

การศึกษาเปรียบเทียบคุณภาพของแบบทดสอบเลือกตอบเมื่อตรวจด้วยวิธีการให้คะแนน
ความรู้บางส่วนของแบบทดสอบวัดผลสัมฤทธิ์คณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3

ปริญญานิพนธ์
ของ
สมจวน มานะก่อ

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต สาขาวิชาการวัดผลการศึกษา

พฤษภาคม 2551

การศึกษาเปรียบเทียบคุณภาพของแบบทดสอบเลือกตอบเมื่อตรวจด้วยวิธีการให้คะแนน
ความรู้บางส่วน of แบบทดสอบวัดผลสัมฤทธิ์คณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3

ปริญญานิพนธ์
ของ
สมจวน มานะก่อ

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต สาขาวิชาการวัดผลการศึกษา

พฤษภาคม 2551

ลิขสิทธิ์เป็นของมหาวิทยาลัยศรีนครินทรวิโรฒ

การศึกษาเปรียบเทียบคุณภาพของแบบทดสอบเลือกตอบเมื่อตรวจด้วยวิธีการให้คะแนน
ความรู้บางส่วนของแบบทดสอบวัดผลสัมฤทธิ์คณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3

บทคัดย่อ
ของ
สมจวน มานะก่อ

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต สาขาวิชาการวัดผลการศึกษา
พฤษภาคม 2551

สมจวน มานะก่อ. (2551). การศึกษาเปรียบเทียบคุณภาพของแบบทดสอบเลือกตอบ

เมื่อตรวจด้วยวิธีการให้คะแนนความรู้บางส่วนของแบบทดสอบวัดผลสัมฤทธิ์คณิตศาสตร์
ชั้นมัธยมศึกษาปีที่ 3. ปริญญาโท กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ:

บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. คณะกรรมการควบคุม:

ผู้ช่วยศาสตราจารย์ ดร.องอาจ นัยพัฒน์, ผู้ช่วยศาสตราจารย์ ดร. นิคม ดังคะพิภพ.

การวิจัยครั้งนี้ มีจุดประสงค์เพื่อเปรียบเทียบคุณภาพของแบบทดสอบเลือกตอบเมื่อตรวจด้วยวิธีการให้คะแนนความรู้บางส่วนของแบบทดสอบวัดผลสัมฤทธิ์คณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 เมื่อตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็นกับการตรวจให้คะแนนตามความยากประจำชั้น กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 สำนักงานเขตพื้นที่การศึกษาปราจีนบุรี ภาคเรียนที่ 2 ปีการศึกษา 2549 จำนวน 29 โรงเรียน จำนวน 1,007 คน ได้มาโดยการสุ่มแบบหลายขั้นตอน (Multi - stage random Sampling) เครื่องมือที่ใช้เป็นแบบทดสอบวัดผลสัมฤทธิ์คณิตศาสตร์ จำนวน 40 ข้อ มีความเชื่อมั่นจากวิธีให้คะแนนของการทดสอบความน่าจะเป็นและจากวิธีให้คะแนนตามความยากประจำชั้น เท่ากับ 0.92 และ 0.91 ตามลำดับ

ผลการวิจัยพบว่า

1. ความเที่ยงตรงตามสภาพ (Concurrent Validity) ของแบบทดสอบวัดผลสัมฤทธิ์คณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 เมื่อตรวจด้วยวิธีการให้คะแนนความรู้บางส่วน โดยพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ (r) ระหว่างคะแนนสอบกลางภาคกับคะแนนที่ได้มาจากการตรวจด้วยวิธีของการทดสอบความน่าจะเป็นกับวิธีการตรวจให้คะแนนตามความยากประจำชั้นมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ใช้สถิติไคสแควร์ (χ^2) ทดสอบความแตกต่างโดยใช้คะแนนพิชเชอร์ที่เปรียบเทียบในภาพรวม พบว่าความเที่ยงตรงตามสภาพไม่ขึ้นอยู่กับวิธีการตรวจให้คะแนน

2. ฟังก์ชันสารสนเทศของข้อสอบ (Item Information) วิธีการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็น ให้ค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงที่สุดในระดับความสามารถ -1 มีค่าเท่ากับ .097 โดยเฉลี่ยสูงกว่าวิธีการตรวจให้คะแนนตามความยากประจำชั้นที่มีค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงที่สุดในระดับความสามารถ -1 มีค่าเท่ากับ .092 และค่าประสิทธิภาพสัมพัทธ์ (Relative Efficiency: $RE(\theta)$) ที่ระดับความสามารถ (θ) -1.4 ถึง 0.6 วิธีการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็นมีประสิทธิภาพสูงกว่าวิธีการตรวจให้คะแนนด้วยวิธีการให้คะแนนตามความยากประจำชั้น

3. ค่าฟังก์ชันสารสนเทศแบบทดสอบ (Test Information) การตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็น มีค่าเท่ากับ 4.895 สูงกว่าการตรวจให้คะแนนตามความยากประจำชั้น มีค่าเท่ากับ 4.630

A COMPARATIVE STUDY OF MULTIPLE CHOICE TEST QUALITY
WHEN SCORING WITH THE PARTIAL CREDIT METHODS OF MATHEMATICS TEST
FOR MATTAYOMSUKSA 3 STUDENTS

AN ABSTRACT
BY
SOMCHUAN MANAKOR

Presented in Partial Fulfillment of the Requirements for the
Master of Education Degree in Educational Measurement
at Srinakharinwirot University

May 2008

Somchuan Manakor. (2008). *The Comparative Study of Multiple Choice Test Quality When Scoring With Partial Credit Method of Mathematic Test for Mathayom Suksa III Students*. Master thesis. M.Ed. (Educational Measurement). Bangkok: Graduate School, Srinakharinwirot University.
Advisor Committee: Asst. Prof. Dr. Ong-Art Naiyapatana, Asst. Prof. Dr. Nikom Tangkapipop.

This study aimed to compare multiple choice test quality when scoring with partial credit method of mathematic test for Mathayom Suksa III students by partial credit method with probability testing and step difficulties. The subjects comprised 1,007 Mathayom Suksa III students in 29 schools under Prachinburi Educational Area during the 2nd semester of 2007 academic year, through multi-stage random sampling. The instruments for collecting data included mathematic test including process skill with 40 items. The probability testing and step difficulties with reliability of .92 and .91 respectively.

The results revealed as follows:

1. According to concurrent validity of mathematic test for Mathayom Suksa III students when scoring with partial credit method through probability testing and step difficulties, it was found that when compared the differences by χ^2 , the concurrent validity was not relied on scoring method and there were no significant differences.
2. Item information function (IIF), probability testing method was significant higher than scoring through step difficulties. In accordance with item information functioning (IIF) as to examinee ability, it was found that those of scoring through probability testing were significant higher with -1 (IIF = .097) than those of scoring through step difficulties with -1 (IIF = .092) and relative efficiency [RE (θ)] at -1.4 to reach 0.6 probability testing method was efficiency higher than scoring through step difficulties.
3. Test information function (TIF), it was found that those of scoring through probability testing were significant higher (TIF = 4.895) than those of scoring through step difficulties (δ = 4.630).

ปริญญานิพนธ์

เรื่อง

การศึกษาเปรียบเทียบคุณภาพของแบบทดสอบเลือกตอบเมื่อตรวจด้วยวิธีการให้คะแนน
ความรู้บางส่วนของแบบทดสอบวัดผลสัมฤทธิ์คณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3

ของ

สมจวน มานะก่อ

ได้รับอนุมัติจากบัณฑิตวิทยาลัยให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาการศึกษามหาบัณฑิต สาขาวิชาการวัดผลการศึกษา

ของมหาวิทยาลัยศรีนครินทรวิโรฒ

.....คณบดีบัณฑิตวิทยาลัย

(รองศาสตราจารย์ ดร.สมชาย สันติวัฒนกุล)

วันที่..... เดือน.....พ.ศ. 2551

คณะกรรมการควบคุมปริญญานิพนธ์

คณะกรรมการสอบปากเปล่า

.....ประธาน

(ผู้ช่วยศาสตราจารย์ ดร.องอาจ นัยพัฒน์)

.....ประธาน

(รองศาสตราจารย์ชูศรี วงศ์รัตนะ)

.....กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร. นิคม ตั้งคะพิภพ)

.....กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.องอาจ นัยพัฒน์)

.....กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.นิคม ตั้งคะพิภพ)

.....กรรมการ

(รองศาสตราจารย์นิภา ศรีไพโรจน์)

ประกาศขอบคุณ

ปริญญานิพนธ์ฉบับนี้ สำเร็จได้ด้วยความอนุเคราะห์อย่างยิ่ง จากผู้ช่วยศาสตราจารย์ ดร.องอาจ นัยพัฒน์ ประธานกรรมการควบคุมปริญญานิพนธ์ ผู้ช่วยศาสตราจารย์ ดร.นิคม ตังคะภิภพ กรรมการควบคุมปริญญานิพนธ์ที่ได้กรุณาช่วยเหลือ ให้คำแนะนำปรึกษา ทั้งยังได้แก้ไขข้อบกพร่องต่างๆ ด้วยความเอาใจใส่เป็นอย่างดีเสมอมา

ขอกราบขอบพระคุณ รองศาสตราจารย์ฐศรี วงศ์รัตนะ และรองศาสตราจารย์ภา ศรีไพโรจน์ ที่ได้กรุณาเป็นกรรมการในการสอบปากเปล่าปริญญานิพนธ์ และให้คำปรึกษา ปริญญานิพนธ์ฉบับนี้สมบูรณ์ยิ่งขึ้นและที่สำคัญยิ่งผู้วิจัยขอกราบระลึกถึงพระคุณอาจารย์ ภาควิชาการวัดผลและวิจัยการศึกษาทุกท่านที่ได้อบรมสั่งสอนให้ความรู้ ซึ่งเป็นประโยชน์ต่อการ ทำปริญญานิพนธ์ฉบับนี้จนสำเร็จลุล่วง

ขอกราบขอบพระคุณ อาจารย์ณรงค์พจน์ มีสรลักษณ์ อาจารย์เกดิษฐ์ จันทขจร อาจารย์สุกัลยา ฉายสุวรรณ รองศาสตราจารย์สัญญา รัตนวรารักษ์ และอาจารย์วิรัช สุกใสที่กรุณา ตรวจสอบคุณภาพเครื่องมือที่ใช้ในการวิจัย รวมทั้งให้ข้อเสนอแนะในการปรับปรุงเครื่องมือที่ใช้ ในการเก็บข้อมูล

ขอกราบขอบพระคุณ ท่านผู้บริหาร คุณครูและนักเรียนกลุ่มตัวอย่างของโรงเรียนต่าง ๆ ในเขตพื้นที่การศึกษาปราจีนบุรี ที่ให้ความร่วมมือในการเก็บรวบรวมข้อมูลในการวิจัยเป็นอย่างดี

ขอกราบขอบพระคุณพี่ ๆ เพื่อน ๆ น้อง ๆ ทั้งที่ทำงานและมหาวิทยาลัยทุกคนที่เป็น กำลังใจและคอยช่วยเหลือผู้วิจัยเสมอมา

ขอกราบขอบพระคุณ คุณแม่สีดา มานะก่อ พี่ ๆ น้องและครอบครัวที่ให้กำลังใจ ห่วงใย เป็นแรงบันดาลใจ ทำให้ประสบความสำเร็จในการศึกษาในการวิจัยครั้งนี้

คุณค่าและประโยชน์จากปริญญานิพนธ์ผู้วิจัยขอมอบแด่ทุกท่านที่ทำให้ผู้วิจัย ประสบความสำเร็จในวันนี้

สมจวน มานะก่อ

สารบัญ

บทที่	หน้า
1 บทนำ.....	1
ภูมิหลัง.....	1
วัตถุประสงค์การวิจัย	8
ความสำคัญของวิจัย	8
ขอบเขตการวิจัย	9
กรอบแนวคิดการวิจัย	10
สมมติฐานการวิจัย	11
นิยามศัพท์เฉพาะ.....	11
2 เอกสารและงานวิจัยที่เกี่ยวข้อง.....	14
ตอนที่ 1 ทฤษฎีการตอบสนองข้อสอบ.....	14
แนวคิดพื้นฐานทฤษฎีการตอบสนองข้อสอบ.....	14
หลักการของทฤษฎีการตอบสนองข้อสอบ.....	16
ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ.....	16
โมเดลการตอบสนองข้อสอบ.....	17
คุณสมบัติของความไม่แปรเปลี่ยนของค่าพารามิเตอร์.....	29
ฟังก์ชันสารสนเทศของข้อสอบและแบบทดสอบ.....	30
ตอนที่ 2 วิธีการตรวจให้คะแนน.....	35
ตอนที่ 3 ความรู้เกี่ยวกับแบบทดสอบเลือกตอบ.....	46
ตอนที่ 4 การตรวจสอบคุณภาพของแบบทดสอบ.....	51
ตอนที่ 5 งานวิจัยที่เกี่ยวข้อง.....	58
3 วิธีดำเนินการวิจัย.....	70
ประชากรและกลุ่มตัวอย่าง.....	70
การสร้างเครื่องมือที่ใช้ในการวิจัยและขั้นตอนการสร้างเครื่องมือ.....	73
วิธีการดำเนินการทดลองและเก็บรวบรวมข้อมูล.....	79
การวิเคราะห์ข้อมูลและสถิติที่ใช้ในการวิเคราะห์ข้อมูล.....	80

สารบัญ(ต่อ)

บทที่	หน้า
4 ผลการวิจัย.....	85
สัญลักษณ์ที่ใช้ในการวิเคราะห์ข้อมูล.....	85
การเสนอผลการวิเคราะห์ข้อมูล.....	85
ผลการเปรียบเทียบความเที่ยงตรงตามสภาพระหว่างวิธีการตรวจให้ คะแนนความรู้บางส่วนแบบด้วยวิธีการของการทดสอบความ น่าจะเป็นกับการตรวจให้คะแนนตามความยากประจำชั้น.....	90
ผลการเปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบฟังก์ชันสารสนเทศ ของแบบทดสอบระหว่างวิธีการตรวจให้คะแนน ความรู้บางส่วน ด้วยวิธีการของการทดสอบความน่าจะเป็นกับการตรวจให้คะแนน ตามความยากประจำชั้น.....	91
5 สรุป อภิปรายผลและข้อเสนอแนะ.....	98
การวิเคราะห์ข้อมูล.....	100
การสรุปผลการวิจัย.....	100
อภิปรายผล.....	101
ข้อเสนอแนะ.....	103
บรรณานุกรม.....	104
ภาคผนวก.....	111
ภาคผนวก ก คุณภาพของเครื่องมือที่ใช้ในการวิจัย.....	112
ภาคผนวก ข ตัวอย่างเครื่องมือที่ใช้ในการวิจัย.....	117
ภาคผนวก ค แผนภาพสารสนเทศข้อสอบ.....	124
ภาคผนวก ง รายงานผู้เชี่ยวชาญ.....	153
ประวัติย่อผู้วิจัย.....	155

บัญชีตาราง

ตาราง	หน้า
1 การคำนวณค่าฟังก์ชันสารสนเทศ ค่าสารสนเทศสูงสุดของข้อสอบและตำแหน่งที่มี θ ที่มีสารสนเทศของข้อสอบสูงสุด สำหรับโมเดลการตอบสนองข้อสอบ โลจิสติก 1 , 2 , 3 พารามิเตอร์.....	31
2 การเลือกตอบอย่างเสรี (Free Choice Test) และจะได้คะแนนตามเงื่อนไขถ้าข้อใด เว้นว่างไว้จะไม่ได้คะแนน.....	38
3 ข้อมูลหน่วยสมาชิกของประชากร ชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2549.....	71
4 ข้อมูลการสุ่มตัวอย่างจำนวนนักเรียนชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2549.....	72
5 เฉลยและอธิบายการให้คะแนนตามวิธีการตรวจให้คะแนน.....	74
6 การวิเคราะห์เนื้อหาในการสร้างข้อสอบ.....	77
7 ค่าเฉลี่ยและค่าส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบจำแนกตามวิธีการตรวจ ให้คะแนน.....	86
8 ผลการวิเคราะห์องค์ประกอบของแบบทดสอบวัดผลสัมฤทธิ์คณิตศาสตร์.....	87
9 ความเที่ยงตรงตามสภาพจากการสอบจำแนกตามวิธีการตรวจให้คะแนน.....	90
10 ฟังก์ชันสารสนเทศของข้อสอบ จำแนกตามรายข้อ และวิธีการตรวจให้คะแนน.....	91
11 ฟังก์ชันสารสนเทศของข้อสอบเฉลี่ย จำแนกตามความสามารถของผู้สอบ และวิธีการตรวจให้คะแนน.....	93
12 เปรียบเทียบประสิทธิภาพสัมพัทธ์ : $RE(\theta)$ ณ ตำแหน่ง θ เดียวกัน.....	94
13 ฟังก์ชันสารสนเทศของแบบทดสอบ ตามวิธีการตรวจให้คะแนน.....	96
14 ผลการตรวจสอบความเที่ยงตรงด้านเนื้อหา (Content Validity) โดยผู้เชี่ยวชาญ...	112
15 ค่าความยาก ค่าอำนาจจำแนกรายข้อ.....	115

บัญชีภาพประกอบ

ภาพประกอบ	หน้า
1 กรอบแนวคิดการวิจัย.....	11
2 โค้งลักษณะข้อสอบแบบ 1 พารามิเตอร์.....	18
3 โค้งลักษณะข้อสอบแบบ 2 พารามิเตอร์.....	19
4 โค้งลักษณะข้อสอบแบบ 3 พารามิเตอร์.....	20
5 โค้งรายการคำตอบของข้อคำถามที่มี 4 รายการคำตอบที่วิเคราะห์โดยใช้ PCM....	25
6 โค้งลักษณะข้อสอบและการแจกแจงความสามารถของผู้สอบ 2 กลุ่ม.....	29
7 ค่าพารามิเตอร์ของผู้สอบที่คำนวณได้จากแบบสอบ 2 ฉบับที่มีความแตกต่างกัน...	30
8 โค้งสารสนเทศของข้อสอบ.....	32
9 ขั้นตอนในการสร้างแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์.....	76
10 ค่าไอเกน (Eigen) ของการตรวจให้คะแนนของการทดสอบความน่าจะเป็น.....	89
11 ค่าไอเกน (Eigen) ของการตรวจให้คะแนนตามความยากประจำชั้น.....	89
12 เปรียบเทียบโค้งสารสนเทศของข้อสอบของการให้คะแนน 2 วิธี.....	95
13 ฟังก์ชันสารสนเทศของแบบทดสอบจำแนกตามวิธีการตรวจให้คะแนนด้วยวิธีการ ของการทดสอบความน่าจะเป็น.....	97
14 ฟังก์ชันสารสนเทศของแบบทดสอบจำแนกตามวิธีการตรวจให้คะแนนด้วยวิธีตาม ความยากประจำชั้น.....	97
15 โค้งลักษณะข้อสอบที่มีการตรวจให้คะแนนด้วยวิธีการของการทดสอบ ความน่าจะเป็น.....	124

บทที่ 1

บทนำ

ภูมิหลัง

การสอบทางการศึกษาส่วนใหญ่เป็นการตรวจสอบความสามารถของผู้เรียนอันเป็นผลจากการเรียนรู้ที่เกิดขึ้น เพื่ออธิบายความสามารถในปัจจุบันหรือทำนายความสามารถ ความพร้อมของการเรียนรู้ในอนาคต ความสามารถของบุคคลเป็นคุณลักษณะภายในที่ต้องอาศัยวิธีการวัดทางอ้อม โดยการตอบสนองต่อสิ่งเร้าที่จัดขึ้นแล้วนำผลการตอบสนองที่ได้ ทำการสรุปอ้างอิงไปยังคุณลักษณะภายในสิ่งที่มีวัด (คณิต ไชมุขต์. 2548: 1) คะแนนจึงเป็นแหล่งสำคัญของการอ้างอิงในการวัดทางการศึกษาเพราะคะแนนเป็นผลการตอบข้อสอบซึ่งไม่เพียงแต่เป็นฟังก์ชันที่เกี่ยวกับข้อสอบ งานที่กำหนดหรือเงื่อนไขของสิ่งเร้าต่าง ๆ เท่านั้น แต่ยังเกี่ยวกับวิธีการตอบของผู้สอบและเกี่ยวกับบริบทของการวัดด้วย การอ้างอิงถึงความเที่ยงตรงและความเชื่อมั่นของการวัดใด ๆ จึงต้องอ้างอิงจากคะแนนนี้ (พรทิพย์ ไชยโส. 2534 อ้างอิงจาก Messick cited in Linn. 1981) อย่างไรก็ตามคะแนนที่วัดได้ทางการศึกษาเป็นปริมาณที่รวมเอาความคลาดเคลื่อนที่เกิดจากแหล่งต่าง ๆ ทั้งข้อสอบที่ใช้ลักษณะของผู้สอบ สถานการณ์การสอบเข้าไว้ด้วยทุกครั้งทีวัด คะแนนที่ได้จึงไม่ใช่ค่าที่แท้จริงของสิ่งที่มีวัด (Lord and Novick. 1968) การให้คะแนนเป็นองค์ประกอบสำคัญของการวัดนอกเหนือจากวิธีการตอบ (พรทิพย์ ไชยโส. 2534: 21) ปัจจุบันได้มีนักทฤษฎีวัดผลอ้างอิงถึงความสัมพันธ์ระหว่างความสามารถที่แท้จริงกับลักษณะการตอบสนองข้อสอบ เพื่อทำนายลักษณะที่ซ่อนอยู่ในตัว เรียกว่า ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory : IRT) เป็นทฤษฎีที่มีพื้นฐานความเชื่อว่าพฤติกรรมตอบสนองต่อข้อสอบของผู้สอบเป็นสิ่งสังเกตได้โดยตรงว่าถูกหรือผิดจะถูกกำหนดโดยคุณลักษณะภายใน (Trait) หรือความสามารถ (Ability) ที่อยู่ในตัวบุคคลซึ่งเป็นสิ่งที่ไม่สามารถสังเกตได้โดยตรง สามารถอธิบายความสัมพันธ์ของคุณลักษณะภายในหรือความสามารถที่มีอยู่ในตัวบุคคลกับพฤติกรรมตอบสนองข้อสอบโดยใช้โค้งลักษณะข้อสอบ (Item Characteristic Curve : ICC) ซึ่งกำหนดลักษณะของข้อสอบด้วยพารามิเตอร์ความยาก (b) ค่าอำนาจจำแนก (a) และโอกาสการเดา (c) มีลักษณะเป็นฟังก์ชันทางคณิตศาสตร์ เรียกว่า ฟังก์ชันโลจิสติก (Logistic Function) หรือใกล้เคียงกับฟังก์ชันปกติสะสม (Normal Ogive Function) (ศิริชัย กาญจนวาสี. 2545: 45 – 46) ฟังก์ชันโลจิสติก (Logistic Function) มีลักษณะของสูตรทางคณิตศาสตร์ วิธีการคำนวณง่าย สะดวกกว่าและยังมีความทนทานต่อความคลาดเคลื่อนที่เกิดขึ้นกับผู้สอบที่มีความสามารถสูงจะตอบข้อสอบได้ดีกว่า (คณิต ไชมุขต์. 2548: 5) ตลอดจนความไม่แปรเปลี่ยนไปตามกลุ่มตัวอย่างของค่าพารามิเตอร์ของข้อสอบ ทำให้ค่าสารสนเทศเหมาะสมที่จะใช้เป็นตัวชี้วัดคุณภาพของข้อสอบ

และแบบทดสอบแทนการหาค่าความเชื่อมั่นและความคลาดเคลื่อนมาตรฐานของการวัดตามทฤษฎีการทดสอบแบบดั้งเดิม (คณิต ไช่มุกต์. 2548: 10 อ้างอิงจาก Hambleton. 1965) จากการวิเคราะห์ด้วยทฤษฎีทดสอบแบบดั้งเดิม (Classical Test Theory : CTT) พบจุดอ่อนบางประการที่ไม่สามารถแก้ปัญหาโอกาสในการเดา ความคลาดเคลื่อนในการวัดและไม่ได้ให้ความสนใจต่อสถานการณ์หรือเงื่อนไขในการวัด ตลอดจนปฏิสัมพันธ์ระหว่างผู้สอบกับเงื่อนไขในการวัดที่สามารถส่งผลต่อความน่าเชื่อถือของผลการวัด ค่าพารามิเตอร์ของข้อสอบจะผันแปรตามกลุ่มผู้สอบและคะแนนที่สังเกตได้หรือค่าประมาณความสามารถของผู้ตอบไม่เป็นอิสระหรือขึ้นอยู่กับข้อสอบ และแบบทดสอบที่นำมาใช้ (ศิริชัย กาญจนวาสี. 2545: 45) ซึ่ง แฮมเบลตัน (Hambleton and Swaminathan.1985) แสดงให้เห็นว่าค่าความยากและค่าอำนาจจำแนก แปรเปลี่ยนไปตามกลุ่มผู้สอบ การเปรียบเทียบความสามารถระหว่างบุคคลจะเปรียบเทียบได้ต่อเมื่อใช้แบบทดสอบฉบับเดียวกันและความคลาดเคลื่อนมาตรฐานในการวัดของแต่ละบุคคลจะเท่ากัน การนำทฤษฎีตอบสนองข้อสอบ (Item Response Theory : IRT) ไปใช้ในการแก้ปัญหการวัดที่สำคัญ ๆ โดยเฉพาะการแก้ปัญหาที่ไม่สามารถแก้ไขด้วยทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory : CTT) เนื่องจากทฤษฎีการตอบสนองข้อสอบสามารถประมาณค่าความสามารถหรือคุณลักษณะของผู้ตอบเป็นอิสระจากชุดของข้อคำถามไม่ว่าจะเลือกกลุ่มตัวอย่างย่อยกลุ่มใดมาเพื่อใช้ในการประมาณค่าพารามิเตอร์ ข้อความค่าพารามิเตอร์นั้นจะคงที่ (องอาจ นัยพัฒน์. 2535: 3) คุณสมบัตินี้เป็นประโยชน์อย่างมากสำหรับการเลือกโมเดลเพื่อนำ ไปใช้ในการสร้างและพัฒนาแบบทดสอบในปัจจุบัน นอกจากการเลือกโมเดลที่เหมาะสมแล้ว วิธีการประมาณค่าพารามิเตอร์ของข้อคำถามหรือข้อความ และคุณลักษณะภายในตัวบุคคลที่ต้องการวัดเป็นอีกกระบวนการหนึ่งที่ต้องเลือกให้เหมาะสมกับสภาพการวัด ไวท์และมาสเตอร์ (Wright and Masters.1982) ได้พัฒนาและขยายแนวคิดการประมาณค่าพารามิเตอร์ที่ได้จากการตรวจให้คะแนนแบบสองค่า (Dichotomous) มาเป็นการประมาณค่าพารามิเตอร์ที่ได้จากการตรวจให้คะแนนคะแนนแบบหลายค่าหรือการตรวจให้คะแนนมากกว่า 2 ค่า (Polytomous Item Response Theory) รวมถึงการตรวจให้คะแนนความรู้บางส่วน (Partial Credit Model : PCM) ที่พัฒนา โดยมาสเตอร์ (Master. 1982) ในการวิเคราะห์ข้อสอบหรือข้อคำถามที่มีกระบวนการตอบหลายลำดับขั้น (Ordered Response Categories) ซึ่งจำเป็นต้องมีการตรวจให้คะแนนการตอบถูกต้องหรือตอบถูกบางส่วนในแต่ละลำดับขั้นของกระบวนการตอบ (Ordered Performance Levels) ใช้กับข้อสอบวัดผลสัมฤทธิ์ทางคณิตศาสตร์ ซึ่งมีการให้คะแนนแบบคำตอบถูกบางส่วน มีลักษณะเป็นโมเดลที่พัฒนาขยายต่อจากโมเดลการตอบสนองข้อสอบที่มี 1 พารามิเตอร์ (1 – parameter Model) ที่คล้ายโมเดลของราสช์ (Rasch Model) และใช้หลักการคำนวณความน่าจะเป็นของการตอบแต่ละระดับขั้นการตอบโดยตรงแบบขั้นตอนเดียว (Direct IRT

Method) ลักษณะข้อคำถามแต่ละข้อ (i) อธิบายด้วยค่าพารามิเตอร์ระดับความยาก ระหว่างรายการคำตอบที่อยู่ถัดไป (item step difficulty : δ_{ij}) โดยทุกข้อมีค่าพารามิเตอร์ความชันเท่ากันข้อคำถาม i มีคะแนน $x = 0.1 \dots, m_i$ โดยมีจำนวนรายการคำตอบ $K_i = M_i + 1$ (ศิริชัย กาญจนวาสี. 2545: 83) เช่น ข้อสอบคณิตศาสตร์ข้อหนึ่งต้องการคำตอบของ $\sqrt{(9/0.3) - 5} = ?$ ในการตรวจให้คะแนนได้ กำหนดลำดับขั้นของคะแนนผลการตอบดังนี้ (Wright and Masters. 1982: 41)

เมื่อไม่สามารถตอบขึ้น (step) ได้ถูกต้อง	ได้	0	คะแนน
ทราบว่า $9 / 0.3 = 30$ (ผ่านขั้นที่ 1)	ได้	1	คะแนน
ทราบว่า $30 - 5 = 25$ (ผ่านขั้นที่ 2)	ได้	2	คะแนน
ทราบว่า $\sqrt{25} = 5$ (ผ่านขั้นที่ 3)	ได้	3	คะแนน

จะเห็นว่าข้อสอบนี้มีการให้คะแนนความรู้บางส่วน โดยกำหนดลำดับขั้นของคะแนนการตอบ (Ordered Performance Levels) ไว้ 4 ระดับ คือ ให้ 0, 1, 2, 3 คะแนน เนื่องจากการคำนวณหาคำตอบของข้อสอบนี้ ผู้ตอบที่จะตอบในขั้นท้าย ๆ ได้ถูกต้องจะต้องสามารถผ่านหรือมีความรู้จนสามารถตอบถูกในขั้นแรก ๆ มาแล้ว (ยกเว้นตอบถูกโดยโอกาสบังเอิญ) ดังนั้น การที่ผู้ตอบตอบได้ 1 คะแนน แสดงว่าสามารถผ่านหรือตอบข้อสอบได้ถูกต้องในขั้นที่ 1 คือ ทราบว่า $9 / 0.3 = 30$ และการที่ได้ 2 คะแนน แสดงว่า สามารถทำข้อสอบได้ถูกต้องต่อไปอีกหนึ่งขั้น คือ ทราบว่า $30 - 5 = 25$ และการที่ได้ 3 คะแนน ก็อาศัยหลักการในทำนองเดียวกัน การให้คะแนนความรู้บางส่วนของผู้ตอบแสดงถึงตำแหน่งความรู้จริงของผู้ตอบซึ่งมีส่วนช่วยลดความคลาดเคลื่อนจากแหล่งการเดาสุ่มของผู้ตอบทำให้คะแนนมีความเที่ยงตรงมากขึ้น ขณะเดียวกันก็ส่งผลต่อความเชื่อมั่นของกระบวนการวัด (พรทิพย์ ไชยโส. 2534: 9 อ้างอิงจาก Frary. 1980) คะแนนความรู้บางส่วนแบ่งเป็นประเภทใหญ่ ดังนี้ (Simon, Budescu and Nevo. 1997: 65 – 88)

1. การให้น้ำหนักแก่ข้อสอบที่แตกต่างกัน (Differential Item Weighting) ลักษณะข้อสอบที่มีคุณภาพจะได้ให้น้ำหนักมากกว่าข้อสอบที่ไม่มีคุณภาพ โดยใช้ผลที่ได้จากการวิเคราะห์ข้อสอบเป็นเกณฑ์ คือ ค่าความยาก ความสามารถ ความเชื่อมั่น ความแปรปรวนหรือการพิจารณาของผู้เชี่ยวชาญ หวังและสแตนเลย์ (Wang and Stanley. 1970: 21 – 35) ให้ข้อสรุปว่าการให้น้ำหนักแก่ข้อสอบโดยใช้ทฤษฎีการวัดแบบดั้งเดิม ทำให้น้ำหนักของคะแนนแต่ละข้อมีค่าไม่คงที่เปลี่ยนแปลงไปตามกลุ่มตัวอย่าง จึงควรให้คะแนนตามทฤษฎีการตอบสนองของข้อสอบ เพื่อกำหนดน้ำหนักคะแนนที่เหมาะสม

2. การให้น้ำหนักแก่ตัวเลือกที่แตกต่างกัน (Differential Option Weighting) ซึ่ง หวังและสแตนเลย์ (Wang and Stanley. 1970: 663 – 705) พบว่า การให้น้ำหนักคะแนนแก่ตัวเลือกที่แตกต่างกันมีประโยชน์บ้างในแง่ที่ช่วยเพิ่มค่าความเที่ยงตรงภายในการเปรียบเทียบค่าความ

เที่ยงตรงและความเชื่อมั่นระหว่างแบบการให้น้ำหนักแก่ตัวเลือกกับแบบธรรมดา พบว่ามีค่าไม่คงที่ของความแตกต่างอย่างมีนัยสำคัญ วิธีการเหล่านี้จึงไม่เป็นที่นิยม เนื่องจากการพัฒนาเพื่อให้ได้มาซึ่งน้ำหนักใช้ทรัพยากรมาก การคำนวณการให้คะแนนมีวิธีการยุ่งยากและยากในการอธิบายและตัดสินใจคะแนนแก่ผู้สอบ

3. การเปลี่ยนแปลงโครงสร้างของข้อสอบ (Changing the Item Structure) วิธีนี้มีรูปแบบของโครงสร้างข้อสอบ คำอธิบายที่แตกต่างจากแบบเลือกตอบธรรมดา ดังนี้

ข้อสอบที่มี 2 ตัวเลือก คือ ถูกและผิด ผู้สอบจะต้องเลือกคำตอบอย่างใดอย่างหนึ่งคะแนนที่ได้ ได้จากผลรวมของจำนวนคำตอบที่ถูกต้อง เนื่องจากการตอบแต่ละข้อใช้เวลาน้อย จึงต้องเพิ่มจำนวนข้อให้มากขึ้นกว่าข้อสอบแบบเลือกตอบธรรมดา

ข้อสอบที่มีตัวเลือกถูกหลาย ๆ ตัวเลือก (มากกว่า 1) โดยต้องมีคำชี้แจงอธิบายให้แก่ผู้สอบซึ่งอาจจะระบุจำนวนตัวเลือกที่ถูกหรือไม่ระบุก็ได้ คะแนนที่ได้เท่ากับผลรวมของการระบุตัวเลือกถูกได้ถูกต้อง ในบางกรณีมีการหักคะแนนการตอบที่ตอบตัวเลือกผิดโดยมีจุดประสงค์เพื่อลดการเดาและประเมินความรู้บางส่วน แบบทดสอบลักษณะนี้ได้มีการพัฒนาต่อมา เช่น แบบทดสอบที่มีตัวเลือกถูกทุกข้อ ผิดทุกข้อ หรือถูกเฉพาะตัวเลือกหนึ่งกับอีกตัวเลือกหนึ่ง

ข้อสอบที่มีกลุ่มของข้อสอบและกลุ่มของคำตอบที่เรียกว่าข้อสอบแบบจับคู่ โดยผู้สอบต้องเลือกคำตอบที่ถูกคู่กับคำถามที่กำหนดให้ ถ้าจำนวนคำถามเท่ากับจำนวนคำตอบ เรียกว่าจับคู่อย่างง่าย แต่ถ้าจำนวนคำตอบมากกว่าจำนวนคำถาม เรียกว่าจับคู่แบบพหุ ความแตกต่างของ 2 แบบ คือ การจับคู่แบบพหุช่วยลดโอกาสในการเดามากกว่า เนื่องจากมีจำนวนตัวเลือกมากกว่า

4. การเปลี่ยนแปลงวิธีการตอบ (Changing the Response Method) เป็นวิธีการตอบโดยใช้น้ำหนักที่ให้ผู้สอบประเมินตนเอง เป็นวิธีที่ดีที่สุดในการได้มาซึ่งสารสนเทศเกี่ยวกับความรู้ของผู้สอบในแต่ละข้อ ผู้สอบจะให้สารสนเทศอย่างสมบูรณ์เกี่ยวกับน้ำหนัก วิธีการให้คะแนนและการควบคุมระดับความมั่นใจจะช่วยลดการเดาและความคลาดเคลื่อนในการวัด วิธีนี้ใช้รูปแบบข้อสอบเหมือนแบบทดสอบชนิดเลือกตอบทั่วไปเพียงแต่คำสั่ง วิธีการตอบและการให้คะแนนแตกต่างจากแบบทดสอบชนิดเลือกตอบธรรมดา แต่ละวิธีการไม่ยุ่งยาก แบ่งได้ 6 วิธี (เอมอร์ จังศิริพรปกรณ์. 2546: 24) 1. วิธีการแก้การเดา (Correction for guessing) 2. วิธีการให้คะแนนจากการตัดตัวลง (Elimination Scoring) เป็นวิธีการให้ผู้สอบตัดตัวลงออกไปได้บ้างอย่างถูกต้องแม้จะไม่ทราบว่าคำตอบถูกต้องคืออะไร (Coombs, Milhlland and Womer.1956: 13 -37) ซึ่งเอมอร์ จังศิริพรปกรณ์ (2546: 84) พบว่า วิธีการให้คะแนนจากการตัดตัวลงที่เปิดโอกาสให้ตอบตัดตัวลง 1 ตัวหรือตัดตัวลง 2 ตัวก็ได้ทำให้ได้คะแนนรู้บางส่วนมากกว่าวิธีการเลือกชุดตัวถูก 3. วิธีการของการทดสอบความน่าจะเป็น (Probability Testing) เป็นวิธีที่มีความยืดหยุ่นและเปิดโอกาสให้ผู้สอบแสดง

ความรู้บางส่วนมากที่สุดโดยรายงานความน่าจะเป็นในตัวเลือกที่ผู้สอบคิดว่าเป็นคำตอบที่ถูกโดยมีตำแหน่งคะแนนที่เป็นไปได้ 101 ตำแหน่งและความเชื่อมั่นสูงกว่าวิธีเลือกตอบธรรมดา (Hambleton, Robert and Taub. 1970) โคลเลอร์ (Koehler. 1971) พบว่า มีความไม่แตกต่างกันของความเที่ยงตรงและความเชื่อมั่นระหว่างวิธีประเมิน (Number Correct: NC) และวิธีของการทดสอบความน่าจะเป็น (Probability Testing: PT) ซึ่งยังไม่พบข้อสรุปที่ชัดเจน 4. วิธี การแสดงความมั่นใจ (Confidence Marking) (เอมอร์ จังศิริพรปกรณ์. 2546: 5) พบว่า วิธีการนี้ให้ความเชื่อมั่นของแบบทดสอบสูงกว่าการให้คะแนนแบบ 0, 1 แต่วิธีการนี้ต้องการเวลาตอบและการให้คะแนนมากกว่าเดิม 5. วิธีการจัดอันดับอย่างสมบูรณ์ (Complete Ordering) เป็นวิธีเฉพาะของการทดสอบความน่าจะเป็น (PT) คือ แทนที่จะแสดงให้เห็นความน่าจะเป็นในแต่ละตัวเลือกผู้สอบจะต้องจัดอันดับของตัวเลือกซึ่งทำให้ง่ายต่อการนำไปใช้ทำให้สามารถจำแนกระดับความรู้ออกเป็น 3 ระดับ ดังนี้ Full Knowledge คือ การจัดอันดับให้คำตอบถูกอยู่ในตำแหน่งที่สูงที่สุด Partial Knowledge คือการจัดอันดับให้คำตอบถูกอยู่ในตำแหน่งกลาง และ Absence of Knowledge คือ การจัดอันดับให้คำตอบถูกอยู่ในตำแหน่งต่ำที่สุด 6. วิธีการจัดอันดับเป็นบางส่วน (Partial Ordering) เป็นวิธีที่ผสมผสานระหว่างวิธีการให้คะแนนแบบตัดตัวลง (ET) และวิธีการจัดอันดับอย่างสมบูรณ์ (CO) โดยให้ผู้สอบจัดอันดับตัวเลือกที่ไม่สามารถตัดออกได้

การที่จะทำให้ทราบว่าคะแนนที่ได้จากการสอบเกิดจากความรู้ที่แท้จริงวิธีการตรวจให้คะแนนจึงมีบทบาทสำคัญเพื่อให้คะแนนที่ได้จากแบบทดสอบน่าเชื่อถือมากกว่านี้ การตรวจให้คะแนนแบบสองค่า (binary) เช่น การตรวจให้คะแนนแบบ 0 หรือ 1 ถูกหรือผิด เห็นด้วยหรือไม่เห็นด้วยเป็นวิธีการตรวจให้คะแนนกันอย่างแพร่หลายในปัจจุบัน เพราะสามารถวัดความรู้ความสามารถและวัดผลสัมฤทธิ์ได้อย่างกว้างขวาง เนื่องจากมีความเป็นปรนัยในการตรวจให้คะแนนใช้เวลาตรวจน้อย แต่ยังมีข้อจำกัดที่สำคัญ คือ เปิดโอกาสให้มีการตอบถูกโดยการเดาปัญหาการเดาสุ่ม จึงเป็นแหล่งความคลาดเคลื่อนในการวัด นอกจากนี้ ผลการวัดที่ได้ไม่สามารถให้สารสนเทศ ที่เพียงพอที่จะจำแนกผู้ตอบในระดับต่าง ๆ ของความรู้ระหว่างผู้รู้จริง (full knowledge) ผู้ที่มีความรู้บางส่วน (partial knowledge) และผู้ที่ไม่รู้ (absence of knowledge) (เอมอร์ จังศิริพรปกรณ์. 2544: 21) วิธีการแก้ปัญหาเพื่อลดโอกาสในการเดาและเพิ่มสารสนเทศเกี่ยวกับการตอบข้อสอบเลือกตอบให้มากที่สุดโดยใช้วิธีการให้คะแนนที่เป็นองค์ประกอบที่สำคัญของกระบวนการวัดสำหรับวิธีการให้คะแนนที่เหมาะสมจะทำให้คุณสมบัติของการวัดในประเด็นที่เกี่ยวข้องกับความเที่ยงตรงและความเชื่อมั่นสูงขึ้น การให้คะแนนความรู้บางส่วน (Partial Credit) ของผู้ตอบแสดงถึงตำแหน่งจริงของความรู้ผู้ตอบ ซึ่งมีส่วนช่วยลดความคลาดเคลื่อนจากแหล่งการเดาของผู้ตอบทำให้คะแนนมีความเที่ยงตรงมากขึ้นในขณะเดียวกันก็ส่งผลต่อความเชื่อมั่นของ

กระบวนการวัดด้วย (พรทิพย์ ไชยโส. 2534: 9 อ้างอิงจาก Frary. 1980) การวัดผลประเมินผลการเรียนรู้คณิตศาสตร์นั้นช่วยให้สารสนเทศเกี่ยวกับผู้เรียนซึ่งสามารถทำได้โดยใช้เครื่องมือและวิธีการที่เหมาะสม เครื่องมือที่ใช้ในการวัดผลประเมินผลคณิตศาสตร์ที่สอดคล้องกับความรู้ ความคิด ทักษะ กระบวนการและคุณลักษณะที่พึงประสงค์มี 2 ประเภท คือ แบบทดสอบและภาระงานที่ได้รับมอบหมาย (กระทรวง ศึกษาธิการ. 2546: 13) แบบทดสอบวิชาคณิตศาสตร์ที่ใช้อย่างแพร่หลายในปัจจุบันเป็นแบบทดสอบแบบปรนัยและอัตนัย (Scoring, objectivity and subjectivity) โดยเฉพาะแบบทดสอบเลือกตอบที่ใช้วัดได้ครอบคลุมทั้งด้านความรู้ ความคิด หลักการ ทฤษฎี การตัดสินใจ การประเมินตัวแปร การแปลความหมายข้อมูล การแสดงความเข้าใจในธรรมชาติของคณิตศาสตร์ ตลอดจนความสามารถด้านทักษะกระบวนการทางคณิตศาสตร์ การให้คะแนนของแบบทดสอบเลือกตอบคณิตศาสตร์ ส่วนใหญ่จะพิจารณาจากความถูกต้องของการเลือกตอบเป็นสำคัญ คือ ตอบถูกได้ 1 คะแนน ตอบผิดได้ 0 คะแนน เนื่องจากแบบทดสอบเลือกตอบมีความเป็นปรนัยสูงการตรวจให้คะแนนง่ายและได้ผลตรงกัน สะดวกใช้เวลาน้อย แต่ยากต่อการสร้างคำถามที่ชัดเจน เป็นปรนัยตรงประเด็นและคำถามที่วัดความคิดระดับสูงและทักษะกระบวนการ ใช้วัดผลการเรียนรู้ได้ในระดับความรู้ ความจำและวัดได้เพียงด้านใดด้านหนึ่งเท่านั้น (กระทรวงศึกษาธิการ. 2546: 35) ซึ่งตรงกับคำกล่าวของ กรอนลันด์ (Gronlund. 1982: 73) ว่าข้อสอบปรนัยเหมาะสมกับการวัดความรู้ ความจำ ความเข้าใจ การนำไปใช้ไม่เหมาะในการสังเคราะห์ ประเมินค่า ใช้ข้อสอบจำนวนมากข้อครอบคลุมเนื้อหาได้กว้างขวาง สมบูรณ์เป็นตัวแทนที่ดีของเนื้อหาทั้งหมดทำให้ยากและเสียเวลาในการเขียนข้อสอบ การให้คะแนนมีความเป็นปรนัยและความเชื่อมั่นสูงเพราะคะแนนที่ได้มาจากความสามารถในการอ่านและการเดา ในขณะที่แบบทดสอบแบบอัตนัยซึ่ง กรอนลันด์ (Gronlund. 1982: 73) กล่าวว่า เป็นข้อสอบที่ให้เสรีภาพเป็นอย่างมากในการตอบ สามารถวัดพฤติกรรมขั้นสูงของนักเรียนได้ดี สร้างง่าย ส่งเสริมการเขียนและการแสดงความคิดเห็น ลดค่าการเดา ไม่เหมาะในการวัดความรู้ความจำ ไม่สามารถเป็นตัวแทนของเนื้อหาทั้งหมดได้ ยากในการเขียนข้อสอบที่ดี ให้คะแนนยาก ความเชื่อมั่นต่ำและผู้ตอบสามารถเขียนตอบตลวงตอบว่ารู้ได้ แม้ว่าข้อสอบจะมีลักษณะยั่วให้ผู้ตอบจัดกระบวนการผสมผสานความคิดของตนเองได้ ซึ่งตรงกับคำกล่าวของ สติกกินส์ (Striggins. 1994: 154-155) ว่าข้อสอบอัตนัยเป็นข้อสอบที่มีปัญหาการสร้างและการตรวจให้คะแนน กล่าวคือ การสร้างมีเป้าหมายไม่ชัดเจนไม่รู้จะเน้นความรู้ใดผิดวัตถุประสงค์ในการใช้ข้อสอบ การเขียนตอบของนักเรียนไม่มีประสิทธิภาพ เรื่องที่ถามไม่เหมาะสมคำถามไม่มีคุณภาพ การให้คะแนนมีเกณฑ์ที่ไม่เหมาะสม ผู้ตรวจให้คะแนนไม่มีทักษะในการตรวจให้คะแนนใช้เวลาตรวจมาก แม้ว่าข้อสอบอัตนัยจะให้เสรีภาพในการตอบ สามารถวัดสมรรถภาพทางสมองขั้นสูงและเกิดการผสมผสานทางความคิด กระบวนการคิด อธิบายจุดยืนของตนเองได้ดี แต่ยังขาดความ

เที่ยงตรงด้านเนื้อหา เสียเวลาในการตรวจให้คะแนน คะแนนมีความเชื่อมั่นต่ำและคะแนนที่ได้อาจไม่สามารถบอกระดับผลสัมฤทธิ์ทางการเรียนที่แท้จริงได้ ในขณะที่ข้อสอบปรนัยหรือแบบทดสอบแบบเลือกตอบที่มีการตรวจให้คะแนนแบบ ตอบถูกให้ 1 ตอบผิดให้ 0 มีความเป็นปรนัยในการตรวจให้คะแนน ใช้เวลาตรวจน้อย เป็นตัวแทนที่ดีของเนื้อหาทั้งหมดการให้คะแนนมีความเป็นปรนัยและความเชื่อมั่นสูง

จากข้อค้นพบต่าง ๆ ทำให้ผู้วิจัยมีความสนใจในการตรวจให้คะแนนของคณิตศาสตร์ที่สามารถวัดสมรรถภาพทางสมองขั้นสูงได้ ไม่เปิดโอกาสในการเดา ใช้เวลาตรวจน้อย มีเกณฑ์การให้คะแนนที่เหมาะสม แบบทดสอบครอบคลุมเนื้อหาและสามารถให้สารสนเทศแก่ผู้สอบเพียงพอที่จะจำแนกผู้ตอบในระดับต่าง ๆ ของความรู้ระหว่างผู้รู้จริง ผู้ที่มีความรู้บางส่วนและผู้ที่ไม่รู้ได้ (เอมอร จังศิริพรปกรณ์.2544: 21) ตลอดจนจากการศึกษาทฤษฎีและงานวิจัยที่ผ่านมา พบว่าส่วนใหญ่จะเปรียบเทียบคุณภาพของแบบทดสอบในด้านความเชื่อมั่นและความเที่ยงตรง จากการให้คะแนนความรู้บางส่วนหลาย ๆ วิธี โดยใช้ทฤษฎีการวัดแบบดั้งเดิม ซึ่งมีข้อจำกัดในเรื่องแหล่งความคลาดเคลื่อนในการวัดและค่าสถิติที่ได้แปรเปลี่ยนไปตามกลุ่มตัวอย่างที่ใช้ ผู้วิจัยเห็นว่าทฤษฎีการตอบสนองข้อสอบที่มีการตรวจให้คะแนนมากกว่า 2 ค่า (Polytomous Item Response Model) โดยเฉพาะการให้คะแนนความรู้บางส่วน (Partial Credit Model : PCM) ที่เป็นโมเดลที่แสดงถึงการให้คะแนนที่ใกล้เคียงกับความรู้ที่แท้จริงของผู้ตอบ (พรทิพย์ ไชยโส. 2534: 3) สามารถอธิบายถึงความสัมพันธ์ระหว่างความสามารถของผู้สอบและโอกาสของการตอบได้ถูกต้องด้วยฟังก์ชันทางคณิตศาสตร์ที่นำเสนอในรูปแบบโค้งลักษณะข้อสอบ จากวิธีการประมาณค่าพารามิเตอร์ความสามารถของผู้ตอบ (สมถวิล วิจิตรวรรณ. 2542: 22) ตลอดจนมีความเหมาะสมในการวัดความสามารถทางสติปัญญา เช่น วัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ ที่เชื่อว่าผลการตอบสนองข้อสอบขึ้นอยู่กับลักษณะของข้อสอบ และความสามารถของผู้สอบ (ศิริชัย กาญจนวาสี. 2545: 75) ด้วยวิธีการเปรียบเทียบความเที่ยงตรงตามสภาพวิธีการให้คะแนนความรู้บางส่วนที่ใช้กับแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ด้วยการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็น (Probability Testing) กับการตรวจให้คะแนนตามความยากประจำข้อ ว่าวิธีใดให้ความเที่ยงตรงตามสภาพหรือความสามารถที่แท้จริง ฟังก์ชันสารสนเทศของแบบทดสอบ ฟังก์ชันสารสนเทศของข้อสอบ มากที่สุดและวิธีการตรวจให้คะแนนต่างกันให้ผลการวัดที่สอดคล้องกันหรือไม่ ผู้วิจัยคาดว่า ผลที่ได้จากการวิจัยนี้จะทำให้ได้ความรู้ในการที่จะใช้เครื่องมือการวัดผลทางการศึกษาให้มีคุณภาพในกระบวนการวัด ตลอดจนนำแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ไปวินิจฉัยข้อบกพร่องหรือให้สารสนเทศแก่นักเรียนชั้นมัธยมศึกษาปีที่ 3 เพื่อเป็นข้อมูลย้อนกลับในการจัดการเรียนรู้และตัดสินผลการเรียนรู้ของนักเรียน

วัตถุประสงค์ของการวิจัย

1. เพื่อเปรียบเทียบความเที่ยงตรงตามสภาพ (Concurrent Validity) ของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ที่มีวิธีการตรวจให้คะแนนความรู้บางส่วน ด้วยวิธีการของการทดสอบความน่าจะเป็นกับการให้คะแนนตามความยากประจำชั้น
2. เพื่อเปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบ (Item Information) และฟังก์ชันสารสนเทศของแบบทดสอบ (Test Information) ด้วยวิธีการตรวจให้คะแนนความรู้บางส่วน ด้วยวิธีการของการทดสอบความน่าจะเป็นกับการให้คะแนนตามความยากประจำชั้น

ความสำคัญของการวิจัย

เพื่อที่จะอธิบายถึงความสัมพันธ์ของคุณลักษณะภายในหรือความสามารถที่มีอยู่ในตัวบุคคลของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ด้วยการประมาณค่าพารามิเตอร์ของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ และเป็นแบบทดสอบเลือกตอบที่มีความเป็นปรนัยและคะแนนมีความเชื่อมั่นสูงต่อการตรวจให้คะแนน ผู้วิจัยต้องการศึกษาค่าความเที่ยงตรงตามสภาพและฟังก์ชันสารสนเทศของแบบทดสอบและข้อสอบจากการตรวจให้คะแนนให้ความรู้บางส่วนด้วยวิธีการให้คะแนนวิธีการของการทดสอบความน่าจะเป็น (Probability Testing) กับวิธีการให้คะแนนตามความยากประจำชั้น (Step Difficulties) โดยให้ความสำคัญทั้งในเชิงทฤษฎีและการปฏิบัติ

1. ในทางทฤษฎีเพื่อจะทำให้ทราบถึงความสามารถที่แท้จริงของผู้สอบและความสามารถในการประมาณค่าพารามิเตอร์เกี่ยวกับแบบทดสอบ รวมทั้งเป็นการเพิ่มสารสนเทศเกี่ยวกับการตอบข้อสอบเลือกตอบคณิตศาสตร์ให้มากขึ้น และนำมาซึ่งการพัฒนาแนวคิดเกี่ยวกับการหาค่าคะแนนความสามารถที่แท้จริงจากการใช้วิธีการตอบและการตรวจให้คะแนนต่าง ๆ ตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory)

2. ในทางปฏิบัติ เพื่อช่วยพัฒนาการวัดผลประเมินผลคณิตศาสตร์ที่ส่วนใหญ่เครื่องมือวัดเป็นแบบทดสอบเลือกตอบ ที่มีวิธีการให้คะแนนแบบ 0,1 ซึ่งไม่สามารถจำแนกนักเรียนได้ตามระดับของความรู้ระหว่างผู้รู้จริง ผู้ที่มีความรู้บางส่วนและผู้ที่ไม่รู้ แก้ปัญหาการเดาสุ่มและลดระยะเวลาในการตรวจให้คะแนน ทำให้มีแบบทดสอบและข้อสอบที่มีค่าความเที่ยงตรงตามสภาพและฟังก์ชันสารสนเทศสูง อันเป็นผลดีต่อวิธีการวัดผล ประเมินผลของครูผู้สอนที่จะได้ทราบถึงสารสนเทศเกี่ยวกับผู้เรียน ทราบถึงความสามารถที่แท้จริงของผู้เรียน ตลอดจนสามารถวินิจฉัยตัดสินใจในการตัดสินผลการเรียนรู้อันหนึ่งของนักเรียนตามระดับความรู้ ความสามารถของนักเรียนได้อย่าง

เหมาะสมและนำสารสนเทศไปพัฒนาผู้เรียนทั้งด้านกระบวนการจัดการเรียนการเรียนรู้ให้ผู้เรียนให้เกิดความเหมาะสม

ขอบเขตของการวิจัย

ประชากร

ประชากรที่ใช้ในการวิจัยครั้งนี้ เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 2 ปีการศึกษา 2549 สำนักงานเขตพื้นที่การศึกษาปราจีนบุรี จำนวน 97 โรงเรียน จำนวน 199 ห้องเรียน นักเรียน 6,803 คน (สำนักงานเขตพื้นที่การศึกษาปราจีนบุรี. 2549)

กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3ที่กำลังศึกษาในภาคเรียนที่ 2 ปีการศึกษา 2549 สำนักงานเขตพื้นที่การศึกษาปราจีนบุรี จำนวน 97 โรงเรียน จำนวนห้องเรียน 199 ห้องเรียน จำนวนนักเรียนทั้งหมด 6,803 คน ซึ่งได้มาโดยการสุ่มแบบหลายขั้นตอน (Multi - stage random Sampling) โดยการแบ่งเขตพื้นที่การศึกษาปราจีนบุรีเป็นอำเภอ จำนวน 7 อำเภอ จำนวน 3 อำเภอ 3,287 คน สุ่มได้นักเรียนจำนวน 593 คน แต่ในการวิเคราะห์ข้อมูลตามทฤษฎีตอบสนองข้อสอบ (Item Response Model) ซึ่งการประมาณค่าพารามิเตอร์กลุ่มตัวอย่างต้องใหญ่เพียงพอที่จะทำให้ความคลาดเคลื่อนมาตรฐานของการประมาณค่าอยู่ในระดับที่ยอมรับได้ตามเป้าหมายของการนำไปใช้ (ศิริชัย กาญจนวาสี. 2545: 100) แล้วสุ่มแบบแบ่งชั้นตามขนาดของโรงเรียน คือโรงเรียนขนาดใหญ่ จำนวน 3 โรงเรียน ขนาดกลาง จำนวน 13 โรงเรียน ขนาดเล็ก จำนวน 28 โรงเรียน สุ่มได้โรงเรียนขนาดใหญ่ จำนวน 1 โรงเรียน ขนาดกลาง จำนวน 4 โรงเรียน ขนาดเล็ก จำนวน 6 โรงเรียน ได้กลุ่มตัวอย่างจำนวน 1,007 คน

ตัวแปร

ตัวแปรอิสระ ได้แก่วิธีตรวจให้คะแนนความรู้บางส่วน 2 วิธี ดังนี้

1. วิธีตรวจให้คะแนนวิธีการของการทดสอบความน่าจะเป็น (Probability Testing)
2. วิธีตรวจให้คะแนนตามความยากประจำชั้น (Step Difficulties)

ตัวแปรตาม

1. ความเที่ยงตรงตามสภาพ (Concurrent Validity) ของแบบทดสอบ
2. ฟังก์ชันสารสนเทศของข้อสอบ (Item Information) ฟังก์ชันสารสนเทศของแบบทดสอบ (Test Information)

เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูล

แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ทักษะกระบวนการคิดคำนวณ การบวก การลบ การคูณและการหารจำนวนจริง เลขยกกำลัง ชั้นมัธยมศึกษาปีที่ 3 จำนวน 40 ข้อ

กรอบแนวคิดในการวิจัย

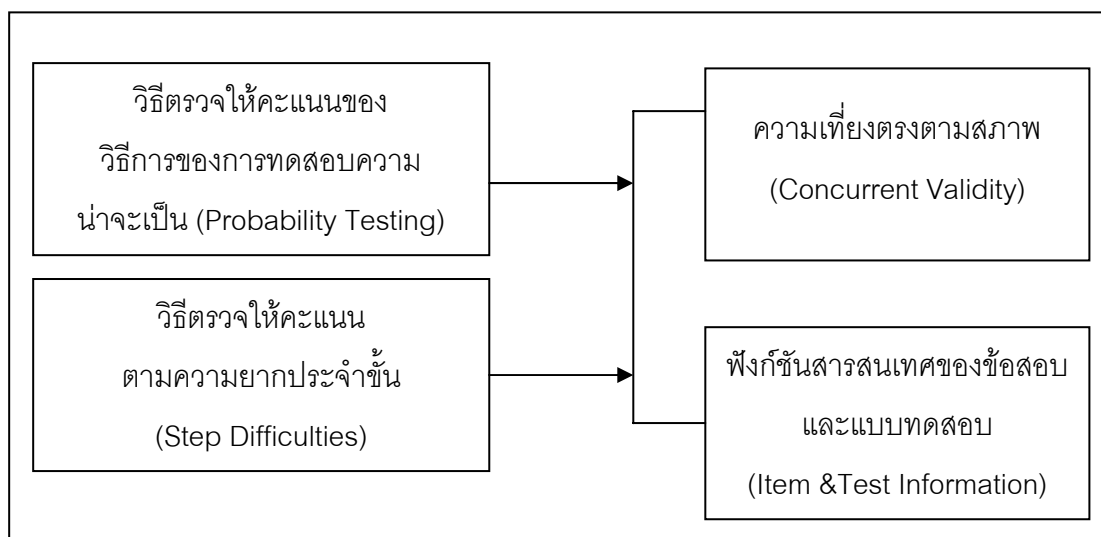
ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory : IRT) นี้มีพื้นฐานความเชื่อว่า พฤติกรรมการตอบสนองข้อสอบต่อข้อสอบซึ่งเป็นสิ่งที่สังเกตได้โดยตรงว่าถูกหรือผิด จะถูกกำหนดโดยคุณลักษณะภายในหรือความสามารถที่มีอยู่ในตัวบุคคล ซึ่งเป็นสิ่งที่ไม่สามารถสังเกตได้โดยตรง ฟังก์ชันสำหรับการอธิบายความสัมพันธ์ระหว่างความน่าจะเป็นของการตอบข้อสอบได้ถูกต้องหรือการเลือกการตอบกับความสามารถของผู้ตอบหรือคุณลักษณะภายในของผู้ตอบและคุณลักษณะของข้อสอบหรือรายการคำตอบโดยใช้โมเดล 2 ประเภท ตามลักษณะการตรวจให้คะแนนคำตอบ คือ โมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนน 2 ค่า (Dichotomous IRT Model) และโมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่า 2 ค่า (Polytomous IRT Models) ปัจจุบันมีรูปแบบของเครื่องมือที่นิยมใช้กัน เช่น มาตรฐานค่า (Rating Scale) การตรวจให้คะแนนความรู้บางส่วน (Partial Credit) การตรวจให้คะแนนตามลำดับขั้นตอนของรายการหลายคำตอบ (Ordered response Categories) ซึ่งการตรวจให้คะแนนความรู้บางส่วน (Partial Credit Model : PCM) พัฒนาโดยมาสเตอร์ (Masters. 1982) และ ล่าสุด คือ มูรากิ (Muraki. 1992) ได้พัฒนาโมเดลรวมค่าพารามิเตอร์ความชันของโค้งลักษณะข้อคำถามแต่ละข้อ เรียกว่า Generalized Credit Model : GPCM มีลักษณะเป็นโมเดลสำหรับการวิเคราะห์ข้อสอบหรือข้อคำถามที่มีกระบวนการตอบหลายลำดับขั้น ซึ่งจำเป็นต้องมีการตรวจให้คะแนนการตอบถูกต้องหรือตอบถูกบางส่วนในแต่ละลำดับขั้นของกระบวนการตอบ โดยยอมให้ข้อคำถามแต่ละข้อสามารถมีค่าพารามิเตอร์ความชันแตกต่างกันได้ ใช้หลักการคำนวณความน่าจะเป็นของการตอบแต่ละลำดับขั้นการตอบโดยตรงแบบขั้นตอนเดียว (Direct IRT Model) ลักษณะคำถามในแต่ละข้อ (i) อธิบายด้วยค่าพารามิเตอร์ระดับความยากของขั้นการตอบ (δ_{ij}) และค่าพารามิเตอร์ความชันของข้อคำถาม (α_j) วิธีการให้คะแนนความรู้บางส่วนที่ผู้วิจัยให้ความสนใจ คือ

1. การให้น้ำหนักแก่ตัวเลือกที่แตกต่างกัน (Differential Option Weighting) ด้วยวิธีการให้คะแนนตรวจให้คะแนนตามความยากประจำขั้นของคะแนนผลการตอบโดยกำหนดความยากประจำขั้นของการตอบ (Step Difficulties)

2. การเปลี่ยนแปลงวิธีการตอบ (Changing the response method) ด้วยวิธีการของการทดสอบความน่าจะเป็น (Probability Testing) เป็นวิธีที่มีความยืดหยุ่นเปิดโอกาสให้ผู้สอบแสดง

ความรู้บางส่วนได้มากที่สุด โดยรายงานความน่าจะเป็นในตัวเลือกที่ผู้สอบคิดว่าเป็นคำตอบที่ถูกโดยมีตำแหน่งคะแนนที่เป็นไปได้ 4 ตำแหน่ง คือ 1 – 4 คะแนน เนื่องจากโมเดลดังกล่าวไม่เชื่อมโยงกับข้อตกลงเบื้องต้น การประมาณค่าพารามิเตอร์มีค่าอำนาจจำแนกรายข้อในฟังก์ชันด้วยจึงเหมาะสมในการวัดความสามารถทางสติปัญญา

แผนภาพกรอบแนวคิดการวิจัย



ภาพประกอบ 1 กรอบแนวคิดการวิจัย

สมมติฐานของการวิจัย

1. ค่าความเที่ยงตรงตามสภาพ (Concurrent Validity) ระหว่างการตรวจให้คะแนนความรู้บางส่วนด้วยวิธีการตรวจให้คะแนนตามความยากประจำขั้น (Step Difficulties) กับวิธีการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็น (Probability Testing) แตกต่างกัน
2. ค่าฟังก์ชันสารสนเทศแบบทดสอบ (Test Information) และฟังก์ชันสารสนเทศข้อสอบ (Item Information) ระหว่างการตรวจให้คะแนนความรู้บางส่วนด้วยวิธีการตรวจให้คะแนนตามความยากประจำขั้น (Step Difficulties) กับวิธีการตรวจให้คะแนนวิธีการของการทดสอบความน่าจะเป็น (Probability Testing) แตกต่างกัน

นิยามศัพท์เฉพาะ

คะแนนความรู้บางส่วน หมายถึง คะแนนที่แสดงถึงตำแหน่งจริงของความรู้ของผู้ตอบจากการทำข้อสอบวิชาคณิตศาสตร์ – ทักษะการคิดคำนวณ ด้วยวิธีการตรวจให้คะแนนวิธีการของการทดสอบความน่าจะเป็น (Probability Testing) และตรวจให้คะแนนตามความยากประจำขั้นด้วยการวิเคราะห์จากโปรแกรม MULTILOG 7.03 For Window (David Thissen, Wen-Hung Chen and Darrell Bock. 2003)

ความเที่ยงตรงตามสภาพ (Concurrent Validity) หมายถึง คุณสมบัติของแบบทดสอบที่สามารถทำหน้าที่วัดสภาพความสามารถที่เป็นจริงของผู้สอบ จากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 โดยคำนวณจากค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนดิบของการสอบกลางภาคของกลุ่มตัวอย่าง ในภาคเรียนที่สอง ปีการศึกษา 2549 กับคะแนนของแบบทดสอบวัดผลสัมฤทธิ์คณิตศาสตร์ที่มีการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็นและวิธีการตรวจให้คะแนนตามความยากประจำชั้น

ฟังก์ชันสารสนเทศของแบบทดสอบ (Test Information) หมายถึง ค่าความแม่นยำในการประมาณค่าความสามารถจริงของผู้สอบจากการตอบแบบทดสอบนั้น คำนวณได้จากผลรวมของค่าฟังก์ชันสารสนเทศของข้อสอบทุกข้อในแบบทดสอบทั้งฉบับถ้าสูงที่ระดับใดแสดงว่ามีความแม่นยำสูง

ฟังก์ชันสารสนเทศของข้อสอบ (Item Information) หมายถึง ค่าความแม่นยำในการประมาณค่าความสามารถของผู้สอบจากการตอบข้อสอบในข้อนั้น ๆ คำนวณได้จากอัตราส่วนระหว่างค่าอนุพันธ์ของการตอบได้คะแนนในระดับที่มีความรู้เต็มที่กับผลคูณของการตอบในระดับของความรู้อยู่บางส่วน

ค่าพารามิเตอร์ของข้อสอบ หมายถึง โค้งลักษณะข้อสอบ 1 พารามิเตอร์ ได้แก่ ค่าความยาก ที่เปลี่ยนไปตามระดับความสามารถของผู้สอบ

ประสิทธิภาพสัมพัทธ์ (Relative Efficiency) อัตราส่วนระหว่างค่าฟังก์ชันสารสนเทศของแบบทดสอบที่มีการตรวจให้คะแนนต่างกัน ณ ตำแหน่งความสามารถ (θ) เดียวกัน

วิธีการให้คะแนนวิธีการของการทดสอบความน่าจะเป็น (Probability Testing) หมายถึง การให้คะแนนตัวเลือกทุกตัว ที่พิจารณานำหนักคะแนนจากลำดับขั้นของกระบวนการหาคำตอบโดยที่ข้อสอบรายข้อมีทักษะการคิดหลายขั้นตอน ในขั้นที่สูงกว่าหรือขั้นที่ได้คะแนนมากกว่า ต้องอาศัยการตอบในขั้นที่ต่ำกว่าหรือได้คะแนนน้อยกว่า แบ่งช่วงคะแนนความรู้ออกเป็น 5 ระดับ ดังนี้

- 4 คะแนน หมายถึง ตอบข้อที่ใช้ลำดับขั้นของกระบวนการหาคำตอบในขั้นที่ 1 – 4
- 3 คะแนน หมายถึง ตอบข้อที่ใช้ลำดับขั้นของกระบวนการหาคำตอบในขั้นที่ 1 – 3
- 2 คะแนน หมายถึง ตอบข้อที่ใช้ลำดับขั้นของกระบวนการหาคำตอบในขั้นที่ 1 – 2
- 1 คะแนน หมายถึง ตอบข้อที่ใช้ลำดับขั้นของกระบวนการหาคำตอบในขั้นที่ 1
- 0 คะแนน หมายถึง ไม่ตอบ

วิธีการให้คะแนนตามความยากประจำขั้น (Step Difficulties) หมายถึง การตรวจให้คะแนนตัวเลือกทุกตัว ที่มีการกำหนดน้ำหนักคะแนนตามความยากประจำขั้นของแต่ละตัวเลือกไว้แล้ว พิจารณาความยากประจำขั้นตามทฤษฎีของโครงสร้างความรู้ ทางคณิตศาสตร์เรียนรู้จากง่ายไปยาก คือเรียนรู้จากการบวก การลบ การคูณ การหาร จำนวนเต็ม เศษส่วน เลขยกกำลัง ตามลำดับ แบ่งช่วงคะแนนความยากประจำขั้นเป็น 5 ระดับ ดังนี้

- 4 คะแนน หมายถึง ตอบข้อที่ใช้ทักษะการคิดคำนวณในขั้นยากมากที่สุด
- 3 คะแนน หมายถึง ตอบข้อที่ใช้ทักษะการคิดคำนวณในขั้นยากมาก
- 2 คะแนน หมายถึง ตอบข้อที่ใช้ทักษะการคิดคำนวณในขั้นยากปานกลาง
- 1 คะแนน หมายถึง ตอบข้อที่ใช้ทักษะการคิดคำนวณในขั้นยากน้อย
- 0 คะแนน หมายถึง ไม่ตอบ

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

ผู้วิจัยนำเสนอผลการศึกษาเอกสารที่เกี่ยวข้องกับแนวคิด ดังนี้

ตอนที่ 1 ทฤษฎีการตอบสนองข้อสอบ

ตอนที่ 2 วิธีการตรวจให้คะแนน

ตอนที่ 3 การศึกษาเกี่ยวกับแบบทดสอบเลือกตอบ

ตอนที่ 4 การตรวจสอบคุณภาพของแบบทดสอบ

ตอนที่ 5 งานวิจัยที่เกี่ยวข้อง

ตอนที่ 1 ทฤษฎีการตอบสนองข้อสอบ

1.1 แนวคิดพื้นฐานทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory : IRT) เป็นทฤษฎีการวัดที่อธิบายความสัมพันธ์ระหว่างคุณลักษณะภายในหรือความสามารถที่มีอยู่ในตัวบุคคลกับพฤติกรรม การตอบสนองข้อสอบของบุคคลนั้นว่ามีโอกาสตอบข้อสอบถูกมาน้อยเพียงไร เป็นทฤษฎีที่ได้แก้ไขข้อจำกัดหลายประการของทฤษฎีการวัดแบบดั้งเดิม (Classical Test Theory) มีพัฒนาการเริ่มต้นที่ (Hambleton and Swaminatan. 1985) บิเนตและไซมอน (Binet and Simon) เป็นนักจิตวิทยาได้สร้างกราฟที่แสดงความสัมพันธ์ระหว่างทักษะการคิดของเด็กกับอายุซึ่งเป็นที่มาของโค้งคุณลักษณะ ในระยะเวลาต่อมา ต่อมาในปี ค.ศ.1936 ริชาร์ดสัน (Richardson) ได้แสดงความสัมพันธ์ระหว่าง ทฤษฎีตอบสนองข้อสอบกับทฤษฎีการทดสอบแบบดั้งเดิมและหาวิธีในการประมาณค่าพารามิเตอร์ ตามทฤษฎีตอบสนองข้อสอบ ค.ศ.1943 ลอว์ลี (Lawley) ได้เสนอบทความเกี่ยวกับปัญหาการสร้าง และการเลือกข้อสอบโดยเสนอแนวคิดในรูปของโมเดล Normal ogive Model ในปี ค.ศ.1950 ราสช์ (Rasch) นักคณิตศาสตร์ชาวเดนมาร์กได้เสนอโมเดลราสช์ (Rasch Model) แบบ 1 พารามิเตอร์โดยมีแนวคิดว่าคุณสมบัติของข้อสอบเป็นสิ่งเดียวที่มีอิทธิพลต่อการตอบสนองข้อสอบ ต่อมาปี ค.ศ. 1952 ลอร์ด (Lord) ได้ขยายแนวคิดทฤษฎีตอบสนองข้อสอบ โดยเสนอ Normal ogive function แบบ 2 พารามิเตอร์ที่มีการเพิ่มพารามิเตอร์ อำนาจจำแนก ในปี ค.ศ.1968 เบิร์นบอม (Birbaum) ได้พัฒนาโมเดลโลจิสติก (Logistic Model) ที่เป็นฟังก์ชันที่สามารถคำนวณได้ง่ายขึ้น ในปี ค.ศ. 1974 ลอร์ด (Lord) ได้เสนอแนวคิดโมเดลโลจิสติกแบบ 3 พารามิเตอร์ โดยเพิ่มพารามิเตอร์ในโอกาสการเดาข้อสอบในการประมาณค่าพารามิเตอร์ของผู้สอบด้วย

โมเดลการตอบสนองข้อสอบในระยะเวลาแรก ๆ ใ้กับแบบทดสอบที่ให้คะแนนรายข้อ เป็น 2 ระดับหรือการให้คะแนนแบบ 0 – 1 เรียกว่า โมเดลการตอบสนองข้อสอบแบบทวิภาค (Dichotomous Item Response Model) ต่อมาได้มีผู้พัฒนาเพื่อใ้กับแบบทดสอบและแบบวัด เจตคติที่ตรวจให้คะแนนเป็นแบบมาตราประมาณค่าหรือแบบพหุภาค (Polytomous) โมเดลในแนวนนี้ เรียกว่า (Polytomous Item Response Model) (Murki. 1992, 1993) โดยมีความเชื่อที่สำคัญ ดังนี้

ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory : IRT) นี้มีพื้นฐานความเชื่อว่า พฤติกรรมการตอบสนองต่อข้อสอบของผู้สอบซึ่งเป็นสิ่งที่สังเกตได้โดยตรงว่าถูกหรือผิดจะถูกกำหนด โดยคุณลักษณะภายในหรือความสามารถที่มีอยู่ในตัวบุคคล ซึ่งเป็นสิ่งที่ไม่สามารถสังเกตได้โดยตรง ทฤษฎีนี้อธิบายความสัมพันธ์ของคุณลักษณะภายใน หรือความสามารถที่มีในตัวบุคคล (Latent trait or ability) กับพฤติกรรมการตอบสนองข้อสอบ โดยใช้โค้งลักษณะข้อสอบ (Item Characteristic Curve ; ICC) ซึ่งมีการกำหนดลักษณะของข้อสอบด้วยพารามิเตอร์ความยาก (b) อำนาจจำแนก (a) และโอกาสการเดา (c) ทฤษฎีการตอบสนองข้อสอบ (IRT) จึงอยู่บนพื้นฐานความคิดที่สำคัญ 2 ประการ คือ 1) ผลการตอบข้อสอบหรือข้อคำถามของผู้ตอบ สามารถอธิบายได้ด้วยความสามารถที่มีอยู่ ภายในของผู้ตอบ 2) ความสัมพันธ์ระหว่างผลการตอบข้อสอบกับความสามารถที่มีอยู่ภายใน สามารถอธิบายได้ด้วยฟังก์ชันลักษณะข้อสอบหรือโค้งลักษณะข้อสอบ (ICC) มีลักษณะเป็นฟังก์ชัน ทางคณิตศาสตร์ เรียกว่า ฟังก์ชันโลจิสติก (Logistic function) หรือใกล้เคียงฟังก์ชันปกติสะสม (Normal ogive function) ซึ่งเรียกว่า โมเดลการตอบสนองข้อสอบซึ่งมีหลายรูปแบบเช่น โมเดล 1 , 2 , 3 พารามิเตอร์

สำหรับการเลือกโมเดลของทฤษฎีการตอบสนองข้อสอบมาใช้นั้น จะต้องเลือกให้ เหมาะสมกับสถานการณ์ต่าง ๆ ตามข้อตกลงเบื้องต้นของแต่ละโมเดล โดยทฤษฎีนี้มีจุดเด่นหลาย ประการ ดังนี้ (กนกวรรณ รัตนธ. 2544 อ้างอิงจาก Hambleton and Swaminathan. 1985)

1. ค่าพารามิเตอร์ของข้อสอบเป็นอิสระจากกลุ่มผู้สอบที่ใช้ในการประมาณค่า คือ ค่าพารามิเตอร์ของข้อสอบแต่ละข้อ ได้แก่ ค่าความยาก (b) ค่าอำนาจจำแนก (a) และค่าการเดา (c) เป็นค่าคงที่ไม่แปรเปลี่ยนไปตามกลุ่มของผู้สอบ ถึงแม้ว่าผู้สอบต่างกลุ่มกัน ค่าพารามิเตอร์ของข้อสอบจะมีค่าคงเดิมเสมอและในทำนองเดียวกันค่าพารามิเตอร์ความสามารถของผู้สอบเป็นอิสระจาก ข้อสอบที่ใช้ในการประมาณค่าจะไม่แปรเปลี่ยนไปตามค่าพารามิเตอร์ของข้อสอบ ไม่ว่าผู้สอบจะทำ ข้อสอบข้อใดหรือได้ทำข้อสอบเมื่อใดก็ตาม

2. การเปรียบเทียบความสามารถของผู้สอบ จะไม่ขึ้นอยู่กับคำถามในแบบสอบ ถึงแม้ว่า จะใช้ข้อคำถามต่างกัน ก็สามารถนำค่าความสามารถมาเปรียบเทียบกันได้ ทั้งนี้เนื่องจากค่า ความสามารถที่ประมาณค่าได้นั้น เป็นคะแนนโลจิท (Logit) ซึ่งอยู่ในมาตรวัดเดียวกัน

3. การรายงานคุณภาพของข้อสอบในรูปค่าอินฟอร์เมชัน (information) สามารถรายงานได้ทั้งเป็นรายข้อและทั้งฉบับ ค่าอินฟอร์เมชัน เป็นตัวบ่งชี้ถึงความถูกต้องแม่นยำ (accuracy) ในการประมาณค่าความสามารถ สามารถนำมาใช้แทนค่าความเชื่อมั่นและความคลาดเคลื่อนมาตรฐานในการวัดได้

1.2 หลักการของทฤษฎีตอบสนองข้อสอบ

ทฤษฎีการตอบข้อสอบ (Item Response Model) เชื่อมเกี่ยวกับค่าพารามิเตอร์ของข้อสอบ (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 and Swaminathan, 1985: 9) นักวัดผลพยายามหาความสัมพันธ์ระหว่างผลที่ได้จากการตอบแบบทดสอบ หรือคะแนน กับปริมาณความสามารถของผู้ตอบแต่ละคน เพื่อเขียนมาเป็นโมเดลทางคณิตศาสตร์ ความสัมพันธ์ระหว่างปริมาณของการทำข้อสอบหรือคะแนนของผู้สอบ (Test Performance) กับปริมาณความสามารถของผู้สอบ (Ability) สามารถเขียนในรูปของความสัมพันธ์ทั่วไป (Lord, 1982; อ้างอิงจาก ตรึงใจ พูนผลอำนวย, 2534)

$$P = f(U_i | \theta_1, \theta_2, \theta_3, \dots, \theta_k; \beta_j)$$

เมื่อ P แทน ผลการสอบ

f แทน ฟังก์ชัน

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

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

β_j แทน ค่าพารามิเตอร์ของข้อสอบที่ j

1.3 ข้อตกลงเบื้องต้นของทฤษฎีตอบสนองข้อสอบ

1.3.1 ความเป็นมิติเดียว (Unidimension) หมายถึง ข้อสอบแต่ละข้อในแบบทดสอบจะต้องวัดความสามารถหรือคุณลักษณะเดียวกัน (Unidimensionality) หรือข้อสอบเหล่านี้มีความเป็นเอกพันธ์ (Homogeneous) การกำหนดเช่นนี้ก็เพื่อให้รูปแบบของทฤษฎีนี้มีความหมายซับซ้อนน้อยลง และง่ายต่อการแปลความหมายของคะแนนที่ได้จากแบบทดสอบ วิธีการตรวจสอบว่าแบบ

ทดสอบนั้นวัดมิติเดียวกันหรือไม่นั้นทำได้โดยการวิเคราะห์องค์ประกอบ (Factor Analysis) แล้วสังเกตค่าไอเก้น (Eigen value) ค่าสูงสุดว่าแตกต่างจากค่าอื่น ๆ อย่างชัดเจนหรือไม่ ดังที่แฮมเบิลตันและสวามินาธาน (Hambleton and Swaminathan, 1985: 16 – 17) กล่าวไว้ว่า ข้อตกลงนี้ไม่เข้มงวดนัก ขอให้มึลักษณะเด่นที่จะวัดองค์ประกอบใดองค์ประกอบหนึ่งก็ใช้ได้

1.3.2 ความเป็นอิสระในการตอบข้อสอบ (Local Independence) หมายถึง โอกาสในการตอบข้อสอบแต่ละข้อได้ถูกต้องเป็นอิสระจากกัน นั่นคือ การตอบข้อสอบข้อใดข้อหนึ่งได้ถูกหรือผิดจะไม่มีผลต่อการตอบข้ออื่น ๆ

1.3.3 โค้งลักษณะข้อสอบ (Item Characteristic Curve) เป็นฟังก์ชันทางคณิตศาสตร์ที่แสดงถึงความสัมพันธ์ระหว่างโอกาสในการตอบข้อสอบข้อนั้นถูกต้องกับระดับความสามารถที่วัดได้โดยชุดของข้อสอบหรือแบบทดสอบนั้น (Hambleton and Swaminathan, 1985: 9) จะเห็นว่าโอกาสผู้สอบตอบข้อสอบถูกจะขึ้นอยู่กับโค้งลักษณะของข้อสอบในแต่ละข้อมีคุณสมบัติไม่แปรเปลี่ยนไปตามกลุ่มตัวอย่างหรือตัวผู้สอบ ดังนั้นโอกาสการตอบข้อสอบถูกจึงไม่แปรเปลี่ยนด้วย

ผู้วิจัยอธิบายคุณลักษณะเบื้องต้นของแบบทดสอบตามทฤษฎีการตอบข้อสอบ (Item Response Model) ด้วยการแสดงความเป็นมิติเดียว (Unidimension) โดยการวิเคราะห์องค์ประกอบ (Factor Analysis) แล้วสังเกตค่าไอเก้น (Eigen value)

1.4 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 ไปยังไม่เฉพาะเจาะจงจึงมีผู้สนใจศึกษาและพัฒนาเพื่อให้ได้สมการที่สามารถอธิบายลักษณะเฉพาะของข้อสอบแต่ละข้อที่สามารถอธิบายความสัมพันธ์ของสองตัวแปรนี้ได้เด่นชัดขึ้น

คุณลักษณะของโมเดลการตอบสนองข้อสอบเป็นระบบความสัมพันธ์ระหว่างโอกาสตอบข้อสอบถูก (P_i) กับความสามารถที่มีอยู่ภายในผู้ตอบ (θ) ในรูปของโค้งลักษณะข้อสอบ (ICC) ซึ่งมี

ลักษณะเป็นฟังก์ชันโลจิสติก (Logistic function) หรือฟังก์ชันปกติสะสม (Normal ogive function) บางครั้งเรียกว่าโมเดลโลจิสติกหรือโมเดลปกติสะสมใช้ฟังก์ชันปกติสะสมแสดงความสัมพันธ์ระหว่างผลการตอบข้อสอบกับความสามารถของผู้สอบ ส่วนโมเดลโลจิสติกใช้ฟังก์ชัน โลจิสติกแสดงความสัมพันธ์ระหว่างผลการตอบกับความสามารถ ซึ่งฟังก์ชันทั้งสองให้ผลลัพธ์ของการประมาณค่าใกล้เคียงกันมากแต่ฟังก์ชันโลจิสติกมีลักษณะของสูตรทางคณิตศาสตร์ และวิธีคำนวณง่าย สะดวกกว่า นอกจากนี้โมเดลโลจิสติกยังมีความทนทานต่อความคลาดเคลื่อนที่เกิดขึ้นกับผู้สอบที่มีความสามารถสูงจะตอบข้อสอบได้ดีกว่า ทำให้โมเดลโลจิสติกเป็นที่นิยมนำไปใช้จริง

1.4.1 โมเดลการตอบสนองข้อสอบแบบทวิภาค (Dichotomous Item

Response Model) โค้งลักษณะข้อสอบแสดงความสัมพันธ์ระหว่างระดับความสามารถกับโอกาสการตอบข้อสอบได้ถูกต้อง ทำให้เกิดโมเดลของโค้งลักษณะของข้อสอบขึ้นหลายโมเดล แต่รูปแบบที่นิยมใช้กันแพร่หลาย คือ รูปแบบโลจิสติก (Logistic Model)

1.4.1.1 โมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ หรือราส์ชโมเดล (One-Parameter Model, Rasch model) มีข้อตกลงว่าข้อสอบทุกข้อมีค่าอำนาจจำแนกเท่ากัน ดังสมการ

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

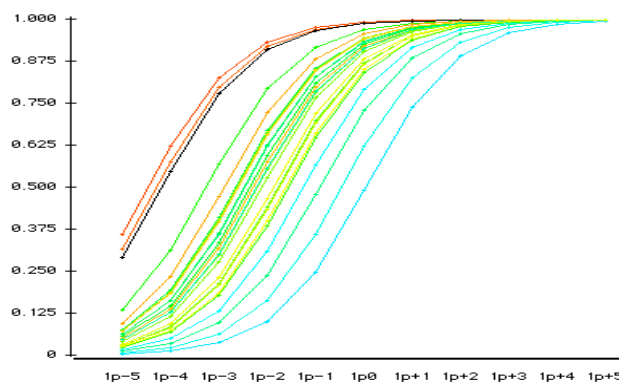
เมื่อ $P_i(\theta)$ = ความน่าจะเป็นที่ผู้ตอบซึ่งมีความสามารถ θ จะตอบข้อสอบที่ i ได้ถูกต้อง

b_i = ค่าพารามิเตอร์ความยากของข้อสอบที่ i ซึ่งเป็นค่าที่แสดงตำแหน่งของ

ICC ณ จุด θ ที่มีโอกาสตอบข้อสอบถูก 0.50

D = ค่าคงที่ของ scale factor เพื่อปรับค่าฟังก์ชันโลจิสติกให้ใกล้เคียงกับค่าที่ได้จากฟังก์ชันรูปตัวเอสปกติ มีค่าประมาณ 1.7

$e = 2.718$



ภาพประกอบ 2 โค้งลักษณะข้อสอบแบบ 1 พารามิเตอร์

ที่มา: John Behrens. <http://www.creative-wisdom.com/computer/sas/IRT.html>

โค้งลักษณะข้อสอบ (ICC) สำหรับโมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ มีลักษณะที่สำคัญดังนี้

b_i = มีค่าแปรเปลี่ยนตามลักษณะของข้อสอบแต่ละข้อ

a_i = มีค่าคงที่

$c_i = 0$

1.4.1.2 โมเดลการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ (Two - Parameter Model) มีโค้งลักษณะข้อสอบที่เขียนด้วยฟังก์ชันโลจิสติก ดังสมการ (Hambleton, Swaminathan and Roger. 1991: 14)

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

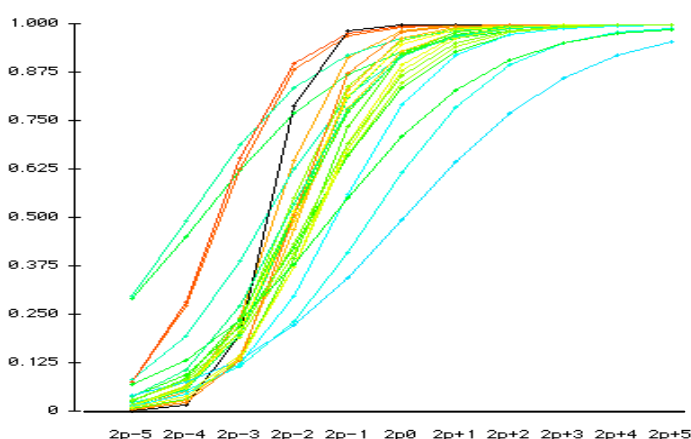
เมื่อ $P_i(\theta)$ = ความน่าจะเป็นที่ผู้ตอบซึ่งมีความสามารถ θ จะตอบข้อสอบที่ i ได้ถูกต้อง

b_i = ค่าพารามิเตอร์ความยากของข้อสอบที่ i ซึ่งเป็นค่าที่แสดงตำแหน่งของ ICC ณ จุด θ ที่มีโอกาสตอบข้อสอบถูก 0.50

a_i = ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบที่ i ซึ่งเป็นค่าความชันของ ICC ณ ตำแหน่ง b_i

$e = 2.718$

$D = 1.70$



ภาพประกอบ 3 โค้งลักษณะข้อสอบแบบ 2 พารามิเตอร์

ที่มา: John Behrens. <http://www.creative-wisdom.com/computer/sas/IRT.html>

โค้งลักษณะข้อสอบ (ICC) สำหรับโมเดลการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ มีลักษณะที่สำคัญดังนี้

b_i = มีค่าแปรเปลี่ยนตามลักษณะของข้อสอบแต่ละข้อ

a_i = มีค่าแปรเปลี่ยนตามลักษณะของข้อสอบแต่ละข้อ

$c_i = 0$

1.4.1.3 โมเดลการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ (Three – Parameter Model) มีโค้งลักษณะข้อสอบที่เขียนด้วยฟังก์ชันโลจิสติก ดังสมการ (Hambleton, Swaminathan and Roger. 1991: 18)

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

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

b_i = ค่าพารามิเตอร์ความยากของข้อสอบที่ i ซึ่งเป็นค่าที่แสดงตำแหน่งของ

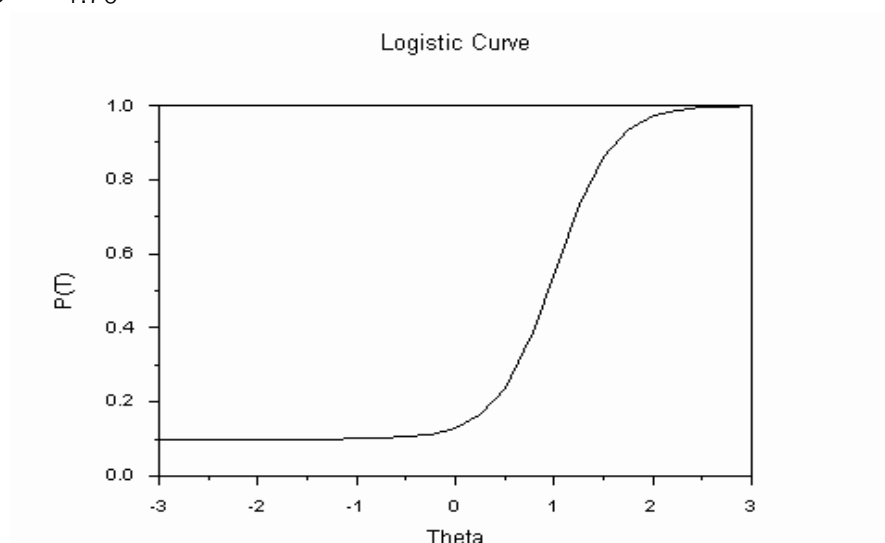
ICC ณ จุด θ ที่มีโอกาสตอบข้อสอบถูก $\frac{1 + C_i}{2}$

a_i = ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบที่ i ซึ่งเป็นค่าความชันของ ICC ณ ตำแหน่ง b_i

c_i = ค่าพารามิเตอร์โอกาสเดาข้อสอบได้ถูก

$e = 2.718$

$D = 1.70$



ภาพประกอบ 4 โค้งลักษณะข้อสอบแบบ 3 พารามิเตอร์

ที่มา: Michael T. Brannick. <http://luna.cas.usf.edu/~mbrannic/files/pmet/irt.htm>

โค้งลักษณะข้อสอบ (ICC) สำหรับโมเดลการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ มีลักษณะที่สำคัญดังนี้

- b_i = มีค่าแปรเปลี่ยนตามลักษณะของข้อสอบแต่ละข้อ
- a_i = มีค่าแปรเปลี่ยนตามลักษณะของข้อสอบแต่ละข้อ
- c_i = มีค่าแปรเปลี่ยนตามลักษณะของข้อสอบแต่ละข้อ

1.4.2 โมเดลการตอบสนองข้อสอบแบบพหุวิภาค (Polytomous Item

Response Model) ทฤษฎีการตอบสนองของข้อสอบ (IRT) เสนอฟังก์ชันสำหรับอธิบายความสัมพันธ์ระหว่างความน่าจะเป็นของการตอบข้อสอบได้ถูกต้อง (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 ถูกหรือผิด เห็นด้วยหรือไม่เห็นด้วย เป็นต้น ในขณะที่เครื่องมือทางการศึกษาและจิตวิทยาหลายประเภท แต่ละข้อคำถามมักมีหลายรายการคำตอบที่กำหนดลำดับหรือน้ำหนักคะแนนต่างกัน เพราะน่าจะให้สารสนเทศและความเชื่อมั่นจากการตอบที่สูงกว่าการตรวจให้คะแนนแบบ 2 ค่า เช่น แบบวัดเจตคติ แบบวัดบุคลิกภาพ แบบวัดความสนใจในอาชีพ เป็นต้น จึงได้มีการพัฒนาโมเดลการตอบสนองข้อสอบสำหรับใช้กับการตรวจให้คะแนนรายข้อมากกว่า 2 ค่า ซึ่งปัจจุบันมีรูปแบบของเครื่องมือที่นิยมใช้กัน เช่น มาตรฐานค่า (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)

Graded – Response Model

ซามิจิมา (Samejima. 1996) ได้พัฒนา Graded Response Model (GRM) สำหรับใช้กับแบบทดสอบหรือแบบวัดที่แต่ละข้อคำถามมีรายการคำตอบแบบมาตราเรียงลำดับ (Ordered – response categories) โดยแต่ละข้ออาจมีจำนวนรายการคำตอบที่แตกต่างกันได้ เช่น มาตราประมาณค่า (Rating Scale) ที่ไม่จำเป็นต้องมีรายการคำตอบเท่ากันทุกข้อ การตรวจให้คะแนนความรู้บางส่วนที่แต่ละข้อมีจำนวนลำดับขั้นของการให้คะแนนแตกต่างกัน เช่น มาตราวัดของ Likert หรือมาตราวัดแบบนัยจำแนกและมีการใช้สารสนเทศจากโค้งลักษณะการตอบแต่ละตัวเลือกด้วยการที่โมเดลได้ชื่อว่า difference models เพราะการคำนวณความน่าจะเป็นของผู้สอบที่ตอบตัวเลือกที่ K ได้ถูกต้องนั้นต้องคำนวณจากผลต่างระหว่างความน่าจะเป็นจากตัวเลือกที่ i และตัวเลือกที่ $k - 1$

Graded - Response Model (GRM) มีลักษณะเป็นโมเดลทั่วไปของโมเดลการตอบสนองข้อสอบที่มี 2 พารามิเตอร์ (2- Parameter Model) และใช้หลักการคำนวณความน่าจะเป็นของการตอบแต่ละรายการคำตอบแบบ 2 ขั้นตอน (Indirect IRT Model) โดยขั้นตอนแรกคำนวณหาค่าความชันรวมของแต่ละข้อคำถาม จากนั้นจึงคำนวณหาค่าพารามิเตอร์ของแต่ละรายการคำตอบในแต่ละข้อคำถาม คำถามแต่ละข้อ (i) อธิบายได้ด้วยค่าความชันรวมของข้อคำถาม (Common item slope parameter, α_i) และค่า threshold ของแต่ละรายการคำตอบ (Category threshold parameter, β_{ij}) เมื่อ $j = 1, \dots, m_i$ เป็นจำนวนของ threshold ข้อ i จำนวนรายการคำตอบของข้อ i (K_i) $= m_i + 1$

การวิเคราะห์ตามโมเดล GRM จึงมีเป้าหมายเพื่อการประมาณค่า α_i และตำแหน่งของ β_{ij} ของผู้ตอบที่มีคุณลักษณะ (θ) บนสเกลที่ต่อเนื่องกัน โดยใช้สูตร ดังนี้

$$P_{ix}^*(\theta) = \frac{\exp [\alpha_i (\theta - \beta_{ij})]}{1 + \exp [\alpha_i (\theta - \beta_{ij})]}$$

เมื่อ $x = j = 1, \dots, m_i$

$P_{ix}(\theta)$ = ความน่าจะเป็นที่ผู้ตอบซึ่งมีคุณลักษณะระดับ θ จะตอบข้อที่ i ด้วยการเลือกรายการคำตอบที่ x เมื่อ $x = 1, 2, \dots, m_i$

α_i = ค่าพารามิเตอร์ความชันรวม (slope parameter) ของข้อที่ i

β_{ij} = ค่าพารามิเตอร์ Threshold ของแต่ละรายการคำตอบ (threshold parameter) ของข้อที่ i

Modified Graded – Response Model

มูรากิ (Muraki. 1992) ได้ปรับปรุงโมเดล GRM จึงเรียกว่า Modified Graded - Response Model (M-GRM) เพื่อให้สะดวกแก่การใช้กับแบบวัดประเภทมาตราประมาณค่า (Rating Scale) ที่นิยมให้มีจำนวนรายการคำตอบที่เท่ากันหรือรูปแบบการตอบให้คงที่สำหรับทุกข้อคำถาม เช่น รูปแบบการตอบเป็นมาตราประมาณค่า 5 ระดับของลิเคิร์ต (Likert rating scale) เป็นต้น

(M-GRM) มีลักษณะเป็นโมเดลเฉพาะของโมเดล GRM โดยสามารถนำไปใช้กับข้อคำถามที่มีรายการคำตอบแบบมาตราเรียงลำดับ ที่มีจำนวนรายการคำตอบเท่ากันทุกข้อคำถามหรือมีรูปแบบการตอบคงที่เหมือนกันทุกข้อ สำหรับการคำนวณหาความน่าจะเป็นของการตอบแต่ละรายการคำตอบใช้วิธี 2 ขั้นตอน (Indirect Method) เหมือนโมเดล GRM โมเดล ใน M – GRM ลักษณะข้อคำถามแต่ละข้อ (i) อธิบายได้ด้วยค่าความชันร่วมของข้อคำถาม 1 ค่า (Common item slope parameter, α_i) และ threshold parameter (β_{ij}) ของแต่ละรายการคำตอบจำนวน m_i ค่า แต่ β_{ij} สามารถแยกได้เป็น 2 ส่วนคือ ค่าความยากของข้อ (Item location parameter, b_i) และ ชุดของ threshold parameter สำหรับแต่ละรายการคำตอบทั้งฉบับ (c_j) ดังนี้ $\beta_{ij} = b_i - c_j$ โค้งคุณลักษณะปฏิบัติการ (OCC) ของโมเดลของโมเดล M – GRM สามารถเขียนได้ดังสมการ

$$P_{ix}^*(\theta) = \frac{\exp[\alpha_i(\theta - (b_i - c_j))]}{1 + \exp[\alpha_i(\theta - (b_i - c_j))]}$$

หรือ

$$P_{ix}^*(\theta) = \frac{\exp[\alpha_i(\theta - (b_i + c_j))]}{1 + \exp[\alpha_i(\theta - (b_i + c_j))]}$$

เมื่อ $P_{ix}^*(\theta)$ = ความน่าจะเป็นที่ผู้ตอบซึ่งมีคุณลักษณะ θ จะตอบข้อ i ด้วยการเลือกรายการคำตอบที่ x หรือสูงกว่าเมื่อ $x = 1, 2, \dots, m_i$

α_i = ค่าพารามิเตอร์ความชันร่วมของข้อคำถามที่ i

b_i = ค่าพารามิเตอร์ตำแหน่งของโค้ง OCC หรือค่าความยากของข้อคำถามที่ i

c_j = ค่าพารามิเตอร์ threshold สำหรับแต่ละรายการคำตอบทั้งฉบับ

เมื่อพิจารณาสมการ M-GRM ซึ่งเป็นโมเดลที่มีความเฉพาะ เพราะมีสมมติฐานว่าขอบเขตของแต่ละรายการคำตอบมีระยะห่างจากกันและกันเท่ากัน ซึ่งแตกต่างจาก GRM ที่ระยะดังกล่าวปล่อยให้ค่าอิสระที่แตกต่างกันได้ระหว่างข้อคำถาม ผลที่ตามมาทำให้โมเดล M-GRM มีจำนวนค่าพารามิเตอร์ที่ต้องคำนวณน้อยกว่าโมเดล GRM และมีข้อดีตรงที่การคำนวณค่า b_i แยกออกจาก β_{ij} แต่โมเดล GRM จะสะดวกกว่าในกรณีที่แต่ละข้อคำถามมีจำนวนรายการคำตอบแตกต่างกัน ทำให้สามารถนำค่า b_i มาใช้เรียงลำดับความยากของข้อคำถาม รวมทั้งค่า c_j ใช้บอกระยะห่างของสเกล ทางจิตวิทยาคล้ายกับค่าน้ำหนักสเกลตามวิธีของ เทอร์สโตน (Thurstone's method)

ค่าพารามิเตอร์ความชันร่วมของคำถาม (α) ตาม M-GRM จะมีความผันแปรน้อยกว่าเมื่อเทียบกับ GRM ตามโปรแกรม PARSCALE (Muraki, 1993) จะมีการตรวจสอบความสอดคล้องของโมเดลข้อมูลด้วยการทดสอบ χ^2 -test ซึ่งเป็นค่าสถิติที่บ่งชี้ได้ว่าข้อสอบหรือข้อคำถามใดบ้างที่ทำการวิเคราะห์ได้เหมาะสมด้วย M-GRM

Partial Credit Model

มาสเตอร์ (Masters, 1982) ได้พัฒนา Partial Credit Model (PCM) โดยการขยายแนวคิดมาจาก ราสส์ โมเดล (Rasch Model) ซึ่งเป็นโมเดลที่ใช้ในการประมาณค่าพารามิเตอร์สำหรับข้อสอบที่ให้คะแนนแบบสองค่า (0 – 1) สำหรับใช้ในการวิเคราะห์ข้อสอบหรือข้อคำถามที่มีกระบวนการตอบหลายลำดับขั้น ซึ่งจำเป็นต้องมีการตรวจให้คะแนนการตอบถูกต้องหรือตอบถูกบางส่วนในแต่ละลำดับขั้นของกระบวนการตอบ เช่น ข้อสอบวัดผลสัมฤทธิ์ทางคณิตศาสตร์ ซึ่งมีการให้คะแนนแบบคำตอบถูกบางส่วน แบบวัดเจตคติ บุคลิกภาพ เชาว์ปัญญา ที่มีการให้คะแนนคำตอบเป็นลำดับขั้น PCM มีลักษณะเป็นโมเดลที่พัฒนาขยายต่อจากโมเดลการตอบสนองข้อสอบที่มี 1 พารามิเตอร์ (1-parameter Model) จึงมีลักษณะพารามิเตอร์มาตรฐานคล้ายโมเดลของ (Rasch Model) และใช้หลักการคำนวณความน่าจะเป็นของการตอบแต่ละระดับขั้นการตอบโดยตรงแบบขั้นตอนเดียว (Direct IRT Method)

ลักษณะข้อคำถามแต่ละข้อ (i) อธิบายด้วยค่าพารามิเตอร์ระดับความยากระหว่างรายการคำตอบที่อยู่ถัดไป (item step difficulty, δ_{ij}) โดยทุกข้อมีค่าพารามิเตอร์ความชันเท่ากัน (equal slope) สมมุติว่าข้อคำถาม i มีคะแนน $x = 0, 1, \dots, m_i$ โดยมีจำนวนรายการคำตอบ $K_i = M_i + 1$ เช่น รายการข้อคำถามที่มี 4 รายการคำตอบ ไม่เคย (0) น้อย (1) ปานกลาง (2) มาก (3) เป็นต้น ในการตอบคำถามข้อนี้ ผู้ตอบจะเลือกรายการสูงสุดต้องคิด 3 ขั้นตอนโดยทำการตัดสินใจ ขั้นที่ 1 เลือกระหว่างไม่เคย (0) กับน้อย (1) ขั้นที่ 2 เลือกระหว่างน้อย (1) กับปานกลาง (2) และขั้นที่ 3 เลือก

ระหว่างปานกลาง (2) กับมาก (3) ค่า δ_{ij} ไม่ได้บอกถึง θ ของผู้ตอบที่มีโอกาส 0.50 ในการเลือกรายการคำตอบที่อยู่เหนือถัดไปอย่างโมเดล GRM แต่ δ_{ij} แสดงถึงระดับความยากสัมพัทธ์ของแต่ละขั้นตอนการตอบ ภายในคำถามข้อเดียวกัน แต่ละขั้นตอนอาจง่ายกว่าหรือยากกว่าบางขั้นตอน δ_{ij} ยังแสดงถึงตำแหน่ง θ ที่ไค้การเลือกรายการคำตอบตัดกัน จึงบ่งบอกถึงผู้ตอบที่มีคุณลักษณะ θ จะมีโอกาสเลือกรายการคำตอบใดมากกว่ากัน

สำหรับ $x = j$ รายการคำตอบไค้ฟังก์ชันการเลือกรายการคำตอบของโมเดล PCM เขียนได้ดังนี้

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

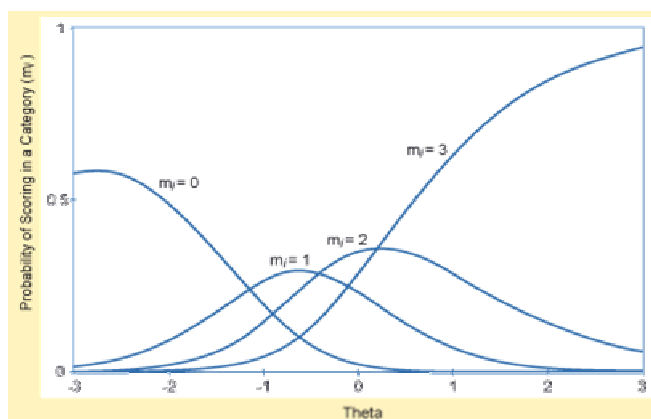
$$\text{เมื่อ } \sum_{j=0}^0 (\theta - \delta_{ij}) = 0$$

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

δ_{ij} = ค่าพารามิเตอร์ระดับความยากของขั้นตอนที่ j ในข้อ i (Item step difficulty)

เมื่อ $j = 1, 2, \dots, m_i$ ค่าที่สูงแสดงถึงขั้นตอนการตอบนั้นมีความยากสัมพัทธ์สูงกว่าขั้นอื่น

มาสเตอร์ (ศิริชัย กาญจนวาสี. 2545: 85 อ้างอิงจาก Masters. 1985) ได้สาธิตการใช้โมเดล PCM สำหรับวิเคราะห์ข้อคำถามที่มีคำตอบ 4 รายการ ดังนี้



ภาพประกอบ 5 ไค้รายการคำตอบของข้อคำถามที่มี 4 รายการคำตอบที่วิเคราะห์โดยใช้ PCM

ที่มา: John Behrens. <http://www.creative-wisdom.com/computer/sas/IRT.html>

ความสัมพันธ์ระหว่างคุณลักษณะ θ กับระดับขั้นของรายการคำตอบแต่ละข้อคำถามตามโมเดล PCM สามารถนำมาคำนวณคะแนนที่คาดหวัง (Expected score) หรือคะแนนจริง (True score) ของแต่ละข้อคำถามได้ตามสูตรดังนี้

$$E(X) = \sum_{x=0}^{m_i} xP_x(\theta)$$

โค้งของฟังก์ชัน ดังกล่าวแสดงถึงการแจกแจงคะแนนรายข้อที่คาดหวังของผู้ตอบ ที่มีคุณลักษณะ θ เมื่อรวมโค้งของทุกข้อเข้าด้วยกันทั้งฉบับ จะทำให้ได้โค้งการแจกแจงคะแนนดิบรวมที่คาดหวังของผู้ตอบที่มีคุณลักษณะ θ ข้อดีของโมเดล PCM ซึ่งคล้ายกับโมเดลของราล์ซ คือ คะแนนดิบที่คาดหวังเป็นค่าสถิติที่เพียงพอสำหรับการคำนวณค่าของคุณลักษณะ θ ดังนั้นคะแนนดิบของข้อคำถามชุดเดียวกันที่เป็นไปตามโมเดล PCM จะสอดคล้องหรือสมมูลกับตำแหน่งคุณลักษณะของผู้ตอบบนสเกลของ θ แต่อย่างไรก็ตามคงต้องอยู่บนพื้นฐานว่าข้อคำถามทุกข้อจะต้องมีความสัมพันธ์ที่เท่าเทียมกันกับคุณลักษณะ θ

Generalized Partial Credit Model

มูรากิ (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 จึงเขียนได้โดยแทนค่าพารามิเตอร์ความชันลงในสูตรของโมเดล PCM ดังนี้

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

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

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

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

α_i = ค่าพารามิเตอร์ความชันของข้อคำถามที่ i

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

Rating Scale Model (RSM)

มีผู้พัฒนาโมเดลที่เรียกว่า Rating Scale Model หลายลักษณะซึ่งมีความหมายแตกต่างกันในแง่ของความสลับซับซ้อนอย่างหลายแนวคิด ในที่นี้จะขอนำเสนอ Rating Scale Model (RSM) ตามแนวคิดของ แอนดริช (Andrich, 1978)

RSM เป็นโมเดลที่คล้าย PCM อยู่บนพื้นฐานของโมเดลการตอบสนองข้อสอบที่มี 1 พารามิเตอร์ (1 - Parameter Model) และใช้หลักการคำนวณความน่าจะเป็นของการตอบแต่ละระดับขั้นการตอบโดยตรงแบบขั้นตอนเดียว (Direct IRT Method) โมเดล RSM มีความคล้ายโมเดล PCM แต่ก็มีแตกต่างที่สำคัญอยู่หลายประการใน RSM มีลักษณะคำถามแต่ละข้อ (i) อธิบายด้วยค่าพารามิเตอร์ตำแหน่งของข้อคำถาม (Location parameter; λ_i) ซึ่งสะท้อนถึงความยากง่ายสัมพัทธ์ของแต่ละข้อ และค่าพารามิเตอร์ Threshold ของรายการคำตอบ (δ_j) ซึ่งเป็นค่าร่วมกันของทุกข้อโดย $j = k - 1$ รายการคำตอบ โมเดล RSM มีข้อตกลงเบื้องต้นว่าทุกข้อคำถามมีค่าพารามิเตอร์ความชันเท่ากันหรือมีค่าอำนาจจำแนกเท่ากัน (มีค่าพารามิเตอร์ระดับความยากของขั้นการตอบร่วมกันทั้งฉบับ) และคะแนนดิบรวมเป็นค่าสถิติที่เพียงพอสำหรับการคำนวณค่าคุณลักษณะ θ ของผู้ตอบ ในขณะที่โมเดล PCM ยอมให้ค่าระดับความยากของแต่ละขั้นการตอบมีความผันแปรได้ระหว่างข้อคำถาม

ใน RSM ค่าพารามิเตอร์ ของแต่ละรายการคำตอบถูกจำแนกออกเป็น 2 ส่วนประกอบด้วย λ_i และ δ_j โดย $(\delta_{ij} = (\lambda_i + \delta_j))$ ดังนั้น โค้งฟังก์ชันการเลือกรายการคำตอบของโมเดล RSM เขียนได้ดังนี้

$$P_x(\theta) = \frac{\exp \left\{ \sum_{j=0}^x [\theta - (\lambda_i + \delta_j)] \right\}}{\sum_{x=0}^m \exp \left\{ \sum_{j=0}^x [\theta - (\lambda_i + \delta_j)] \right\}}$$

$$\text{เมื่อ } \sum_{j=0}^0 [\theta - (\lambda_i + \delta_j)] = 0$$

$P_{ix}(\theta)$ = ความน่าจะเป็นที่ผู้ตอบซึ่งมีคุณลักษณะ θ จะเลือกหรือสามารถทำรายการคำตอบขั้นที่ x จากจำนวน M ขั้น (Step)

λ_i = ค่าพารามิเตอร์ตำแหน่งของข้อคำถาม i (Location parameter)

δ_j = ค่าพารามิเตอร์ Threshold ของขั้นคำตอบที่ j (Category intersection Parameter) ซึ่งเป็นค่าร่วมกันของทุกข้อคำถาม

Nominal Response Model

บอค (Bock, 1972) ได้พัฒนา Nominal Response Model (NRM) สำหรับใช้วิเคราะห์ข้อสอบหรือข้อคำถามที่รายการคำตอบไม่จำเป็นต้องถูกจัดเรียงลำดับ เช่น ข้อสอบแบบหลายตัวเลือก (Multiple choices) ข้อคำถามวัดเจตคติ ข้อคำถามสำหรับประเมินบุคลิกลักษณะ เป็นต้น Special cases NRM มีลักษณะเป็นโมเดลทั่วไปที่ใช้หลักการคำนวณความน่าจะเป็นของการตอบแต่ละรายการคำตอบ โดยตรงแบบขั้นตอนเดียว (Direct IRT Model) โมเดลที่คำนวณความน่าจะเป็นแบบขั้นตอนเดียวที่กล่าวมาข้างต้นต่างเป็นโมเดลลักษณะเฉพาะ (Special case) ของโมเดล NRM ซึ่งลักษณะข้อคำถามแต่ละข้อ (i) อธิบายได้ด้วยค่าพารามิเตอร์ความชันของแต่ละรายการคำตอบ (Slope of the trace lines ; α_{ix}) และค่าพารามิเตอร์จุดตัดของแต่ละรายการคำตอบ (Intercept parameter, c_{ix}) โอกาสในการตอบรายการคำตอบ X ($X=0,1,\dots,m_i$) ตามโมเดล สามารถเขียนเป็นฟังก์ชันได้ดังนี้

$$P_{ix}(\theta) = \frac{\exp(\alpha_{ix}\theta + c_{ix})}{\sum_{x=0}^{m_i} \exp(\alpha_{ix}\theta + c_{ix})}$$

เมื่อ $P_{ix}(\theta)$ = ความน่าจะเป็นที่ผู้ตอบซึ่งมีคุณลักษณะ θ จะเลือกรายการคำตอบที่ x ในเมื่อ $x = 0, 1, \dots, m_i$

α_i = ค่าพารามิเตอร์ความชันของแต่ละรายการคำตอบ (Slope parameter)

c_{ix} = ค่าพารามิเตอร์จุดตัดของแต่ละรายการคำตอบ (intercept parameter)

เพื่อให้สามารถประมาณค่าพารามิเตอร์ในสมการได้ จึงจำเป็นต้องกำหนดเงื่อนไขให้ $\sum \alpha_{ix} = \sum c_{ix} = 0$ บางกรณีมีการกำหนดให้ค่าพารามิเตอร์ของรายการคำตอบต่ำสุด $\alpha_{ix} = c_{ix} = 0$ ตามโมเดล NRM จะมีการประมาณค่าพารามิเตอร์ α_{ix} และ c_{ix} สำหรับแต่ละรายการคำตอบซึ่งมีทั้งหมดจำนวน m_i+1 รายการคำตอบในแต่ละข้อ

การวิจัยด้านการวัดผลการศึกษาเกี่ยวกับทฤษฎีการตอบสนองของรายชื่อนอกจากจะเป็นการพัฒนาโมเดลและตรวจสอบโมเดลแล้ว ยังมีการวิจัยเกี่ยวกับการสร้างมาตรวัดคะแนนความสามารถ (ability scores) ของผู้สอบมีการพัฒนาคะแนนความสามารถในรูปฟังก์ชันของพารามิเตอร์ความสามารถของผู้สอบรูปต่าง ๆ โดยมีการกำหนดน้ำหนักคะแนนแบบต่างๆและมีการวิจัยเกี่ยวกับการพัฒนาวิธีประมาณค่าพารามิเตอร์ความสามารถแบบต่าง ๆ ด้วย

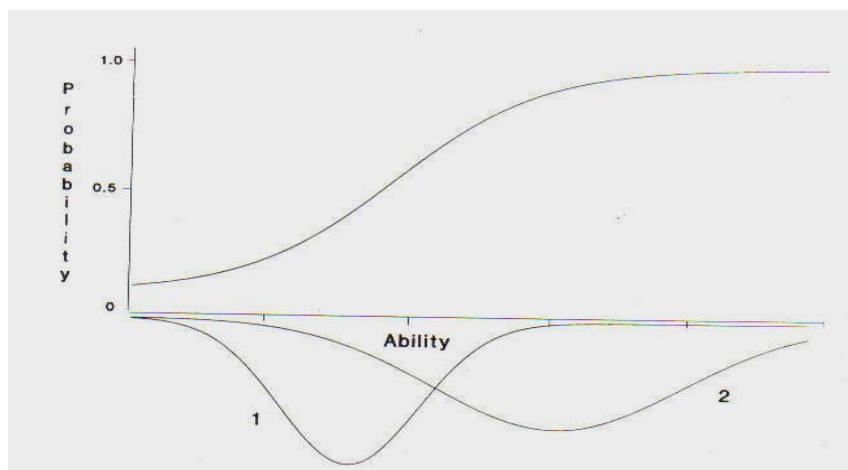
การวิจัยด้านการวัดผลการศึกษาเกี่ยวกับทฤษฎีการตอบสนองของรายชื่อนอกจากจะเป็นการพัฒนาโมเดลและตรวจสอบโมเดลแล้ว ยังมีการวิจัยเกี่ยวกับการสร้างมาตรวัดคะแนนความสามารถ (ability scores)ของผู้สอบมีการพัฒนาคะแนนความสามารถในรูปฟังก์ชันของพารามิเตอร์ความสามารถของผู้สอบรูปต่าง ๆ โดยมีการกำหนดน้ำหนักคะแนนแบบต่างๆและมีการวิจัยเกี่ยวกับการพัฒนาวิธีประมาณค่าพารามิเตอร์ความสามารถแบบต่าง ๆ ด้วย

1.5 คุณสมบัติของความไม่แปรเปลี่ยนของค่าพารามิเตอร์ (Invariance)

เมื่อ IRT Model สอดคล้อง (fit) กับข้อมูลจะทำให้เกิดความไม่แปรเปลี่ยนของพารามิเตอร์ของข้อสอบ (Item parameter) และพารามิเตอร์ความสามารถของผู้สอบ (Ability parameter) ซึ่งเป็นคุณสมบัติสำคัญของ IRT

1.5.1 ความไม่แปรเปลี่ยนของพารามิเตอร์ข้อสอบ (Item Invariance)

- ค่าพารามิเตอร์ของข้อสอบไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบ
- ICC ลักษณะเดียวกัน (a, b, c) สำหรับทุกกลุ่มความสามารถของผู้สอบ แสดงว่ามีความคงที่ข้ามกลุ่มผู้สอบ



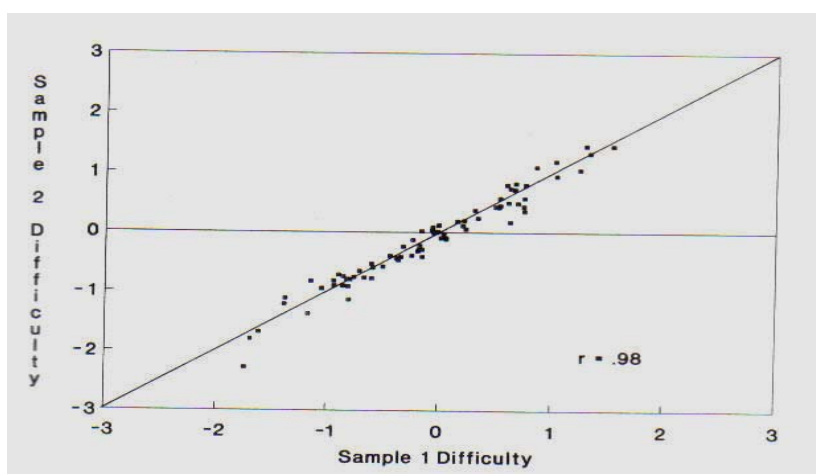
ภาพประกอบ 6 โค้งลักษณะข้อสอบและการแจกแจงความสามารถของผู้สอบ 2 กลุ่ม

ที่มา: Hambleton, Swaminathan and Roger. 1991: 8

1.5.2 ความไม่แปรเปลี่ยนของพารามิเตอร์ความสามารถของผู้สอบ(Ability

Invariance)

- ค่าพารามิเตอร์ของผู้สอบไม่แปรเปลี่ยนไปตามชุดของข้อสอบ
- เมื่อนำข้อสอบต่างชุด (ทุกข้อมุ่งวัดสิ่งเดียวกัน) เช่นข้อสอบชุดค่อนข้างง่ายกับข้อสอบชุดค่อนข้างยาก มาสอบวัดผู้สอบกลุ่มเดียวกัน ค่า θ ที่ประมาณได้ จากข้อสอบทั้งสองชุดมีความแตกต่างกัน ไม่เกิน SEE แสดงว่าการประมาณค่าความสามารถมีความคงที่ข้ามชุดของข้อสอบ



ภาพประกอบ 7 ค่าพารามิเตอร์ของผู้สอบที่คำนวณได้จากแบบทดสอบ 2 ฉบับที่มีความแตกต่างกัน

ที่มา : Hambleton, Swaminathan and Roger. 1991: 25

1.6 ฟังก์ชันสารสนเทศของข้อสอบและแบบทดสอบ (Item and Test Information)

1.6.1 ฟังก์ชันสารสนเทศของข้อสอบ (Item Information) ฟังก์ชันสารสนเทศของข้อสอบเป็นดัชนีผสมที่สร้างจากดัชนีคุณลักษณะของข้อสอบหลายลักษณะประกอบด้วยค่าพารามิเตอร์ความยากค่าพารามิเตอร์อำนาจจำแนกและค่าความแปรปรวนของคะแนนรายข้อเพื่อใช้บ่งชี้คุณภาพของข้อสอบ (ศิริชัย กาญจนวาสี. 2545: 54 อ้างอิงจาก Birnbaum. 1968)

$$I_i(\theta) = \frac{[P'_i(\theta)]^2}{P_i(\theta)Q_i(\theta)} \quad , i = 1, 2, \dots, k$$

เมื่อ $I_i(\theta)$ = ค่าฟังก์ชันสารสนเทศหรือค่าสารสนเทศที่ได้รับจากข้อสอบที่ i สำหรับตอบที่มี
ความสามารถ θ

$$P'_i(\theta) = P'_i = \text{ความชันของฟังก์ชันการตอบสนองของข้อสอบที่ } i \text{ ณ ตำแหน่ง ความสามารถ } \theta$$

$$P_i(\theta) = P_i = \text{ความน่าจะเป็นที่ผู้ตอบซึ่งมีความสามารถ } \theta \text{ จะตอบข้อสอบที่ } i \text{ ได้ถูกต้อง}$$

$$Q_i = Q_i = 1 - P_i(\theta)$$

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

ตาราง 1 การคำนวณค่าฟังก์ชันสารสนเทศ ค่าสารสนเทศสูงสุดของข้อสอบและตำแหน่ง θ ที่มีสารสนเทศของข้อสอบสูงสุด สำหรับโมเดลการตอบสนองข้อสอบโลจิสติก 1, 2, 3 พารามิเตอร์

ค่าประมาณ	1 - parameter	2 - parameter	3 - parameter
$I_i(\theta)$	$D^2 P_i Q_i$	$D^2 a_i^2 P_i Q_i$	$D^2 a_i^2 Q_i (P_i - c_i)^2 / (1 - c_i)^2$
P'_i	$D P_i Q_i$	$D a_i^2 P_i Q_i$	$D a_i Q_i (P_i - c_i) / (1 - c_i)$
$I_i(\theta)_{\max}$	$\frac{1}{4} D^2$	$\frac{1}{4} D^2 a_i^2$	$D^2 a_i^2 \frac{[1 - 20c_i - 8c_i^2 + (1 + 8c_i)^{3/2}]}{8(1 - c_i^2)}$
$\theta_{\max}$	b_i	b_i	$b_i + \frac{1}{Da_i} \left[\ln 1 + \frac{(1 + c_i)^{1/2}}{2} \right]$

ที่มา : ศิริชัย กาญจนวาสี. 2545: 55

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 $\theta_{\max}$ ถ้า $C_i = 0$, $\theta_{\max} = b$ แต่ถ้า $C_i > 0$, $\theta_{\max} > b$

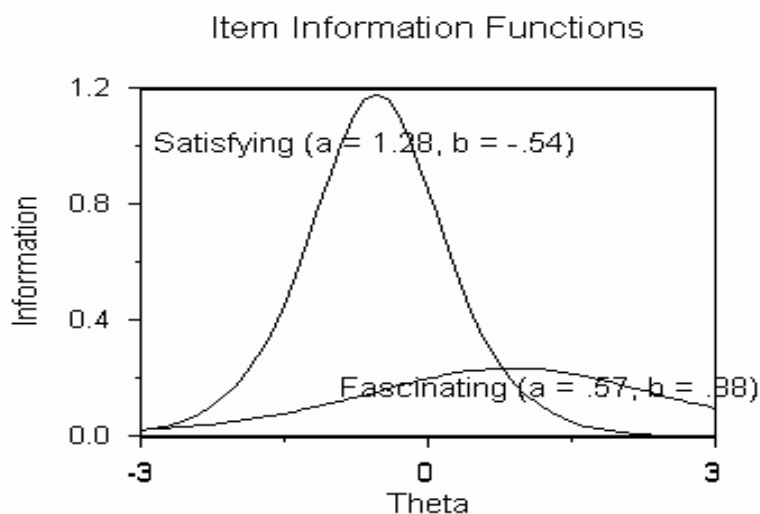
1.6.2 ฟังก์ชันสารสนเทศของแบบทดสอบ (Test Information) การวิเคราะห์ตาม IRT จะใช้แบบแผนการตอบสนองแบบทดสอบเป็นรายชื่อในการประมาณค่าความสามารถของผู้สอบ ดังนั้น การประเมินคุณภาพของแบบทดสอบ จึงสามารถพิจารณาจากความถูกต้องแม่นยำในการประมาณค่าความสามารถของผู้ตอบโดยใช้ดัชนีที่เรียกว่า สารสนเทศของแบบทดสอบ [Test Information; $I(\theta)$] ซึ่งเป็นค่าฟังก์ชันสารสนเทศของแบบทดสอบอันเกิดจากผลรวมเชิงพีชคณิตของค่าฟังก์ชันสารสนเทศของข้อสอบแต่ละข้อรวมเข้าด้วยกันทั้งฉบับ ณ ตำแหน่ง θ เดียวกันดังสูตร

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

เมื่อ $I_i(\theta)$ = ค่าฟังก์ชันสารสนเทศหรือค่าที่ได้รับจากแบบทดสอบสำหรับผู้ตอบที่มีความสามารถ θ

ค่าฟังก์ชันสารสนเทศของข้อสอบแต่ละข้อ จึงมีส่วนอย่างเป็นอิสระจากกันต่อค่าสารสนเทศของแบบทดสอบ ลักษณะเช่นนี้ไม่ได้เกิดกับทฤษฎีการทดสอบแบบดั้งเดิม เนื่องจากค่าความยากและค่าอำนาจจำแนกของข้อสอบแต่ละข้อต่างก็ส่งผลต่อค่าความเที่ยงของแบบทดสอบทั้งฉบับแต่ไม่สามารถคำนวณค่าของแต่ละข้อได้อย่างเป็นอิสระจากกัน ดังนั้นคะแนนที่ได้จึงขึ้นอยู่กับลักษณะเฉพาะของกลุ่มข้อสอบ และแบบทดสอบเฉพาะฉบับที่เลือกมาใช้

เนื่องจากค่าสารสนเทศมีความสัมพันธ์ผกผันกับความคลาดเคลื่อนมาตรฐานของการประมาณค่า ดังนั้นถ้าค่าสารสนเทศของแบบทดสอบมีค่าสูงในช่วง θ ใดก็就会有ความถูกต้องแม่นยำสูงในการประมาณค่าความสามารถของผู้ตอบในช่วง θ นั้น ๆ โดยมีความคลาดเคลื่อนมาตรฐานของการประมาณค่าต่ำ



ภาพประกอบ 8 โค้งสารสนเทศของข้อสอบ

ที่มา: Michael T. Brannick. <http://luna.cas.usf.edu/~mbrannic/files/pmet/irt.htm>

1.6.3 ประสิทธิภาพสัมพัทธ์ (Relative Efficiency)

1.6.3.1 การเปรียบเทียบค่าฟังก์ชันสารสนเทศระหว่างแบบทดสอบที่วัดคุณลักษณะเดียวกันสามารถทำได้โดยคำนวณประสิทธิภาพสัมพัทธ์ (Relative Efficiency; $RE(\theta)$) ซึ่งเป็นอัตราส่วนระหว่างค่าฟังก์ชันสารสนเทศของแบบทดสอบต่างฉบับ ณ ตำแหน่ง θ เดียวกันดังนี้

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

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

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

การแปลความหมายของค่าดัชนีประสิทธิภาพสัมพัทธ์ สามารถนำไปช่วยคัดเลือกแบบทดสอบได้ดังนี้

$RE(\theta) = 1$ แสดงว่าแบบทดสอบทั้งสองฉบับมีประสิทธิภาพเท่ากันสำหรับผู้ตอบที่มีระดับความสามารถ θ

$RE(\theta) > 1$ แสดงว่าแบบทดสอบฉบับ A มีประสิทธิภาพสูงกว่าแบบทดสอบฉบับ B สำหรับ ผู้ตอบที่มีระดับความสามารถ 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

1.6.3.2 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 Y ณ ทุกตำแหน่ง θ

1.6.4 การประมาณค่าสารสนเทศของข้อสอบและแบบทดสอบ การประมาณค่าสารสนเทศของข้อสอบ (Item Information Function, IIF) หรือ $I_i(\theta)$ และฟังก์ชันสารสนเทศของแบบทดสอบ (Test Information Function, TIF) หรือ $I(\theta)$ โมเดลการตอบสนองข้อสอบแบบตรรกให้คะแนน 2 ค่า สามารถนำแนวคิดมาประยุกต์ใช้กับโมเดลการตอบสนองข้อสอบแบบตรรกให้คะแนนมากกว่า 2 ค่า ได้เช่นเดียวกัน กล่าวคือ เมื่อพิจารณาข้อสอบหรือข้อคำถามแต่ละข้อ เราสามารถแปลงโค้งรายการคำตอบให้เป็นโค้งสารสนเทศของข้อสอบหรือข้อคำถามได้ จากโค้งสารสนเทศของข้อสอบ เมื่อนำมารวมกัน (ณ ตำแหน่ง θ เดียวกัน) ทำให้ได้โค้งสารสนเทศของแบบทดสอบสามารถเขียนในรูปของฟังก์ชันสารสนเทศของข้อสอบและแบบทดสอบได้ (Dodd, De Ayala and Koah. 1995) ดังสมการ

ฟังก์ชันสารสนเทศของข้อสอบ (Item Information Function, IIF) หรือ $I_i(\theta)$

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

จาก IIF สามารถนำมาคำนวณ ดังนี้

ฟังก์ชันสารสนเทศของแบบทดสอบ (Test Information Function, TIF) หรือ $I(\theta)$

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

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

1.6.5 ความคลาดเคลื่อนมาตรฐานของการประมาณค่า (Standard Error of Estimation) $[SE(\theta)]$ เป็นค่าส่วนเบี่ยงเบนมาตรฐานของการแจกแจงความน่าจะเป็นของค่าประมาณความสามารถที่แท้จริง (θ) ซึ่งเป็นค่าสัดส่วนผกผันกับความถูกต้องแม่นยำของการประมาณค่าความสามารถหรือค่าฟังก์ชันสารสนเทศของแบบทดสอบ ดังสูตร (ศิริชัย กาญจนวาสี 2545: 59) ดังนี้

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

เมื่อ $SE(\theta)$ แทน ความคลาดเคลื่อนมาตรฐานของการประมาณค่าสำหรับ
ผู้ตอบที่มี

ความสามารถ θ

$I(\theta)$ แทน สารสนเทศของแบบทดสอบสำหรับผู้ตอบที่มีความสามารถ θ

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

การวิจัยด้านการวัดผลการศึกษาเกี่ยวกับทฤษฎีการตอบสนองรายข้อ นอกจากจะเป็นการพัฒนาโมเดลและตรวจสอบโมเดลแล้ว ยังมีการวิจัยเกี่ยวกับการสร้างมาตรวัดคะแนนความสามารถ (ability score) ของผู้สอบมีการพัฒนาคะแนนความสามารถในรูปฟังก์ชันของพารามิเตอร์ความสามารถของผู้สอบรูปต่าง ๆ โดยมีการกำหนดน้ำหนักคะแนนแบบต่าง ๆ และมีการวิจัยเกี่ยวกับการพัฒนาวิธีการประมาณค่าพารามิเตอร์ความสามารถต่าง ๆ ด้วย

ทฤษฎี IRT ที่ใช้ในการวิเคราะห์ข้อสอบที่มีการตรวจให้คะแนนหลายค่า เช่น โมเดลเกรด เรสพอนส์ โมเดลเจเนอรัลไลซ์ พาเชียล เครดิต โมเดลพาเชียล เครดิต เป็นโมเดลที่รู้จักกันแพร่หลายและให้ผลการวิเคราะห์ที่คล้ายคลึงกันเป็นส่วนใหญ่ทั้งในแบบทดสอบที่ข้อสอบตรวจให้คะแนนหลายค่าและแบบทดสอบที่ข้อสอบอยู่ในรูปผลรวม ผู้วิจัยมีการประยุกต์ใช้โมเดลที่มี 1 พารามิเตอร์ คือ โมเดลพาเชียล เครดิต ในแบบทดสอบเลือกตอบที่มีคำตอบถูกหลายคำตอบและให้ผู้สอบตัดสินทุกตัวเลือก

ตอนที่ 2 วิธีการตรวจให้คะแนน

วิธีการตรวจให้คะแนนแบบ 0 – 1 มีข้อบกพร่อง หลายประการ (อนันต์ ศรีโสภณ. 2532: 11) เช่น เป็นการยั่วยุหรือส่งเสริมให้นักเรียนที่ไม่มีความรู้ ความสามารถเดามากขึ้น แม้ว่าจะมีสูตรแก้การเดาก็มิได้ทำให้คะแนนหลังจากแก้การเดาแล้ว แทนความรู้ความสามารถที่แท้จริงของนักเรียนได้ การที่นักเรียนทำถูกได้ 1 คะแนน และทำผิดหรือไม่ทำเลยได้ 0 คะแนน ความจริงนักเรียนบางคนอาจมีความรู้ความสามารถไม่สมบูรณ์แต่ก็ไม่ถึงกับไม่รู้เลย คือ มีความรู้บางส่วนวิธีการแบบนี้ไม่สามารถวัดความรู้ความสามารถบางส่วนเหล่านั้นออกมาได้ อนึ่งข้อสอบแต่ละข้อ

ระดับความยากไม่เท่ากัน คือ บางข้อยากมาก บางข้อยากน้อยแต่ก็ให้น้ำหนักของคะแนนเท่ากันหมดทุกข้อ คือ นักเรียนตอบถูกไม่ว่าข้อนั้นจะยากหรือง่ายก็ได้ 1 คะแนนเท่ากัน คะแนนรวมที่ได้จึงไม่สอดคล้องกับความรู้ความสามารถที่แท้จริงของนักเรียน

จากข้อบกพร่องของวิธีการตรวจให้คะแนนแบบ 0 – 1 ดังกล่าว จึงได้มีนักวัดผลการศึกษาหลายท่านพยายามคิดหาวิธีเพื่อพัฒนาวิธีการตรวจให้คะแนนข้อสอบเลือกตอบให้มีประสิทธิภาพยิ่งขึ้น และเพื่อแสดงถึงตำแหน่งความรู้จริงของผู้ตอบช่วยลดความคลาดเคลื่อนจากการเดาสุ่มของผู้ตอบ ทำให้คะแนนมีความเที่ยงตรงและความเชื่อมั่นมากขึ้น (เอมอร์ จังศิริพรปกรณ์. 2546: 12) แบ่งวิธีการให้คะแนนความรู้บางส่วน เป็นประเภทใหญ่ ๆ 4 วิธี ดังนี้

2.1 การให้น้ำหนักแก่ข้อสอบที่แตกต่างกัน (Differential Item Weighting)

2.2 การให้น้ำหนักแก่ตัวเลือกที่แตกต่างกัน (Differential Option Weighting)

2.3 การเปลี่ยนแปลงโครงสร้างของข้อสอบ (Changing the Item Structure)

2.4 การเปลี่ยนแปลงวิธีการตอบ (Changing the Response Method)

โดยแต่ละวิธีมีรายละเอียด ดังนี้

2.1 การให้น้ำหนักแก่ข้อสอบที่แตกต่างกัน (Differential Item Weighting)

หลักการของวิธีนี้ คือ ข้อสอบที่มีคุณภาพจะได้น้ำหนักมากกว่าข้อสอบที่ไม่มีคุณภาพโดยใช้ผลที่ได้จากการวิเคราะห์ข้อสอบเป็นเกณฑ์ คือ ค่าความยาก ความสามารถ ความเชื่อมั่น ความแปรปรวนหรือการพิจารณาของผู้เชี่ยวชาญ หวังและสแตนเลย์ (Wang and Stanley. 1970: 663 – 705) ให้ข้อสรุปว่าการให้น้ำหนักแก่ข้อสอบโดยใช้ทฤษฎีการวัดแบบดั้งเดิมทำให้น้ำหนักของคะแนนแต่ละข้อมีค่าไม่คงที่เปลี่ยนแปลงไปตามกลุ่มตัวอย่าง จึงควรให้คะแนนตามทฤษฎีการตอบสนองของข้อสอบ การกำหนดน้ำหนักคะแนนที่เหมาะสมรายข้อนี้ยังวางอยู่บนพื้นฐานของวิธีการตอบและการให้คะแนนเฉพาะข้อที่ตอบถูกแบบเดิม ไม่ได้มีการพิจารณาให้คะแนนความรู้บางส่วนแก่ผู้ตอบ

2.2 การให้น้ำหนักแก่ตัวเลือกที่แตกต่างกัน (Differential Option Weighting)

หลักการของวิธีนี้ คือ ความรู้บางส่วนสามารถถูกวัดได้จากการให้น้ำหนักคะแนนแก่ตัวเลือกแบ่งวิธีการได้ 2 วิธี วิธีการแรกเป็นการให้น้ำหนักแก่ตัวเลือกโดยใช้วิจารณ์ญาณของผู้เชี่ยวชาญหรือจากทฤษฎีของโครงสร้างความรู้ วิธีที่สองเป็นการให้น้ำหนักจากข้อมูลเชิงประจักษ์ที่ผ่านมาในอดีตหรือปัจจุบันโดยอาศัยสิ่งที่น่าสนใจของตัวเลือก เช่น มาตรฐานเฉลี่ยของแต่ละตัวเลือกที่ผู้สอบเลือกความสัมพันธ์ระหว่างตัวเลือกที่ผู้สอบเลือกกับคะแนนรวมทั้งหมด

2.2.1 วิธีการตรวจให้คะแนนของเดวิสและฟิฟเฟอร์ (Davis and Fifer. 1959: 160 - 161) (อนันต์ ศรีโสภณ. 2532: 13) ได้ทดลองเกี่ยวกับ ค่าความเชื่อมั่นและค่าความเที่ยงตรงของการ

ตรวจให้คะแนนแบบทดสอบความถนัดและแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน โดยให้คะแนนทุกตัวเลือก กลุ่มตัวอย่างที่ใช้เป็นนักเรียนศิษย์การบินที่ Lakeland air Force Base จำนวน 65 คน วิธีการตรวจให้คะแนนมี 2 วิธี คือ วิธีแบบธรรมดา หรือแบบ 0 - 1 และวิธีการให้น้ำหนักทุกตัวเลือก โดยให้ผู้เชี่ยวชาญเป็นผู้กำหนดน้ำหนักตัวเลือกแต่ละตัว พบว่าวิธีการตรวจให้คะแนนของเดวิส และฟิฟเฟอร์ เป็นวิธีหนึ่งที่จะช่วยให้คะแนนที่ได้จากแบบทดสอบมีความเชื่อมั่นสูงขึ้น แต่มีข้อยุ่งยากในการปฏิบัติมาก

2.2.2 วิธีการตรวจให้คะแนนแฟรรี่ (The proposed new method of Frary) (Frary. 1989) วิธีนี้เสนอโดยแฟรรี่เขาต้องการจะปรับวิธีการให้คะแนนของครอสใหม่จากช่วงคะแนน 1 ถึง 7 เป็น -1 ถึง 1 คะแนน เพราะวิธีของครอสในกรณีที่ไม่มีตัวเลือกใดเลย ผู้ตอบจะได้คะแนนเท่ากับจำนวนตัวเลือกในข้อสอบ แฟรรี่ เสนอว่า เขาควรได้ 0 คะแนน และใช้ระบบการแปลงคะแนนเชิงเส้นจากคะแนนของครอสใหม่ และได้วิเคราะห์ผลของการให้คะแนนแก่ความรู้บางส่วน การลงโทษผู้ผิดและการเดาจากคะแนนเฉลี่ยรายชื่อของวิธีการให้คะแนนแบบต่างๆ ให้ข้อเสนอว่า วิธีแบบแฟรรี่ เป็นวิธีการให้คะแนนที่เหมาะสมกับการให้คะแนนความรู้บางส่วน เช่นเดียวกับวิธีการให้คะแนนของคูมส์และอาร์โนลด์ที่จะช่วยให้ความเที่ยงตรงของการวัดสูงขึ้น

เอมอร์ จังศิริพรกรณ์ (2546) พบว่าการให้น้ำหนักคะแนนแก่ตัวเลือกมีประโยชน์บ้างในแง่ที่ช่วยเพิ่มค่าความเที่ยงตรงภายใน การเปรียบเทียบความเชื่อมั่นและความเที่ยงตรงระหว่างแบบการให้น้ำหนักแก่ตัวเลือกกับแบบธรรมดา พบว่า มีค่าไม่คงที่ของความแตกต่างอย่างมีนัยสำคัญ วิธีการเหล่านี้จึงไม่เป็นที่นิยม เนื่องจากการพัฒนาเพื่อให้ได้มาซึ่งน้ำหนักใช้ทรัพยากรมาก การคำนวณการให้คะแนนมีวิธีการยุ่งยากและยากในการอธิบายและตัดสินคะแนนแก่ผู้สอบ

2.3 การเปลี่ยนแปลงโครงสร้างของข้อสอบ (Changing the Item Structure)

วิธีนี้มีรูปแบบของโครงสร้างข้อสอบและ/หรือคำอธิบายที่แตกต่างจากแบบเลือกตอบธรรมดา ดังนี้

2.3.1 ข้อสอบที่มี 2 ตัวเลือก คือ ถูกและผิด ผู้ตอบจะต้องเลือกคำตอบอย่างใดอย่างหนึ่งคะแนนที่ได้ได้จากผลรวมของจำนวนคำตอบที่ถูกต้อง เนื่องจากการตอบแต่ละข้อใช้เวลา น้อยจึงต้องเพิ่มจำนวนข้อให้มากขึ้นมากกว่าข้อสอบแบบเลือกตอบธรรมดา

2.3.2 ข้อสอบที่มีตัวเลือกถูกหลาย ๆ ตัวเลือก โดยต้องมีคำชี้แจงอธิบายให้แก่ผู้สอบ ซึ่งอาจจะระบุจำนวนตัวเลือกที่ถูกหรือไม่ระบุก็ได้ คะแนนที่ได้เท่ากับผลรวมของการระบุตัวเลือกถูกได้ถูกต้อง ในบางกรณีมีการหักคะแนนการตอบที่ตอบตัวเลือกผิด โดยมีจุดประสงค์เพื่อลดการเดา และประเมินความรู้บางส่วน เช่น

(1) วิธีการให้คะแนนของเดรสเซลและสมิท วิธีการให้คะแนนวิธีนี้ (Dressel and Schmidt. 1953: 574 - 595) ได้ศึกษาวิธีการตอบ และวิธีการตรวจให้คะแนนที่ต่างกันเมื่อปรับปรุงอำนาจจำแนกของแบบทดสอบเลือกตอบโดยมีวิธีการตอบ 4 วิธี ดังนี้

1. เปิดโอกาสให้ผู้เลือกตอบอย่างเสรี (Free Choice Test) เป็นแบบทดสอบที่มี 5 ตัวเลือก โดยในแต่ละข้อผู้ตอบจะเลือกตอบกี่ตัวเลือกก็ได้และจะได้คะแนนตามเงื่อนไขถ้าข้อใดเว้นว่างไว้จะไม่ได้คะแนน

ตาราง 2 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. แบบบอกระดับความมั่นใจในการตอบ (A Degree of certainty Test) เป็นแบบที่ผู้เข้าสอบจะต้องแสดงระดับความมั่นใจในการตอบซึ่งมี 4 ระดับ แต่คำถามแต่ละข้อให้คำตอบได้เพียงคำตอบเดียว

3. แบบมีคำตอบถูกมากกว่าหนึ่ง (A Multiple – Answer Test) แบบทดสอบชนิดนี้จะมีคำตอบถูกอยู่หลายคำตอบผู้เข้าสอบจะต้องทำเครื่องหมายทุกข้อที่ถูก

4. แบบมีคำตอบถูกสองคำตอบ (A Two Multiple – Answer Test) ผู้เข้าสอบจะต้องเลือกตอบเพียงสองคำตอบเท่านั้น

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

2.3.3 ข้อสอบที่มีกลุ่มของข้อสอบและกลุ่มของคำตอบที่เรียกว่าข้อสอบแบบจับคู่ โดยผู้สอบต้องเลือกคำตอบที่ถูกคู่กับคำถามที่กำหนดให้ ถ้าจำนวนคำถามเท่ากับจำนวนคำตอบ เรียกว่า การจับคู่อย่างง่าย (simple matching) ถ้าจำนวนคำตอบมากกว่าจำนวนคำถาม เรียกว่า การจับคู่แบบพหุ (multiple matching) (Gulliksen. 1986) ความแตกต่างของสองแบบ คือ การจับคู่แบบพหุ ช่วยลดโอกาสในการเดามากกว่าแต่ข้อจำกัดที่สำคัญของวิธีการนี้ คือ ความยากในการสร้างข้อสอบให้มีความเหมาะสม พบว่า การจับคู่แบบพหุให้ค่าความเที่ยงตรงและค่าความเชื่อมั่นมากกว่าแบบทดสอบเลือกตอบธรรมดา ความเที่ยงตรงและค่าความเชื่อมั่นเพิ่มขึ้นเมื่อสัดส่วนของจำนวนตัวเลือกเพิ่มขึ้นและต้องการใช้เวลาในการทำข้อสอบมากกว่าปกติ

2.4 การเปลี่ยนแปลงวิธีการตอบ (Changing the Response Method)

วิธีการนี้แตกต่างจากวิธีการที่กล่าวมา คือ ใช้น้ำหนักที่ให้ผู้สอบเองซึ่งสะท้อนถึงความรู้ที่มีอยู่ในตัวผู้สอบ การประเมินตนเองนี้เป็นวิธีที่ดีที่สุดในการได้มาซึ่งสารสนเทศเกี่ยวกับความรู้ของผู้สอบในแต่ละข้อสอบจะให้สารสนเทศอย่างสมบูรณ์เกี่ยวกับน้ำหนัก วิธีการให้คะแนนและการควบคุมระดับความมั่นใจในการให้ ซึ่งจะช่วยลดการเดาและความคลาดเคลื่อนในการวัด วิธีนี้ใช้รูปแบบข้อสอบเหมือนแบบทดสอบชนิดเลือกตอบทั่วไปเพียงแต่คำสั่ง วิธีการตอบและการให้คะแนนแตกต่างจากแบบทดสอบชนิดเลือกตอบธรรมดาแต่ละวิธีการไม่ยุ่งยาก แบ่งได้ 6 วิธี ดังนี้ (เอมอร์ จังศิริพรปกรณ. 2546: 14 – 19)

2.4.1 วิธีการแก้การเดา (Correction for guessing)

วิธีการนี้เป็นวิธีการที่ผู้สอบทำข้อสอบเหมือนกับแบบทดสอบชนิดเลือกตอบธรรมดา แตต่างกันที่การให้คะแนน วิธีการแก้การเดา มีวิธีที่น่าสนใจ ดังนี้

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

- วิธีตรวจให้คะแนนของลอร์ด (Lord. 1975: 7 - 11) โดยใช้สูตรการให้คะแนนเพื่อแก้การเดาคำตอบ คือ

$$X_c = R - W/(K - 1)$$

เมื่อ X_c คือ คะแนนที่ได้หลังจากการแก้การเดา

R คือ จำนวนข้อสอบที่ผู้สอบตอบถูก

W คือ จำนวนข้อสอบที่สอบตอบผิด

วิธีแก้การเดาสัมพันธ์กับแบบทดสอบความเร็วที่ผู้ตอบไม่ทัน แต่ถ้าเป็นทดสอบที่ไม่จำกัดเวลา ผู้ตอบอาจไม่ได้ตอบอย่างเดาสุ่มเพียงอย่างเดียวผู้ตอบอาจมีความรู้บางส่วนในการตัด

ตัวลงบางตัวออกได้ ดังนั้นการใช้สูตรแก้การเดาอาจไม่เหมาะสมเพราะควรที่จะพิจารณาถึงการตอบโดยมีความรู้บางส่วนของผู้ตอบด้วย

- วิธีการตรวจให้คะแนนของเดวิส (Davis. 1966: 79) วิธีนี้ต้องใช้สูตรการตรวจให้คะแนนเพื่อแก้การเดาของนักเรียน 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 บางครั้งไม่ได้เกิดจากการเดาแต่เป็นเพราะได้ความรู้ที่ผิดหรือข้อสอบข้อนั้นมีตัวลงที่สมเหตุสมผลจนทำให้ผู้สอบเข้าใจผิดไปได้

(2) วิธีการให้รางวัลแก่ข้อที่ไม่ตอบ

- วิธีการตรวจให้คะแนนของอีเบล (Ebel. 1972: 178) วิธีนี้ตรวจให้คะแนนวิธีนี้เสนอขึ้นเพื่อแก้การเดาหรือการตอบถูกโดยบังเอิญของผู้ตอบ โดยไม่หักคะแนนข้อที่ผิดแต่จะเพิ่มคะแนนให้กับข้อที่ไม่ตอบ โดยถือว่าข้อที่ผู้ตอบเว้นไว้นั้นเป็นข้อที่ผู้ตอบไม่รู้หรือทำไม่ทัน แต่ถ้าผู้ตอบตอบโดยการเดาหรือทำเครื่องหมายลงในกระดาษคำตอบโดยที่ไม่ได้อ่านข้อสอบ สูตรการคะแนนมีดังนี้

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

เมื่อ	X_c	แทน	คะแนนที่แก้การเดา
	R	แทน	จำนวนข้อที่ตอบถูก
	0	แทน	จำนวนข้อที่เว้นไว้
	C	แทน	จำนวนตัวเลือกทั้งหมดในข้อสอบแต่ละข้อ

2.4.2 วิธีการให้คะแนนจากการตัดตัวลง (Elimination Scoring)

วิธีการนี้เป็นวิธีการให้ผู้สอบตัดตัวลงออกไปได้บ้างอย่างถูกต้องแม้จะไม่ทราบคำตอบที่ถูกต้องคืออะไร มีวิธีการและงานที่สนับสนุน ดังนี้

(1) วิธีการตอบและให้คะแนนแบบคัมป์ (The Coombs Response / Scoring method) (Coombs, Milholland and Womer. 1956) วิธีการนี้จำแนกระดับความรู้ของผู้สอบออกเป็น 5 ประเภท ดังนี้

- 1) มีความรู้เต็มที่ (Full knowledge) คือ ผู้ที่สามารถตัดตัวลงทั้งหมดออกได้
- 2) มีความรู้บางส่วน (Partial knowledge) คือ ผู้ที่สามารถตัดตัวลงบางตัวออกได้แต่ไม่ทั้งหมด
- 3) มีความรู้ผิดบางส่วน (Partial misinformation) คือ ผู้ที่สามารถตัดตัวลงบางตัวได้บ้างและตัดตัวถูกออกด้วย
- 4) มีความรู้ผิดเต็มที่ (Full misinformation) คือ ผู้ที่ตัดตัวถูกออกเพียงตัวเดียว
- 5) ไม่มีความรู้ (Absence knowledge) คือ ผู้ที่ไม่ตอบหรือตัดตัวเลือกทุกตัวออกหมด

วิธีการตอบและการให้คะแนนของคัมป์นี้ใช้รูปแบบของแบบทดสอบเลือกตอบที่มีคำตอบถูกเพียงคำตอบเดียวโดยให้ผู้สอบตัดตัวลงที่ไม่ใช่คำตอบที่แน่ใจออกไปมากที่สุด การให้คะแนนผู้ตอบจะได้รับ 1 คะแนน ในแต่ละตัวลงที่ตัดออกอย่างถูกต้องและในกรณีที่ผู้สอบมีความรู้ที่ผิดไปตัดตัวถูกออกโดยเข้าใจว่าเป็นตัวลงจะถูกตัดคะแนน $K - 1$ โดยที่ K คือ จำนวนตัวเลือก ดังนั้นช่วงคะแนนในแต่ละข้ออยู่ระหว่าง $-(K - 1)$ ถึง $(K - 1)$

(2) วิธีการตอบและการให้คะแนนแบบอาร์โนลด์ (The Arnold Scoring / Response Method) (Arnold and Arnold. 1970: 8 - 13) เป็นการให้คะแนนแบบทดสอบเลือกตอบซึ่งยอมให้คะแนนความรู้บางส่วนและควบคุมคะแนนที่เพิ่มขึ้นโดยโอกาสการเดาเป็น 0 ซึ่งเป็นการพัฒนามาจากทฤษฎีเกมเบื้องต้น (Game theory) โดยให้ผู้ตอบตัดตัวเลือกที่แน่ใจว่าเป็นตัวลงหรือคำตอบที่ผิดออกไปให้มากที่สุดเท่าที่เขารู้ สามารถพัฒนาสูตรการให้คะแนน ดังนี้

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

โดยที่ C_d คือ คะแนนเมื่อสามารถตัดตัวลงบางตัวออกไปได้และคะแนนที่เพิ่มขึ้นจากการเดาเป็น 0

D คือ จำนวนตัวลงที่ตัดออกได้ถูกต้อง

K คือ คะแนนลงโทษในกรณีที่ผู้สอบจัดตัวเลือกที่ถูกเข้าใจเป็นตัวลงในกรณีของ 4 ตัวเลือก คะแนนที่ได้เป็นผลจากการตอบของผู้สอบ ดังนี้

- คะแนน C_0 ถ้าผู้เข้าสอบไม่สามารถตัดตัวลงใด ๆ ได้เลย
- คะแนน C_1 ถ้าผู้เข้าสอบไม่สามารถตัดตัวลงออกได้ 1 ตัว
- คะแนน C_2 ถ้าผู้เข้าสอบไม่สามารถตัดตัวลงออกได้ 2 ตัว
- คะแนน C_3 ถ้าผู้เข้าสอบไม่สามารถตัดตัวลงออกได้ 3 ตัว
- คะแนน $-p$ ถ้าผู้สอบตัดตัวลงถูกต้องเพราะเข้าใจว่าเป็นตัวลง

(3) วิธีการตอบและให้คะแนนแบบครอส (The Cross Response/ Method, CRS) วิธีการนี้การตอบให้ผู้ตอบเลือกชุดย่อยของคำตอบที่เป็นตัวลงออก เช่นเดียวกับวิธีการของ คูมพ์และอาร์โนลด์ แต่การให้คะแนนแตกต่างกัน คือวิธีการให้คะแนนของครอสจะให้คะแนนรายข้อเฉลี่ยต่ำสุด เมื่อผู้ตอบตัดตัวเลือกที่เป็นคำตอบโดยเข้าใจว่าเป็นตัวลงและตัดตัวลงอีก $K - 2$ ตัว โดยเข้าใจว่าเป็นชุดของตัวเลือกที่ผิด การให้คะแนนแบบครอสผู้สอบจะได้ 2 คะแนน ในแต่ละตัวลงที่เขาตัดออกได้อย่างถูกต้องและจะได้ 1 คะแนน สำหรับแต่ละตัวเลือกที่ไม่ได้ตัดตัวลง เมื่อคำตอบถูกยังไม่ได้ถูกตัด แต่เมื่อใดที่ผู้ตอบตัดตัวเลือกที่ถูกด้วย misinformation จะได้คะแนนเฉพาะตัวเลือกที่ไม่ได้ตัดตัวเลือกละ 1 คะแนน ในกรณีที่ข้อสอบมี 4 ตัวเลือก คะแนนที่ยอมให้เกิดขึ้นได้มีค่าตั้งแต่ 1 – 7 คะแนน และคะแนนต่ำสุดจะเกิดจากการที่ผู้ตอบตัดตัวเลือกที่ถูกต้องหลังจากตัดตัวลงมาแล้ว 2 ตัว (เอมอร์ จังศิริพรปกรณ์. 2546) เมื่อเปรียบเทียบการให้คะแนนแบบครอสกับแบบอื่น ๆ พบว่าคะแนนแบบครอสให้ค่าความเที่ยงตรงต่ำกว่าวิธีการของคูมพ์และอาร์โนลด์

(4) วิธีการเลือกชุดย่อยของคำตอบ (The Subset Selection Method, SST)

1) เดรสเซลและสมิธ (Dressel and Schmidt. 1953: 574 - 595) เป็นวิธีที่ตรงกันข้ามกับวิธีการให้คะแนนแบบตัดตัวลง (ET) วิธีนี้ให้ผู้สอบเลือกชุดย่อยของคำตอบที่เขาเห็นว่าน่าจะมีตัวเลือกที่เป็นคำตอบถูกรวมอยู่ด้วย การให้คะแนนขึ้นอยู่กับจำนวนตัวเลือกที่มีอยู่ในชุดย่อยของคำตอบ ยิ่งน้อยเท่าใดคะแนนจะได้เพิ่มขึ้น ในกรณีที่ตัวเลือกที่ถูกต้องรวมอยู่ในชุดย่อยของคำตอบนั้นด้วย แต่ถ้าชุดย่อยนั้นไม่มีคำตอบที่ถูกต้องอยู่ด้วยคะแนนจะถูกหักออกไปเท่ากับขนาดของจำนวนตัวเลือกที่ผู้ตอบเลือก ในกรณีที่ผู้ตอบไม่เลือกตัวเลือกใดเลยจะได้ 0 คะแนน

2) วิธีการตรวจให้คะแนนของกิบบอนส์และคณะ กิบบอนส์ โอลกินและโซเบล (Gibbon, Olkin and Sobel. 1997: 259 - 270) เสนอวิธีตรวจให้คะแนนที่เรียกว่า เทคนิคการเลือกกลุ่มย่อย (Subset Selection Technique; SST) ในแบบทดสอบเลือกตอบที่มีตัวเลือกถูกเพียงตัวเลือกเดียว โดยผู้ตอบสามารถเลือกตัวเลือกได้ตั้งแต่ 1 ตัว จนกระทั่งถึง $1 - k$ ตัวเลือกที่ผู้ตอบคาดว่าจะมีคำตอบที่ถูกหรือตัวเลือกที่ถูกรวมอยู่ด้วย เมื่อ k คือ จำนวนตัวเลือกทั้งหมดของข้อสอบแต่ละข้อ การจะได้คะแนนของผู้ตอบขึ้นอยู่กับจำนวนตัวเลือกที่ผู้ตอบเลือก ยิ่งจำนวนตัวเลือกเพิ่มขึ้นค่าของคะแนนจะลดลง ส่วนการให้คะแนนยึดหลักว่าคะแนนที่คาดหวังที่ได้จากการ

เดาสุ่มของผู้ตอบในแต่ละข้อเป็น 0 เช่น กรณีที่แบบทดสอบเลือกตอบมี 5 ตัวเลือก และผู้ตอบอย่าง สุ่มมา 2 ตัวเลือก ความน่าจะเป็นที่ตัวเลือกทั้ง 2 จะถูกคือ $2/5$ และความน่าจะเป็นที่จะตอบผิด คือ $3/5$ ดังนั้นคะแนนที่คาดหวัง คือ

$$E(s) = 3(2/5) - 2(3/5) = 0$$

ในกรณีนี้ถ้าตอบถูก หมายถึง มีคำตอบที่ถูกรวมอยู่จะได้ 3 คะแนน แต่ถ้าตอบผิด คือไม่มีคำตอบที่ถูกรวมอยู่ด้วย จะได้ -2 คะแนน โดยรูปแบบการให้คะแนนสามารถแสดงให้เห็นดังตารางต่อไป

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

จะเห็นว่า ในกรณีที่ผู้ตอบเลือกตอบแล้วไม่มีตัวเลือกที่ถูกรวมอยู่ด้วย คะแนนที่ได้จะมีค่า เป็นลบวิธีการนี้ เป็นวิธีลดค่าการเดาและยังสามารถวัดส่วนของความรู้ย่อย (Partial Information) ของผู้ตอบในเรื่องนั้น ๆ

ทั้ง ET และ SST มีแนวโน้มในการเพิ่มคุณภาพของการจำแนกระดับความรู้ของผู้สอบและ ลดการเดา แต่อย่างไรก็ตามวิธีการยังมีความยุ่งยากและต้องใช้เวลาทำนานกว่าแบบทดสอบ เลือกตอบธรรมดาซึ่งยังเป็นข้อจำกัดของการให้คะแนนและต้องมีการพัฒนาต่อไป

2.4.3 วิธีการของการทดสอบความน่าจะเป็น (Probability Testing)

วิธีนี้เป็นวิธีที่มีความยืดหยุ่นและเปิดโอกาสให้ผู้สอบแสดงความรู้บางส่วน มากที่สุด โดยรายงานความน่าจะเป็นในตัวเลือกที่ผู้สอบคิดว่าเป็นคำตอบที่ถูกโดยมีตำแหน่งคะแนน ที่เป็นไปได้ 101 ตำแหน่งช่วง (0 – 100) โดยแบ่งระดับความรู้เป็น 5 ประเภท ดังนี้ วอลสแตนส์ บุคเดซี และซวิก (Wallsten, Budescu and Zwick. 1993: 176 - 190)

- 1) มีความรู้เต็มที่ (Full knowledge) คือ ให้ความน่าจะเป็น = 1 แก่คำตอบที่ถูกต้อง
- 2) มีความรู้บางส่วน (Partial knowledge) คือ ให้ความน่าจะเป็นมากกว่า 0 แต่ไม่ถึง 1 ($0 < p < 1$) แก่คำตอบที่ถูกต้อง

- 3) มีความรู้ผิดบางส่วน (Partial misinformation) คือ ให้ความน่าจะเป็นมากกว่า 0 แต่ไม่ถึง 1 ($0 < p < 1$) แก่คำตอบที่ผิด
- 4) มีความรู้ผิดเต็มที่ (Full misinformation) คือ ให้ความน่าจะเป็นเท่ากับ 0 แก่คำตอบที่ถูกต้อง
- 5) ไม่มีความรู้ (Absence knowledge) คือ ให้ความน่าจะเป็นแก่ทุกตัวเลือกเท่ากัน หรือไม่ตอบ

ไมเคิล (Michael. 1968) พบว่า วิธีการให้คะแนนความน่าจะเป็น มีความเที่ยงตรงสูงกว่าวิธีเลือกตอบธรรมดา

แฮมเบิลตัน โรเบิร์ตและทาวบ์ (Hambleton, Robert and Taub. 1970) รายงานว่า ความเชื่อมั่นสูงกว่าวิธีเลือกตอบธรรมดา

โคเลเลอร์ (Koehler. 1971) พบว่า มีความไม่แตกต่างกันของความเที่ยงตรงและความเชื่อมั่นระหว่าง NC และ PT ซึ่งยังไม่พบข้อสรุปที่ชัดเจน

2.4.4 วิธีการแสดงความมั่นใจ (Confidence Marking)

1) เดรสเซล และสมิท (Dressel and Schmidt. 1953: 574 - 595) โดยที่ผู้สอบถูกถามให้แสดงความมั่นใจในคำตอบที่คิดว่าถูกโดยใช้ C – point scale (โดยทั่วไป $3 \leq c \leq 5$) หรืออาจเป็นการบอกความมั่นใจเป็นภาษา คือ ไม่น่าจะ, ค่อนข้างแน่ใจ, แน่ใจมาก ซึ่งแบ่งระดับความรู้ออกเป็น 5 ระดับ

- 1) Full knowledge คือ การเลือกคำตอบได้ถูกต้องด้วยความมั่นใจระดับสูงสุด
- 2) Partial knowledge คือ การเลือกคำตอบได้ถูกต้องด้วยความมั่นใจระดับต่ำ
- 3) Partial misinformation คือ การเลือกคำตอบผิด ด้วยความมั่นใจระดับต่ำ
- 4) Full misinformation คือ การเลือกคำตอบผิด ด้วยความมั่นใจระดับสูงสุด
- 5) Absence knowledge คือ การไม่เลือกคำตอบเว้นว่างไว้

ผลการวิจัยพบว่า วิธีการนี้ให้ความเชื่อมั่นของแบบทดสอบสูงกว่าการให้คะแนนแบบ 0,1 แต่วิธีการนี้ต้องการเวลาตอบและการให้คะแนนมากกว่าเดิม

2) อีเชอร์แนชท์ (Echternach. 1972: 217) วิธีการตรวจให้คะแนนโดยการบอกระดับความมั่นใจ ใช้กับแบบทดสอบถูกผิด ต่อมาก็เป็นแบบทดสอบเลือกตอบ วิธีการให้คะแนนแบบนี้จะศึกษาจากความรู้สึกของผู้ตอบที่มีต่อการทำข้อสอบว่าในการเลือกตอบแต่ละข้อนั้นผู้ตอบมีความมั่นใจมากน้อยเพียงใดว่าคำตอบที่เลือกจะเป็นคำตอบที่ถูกต้องซึ่งแม้จะมีผู้ใดกำหนดน้ำหนักคะแนนต่างกันออกไปแต่หลักการร่วมกันคือข้อที่ตอบถูกด้วยความมั่นใจจะกำหนดน้ำหนักคะแนนให้

มากกว่าข้อที่ถูกแต่ไม่มั่นใจและการกำหนดน้ำหนักคะแนนนี้จะปรับคะแนนให้ติดลบสำหรับข้อที่ตอบผิด

2.4.5 วิธีการจัดอันดับอย่างสมบูรณ์ (Complete Ordering)

เป็นวิธีเฉพาะของ การทดสอบความน่าจะเป็น (PT) คือแทนที่จะแสดงให้เห็นความน่าจะเป็นในแต่ละตัวเลือกผู้สอบจะต้องจัดอันดับของตัวเลือก ซึ่งทำให้ง่ายต่อการนำไปใช้ทำให้สามารถจำแนกระดับความรู้ออกเป็น 3 ระดับ ดังนี้

- 1) Full Knowledge คือ การจัดอันดับให้คำตอบถูกอยู่ในตำแหน่งที่สูงที่สุด
- 2) Partial Knowledge คือการจัดอันดับให้คำตอบถูกอยู่ในตำแหน่งกลาง
- 3) Absence of Knowledge คือ การจัดอันดับให้คำตอบถูกอยู่ในตำแหน่งต่ำที่สุด

2.4.6 วิธีการจัดอันดับเป็นบางส่วน (Partial Ordering)

เป็นวิธีที่ผสมผสานระหว่าง วิธีการให้คะแนนแบบตัดตัวลง (ET) และวิธีการจัดอันดับอย่างสมบูรณ์ (CO) โดยให้ผู้สอบจัดอันดับตัวเลือกที่ไม่สามารถตัดออกได้ วิธีนี้ เสนอโดย ดี ฟิเนตตี (De Finetti. 1965)

- 1) Full knowledge คือ การจัดอันดับแก่คำตอบถูกในตำแหน่งสูงที่สุด
- 2) Partial knowledge คือ การจัดอันดับแก่หลายตัวเลือก โดยคำตอบถูก ถูกจัดอันดับในตำแหน่งสูงที่สุด
- 3) Partial misinformation คือ การจัดอันดับแก่คำตอบถูกในตำแหน่งที่ต่ำกว่าตัวลง
- 4) Full misinformation คือ การจัดอันดับตำแหน่งตัวลง โดยตัดคำตอบถูกออก
- 5) Absence knowledge คือ การจัดอันดับแก่ตัวถูกในตำแหน่งต่ำที่สุดหรือไม่ตอบ

วิธีการตรวจให้คะแนนของอนันต์ ศรีโสภา

วิธีนี้ อนันต์ ศรีโสภา (2532: 8 – 9) ได้เสนอวิธีการที่ให้ผู้ตอบ ตอบทั้งตัวเลือกที่ถูกและตัวเลือกที่ผิดโดยให้ทำเครื่องหมายบวก (+) ลงบนตัวเลือกที่ถูกและทำเครื่องหมายลบ (-) ลงบนตัวเลือกที่ผิดและเว้นว่างไว้ถ้าไม่แน่ใจหรือไม่ตอบ วิธีการให้คะแนนจะให้ +1 แก่ตัวเลือกที่ถูก -1 แก่ตัวเลือกที่ผิดและ 0 แก่ตัวเลือกที่ไม่ตอบแล้วคุณคำตอบของผู้ตอบด้วยค่า +1 , -1 หรือ 0 คะแนนของข้อสอบแต่ละข้อจึงเป็นผลบวกของผลคูณตัวเลขเหล่านี้และคะแนนรวมของแบบทดสอบวิชานี้จึงเป็นผลบวกของคะแนนแต่ละข้อเข้าด้วยกัน คะแนนในแต่ละข้อจะมีค่าอยู่ระหว่าง +5 ถึง -5 ในกรณีที่มี 5 ตัวเลือก

วิธีการตรวจให้คะแนนแบบให้คะแนนทุกตัวเลือก (MTF)

วิธีการตรวจให้คะแนนวิธีนี้ อัลบานีสและคนอื่น ๆ (Albanese and others. 1979: 945 - 950) ได้เสนอไว้ในการให้ผู้ตอบเลือกตอบแบบทดสอบเลือกตอบถูกผิดหลายตัวเลือกโดยในข้อสอบ

หนึ่งข้อจะมีตัวเลือกที่ถูกและที่ผิดมากกว่าหนึ่งตัวเลือกแล้วให้ผู้ตอบพิจารณาตัวเลือกที่ผิดและที่ถูก ซึ่งวิธีการตรวจจะให้คะแนน 1 คะแนน ต่อตัวเลือกเมื่อผู้ตอบตอบได้ตรงกับสภาพจริงของตัวเลือกนั้น และให้ 0 คะแนนต่อตัวเลือกที่ผู้ตอบตอบไม่ตรงกับสภาพจริงของตัวเลือกนั้น

จากการศึกษาทฤษฎีและงานวิจัยที่ผ่านมา พบว่า ส่วนใหญ่จะเปรียบเทียบคุณภาพของแบบทดสอบในด้านความเชื่อมั่นและความเที่ยงตรง จากการให้คะแนนความรู้บางส่วนหลาย ๆ วิธี โดยให้ทฤษฎีการวัดแบบดั้งเดิม ซึ่งมีข้อจำกัดในเรื่องแหล่งความคลาดเคลื่อนในการวัดและค่าสถิติที่ได้แปรเปลี่ยนไปตามกลุ่มตัวอย่างที่ใช้ ในปัจจุบันนักวัดผลได้พัฒนาทฤษฎีการตอบสนองข้อสอบ (IRT) เพื่อแก้ปัญหาการวิเคราะห์ตามแนวทฤษฎีแบบดั้งเดิม การวิจัยด้านการวัดผลการศึกษาเกี่ยวกับทฤษฎีการตอบสนองรายข้อนอกจากจะเป็นการพัฒนาโมเดลและตรวจสอบโมเดลแล้วยังมีการวิจัยเกี่ยวกับการสร้างมาตรวัดคะแนนความสามารถ (ability scores) ของผู้สอบ มีการพัฒนาคะแนนความสามารถในรูปฟังก์ชันของพารามิเตอร์ความสามารถของผู้สอบรูปต่าง ๆ โดยมีการกำหนดน้ำหนักคะแนนแบบต่างๆและมีการวิจัยเกี่ยวกับการพัฒนาวิธีประมาณค่าพารามิเตอร์ความสามารถแบบต่าง ๆ ด้วย

ตอนที่ 3 ความรู้เกี่ยวกับแบบทดสอบเลือกตอบ

แบบทดสอบ (Test)

ประเภทของแบบทดสอบสามารถแบ่งประเภทออกได้หลายลักษณะขึ้นอยู่กับเกณฑ์ที่จะใช้

3.1 แบ่งตามสมรรถภาพที่จะวัด แบ่งเป็น 3 ประเภท

3.1.1 แบบทดสอบวัดผลสัมฤทธิ์ (Achievement Test) หมายถึง แบบทดสอบที่วัดสมรรถภาพสมองด้านต่าง ๆ ที่นักเรียนได้รับการเรียนรู้ผ่านมาแล้วว่ามีอยู่เท่าใด แบบทดสอบแบ่งออกเป็น 2 ชนิด

- แบบทดสอบที่ครูสร้างขึ้น หมายถึง แบบทดสอบที่มุ่งวัดผลสัมฤทธิ์ของผู้เรียนเฉพาะกลุ่มที่ครูสอน
- แบบทดสอบมาตรฐาน หมายถึง แบบทดสอบที่มุ่งวัดผลสัมฤทธิ์มีจุดมุ่งหมายเพื่อเปรียบเทียบคุณภาพต่าง ๆ ของนักเรียนที่ต่างกลุ่มกัน เช่น แบบทดสอบมาตรฐานระดับชาติ

3.1.2 แบบทดสอบวัดความถนัด (Aptitude Test) หมายถึง แบบทดสอบที่มุ่งวัดสมรรถภาพสมองของผู้เรียน

3.1.2.1 แบบทดสอบวัดความถนัดทางการเรียน หมายถึง แบบทดสอบที่มุ่งวัดความถนัดทางด้านวิชาการต่าง ๆ เช่น ด้านภาษา

3.1.2.2 แบบทดสอบวัดความถนัดเฉพาะ หมายถึง แบบทดสอบที่มุ่งวัดความถนัดเฉพาะที่เกี่ยวกับงานอาชีพต่าง ๆ หรือความสามารถพิเศษ เช่น ความสามารถทางดนตรี

3.1.3 แบบทดสอบวัดบุคลิกภาพทางสังคม หมายถึง แบบทดสอบที่ใช้วัดบุคลิกภาพและการปรับตัวให้เข้ากับ สังคม ซึ่งเป็นเรื่องที่วัด ได้ยาก ผลที่ได้ไม่คงที่แน่นอน

3.1.3.1 แบบทดสอบวัดเจตคติที่มีต่อบุคคล สิ่งของ เรื่องราว

3.1.3.2 แบบทดสอบวัดความสนใจที่มีต่ออาชีพ การศึกษา

3.1.3.3 แบบทดสอบวัดการปรับตัว เช่น การปรับตัวเข้ากับเพื่อน ๆ

3.2 แบ่งตามลักษณะการตอบ

3.2.1 แบบทดสอบภาคปฏิบัติ หมายถึง แบบทดสอบที่ให้นักเรียนลงมือปฏิบัติจริง

3.2.2 แบบทดสอบข้อเขียน หมายถึง แบบทดสอบที่ใช้การเขียนตอบ

3.2.3 แบบทดสอบปากเปล่า หมายถึง แบบทดสอบที่ใช้การพูดโต้ตอบแทนการเขียน

3.3 แบ่งตามเวลาที่กำหนดให้ตอบ

3.3.1 แบบทดสอบที่จำกัดเวลาการตอบ หมายถึง แบบทดสอบที่ใช้เวลาน้อย

3.3.2 แบบทดสอบที่ไม่จำกัดเวลาการตอบ หมายถึง แบบทดสอบที่ใช้เวลาตอบมาก

3.4. แบ่งตามจำนวนผู้เข้าสอบ

3.4.1 แบบทดสอบรายบุคคล หมายถึง การสอบที่แต่ละคนมักเป็นการสอบภาคปฏิบัติ

3.4.2 แบบทดสอบเป็นชั้นหรือเป็นหมู่ หมายถึง การสอบที่ละหลาย ๆ คน

3.5. แบ่งตามสิ่งเร้าของการถาม

3.5.1 แบบทดสอบทางภาษา หมายถึง แบบทดสอบที่ต้องอาศัยภาษาของสังคมนั้น ๆ เป็นหลัก ใช้กับผู้ที่สามารถอ่านออกเขียนได้

3.5.2 แบบทดสอบที่ไม่ใช้ภาษา หมายถึง แบบทดสอบที่ใช้รูปภาพ สัญลักษณ์หรือตัวเลข

3.6 แบ่งตามลักษณะของการใช้ประโยชน์

3.6.1 แบบทดสอบย่อย หมายถึง แบบทดสอบประจำบท หรือหน่วยการเรียนรู้

3.6.2 แบบทดสอบรวม หมายถึง แบบทดสอบสรุปเนื้อหาที่เรียนผ่านมาตลอดภาคเรียน

3.7 แบ่งตามเนื้อหาของข้อสอบในฉบับ

3.7.1 แบบทดสอบอัตนัย หมายถึง แบบทดสอบที่มีเฉพาะคำถามนักเรียนต้องคิดหาคำตอบเอง

3.7.2 แบบทดสอบปรนัย หมายถึง แบบทดสอบที่มีทั้งคำถามและคำตอบเฉพาะคงที่

การสร้างแบบทดสอบเลือกตอบ

ข้อสอบเลือกตอบ (Multiple Choice Test) ลักษณะทั่วไปคำถามแบบเลือกตอบจะประกอบด้วยตัวเลือกที่เป็นคำตอบถูกและตัวเลือกที่เป็นตัวลวง และคำถามแบบเลือกตอบที่ดี นิยมใช้ตัวเลือกที่ใกล้เคียงกันดูเผิน ๆ จะเห็นว่าทุกตัวเลือกถูกหมด

หลักในการสร้าง

1. เขียนตอนนำให้เป็นประโยคคำถามสมบูรณ์ อาจใส่เครื่องหมายปริศนา (?)
2. เน้นเรื่องจะถามให้ชัดเจนและตรงจุดไม่คลุมเครือ
3. ควรถามในเรื่องที่มีคุณค่าต่อการวัด
4. หลีกเลี่ยงคำถามปฏิเสธ

ขั้นตอนการสร้างแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน

การสร้างแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนมี ขั้นตอนการสร้าง สรุปได้ 3 ขั้นตอน คือ

ขั้นตอนที่ 1 การวางแผนสร้างข้อสอบ

ขั้นตอนที่ 2 การลงมือสร้างข้อสอบ

ขั้นตอนที่ 3 การตรวจสอบคุณภาพข้อสอบก่อนไปใช้

รายละเอียดในการสร้างแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนในแต่ละขั้นตอน มีดังนี้

ขั้นตอนที่ 1 การวางแผนสร้างข้อสอบ ประกอบด้วย

1. ศึกษาวิธีสร้างแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน จากเอกสารและงานวิจัยต่าง ๆ ขั้นตอนนี้เป็นขั้นตอนที่ผู้สร้างแบบทดสอบ ต้องทำการศึกษาค้นคว้าวิธีการสร้างแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนว่ามีแบบใดบ้าง แบบทดสอบแต่ละชนิดมีวิธีการสร้าง และมีข้อดีข้อจำกัดอย่างไร
2. การกำหนดจุดมุ่งหมายของการใช้แบบทดสอบ ในการสร้างแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน ก่อนจะเริ่มเขียนข้อสอบ จะต้องกำหนดจุดมุ่งหมายของการใช้แบบทดสอบ ให้ชัดเจนว่าจะวัดไปเพื่ออะไร จะได้เขียนข้อสอบให้เหมาะสมและสอดคล้องกับจุดมุ่งหมายนั้น
3. การกำหนดเนื้อหาและพฤติกรรมที่ต้องการวัดในตารางวิเคราะห์หลักสูตรผู้สร้างข้อสอบ จะต้องกำหนดขอบเขตเนื้อหา มาตรฐานการเรียนรู้ และสาระการเรียนรู้และพฤติกรรมที่จะวัดในด้านพุทธิพิสัย ได้แก่ ความรู้ ความจำ ความเข้าใจ การนำไปใช้ การวิเคราะห์ การสังเคราะห์และการประเมินค่า
4. การกำหนดลักษณะของข้อสอบ และส่วนอื่น ๆ ที่เกี่ยวข้องกับการสอบ แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนจะเป็นแบบทดสอบอิงเกณฑ์หรืออิงกลุ่มก็ได้ ซึ่งลักษณะข้อสอบจะเป็นแบบปรนัยหรืออัตนัยก็ได้ หรือลักษณะข้อสอบจะเป็นทั้งปรนัยและอัตนัยรวมกันก็ได้ ทั้งนี้ผู้สร้างข้อสอบอาจใช้เกณฑ์ต่อไปนี้ กำหนดลักษณะข้อสอบ ซึ่งได้แก่

4.1 วัตถุประสงค์ของการวัดและประเมินผล

4.2 ระดับพฤติกรรมการเรียนรู้ที่จะวัด

4.3 ลักษณะหรือคุณสมบัติผู้เข้าสอบ

4.4 จำนวนผู้สอบ

4.5 ระยะเวลาที่ใช้ในการสร้างข้อสอบ ดำเนินการสอบ และตรวจข้อสอบ

4.6 ความเป็นอิสระในการตอบ

ส่วนการกำหนดส่วนอื่น ๆ ที่เกี่ยวข้องกับการสอบ ได้แก่ ความยาวของแบบทดสอบหรือจำนวนข้อของข้อสอบและคะแนน ระยะเวลาที่ให้ทำแบบทดสอบ วิธีดำเนินการสอบวิธีการตรวจให้คะแนน การแปลความหมายของคะแนน ตลอดจนค่าใช้จ่ายทั้งหมดที่เกี่ยวข้องกับการสอบ

ขั้นตอนที่ 2 การลงมือสร้างข้อสอบ

1. สร้างข้อสอบ ผู้สร้างข้อสอบลงมือสร้างแบบทดสอบ ตามรายละเอียดในตารางวิเคราะห์หลักสูตร ตามลักษณะของข้อสอบ คำนึงถึงความยากของแบบทดสอบ ระยะเวลาที่ใช้สอบ คะแนนและการตรวจให้คะแนนด้วย

2. ตรวจสอบข้อสอบ ผู้สร้างต้องทบทวน ตรวจสอบข้อสอบ เพื่อให้ข้อสอบที่สร้างขึ้นมานั้นมีความถูกต้อง ครบถ้วนตามรายละเอียดที่กำหนดไว้ในตารางวิเคราะห์หลักสูตรแล้ว จัดพิมพ์เป็นฉบับทดลองเพื่อนำไปใช้ต่อไป

ขั้นตอนที่ 3 การตรวจสอบคุณภาพข้อสอบก่อนนำไปใช้

1. นำแบบทดสอบที่ผู้วิจัยสร้างขึ้น ไปให้ผู้เชี่ยวชาญด้านเนื้อหา และด้านวัดผลการศึกษา จำนวน 3 - 5 ท่าน ตรวจสอบความเชื่อมั่นตรงด้านเนื้อหา (Content validity) โดยให้ผู้เชี่ยวชาญพิจารณาว่าข้อสอบแต่ละข้อนั้นสร้างได้ถูกต้อง และเหมาะสมเพียงใด พิจารณาความสอดคล้องของข้อสอบกับจุดประสงค์การเรียนรู้หรือเนื้อหาตามตารางวิเคราะห์หลักสูตรหรือไม่โดยใช้เกณฑ์การประเมินดังนี้

+1 หมายถึง แน่ใจว่าข้อสอบวัดจุดประสงค์ข้อนั้น

0 หมายถึง ไม่แน่ใจว่าข้อสอบวัดจุดประสงค์ข้อนั้น

-1 หมายถึง แน่ใจว่าข้อสอบไม่วัดจุดประสงค์ข้อนั้น

นำข้อมูลที่ได้ หาค่าความสอดคล้อง (IOC) และคัดเลือกข้อสอบที่มีค่า IOC ตั้งแต่ 0.50 ขึ้นไปจัดพิมพ์เป็นแบบทดสอบฉบับใหม่

2. ทดลองสอบ นำแบบทดสอบที่ได้ปรับปรุง แก้ไขแล้ว ไปทดลองสอบ (Try out) กับนักเรียนที่มีลักษณะคล้ายคลึง หรือนักเรียนที่เพิ่งเคยเรียนในเรื่องนั้น ๆ จำนวนตั้งแต่ 30 คนขึ้นไป

3. วิเคราะห์หาคุณภาพข้อสอบ นำผลการสอบมาวิเคราะห์หาค่าความยาก และ ค่าอำนาจจำแนก เป็นรายข้อ แล้วคัดเลือกข้อสอบที่ใช้ได้ คือ มีค่าความยากง่าย (p) ระหว่าง 0.20 - 0.80 และมีค่าอำนาจจำแนก (r) ตั้งแต่ 0.20 ขึ้นไปจากนั้นนำข้อสอบที่ได้คัดเลือกแล้วจัดพิมพ์เป็นแบบทดสอบฉบับใหม่ นำไปทดลองสอบกับนักเรียนที่มีลักษณะคล้ายคลึง หรือนักเรียนที่เพิ่งเคยเรียนในเรื่องนั้น ๆ จำนวนตั้งแต่ 30 คนขึ้นไป เพื่อหาค่าความเชื่อมั่น

4. จัดพิมพ์เป็นแบบทดสอบฉบับจริง เพื่อนำไปใช้กับกลุ่มเป้าหมายต่อไป

แบบทดสอบเลือกตอบเป็นแบบทดสอบที่ได้รับความนิยมใช้กันมากเพราะสามารถวัดความรู้ความสามารถได้ตั้งแต่ระดับต้นจนถึงระดับสูง คือ วัดความรู้ความเข้าใจ การนำไปใช้ การวิเคราะห์ การสังเคราะห์และการประเมินค่า

แบบทดสอบเลือกตอบเป็นแบบทดสอบที่กำหนดคำตอบไว้ในแต่ละข้อโดยคำตอบนั้นมีทั้งตัวถูกและตัวผิด ผู้สอบจะต้องเลือกเอาคำตอบที่เห็นว่าถูกต้องจากคำตอบที่กำหนดให้เท่านั้น แบบทดสอบโดยทั่วไปประกอบด้วย ส่วนที่เป็นปัญหาหรือคำถามหรือตอนหน้า (Problem or Question or Stem) และส่วนที่เป็นตอนตอบหรือตัวเลือก (Alternative or Choice) ตัวเลือกที่ไม่ถูกต้องเรียกว่าตัวลวง หรือตัวผิด (Foils or Distractors)

การประยุกต์ IRT สำหรับการสร้างแบบทดสอบ

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

การสร้างแบบทดสอบทั่วไป ในกรณีที่มีคลังข้อสอบขนาดใหญ่ ซึ่งเก็บข้อสอบที่มีคุณภาพสูงไว้จำนวนมาก การคัดเลือกข้อสอบโดยพิจารณาจากสารสนเทศ เพื่อนำมาสร้างเป็นแบบทดสอบตามเป้าหมายที่ต้องการ มีขั้นตอนการคัดเลือก ดังนี้ (ศิริชัย กาญจนวาสี, 2550: 80-81 อ้างอิงจาก Lord, 1977b)

1. กำหนดรูปลักษณะของฟังก์ชันสารสนเทศเป้าหมายที่ต้องการ (Target information function)
2. เลือกข้อสอบจากคลังข้อสอบให้ได้ข้อสอบที่มีสารสนเทศของข้อสอบ $[I_j(\theta)]$ ที่จะช่วยเติมเต็มให้เกิดรูปลักษณะตามฟังก์ชันสารสนเทศเป้าหมาย
3. หลังจากเลือกข้อสอบเพิ่มเข้ามาทีละข้อ คำนวณค่าฟังก์ชันสารสนเทศของแบบทดสอบ $[I(\theta)]$
4. เลือกข้อสอบข้อต่อไปตาม 2 และคำนวณ $I(\theta)$ ตาม 3 จนได้รูปลักษณะบรรลุตามฟังก์ชันสารสนเทศตามเป้าหมาย

ตอนที่ 4 การตรวจสอบคุณภาพของแบบทดสอบ

4.1 อำนาจจำแนก (Discrimination Index)

นักการศึกษาหลายท่านได้ให้ความหมายของอำนาจจำแนกไว้ว่า อำนาจจำแนก หมายถึง ความสามารถหรือคุณสมบัติของข้อสอบ แบบทดสอบ หรือแบบวัด ในการจำแนกแยกแยะคน 2 กลุ่มได้ นั่นคือ แยกคนที่มีคุณลักษณะนั้นสูง กับคนที่มีคุณสมบัติที่ต่ำได้ (ล้วน สายยศ และอังคณา สายยศ, 2543:299) นั่นคือ คุณลักษณะรายข้อที่สามารถจำแนกแยกแยะความแตกต่างระหว่างบุคคลตามลักษณะที่ต้องการวัดได้

การหาค่าอำนาจจำแนกมีหลายวิธีขึ้นกับธรรมชาติของคะแนนที่ได้จากการวัดนั้นๆ ดังนี้

วิธีที่ 1 ดัชนีพอยท์ไบซีเรียล (Point – Biserial Index) ดัชนีแบบนี้เป็นลักษณะสหสัมพันธ์ระหว่างตัวแปร 2 ตัวหรือคะแนน 2 กลุ่ม แต่มีข้อตกลงว่า คะแนนกลุ่มหนึ่งเป็นแบบต่อเนื่อง (Continuous Variable) อีกกลุ่มหนึ่งเป็นแบบไม่ต่อเนื่องมี 2 กลุ่ม (Dichotomous Variable) การให้คะแนนทำถูกต้องได้ 1 คะแนน และทำผิดได้ 0 คะแนนเท่านั้น

วิธีที่ 2 ดัชนีสหสัมพันธ์เพียร์สัน มีข้อตกลงว่า กรณีตัวเลือกเป็นคะแนนแบบช่วงเท่ากัน เช่น 1, 2, 3 หรือ 1, 2, 3, 4 หรือมากกว่านั้นก็ใช้ได้ ด้านคะแนนมากมักจะเป็นลักษณะเห็นด้วยอย่างมากหรือมีคุณลักษณะนั้นอยู่อย่างมาก เมื่อผู้ตอบเลือกตอบตัวเลือกที่มีคะแนนมาก ย่อมได้คะแนนรวมมากด้วย หรือผู้ตอบเลือกตอบตัวเลือกที่มีคะแนนน้อย ย่อมได้คะแนนรวมน้อยด้วย ลักษณะของคะแนน 2 อย่างขึ้นลงตามกัน แสดงว่าข้อนั้นจำแนกได้ แต่ถ้าไม่ขึ้นลงตามกันแสดงว่าค่าของอำนาจจำแนกไม่ดี หรืออาจขึ้นลงกลับกันแปลว่าเป็นข้อที่ไม่ดี ไม่ควรนำมาใช้หรือควรนำมาปรับปรุง และควรตรวจเช็คการให้คะแนนให้ดี อาจเป็นข้อเขียนแสดงความคิดเห็นทางลบเวลาให้คะแนนก็ต้องกลับกัน

วิธีที่ 3 อำนาจจำแนกจากการทดสอบที (t - test Index) การใช้ดัชนีนี้เสนอโดย A. L. Edwards ในปี 1957 ใช้ในกรณีคะแนนแสดงความรู้สึกแต่ละข้อมีมากกว่า 1 คะแนน แต่ละข้อควรให้คะแนนเหมือนกัน นั่นคือ ถ้า 3 ก็คะแนน 3 เหมือนกันหมด ถ้าข้อละ 5 คะแนนก็ให้ 5 คะแนนเหมือนกันหมด โดยหลักการก็คือ พยายามหาความแตกต่างของคะแนนกลุ่มที่ได้คะแนนสูงกับกลุ่มที่ได้คะแนนต่ำ ว่าทำข้อนั้น ๆ ได้คะแนนไปตามสภาพความเป็นจริงหรือไม่ ตามทฤษฎีผู้ที่ได้คะแนนรวมสูงควรทำข้อนั้นได้คะแนนสูง ผู้ที่ได้คะแนนรวมต่ำควรทำข้อนั้นได้คะแนนต่ำ ถ้าแบบนี้ความแตกต่างของคะแนนเฉลี่ยกลุ่มสูงและกลุ่มต่ำก็จะต่างกัน ข้อนั้นก็ถือว่าจำแนกคนได้ แต่ในทางปฏิบัติ คะแนนกลุ่มสูงกับคะแนนกลุ่มต่ำอาจไม่แตกต่างกัน หรือนัยกลับกันคือ กลุ่มต่ำอาจอยู่สูงกว่ากลุ่มสูง กรณีนี้อำนาจจำแนกจะใช้ไม่ได้ อำนาจจำแนกแบบที (t - test Index) ที่ควรยอมรับคือ 1.75 ตามแนว Edwards หมายถึง ถ้าคำนวณค่าอำนาจจำแนกได้ตั้งแต่ 1.75 ขึ้นไป ถือว่าข้อนั้นมีอำนาจจำแนกใช้ได้ แต่ถ้าค่า t น้อยกว่า 1.75 ถือว่าใช้ไม่ได้ ต้องแก้ไขปรับปรุง ในกรณีค่า t เป็นลบ (-) และค่าตัวเลขจะสูงกว่าเกณฑ์ ถือว่าใช้ไม่ได้เพราะเป็นผลกลับกัน ถ้าไม่ยึดตามเกณฑ์ของ Edwards ก็สามารถหาค่า t แล้วไปเปิดตารางทดสอบค่า t และ df กำหนดนัยสำคัญ และเป็นการทดสอบทิศทางเดียว ถ้าผลออกมาค่า t มีระดับนัยสำคัญที่ระดับที่ต้องการค่าอำนาจจำแนกนั้นก็สามารถนำไปใช้ได้ (ล้วน สายยศ และอังคณา สายยศ. 2543:302-306)

ในการวิจัยครั้งนี้ ผู้วิจัยหาค่าอำนาจจำแนกของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ด้วยสูตรสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน

4.2 การแสดงหลักฐานความเชื่อมั่น (Reliability)

นักการศึกษาหลายท่านได้ให้ความหมายของความเชื่อมั่นว่า ความเชื่อมั่น เป็นความคงที่หรือความคงเส้นคงวาของแบบวัดหรือแบบทดสอบที่สามารถวัดบุคคลเดียวกัน ในเรื่องเดียวกัน ต่างเวลาและต่างโอกาสกัน ก็ให้ผลเหมือนเดิมไม่ว่าจะวัดหรือทดสอบกี่ครั้งก็ตาม (ล้วน สายยศ และอังคณา สายยศ. 2543: 209) ความเชื่อมั่น ของแบบทดสอบวัดผลสัมฤทธิ์คณิตศาสตร์ หมายถึง คุณสมบัติของแบบทดสอบที่ทดสอบทักษะการคิดคำนวณ นักเรียนชั้นมัธยมศึกษาปีที่ 3 ได้คงที่แน่นอน

ล้วน สายยศ และอังคณา สายยศ (2541:222-232) ได้กล่าวถึงวิธีการหาค่าความเชื่อมั่นของแบบทดสอบไว้ ดังนี้

1. การวัดความแน่นอน (Stability) เป็นการวัดว่าสอบกับเด็กคนเดิม 2 ครั้ง ด้วยแบบทดสอบเดิมคะแนนจะเปลี่ยนแปลงหรือคงเส้นคงวาเพียงใด วิธีการก็คือ เตรียมแบบทดสอบที่ต้องการหาความเชื่อมั่นให้เรียบร้อย แล้วเอาไปสอบกับกลุ่มตัวอย่าง 2 ครั้ง ช่วงการสอบควรอยู่

ระหว่าง 4 – 5 วัน ดังนั้นเด็กคนหนึ่งๆ จะได้คะแนนจากการสอบเดียวกัน 2 ชุด แล้วนำไปหาความเชื่อมั่นจากสูตรของเพียร์สัน

2. การหาความเชื่อมั่นโดยใช้ข้อสอบคู่ขนาน ต้องสร้างแบบทดสอบหนึ่งๆ มีจำนวน 2 ฉบับที่มีคุณภาพเหมือนกันทุกประการ เป็นต้นว่าเนื้อหาเหมือนกัน รูปแบบเหมือนกัน ความยาว ความยาก คะแนนเฉลี่ย และความเบี่ยงเบนมาตรฐานทั้งสองฉบับต้องเท่ากันด้วย เรียกว่า แบบทดสอบหรือข้อสอบคู่ขนาน (Parallel form) การประมาณค่าด้วยวิธีนี้อาศัยแนวคิดที่ว่า แบบทดสอบที่สร้างทั้งสองฉบับจะเป็นตัวแทนของคุณลักษณะที่ต้องการวัด แล้วนำแบบทดสอบทั้งสองฉบับไปทดสอบกับนักเรียนกลุ่มเดียวกัน แล้วนำคะแนนที่ได้จากการทดสอบมาคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์โดยใช้สูตรของเพียร์สัน (บุญเชิด ภิญโญนนตพงษ์. 2527)

3. การหาค่าความเชื่อมั่นแบบแบ่งครึ่งฉบับ (Split – half) เป็นการนำเอาแบบทดสอบฉบับหนึ่งมาแบ่งออกเป็น 2 ฉบับ โดยมีหลักการว่า แต่ละฉบับที่แบ่งจะต้องคลุมเนื้อหา ความยากง่าย จากยากน้อยไปหายากมากตั้งแต่ข้อแรกจนถึงข้อสุดท้าย เมื่อเป็นดังนี้จึงแบ่งแบบทดสอบออกเป็นสองฉบับโดยอาศัยแยกข้อคู่และข้อคี่ แล้วนำคะแนนจากข้อคู่และข้อคี่มาหาสหพันธ์กันแบบธรรมดา ต่อจากนั้นจึงใช้สูตรขยายความเชื่อมั่นทั้งฉบับอีก เรียกว่า สูตรสเปียร์แมน – บราวน์ (Spearman – Brown)

4. การหาความเชื่อมั่นแบบคูเดอร์ – ริชาร์ดสัน (Kuder – Richardson) ในปี ค.ศ.1937 คูเดอร์และริชาร์ดสันได้พัฒนาสูตรที่หาค่าความเชื่อมั่นได้ง่ายเข้า โดยเครื่องมือที่จะหาความเชื่อมั่นโดยวิธีนี้จะต้องมีลักษณะที่วัดองค์ประกอบร่วมกัน และคะแนนแต่ละข้อต้องอยู่ในลักษณะที่ทำถูกต้อง 1 คะแนน ทำผิดได้ 0 คะแนนเท่านั้น ถ้าตรวจให้คะแนนนอกเหนือจากนี้จะใช้วิธีการนี้หาค่าความเชื่อมั่นไม่ได้และวิธีนี้มีสูตรหาความเชื่อมั่นอยู่ 2 สูตร คือ สูตร KR.20 และ KR.21 (ล้วน สายยศ และ อังคณา สายยศ. 2543:215)

5. การหาความเชื่อมั่นโดยวิธีของครอนบาค (Cronbach alpha procedure) ครอนบาคได้พัฒนาสูตรหาความเชื่อมั่นในรูปสัมประสิทธิ์แอลฟา (α – Coefficient) ในปี ค.ศ.1951 โดยพัฒนามาจากสูตร KR.20 ทั้งนี้เป็นเพราะว่าจะได้ใช้หาค่าความเชื่อมั่นกับเครื่องมือที่ไม่ได้ตรวจให้คะแนนเป็น 1 กับ 0 จะตรวจให้คะแนนลักษณะใดก็ได้ เช่น ถ้าทำถูกต้องคะแนนเป็น 10, 8 หรือในลักษณะแบบสอบถามที่ให้คะแนนแต่ละข้อเป็น 3, 2, 1 หรือ 5, 4, 3, 2, 1 ก็ได้ (ล้วน สายยศ และอังคณา สายยศ. 2543:218)

ในการวิจัยครั้งนี้ ผู้วิจัยหาค่าความเชื่อมั่นของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ด้วยวิธีสัมประสิทธิ์แอลฟา (α – Coefficient) ของครอนบาค

4.3 การแสดงหลักฐานความเที่ยงตรง (Validity)

นักการศึกษาหลายท่านได้ให้ความหมายของความเที่ยงตรงไว้สอดคล้องกันว่า ความเที่ยงตรง เป็นคุณสมบัติของแบบวัดหรือแบบทดสอบที่สามารถวัดหรือประเมินผลได้ตรงตามลักษณะหรือจุดประสงค์ที่ต้องการได้อย่างถูกต้องและแม่นยำ (Bloom. 1967: 468; Gronlund. 1976: 65; Anastasi. 1988: 139; อนันต์ ศรีโสภณ. 2525: 43; ล้วน สายยศ และอังคณา สายยศ. 2543: 246; บุญเชิด ภิบุญอนันตพงษ์. 2545: 89) ความเที่ยงตรงเป็นคุณสมบัติที่บ่งบอกถึงคุณภาพของเครื่องมือที่สำคัญที่สุด (American Psychological Association. 1985:9) ซึ่งหมายถึงการตัดสินใจเชิงประเมินที่มีลักษณะบูรณาการเกี่ยวกับระดับความเข้มข้นของการแสดงหลักฐานเชิงประจักษ์และเชิงทฤษฎีเพื่อสนับสนุนความพอเพียงและความเหมาะสมของการอ้างอิง และการใช้ผลรวมของคะแนนการทดสอบ (Messick. 1989:13) ซึ่งความหมายใกล้เคียงกับการให้ความหมายของ สมาคมจิตวิทยาของสหรัฐอเมริกา (American Psychological Association. 1985: 9) ให้ความหมายว่า ความเที่ยงตรงเป็นความเหมาะสม (Appropriateness) การมีความหมาย (Meaningfulness) และสามารถประโยชน์ได้ (Usefulness) ของการอ้างอิงที่เกิดจากคะแนนของแบบทดสอบ หรือความเหมาะสมของการแปลความหมายคะแนนจากคะแนนแบบทดสอบ

คุณลักษณะที่แสดงว่าแบบทดสอบมีคุณภาพดีทั้งฉบับได้แก่ ความเที่ยงตรง หรือ(Validity) และความเชื่อมั่นของแบบทดสอบ (Reliability) ความเที่ยงตรงที่สำคัญมากของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน คือความเที่ยงตรงตามเนื้อหา ซึ่งจะใช้วิธีการพิจารณาจุดมุ่งหมายหรือการพิจารณาตารางวิเคราะห์หลักสูตร โดยการให้ผู้เชี่ยวชาญร่วมกันพิจารณาแล้วนำผลมาคำนวณค่าดัชนีความสอดคล้อง (IOC) ส่วนการหาค่าความเชื่อมั่นจะต้องคำนวณหาจากสูตรต่าง ๆ

การหาความเที่ยงตรงของแบบทดสอบ แบบทดสอบที่ใช้ในการประเมินผลสัมฤทธิ์ทางการเรียน จะพิจารณาดังนี้

1. การประเมินระหว่างเรียน จะพิจารณาข้อสอบที่สร้างขึ้นนั้นวัดได้ตรงตามเนื้อหาและตรงตามพฤติกรรมที่ระบุไว้ในจุดมุ่งหมายที่ต้องการวัด เขียนถูกต้องตามหลักการเขียนข้อสอบ ภาษาและสำนวนที่ใช้เหมาะสมกับวัยของผู้สอบและมีความเป็นปรนัย

2. การประเมินปลายภาคเรียน จะพิจารณาว่าข้อสอบที่สร้างขึ้นทั้งฉบับนั้นมีเนื้อหาพฤติกรรม และจำนวนข้อตรงตามตารางวิเคราะห์หลักสูตรที่สร้างขึ้นหรือไม่ ข้อสอบเขียนถูกต้องตามหลักการเขียนข้อสอบ และมีความเป็นปรนัย

ความเที่ยงตรงของแบบทดสอบหรือเครื่องมือวัด เป็นค่าที่แสดงหลักฐานว่า สิ่งที่วัดกับสิ่งที่ต้องการวัดสอดคล้องกัน ส่วนการตรวจสอบความเที่ยงตรง (Validation) เป็นกระบวนการที่หาหลักฐานมาสนับสนุนความเที่ยงตรง การตรวจสอบความเที่ยงตรง มี 2 ระยะ คือ ระยะก่อนรวบรวม

ข้อมูลกับระยะหลังเก็บข้อมูลจริง การตรวจสอบความเที่ยงตรงก่อนรวบรวมข้อมูลจริง คือการพิจารณาตัดสินโดยผู้ทรงคุณวุฒิหรือผู้เชี่ยวชาญในสาขาวิชาที่เกี่ยวข้องกับสิ่งที่ต้องการวัด ส่วนการตรวจสอบความเที่ยงตรงหลังการเก็บข้อมูล เป็นการใช้ข้อมูลที่ไปวัดมาคำนวณค่าทางสถิติ หรือค่าดัชนีชี้คุณภาพของเครื่องมือตามเกณฑ์สากล เรียกว่าการหาความเที่ยงตรงเชิงประจักษ์ ความเที่ยงตรงแบบนี้ ได้แก่ ความเที่ยงตรงตามโครงสร้าง ความตรงตามสภาพ ความเที่ยงตรงตามพยากรณ์ วิธีการทางสถิติที่ใช้มักเป็น การวิเคราะห์ความสัมพันธ์ระหว่างตัวแปร การวิเคราะห์องค์ประกอบ (Factor analysis) การหาความเที่ยงตรงเชิงจำแนก (Discriminant validity) การตรวจสอบความเที่ยงตรงตามเกณฑ์ภายนอกที่เกี่ยวข้อง (Criterion related validation) ฯลฯ

สำหรับการตรวจสอบความเที่ยงตรงตามเนื้อหา มักทำก่อนเก็บรวบรวมข้อมูลจริง นิยมใช้กลุ่มบุคคลประมาณ 5 คน ที่มีความรู้ในเนื้อหาสาระทฤษฎีทางจิตวิทยาเป็นอย่างดี ให้ทำหน้าที่พิจารณาความสอดคล้องของเครื่องมือวัด กับจุดมุ่งหมาย เนื้อหา ทฤษฎี ดังนี้

1. ความครบถ้วนสมบูรณ์ ตรงตามทฤษฎีที่เกี่ยวข้อง
2. จำนวนข้อเป็นสัดส่วนกับประเด็นหลัก และประเด็นย่อยหรือไม่
3. ประเด็นย่อยเป็นสัดส่วนกับประเด็นหลักหรือไม่
4. ประเด็นหลักครบถ้วน สมบูรณ์ และสอดคล้องกับทฤษฎีที่อ้างหรือไม่
5. ข้อความที่เขียนสะท้อนทฤษฎีที่อ้างได้ถูกต้องหรือไม่

คุณภาพของเครื่องมือด้านความเที่ยงตรง (Validity) หมายถึง การวัดได้ในสิ่งที่ต้องการวัด วัดได้ตรงตามวัตถุประสงค์ซึ่งความเที่ยงตรงแบ่งได้ 4 ประเภท (ล้วน สายยศ และอังคณา สายยศ, 2538) มีหลายประเภท ดังนี้

ความเที่ยงตรงตามเนื้อหา (Content Validity)

ความเที่ยงตรงตามเนื้อหา (Content Validity) หมายถึงการที่เครื่องมือวัดมีข้อความตรงกับสิ่งที่วัด ตรงตามเนื้อหาของสิ่งที่ต้องการศึกษา โดยมีวิธีการหาดังนี้ คือตรวจสอบความสอดคล้องของข้อความกับจุดประสงค์ที่ต้องการจะวัดจากผู้เชี่ยวชาญและ ตัวผู้วิจัยเอง ซึ่งขั้นตอนการหาความเที่ยงตรงด้านเนื้อหามีรายละเอียดดังนี้

1. ผู้วิจัยทำการหาคุณภาพเองโดยเอาเกณฑ์เอกสารที่เป็นเนื้อหาามาพิจารณาว่าตรงตามวัตถุประสงค์หรือไม่ตรงประเด็นในเรื่องที่เราต้องการวัดหรือไม่หรือพิจารณาจากตารางวิเคราะห์หลักสูตร
2. ผู้เชี่ยวชาญทำการหาคุณภาพ มีวิธีการดังนี้

1) ให้ผู้เชี่ยวชาญการวิจัยอย่างน้อย 3 คน ช่วยประเมินเป็นรายบุคคลว่า ข้อความของเครื่องมือแต่ละข้อนั้น สามารถวัดได้ตรงกับเนื้อหาในเรื่องที่เราวิจัยหรือไม่ โดยให้คะแนนตามเกณฑ์ดังนี้

- ข้อความสามารถวัดได้ตามเนื้อหาหรือประเด็นให้ +1
- ข้อความที่ไม่แน่ใจว่าวัดได้ตามเนื้อหาหรือประเด็นให้ 0
- ข้อความไม่สามารถวัดได้ตามเนื้อหาหรือประเด็นให้ - 1

2) นำคะแนนของผู้เชี่ยวชาญทุกคนที่ประเมิน มากรอกลงในแบบวิเคราะห์ความสอดคล้องของข้อความกับวัตถุประสงค์หรือเกณฑ์ เพื่อหาค่าเฉลี่ยสำหรับข้อความแต่ละข้อ

ความเที่ยงตรงตามโครงสร้าง(Construct Validity)

ความเที่ยงตรงตามโครงสร้าง(Construct Validity) หมายถึง การที่เครื่องมือนั้น สามารถวัดได้ ตรงตามลักษณะหรือทฤษฎีที่วางไว้ ความเที่ยงตรงเชิงโครงสร้างสามารถหาได้หลายวิธีดังนี้

1. หาค่าสหสัมพันธ์ของคะแนนจากเครื่องมือของงานวิจัยกับคะแนนจากเครื่องมือต่างชุดกันที่วัดคุณลักษณะเดียวกัน ซึ่งมีการหาแล้วว่าสามารถวัดคุณลักษณะที่จะวัดนั้นได้จริง โดยหาค่าสัมประสิทธิ์สหสัมพันธ์ การหาค่าความเชื่อมั่นจากเครื่องมือที่เป็นคู่ขนานกัน

2. ใช้กลุ่มที่รู้ชัดอยู่แล้ว (Known – Group Technique) เป็นวิธีเปรียบเทียบคะแนนเฉลี่ยระหว่างกลุ่มที่มีลักษณะที่ต้องการจะวัด กับกลุ่มที่ไม่มีคุณลักษณะของสิ่งที่ต้องการจะวัด ถ้ากลุ่มที่มีคุณลักษณะที่ต้องการจะวัดได้คะแนนเฉลี่ยสูงกว่ากลุ่มที่ไม่มี คุณลักษณะที่ต้องการจะวัด แสดงว่าเครื่องมือมีความ เที่ยงตรงเชิงโครงสร้าง

3. ใช้วิธีวิเคราะห์องค์ประกอบ (Factor Analysis) โดยการหาค่าสหสัมพันธ์ภายใน (Inter correlation) ของข้อคำถามแต่ละข้อว่าสามารถวัดองค์ประกอบร่วมกันหรือไม่ ถ้าเครื่องมือ นั้นมี 1 องค์ประกอบ แสดงว่ามีความเที่ยงตรงเชิงโครงสร้าง

4. ให้ผู้เชี่ยวชาญวิเคราะห์หาค่าความสอดคล้องระหว่างข้อความหรือข้อคำถามกับ เกณฑ์หรือทฤษฎี ซึ่งเหมือนกับการหาค่าความเที่ยงตรง เชิงเนื้อหา (IOC)

ความเที่ยงตรงตามสภาพ (Concurrent Validity)

ความเที่ยงตรงตามสภาพ (Concurrent Validity) หมายถึง คุณสมบัติของแบบทดสอบที่สามารถทำหน้าที่วัดสภาพความสามารถที่เป็นจริงของผู้สอบ จากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 โดยคำนวณจากค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนดิบของการสอบกลางภาค กับคะแนนของแบบทดสอบวัดผลสัมฤทธิ์คณิตศาสตร์ที่มีการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็นและให้คะแนนตามความยากประจำชั้น

สัมประสิทธิ์สหสัมพันธ์ในทางบวกที่สูงแสดงถึงคะแนนจากแบบทดสอบสามารถใช้เป็นตัวบ่งชี้ที่ดีของสถานภาพของลักษณะที่มุ่งวัดนั้น (ศิริชัย กาญจนวาสี. 2544: 83) เช่น บุคคลใดที่ทำข้อสอบมนุษยสัมพันธ์ได้คะแนนสูงก็น่าจะมีพฤติกรรมของผู้มีมนุษยสัมพันธ์ดีแต่หากว่าความเป็นจริงแล้วเขาไม่มีพฤติกรรมด้านมนุษยสัมพันธ์ ก็แสดงว่าเครื่องมือนี้ไม่มีความเที่ยงตรงตามสภาพ

วิธีการหา ใช้วิธีการหาค่าสถิติมักใช้ค่าสหสัมพันธ์ หาได้ทั้งรายข้อและทั้งฉบับ เป็นความเที่ยงตรงของคะแนนที่ได้จากเครื่องมือวัด

ความเที่ยงตรงตามการทำนาย (Predictive Validity)

ความเที่ยงตรงตามการทำนาย (Predictive Validity) หมายถึง เครื่องมือนั้นสามารถคาดคะเน พยากรณ์ สภาพหรือการกระทำในอนาคตได้ โดยพิจารณาได้จาก

1. ความสัมพันธ์ระหว่างคะแนนจากการทำเครื่องมือวิจัยกับเกณฑ์ในอนาคต โดยนำเอาผลการทดสอบไปพยากรณ์ผลความสำเร็จในอนาคต โดยหาค่าสัมประสิทธิ์ สหสัมพันธ์ เช่น การคัดเลือกบุคคลเข้าทำงานในสาขาอาชีพต่าง ๆ หากเขามีคะแนนสูงก็แสดงว่า เขาน่าจะทำงานที่มีประสิทธิภาพ

2. เมื่อได้ผลจากในขั้นที่ 1 แล้วก็ให้คำนวณหาค่าการพยากรณ์ โดยเอาคะแนนของตัวแปรต้น หรือตัวแปรอิสระ ซึ่งถือว่าเป็นตัวพยากรณ์ (Predictor) และตัวแปรตามซึ่งถือว่าเป็นเกณฑ์ (Criteria)

ทั้งความเที่ยงตรงตามสภาพและความเที่ยงตรงตามการทำนาย มีหลักการเหมือนกันคือใช้เกณฑ์ภายนอกเพื่อหาความเที่ยงตรง ส่วนความแตกต่างระหว่าง 2 ประเภทนี้คือ เรื่องของเวลา

สำหรับปัจจัยที่ส่งผลต่อความเที่ยงตรงของเครื่องมือ ดังนี้

ค่าความเที่ยงตรงของแบบทดสอบเป็นสิ่งสำคัญมากสำหรับการวัดผลประเมินผล และมีปัจจัยหลายอย่างที่มีอิทธิพลต่อค่าความเที่ยงตรงของแบบทดสอบทั้งทางตรงและทางอ้อม ได้แก่ (Gronlund, 1976: 98 -102)

1. ปัจจัยที่เกิดขึ้นจากผู้สอบ เช่น

1.1 การเดาของผู้สอบในการทดสอบ เช่น การบังคับให้ผู้สอบเดาในการทดสอบจะทำให้ค่าความเที่ยงตรงของแบบทดสอบเพิ่มขึ้น

1.2 นิสัยในการตอบข้อสอบของผู้สอบ เช่น บางคนทำข้อสอบไม่ได้ จะเลือกตอบข้อ ค. ทุกข้อ ซึ่งทำให้ไม่ทราบว่าเป็นแบบทดสอบนั้นวัดได้ในสิ่งที่จะวัดหรือไม่

1.3 ความไม่พร้อมทางร่างกายและจิตใจของผู้สอบ เช่น ป่วย วิตกกังวล ฯลฯ

2. ปัจจัยที่เกิดจากการบริหารการสอบ เช่น

2.1 การจัดห้องสอบไม่เหมาะสม เช่น ห้องสอบร้อนเกินไป แสงสว่างไม่พอมีเสียงดังรบกวน กรรมการคุมสอบแนะนำการสอบไม่ดีพอ การดำเนินการสอบไม่ยุติธรรม ฯลฯ ปัจจัยต่างๆ เหล่านี้จะทำให้ค่าความเที่ยงตรงของแบบทดสอบต่ำ

2.2 การให้คะแนนที่ไม่เป็นปรนัยจะทำให้แบบทดสอบมีค่าความเที่ยงตรงไม่แน่นอน

2.3 ระยะเวลาในการทดสอบ แบบทดสอบที่กำหนดเวลาในการสอบพอเหมาะจะมีค่าความเที่ยงตรงสูงสุด แต่ถ้าให้เวลามากเกินไปและน้อยเกินไปค่าความเที่ยงตรงจะลดลง (สุพัฒน์ สุกมลสันต์. 2530: 60)

3. ปัจจัยที่เกิดจากแบบทดสอบ เช่น

3.1 คำสั่งในการทำแบบทดสอบไม่ชัดเจน

3.2 คำศัพท์และโครงสร้างของภาษายากเกินไป

3.3 ข้อสอบไม่เป็นปรนัย

3.4 รูปแบบของแบบทดสอบไม่เหมาะสม

3.5 แบบทดสอบสั้นเกินไป

4. ปัจจัยที่เกิดจากการสอนของผู้สอน

5. ปัจจัยที่เกิดจากกลุ่มตัวอย่างได้แก่ จำนวนผู้สอบ เพศ อายุ พื้นฐานการศึกษาพื้นฐานทางสังคม ฯลฯ ปัจจัยต่าง ๆ เหล่านี้มีส่วนทำให้แบบทดสอบมีค่าความเที่ยงตรงแตกต่างกันได้

ความสัมพันธ์ระหว่างค่าความเที่ยงตรงและค่าความเชื่อมั่นของแบบทดสอบ

แบบทดสอบที่มีค่าความเที่ยงตรงต่ำจำเป็นจะต้องมีค่าความเชื่อมั่นต่ำเสมอ แต่แบบทดสอบที่มีค่าความเชื่อมั่นสูงอาจมีค่าความเที่ยงตรงต่ำก็ได้ เพราะแบบทดสอบนั้นไม่ได้วัดในสิ่งที่ต้องการวัดตามต้องการ แต่ค่าสูงสุดของค่าความเที่ยงตรงจะมีได้ไม่เกินรากที่สองของค่าความเชื่อมั่นของแบบทดสอบฉบับเดียวกัน

ตอนที่ 5 งานวิจัยที่เกี่ยวข้อง

5.1 งานวิจัยที่เกี่ยวข้องกับวิธีการตรวจให้คะแนน

สุนิสา จุ้ยม่วงศรี (2546) ได้ศึกษาผลการเทียบคะแนนข้อสอบที่ตรวจให้คะแนนแบบหลายค่า โดยมีจุดมุ่งหมายเพื่อศึกษาผลการเทียบคะแนนที่ข้อสอบตรวจให้คะแนนหลายค่าตามวิธีการเทียบคะแนนในทฤษฎีการตอบสนองข้อสอบและทฤษฎีการทดสอบแบบมาตรฐานเดิมผลการเทียบคะแนนพิจารณาความคล้ายคลึงกันของการแจกแจงคะแนนสะสม ดังนั้นคะแนนความแตกต่างและดัชนีความแตกต่าง (C) จากกลุ่มสอบทาน กลุ่มตัวอย่างที่ใช้ในการวิจัย เป็นนักเรียนชั้นมัธยม

ศึกษาปีที่ 2 ภาคเรียนที่ 1 ปีการศึกษา 2545 ของโรงเรียนในสังกัดกรุงเทพมหานคร ประกอบด้วยกลุ่มตัวอย่างเทียบคะแนน จำนวน 2,312 คน เป็นกลุ่มที่ใช้ทำนายผลการสอบไปวิเคราะห์เทียบคะแนนแบบทดสอบ และกลุ่มตัวอย่างสอบทานผล จำนวน 589 คน เป็นกลุ่มที่ใช้เพื่อประเมินคุณภาพของวิธีการให้คะแนน

ผลการวิจัยพบว่า การสร้างตารางเปรียบเทียบคะแนนแบบทดสอบ จากการเทียบคะแนนทั้ง 5 วิธี 8 รูปแบบ จำแนกเป็นการเทียบคะแนนตามทฤษฎีการทดสอบมาตรฐานเดิม ได้แก่วิธีการเทียบคะแนนเชิงเส้นตรงและอีควิเปอร์เซ็นต์ไทล์ เป็นการเทียบคะแนนดิบจากแบบทดสอบฉบับ Y ปรับสู่สเกลแบบทดสอบฉบับ X และการเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบจำแนกตามวิธีการวิเคราะห์ค่าพารามิเตอร์ ซึ่งมีทั้งวิเคราะห์แยกฉบับและวิเคราะห์พร้อมกันทั้ง 3 วิธี แต่ละวิธีเทียบคะแนนโดยใช้คะแนนจริงและคะแนนสังเกตจากแบบทดสอบฉบับ Y มีค่าสูงกว่าคะแนนของแบบทดสอบฉบับ X ในช่วงคะแนนที่แตกต่างกันในแต่ละวิธีการเทียบคะแนน

การวิเคราะห์คะแนนความคล้ายคลึงกันของการแจกแจงคะแนนสะสม พบว่า ไม่มีความแตกต่างกัน แสดงว่าสัดส่วนความถี่สะสมของคะแนนแปลงกับคะแนนจากแบบทดสอบเป้าหมายในแต่ละระดับมีค่าเท่ากัน

การวิเคราะห์ความแตกต่างของคะแนนแปลง พบว่า การวิเคราะห์ด้วยดัชนีความแตกต่างทั้ง 5 ดัชนี ได้แก่ดัชนี RMSD ถ่วงและไม่ถ่วงน้ำหนักดัชนี MAD ถ่วงและไม่ถ่วงน้ำหนักและดัชนี MSD ไม่ถ่วงน้ำหนัก สรุปได้ว่า การเทียบคะแนนจริงวิธีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยากมีฟังก์ชันการเทียบคะแนนที่คล้ายคลึงกับฟังก์ชันคะแนนพื้นฐานมากกว่าการเทียบคะแนนวิธีอื่น โดยคล้ายคลึงกับฟังก์ชันการเทียบคะแนนอีควิเปอร์เซ็นต์ไทล์มากกว่าการเทียบคะแนนเชิงเส้นตรง จากการวิเคราะห์ MAD ที่ถ่วงน้ำหนัก สรุปได้ว่าการเทียบคะแนนสังเกตวิธีปรับค่าพารามิเตอร์พร้อมกันมีฟังก์ชันมีฟังก์ชันคล้ายคลึงกับการเทียบคะแนนพื้นฐานโดยเฉพาะมีความคล้ายคลึงกับฟังก์ชันการเทียบคะแนนอีควิเปอร์เซ็นต์ไทล์มากกว่าฟังก์ชันการเทียบคะแนนเชิงเส้นตรง และฟังก์ชันการเทียบคะแนนสังเกตวิธีลักษณะแบบทดสอบมีความแตกต่างกับฟังก์ชันการเทียบคะแนนเชิงเส้นตรงมากที่สุด

การเทียบคะแนนจริงทุกวิธีมีฟังก์ชันการเทียบคะแนนที่คล้ายคลึงกับฟังก์ชันการเทียบคะแนนพื้นฐานมากกว่าการเทียบคะแนนสังเกตยกเว้นการเทียบคะแนนสังเกตวิธีปรับค่าพารามิเตอร์พร้อมกับฟังก์ชันการเทียบคะแนนที่คล้ายคลึงกับการเทียบคะแนนอีควิเปอร์เซ็นต์ไทล์

การประเมินคุณภาพของการเทียบคะแนนโดยใช้ค่าดัชนีความแตกต่างแต่ละวิธี คือ การเทียบคะแนนเชิงเส้นตรงอีควิเปอร์เซ็นต์ไทล์ การเทียบคะแนนจริง และคะแนนสังเกตวิธีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยาก การเทียบคะแนนจริงและคะแนนสังเกต วิธีโค้ง

ลักษณะแบบทดสอบและการเทียบคะแนนจริงและคะแนนสังเกตวิธีปรับค่าพารามิเตอร์พร้อมกันมีค่าเท่ากับ 0.192 0.179 0.181 0.174 0.182 0.173 0.184 และ 0.173 ตามลำดับ ซึ่งวิธีการเทียบคะแนนเชิงเส้นตรงให้ความคลาดเคลื่อนรวมสูงสุด วิธีการปรับค่าพารามิเตอร์พร้อมกันและวิธีโค้งลักษณะแบบทดสอบเมื่อเทียบคะแนนสังเกตให้ความคลาดเคลื่อนรวมต่ำสุดและเมื่อประเมินด้วยเกณฑ์ของปีเตอร์เซ็นและคณะ พบว่า ผลการเทียบคะแนนทุกวิธี อยู่ในระดับที่น่าพอใจอย่างมาก

การเปรียบเทียบความแตกต่างสัมบูรณ์ระหว่างคะแนนแปลงจากการเทียบคะแนนกับคะแนนเกณฑ์ในกลุ่มสอบทานผล ไม่แตกต่างกัน

ฉวีวรรณ บุญมั่ง (2540) ได้ศึกษาการเปรียบเทียบความสัมพันธ์ระหว่างคะแนนความสามารถจากการวิเคราะห์โดยทฤษฎีการตอบสนองข้อสอบกับคะแนนที่ได้จากการตอบและตรวจให้คะแนน 4 วิธี กลุ่มตัวอย่างเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 โรงเรียนสังกัดสำนักงานการประถมศึกษาอำเภอบรบือ จังหวัดมหาสารคาม ภาคเรียนที่ 1 ปีการศึกษา 2540 จำนวน 496 คน พบว่า คะแนนความสามารถมีความสัมพันธ์กับคะแนนที่ได้จากวิธี 0 – 1 สูงสุดเท่ากับ 0.511 สัมพันธ์กับคะแนนที่ได้จากวิธีของคูมบ์ ต่ำสุดเท่ากับ 0.3896 และสัมพันธ์กับคะแนนที่ได้จากวิธีของกิบบอนส์และคณะ และวิธีของอนันต์ ศรีโสภา ใกล้เคียงกันค่าสัมประสิทธิ์สหสัมพันธ์ทุกค่าเป็นบวก และมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 เมื่อเปรียบเทียบความแตกต่างของค่าสัมประสิทธิ์สหสัมพันธ์พบว่าไม่แตกต่างกัน

ผลการเปรียบเทียบค่าจากการวิเคราะห์โดยทฤษฎีการตอบสนองข้อสอบกับสัดส่วนการเดาที่ได้จากการตอบและตรวจให้คะแนน 4 วิธี พบว่า ค่าการเดาแตกต่างกับสัดส่วนการเดาที่ได้จากวิธีของ 0 – 1 วิธีของคูมบ์ วิธีของกิบบอนส์และคณะ และวิธีของอนันต์ ศรีโสภา และสัดส่วนการเดาที่ได้จากวิธีของคูมบ์แตกต่างกับวิธีของ 0 – 1 และวิธีของกิบบอนส์และคณะอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนสัดส่วนการเดาที่ได้จากการตอบและตรวจให้คะแนนผู้อื่น ๆ ไม่แตกต่างกัน

ผลการเปรียบเทียบค่าการเดาจากการวิเคราะห์โดยทฤษฎีการตอบสนองข้อสอบกับสัดส่วนการเดาที่ได้จากการตอบและตรวจให้คะแนนวิธีของคูมบ์ วิธีของกิบบอนส์และคณะและวิธีของอนันต์ ศรีโสภา พบว่า ค่าการเดาแตกต่างกับสัดส่วนการเดาที่ได้จากวิธีของคูมบ์ และวิธีของกิบบอนส์และคณะ สัดส่วนการเดาที่ได้จากวิธีของคูมบ์แตกต่างกับวิธีของกิบบอนส์และคณะและวิธีของอนันต์ ศรีโสภา อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนค่าการเดากับสัดส่วนการเดาที่ได้จากวิธีของอนันต์ ศรีโสภา

พรทิพย์ ไชยโส (2534) ได้พัฒนาสูตรการให้คะแนนแบบทดสอบเลือกตอบสำหรับการให้คะแนนความรู้บางส่วนของผู้ตอบ การประยุกต์ใช้วิธีการของอาโนลด์และวิธีการของแฮมดาน

กลุ่มตัวอย่างที่ใช้เป็นนักเรียนชั้น ม.1 ม.2 และ ม.3 จากโรงเรียนสาธิตมหาวิทยาลัยเกษตรศาสตร์ ภาควิชาที่ 1 ปีการศึกษา 2532 จำนวน 731 คน นักเรียนชาย 391 คน นักเรียนหญิง 340 คน

ผลการวิจัยด้านความเที่ยงตรงเชิงทฤษฎี พบว่า สูตรการให้คะแนนที่ประยุกต์จากวิธีของแฮมมิดาน สูตรการให้คะแนนของแฮมมิดานและสูตรการให้คะแนนแบบประเพณีนิยม มีความเที่ยงตรงเชิงทฤษฎีไม่แตกต่างกัน แต่เมื่อเทียบกับสูตรการให้คะแนนที่ประยุกต์จากวิธีของอาโนลด์และสูตรการให้คะแนนของอาโนลด์ มีความเที่ยงตรงเชิงทฤษฎีสูงกว่า สูตรการให้คะแนนที่ประยุกต์จากวิธีของอาโนลด์มีความเที่ยงตรงเชิงทฤษฎีต่ำกว่าสูตรการให้คะแนนของอาโนลด์

ด้านความเที่ยงตรงตามเกณฑ์ สูตรการให้คะแนนสูตรต่าง ๆ มีค่าความเที่ยงตรงตามเกณฑ์ไม่แตกต่างกันอย่างมีนัยสำคัญ โดยสูตรการให้คะแนนที่ประยุกต์มาจากวิธีของแฮมมิดานมีความเที่ยงตรงตามเกณฑ์สูงสุด เท่ากับ 0.624 ไม่ว่าเกณฑ์จะให้คะแนนด้วยวิธีแบบประเพณีนิยมหรือให้คะแนนตามความถูกต้องของการใช้เหตุผล ส่วนสูตรการให้คะแนนที่ประยุกต์จากวิธีของอาโนลด์มีความเที่ยงตรงตามเกณฑ์ต่ำสุด เท่ากับ 0.495 ไม่ว่าเกณฑ์จะให้คะแนนด้วยวิธีแบบประเพณีนิยมหรือให้ตามระดับความถูกต้องของเหตุผลที่ใช้ สูตรการให้คะแนนแบบประเพณีนิยมมีค่าเที่ยงตรงตามเกณฑ์ เท่ากับ 0.618 สูตรการให้คะแนนของแฮมมิดาน ค่าเที่ยงตรงตามเกณฑ์ เท่ากับ 0.612 และสูตรการให้คะแนนของอาโนลด์มีค่าเที่ยงตรงตามเกณฑ์เท่ากับ 0.557

การวิเคราะห์ค่าความเชื่อมั่นแบบสอดคล้องภายใน ไม่แตกต่างกัน ดังนี้ สูตรการให้คะแนนที่ประยุกต์จากวิธีของอาโนลด์ให้คะแนนมีค่าความเชื่อมั่นสูงสุด เท่ากับ 0.879 สูตรการให้คะแนนของอาโนลด์ เท่ากับ 0.814 สูตรการให้คะแนนที่ประยุกต์จากวิธีของแฮมมิดาน เท่ากับ 0.788 สูตรการให้คะแนนแบบประเพณีนิยม เท่ากับ 0.787 และสูตรการให้คะแนนของแฮมมิดาน เท่ากับ 0.781 ตามลำดับ เมื่อเปรียบเทียบความเชื่อมั่นของสูตรการให้คะแนนในด้านประสิทธิภาพด้านความยาวของแบบทดสอบ พบว่า สูตรการให้คะแนนที่ประยุกต์จากวิธีของอาโนลด์จะมีประสิทธิภาพด้านความยาวสูงกว่าสูตรการให้คะแนนที่ประยุกต์จากวิธีของแฮมมิดาน สูตรการให้คะแนนของแฮมมิดาน สูตรการให้คะแนนของอาโนลด์ และสูตรการให้คะแนนแบบประเพณีนิยม ตามลำดับ

จากการศึกษา พบว่า ไม่สามารถให้ข้อสรุปได้ชัดเจนถึงคุณภาพของสูตรการให้คะแนนที่พัฒนาขึ้น ถึงแม้ว่าคุณภาพเหนือกว่าสูตรการให้คะแนนอื่นทั้ง 3 เกณฑ์ คือความตรงเชิงทฤษฎี ความเที่ยงตรงตามเกณฑ์และความเชื่อมั่นแบบสอดคล้องภายใน เนื่องจากสูตรการให้คะแนนแบบประเพณีนิยมมีความเที่ยงตรงเชิงทฤษฎี ความเที่ยงตรงตามเกณฑ์และความเชื่อมั่นแบบสอดคล้องภายในไม่ยิ่งหย่อนกว่าสูตรการให้คะแนนอื่น ๆ มากนัก กลับมีความเที่ยงตรงเชิงทฤษฎีและความเที่ยงตรงตามเกณฑ์ไม่แตกต่างจากสูตรการให้คะแนนที่ประยุกต์จากวิธีของแฮมมิดาน

พินิจ อุไรรักษ์ (2533) ศึกษาผลของวิธีการให้คะแนนที่มีต่อคะแนนสอบ ค่าความเชื่อมั่น และค่าความเที่ยงตรงจากวิธีการให้คะแนน 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. ค่าความเชื่อมั่นของแบบทดสอบพบว่าวิธีการให้น้ำหนักคะแนนรายข้อเท่ากัน วิธีการให้น้ำหนักคะแนนรายข้อต่างกันตามระดับความมั่นใจในการตอบ วิธีการให้น้ำหนักคะแนนรายข้อต่างกันตามค่าพารามิเตอร์ตามสูตรของลอร์ด และวิธีการให้คะแนนรายข้อต่างกันตามค่าความสามารถของผู้สอบ มีค่าเท่ากับ .075
3. วิธีการให้น้ำหนักคะแนนรายข้อต่างกันตามระดับความมั่นใจในการตอบวิธีการให้น้ำหนักคะแนนรายข้อต่างกันตามค่าพารามิเตอร์ตามสูตรของลอร์ด และวิธีการให้คะแนนรายข้อต่างกันตามค่าความสามารถของผู้สอบ สูงกว่าวิธีการให้น้ำหนักคะแนนรายข้อเท่ากัน

กาญจนา ศิริวัฒนาพงษ์ (2520) ได้ศึกษาเปรียบเทียบวิธีการตอบและตรวจให้คะแนนแบบทดสอบปรนัยเลือกตอบที่มีลักษณะต่างกัน คือแบบทดสอบที่มีตัวเลือกถูกตัวเดียว แบบตัวเลือกถูกที่สุดเพียงตัวเดียวและแบบตัวเลือกถูกหลายตัว ที่มีลักษณะต่างกัน โดยใช้วิธีการตรวจให้คะแนนต่างกัน 3 วิธี คือ วิธี 0 - 1 วิธีของคัมป์ส วิธีของอนันต์ โดยใช้แบบทดสอบวัดผลสัมฤทธิ์วิทยาศาสตร์ กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 โรงเรียนจังหวัดปทุมธานี 810 คน ผลการศึกษาพบว่า

1. เมื่อตอบและตรวจด้วยวิธี 0 - 1 และวิธีของคัมป์ส ค่าความเชื่อมั่น ค่าความเที่ยงตรงและค่าอำนาจจำแนกไม่แตกต่างกัน แต่เมื่อตอบด้วยวิธีอนันต์ ค่าความเชื่อมั่นของแบบทดสอบที่มีตัวเลือกถูกเพียงตัวเดียวสูงกว่าตัวเลือกถูกที่สุดและตัวเลือกถูกหลายตัว ค่าความเที่ยงตรงและค่าอำนาจจำแนกไม่แตกต่างกัน
2. แบบทดสอบที่มีตัวเลือกถูกเพียงตัวเดียวและตัวเลือกถูกที่สุดเพียงตัวเดียว เมื่อตอบและตรวจให้คะแนนโดยวิธีของอนันต์ ให้ค่าความเชื่อมั่นและค่าความเที่ยงตรงสูงกว่าวิธีการตรวจให้คะแนนของคัมป์สแต่ค่าความเที่ยงตรงและค่าอำนาจจำแนกไม่แตกต่างกัน ส่วนแบบทดสอบที่มีตัวเลือกถูกหลายตัวค่าความเที่ยงตรงและค่าอำนาจจำแนกไม่แตกต่างกัน
3. เมื่อตรวจให้คะแนนทั้ง 3 วิธี แบบทดสอบปรนัยเลือกตอบที่มีตัวเลือกถูกเพียงตัวเดียวและตัวเลือกถูกหลายตัวมีการเดาน้อยกว่าตัวเลือกถูกที่สุดเพียงตัวเดียว ส่วนแบบทดสอบที่มี

ตัวเลือกถูกเพียงตัวเดียวมีค่าการเดาน้อยกว่าตัวเลือกถูกหลายตัว และเมื่อตรวจให้คะแนนโดยวิธีของอนันต์และของคูมบัสไม่พบว่าแตกต่างกัน

4. แบบทดสอบปรนัยเลือกตอบทั้ง 3 ลักษณะเมื่อตอบและตรวจให้คะแนนด้วยวิธีของอนันต์ มีการเดาน้อยกว่าเมื่อตอบและตรวจให้คะแนนด้วย วิธี 0 – 1 และวิธีของคูมบัส แต่ไม่พบความแตกต่างในแบบทดสอบที่มีตัวเลือกถูกที่สุดเพียงตัวเดียว

ไซมอน บุคชี และนีโว (Simon, Budescu and Nevo. 1997) ได้เปรียบเทียบวิธีการตอบสนองข้อสอบ เพื่อให้คะแนนความรู้บางส่วน 6 วิธีกับวิธีเลือกตอบแบบประเพณีนิยม เก็บข้อมูลกับนักเรียน 2,989 คนที่สอบคัดเลือกเข้ามหาวิทยาลัยในอิสราเอล ระหว่างเดือนเมษายน 1985 – กรกฎาคม 1986 โดยใช้แบบทดสอบ 4 วิชา คือ General Knowledge 30 ข้อ วิชา General Reasoning 22 ข้อ วิชา Figural Reasoning 18 ข้อ และวิชา Mathematical Reasoning 21 ข้อ แต่ละวิชามีวิธีการตอบ 7 วิธี ได้แก่ 1. วิธีแก้การเดา (Correction for guessing, GC) 2. วิธีการตัดตัวลวง (Elimination Scoring, ET) 3. วิธีการทดสอบความน่าจะเป็น (Probability Testing, PT) 4. วิธีการแสดงความมั่นใจ (Confidence Marking, CM) 5. วิธีการจัดอันดับอย่างสมบูรณ์ (Complete Ordering, CO) 6. วิธีการจัดอันดับเป็นบางส่วน (Partial Ordering, PO) และ 7. วิธีประเพณีนิยม

อัลเบิร์ต (Albert. 1970) ได้ศึกษาวิธีการตรวจให้คะแนนความรู้บางส่วนวิชาคณิตศาสตร์ โดยใช้โจทย์ปัญหาคณิตศาสตร์ จำนวน 25 ข้อ กับกลุ่มตัวอย่างที่เป็นนักเรียนเกรด 11 จำนวน 2 กลุ่ม กลุ่มละ 146 คน มีวิธีการตรวจให้คะแนน 2 ลักษณะ คือ

1. แบบทดสอบเลือกตอบที่มีตัวเลือกถูกเพียงตัวเดียว มีการให้คะแนน 3 วิธี คือ วิธี 0 – 1 วิธีกำหนดน้ำหนักตัวเลือกโดยดูจากโอกาสผิดพลาดมากหรือน้อย และวิธีกำหนดน้ำหนักตัวเลือกโดยครูเป็นผู้กำหนด

2. แบบขยายคำตอบ เป็นแบบที่มีการแก้ปัญหาหลายขั้นตอน มีการตรวจให้คะแนน 2 วิธี คือ วิธีการให้คะแนนขั้นละ 1 คะแนนในการตอบที่ถูก และวิธีให้คะแนนขั้นละ 1 คะแนนถ้านักเรียนเริ่มต้นจากข้อมูลที่ถูกต้องถึงคำตอบที่ถูกต้อง

ผลการศึกษาพบว่า

1. แบบขยายคำตอบสามารถพิสูจน์ความรู้บางส่วนได้
2. แบบขยายคำตอบให้ค่าอำนาจจำแนกสูงกว่าแบบธรรมดา
3. แบบขยายคำตอบให้ค่าความเชื่อมั่นสูงกว่าแบบธรรมดา
4. แบบขยายคำตอบให้ค่าความเที่ยงตรงตามสภาพสูงกว่าแบบธรรมดาและสามารถหาความคลาดเคลื่อนที่เกิดจากการเดาได้

จากการศึกษาจะเห็นว่า แบบขยายคำตอบซึ่งเป็นตรวจให้คะแนนความรู้บางส่วนของนักเรียนในการตอบจะทำให้ค่าอำนาจจำแนกและค่าความเชื่อมั่นสูง เปอร์เซนต์การเดาลดลงและยังให้ค่าความเที่ยงตรงตามสภาพสูงด้วย

เดรสเซลและชมิท (Dressel and Schmidt. 1953) ได้ศึกษาวิธีการตอบและวิธีการตรวจให้คะแนนที่ต่างกันเพื่อปรับปรุงอำนาจจำแนกของแบบทดสอบเลือกตอบ โดยมีวิธีการตอบ 4 วิธี คือ

1. แบบเลือกตอบเสรี เป็นแบบทดสอบเลือกตอบที่มี 5 ตัวเลือก ผู้สอบจะเลือกก็ได้ตัวเลือกก็ได้
2. แบบบอกระดับความมั่นใจในการตอบ เป็นแบบที่ผู้เข้าสอบจะต้องแสดงระดับความมั่นใจในการตอบ ซึ่งมี 4 ระดับ แต่คำถามแต่ละข้อให้ตอบเพียงคำตอบเดียว
3. แบบคำตอบถูกมากกว่าหนึ่ง แบบทดสอบชนิดนี้จะมีคำตอบถูกอยู่หลายคำตอบผู้เข้าสอบจะต้องทำเครื่องหมายทุกข้อที่ถูก
4. แบบมีคำตอบถูกสองคำตอบ ผู้เข้าสอบจะเลือกตอบได้เพียงสองคำตอบเท่านั้น

ผลการศึกษาพบว่า การข้อสอบของนักเรียนเก่ง (พิจารณาจากคะแนนสอบ) จะแตกต่างกับนักเรียนปานกลางและอ่อน อย่างมีนัยสำคัญทางสถิติ เมื่อใช้แบบทดสอบเลือกตอบชนิดตอบเสรีคือกลุ่มเก่งจะทำเครื่องหมายน้อยกว่ากลุ่มอ่อนทั้งในข้อสอบที่มีระดับความยากสูง ปานกลางและต่ำในทางตรงกันข้ามแบบทดสอบชนิดทดสอบความมั่นใจไม่ได้จำแนกนักเรียนเก่ง ปานกลาง อ่อนได้ดีกว่าวิธีแรก เพราะเมื่อพิจารณาการตอบข้อสอบแล้วพบว่า ผู้เข้าสอบตอบข้อสอบที่มีความยากปานกลาง และยากมากด้วยระดับความมั่นใจเดียวกัน เดรสเซลและชมิท ได้สรุปว่า องค์ประกอบของความมั่นใจที่วัดโดยการให้คำตอบแบบเสรีนั้นแตกต่างกับที่วัดโดยผู้เข้าสอบบอกระดับความมั่นใจ

5.2 งานวิจัยที่เกี่ยวกับการหาคุณภาพข้อสอบ

ปรมินทร์ อริเดช (2547) ได้ศึกษาการเปรียบเทียบฟังก์ชันสารสนเทศในการใช้โมเดลโลจิสติก จีอาร์เอ็มและจีพีซีเอ็มของมาตรวัดเจตคติแบบลิเคิร์ตและมาตรวัดแบบตัวเลือกบังคับตอบที่มีวิธีการให้คะแนนแบบสองค่าและแบบหลายค่า กลุ่มตัวอย่างเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 จังหวัดเชียงราย ผลการวิจัยพบว่า ในมาตรวัดเจตคติแบบลิเคิร์ตการตรวจให้คะแนนแบบสองค่าที่วิเคราะห์ด้วยโมเดลโลจิสติกให้ค่าเฉลี่ยฟังก์ชันสารสนเทศสูงกว่าการตรวจให้คะแนนแบบหลายค่าที่วิเคราะห์ตาม GRM GCPM และ GPCM ที่มีการยุบรวมลำดับชั้นคะแนนใหม่ แต่ถ้าพิจารณาโค้งฟังก์ชันสารสนเทศจะพบว่า การวิเคราะห์ด้วย GPCM ที่มีการลำดับชั้นคะแนนใหม่ จะให้ค่าฟังก์ชันสูงสุดและพิจารณาจากมาตรวัดแบบลิเคิร์ตแบบซึ่งมีค่าตรงกลางและไม่มีค่าตรงกลาง การตรวจให้คะแนนแบบไม่มีค่าตรงกลางจะให้ค่าเฉลี่ยฟังก์ชันสารสนเทศสูงกว่าการตรวจให้คะแนนแบบมีค่าตรงกลาง ในมาตรวัดความมีระเบียบวินัยแบบตัวเลือกบังคับตอบ การให้คะแนนแบบหลายค่า

ที่วิเคราะห์ด้วย GRM ให้ค่าเฉลี่ยฟังก์ชันสารสนเทศสูงสุด รองลงมา คือ GPCM ที่มีการยุบรวมลำดับขั้นคะแนนใหม่ GPCM ซึ่งสูงกว่าการให้คะแนนแบบสองค่าที่วิเคราะห์ด้วยโมเดลโลจิสติก 3 , 2 และ 1 พารามิเตอร์ ตามลำดับ เมื่อพิจารณามาตรวัดความมีระเบียบวินัยแบบตัวเลือกบังคับตอบที่มีการให้คะแนนแบบสองค่า ที่มีวิธีกำหนดน้ำหนักคะแนนแตกต่างกัน 3 แบบ พบว่า การวิเคราะห์ด้วยโมเดลโลจิสติก ด้วยวิธีการให้คะแนนแบบที่ 3 ให้คะแนนเฉลี่ยฟังก์ชันสารสนเทศสูงสุด รองลงมา คือ 2 และ 1 พารามิเตอร์

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

ผลการวิจัยพบว่า

1. ผลการเปรียบเทียบความตรงตามสภาพ ระหว่างวิธีการตรวจให้คะแนนที่แตกต่างกันจากการศึกษาความตรงตามสภาพตามวิธีประยุกต์การให้คะแนนของคูมบี้ วิธีประยุกต์การให้คะแนนของเดรสเซลและสมิทและวิธีประเพณีนิยม แล้วนำมาเปรียบเทียบความแตกต่างโดยใช้คะแนนฟิชเชอร์ซี เปรียบเทียบในภาพรวมโดยใช้สถิติทดสอบไคร์สแควร์ (พบว่าค่าความตรงตามสภาพไม่ขึ้นอยู่กับวิธีการตรวจให้คะแนนในแต่ละวิธี และเมื่อนำมาเปรียบเทียบความแตกต่างกันเป็นรายคู่ ด้วยสถิติทดสอบซี (Z) พบว่า ไม่มีความแตกต่างกันของความตรงสภาพในแต่ละคู่อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05

2. ผลการเปรียบเทียบความยากและอำนาจจำแนกระหว่างวิธีการตรวจให้คะแนนที่แตกต่างกันผลการเปรียบเทียบความแตกต่างของค่าเฉลี่ยของค่าอำนาจจำแนกและค่าเฉลี่ยของค่าความยาก ด้วยสถิติ F พบว่า ค่าเฉลี่ยของค่าอำนาจจำแนกตามวิธีการตรวจให้คะแนนมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ($F = 85.369$, $p\text{-value} < .000$) และเมื่อนำมาเปรียบเทียบความแตกต่างรายคู่ด้วยวิธีของ Scheffe พบว่าวิธีประยุกต์การให้คะแนนของคูมบี้ มีค่าอำนาจจำแนกเฉลี่ยสูงกว่าวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิทอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ในขณะที่เดียวกันวิธีประเพณีนิยมมีค่าอำนาจจำแนกเฉลี่ยสูงกว่าวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิทอย่างมีนัยสำคัญทางสถิติ เช่นเดียวกัน ส่วนค่าเฉลี่ยของความยากตามวิธีการตรวจให้คะแนนมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ($F = 20.592$, $p\text{-value} < .000$) และเมื่อนำมาเปรียบเทียบความแตกต่างรายคู่ด้วยวิธีของ Scheffe พบว่าวิธีประยุกต์การให้คะแนนของคูมบี้ มีค่าความยากเฉลี่ยต่ำกว่าวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท

และวิธีประเพณีนิยมอย่างมีนัยสำคัญทางสถิติ ในขณะที่เดียวกันวิธีประเพณีนิยมมีค่าความยากเฉลี่ยไม่แตกต่างจากวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิทอย่างมีนัยสำคัญทางสถิติ

3. ผลการเปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบสอบและอัตราส่วนสารสนเทศ ระหว่างวิธีการตรวจให้คะแนนที่แตกต่างกัน

วิธีประยุกต์การให้คะแนนของคัมป์มีค่าฟังก์ชันสารสนเทศของข้อสอบโดยเฉลี่ยสูงกว่าวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิทและวิธีประเพณีนิยมอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ส่วนวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท มีค่าฟังก์ชันสารสนเทศของข้อสอบโดยเฉลี่ยสูงกว่าวิธีประเพณีนิยมอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 เมื่อเปรียบเทียบค่าฟังก์ชันสารสนเทศของข้อสอบจำแนกตามค่าความสามารถของผู้สอบตามวิธีการตรวจให้คะแนน พบว่า วิธีประยุกต์การให้คะแนนของคัมป์ ให้ค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงที่สุดในระดับความสามารถ - 1 และมีค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยเท่ากับ 0.673 ซึ่งสูงกว่าวิธีอื่นๆ ในขณะที่วิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท ให้ค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงที่สุดในระดับความสามารถ 0 มีค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยเท่ากับ 0.395 ส่วนวิธีประเพณีนิยมให้ค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงที่สุดในระดับความสามารถ 0 มีค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยเท่ากับ 0.187

วิธีประยุกต์การให้คะแนนของคัมป์ ให้ค่าฟังก์ชันสารสนเทศของแบบสอบ สูงกว่าวิธีอื่นๆ รองลงมาคือ วิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท ส่วนวิธีประเพณีนิยมให้ค่าฟังก์ชันสารสนเทศของแบบสอบต่ำที่สุด ส่วนอัตราส่วนสารสนเทศเฉลี่ยระหว่างวิธีประยุกต์การให้คะแนนของคัมป์กับวิธีประเพณีนิยม มีค่าสูงสุด รองลงมาคืออัตราส่วนสารสนเทศเฉลี่ยระหว่างวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิทกับวิธีประเพณีนิยม ซึ่งแสดงให้เห็นว่า วิธีประยุกต์การให้คะแนนของคัมป์มีประสิทธิภาพสัมพัทธ์สูงที่สุด รองลงมา คือ วิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท ส่วนวิธีประเพณีนิยมมีประสิทธิภาพสัมพัทธ์ต่ำที่สุด

สุรัชย์ มีชาญ (2539) เปรียบเทียบความยากประจำชั้นและประสิทธิภาพในการประมาณค่าเจตคติของมาตรวัดเจตคติแบบลิเคิร์ทที่มีรูปแบบของการตอบ จำนวนของลำดับขั้นและทิศทางของการเรียงลำดับแตกต่างกันด้วยพาเซี่ยลเครดิทโมเดลของ มาสเตอร์ กลุ่มตัวอย่างที่ใช้เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2538 ของโรงเรียนสังกัดกรมสามัญศึกษาจังหวัดลพบุรีและสิงห์บุรี จำนวน 3,226 คน

ผลการวิจัยพบว่า มาตรวัดเจตคติมีจำนวนของลำดับขั้นเพียง 4 ระดับ มีแนวโน้มว่าหากมาตรวัดนั้นใช้รูปแบบของการตอบและทิศทางของการเรียงลำดับขั้นแตกต่างกันค่าเฉลี่ยความยากประจำชั้นที่ 1 และชั้นที่ 3 ของมาตรวัด จะมีความแตกต่างกัน แต่ค่าเฉลี่ยความยากประจำชั้นที่ 2 จะ

ไม่แตกต่างกัน และเมื่อมาตรวัดเหล่านั้นมีจำนวนของลำดับชั้นเพิ่มขึ้นเป็น 5 ระดับ ค่าเฉลี่ยความยากประจำชั้นของทุกชั้นในมาตรวัดเหล่านั้นจะไม่แตกต่างกัน ในการศึกษาความสัมพันธ์ของค่าความยากประจำชั้นของมาตรวัดเจตคติแต่ละฉบับ พบว่า จำนวนลำดับชั้น 4 ระดับ มาตรวัดฉบับที่ 1, 3 และ 4 มีค่าความยากประจำชั้นที่ 1 มีความสัมพันธ์กันทางบวกทุกคู่ ในกรณีมาตรวัดมีจำนวนของลำดับชั้น 5 ระดับ ฉบับที่ 3 กับ ฉบับที่ 7 และฉบับที่ 4 กับฉบับที่ 7 มีค่าความยากประจำชั้นที่ 1 ไม่สัมพันธ์กัน

การเปรียบเทียบค่าฟังก์ชันอินฟอร์เมชันของมาตรวัด แต่ละฉบับพบว่า มาตรวัดที่มีจำนวนลำดับชั้น 5 ระดับ มีค่าอินฟอร์เมชันสูงกว่ามาตรวัดที่มีจำนวนลำดับชั้น 4 ระดับ ทุกคู่ของการเปรียบเทียบตลอดช่วงเจตคติ มาตรวัดที่มีจำนวนลำดับชั้นเท่ากันมีค่าเฉลี่ยของ $RMSE_n$ ใกล้เคียงกัน โดยมาตรวัดที่มีจำนวนลำดับ 4 ระดับ มีค่าเฉลี่ย $RMSE_n$ ไม่แตกต่างกันทุกคู่ยกเว้น ฉบับที่ 1 และฉบับที่ 5 ขณะที่มาตรวัดที่มีจำนวนลำดับชั้น 5 ระดับ มีค่าเฉลี่ย $RMSE_n$ ไม่แตกต่างในทุกคู่ และมีความสูงกว่ามาตรวัดที่มีจำนวนลำดับชั้น 4 ระดับ

การเปรียบเทียบค่าสัดส่วนของผู้ตอบที่มีความเหมาะสมกับโมเดลระหว่างกลุ่มผู้ตอบมาตรวัดเจตคติต่างฉบับกัน พบว่า ค่าสัดส่วนของผู้ตอบที่เหมาะสมกับโมเดลสูงสุด คือ ฉบับที่ 3 รองลงมา คือฉบับที่ 7 ฉบับที่ 4 ฉบับที่ 8 ฉบับที่ 1 ฉบับที่ 2 ฉบับที่ 6 และฉบับที่ 5 ตามลำดับ

ธนวัฒน์ แสนสุข (2538) ได้ศึกษาการใช้ จีอาร์เอ็ม จีพีซีเอ็มและโมเดลโลจิสติกในการเปรียบเทียบฟังก์ชันสารสนเทศของแบบวัดที่มีวิธีการให้คะแนนต่างกัน โดยที่แบบวัดมีการตรวจให้คะแนนแบบทวิภาค (0,1) และแบบพหุภาค (1,2,3,4) ทำการตรวจสอบความสมบูรณ์ของข้อมูลโดยวิเคราะห์ ด้วยโปรแกรม BIGSTEPS และวิเคราะห์หาค่าฟังก์ชันสารสนเทศด้วยโปรแกรม MULTILOG และ PARSCALE ผลการวิจัย พบว่า การตรวจให้คะแนนแบบพหุภาคเมื่อวิเคราะห์ตาม GRM ให้ค่าฟังก์ชันสารสนเทศสูงกว่าการให้คะแนนแบบทวิภาค สำหรับการตรวจให้คะแนนแลลพหุภาควิเคราะห์ตาม GPCM และแบบทวิภาคยังไม่สามารถสรุปได้ว่าวิธีใดให้ค่าฟังก์ชันสารสนเทศสูงกว่ากัน

องอาจ นัยพัฒน์ (2535) ได้ศึกษาประสิทธิภาพของผลการประมาณค่าพารามิเตอร์ของพาเชียล เครดิต โมเดล (CREDIT) ระหว่างวิธีพรีออกซ์ (PROX) และวิธียูคอน (UCON) ในแบบทดสอบวัดเจตคติแบบลิเคิร์ทสเกล จำนวน 50 รายการ กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ปีการศึกษา 2534 เขตพระโขนง กรุงเทพมหานคร จำนวน 500 คน ผลการวิจัย พบว่า ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยที่เป็นผลมาจากการประมาณค่าเจตคติของผู้ตอบด้วยวิธียูคอนมีค่าต่ำกว่าวิธีพรีออกซ์ อย่างมีนัยสำคัญทางสถิติที่ระดับ.01 ส่วนค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยที่เป็นผลมาจากการประมาณค่าความยากประจำชั้นของข้อความ

ระหว่างวิธีทั้งสองมีค่าไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ค่าความเที่ยงตรงเชิงสภาพของการประมาณค่าพารามิเตอร์เจตคติของผู้ตอบที่ประมาณค่าด้วยวิธีพรีออกซ์มีค่าสูงกว่าวิธียูคอนอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบที่ประมาณค่าพารามิเตอร์ด้วยวิธีพรีออกซ์สูงกว่าวิธียูคอนในช่วงของเจตคติระดับปานกลาง ส่วนในช่วงเจตคติระดับต่ำและสูงพบว่าวิธียูคอนสูงกว่า ด้านความสัมพันธ์ระหว่างเจตคติที่เป็นผลมาจากการประมาณค่าพารามิเตอร์ด้วยวิธีพรีออกซ์และวิธียูคอน มีความสัมพันธ์ในทางบวกซึ่งกันและกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ทำนองเดียวกับความยากประจำชั้นของข้อความที่ประมาณค่าด้วยวิธีทั้งสอง ก็พบว่าความสัมพันธ์ในทางบวกในระดับสูงมาก

อนันต์ ศรีโสภา (2532) ได้ศึกษาเปรียบเทียบประสิทธิภาพของแบบทดสอบที่ใช้วิธีการตอบและการตรวจให้คะแนนแบบใหม่ แบบของคัมบัส และแบบ 0 – 1 โดยใช้แบบทดสอบวิชาวิจัยเบื้องต้น กลุ่มตัวอย่างที่ใช้เป็นนักศึกษาวิทยาลัยพยาบาลกรุงเทพ ชั้นปีที่ 4 จำนวน 144 คน แบบออกเป็น 3 กลุ่ม กลุ่มละ 48 คน แต่ละกลุ่มให้ตอบและตรวจให้คะแนนด้วยวิธีของอนันต์ แบบของคัมบัส และแบบ 0 – 1 ตามลำดับ ผลการศึกษาพบว่า วิธีการตอบและตรวจให้คะแนนแบบใหม่มีค่าการเดาน้อยกว่าวิธี 0 – 1 อย่างมีนัยสำคัญเมื่อเปรียบเทียบกับวิธีของคัมบัส จำนวนข้อถูกและข้อผิดของวิธีการตรวจให้คะแนนแบบใหม่แตกต่างจากวิธี 0 – 1 อย่างมีนัยสำคัญแต่ไม่แตกต่างจากวิธีคัมบัส การค้นพบนี้สรุปได้ว่า คะแนนการเดา คะแนนข้อถูกคะแนนข้อผิด และคะแนนถูกบางส่วนของวิธีการใหม่ไม่แตกต่างจากวิธีของคัมบัส แต่แตกต่างจากวิธีการตรวจให้คะแนนแบบ 0 -1 อย่างมีนัยสำคัญ ค่าความเชื่อมั่นของการตรวจให้คะแนนแบบใหม่สูงกว่าวิธีของคัมบัสอย่างมีนัยสำคัญ ส่วนค่าความเที่ยงตรงของแบบทดสอบที่มีตรวจให้คะแนนแบบใหม่ วิธีของคัมบัส ไม่แตกต่างกันอย่างมีนัยสำคัญ วิธีการตรวจให้คะแนนแบบใหม่สามารถปรับปรุงให้การสอบโดยใช้แบบทดสอบเลือกตอบมีประสิทธิภาพสูงขึ้นและมีคุณค่าแก่การศึกษา

สมิธ (Smith.1987) ได้ศึกษาเปรียบเทียบความตรงสภาพและความเชื่อมั่นโดยใช้สัมประสิทธิ์แอลฟาระหว่างแบบทดสอบที่มีการให้คะแนนแบบทวิภาค (Dichotomous, 0 -1) และแบบทดสอบที่มีการให้คะแนนแบบพหุภาค (Polytomous) โดยเก็บข้อมูลจากนักศึกษาที่มาสอบความ ถนัดทางภาษา จำนวน 1,399 คน ให้ทำแบบทดสอบทางภาษาที่ใช้โดยทั่วไปเพื่อนำไปเป็นเกณฑ์ ในการหาความตรงตามสภาพกับแบบทดสอบภาษาที่ใช้ในการศึกษารั้งนี้ มี 2 ชุด ชุดแรกมีการให้คะแนนแบบทวิภาค (Dichotomous) และชุดที่ 2 ให้คะแนนแบบพหุภาค (Polytomous) โดยที่แต่ละตัวเลือกมีคะแนนแต่ละข้อไม่เท่ากันโดยมีหลักการให้คะแนนความรู้บางส่วนจากทฤษฎีโครงสร้างความรู้ และมีการแบ่งกลุ่มตัวอย่างออกเป็นสองกลุ่ม กลุ่มที่ 1 จำนวน 400 คน ทดสอบตามปกติ กลุ่มที่ 2 จำนวน 999 คน ทดสอบความตรงข้ามกลุ่ม โดยแบ่งเป็นกลุ่มที่มีความสามารถ

สูงและกลุ่มที่มีความสามารถต่ำ ผลการวิจัยพบว่า การให้คะแนนแบบพหุวิภาค (Polytomous) ค่าความตรงสภาพและความเชื่อมั่นสูงกว่าการให้คะแนนแบบทวิภาค (Dichotomous) อย่างมีนัยสำคัญทางสถิติทั้งกลุ่มที่มีความสามารถสูงและกลุ่มที่มีความสามารถต่ำ

จากเอกสารและงานวิจัยที่เกี่ยวข้องกับการตรวจให้คะแนนของแบบทดสอบยังไม่ทราบแน่ชัดว่าวิธีการตรวจให้คะแนนแบบใดจะทำให้ได้แบบทดสอบที่มีคุณภาพ ประหยัดเวลา และลดความยุ่งยากในการสร้างแบบทดสอบ ผู้วิจัยจึงสนใจที่จะศึกษาคุณภาพของแบบทดสอบเลือกตอบเมื่อตรวจด้วยวิธีการให้คะแนนความรู้บางส่วนว่าวิธีตรวจให้คะแนนวิธีใดให้ค่าความเที่ยงตรงตามสภาพและค่าฟังก์ชันสารสนเทศของแบบทดสอบและข้อสอบสูงที่สุด จากวิธีการตรวจให้คะแนนความรู้บางส่วน 2 ลักษณะ คือ

1. ด้านการให้น้ำหนักแก่ตัวเลือกที่แตกต่างกัน (Differential Option Weighting) ด้วยวิธีการให้คะแนนตรวจให้คะแนนตามความยากประจำขั้นของคะแนนผลการตอบโดยกำหนดความยากประจำขั้นของการตอบ (Step Difficulties)

2. การเปลี่ยนแปลงวิธีการตอบ (Changing the response method) ด้วยวิธีการของการทดสอบความน่าจะเป็น (Probability Testing) เป็นวิธีที่มีความยืดหยุ่นละเปิดโอกาสให้ผู้สอบแสดงความรู้บางส่วนได้มากที่สุด โดยรายงานความน่าจะเป็นในตัวเลือกที่ผู้สอบคิดว่าเป็นคำตอบที่ถูกโดยมีตำแหน่งคะแนนที่เป็นไปได้ 4 ตำแหน่ง คือ 1 – 4 คะแนน เนื่องจากโมเดลดังกล่าวไม่เชื่อมโยงกับข้อตกลงเบื้องต้น การประมาณค่าพารามิเตอร์มีค่าอำนาจจำแนกรายข้อในฟังก์ชันด้วยจึงเหมาะสมในการวัดความสามารถทางสติปัญญา

บทที่ 3

วิธีการดำเนินงานวิจัย

ในการวิจัยครั้งนี้มีขั้นตอนการดำเนินการวิจัย ดังนี้

ตอนที่ 1 ประชากรและกลุ่มตัวอย่าง

ตอนที่ 2 เครื่องมือที่ใช้ในการวิจัย

ตอนที่ 3 การเก็บรวบรวมข้อมูล

ตอนที่ 4 การวิเคราะห์ข้อมูล

ตอนที่ 1 ประชากรและกลุ่มตัวอย่าง

ประชากร

ประชากรที่ใช้ในการวิจัยครั้งนี้ เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ที่กำลังศึกษาในภาคเรียนที่ 2 ปีการศึกษา 2549 สำนักงานเขตพื้นที่การศึกษาปราจีนบุรี จำนวน 97 โรงเรียน จำนวน 199 ห้องเรียน นักเรียน 6,803 คน (สำนักงานเขตพื้นที่การศึกษาปราจีนบุรี. 2549)

กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ เป็นนักเรียนชั้นเรียนชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 2 ปีการศึกษา 2549 สำนักงานเขตพื้นที่การศึกษาปราจีนบุรี จำนวน 1,007 คน ได้มาโดยการสุ่มแบบหลายขั้นตอน (Multi - stage random Sampling) โดยกำหนดขนาดกลุ่มตัวอย่างประชากร จากสูตรการคำนวณของ มยุรี ศรีชัย สุ่มได้นักเรียนจำนวน 593 คน แต่ในการวิเคราะห์ข้อมูลตามทฤษฎีตอบสนองข้อสอบ (Item Response Model) ซึ่งประมาณค่าพารามิเตอร์ด้วยควรจะเป็นสูงสุดขอบ (Marginal Maximum likelihood : MML) ควรใช้กลุ่มตัวอย่างประมาณ 1,000 คน (ธนวัฒน์ แสนสุข. 2538; อ้างอิงจาก Zwinderman and Wollenberg. 1990) และการประมาณค่าพารามิเตอร์โดยใช้โปรแกรม MULTILOG ควรใช้ขนาดกลุ่มตัวอย่างไม่ต่ำกว่า 500 คน ควรกำหนดให้มีขนาดใหญ่พอที่จะทำให้ความคลาดเคลื่อนมาตรฐานของการประมาณค่าพารามิเตอร์มีขนาดเล็กถึงระดับที่ยอมรับได้ตามเป้าหมายของการนำไปใช้ในทางปฏิบัติ (ศรีชัย กาญจนวาสี. 2545: 100)

สูตรการคำนวณ การกำหนดขนาดกลุ่มตัวอย่าง (มยุรี ศรีชัย. 2538: 104)

$$n = \frac{NZ_{\alpha}^2 \sigma^2}{(N-1)E^2 + Z_{\alpha}^2 \sigma^2}$$

โดยที่ n หมายถึง ขนาดกลุ่มตัวอย่างที่พอเพียงที่จะเป็นตัวแทนของประชากร

N หมายถึง ขนาดของประชากร

E หมายถึง ความคลาดเคลื่อนที่ยอมรับได้ในการวิจัยครั้งนี้ยอมให้เกิดได้ 5% ($E = 1$)

σ^2 หมายถึง ความแปรปรวนของประชากร ($\sigma = 13$)

Z หมายถึง ระดับความเชื่อมั่น 95 % ($\alpha = .05, Z_{\alpha/2} = 1.96$)

การสุ่มกลุ่มตัวอย่างมีขั้นตอน ดังนี้

1. สํารวจข้อมูลหน่วยสมาชิกของประชากรชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2549 จากสำนักงานเขตพื้นที่การศึกษาปราจีนบุรี จัดทำกรอบการสุ่ม (Sampling Frame) โดยอาศัยลักษณะการแบ่งเขตการปกครองของจังหวัดปราจีนบุรี มีจำนวน 97 โรงเรียน จำนวน 199 ห้องเรียน จำนวน 6,803 คน ดังตาราง

ตาราง 3 ข้อมูลหน่วยสมาชิกของประชากร ชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2549

ที่	ชื่ออำเภอ	จำนวนทั้งหมด		
		โรงเรียน	ห้องเรียน	นักเรียน
1	เมือง	22	58	2134
2	กบินทร์บุรี	35	58	2031
3	ประจันตคาม	12	21	611
4	บ้านสร้าง	5	8	231
5	ศรีมโหสถ	2	6	224
6	ศรีมหาโพธิ์	11	31	1030
7	นาดี	10	17	542
รวม		97	199	6803

2. สุ่มแบบแบ่งกลุ่ม โดยแบ่งเขตพื้นที่การศึกษาปราจีนบุรีเป็น 7 อำเภอ 97 โรงเรียน สุ่มได้มา 3 อำเภอ ได้แก่ อำเภอเมือง อำเภอประจันตคาม และอำเภอนาดี จำนวน 44 โรงเรียน จำนวน 54 ห้องเรียน จำนวน 3,287 คน

3. สุ่มโรงเรียน โดยวิธีการสุ่มแบบแบ่งชั้น (Stratified Random Sampling) โดยมีขนาดโรงเรียนเป็นชั้น (Strata) มีโรงเรียนเป็นหน่วยสุ่ม (Sampling Unit)

3.1 จำแนกโรงเรียนที่มีอยู่ เป็น 3 ขนาด (ขนาดใหญ่ , ขนาดกลาง และขนาดเล็ก)

จากการสำรวจพบว่า

- (1) ขนาดใหญ่ พิจารณาจากจำนวนนักเรียนทั้งหมดมากกว่า 1,500 คน จำนวน 3 โรงเรียน จำนวนนักเรียน 1,293 คน
- (2) ขนาดกลาง พิจารณาจากจำนวนนักเรียนทั้งหมดตั้งแต่ 500 – 1,499 คน 13 โรงเรียน จำนวนนักเรียน 1,053 คน
- (3) ขนาดเล็ก พิจารณาจากจำนวนนักเรียนทั้งหมดน้อยกว่า 500 คน 28 โรงเรียน จำนวนนักเรียน 941 คน

3.2 สุ่มโรงเรียนตามสัดส่วนของโรงเรียนแต่ละขนาด โดยใช้วิธีการสุ่มอย่างง่ายได้ โรงเรียนขนาดใหญ่ 1 โรงเรียน โรงเรียนขนาดกลาง 4 โรงเรียน โรงเรียนขนาดเล็ก 6 โรงเรียน ดังนี้

ตาราง 4 ข้อมูลการสุ่มจำนวนนักเรียน ของประชากร ชั้นมัธยมศึกษาปีที่ 3

โรงเรียน	จำนวนห้องเรียน	จำนวนนักเรียน
โรงเรียนขนาดใหญ่		
1. โรงเรียนปราจีนกัลยาณี	12	455
โรงเรียนขนาดกลาง		
1. โรงเรียนบ้านแหลมหิน	2	65
2. โรงเรียนไทยรัฐวิทยา	2	111
3. โรงเรียนอนุบาลประจันตคาม	2	63
4. โรงเรียนมณีเสวตอุปถัมภ์	3	120
รวม	9	359
โรงเรียนขนาดเล็ก		
1. โรงเรียนประชารัฐวิทยา	1	39
2. โรงเรียนกองทัพบกอุปถัมภ์	1	44
3. โรงเรียนสุวรรณวิทยา	2	54
4. โรงเรียนวัดบ้านหอย	1	31
5. โรงเรียนบ้านบุสูง	1	25
รวม	6	193
รวมทั้งสิ้น	28	1,007

ตอนที่ 2 การสร้างเครื่องมือที่ใช้ในการวิจัยและขั้นตอนการสร้างเครื่องมือ

เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยครั้งนี้ คือ แบบทดสอบวัดผลสัมฤทธิ์ทางคณิตศาสตร์ สาระทักษะกระบวนการคิดคำนวณ เกี่ยวกับการบวก การลบ การคูณ การหาร จำนวนจริงและเลขยกกำลัง ชั้นมัธยมศึกษาปีที่ 3 ที่ผู้วิจัยสร้างขึ้น มีวิธีการตรวจให้คะแนน 2 วิธี จากแบบทดสอบชุดเดียวกัน มีลักษณะเป็นข้อสอบเลือกตอบ จำนวน 40 ข้อ โดยมีวิธีการให้คะแนน ดังนี้

วิธีการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็น (Probability Testing) หมายถึง การให้คะแนนตัวเลือกทุกตัว ที่พิจารณาน้ำหนักคะแนนจากลำดับขั้นของกระบวนการหาคำตอบ โดยที่ข้อสอบรายข้อมีทักษะการคิดหลายขั้นตอน ในขั้นที่สูงกว่าหรือขั้นที่ได้คะแนนมากกว่าต้องอาศัยการตอบในขั้นที่ต่ำกว่าหรือได้คะแนนน้อยกว่า แบ่งช่วงคะแนนความรู้ออกเป็น 5 ระดับ ดังนี้

- 4 คะแนน หมายถึง ตอบข้อที่ใช้ลำดับขั้นของกระบวนการหาคำตอบในขั้นที่ 1 – 4
- 3 คะแนน หมายถึง ตอบข้อที่ใช้ลำดับขั้นของกระบวนการหาคำตอบในขั้นที่ 1 – 3
- 2 คะแนน หมายถึง ตอบข้อที่ใช้ลำดับขั้นของกระบวนการหาคำตอบในขั้นที่ 1 – 2
- 1 คะแนน หมายถึง ตอบข้อที่ใช้ลำดับขั้นของกระบวนการหาคำตอบในขั้นที่ 1
- 0 คะแนน หมายถึง ไม่ตอบ

วิธีการตรวจให้คะแนนตามความยากประจำขั้น (Step Difficulties) หมายถึง การตรวจให้คะแนนตามตัวเลือกที่มีการกำหนดน้ำหนักความยากประจำขั้นของแต่ละตัวเลือกไว้แล้ว พิจารณาน้ำหนักความยากประจำขั้นตามทฤษฎีของโครงสร้างความรู้ ทางคณิตศาสตร์เรียนรู้ จากง่ายไปยาก คือเรียนรู้จากการบวก การลบ การคูณ การหาร จำนวนเต็ม เศษส่วน เลขยกกำลัง ตามลำดับ แบ่งช่วงคะแนนความยากประจำขั้นเป็น 5 ระดับ ดังนี้

- 4 คะแนน หมายถึง ตอบข้อที่ใช้ทักษะการคิดคำนวณในขั้นยากมากที่สุด
- 3 คะแนน หมายถึง ตอบข้อที่ใช้ทักษะการคิดคำนวณในขั้นยากมาก
- 2 คะแนน หมายถึง ตอบข้อที่ใช้ทักษะการคิดคำนวณในขั้นยากปานกลาง
- 1 คะแนน หมายถึง ตอบข้อที่ใช้ทักษะการคิดคำนวณในขั้นยากน้อย
- 0 คะแนน หมายถึง ไม่ตอบ

ตัวอย่างแบบทดสอบ

คำชี้แจง

1. แบบทดสอบฉบับนี้มีจำนวน 40 ข้อ ใช้เวลา 60 นาที
2. แบบทดสอบฉบับนี้เป็นแบบทดสอบแบบเลือกตอบ ชนิด 4 ตัวเลือก
3. ให้นักเรียนเขียนเครื่องหมายกากบาท (X) ลงในช่องที่ตรงกับตัวเลือกที่สามารถคิดคำนวณได้อย่างมั่นใจที่สุดเพียงข้อเดียว

$$(0) \quad \sqrt{(9/0.3) - 175} = ?$$

ก. 900

ข. 625

ค. 30

ง. 25

ตาราง 5 เกลยและอธิบายการให้คะแนนตามวิธีการตรวจให้คะแนน

ตอบ	วิธีการตรวจให้คะแนนของการทดสอบความน่าจะเป็น
ก. 900	ได้ 2 คะแนน เนื่องจาก เป็นคำตอบที่ใช้ลำดับขั้นของกระบวนการหาคำตอบในขั้นที่ 1 – 2
ข. 625	ได้ 3 คะแนน เนื่องจาก เป็นคำตอบที่ใช้ลำดับขั้นของกระบวนการหาคำตอบในขั้นที่ 1 – 3
ค. 30	ได้ 1 คะแนน เนื่องจาก เป็นคำตอบที่ใช้ลำดับขั้นของกระบวนการหาคำตอบในขั้นที่ 1
ง. 25	ได้ 4 คะแนน เนื่องจาก เป็นคำตอบที่ใช้ลำดับขั้นของกระบวนการหาคำตอบในขั้นที่ 1 – 4
ตอบ	วิธีการตรวจให้คะแนนตามความยากประจำขั้น
ก. 900	ได้ 3 คะแนน เป็นทักษะการคิดในขั้นยากมาก โดยพิจารณาจาก $9 \div \frac{3}{10} = 9 \times \frac{10}{3} = 30$
ข. 625	ได้ 2 คะแนน เป็นทักษะการคิดในขั้นยากปานกลาง โดยพิจารณาจาก $30^2 = 30 \times 30 = 900$
ค. 30	ได้ 1 คะแนน เป็นทักษะการคิดในขั้นยากน้อย โดยพิจารณาจาก $900 - 275 = 625$
ง. 25	ได้ 4 คะแนน เป็นทักษะการคิดในขั้นยากมากที่สุด โดยพิจารณาจาก $\sqrt{625} = \sqrt{25 \times 25} = 25$

(00) ค่าของ $\left[\frac{(139 - 94) \times 0.8}{2} \right]^2$ ตรงกับข้อใด

ก. 18

ข. 324

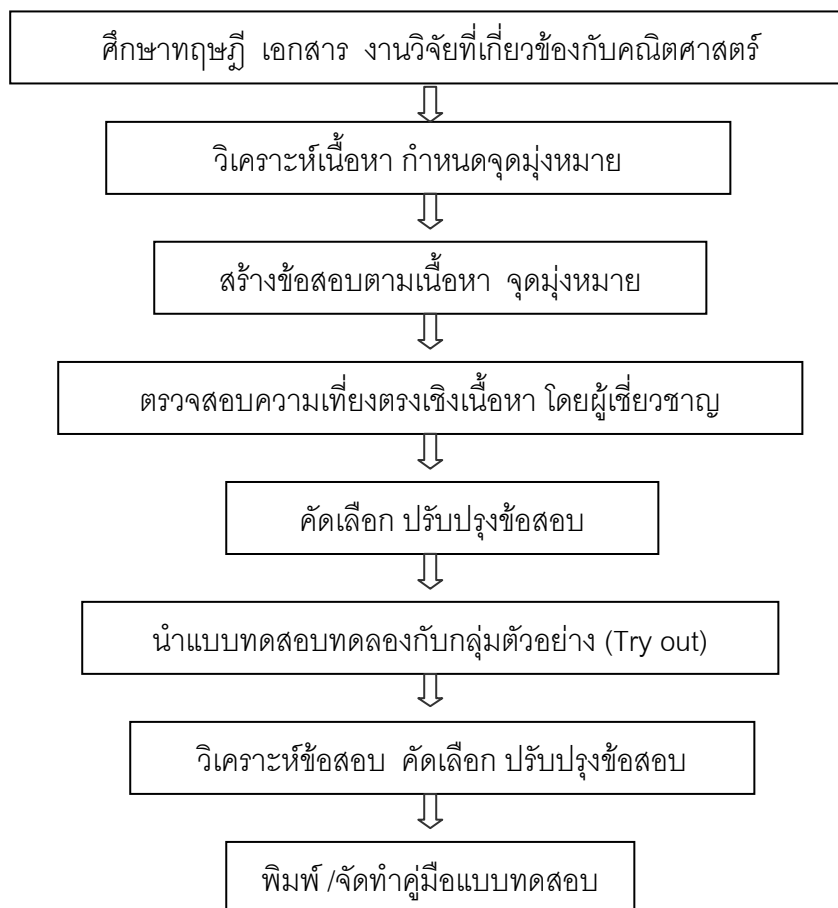
ค. 45

ง. 36

ตาราง 5 (ต่อ) เฉลยและอธิบายการให้คะแนนตามวิธีการตรวจให้คะแนน

ตอบ	วิธีการตรวจให้คะแนนของการทดสอบความน่าจะเป็น	
ก. 18	ได้ 2 คะแนน	เนื่องจาก เป็นคำตอบที่ใช้ลำดับชั้นของกระบวนการหาคำตอบในชั้นที่ 1 – 2
ข. 324	ได้ 4 คะแนน	เนื่องจาก เป็นคำตอบที่ใช้ลำดับชั้นของกระบวนการหาคำตอบในชั้นที่ 1 – 4
ค. 45	ได้ 1 คะแนน	เนื่องจาก เป็นคำตอบที่ใช้ลำดับชั้นของกระบวนการหาคำตอบในชั้นที่ 1
ง. 36	ได้ 3 คะแนน	เนื่องจาก เป็นคำตอบที่ใช้ลำดับชั้นของกระบวนการหาคำตอบในชั้นที่ 1 – 3
ตอบ	วิธีการตรวจให้คะแนนตามความยากประจำชั้น	
ก. 18	ได้ 3 คะแนน	ใช้ทักษะการคิดในชั้นยากมาก โดยพิจารณาจาก $36 \div 2 = 18$
ข. 324	ได้ 4 คะแนน	ใช้ทักษะการคิดในชั้นยากมากที่สุดโดยพิจารณาจาก $18^2 = 18 \times 18 = 324$
ค. 45	ได้ 1 คะแนน	ใช้ทักษะการคิดในชั้นยากน้อย โดยพิจารณาจาก $139 - 94 = 45$
ง. 36	ได้ 2 คะแนน	ใช้ทักษะการคิดในชั้นยากปานกลาง โดยพิจารณาจาก $45 \times 0.8 = 36$

การดำเนินการสร้างเครื่องมือวัดแบบทดสอบผลสัมฤทธิ์ทางคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 มีขั้นตอนการสร้างดังนี้



ภาพประกอบ 9 ขั้นตอนในการสร้างแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์

ในการสร้างเครื่องมือวัดแบบทดสอบผลสัมฤทธิ์ทางคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 3 อาศัยเอกสารหลักสูตรและการวิเคราะห์หลักสูตร เพื่อที่จะระบุจุดมุ่งหมายในการชี้แนวทางการเลือกตัวแทนของเนื้อหาที่จะออกข้อสอบว่าออกข้อสอบอย่างไรจึงจะเป็นตัวแทนที่ดีของเนื้อหา วิธีการนี้จะเป็นหลักประกันได้ว่าแบบทดสอบนั้นจะมีความเที่ยงตรงด้านเนื้อหา (Content validity) มีการดำเนินการสร้างแบบทดสอบ ดังนี้

1. ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการสร้างแบบทดสอบวัดผลสัมฤทธิ์ทางคณิตศาสตร์เพื่อเป็นแนวทางการสร้างแบบทดสอบ

2. ศึกษาหลักสูตรคู่มือครูและแบบเรียนวิชาคณิตศาสตร์ ช่วงชั้นที่ 3 ชั้นมัธยมศึกษาปีที่ 3 เพื่อวิเคราะห์เนื้อหา จุดประสงค์การเรียนรู้ สาระทักษะการคิดคำนวณเกี่ยวกับการบวก การลบ การคูณ การหาร จำนวนจริงและเลขยกกำลัง

3. สร้างตารางวิเคราะห์เนื้อหา โดยกำหนดเนื้อหา จุดประสงค์เชิงพฤติกรรมให้สอดคล้องกับพฤติกรรมที่ต้องการวัด

ตาราง 6 การวิเคราะห์เนื้อหาในการสร้างข้อสอบวัดทักษะกระบวนการคิดคำนวณ

เนื้อหา	จำนวนข้อ	ข้อที่
การบวก การลบ การคูณ การหารจำนวนจริง	30	1,2,5,14,15,16,17,18,19,20,21,22,23,24,25,26, 31,32,40,41,42,43,44,45,47,48,49,51,52,55,
เลขยกกำลัง	30	3,4,6,7,8,9,10,11,12,13,27,28,29,30,33,34, 35,36,37,38,39,43,46,50,54,56,57,58,59,60

4. สร้างข้อสอบตามเนื้อหาได้แบบทดสอบวัดผลสัมฤทธิ์คณิตศาสตร์ ด้านทักษะกระบวนการคิดคำนวณเกี่ยวกับการบวก การลบ การคูณและการหาร จำนวนจริง เลขยกกำลัง ชนิดเลือกตอบ 4 ตัวเลือก จำนวน 60 ข้อ

6. ตรวจสอบคุณภาพเครื่องมือ

6.1 นำเครื่องมือที่สร้างขึ้นให้ผู้เชี่ยวชาญ จำนวน 5 คน ตรวจสอบคุณภาพเบื้องต้น ด้านความเที่ยงตรงด้านเนื้อหา (Content Validity) เพื่อให้ผู้เชี่ยวชาญประเมินว่าข้อสอบแต่ละข้อนั้น สามารถวัดได้ตรงกับเนื้อหาหรือไม่ และกำหนดค่าน้ำหนักคะแนนของตัวเลือกตามวิธีการตรวจให้คะแนนทั้งสองวิธี การหาค่าดัชนีความสอดคล้องเป็นรายข้อ (Index of Congruency: IOC) โดยใช้สูตร (บุญเขต ภิญญอนันตพงษ์. 2545: 179)

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

IOC แทน ดัชนีความสอดคล้อง

ΣR แทน ผลรวมของคะแนนความคิดเห็นของผู้เชี่ยวชาญแต่ละคน

N แทน จำนวนผู้เชี่ยวชาญ

คะแนนค่าดัชนีความสอดคล้องระหว่างเนื้อหากับข้อสอบรายข้อของผู้เชี่ยวชาญ ดังนี้

+ 1 หมายถึง แน่ใจว่าสอดคล้องกับเนื้อหาและวิธีการให้คะแนน

0 หมายถึง ไม่แน่ใจสอดคล้องกับเนื้อหาและวิธีการให้คะแนน

- 1 หมายถึง แน่ใจว่าไม่สอดคล้องกับเนื้อหาและวิธีการให้คะแนน

ได้ค่าดัชนีความสอดคล้อง (IOC) ตั้งแต่ - 0.4 ถึง 1.0 นำมาปรับปรุง แก้ไขข้อสอบให้เหมาะสมตามคำแนะนำของอาจารย์ที่ควบคุมปริญญานิพนธ์และผู้เชี่ยวชาญ จำนวน 5 คน แล้วคัดเลือกข้อสอบ จำนวน 47 ข้อ โดยมีค่าดัชนีความสอดคล้อง ตั้งแต่ .50 ขึ้นไป

6.2 นำแบบทดสอบที่ปรับปรุงตามคำแนะนำของอาจารย์ผู้ควบคุมปริญญานิพนธ์และผู้เชี่ยวชาญ ไปทดลองใช้กับนักเรียนชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 2 ปีการศึกษา 2549 จำนวน 100 คน ประกอบด้วยนักเรียนโรงเรียนอนุบาลนาดี จำนวน 30 คน โรงเรียนทุ่งใหญ่วิทยาคาร จำนวน 50 คน และโรงเรียนเตรียมอุดมศึกษาน้อมเกล้าฯ กบินทร์บุรี จำนวน 20 คน ระหว่างวันที่ 2 – 10 ธันวาคม 2549

6.3 นำแบบทดสอบมาตรวจให้คะแนนตามวิธีการให้คะแนน

6.4 นำผลการทดสอบ วิเคราะห์คุณภาพข้อสอบ เพื่อหาค่าความยากโดยพิจารณาค่าความยากจากค่าเฉลี่ยของคะแนนของข้อนั้นๆ ตามวิธีการให้คะแนนทั้งสองวิธี โดยแบ่งระดับความยากออกเป็น 3 ระดับ ดังนี้

1.00 - 1.50	แทน	ข้อสอบยาก
1.51 - 2.59	แทน	ข้อสอบยากปานกลาง
2.60 - 4.00	แทน	ข้อสอบง่าย

ผลการวิเคราะห์ พบว่า มีค่าความยากรายข้อจากการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็น อยู่ระหว่าง 2.27 ถึง 3.67 เป็นความยากที่อยู่ในระดับปานกลางถึงง่าย และจากการตรวจให้คะแนนตามความยากประจำชั้น อยู่ระหว่าง 2.07 ถึง 3.97 เป็นความยากที่อยู่ในระดับปานกลางถึงง่าย

6.5 หาค่าอำนาจจำแนกของข้อสอบเป็นรายข้อ ตามวิธีตรวจให้คะแนนทั้ง 2 วิธี เพื่อคัดเลือกข้อสอบที่มีค่าอำนาจจำแนกสูง พิจารณาค่าอำนาจจำแนกจากค่าสหสัมพันธ์ระหว่างคำถามเป็นรายข้อกับคะแนนรวม (Item – total correlation) วิเคราะห์โดยใช้โปรแกรม SPSS คัดเลือกข้อสอบที่มีค่าอำนาจจำแนกตามวิธีการตรวจให้คะแนน ดังนี้

วิธีการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็น มีค่าอำนาจจำแนกรายข้ออยู่ระหว่าง .21 ถึง .72

วิธีการตรวจให้คะแนนตามความยากประจำชั้นมีค่าอำนาจจำแนกรายข้อระหว่าง .21 ถึง .91

คัดเลือกข้อสอบโดยพิจารณาค่าอำนาจจำแนกที่มีค่าตั้งแต่ .21 ขึ้นไป จากวิธีการตรวจให้คะแนนทั้งสองวิธีไว้ จำนวน 40 ข้อ

6.6 นำคะแนนที่ได้มาวิเคราะห์หาค่าความเชื่อมั่น จากสูตรสัมประสิทธิ์แอลฟา (α – Coefficient) ของครอนบัค ความเชื่อมั่นที่ได้จากการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็น เท่ากับ 0.92 และจากการตรวจให้คะแนนตามความยากประจำชั้น เท่ากับ 0.91

7. ปรับปรุงข้อสอบและคัดเลือกข้อสอบที่มีคุณภาพในเกณฑ์ตามทฤษฎีตอบสนองข้อสอบให้เหมาะสม จัดพิมพ์และนำไปใช้กับนักเรียนกลุ่มตัวอย่าง

ตอนที่ 3 วิธีการดำเนินการทดลองและเก็บรวบรวมข้อมูล

ระยะเวลาในการเก็บรวบรวมข้อมูล

การดำเนินการเก็บรวบรวมข้อมูลการวิจัย ผู้วิจัยหรือตัวแทนผู้วิจัยใช้เวลาในการรวบรวมข้อมูลระหว่างวันที่ 22 มกราคม – 16 กุมภาพันธ์ 2550

ขั้นตอนการเก็บรวบรวมข้อมูล

1. ติดต่อโรงเรียนที่ใช้เป็นกลุ่มตัวอย่าง จำนวน 29 โรงเรียนด้วยตนเอง ในช่วงวันที่ 18 มกราคม ถึง 16 กุมภาพันธ์ 2550 เพื่อกำหนด วัน เวลา ในการสอบ
2. เตรียมข้อสอบไว้เพียงพอกับความต้องการของนักเรียนที่สอบในแต่ละครั้ง และวางแผนในการดำเนินการสอบล่วงหน้า
3. ดำเนินการสอบด้วยตนเอง จำนวน 5 โรงเรียน ตัวแทนผู้วิจัย ดำเนินการสอบแทนจำนวน 5 โรงเรียน

3.1 รายชื่อโรงเรียนที่ดำเนินการเก็บข้อมูลด้วยตนเอง

วันที่ 22 มกราคม 2550 โรงเรียนโรงเรียนมณีเสวตอุปถัมภ์

วันที่ 23 มกราคม 2550 โรงเรียนบ้านนุสูง

วันที่ 1 กุมภาพันธ์ 2550 โรงเรียนบ้านแหลมหินและโรงเรียนกองทัพบกอุปถัมภ์

วันที่ 2 กุมภาพันธ์ 2550 โรงเรียนไทยรัฐวิทยา

3.2 รายชื่อโรงเรียนที่ตัวแทนผู้วิจัย ดำเนินการสอบ

วันที่ 22 มกราคม 2550 โรงเรียนอนุบาลประจันตคาม

วันที่ 6 กุมภาพันธ์ 2550 โรงเรียนพระราชรัฐวิทยา

วันที่ 7 – 9 กุมภาพันธ์ 2550 โรงเรียนปราจีนกัลยาณี

วันที่ 15 กุมภาพันธ์ 2550 โรงเรียนสุวรรณวิทยา

วันที่ 16 กุมภาพันธ์ 2550 โรงเรียนวัดบ้านหอย

4. ผู้วิจัย ตัวแทนผู้วิจัย อธิบายให้นักเรียนเข้าใจจุดประสงค์ของการทำแบบทดสอบ
5. ทำการทดสอบกับนักเรียนกลุ่มตัวอย่าง

ตอนที่ 4 การวิเคราะห์ข้อมูลและสถิติที่ใช้ในการวิเคราะห์ข้อมูล

4.1 การจัดกระทำข้อมูล

1. ในการทดลอง (Try out) เครื่องมือ มีขั้นตอน ดังนี้

1.1 ตรวจสอบความสมบูรณ์ของการตอบแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนจากกระดาษคำตอบของนักเรียนก่อน

1.2 นำแบบทดสอบมาตรวจให้คะแนนตามเกณฑ์ที่กำหนดไว้

1.3 วิเคราะห์หาค่าอำนาจจำแนกรายข้อ โดยพิจารณาจากค่าสหสัมพันธ์ระหว่างคำถามเป็นรายข้อกับคะแนนรวม (Item – total correlation)

1.4 หาค่าความยากโดยพิจารณาจากค่าเฉลี่ยของคะแนนรายข้อ

2. การวิเคราะห์ข้อมูลเพื่อตอบคำถามการวิจัย มีขั้นตอน ดังนี้

2.1 วิเคราะห์ข้อมูลขั้นต้น (preliminary data analysis) มีขั้นตอน ดังนี้

2.1.1 นำข้อมูลที่รวบรวมได้มาตรวจสอบความสมบูรณ์ของการตอบจากกระดาษคำตอบของนักเรียนว่ามีสภาพเป็นอย่างไร

2.1.2 ตรวจให้คะแนนตามเกณฑ์กำหนด

2.1.3 วิเคราะห์ค่าสถิติพื้นฐาน ได้แก่ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน

2.1.4 วิเคราะห์ข้อตกลงเบื้องต้นของทฤษฎีตอบสนองข้อสอบ ด้านความเป็นมิติเดียว (Unidimension) ของแบบทดสอบ โดยการวิเคราะห์องค์ประกอบ (Factor Analysis) แล้วสังเกตค่าไอเก้น (Eigen value) ค่าสูงสุดว่าแตกต่างจากค่าอื่น ๆ อย่างชัดเจนหรือไม่ (สุรชัย มีชาญ. 2539: 25; อ้างอิงจาก Hambleton and Trumb. 1973: 195 – 211)

2.2 นำข้อมูลที่ผ่านการวิเคราะห์ขั้นต้นมาทำการหาคุณภาพข้อสอบตามทฤษฎีตอบสนองข้อสอบ ในโมเดลพาเชียล เครดิต (Partial Credit Model) โดยใช้โปรแกรม MULTILOG 7.03 For Window (David Thissen, Wen-Hung Chen and Darrell Bock. 2003) เพื่อตอบคำถามการวิจัย

4.2 สถิติที่ใช้ในการวิเคราะห์ข้อมูล

1. สถิติที่ใช้ในการหาคุณภาพเครื่องมือ

1.1 หาค่าความเที่ยงตรงเชิงเนื้อหา โดยพิจารณาหาค่าดัชนีความสอดคล้อง (Index of Congruency : IOC) (บุญเชิด ภิญโญนนตพงษ์. 2545: 179)

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

เมื่อ IOC แทน ดัชนีความสอดคล้อง
 $\sum R$ แทน ผลรวมของความคิดเห็นของผู้เชี่ยวชาญทั้งหมด
 N แทน จำนวนผู้เชี่ยวชาญทั้งหมด

1.2 วิเคราะห์ข้อคำถามเป็นรายข้อ (Item Analysis) เพื่อหาค่าอำนาจจำแนก โดยพิจารณาจากค่าสหสัมพันธ์ระหว่างข้อคำถามเป็นรายข้อกับคะแนนรวม (Item – total correlation) โดยใช้โปรแกรม SPSS version 10.0 คำนวณโดยใช้สูตรสัมประสิทธิ์สหสัมพันธ์ (บุญเชิด ภิญโญนนตพงษ์. 2545: 84)

$$r = \frac{N\sum XY - \sum X\sum Y}{\sqrt{[N\sum X^2 - (\sum X)^2][N\sum Y^2 - (\sum Y)^2]}}$$

เมื่อ r แทน สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน
 N แทน จำนวนกลุ่มตัวอย่าง
 $\sum X$ แทน ผลรวมของคะแนนรายข้อ (item)
 $\sum Y$ แทน ผลรวมของคะแนนรวม (total) ที่ไม่รวมข้อที่นำมาคำนวณ
 $\sum XY$ แทน ผลรวมของผลคูณของคะแนน X และคะแนน Y
 $\sum X^2$ แทน ผลรวมกำลังสองของคะแนน X
 $\sum Y^2$ แทน ผลรวมกำลังสองของคะแนน Y

1.3 ค่าความยาก (Difficulty) โดยพิจารณาค่าเฉลี่ยของคะแนนรายข้อ

$$\bar{X} = \frac{\sum X}{N}$$

เมื่อ $\bar{X}$ แทน ดัชนีค่าความยาก
 $\sum X$ แทน ผลรวมของคะแนนแต่ละข้อ
 N แทน จำนวนนักเรียนที่ทำข้อสอบ

1.4 ความเชื่อมั่น (Reliability) โดยวิธีสัมประสิทธิ์แอลฟา (α – Coefficient) ของครอนบัค (Cronbach alpha procedure) (ล้วน สายยศและอังคณา สายยศ. 2543: 218)

$$\alpha = \frac{k}{k-1} \left\{ 1 - \frac{\sum \sigma_i^2}{\sigma^2} \right\}$$

เมื่อ α แทน สัมประสิทธิ์ความเชื่อมั่น
 k แทน จำนวนข้อสอบ
 σ_i^2 แทน คะแนนความแปรปรวนเป็นรายข้อ

4.3 การวิเคราะห์ข้อมูลในการวิจัย

ข้อมูลที่ได้รับจากการสอบของนักเรียนนำมาตรวจให้คะแนนตามเกณฑ์ที่กำหนดแล้ว นำคะแนนที่ได้ในแบบทดสอบการวิเคราะห์ข้อมูล ดังนี้

1. วิเคราะห์ข้อมูลเบื้องต้นจากการหาค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน
2. ตรวจสอบคุณสมบัติการวัดมิติเดียวของแบบทดสอบของการตรวจให้คะแนนแต่ละวิธีด้วยการวิเคราะห์องค์ประกอบโดยสัณฐานองค์ประกอบหลัก (Principle Component: PC) โดยใช้โปรแกรมสำเร็จรูป SPSS
3. วิเคราะห์ค่าความเที่ยงตรงตามสภาพ จากค่าสัมประสิทธิ์สหสัมพันธ์ (r) ระหว่างคะแนนสอบกลางภาคซึ่งเป็นคะแนนดิบของกลุ่มตัวอย่าง ในภาคเรียนที่สอง ปีการศึกษา 2549 กับคะแนนของวิธีการตรวจให้คะแนนด้วยวิธีของการทดสอบความน่าจะเป็นและคะแนนของวิธีการตรวจให้คะแนนตามความยากประจำชั้น แล้วแปลงค่าความเที่ยงตรงตามสภาพจากค่าสัมประสิทธิ์สหสัมพันธ์ เป็นคะแนนฟิชเชอร์ซี (Z_r) โดยใช้สูตร (Glass and Stanley. 1970: 265)

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

เมื่อ Z_r แทน คะแนนฟิชเชอร์ซี
 R แทน ค่าสัมประสิทธิ์สหสัมพันธ์

4. เปรียบเทียบความเป็นอิสระของความเที่ยงตรงตามสภาพในภาพรวม โดยใช้สถิติทดสอบ ไคว์สแควร์ (χ^2) (Glass and Stanley.1970: 267) มีสูตร ดังนี้

$$\chi^2_{j=1;\alpha} = \sum (Z_{rj} - Z_r)^2 (n_j - 3)$$

เมื่อ	χ^2	แทน	สถิติทดสอบไคว์สแควร์
	Z_{rj}	แทน	คะแนนพิชเชอร์รี่ ของกลุ่มผู้สอบ j
	n_j	แทน	จำนวนผู้สอบ

5. วัดคุณภาพของข้อสอบและแบบทดสอบจากฟังก์ชันสารสนเทศของข้อสอบ และฟังก์ชันสารสนเทศของแบบทดสอบ โดยวิเคราะห์ข้อสอบรายข้อตามทฤษฎีการตอบข้อสอบ (IRT) แบบพหุภาคีที่ใช้โมเดลการให้คะแนนความรู้บางส่วน (Partial Credit Model: PCM) โดยใช้โปรแกรม MULTILOG 7.03 For Window (David Thissen, Wen-Hung Chen and Darrell Bock. 2003)

5.1 คำนวณค่าฟังก์ชันสารสนเทศของข้อสอบ (Dodd, De Ayala and Koch.1995) จากสูตร

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

เมื่อ	$I_i(\theta)$	แทน	สารสนเทศของข้อสอบข้อที่ i ที่ระดับความสามารถ θ
	$P'_{ix}(\theta)$	แทน	ความน่าจะเป็นหรือโอกาสสำหรับคนที่มีความสามารถ θ จะตอบข้อ i ได้คะแนน 1 มากกว่าคะแนน x
	$P_{ix}(\theta)$	แทน	ความน่าจะเป็นหรือโอกาสของคนที่มีความสามารถ θ จะตอบข้อสอบ i ได้คะแนน x

5.2 ค่าฟังก์ชันสารสนเทศของแบบทดสอบ

จาก IIF สามารถนำมาคำนวณ TIF

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

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

5.3 ค่าความคลาดเคลื่อนมาตรฐานของแบบทดสอบ

จาก TIF สามารถนำมาคำนวณ $SE(\theta)$

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

เมื่อ $SE(\theta)$ แทน ความคลาดเคลื่อนมาตรฐานของการประมาณค่า
 สำหรับผู้ตอบที่มีความสามารถ θ
 $I(\theta)$ แทน สารสนเทศของแบบทดสอบสำหรับผู้ตอบที่มี
 ความสามารถ θ

5.4 ค่าความประสิทธิภาพสัมพัทธ์ (Relative Efficiency : $RE(\theta)$) จากสูตร

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

เมื่อ $I_A(\theta)$ แทน ค่าฟังก์ชันสารสนเทศของแบบทดสอบฉบับ A ณ ตำแหน่ง
 บนสเกลความสามารถร่วมกันระดับ θ
 $I_B(\theta)$ แทน ค่าฟังก์ชันสารสนเทศของแบบทดสอบฉบับ B ณ ตำแหน่ง
 บนสเกลความสามารถร่วมกันระดับ θ

การแปลความหมายของค่าดัชนีประสิทธิภาพสัมพัทธ์ สามารถนำไปช่วยคัดเลือก
 แบบทดสอบได้ดังนี้

$RE(\theta) = 1$ แสดงว่าแบบทดสอบทั้งสองฉบับมีประสิทธิภาพเท่ากันสำหรับผู้ตอบที่มี
 ระดับความสามารถ θ

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

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

บทที่ 4

ผลการวิเคราะห์ข้อมูล

สัญลักษณ์ที่ใช้ในการวิเคราะห์ข้อมูล

การเสนอผลการวิเคราะห์ข้อมูลในการวิจัยครั้งนี้ ผู้วิจัยได้กำหนดสัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูลเพื่อสื่อความหมายในการเสนอผลงานวิจัยให้เข้าใจตรงกัน ดังนี้

Min	แทน	คะแนนต่ำที่สุด
Max	แทน	คะแนนสูงสุด
Mean	แทน	คะแนนเฉลี่ย
SD	แทน	ค่าความเบี่ยงเบนมาตรฐาน
k	แทน	คะแนนเต็ม
r	แทน	ค่าสัมประสิทธิ์สหสัมพันธ์
Z_r	แทน	คะแนนพิชเชอร์ซี
χ^2	แทน	สถิติทดสอบไคว์สแควร์
θ	แทน	ความสามารถ
$SE(\theta)$	แทน	ความคลาดเคลื่อนมาตรฐานของการประมาณค่าสำหรับผู้ตอบที่มีระดับความสามารถ θ
$RE(\theta)$	แทน	ค่าประสิทธิภาพสัมพัทธ์สำหรับผู้ตอบที่มีระดับความสามารถ θ
IIF	แทน	ค่าฟังก์ชันสารสนเทศของข้อสอบ
TIF	แทน	ค่าฟังก์ชันสารสนเทศของแบบทดสอบ

การเสนอผลการวิเคราะห์ข้อมูล

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบความเที่ยงตรงตามสภาพ (Concurrent Validity) ของข้อสอบ ฟังก์ชันสารสนเทศของข้อสอบ (Item Information) และฟังก์ชันสารสนเทศของแบบทดสอบ (Test Information) จากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 3 ที่ตรวจให้คะแนนความรู้บางส่วน ด้วยวิธีการของการทดสอบความน่าจะเป็นกับการตรวจให้คะแนนตามความยากประจำชั้น ผู้วิจัยได้เสนอผลการวิเคราะห์ข้อมูลเรียงตามลำดับขั้นตอน ดังนี้

ตอนที่ 1 ผลการวิเคราะห์ข้อมูลเบื้องต้น

ตอนที่ 2 ผลการเปรียบเทียบความเที่ยงตรงตามสภาพระหว่างวิธีการตรวจให้คะแนน
ด้วยวิธีการของการทดสอบความน่าจะเป็นกับวิธีการตรวจให้คะแนน
ตามความยากประจำชั้น

ตอนที่ 3 ผลการเปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของ
แบบทดสอบระหว่างวิธีการตรวจให้คะแนนด้วยวิธีการของการทดสอบ
ความน่าจะเป็นกับการตรวจให้คะแนนตามความยากประจำชั้น

ตอนที่ 1 ผลการวิเคราะห์ข้อมูลเบื้องต้น

1. ค่าสถิติพื้นฐานของคะแนนที่ได้จากการตอบแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน
คณิตศาสตร์ ค่าเฉลี่ยและค่าส่วนเบี่ยงเบนมาตรฐานของคะแนนดิบจากแบบทดสอบวัดผลสัมฤทธิ์
คณิตศาสตร์ จำนวน 40 ข้อ ที่มีวิธีการตรวจแตกต่างกัน 2 วิธี จากแบบทดสอบชุดเดียวกัน คือ
วิธีการของการทดสอบความน่าจะเป็นกับการให้คะแนนตามความยากประจำชั้น ซึ่งทั้ง 2 วิธี มี
คะแนนสูงสุด เท่ากับ 4 ต่ำสุดเท่ากับ 0 ในแต่ละข้อ รวมคะแนนเต็มเท่ากับ 160 คะแนน ดังตาราง 7
ตาราง 7 ค่าเฉลี่ยและค่าส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบจำแนกตามวิธีการตรวจให้คะแนน

วิธีการตรวจให้คะแนน	จำนวน ผู้สอบ	k	คะแนน			
			Min	Max	Mean	SD
วิธีการของการทดสอบความน่าจะเป็น	1,007	160	75	147	109.99	14.70
วิธีการให้คะแนนตามความยากประจำชั้น	1,007	160	72	131	95.61	8.72

จากตาราง 6 พบว่า ค่าเฉลี่ยการตรวจคะแนนความรู้บางส่วนด้วยวิธีการของการทดสอบ
ความน่าจะเป็น มีค่าเฉลี่ย เท่ากับ 109.99 สูงกว่าการตรวจให้คะแนนตามความยากประจำชั้นที่มี
ค่าเฉลี่ยเท่ากับ 95.61 และส่วนเบี่ยงเบนมาตรฐานการตรวจให้คะแนนด้วยวิธีการของการทดสอบความ
น่าจะเป็น เท่ากับ 14.70 มีค่าสูงกว่าการตรวจให้คะแนนตามความยากประจำชั้นที่มีค่าเท่ากับ 8.72

2. ผลการวิเคราะห์องค์ประกอบเพื่อตรวจสอบความเป็นมิติเดียว

เพื่อตรวจสอบว่าแบบทดสอบวัดผลสัมฤทธิ์คณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 มีคุณสมบัติ
คุณสมบัติเป็นมิติเดียวหรือไม่ ผู้วิจัยจึงนำคะแนนของการตรวจให้คะแนนแต่ละวิธีมาทำการ

วิเคราะห์องค์ประกอบ โดยการสกัดองค์ประกอบหลัก (Principle Component Analysis) แบบหมุนแกน ด้วยวิธีแวนิแมกซ์ (Varimax) โดยใช้โปรแกรมสำเร็จรูป SPSS ปรากฏผล ดังตาราง 8

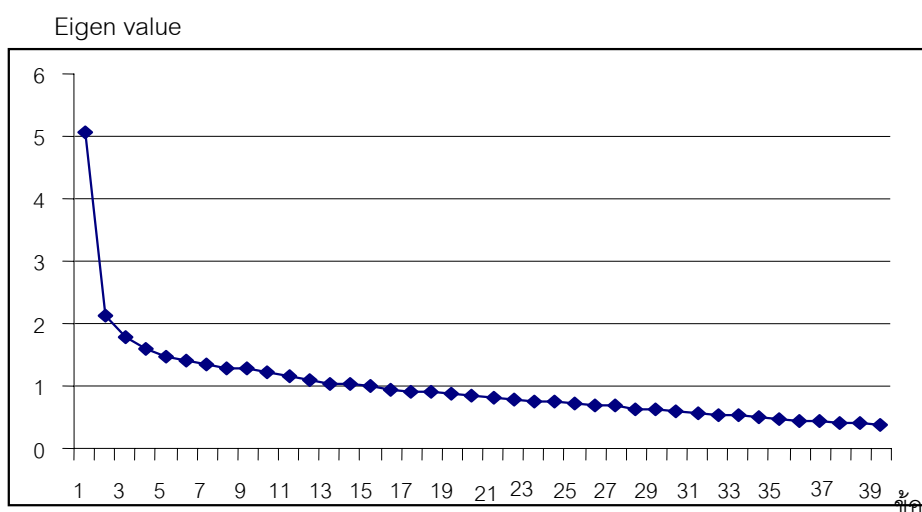
ตาราง 8 ผลการวิเคราะห์องค์ประกอบของแบบทดสอบวัดผลสัมฤทธิ์คณิตศาสตร์ ที่มีการตรวจให้คะแนนความรู้บางส่วน ด้วยวิธีการของการทดสอบความน่าจะเป็นกับการตรวจให้คะแนนตามความยากประจำชั้น

องค์ประกอบที่	ค่าไอเกน (Eigen Value) และเปอร์เซ็นต์ของความแปรปรวน			
	วิธีการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็น		วิธีการตรวจให้คะแนนตามความยากประจำชั้น	
	Eigen	% of V	Eigen	% of V
1	5.060	12.650	3.976	9.941
2	2.113	5.283	2.136	5.341
3	1.771	4.428	1.708	4.270
4	1.608	4.021	1.589	3.971
5	1.466	3.665	1.499	3.748
6	1.397	3.491	1.490	3.725
7	1.333	3.333	1.363	3.407
8	1.274	3.185	1.326	3.314
9	1.271	3.178	1.272	3.180
10	1.219	3.048	1.243	3.108
11	1.162	2.905	1.186	2.966
12	1.088	2.719	1.155	2.887
13	1.043	2.609	1.125	2.812
14	1.018	2.545	1.053	2.633
15	1.013	2.533	1.045	2.611
16	.938	2.346	1.014	2.535
17	.918	2.295	.980	2.449
18	.898	2.244	.927	2.317
19	.870	2.176	.901	2.253
20	.832	2.079	.854	2.135
21	.800	2.001	.829	2.072
22	.794	1.986	.813	2.032

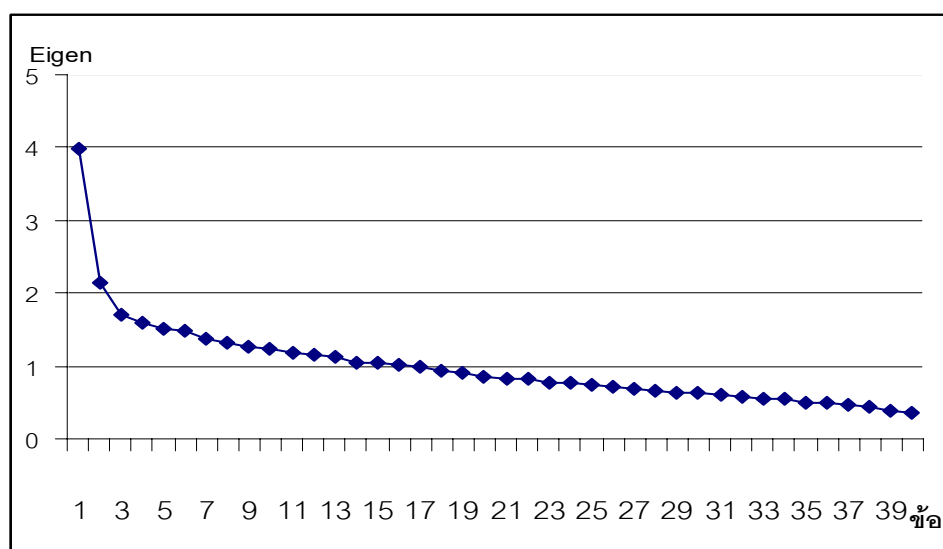
ตาราง 8 (ต่อ)

องค์ประกอบที่	ค่าไอเกน (Eigen Value) และเปอร์เซ็นต์ของความแปรปรวน			
	วิธีการตรวจให้คะแนน		วิธีการตรวจให้คะแนน	
	ด้วยวิธีของการทดสอบความน่าจะเป็น		ตามความยากประจำชั้น	
	Eigen	% of V	Eigen	% of V
23	.756	1.890	.781	1.953
24	.750	1.875	.771	1.927
25	.710	1.776	.751	1.876
26	.691	1.727	.711	1.778
27	.675	1.687	.680	1.699
28	.632	1.581	.659	1.648
29	.614	1.536	.630	1.576
30	.583	1.458	.622	1.556
31	.559	1.398	.600	1.499
32	.538	1.346	.574	1.435
33	.522	1.304	.563	1.409
34	.508	1.271	.538	1.344
35	.480	1.200	.498	1.245
36	.449	1.124	.484	1.210
37	.438	1.094	.464	1.160
38	.419	1.048	.442	1.104
39	.414	1.034	.395	.988
40	.373	.933	.354	.884

นำค่าไอเกน (Eigen value) ของวิธีการให้คะแนนไปเขียนกราฟจะได้กราฟตาม
ภาพประกอบ 10 และ 11 ดังนี้



ภาพประกอบ 10 ค่าไอเกน(Eigen value) การตรวจให้คะแนนวิธีการของการทดสอบความน่าจะเป็น



ภาพประกอบ 11 ค่าไอเกน (Eigen value) การตรวจให้คะแนนตามความยากประจำชั้น

จากตาราง 8 ภาพประกอบ 10 และ 11 พบว่า แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน คณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 3 ด้วยค่าไอเกน (Eigen value) ของวิธีการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็น ค่าไอเกนของตัวประกอบ ตัวที่ 1 เท่ากับ 5.060 สูงกว่าค่าไอเกนของตัวประกอบ ตัวที่ 2 และค่าไอเกนของตัวประกอบที่เหลือมีค่าใกล้เคียงกัน ซึ่งแตกต่างกันอย่างชัดเจน และค่าไอเกนของวิธีการตรวจให้คะแนนตามความยากประจำชั้น ค่าไอเกนของตัวประกอบ ตัวที่ 1 เท่ากับ 3.976 สูงกว่าค่าไอเกนของตัวประกอบ ตัวที่ 2 และค่าไอเกนของตัวประกอบที่เหลือมีค่าใกล้เคียงกันซึ่งแตกต่างกันอย่างชัดเจนเช่นเดียวกัน ค่าไอเกนที่แสดงเห็นถึงความแตกต่างอย่างชัดเจน ถือได้ว่าแบบทดสอบนั้นมีความเป็นมิติเดียว (ผจงจิต อินทสุวรรณ.

2545: 17 - 19) สะท้อนให้เห็นว่าคะแนนของผลการตอบแบบทดสอบวัดผลสัมฤทธิ์คณิตศาสตร์มีความเป็นมิตินี้เดียวค่อนข้างเด่นชัดสามารถนำไปประมาณค่าพารามิเตอร์และวิเคราะห์ตามหลักการของ Partial Credit Model ได้

ตอนที่ 2 ผลการเปรียบเทียบความเที่ยงตรงตามสภาพระหว่างวิธีการตรวจให้คะแนนความรู้บางส่วนแบบด้วยวิธีการของการทดสอบความน่าจะเป็นกับการตรวจให้คะแนนตามความยากประจำชั้น

การคำนวณความเที่ยงตรงตามสภาพของแบบทดสอบที่มีวิธีการตรวจให้คะแนนที่แตกต่างกัน พิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์ (r) ระหว่างคะแนนสอบกลางภาคซึ่งเป็นคะแนนดิบวิชาคณิตศาสตร์ของกลุ่มตัวอย่างในภาคเรียนที่ 2 ปีการศึกษา 2549 กับคะแนนของวิธีการตรวจให้คะแนนด้วยวิธีของการทดสอบความน่าจะเป็นและคะแนนของวิธีการตรวจให้คะแนนตามความยากประจำชั้น แล้วแปลงค่าความเที่ยงตรงตามสภาพจากค่าสัมประสิทธิ์สหสัมพันธ์เป็นคะแนนพิชเชอร์ซี (Z_r) โดยใช้สูตร (Glass and Stanley, 1970) แล้วนำมาเปรียบเทียบความแตกต่างค่าความเที่ยงตรงตามสภาพในภาพรวมโดยใช้สถิติทดสอบไคว์สแควร์ (χ^2) ได้ผลการวิเคราะห์แสดงผลดังตาราง 9

ตาราง 9 ความเที่ยงตรงตามสภาพจากการสอบจำแนกตามวิธีการตรวจให้คะแนน

วิธีการตรวจให้คะแนน	จำนวนผู้สอบ	ความเที่ยงตรงตามสภาพ		χ^2
		r	Z_r	
วิธีการของการทดสอบความน่าจะเป็น	1,007	.515**	.570	4.37
วิธีการให้คะแนนตามความยากประจำชั้น	1,007	.465**	.504	

** p-value < .05

ผลการวิเคราะห์ความเที่ยงตรงตามสภาพ ตามตาราง 9 พบว่า ความเที่ยงตรงตามสภาพที่พิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์ (r) ระหว่างคะแนนสอบกลางภาคซึ่งเป็นคะแนนดิบของกลุ่มตัวอย่าง ในภาคเรียนที่สอง ปีการศึกษา 2549 กับคะแนนที่ได้มาจากการตรวจให้คะแนน 2 วิธี มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เมื่อเปรียบเทียบความแตกต่างโดยใช้

คะแนนฟิชเชอร์ซีเปรียบเทียบในภาพรวม โดยใช้สถิติทดสอบไคสแควร์ (χ^2) พบว่า ความเที่ยงตรงตามสภาพไม่ขึ้นอยู่กับวิธีการตรวจให้คะแนนในแต่ละวิธี

ตอนที่ 3 ผลการเปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบฟังก์ชันสารสนเทศของแบบทดสอบระหว่างวิธีการตรวจให้คะแนนความรู้บางส่วนด้วยวิธีการของการทดสอบความน่าจะเป็นกับการตรวจให้คะแนนตามความยากประจำชั้น

3.1 การเปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบตามวิธีการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็นกับการให้คะแนนตามความยากประจำชั้น จำแนกตามรายข้อ (40 ข้อ) แสดงผลดังตาราง 10

ตาราง 10 ฟังก์ชันสารสนเทศสูงสุดของข้อสอบ จำแนกตามรายข้อและวิธีการตรวจให้คะแนน

ข้อ	ฟังก์ชันสารสนเทศของข้อสอบ(IIF)ที่ตรวจให้คะแนนความรู้บางส่วน	
	ด้วยวิธีการของการทดสอบความน่าจะเป็น	วิธีการตรวจให้คะแนนตามความยากประจำชั้น
1	0.843	0.016
2	0.278	0.110
3	0.020	0.103
4	0.006	0.006
5	0.095	0.220
6	0.044	0.016
7	0.002	0.024
8	0.535	0.037
9	0.015	0.022
10	0.008	0.014
11	0.027	0.047
12	0.009	0.018
13	0.006	0.068
14	0.286	0.261
15	0.072	0.007
16	0.074	0.030
17	0.017	0.039

ตาราง 10 (ต่อ) ฟังก์ชันสารสนเทศของข้อสอบ จำแนกตามรายข้อ และวิธีการตรวจให้คะแนน

ข้อ	ฟังก์ชันสารสนเทศของข้อสอบ (IIF)	
	วิธีการตรวจให้คะแนนความรู้บางส่วน	วิธีการตรวจให้คะแนนตาม
	แบบของการทดสอบความน่าจะเป็น	ความยากประจำข้อ
18	0.086	0.015
19	0.023	0.482
20	0.010	0.044
21	0.015	0.296
22	0.011	0.057
23	0.008	0.011
24	0.011	0.016
25	0.600	0.031
26	0.005	0.019
27	0.020	0.029
28	0.056	0.019
29	0.492	0.500
30	0.013	0.002
31	0.011	0.016
32	0.061	0.017
33	0.012	0.083
34	0.020	0.513
35	0.020	0.019
36	0.017	0.007
37	0.004	0.030
38	0.029	0.244
39	0.006	0.027
40	0.109	0.131

จากตาราง 10 พบว่า การตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็น ข้อ 1, 2, 8, 14, 15, 16, 18, 25, 28, 30, 32 ให้ค่าฟังก์ชันสารสนเทศของข้อสอบสูงกว่า การตรวจให้คะแนนตามความยากประจำข้อ และการตรวจให้คะแนนตามความยากประจำข้อ ข้อที่ 3, 4, 5,

6, 7, 9, 10, 11, 12, 13, 17, 19, 20, 21, 22, 23, 24, 26, 27, 29, 31, 31, 33, 34, 35, 36, 37, 38, 39, 40 ให้ค่าฟังก์ชันสารสนเทศของข้อสอบสูงกว่าการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็น

3.2 การเปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบจำแนกตามคะแนนความสามารถของผู้สอบ เป็นการเปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยตามวิธีการตรวจให้คะแนนทั้ง 2 วิธี ผลการวิเคราะห์ที่ได้แสดงผล ดังตาราง 11

ตาราง 11 ฟังก์ชันสารสนเทศของข้อสอบเฉลี่ย จำแนกตามความสามารถ (θ) ของผู้สอบ และวิธีการตรวจให้คะแนน

ความสามารถ (θ)	วิธีการตรวจให้คะแนน	ฟังก์ชันสารสนเทศของข้อสอบเฉลี่ย				SE (θ)
		Min	Max	Mean	SD	
-3.00	วิธีการของการทดสอบความน่าจะเป็น	.002	.207	.048	.058	2.843
	วิธีการให้คะแนนตามความยากประจำชั้น	.002	.221	.060	.063	3.308
-2.00	วิธีการของการทดสอบความน่าจะเป็น	.002	.548	.077	.128	4.010
	วิธีการให้คะแนนตามความยากประจำชั้น	.002	.423	.082	.111	1.851
-1.00	วิธีการของการทดสอบความน่าจะเป็น	.002	.829	.097	.185	4.827
	วิธีการให้คะแนนตามความยากประจำชั้น	.002	.506	.092	.135	6.043
0.00	วิธีการของการทดสอบความน่าจะเป็น	.002	.792	.095	.180	4.778
	วิธีการให้คะแนนตามความยากประจำชั้น	.002	.493	.090	.132	1.940
1.00	วิธีการของการทดสอบความน่าจะเป็น	.002	.444	.072	.117	3.825
	วิธีการให้คะแนนตามความยากประจำชั้น	.002	.337	.074	.095	1.848
2.00	วิธีการของการทดสอบความน่าจะเป็น	.002	.188	.046	.053	2.760
	วิธีการให้คะแนนตามความยากประจำชั้น	.002	.163	.052	.047	1.816
3.00	วิธีการของการทดสอบความน่าจะเป็น	.002	.232	.034	.039	2.263
	วิธีการให้คะแนนตามความยากประจำชั้น	.002	.089	.036	.022	1.764

จากตาราง 11 พบว่า ค่าฟังก์ชันสารสนเทศของข้อสอบจำแนกตามค่าความสามารถของผู้สอบตามวิธีการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็น ให้ค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงที่สุดที่ระดับความสามารถ -1 มีค่าเท่ากับ .097 โดยเฉลี่ยสูงกว่าวิธีการตรวจให้คะแนนตามความยากประจำชั้น ที่มีค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงที่สุดในระดับความสามารถ -1 มีค่าเท่ากับ .092

3.3 การเปรียบเทียบประสิทธิภาพสัมพัทธ์: RE (θ) ณ ตำแหน่ง θ เดียวกัน

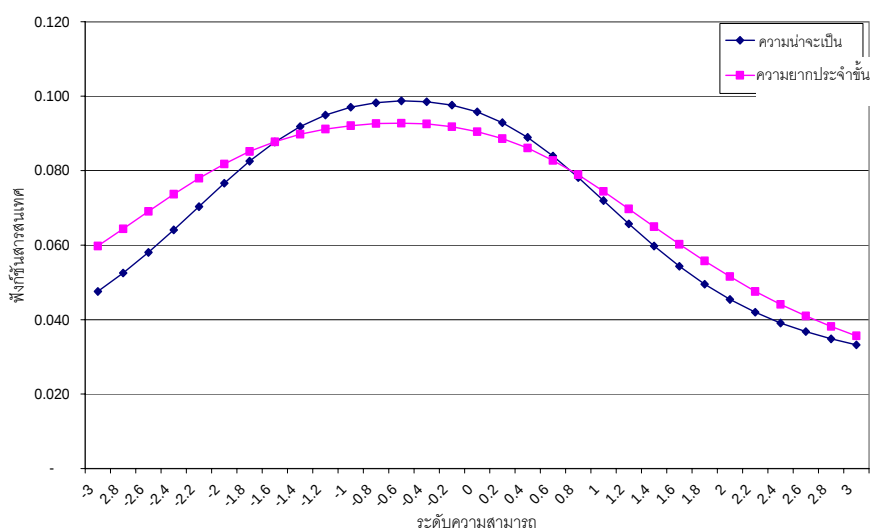
ตาราง 12 ผลการเปรียบเทียบประสิทธิภาพสัมพัทธ์ : RE (θ) ณ ตำแหน่ง θ เดียวกัน

ค่าความสามารถ (θ)	ค่าฟังก์ชันสารสนเทศของวิธีการตรวจให้คะแนน		ค่าประสิทธิภาพสัมพัทธ์ : RE(θ)
	วิธีการของการทดสอบ	การตรวจให้คะแนน	
	ความน่าจะเป็น	ตามความยากประจำชั้น	
-3.0	0.048	0.060	0.796
-2.8	0.053	0.064	0.816
-2.6	0.058	0.069	0.840
-2.4	0.064	0.074	0.870
-2.2	0.070	0.078	0.903
-2.0	0.077	0.082	0.937
-1.8	0.083	0.085	0.969
-1.6	0.088	0.088	0.999
-1.4	0.092	0.090	1.023
-1.2	0.095	0.091	1.041
-1.0	0.097	0.092	1.054
-0.8	0.098	0.093	1.060
-0.6	0.099	0.093	1.065
-0.4	0.099	0.093	1.064
-0.2	0.098	0.092	1.063
0	0.096	0.090	1.059
0.2	0.093	0.089	1.049
0.4	0.089	0.086	1.033
0.6	0.084	0.083	1.015

ตาราง 12 ผลการเปรียบเทียบประสิทธิภาพสัมพัทธ์ : $RE(\theta)$ ณ ตำแหน่ง θ เดียวกัน (ต่อ)

ค่าความ สามารถ (θ)	ค่าฟังก์ชันสารสนเทศของวิธีการตรวจให้คะแนน		ค่าประสิทธิภาพ สัมพัทธ์ : $RE(\theta)$
	วิธีการของการทดสอบ	การตรวจให้คะแนน	
	ความน่าจะเป็น	ตามความยากประจำชั้น	
0.8	0.078	0.079	0.990
1.0	0.072	0.074	0.967
1.2	0.066	0.070	0.943
1.4	0.060	0.065	0.920
1.6	0.054	0.060	0.902
1.8	0.050	0.056	0.888
2.0	0.045	0.052	0.881
2.2	0.042	0.048	0.883
2.4	0.039	0.044	0.887
2.6	0.037	0.041	0.897
2.8	0.035	0.038	0.913
3.0	0.033	0.036	0.931

โค้งสารสนเทศของข้อสอบของการให้คะแนน 2 วิธี



ภาพประกอบ 12 เปรียบเทียบโค้งสารสนเทศของข้อสอบของการให้คะแนน 2 วิธี

จากตาราง 12 และภาพประกอบ 12 พบว่า ค่าประสิทธิภาพสัมพัทธ์ (Relative Efficiency : $RE(\theta)$) ที่ระดับความสามารถ (θ) -1.4 ถึง 0.6 วิธีการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็นมีประสิทธิภาพสูงกว่าวิธีการตรวจให้คะแนนด้วยวิธีการให้คะแนนตามความยากประจำชั้น เหมาะสำหรับการใช้ตรวจให้คะแนนกับผู้สอบที่มีระดับความสามารถ (θ) -1.4 ถึง 0.6 มีค่าฟังก์ชันสารสนเทศของข้อสอบสูงสุดที่ระดับความสามารถ -0.6 มีค่าเท่ากับ .099 และที่ระดับความสามารถ (θ) -3.0 ถึง -1.6 และ 0.8 ถึง 3.0 วิธีการตรวจให้คะแนนตามความยากประจำชั้นมีประสิทธิภาพสูงกว่าวิธีการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็น เหมาะสำหรับการใช้ตรวจให้คะแนนกับผู้สอบที่มีระดับความสามารถ (θ) -3.0 ถึง -1.6 และ 0.8 ถึง 3.0 มีค่าฟังก์ชันสารสนเทศของข้อสอบสูงสุดที่ระดับความสามารถ -0.6 มีค่าเท่ากับ .093

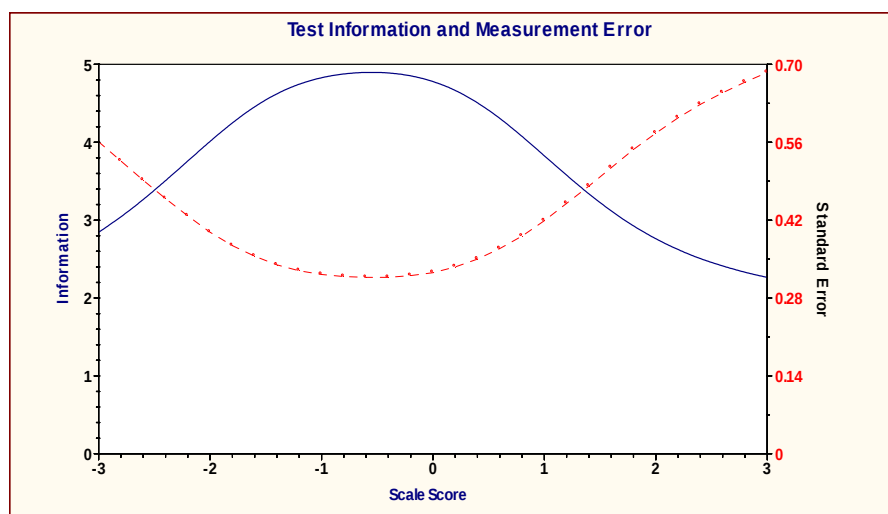
ค่าฟังก์ชันสารสนเทศของข้อสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ที่มีการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็นมีประสิทธิภาพ เท่ากับ 0.099 สูงกว่าวิธีการตรวจให้คะแนนตามความยากประจำชั้นมีประสิทธิภาพ เท่ากับ 0.093 ณ ตำแหน่งความสามารถ (θ) - 0.6

3.4 ผลการเปรียบเทียบฟังก์ชันสารสนเทศของแบบทดสอบ ตามวิธีการตรวจให้คะแนนที่แตกต่างกัน

ตาราง 13 ฟังก์ชันสารสนเทศของแบบทดสอบ (TIF) ตามวิธีการตรวจให้คะแนน

วิธีการตรวจให้คะแนนความรู้บางส่วน	จำนวนผู้สอบ	ฟังก์ชันสารสนเทศของแบบทดสอบ(TIF)	SE (θ)
วิธีการของการทดสอบความน่าจะเป็น	1007	4.895	0.452
วิธีการให้คะแนนตามความยากประจำชั้น	1007	4.630	0.464

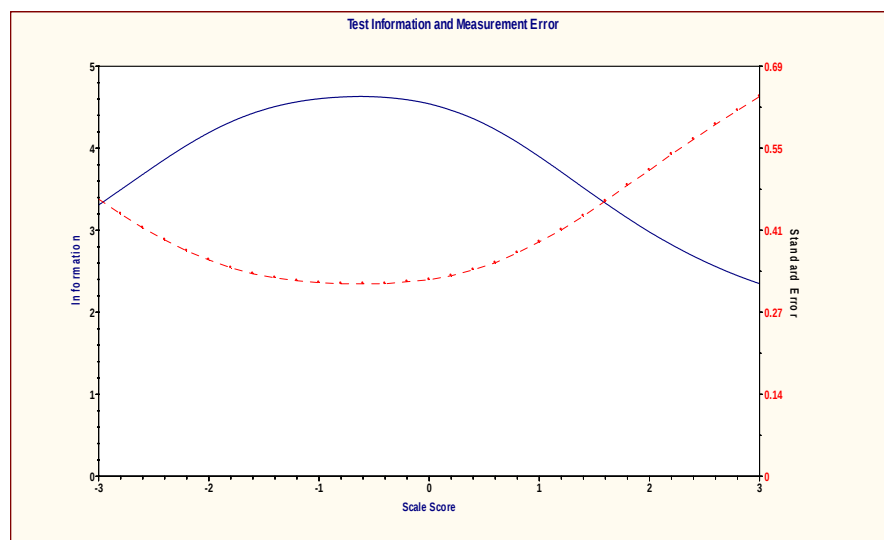
จากตาราง 13 พบว่า ฟังก์ชันสารสนเทศของแบบทดสอบตามวิธีการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็น มีค่าเท่ากับ 4.895 และค่าความคลาดเคลื่อนมาตรฐาน เท่ากับ 0.452 สูงกว่า วิธีการตรวจให้คะแนนตามความยากประจำชั้นให้ค่าฟังก์ชันสารสนเทศของแบบทดสอบ มีค่าเท่ากับ 4.630 และค่าความคลาดเคลื่อนมาตรฐาน เท่ากับ 0.464



เส้นทึบ : โค้งสารสนเทศของแบบทดสอบ อ่านค่าจากทางด้านซ้าย

เส้นประ : โค้งความคลาดเคลื่อนมาตรฐานของแบบทดสอบอ่านค่าจากทางด้านขวา

ภาพประกอบ 13 ฟังก์ชันสารสนเทศของแบบทดสอบ (TIF) จำแนกตามวิธีการตรวจให้คะแนน
ด้วยวิธีการของการทดสอบความน่าจะเป็น



เส้นทึบ : โค้งสารสนเทศของแบบทดสอบ อ่านค่าจากด้านซ้าย

เส้นประ : โค้งความคลาดเคลื่อนมาตรฐานของแบบทดสอบ อ่านค่าจากด้านขวา

ภาพประกอบ 14 ฟังก์ชันสารสนเทศของแบบทดสอบ (TIF) จำแนกตามวิธีการตรวจให้คะแนนด้วย
วิธีตามความยากประจำชั้น

จากภาพประกอบ 13 และ 14 พบว่าค่าสารสนเทศของการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็นและการตรวจให้คะแนนตามความยากประจำชั้นมีความผูกพันกับความคลาดเคลื่อนมาตรฐาน ณ ระดับความสามารถ (θ) -0.6

บทที่ 5

บทสรุป อภิปรายและข้อเสนอแนะ

การวิจัยครั้งนี้ ต้องการศึกษาวิธีการตรวจให้คะแนนความรู้บางส่วน จากการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็นกับวิธีการตรวจให้คะแนนตามความยากประจำชั้นของแบบทดสอบชุดเดียวกัน เพื่อเปรียบเทียบความเที่ยงตรงตามสภาพ (Concurrent Validity) ฟังก์ชันสารสนเทศของข้อสอบ (Item Information) และฟังก์ชันสารสนเทศของแบบทดสอบ (Test Information) จากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ชนิดเลือกตอบชั้นมัธยมศึกษาปีที่ 3 ที่ตรวจให้คะแนนความรู้บางส่วน ด้วยวิธีการของการทดสอบความน่าจะเป็นกับการตรวจให้คะแนนตามความยากประจำชั้น กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 2 ปีการศึกษา 2549 สำนักงานเขตพื้นที่การศึกษาปราจีนบุรี จำนวน 97 โรงเรียน จำนวน 199 ห้องเรียน จำนวนนักเรียนทั้งหมด 6,803 คน ซึ่งได้มาโดยการสุ่มแบบหลายขั้นตอน (Multi - stage random Sampling) โดยการแบ่งเขตพื้นที่การศึกษาปราจีนบุรีเป็นอำเภอ จำนวน 7 อำเภอ สุ่มได้จำนวน 3 อำเภอ 3,287 คน สุ่มได้นักเรียนจำนวน 593 คน แต่ในการวิเคราะห์ข้อมูลตามทฤษฎีตอบสนองข้อสอบ (Item Response Model) ซึ่งการประมาณค่าพารามิเตอร์ กลุ่มตัวอย่างต้องใหญ่เพียงพอที่จะทำให้ความคลาดเคลื่อนมาตรฐานของการประมาณค่าอยู่ในระดับที่ยอมรับได้ตามเป้าหมายของการนำไปใช้ (ศิริชัย กาญจนวาสี. 2545: 100) แล้วสุ่มแบบแบ่งชั้นตามขนาดของโรงเรียน คือโรงเรียนขนาดใหญ่ จำนวน 3 โรงเรียน ขนาดกลาง จำนวน 13 โรงเรียน ขนาดเล็ก จำนวน 28 โรงเรียน สุ่มได้โรงเรียนขนาดใหญ่ จำนวน 1 โรงเรียน ขนาดกลาง จำนวน 4 โรงเรียน ขนาดเล็ก จำนวน 6 โรงเรียน ได้กลุ่มตัวอย่าง จำนวน 1,007 คน มีขั้นตอนวิธีการดำเนินการวิจัยดังนี้

ขั้นที่ 1 การดำเนินการสร้างเครื่องมือเป็นแบบทดสอบวัดผลสัมฤทธิ์ทางคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 3 ผู้วิจัยดำเนินการตามลำดับ ดังนี้

1. ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการสร้างแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์เพื่อเป็นแนวทางการสร้างแบบทดสอบ
2. ศึกษาหลักสูตรคู่มือครูและแบบเรียนวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 3 เพื่อวิเคราะห์เนื้อหา จุดประสงค์การเรียนรู้ สาระทักษะการคิดคำนวณเกี่ยวกับการบวก การลบ การคูณ และการหาร จำนวนจริง เลขยกกำลัง
3. สร้างตารางวิเคราะห์เนื้อหา โดยกำหนดเนื้อหา จุดประสงค์เชิงพฤติกรรมให้สอดคล้องกับพฤติกรรมที่ต้องการวัด

4. สร้างข้อสอบตามเนื้อหาที่กำหนดไว้ ในด้านสาระทักษะกระบวนการคิดคำนวณ การบวก การลบ การคูณและการหาร จำนวนจริง เลขยกกำลัง ชนิดเลือกตอบ 4 ตัวเลือก จำนวน 60 ข้อ

ขั้นที่ 2 หาคุณภาพของแบบทดสอบ ดังนี้

1. นำแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ให้ผู้เชี่ยวชาญจำนวน 5 คน พิจารณาความสอดคล้องระหว่างข้อสอบกับเนื้อหาและให้ค่าน้ำหนักคะแนน แต่ละวิธีของการตรวจให้คะแนน ได้ค่าดัชนีความสอดคล้อง (IOC) โดยเฉลี่ย เท่ากับ .72 คัดเลือกข้อที่มีค่า IOC ตั้งแต่ .50 ขึ้นไป จำนวน 47 ข้อ

2. ปรับปรุงข้อสอบตามข้อเสนอแนะของผู้เชี่ยวชาญแล้วจัดพิมพ์

3. นำแบบทดสอบ จำนวน 47 ข้อ ไปทดสอบกับประชากรที่ไม่ใช่กลุ่มตัวอย่าง จำนวน 100 คน ประกอบด้วยนักเรียนชั้นมัธยมศึกษาปีที่ 3 โรงเรียนอนุบาลนาดี จำนวน 30 คน โรงเรียนทุ่งใหญ่วิทยาคาร จำนวน 50 คน และโรงเรียนเตรียมอุดมศึกษาน้อมเกล้าฯ กบินทร์บุรี จำนวน 20 คน ปีการศึกษา 2549 ระหว่าง วันที่ 2 – 10 ธันวาคม 2549

4. วิเคราะห์คุณภาพข้อสอบ เพื่อหาค่าความยากโดยพิจารณาค่าความยากจากค่าเฉลี่ยของคะแนนของข้อนั้นๆ ตามวิธีการให้คะแนนทั้งสองวิธี พบว่า มีค่าความยากรายข้อจากการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็น อยู่ระหว่าง 2.27 ถึง 3.67 เป็นความยากที่อยู่ในระดับปานกลางถึงง่าย และจากการตรวจให้คะแนนตามความยากประจำข้อ อยู่ระหว่าง 2.07 ถึง 3.97 เป็นความยากที่อยู่ในระดับปานกลางถึงง่าย

5. หาค่าอำนาจจำแนกจากวิธีการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็น มีค่าอำนาจจำแนกอยู่ระหว่าง ระหว่าง .21 ถึง .72 ค่าอำนาจจำแนกจากวิธีการตรวจให้คะแนนตามความยากประจำข้อ อยู่ระหว่าง .21 ถึง .91 และคัดเลือกข้อสอบจากวิธีการตรวจให้คะแนนทั้งสองวิธี ที่มีค่าอำนาจจำแนกตั้งแต่ .20 ขึ้นไป จำนวน 40 ข้อ

6. วิเคราะห์หาค่าความเชื่อมั่น จากสูตรสัมประสิทธิ์แอลฟา (α – Coefficient) ของครอนบัค ความเชื่อมั่นที่ได้จากการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็น เท่ากับ 0.92 และความเชื่อมั่นที่ได้จากการตรวจให้คะแนนตามความยากประจำข้อ เท่ากับ 0.91

ขั้นที่ 3 การเก็บรวบรวมข้อมูล

ดำเนินการเก็บรวบรวมข้อมูลการวิจัย ผู้วิจัยหรือตัวแทนผู้วิจัยใช้เวลาในการรวบรวมข้อมูล ระหว่างวันที่ 22 มกราคม – 16 กุมภาพันธ์ 2550 โดยใช้แบบทดสอบชุดเดียว จำนวน 40 ข้อ

การวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลที่ได้รับจากการสอบของนักเรียนกลุ่มตัวอย่าง ชั้นมัธยมศึกษาปีที่ 3 นำมาตรวจให้คะแนนตามวิธีการตรวจให้คะแนนทั้งสองวิธี แล้วนำคะแนนที่ได้ในแบบทดสอบมาทำการวิเคราะห์ข้อมูล ดังนี้

1. วิเคราะห์ค่าความเที่ยงตรงตามสภาพ จากค่าสัมประสิทธิ์สหสัมพันธ์ (r) และแปลงค่าความเที่ยงตรงตามสภาพจากค่าสัมประสิทธิ์สหสัมพันธ์ เป็นคะแนนพิชเชอร์ซี (Z) เปรียบเทียบความเป็นอิสระของความเที่ยงตรงสภาพในภาพรวม โดยใช้สถิติทดสอบไคสแควร์ (χ^2)

2. คำนวณค่าฟังก์ชันสารสนเทศของข้อสอบและแบบทดสอบ

3. เปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบทดสอบ ระหว่างวิธีการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็น กับวิธีการตรวจให้คะแนนตามความยากประจำชั้น

สรุปผลการวิจัย

1. ความเที่ยงตรงตามสภาพ (Concurrent Validity) ของแบบทดสอบวัดผลสัมฤทธิ์คณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ที่มีการตรวจให้คะแนนความรู้บางส่วน พิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์ (r) ระหว่างคะแนนสอบกลางภาคซึ่งเป็นคะแนนดิบของกลุ่มตัวอย่าง ในภาคเรียนที่สอง ปีการศึกษา 2549 กับคะแนนที่ได้มาจากการตรวจให้คะแนน 2 วิธี มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เมื่อเปรียบเทียบความแตกต่างโดยใช้คะแนนพิชเชอร์ซีเปรียบเทียบในภาพรวม โดยใช้สถิติทดสอบไคสแควร์ (χ^2) พบว่า ความเที่ยงตรงตามสภาพไม่ขึ้นอยู่กับวิธีการตรวจให้คะแนนในแต่ละวิธี

2. ฟังก์ชันสารสนเทศของข้อสอบ (IIF) วิธีการตรวจให้คะแนนความรู้บางส่วนด้วยวิธีการของการทดสอบความน่าจะเป็น จำแนกตามรายข้อ ข้อ 1, 2, 8, 14, 15, 16, 18, 25, 28, 30 และ 32 ให้ค่าฟังก์ชันสารสนเทศของข้อสอบสูงกว่าการตรวจให้คะแนนตามความยากประจำชั้น และฟังก์ชันสารสนเทศของข้อสอบ เมื่อตรวจให้คะแนนตามความยากประจำชั้นสูงกว่าวิธีการของการทดสอบความน่าจะเป็น ข้อที่ 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 17, 19, 20, 21, 22, 23, 24, 26, 27, 29, 31, 31, 33, 34, 35, 36, 37, 38, 39 และ 40

จำแนกตามค่าความสามารถ (θ) ของผู้สอบตามวิธีการให้คะแนน พบว่า วิธีการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็น ให้ค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงที่สุดในระดับความสามารถ -1 มีค่าเท่ากับ .097 โดยเฉลี่ยสูงกว่าวิธีการตรวจให้คะแนนตาม

ความยากประจำชั้น ที่มีค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงที่สุดในระดับความสามารถ -1 มีค่าเท่ากับ .092

ค่าฟังก์ชันสารสนเทศของข้อสอบที่มีการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็นมีประสิทธิภาพ เท่ากับ 0.099 สูงกว่าวิธีการตรวจให้คะแนนตามความยากประจำชั้นมีประสิทธิภาพ เท่ากับ 0.093 ณ ตำแหน่งความสามารถ (θ) สูงสุด - 0.6

3. ฟังก์ชันสารสนเทศของแบบทดสอบตามวิธีการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็น มีค่าเท่ากับ 4.895 และค่าความคลาดเคลื่อนมาตรฐาน เท่ากับ 0.452 สูงกว่า วิธีการตรวจให้คะแนนตามความยากประจำชั้นให้ค่าฟังก์ชันสารสนเทศของแบบทดสอบมีค่าเท่ากับ 4.630 และค่าความคลาดเคลื่อนมาตรฐาน เท่ากับ 0.464

อภิปรายผลการวิจัย

ในการวิจัยครั้งนี้ มีความมุ่งหมาย ในการเปรียบเทียบความเที่ยงตรงตามสภาพ (Concurrent Validity) และเปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบและแบบทดสอบของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 3 ที่มีวิธีการตรวจให้คะแนนความรู้บางส่วน

1. จากการศึกษาค่าความเที่ยงตรงตามสภาพ (Concurrent Validity) ของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ที่มีการตรวจให้คะแนนความรู้บางส่วน 2 วิธี คือ วิธีการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็นกับการตรวจให้คะแนนตามความยากประจำชั้น โดยพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ (r) ระหว่างคะแนนสอบกลางภาคซึ่งเป็นคะแนนดิบของกลุ่มตัวอย่าง ในภาคเรียนที่สอง ปีการศึกษา 2549 กับคะแนนที่ได้มาจากการตรวจให้คะแนน 2 วิธี มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เมื่อเปรียบเทียบความแตกต่างโดยใช้คะแนนฟิชเชอร์ซีเปรียบเทียบในภาพรวม โดยใช้สถิติทดสอบไคสแควร์ (χ^2) พบว่าความเที่ยงตรงตามสภาพไม่ขึ้นอยู่กับวิธีการตรวจให้คะแนนในแต่ละวิธี ทั้งนี้อาจเนื่องจากการตรวจให้คะแนนของทั้งสองวิธีมีการให้ค่าน้ำหนักคะแนนเท่ากันจึงทำให้ค่าความเที่ยงตรงตามสภาพของทั้งสองวิธีไม่แตกต่างกัน แต่อย่างไรก็ตาม พบว่า การให้คะแนนด้วยวิธีของการทดสอบความน่าจะเป็นให้ผลสูงกว่าการตรวจให้คะแนนตามความยากประจำชั้นเล็กน้อย ในแง่ของค่าเฉลี่ยของคะแนนที่ตรวจด้วยวิธีของการทดสอบความน่าจะเป็น เฉลี่ยเท่ากับ 109.99 สูงกว่าการตรวจให้คะแนนตามความยากประจำชั้นเล็กน้อย โดยเฉลี่ยเท่ากับ 95.61 ทั้งนี้อาจเป็นผลมาจากวิธีการให้คะแนนที่ตรวจด้วยวิธีของการทดสอบความน่าจะเป็น เป็นการให้คะแนนจากการตอบที่ผ่านกระบวนการคิดของผู้สอบตามลำดับขั้นของทักษะการคิดอย่างมีระบบในการใช้ความสามารถหาคำตอบแต่ละข้อที่มีการคิด

หลายขั้นตอนโดยที่ในขั้นที่สูงกว่าหรือขั้นที่ได้คะแนนมากกว่าต้องอาศัยการตอบในขั้นที่ต่ำกว่าหรือได้คะแนนน้อยกว่า ในขณะที่การตรวจให้คะแนนตามความยากประจำชั้นไม่อาศัยการตอบโดยอาศัยลำดับขั้นของการคิดแต่เป็นการหาคำตอบจากประโยคสัญลักษณ์แต่ละตอนมาคิดคำนวณในชั้นสามารถที่ตอบได้ แสดงให้เห็นถึงความเหมาะสมของวิธีการตรวจให้คะแนนความรู้บางส่วนแต่ละวิธี

2. พังค์ชันสารสนเทศของข้อสอบจำแนกตามค่าความสามารถของผู้สอบตามวิธีการให้คะแนน พบว่า ค่าพังค์ชันสารสนเทศของข้อสอบเฉลี่ยสูงที่สุดในระดับความสามารถ -1 ทั้งสองวิธีการตรวจให้คะแนน โดยที่วิธีการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็นมีค่าเท่ากับ .097 สูงกว่าวิธีการตรวจให้คะแนนตามความยากประจำชั้นที่มีค่าพังค์ชันสารสนเทศของข้อสอบเท่ากับ .092 และค่าประสิทธิภาพสัมพัทธ์ (Relative Efficiency : RE (θ)) ที่ระดับความสามารถ (θ) -1.4 ถึง 0.6 วิธีการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็นมีประสิทธิภาพสูงกว่าวิธีการตรวจให้คะแนนด้วยวิธีการให้คะแนนตามความยากประจำชั้น เหมาะสำหรับการใช้ตรวจให้คะแนนกับผู้สอบที่มีระดับความสามารถ (θ) -1.4 ถึง 0.6 และมีค่าพังค์ชันสารสนเทศของข้อสอบสูงสุดที่ระดับความสามารถ (θ) -0.6 มีค่าเท่ากับ .099 และที่ระดับความสามารถ (θ) -3.0 ถึง -1.6 และ 0.8 ถึง 3.0 วิธีการตรวจให้คะแนนตามความยากประจำชั้นมีประสิทธิภาพสูงกว่าวิธีการให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็น เหมาะสำหรับการใช้ตรวจให้คะแนนกับผู้สอบที่มีระดับความสามารถ (θ) -3.0 ถึง -1.6 และ 0.8 ถึง 3.0 มีค่าพังค์ชันสารสนเทศของข้อสอบสูงสุดที่ระดับความสามารถ -0.6 มีค่าเท่ากับ .093 จากผลการวิจัย พบว่า ค่าพังค์ชันสารสนเทศของวิธีการตรวจให้คะแนนของข้อสอบทั้งสองวิธีบอกถึงความสามารถ (θ) ของผู้สอบสูงสุดอยู่ในช่วง -1.4 ถึง 0.6 เป็นช่วงความสามารถที่ค่อนข้างต่ำ ดังนั้น การประเมินคุณภาพของแบบทดสอบที่ตรวจด้วยการให้คะแนนต่างกัน คือ วิธีการของการทดสอบความน่าจะเป็นและวิธีการให้คะแนนตามความยากประจำชั้น สามารถพิจารณาจากความถูกต้องแม่นยำในการประมาณความสามารถของผู้สอบ

3. พังค์ชันสารสนเทศของแบบทดสอบตามวิธีการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็น มีค่าเท่ากับ 4.895 และค่าความคลาดเคลื่อนมาตรฐาน เท่ากับ 0.452 สูงกว่า วิธีการตรวจให้คะแนนตามความยากประจำชั้นให้ค่าพังค์ชันสารสนเทศของแบบทดสอบมีค่าเท่ากับ 4.630 และค่าความคลาดเคลื่อนมาตรฐาน เท่ากับ 0.464 ณ ระดับความสามารถ (θ) -0.6 แสดงให้เห็นว่า วิธีการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็นกับวิธีการตรวจให้คะแนนตามความยากประจำชั้นไม่แตกต่างกัน สามารถเลือกใช้วิธีการตรวจให้คะแนนวิธีใดวิธีหนึ่งให้เหมาะสมกับจุดประสงค์การสอบและความสามารถผู้สอบได้

ผลจากการวิจัยในครั้งนี้พบว่า การตรวจให้คะแนนความรู้บางส่วน (Partial Credit Model: PCM) ด้วยวิธีการทดสอบความน่าจะเป็นกับวิธีการตรวจให้คะแนนตามความยากประจำชั้นพบว่า ทั้งสองวิธีให้ความเที่ยงตรงตามสภาพหรือความสามารถที่แท้จริง ฟังก์ชันสารสนเทศของแบบทดสอบ ฟังก์ชันสารสนเทศของข้อสอบ ไม่แตกต่างกัน และสามารถนำแบบทดสอบไปวินิจฉัยข้อบกพร่องของผู้เรียนได้

ข้อเสนอแนะ

ข้อเสนอแนะสำหรับการวิจัยครั้งต่อไป

1. การสร้างข้อสอบลักษณะเช่นนี้ สามารถนำไปพัฒนาเป็นแบบทดสอบวินิจฉัยผู้เรียนต่อไปได้

2. ควรนำข้อสอบที่มีคุณภาพมาวัดเพื่อควบคุมตัวแปรแทรกซ้อน

ข้อเสนอแนะสำหรับการพัฒนาการศึกษา

1. ควรมีการทดสอบด้วยการให้คะแนนความรู้บางส่วน เพื่อเป็นข้อมูลย้อนกลับในการจัดการเรียนรู้และตัดสินผลการเรียนรู้ของนักเรียน

2. ควรนำแบบทดสอบที่มีการตรวจให้คะแนนความรู้บางส่วนวินิจฉัยนักเรียนว่าเป็นผู้ที่มีความรู้ระดับใด นำไปสู่การพัฒนานักเรียนให้เกิดความรอบรู้

3. ครูผู้สอนสามารถเลือกใช้ข้อสอบที่เหมาะสมกับความสามารถของนักเรียนเพื่อไม่ให้ผู้สอบเกิดความเบื่อหน่าย เกิดการเดาในคำตอบที่ยากเกินไปและอาจเกิดความสับสนสำหรับผู้ที่มีความสามารถสูงก่อให้เกิดความผิดพลาดในการแปลผล

บรรณานุกรม

บรรณานุกรม

- กนกวรรณ รัตนธน. (2544). การเปรียบเทียบความตรงตามสภาพและความคงที่ในการประมาณค่าความสามารถของผู้สอบในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์เมื่อระดับความสามารถของผู้สอบและอัตราการใช้ข้อสอบซ้ำแตกต่างกัน. วิทยานิพนธ์ครุศาสตรมหาบัณฑิต. กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.
- กระทรวงศึกษาธิการและสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี. (2546). คู่มือวัดผลประเมินผลคณิตศาสตร์. กรุงเทพฯ: กระทรวงฯ.
- กาญจนา ศิริวัฒนาพงษ์. (2520). การศึกษาเปรียบเทียบวิธีการตอบและการตรวจให้คะแนนแบบสอบปรนัยชนิดเลือกตอบที่มีลักษณะแตกต่างกัน. วิทยานิพนธ์. กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ.
- คณิต ไช้มุกต์. (2534). การพัฒนาเกณฑ์ในการเลือกข้อสอบที่เหมาะสมกับความสามารถของผู้สอบ. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย.
- ฉวีวรรณ บุญมั่ง. (2540). ได้ศึกษาการเปรียบเทียบความสัมพันธ์ระหว่างคะแนนความสามารถจากการวิเคราะห์โดยทฤษฎีการตอบสนองข้อสอบกับคะแนนที่ได้จากการตอบและตรวจให้คะแนน 4 วิธี. วิทยานิพนธ์. สำนักวิทยบริการ มหาวิทยาลัยมหาสารคาม.
- ตริงใจ พูลผลอำนวย. (2534). การพัฒนาแบบสอบเฉพาะบุคคลในวิชาวิทยาศาสตร์สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3. วิทยานิพนธ์ ค.ด. (การทดสอบและการวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, อุดลำนาค.
- ธนวัฒน์ แสนสุข. (2538). การใช้จีอาร์เอ็ม จีพีซีเอ็มและโมเดลโลจิสติกในการเปรียบเทียบฟังก์ชันสารสนเทศของแบบวัดที่มีการให้คะแนนต่างกัน. วิทยานิพนธ์ ค.ม. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- ผจจจิต อินทสุวรรณ. (2545). ทฤษฎีการตอบสนองข้อคำถาม. กรุงเทพฯ: สถาบันวิจัยพฤติกรรมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.
- พรทิพย์ ไชยโส. (2534). การพัฒนาสูตรการให้คะแนนแบบสอบเลือกตอบสำหรับความรู้บางส่วนของผู้ตอบ: การประยุกต์ใช้วิธีการของอาร์โนลด์และวิธีการของแฮมมิง. วิทยานิพนธ์ ค.ด. (วิจัยการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.

- พินิจ อุไรรักษ์. (2533). ผลของวิธีการให้คะแนนที่มีต่อคะแนนสอบค่าสัมประสิทธิ์ ความเที่ยงและค่าสัมประสิทธิ์ความเที่ยง. วิทยานิพนธ์ ค.ม.(การวัดและการประเมินผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. อัดสำเนา.
- บุญเชิด ภิญญอนันตพงษ์. (2527). การทดสอบแบบอิงเกณฑ์ : แนวคิดและวิธีการ. กรุงเทพฯ: โอเดียนสโตร์.
- _____. (2545). “คุณภาพของเครื่องมือวัด” หน่วยที่ 3. ประมวลสาระชุดวิชาการพัฒนาเครื่องมือสำหรับการประเมินการศึกษา. นนทบุรี: มหาวิทยาลัยสุโขทัยธรรมาธิราช.
- ปรมินทร์ อริเดช. (2547). การเปรียบเทียบฟังก์ชันสารสนเทศในการใช้ โมเดลโลจิสติก จีอาร์เอ็ม และจีพีซีเอ็มของมาตรวัดเจตคติแบบลิเคิร์ตและมาตรวัดแบบตัวเลือกบังคับตอบที่มีการให้คะแนนแบบสองค่าและแบบหลายค่า. วิทยานิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- มยุรี ศรีชัย. (2538). เทคนิคการสุ่มตัวอย่าง. กรุงเทพฯ ฯ หจก. วิเจ พรีนติ้ง.
- ศิริชัย กาญจนวาสี. (2545). ทฤษฎีการทดสอบแนวใหม่. กรุงเทพฯ ฯ คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.
- _____. (2550). ทฤษฎีการทดสอบแนวใหม่. พิมพ์ครั้งที่ 3 . กรุงเทพฯ ฯ คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.
- สุนิสา จัยม่วงศรี. (2546). การศึกษาผลของการเทียบคะแนนข้อสอบที่ตรวจให้คะแนนแบบหลายค่า. วิทยานิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- สุรัช มีชาญ. (2539). การเปรียบเทียบความยากประจำชั้นและประสิทธิภาพในการประมาณค่าเจตคติของมาตรวัดเจตคติแบบลิเคิร์ตที่มีรูปแบบของการตอบจำนวนของลำดับขั้นและทิศทางของการเรียงลำดับขั้นแตกต่างกันด้วยพาเซี่ยลเครดิติ โมเดลของมาสเตอร์. วิทยานิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ.
- สุพัฒน์ สุขมลสันต์. (2530, กันยายน - ธันวาคม). "นานาปัจจัยที่ส่งผลต่อค่าความเชื่อมั่นและค่าความเที่ยงตรงของแบบทดสอบ." วิทยาทวารวิจัย. 2 (3): 52 - 62.
- สำนักงานเขตพื้นที่การศึกษาปราจีนบุรี. (2549). จำนวนข้อมูลนักเรียนรายบุคคล(OBEC_SMIS) ประจำปีการศึกษา 2549. ปราจีนบุรี: กลุ่มนโยบายและแผน. สำนักงาน.
- ลำเจิง บุญเรืองรัตน์. (2529, มกราคม - เมษายน). “การพัฒนาทฤษฎีลาเท้นเทรทเพื่อการวิเคราะห์ข้อสอบ” วารสารการวัดผลการศึกษา. 7(21): 41 - 66.

- สมถวิล วิจิตรวรรณ. (2542, พฤษภาคม - สิงหาคม). "การประมาณค่าพารามิเตอร์ในทฤษฎีการตอบสนองข้อสอบ" *วารสารการวัดการศึกษา*. 20 (58): 21 - 27.
- ล้วน สายยศและอังคณา สายยศ. (2540). *สถิติวิทยาการวิจัย*. พิมพ์ครั้งที่ 3. กรุงเทพฯ: สุวีริยาสาส์น.
- _____. (2543). *การวัดด้านจิตพิสัย*. กรุงเทพฯ: สุวีริยาสาส์น.
- อนันต์ ศรีโสภ. (2532). *การทดลองรูปแบบใหม่ของข้อสอบแบบเลือกตอบ*. มหาวิทยาลัยศรีนครินทรวิโรฒ.
- องอาจ นัยพัฒน์. (2535). *การศึกษาประสิทธิภาพของผลการประมาณค่าพารามิเตอร์ของพาเชียลเครดิต โมเดล (CREDIT) ระหว่างวิธีพรีออกซ์ (PROX) และวิธียูคอน (UCON) ในแบบทดสอบวัดเจตคติแบบลิเคิร์ทสเกล*. ปรินญานิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- เอมอร จังศิริพรปกรณ์. (2544, พฤษภาคม - สิงหาคม). "การให้คะแนนความรู้บางส่วนทำได้อย่างไร" *วารสารการวัดผลการศึกษา*. 23 (67): 21 - 31.
- _____. (2546). *การเปรียบเทียบคุณภาพของแบบเลือกตอบเมื่อตรวจด้วยวิธีการให้คะแนนความรู้บางส่วนกับวิธีประเพณีนิยม*. กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.
- อุทัยวรรณ สายพัฒน์. (2547). *การเปรียบเทียบประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่าระหว่างวิธี GMH และวิธี Polytomous SIBTEST*. ปรินญานิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- Albanese, Mark A., Thomas H. Kent and Douglas R. Whitnay. (1979, December). "Cluing in the multiple – choice test items with combinations of correct response." *Journal of Medical Education*. 54: 945 - 950.
- Albert, Wilfred Georgrn. (1970, October). "Scoring for partial Knowledge in mathematics testing; A study of modification and extension of multiple - choice item applied to the testing of achievement in mathematics," *Dessertation Abstracts International*. 31(4): 1619A- 1620A.
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561- 573.
- Bock, R.D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37.

- Coombs, C.H., Milholland, J.E. and Womer, F.B. (1956). The assessment of partial knowledge. *Education and Psychological Measurement*, 16, 13 - 37.
- David Thissen, Wen-Hung Chen and Darrell Bock. (2003). *MULTILOG 7.03 - Analysis of Multiple -Category Response Data*. Scientific Software International, Inc. Lincolnwood, USA.
- Davis Frederick B. (1966). *Education Measurement and Their Interpretation*. Belmont: Wadsworth Publishing Co.
- Davis Frederick B. and Gordon Fifer. (1959, summer). "The effect on the test reliability and validity of scoring aptitude and achievement tests with weight 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.
- Dressel, P.L. and John Schmidt. (1953). "Some Modification of the Multiple – Choice Item," *Education and Psychological Measurement*. 13; 574 - 595.
- Dodd, B.G., De Ayala, R.J. & Koch, W.R. (1995). Computerized adaptive testing with polytomous items. *Applied Psychological Measurement*, 19, 5 - 22.
- Ebel, R.L. (1972). *Essential of Education Measurement*. 2nd. Englewood Cliffs.
- Echternach, R.M. (1972). "A Comparison of Five Different Scoring Functions For Confidence Test," *Journal of Education Measurement*. 7: 217 - 221.
- Frary, R.B. (1989). Partial-credit scoring methods for multiple-choice tests. *Applied Measurement in Education*, 2, 79 - 96.
- Gibson, J.d., Olkin, I. & Sobel, M. (1997). A subset selection technique for scoring items on a multiple choice test. *Psychometrika*, 44, 259 - 270.
- Glass, G.V. and Stanley, J.C. (1970). *Statistical Methods in Education and Psychology*. Englewood Cliffs, NJ.: Prentice – Hall.
- Gronlund, N.E. (1982). *Constructing achievement tests*. 3rd ed. Englewood Cliffs, NJ: Prentice-Hall.
- Gulliksen, H. (1986). Perspective on educational measurement. *Applied Psychological Measurement*, 10, 109 - 230.

- Hambleton, R.K., Robert, D.M. and 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. and Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*, Kluwer/Nijhoff Publishing, Boston.
- Hambleton, R.K., Swaminathan, H. & Roger, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park: Sage.
- 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.
- Michael T. Brannick. (2007). *Item Response Theory*. <http://luna.cas.usf.edu/~mbrannic/files/pmet/irt.htm>. Retrieved June 21, 2006.
- John Behrens. (2007). *The idea of using Data Desk to visualize ICC graphs is originated*. <http://www.creative-wisdom.com/computer/sas/IRT.html>. Retrieved June 21, 2006.
- Lord, F.M. (1975). Formula scoring and number right scoring. *Journal of Educational Measurement*, 12, 7 - 11.
- Lord, F.M. (1980). *Applications of item response theory to practical testing problems*. New Jersey: Lawrence Erlbaum Associates, Inc.
- Lord, F.M. and Novick, M.R. (1968). *Statistical Theories of Mental Test Scores*, Addison - Wesley. Publishing Company. Reading, Massachusetts.
- Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149 -174.
- Michael, J.C. (1968). The reliability of a multiple choice examination under various test-taking instructions. *Journal of Educational Measurement*, 5, 307 - 314.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 14, 59 - 71.
- Muraki, E. (1993). Information functions of the generalized partial credit model. *Applied Psychological Measurement*, 17, 351-363.

- Samejima, F. (1996). Graded response model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory*. 85 -100. New York: Springer.
- Simon,A.B., Budscu,D.V. and Nevo,H.B. (1997). A comparative study of measurement procedure. *Psychometrika*.
- Smith,R.M. (1987). Assessing partial knowledge in vocabulary. *Journal of Educationnal Measuerment*. 24, 217 - 231.
- Stiggins, Richard J. (1994). *Student - Centered Classroom Assessment*. New york: Maxwell Macmillan College Publishing Company.
- Traub,R.D. and Hambleton,R.K. (1972). *The Effectof scoring instructions and degree of speededness on the validity and reliability of multiple choice test*. Education and Psychological Measurement, 29, 2847 - 861.
- Wallsten, T.S,. Budescu, D.V., and Zwick, R. (1993). Comparing the calibration and coherence of numerical and verbal probabilistic judgments. *Management Science*, 39, 176 -190
- Wang, M.D. and Stanley,J.C. (1970). Differential weighting: A review of methods and empirical studies. *Review of Educational Research*.
- Wright, B.D. and Masters, G.N. (1982). *Rating Scale Analysis*. Chicago: MESA Press.

ภาคผนวก

ภาคผนวก ก
คุณภาพของเครื่องมือที่ใช้ในการวิจัย

ตาราง 14 ผลการตรวจสอบความเที่ยงตรงด้านเนื้อหา (Content Validity) โดยผู้เชี่ยวชาญ

ข้อ	ผู้เชี่ยวชาญคนที่					IOC	การคัดเลือก
	1	2	3	4	5		
1	1	1	1	1	1	1	คัดเลือก
2	0	-1	0	0	1	0	ไม่ใช่
3	1	0	0	1	1	0.6	คัดเลือก
4	1	1	0	1	1	0.8	คัดเลือก
5	-1	-1	0	0	0	-0.4	ไม่ใช่
6	1	1	0	1	1	0.8	คัดเลือก
7	0	-1	-1	1	1	0	ไม่ใช่
8	1	1	1	0	1	0.8	คัดเลือก
9	1	1	1	1	1	1	คัดเลือก
10	1	0	0	1	0	0.4	ไม่ใช่
11	1	1	1	1	1	1	คัดเลือก
12	1	0	0	0	1	0.4	ไม่ใช่
13	1	1	0	1	1	0.8	คัดเลือก
14	0	1	0	0	1	0.4	ไม่ใช่
15	1	1	1	1	1	1	คัดเลือก
16	1	1	1	1	1	1	คัดเลือก
17	0	1	1	1	1	0.8	คัดเลือก
18	1	1	1	1	1	1	คัดเลือก
19	1	0	1	1	1	0.8	คัดเลือก
20	1	0	0	1	0	0.4	ไม่ใช่
21	1	1	1	1	1	1	คัดเลือก
20	0	1	0	1	1	0.6	คัดเลือก
23	1	0	1	0	0	0.4	ไม่ใช่
24	1	1	0	0	0	0.4	ไม่ใช่
25	0	0	0	0	1	0.2	ไม่ใช่

ตาราง 14 ผลการตรวจสอบความเที่ยงตรงด้านเนื้อหา (Content Validity) โดยผู้เชี่ยวชาญ (ต่อ)

ข้อ	ผู้เชี่ยวชาญคนที่					IOC	การคัดเลือก
	1	2	3	4	5		
26	1	0	0	1	1	0.6	คัดเลือก
27	1	1	1	1	1	1	คัดเลือก
28	1	1	1	1	1	1	คัดเลือก
29	1	0	1	1	1	0.8	คัดเลือก
30	0	0	0	1	-1	0	ไม่ใช่
31	0	1	1	1	0	0.6	คัดเลือก
32	1	1	1	1	1	1	คัดเลือก
33	1	1	1	1	1	1	คัดเลือก
34	1	1	1	1	1	1	คัดเลือก
35	1	0	1	1	1	0.8	คัดเลือก
36	0	1	1	1	1	0.8	คัดเลือก
37	0	1	0	0	1	0.4	ไม่ใช่
38	0	1	1	1	1	0.8	คัดเลือก
39	1	1	1	1	1	1	คัดเลือก
40	1	0	1	1	1	0.8	คัดเลือก
41	1	1	1	1	1	1	คัดเลือก
42	1	1	1	1	1	1	คัดเลือก
43	1	1	1	1	1	1	คัดเลือก
44	0	1	1	1	0	0.6	คัดเลือก
45	0	1	1	1	0	0.6	คัดเลือก
46	1	1	1	1	1	1	คัดเลือก
47	1	0	1	1	1	0.8	คัดเลือก
48	1	1	1	1	1	1	คัดเลือก
49	1	1	1	1	1	1	คัดเลือก
50	1	0	1	1	1	0.8	คัดเลือก

ตาราง 14 ผลการตรวจสอบความเที่ยงตรงด้านเนื้อหา (Content Validity) โดยผู้เชี่ยวชาญ (ต่อ)

ข้อ	ผู้เชี่ยวชาญคนที่					IOC	การคัดเลือก
	1	2	3	4	5		
51	0	1	1	1	1	0.8	คัดเลือก
52	0	0	1	1	1	0.6	คัดเลือก
53	0	1	0	0	1	0.4	ไม่ใช่
54	1	1	0	0	1	0.6	คัดเลือก
55	1	0	1	1	1	0.8	คัดเลือก
56	1	1	1	1	1	1	คัดเลือก
57	1	1	1	1	1	1	คัดเลือก
58	1	0	1	0	1	0.6	คัดเลือก
59	1	1	1	1	1	1	คัดเลือก
60	1	1	1	1	1	1	คัดเลือก

ตาราง 15 ค่าความยาก ค่าอำนาจจำแนกรายข้อ ตามวิธีการตรวจให้คะแนน

ข้อ	ค่าความยาก		ค่าอำนาจจำแนก		ผลการ คัดเลือก
	วิธีการของความ	ตามความยาก	วิธีการของความ	ตามความ	
	น่าจะเป็น	ประจำชั้น	น่าจะเป็น	ยากประจำชั้น	
1	2.67	3.00	0.54	0.29	คัดเลือก
2	2.87	2.44	0.58	0.51	คัดเลือก
3	2.27	2.99	0.41	0.54	คัดเลือก
4	3.97	2.93	0.39	-0.14	ไม่ใช้
5	2.41	2.50	0.49	0.46	คัดเลือก
6	3.17	3.67	0.19	-0.09	ไม่ใช้
7	3.53	3.07	0.21	0.27	คัดเลือก
8	2.63	2.91	0.35	0.44	คัดเลือก
9	2.64	2.59	0.61	0.26	คัดเลือก
10	3.03	3.06	0.38	0.11	ไม่ใช้
11	2.63	2.75	0.51	0.21	คัดเลือก
12	2.41	2.80	0.49	0.46	คัดเลือก
13	3.11	2.72	0.21	0.23	คัดเลือก
14	2.88	2.27	0.28	0.09	ไม่ใช้
15	2.96	3.21	0.56	0.49	คัดเลือก
16	2.69	2.56	0.55	0.55	คัดเลือก
17	2.22	2.67	0.49	0.52	คัดเลือก
18	2.66	2.85	0.51	0.23	คัดเลือก
19	2.27	2.86	0.43	0.46	คัดเลือก
20	2.18	2.62	0.54	0.86	คัดเลือก
21	2.16	2.96	0.44	0.53	คัดเลือก
22	2.53	2.56	0.56	0.41	คัดเลือก
23	2.70	3.37	0.29	0.41	คัดเลือก
24	3.86	2.67	0.59	0.91	คัดเลือก

ตาราง 15 (ต่อ) ค่าความยาก ค่าอำนาจจำแนกรายข้อ ตามวิธีการตรวจให้คะแนน

ข้อ	ค่าความยาก		ค่าอำนาจจำแนก		ผลการ คัดเลือก
	วิธีการของความ น่าจะเป็น	ตามความยาก ประจำชั้น	วิธีการของความ น่าจะเป็น	ตามความ ยากประจำชั้น	
25	2.34	2.81	0.52	0.50	คัดเลือก
26	2.22	2.52	0.49	0.44	คัดเลือก
27	2.51	2.79	0.72	0.44	คัดเลือก
28	2.89	2.94	0.41	0.65	คัดเลือก
29	2.07	2.77	0.56	0.66	คัดเลือก
30	2.14	3.01	0.44	0.35	คัดเลือก
31	2.99	2.61	0.47	0.30	คัดเลือก
32	2.44	2.73	0.57	0.58	คัดเลือก
33	3.97	2.79	0.53	0.49	คัดเลือก
34	2.40	3.11	0.54	0.17	ไม่ใช้
35	2.63	2.65	0.49	0.45	คัดเลือก
36	2.70	2.77	0.49	-0.14	ไม่ใช้
37	3.86	2.92	0.30	0.52	คัดเลือก
38	2.46	2.95	0.48	0.49	คัดเลือก
39	2.39	2.98	0.36	0.41	คัดเลือก
40	2.34	3.38	0.29	-0.03	ไม่ใช้
41	2.22	2.69	0.31	0.44	คัดเลือก
42	2.39	2.79	0.29	0.44	คัดเลือก
43	2.34	2.86	0.26	0.65	คัดเลือก
44	2.32	2.48	0.50	0.66	คัดเลือก
45	2.34	2.81	0.45	0.44	คัดเลือก
46	2.22	2.52	0.39	0.65	คัดเลือก
47	2.51	2.79	0.63	0.64	คัดเลือก

ภาคผนวก ข
ตัวอย่างเครื่องมือที่ใช้ในการวิจัย

แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3
ทักษะการคิดคำนวณ การบวก การลบ การคูณและการหาร จำนวนจริง เลขยกกำลัง
กลุ่มสาระการเรียนรู้คณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3

คำชี้แจง

1. แบบทดสอบฉบับนี้มีจำนวน 40 ข้อ ใช้เวลา 60 นาที
2. แบบทดสอบฉบับนี้เป็นแบบทดสอบแบบเลือกตอบ ชนิด 4 ตัวเลือก
3. ให้นักเรียนเขียนเครื่องหมายกากบาท (X) ลงในช่องที่ตรงกับตัวเลือกที่สามารถคิดคำนวณได้อย่างมั่นใจที่สุดเพียงข้อเดียว

ตัวอย่างแบบทดสอบ

$$(0) \sqrt{(9/0.3)^2 - 175} = ?$$

เฉลย

ตอบ	ได้คะแนนจากวิธีการตรวจให้คะแนนของ การทดสอบความน่าจะเป็น	ได้คะแนนจากวิธีการตรวจให้คะแนน ตามความยากประจำชั้น
ก. 900	2 คะแนน เป็นคำตอบที่ใช้ลำดับขั้นของ กระบวนการหาคำตอบในขั้นที่ 1-2	3 คะแนน เป็นทักษะการคิดในชั้นยากมาก
ข. 625	3 คะแนน เป็นคำตอบที่ใช้ลำดับขั้นของ กระบวนการหาคำตอบในขั้นที่ 1-3	2 คะแนน เป็นทักษะการคิดในชั้นยากปานกลาง
ค. 30	1 คะแนน เป็นคำตอบที่ใช้ลำดับขั้นของ กระบวนการหาคำตอบในขั้นที่ 1	1 คะแนน เป็นทักษะการคิดในชั้นยากน้อย
ง. 25	4 คะแนน เป็นคำตอบที่ใช้ลำดับขั้นของ กระบวนการหาคำตอบในขั้นที่ 1-4	4 คะแนน เป็นทักษะการคิดในชั้นยากมากที่สุด

1. $\frac{(\sqrt{200} \times \sqrt{8}) - 60}{4} = ?$

ก. 40

ข. -20

ค. $\sqrt{1,600}$

ง. -5

2. $\frac{(-2)^2 \times (-4)^4}{8} = ?$

ก. 256

ข. 4

ค. 128

ง. 1,024

3. $\frac{4^2}{5^3} + \frac{4^3}{5^3} = ?$

ก. $\frac{16}{125}$

ข. $\frac{64}{125}$

ค. $\frac{80}{125}$

ง. $\frac{16}{25}$

4. $\frac{2^3 + 5^2}{3^2} - \frac{3}{(6^2 - 2^5)^2} = ?$

ก. $\frac{3}{16}$

ข. $\frac{501}{144}$

ค. $\frac{33}{9}$

ง. $3\frac{23}{48}$

5. $\frac{(-1)^3 (-2)^3}{(2^4)(-1)^2} \times 3 = ?$

ก. 16

ข. 8

ค. $1\frac{1}{2}$

ง. $\frac{24}{16}$

6. $\frac{(2 \times 5)^3}{2^3} = ?$

ก. 125

ข. 10^3

ค. 1,000

ง. 8

7. $\left(\frac{-3^3}{5^3}\right)\left(\frac{-5}{3}\right)^2 + \frac{3}{5} = ?$

ก. $\frac{25}{9}$

ข. 0

ค. $-\frac{3}{5}$

ง. $\frac{-27}{125}$

8. $\left(\frac{-4}{5}\right)^2 \times 10^3 = ?$

ก. 1,000

ข. 40

ค. $\frac{16}{25}$

ง. 640

9. $\left(3\sqrt{\frac{18000}{20}}\right) + 100 = ?$

ก. $\sqrt{900}$

ข. 190

ค. 90

ง. 30

10. $\frac{\sqrt{12} \times \sqrt{48}}{6} = ?$

ก. 4

ข. 24

ค. $4\sqrt{3}$

ง. $2\sqrt{3}$

11. $\left(\frac{224 \times 2}{56}\right) \div \left(\frac{20}{3} \times \frac{4}{5}\right) = ?$

ก. 8

ข. $\frac{3}{2}$

ค. $1\frac{1}{2}$

ง. $\frac{16}{3}$

12. $36 + \left(264 \times \frac{7}{22} \times \frac{1}{6}\right) = ?$

ก. 2

ข. 50

ค. 12

ง. 14

13. $(2\sqrt{108} - \sqrt{27}) \times \sqrt{3} = ?$

ก. $3\sqrt{3}$

ข. $9\sqrt{3}$

ค. 27

ง. $12\sqrt{3}$

14. $\left(\frac{1}{2} - \frac{1}{4}\right) \div \left(\frac{3}{5} \times 3\frac{1}{8}\right) = ?$

ก. $\frac{2}{4}$

ข. $\frac{1}{4}$

ค. $\frac{15}{8}$

ง. $\frac{2}{15}$

15. $\sqrt{25^2 - 7^2} = ?$

ก. $\sqrt{576}$

ข. $\sqrt{49}$

ค. $\sqrt{625}$

ง. 24

16. $(125)^{\frac{2}{3}} + 64$

ก. 5^3

ข. 89

ค. 5^2

ง. 25

17. $(625)^{\frac{3}{4}} + 5^2 = ?$

ก. 5^4

ข. 25

ค. 150

ง. 125

18. $(64)^{\frac{5}{6}} \times 8 = ?$

ก. 2^3

ข. 2^6

ค. 256

ง. 2^3

19. $\left(\frac{69 \times 25}{3}\right) - \frac{108}{3} = ?$

ก. 575

ข. 419

ค. 23

ง. 156

20. $\sqrt[3]{8^4} \times \sqrt[6]{8^4} = ?$

ก. 64

ข. 2^2

ค. 2^6

ง. 2^4

21. $(64)^{\frac{5}{6}} \times (3^4)^{\frac{3}{4}} = ?$

ก. 2^5

ข. 864

ค. 27

ง. 32

22. $\frac{(625)^{\frac{1}{2}}}{125} = ?$

ก. 5^2

ข. $\frac{1}{5}$

ค. 5^{-1}

ง. 25

23. $\sqrt[3]{\frac{250 \times (0.1)^3}{2}} = ?$

ก. 5^3

ข. $\sqrt[3]{0.5^3}$

ค. 0.5

ง. 125

24. $\left(\frac{1}{2} \times \frac{48}{\sqrt{3}}\right)^2 = ?$

ก. 576

ข. 192

ค. 3

ง. $\frac{24^2}{\sqrt{3}^2}$

25. $\left(\frac{75}{125} + \frac{100}{125}\right) \times 45 + 845 = ?$

ก. $\frac{175}{125}$

ข. 63

ค. $\frac{7}{5}$

ง. 908

26. $\sqrt{0.0003} \times 25 = ?$

ก. $\sqrt{3 \times 10^{-4}}$

ข. $\frac{\sqrt{3}}{100}$

ค. $\frac{\sqrt{3}}{4}$

ง. $\frac{\sqrt{3}}{\sqrt{100}^2}$

27. $-\sqrt{25} \times \frac{96}{108} = ?$

ก. -5

ข. $-4\frac{4}{9}$

ค. $-\frac{40}{9}$

ง. $\frac{8}{9}$

28. $\sqrt{(-2)^4} \times \sqrt{3^{-2}} = ?$

ก. 4

ข. $\sqrt{16}$

ค. $\frac{1}{3}$

ง. $1\frac{1}{3}$

29. $\frac{\sqrt{300}}{\sqrt{3}} \times 5^2 = ?$

ก. 10

ข. 25

ค. $\sqrt{100}$

ง. 250

30. $\frac{3\sqrt{18000}}{\sqrt{20}} \times \frac{1}{9} = ?$

ก. $\sqrt{900}$

ข. 90

ค. 10

ง. 30

31. $\sqrt{\frac{25}{81}} \times 36 = ?$

ก. $3\frac{1}{3}$

ข. 6

ค. $\frac{5}{9}$

ง. $\frac{30}{9}$

32. $\left(1\frac{5}{6} + \frac{9}{5}\right) \div \frac{6}{5} = ?$

ก. $\frac{11}{16}$

ข. $\frac{109}{36}$

ค. $3\frac{1}{36}$

ง. $\frac{109}{30}$

33. $\left(\frac{2}{3} - \frac{3}{15}\right) \times 7^{-1} = ?$

ก. $\frac{1}{15}$

ข. $\frac{1}{7}$

ค. $\frac{7}{15}$

ง. $\frac{10}{15}$

34. $\frac{1}{2} \times (65 + 95) \times \sqrt{81} = ?$

ก. 720

ข. 9

ค. 160

ง. 80

35. $5\sqrt{27} - 7\sqrt{12} + 3\sqrt{48} = ?$

ก. $12\sqrt{3}$

ข. $13\sqrt{3}$

ค. $14\sqrt{3}$

ง. $15\sqrt{3}$

36. $(2^2)^4 + 3^6 = ?$

ก. 985

ข. 2^8

ค. 729

ง. 256

37. $\left(51 \times \frac{5}{17}\right)^2 - \sqrt{64} = ?$

ก. 225

ข. 8

ค. 217

ง. 15

38. $\left(10 - \frac{8}{4}\right)^2 + 314 = ?$

ก. 8^2 ข. $\frac{40}{4}$

ค. 378

ง. 64

39. $\left(1\frac{1}{3} \times 3^3\right) \times \frac{5}{9} = ?$

ก. 20

ข. 36

ค. 27

ง. $\frac{4}{3}$

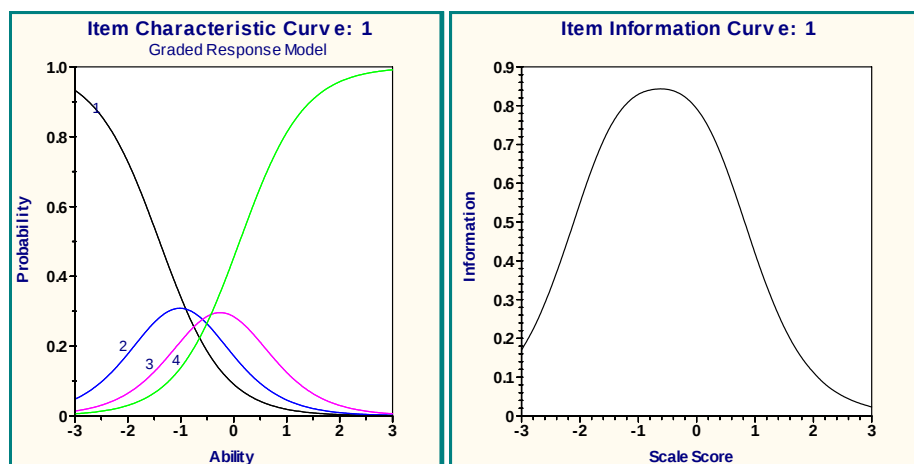
40. $\frac{(265 - 256)^2}{2^4} \times \frac{24}{108} = ?$

ก. 9^2 ข. $\frac{81}{16}$ ค. $1\frac{1}{8}$ ง. $\frac{9}{8}$

ภาคผนวก ค

โค้งลักษณะข้อสอบและสารสนเทศข้อสอบที่มีการตรวจให้คะแนนความรู้บางส่วน

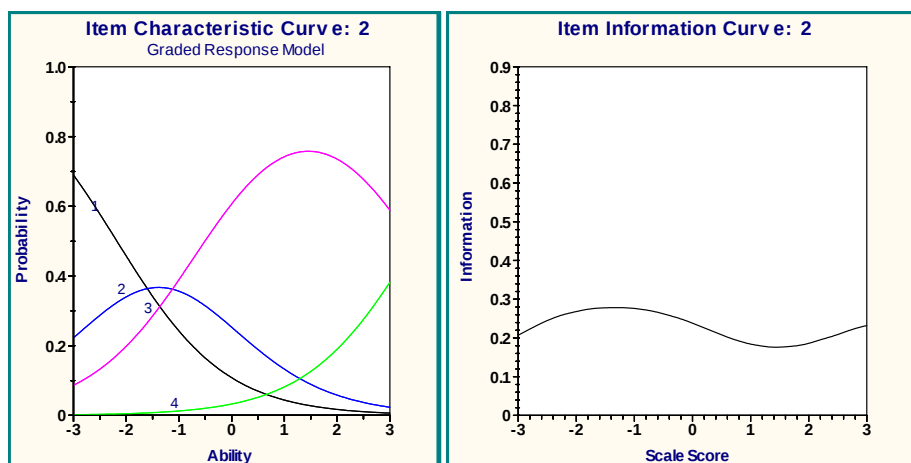
ภาพประกอบ 15 โค้งลักษณะข้อสอบและสารสนเทศข้อสอบที่มีการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็น



Category legends

Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

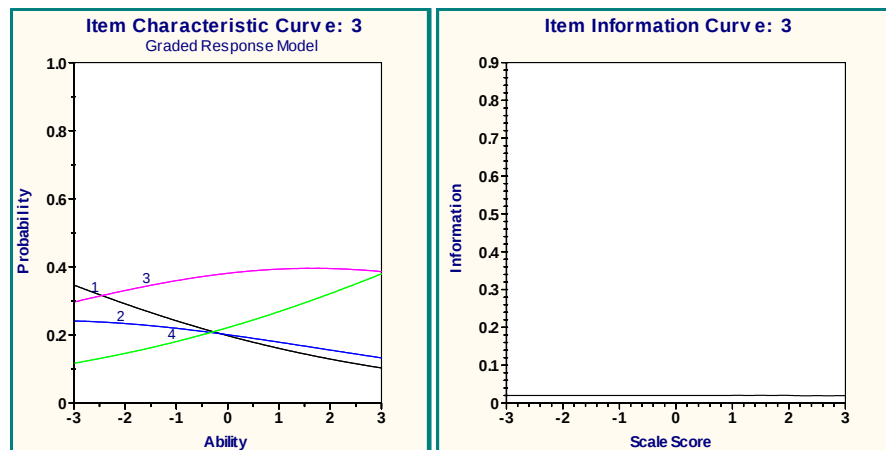
Item: 1



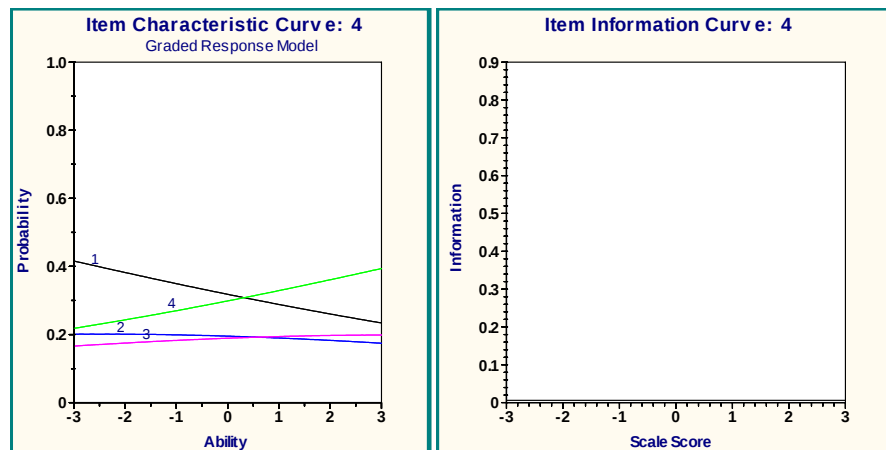
Category legends

Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

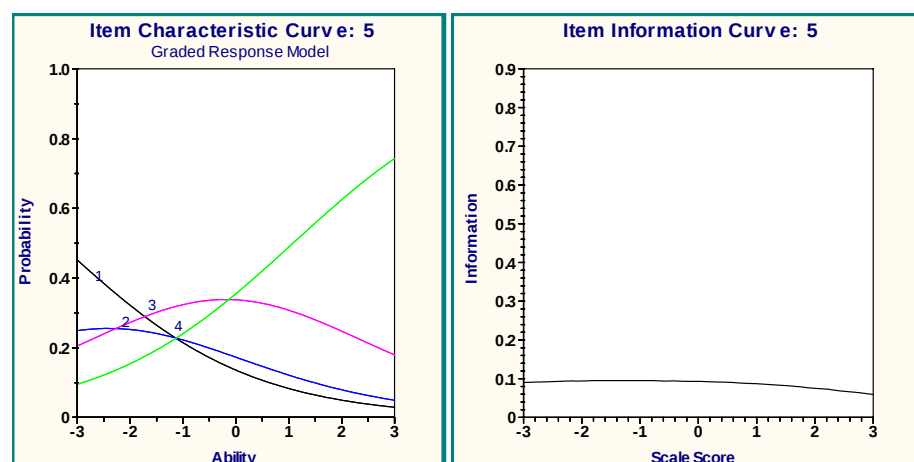
Item: 2

**Category legends****Item: 3**

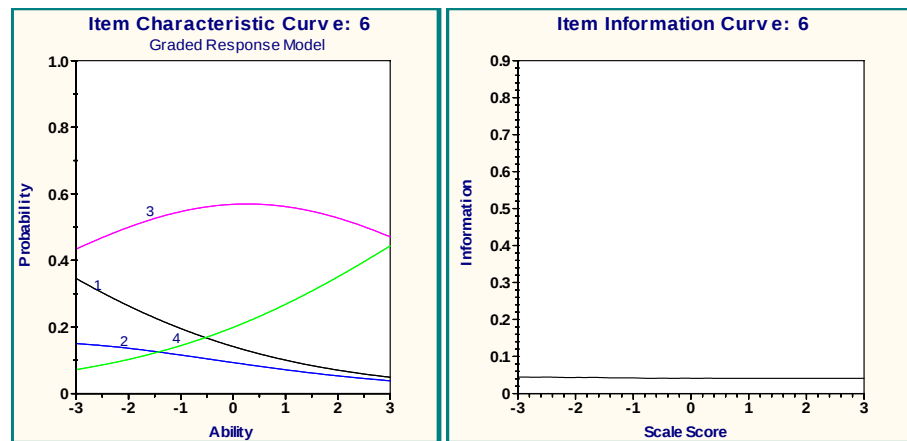
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

**Category legends****Item: 4**

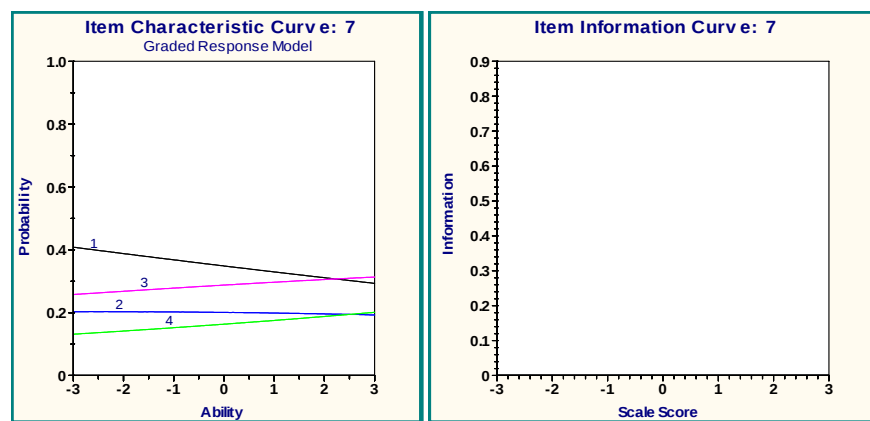
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

**Category legends****Item: 5**

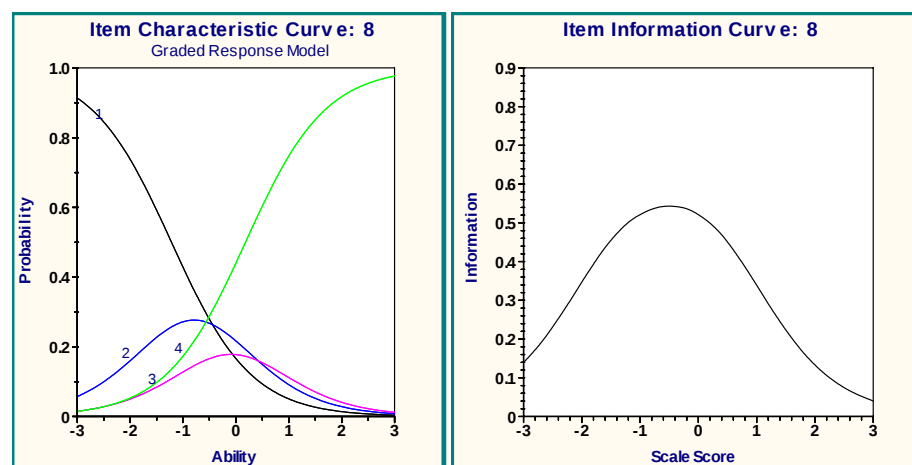
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

**Category legends**

Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

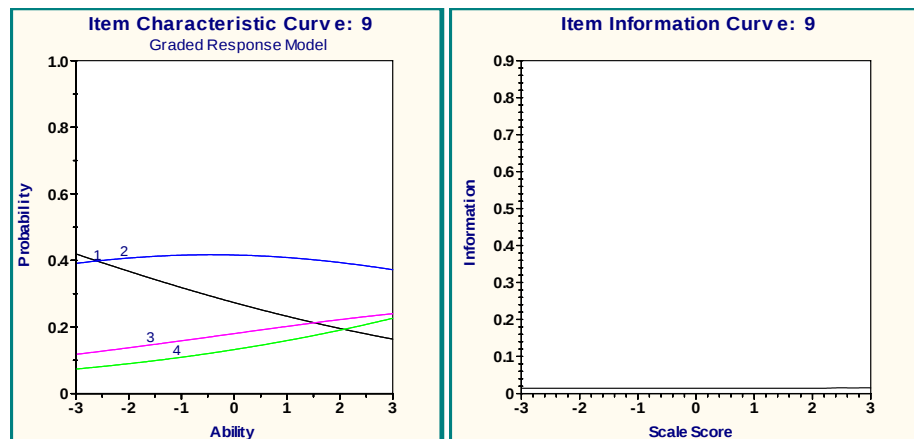
Item: 6**Category legends**

Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

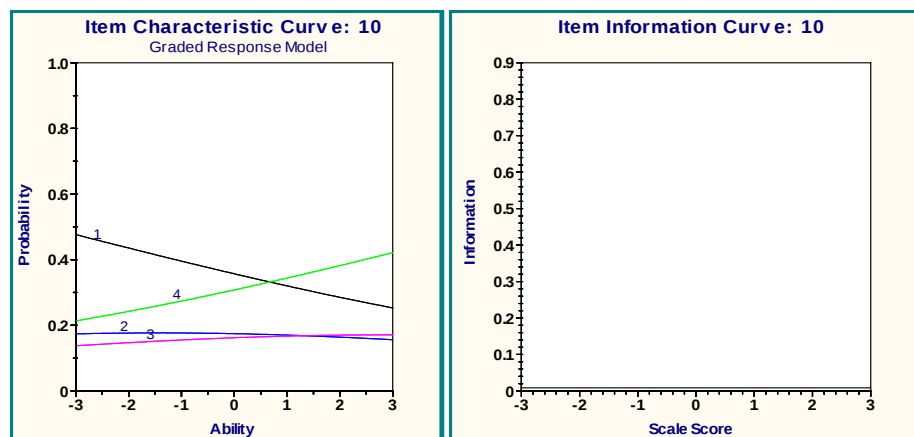
Item: 7**Category legends**

Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

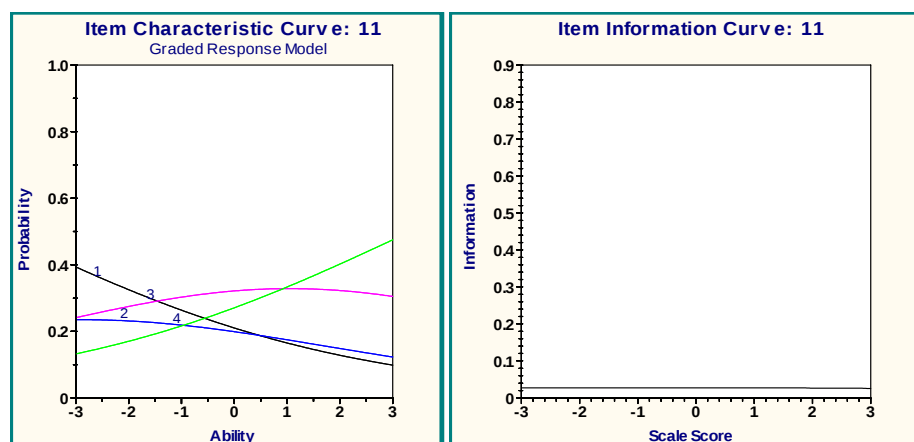
Item: 8

**Category legends**

Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

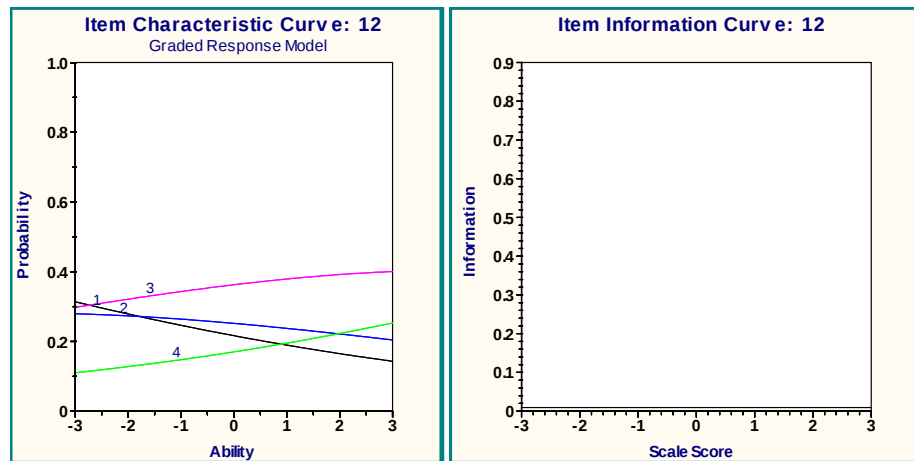
Item: 9**Category legends**

Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

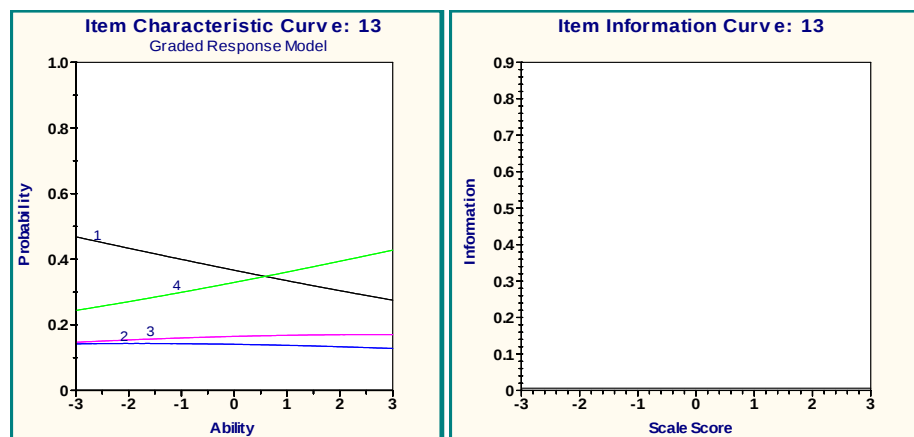
Item: 10**Category legends**

Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

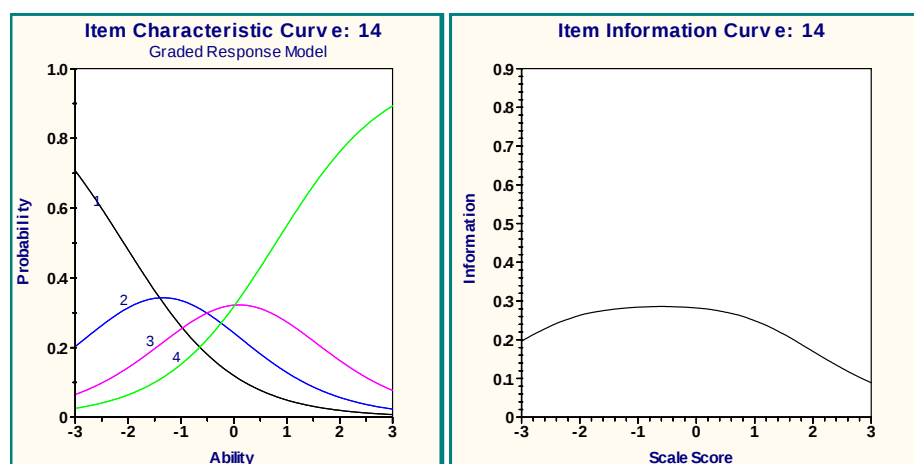
Item: 11

**Category legends****Item: 12**

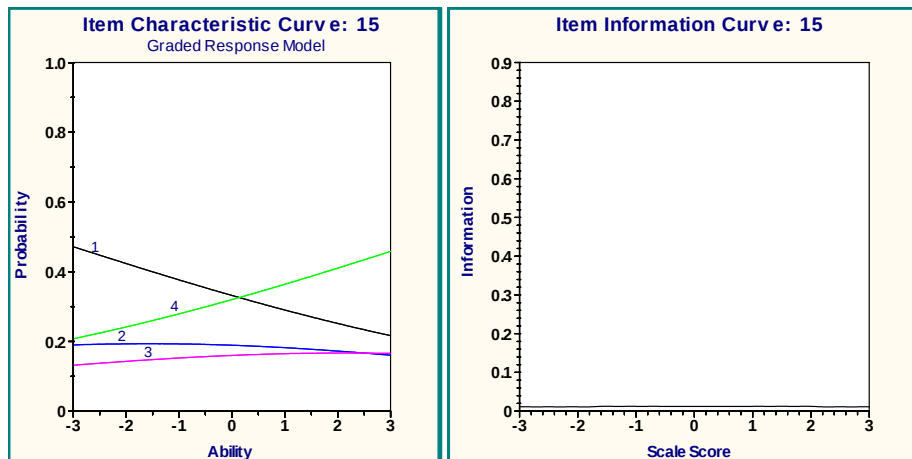
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

**Category legends****Item: 13**

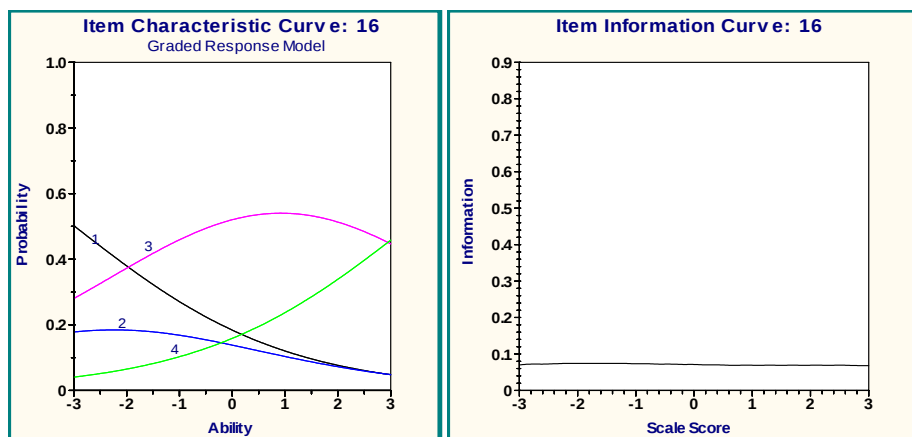
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

**Category legends****Item: 14**

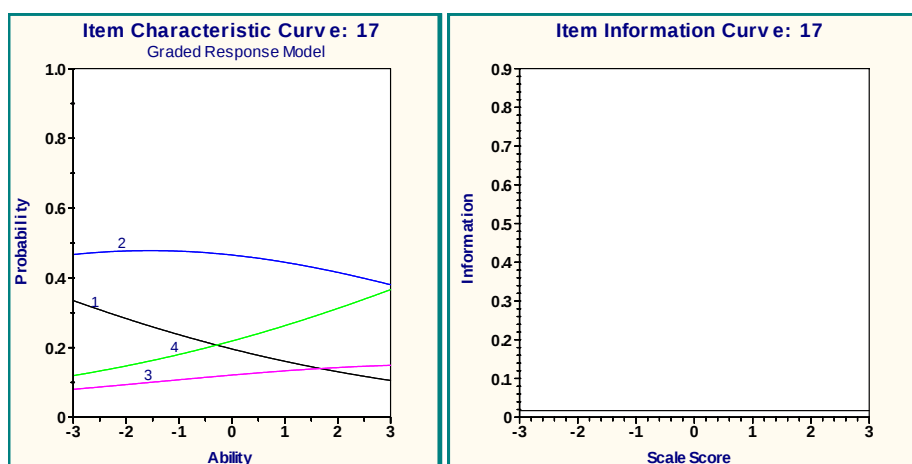
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

**Category legends****Item: 15**

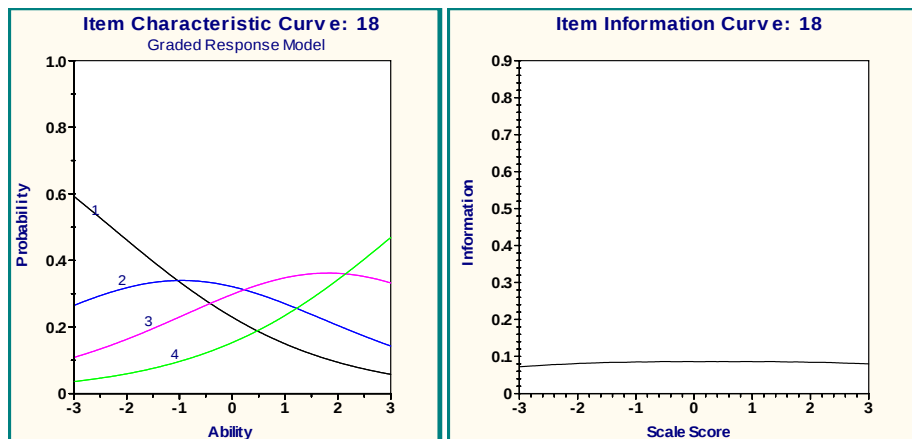
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

**Category legends****Item: 16**

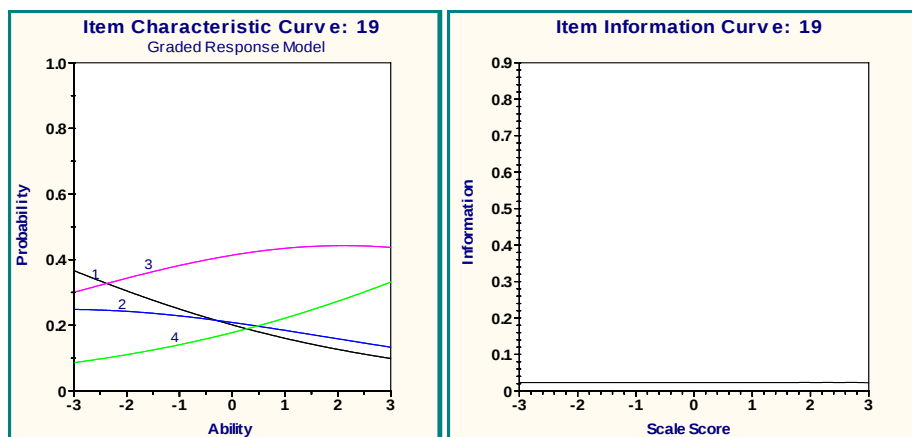
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

**Category legends****Item: 17**

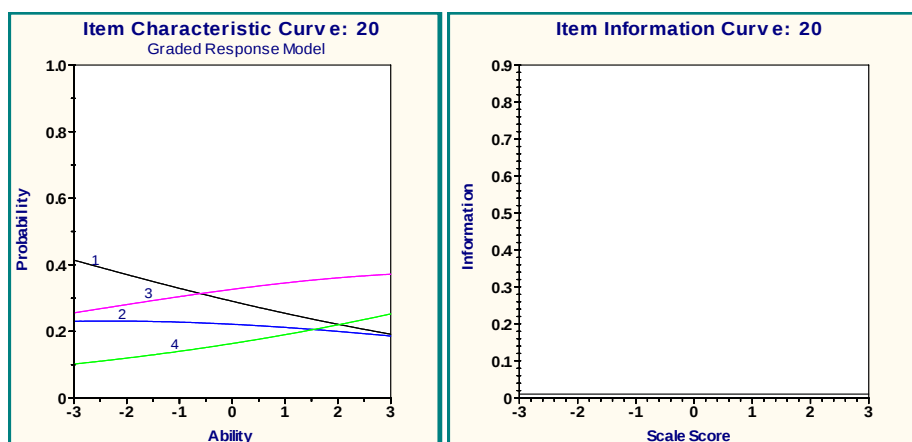
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

**Category legends**

Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

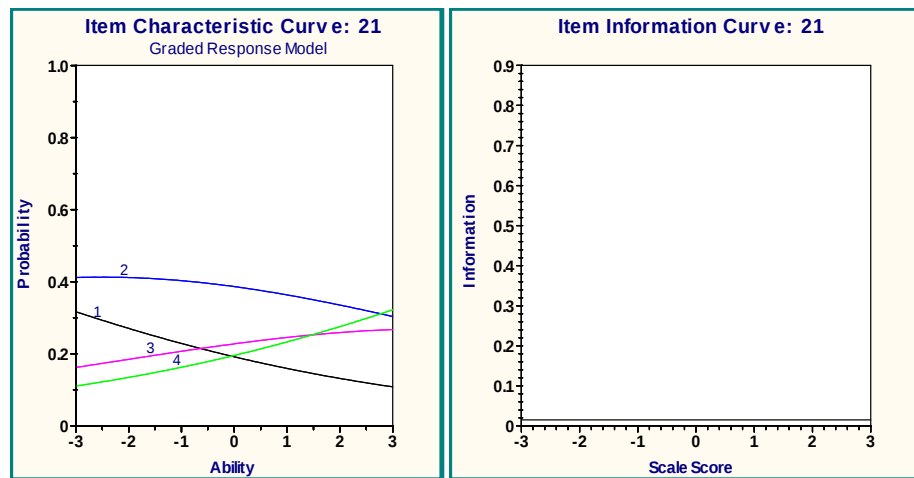
Item: 18**Category legends**

Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

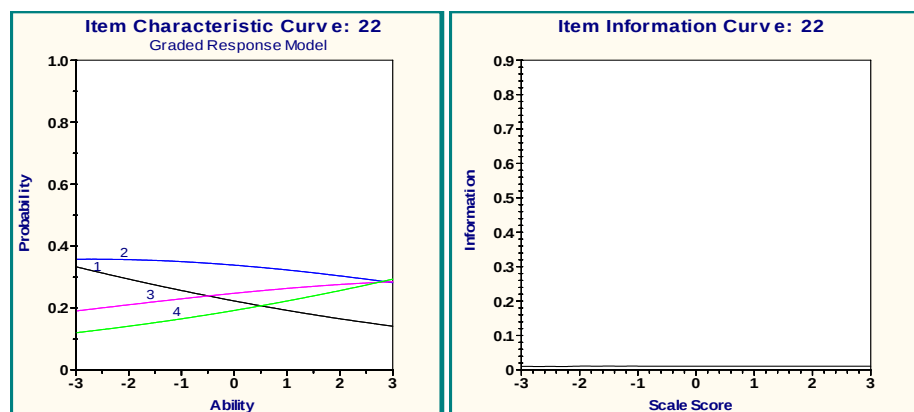
Item: 19**Category legends**

Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

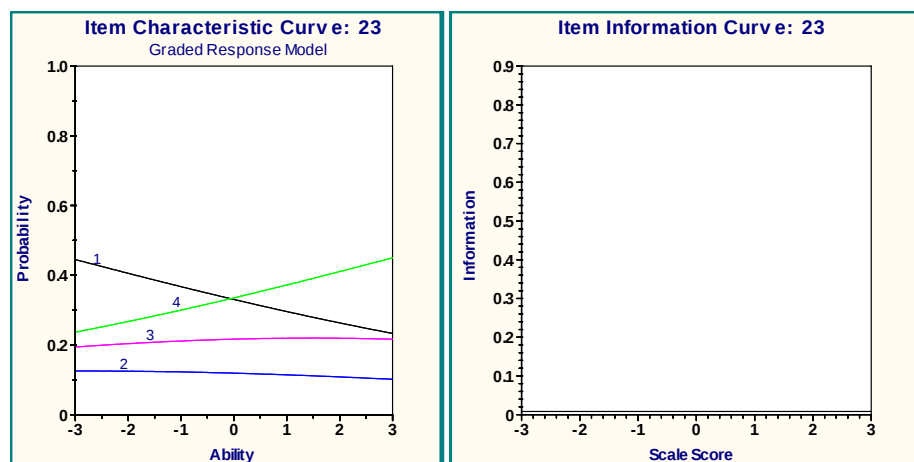
Item: 20

**Category legends****Item: 21**

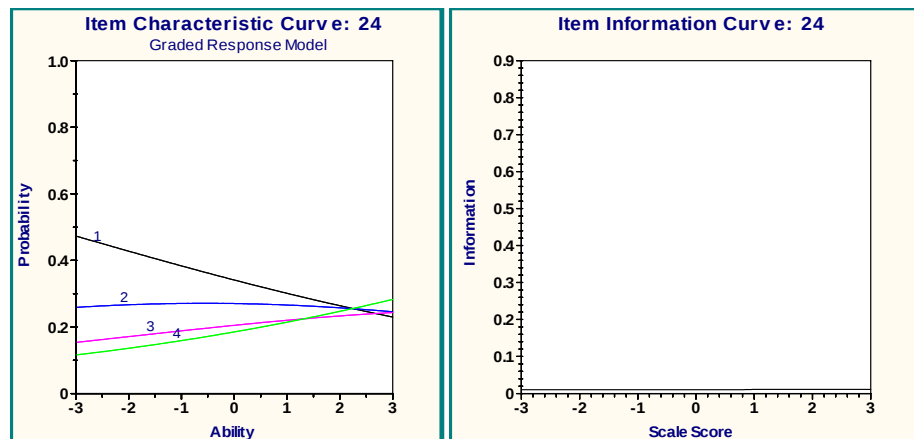
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

**Category legends****Item: 22**

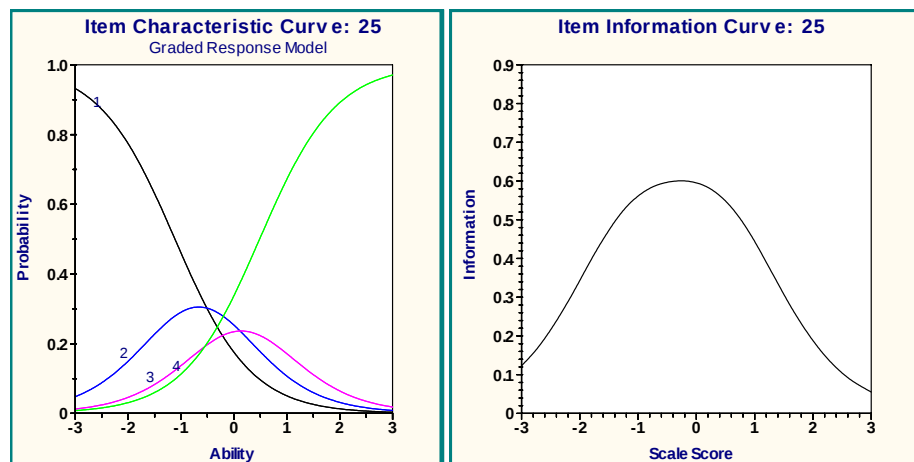
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

**Category legends****Item: 23**

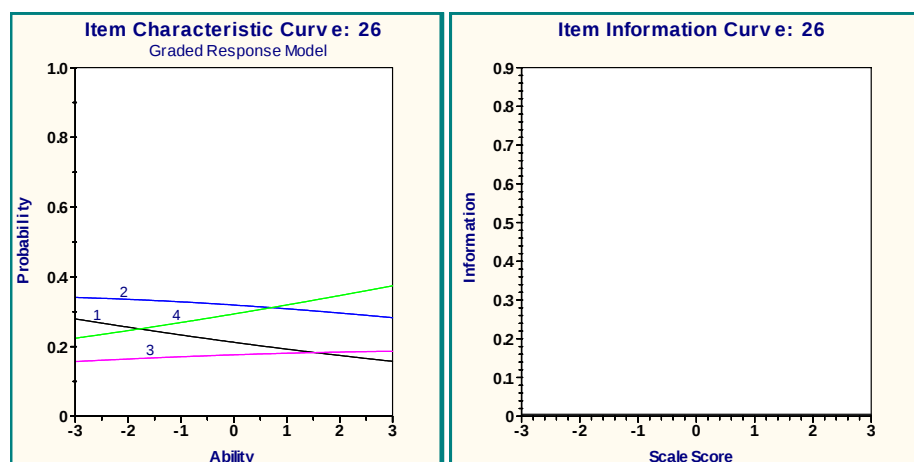
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

**Category legends**

Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

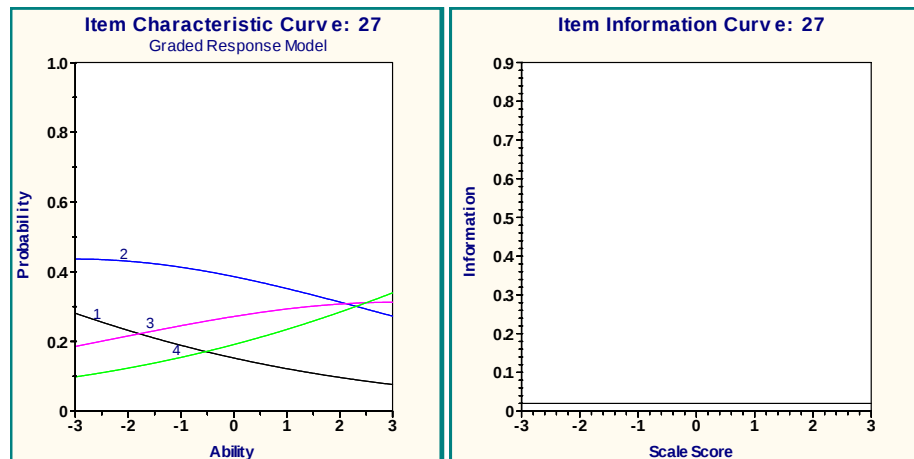
Item: 24**Category legends**

Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

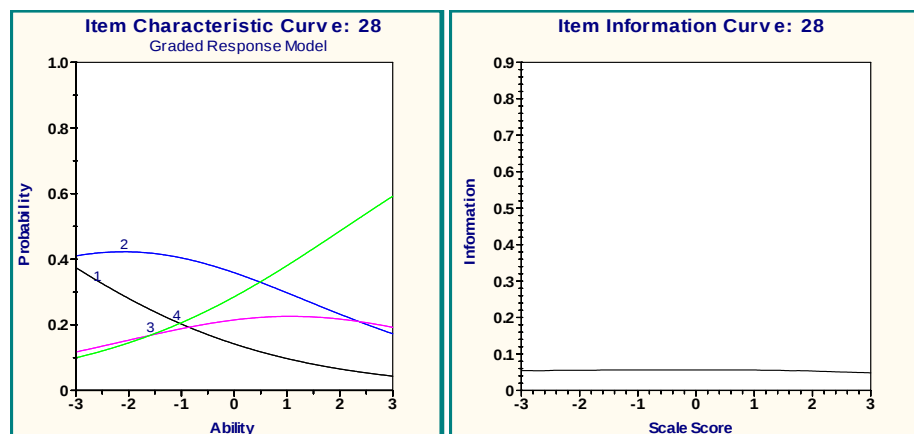
Item: 25**Category legends**

Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

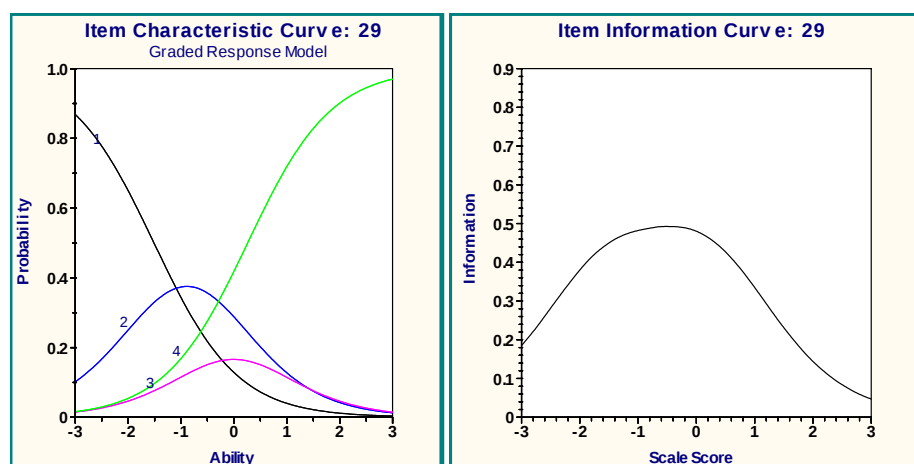
Item: 26

**Category legends****Item: 27**

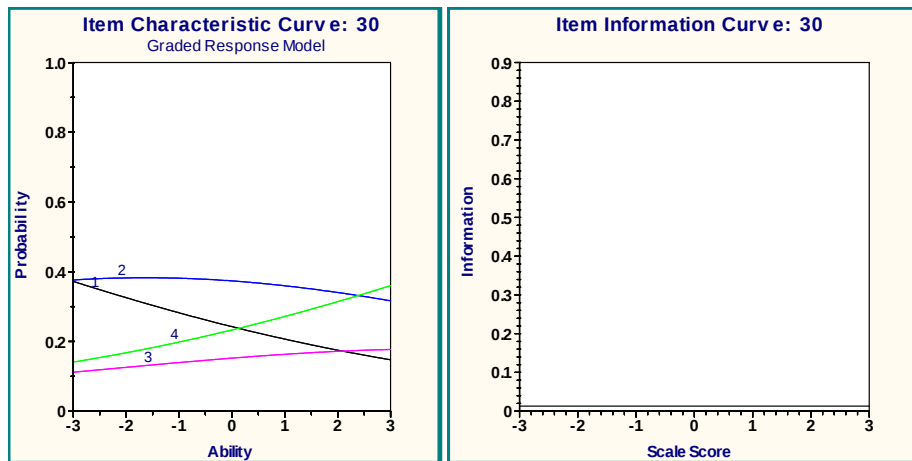
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

**Category legends****Item: 28**

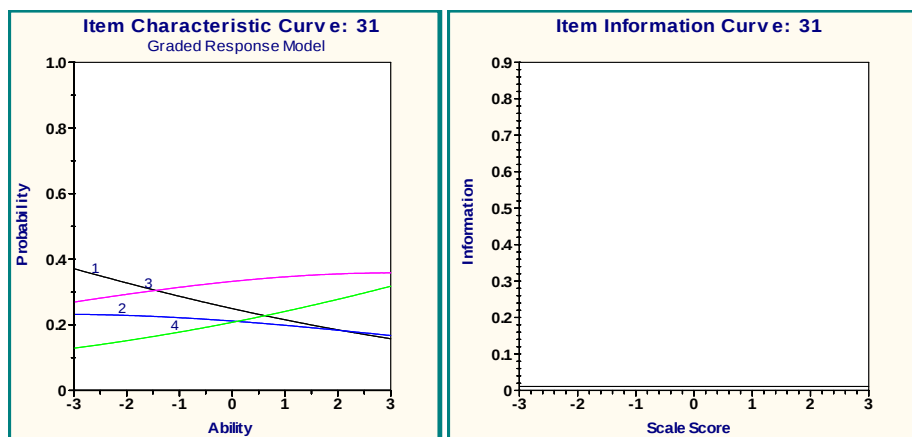
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

**Category legends****Item: 29**

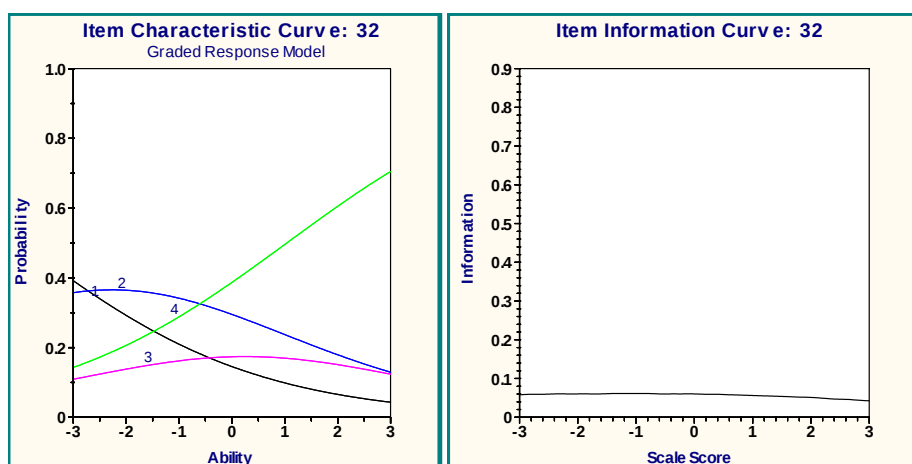
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

**Category legends****Item: 30**

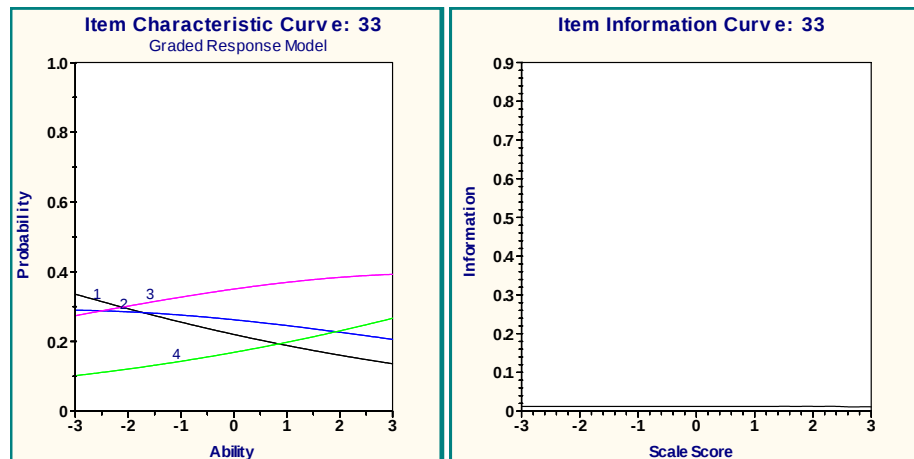
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

**Category legends****Item: 31**

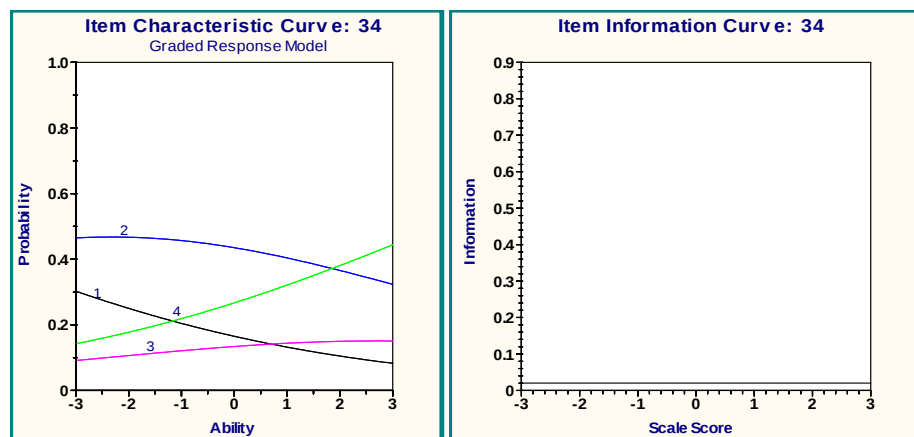
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

**Category legends****Item: 32**

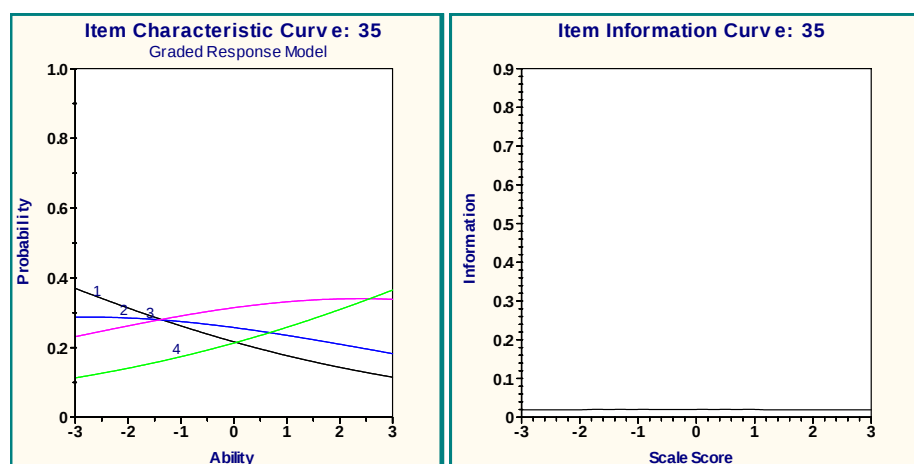
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

**Category legends****Item: 33**

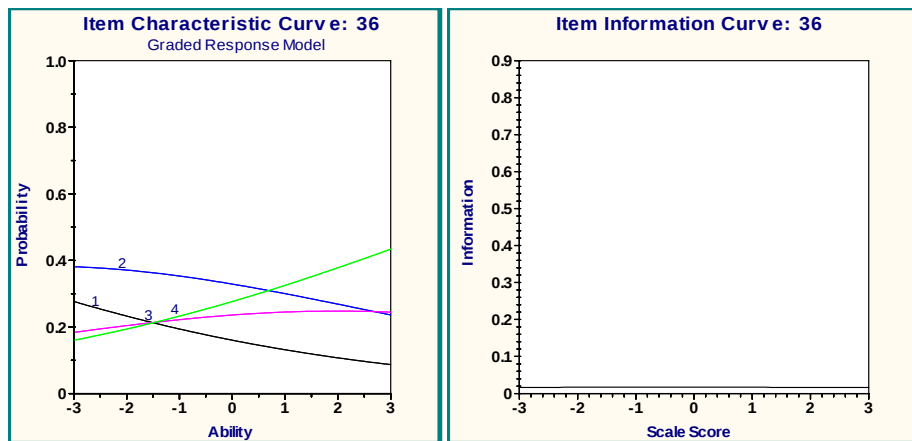
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

**Category legends****Item: 34**

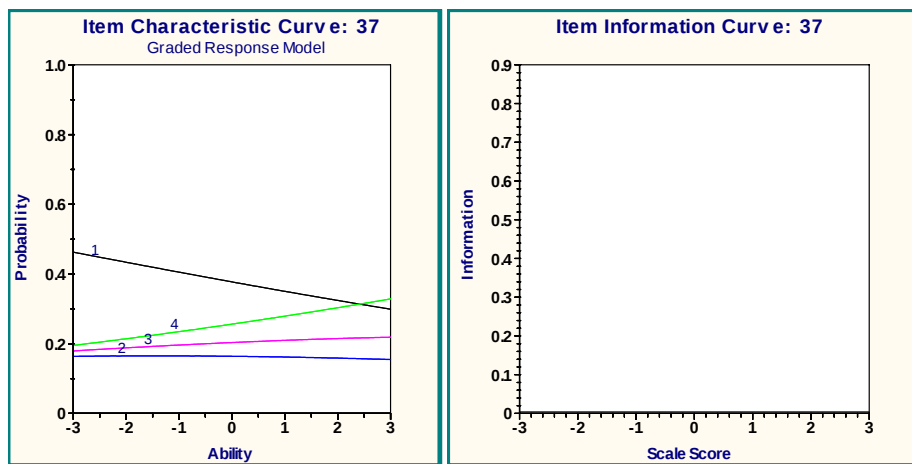
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

**Category legends****Item: 35**

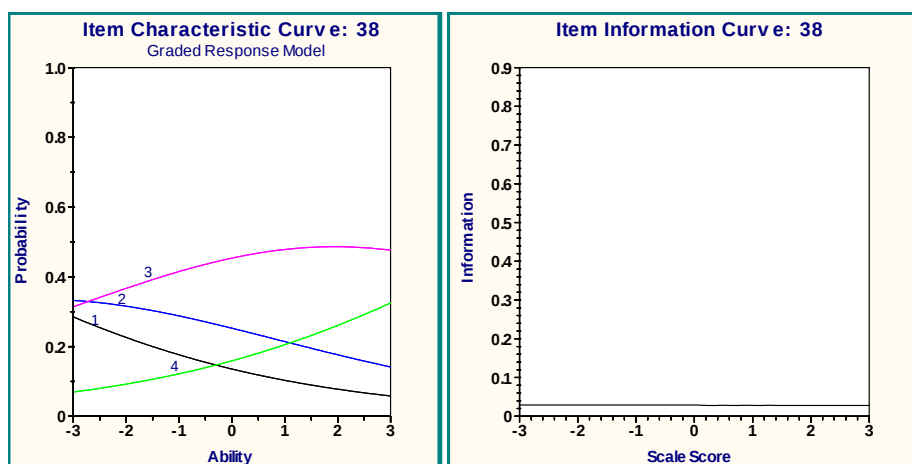
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

**Category legends**

Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

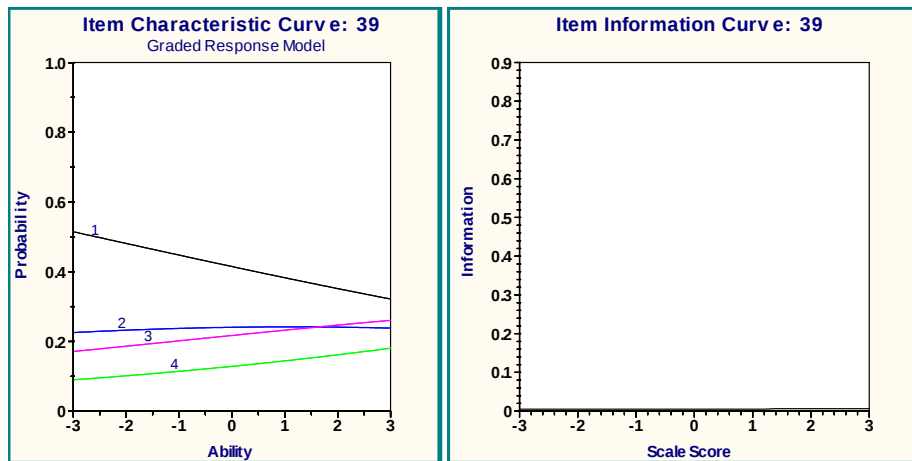
Item: 36**Category legends**

Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

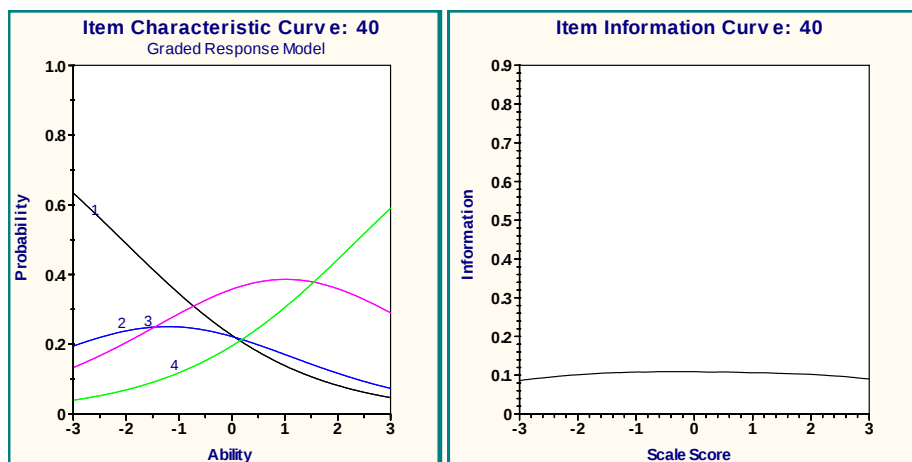
Item: 37**Category legends**

Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

Item: 38

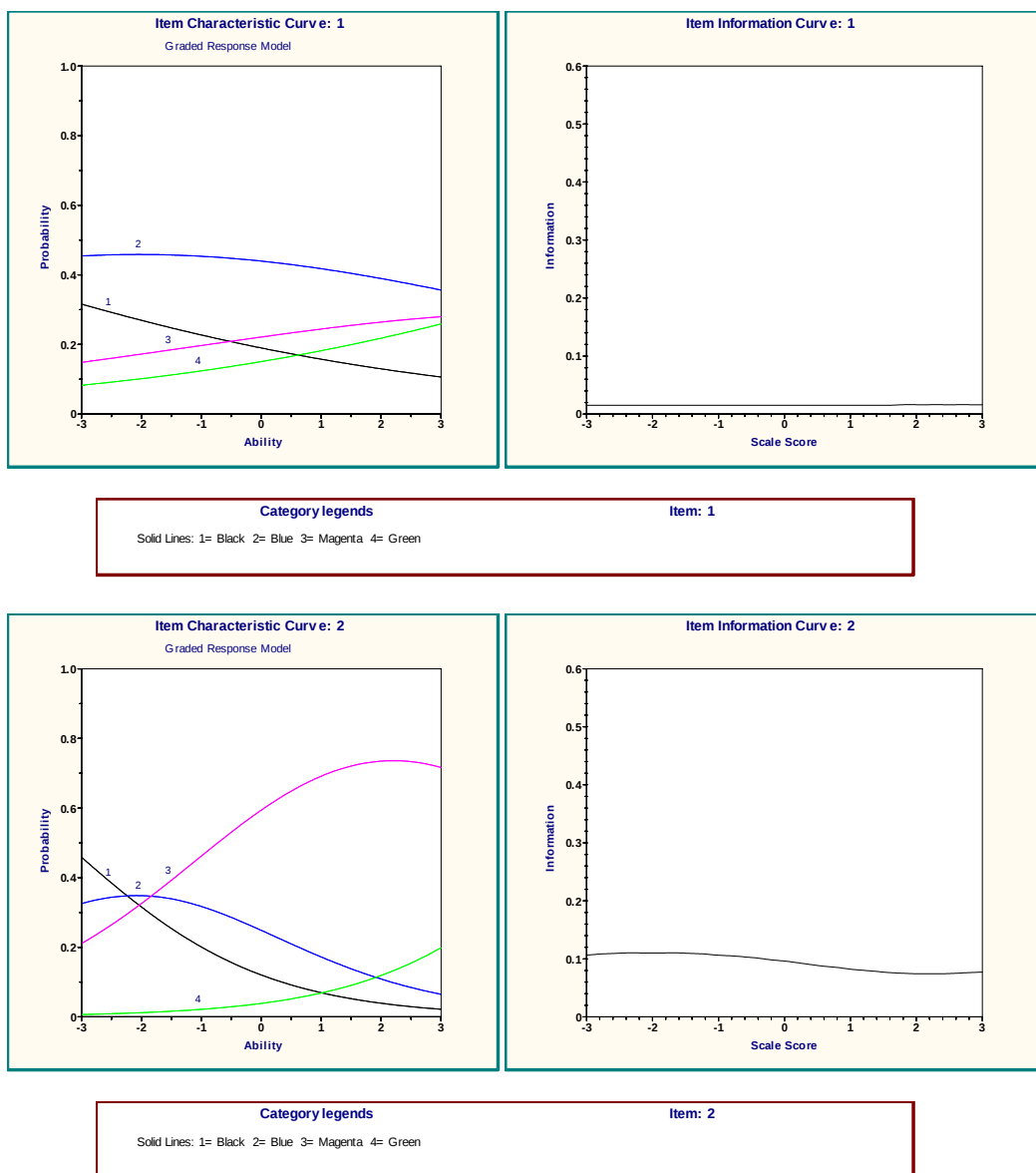
**Category legends****Item: 39**

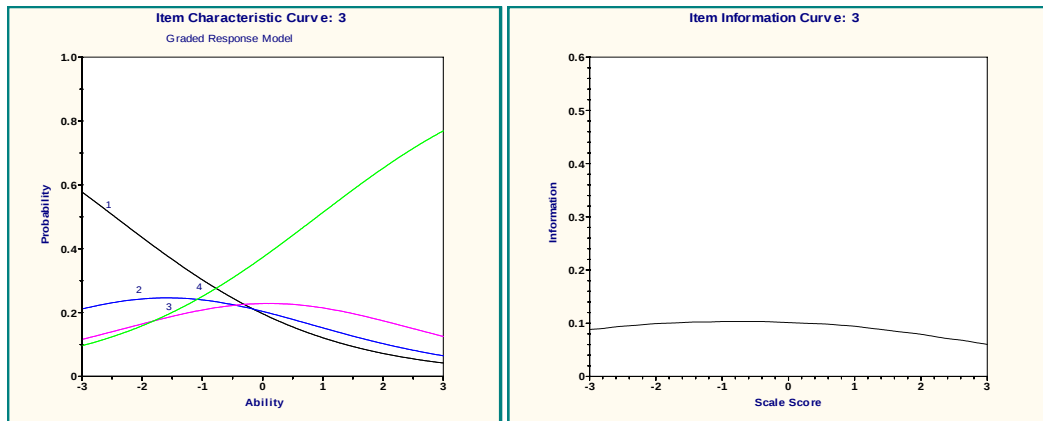
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

**Category legends****Item: 40**

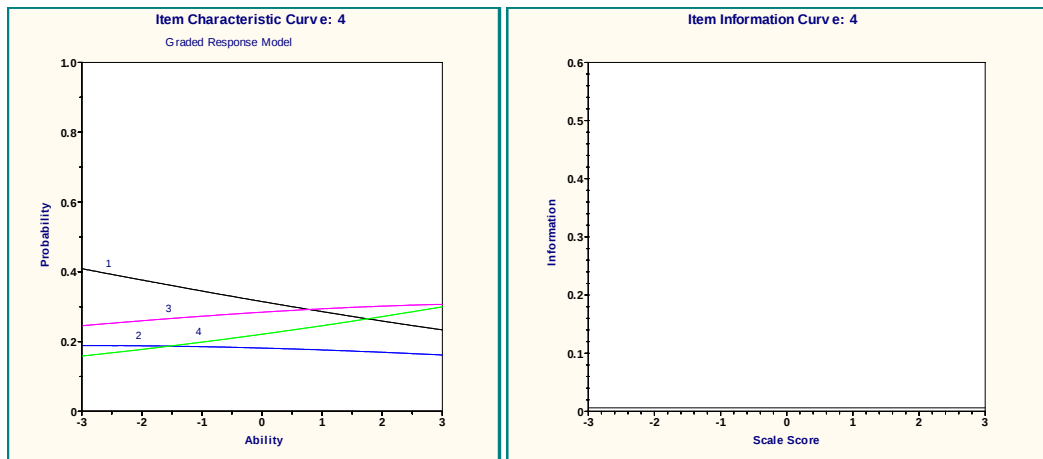
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green

ภาพประกอบ 15 โค้งลักษณะข้อสอบและสารสนเทศข้อสอบที่มีการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็น

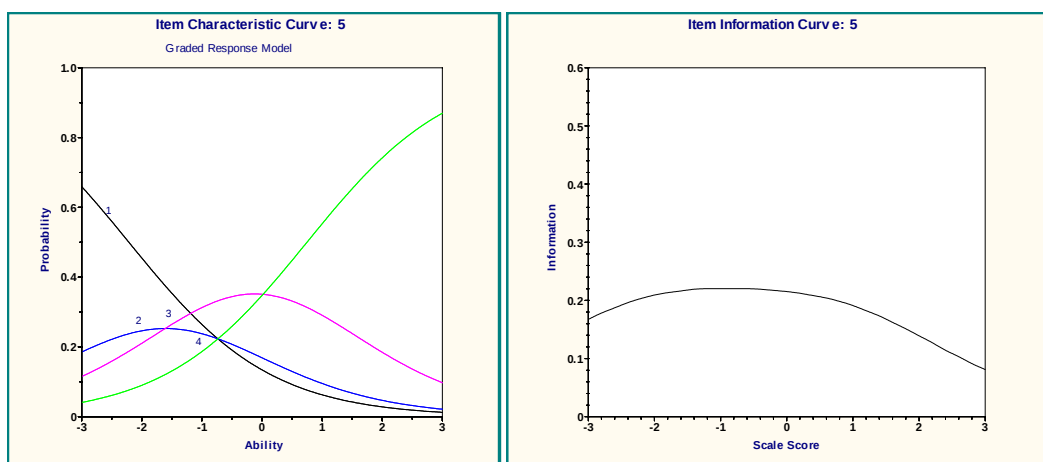




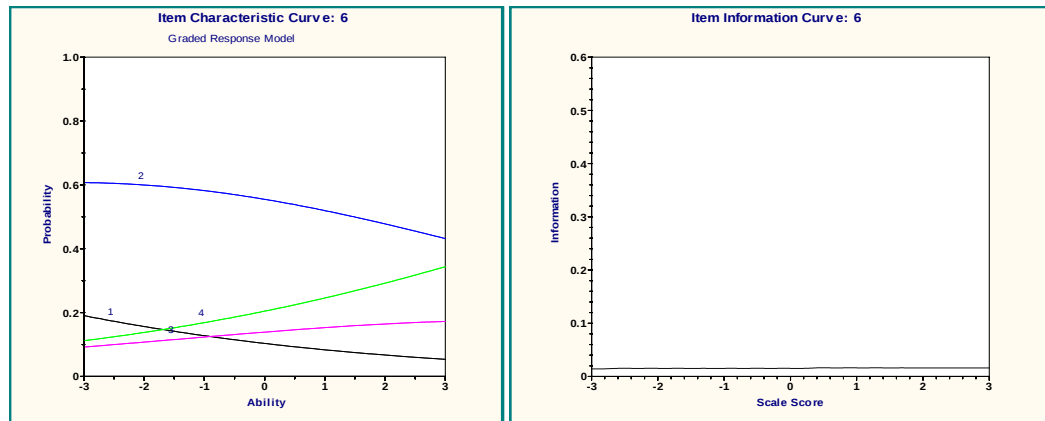
Category legends	Item: 3
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	



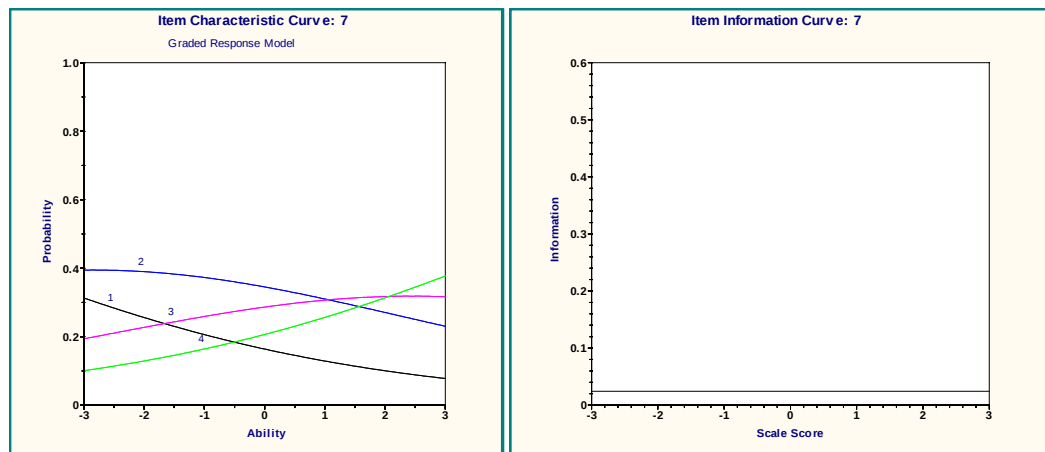
Category legends	Item: 4
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	



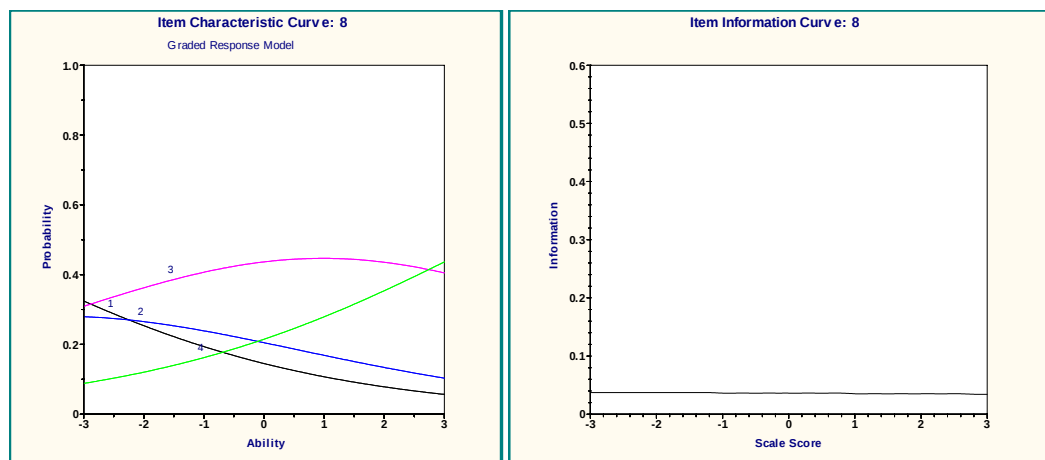
Category legends	Item: 5
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	



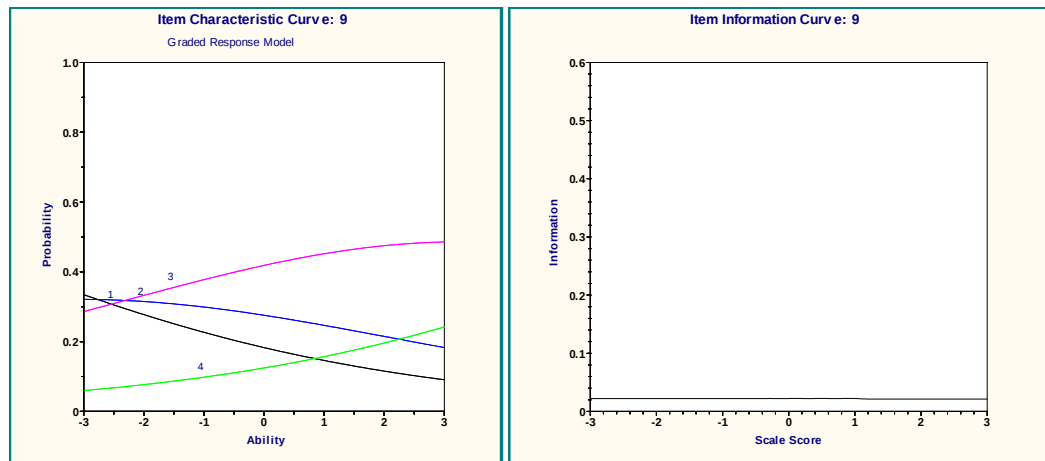
Category legends Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	Item: 6
--	----------------



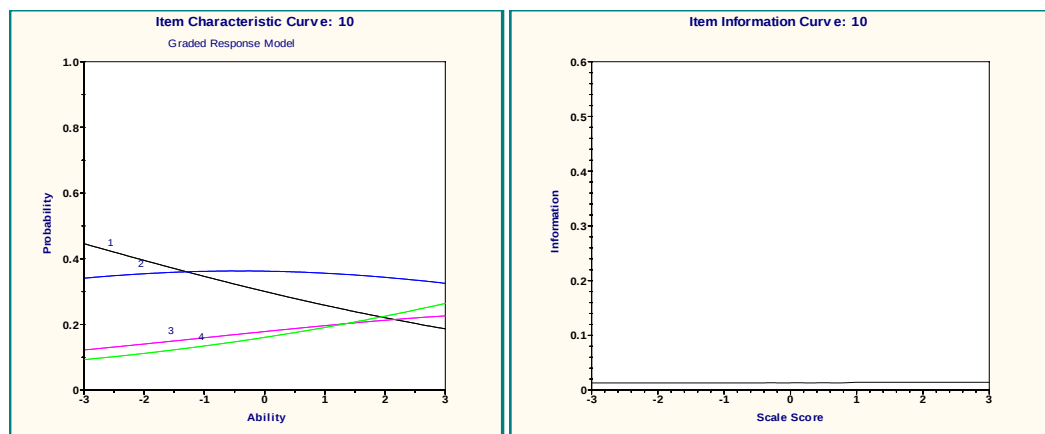
Category legends Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	Item: 7
--	----------------



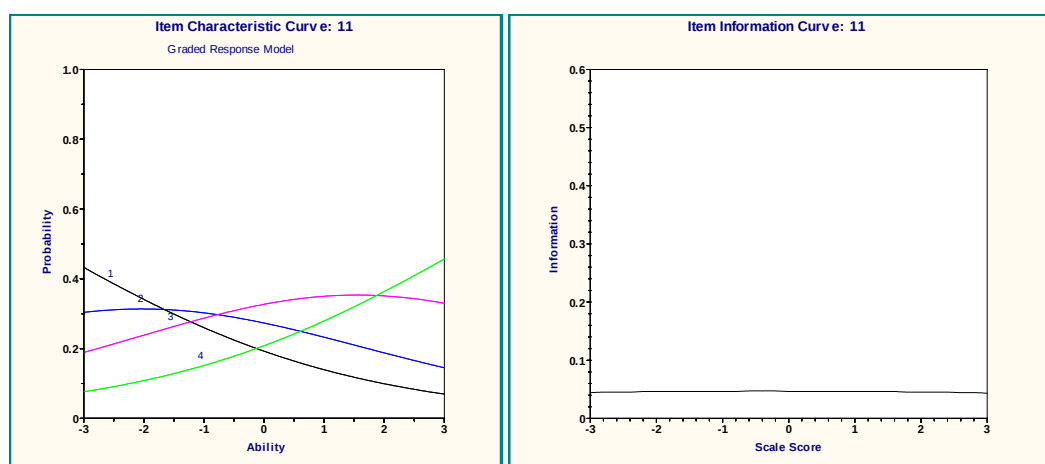
Category legends Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	Item: 8
--	----------------



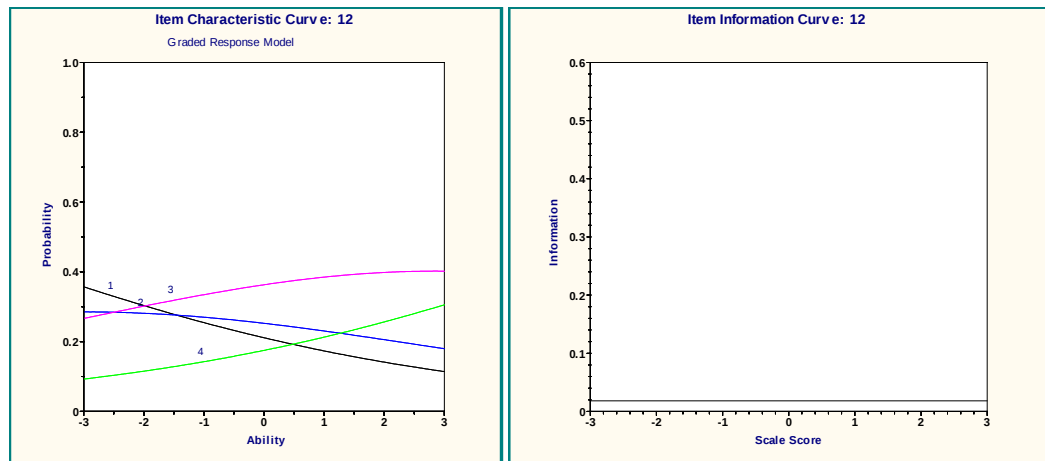
Category legends Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	Item: 9
--	----------------



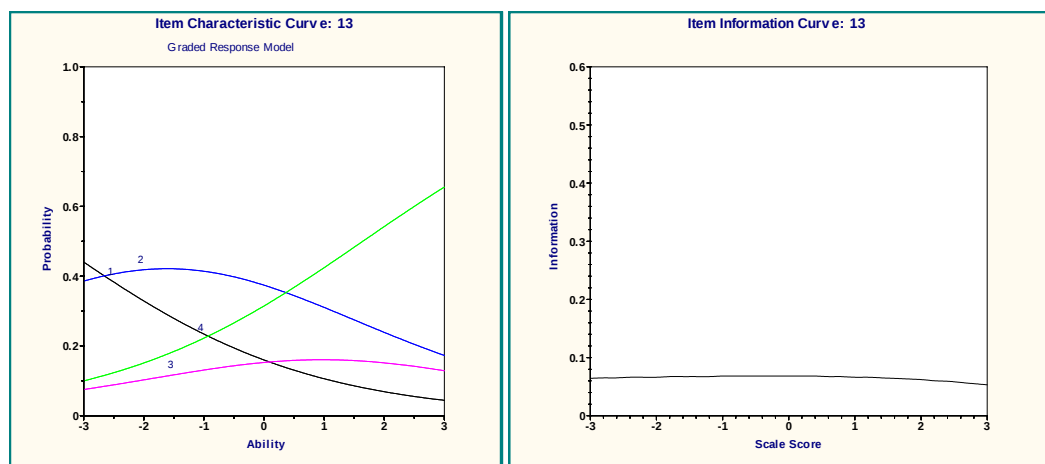
Category legends Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	Item: 10
--	-----------------



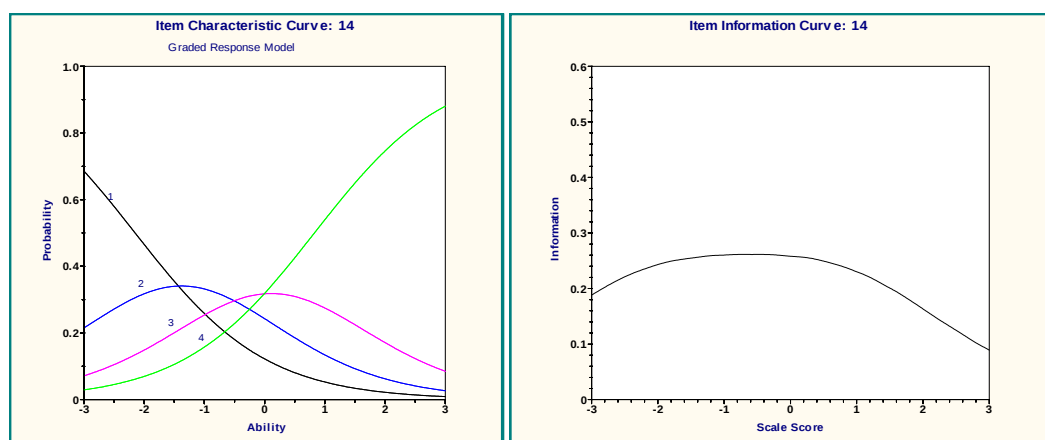
Category legends Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	Item: 11
--	-----------------



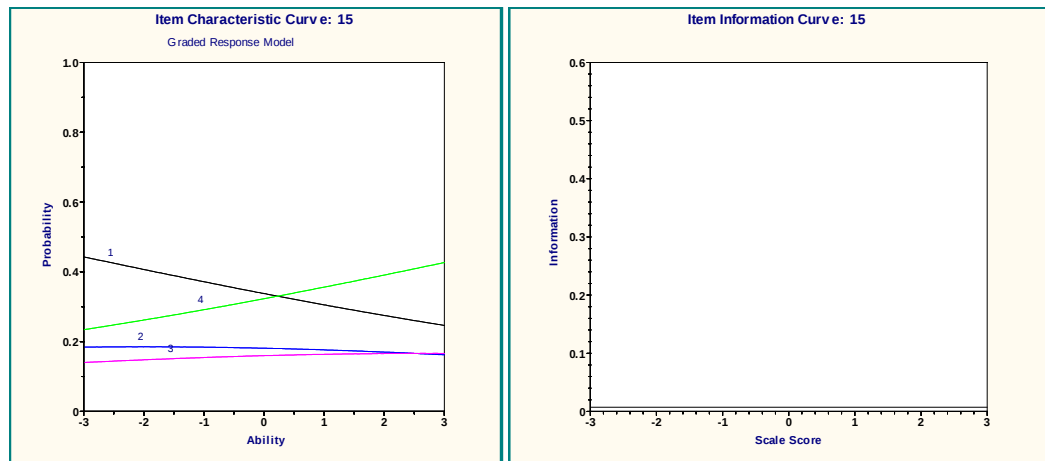
Category legends Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	Item: 12
--	-----------------



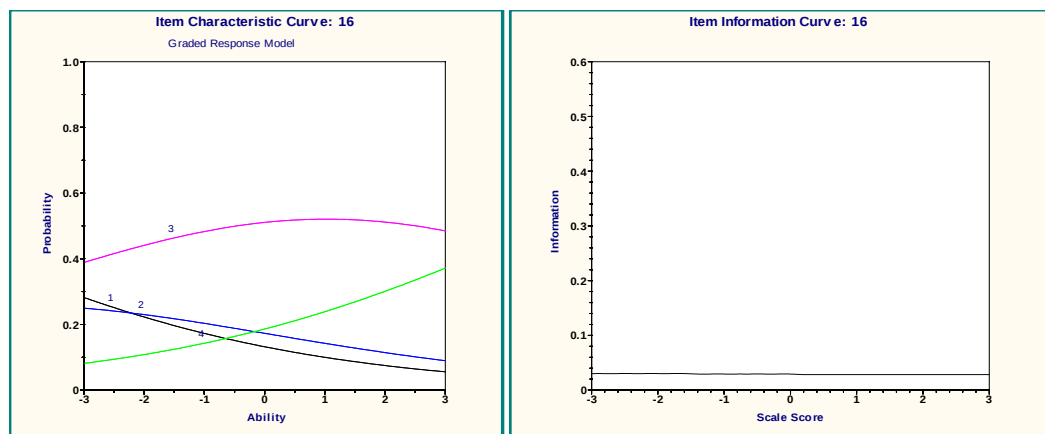
Category legends Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	Item: 13
--	-----------------



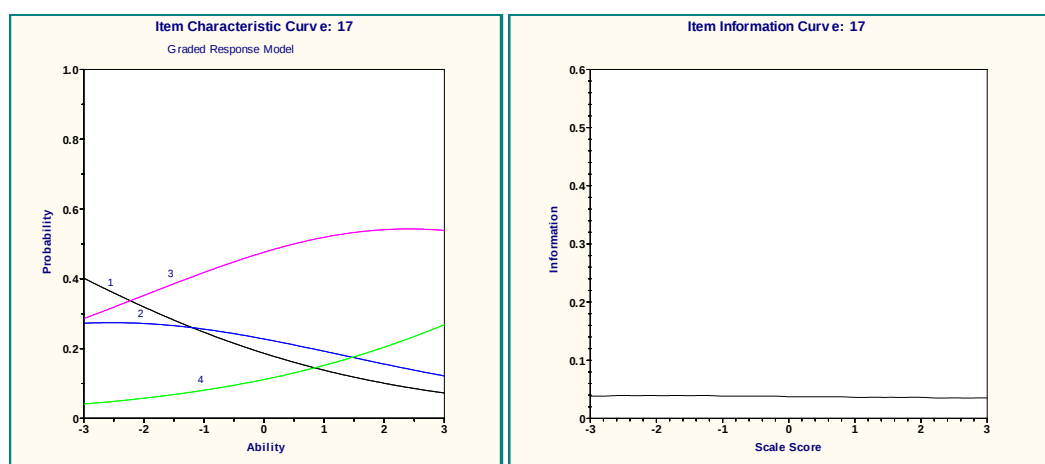
Category legends Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	Item: 14
--	-----------------



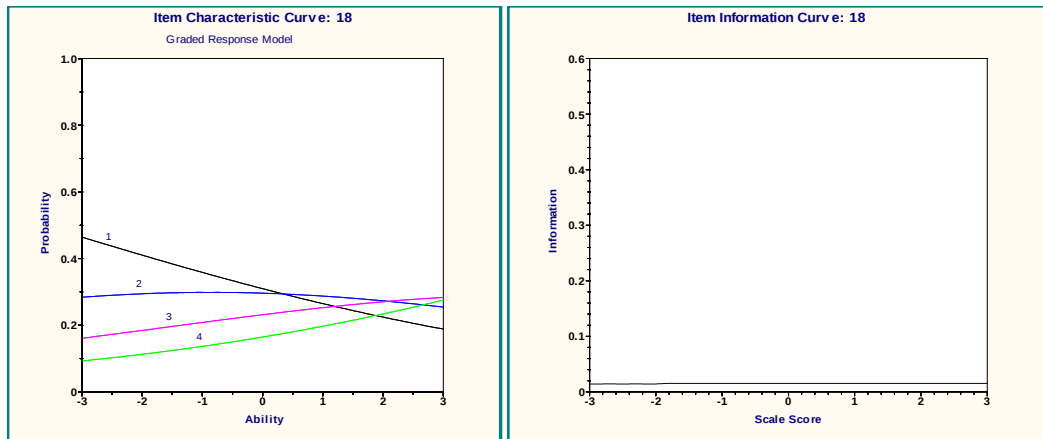
Category legends	Item: 15
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	



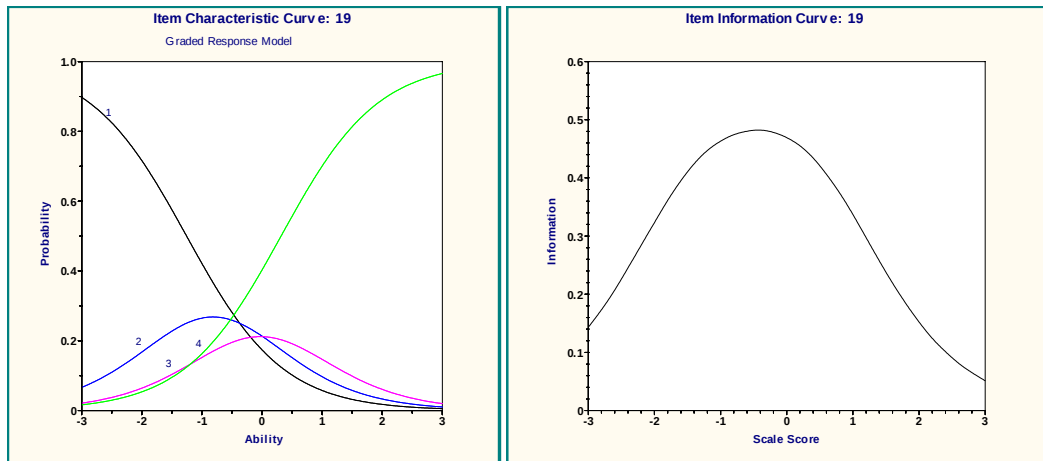
Category legends	Item: 16
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	



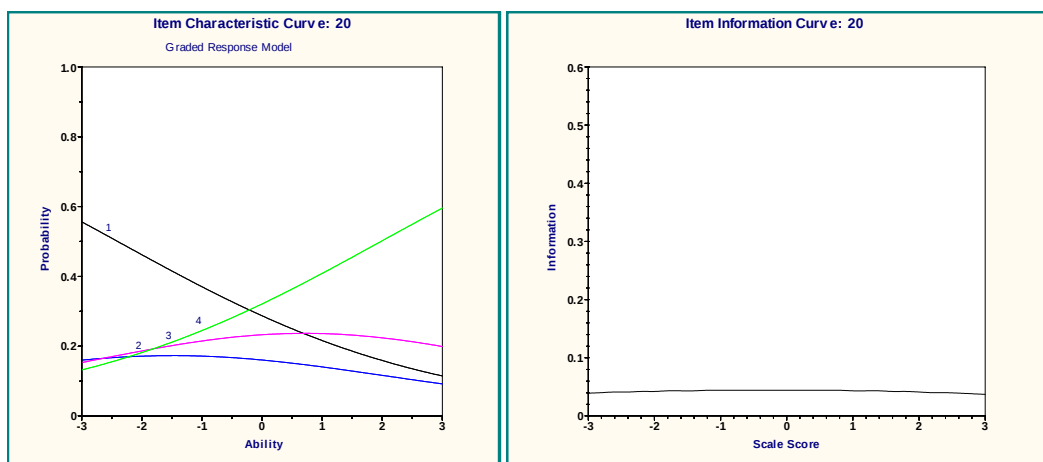
Category legends	Item: 17
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	



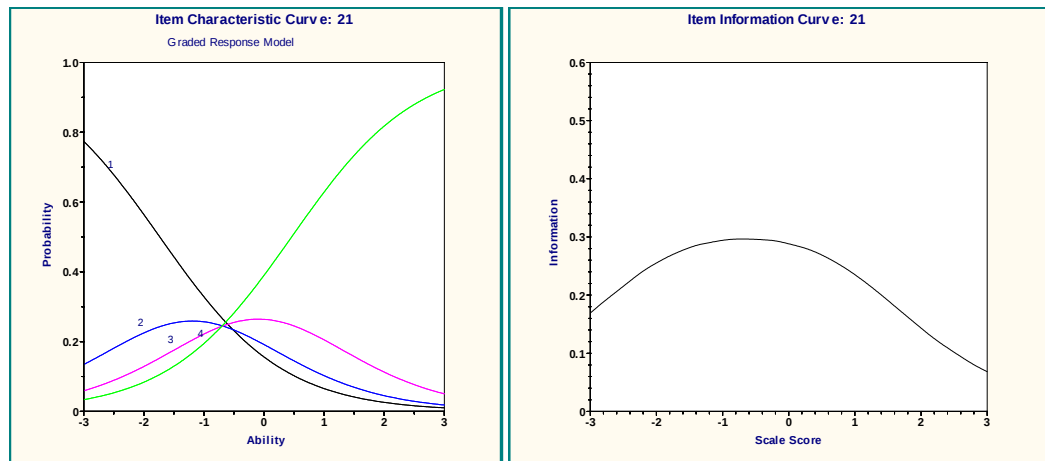
Category legends Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	Item: 18
--	-----------------



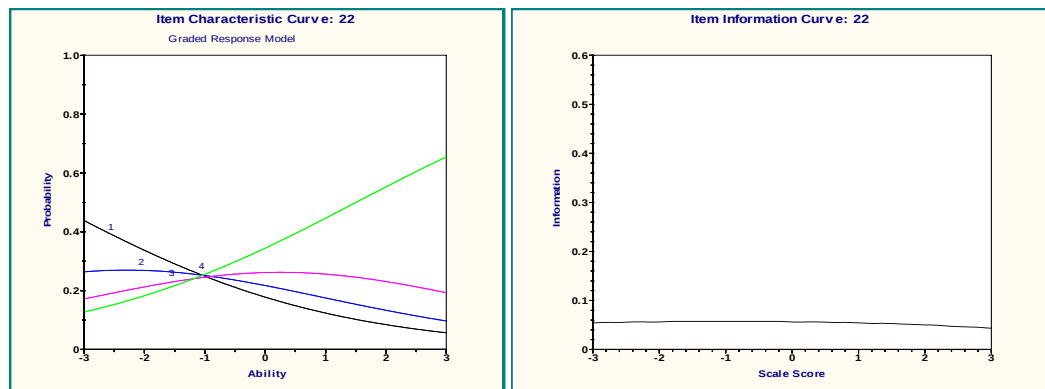
Category legends Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	Item: 19
--	-----------------



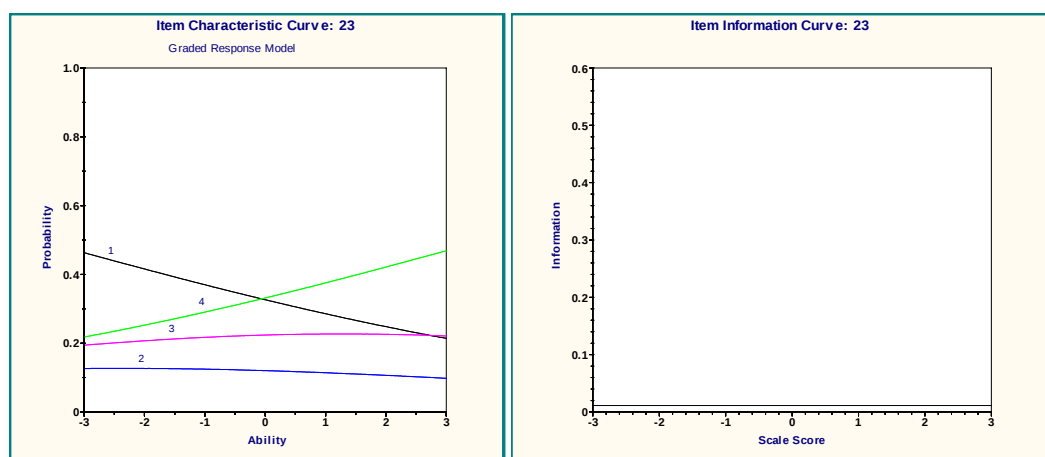
Category legends Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	Item: 20
--	-----------------



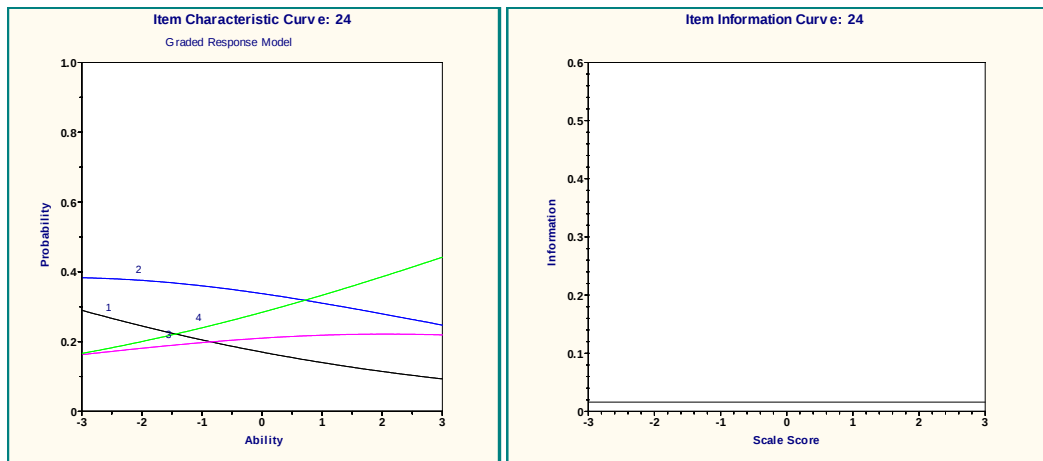
Category legends	Item: 21
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	



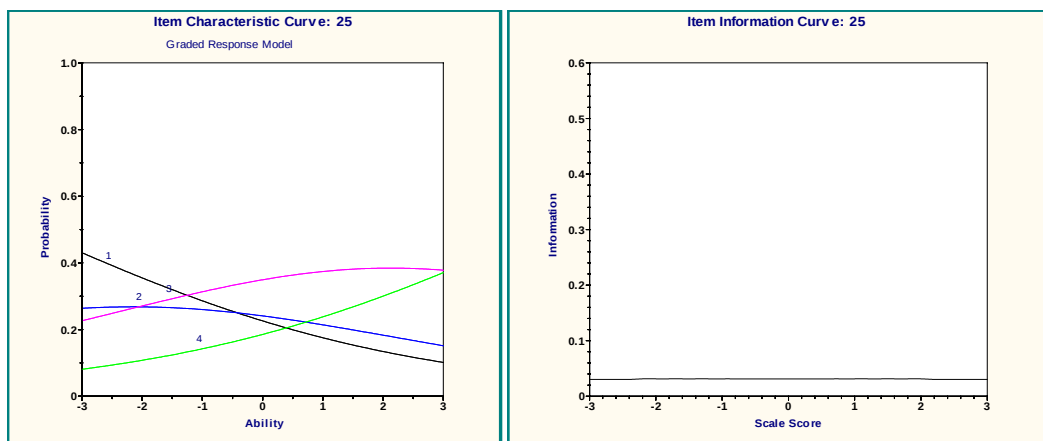
Category legends	Item: 22
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	



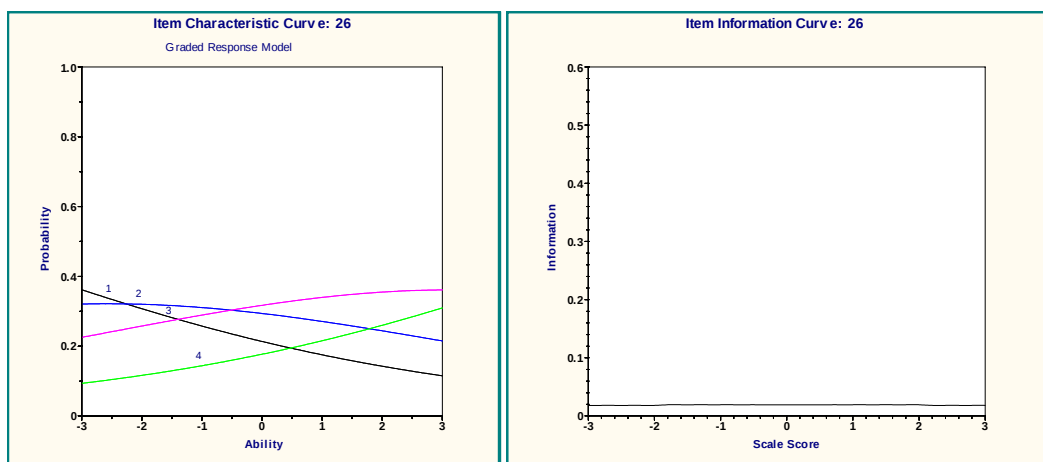
Category legends	Item: 23
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	



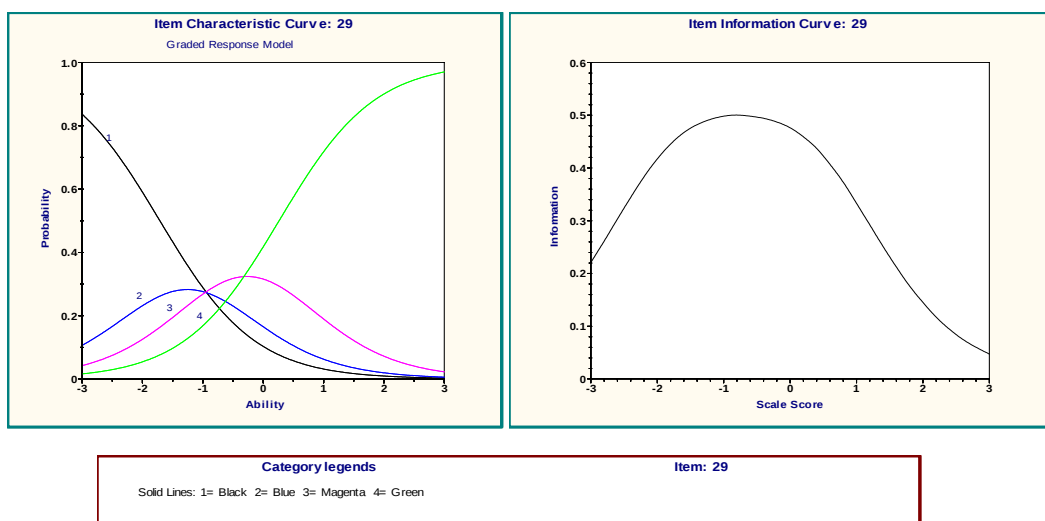
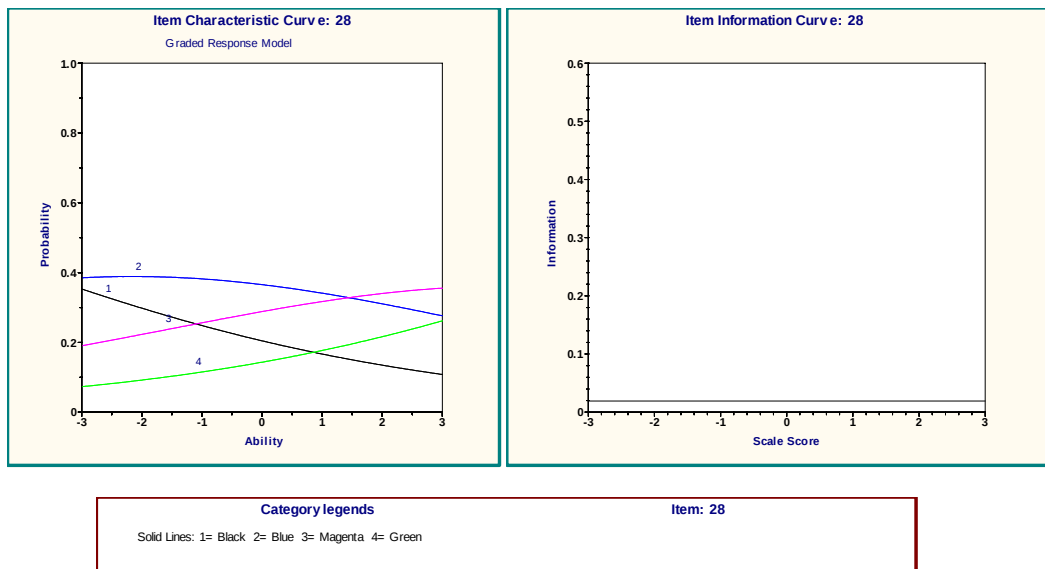
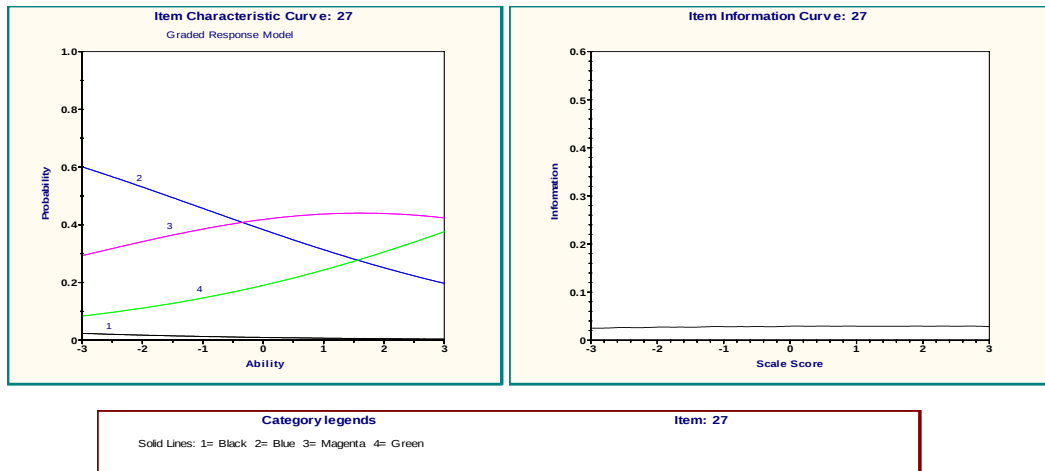
Category legends	Item: 24
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	

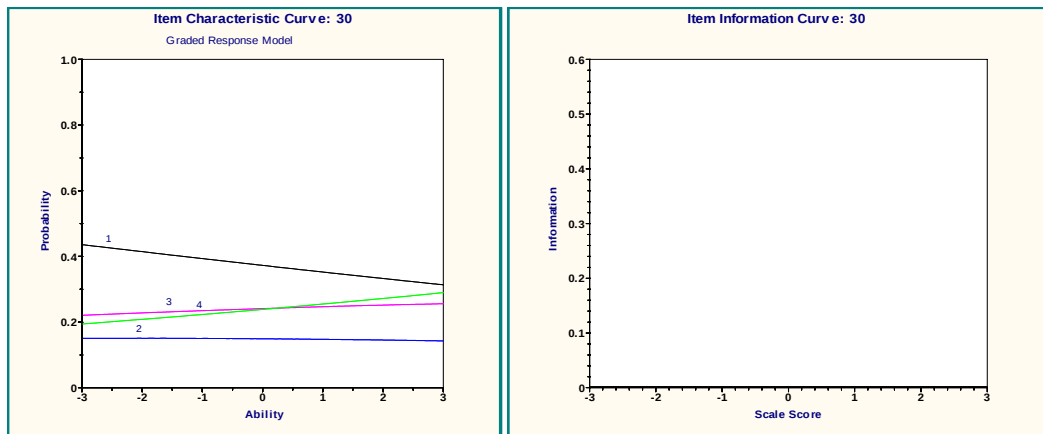


Category legends	Item: 25
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	

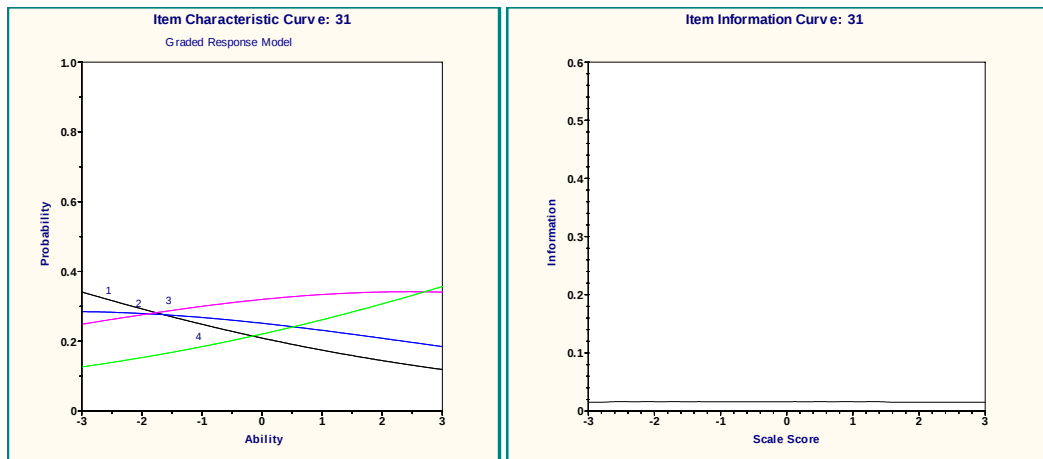


Category legends	Item: 26
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	

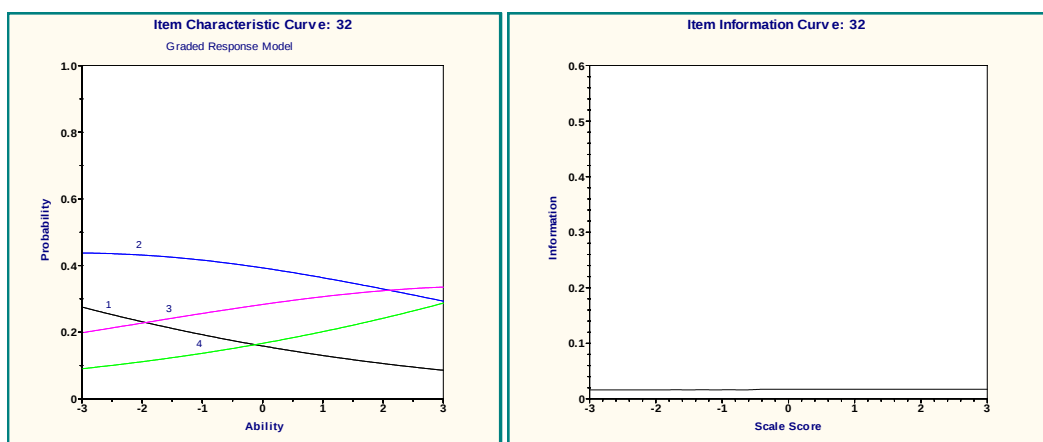




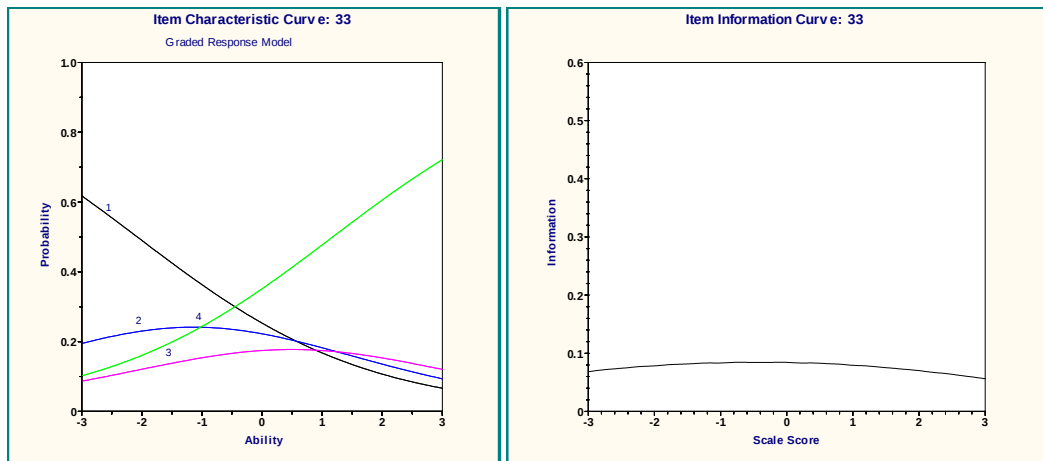
Category legends	Item: 30
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	



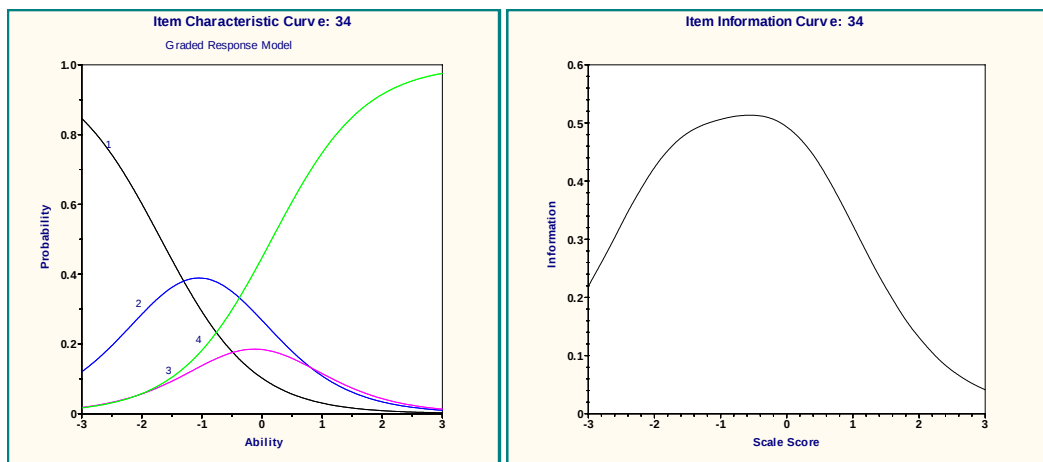
Category legends	Item: 31
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	



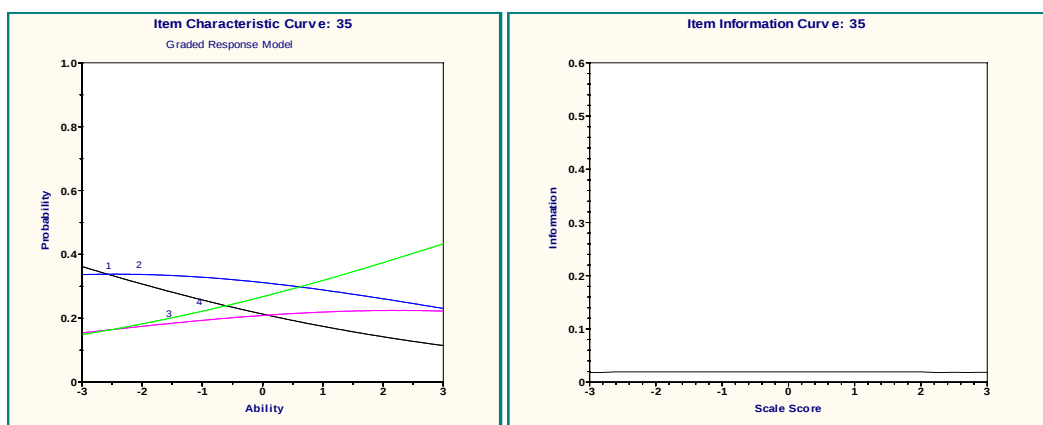
Category legends	Item: 32
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	



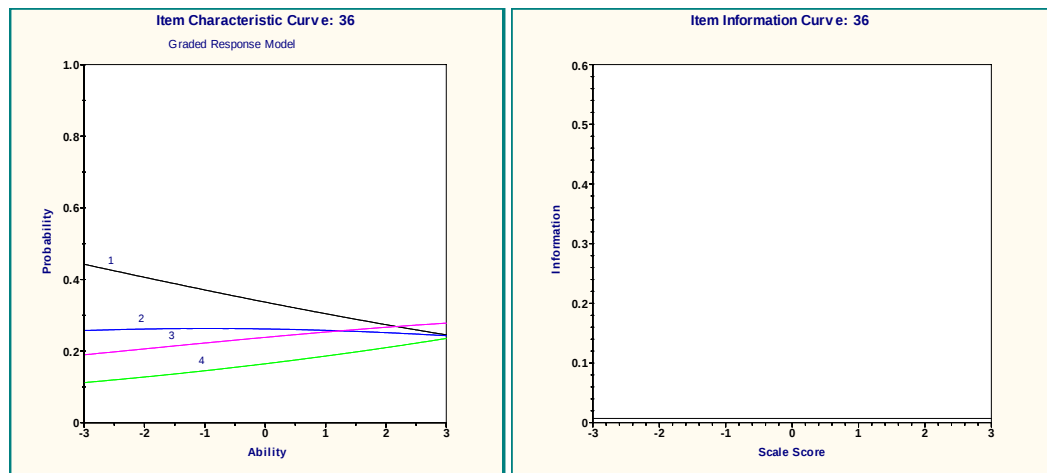
Category legends Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	Item: 33
--	-----------------



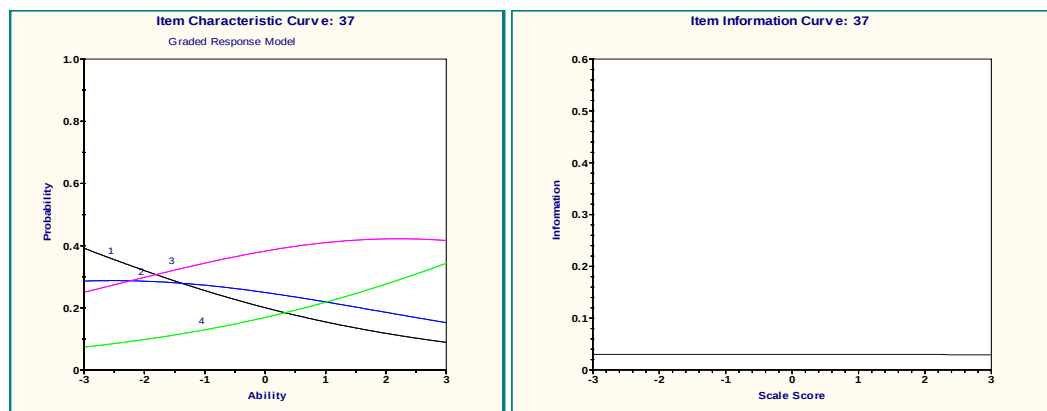
Category legends Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	Item: 34
--	-----------------



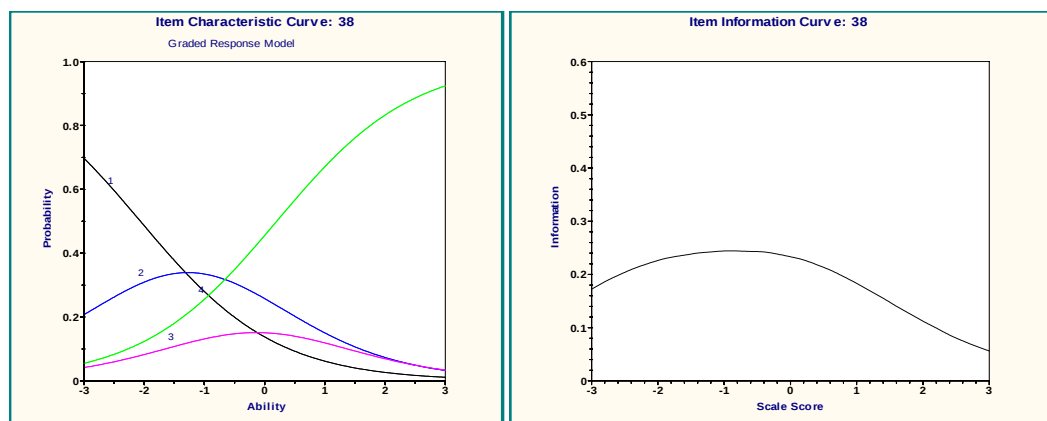
Category legends Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	Item: 35
--	-----------------



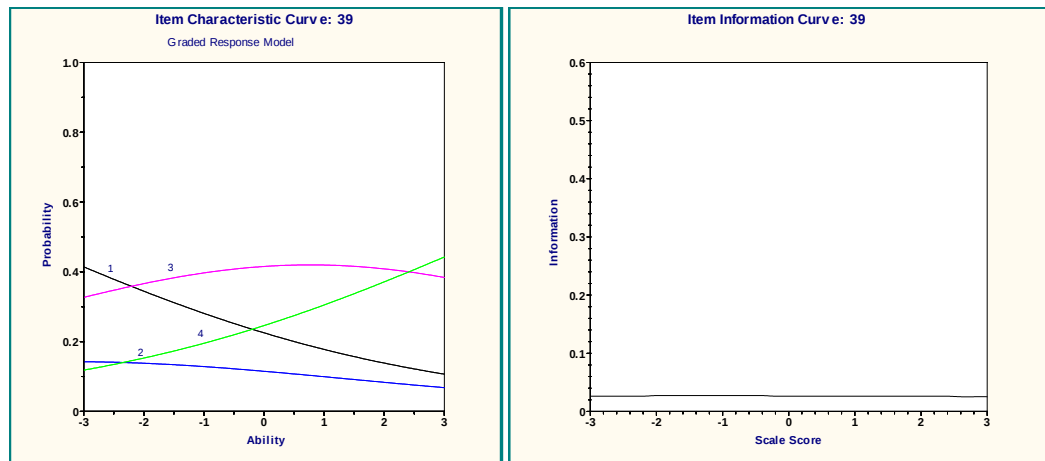
Category legends	Item: 36
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	



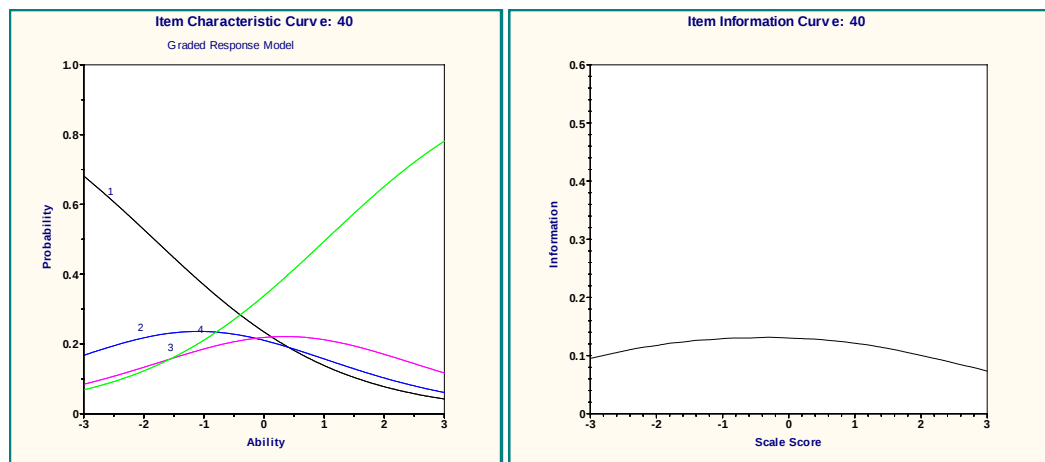
Category legends	Item: 37
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	



Category legends	Item: 38
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	



Category legends	Item: 39
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	



Category legends	Item: 40
Solid Lines: 1= Black 2= Blue 3= Magenta 4= Green	

ภาคผนวก ง
รายนามผู้เชี่ยวชาญ

รายนามผู้เชี่ยวชาญ

นายณรงค์พจน์ มีสรลักษณ์	ศึกษานิเทศก์ฝ่ายวัดผลและประเมินผล สำนักงานเขตพื้นที่การศึกษาปราจีนบุรี จังหวัดปราจีนบุรี
อาจารย์เกดิษฐ์ จันทขจร	หัวหน้ากลุ่มสาระคณิตศาสตร์ โรงเรียนสาธิตมหาวิทยาลัยศรีนครินทรวิโรฒปทุมวัน
อาจารย์สุกัลยา ฉายสุวรรณ	ครู กลุ่มสาระการเรียนรู้คณิตศาสตร์ โรงเรียนสาธิตมหาวิทยาลัยศรีนครินทรวิโรฒปทุมวัน
รองศาสตราจารย์สัญญา รัตนวรารักษ์	ครู กลุ่มสาระการเรียนรู้คณิตศาสตร์ โรงเรียนสาธิตมหาวิทยาลัยศรีนครินทรวิโรฒปทุมวัน
อาจารย์วิรัช สุกใส	ครูชำนาญการ สาระการเรียนรู้คณิตศาสตร์ โรงเรียนมณีเสวตอุบลรัตน์

ประวัติย่อผู้วิจัย

ประวัติย่อผู้วิจัย

ชื่อ ชื่อสกุล	นางสาวสมจวน มานะก่อ
วันเดือนปีเกิด	1 มิถุนายน 2510
สถานที่เกิด	อำเภอประจันตคาม จังหวัดปราจีนบุรี
สถานที่อยู่ปัจจุบัน	31 หมู่ 4 ตำบลหนองแสง อำเภอประจันตคาม จังหวัดปราจีนบุรี 25130
ตำแหน่งหน้าที่การงานในปัจจุบัน	ครูชำนาญการ คศ. 2
สถานที่ทำงานปัจจุบัน	โรงเรียนอนุบาลนาดี ตำบลนาดี อำเภอนาดี จังหวัดปราจีนบุรี
ประวัติการศึกษา	
พ.ศ. 2529	มัธยมศึกษาปีที่ 6 โรงเรียนประจันตราษฎ์บำรุง อำเภอประจันตคาม จังหวัดปราจีนบุรี
พ.ศ. 2533	กศ.บ. สังคมศึกษา มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร
พ.ศ. 2551	กศ.ม. การวัดผลการศึกษา มหาวิทยาลัยศรีนครินทรวิโรฒ