาพ และเมื่อนำมาเปรียบเทียบความแตกต่างของค่าความตรงตามสภาพ โดยใช้คะแนนฟิชเชอร์ซี (Z_r) พบว่า ค่าความตรงตามสภาพขึ้นอยู่กับวิธีการให้คะแนนในแต่ละวิธี โดยใช้สถิติซี (Z) พบว่า ความตรงตามสภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา สูงกว่า แบบทดสอบตามวิธีการให้คะแนนแบบพิจารณาความมั่นใจ

อภิปรายผล

1. ผลการวิเคราะห์คุณภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ ในด้านฟังก์ชันสารสนเทศ ความยากและอำนาจจำแนกความตรงตามสภาพ

ในการวิเคราะห์คุณภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ โดยวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (Item response

theory) แบบทวิภาค ที่ใช้โมเดลโลจิสติก (Logistic) 2 พารามิเตอร์ และโมเดลการให้คะแนนความรู้บางส่วน (Genralized partial credit model: GPCM) จากโปรแกรม IRTPRO ที่สามารถประมาณค่าฟังก์ชันสารสนเทศ ความยากและอำนาจจำแนกความตรงตามสภาพ ของแบบทดสอบ ที่มีวิธีการตอบและตรวจให้คะแนนทั้ง 2 วิธี พบว่า ค่าฟังก์ชันสารสนเทศแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ ส่วนใหญ่จะให้ค่าฟังก์ชันสารสนเทศจำแนกรายข้อสูงกว่า แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา และแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา มีค่าฟังก์ชันสารสนเทศแบบทดสอบสูงสุดเมื่อผู้สอบมีความสามารถระดับปานกลาง ($\theta = 0.0$) และแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ มีค่าฟังก์ชันสารสนเทศแบบทดสอบสูงสุดเมื่อผู้สอบมีความสามารถระดับค่อนข้างสูง ($\theta = 0.8$) สอดคล้องกับงานวิจัยของ นิตยา พลดงนอก (2540) ได้ศึกษาพัฒนาแบบวัดความเข้าใจในการอ่านภาษาไทยของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้ทฤษฎีการตอบสนองข้อสอบ พบว่า ค่าฟังก์ชันสารสนเทศของแบบวัดตามทฤษฎีการตอบสนองข้อสอบ จะมีประสิทธิภาพสูงสุดเมื่อใช้ทดสอบกับผู้ที่มีค่าระดับความสามารถระดับ ($\theta = 0.6$)

จากผลการวิเคราะห์ค่าความยากและค่าอำนาจจำแนกของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ พบว่า ค่าอำนาจจำแนกของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดามีค่าสูงกว่าแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ ค่าความยากของแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจค่าสูงกว่าแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดามี จากค่าอำนาจจำแนกของแบบทดสอบเลือกตอบที่ให้มีการตรวจให้คะแนนความรู้สมบูรณ์มีค่าสูงกว่าการตรวจให้คะแนนความรู้บางส่วน ทำให้คะแนนที่ประมาณค่าความสามารถของผู้สอบได้ใกล้เคียงกับความสามารถที่แท้จริง ทั้ง ๆ ที่การตรวจให้คะแนนความรู้บางส่วนเป็นวิธีที่ได้ข้อมูลของผู้สอบมาพิจารณาความสามารถมากกว่าการตรวจให้คะแนนความรู้สมบูรณ์ อาจเนื่องมาจากกลุ่มตัวอย่างส่วนใหญ่มักไม่นิยมตอบแบบทดสอบที่มีการตรวจให้คะแนนความรู้บางส่วน แต่ผู้สอบคุ้นชินกับทำแบบทดสอบที่มีการตรวจให้คะแนนความรู้สมบูรณ์ เพราะการตอบและการตรวจให้คะแนนด้วยวิธีการตรวจให้คะแนนความรู้สมบูรณ์ในกลุ่มที่มีความสามารถต่ำถึงปานกลางจะตอบทุกข้ออาจจะทำให้ได้คะแนนบ้างจากการเดาสุ่ม หรือถ้าผู้ตอบไม่แน่ใจเลยก็จะไม่ตอบ ทำให้ข้อนั้นไม่ได้คะแนน และอาจจะมีผลให้ค่าอำนาจจำแนกของการตรวจให้คะแนนความรู้สมบูรณ์สูงกว่าการตรวจให้คะแนนความรู้บางส่วน ส่วนค่าความยากของการตรวจให้คะแนนความรู้บางส่วนสูงกว่าการตรวจให้คะแนนความรู้สมบูรณ์ เนื่องจากเปิดโอกาสให้ผู้สอบที่มีความสามารถระดับปานกลางถึงสูงได้คะแนนในกรณีที่มีความรู้บางส่วนได้ โดยไม่ต้องเดา ทำให้กลุ่มที่มีความสามารถต่ำและ

ปานกลางไม่มีโอกาสที่จะได้คะแนนบางส่วนจากข้อสอบ แต่ได้คะแนนเต็มเมื่อตอบถูกและได้คะแนน 0 เมื่อตอบผิด ซึ่งไม่ต่างจากการตรวจให้คะแนนความรู้สมบูรณ์ ที่คะแนนที่ได้อาจจะเกิดจากการเดามากกว่าจะเป็นคะแนนความสามารถที่แท้จริง ซึ่งสอดคล้องกับงานวิจัย ของ รัตนา ไชยตรี (2546) ศึกษาการเปรียบเทียบความยากและอำนาจจำแนก ระหว่างแบบสอบที่มีวิธีการตอบและตรวจให้คะแนนทั้ง 3 รูปแบบ พบว่า แบบสอบที่มีการตอบและตรวจให้คะแนนโดยวิธีการตอบโดยบอกระดับความมั่นใจมีค่าความยากเฉลี่ยสูงสุด เมื่อเทียบกับวิธีอื่น ๆ รองลงมาคือ วิธีการประยุกต์การให้คะแนนของคูมบ์ส และวิธีประยุกต์การให้คะแนนของเดรสเซลและชมิดท์ ตามลำดับโดยแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจมีการให้คะแนนแตกต่างจากแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาจึงส่งผลค่าความยากและค่าอำนาจจำแนก สอดคล้องกับ ณัฐภรณ์ เลขะวัฒนพงษ์ (2560) ได้ศึกษาการเปรียบเทียบค่าความยากและค่าอำนาจจำแนก จากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชา เคมี โดยวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (Item response theory) โมเดลการให้คะแนนความรู้บางส่วน (Genralized partial credit model: GPCM) ด้วยโปรแกรม IRTPRO พบว่า วิธีการให้คะแนนความรู้บางส่วนที่ต่างกันส่งผลต่อความยากของข้อสอบแตกต่างกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งสอดคล้องกับงานวิจัย ของ รัตนา ไชยตรี (2546; เอมอร จังศิริพรปกรณ์, 2545) เนื่องจากแต่ละวิธีมีวิธีการตอบและการให้คะแนนที่แตกต่างกัน จึงส่งผลต่อการพิจารณาเลือกคำตอบโดยที่การให้คะแนนความรู้บางส่วนเปิดโอกาสให้ผู้สอบได้คะแนนในกรณีที่มีความรู้บางส่วนได้โดยไม่ต้องเดา ทำให้กลุ่มที่มีความสามารถต่ำและปานกลางมีโอกาสที่จะได้คะแนนจากแบบทดสอบแต่ละข้อบ้าง (เอมอร จังศิริพรปกรณ์, 2545)

จากผลการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ (r) ระหว่างคะแนนจากแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาเคมี ภาคเรียนที่ 2 ชั้นมัธยมศึกษาปีที่ 4 ที่ได้จากวิธีการตอบ และการให้คะแนนในแต่ละวิธี กับระดับผลการเรียนวิชาเคมี ภาคเรียนที่ 1 ชั้นมัธยมศึกษาปีที่ 4 จำนวน 658 คน เป็นคะแนนฟิชเชอร์ซี (Z_r) เมื่อนำมาเปรียบเทียบความแตกต่างกันโดยใช้สถิติซี (Z) พบว่าค่าความตรงตามสภาพที่พิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์ (r) ระหว่างคะแนนจากแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาเคมี ภาคเรียนที่ 2 ชั้นมัธยมศึกษาปีที่ 4 ที่ได้จากวิธีการตอบและการให้คะแนนในแต่ละวิธี พบว่า แต่ละรูปแบบการตอบและตรวจให้คะแนนมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่าแบบทดสอบตามวิธีการให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ มีความตรงตามสภาพ และเมื่อนำมาเปรียบเทียบความแตกต่างของค่าความตรงตามสภาพ โดยใช้คะแนนฟิชเชอร์ซี (Z_r) ค่าความตรงตามสภาพขึ้นอยู่กับวิธีการให้คะแนนในแต่ละวิธี โดยใช้สถิติซี (Z) พบว่า ความตรงตามสภาพของ

แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา สูงกว่า แบบทดสอบตามวิธีการให้คะแนนแบบพิจารณาความมั่นใจ สอดคล้องกับงานวิจัยของ รัตนา ไชยตรี (2546) ศึกษาความตรงตามสภาพตามวิธีประยุกต์การให้คะแนนของคูมบ์ส วิธีการให้คะแนนเดรสเซลและชมิดท์ และวิธีตอบโดยบอกระดับความมั่นใจ แล้วนำมาเปรียบเทียบความแตกต่างโดยใช้คะแนนฟิชเชอร์ซี เปรียบเทียบในภาพรวมโดยใช้สถิติทดสอบไคว์สแควร์ พบว่า ค่าความตรงตามสภาพขึ้นอยู่กับการตรวจให้คะแนนในแต่ละวิธี

2. เปรียบเทียบคุณภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจในด้านดังต่อไปนี้ ฟังก์ชันสารสนเทศ ความตรงตามสภาพ

จากผลการเปรียบเทียบฟังก์ชันสารสนเทศแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจจำแนกตามระดับความสามารถของผู้สอบ พบว่า แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา มีค่าฟังก์ชันสารสนเทศแบบทดสอบสูงเมื่อผู้สอบมีความสามารถระดับต่ำถึงปานกลาง (θ = -2.8 ถึง 0.0) และแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ มีค่าฟังก์ชันสารสนเทศแบบทดสอบสูงเมื่อผู้สอบมีความสามารถระดับค่อนข้างสูงถึงสูงมาก (θ = 0.8 ถึง 2.8) เมื่อใช้วิธีการตรวจให้คะแนนความรู้บางส่วนจะมีค่าฟังก์ชันสารสนเทศแบบทดสอบสูงกว่าเมื่อใช้วิธีการตรวจให้คะแนนความรู้สมบูรณ์ จึงทำให้วิธีการตรวจให้คะแนนความรู้บางส่วนเหมาะสมกว่าวิธีการตรวจให้คะแนนความรู้สมบูรณ์ที่ระดับความสามารถปานกลางถึงสูง และผู้สอบที่มีระดับความสามารถค่อนข้างต่ำมีโอกาสได้คะแนนจากวิธีการตรวจให้คะแนนความรู้สมบูรณ์ จึงมีความเหมาะสมกว่าการตรวจให้ความรู้บางส่วน สอดคล้องกับ เอมอร จังพรศิริพรปกรณ์ (2545, หน้า ง) ได้ศึกษาเปรียบเทียบคุณภาพของแบบทดสอบในด้านฟังก์ชันสารสนเทศของข้อสอบและแบบสอบ โดยใช้ทฤษฎีการตอบสนองข้อสอบเมื่อตรวจให้คะแนนแบบประเพณีนิยมกับวิธีให้คะแนนความรู้บางส่วน พบว่า ค่าฟังก์ชันสารสนเทศของข้อสอบและแบบสอบและอัตราส่วนสารสนเทศเฉลี่ย วิธีประยุกต์การให้คะแนนของคูมบ์สสูงสุด รองลงมาคือ วิธีประยุกต์การให้คะแนนของเดรสเซลและชมิดท์ และแบบประเพณีนิยมตามลำดับ ซึ่งไม่สอดคล้องกับ ธนวัฒน์ แสนสุข (2539) ศึกษาการใช้จีอาร์เอ็ม จีพีซีเอ็ม และโมเดลโลจิสติกในการเปรียบเทียบฟังก์ชันสารสนเทศของแบบวัดที่มีการให้คะแนนต่างกัน พบว่า การตรวจให้คะแนนแบบพหุวิภาคที่วิเคราะห์ตาม GRM ให้ค่าสารสนเทศเฉลี่ยสูงกว่าการตรวจให้คะแนนแบบทวิวิภาคเมื่อวิเคราะห์ตามโมเดลโลจิสติก 1, 2 และ3 พารามิเตอร์ แต่การตรวจให้คะแนนแบบพหุวิภาคที่วิเคราะห์ตามGPCM พบว่า ให้ค่าสารสนเทศเฉลี่ยต่ำกว่าการตรวจให้คะแนนแบบทวิวิภาคที่วิเคราะห์ตามโมเดลโลจิสติก 1, 2 และ3 พารามิเตอร์

จากผลการเปรียบเทียบค่าประสิทธิภาพสัมพัทธ์ RE (θ) ระหว่างค่าฟังก์ชันสารสนเทศระหว่างแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ พบว่า แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดามีค่าประสิทธิภาพสัมพัทธ์ RE (θ) สูงกว่าแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ ที่มีความสามารถระดับต่ำถึงปานกลาง (θ = -2.8 ถึง 0.0) และแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจมีค่าประสิทธิภาพสัมพัทธ์ RE (θ) สูงกว่า แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา ที่มีความสามารถระดับปานกลางถึงสูง (θ = 0.4 ถึง 2.8) ซึ่งเป็นไปตามสมมติฐานที่ตั้งไว้ สอดคล้องกับ ศิริชัย กาญจนวาสี (2548) ได้มีการเปรียบเทียบคุณภาพแบบสอบเลือกตอบการตรวจให้คะแนนแบบประเพณีนิยม และให้คะแนนความรู้บางส่วน พบว่า วิธีการตรวจให้คะแนนความรู้บางส่วนมีค่าฟังก์ชันสารสนเทศเฉลี่ยของข้อสอบและค่าฟังก์ชันสารสนเทศของแบบสอบรวมสูงสุด โดยมีประสิทธิภาพสัมพัทธ์สูงกว่าแบบสอบเลือกตอบรูปแบบอื่น 1.6 เท่า ซึ่งแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาเหมาะสมที่จะนำไปใช้กับผู้สอบที่มีความสามารถระดับ ต่ำ ถึงปานกลาง และแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ เหมาะสมที่จะนำไปใช้ผู้สอบที่มีความสามารถปานกลางถึงสูง สอดคล้องกับ ศิริชัย กาญจนวาสี (2545) กล่าวว่า ค่าสารสนเทศของแบบสอบมีค่าสูงในช่วงความสามารถใดก็จะมีความถูกต้องแม่นยำสูงในการประมาณค่าความสามารถของผู้ตอบในช่วงความสามารถนั้น ๆ เช่นเดียวกับ (Pajares & Miller, 1994 อ้างถึงใน นริศรา อุปกูล, 2539) การศึกษาของ ปาจาเรส และมิลเลอร์ ที่ทำการศึกษาบทบาทของความเชื่อในประสิทธิภาพของตนเอง (วัดได้จากความมั่นใจในการแก้ปัญหาทางคณิตศาสตร์) และความเชื่อเกี่ยวกับอัตมโนทัศน์ ต่อความสามารถในการแก้ปัญหาทางคณิตศาสตร์ ผลการศึกษาพบว่า ความมั่นใจเป็นตัวแปรที่ส่งผลกระทบทั้งทางตรงและทางอ้อมต่อความสามารถในการแก้ปัญหาทางคณิตศาสตร์ นอกจากนี้ ปาจาเรส และมิลเลอร์ ยังได้กล่าวว่า การที่บุคคลมีความมั่นใจในสิ่งใดสิ่งหนึ่งนั้น จะเป็นแรงผลักดันให้บุคคลนั้นแสดงพฤติกรรมออกมาเพื่อที่จะกระทำการใด ๆ ให้ไปสู้ความสำเร็จตามที่ตั้งเป้าหมายเอาไว้ ดังนั้น ความมั่นใจจึงเป็นตัวแปรที่มีผลกระทบต่อการกระทำใด ๆ รวมทั้งการตอบแบบทดสอบแบบเลือกตอบด้วย ซึ่งไม่สอดคล้องกับงานวิจัยของ ปรมินทร์ อริเดช (2539) ศึกษาการใช้จีอาร์เอ็ม จีพีซีเอ็ม และโมเดลโลจิสติกในการเปรียบเทียบฟังก์ชันสารสนเทศของมาตรประมาณค่าที่มีการให้คะแนนแบบทวิวิภาคและแบบพหุวิภาค พบว่า การตรวจให้คะแนนแบบพหุวิภาคที่วิเคราะห์ตาม GPCM ให้ค่าฟังก์ชันสารสนเทศเฉลี่ยต่ำกว่าการตรวจให้คะแบบแบบทวิวิภาคที่วิเคราะห์ตามโมเดลโลจิสติก 3, 2 และ 1 พารามิเตอร์

จากการเปรียบเทียบความตรงตามสภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ พบว่า รูปแบบการตอบและตรวจให้คะแนนมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และความตรงตามสภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจแตกต่างกัน ขึ้นอยู่กับวิธีการให้คะแนนในแต่ละวิธี โดยความตรงตามสภาพของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา สูงกว่า แบบทดสอบตามวิธีการให้คะแนนแบบพิจารณาความมั่นใจ การตรวจให้คะแนนความรู้สมบูรณ์เป็นวิธีเปิดโอกาสให้ผู้สอบได้คะแนนบางส่วน คะแนนที่ได้จะเป็นแบบ 0-1 ทำให้ได้ค่าความตรงตามสภาพสูงกว่าการตรวจให้คะแนนความรู้สมบูรณ์ ซึ่งเป็นไปตามสมมติฐานที่ตั้งไว้ สอดคล้องกับงานวิจัยของ Kansup and Hakstian (1975) พบว่า วิธีการให้น้ำหนักคะแนนรายข้อเท่ากันกับวิธีการให้น้ำหนักคะแนนรายข้อต่างกันตามระดับความมั่นใจในการตอบให้ค่าความตรงของแบบสอบต่างกัน อย่างมีนัยสำคัญ และสอดคล้องกับ ฉัตรศิริ ปิยะพิมลสิทธิ์ (2540, บทคัดย่อ) ได้ศึกษาเปรียบเทียบความเชื่อมั่นและความเที่ยงตรงของแบบทดสอบเลือกตอบความสามารถในการอ่านภาษาไทย ซึ่งรูปแบบต่างกัน 3 แบบ พบว่า ความเที่ยงตรงตามสภาพของแบบทดสอบเลือกตอบแบบตัวเลือกธรรมดา แบบตัวเลือกซ้อน แบบตัวเลือกถูกผิดที่ตรวจให้คะแนนด้วยวิธี 0-1 และตรวจให้คะแนนทุกตัวเลือก มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

และมีข้อค้นพบประการหนึ่งคือ จากการคำนวณความตรงตามสภาพ โดยใช้ระดับผลการเรียน ซึ่งเป็นเกณฑ์ที่ผู้สอนกำหนดให้ตามระดับความสามารถของผู้สอบ โดยมีแบบทดสอบเป็นส่วนหนึ่งที่ใช้ในการกำหนดระดับผลการเรียนเป็นเกณฑ์โดยแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา จึงมีความสัมพันธ์ค่อนข้างสูง โดยแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจมีแบบทดสอบหลายข้อที่ผู้สอบตอบผิดด้วยความมั่นใจที่สูง ซึ่งแปลว่าผู้สอบขาดความรู้ในเรื่องที่ตอบ ซึ่งสอดคล้อง นริศรา อุปกูล (2539) ได้ศึกษาองค์ประกอบเชิงสาเหตุด้านตัวนักเรียน แบบการคิด คุณภาพการสอน ที่มีผลต่อความมั่นใจในการตอบแบบสอบเลือกตอบและผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ ของนักเรียนชั้นประถมศึกษา ปีที่ 6 พบว่า ตัวแปรที่มีผลทางตรงอย่างเดียว ต่อความมั่นใจในการตอบแบบสอบเลือกตอบ ได้แก่ แบบการคิด แรงจูงใจใฝ่สัมฤทธิ์ และความรู้พื้นฐานเดิมทางคณิตศาสตร์ หมายความว่า ผู้ที่มีแบบการคิดแบบอิสระ มีแรงจูงใจใฝ่สัมฤทธิ์สูง และมีความรู้พื้นฐานเดิมทางคณิตศาสตร์ดี จะมีความมั่นใจในการการตอบแบบสอบเลือกตอบสูง ในทางกลับกัน ผู้ที่มีแบบการคิดแบบไม่อิสระ มีแรงจูงใจใฝ่สัมฤทธิ์ต่ำ และไม่มีความรู้พื้นฐานเดิมทางคณิตศาสตร์ จะมีความมั่นใจในการตอบแบบสอบแบบเลือกตอบต่ำ Jaradat and Tollefson (1988) ได้ศึกษาคุณภาพของแบบสอบโดยใช้ทฤษฎีการวัดแบบดั้งเดิม ซึ่งเปรียบเทียบค่าความตรง ค่าความเที่ยง และการให้เกรดระหว่างวิธีการให้คะแนน 4 วิธีของ

ข้อสอบแบบเลือกตอบ ได้แก่ วิธีประเพณีนิยม (FR) วิธีตัดตัวลวง (ET) วิธีเลือกชุดตัวถูก (SS) และวิธีการแก้การเดา (GC) โดยศึกษาในกลุ่มนักศึกษา 54 คน ในวิชาการวัดผลการศึกษา ข้อสอบเลือกตอบ มี 4 ตัวเลือก ใช้คะแนนสอบกลางภาค (MT) คะแนน จากวิธีประเพณีนิยม (FR) และคะแนนโครงงานในชั้นเรียนเป็นเกณฑ์ในการศึกษาความตรง ผลการวิจัย พบว่า ค่าความตรงและค่าความเที่ยงของ 4 วิธี ไม่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ส่วนการให้เกรด พบว่า วิธีตัดตัวลวงและวิธีเลือกชุดตัวถูก ทำให้ผู้สอบได้เกรด A และ B สูงกว่าวิธีแก้การเดา และสอดคล้องกับงานวิจัยของ รัตนา ไชยตรี (2546) ศึกษาความตรงตามสภาพตามวิธีประยุกต์การให้คะแนนของคูมบ์ส วิธีประยุกต์การให้คะแนนของเดรสเซลและชมิดท์ และวิธีการตอบโดยบอกระดับความมั่นใจ พบว่า ค่าความตรงตามสภาพขึ้นอยู่กับวิธีการตรวจให้คะแนนแต่ละวิธี และเมื่อนำมาเปรียบเทียบความแตกต่างกันเป็นรายคู่ด้วยสถิติทดสอบ (Z) พบว่า ความตรงตามสภาพของวิธีการประยุกต์การให้คะแนนของเดรสเซลและชมิดท์ สูงกว่าวิธีประยุกต์การให้คะแนนของคูมบ์สและวิธีการตอบโดยบอกระดับความมั่นใจอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

ข้อเสนอแนะ

ข้อเสนอแนะในการนำไปใช้

1. การใช้แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาเคมี แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาโดยวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (Item response theory) แบบทวิภาค ที่ใช้โมเดลโลจิสติก (Logistic) 2 พารามิเตอร์ จะให้ค่าประสิทธิภาพสัมพัทธ์สูงกว่า โดยเฉพาะผู้สอบที่มีระดับความสามารถต่ำ การตรวจให้คะแนนดังกล่าวเหมาะสมกว่า

2. การใช้แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาเคมี แบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจโดยวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (Item response theory) โมเดลการให้คะแนนความรู้บางส่วน (Genralized partial credit model: GPCM) จะให้ค่าประสิทธิภาพสัมพัทธ์สูงกว่า โดยเฉพาะผู้สอบที่ระดับความสามารถค่อนข้างสูง การตรวจให้คะแนนดังกล่าวเหมาะสมกว่า

3. แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาเหมาะสมที่จะนำไปใช้กับผู้สอบที่ระดับอ่อนถึงปานกลาง และแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจเหมาะสมที่จะนำไปใช้กับผู้สอบที่ระดับค่อนข้างเก่งถึงเก่งมาก

4. ควรตรวจสอบระดับความสามารถผู้สอบก่อน เพื่อเลือกวิธีการตอบและให้คะแนนเหมาะสมแก่ระดับความสามารถผู้สอบ

ข้อเสนอแนะในการวิจัย

1. เพื่อเป็นการเพิ่มประสิทธิภาพในการจัดการเรียนการสอน ครูผู้สอนควรนำผลการทดสอบไปใช้ในแนวทางในการวางแผนการจัดการเรียนการสอนที่สอดคล้องกับระดับความสามารถของผู้เรียน

2. ควรทำการศึกษาเพื่อตรวจสอบคุณภาพของวิธีการให้คะแนนความรู้บางส่วนในวิชาอื่น ๆ เช่น วิชาภาษาไทย วิชาภาษาอังกฤษ วิชาสังคมศึกษา เป็นต้น เพราะบริบทของแต่ละวิชาไม่เหมือนกัน ซึ่งวิธีการให้คะแนนความรู้บางส่วนแต่ละวิธีอาจจะเหมาะสมกับแต่ละวิชาแตกต่างกัน

3. ควรมีการเปรียบเทียบแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจกับแบบทดสอบวิธีอื่น ๆ เช่น แบบทดสอบหลายตัวเลือก, วิธีประยุกต์การให้คะแนนของคูมบ์สและวิธีประยุกต์การให้คะแนนของเดรสเซลและชมิดท์

4. ควรเว้นระยะห่างในการทำแบบทดสอบมากกว่า 7 วัน เพื่อลดปัญหาในการที่ผู้สอบสามารถคาดเดาแบบทดสอบได้

บรรณานุกรม

กนกวรรณ รัตนธน. (2544). *การเปรียบเทียบความตรงตามสภาพและความคงที่ในการประมาณค่าความสามารถของผู้สอบในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์เมื่อระดับความสามารถของผู้สอบและอัตราการใช้ข้อสอบซ้ำแตกต่างกัน*. วิทยานิพนธ์ครุศาสตรมหาบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา, คณะครุศาสตร์, จุฬาลงกรณ์มหาวิทยาลัย.

กาญจนา ศิริวัฒนาพงษ์. (2520). *การศึกษาเปรียบเทียบวิธีการตอบและการตรวจให้คะแนนแบบสอบปรนัยชนิดเลือกตอบที่มีลักษณะแตกต่างกัน*. ปริญญานิพนธ์การศึกษามหาบัณฑิต, สาขาวิชาการวัด ประเมินและวิจัยการศึกษา, บัณฑิตวิทยาลัย, มหาวิทยาลัยศรีนครินทรวิโรฒ.

เกริกชัย ฮวบเจริญ. (2525). *การเปรียบเทียบการตอบและการให้คะแนนแบบใหม่ แบบวิธีทดสอบความมั่นใจ และแบบธรรมดาที่มีผลต่อค่าความเชื่อมั่นและคะแนนการเดาของแบบทดสอบเลือกตอบ*. วิทยานิพนธ์ครุศาสตรมหาบัณฑิต, สาขาวิชาวิจัยการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.

จักรกฤษณ์ สำราญใจ. (2530). *ประสิทธิภาพเชิงสัมพัทธ์ของข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือก เทียบกับข้อสอบเลือกตอบ ตอบชนิดแบบฉบับในแบบสอบผลสัมฤทธิ์*. วิทยานิพนธ์ครุศาสตรดุษฎีบัณฑิต, สาขาวิชาวิจัยการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.

จินดา โตอนันต์. (2526). *การเทียบคุณลักษณะของแบบสอบสัมฤทธิผลแบบสอบเลือกตอบถูกผิดทุกตัวเลือก และแบบสอบเลือกตอบตัวเลือกเดีย*. วิทยานิพนธ์ครุศาสตรดุษฎีบัณฑิต, สาขาวิชาวิจัยการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.

ฉวีวรรณ บุญมั่ง. (2540). *การเปรียบเทียบความสัมพันธ์ระหว่างคะแนนความสามารถจากการวิเคราะห์โดยทฤษฎีการตอบสนองข้อสอบกับคะแนนที่ได้จากการตอบและตรวจให้คะแนน 4 วิธี*. วิทยานิพนธ์การศึกษามหาบัณฑิต, สาขาวิชาการวัดผลการศึกษา, บัณฑิตวิทยาลัย, มหาวิทยาลัยมหาสารคาม.

ฉัตรศิริ ปิยะพิมลสิทธิ์. (2540). *การศึกษาเปรียบเทียบการตอบและการตรวจให้คะแนนแบบใหม่และวิธีทดสอบความมั่นใจและแบบธรรมดาที่มีต่อค่าความเชื่อมั่นและคะแนนการเดาของแบบทดสอบเลือกตอบ*. ปริญญานิพนธ์การศึกษามหาบัณฑิต, สาขาวิชาการวัดประเมินและวิจัยการศึกษา, บัณฑิตวิทยาลัย, มหาวิทยาลัยศรีนครินทรวิโรฒ.

ชวาล แพรัตกุล. (2518). *เทคนิคการวัดผล* (พิมพ์ครั้งที่ 6). กรุงเทพฯ: วัฒนาพานิช.

ชวาล แพรัตกุล. (2520). *เทคนิคการเขียนข้อสอบ*. กรุงเทพฯ: พิทักษ์อักษร.

ณัฐภรณ์ เลขะวัฒนพงษ์. (2560). *การเปรียบเทียบคุณภาพของวิธีการตรวจให้คะแนนความรู้บางส่วน: การพัฒนาวิธีประยุกต์การให้คะแนนแบบตัดตัวลวง-เลือกตัวถูก.* วิทยานิพนธ์ครุศาสตรมหาบัณฑิต, สาขาวิชาวิจัยการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.

ณัฏฐภรณ์ หลาวทอง. (2551). *การวัดและประเมินผลการศึกษา.* กรุงเทพฯ: คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

ดำรง ศริรเจริญ. (2529). *การวัดผลแบบอิงเกณฑ์*. กรุงเทพฯ: ภาควิชาพื้นฐานการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ

ตรึงใจ พูลผลอำนวย. (2534). *การพัฒนาแบบสอบเฉพาะบุคคลในวิชาวิทยาศาสตร์ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3*. วิทยานิพนธ์ครุศาสตรมหาบัณฑิต, สาขาวิชาวิจัยการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.

ทวี ทองคำ. (2526). *การเปรียบเทียบค่าความเที่ยง ความตรงและอำนาจจำแนกของแบบสอบชนิดเลือกตอบที่ใช้คำสั่งและวิธีการให้คะแนนที่ต่างกัน*. วิทยานิพนธ์ครุศาสตรมหาบัณฑิต, สาขาวิชาวิจัยการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.

ธนวัฒน์ แสนสุข. (2539). การใช้จีอาร์เอ็ม จีพีซีเอ็ม และโมเดลโลจิสติกในการเปรียบเทียบฟังก์ชั่นสารสนเทศของแบบวัดที่มีวิธีการให้คะแนนต่างกัน. *วารสารวิธีวิทยาการวิจัย, 8*(1), 82-94.

นริศรา อุปกูล. (2539). *องค์ประกอบเชิงสาเหตุด้านตัวนักเรียนแบบการคิด คุณภาพการสอนที่มีผลต่อความมั่นใจในการตอบแบบสอบแบบเลือกตอบ และผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ ของนักเรียนชั้นประถมศึกษา ปีที่ 6.* วิทยานิพนธ์ครุศาสตรมหาบัณฑิต, สาขาวิชาวิจัยการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.

นิตยา พลดงนอก. (2540). *การพัฒนาแบบวัดความเข้าใจในการอ่านภาษาไทยของนักเรียนชั้นประถมศึกษาปีที่ 6*. วิทยานิพนธ์ครุศาสตรมหาบัณฑิต, สาขาวิชาวิจัยการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.

บุญเชิด ภิญโญอนันตพงษ์. (2527). *การทดสอบแบบอิงเกณฑ์: แนวคิดและวิธีการ*. กรุงเทพฯ: โอเดียนสโตร์.

บุญเชิด ภิญโญอนันตพงษ์. (2544). การให้คะแนนความรู้บางส่วนทำได้อย่างไร. *วารสารการวัดผลการศึกษา, 23*(67), 21-31.

บุญเชิด ภิญโญอนันตพงษ์. (2545). “คุณภาพของเครื่องมือวัด” หน่วยที่ 3. ใน *ประมวลสาระชุดวิชาการพัฒนาเครื่องมือสำหรับการประเมินการศึกษา*. นนทบุรี: สำนักพิมพ์มหาวิทยาลัยสุโขทัยธรรมาธิราช.

เบญจพร ยนตจักวิถี. (2539). *การเปรียบเทียบผลการวิเคราะห์ข้อสอบวัดผลสัมฤทธิ์วิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 ระหว่างทฤษฎีการทดสอบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบ*. วิทยานิพนธ์ศึกษาศาสตรมหาบัณฑิต, สาขาวิชาวิจัยการศึกษา, บัณฑิตวิทยาลัย, มหาวิทยาลัยสงขลานครินทร์.

ปรมินทร์ อริเดช. (2547). *การเปรียบเทียบฟังก์ชันสารสนเทศในการใช้ โมเดลโลจีสติก จีอาร์เอ็ม และจีพีซีเอ็มของมาตรวัดเจตคติแบบลิเคิร์ตและมาตรวัดแบบตัวเลือกบังคับตอบที่มีการให้คะแนนแบบสองค่าและแบบหลายค่า*. ปริญญานิพนธ์การศึกษาดุษฎีบัณฑิต, สาขาวิชาการทดสอบและวัดผลการศึกษา, บัณฑิตวิทยาลัย, มหาวิทยาลัยศรีนครินทรวิโรฒ.

ปราณี ร่วมทอง. (2528). *การเทียบคุณลักษณะของแบบสอบสัมฤทธิ์ผลแบบสอบเลือกตอบถูกผิดทุกตัวเลือก และแบบสอบเลือกตอบตัวเลือกเดียว*. วิทยานิพนธ์ครุศาสตรมหาบัณฑิต, สาขาวิชาวิจัยการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.

พัชรี มีวรรณ์. (2540). *ผลของจำนวนตัวเลือกที่มีต่อค่าฟังก์ชันสารสนเทศของแบบสอบเลือกตอบแบบถูกผิดในระดับการศึกษาที่ต่างกัน*. วิทยานิพนธ์ศึกษาศาสตรมหาบัณฑิต, สาขาวิชาวิจัยการศึกษา, บัณฑิตวิทยาลัย, มหาวิทยาลัยสงขลานครินทร์.

พิชิต ฤทธิ์จรูญ. (2544). *แนวทางการวิจัยเพื่อพัฒนาการเรียนรู้ บทบาทครูกับการวิจัยในชั้นเรียน*. กรุงเทพฯ: พริกหวาน กราฟิค.

พินิจ อุไรรักษ์. (2533). *ผลของวิธีการให้คะแนนที่มีต่อคะแนนสอบค่าสัมประสิทธิ์ ความเที่ยงและค่าสัมประสิทธิ์ความเที่ยง*. วิทยานิพนธ์ครุศาสตรมหาบัณฑิต, สาขาวิชาการวัดและการประเมินผลการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.

พิมพ์อุษา เจริญยศ. (2538). *การเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบเลือกตอบเมื่อลดจำนวนข้อโดยใช้เกณฑ์ค่าสถิติกับใช้เกณฑ์ดุลพินิจของครู*. ปริญญานิพนธ์การศึกษามหาบัณฑิต, สาขาวิชาการวัดผลการศึกษา, บัณฑิตวิทยาลัย, มหาวิทยาลัยศรีนครินทรวิโรฒ.

เพ็ญศรี สว่างเนตร. (2520). *ความเที่ยงของแบบสอบชนิดเลือกตอบด้วยเทคนิคการให้คะแนนที่ต่างกัน*. วิทยานิพนธ์ครุศาสตรมหาบัณฑิต, สาขาวิชาวิจัยการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.

รัตนา ไชยตรี. (2546). *การเปรียบเทียบคุณภาพและอัตราความคลาดเคลื่อนของการกำหนดเกรดแบบอิงกลุ่ม และอิงเกณฑ์ เมื่อใช้แบบสอบเลือกตอบที่มีการตอบ และการตรวจให้คะแนนแบบความรู้บางส่วนด้วยวิธีแตกต่างกัน.* วิทยานิพนธ์ครุศาสตรมหาบัณฑิต, สาขาวิชาวิจัยและจิตวิทยาการศึกษา, คณะครุศาสตร์, จุฬาลงกรณ์มหาวิทยาลัย.

รุจิรา ขาวสะอาด. (2543). *ผลของการให้น้ำหนักตามระดับความมั่นใจที่มีต่อความเที่ยง และค่าความตรงของแบบสอบประเภทเลือกตอบที่มีรูปแบบการเขียนตัวเลือกต่างกัน*. วิทยานิพนธ์ครุศาสตรมหาบัณฑิต, สาขาวิชาวิจัยการศึกษา, คณะครุศาสตร์, จุฬาลงกรณ์มหาวิทยาลัย.

ล้วน สายยศ และอังคณา สายยศ. (2538). *เทคนิคการวิจัยทางการศึกษา* (พิมพ์ครั้งที่ 4). กรุงเทพฯ: สุวีริยาสาส์น.

ล้วน สายยศ และอังคณา สายยศ. (2539). *เทคนิคการวัดผลการเรียนรู้*. กรุงเทพฯ: ชมรมเด็ก.

ล้วน สายยศ และอังคณา สายยศ. (2540). *สถิติวิทยาการวิจัย* (พิมพ์ครั้งที่ 3). กรุงเทพฯ: สุวีริยสาส์น.

ล้วน สายยศ และอังคณา สายยศ. (2543). *การวัดด้านจิตพิสัย*. กรุงเทพฯ: สุวีริยาสาส์น.

ศราวุธ รัตนะ. (2562). *การสร้างแบบทดสอบวินิจฉัยกระบวนการให้เหตุผลทางคณิตศาสตร์สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 4 จังหวัดระยอง*. วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต, สาขาวิชาวิจัย วัดผลและสถิติการศึกษา, คณะศึกษาศาสตร์, มหาวิทยาลัยบูรพา.

ศิริชัย กาญจนวาสี. (2545). *ทฤษฎีการทดสอบแนวใหม่*. กรุงเทพฯ: คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนวาสี. (2548). การเปรียบเทียบคุณภาพของแบบสอบเลือกตอบที่มีจำนวนตัวถูกวิธีการตอบและวิธีการตรวจให้คะแนนที่ต่างกัน. *วารสารวิธีวิทยาการวิจัย, 18*(2), 181-192.

ศิริชัย กาญจนวาสี. (2550). *ทฤษฎีการทดสอบแนวใหม่* (พิมพ์ครั้งที่ 3). กรุงเทพฯ: คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนวาสี. (2556). *ทฤษฎีการทดสอบแบบดั้งเดิม*. กรุงเทพฯ: คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

สมนึก ภัททิยธนี. (2553). *การวัดผลการศึกษา* (พิมพ์ครั้งที่ 7). กาฬสินธุ์: ประสานการพิมพ์.

สำนักวิชาการและมาตรฐานการศึกษา. (2553). *แนวปฏิบัติการวัดและการประเมินผลการเรียนรู้ตามหลักสูตรแกนกลาง การศึกษาขั้นพื้นฐาน พุทธศักราช 2551*. กรุงเทพฯ: สหกรณ์การเกษตรแห่งประเทศไทย จำกัด.

สำราญ มีแจ้ง. (2525). *ผลของการใช้คำสั่งและการให้คะแนนที่ต่างกันต่อค่าความเที่ยง ความตรงและอำนาจจำแนกของแบบทดสอบชนิดเลือกตอบ*. วิทยานิพนธ์ครุศาสตรมหาบัณฑิต, สาขาวิชาวิจัยการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.

สำเริง บุญเรืองรัตน์. (2529). การพัฒนาทฤษฎีลาเท้นเทรทเพื่อการวิเคราะห์ข้อสอบ. *วารสารการวัดผลการศึกษา, 7*(21), 41-66.

สุพจน์ เกิดสุวรรณ. (2545). *การพัฒนาการวัดความรู้บางส่วนของผู้ตอบแบบสอบเลือกตอบ*. วิทยานิพนธ์ครุศาสตรดุษฎีบัณฑิต, สาขาวิชาวิจัยการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.

สุพัฒน์ สุกมลสันต์. (2530). นานาปัจจัยที่ส่งผลต่อค่าความเชื่อมั่นและค่าความเที่ยงตรงของแบบทดสอบ. *วิธีวิทยาการวิจัย, 2*(3), 52-62.

สุพัฒน์ สุกมลสันต์. (2539). *ธนาคารข้อทดสอบและการทดสอบปรับเปลี่ยนด้วยคอมพิวเตอร์*. กรุงเทพฯ: พิมพ์ดี.

สุรีพร อนุศาสนนันท์. (2550). *การเปรียบเทียบคุณภาพของการกำหนดมาตรฐานระหว่างวิธีแองกอฟที่ได้รับการปรับปรุงกับวิธีบุ๊คมาร์ค.* วิทยานิพนธ์ครุศาสตรดุษฎีบัณฑิต, สาขาวิชาวิจัยการศึกษา, คณะครุศาสตร์, จุฬาลงกรณ์มหาวิทยาลัย.

สุรีพร อนุศาสนนันท์. (2558). *การวัดและประเมินผลในชั้นเรียน.* ชลบุรี: ภาควิชาวิจัยและจิตวิทยาประยุกต์ คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา.

อรวรรณ ตัณฑ์เจริญรัตน์. (2517). *การศึกษาวิธีการตอบและการตรวจให้คะแนนแก่ข้อสอบปรนัยแบบเลือกตอบ*. ปริญญานิพนธ์การศึกษามหาบัณฑิต, สาขาวิชาการวัดผลการศึกษา, บัณฑิตวิทยาลัย, มหาวิทยาลัยศรีนครินทรวิโรฒ.

อวยพร เรืองตระกูล. (2544). *การพัฒนาและวิเคราะห์คุณภาพของวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบ*. วิทยานิพนธ์ครุศาสตรดุษฎีบัณฑิต, สาขาวิชาวิจัยการศึกษา, คณะครุศาสตร์, จุฬาลงกรณ์มหาวิทยาลัย.

อนันต์ ศรีโสภา. (2532). *การทดลองรูปแบบใหม่ของข้อสอบแบบเลือกตอบ*. กรุงเทพฯ: สำนักพิมพ์มหาวิทยาลัยศรีนครินทรวิโรฒ.

อุทัยวรรณ สายพัฒนะ. (2547). *การเปรียบเทียบประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่าระหว่างวิธี GMH และวิธี Polytomous SIBTEST.* ปริญญานิพนธ์การศึกษาดุษฎีบัณฑิต, สาขาวิชาการทดสอบและวัดผลการศึกษา, บัณฑิตวิทยาลัย, มหาวิทยาลัยศรีนครินทรวิโรฒ.

เอมอร จังศิริพรปกรณ์. (2544). การให้คะแนนความรู้บางส่วนทำได้อย่างไร. *วารสารการวัดผลการศึกษา, 23*(67), 21-31.

เอมอร จังศิริพรปกรณ์. (2545). *การเปรียบเทียบคุณภาพของแบบสอบเลือกตอบ เมื่อตรวจด้วยวิธีการให้คะแนนความรู้บางส่วนกับวิธีประเพณีนิยม.* กรุงเทพฯ: กองทุนรัชดาภิเษกสมโภช จุฬาลงกรณ์มหาวิทยาลัย.

เอมอร จังศิริพรปกรณ์. (2546). *การเปรียบเทียบคุณภาพของแบบเลือกตอบเมื่อตรวจด้วยวิธีการให้คะแนนความรู้บางส่วนกับวิธีประเพณีนิยม*. กรุงเทพฯ: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

เอมอร จังศิริพรปกรณ์. (2548). *การเปรียบเทียบคุณภาพของแบบสอบระหว่างแบบสอบเลือกตอบที่มีจำนวนตัวถูกตัวเดียวกับตัวถูกมากกว่า 1 ตัว เมื่อตรวจด้วยวิธีการให้คะแนนความรู้บางส่วน*. กรุงเทพฯ: กองทุนรัชดาภิเษกสมโภช จุฬาลงกรณ์มหาวิทยาลัย.

Abu-Sayf, F. K. (1975). Relative effectiveness of the conventional formula score. *Journal of Educational Reasearch, 69*, 160-162.

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan Publishing Company.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*, 561-573.

Angoff, W. H. (1953). Test reliability and effective test length. *Psychometrika*, *18*, 1-14.

Arnold, J. C., & Arnold, P. L. (1970). On scoring multiple-choice exams allowing for partial knowledge. *The Journal of Experimental Education, 39*, 8-13.

Berk, R. A. (1980). *A guide to criterion-referenced test construction.* London: The John Hopkins University Press.

Behrens, J. (1997). *Using Perl, Winsteps, Bilog, SAS, and DataDesk to visualize item characteristic curves*. Retrieved from http://www.creative-wisdom.com/computer/sas/IRT.html

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397-479). Reading, MA: Addison-Wesley.

Bloom, B. S. (1967). *Evaluation of learning in secondary school.* New York: McGraw-Hill.

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika, 37*, 29-51.

Brandon, P. R. (2004). Conclusions about frequently studied modified Angoff standard setting topics. *Applied Measurement in Education, 17*, 59-88.

Brannick, M. T. (2007). *Item response theory*. Retrieved from http://luna.cas.usf.edu/~mbrannic/files/pmet/irt.html

Brennan, P. A., Mednick, R. R., & Mednick, S. A. (1974). Parental psychopathology, congenital factor, and violence. In S. Hodgins (Ed.)., *Mental disorder and crime* (pp. 244-261). Thosand Oaks, CA: Sage.

Busch, M. I., & Jaeger, R. M. (1990). Influence of type of judge, normative information, and discussion on standards recommended for the national teacher examination. *Journal of Educational Measurement, 27*, 145-163.

Chohen, A. S., Kane, M. T., & Crooks, T. J. (1999). A generalized examinee-centered method for setting standards on achievement tests. *Applied Measurement in Education, 12*(4), 343-366.

Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. London: Sage.

Coombs, C. H., Milholland, J. E., & Worner, F.B. (1956). The assessment of partial knowledge. *Educational and Psychological Measurement, 16*(1), 13-37.

Davis, I. (1966). Numic consonantal correspondences. *International Journal of American Linguistics, 32*, 124-140.

Davis, F. B., & Fifer, G. (1959). The effect on test reliability and validity of scoring aptitude and achievement tests with weights for every choice. *Educational and Psychological Measurement, 19,* 159-170.

De Finetti, B. (1965). Methods for discriminating levels of partial knowledge concerning a test item. *British Journal of Mathematical and Statistical Psychology, 18*, 87-123.

De Ayala, R. J., Dodd, B. G., & Koch, E. R. (1992). A comparison of the partial credit and graded response models in computerized adaptive testing. *Applied Measurement in Education, 5*, 17-34.

Dodd, B. G., & Koch, W. R. (1989). Operational characteristics of adaptive testing procedures using the graded response model. *Applied Psychological Measurement, 13*(2), 129-143.

Dressel, P. L., & Schmidt, J. (1953). Some modification of the multiple-choice item. *Education and Psychological Measurement*, *13*, 574-595.

Ebel, R. L. (1965). *Measuring educational achievement.* Englewood Cliffs, NJ: Prentice-Hall.

Ebel, R. L. (1972). *Essential of education measurement* (2nd ed.). Englewood Cliffs, NJ: Principle-Hall.

Ebel, R. L., & Fisbie, D. A. (1986). *Essentials of educational measurement.* Englewood Cliffs, NJ: Principle-Hall.

Echternacth, R. M. (1972). A comparison of five different scoring functions for confidence test. *Journal of Education Measurement, 7*, 217-221.

Fehrmann, M. L., Woehr, D. J., & Arthur, W. (1991). An empirical comparison of cutoff score methods for content-related and criterion-related validity settings. *Educational and Psychological Measurement, 51*, 1029-1039.

Field, A. (2000). *Discovering statistics using SPSS for Windows*. London. Thousand Oaks, New Delhi: Sage.

Frary, R.B. (1989). Partial credit scoring methods for multiple-choice tests. *Applied Measurement in Education, 2*, 79-96.

Gibbons, J. D., Olkin, I., & Sobel, M. (1979). A subset selection technique for scoring items on a multiple choice test. *Psychometrika, 44*, 259-270.

Glass, G. V. (1978). Standards and criteria. *Journal of Educational Measurement, 15*, 237-261.

Glass, G. V., & Stanley, J. C. (1970). *Statistical methods in education and psychology.* Englewood Cliffs, NJ: Printice-Hall

Gritten, F., & Johnson, D. M. (1941). Individual differences in judging multiple choice question. *Journal of Education Psychology, 32*, 423-430.

Gronlund, N. E. (1976). *Measurement and evaluation in teaching*. New York: Macmillan.

Gronlund, N. E. (1981). *Measurement and evaluation in teaching* (4th ed.). New York: Collier Macmillan.

Gronlund, N. E., & Linn, R. L. (1990). *Measurement and evaluation in teaching* (6th ed.). New York: Macmillan.

Gulliksen, H. (1986). Perspective on educational measurement. *Applied Psychological Measurement, 10*, 109-132.

Hambleton, R. K., & Plake, B. S. (1995). Using *an* extended Angoff procedure *to* set standards *on* complex performance assessments. *Applied Measurement in Education, 8*(1), 41-55.

Hambleton, R. K., Robert, D. M., & Taub, 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.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications.* Boston, MA: Kluwer Academic.

Hambleton, R. K., Swaminathan, H., & Roger, H. J. (1991). *Fundamentals of item response theory*. Newbury Park: Sage.

Harris, C. W. (1972). An interpretation of Livingston's reliability coefficient for criterion-referenced tests. *Journal* of *Educational Measurement*, *9*, 27-29.

Hess, N. K., Singer, P. A., Trinh, K., Nikkhoy, M., & Bernstein, S. I. (2007). Transcriptional regulation of the Drosophila melanogaster muscle myosin heavy-chain gene. *Gene Expr. Patterns, 7*, 413-422.

Hoyt, C. (1941). Test reliability estimated by analysis of variance. *Psychometrika, 6*, 153-16.

Hurtz, G. M., & Auerbach, M. A. (2003). A meta-analysis of the effects of modifications to the Angoff method on cutoff scores and judgment consensus. *Educational and Psychological Measurement, 63*, 584-601.

Impara, J. C., & Plake, B. S. (1997). Standard-setting: An alternative approach. *Journal of Educational Measurement, 34*, 353-366.

Jaradat, D., & Tollefson, N. (1988). The impact of the scoring procedures for multiple-choice items on test reliability, validity and grading. *Educational and Psychological Measurement, 48*, 627-635.

Kane, M. (1994). Validating the performance standards associated with passing score. *Review of Educational Research, 64*(3), 425-461.

Kansup, W., & Hakstian, A. R. (1975). A comparison of several methods of assessing partial knowledge in multiple-choice test. *Journal of Educational Measurement, 12*, 219-230.

Koehler, R. A. (1971). A comparison of the validities of conventional choice testing and various confidence marking procedures. *Journal of Educational Measurement, 8*, 297-303.

Livingston, S. A. (1972). A criterion-referenced application of classical test theory. *Journal of Educational Measurement, 9*, 13-26.

Livingston, S., & Zieky, M. (1982). *Passing scores a manual for setting standards of performance on educational and occupational tests.* Princeton, NJ: Princeton: Educational Testing Services.

Lord, F. M. (1975). Formula scoring and number right scoring. *Journal of Educational Measurement, 12*, 7-11.

Lord, F. M. (1980). *Application of item response theory to practical testing problems.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Lovett, H. T. (1978). The effect of violating the assumption of edual item means in estimating the Livingston coefficient. *Educational and Psychological Measurement, 38*, 239-251.

Marion, S., & Pellegrino, J. (2006). A validity framework for evaluating the technical quality of alternate assessments. *Educational Measurement: Issues and Practice, 25*(4), 47-57.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149-174.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 14*, 59-71.

Muraki, E. (1993). Information function of generalized partial credit model. *Applied Psychological Measurement, 17*(4), 351-363.

Norcini, J. J. (2003). Setting standards on educational tests. *Medical Education, 37*(5), 464-469.

Pugh, R. C., & Brunza, J. J. (1975). Effects of a confidence-weighted scoring system on measures of test reliability and validity. *Educational and Psychological Measurement, 35*, 73-78.

Raju, N. B. (1982). Easy methods for fluorescent staining of Neurospora nuclei. *Neurospora Newsl, 29*, 24-26.

Samejima, F. (1996). Graded response model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 85-100). New York: Springer.

Smith, R. M. (1987). Assessing partial knowledge in vocabulary. *Journal of Educational Measurement, 24*, 217-231.

Stanley, J. C., & Wang, M. D. (1970). Weighting test items and test item options, an overview of the analytical and empirical literature. *Educational and Psychological Measurement*, *30*, 21-35.

Thissen, D. M. (1976). Information in wrong responses to the raven progressive matrices. *Journal of Educational Measurement,14*, 201-214.

Tollefson, N., & Chung, J. M. (1986). *A comparison of two methods of assessing partial knowledge on multiple choice.* Lawrence, KS: Kansas University Press.

Towles-Reeves, E. (2008). *Standard setting approaches for alternate assessment.* Retrieved from http://www.ccsso.org/content/PDFs/190_ Towles-Reeves.pdf

Wallsten, T. S., Budescu, D. V., & Zwick, R. (1993). Comparing the calibration and coherence of numerical and verbal probabilistice judgments. *Management Science*, *39*, 176-190.

Wang, M. D., & Stanley, J. C.(1970). Weighting test items and test-item options, an overview of the analytical and empirical literature. *Educational and Psychological Measurement, 30*(1), 21-35.

Waugh, K. C., & Gronlund, N. E. (2013). *Assessment of student achievement.* New York: Pearson Education.

Zieky, J. M. (2008). *Cut scores: A manual for setting standards of performance on educational and occupational test*. Princeton, NJ: Educational Testing Service.

Zieky, J. M., & Perie, M. (2006). *A primer on setting cutscores on tests of educational achievement*. Princeton, NJ: Educational Testing Service.

ภาคผนวก

ภาคผนวก ก

รายนามผู้เชี่ยวชาญตรวจสอบเครื่องมือในการวิจัย

รายชื่อผู้ทรงคุณวุฒิตรวจสอบเครื่องมือการวิจัย

1. ดร.ภัทรภร ชัยประเสริฐ	คณาจารย์ภาควิชาการจัดการเรียนรู้ คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา
2. นายธนายุต จันทราเขต	ครูชำนาญการพิเศษ โรงเรียนชลราษฎรอำรุง เชี่ยวชาญการสอนรายวิชา เคมี
3. นางพนา ชุมเกษียร	ครูชำนาญการพิเศษ โรงเรียนสิงห์สมุทร เชี่ยวชาญการสอนรายวิชา เคมี
4. นางปวีณา พันทอง	ครูชำนาญการพิเศษ โรงเรียนสิงห์สมุทร เชี่ยวชาญการสอนรายวิชา เคมี
5. นางสาวอรรถการณ์ ทองแดงเจือ	ครูโรงเรียนชัญบุรี เชี่ยวชาญการสอนรายวิชา เคมี

ภาคผนวก ข

คู่มือการใช้การกำหนดคะแนนจุดตัด ด้วยวิธีแองกอฟที่ได้รับการปรับปรุง

คู่มือการใช้
การกำหนดคะแนนจุดตัด ด้วยวิธีแองกอฟที่ได้รับการปรับปรุง

โดย
นางสาวณพานันท์ ยมจินดา
คณะศึกษาศาสตร์ สาขา วิจัย วัดผลและสถิติการศึกษา
หลักสูตร วิทยาศาสตรมหาบัณฑิต (วท.ม.)
อาจารย์ที่ปรึกษาวิทยานิพนธ์
อาจารย์ที่ปรึกษาหลัก ผู้ช่วยศาสตราจารย์ ดร.สุรีพร อนุศาสนนันท์
อาจารย์ที่ปรึกษาร่วม ผู้ช่วยศาสตราจารย์ ดร.สมพงษ์ ปั้นหุ่น

คำนิยามปฏิบัติการ

1. คะแนนจุดตัด (Cut score) หมายถึง จุดคะแนนที่ใช้เป็นจุดแบ่งนักเรียนในวิชาเคมี ชั้นมัธยมศึกษาปีที่ 4 ภาคเรียนที่ 2 ตามความสามารถของนักเรียน งานวิจัยนี้แบ่งคะแนนจุดตัดออกเป็น 4 จุด คือ 1) จุดคะแนนที่แบ่งความสามารถนักเรียนที่มีความสามารถอ่อน ออกจากนักเรียนระดับตก 2) จุดคะแนนที่แบ่งความสามารถนักเรียนที่มีความสามารถอ่อน ออกจากนักเรียนระดับพอใช้ 3) จุดคะแนนที่แบ่งความสามารถนักเรียนที่มีความสามารถพอใช้ ออกจากนักเรียนระดับดี 4) จุดคะแนนที่แบ่งความสามารถนักเรียนที่มีความสามารถดี ออกจากนักเรียนระดับดีมาก

2. วิธีการแองกอฟที่ได้รับการปรับปรุง (Modified Angoff method) หมายถึง วิธีการกำหนดคะแนนจุดตัด โดยให้ผู้ตัดสินประมาณระดับความสามารถขั้นต่ำของผู้สอบในข้อสอบแต่ละข้อจนครบทุกข้อ หลังจากนั้นผู้ตัดสินแต่ละคนร่วมกันอภิปรายผลการตัดสินของแต่ละคน เมื่อผู้ตัดสินอภิปรายเสร็จก็เปิดโอกาสให้ตัดสินคะแนนอีกครั้งหนึ่ง ผลการตัดสินที่ได้จากผู้ตัดสินแต่ละคนในรอบที่สองนำมาหาค่าเฉลี่ยซึ่งจะเป็นจุดตัดคะแนน

3. มาตรฐาน (Standard) หมายถึง ระดับที่บ่งบอกคุณภาพของนักเรียนการทำแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชา เคมี ชั้นมัธยมศึกษาปีที่ 4 ในการศึกษาครั้งนี้ได้แบ่งระดับของนักเรียนเป็น 5 ระดับ ได้แก่ 1) ระดับดีมาก (A) 2) ระดับดี (B) 3) ระดับพอใช้ (C) 4) ระดับอ่อน (D) 5) ระดับตก (F)

การเตรียมการก่อนการสัมมนา

1. จัดเตรียมรายชื่อผู้ตัดสินที่มีคุณสมบัติตามที่ต้องการ เช่น เป็นผู้มีประสบการณ์สอนในรายวิชาเคมี ไม่น้อยกว่า 2 ปี

2. จัดทำหนังสือเชิญคุณครูที่มีคุณสมบัติดังกล่าวมาเป็นผู้ตัดสิน

3. จัดเตรียมเอกสารการสัมมนาดังต่อไปนี้

3.1 คู่มือการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟที่ได้รับการปรับปรุง

3.2 ค่าความยากของข้อสอบ

3.3 แบบฟอร์มประมาณค่าเปอร์เซ็นต์การตอบข้อสอบแต่ละข้อถูกของผู้สอบที่มีความสามารถระดับต่าง ๆ เช่น ถ้าต้องการแบ่งเกรดออกเป็น 5 ระดับ คือ ระดับดีมาก (A) ระดับดี (B) ระดับพอใช้ (C) ระดับอ่อน (D) และระดับตก (F) ผู้ตัดสินก็ต้องทำการตัดสินคะแนนจุดตัด 4 ระดับ คือ ระดับดีมาก ระดับดี ระดับพอใช้ ระดับอ่อน ดังนั้นผู้ตัดสินแต่ละคนก็จะได้รับแบบฟอร์มประมาณค่าเปอร์เซ็นต์การตอบข้อสอบแต่ละข้อของผู้สอบที่มีความสามารถทั้ง 4 ระดับ คือผู้ตัดสินแต่ละคนจะได้รับแบบฟอร์มคนละ 4 ฉบับ

3.4 แบบฟอร์มเขียนคำบรรยายการปฏิบัติแต่ละระดับ

3.5 คำอธิบายลักษณะความสามารถในวิชานั้นของนักเรียนในแต่ละระดับ เช่น นักเรียนที่มีความสามารถระดับดีมาก หมายถึง นักเรียนที่มีความสามารถถึงระดับที่คาดหวังขั้นสูงสำหรับการปฏิบัติของนักเรียนในวิชา เคมี ชั้นมัธยมศึกษาปีที่ 4 ตามหลักสูตรกระทรวงศึกษาธิการ

3.6 ค่าความยากของข้อสอบแต่ละข้อ

3.7 แบบฟอร์มแสดงค่าเฉลี่ยผลการตัดสินของผู้ตัดสินแต่ละระดับ

ขั้นตอนการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟที่ได้รับการปรับปรุง

1. ผู้ตัดสิน 5 คน ร่วมกันอบรมเชิงปฏิบัติการเพื่อรับฟังวิธีการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟที่ได้รับการปรับปรุง

2. แจกคำอธิบายลักษณะความสามารถของนักเรียนในแต่ละระดับ เพื่อให้ผู้ตัดสินเข้าใจตรงกัน และแจกค่าความยากง่ายของข้อสอบแต่ละข้อ

3. ดำเนินการกำหนดเปอร์เซ็นต์การตอบข้อสอบแต่ละข้อถูก จากผู้สอบที่มีความสามารถแต่ละระดับ มี 3 ขั้นตอน ดังนี้

ขั้นที่ 1 รอบที่ 1 ดำเนินการตัดสินเป็นรายบุคคล ดังนี้

ผู้ตัดสินแต่ละคนพิจารณาข้อสอบทีละข้อตั้งแต่ข้อสอบข้อที่ 1 เพื่อกำหนดคะแนนจุดตัดโดยตัดสินว่า

1. กลุ่มผู้ที่มีความสามารถระดับดีมาก (A) มีโอกาสตอบข้อข้อนั้นถูกกี่เปอร์เซ็นต์ หมายความว่า ผู้ตัดสินคาดว่า นักเรียนที่มีความสามารถระดับดีมาก (A) ประมาณกี่คนจากทั้งหมด 100 คน สามารถตอบข้อสอบข้อนั้นถูก จากนั้นกรอกตัวเลขลงในแบบประมาณเปอร์เซ็นต์การตอบข้อสอบแต่ละข้อถูกที่มีความสามารถระดับดีมาก (A) ในช่องเปอร์เซ็นต์การตอบถูก รอบที่ 1 (รายบุคคล) จากนั้นพิจารณาข้อสอบข้อถัดไปเรื่อย ๆ จนครบจำนวนข้อสอบทั้งหมด

2. กลุ่มผู้ที่มีความสามารถระดับดี (B) มีโอกาสตอบข้อข้อนั้นถูกกี่เปอร์เซ็นต์ หมายความว่า ผู้ตัดสินคาดว่า นักเรียนที่มีความสามารถระดับดี (B) ประมาณกี่คนจากทั้งหมด 100 คน สามารถตอบข้อสอบข้อนั้นถูก จากนั้นกรอกตัวเลขลงในแบบประมาณเปอร์เซ็นต์การตอบข้อสอบแต่ละข้อถูกที่มีความสามารถระดับดี (B) ในช่องเปอร์เซ็นต์การตอบถูก รอบที่ 1 (รายบุคคล) จากนั้นพิจารณาข้อสอบข้อถัดไปเรื่อย ๆ จนครบจำนวนข้อสอบทั้งหมด

3. กลุ่มผู้ที่มีความสามารถระดับพอใช้ (C) มีโอกาสตอบข้อข้อนั้นถูกกี่เปอร์เซ็นต์ หมายความว่า ผู้ตัดสินคาดว่า นักเรียนที่มีความสามารถระดับพอใช้ (C) ประมาณกี่คนจากทั้งหมด 100 คน สามารถตอบข้อสอบข้อนั้นถูก จากนั้นกรอกตัวเลขลงในแบบประมาณเปอร์เซ็นต์การตอบข้อสอบแต่ละข้อถูกที่มีความสามารถระดับพอใช้ (C) ในช่องเปอร์เซ็นต์การตอบถูก รอบที่ 1 (รายบุคคล) จากนั้นพิจารณาข้อสอบข้อถัดไปเรื่อย ๆ จนครบจำนวนข้อสอบทั้งหมด

4. กลุ่มผู้ที่มีความสามารถระดับอ่อน (D) มีโอกาสตอบข้อข้อนั้นถูกกี่เปอร์เซ็นต์ หมายความว่า ผู้ตัดสินคาดว่า นักเรียนที่มีความสามารถระดับอ่อน (D) ประมาณกี่คนจากทั้งหมด 100 คน สามารถตอบข้อสอบข้อนั้นถูก จากนั้นกรอกตัวเลขลงในแบบประมาณเปอร์เซ็นต์การตอบข้อสอบแต่ละข้อถูกที่มีความสามารถระดับอ่อน (D) ในช่องเปอร์เซ็นต์การตอบถูก รอบที่ 1 (รายบุคคล) จากนั้นพิจารณาข้อสอบข้อถัดไปเรื่อย ๆ จนครบจำนวนข้อสอบทั้งหมด

ขั้นที่ 2 รอบที่ 2 ดำเนินการแบ่งผู้ตัดสินเป็นกลุ่มย่อย 2 กลุ่ม ดังนี้

1. ในแต่ละกลุ่มย่อย ผู้ตัดสินร่วมกันอภิปรายผลการตัดสินของละคน โดยเฉพาะผู้ตัดสินที่มีผลตัดสินต่ำสุด และสูงสุดให้อธิบายแสดงความคิดเห็นถึงเหตุผลที่ให้ผลการตัดสินต่ำสุด และสูงสุด และระหว่างการอภิปรายผู้อำนวยความสะดวกจะแสดงรูปการแจกแจงผลการตัดสินของผู้ตัดสินแต่ละคน

2. เมื่อผู้ตัดสินอภิปรายเสร็จ ก็เปิดโอกาสให้กำหนดคะแนนจุดตัดใหม่อีกครั้งหนึ่ง

3. นำผลการตัดสินของผู้ตัดสินทุกคนจากแต่ละกลุ่ม โดยผู้ตัดสินทุกคนจะมีคะแนนจุดตัด 4 ระดับ คือ 1) ระดับดีมาก (A) 2) ระดับดี (B) 3) ระดับพอใช้ (C) 4) ระดับอ่อน (D) นำคะแนนจุดตัดของผู้ตัดสินแต่ละคนมาหาค่าเฉลี่ยในแต่ละระดับ จะได้ค่าเฉลี่ยของคะแนนจุดตัด 4 จุด หรือมีมาตรฐาน 5 ระดับ

ขั้นที่ 3 รอบที่ 3 ดำเนินการรวมผู้ตัดสินทั้งสองกลุ่มย่อยเป็นกลุ่มใหญ่เพียงกลุ่มเดียว ดำเนินการดังนี้

1. นำผู้ตัดสินทั้งสองกลุ่มย่อยเป็นกลุ่มใหญ่เพียงกลุ่มเดียวกัน ผู้อำนวยความสะดวกแสดงค่าเฉลี่ยคะแนนจุดตัดของแต่ละกลุ่มให้ผู้ตัดสินพิจารณา เปิดโอกาสให้ผู้ตัดสินอภิปรายร่วมกัน
2. เปิดโอกาสให้ผู้ตัดสินทำการตัดสินคะแนนจุดตัดทั้ง 4 ระดับครั้งอย่างอิสระจากกัน
3. นำผลการตัดสินของผู้ตัดสินแต่ละคนมาหาค่าเฉลี่ยในแต่ละระดับ
4. เมื่อได้คะแนนจุดตัด 4 ระดับ ให้ผู้ตัดสินร่วมกันเขียนบรรยายระดับการปฏิบัติการของนักเรียนในประเด็นความสามารถของนักเรียน ณ ระดับคะแนนจุดตัดของระดับนั้น รวม 5 ระดับ

แบบประมาณเปอร์เซ็นต์การตอบข้อสอบแต่ละข้อถูกจากผู้สอบที่มีความสามารถระดับดีมาก (A)

ชื่ออาจารย์ ……..…….. โรงเรียน …….

คำชี้แจง

แบบทดสอบที่แนบมาพร้อมกันนี้เป็นแบบทดสอบปรนัยชนิดเลือกตอบ 4 ตัวเลือก แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชา เคมี ชั้นมัธยมศึกษาปีที่ 4 จำนวน 30 ข้อ

ขอให้อาจารย์ประมาณว่า นักเรียนระดับชั้นมัธยมศึกษาปีที่ 4 ที่มีความสามารถระดับดีมาก (A) มีโอกาสตอบข้อสอบข้อนั้นถูกกี่เปอร์เซ็นต์ แล้วกรอกลงในช่องเปอร์เซ็นต์การตอบถูกไว้ให้สำหรับข้อสอบทุกข้อ เปอร์เซ็นต์การตอบถูกมีค่าตั้งแต่ 0 ถึง 100 ดังนั้น คือ

0 หมายถึง โอกาสที่นักเรียนตอบข้อสอบข้อนั้นถูกน้อยที่สุด

100 หมายถึง โอกาสที่นักเรียนตอบข้อสอบข้อนั้นถูกมากที่สุด

ตัวอย่าง

(1) จำนวนโมเลกุลของสารใดมีค่ามากที่สุด (ความรู้/ ความจำ)

ก. ก๊าซไฮโดรเจนหนัก 40 กรัม

ข. ก๊าซ CH_4 0.1 โมล

ค. ก๊าซออกซิเจนหนัก 10 กรัม

ง. ก๊าซคลอรีน 10^{23} อะตอม

1. ถ้าข้อ (1) อาจารย์คิดว่านักเรียนที่มีความสามารถระดับดีมาก (A) ประมาณ 50 คน สามารถตอบข้อสอบข้อนี้ถูกต้อง ดังนั้นเปอร์เซ็นต์ที่จะตอบข้อสอบข้อนี้ถูกจะเท่ากับ 50% ให้กรอกลงในช่องเปอร์เซ็นต์การตอบถูกที่เตรียมไว้ในช่องรอบที่ 1

2. ดำเนินการร่วมกลุ่มย่อย เพื่ออภิปรายผลการตัดสินของแต่ละคน โดยเฉพาะผู้ตัดสินที่มีผลการตัดสินสูงสุดและต่ำสุด จากนั้นให้ตัดสินคะแนนอีกครั้ง ถ้าต้องการเปลี่ยนเปอร์เซ็นต์การตอบถูกเป็น 60% ในกรอกลงในช่องเปอร์เซ็นต์การตอบถูกที่เตรียมไว้ในช่องที่ 2

3. นำผู้ตัดสินทั้งหมดร่วมเป็นกุล่มใหญ่ อภิปรายผลการตัดสินร่วมกันอีกครั้ง ในครั้งนี้จะมีค่าเฉลี่ยคะแนนจุดตัดให้พิจารณาอีกครั้ง จากนั้นให้ตัดสินคะแนนเป็นครั้งสุดท้าย ถ้าอาจารย์ไม่เปลี่ยนเปอร์เซ็นต์การตอบถูก ให้อาจารย์กรอกลงในช่องเปอร์เซ็นต์การตอบถูกที่เตรียมไว้ในช่องรอบที่ 3 ดังนี้

ข้อที่	เปอร์เซ็นต์การตอบถูก		
	รอบที่ 1 (รายบุคคล)	**รอบที่ 2 (กลุ่มย่อย)**	**รอบที่ 3 (กลุ่มใหญ่)**
1	50	60	60

ชื่ออาจารย์ ..... โรงเรียน .....

ข้อที่	เปอร์เซ็นต์การตอบถูก		
	รอบที่ 1 (รายบุคคล)	**รอบที่ 2 (กลุ่มย่อย)**	**รอบที่ 3 (กลุ่มใหญ่)**
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			

ข้อที่	เปอร์เซ็นต์การตอบถูก		
	รอบที่ 1 (รายบุคคล)	**รอบที่ 2 (กลุ่มย่อย)**	**รอบที่ 3 (กลุ่มใหญ่)**
25			
26			
27			
28			
29			
30			

แบบประมาณเปอร์เซ็นต์การตอบข้อสอบแต่ละข้อถูกจากผู้สอบที่มีความสามารถระดับดี (B)

ชื่ออาจารย์ …….. โรงเรียน …….

คำชี้แจง

แบบทดสอบที่แนบมาพร้อมกันนี้เป็นแบบทดสอบปรนัยชนิดเลือกตอบ 4 ตัวเลือก แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชา เคมี ชั้นมัธยมศึกษาปีที่ 4 จำนวน 30 ข้อ

ขอให้อาจารย์ประมาณว่า นักเรียนระดับชั้นมัธยมศึกษาปีที่ 4 ที่มีความสามารถระดับดี (B) มีโอกาสตอบข้อสอบข้อนั้นถูกกี่เปอร์เซ็นต์ แล้วกรอกลงในช่องเปอร์เซ็นต์การตอบถูกไว้ให้สำหรับข้อสอบทุกข้อ เปอร์เซ็นต์การตอบถูกมีค่าตั้งแต่ 0 ถึง 100 ดังนั้น คือ

0 หมายถึง โอกาสที่นักเรียนตอบข้อสอบข้อนั้นถูกน้อยที่สุด

100 หมายถึง โอกาสที่นักเรียนตอบข้อสอบข้อนั้นถูกมากที่สุด

ตัวอย่าง

(1) จำนวนโมเลกุลของสารใดมีค่ามากที่สุด (ความรู้/ ความจำ)

ก. ก๊าซไฮโดรเจนหนัก 40 กรัม

ข. ก๊าซ CH_4 0.1 โมล

ค. ก๊าซออกซิเจนหนัก 10 กรัม

ง. ก๊าซคลอรีน 10^{23} อะตอม

1. ถ้าข้อ (1) อาจารย์คิดว่านักเรียนที่มีความสามารถระดับดี (B) ประมาณ 50 คน สามารถตอบข้อสอบข้อนี้ถูกต้อง ดังนั้นเปอร์เซ็นต์ที่จะตอบข้อสอบข้อนี้ถูกจะเท่ากับ 50% ให้กรอกลงในช่องเปอร์เซ็นต์การตอบถูกที่เตรียมไว้ในช่องรอบที่ 1

2. ดำเนินการร่วมกลุ่มย่อย เพื่ออภิปรายผลการตัดสินของแต่ละคน โดยเฉพาะผู้ตัดสินที่มีผลการตัดสินสูงสุดและต่ำสุด จากนั้นให้ตัดสินคะแนนอีกครั้ง ถ้าต้องการเปลี่ยนเปอร์เซ็นต์การตอบถูกเป็น 60% ในกรอกลงในช่องเปอร์เซ็นต์การตอบถูกที่เตรียมไว้ในช่องที่ 2

3. นำผู้ตัดสินทั้งหมดร่วมเป็นกุล่มใหญ่ อภิปรายผลการตัดสินร่วมกันอีกครั้ง ในครั้งนี้จะมีค่าเฉลี่ยคะแนนจุดตัดให้พิจารณาอีกครั้ง จากนั้นให้ตัดสินคะแนนเป็นครั้งสุดท้าย ถ้าอาจารย์ไม่เปลี่ยนเปอร์เซ็นต์การตอบถูก ให้อาจารย์กรอกลงในช่องเปอร์เซ็นต์การตอบถูกที่เตรียมไว้ในช่องรอบที่ 3 ดังนี้

ข้อที่	เปอร์เซ็นต์การตอบถูก		
	รอบที่ 1 (รายบุคคล)	รอบที่ 2 (กลุ่มย่อย)	รอบที่ 3 (กลุ่มใหญ่)
1	50	60	60

ชื่ออาจารย์ ..... โรงเรียน .....

ข้อที่	เปอร์เซ็นต์การตอบถูก		
	รอบที่ 1 (รายบุคคล)	**รอบที่ 2 (กลุ่มย่อย)**	**รอบที่ 3 (กลุ่มใหญ่)**
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			

ข้อที่	เปอร์เซ็นต์การตอบถูก		
	รอบที่ 1 (รายบุคคล)	**รอบที่ 2 (กลุ่มย่อย)**	**รอบที่ 3 (กลุ่มใหญ่)**
25			
26			
27			
28			
29			
30			

แบบประมาณเปอร์เซ็นต์การตอบข้อสอบแต่ละข้อถูกจากผู้สอบที่มีความสามารถระดับพอใช้ (C)
ชื่ออาจารย์ …….. โรงเรียน …….
คำชี้แจง

แบบทดสอบที่แนบมาพร้อมกันนี้เป็นแบบทดสอบปรนัยชนิดเลือกตอบ 4 ตัวเลือก แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชา เคมี ชั้นมัธยมศึกษาปีที่ 4 จำนวน 30 ข้อ

ขอให้อาจารย์ประมาณว่า นักเรียนระดับชั้นมัธยมศึกษาปีที่ 4 ที่มีความสามารถระดับพอใช้ (C) มีโอกาสตอบข้อสอบข้อนั้นถูกกี่เปอร์เซ็นต์ แล้วกรอกลงในช่องเปอร์เซ็นต์การตอบถูกไว้ให้สำหรับข้อสอบทุกข้อ เปอร์เซ็นต์การตอบถูกมีค่าตั้งแต่ 0 ถึง 100 ดังนั้น คือ

0 หมายถึง โอกาสที่นักเรียนตอบข้อสอบข้อนั้นถูกน้อยที่สุด

100 หมายถึง โอกาสที่นักเรียนตอบข้อสอบข้อนั้นถูกมากที่สุด

ตัวอย่าง

(1) จำนวนโมเลกุลของสารใดมีค่ามากที่สุด (ความรู้/ ความจำ)

ก. ก๊าซไฮโดรเจนหนัก 40 กรัม

ข. ก๊าซ CH_4 0.1 โมล

ค. ก๊าซออกซิเจนหนัก 10 กรัม

ง. ก๊าซคลอรีน 10^{23} อะตอม

1. ถ้าข้อ (1) อาจารย์คิดว่านักเรียนที่มีความสามารถระดับพอใช้ (C) ประมาณ 50 คน สามารถตอบข้อสอบข้อนี้ถูกต้อง ดังนั้นเปอร์เซ็นต์ที่จะตอบข้อสอบข้อนี้ถูกจะเท่ากับ 50% ให้กรอกลงในช่องเปอร์เซ็นต์การตอบถูกที่เตรียมไว้ในช่องรอบที่ 1

2. ดำเนินการร่วมกลุ่มย่อย เพื่ออภิปรายผลการตัดสินของแต่ละคนโดยเฉพาะผู้ตัดสินที่มีผลการตัดสินสูงสุดและต่ำสุด จากนั้นให้ตัดสินคะแนนอีกครั้ง ถ้าต้องการเปลี่ยนเปอร์เซ็นต์การตอบถูกเป็น 60% ในกรอกลงในช่องเปอร์เซ็นต์การตอบถูกที่เตรียมไว้ในช่องที่ 2

3. นำผู้ตัดสินทั้งหมดร่วมเป็นกุล่มใหญ่ อภิปรายผลการตัดสินร่วมกันอีกครั้ง ในครั้งนี้จะมีค่าเฉลี่ยคะแนนจุดตัดให้พิจารณาอีกครั้ง จากนั้นให้ตัดสินคะแนนเป็นครั้งสุดท้าย ถ้าอาจารย์ไม่เปลี่ยนเปอร์เซ็นต์การตอบถูก ให้อาจารย์กรอกลงในช่องเปอร์เซ็นต์การตอบถูกที่เตรียมไว้ในช่องรอบที่ 3 ดังนี้

ข้อที่	เปอร์เซ็นต์การตอบถูก		
	รอบที่ 1 (รายบุคคล)	รอบที่ 2 (กลุ่มย่อย)	รอบที่ 3 (กลุ่มใหญ่)
1	50	60	60

ชื่ออาจารย์ …….. โรงเรียน …….

ข้อที่	เปอร์เซ็นต์การตอบถูก		
	รอบที่ 1 (รายบุคคล)	**รอบที่ 2 (กลุ่มย่อย)**	**รอบที่ 3 (กลุ่มใหญ่)**
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			

ข้อที่	เปอร์เซ็นต์การตอบถูก		
	รอบที่ 1 (รายบุคคล)	**รอบที่ 2 (กลุ่มย่อย)**	**รอบที่ 3 (กลุ่มใหญ่)**
25			
26			
27			
28			
29			
30			

แบบประมาณเปอร์เซ็นต์การตอบข้อสอบแต่ละข้อถูกจากผู้สอบที่มีความสามารถระดับอ่อน (D)
ชื่ออาจารย์ …….. โรงเรียน …….
คำชี้แจง

แบบทดสอบที่แนบมาพร้อมกันนี้เป็นแบบทดสอบปรนัยชนิดเลือกตอบ 4 ตัวเลือก แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชา เคมี ชั้นมัธยมศึกษาปีที่ 4 จำนวน 30 ข้อ

ขอให้อาจารย์ประมาณว่า นักเรียนระดับชั้นมัธยมศึกษาปีที่ 4 ที่มีความสามารถระดับอ่อน (D) มีโอกาสตอบข้อสอบข้อนั้นถูกกี่เปอร์เซ็นต์ แล้วกรอกลงในช่องเปอร์เซ็นต์การตอบถูกไว้ให้สำหรับข้อสอบทุกข้อ เปอร์เซ็นต์การตอบถูกมีค่าตั้งแต่ 0 ถึง 100 ดังนั้น คือ

0 หมายถึง โอกาสที่นักเรียนตอบข้อสอบข้อนั้นถูกน้อยที่สุด

100 หมายถึง โอกาสที่นักเรียนตอบข้อสอบข้อนั้นถูกมากที่สุด

ตัวอย่าง

(1) จำนวนโมเลกุลของสารใดมีค่ามากที่สุด (ความรู้/ ความจำ)

ก. ก๊าซไฮโดรเจนหนัก 40 กรัม

ข. ก๊าซ CH_4 0.1 โมล

ค. ก๊าซออกซิเจนหนัก 10 กรัม

ง. ก๊าซคลอรีน 10^{23} อะตอม

1. ถ้าข้อ (1) อาจารย์คิดว่านักเรียนที่มีความสามารถระดับอ่อน (D) ประมาณ 50 คน สามารถตอบข้อสอบข้อนี้ถูกต้อง ดังนั้นเปอร์เซ็นต์ที่จะตอบข้อสอบข้อนี้ถูกจะเท่ากับ 50% ให้กรอกลงในช่องเปอร์เซ็นต์การตอบถูกที่เตรียมไว้ในช่องรอบที่ 1

2. ดำเนินการร่วมกลุ่มย่อย เพื่ออภิปรายผลการตัดสินของแต่ละคนโดยเฉพาะผู้ตัดสินที่มีผลการตัดสินสูงสุดและต่ำสุด จากนั้นให้ตัดสินคะแนนอีกครั้ง ถ้าต้องการเปลี่ยนเปอร์เซ็นต์การตอบถูกเป็น 60% ในกรอกลงในช่องเปอร์เซ็นต์การตอบถูกที่เตรียมไว้ในช่องที่ 2

3. นำผู้ตัดสินทั้งหมดร่วมเป็นกุล่มใหญ่ อภิปรายผลการตัดสินร่วมกันอีกครั้ง ในครั้งนี้จะมีค่าเฉลี่ยคะแนนจุดตัดให้พิจารณาอีกครั้ง จากนั้นให้ตัดสินคะแนนเป็นครั้งสุดท้าย ถ้าอาจารย์ไม่เปลี่ยนเปอร์เซ็นต์การตอบถูก ให้อาจารย์กรอกลงในช่องเปอร์เซ็นต์การตอบถูกที่เตรียมไว้ในช่องรอบที่ 3 ดังนี้

ข้อที่	เปอร์เซ็นต์การตอบถูก		
	รอบที่ 1 (รายบุคคล)	รอบที่ 2 (กลุ่มย่อย)	รอบที่ 3 (กลุ่มใหญ่)
1	50	60	60

ชื่ออาจารย์ …….. โรงเรียน …….

ข้อที่	เปอร์เซ็นต์การตอบถูก		
	รอบที่ 1 (รายบุคคล)	**รอบที่ 2 (กลุ่มย่อย)**	**รอบที่ 3 (กลุ่มใหญ่)**
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			

ข้อที่	เปอร์เซ็นต์การตอบถูก		
	รอบที่ 1 (รายบุคคล)	**รอบที่ 2 (กลุ่มย่อย)**	**รอบที่ 3 (กลุ่มใหญ่)**
25			
26			
27			
28			
29			
30			

ภาคผนวก ค

แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชา เคมี ชั้นมัธยมศึกษาปีที่ 4 ภาคเรียน 2

แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาเคมี
กลุ่มสาระการเรียนรู้วิทยาศาสตร์ เรื่อง ปริมาณสารสัมพันธ์

.....

คำชี้แจง ห้ามขีดเขียนหรือทำเครื่องหมายใด ๆ ลงในแบบทดสอบ กรุณาทดเลขในกระดาษคำตอบ

ผลการเรียนรู้ที่ 1 สามารถอธิบายความหมายของมวลอะตอม มวลโมเลกุล คำนวณหามวลอะตอมของธาตุได้

1. ธาตุ A มีมวลอะตอม 20 หมายความว่าอย่างไร (ความรู้/ ความจำ)
   ก. ธาตุ A 1 อะตอมหนัก 20เท่าของ 6.02×10^{23} กรัม
   ข. ธาตุ A 1 อะตอมหนัก 20เท่าของ ไฮโดรเจน 1 อะตอม
   ค. ธาตุ A 1 อะตอมหนัก 20 กรัม
   ง. ธาตุ A 1 อะตอมมีมวล $20 \times 1.66 \times 10^{-24}$ กรัม

ผลการเรียนรู้ที่ 2 สามารถอธิบายความหมายของมวลโมเลกุลของสารมวลของธาตุ 1 อะตอม มวลอะตอมเฉลี่ยของธาตุ และมวลของสาร 1 โมเลกุลได้

2. สารประกอบ X มีมวลโมเลกุล 44 ถ้ามี 2 โมเลกุลจะหนักเท่าใด (ความเข้าใจ)
   ก. $2 \times 1.66 \times 10^{-24}$ g
   ข. $44 \times 1.66 \times 10^{-24}$ g
   ค. $46 \times 1.66 \times 10^{-24}$ g
   ง. $88 \times 1.66 \times 10^{-24}$ g
3. ธาตุแคลเซียมมวลอะตอม 40 ข้อใดถูกต้อง (ความเข้าใจ)
   ก. ธาตุแคลเซียม 1 อะตอม หนักเป็น 40 เท่า ของ $\frac{1}{12}$ มวลของคาร์บอน 12, 12 อะตอม
   ข. ธาตุแคลเซียม 1 อะตอม หนักเป็น 40 เท่า ของ $\frac{1}{16}$ มวลของออกซิเจน, 1 อะตอม
   ค. ธาตุแคลเซียม 1 อะตอม หนัก $40 \times 1.66 \times 10^{-24}$ กิโลกรัม
   ง. ธาตุแคลเซียม 1 อะตอม หนัก 6.64×10^{-23} กรัม
4. ธาตุคาร์บอนมีไอโซโทปที่เสถียรในธรรมชาติ 2 ชนิด คือ C-12 มีมวลอะตอม 12.000 เปอร์เซ็นต์ในธรรมชาติ 98.9% และ C-13 มีมวลอะตอม 13.003 ถ้ามวลอะตอมเฉลี่ยของไอโซโทปของคาร์บอนเป็น 12.011 เปอร์เซ็นต์ในธรรมชาติของไอโซโทป C-13 เป็นเท่าใด (ความรู้/ ความจำ)
   ก. 0.9 %
   ข. 1.0 %
   ค. 1.1 %
   ง. 2.0 %

ผลการเรียนรู้ที่ 3 สามารถอธิบายความสัมพันธ์ระหว่างโมล มวล จำนวนอนุภาคและปริมาตรของก๊าซได้

5. จำนวนโมเลกุลของสารใดมีค่ามากที่สุด (ความรู้/ ความจำ)
   ก. ก๊าซไฮโดรเจนหนัก 40 กรัม
   ข. ก๊าซ CH_4 0.1 โมล
   ค. ก๊าซออกซิเจนหนัก 10 กรัม
   ง. ก๊าซคลอรีน 10^{23} อะตอม
6. ข้อความใดต่อไปนี้ไม่ใช่เป็นสมบัติของก๊าซใด ๆ ที่มีปริมาตร 22.4 dm^3 ที่อุณหภูมิและความดันมาตรฐาน (ความรู้/ ความจำ)
   ก. จำนวนโมเลกุล = 6.02×10^{23} โมเลกุล
   ข. จำนวนโมเลกุลของก๊าซนี้เท่ากับจำนวนของไฮโดรเจนที่อุณหภูมิและความดันเดียวกัน
   ค. มวล = มวลโมเลกุลคิดเป็นกรัม
   ง. จำนวนโมล = 6.02×10^{23} โมล

ผลการเรียนรู้ที่ 4 สามารถคำนวณความสัมพันธ์ระหว่างโมล มวล จำนวนอนุภาคและปริมาตรของก๊าซได้

7. ก๊าซชนิดหนึ่งมีสูตรโมเลกุล C_3H_8 ถ้าก๊าซนี้หนัก 2 กรัมจะมีปริมาตรเท่าใดที่ STP (C = 12, H = 1) (การนำไปใช้)
   ก. 2.42 ลิตร ข. 1.96 ลิตร
   ค. 1.08 ลิตร ง. 1.87 ลิตร
8. ก๊าซ A มีสูตรโมเลกุลเป็น A_3 ถ้าก๊าซนี้มีอยู่ 6.02×10^{23} อะตอมจะมีปริมาตรเท่าใดที่ STP (การนำไปใช้)
   ก. 7.47 dm^3 ข. 11.20 dm^3
   ค. 14.94 dm^3 ง. 22.40 dm^3
9. คริปตอนมีปริมาตร 44.8 dm^3 ที่อุณหภูมิและความดันมาตรฐาน จะมีมวลเท่าใด (คริปตอนมีมวลอะตอม = 84) (การนำไปใช้)
   ก. 22.4 g ข. 44.8 g
   ค. 84.0 g ง. 168.0 g
10. นำสารต่อไปนี้มาอย่างละ 10 กรัม สารใดมีจำนวนโมเลกุลมากที่สุด (การนำไปใช้)
    ก. CO_2 ข. H_2O
    ค. NH_3 ง. SO_2
11. ข้อใดกล่าวไม่ถูกต้อง (การนำไปใช้)
    ก. น้ำ 1 โมล หนัก 18 กรัม
    ข. ก๊าซคาร์บอนไดออกไซด์ 0.5 โมล มีปริมาตร 11.2 dm^3 ที่ STP
    ค. โซเดียมอะตอม หนัก 69 กรัม มี 6.02×10^{23} อะตอม
    ง. ก๊าซไฮโดรเจน 44.8 dm^3 ที่STP มี 1.204×10^{24} อะตอม

ผลการเรียนรู้ที่ 5 สามารถอธิบายสมบัติบางประการของสารละลาย ทำการทดลอง เตรียมสารละลายได้

12. สารละลายมีลักษณะอย่างไร (ความรู้/ ความจำ)
    ก. เป็นสารที่มีสถานะเป็นของเหลว
    ข. สารที่มองเห็นเนื้อสารแตกต่างกัน
    ค. สารที่ประกอบขึ้นด้วยสารอย่างน้อย 2 ชนิด
    ง. สารที่เป็นส่วนประกอบมีสถานะเป็นของแข็งและของเหลวเท่านั้น

13. สารในกลุ่มใดต่อไปนี้แตกต่างจากสารกลุ่มอื่น (ความรู้/ ความจำ)
ก. เหล็ก ทองแดง สังกะสี
ข. อากาศ ก๊าซหุงต้ม น้ำเกลือ
ค. เหรียญบาท เหล็กกล้า ทองสัมฤทธิ์
ง. น้ำอัดลม แอลกอฮอล์ล้างแผล ทิงเจอร์ไอโอดีน

14. เมื่อต้องการสารละลายเอทานอลเข้มข้นร้อยละ 15 โดยปริมาตร ข้อใดกล่าวถูกต้อง (ความรู้/ ความจำ)
ก. ใส่เอทานอล 15 cm^3 น้ำ 100 cm^3
ข. ใส่เอทานอล 15 cm^3 น้ำ 85 cm^3
ค. ใส่เอทานอล 15 cm^3 น้ำ 115 cm^3
ง. ใส่เอทานอล 15 cm^3 น้ำ 15 cm^3

15. ถ้าต้องการเตรียมสารละลาย $KMnO_4$ เข้มข้น 0.1 mol/ dm^3 ต้องใช้ $KMnO_4$ กี่กรัม (K = 39, Mn = 55, O = 16) (ความรู้/ ความจำ)
ก. 15 กรัม ข. 15.8 กรัม
ค. 150 กรัม ง. 158 กรัม

ผลการเรียนรู้ที่ 6 สามารถคำนวณความเข้มข้นของสารละลายในหน่วยร้อยละ โมลาริตี โมแลล ส่วนในล้านส่วน ส่วนในพันล้านส่วนและเศษส่วนโมลได้

16. ชั่ง NaCl 2.925 กรัม ในสารละลาย 500 cm^3 จงหาความเข้มข้นของสารละลาย NaCl (Na = 23, Cl = 35.5) (การนำไปใช้)
ก. 0.01 mol/ dm^3 ข. 0.1 mol/ dm^3
ค. 1 mol/ dm^3 ง. 10 mol/ dm^3

17. ในสารละลายชนิดหนึ่งประกอบด้วย เอทานอล (C_2H_6O) 92 กรัม และน้ำ 180 กรัม จงหาความเข้มข้นเป็นโมลต่อกิโลกรัม (C = 12,O = 16,H = 1) (การนำไปใช้)
ก. 1 mol/ kg ข. 10.11 mol/ kg
ค. 11.11 mol/ kg ง. 12.11 mol/ kg

18. ละลาย NaOH จำนวน 20 กรัม ลงในน้ำได้สารละลาย 500 cm^3 แบ่งสารละลายจำนวนนี้มา 100 cm^3 แล้วเติมน้ำให้มีปริมาตร 250 cm^3 และสารละลายจำนวนนี้เมื่อนำมาเคี่ยวจนเหลือปริมาตร 100 cm^3 จะมีความเข้มข้นเท่าใด (Na = 23, H = 1, O = 16) (การนำไปใช้)
ก. 0.5 mol/ L ข. 0.1 mol/ L
ค. 0.2 mol/ L ง. 1.0 mol/ L

ผลการเรียนรู้ที่ 7 สามารถอธิบายและทำการทดลองเกี่ยวกับกฎทรงมวลของสารและกฎสัดส่วนคงที่ได้

19. การจัดการทดลองเพื่อยืนยันกฎทรงมวลนั้น ควรจะต้องระวังสิ่งใดเป็นพิเศษในเรื่องต่อไปนี้ (ความรู้/ ความจำ)
ก. การผสมเพื่อทำให้สารทำปฏิกิริยากัน
ข. การกำหนดปริมาณของสารที่ใช้ให้เหมาะสม
ค. การควบคุมการถ่ายเทพลังงานระหว่างระบบกับสิ่งแวดล้อม
ง. การป้องกันไม่ให้ก๊าซที่อาจเกิดขึ้นหนีหายไป

20. สารบริสุทธ์ 1 และ 2 ประกอบด้วย ธาตุ A และ B สาร 1 หนัก 4.50 กรัม ประกอบด้วย ธาตุ A 1.8 กรัม และจากการวิเคราะห์สาร 2 พบว่า ประกอบด้วยธาตุ A และ B ร้อยละ 40 และ 60 ตามลำดับ จากข้อมูลที่กำหนดให้นี้ ข้อใดถูกต้อง (ความเข้าใจ)

ก. อธิบายถึงกฎทรงมวล

ข. พิสูจน์ให้เห็นว่าทั้ง 1 และ 2 เป็นของผสม

ค. สูตรของ 1 และ 2 คือ AB_3 และ A_2B_3 ตามลำดับ

ง. อธิบายถึงกฎสัดส่วนคงที่

ผลการเรียนรู้ที่ 8 สามารถอธิบายและทำการทดลองความสัมพันธ์ของปริมาตรก๊าซในสมการเคมี ใช้กฎของเกย์ลุสแซก กฎของอาโวกาโดรได้

21. สมมุติ คาร์บอน (C) 3 กรัม สามารถรวมตัวกับออกซิเจน (O) 8 กรัม แล้วเกิดเป็นสารประกอบชนิดหนึ่ง หากต้องการสารประกอบชนิดเดียวกันนี้โดยเตรียมจากคาร์บอน 15 กรัม ต้องใช้ออกซิเจนกี่กรัมจึงจะทำปฏิกิริยากันพอดี และสารประกอบที่เกิดจะมีมวล กี่กรัม (ตอบตามลำดับ) (การวิเคราะห์)

ก. 40, 55 ข. 20, 35

ค. 30, 45 ง. 50, 65

ผลการเรียนรู้ที่ 9 สามารถคำนวณหาปริมาตรของก๊าซในปฏิกิริยาเคมีได้

22. นำก๊าซชนิดหนึ่งซึ่งมีสูตรโมเลกุล A_2 ทำปฏิกิริยากับก๊าซอีกชนิดหนึ่งซึ่งมีสูตรโมเลกุล B_2 ได้ก๊าซที่มีสูตรโมเลกุล AB_3 ดังสมการ

$A_2 + 3\ B_2 \longrightarrow 2AB_3$ ถ้าใช้ก๊าซ A_2 35 ลูกบาศก์เซนติเมตร ทำปฏิกิริยากับ B_2 81 ลูกบาศก์เซนติเมตร ถามว่าก๊าซ A_2 กับ B_2 ก๊าซไหนจะถูกใช้หมดก่อนกัน (การนำไปใช้)

ก. A_2 ข. B_2

ค. หมดพร้อมกัน ง. ไม่มีตัวใดหมด

23. นำ H_2 4 dm^3 และ O_2 3 dm^3 มาทำปฏิกิริยากันที่อุณหภูมิ 100°C ดังสมการ

$2H_2\ (g) + O_2\ (g) \longrightarrow 2H_2O\ (g)$

เมื่อปฏิกิริยาเกิดอย่างสมบูรณ์จะมีก๊าซในระบบกี่ dm^3 (การนำไปใช้)

ก. 4 ข. 5

ค. 6 ง. 8

ผลการเรียนรู้ที่ 10 สามารถอธิบายเกี่ยวกับ สูตร และสมการเคมีได้

24. ข้อใดกล่าวถูกต้องเกี่ยวกับสูตรเอมพิริคัล (ความรู้/ ความจำ)

ก. เป็นสูตรโมเลกุลของสาร

ข. เป็นสูตรที่แสดงอัตราส่วนอย่างต่ำของสาร

ค. เป็นสูตรที่แสดงการจัดเรียงของสาร

ง. ถูกทุกข้อ

25. สารประกอบในข้อใดต่อไปนี้มีสูตรเอมพิริคัลไม่เป็น CH_2O (ความเข้าใจ)

ก. กรดอะซิติก CH_3COOH

ข. เอทานอล CH_3CH_2OH

ค. กลูโคส $C_6H_{12}O_6$

ง. เมทิลฟลอเมต $HCOOCH_3$

ผลการเรียนรู้ที่ 11 สามารถคำนวณเกี่ยวกับสูตรและสมการเคมีได้

26. สารประกอบชนิดหนึ่งประกอบด้วย C = 1.2 g, H = 0.2 g, และO = 1.6 g ถ้ามวลโมเลกุลของสารนี้ 60 สูตรโมเลกุลของสารนี้คือข้อใด (การนำไปใช้)

ก. $C_2H_2O_4$ ข. C_2H_6O

ค. CH_2O ง. $C_2H_4O_2$

27. โลหะ M มีมวลอะตอม 200 เมื่อนำมาเผากับกำมะถันที่มีมากเกินพอปรากฏว่าM หนัก 16.00 กรัม ให้สารประกอบหนัก 19.84 กรัม สูตรเอมพิริคัลของสารประกอบนี้คือข้อใด (การนำไปใช้)

ก. M_2S_3 ข. M_3S_2

ค. MS ง. M_2S

ผลการเรียนรู้ที่ 12 สามารถใช้ความสัมพันธ์ระหว่างจำนวนโมล กับมวลหรือปริมาตรของสาร คำนวณหาค่าโมล มวลหรือปริมาตรของสาร สมการเคมี และหาค่าร้อยละของผลที่ได้จากการทดลองได้

28. ถ้านำเบนซีน (C_6H_6) 15.6 กรัม มาทำปฏิกิริยากับกรดไนตริก (HNO_3) จำนวนมากเกินพอ พบว่า เกิดไนโตรเบนซีน ($C_6H_5NO_2$) 18.0 กรัม ดังสมการ

จงหาผลได้ร้อยละ (การนำไปใช้)

C_6H_6 (l) + HNO_3 (aq) $\longrightarrow$ $C_6H_5NO_2$ (l) + H_2O (l)

ก. 50.0 ข. 62.5

ค. 73.2 ง. 78.3

29. จากปฏิกิริยา C_3H_8 (g) +5O_2 (g) $\rightarrow$ 3CO_2 (g) +4 H_2O (g) หากต้องการ CO_2 44 กรัม ต้องใช้ C_3H_8 กี่ลิตร (ความเข้าใจ)

ก. 6.0 ข. 6.8

ค. 7.5 ง. 8.8

30. ในการเผา $KClO_3$ จะเกิดปฏิกิริยาดังนี้ $2KClO_3 \longrightarrow 2KCl + 3O_2$ ถ้าเผา $KClO_3$ จำนวน 12.26 กรัม จะได้ก๊าซ O_2กี่ลิตรที่ (STP) (K = 39,Cl = 35.5,O = 16) (การวิเคราะห์)

ก. 1.12 ข. 2.24

ค. 3.36 ง. 4.48

ภาคผนวก ง

ผลการวิเคราะห์ฟังก์ชันสารสนเทศของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาเคมี เรื่อง ปริมาณสารสัมพันธ์ เป็นแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา

IRTPRO Version 4.2
Output generated by IRTPRO estimation engine Version 5.20 (64-bit)

Project:	
Description:	
Date:	15 April 2020
Time:	02: 52 PM

Table of Contents

2PL Model Item Parameter Estimates for Group 1, logit: $a\theta + c$ or $a(\theta - b)$
Summed-Score Based Item Diagnostic Tables and X^2s for Group 1
Group Parameter Estimates
Summed Score to Scale Score Conversion Table
Marginal fit (X^2) and Standardized LD X^2 Statistics for Group 1
Item Information Function Values for Group 1 at 15 Values of θ from -2.8 to 2.8
Likelihood-based Values and Goodness of Fit Statistics
Summary of the Data and Control Parameters

2PL Model Item Parameter Estimates for Group 1, logit: $a\theta + c$ or $a(\theta - b)$ **(Back to TOC)**

Item	Label	*a*	*s.e.*	*c*	*s.e.*	*b*	*s.e.*
1	VAR001	[2] 0.13	0.09	[1] 0.17	0.08	-1.35	1.09
2	VAR002	[4] 0.68	0.13	[3] 1.27	0.10	-1.86	0.33
3	VAR003	[6] 0.24	0.09	[5] -1.01	0.09	4.22	1.66
4	VAR004	[8] 0.40	0.09	[7] 0.07	0.08	-0.18	0.20
5	VAR005	[10] 1.25	0.14	[9] 0.00	0.10	-0.00	0.08
6	VAR006	[12] 1.85	0.20	[11] 0.03	0.12	-0.02	0.06
7	VAR007	[14] 1.08	0.13	[13] 0.18	0.10	-0.17	0.09
8	VAR008	[16] 0.60	0.10	[15] -1.18	0.10	1.98	0.33
9	VAR009	[18] 1.23	0.16	[17] 0.84	0.11	-0.68	0.10
10	VAR010	[20] 1.21	0.13	[19] -0.53	0.10	0.44	0.09
11	VAR011	[22] 1.36	0.15	[21] -0.10	0.10	0.07	0.08
12	VAR012	[24] 1.58	0.18	[23] 0.38	0.11	-0.24	0.07
13	VAR013	[26] 0.18	0.09	[25] -0.19	0.08	1.06	0.67
14	VAR014	[28] 0.76	0.10	[27] -0.39	0.09	0.52	0.13
15	VAR015	[30] 1.22	0.15	[29] 0.70	0.11	-0.57	0.09
16	VAR016	[32] 0.92	0.12	[31] 0.35	0.09	-0.38	0.10
17	VAR017	[34] 0.73	0.10	[33] -0.24	0.08	0.33	0.12
18	VAR018	[36] 0.35	0.09	[35] -0.55	0.08	1.57	0.45
19	VAR019	[38] 1.43	0.15	[37] -1.45	0.13	1.02	0.11
20	VAR020	[40] 1.13	0.13	[39] 0.13	0.10	-0.12	0.08
21	VAR021	[42] 0.60	0.10	[41] -0.47	0.09	0.78	0.18
22	VAR022	[44] 0.72	0.10	[43] -0.88	0.09	1.22	0.20
23	VAR023	[46] 0.35	0.09	[45] -0.54	0.08	1.54	0.45
24	VAR024	[48] 1.64	0.17	[47] -0.81	0.12	0.50	0.08
25	VAR025	[50] 1.59	0.17	[49] 0.04	0.11	-0.02	0.07
26	VAR026	[52] 1.87	0.20	[51] -0.38	0.12	0.20	0.06
27	VAR027	[54] 0.74	0.10	[53] -0.61	0.09	0.82	0.15
28	VAR028	[56] 0.60	0.10	[55] -0.31	0.08	0.51	0.16
29	VAR029	[58] 0.95	0.12	[57] -0.37	0.09	0.39	0.10
30	VAR030	[60] 0.67	0.10	[59] -0.04	0.08	0.06	0.13

Summed-Score Based Item Diagnostic Tables and X^2s for Group 1 **(Back to TOC)**

S-X^2 Item Level Diagnostic Statistics

Item	Label	X^2	*d.f.*	Probability
1	VAR001	35.99	22	0.0304
2	VAR002	30.57	19	0.0449
3	VAR003	70.23	22	0.0001

4	VAR004	29.91	23	0.1517
5	VAR005	26.44	19	0.1180
6	VAR006	26.26	18	0.0936
7	VAR007	70.15	21	0.0001
8	VAR008	57.29	21	0.0001
9	VAR009	25.99	19	0.1302
10	VAR010	80.18	20	0.0001
11	VAR011	54.38	19	0.0001
12	VAR012	42.54	18	0.0009
13	VAR013	35.44	22	0.0348
14	VAR014	52.18	22	0.0003
15	VAR015	48.84	19	0.0002
16	VAR016	36.95	21	0.0170
17	VAR017	59.38	22	0.0001
18	VAR018	41.54	21	0.0048
19	VAR019	53.58	20	0.0001
20	VAR020	41.96	20	0.0028
21	VAR021	76.73	22	0.0001
22	VAR022	34.71	21	0.0303
23	VAR023	37.53	22	0.0206
24	VAR024	48.63	20	0.0003
25	VAR025	65.43	18	0.0001
26	VAR026	40.34	18	0.0019
27	VAR027	43.67	21	0.0026
28	VAR028	78.72	22	0.0001
29	VAR029	53.84	21	0.0001
30	VAR030	56.15	22	0.0001

Group Parameter Estimates **(Back to TOC)**

Group	Label	μ	*s.e.*	σ^2	*s.e.*	σ	*s.e.*
1	Group 1	0.00	-----	1.00	-----	1.00	-----

Summed Score to Scale Score Conversion Table **(Back to TOC)**

Summed Score	EAP[θ|x]	SD[θ|x]	Modeled Proportion

Marginal reliability of the scaled scores for summed scores = 1.00000

Marginal fit (X^2) and Standardized LD X^2 Statistics for Group 1 **(Back to TOC)**

Item	Label	Marginal X^2	1	2	3	4	5	6	7	8	9	10
1	VAR001	0.0										
2	VAR002	0.0	32.3									
3	VAR003	0.0	7.3	-0.2								
4	VAR004	0.0	-0.2	7.2	0.9							
5	VAR005	0.4	-0.4	4.4	0.1	7.9						
6	VAR006	1.4	0.9	0.7	1.8	0.3	0.6					
7	VAR007	0.2	2.7	-0.5	-0.0	-0.1	0.3	5.9				
8	VAR008	0.0	0.7	-0.7	0.1	0.2	-0.1	0.3	-0.1			
9	VAR009	0.2	1.6	2.0	-0.4	-0.6	2.7	0.8	2.4	-0.1		
10	VAR010	0.3	-0.2	1.4	2.0	4.2	-0.1	1.4	0.0	5.9	-0.3	
11	VAR011	0.5	0.5	0.5	0.9	0.1	-0.1	0.5	1.6	-0.3	5.5	-0.2
12	VAR012	0.8	0.5	0.6	2.1	1.2	1.1	0.5	1.5	2.6	0.4	9.2
13	VAR013	0.0	0.3	-0.7	0.5	0.2	0.5	0.3	1.4	-0.2	-0.5	0.4
14	VAR014	0.0	-0.6	-0.6	-0.6	2.9	1.9	4.0	-0.2	-0.6	4.5	-0.0
15	VAR015	0.2	1.0	1.0	-0.5	0.9	-0.4	0.4	2.2	-0.4	-0.4	4.1
16	VAR016	0.1	4.1	7.3	-0.1	-0.6	0.3	1.1	0.8	-0.6	-0.4	1.9

17	VAR017	0.0	1.5	-0.6	-0.2	4.0	0.3	0.3	0.7	3.7	0.1	-0.5
18	VAR018	0.0	1.3	0.1	-0.3	5.0	4.4	1.4	-0.6	-0.7	0.2	-0.5
19	VAR019	0.1	-0.1	-0.5	0.3	0.2	0.3	1.1	1.7	-0.4	4.0	1.3
20	VAR020	0.3	0.3	-0.5	1.1	-0.4	-0.3	0.3	0.5	-0.3	1.4	-0.3
21	VAR021	0.0	-0.2	-0.1	0.3	2.0	0.5	6.9	3.6	0.3	-0.5	-0.5
22	VAR022	0.0	1.2	-0.6	-0.7	-0.7	1.8	0.4	3.9	-0.7	-0.6	0.3
23	VAR023	0.0	8.6	5.1	1.5	0.7	-0.4	3.6	0.0	-0.6	0.5	-0.5
24	VAR024	0.7	1.0	0.6	-0.2	1.1	-0.1	7.9	4.9	0.2	5.1	-0.1
25	VAR025	0.9	0.2	4.5	0.0	2.2	9.0	1.8	1.2	0.8	-0.0	0.3
26	VAR026	1.4	1.5	1.4	4.0	0.3	1.3	1.0	1.8	0.3	3.6	0.6
27	VAR027	0.0	-0.3	1.7	3.0	1.9	0.7	2.2	0.3	-0.4	2.4	0.7
28	VAR028	0.0	-0.7	3.7	0.8	-0.7	0.1	6.9	1.7	-0.4	0.5	2.5
29	VAR029	0.1	-0.6	1.3	0.7	0.2	5.1	3.2	1.1	3.3	0.7	-0.5
30	VAR030	0.0	3.1	6.0	-0.6	3.5	-0.0	0.7	-0.4	1.4	0.4	1.4

Item	Label	Marginal X^2	11	12	13	14	15	16	17	18	19	20
11	VAR011	0.5										
12	VAR012	0.8	0.2									
13	VAR013	0.0	-0.3	0.3								
14	VAR014	0.0	-0.2	0.2	-0.0							
15	VAR015	0.2	-0.1	0.3	0.9	5.4						
16	VAR016	0.1	2.7	-0.1	-0.4	2.2	1.1					
17	VAR017	0.0	0.1	3.8	-0.3	0.5	-0.5	2.2				
18	VAR018	0.0	5.0	1.4	-0.3	-0.3	1.5	2.2	0.8			
19	VAR019	0.1	1.5	0.6	1.5	0.2	6.0	1.7	2.2	1.1		
20	VAR020	0.3	0.4	3.8	0.7	0.7	-0.2	1.6	1.1	1.9	-0.1	
21	VAR021	0.0	0.9	0.0	-0.7	-0.2	1.1	0.3	0.4	4.4	0.2	2.1
22	VAR022	0.0	0.7	1.1	3.0	0.9	-0.0	-0.6	-0.1	-0.1	-0.6	-0.5
23	VAR023	0.0	-0.3	2.1	-0.7	4.7	1.5	0.2	7.3	0.4	-0.6	0.6
24	VAR024	0.7	0.0	3.2	3.8	-0.1	-0.1	1.1	1.5	1.6	5.4	2.1
25	VAR025	0.9	0.1	0.2	0.1	-0.1	0.5	1.4	1.2	1.5	0.7	0.4
26	VAR026	1.4	1.5	8.7	1.2	0.7	2.9	2.2	0.4	0.3	0.3	3.6
27	VAR027	0.0	0.0	3.9	0.5	0.3	2.6	0.5	-0.1	-0.5	-0.5	-0.4
28	VAR028	0.0	-0.3	5.3	0.7	2.9	7.4	-0.6	4.4	-0.2	3.3	0.2
29	VAR029	0.1	-0.0	1.1	-0.2	3.8	-0.5	-0.3	0.9	-0.5	-0.6	-0.4
30	VAR030	0.0	3.9	-0.1	-0.1	1.1	2.4	-0.4	2.4	1.8	8.1	1.8

Item	Label	Marginal X^2	21	22	23	24	25	26	27	28	29
21	VAR021	0.0									
22	VAR022	0.0	9.9								
23	VAR023	0.0	-0.5	-0.4							
24	VAR024	0.7	1.4	-0.2	0.1						
25	VAR025	0.9	-0.1	4.1	-0.0	0.3					
26	VAR026	1.4	1.7	2.6	1.7	2.5	0.6				
27	VAR027	0.0	-0.6	1.5	1.9	0.4	-0.0	1.4			
28	VAR028	0.0	6.0	4.3	4.9	0.2	2.5	0.5	0.4		
29	VAR029	0.1	1.4	-0.3	4.1	-0.1	5.1	0.7	-0.5	1.4	
30	VAR030	0.0	1.6	-0.3	2.2	-0.1	-0.1	2.2	0.1	2.9	-0.5

Item Information Function Values for Group 1 at 15 Values of θ from -2.8 to 2.8 **(Back to TOC)**

Item	Label	θ: -2.8	-2.4	-2.0	-1.6	-1.2	-0.8	-0.4	-0.0	0.4	0.8	1.2	1.6	2.0	2.4	2.8
1	VAR001	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	VAR002	0.10	0.11	0.12	0.12	0.11	0.10	0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.02
3	VAR003	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
4	VAR004	0.03	0.03	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.03	0.03	0.03
5	VAR005	0.04	0.07	0.11	0.16	0.23	0.31	0.37	0.39	0.37	0.31	0.23	0.16	0.11	0.07	0.04
6	VAR006	0.02	0.04	0.08	0.17	0.31	0.53	0.76	0.85	0.74	0.51	0.29	0.16	0.08	0.04	0.02
7	VAR007	0.06	0.09	0.12	0.17	0.22	0.26	0.29	0.29	0.26	0.22	0.18	0.13	0.09	0.06	0.04
8	VAR008	0.02	0.02	0.03	0.03	0.04	0.05	0.06	0.06	0.07	0.08	0.08	0.09	0.09	0.09	0.08
9	VAR009	0.10	0.15	0.21	0.28	0.34	0.38	0.37	0.32	0.25	0.18	0.12	0.08	0.05	0.03	0.02
10	VAR010	0.03	0.04	0.07	0.11	0.16	0.22	0.29	0.34	0.37	0.35	0.30	0.23	0.17	0.11	0.08
11	VAR011	0.04	0.06	0.10	0.16	0.24	0.33	0.42	0.46	0.44	0.37	0.27	0.18	0.12	0.07	0.04
12	VAR012	0.04	0.08	0.14	0.23	0.37	0.52	0.62	0.61	0.49	0.34	0.21	0.12	0.07	0.04	0.02
13	VAR013	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
14	VAR014	0.04	0.05	0.06	0.08	0.10	0.11	0.13	0.14	0.14	0.14	0.13	0.12	0.11	0.09	0.07
15	VAR015	0.09	0.13	0.19	0.26	0.32	0.36	0.37	0.33	0.27	0.20	0.14	0.09	0.06	0.04	0.02
16	VAR016	0.07	0.10	0.13	0.16	0.18	0.20	0.21	0.21	0.19	0.16	0.13	0.10	0.08	0.06	0.04
17	VAR017	0.04	0.06	0.07	0.08	0.10	0.11	0.13	0.13	0.13	0.13	0.12	0.11	0.09	0.08	0.07
18	VAR018	0.02	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
19	VAR019	0.01	0.02	0.03	0.05	0.08	0.13	0.21	0.31	0.42	0.50	0.50	0.43	0.32	0.22	0.14
20	VAR020	0.06	0.08	0.12	0.17	0.23	0.28	0.31	0.32	0.30	0.25	0.19	0.14	0.10	0.07	0.04
21	VAR021	0.03	0.04	0.05	0.06	0.06	0.07	0.08	0.09	0.09	0.09	0.09	0.08	0.08	0.07	0.06
22	VAR022	0.03	0.03	0.04	0.05	0.07	0.08	0.09	0.11	0.12	0.13	0.13	0.13	0.12	0.11	0.10
23	VAR023	0.02	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
24	VAR024	0.01	0.02	0.04	0.08	0.15	0.26	0.41	0.57	0.67	0.63	0.49	0.32	0.19	0.11	0.06
25	VAR025	0.03	0.06	0.10	0.18	0.29	0.44	0.58	0.63	0.56	0.42	0.28	0.17	0.09	0.05	0.03
26	VAR026	0.01	0.03	0.06	0.11	0.22	0.41	0.65	0.85	0.85	0.65	0.41	0.22	0.11	0.06	0.03
27	VAR027	0.03	0.04	0.05	0.07	0.08	0.10	0.11	0.13	0.13	0.14	0.13	0.13	0.11	0.10	0.08
28	VAR028	0.04	0.05	0.05	0.06	0.07	0.08	0.08	0.09	0.09	0.09	0.09	0.08	0.07	0.07	0.06
29	VAR029	0.04	0.06	0.08	0.10	0.13	0.17	0.20	0.22	0.22	0.22	0.19	0.16	0.13	0.10	0.08
30	VAR030	0.05	0.06	0.07	0.08	0.09	0.10	0.11	0.11	0.11	0.10	0.10	0.09	0.08	0.06	0.05
	Test Information:	2.12	2.57	3.21	4.11	5.30	6.70	8.03	8.75	8.47	7.38	5.98	4.69	3.68	2.93	2.41
	Expected s.e.:	0.69	0.62	0.56	0.49	0.43	0.39	0.35	0.34	0.34	0.37	0.41	0.46	0.52	0.58	0.64

Marginal Reliability for Response Pattern Scores: 0.84

Likelihood-based Values and Goodness of Fit Statistics **(Back to TOC)**

Statistics based on the loglikelihood	
-2loglikelihood:	23988.82
Akaike Information Criterion (AIC):	24108.82
Bayesian Information Criterion (BIC):	24378.18

Statistics based on the full item x item x ... classification
The table is too sparse to compute the general multinomial goodness of fit statistics.

Statistics based on one-and two-way marginal tables
M_2 statistics not available for this estimation method.

Summary of the Data and Control Parameters **(Back to TOC)**

Sample Size	658
Number of Items	30
Number of Dimensions	1

Item	Label	Categories	Model
1	VAR001	2	2PL
2	VAR002	2	2PL
3	VAR003	2	2PL
4	VAR004	2	2PL
5	VAR005	2	2PL
6	VAR006	2	2PL
7	VAR007	2	2PL
8	VAR008	2	2PL
9	VAR009	2	2PL
10	VAR010	2	2PL
11	VAR011	2	2PL
12	VAR012	2	2PL
13	VAR013	2	2PL
14	VAR014	2	2PL
15	VAR015	2	2PL
16	VAR016	2	2PL
17	VAR017	2	2PL
18	VAR018	2	2PL
19	VAR019	2	2PL
20	VAR020	2	2PL
21	VAR021	2	2PL
22	VAR022	2	2PL
23	VAR023	2	2PL
24	VAR024	2	2PL
25	VAR025	2	2PL
26	VAR026	2	2PL
27	VAR027	2	2PL
28	VAR028	2	2PL
29	VAR029	2	2PL
30	VAR030	2	2PL

Parameter Estimation Control Values

Bock-Aitkin EM Algorithm		
Maximum number of cycles:	500	
Convergence criterion:	1.00e-005	
Maximum number of M-step iterations:	50	
Convergence criterion for iterative M-steps:	1.00e-006	
Number of rectangular quadrature points:	49	
Minimum, Maximum quadrature points:	-6.00	6.00
SEM algorithm tolerance:	1.00e-003	
Standard error computation algorithm:	Supplemented EM	

Miscellaneous Control Values

Print parameter numbers?	Yes
Z tolerance, max. abs. logit value:	50.00
Number of processor cores used:	1
Number of cycles completed:	49
Maximum parameter change:	0.00e+000
Number of free parameters:	60

Processing times (in seconds)

E-step computations:	0.09
M-step computations:	0.05
Standard error computations:	1.63

Goodness-of-fit statistics:	0.19
Total:	1.96

Output Files

HTML results and control parameters:	C: \Users\SURFACE\Desktop\phung\dataIRTPRO_bi658.Test1-irt.htm

Convergence and Numerical Stability

Engine status:	Normal termination
SEM algorithm status:	Normal
First-order test:	Convergence criteria satisfied
Condition number of information matrix:	1.18e+001
Second-order test:	Solution is a possible local maximum

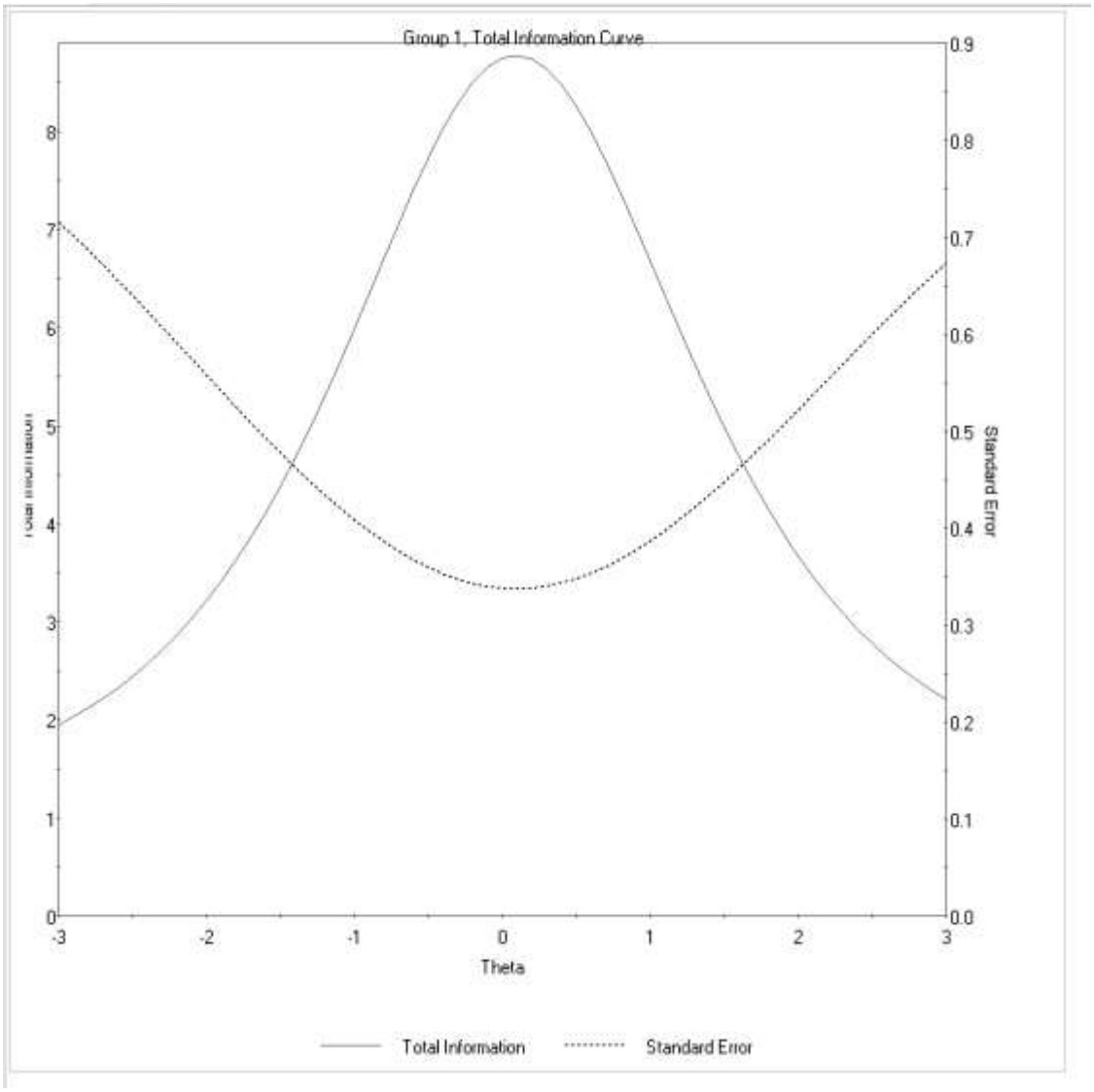
Summed Score to Scale Score Conversion Table **(Back to TOC)**

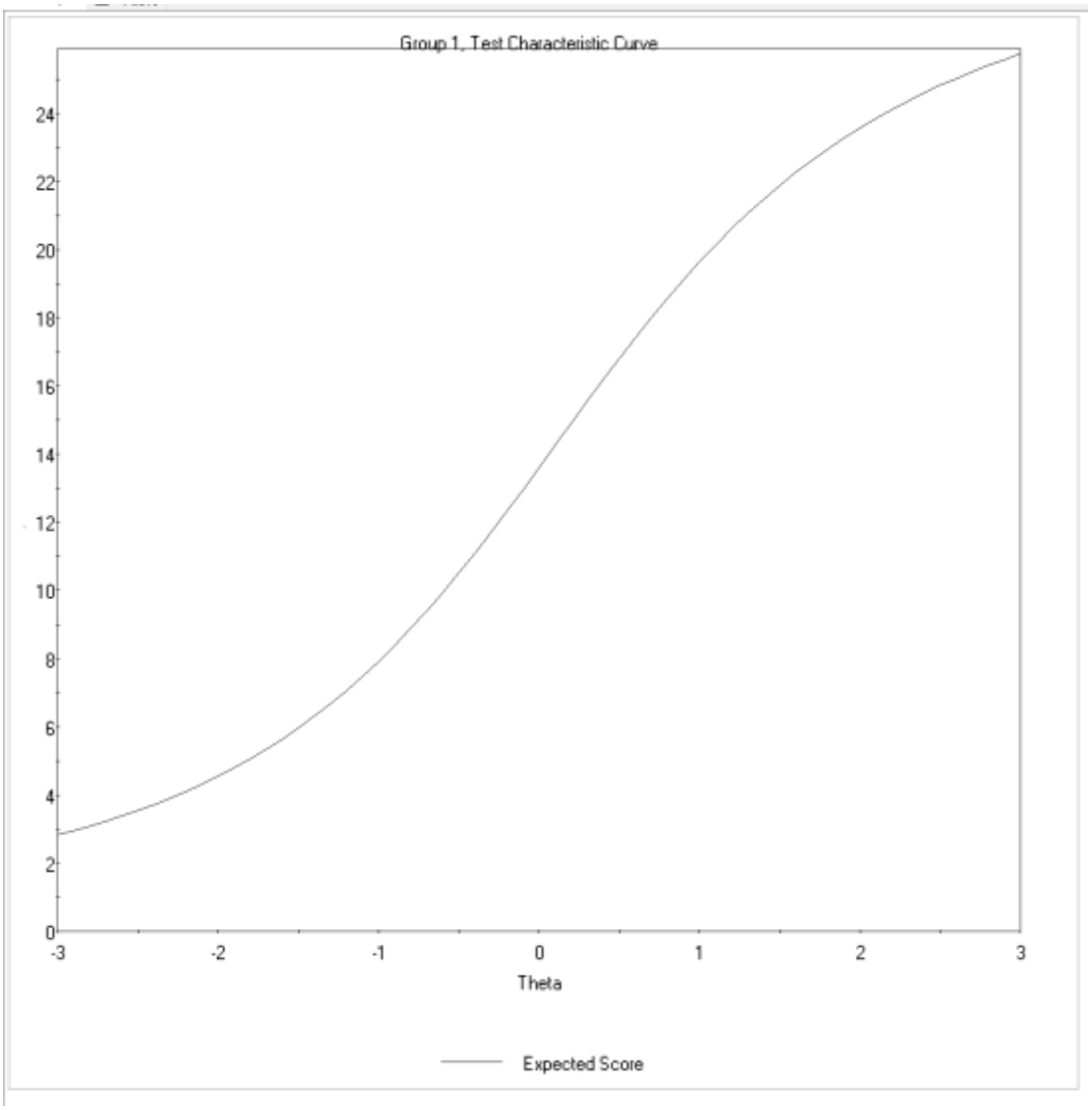
Summed Score	EAP[θ|x]	SD[θ|x]	Modeled Proportion
0	-2.414	0.615	0.0004730
1	-2.214	0.599	0.0025650
2	-2.007	0.578	0.0072022
3	-1.799	0.555	0.0142323
4	-1.595	0.530	0.0225895
5	-1.397	0.505	0.0310184
6	-1.208	0.482	0.0385975
7	-1.027	0.461	0.0448657
8	-0.855	0.442	0.0497184
9	-0.692	0.426	0.0532566
10	-0.535	0.413	0.0556670
11	-0.384	0.403	0.0571492
12	-0.237	0.396	0.0578771
13	-0.093	0.391	0.0579829
14	0.049	0.389	0.0575517
15	0.191	0.390	0.0566217
16	0.333	0.393	0.0551881
17	0.477	0.399	0.0532082
18	0.624	0.407	0.0506092
19	0.775	0.418	0.0473021
20	0.930	0.432	0.0432021
21	1.091	0.448	0.0382625
22	1.258	0.466	0.0325206
23	1.431	0.486	0.0261545
24	1.609	0.507	0.0195290
25	1.792	0.528	0.0131941
26	1.978	0.549	0.0077861
27	2.166	0.570	0.0038177
28	2.352	0.589	0.0014422
29	2.535	0.606	0.0003684
30	2.711	0.620	0.0000470

Marginal reliability of the scaled scores for summed scores = 0.81038

Summary of the Data and Control Parameters **(Back to TOC)**

Sample Size	658
Number of Items	30
Number of Dimensions	1


Group 1, Total Information Curve
Standard Error
Theta
Total Information
Standard Error


Group 1, Test Characteristic Curve
24
22
20
18
16
14
12
10
8
6
4
2
0
-3
-2
-1
0
1
2
3
Theta
Expected Score

ภาคผนวก จ

ผลการวิเคราะห์ฟังก์ชันสารสนเทศของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาเคมี
เรื่อง ปริมาณสารสัมพันธ์ เป็นแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ

IRTPRO Version 4.2
Output generated by IRTPRO estimation engine Version 5.20 (64-bit)

Project:	
Description:	
Date:	15 April 2020
Time:	02: 58 PM

Table of Contents

Nominal Model Slopes and Scoring Function Contrasts for Group 1, logit: ($a\,s_k\theta + c_k$); **s = Tα**
Nominal Model Scoring Function Values for Group 1, logit: ($a\,s_k\theta + c_k$); **s = Tα**
Nominal Model Intercept Contrasts for Group 1, logit: ($a\,s_k\theta + c_k$); **c = Tγ**
Original (Bock, 1972) Parameters, Nominal Items for Group 1, logit: ($a_k\theta + c_k$)
Summed-Score Based Item Diagnostic Tables and X^2s for Group 1
Group Parameter Estimates
Summed Score to Scale Score Conversion Table
Marginal fit (X^2) and Standardized LD X^2 Statistics for Group 1
Item Information Function Values for Group 1 at 15 Values of θ from -2.8 to 2.8
Likelihood-based Values and Goodness of Fit Statistics
Summary of the Data and Control Parameters

GPC Model Item Parameter Estimates, logit: a[k (θ-b) + Σd_k]

Item	Label		a	s.e.	b	s.e.	d_1	d_2	s.e.	d_3	s.e.	d_4	s.e.
1	VAR001	1	0.14	0.04	2.10	0.59	0.00	-15.13	4.30	16.57	4.61	-1.44	0.74
2	VAR002	5	0.36	0.05	-0.18	0.12	0.00	-2.63	0.60	4.18	0.72	-1.55	0.28
3	VAR003	9	0.16	0.04	4.60	1.14	0.00	-13.19	3.69	12.49	3.49	0.70	0.88
4	VAR004	13	0.23	0.04	1.66	0.31	0.00	-3.85	0.90	3.89	0.91	-0.03	0.46
5	VAR005	17	0.56	0.06	0.95	0.11	0.00	-1.71	0.34	2.36	0.36	-0.65	0.19
6	VAR006	21	0.73	0.07	1.01	0.10	0.00	-0.91	0.24	2.09	0.29	-1.18	0.16
7	VAR007	25	0.47	0.05	1.13	0.15	0.00	-1.99	0.42	3.44	0.51	-1.45	0.24
8	VAR008	29	0.23	0.04	3.14	0.55	0.00	-8.07	1.76	5.39	1.36	2.68	0.88
9	VAR009	33	0.50	0.06	0.19	0.09	0.00	-1.54	0.34	1.71	0.34	-0.16	0.20
10	VAR010	37	0.55	0.06	1.27	0.13	0.00	-1.87	0.37	1.83	0.35	0.05	0.24
11	VAR011	41	0.64	0.06	0.86	0.10	0.00	-2.11	0.37	2.70	0.38	-0.59	0.17
12	VAR012	45	0.61	0.06	0.51	0.08	0.00	-1.62	0.31	1.78	0.30	-0.16	0.17
13	VAR013	49	0.15	0.04	2.68	0.68	0.00	-11.37	3.10	10.11	2.80	1.26	0.83
14	VAR014	53	0.26	0.04	1.97	0.31	0.00	-5.76	1.14	5.29	1.06	0.47	0.47
15	VAR015	57	0.45	0.05	0.71	0.12	0.00	-0.93	0.31	2.13	0.37	-1.20	0.23
16	VAR016	61	0.45	0.05	0.60	0.11	0.00	-2.47	0.46	2.75	0.46	-0.28	0.22
17	VAR017	65	0.40	0.05	1.32	0.16	0.00	-3.00	0.55	3.21	0.55	-0.21	0.28
18	VAR018	69	0.17	0.04	4.36	1.07	0.00	-5.41	1.63	6.76	1.92	-1.36	0.83
19	VAR019	73	0.63	0.07	1.68	0.14	0.00	-2.56	0.47	2.62	0.45	-0.06	0.24
20	VAR020	77	0.38	0.05	1.36	0.18	0.00	-1.89	0.45	3.02	0.53	-1.13	0.30
21	VAR021	81	0.26	0.04	1.98	0.31	0.00	-5.35	1.06	4.31	0.93	1.04	0.51
22	VAR022	85	0.34	0.05	2.22	0.28	0.00	-3.98	0.77	3.50	0.71	0.48	0.42
23	VAR023	89	0.18	0.04	4.18	0.98	0.00	-5.37	1.56	6.92	1.88	-1.55	0.80
24	VAR024	93	0.62	0.06	1.22	0.11	0.00	-2.73	0.46	2.88	0.44	-0.16	0.21
25	VAR025	97	0.60	0.06	0.76	0.09	0.00	-2.09	0.37	2.45	0.37	-0.36	0.18
26	VAR026	101	0.76	0.08	0.83	0.08	0.00	-1.41	0.28	1.36	0.26	0.06	0.18
27	VAR027	105	0.32	0.05	2.08	0.28	0.00	-4.02	0.79	3.84	0.76	0.17	0.41
28	VAR028	109	0.32	0.05	1.97	0.28	0.00	-3.00	0.65	3.61	0.70	-0.61	0.38
29	VAR029	113	0.43	0.05	1.41	0.16	0.00	-1.80	0.40	1.35	0.39	0.45	0.33
30	VAR030	117	0.36	0.05	1.30	0.18	0.00	-2.41	0.51	2.69	0.52	-0.28	0.31

Nominal Model Slopes and Scoring Function Contrasts for Group 1, logit: ($a\ s_k\theta + c_k$); s = Tα **(Back to TOC)**

Item	Label		*a*	*s.e.*	Contrasts	α_1	*s.e.*	α_2	*s.e.*	α_3	*s.e.*
1	VAR001	1	0.14	0.04	Trend	1.00	-----	0.00	-----	0.00	-----
2	VAR002	5	0.36	0.05	Trend	1.00	-----	0.00	-----	0.00	-----
3	VAR003	9	0.16	0.04	Trend	1.00	-----	0.00	-----	0.00	-----
4	VAR004	13	0.23	0.04	Trend	1.00	-----	0.00	-----	0.00	-----
5	VAR005	17	0.56	0.06	Trend	1.00	-----	0.00	-----	0.00	-----
6	VAR006	21	0.73	0.07	Trend	1.00	-----	0.00	-----	0.00	-----
7	VAR007	25	0.47	0.05	Trend	1.00	-----	0.00	-----	0.00	-----
8	VAR008	29	0.23	0.04	Trend	1.00	-----	0.00	-----	0.00	-----
9	VAR009	33	0.50	0.06	Trend	1.00	-----	0.00	-----	0.00	-----
10	VAR010	37	0.55	0.06	Trend	1.00	-----	0.00	-----	0.00	-----
11	VAR011	41	0.64	0.06	Trend	1.00	-----	0.00	-----	0.00	-----
12	VAR012	45	0.61	0.06	Trend	1.00	-----	0.00	-----	0.00	-----
13	VAR013	49	0.15	0.04	Trend	1.00	-----	0.00	-----	0.00	-----
14	VAR014	53	0.26	0.04	Trend	1.00	-----	0.00	-----	0.00	-----
15	VAR015	57	0.45	0.05	Trend	1.00	-----	0.00	-----	0.00	-----
16	VAR016	61	0.45	0.05	Trend	1.00	-----	0.00	-----	0.00	-----
17	VAR017	65	0.40	0.05	Trend	1.00	-----	0.00	-----	0.00	-----
18	VAR018	69	0.17	0.04	Trend	1.00	-----	0.00	-----	0.00	-----
19	VAR019	73	0.63	0.07	Trend	1.00	-----	0.00	-----	0.00	-----
20	VAR020	77	0.38	0.05	Trend	1.00	-----	0.00	-----	0.00	-----
21	VAR021	81	0.26	0.04	Trend	1.00	-----	0.00	-----	0.00	-----
22	VAR022	85	0.34	0.05	Trend	1.00	-----	0.00	-----	0.00	-----
23	VAR023	89	0.18	0.04	Trend	1.00	-----	0.00	-----	0.00	-----
24	VAR024	93	0.62	0.06	Trend	1.00	-----	0.00	-----	0.00	-----
25	VAR025	97	0.60	0.06	Trend	1.00	-----	0.00	-----	0.00	-----
26	VAR026	101	0.76	0.08	Trend	1.00	-----	0.00	-----	0.00	-----
27	VAR027	105	0.32	0.05	Trend	1.00	-----	0.00	-----	0.00	-----
28	VAR028	109	0.32	0.05	Trend	1.00	-----	0.00	-----	0.00	-----
29	VAR029	113	0.43	0.05	Trend	1.00	-----	0.00	-----	0.00	-----
30	VAR030	117	0.36	0.05	Trend	1.00	-----	0.00	-----	0.00	-----

Nominal Model Scoring Function Values for Group 1, logit: ($a\ s_k\theta + c_k$); s = Tα **(Back to TOC)**

Item	Category	s_1	s_2	s_3	s_4
1	VAR001	0.00	1.00	2.00	3.00
2	VAR002	0.00	1.00	2.00	3.00
3	VAR003	0.00	1.00	2.00	3.00
4	VAR004	0.00	1.00	2.00	3.00
5	VAR005	0.00	1.00	2.00	3.00
6	VAR006	0.00	1.00	2.00	3.00
7	VAR007	0.00	1.00	2.00	3.00
8	VAR008	0.00	1.00	2.00	3.00
9	VAR009	0.00	1.00	2.00	3.00
10	VAR010	0.00	1.00	2.00	3.00
11	VAR011	0.00	1.00	2.00	3.00
12	VAR012	0.00	1.00	2.00	3.00
13	VAR013	0.00	1.00	2.00	3.00
14	VAR014	0.00	1.00	2.00	3.00
15	VAR015	0.00	1.00	2.00	3.00
16	VAR016	0.00	1.00	2.00	3.00
17	VAR017	0.00	1.00	2.00	3.00
18	VAR018	0.00	1.00	2.00	3.00

19	VAR019	0.00	1.00	2.00	3.00
20	VAR020	0.00	1.00	2.00	3.00
21	VAR021	0.00	1.00	2.00	3.00
22	VAR022	0.00	1.00	2.00	3.00
23	VAR023	0.00	1.00	2.00	3.00
24	VAR024	0.00	1.00	2.00	3.00
25	VAR025	0.00	1.00	2.00	3.00
26	VAR026	0.00	1.00	2.00	3.00
27	VAR027	0.00	1.00	2.00	3.00
28	VAR028	0.00	1.00	2.00	3.00
29	VAR029	0.00	1.00	2.00	3.00
30	VAR030	0.00	1.00	2.00	3.00

Nominal Model Intercept Contrasts for Group 1, logit: ($a\ s_k\theta + c_k$); c = Tγ **(Back to TOC)**

Item	Label	Contrasts	γ_1	*s.e.*	γ_2	*s.e.*	γ_3	*s.e.*
1	VAR001	Trend	[2] -0.29	0.04	[3] -1.10	0.13	[4] -1.33	0.12
2	VAR002	Trend	[6] 0.06	0.04	[7] -0.22	0.11	[8] -0.86	0.09
3	VAR003	Trend	[10] -0.75	0.05	[11] -1.31	0.15	[12] -1.18	0.13
4	VAR004	Trend	[14] -0.38	0.04	[15] -0.51	0.10	[16] -0.52	0.08
5	VAR005	Trend	[18] -0.53	0.05	[19] -0.34	0.12	[20] -0.77	0.09
6	VAR006	Trend	[22] -0.74	0.07	[23] 0.11	0.13	[24] -0.89	0.09
7	VAR007	Trend	[26] -0.53	0.05	[27] -0.15	0.12	[28] -0.93	0.09
8	VAR008	Trend	[30] -0.73	0.05	[31] -1.45	0.14	[32] -0.73	0.12
9	VAR009	Trend	[34] -0.09	0.04	[35] -0.40	0.10	[36] -0.49	0.08
10	VAR010	Trend	[38] -0.69	0.06	[39] -0.61	0.12	[40] -0.58	0.09
11	VAR011	Trend	[42] -0.55	0.06	[43] -0.56	0.13	[44] -0.99	0.10
12	VAR012	Trend	[46] -0.31	0.05	[47] -0.51	0.11	[48] -0.63	0.09
13	VAR013	Trend	[50] -0.39	0.04	[51] -1.06	0.12	[52] -0.85	0.10
14	VAR014	Trend	[54] -0.51	0.04	[55] -0.93	0.12	[56] -0.79	0.10
15	VAR015	Trend	[58] -0.32	0.05	[59] 0.07	0.11	[60] -0.56	0.07
16	VAR016	Trend	[62] -0.27	0.04	[63] -0.57	0.11	[64] -0.71	0.09
17	VAR017	Trend	[66] -0.53	0.05	[67] -0.65	0.12	[68] -0.75	0.09
18	VAR018	Trend	[70] -0.75	0.05	[71] -0.40	0.12	[72] -0.67	0.09
19	VAR019	Trend	[74] -1.06	0.08	[75] -0.92	0.16	[76] -0.96	0.12
20	VAR020	Trend	[78] -0.51	0.05	[79] -0.16	0.11	[80] -0.66	0.08
21	VAR021	Trend	[82] -0.52	0.04	[83] -0.98	0.12	[84] -0.66	0.10
22	VAR022	Trend	[86] -0.76	0.05	[87] -0.88	0.13	[88] -0.69	0.10
23	VAR023	Trend	[90] -0.75	0.05	[91] -0.39	0.12	[92] -0.72	0.09
24	VAR024	Trend	[94] -0.76	0.07	[95] -0.92	0.15	[96] -1.03	0.12
25	VAR025	Trend	[98] -0.46	0.05	[99] -0.60	0.12	[100] -0.86	0.10
26	VAR026	Trend	[102] -0.63	0.06	[103] -0.64	0.13	[104] -0.59	0.10
27	VAR027	Trend	[106] -0.66	0.05	[107] -0.77	0.12	[108] -0.70	0.10
28	VAR028	Trend	[110] -0.62	0.05	[111] -0.44	0.12	[112] -0.66	0.09
29	VAR029	Trend	[114] -0.61	0.05	[115] -0.56	0.12	[116] -0.34	0.09
30	VAR030	Trend	[118] -0.47	0.05	[119] -0.45	0.11	[120] -0.57	0.08

Original (Bock, 1972) Parameters, Nominal Items for Group 1, logit: ($a_k\theta + c_k$) **(Back to TOC)**

Item	Label	Category	1	2	3	4
1	VAR001	*a*	0.00	0.14	0.28	0.42
		c	-0.00	-2.40	-0.38	-0.88
2	VAR002	*a*	0.00	0.36	0.71	1.07
		c	0.00	-0.87	0.68	0.19
3	VAR003	*a*	0.00	0.16	0.33	0.49
		c	-0.00	-2.91	-1.62	-2.26
4	VAR004	*a*	0.00	0.23	0.46	0.70

		c	-0.00	-1.28	-0.76	-1.15
5	VAR005	*a*	0.00	0.56	1.12	1.69
		c	-0.00	-1.49	-0.70	-1.60
6	VAR006	*a*	0.00	0.73	1.47	2.20
		c	0.00	-1.41	-0.62	-2.22
7	VAR007	*a*	0.00	0.47	0.93	1.40
		c	-0.00	-1.45	-0.38	-1.58
8	VAR008	*a*	0.00	0.23	0.47	0.70
		c	-0.00	-2.61	-2.09	-2.20
9	VAR009	*a*	0.00	0.50	0.99	1.49
		c	-0.00	-0.86	-0.11	-0.28
10	VAR010	*a*	0.00	0.55	1.09	1.64
		c	-0.00	-1.71	-1.41	-2.07
11	VAR011	*a*	0.00	0.64	1.27	1.91
		c	-0.00	-1.89	-0.71	-1.64
12	VAR012	*a*	0.00	0.61	1.23	1.84
		c	-0.00	-1.30	-0.52	-0.93
13	VAR013	*a*	0.00	0.15	0.29	0.44
		c	-0.00	-2.05	-0.96	-1.17
14	VAR014	*a*	0.00	0.26	0.52	0.78
		c	-0.00	-2.00	-1.14	-1.53
15	VAR015	*a*	0.00	0.45	0.91	1.36
		c	0.00	-0.75	-0.10	-0.97
16	VAR016	*a*	0.00	0.45	0.90	1.35
		c	-0.00	-1.38	-0.41	-0.81
17	VAR017	*a*	0.00	0.40	0.81	1.21
		c	-0.00	-1.74	-0.98	-1.60
18	VAR018	*a*	0.00	0.17	0.34	0.52
		c	-0.00	-1.68	-1.27	-2.25
19	VAR019	*a*	0.00	0.63	1.27	1.90
		c	-0.00	-2.68	-2.09	-3.19
20	VAR020	*a*	0.00	0.38	0.76	1.13
		c	-0.00	-1.23	-0.60	-1.54
21	VAR021	*a*	0.00	0.26	0.53	0.79
		c	-0.00	-1.94	-1.32	-1.57
22	VAR022	*a*	0.00	0.34	0.69	1.03
		c	-0.00	-2.13	-1.69	-2.29
23	VAR023	*a*	0.00	0.18	0.36	0.54
		c	-0.00	-1.71	-1.22	-2.25
24	VAR024	*a*	0.00	0.62	1.24	1.85
		c	-0.00	-2.44	-1.42	-2.27
25	VAR025	*a*	0.00	0.60	1.21	1.81
		c	-0.00	-1.72	-0.70	-1.38
26	VAR026	*a*	0.00	0.76	1.51	2.27
		c	-0.00	-1.70	-1.30	-1.89
27	VAR027	*a*	0.00	0.32	0.63	0.95
		c	-0.00	-1.93	-1.37	-1.98
28	VAR028	*a*	0.00	0.32	0.63	0.95
		c	-0.00	-1.57	-1.05	-1.86
29	VAR029	*a*	0.00	0.43	0.86	1.29
		c	-0.00	-1.38	-1.41	-1.82
30	VAR030	*a*	0.00	0.36	0.73	1.09
		c	-0.00	-1.35	-0.85	-1.42

Summed-Score Based Item Diagnostic Tables and X^2s for Group 1 **(Back to TOC)**

S-X^2 Item Level Diagnostic Statistics

Item	Label	X^2	*d.f.*	Probability
1	VAR001	174.86	118	0.0005
2	VAR002	148.69	125	0.0728
3	VAR003	134.17	98	0.0090
4	VAR004	189.28	137	0.0021
5	VAR005	144.86	125	0.1080
6	VAR006	124.99	112	0.1890
7	VAR007	148.32	120	0.0406
8	VAR008	95.70	103	0.6829
9	VAR009	149.64	122	0.0451
10	VAR010	130.54	122	0.2818
11	VAR011	126.43	113	0.1828
12	VAR012	128.94	123	0.3385
13	VAR013	130.34	129	0.4509
14	VAR014	140.28	123	0.1364
15	VAR015	145.33	131	0.1850
16	VAR016	121.76	128	0.6391
17	VAR017	153.30	127	0.0559
18	VAR018	132.84	115	0.1220
19	VAR019	103.53	89	0.1389
20	VAR020	152.56	131	0.0957
21	VAR021	150.27	124	0.0542
22	VAR022	112.95	115	0.5372
23	VAR023	116.72	117	0.4906
24	VAR024	127.46	104	0.0589
25	VAR025	151.95	120	0.0258
26	VAR026	129.97	111	0.1052
27	VAR027	133.73	119	0.1681
28	VAR028	154.78	125	0.0364
29	VAR029	149.09	130	0.1206
30	VAR030	182.12	134	0.0036

Group Parameter Estimates **(Back to TOC)**

Group	Label	μ	*s.e.*	σ^2	*s.e.*	σ	*s.e.*
1	Group 1	0.00	-----	1.00	-----	1.00	-----

Summed Score to Scale Score Conversion Table **(Back to TOC)**

Summed Score	EAP[θ|x]	SD[θ|x]	Modeled Proportion

Marginal reliability of the scaled scores for summed scores = 1.00000

Marginal fit (X^2) and Standardized LD X^2 Statistics for Group 1 **(Back to TOC)**

Item	Label	Marginal X^2	1	2	3	4	5	6	7	8	9	10
1	VAR001	0.0										
2	VAR002	0.0	71.5									
3	VAR003	0.0	8.5	5.0								
4	VAR004	0.0	12.5	14.6	6.5							
5	VAR005	0.1	7.3	6.6	3.0	12.8						
6	VAR006	0.2	4.0	4.6	2.9	5.5	13.0					
7	VAR007	0.0	13.0	10.3	2.3	10.5	3.3	1.9				
8	VAR008	0.0	1.7	4.4	8.1	2.8	2.7	1.5	4.1			
9	VAR009	0.1	7.9	9.3	5.5	4.8	7.1	8.0	10.4	6.1		

10	VAR010	0.1	-0.5	5.2	0.3	3.1	1.7	3.7	7.6	7.5	4.6
11	VAR011	0.2	8.4	7.8	6.9	5.9	4.2	11.0	2.0	4.6	14.0	2.3
12	VAR012	0.2	7.9	2.5	3.4	2.3	6.4	7.5	3.9	4.1	14.4	6.5
13	VAR013	0.0	4.0	6.4	1.1	6.0	4.6	3.8	8.5	1.9	3.8	3.9
14	VAR014	0.0	4.6	3.9	2.8	3.2	8.4	3.6	3.7	3.3	8.9	3.6
15	VAR015	0.0	7.0	8.3	6.4	8.7	4.3	13.5	6.4	8.6	10.8	14.2
16	VAR016	0.0	8.5	6.7	3.8	6.4	3.9	4.7	3.2	3.7	10.6	4.3
17	VAR017	0.0	1.5	1.8	2.2	5.1	0.8	2.1	3.4	0.7	7.7	3.5
18	VAR018	0.0	9.3	5.0	1.7	3.0	4.3	6.0	1.7	4.4	2.2	0.5
19	VAR019	0.1	2.1	1.9	1.8	11.1	1.8	4.9	4.8	0.9	4.0	4.4
20	VAR020	0.0	3.7	6.4	1.8	4.2	5.9	10.1	7.5	3.2	10.9	3.5
21	VAR021	0.0	1.0	2.3	1.0	3.8	1.9	11.9	3.6	5.2	12.5	3.6
22	VAR022	0.0	5.7	4.0	3.3	3.9	8.4	2.9	6.8	0.5	3.3	8.7
23	VAR023	0.0	5.3	11.7	5.6	6.6	1.2	3.6	4.7	2.7	3.8	2.4
24	VAR024	0.1	4.9	4.3	1.7	2.1	2.4	15.6	7.9	3.5	2.8	2.2
25	VAR025	0.1	4.6	2.2	1.9	8.4	6.6	6.6	4.0	1.1	10.3	1.0
26	VAR026	0.2	8.0	10.0	4.8	4.8	5.6	11.4	9.1	4.9	9.1	3.8
27	VAR027	0.0	5.3	5.0	6.8	9.4	6.3	2.0	4.4	2.4	4.1	2.5
28	VAR028	0.0	2.3	4.9	5.1	3.1	-0.5	-0.3	8.3	2.2	8.9	6.3
29	VAR029	0.0	3.7	12.7	3.3	3.8	11.0	3.4	6.1	9.0	5.9	3.5
30	VAR030	0.0	1.8	8.6	4.3	5.4	1.1	2.1	4.9	4.1	6.1	4.6

Item	Label	Marginal X^2	11	12	13	14	15	16	17	18	19	20
11	VAR011	0.2										
12	VAR012	0.2	9.1									
13	VAR013	0.0	0.1	6.6								
14	VAR014	0.0	1.8	12.1	3.6							
15	VAR015	0.0	7.0	9.8	4.8	13.2						
16	VAR016	0.0	14.7	5.0	3.1	4.1	4.8					
17	VAR017	0.0	6.0	3.7	2.6	-0.1	7.9	9.1				
18	VAR018	0.0	5.3	20.0	3.8	0.8	10.8	7.0	1.1			
19	VAR019	0.1	2.7	7.0	5.6	1.6	2.1	0.7	1.9	2.5		
20	VAR020	0.0	6.7	4.7	2.0	8.4	13.8	3.6	4.7	1.9	3.1	
21	VAR021	0.0	5.1	9.6	2.4	6.3	8.3	2.4	1.0	3.1	1.4	11.0
22	VAR022	0.0	4.7	12.1	3.9	2.8	3.3	2.0	3.3	5.2	1.8	1.7
23	VAR023	0.0	7.4	5.7	4.0	9.0	4.6	5.7	5.5	7.0	1.5	9.4
24	VAR024	0.1	-0.2	6.1	9.2	7.4	8.3	0.8	2.3	0.7	8.0	6.5
25	VAR025	0.1	1.7	5.7	4.2	3.0	6.8	7.9	2.9	2.8	0.7	9.1
26	VAR026	0.2	10.0	6.2	2.2	2.6	11.0	2.5	1.2	5.8	3.4	16.6
27	VAR027	0.0	5.0	16.1	5.8	1.1	8.6	3.9	4.8	7.3	0.7	7.5
28	VAR028	0.0	2.9	2.9	5.2	5.8	6.8	4.7	2.1	1.4	3.6	8.3
29	VAR029	0.0	9.0	2.3	5.3	10.6	10.4	4.1	4.1	2.0	2.8	10.7
30	VAR030	0.0	5.5	4.3	8.6	5.1	7.8	7.4	2.5	2.1	5.1	1.0

Item	Label	Marginal X^2	21	22	23	24	25	26	27	28	29
21	VAR021	0.0									
22	VAR022	0.0	16.5								
23	VAR023	0.0	4.5	1.4							
24	VAR024	0.1	0.2	1.7	3.2						
25	VAR025	0.1	2.7	2.7	1.1	1.1					
26	VAR026	0.2	7.0	3.6	1.7	9.5	5.9				
27	VAR027	0.0	4.0	4.6	2.0	-0.4	1.6	10.9			
28	VAR028	0.0	1.6	3.5	11.3	5.3	4.5	2.3	9.3		

29	VAR029	0.0	4.7	2.5	15.0	4.4	7.5	6.1	6.7	8.1	
30	VAR030	0.0	2.1	-0.3	8.1	3.5	3.9	9.0	3.1	14.7	12.5

Item Information Function Values for Group 1 at 15 Values of θ from -2.8 to 2.8 **(Back to TOC)**

Item	Label	θ: -2.8	-2.4	-2.0	-1.6	-1.2	-0.8	-0.4	-0.0	0.4	0.8	1.2	1.6	2.0	2.4	2.8
1	VAR001	0.02	0.02	0.02	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
2	VAR002	0.11	0.12	0.14	0.15	0.16	0.16	0.16	0.15	0.13	0.12	0.10	0.09	0.08	0.07	0.06
3	VAR003	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.03	0.04	0.04	0.04	0.04
4	VAR004	0.03	0.04	0.04	0.05	0.06	0.06	0.07	0.07	0.07	0.08	0.08	0.08	0.07	0.07	0.07
5	VAR005	0.04	0.06	0.09	0.12	0.18	0.24	0.31	0.38	0.42	0.41	0.37	0.31	0.25	0.19	0.15
6	VAR006	0.03	0.05	0.08	0.13	0.21	0.31	0.44	0.55	0.61	0.59	0.51	0.41	0.31	0.24	0.19
7	VAR007	0.05	0.07	0.10	0.13	0.16	0.20	0.23	0.26	0.27	0.26	0.24	0.21	0.18	0.15	0.13
8	VAR008	0.01	0.02	0.02	0.03	0.03	0.04	0.04	0.05	0.06	0.07	0.08	0.08	0.09	0.09	0.10
9	VAR009	0.08	0.11	0.15	0.20	0.25	0.30	0.33	0.34	0.33	0.29	0.24	0.19	0.15	0.11	0.09
10	VAR010	0.03	0.04	0.05	0.08	0.11	0.16	0.23	0.30	0.38	0.43	0.43	0.39	0.33	0.26	0.19
11	VAR011	0.03	0.05	0.08	0.13	0.19	0.29	0.40	0.50	0.55	0.53	0.46	0.36	0.27	0.20	0.15
12	VAR012	0.05	0.07	0.11	0.17	0.24	0.34	0.44	0.52	0.53	0.48	0.39	0.30	0.22	0.16	0.11
13	VAR013	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
14	VAR014	0.03	0.04	0.04	0.05	0.06	0.07	0.08	0.09	0.09	0.10	0.10	0.10	0.10	0.10	0.09
15	VAR015	0.07	0.09	0.12	0.15	0.18	0.21	0.23	0.24	0.24	0.23	0.21	0.19	0.16	0.13	0.11
16	VAR016	0.06	0.08	0.11	0.14	0.18	0.23	0.26	0.28	0.29	0.27	0.24	0.20	0.16	0.13	0.10
17	VAR017	0.04	0.05	0.07	0.09	0.11	0.14	0.17	0.20	0.22	0.24	0.23	0.22	0.20	0.17	0.14
18	VAR018	0.02	0.02	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.03	0.03	0.04	0.04	0.04	0.04
19	VAR019	0.01	0.02	0.03	0.04	0.06	0.10	0.16	0.25	0.37	0.50	0.59	0.60	0.52	0.40	0.29
20	VAR020	0.05	0.06	0.07	0.09	0.11	0.13	0.15	0.17	0.18	0.18	0.18	0.17	0.15	0.14	0.12
21	VAR021	0.03	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.10	0.11	0.11	0.11	0.11	0.10
22	VAR022	0.02	0.03	0.03	0.04	0.06	0.07	0.09	0.11	0.13	0.15	0.17	0.18	0.18	0.18	0.16
23	VAR023	0.02	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.03	0.04	0.04	0.04	0.04	0.04	0.04
24	VAR024	0.02	0.03	0.05	0.07	0.11	0.18	0.27	0.39	0.50	0.57	0.56	0.48	0.37	0.27	0.19
25	VAR025	0.04	0.06	0.09	0.13	0.20	0.29	0.39	0.47	0.52	0.49	0.42	0.33	0.25	0.18	0.13
26	VAR026	0.02	0.03	0.06	0.09	0.16	0.26	0.43	0.63	0.80	0.84	0.71	0.52	0.35	0.22	0.15
27	VAR027	0.03	0.03	0.04	0.05	0.06	0.08	0.09	0.11	0.12	0.14	0.15	0.15	0.15	0.14	0.13
28	VAR028	0.03	0.04	0.05	0.06	0.07	0.08	0.10	0.11	0.12	0.13	0.14	0.14	0.13	0.13	0.12
29	VAR029	0.03	0.04	0.06	0.07	0.10	0.13	0.16	0.20	0.24	0.26	0.27	0.27	0.24	0.21	0.17
30	VAR030	0.04	0.05	0.07	0.08	0.10	0.12	0.15	0.16	0.18	0.19	0.18	0.17	0.16	0.14	0.12
	Test Information:	2.07	2.42	2.88	3.51	4.33	5.38	6.59	7.77	8.61	8.82	8.34	7.42	6.36	5.37	4.54
	Expected s.e.:	0.69	0.64	0.59	0.53	0.48	0.43	0.39	0.36	0.34	0.34	0.35	0.37	0.40	0.43	0.47

Marginal Reliability for Response Pattern Scores: 0.84

Likelihood-based Values and Goodness of Fit Statistics **(Back to TOC)**

Statistics based on the loglikelihood	
-2loglikelihood:	42119.31
Akaike Information Criterion (AIC):	42359.31
Bayesian Information Criterion (BIC):	42898.01

Statistics based on the full item x item x ... classification
The table is too sparse to compute the general multinomial goodness of fit statistics.

Statistics based on one-and two-way marginal tables
M_2 statistics not available for this estimation method.

Summary of the Data and Control Parameters **(Back to TOC)**

Sample Size	658
Number of Items	30
Number of Dimensions	1

Item	Label	Categories	Model	T_a	T_c
1	VAR001	4	Nominal	Trend	Trend
2	VAR002	4	Nominal	Trend	Trend
3	VAR003	4	Nominal	Trend	Trend
4	VAR004	4	Nominal	Trend	Trend
5	VAR005	4	Nominal	Trend	Trend
6	VAR006	4	Nominal	Trend	Trend
7	VAR007	4	Nominal	Trend	Trend
8	VAR008	4	Nominal	Trend	Trend
9	VAR009	4	Nominal	Trend	Trend
10	VAR010	4	Nominal	Trend	Trend
11	VAR011	4	Nominal	Trend	Trend
12	VAR012	4	Nominal	Trend	Trend
13	VAR013	4	Nominal	Trend	Trend
14	VAR014	4	Nominal	Trend	Trend
15	VAR015	4	Nominal	Trend	Trend
16	VAR016	4	Nominal	Trend	Trend
17	VAR017	4	Nominal	Trend	Trend
18	VAR018	4	Nominal	Trend	Trend
19	VAR019	4	Nominal	Trend	Trend
20	VAR020	4	Nominal	Trend	Trend
21	VAR021	4	Nominal	Trend	Trend
22	VAR022	4	Nominal	Trend	Trend
23	VAR023	4	Nominal	Trend	Trend
24	VAR024	4	Nominal	Trend	Trend
25	VAR025	4	Nominal	Trend	Trend
26	VAR026	4	Nominal	Trend	Trend
27	VAR027	4	Nominal	Trend	Trend
28	VAR028	4	Nominal	Trend	Trend
29	VAR029	4	Nominal	Trend	Trend
30	VAR030	4	Nominal	Trend	Trend

Parameter Estimation Control Values

Bock-Aitkin EM Algorithm		
Maximum number of cycles:	500	
Convergence criterion:	1.00e-005	
Maximum number of M-step iterations:	50	
Convergence criterion for iterative M-steps:	1.00e-006	
Number of rectangular quadrature points:	49	
Minimum, Maximum quadrature points:	-6.00	6.00
SEM algorithm tolerance:	1.00e-003	
Standard error computation algorithm:	Supplemented EM	

Miscellaneous Control Values

Print parameter numbers?	Yes
Z tolerance, max. abs. logit value:	50.00
Number of processor cores used:	1
Number of cycles completed:	70
Maximum parameter change:	0.00e+000
Number of free parameters:	120

Processing times (in seconds)

E-step computations:	0.11
M-step computations:	0.24
Standard error computations:	2.20
Goodness-of-fit statistics:	0.16
Total:	2.71

Output Files

HTML results and control parameters:	C: \Users\SURFACE\Desktop\phung\dataIRTPRO_poly658.Test1-irt.htm

Convergence and Numerical Stability

Engine status:	Normal termination
SEM algorithm status:	Normal
First-order test:	Convergence criteria satisfied
Condition number of information matrix:	3.15e+001
Second-order test:	Solution is a possible local maximum

Summed Score to Scale Score Conversion Table **(Back to TOC)**

Summed Score	EAP[θ|x]	SD[θ|x]	Modeled Proportion
0	-2.474	0.644	0.0002493
1	-2.326	0.628	0.0005920
2	-2.206	0.618	0.0013680
3	-2.086	0.606	0.0024116
4	-1.970	0.594	0.0038578
5	-1.858	0.582	0.0056095
6	-1.749	0.571	0.0076171
7	-1.643	0.559	0.0098036
8	-1.540	0.548	0.0120702
9	-1.439	0.536	0.0143414
10	-1.342	0.525	0.0165395
11	-1.247	0.514	0.0186069
12	-1.155	0.504	0.0205000
13	-1.065	0.494	0.0221888
14	-0.978	0.484	0.0236574
15	-0.894	0.475	0.0248999
16	-0.812	0.466	0.0259188
17	-0.732	0.457	0.0267228
18	-0.654	0.449	0.0273246
19	-0.578	0.442	0.0277398
20	-0.504	0.435	0.0279850
21	-0.432	0.428	0.0280776
22	-0.361	0.422	0.0280344
23	-0.292	0.416	0.0278718
24	-0.225	0.410	0.0276050
25	-0.158	0.405	0.0272478
26	-0.093	0.401	0.0268131
27	-0.030	0.396	0.0263122
28	0.033	0.392	0.0257553
29	0.095	0.389	0.0251514
30	0.156	0.386	0.0245083
31	0.216	0.383	0.0238328
32	0.275	0.380	0.0231310
33	0.334	0.378	0.0224079

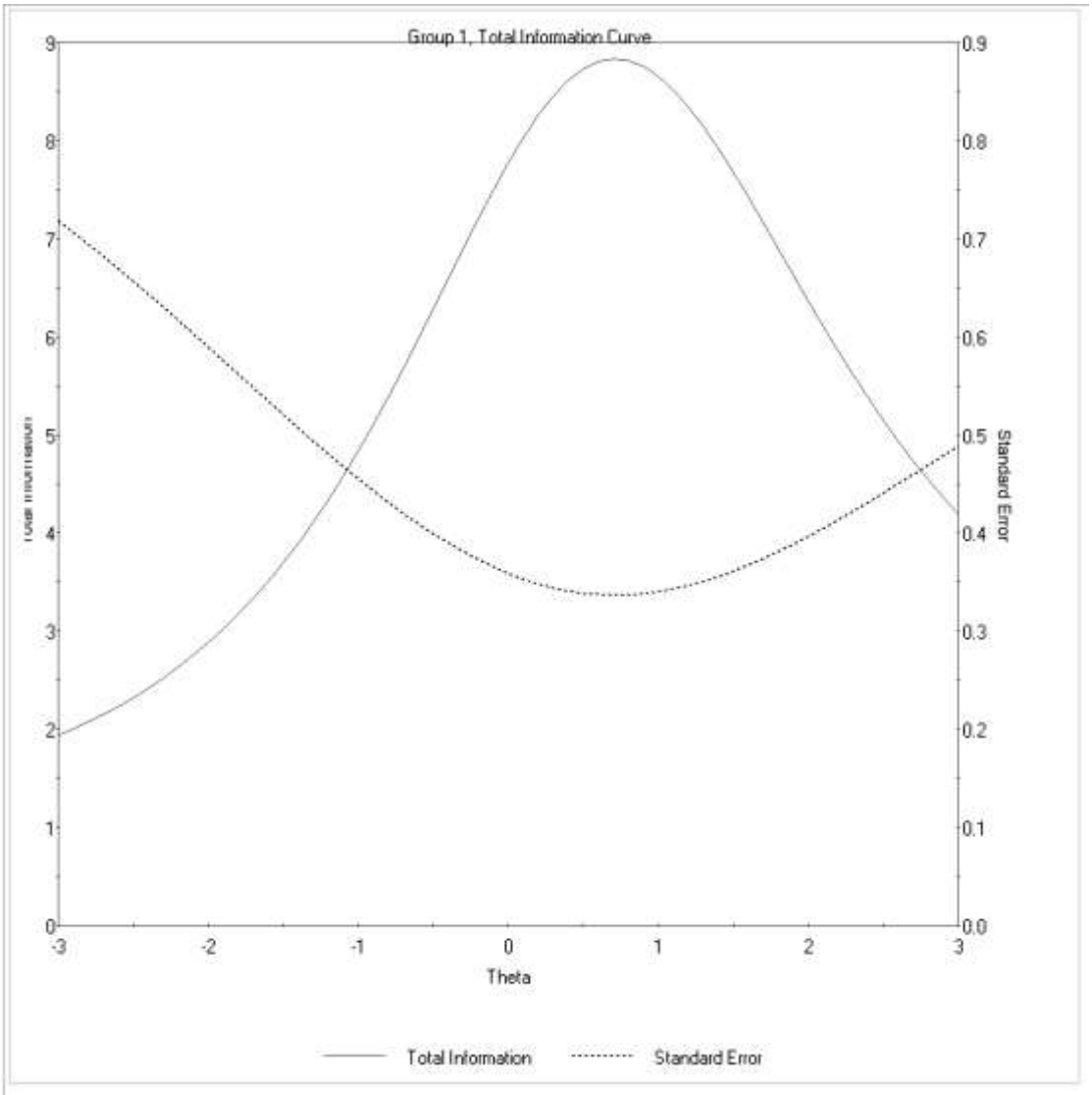
34	0.392	0.376	0.0216679
35	0.450	0.375	0.0209149
36	0.507	0.373	0.0201520
37	0.564	0.373	0.0193822
38	0.620	0.372	0.0186078
39	0.677	0.371	0.0178309
40	0.733	0.371	0.0170534
41	0.789	0.372	0.0162769
42	0.845	0.372	0.0155030
43	0.901	0.373	0.0147330
44	0.957	0.374	0.0139682
45	1.013	0.375	0.0132099
46	1.069	0.376	0.0124594
47	1.125	0.378	0.0117180
48	1.182	0.380	0.0109871
49	1.239	0.382	0.0102679
50	1.296	0.385	0.0095620
51	1.353	0.387	0.0088708
52	1.411	0.390	0.0081960
53	1.470	0.393	0.0075392
54	1.529	0.396	0.0069021
55	1.588	0.400	0.0062865
56	1.649	0.404	0.0056941
57	1.709	0.408	0.0051267
58	1.771	0.412	0.0045861
59	1.833	0.416	0.0040738
60	1.896	0.421	0.0035914
61	1.959	0.425	0.0031403
62	2.024	0.430	0.0027217
63	2.089	0.435	0.0023362
64	2.155	0.441	0.0019847
65	2.222	0.446	0.0016671
66	2.290	0.452	0.0013834
67	2.358	0.457	0.0011328
68	2.428	0.463	0.0009144
69	2.499	0.469	0.0007266
70	2.571	0.475	0.0005677
71	2.644	0.481	0.0004354
72	2.718	0.488	0.0003272
73	2.793	0.494	0.0002406
74	2.870	0.500	0.0001726
75	2.947	0.507	0.0001206
76	3.026	0.514	0.0000818
77	3.107	0.520	0.0000537
78	3.189	0.527	0.0000340
79	3.273	0.534	0.0000207
80	3.358	0.540	0.0000120
81	3.445	0.547	0.0000066
82	3.535	0.554	0.0000034
83	3.626	0.560	0.0000016
84	3.720	0.567	0.0000007
85	3.818	0.574	0.0000003
86	3.919	0.580	0.0000001
87	4.024	0.586	0.0000000
88	4.132	0.592	0.0000000

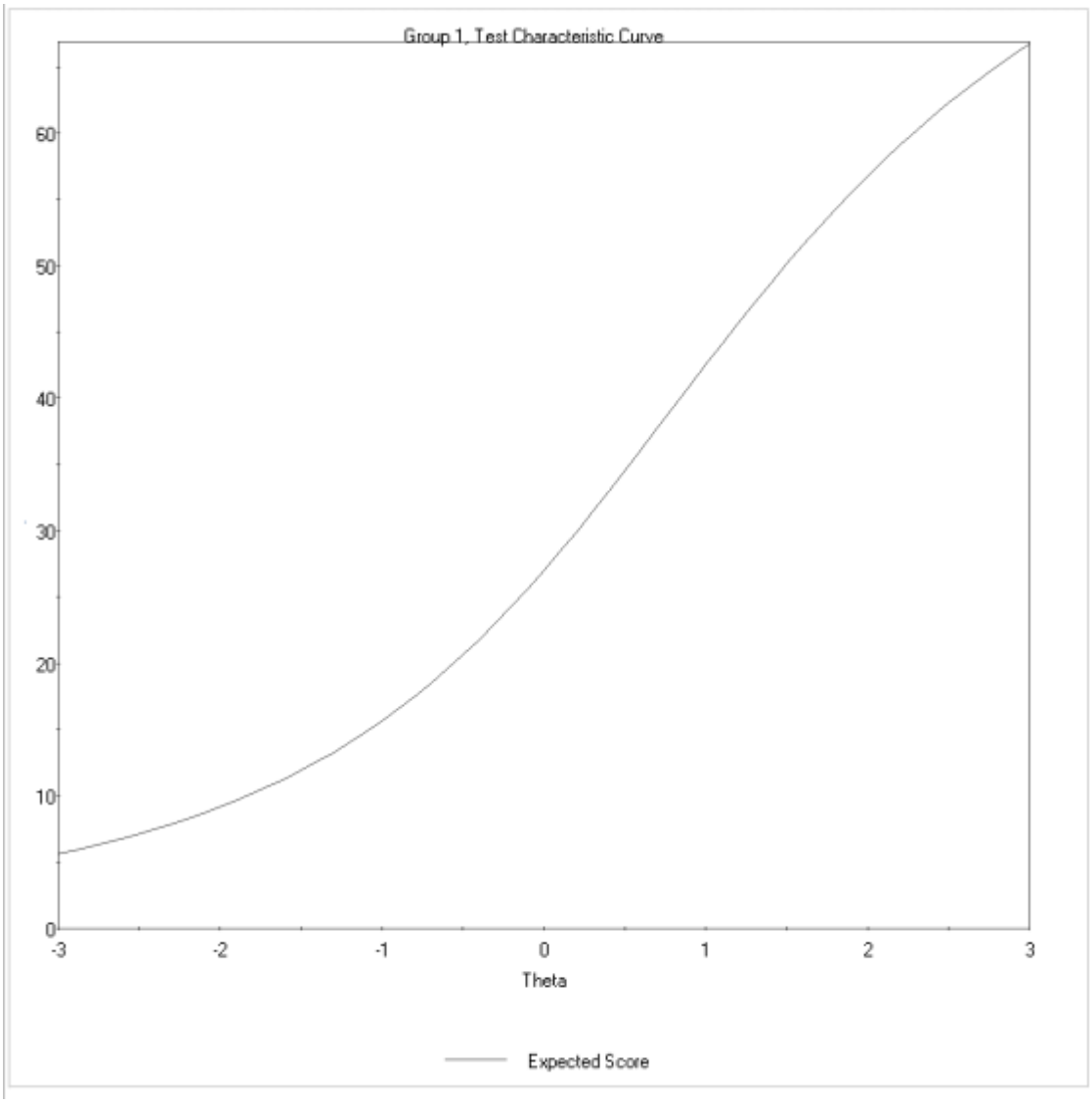
89	4.242	0.597 0.0000000
90	4.352	0.601 0.0000000

Marginal reliability of the scaled scores for summed scores = 0.81867

Summary of the Data and Control Parameters **(Back to TOC)**

Sample Size	658
Number of Items	30
Number of Dimensions	1




Group 1, Test Characteristic Curve
60
50
40
30
20
10
0
-3
-2
-1
0
1
2
3
Theta
Expected Score

ภาคผนวก ฉ

คู่มือการใช้แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนน
แบบพิจารณาความมั่นใจ

คู่มือการใช้แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ

โดยผู้วิจัยสร้างขึ้นเพื่อเป็นแนวทางในการนำไปใช้สำหรับผู้สอนที่สนใจเพื่อให้บรรลุตามวัตถุประสงค์ ซึ่งรายละเอียดประกอบด้วย

1. วัตถุประสงค์
2. โครงสร้างการวิเคราะห์หน่วยการเรียนรู้ ผลการเรียนรู้ และจำนวนข้อสอบตามพฤติกรรมที่ต้องการวัด
3. ลักษณะของแบบทดสอบ
4. การพัฒนาแบบทดสอบ
5. คุณภาพของแบบทดสอบ
6. เวลาที่ใช้ในการสอบ
7. วิธีดำเนินการสอบ
8. การตรวจให้คะแนน
9. ข้อเสนอแนะในการนำไปใช้

สำหรับคู่มือการใช้แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ มีรายละเอียด ดังนี้

วัตถุประสงค์

แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจที่สร้างขึ้นเพื่อ ได้แบบทดสอบเลือกตอบที่มีวิธีการตอบและตรวจให้คะแนนทั้ง 2 วิธี ที่ได้มาตรฐานไปใช้ในห้องเรียน เป็นข้อมูลในการตัดสินใจของครูเพื่อเลือกใช้แบบทดสอบที่เหมาะสม โดยเฉพาะอย่างยิ่งเพื่อแก้ปัญหาแบบทดสอบที่มีการเดา ทำให้ทราบถึงความสามารถที่แท้จริงของผู้สอบและความสามารถในการประมาณค่าพารามิเตอร์เกี่ยวกับแบบทดสอบ รวมทั้งเป็นการเพิ่มสารสนเทศเกี่ยวกับการตอบแบบทดสอบแบบเลือกตอบให้มากขึ้น และนำมาซึ่งการพัฒนาแนวคิดเกี่ยวกับการหาค่าคะแนนความสามารถที่แท้จริงจากการใช้วิธีการตอบและการตรวจให้คะแนนแบบต่าง ๆ

โครงสร้างการวิเคราะห์หน่วยการเรียนรู้ ผลการเรียนรู้ และจำนวนข้อสอบ ตามพฤติกรรมที่ต้องการวัด

หน่วยการเรียนรู้	ผลการเรียนรู้	พฤติกรรมที่ต้องการวัด				รวม
		ความรู้/ความจำ	ความเข้าใจ	การนำไปใช้	การวิเคราะห์	
1. มวลอะตอม และมวลโมเลกุล	1. สามารถอธิบายความหมายของมวลอะตอม มวลโมเลกุล คำนวณหามวลอะตอมของธาตุได้	(2)				(2)
	2. สามารถอธิบายความหมายของมวลโมเลกุลของสารมวลของธาตุ 1 อะตอม มวลอะตอมเฉลี่ยของธาตุ และมวลของสาร 1 โมเลกุลได้	(7)				(7)
2. โมล	3. สามารถอธิบายความสัมพันธ์ระหว่างโมล มวล จำนวนอนุภาคและปริมาตรของแก๊สได้		(2)			(2)
	1. สามารถคำนวณความสัมพันธ์ระหว่างโมล มวล จำนวนอนุภาคและปริมาตรของแก๊สได้		(8)			(8)
3. สารละลายและความเข้มข้น	3. สามารถอธิบายสมบัติบางประการของสารละลาย ทำการทดลอง เตรียมสารละลายได้	(1)		(1)	(1)	(3)

หน่วยการเรียนรู้	ผลการเรียนรู้	พฤติกรรมที่ต้องการวัด				รวม
		ความรู้/ความจำ	ความเข้าใจ	การนำไปใช้	การวิเคราะห์	
	4. สามารถคำนวณความเข้มข้นของสารละลายในหน่วยร้อยละ โมลาริตี โมแลล ส่วนในล้านส่วน ส่วนในพันล้านส่วนและเศษส่วนโมลได้		(4)			(4)
	8. สามารถอธิบายและทำการทดลองเกี่ยวกับกฎทรงมวลของสารและกฎสัดส่วนคงที่ได้		(1)		(1)	(2)
	9. สามารถอธิบายและทำการทดลองความสัมพันธ์ของปริมาตรแก๊สในสมการเคมี ใช้กฎของเกย์ลุสแซก กฎของอาโวกาโดรได้				(1)	(1)
	10. สามารถคำนวณหาปริมาตรของแก๊สในปฏิกิริยาเคมีได้		(2)			(2)
4. การคำนวณเกี่ยวกับสูตรเคมี	11. สามารถอธิบายเกี่ยวกับสูตร และสมการเคมีได้	(1)				(1)
	12. สามารถคำนวณเกี่ยวกับสูตร และสมการเคมีได้		(3)			(3)
	13. สามารถใช้ความสัมพันธ์ระหว่างจำนวนโมล กับมวลหรือปริมาตรของสาร คำนวณหาค่าโมล มวลหรือปริมาตรของสาร สมการเคมี และหาค่าร้อยละของผลที่ได้จาการทดลองได้		(2)			(2)

ลักษณะของแบบทดสอบ

แบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชา เคมี กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 4 เรื่อง ปริมาณสารสัมพันธ์ ซึ่งเป็นแบบทดสอบชนิดเลือกตอบ 4 ตัวเลือก 1 ฉบับ จำนวน 30 ข้อ ที่มีวิธีการตอบและตรวจให้คะแนน 2 วิธี ได้แก่

1. แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา (Multiple choice) หมายถึง แบบทดสอบที่แต่ละข้อคำถามประกอบด้วย ตอนนำ หรือ ตัวคำถาม (Stem or lead or problem) กับตัวเลือก (Choice or option or alternative or response) ในแต่ละข้อคำถาม มีตัวเลือกมากกว่า 2 ตัวเลือก สำหรับการศึกษาครั้งนี้มี 4 ตัวเลือก และมีตัวเลือกถูกเพียงตัวเลือกเดียว เพื่อให้ผู้สอบตอบสนองออกมา สังเกตได้และนับจำนวนได้ โดยมีวิธีการตรวจให้คะแนนแบบธรรมดา คือวิธีการให้ผู้สอบเลือกคำตอบที่ถูกที่สุดเพียงข้อเดียว

2. แบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ (Confidential making) หมายถึง แบบทดสอบที่มีคำชี้แจงให้นักเรียนทำเครื่องหมายแก่ตัวเลือกที่ถูกที่สุดเพียงตัวเลือกเดียว พร้อมทั้งบอกความมั่นใจในการตอบตัวเลือกนั้นด้วยว่ามาก ปานกลาง หรือน้อย โดยมีวิธีการตรวจให้คะแนนความมั่นใจ คือการคาดคะเนของผู้สอบว่าคำตอบที่เลือกนั้นมีโอกาสเป็นคำตอบที่ถูกต้อง โดยระบุความมั่นใจในการตอบว่า มาก ปานกลาง หรือน้อยอย่างใดอย่างหนึ่ง

การพัฒนาแบบทดสอบ

ตอนที่ 1 ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการสร้างแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน ศึกษาหลักสูตรคู่มือครูและแบบเรียนวิชาเคมี กำหนดตารางโครงสร้างการวิเคราะห์แบบทดสอบ ตามผลการเรียนรู้ และกำหนดจำนวนข้อสอบที่ใช้จริงจากจำนวนข้อสอบที่ออกทั้งหมดจากน้ำหนักเป็นร้อยละตามเนื้อหา ผลการเรียนรู้ โดยจำแนกตามพฤติกรรมที่ต้องการวัด แล้วนำมาสร้างแบบทดสอบวัดผลสัมฤทธิ์วิชาเคมี จำนวน 1 ฉบับ ชนิดเลือกตอบ 4 ตัวเลือก จำนวน 40 ข้อ โดยมีวิธีการตอบและตรวจให้คะแนน 2 วิธี

ตอนที่ 2 นำเครื่องมือที่สร้างขึ้นให้ผู้เชี่ยวชาญ จำนวน 5 คน ตรวจสอบคุณภาพเบื้องต้นด้านความตรงเชิงเนื้อหา (Content validity) หาค่าดัชนีความสอดคล้องเป็นรายข้อ (Index of congruency: IOC) และตัดข้อสอบที่ไม่ผ่านเกณฑ์ ได้แบบทดสอบทั้งสิ้นจำนวน 30 ข้อ

ตอนที่ 3 นำแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชา เคมี 1 ฉบับ จำนวน 30 ข้อ ที่มีวิธีการตอบและตรวจให้คะแนนทั้ง 2 วิธี นำไปทดลองใช้ครั้งที่ 1 นักเรียนกลุ่มตัวอย่างจำนวน 50 คน คัดเลือกและปรับปรุงแก้ไขข้อสอบก่อนนำไปทดลองใช้ครั้งที่ 2 กับ นักเรียกลุ่มตัวอย่าง 210 คน เพื่อหาคุณภาพแบบทดสอบ วิเคราะห์ค่าความยากและค่าอำนาจจำแนก (Brennan)

และค่าความเที่ยงของความสอดคล้องภายใน (KR-20) โดยกำหนดให้เว้นระยะเวลาในการทำแบบทดสอบที่มีวิธีการตอบและตรวจให้คะแนนแต่ละวิธีอย่างน้อย 7 วัน

ตอนที่ 4 กำหนดคะแนนจุดตัดของแบบทดสอบอิงเกณฑ์ เลือกใช้การกำหนดคะแนนจุดตัดตามโดยวิธีของ Angoff เป็นวิธีที่กำหนดคะแนนจุดตัดของผู้เชี่ยวชาญในรายวิชา เคมี จำนวน 5 คน ซึ่งก็เป็นชุดเดียวกับที่พิจารณาหาค่าความตรงตามเชิงเนื้อหา (Content validity)

ตอนที่ 5 นำแบบทดสอบที่ได้ทดลองใช้ครั้งที่ 2 กับ นักเรียกลุ่มตัวอย่าง 210 คน แล้วนำมาวิเคราะห์หาคุณภาพแบบทดสอบ วิเคราะห์ค่าความยากและค่าอำนาจจำแนก (Brennan) และค่าความเที่ยงอิงเกณฑ์วิธีของ Livingston เพื่อได้แบบทดสอบที่มีคุณภาพในการนำไปใช้จริงกับนักเรียนกลุ่มตัวอย่างชั้นมัธยมศึกษาปีที่ 4 ภาคเรียนที่ 2 ปีการศึกษา 2560 จำนวน 658 คน

คุณภาพของแบบทดสอบ

1. คุณภาพของแบบทดสอบในด้านค่าความยาก (p) ค่าอำนาจจำแนก (r) ทดลองใช้ครั้งที่ 1 และครั้งที่ 2 ของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ พบว่า ในทดลองใช้ครั้งที่ 1 ผู้วิจัยได้คัดเลือกแบบทดสอบที่มีค่าความยากตั้งแต่ 0.20 ถึง 0.78 และคัดเลือกแบบทดสอบที่มีค่าอำนาจจำแนก (Brennan) ตั้งแต่ 0.21 ถึง 0.77 ข้อที่มีความยากต่ำกว่า 0.20 และค่าอำนาจจำแนก (Brennan) ต่ำกว่า 0.20 นำมาปรับปรุงข้อสอบจำนวน 10 ข้อ ได้แก่ข้อ1, 2, 3, 8, 13, 18, 23, 27, 28 และ 30 เพื่อนำไปทดลองใช้ครั้งที่ 2 แล้วหาคุณภาพแบบทดสอบเพื่อให้ได้แบบทดสอบไปใช้จริง จากการหาคุณภาพแบบทดสอบโดยวิเคราะห์หาค่าความยากและค่าอำนาจจำแนก (Brennan) จากการนำไปทดลองใช้ครั้งที่ 2 พบว่า แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ มีค่าความยากอยู่ระหว่าง 0.20 ถึง 0.73 และ 0.25 ถึง 0.76 ตามลำดับ ข้อสอบผ่านเกณฑ์ทุกข้อ และค่าอำนาจจำแนก (Brennan) มีค่าอยู่ระหว่าง 0.25 ถึง 0.73 และ 0.24 ถึง 0.38 ตามลำดับ ข้อสอบผ่านเกณฑ์ทุกข้อ ได้แบบทดสอบจำนวน 30 ข้อ เพื่อนำไปใช้จริง

2. คุณภาพของแบบทดสอบแบบทดสอบตามวิธีการให้คะแนนของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ นำแบบทดสอบไปหาคะแนนจุดตัด โดยวิธีของ Angoff คะแนนจุดตัดของแบบทดสอบตามวิธีการให้คะแนนของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ เป็นวิธีที่กำหนดคะแนนจุดตัดของผู้เชี่ยวชาญในรายวิชา เคมี จำนวน 5 คน ซึ่งก็เป็นชุดเดียวกับที่พิจารณาหาค่าความตรงตามเชิงเนื้อหา (Content validity) พบว่า ผลการพิจารณาคะแนนจุดตัดของแบบทดสอบ

ตามวิธีการให้คะแนนของแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจ โดยวิธี Angoff ทั้งฉบับ มีคะแนนจุดตัดเท่ากับ 16.87

3. คุณภาพแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาและให้คะแนนแบบพิจารณาความมั่นใจทั้งฉบับจากการทดลองใช้ครั้งที่ 1 พบว่า มีส่วนเบี่ยงเบนมาตราฐาน (SD) มีค่าเท่ากับ 4.79 และ 5.42 ตามลำดับ ความคลาดเคลื่อนของแบบทดสอบ (SEM) มีค่าเท่ากับ 2.30 และ 3.06 ตามลำดับ ความเที่ยงของความสอดคล้องภายใน (rKR-20) มีค่าเท่ากับ 0.77 และ 0.68 ตามลำดับ จากการการนำไปทดลองใช้ครั้งที่ 2 พบว่า มีส่วนเบี่ยงเบนมาตราฐาน (SD) มีค่าเท่ากับ 5.01 และ 5.22 ตามลำดับ ความคลาดเคลื่อนของแบบทดสอบ (SEM) มีค่าเท่ากับ 1.94 และ 2.09 ตามลำดับ ความเที่ยงของความสอดคล้องภายใน (rKR-20) มีค่าเท่ากับ 0.85 และ 0.84 ตามลำดับ ค่าความเที่ยงแบบอิงเกณฑ์ (rcc) มีค่าเท่ากับ 0.97 และ 0.84 ตามลำดับ

เวลาที่ใช้ในการสอบ

ใช้การจับเวลา เพื่อหาเวลาที่ใช้ในการทำแบบทดสอบโดยเฉลี่ย ใน 1 คาบเรียน ได้เวลาโดยเฉลี่ย 51 นาที แต่มีข้อจำกัดคือ ถ้าไม่กำหนดเวลาในการทำแบบทดสอบอาจทำให้การดำเนินการสอบต้องใช้เวลามากเกินไป ดังนั้นเวลาที่ใช้ในการทำแบบทดสอบ ผู้ที่นำไปใช้สามารถยืดหยุ่นได้ตามความเหมาะสม

วิธีดำเนินการสอบ

ดำเนินการสอบกับนักเรียน เตรียมแบบทดสอบให้เพียงพอกับจำนวนนักเรียนที่เข้าสอบในแต่ละครั้ง และวางแผนดำเนินการสอบ ผู้คุมสอบอธิบายให้นักเรียนกลุ่มตัวอย่างเข้าใจวัตถุประสงค์ของแบบทดสอบและวิธีการตอบก่อนเริ่มต้นลงมือทำ โดยกำหนดให้เว้นระยะเวลาในการทำแบบทดสอบที่มีวิธีการตอบและตรวจให้คะแนนแต่ละวิธีอย่างน้อย 7 วัน ตรวจดูความเรียบร้อยในการสอบตลอดจนปัญหาต่าง ๆ ที่อาจจะเกิดขึ้นในขณะสอบ ถ้านักเรียนทำเสร็จก่อนเวลาที่กำหนดอนุญาตให้ออกจากห้องสอบได้ และเก็บรวบรวมกระดาษคำตอบ ให้ถูกต้องครบถ้วนและเรียบร้อย

การตรวจให้คะแนน

นำกระดาษคำตอบของนักเรียนมาตรวจให้คะแนนทั้ง 2 วิธี โดยแบ่งเป็น

1. แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา

ข้อสอบข้อที่ (0) คำตอบที่ถูกต้อง คือ ข้อ ข จะได้คะแนน 1 คะแนน

ข้อ	ตัวเลือก				คะแนน
	ก	ข	ค	ง	
0		✘			1

ข้อสอบข้อที่ (0) คำตอบที่ถูกต้อง คือ ข้อ ข ถ้าผู้สอบตอบข้อ ก จะได้คะแนน 0 คะแนน

ข้อ	ตัวเลือก				คะแนน
	ก	ข	ค	ง	
0	✘				0

เกณฑ์การให้คะแนนแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา ดังนี้

ผู้สอบเลือกตัวเลือก “ถูก” ได้คะแนนเท่ากับ 1

ผู้สอบเลือกตัวเลือก “ผิด” ได้คะแนนเท่ากับ 0

2. แบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ

ข้อสอบข้อที่ (0) คำตอบที่ถูกต้อง คือ ข้อ ข ผู้สอบตอบข้อ ข และระบุระดับความมั่นใจมาก จะได้คะแนน 3 คะแนน

ข้อ	ตัวเลือก				ระดับความมั่นใจในการตอบ			คะแนน
	ก	ข	ค	ง	มาก	ปานกลาง	น้อย	
0		✘			✘			3

ข้อสอบข้อที่ (0) คำตอบที่ถูกต้อง คือ ข้อ ข ถ้าผู้สอบตอบข้อ ข และระบุระดับความมั่นใจปานกลางจะได้คะแนน 2 คะแนน

ข้อ	ตัวเลือก				ระดับความมั่นใจในการตอบ			คะแนน
	ก	ข	ค	ง	มาก	ปานกลาง	น้อย	
0		✘				✘		2

ข้อสอบข้อที่ (0) คำตอบที่ถูกต้อง คือ ข้อ ข ถ้าผู้สอบตอบข้อ ข และระบุระดับความมั่นใจน้อย จะได้คะแนน 1 คะแนน

ข้อ	ตัวเลือก				ระดับความมั่นใจในการตอบ			คะแนน
	ก	ข	ค	ง	มาก	ปานกลาง	น้อย	
0		✘					✘	1

ข้อสอบข้อที่ (0) คำตอบที่ถูกต้อง คือ ข้อ ข ถ้าผู้สอบตอบผิดคือไม่ได้เลือกข้อ ข จะได้คะแนน 0 คะแนน ไม่ว่าจะระบุระดับความมั่นใจในการตอบระดับใด

ข้อ	ตัวเลือก				ระดับความมั่นใจในการตอบ			คะแนน
	ก	ข	ค	ง	มาก	ปานกลาง	น้อย	
0	✘				✘			0

เกณฑ์การให้คะแนนแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ

ผู้สอบเลือกตัวเลือก “ถูก” ในระดับความมั่นใจ “มาก” ได้คะแนนเท่ากับ 3

ผู้สอบเลือกตัวเลือก “ถูก” ในระดับความมั่นใจ “ปานกลาง” ได้คะแนนเท่ากับ 2

ผู้สอบเลือกตัวเลือก “ถูก” ในระดับความมั่นใจ “น้อย” ได้คะแนนเท่ากับ 1

ผู้สอบเลือกตัวเลือก “ผิด” ในระดับความมั่นใจ “มาก” ได้คะแนนเท่ากับ 0

ผู้สอบเลือกตัวเลือก “ผิด” ในระดับความมั่นใจ “ปานกลาง” ได้คะแนนเท่ากับ 0

ผู้สอบเลือกตัวเลือก “ผิด” ในระดับความมั่นใจ “น้อย” ได้คะแนนเท่ากับ 0

ข้อเสนอแนะในการนำไปใช้

1. การใช้แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาเคมี แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดาจะให้ค่าประสิทธิภาพสัมพัทธ์สูงกว่า โดยเฉพาะผู้สอบที่มีระดับความสามารถต่ำ การตรวจให้คะแนนดังกล่าวเหมาะสมกว่า

2. การใช้แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาเคมี แบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ จะให้ค่าประสิทธิภาพสัมพัทธ์สูงกว่า โดยเฉพาะผู้สอบที่ระดับความสามารถค่อนข้างสูง การตรวจให้คะแนนดังกล่าวเหมาะสมกว่า

3. เพื่อเป็นการเพิ่มประสิทธิภาพในการจัดการเรียนการสอน ครูผู้สอนควรนำผลการทดสอบไปใช้ในแนวทางในการวางแผนการจัดการเรียนการสอนที่สอดคล้องกับระดับความสามารถของผู้เรียน

4. ควรทำการศึกษาเพื่อตรวจสอบคุณภาพของวิธีการให้คะแนนความรู้บางส่วนในวิชาอื่น ๆ เช่น วิชาภาษาไทย วิชาภาษาอังกฤษ วิชาสังคมศึกษา เป็นต้น เพราะบริบทของแต่ละวิชาไม่เหมือนกัน ซึ่งวิธีการให้คะแนนความรู้บางส่วนแต่ละวิธีอาจจะเหมาะสมกับแต่ละวิชาแตกต่างกัน

แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาเคมี ชั้นมัธยมศึกษาปีที่ 4
กลุ่มสาระการเรียนรู้วิทยาศาสตร์ เรื่อง ปริมาณสารสัมพันธ์
(แบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ)

คำชี้แจง

1. แบบทดสอบฉบับนี้มีจำนวน 30 ข้อ
2. แบบทดสอบฉบับนี้เป็นแบบทดสอบแบบเลือกตอบ ชนิด 4 ตัวเลือก
3. คำตอบที่ถูกต้องมีเพียงคำตอบเดียว
4. ให้นักเรียนทำเครื่องหมายกากบาท (X) ตรงตัวเลือกที่พิจาณาว่าเป็น **<u>คำตอบที่ถูกต้องที่สุด</u>** เพียงคำตอบเดียว ลงในกระดาษคำตอบ และทำเครื่องหมายกากบาท (X) ลงในช่องระบุความมั่นใจ

ตัวอย่างแบบทดสอบ

ข้อที่ 0 แก๊ส 1 ลิตรที่ STP มีกี่โมล

ก. 0.023 ข. 0.69

ค. 0.045 ง. 1

(คำตอบที่ถูกต้องคือ ข.)

ตัวอย่างกระดาษคำตอบ

	ข้อ	ตัวเลือก				ระดับความมั่นใจ		
		ก	ข	ค	ง	มาก	ปานกลาง	น้อย
	0		X			X		
หรือ	0		X				X	
หรือ	0		X					X

เกณฑ์การให้คะแนนแบบทดสอบเลือกตอบที่ให้คะแนนแบบพิจารณาความมั่นใจ

ผู้สอบเลือกตัวเลือก “ถูก” ในระดับความมั่นใจ “มาก” ได้คะแนนเท่ากับ 3
ผู้สอบเลือกตัวเลือก “ถูก” ในระดับความมั่นใจ “ปานกลาง” ได้คะแนนเท่ากับ 2
ผู้สอบเลือกตัวเลือก “ถูก” ในระดับความมั่นใจ “น้อย” ได้คะแนนเท่ากับ 1
ผู้สอบเลือกตัวเลือก “ผิด” ในระดับความมั่นใจ “มาก” ได้คะแนนเท่ากับ 0
ผู้สอบเลือกตัวเลือก “ผิด” ในระดับความมั่นใจ “ปานกลาง” ได้คะแนนเท่ากับ 0
ผู้สอบเลือกตัวเลือก “ผิด” ในระดับความมั่นใจ “น้อย” ได้คะแนนเท่ากับ 0

โรงเรียน……...…..ชื่อ-นามสกุล……ชั้น…./ …..เลขที่…..

ระดับผลการเรียนวิชาเคมี .....

กระดาษคำตอบ

ข้อ	ตัวเลือก				ระดับความมั่นใจ			ข้อ	ตัวเลือก				ระดับความมั่นใจ		
	ก	ข	ค	ง	มาก	ปานกลาง	น้อย		ก	ข	ค	ง	มาก	ปานกลาง	น้อย
1								16							
2								17							
3								18							
4								19							
5								20							
6								21							
7								22							
8								23							
9								24							
10								25							
11								26							
12								27							
13								28							
14								29							
15								30							

แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน ชั้นมัธยมศึกษาปีที่ 4
กลุ่มสาระการเรียนรู้วิทยาศาสตร์ วิชาเคมี เรื่อง ปริมาณสารสัมพันธ์
(แบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา)

คำชี้แจง

1. แบบทดสอบฉบับนี้มีจำนวน 30 ข้อ
2. แบบทดสอบฉบับนี้เป็นแบบทดสอบแบบเลือกตอบ ชนิด 4 ตัวเลือก
3. ให้นักเรียนทำเครื่องหมายกากบาท (X) ตรงตัวเลือกที่พิจาณาว่าเป็น**คำตอบที่ถูกต้องที่สุด** เพียงคำตอบเดียว ลงในกระดาษคำตอบ

ตัวอย่างแบบทดสอบ

ข้อที่ 0 แก๊ส 1 ลิตรที่ STP มีกี่โมล

ก. 0.023 ข. 0.69

ค. 0.045 ง. 1

(คำตอบที่ถูกต้องคือ ข.)

ตัวอย่างกระดาษคำตอบ

ข้อ	ตัวเลือก			
	ก	ข	ค	ง
0		X		

เกณฑ์การให้คะแนนแบบทดสอบเลือกตอบที่ให้คะแนนแบบธรรมดา ดังนี้

ผู้สอบเลือกตัวเลือก “ถูก” ได้คะแนนเท่ากับ 1

ผู้สอบเลือกตัวเลือก “ผิด” ได้คะแนนเท่ากับ 0

โรงเรียน……...…..ชื่อ-นามสกุล……ชั้น…./ …..เลขที่…..

ระดับผลการเรียนวิชาเคมี .....

กระดาษคำตอบ

ข้อ	ตัวเลือก				ข้อ	ตัวเลือก			
	ก	ข	ค	ง		ก	ข	ค	ง
1					16				
2					17				
3					18				
4					19				
5					20				
6					21				
7					22				
8					23				
9					24				
10					25				
11					26				
12					27				
13					28				
14					29				
15					30				