510.76
ว322ก
2560

ก

การศึกษาเปรียบเทียบคุณภาพของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน วิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 แบบเลือกตอบเมื่อตรวจด้วยวิธีการให้คะแนนความรู้บางส่วนด้วยวิธีที่ต่างกัน

ปริญญานิพนธ์

ของ

วรารัตน์ แสงสุข

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต สาขาวิชาการวิจัยและพัฒนาศักยภาพมนุษย์

สิงหาคม 2560

ลิขสิทธิ์เป็นของมหาวิทยาลัยศรีนครินทรวิโรฒ

A 430144

การศึกษาเปรียบเทียบคุณภาพของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน วิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 แบบเลือกตอบเมื่อตรวจด้วยวิธีการให้คะแนนความรู้บางส่วนด้วยวิธีที่ต่างกัน

บทคัดย่อ

ของ

วรารัตน์ แสงสุข

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต สาขาวิชาการวิจัยและพัฒนาศักยภาพมนุษย์
สิงหาคม 2560

วรารัตน์ แสงสุข. (2560). *การศึกษาเปรียบเทียบคุณภาพของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 แบบเลือกตอบเมื่อตรวจด้วยวิธีการให้คะแนนความรู้บางส่วนด้วยวิธีที่ต่างกัน.* ปริญญานิพนธ์ กศ.ม. (การวิจัยและพัฒนาศักยภาพมนุษย์ แขนงวิชาการทดสอบและวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. อาจารย์ที่ปรึกษาปริญญานิพนธ์: ผู้ช่วยศาสตราจารย์ ดร. เรืองเดช ศิริกิจ, อาจารย์ ดร.วิไลลักษณ์ ลังกา

การวิจัยครั้งนี้มีความมุ่งหมายเพื่อศึกษาฟังก์ชันสารสนเทศของข้อสอบ และฟังก์ชันสารสนเทศของแบบทดสอบด้วยวิธีการตรวจให้คะแนนความรู้บางส่วนโดยใช้วิธีประยุกต์การให้คะแนนของคูมบ์และวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ ได้แก่ นักเรียนชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2559 สำนักงานเขตพื้นที่การศึกษามัธยมศึกษาเขต 2 จังหวัดกรุงเทพมหานคร จำนวน 10 โรงเรียน จำนวน 26 ห้องเรียน จำนวนนักเรียนทั้งหมด 1,200 คน ซึ่งได้มาโดยการสุ่มแบบหลายขั้นตอน (Multi-stage random sampling) เครื่องมือที่ใช้เป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ เรื่องตัวหารร่วมมากและตัวคูณร่วมน้อย จำนวน 25 ข้อ ข้อมูลที่ได้นำมาวิเคราะห์ด้วยสถิติพื้นฐานและวิเคราะห์ค่าฟังก์ชันสารสนเทศของข้อสอบและค่าฟังก์ชันสารสนเทศของแบบทดสอบ โดยวิเคราะห์ตามโมเดล Generalized Partial Credit Model (G-PCM)

ผลการวิจัยสรุปได้ดังนี้

การศึกษาค่าฟังก์ชันสารสนเทศของข้อสอบ พบว่า วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ ให้ค่าฟังก์ชันสารสนเทศของข้อสอบสูงกว่าวิธีประยุกต์การให้คะแนนของคูมบ์

การศึกษาค่าฟังก์ชันสารสนเทศของแบบทดสอบ พบว่า วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ให้ค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบทดสอบเท่ากับ 4.84 ซึ่งสูงกว่าวิธีประยุกต์การให้คะแนนของคูมบ์ที่ให้ค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบทดสอบเท่ากับ 4.16

A COMPARATIVE STUDY OF MULTIPLE CHOICE MATHEMATICS TEST QUALITY WHEN SCORING WITH PARTIAL CREDIT METHOD FOR MATTAYOM 1 STUDENTS

AN ABSTRACT

BY

WARARAT SANGSUK

Presented in Partial Fulfillment of the Requirements for the

Master of Education Degree in Research and Development on Human Potentials

at Srinakharinwirot University

August 2017

Wararat Sangsuk. (2017). *A Comparative Study of Multiple Choice Mathematics Test Quality When Scoring With Partial Credit Method for Mattayom 1 Students.* Master thesis. M.Ed. (Research and development on Human Potentials: Educational Testing and Measurement). Bangkok: Graduate School, Srinakarinwirot University. Advisor Committee: Asst. Prof. Dr. Ruangdach Sirikit , Dr. Wilailak Langka

The research aimed to study the item information function and information function of test when scoring with two different partial knowledge methods, between the modified Coombs approach and the Arnold scoring/response method. Through multi-stage random sampling, the subjects comprised 1,200 Mathayom Suksa 1 students from 26 classrooms, studying in the second semester of 2016 academic year, in 10 schools under the authority of Secondary Educational Service Area Office 2. The instrument was a multiple choice mathematics achievement test in greatest common factors and least common multiples with 25 items. The descriptive statistics, item information function and test information function scoring, by Generalized Partial Credit Model (G-PCM), were analyzed.

The results revealed the following;

According to the item information function (IIF), the Arnold scoring/response method was significantly higher than the modified Coombs approach.

According to test information function (TIF), the Arnold scoring/response method significantly higher (TIF=4.84) than the modified Coombs approach (TIF=4.16).

ปริญญานิพนธ์
เรื่อง
การศึกษาเปรียบเทียบคุณภาพของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1 แบบเลือกตอบเมื่อตรวจด้วยวิธีการให้คะแนนความรู้บางส่วนด้วยวิธีที่ต่างกัน
ของ
วรารัตน์ แสงสุข

ได้รับอนุมัติจากบัณฑิตวิทยาลัยให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาการศึกษามหาบัณฑิต สาขาวิชาการวิจัยและพัฒนาศักยภาพมนุษย์
ของมหาวิทยาลัยศรีนครินทรวิโรฒ

....คณบดีบัณฑิตวิทยาลัย
(ผู้ช่วยศาสตราจารย์ นายแพทย์ฉัตรชัย เอกปัญญาสกุล)
วันที่...10...เดือน...สิงหาคม.......พ.ศ. 2560...

อาจารย์ที่ปรึกษาปริญญานิพนธ์

....ที่ปรึกษาหลัก
(ผู้ช่วยศาสตราจารย์ ดร. เรืองเดช ศิริกิจ)

....ที่ปรึกษาร่วม
(อาจารย์ ดร. วิไลลักษณ์ ลังกา)

คณะกรรมการสอบปากเปล่า

....ประธาน
(ผู้ช่วยศาสตราจารย์ ดร. สุวิมล กฤชคฤหาสน์)

....กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร. เรืองเดช ศิริกิจ)

....กรรมการ
(อาจารย์ ดร. วิไลลักษณ์ ลังกา)

....กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร. สุรชัย มีชาญ)

ประกาศคุณูปการ

ปริญญานิพนธ์ฉบับนี้ สำเร็จลุล่วงได้ด้วยดี เนื่องมาจากผู้วิจัยได้รับความกรุณาอย่างยิ่งจากผู้ช่วยศาสตราจารย์ ดร. เรืองเดช ศิริกิจ อาจารย์ที่ปรึกษาปริญญานิพนธ์หลักและ อาจารย์ ดร. วิไลลักษณ์ ลังกา อาจารย์ที่ปรึกษาปริญญานิพนธ์ร่วม ที่ได้สละเวลาเพื่อให้คำปรึกษา แนะนำแนวทาง ตลอดจนตรวจสอบแก้ไขข้อบกพร่องต่าง ๆจนปริญญานิพนธ์ฉบับนี้มีความถูกต้องสมบูรณ์ ผู้วิจัยรู้สึกซาบซึ้งและขอขอบพระคุณท่านอาจารย์ทั้งสองเป็นอย่างสูงไว้ ณ ที่นี้

ขอขอบพระคุณผู้ช่วยศาสตราจารย์ ดร. สุวิมล กฤชคฤหาสน์ ประธานสอบปริญญานิพนธ์ ผู้ช่วยศาสตราจารย์ ดร. เรืองเดช ศิริกิจ, อาจารย์ ดร.วิไลลักษณ์ ลังกา และ ผู้ช่วยศาสตราจารย์ ดร. สุรชัย มีชาญ กรรมการสอบปริญญานิพนธ์ ที่ได้ให้คำแนะนำเพิ่มเติม ตรวจสอบ และแก้ไขข้อบกพร่อง เพื่อให้ปริญญานิพนธ์ฉบับนี้มีความสมบูรณ์ชัดเจนยิ่งขึ้น ผู้วิจัยขอขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

ขอขอบพระคุณคณาจารย์ประจำภาควิชาการวัดผลและวิจัยการศึกษาทุกท่าน ที่อบรมสั่งสอนและให้ความรู้ต่าง ๆแก่ผู้วิจัย ซึ่งเป็นประโยชน์อย่างมากในการวิจัยครั้งนี้

ขอขอบพระคุณ อาจารย์ ดร.มนตา ตุลย์เมธาการ อาจารย์ ดร.นวรินทร์ ตาก้อนทอง อาจารย์ปานวาสน์ มหาลวเลิศ อาจารย์พรปวีณ์ ศรียะพันธุ์ และอาจารย์ชัยพร นาควิเชียร ที่ให้ความกรุณาเป็นผู้เชี่ยวชาญในการตรวจสอบเครื่องมือวิจัย แนะนำให้ข้อคิดเห็นที่เป็นประโยชน์และแก้ไขข้อบกพร่องต่างๆ เป็นอย่างดี

ขอขอบพระคุณผู้บริหารและครู-อาจารย์ของโรงเรียนที่เป็นกลุ่มตัวอย่างในเขตพื้นที่การศึกษามัธยมศึกษาเขต 2 จังหวัดกรุงเทพมหานคร ที่อำนวยความสะดวกในการทดลองใช้เครื่องมือและเก็บรวบรวมข้อมูล และขอขอบใจนักเรียนชั้นมัธยมศึกษาปีที่ 1 ที่ให้ความร่วมมือในการเก็บรวบรวมข้อมูล

ขอขอบคุณเพื่อน พี่ และน้องสาขาวิชาการวิจัยและพัฒนาศักยภาพมนุษย์และเพื่อนครูโรงเรียนเทพศิรินทร์ร่มเกล้าทุกท่านที่คอยช่วยเหลือและให้กำลังใจกับผู้วิจัยมาโดยตลอด จนกระทั่งปริญญานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี

ขอขอบพระคุณ คุณพ่อการีม และคุณแม่กิตติยา แสงสุข ที่ให้การสนับสนุนการศึกษาของผู้วิจัยในทุก ๆ ด้าน ให้กำลังใจ จนทำให้ประสบความสำเร็จในการศึกษาครั้งนี้

วรารัตน์ แสงสุข

สารบัญ

บทที่		หน้า
1	บทนำ	1
	ภูมิหลัง	1
	ความมุ่งหมายของการวิจัย	4
	ความสำคัญของการวิจัย	4
	ขอบเขตของการวิจัย	4
	นิยามศัพท์เฉพาะ	5
	กรอบแนวคิดในการวิจัย	7
	สมมติฐานของการวิจัย	8
2	เอกสารและงานวิจัยที่เกี่ยวข้อง	9
	ความรู้ทั่วไปเกี่ยวกับทฤษฎีการตอบสนองข้อสอบ	10
	ความเป็นมาของทฤษฎีการตอบสนองข้อสอบ	10
	แนวคิดสำคัญของทฤษฎีการตอบสนองข้อสอบ	10
	ลักษณะเด่นของทฤษฎีการตอบสนองข้อสอบ	12
	หลักการของทฤษฎีการตอบสนองข้อสอบ	13
	โมเดลการตอบสนองข้อสอบ	15
	พารามิเตอร์ในทฤษฎีการตอบสนองข้อสอบ	16
	ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ	17
	ฟังก์ชันสารสนเทศของข้อสอบ	18
	ฟังก์ชันสารสนเทศของแบบทดสอบ	19
	ประสิทธิภาพสัมพัทธ์	20
	ความคลาดเคลื่อนมาตรฐานของการประมาณค่า	21
	ความรู้ทั่วไปเกี่ยวกับการให้คะแนนความรู้บางส่วน	22
	การทำหน้าที่ต่างกันของข้อสอบ	33
	ความหมายของการทำหน้าที่ต่างกันของข้อสอบและการทำหน้าที่ต่างกันของตัวลวง	33
	หลักการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ	34
	วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ	37

สารบัญ (ต่อ)

บทที่ หน้า

2 (ต่อ)

เกณฑ์เปรียบเทียบคุณภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 39

การแก้ปัญหาทางคณิตศาสตร์ 41

การศึกษาเกี่ยวกับแบบทดสอบ 45

แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน 45

แบบทดสอบแบบเลือกตอบ 46

งานวิจัยที่เกี่ยวข้อง 49

งานวิจัยในประเทศ 49

งานวิจัยต่างประเทศ 51

3 วิธีการดำเนินงานวิจัย 53

การกำหนดประชากรและการสุ่มกลุ่มตัวอย่าง 53

การสร้างเครื่องมือที่ใช้ในการวิจัย 55

การเก็บรวบรวมข้อมูล 60

การจัดกระทำข้อมูลและการวิเคราะห์ข้อมูล 61

4 ผลการวิเคราะห์ข้อมูล 66

สัญลักษณ์ที่ใช้ในการวิเคราะห์ข้อมูล 66

การเสนอผลการวิเคราะห์ข้อมูล 66

5 สรุปผล อภิปราย และข้อเสนอแนะ 83

สรุปผลการวิจัย 85

อภิปรายผลการวิจัย 86

ข้อเสนอแนะ 88

บรรณานุกรม 89

สารบัญ (ต่อ)

บทที่ หน้า

ภาคผนวก 93

ภาคผนวก ก 94

ภาคผนวก ข 114

ภาคผนวก ค 122

ภาคผนวก ง 132

ประวัติย่อผู้วิจัย 134

บัญชีตาราง

ตาราง		หน้า
1	วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่มีการตรวจให้คะแนน แบบทวิภาค(Dichotomous DIF) และพหุวิภาค (Polytomous DIF)	38
2	**คุณภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ**	40
3	ข้อมูลการสุ่มห้องเรียนแต่ละโรงเรียนของประชากร ชั้นมัธยมศึกษาปีที่ 1	55
4	การตรวจให้คะแนนโดยใช้วิธีประยุกต์การให้คะแนนของคูมบ์โดยจำนวนตัวเลือกมี 4 ตัวเลือก	57
5	การตรวจให้คะแนนโดยใช้วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์โดยจำนวนตัวเลือกมี 4 ตัวเลือก	58
6	ค่าสถิติพื้นฐานของคะแนนสอบจำแนกตามวิธีการตรวจให้คะแนน	67
7	ค่าไอเกน และร้อยละของความแปรปรวนขององค์ประกอบของแบบทดสอบจำนวน 25 ข้อ เมื่อตรวจด้วยวิธีประยุกต์การให้คะแนนของคูมบ์	68
8	ค่าไอเกน และร้อยละของความแปรปรวนขององค์ประกอบของแบบทดสอบจำนวน 25 ข้อ เมื่อตรวจด้วยวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์	70
9	ค่าพารามิเตอร์อำนาจำแนกของข้อสอบ (a) ที่วิเคราะห์จากคะแนนที่ได้จากวิธีการตอบและตรวจให้คะแนนแบบต่างๆ ทั้ง 2 วิธี โดยวิเคราะห์ตามโมเดล GPCM ...	71
10	ค่าพารามิเตอร์ความยากง่ายของข้อสอบ (b) ที่วิเคราะห์จากคะแนนที่ได้จากวิธีการตอบและตรวจให้คะแนนแบบต่างๆ ทั้ง 2 วิธี โดยวิเคราะห์ตามโมเดล GPCM	73
11	ค่าพารามิเตอร์ระดับความยากของขั้นการตอบ(δ_{ij}) ที่วิเคราะห์จากคะแนนที่ได้จากวิธีการตอบและตรวจให้คะแนนทั้ง 2 วิธี โดยวิเคราะห์ตามโมเดล GPCM ...	74
12	ค่าฟังก์ชันสารสนเทศสูงสุดของข้อสอบ จำแนกตามรายข้อและวิธีการตรวจให้คะแนน	76
13	ค่าฟังก์ชันสารสนเทศของแบบทดสอบที่วิเคราะห์จากคะแนนที่ได้จากวิธีการตอบและตรวจให้คะแนนทั้ง 2 แบบ	77
14	ค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบทดสอบ	79
15	ประสิทธิภาพสัมพัทธ์เฉลี่ยของแบบทดสอบ	80
16	ค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าของแบบทดสอบ	81
17	ผลการตรวจสอบความเที่ยงตรงด้านเนื้อหา(Content Validity) โดยผู้เชี่ยวชาญ...	95
18	ค่าความยากง่าย (p) ค่าอำนาจจำแนกรายข้อ(r) การทำหน้าที่ต่างกันของข้อสอบทั้งตัวถูกและตัวลวง	97

บัญชีตาราง (ต่อ)

ตาราง		หน้า
19	ตารางวิเคราะห์ข้อสอบวิชาคณิตศาสตร์ เรื่องตัวหารร่วมมากและตัวคูณร่วมน้อย..	115
20	ตารางแสดงจำนวนข้อสอบในแบบทดสอบวิชาคณิตศาสตร์ เรื่องตัวหารร่วมมาก และตัวคูณร่วมน้อย	116
21	ค่าสถิติพื้นฐานของข้อมูลเมื่อตรวจด้วยวิธีประยุกต์การให้คะแนนของคูมบ์	123
22	ค่าสถิติพื้นฐานของข้อมูลเมื่อตรวจด้วยวิธีการตอบและการตรวจให้คะแนนแบบ อาร์โนลด์	123
23	ค่าไอเกน และร้อยละของความแปรปรวนขององค์ประกอบของแบบทดสอบเมื่อ ตรวจด้วยวิธีประยุกต์การให้คะแนนของคูมบ์	124
24	ค่าไอเกน และร้อยละของความแปรปรวนขององค์ประกอบของแบบทดสอบเมื่อ ตรวจด้วยวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์	125
25	GPCM Model Item Parameter Estimates, logit: $a[k(\theta - b) + \Sigma d_k]$ วิธี ประยุกต์การให้คะแนนของคูมบ์	126
26	ค่าพารามิเตอร์ระดับความยากของขั้นการตอบ (δ_{ij}) ที่วิเคราะห์จากคะแนนที่ได้ จากวิธีประยุกต์การให้คะแนนของคูมบ์ โดยวิเคราะห์ตามโมเดล GPCM	127
27	GPCM Model Item Parameter Estimates, logit: $a[k(\theta - b) + \Sigma d_k]$ วิธีการ ตอบและการตรวจให้คะแนนแบบอาร์โนลด์	128
28	ค่าพารามิเตอร์ระดับความยากของขั้นการตอบ (δ_{ij}) ที่วิเคราะห์จากคะแนนที่ได้ จากวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ โดยวิเคราะห์ตามโมเดล GPCM	129
29	Item Information Function Values for Group 1 at 15 Values of θ from -2.8 to 2.8 วิธีประยุกต์การให้คะแนนของคูมบ์	130
30	Item Information Function Values for Group 1 at 15 Values of θ from -2.8 to 2.8 วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์	131

บัญชีภาพประกอบ

ภาพประกอบ		หน้า
1	กรอบแนวคิดการวิจัย	8
2	ความไม่แปรเปลี่ยนของค่าพารามิเตอร์ของข้อสอบที่ประมาณค่าจากกลุ่มที่แตกต่างกัน	12
3	ความสัมพันธ์ระหว่างความสามารถกับพฤติกรรมการตอบสนอง	14
4	เส้นโค้งลักษณะข้อสอบของโมเดลโลจิสแบบ 3 พารามิเตอร์	15
5	ขั้นตอนในการสร้างแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์	58
6	ผลการตรวจสอบความเป็นมิติเดียวกันของข้อสอบเมื่อตรวจด้วยวิธีประยุกต์การให้คะแนนของคูมบ์	69
7	ผลการตรวจสอบความเป็นมิติเดียวกันของข้อสอบเมื่อตรวจด้วยวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์	70
8	เปรียบเทียบโค้งฟังก์ชันสารสนเทศของแบบทดสอบที่มีวิธีการตอบและตรวจให้คะแนนทั้ง 2 วิธี	79

บทที่ 1
บทนำ

ภูมิหลัง

คณิตศาสตร์มีบทบาทสำคัญยิ่งต่อการพัฒนาความคิดมนุษย์ ทำให้มนุษย์มีความคิดสร้างสรรค์ คิดอย่างมีเหตุผล เป็นระบบ มีแบบแผน สามารถวิเคราะห์ปัญหาหรือสถานการณ์ได้อย่างถี่ถ้วน รอบคอบ ช่วยให้คาดการณ์ วางแผน ตัดสินใจ แก้ปัญหา และนำไปใช้ในชีวิตประจำวันได้อย่างถูกต้อง เหมาะสม (กระทรวงศึกษาธิการ. 2551: 54) ซึ่งในการจัดการเรียนการสอนวิชาคณิตศาสตร์นั้นจำเป็นอย่างยิ่งที่จะต้องมีการวัดผลและประเมินผลผู้เรียนอยู่ตลอดเวลาเพื่อตรวจสอบความเข้าใจของผู้เรียน และถ้าหากผู้เรียนไม่เข้าใจในการเรียนแล้วนั้น การจัดการเรียนการสอนจะไม่บรรลุถึงเป้าหมายที่วางไว้ ดังนั้นครูผู้สอนจึงมีความจำเป็นจะต้องมีความรู้ความเข้าใจในการวัดและประเมินผล เพื่อประโยชน์กับผู้เรียน ซึ่งการวัดและประเมินผลจะได้ผลดีมีมาตรฐานย่อมขึ้นอยู่กับเครื่องมือที่ใช้เป็นหลักสำคัญ(สุเทพ สันติวรานนท์. 2534: 72)

ธรรมชาติของวิชาคณิตศาสตร์เป็นวิชาที่เกี่ยวกับทักษะกระบวนการ ซึ่งแต่ละเนื้อหาจะเน้นและสอดแทรกการแก้ปัญหาเป็นหลัก เพื่อนำไปใช้ในชีวิตประจำวัน การสร้างเครื่องมือจึงเป็นสิ่งที่ควรให้ความสำคัญเป็นอันดับต้นๆ ในการวัดและประเมินผล ซึ่งการประเมินทักษะและกระบวนการทางคณิตศาสตร์ สามารถทำได้โดยใช้เครื่องมือในการประเมินที่หลากหลาย เช่น แบบทดสอบ แบบสัมภาษณ์ แบบสอบถาม แบบสังเกต แบบบันทึกการเรียนรู้ แบบประเมินพฤติกรรม แฟ้มสะสมงาน โครงงานคณิตศาสตร์ ปัญหา/สถานการณ์จำลอง แบบวัดการปฏิบัติ แบบตรวจรายการ/แบบสำรวจตัวเอง และนิทรรศการ แต่อย่างไรก็ตาม ในทางปฏิบัติการประเมินทางคณิตศาสตร์ยังคงเป็นการประเมินผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ที่เป็นความรู้ความเข้าใจมากกว่าทักษะและกระบวนการทางคณิตศาสตร์ และหากเป็นการประเมินทักษะ การเลือกใช้เครื่องมือก็ยังประเมินความสามารถที่แท้จริงของผู้เรียนได้ไม่มากนัก ผู้สอนส่วนใหญ่นิยมใช้ข้อสอบปรนัยในการวัดและประเมินผลการเรียนรู้คณิตศาสตร์ เนื่องจากมีความสะดวกในการสร้างข้อสอบและการตรวจให้คะแนน โดยอาจมีหลากหลายแบบคือ 1. แบบเลือกตอบ 2.แบบถูก-ผิด 3. แบบจับคู่ 4. แบบเรียงลำดับ 5. แบบเติมคำ ข้อสอบปรนัยทั้งห้าแบบมักถูกนำมาใช้วัดทักษะการแก้ปัญหา ซึ่งตรงกับพฤติกรรมด้านพุทธิพิสัยระดับการนำไปใช้ โดยมักใช้วัดความสามารถในการหาคำตอบสุดท้ายหรือแก้ปัญหามากกว่าวัดกระบวนการระหว่างการแก้ปัญหา อย่างไรก็ตาม การประเมินทักษะและกระบวนการทางคณิตศาสตร์นั้น สามารถใช้ข้อสอบปรนัยได้ทุกแบบ แต่จะประเมินทักษะที่ต้องการได้หรือไม่นั้น ขึ้นอยู่กับคำถามหรือปัญหาที่ถาม หากผู้เรียนจำเป็นต้องใช้ทักษะที่ต้องการวัดในกระบวนการคิดและการทำงานเพื่อตอบคำถามหรือปัญหา ก็นับว่าคำถามหรือปัญหานั้นวัดทักษะที่ต้องการได้ การสร้างข้อสอบปรนัยเพื่อประเมินทักษะและกระบวนการทางคณิตศาสตร์ที่ผู้เรียนจะนำมาใช้เป็นหลัก จากนั้นจึงสร้างคำถามหรือปัญหาภายหลัง (อัมพร ม้าคนอง. 2553: 165-

166) แต่ถึงอย่างไรการสร้างข้อสอบปรนัยก็ยังมีจุดอ่อนที่สำคัญ คือ เปิดโอกาสให้มีการตอบถูกโดยการเดา ปัญหาการตอบแบบเดาสุ่มจึงเป็นแหล่งความคลาดเคลื่อนในการวัด นอกจากนี้ผลการวัดที่ได้ไม่สามารถให้สารสนเทศที่เพียงพอที่จะจำแนกผู้ตอบในระดับต่าง ๆของความรู้ ระหว่างผู้ที่รู้ ผู้ที่มีความรู้บางส่วน และผู้ที่ไม่รู้ (เอมอร จังศิริพรปกรณ์. 2545: 21)

ในปัจจุบันทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) ถือเป็นทฤษฎีการทดสอบแนวใหม่ที่ถูกนำมาใช้ในการพัฒนาเครื่องมือและใช้ในระบบการทดสอบอย่างกว้างขวาง ซึ่งทฤษฎีการตอบสนองข้อสอบ เสนอแนวคิดที่ว่า ความน่าจะเป็นของการตอบข้อสอบได้ถูกต้องขึ้นอยู่กับความสามารถจริงของผู้ตอบ (ศิริชัย กาญจนวาสี. 2555: 7) และโมเดลการตอบสนองข้อสอบที่นำมาใช้ในการทดสอบยุคปัจจุบันที่สำคัญคือแบบตรวจให้คะแนนมากกว่า 2 ค่า ได้แก่ การตรวจให้คะแนนความรู้บางส่วน (Partial Credit) โดยการตรวจให้คะแนนความรู้บางส่วนนั้น เป็นการให้คะแนนที่ใกล้เคียงกับความรู้ที่แท้จริงของผู้ตอบ ซึ่งมีส่วนช่วยลดความคลาดเคลื่อนจากแหล่งการเดาสุ่มของผู้ตอบลงทำให้คะแนนมีความเชื่อมั่นมากขึ้น ในขณะเดียวกันก็ส่งผลต่อความเที่ยงตรงของกระบวนการวัดด้วย(Frary. 1980) จากการศึกษาวิธีการให้คะแนนความรู้บางส่วนสรุปเป็นประเภทใหญ่ๆ ได้ 4 วิธี ดังนี้ (เอมอร จังศิริพรปกรณ์. 2545; อ้างอิงจาก Simon, Budescu; & Nevo. 1997) 1. การให้น้ำหนักแก่ข้อสอบที่แตกต่างกัน (Differential Item Weighting) หลักการของวิธีนี้ คือข้อสอบที่มีคุณภาพสูงจะได้น้ำหนักมากกว่าข้อสอบที่มีคุณภาพต่ำโดยใช้ผลที่ได้จากการวิเคราะห์ข้อสอบเป็นเกณฑ์ คือ ค่าความยาก ความสามารถ ความเที่ยงตรง หรือ การพิจารณาของผู้เชี่ยวชาญ 2. การให้น้ำหนักแก่ตัวเลือกที่แตกต่างกัน (Differential Option Weighting) หลักการของวิธีนี้ คือ ความรู้บางส่วนสามารถถูกวัดได้จากการให้น้ำหนักคะแนนแก่ตัวเลือก วิธีการได้ 2 วิธี คือ วิธีแรกเป็นการให้น้ำหนักแก่ตัวเลือกโดยใช้วิจารณญาณของผู้เชี่ยวชาญ หรือจากทฤษฎีของโครงสร้างความรู้ ส่วนวิธีที่สองเป็นการให้น้ำหนักจากข้อมูลเชิงประจักษ์ที่ผ่านมาในอดีต หรือปัจจุบันโดยอาศัยสิ่งที่น่าสนใจของตัวเลือก 3. การเปลี่ยนแปลงโครงสร้างของข้อสอบ (Changing the Item Structure) วิธีนี้มีรูปแบบของโครงสร้างข้อสอบ หรือ คำอธิบายที่แตกต่างจากแบบเลือกตอบธรรมดา เช่น ข้อสอบที่มี 2 ตัวเลือก คือ ถูก และ ผิด ผู้ตอบจะต้องเลือกคำตอบอย่างใดอย่างหนึ่ง คะแนนที่ได้จากผลรวมของจำนวนคำตอบที่ถูกต้อง เนื่องจากการตอบแต่ละข้อใช้เวลาน้อย จึงต้องเพิ่มจำนวนข้อให้มากขึ้นกว่าข้อสอบแบบเลือกตอบธรรมดา 4. การเปลี่ยนแปลงวิธีการตอบ (Changing the Response Method) คือ ใช้น้ำหนักที่ให้โดยผู้สอบเอง ซึ่งสะท้อนถึงความรู้ที่มีอยู่ในตัวผู้สอบ การประเมินด้วยตนเองนี้เป็นวิธีที่ดีที่สุดในการได้มาซึ่งสารสนเทศเกี่ยวกับความรู้ของผู้สอบในแต่ละข้อ ผู้สอบจะให้สารสนเทศอย่างสมบูรณ์เกี่ยวกับน้ำหนัก วิธีการให้คะแนน และการควบคุมระดับความมั่นใจในการให้ ซึ่งจะช่วยลดการเดาและความคลาดเคลื่อนในการวัด วิธีนี้ใช้รูปแบบข้อสอบเหมือนแบบทดสอบชนิดเลือกตอบทั่วไปเพียงแต่คำสั่ง วิธีการตอบ และการให้คะแนนแตกต่างจากแบบทดสอบชนิดเลือกตอบธรรมดา ซึ่งแต่ละวิธีใช้วิธีการที่ไม่ยุ่งยาก โดยแบ่งเป็นชนิดต่างๆได้ 6 วิธี คือ 1. วิธีการแก้การเดา (Correction for guessing) 2. วิธีการให้คะแนนจากการตัดตัวลวง (Elimination Scoring) 3. วิธีการของการทดสอบความน่าจะเป็น

(Probability Testing) 4. วิธีการแสดงความมั่นใจ (Confidence Marking) 5. วิธีการจัดอันดับอย่างสมบูรณ์ (Complete Ordering) 6. วิธีการจัดอันดับเป็นบางส่วน (Partial Ordering)

วิธีการของการตรวจให้คะแนนความรู้บางส่วนมีผู้คนจำนวนมากให้ความสนใจ ซึ่งได้นำวิธีการตรวจให้คะแนนด้วยวิธีการต่างๆ มาศึกษาโดยเฉพาะวิธีการเปลี่ยนแปลงวิธีการตอบ (Changing the Response Method) เนื่องจากเป็นวิธีที่สะท้อนถึงความรู้ที่มีอยู่ในตัวผู้สอบมากที่สุด การประเมินด้วยตนเองนี้เป็นวิธีที่ดีที่สุดในการได้มาซึ่งสารสนเทศเกี่ยวกับความรู้ของผู้สอบในแต่ละข้อ ดังจะเห็นได้จากงานวิจัยของ สุพจน์ เกิดสุวรรณ์ (2545: 147-161) ได้เปรียบเทียบคุณภาพวิธีการตอบและตรวจให้คะแนนตามวิธีของคูมบ์ วิธีของอาร์โนลด์ วิธีของเดรสเซลและชมิท วิธีที่ประยุกต์จากวิธีของคูมบ์ วิธีที่ประยุกต์จากวิธีของอาร์โนลด์ วิธีที่ประยุกต์จากวิธีของเดรสเซลและชมิท และวิธีการตรวจให้คะแนนแบบประเพณีนิยม ผลการศึกษาพบว่า วิธีที่ประยุกต์จากวิธีของคูมบ์ ให้ค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบทดสอบและประสิทธิภาพสัมพัทธ์เฉลี่ยของแบบทดสอบสูงกว่าวิธีอื่นๆ รองลงมาคือวิธีของอาร์โนลด์ และวิธีที่ประยุกต์จากวิธีของเดรสเซลและชมิดตามลำดับ ส่วนวิธีประเพณีนิยม จะให้ค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบทดสอบและประสิทธิภาพสัมพัทธ์เฉลี่ยของแบบทดสอบต่ำกว่าทุกวิธี และในปีเดียวกัน เอมอร จังศิริพรปกรณ์ (2545: 77-85) ได้เปรียบเทียบคุณภาพของแบบทดสอบเลือกตอบด้วยวิธีการตรวจให้คะแนนแบบประเพณีนิยมกับวิธีการตรวจให้คะแนนความรู้บางส่วน ซึ่งประกอบด้วยวิธีประยุกต์การให้คะแนนของคูมบ์ และวิธีการประยุกต์การให้คะแนนของเดรสเซลและสมิท โดยทั้งสองวิธีการนี้อยู่ในวิธีการให้คะแนนจากการตัดตัวลวง (Elimination Scoring) ผลการศึกษาพบว่า วิธีประยุกต์การให้คะแนนของคูมบ์ มีค่าฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบทดสอบ และอัตราส่วนสารสนเทศเฉลี่ยสูงที่สุด รองลงมาคือวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิทส่วนวิธีประเพณีนิยมมีค่าต่ำสุด และอีกหนึ่งงานวิจัยที่น่าสนใจคือ งานวิจัยของ รัตนา ไชยตรี (2546: 125) ได้เปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบทดสอบ และอัตราส่วนสารสนเทศ ระหว่างวิธีการตอบและตรวจให้คะแนนความรู้บางส่วนที่ประยุกต์วิธีของคูมบ์ ประยุกต์วิธีของเดรสเซลและชมิดท์ และวิธีการตอบโดยบอกระดับความมั่นใจ ผลการวิจัยพบว่า วิธีการตอบและตรวจให้คะแนนความรู้บางส่วนตามวิธีการตอบโดยบอกระดับความมั่นใจ ส่วนใหญ่จะให้ค่าฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบทดสอบ และอัตราส่วนสารสนเทศสูงสุด รองลงมาคือ วิธีการตอบและตรวจให้คะแนนวิธีที่ประยุกต์การให้คะแนนของเดรสเซลและชมิดท์ และวิธีที่ประยุกต์การให้คะแนนของคูมบ์ตามลำดับ

จากข้อมูลต่างๆ เหล่านี้ทำให้ผู้วิจัยสนใจที่จะนำวิธีการของการตรวจให้คะแนนความรู้บางส่วนมาศึกษาเปรียบเทียบเพื่อหาคุณภาพระหว่างวิธีประยุกต์การให้คะแนนของคูมบ์ (Modified Coombs Approach) และวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ (The Arnold Scoring/Response Method) ของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ เพื่อนำสารสนเทศที่ได้มาจำแนกผู้เรียน ระหว่างผู้ที่มีความรู้จริง ผู้ที่มีความรู้บางส่วน และผู้ที่ไม่มีความรู้ (ศิริชัย กาญจนวาสี. 2555: 10) และเพื่อที่จะได้ทราบว่าวิธีการตรวจให้คะแนนแต่ละวิธีเหมาะสมกับ

เด็กกลุ่มใดมากที่สุด ทั้งนี้ทฤษฎีการทดสอบแนวใหม่จึงน่าจะให้ผลการวัดที่ชัดเจนตรงประเด็นมากขึ้น อันจะเป็นประโยชน์ต่อการพัฒนาข้อสอบและแบบทดสอบให้สามารถนำไปใช้ในสถานการณ์ต่าง ๆของการทดสอบได้อย่างมีประสิทธิภาพ สามารถระบุนัยทั่วไปของคะแนนจริง คุณภาพของแบบทดสอบตามเงื่อนไขของการทดสอบ รวมทั้งการประเมินความสามารถที่แท้จริงของผู้สอบ และบรรยายพฤติกรรมการตอบสนองข้อสอบของผู้สอบได้เป็นอย่างดี

ความมุ่งหมายของการวิจัย

เพื่อศึกษาฟังก์ชันสารสนเทศของข้อสอบ (Item Information Function) และฟังก์ชันสารสนเทศของแบบทดสอบ (Test Information Function) ด้วยวิธีการตรวจให้คะแนนความรู้บางส่วนด้วยวิธีประยุกต์การให้คะแนนของคูมบ์(Modified Coombs Approach) และวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์(The Arnold Scoring/Response Method)

ความสำคัญของการวิจัย

ทำให้ทราบถึงสารสนเทศที่ได้จากการตรวจให้คะแนนความรู้บางส่วนระหว่างวิธีประยุกต์การให้คะแนนของคูมบ์ (Modified Coombs Approach) และวิธีการตอบและการตรวจให้คะแนนแบบ อาร์โนลด์ (The Arnold Scoring/Response Method) ว่าจะส่งผลต่อฟังก์ชันสารสนเทศของข้อสอบ (Item Information Function) และฟังก์ชันสารสนเทศของแบบทดสอบ (Test Information Function) อย่างไร และให้ผลสอดคล้องหรือแตกต่างกันหรือไม่ ซึ่งถ้าวิธีการตอบและการตรวจให้คะแนนให้ผลสอดคล้องหรือใกล้เคียงกัน เราก็สามารถเลือกใช้วิธีการได้อย่างหลากหลาย แต่ถ้าวิธีการตอบและการตรวจให้คะแนนให้ผลแตกต่างกัน เราก็สามารถที่จะเลือกใช้วิธีการที่ดีที่มีประสิทธิภาพมากกว่า และนอกจากนี้ผลที่ได้ยังสามารถอธิบายถึงความสัมพันธ์ของคุณลักษณะภายในหรือความสามารถที่แท้จริงที่มีอยู่ในตัวบุคคล ด้วยการประมาณค่าพารามิเตอร์ของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ทั้งนี้เป็นการเพิ่มสารสนเทศเกี่ยวกับการตอบแบบทดสอบเลือกตอบในวิชาคณิตศาสตร์ เพื่อเป็นประโยชน์กับผู้ที่สนใจและนำไปใช้ในการวัดผลทางการศึกษาต่อไป

ขอบเขตของการวิจัย

ประชากรที่ใช้ในการวิจัย

ประชากรที่ใช้ในการวิจัยครั้งนี้ เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2559 สำนักงานเขตพื้นที่การศึกษามัธยมศึกษาเขต 2 จังหวัดกรุงเทพมหานคร จำนวน 52 โรงเรียน จำนวน 548 ห้องเรียน นักเรียน 23,056 คน (สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน. 2559)

กลุ่มตัวอย่างที่ใช้ในการวิจัย

กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2559 สำนักงานเขตพื้นที่การศึกษามัธยมศึกษาเขต 2 จังหวัดกรุงเทพมหานคร จำนวน 10 โรงเรียน จำนวน 26 ห้องเรียน จำนวนนักเรียนทั้งหมด 1,200 คน ซึ่งได้มาโดยการสุ่มแบบหลายขั้นตอน (Multi-stage random sampling)

ตัวแปรที่ศึกษา

ตัวแปรอิสระ ได้แก่ วิธีตรวจให้คะแนนความรู้บางส่วน 2 วิธี ดังนี้

1. วิธีประยุกต์การให้คะแนนของคูมบ์ (Modified Coombs Approach)
2. วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ (The Arnold Scoring/Response Method)

ตัวแปรตาม

1. ฟังก์ชันสารสนเทศของข้อสอบ (Item Information Function)
2. ฟังก์ชันสารสนเทศของแบบทดสอบ (Test Information Function)

นิยามศัพท์เฉพาะ

การตรวจด้วยวิธีการให้คะแนนความรู้บางส่วนด้วยวิธีที่ต่างกัน หมายถึง การตรวจให้คะแนนจากการสอบด้วยวิธีการ 2 วิธีการ วิธีการที่หนึ่ง คือ วิธีประยุกต์การให้คะแนนของคูมบ์ (Modified Coombs Approach) วิธีการที่สอง คือ วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ (The Arnold Scoring/Response Method)

ผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ เรื่องตัวหารร่วมมากและตัวคูณร่วมน้อย หมายถึง คะแนนที่ได้จากการทำแบบทดสอบแบบเลือกตอบเรื่องตัวหารร่วมมากและตัวคูณร่วมน้อยของนักเรียนชั้นมัธยมศึกษาปีที่ 1 ซึ่งประกอบด้วยเนื้อหาที่เกี่ยวกับตัวประกอบ จำนวนเฉพาะ ตัวหารร่วมมากและตัวคูณร่วมน้อย

คะแนนความรู้บางส่วน หมายถึง คะแนนที่แสดงถึงความสามารถที่แท้จริงของผู้ตอบจากการทำแบบทดสอบวิชาคณิตศาสตร์ เรื่อง ตัวหารร่วมมากและตัวคูณร่วมน้อย ด้วยวิธีประยุกต์การให้คะแนนของคูมบ์(Modified Coombs Approach) และวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์(The Arnold Scoring/Response Method)

ฟังก์ชันสารสนเทศของข้อสอบ (Item Information Function)หมายถึง ค่าที่แสดงถึงความถูกต้องแม่นยำในการประมาณความสามารถของผู้สอบจากการตอบข้อสอบในข้อนั้นๆ ซึ่งประกอบด้วย ค่าพารามิเตอร์ความยาก ค่าพารามิเตอร์อำนาจจำแนก และค่าความแปรปรวนของคะแนนแต่ละข้อ

ฟังก์ชันสารสนเทศของแบบทดสอบ (Test Information Function) หมายถึง ค่าที่แสดงถึงความถูกต้องแม่นยำในการประมาณความสามารถของผู้สอบจากการตอบแบบทดสอบนั้น ซึ่งเกิดจากผลรวมของค่าฟังก์ชันสารสนเทศของข้อสอบแต่ละข้อรวมกันทั้งฉบับ

ความคลาดเคลื่อนมาตรฐานของการประมาณค่า (Standard Error of Estimation) หมายถึง ส่วนเบี่ยงเบนมาตรฐานของการแจกแจงความน่าจะเป็น ของค่าประมาณความสามารถที่แท้จริง (θ) ซึ่งเป็นค่าสัดส่วนผกผันกับความถูกต้องแม่นยำของการประมาณค่าความสามารถ หรือค่าฟังก์ชันสารสนเทศของแบบทดสอบ

ค่าประสิทธิภาพสัมพัทธ์เฉลี่ย (ratio of average information: RAI) หมายถึง ค่าที่ได้จากการคำนวณซึ่งเป็นอัตราส่วนระหว่างค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบทดสอบของวิธีการตอบและตรวจให้คะแนนแต่ละวิธี

วิธีประยุกต์การให้คะแนนของคูมบ์(Modified Coombs Approach) หมายถึง วิธีการตอบที่ให้ผู้สอบเลือกตอบโดยการตัดตัวลวงออกไป และให้คะแนนตามจำนวนตัวลวงที่ตัดออกไปได้ถูกต้อง แต่ถ้าตัดตัวเลือกที่ถูกต้องออกไปด้วยคะแนนรายข้อจะได้เท่ากับ 0 ซึ่งคะแนนจะอยู่ในช่วง 0 ถึง $K-1$ คะแนน เมื่อ K คือจำนวนตัวเลือกทั้งหมด

วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์(The Arnold Scoring/Response Method) หมายถึง วิธีการตอบที่ให้ผู้สอบเลือกตอบโดยการตัดตัวลวงออกไป และจะให้คะแนน

ตามสูตร $C_d = (p)(d/(K-d))$

โดยที่ C_d คือคะแนนเมื่อสามารถตัดตัวลวงบางตัวออกไปได้ และคะแนนที่เพิ่มขึ้นจากการเดาเป็น 0

d คือจำนวนตัวลวงที่ตัดออกได้ถูกต้อง

K คือจำนวนตัวเลือก

p คือ $1/(K-1)$

ส่วนในกรณีที่ผู้สอบตัดตัวเลือกที่ถูกเข้าไว้เป็นตัวลวงจะได้คะแนนลงโทษคือ $-p$

การทำหน้าที่ต่างกันของข้อสอบ (Differential Item functioning: DIF) หมายถึง ข้อสอบที่นำไปวัดผู้สอบที่มีความสามารถหรือคุณลักษณะเท่ากันจากต่างกลุ่มกัน แต่มีโอกาสในการตอบถูกแตกต่างกัน เนื่องด้วยมีลักษณะบางอย่างที่แตกต่างกัน ทำให้ได้ผลการตอบข้อสอบที่ถูกต้องไม่เท่ากันทำให้เกิดการได้เปรียบเสียเปรียบจากการตอบข้อสอบเดียวกัน ในงานวิจัยนี้ตรวจสอบลักษณะที่แตกต่างกัน คือ เพศ เพื่อใช้เป็นเกณฑ์ในการคัดเลือกข้อสอบ

การทำหน้าที่ต่างกันของตัวลวง (Differential Distractor functioning: DDF) หมายถึง การตรวจสอบความแตกต่างในการตอบตัวเลือกที่เป็นตัวลวงเมื่อผู้สอบต่างกลุ่มกัน ซึ่งมี

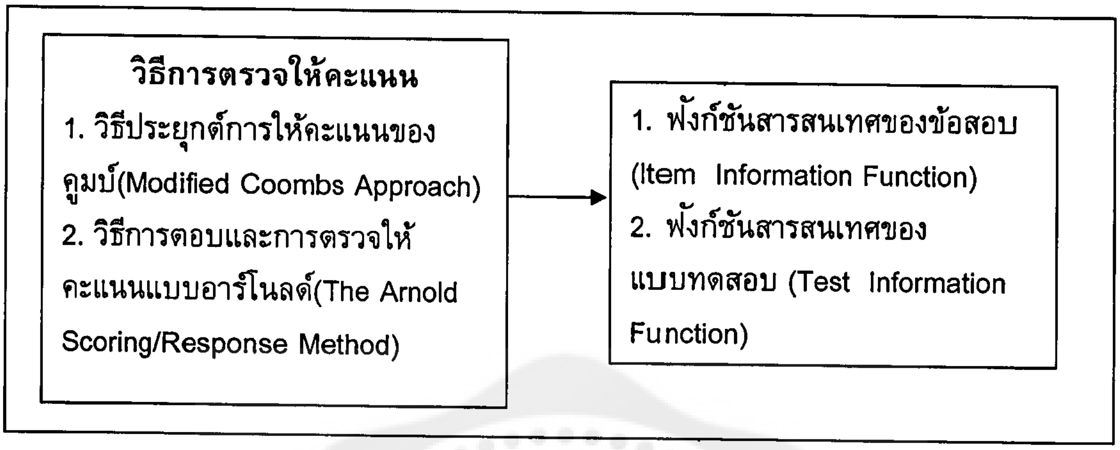
ความสามารถเท่ากัน จะมีโอกาสในการเลือกตัวลวงได้แตกต่างกัน ทำให้ได้ผลการตอบข้อสอบที่ถูกต้องไม่เท่ากันทำให้เกิดการได้เปรียบเสียเปรียบจากการตอบข้อสอบเดียวกัน ในงานวิจัยนี้ตรวจสอบลักษณะที่แตกต่างกัน คือ เพศ เพื่อใช้เป็นเกณฑ์ในการคัดเลือกข้อสอบ

กรอบแนวคิดในการวิจัย

จากการศึกษาวิธีการตอบและการตรวจให้คะแนนความรู้บางส่วนทำให้ทราบว่ามีวิธีการต่างๆ มากมาย ซึ่งสามารถแบ่งเป็นประเภทใหญ่ๆ 4 ประเภท คือ 1. การให้น้ำหนักแก่ข้อสอบที่แตกต่างกัน (Differential Item Weighting) 2. การให้น้ำหนักแก่ตัวเลือกที่แตกต่างกัน (Differential Option Weighting) 3. การเปลี่ยนแปลงโครงสร้างของข้อสอบ (Changing the Item Structure) 4. การเปลี่ยนแปลงวิธีการตอบ (Changing the Response Method) ซึ่งวิธีการตอบและการตรวจให้คะแนนความรู้บางส่วนที่ผู้วิจัยให้ความสนใจ คือ การเปลี่ยนแปลงวิธีการตอบ (Changing the Response Method) ด้วยวิธีการตัดตัวลวง ทั้งนี้ผู้วิจัยได้ศึกษา 2 วิธีคือ วิธีประยุกต์การให้คะแนนของคูมบ์(Modified Coombs Approach)และวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์(The Arnold Scoring/Response Method) เนื่องจากในอดีตคนมักใช้วิธีการตรวจแบบประเพณีนิยมจำนวนมากเพราะง่ายและสะดวกต่อการดำเนินการสอบ แต่วิธีการให้คะแนนแบบประเพณีนิยมยังไม่สามารถแก้ไขปัญหาการเดาสุ่มของผู้สอบได้ ซึ่งจากการศึกษางานวิจัยของสุพจน์ เกิดสุวรรณ์ (2545: 147-161) พบว่า วิธีที่ประยุกต์จากวิธีของคูมบ์ ให้ค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบทดสอบและประสิทธิภาพสัมพัทธ์เฉลี่ยของแบบทดสอบสูงกว่าวิธีอื่นๆ รองลงมาคือวิธีของอาร์โนลด์ ส่วนวิธีประเพณีนิยมจะให้ค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบทดสอบและประสิทธิภาพสัมพัทธ์เฉลี่ยของแบบทดสอบต่ำกว่าทุกวิธี และจากงานวิจัยของ เอมอร จังศิริพรปกรณ์ (2545: 77-85) พบว่า วิธีประยุกต์การให้คะแนนของคูมบ์ มีค่าฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบทดสอบ และอัตราส่วนสารสนเทศเฉลี่ยสูงที่สุด ส่วนวิธีประเพณีนิยมมีค่าต่ำสุด และจากงานวิจัยของ รัดนา ไชยตรี (2546: 125) ได้เปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบทดสอบ และอัตราส่วนสารสนเทศ ระหว่างวิธีการตอบและตรวจให้คะแนนความรู้บางส่วนที่ประยุกต์วิธีของคูมบ์ ประยุกต์วิธีของเดรสเซลและชมิดท์ และวิธีการตอบโดยบอกระดับความมั่นใจ ผลการวิจัยพบว่า วิธีการตอบและตรวจให้คะแนนความรู้บางส่วนตามวิธีการตอบโดยบอกระดับความมั่นใจ ส่วนใหญ่จะให้ค่าฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบทดสอบ และอัตราส่วนสารสนเทศสูงสุด รองลงมาคือ วิธีการตอบและตรวจให้คะแนนวิธีที่ประยุกต์การให้คะแนนของเดรสเซลและชมิดท์ และวิธีที่ประยุกต์การให้คะแนนของคูมบ์ตามลำดับ

ดังนั้นผู้วิจัยจึงต้องการศึกษาว่าระหว่างวิธีการตัดตัวลวงทั้ง 2 วิธี คือ วิธีประยุกต์การให้คะแนนของคูมบ์และวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ วิธีการใดให้ค่าฟังก์ชันสารสนเทศของข้อสอบและค่าฟังก์ชันสารสนเทศของแบบทดสอบสูงกว่ากัน ทั้งยังศึกษาด้วยว่า

วิธีการตัดตัวลวงแบบใดที่จะให้สารสนเทศสูงที่สุด และเพื่อที่จะได้ทราบว่าวิธีการตรวจให้คะแนนวิธีการใดเหมาะสมกับเด็กกลุ่มใดมากที่สุด



ภาพประกอบ 1 กรอบแนวคิดการวิจัย

สมมติฐานในการวิจัย

ค่าฟังก์ชันสารสนเทศของข้อสอบ (Item Information Function) และค่าฟังก์ชันสารสนเทศของแบบทดสอบ (Test Information Function) ระหว่างการตรวจให้คะแนนความรู้บางส่วนด้วยวิธีประยุกต์การให้คะแนนของคูมบ์ (Modified Coombs Approach) และวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ (The Arnold Scoring/Response Method) แตกต่างกัน

บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

ในการวิจัยครั้งนี้ ผู้วิจัยได้ศึกษาแนวคิดและทฤษฎี ตลอดจนงานวิจัยที่เกี่ยวข้อง ซึ่งผู้วิจัยจะนำเสนอประเด็นสำคัญตามลำดับดังต่อไปนี้

1. ความรู้ทั่วไปเกี่ยวกับทฤษฎีการตอบสนองข้อสอบ
    1.1 ความเป็นมาของทฤษฎีการตอบสนองข้อสอบ
    1.2 แนวคิดสำคัญของทฤษฎีการตอบสนองข้อสอบ
    1.3 ลักษณะเด่นของทฤษฎีการตอบสนองข้อสอบ
    1.4 หลักการของทฤษฎีการตอบสนองข้อสอบ
    1.5 โมเดลการตอบสนองข้อสอบ
    1.6 พารามิเตอร์ในทฤษฎีการตอบสนองข้อสอบ
    1.7 ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ
    1.8 ฟังก์ชันสารสนเทศของข้อสอบ
    1.9 ฟังก์ชันสารสนเทศของแบบทดสอบ
    1.10 ประสิทธิภาพสัมพัทธ์
    1.11 ความคลาดเคลื่อนมาตรฐานของการประมาณค่า
2. ความรู้ทั่วไปเกี่ยวกับการให้คะแนนความรู้บางส่วน
3. การทำหน้าที่ต่างกันของข้อสอบ
    3.1 ความหมายของการทำหน้าที่ต่างกันของข้อสอบและการทำหน้าที่ต่างกันของตัวลวง
    3.2 หลักการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ
    3.3 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ
    3.4 เกณฑ์เปรียบเทียบคุณภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ
4. การแก้ปัญหาทางคณิตศาสตร์
5. การศึกษาเกี่ยวกับแบบทดสอบ
    5.1 แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน
    5.2 แบบทดสอบแบบเลือกตอบ
6. งานวิจัยที่เกี่ยวข้อง
    6.1 งานวิจัยในประเทศ
    6.2 งานวิจัยต่างประเทศ

1. ความรู้ทั่วไปเกี่ยวกับทฤษฎีการตอบสนองข้อสอบ

1.1 ความเป็นมาของทฤษฎีการตอบสนองข้อสอบ

ลอว์เลย์ (Lawley. 1943) ได้ให้แนวคิดในการพัฒนาทฤษฎีการตอบสนองข้อสอบในบทความเกี่ยวกับปัญหาของการเลือกข้อสอบและการสร้างแบบทดสอบ ต่อมา ลอร์ด (Lord. 1952) ได้พัฒนาทฤษฎีของคะแนนสอบ ราสช์ (Rasch. 1960) ได้พัฒนาโมเดลการตอบสนองข้อสอบแบบหนึ่งพารามิเตอร์ เบิร์นบอม (Birnbuam. 1968, 1969) ได้พัฒนาโมเดลโลจิสสำหรับทฤษฎีการตอบสนองข้อสอบสำหรับประมาณค่าความสามารถทางสมองของผู้ตอบ ลอร์ดและโนวิค (Lord and Novick. 1968) ได้เขียนตำราเรื่อง "Statistical Theories of Mental Test Scores" ซึ่งให้เนื้อหาที่ลุ่มลึกและครอบคลุมความรู้ทั้งทฤษฎีการทดสอบแบบดั้งเดิม และทฤษฎีการทดสอบแนวใหม่ ซามีจิมา (Samejima. 1969) ได้พัฒนาโมเดลการตอบสนองข้อสอบ ที่สามารถวัดได้หลายคุณลักษณะ ไรท์และสโตน (Wright; & Stone. 1979) ได้ใช้แนวคิดของทฤษฎีการตอบสนองข้อสอบ สำหรับการออกแบบสร้างแบบทดสอบที่ดี ต่อมา ลอร์ด (Lord. 1980) ได้เขียนหนังสือเพื่ออธิบายแนวคิดและประยุกต์ทฤษฎีการตอบสนองข้อสอบสำหรับการทดสอบทั่วไป ในขณะเดียวกันได้มีการสร้างโปรแกรมคอมพิวเตอร์สำหรับประยุกต์ทฤษฎีการตอบสนองข้อสอบสำหรับการวิเคราะห์ข้อสอบและประมาณความสามารถของผู้สอบ เช่น โปรแกรม BICAL (Wright and Mead. 1978) โปรแกรม LOGIST (Wingersky, Barton; & Lord. 1982) โปรแกรม BILOG (Mislevy; & Bock. 1984) เป็นต้น ความพยายามดังกล่าวมีส่วนผลักดันให้ทฤษฎีการตอบสนองข้อสอบ เป็นที่รู้จักแพร่หลายมากขึ้นจนถึงปัจจุบันและมีการนำไปใช้ในการพัฒนาเครื่องมือ และระบบการทดสอบกว้างขวางยิ่งขึ้นตามลำดับ(มหาวิทยาลัยสุโขทัยธรรมาธิราช. 2549: 95)

จากที่ได้กล่าวมาพบว่า ทฤษฎีการตอบสนองข้อสอบได้มีคนให้ความสนใจจำนวนมาก ซึ่งได้มีการพัฒนาอย่างต่อเนื่องในทุกยุคทุกสมัย เริ่มจากการเสนอแนวคิดจนกระทั่งสร้างเป็นโมเดลและได้มีนักวัดผลหลายท่านได้เขียนหนังสือเพื่อเสนอแนวคิด และได้พัฒนาจนในปัจจุบันได้มีโปรแกรมทางคอมพิวเตอร์เพื่อช่วยวิเคราะห์ข้อมูลเพื่อความสะดวกต่อการวัดและประเมินผล

1.2 แนวคิดสำคัญของทฤษฎีการตอบสนองข้อสอบ

มหาวิทยาลัยสุโขทัยธรรมาธิราช (2549: 95-96) ได้กล่าวเกี่ยวกับแนวคิดสำคัญของทฤษฎีการทดสอบแนวใหม่ไว้ว่า ทฤษฎีการทดสอบแนวใหม่มุ่งขยายแนวคิดของทฤษฎีการทดสอบแบบดั้งเดิมให้มีขอบเขตที่กว้างขวาง ชัดเจน และน่าเชื่อถือยิ่งขึ้น โดยพยายามแก้ไขข้อตกลงเบื้องต้นของทฤษฎีการทดสอบแบบดั้งเดิมให้มีความสมจริง สมเหตุสมผล และเป็นที่ยอมรับได้มากขึ้น เช่น การผ่อนคลายข้อตกลงเบื้องต้นเกี่ยวกับความคลาดเคลื่อนของการวัดยอมให้ความคลาดเคลื่อนจากการวัดมีความแตกต่างกันไปตามสถานการณ์ของการวัด เพื่อศึกษาความเชื่อมั่นทั่วไปของแบบทดสอบภายใต้เงื่อนไขต่างๆ ของการวัด (G-Theory) หรือ ยอมให้ความคลาดเคลื่อนจากการวัดมีความแตกต่างกันไปขึ้นกับระดับความสามารถของแต่ละบุคคลและคุณลักษณะของข้อสอบแต่

ละข้อ พร้อมทั้งพยายามวัดคุณลักษณะภายในหรือความสามารถที่แท้จริงของแต่ละบุคคล (IRT) ทฤษฎีการทดสอบแนวใหม่จึงน่าจะให้ข้อสรุปที่ชัดเจนตรงประเด็นมากขึ้น อันจะเป็นประโยชน์ต่อการพัฒนาข้อสอบและแบบทดสอบให้สามารถนำไปใช้ในสถานการณ์ต่างๆ ของการทดสอบได้อย่างมีประสิทธิภาพ สามารถระบุนัยทั่วไปของคะแนนจริง คุณภาพของแบบทดสอบตามเงื่อนไขของการทดสอบ รวมทั้งการประเมินความสามารถที่แท้จริงของผู้สอบและการบรรยายพฤติกรรมการตอบสนองข้อสอบของผู้สอบได้เป็นอย่างดี

ศิริชัย กาญจนวาสี (2555: 7) ได้กล่าวเกี่ยวกับแนวคิดสำคัญของทฤษฎีการตอบสนองข้อสอบว่า ทฤษฎีการตอบสนองข้อสอบ (IRT) ทฤษฎีนี้เสนอแนวคิดที่ว่า ความน่าจะเป็นของการตอบข้อสอบได้ถูกต้องขึ้นอยู่กับความสามารถจริงของผู้ตอบ และคุณลักษณะของข้อสอบอันประกอบด้วยพารามิเตอร์ความยาก อำนาจจำแนกและโอกาสการเดาข้อสอบได้ถูก ระบบความสัมพันธ์ดังกล่าว สามารถแสดงด้วยโมเดลการตอบสนองข้อสอบ ซึ่งอาจเป็นโมเดล 1 พารามิเตอร์ โมเดล 2 พารามิเตอร์ หรือโมเดล 3 พารามิเตอร์ ทฤษฎี IRT ถือว่าค่าพารามิเตอร์ของข้อสอบและความสามารถจริงของผู้สอบมีความสัมพันธ์กัน การประมาณค่าพารามิเตอร์ของข้อสอบได้แก่ ค่าความยาก ค่าอำนาจจำแนก และค่าความน่าจะเป็นในการเดาข้อสอบได้ถูก จึงต้องพิจารณาร่วมกับความสามารถจริงของผู้ตอบ ดังนั้นเมื่อได้กลุ่มผู้ตอบขนาดใหญ่ที่เป็นตัวแทนประชากร การประมาณค่าพารามิเตอร์ของข้อสอบจะต้องกระทำพร้อมๆไปกับการประมาณค่าความสามารถจริงของผู้สอบ จึงจะทำให้ได้ค่าพารามิเตอร์ที่มีนัยทั่วไป มีความน่าเชื่อถือและไม่แปรผันไปตามความสามารถของกลุ่มผู้สอบในด้านความคลาดเคลื่อนของการวัด ทฤษฎี IRT สามารถวิเคราะห์ถึงความคลาดเคลื่อนในการวัดของทั้งข้อสอบแต่ละข้อและแบบทดสอบทั้งฉบับ จำแนกตามระดับความสามารถจริงของผู้ตอบ นั่นคือข้อสอบแต่ละข้อรวมทั้งแบบทดสอบแต่ละฉบับ สามารถมีคุณภาพในด้านการให้สารสนเทศ (ความเชื่อมั่นหรือความน่าเชื่อถือ) แตกต่างกันได้สำหรับผู้สอบที่มีระดับความสามารถจริงแตกต่างกัน สำหรับการเปรียบเทียบคะแนนระหว่างแบบทดสอบ ทฤษฎี IRT ไม่มีข้อจำกัดว่าแบบทดสอบจะต้องเป็นแบบทดสอบคู่ขนานกัน ในทฤษฎี IRT มีเทคนิคที่สามารถเปรียบเทียบคะแนนของแบบทดสอบต่างฉบับที่มีความยากต่างกันได้อย่างมีประสิทธิภาพ ทฤษฎี IRT มีข้อตกลงเบื้องต้นว่า แบบทดสอบมุ่งวัดคุณลักษณะเดียว (Unidimensionality) มีความเป็นอิสระระหว่างข้อสอบ (Independence) โมเดลการตอบสนองข้อสอบมีรูปแบบเป็นฟังก์ชัน โลจิสติค (Logistic function) และแบบทดสอบที่ใช้ต้องไม่เป็นแบบทดสอบประเภทความเร็ว(Speed test)

จากแนวคิดสำคัญเกี่ยวกับทฤษฎีการตอบสนองข้อสอบที่ได้กล่าวมาสรุปได้ว่า ทฤษฎีการตอบสนองข้อสอบ คือ ทฤษฎีการทดสอบแนวใหม่ที่ขยายแนวคิดของทฤษฎีการทดสอบแบบดั้งเดิม ยอมให้ความคลาดเคลื่อนจากการวัดมีความแตกต่างกันไปขึ้นอยู่กับระดับความสามารถของแต่ละบุคคลและคุณลักษณะของข้อสอบแต่ละข้อ และถือว่าค่าพารามิเตอร์ของข้อสอบและความสามารถจริงของผู้สอบมีความสัมพันธ์กัน

1.3 ลักษณะเด่นของทฤษฎีการตอบสนองข้อสอบ

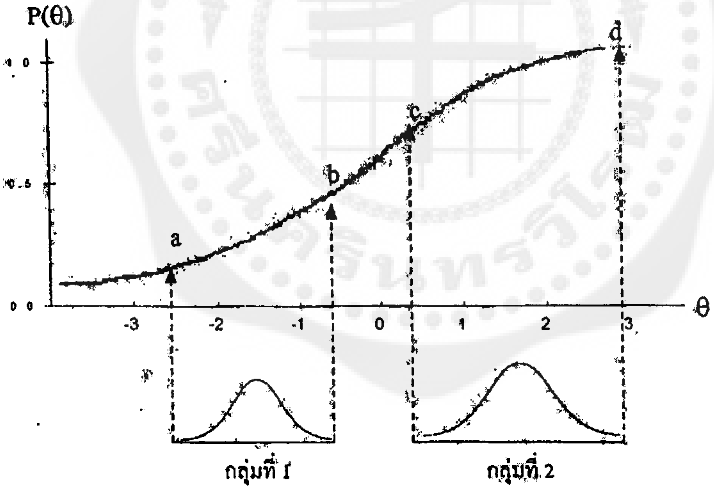
สุรศักดิ์ อมรรัตนศักดิ์ (2544: 181-182) ได้กล่าวถึงลักษณะเด่นของทฤษฎีการตอบสนองต่อข้อสอบว่า ทฤษฎีการตอบสนองมีความเหนือกว่า (Superiority) ทฤษฎีการสอบแบบคลาสสิค ดังนี้

1. ความไม่แปรเปลี่ยนของค่าพารามิเตอร์ (Invariant of Item Parameter) กล่าวคือ ไม่ว่าจะประมาณค่าจากกลุ่มตัวอย่างที่มีความสามารถระดับใดก็ตาม ค่าพารามิเตอร์จะไม่แปรเปลี่ยนไปตามกลุ่มของผู้สอบ นั่นคือค่าความยาก ค่าอำนาจจำแนก และค่าโอกาสในการตอบถูกโดยการเดาจะไม่แปรเปลี่ยนไม่ว่าจะประมาณค่าจากผู้สอบกลุ่มใด

2. จะใช้ข้อสอบกับใครก็ได้ (Person – Free) นั่นคือ ไม่ว่าจะนำข้อสอบไปใช้สอบกับบุคคลกลุ่มใด โค้งลักษณะข้อสอบก็จะคงเดิม

3. จะใช้ข้อสอบข้อใดก็ได้ (Item – Free) ในการประมาณค่าความสามารถแท้ (θ) ของผู้สอบโดยจะใช้ข้อสอบชุดใดจำนวนเท่าใดก็ได้ บางครั้งอาจใช้ข้อสอบเพียง 3-5 ข้อ ก็สามารถประมาณค่าความสามารถแท้ของผู้สอบได้แล้ว

ความไม่แปรเปลี่ยนของค่าพารามิเตอร์นั้นอธิบายได้ด้วยภาพ ดังนี้



ภาพประกอบ 2 ความไม่แปรเปลี่ยนของค่าพารามิเตอร์ของข้อสอบที่ประมาณค่าจากกลุ่มที่แตกต่างกัน

ที่มา: สุรศักดิ์ อมรรัตนศักดิ์. (2544). *ทฤษฎีทางการทดสอบ.* หน้า 181.

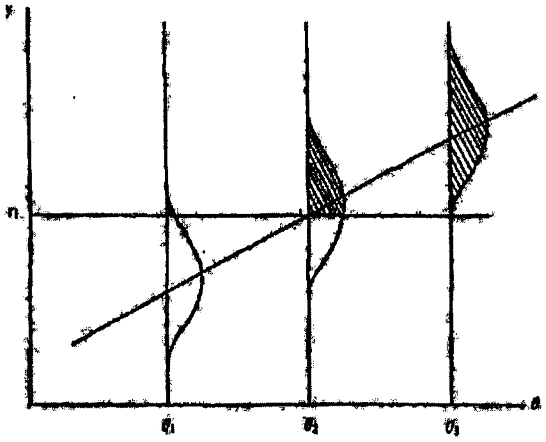
จากภาพแสดงให้เห็นว่าในการประมาณค่าจากกลุ่มผู้สอบที่มีความสามารถต่ำ (กลุ่มที่ 1) เส้นโค้งลักษณะข้อสอบที่ได้จะอยู่ในช่วง a – b เส้นโค้งลักษณะข้อสอบในช่วงนี้จะอธิบายถึงโอกาสในการตอบข้อสอบข้อนี้ถูกของผู้สอบที่มีความสามารถต่ำ และในทำนองเดียวกันเส้นโค้งลักษณะ

ข้อสอบในช่วง c – d ก็จะแสดงถึงโอกาสที่จะตอบข้อสอบนี้ถูกของผู้สอบที่มีความสามารถสูง (กลุ่มที่ 2) ถ้าข้อสอบข้อนี้วัดความสามารถอันเดียวกันของตัวอย่างทั้งสองกลุ่มแล้วเส้นโค้งลักษณะข้อสอบย่อมมีเส้นเดียว เส้นโค้งในช่วง a – b และ c – d ย่อมเป็นส่วนหนึ่งของเส้นโค้งลักษณะข้อสอบข้อนี้ เมื่อเส้นโค้งลักษณะข้อสอบมีได้เส้นเดียว ตัวพารามิเตอร์ซึ่งเป็นตัวกำหนดรูปร่างลักษณะของเส้นโค้งก็ย่อมมีได้ค่าเดียวเช่นกัน หรืออาจกล่าวได้ว่าถ้าพารามิเตอร์ของข้อสอบไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบแล้วข้อสอบที่นำไปสอบกับผู้สอบที่มีความสามารถแตกต่างกันเมื่อนำข้อมูลที่ได้จากแต่ละกลุ่มมาวิเคราะห์จะมีค่าพารามิเตอร์ของข้อสอบเป็นค่าเดียวกัน

จากที่ได้กล่าวมาสรุปได้ว่า ทฤษฎีการตอบสนองข้อสอบเป็นทฤษฎีที่มีจุดเด่นหลายประการที่เหนือกว่าทฤษฎีการทดสอบแบบดั้งเดิมไม่ว่าจะเด่นในเรื่องของพารามิเตอร์ที่จะไม่แปรเปลี่ยนไปตามกลุ่มตัวอย่าง ซึ่งจะทดสอบกับใครก็ได้ หรือจะนำข้อสอบข้อใดไปทดสอบก็ได้ ก็สามารถที่จะวัดประมาณค่าความสามารถที่แท้จริงของแต่ละบุคคลได้

1.4 หลักการของทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบเป็นทฤษฎีที่อธิบายถึงความสัมพันธ์ระหว่างลักษณะหรือความสามารถที่มีอยู่ในตัวบุคคลกับพฤติกรรมการตอบสนองต่อข้อสอบของบุคคลนั้น โดยทฤษฎีนี้มีความเชื่อว่าพฤติกรรมการตอบสนองข้อสอบของบุคคลจะถูกกำหนดโดยลักษณะหรือความสามารถที่มีอยู่ภายในตัวของบุคคล (สุรศักดิ์ อมรรัตนศักดิ์. 2544 ; อ้างอิงจาก Lord; & Novick. 1968) ซึ่งไม่สามารถจะสังเกตได้ ทฤษฎีนี้พยายามที่จะอธิบายความสัมพันธ์ระหว่างลักษณะภายในตัวบุคคลกับพฤติกรรมที่บุคคลตอบสนองต่อข้อสอบ การอธิบายความสัมพันธ์ดังกล่าวแสดงออกมาในรูปแบบจำลองทางคณิตศาสตร์ โดยให้คะแนนที่ได้รับจากการตอบข้อสอบ (y) แทน พฤติกรรมการตอบสนองข้อสอบ ให้ θ แทนลักษณะหรือความสามารถในตัวบุคคล และ r_i เป็นเกณฑ์ที่บอกว่า y แค่ไหนจึงจะทำข้อสอบข้อ i ได้ถูกต้อง ดังนั้น ถ้า $y > r_i$ แสดงว่าทำข้อสอบข้อ i ได้ถูกต้อง และถ้า $y < r_i$ แสดงว่า ทำข้อสอบข้อ i ผิด ความสัมพันธ์ระหว่างความสามารถ (θ) กับพฤติกรรมการตอบสนองข้อสอบ (y) แสดงได้ด้วยภาพดังนี้



ภาพประกอบ 3 ความสัมพันธ์ระหว่างความสามารถกับพฤติกรรมการตอบสนอง

ที่มา: สุรศักดิ์ อมรรัตนศักดิ์. (2544). *ทฤษฎีทางการทดสอบ.* หน้า 179.

จากภาพจะเห็นได้ว่าถ้านำโอกาสที่จะตอบถูก (พื้นที่ส่วนที่แรเงา) ในระดับความสามารถต่างๆมาเขียนกราฟใหม่ จะได้โค้งลักษณะข้อสอบ (ICC) เป็นรูปต่างๆ ทั้งนี้ขึ้นอยู่กับฟังก์ชันทางคณิตศาสตร์และจำนวนพารามิเตอร์ที่จะใช้อธิบาย ฟังก์ชันทางคณิตศาสตร์ที่แสดงความสัมพันธ์ระหว่างระดับความสามารถกับพฤติกรรมการตอบสนองข้อสอบเรียกว่า ฟังก์ชันการตอบสนองข้อสอบ (Item Response Function) (สุรศักดิ์ อมรรัตนศักดิ์. 2544; อ้างอิงจาก Lord. 1980) หรือฟังก์ชันลักษณะข้อสอบ (Item Characteristic Function) (สุรศักดิ์ อมรรัตนศักดิ์. 2544; อ้างอิงจาก Lord and Novick. 1968) ซึ่งสามารถเขียนฟังก์ชันได้ดังนี้

$$P_i(\theta) = \text{Prob}(U_i = 1 \mid \theta) \text{ เมื่อ } U_i = 0, 1$$

จากฟังก์ชันข้างต้นนี้ หมายถึง โอกาสที่ผู้สอบซึ่งมีความหมาย θ จะตอบคำถามข้อ i ได้ถูกต้อง (สุรศักดิ์ อมรรัตนศักดิ์. 2544: 178-180)

จากที่ได้กล่าวมาสรุปได้ว่า ความสามารถที่อยู่ภายในตัวของบุคคลในการทำข้อสอบไม่สามารถที่จะทราบโดยตรงได้ นอกจากดูจากพฤติกรรมการตอบสนองต่อข้อสอบของบุคคลนั้น การอธิบายความสัมพันธ์อธิบายในรูปแบบจำลองทางคณิตศาสตร์ ซึ่งแสดงความสัมพันธ์ระหว่างระดับความสามารถกับพฤติกรรมการตอบสนองต่อข้อสอบ เพื่อง่ายต่อการทำความเข้าใจ

1.5 โมเดลการตอบสนองข้อสอบ (Item Response Model)

1.5.1 คุณลักษณะของโมเดลการตอบสนองข้อสอบ

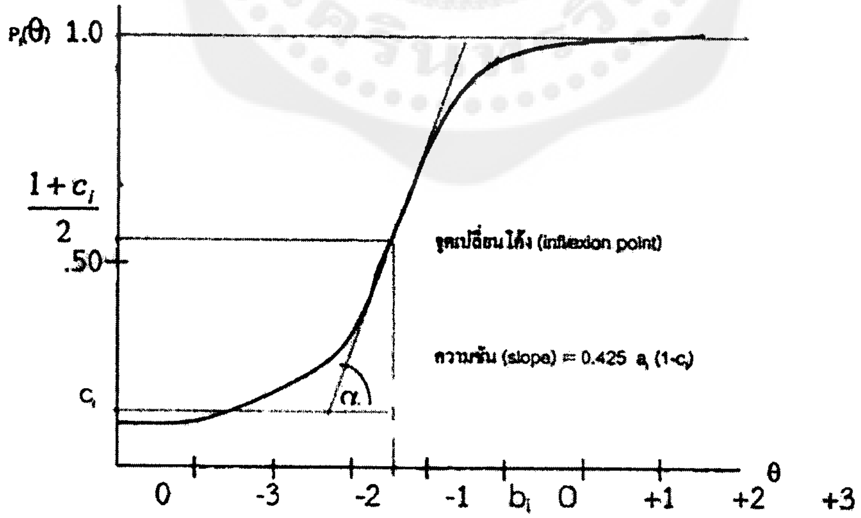
โมเดลการตอบสนองข้อสอบ เป็นระบบความสัมพันธ์ระหว่างโอกาสตอบข้อสอบถูก(P_i) กับความสามารถที่มีอยู่ภายในผู้ตอบ (θ) ในรูปของโค้งลักษณะข้อสอบ (ICC) ซึ่งมีลักษณะเป็นฟังก์ชันโลจิส (Logistic function) หรือฟังก์ชันปกติสะสม (Normal ogive function) บางครั้งอาจเรียกว่า โมเดลโลจิส หรือโมเดลปกติสะสม

โมเดลปกติสะสม ใช้ฟังก์ชันปกติสะสมแสดงความสัมพันธ์ระหว่างผลการตอบข้อสอบกับความสามารถของผู้สอบ ส่วนโมเดลโลจิสใช้ฟังก์ชันโลจิสแสดงความสัมพันธ์ระหว่างผลการตอบกับความสามารถดังกล่าว ซึ่งฟังก์ชันทั้งสองให้ผลลัพธ์ของการประมาณค่าใกล้เคียงกันมาก แต่ฟังก์ชันโลจิสมีลักษณะของสูตรคณิตศาสตร์ และวิธีคำนวณง่ายและสะดวกกว่า นอกจากนี้ โมเดลโลจิสยังมีความทนทานต่อความคลาดเคลื่อนที่เกิดขึ้นกับผู้สอบที่มีความสามารถสูงจะตอบข้อสอบได้ดีกว่า จึงทำให้โมเดลโลจิสเป็นที่นิยมกันมากในการนำไปใช้จริง(มหาวิทยาลัยสุโขทัยธรรมาธิราช. 2549 ; อ้างอิงจาก Lord. 1980)

โมเดลโลจิสแบบสามพารามิเตอร์ (Three-Parameter Logistic Model) และโค้งลักษณะข้อสอบ (ICC) สามารถแสดงได้ดังนี้

โมเดล

$$P_i(\theta) = c_i + \frac{(1 - c_i)}{1+e^{-Da_i(\theta - b_i)}}$$



ภาพประกอบ 4 เส้นโค้งลักษณะข้อสอบของโมเดลโลจิสแบบ 3 พารามิเตอร์

ที่มา: มหาวิทยาลัยสุโขทัยธรรมาธิราช. (2549). *การพัฒนาแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน*. หน้า 120.

1.5.2 ประเภทของโมเดลการตอบสนองข้อสอบ

จากแนวคิดตามทฤษฎีการตอบสนองข้อสอบได้มีการพัฒนาโมเดลหรือแบบจำลองขึ้นมาหลายรูปแบบด้วยกัน โดยแต่ละโมเดลจะมีฟังก์ชันทางคณิตศาสตร์และจำนวนพารามิเตอร์ในฟังก์ชันที่แตกต่างกันซึ่งสามารถจำแนกเป็น 3 ประเภทใหญ่ๆ (มหาวิทยาลัยสุโขทัยธรรมาธิราช. 2549; อ้างอิงจาก Hambleton and Sawaminathan. 1985) ดังนี้

1. โมเดลการตอบสนองข้อสอบที่ใช้กับการตรวจคะแนนรายข้อแบบ 2 ค่า เป็นข้อสอบที่ตรวจให้คะแนนแบบ 0, 1 (ตอบผิดได้ 0 คะแนน ตอบถูกได้ 1 คะแนน) โมเดลประเภทนี้ถูกพัฒนาขึ้นในระยะเริ่มแรก (ค.ศ.1943 – 1968) เช่น สเกลกัตแมนสมบูรณ์, โมเดลระยะห่างแฝง (Latent Distance Model) โมเดลเชิงเส้นตรง เป็นต้น ในระยะต่อมา (ค.ศ.1952 – 1982) ได้มีการพัฒนาโมเดลประเภทนี้เพิ่มขึ้นมา เช่น โมเดลปกติสะสมแบบ 1, 2, 3 พารามิเตอร์ และโมเดลโลจิสแบบ 1, 2, 3, 4 พารามิเตอร์ เป็นต้น

2. โมเดลการตอบสนองข้อสอบที่ใช้กับการตรวจคะแนนรายข้อแบบมากกว่า 2 ค่า (Multichotamous) เช่น โมเดลการตอบสนองแบบนามบัญญัติ, โมเดลการตอบสนองแบบเกรด, โมเดลการให้คะแนนบางส่วน เป็นต้น

3. โมเดลการตอบสนองข้อสอบที่ใช้กับการตรวจคะแนนรายข้อแบบต่อเนื่อง เช่น โมเดลการตอบสนองแบบคะแนนต่อเนื่อง (Samejima. 1972) เป็นต้น

จากที่ได้กล่าวมาสรุปได้ว่า โมเดลการตอบสนองข้อสอบมีลักษณะเป็นฟังก์ชันโลจิส หรือฟังก์ชันปกติสะสม ฟังก์ชันทั้งสองให้ผลของการประมาณค่าใกล้เคียงกัน แต่โมเดลโลจิสเป็นที่นิยมนำไปใช้จริงมากกว่า และจำนวนพารามิเตอร์ในฟังก์ชันสามารถจำแนกเป็น 3 ประเภทใหญ่ๆ คือ โมเดลการตอบสนองข้อสอบที่ใช้กับการตรวจคะแนนรายข้อแบบ 2 ค่า แบบมากกว่า 2 ค่า และการตรวจคะแนนรายข้อแบบต่อเนื่อง

1.6 พารามิเตอร์ในทฤษฎีการตอบสนองข้อสอบ

ความหมายของค่าพารามิเตอร์ (เยาวดี รางชัยกุล วิบูลย์ศรี. 2553: 172-173)

1. ค่าความยาก (b)

ค่าความยาก ได้มาจากค่าความสามารถที่ตรงจุดเปลี่ยนโค้ง (Inflexion Point) ซึ่งเป็นจุดที่โค้งมีความชันมากที่สุด หรือมีความหมายอีกนัยหนึ่งก็คือ ผู้สอบที่มีความสามารถถึงระดับ ณ จุดเปลี่ยนโค้งข้อกระทงนั้น จะมีโอกาสตอบข้อกระทงข้อนี้ถูกอยู่ 0.5 หรือในทางปฏิบัติกล่าวได้ว่า จากจุดบนแกน y ที่แสดงถึงตำแหน่งโอกาสในการตอบข้อกระทงข้อนี้ถูกมีอยู่ 0.5 ถ้าลากเส้นตั้งขนานกับแกน x จนพบกับเส้นโค้ง ซึ่งจะเป็นจุดเปลี่ยนโค้งด้วยนั้น ในทางตรงกันข้ามเมื่อลากเส้นตั้งฉากจากจุดดังกล่าวให้มาจดแกน x ค่าที่วัดได้ในแกน x คือค่าความยากของข้อกระทงนั้นๆ ซึ่งข้อกระทงทั้งหมดในแบบทดสอบฉบับหนึ่งๆ ที่นำมาวิเคราะห์จะมีค่าความยากกระจายอยู่ในแกน x จากค่า $-\alpha$ ถึง $+\alpha$ แต่ในทางปฏิบัตินิยมใช้ช่วง -3 ถึง +3 และแบบทดสอบทั่วไปมักจะมีค่า

b_i อยู่ระหว่าง -2.5 ถึง +2.5 ถ้าค่า b_i อยู่ใกล้ -2.5 แสดงว่าเป็นข้อกระทงที่ง่าย ตรงกันข้าม ถ้าค่า b_i อยู่ใกล้ +2.5 แสดงว่าเป็นข้อกระทงที่ยาก

ในกรณีที่เป็นรูปแบบที่มีค่าพารามิเตอร์ 3 ตัว หรือคำนึงถึงโอกาสในการเดาคำตอบ (c) ค่าความยากซึ่งเริ่มต้นจากจุดบนแกน y นั้น จะใช้จุดตั้งต้นตรงที่โอกาสในการตอบถูก กล่าวคือ ค่าความยากมีค่าเริ่มต้นที่ $\frac{1-c}{2}$

2. ค่าอำนาจจำแนก (a)

ค่าอำนาจจำแนก เป็นสัดส่วนกับค่าความชัน (Slope) ของ $p_t(\theta)$ ที่จุดเปลี่ยนโค้ง หรือที่จุด $\theta = b$ โดยทฤษฎีแล้ว ค่าอำนาจจำแนกจะมีค่าอยู่ในช่วง $-\alpha$ ถึง $+\alpha$ แต่ในทางการนำมาใช้ประโยชน์นั้น ข้อกระทงข้อใดที่มีค่า a ติดลบย่อมแสดงว่า ข้อกระทงข้อนั้นไม่ดี และควรจะต้องถูกตัดออกไป ส่วนข้อกระทงที่มีค่า a สูงขึ้น ย่อมแสดงว่า ความน่าจะเป็นของการตอบข้อกระทงข้อนั้นๆ เพิ่มขึ้น เมื่อระดับความสามารถของผู้สอบสูงขึ้น ในทางปฏิบัติมักจะใช้ข้อกระทงที่มีค่า a_i อยู่ระหว่าง 0.5 ถึง 2.5

3. ค่าการเดา (c)

ค่าการเดา เป็นค่าที่อยู่ปลายโค้งด้านต่ำ (Lower Asymtote) ของข้อกระทง ค่านี้เป็นค่าแทนโอกาสที่คนซึ่งมีความสามารถต่ำ แต่สามารถตอบข้อกระทงข้อนั้นได้ถูกโดยการเดา ค่าการเดาจะอยู่ในช่วง 0.00 ถึง 0.20 เมื่อแบบทดสอบเป็นแบบเลือกตอบที่มี 4 ตัวเลือก ดังนั้นข้อกระทงใดที่มีค่า c มากกว่า 0.30 ขึ้นไปแสดงว่า ข้อกระทงข้อนั้นๆ ไม่ดี ควรพิจารณาตัดออก

จากที่ได้กล่าวมาสรุปได้ว่า ค่าพารามิเตอร์จะไม่แปรเปลี่ยนไปตามกลุ่มของผู้สอบ นั่นคือ ค่าความยาก ค่าอำนาจจำแนก และค่าโอกาสในการตอบถูกโดยการเดาจะไม่แปรเปลี่ยนไม่ว่าจะประมาณค่าจากผู้สอบกลุ่มใดก็ตาม

1.7 ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ

สุรศักดิ์ อมรรัตนศักดิ์. (2544: 180) ได้กล่าวเกี่ยวกับข้อตกลงเบื้องต้นไว้ดังนี้

1. Unidimensionality เป็นการสมมติว่าข้อสอบในแบบทดสอบมีลักษณะเป็นเอกพันธ์ นั่นคือ แบบทดสอบนั้นจะต้องมุ่งวัดความสามารถอย่างใดอย่างหนึ่งเพียงความสามารถเดียว หากไม่กำหนดข้อตกลงเบื้องต้นเช่นนี้จะทำให้แบบจำลองของทฤษฎีมีความยุ่งยาก (Hambleton and Cook. 1977) ส่วนการตรวจสอบว่าข้อมูลจากการสอบเป็นไปตามข้อตกลงเบื้องต้นหรือไม่ อาจทำได้โดยการวิเคราะห์องค์ประกอบ (Factor Analysis)

2. Local Independence เป็นกำหนดข้อตกลงเบื้องต้นเกี่ยวกับความเป็นอิสระใน การตอบสนองต่อข้อสอบ กล่าวคือ การตอบสนองต่อข้อสอบข้อใดข้อหนึ่งของผู้สอบไม่มีผลต่อ การตอบสนองต่อข้อสอบข้ออื่นๆ ในแบบทดสอบ ในการตรวจสอบว่าข้อสอบแต่ละข้อเป็นไปตามข้อตกลงเกี่ยวกับ Local Independence หรือไม่ทำได้โดยอาศัยเทคนิคการวิเคราะห์องค์ประกอบเช่นกัน (Hambleton and Cook. 1977)

3. Item Characteristic Curve เป็นข้อตกลงเกี่ยวกับโค้งลักษณะข้อสอบ กล่าวคือ โค้งลักษณะข้อสอบเป็นกราฟของฟังก์ชันทางคณิตศาสตร์ ที่แสดงถึงความสัมพันธ์ระหว่างโอกาสในการตอบข้อสอบข้อนั้นถูกกับความสามารถหรือลักษณะที่วัดโดยข้อสอบข้อนั้น โค้งลักษณะข้อสอบมีหลายรูปแบบทั้งนี้ขึ้นอยู่กับแบบจำลองที่ใช้อธิบายความสัมพันธ์ดังกล่าว

จากการศึกษาพบว่า ทฤษฎีการตอบสนองต่อข้อสอบขึ้นอยู่กับความเชื่อหรือข้อตกลงเบื้องต้นหลายข้อเกี่ยวกับลักษณะของข้อมูลที่สามารถนำทฤษฎีไปใช้ได้อย่างเหมาะสม แม้ว่าจะมีบางข้อที่ไม่สามารถตรวจสอบได้โดยตรงแต่ก็สามารถเก็บหลักฐานอื่นๆมาช่วยสนับสนุนได้

1.8 ฟังก์ชันสารสนเทศของข้อสอบ (Item Information Function: IIF)

ฟังก์ชันสารสนเทศของข้อสอบ หมายถึง ความถูกต้องแม่นยำในการประมาณค่าความสามารถจริงด้วยผลการตอบข้อสอบข้อนั้นๆ ซึ่งมีค่าเท่ากับอัตราส่วนกำลังสองของความชันของโค้งลักษณะข้อสอบกับความแปรปรวนของข้อสอบข้อนั้น โดยฟังก์ชันสารสนเทศของข้อสอบแต่ละข้อสามารถคำนวณได้จากสูตร

$$I_i(\theta) = \left(P_i'\right)^2 / P_i Q_i$$

เมื่อ $I_i(\theta)$ คือ ฟังก์ชันสารสนเทศของข้อสอบข้อที่ i

P_i คือ ความน่าจะเป็นที่ผู้สอบคนหนึ่งที่มีความสามารถจะตอบข้อสอบข้อที่ i ได้ถูกต้อง

Q_i คือ $1 - P_i$

P_i' คือ ค่าอนุพันธ์ (Derivative) ของโอกาสในการตอบข้อสอบที่ i ถูก หรือความชันของโค้งลักษณะข้อสอบที่ระดับความสามารถ (θ)

จากสมการดังกล่าวแสดงให้เห็นว่า โค้งสารสนเทศของข้อสอบขึ้นอยู่กับความชันของโค้งลักษณะของข้อสอบแต่ละข้อ และความแปรปรวนของคะแนนจากข้อสอบในแต่ละระดับความสามารถ (θ) กล่าวคือ ยิ่งโค้งลักษณะของข้อสอบชันเท่าไรและความแปรปรวนน้อยเท่าไรโค้งสารสนเทศของข้อสอบที่ระดับความสามารถนั้นจะยิ่งสูงขึ้น ซึ่งหมายความว่าข้อสอบข้อนั้นสามารถจำแนกความสามารถของผู้สอบ ณ ระดับความสามารถนั้นได้ดีนั่นเอง (สุรศักดิ์ อมรรัตนศักดิ์. 2544: 192-193)

จากที่ได้กล่าวไปแล้วสรุปได้ว่า โค้งสารสนเทศของข้อสอบมีความสัมพันธ์กับความชันของโค้งลักษณะข้อสอบและความแปรปรวนของคะแนนจากข้อสอบในแต่ละระดับความสามารถ ซึ่งกล่าว

ได้ว่าฟังก์ชันสารสนเทศของข้อสอบ ก็คือความถูกต้องแม่นยำในการประมาณค่าความสามารถจริงด้วยผลการตอบข้อสอบข้อนั้นๆนั่นเอง

1.9 ฟังก์ชันสารสนเทศของแบบทดสอบ (Test Information Function: TIF)

ฟังก์ชันสารสนเทศของแบบทดสอบ หมายถึง ความถูกต้องแม่นยำในการประมาณค่าความสามารถจริงของผู้สอบด้วยผลการตอบแบบทดสอบ ซึ่งเกิดจากผลรวมของค่าฟังก์ชันสารสนเทศของข้อสอบแต่ละข้อเข้าด้วยกัน โดยฟังก์ชันสารสนเทศของแบบทดสอบสามารถคำนวณได้จากสูตร

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

หรือ $$I(\theta) = \sum_{i=1}^{n} \frac{\left(P_i'\right)^2}{P_i Q_i}$$

คุณลักษณะของค่าฟังก์ชันสารสนเทศของแบบทดสอบ

1. ฟังก์ชันสารสนเทศของแบบทดสอบ จะนิยามที่แต่ละจุดของระดับความสามารถ ความแม่นยำในการวัดจะแตกต่างกันที่ระดับความสามารถต่างกัน

2. ฟังก์ชันสารสนเทศของแบบทดสอบจะขึ้นอยู่กับคุณภาพและจำนวนข้อของแบบทดสอบ

3. ฟังก์ชันสารสนเทศของแบบทดสอบ ไม่ได้ขึ้นอยู่กับการนำค่าฟังก์ชันสารสนเทศของข้อสอบเฉพาะบางข้อมารวมกัน เพราะการกระจายของข้อสอบแต่ละข้อจะเป็นอิสระต่อกัน

4. ฟังก์ชันสารสนเทศของแบบทดสอบที่ระดับความสามารถใดๆ จะเป็นสัดส่วนผกผันกับค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ (Standard Error of the Ability Estimates: $SE(\theta)$) ณ ระดับความสามารถนั้น (สุรศักดิ์ อมรรัตนศักดิ์. 2544: 193)

จากที่ได้กล่าวไปแล้วสรุปได้ว่า ฟังก์ชันสารสนเทศของแบบทดสอบจะขึ้นอยู่กับคุณภาพและจำนวนข้อของแบบทดสอบ แต่จะไม่ได้ขึ้นอยู่กับการนำค่าฟังก์ชันสารสนเทศของข้อสอบเฉพาะบางข้อมารวมกัน ซึ่งฟังก์ชันสารสนเทศของแบบทดสอบ กล่าวได้ว่าคือความถูกต้องแม่นยำในการประมาณค่าความสามารถจริงของผู้สอบด้วยผลการตอบแบบทดสอบนั่นเอง

1.10 ประสิทธิภาพสัมพัทธ์ (Relative Efficiency)

1. การเปรียบเทียบค่าฟังก์ชันสารสนเทศระหว่างแบบทดสอบที่วัดคุณลักษณะเดียวกัน สามารถทำได้โดยการคำนวณประสิทธิภาพสัมพัทธ์ (Relative Efficiency; $RE(\theta)$) ซึ่งเป็นอัตราส่วนระหว่างค่าฟังก์ชันสารสนเทศของแบบทดสอบต่างฉบับ ณ ตำแหน่ง θ เดียวกัน

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

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

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

การแปลความหมายของค่าดัชนีประสิทธิภาพสัมพัทธ์ สามารถนำไปช่วยคัดเลือกแบบทดสอบได้ดังนี้

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

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

3) $RE(\theta)<1$ 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 แบบทดสอบฉบับ A มีประสิทธิภาพสูงกว่าแบบทดสอบฉบับ B สำหรับใช้สอบกับผู้สอบที่มีระดับความสามารถ θ หรือ อาจกล่าวได้ว่าแบบทดสอบฉบับ A ทำหน้าที่เหมือนเป็นแบบทดสอบที่มีความยาวกว่าแบบทดสอบฉบับ B 25% หรือแบบทดสอบฉบับ A สามารถตัดข้อออก 25% ก็ยังคงมีความแม่นยำในการประมาณค่าความสามารถ ณ ตำแหน่ง θ ได้ดีเท่ากับแบบทดสอบฉบับ B

2. การเปรียบเทียบค่าฟังก์ชันสารสนเทศเฉลี่ยระหว่างแบบทดสอบที่วัดคุณลักษณะเดียวกัน สามารถทำได้โดยคำนวณประสิทธิภาพสัมพัทธ์เฉลี่ย (RAI) ซึ่งเป็นอัตราส่วนระหว่างค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบทดสอบต่างฉบับ ณ ทุกตำแหน่ง θ ดังนี้

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

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

$\mathrm{AI}(\theta, Y)$ = ค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบทดสอบฉบับ Y ณ ทุกตำแหน่ง θ

(ศิริชัย กาญจนวาสี. 2555: 66-67)

จากที่ได้กล่าวไปแล้วสรุปได้ว่า ประสิทธิภาพสัมพัทธ์สามารถนำไปช่วยในการคัดเลือกแบบทดสอบได้ ทั้งนี้จะพิจารณาจากอัตราส่วนระหว่างค่าฟังก์ชันสารสนเทศของแบบทดสอบต่างฉบับ ณ ตำแหน่ง θ เดียวกัน เพื่อตัดสินใจง่ายขึ้น

1.11 ความคลาดเคลื่อนมาตรฐานของการประมาณค่า (Standard Error of Estimation)

ความคลาดเคลื่อนมาตรฐานของการประมาณค่า $\mathrm{SE}(\theta)$ เป็นค่าส่วนเบี่ยงเบนมาตรฐานของการแจกแจงความน่าจะเป็น ของค่าประมาณความสามารถที่แท้จริง (θ) ซึ่งเป็นค่าสัดส่วนผกผันกับความถูกต้องแม่นยำของการประมาณค่าความสามารถ หรือค่าฟังก์ชันสารสนเทศของแบบทดสอบ ดังสูตร

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

เมื่อ $\mathrm{SE}(\theta)$ = ความคลาดเคลื่อนมาตรฐานของการประมาณค่า สำหรับผู้ตอบที่มีความสามารถ θ

$I(\theta)$ = สารสนเทศที่ได้จากแบบทดสอบ สำหรับผู้ตอบที่มีความสามารถ θ

การสอบทุกครั้งย่อมมีความคลาดเคลื่อนในการประมาณค่าความสามารถ (θ) ความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถ มีค่าเท่ากับส่วนกลับของรากที่สอง

ของค่าฟังก์ชันสารสนเทศของแบบทดสอบ ดังนั้น ค่าฟังก์ชันสารสนเทศของแบบทดสอบมีค่าสูง ความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถจะต่ำ นั่นคือ ถ้าค่าฟังก์ชันสารสนเทศของแบบทดสอบมีค่าสูง ณ ผู้สอบระดับความสามารถ θ ใด (ฟังก์ชันการตอบสนองข้อสอบแต่ละข้อมีความชันสูงและความแปรปรวนต่ำ ณ ระดับ θ นั้น) ความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถของผู้สอบ ณ ระดับ θ นั้นจะต่ำหรือมีความแม่นยำสูงในการประมาณค่าความสามารถของ ผู้สอบ θ นั้น (ศิริชัย กาญจนวาสี. 2534 (ค))

$SE(\theta)$ เป็นค่าที่คล้ายกับความคลาดเคลื่อนมาตรฐานของการวัด (Standard Error of Measurement ; SEM) ใน CTT แต่ต่างกันตรงที่ $SE(\theta)$ มีค่าผันแปรไปตามตำแหน่ง θ ในขณะที่ SEM เป็นค่าคงที่สำหรับผู้ตอบทุกคน

ขนาดของ $SE(\theta)$ มีลักษณะโดยทั่วไปดังนี้

- แบบทดสอบที่มีจำนวนข้อสอบมากขึ้น ขนาดของ $SE(\theta)$ จะลดลง
- แบบทดสอบที่มีข้อสอบที่ a สูง , c ต่ำ ขนาดของ $SE(\theta)$ จะลดลง
- แบบทดสอบที่ประกอบด้วยข้อสอบที่ b ใกล้เคียงกับ θ ของผู้สอบ (ไม่ง่ายหรือยากเกินไป) ขนาดของ $SE(\theta)$ จะลดลง
- $I(\theta)$ ที่มีค่าเกิน 25 จะทำให้ขนาดของ $SE(\theta)$ ค่อนข้างคงที่และมีผลน้อยต่อความคลาดเคลื่อนในการประมาณค่า θ (Green, Yen; & Burket. 1989)

(ศิริชัย กาญจนวาสี. 2555: 68)

จากที่ได้กล่าวไปแล้วสรุปได้ว่า ความคลาดเคลื่อนในการประมาณค่าความสามารถย่อมเกิดขึ้นในการสอบทุกครั้ง และในการสอบมีความจำเป็นที่จะต้องให้มีความคลาดเคลื่อนน้อยที่สุด ซึ่งถ้าค่าฟังก์ชันสารสนเทศของแบบทดสอบมีค่าสูงความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถจะต่ำนั่นเอง และจะส่งผลดีในการทดสอบอีกด้วย

2. ความรู้ทั่วไปเกี่ยวกับการให้คะแนนความรู้บางส่วน

มาสเตอร์ (ศิริชัย กาญจนวาสี. 2555: 96; อ้างอิงจาก Master. 1982) ได้พัฒนา Partial Credit Model (PCM) สำหรับใช้ในการวิเคราะห์ข้อสอบหรือข้อคำถามที่มีกระบวนการตอบหลายลำดับขั้น ซึ่งจำเป็นต้องมีการตรวจให้คะแนนการตอบถูกต้องหรือตอบถูกบางส่วนในแต่ละลำดับขั้นของกระบวนการตอบ เช่น ข้อสอบวัดผลสัมฤทธิ์ทางคณิตศาสตร์ ซึ่งมีการให้คะแนนคำตอบถูกบางส่วน แบบวัดเจตคติ บุคลิกภาพ เชาว์ปัญญา ที่มีการให้คะแนนคำตอบเป็นลำดับขั้น เป็นต้น

การให้คะแนนความรู้บางส่วนของผู้ตอบ หมายถึง การให้คะแนนที่ใกล้เคียงกับความรู้ที่แท้จริงของผู้ตอบ ซึ่งมีส่วนช่วยลดความคลาดเคลื่อนจากแหล่งการเดาสุ่มของผู้ตอบลง ทำให้คะแนนมีความเชื่อมั่นมากขึ้น ในขณะเดียวกันก็ส่งผลต่อความเที่ยงตรงของกระบวนการวัดด้วย (Frary. 1980) จากการศึกษาวิธีการให้คะแนนความรู้บางส่วนสรุปเป็นประเภทใหญ่ๆ ได้ 4 วิธี ดังนี้ (เอมอร จังศิริพรปกรณ์. 2545; อ้างอิงจาก Simon, Budescu; & Nevo. 1997)

1. การให้น้ำหนักแก่ข้อสอบที่แตกต่างกัน (Differential Item Weighting)
2. การให้น้ำหนักแก่ตัวเลือกที่แตกต่างกัน (Differential Option Weighting)
3. การเปลี่ยนแปลงโครงสร้างของข้อสอบ (Changing the Item Structure)
4. การเปลี่ยนแปลงวิธีการตอบ (Changing the Response Method)

1. การให้น้ำหนักแก่ข้อสอบที่แตกต่างกัน (Differential Item Weighting)

หลักการของวิธีนี้ คือข้อสอบที่มีคุณภาพสูงจะได้น้ำหนักมากกว่าข้อสอบที่มีคุณภาพต่ำ โดยใช้ผลที่ได้จากการวิเคราะห์ข้อสอบเป็นเกณฑ์ คือ ค่าความยาก ความสามารถ ความเที่ยงตรง หรือ การพิจารณาของผู้เชี่ยวชาญ Stanley และ Wang (1970) ได้ให้ข้อสรุปว่า การให้น้ำหนักแก่ข้อสอบโดยใช้ทฤษฎีการวัดแบบดั้งเดิม ทำให้น้ำหนักของคะแนนแต่ละข้อมีค่าไม่คงที่ เปลี่ยนแปลงไปตามกลุ่มตัวอย่าง จึงควรให้คะแนนตามทฤษฎีการตอบสนองข้อสอบ (IRT) โดย Lord (1980) สามารถพิสูจน์ได้ว่า เมื่อกำหนดน้ำหนักคะแนนรายข้อที่เหมาะสมกับระดับความสามารถของผู้ตอบแล้ว และรวมคะแนนทุกข้อเข้าด้วยกัน จะทำให้ได้สารสนเทศสูงสุด อย่างไรก็ตามการกำหนดน้ำหนักคะแนนที่เหมาะสมรายข้อนี้ ยังวางอยู่บนพื้นฐานของวิธีการตอบ และการให้คะแนนเฉพาะข้อที่ตอบถูกแบบเดิม ไม่ได้มีการพิจารณาการให้คะแนนความรู้บางส่วนแก่ผู้ตอบในแต่ละตัวเลือก

2. การให้น้ำหนักแก่ตัวเลือกที่แตกต่างกัน (Differential Option Weighting)

หลักการของวิธีนี้ คือ ความรู้บางส่วนสามารถถูกวัดได้จากการให้น้ำหนักคะแนนแก่ตัวเลือก แบ่งวิธีการได้ 2 วิธี คือ วิธีแรกเป็นการให้น้ำหนักแก่ตัวเลือกโดยใช้วิจารณญาณของผู้เชี่ยวชาญ หรือจากทฤษฎีของโครงสร้างความรู้ ส่วนวิธีที่สองเป็นการให้น้ำหนักจากข้อมูลเชิงประจักษ์ที่ผ่านมาในอดีต หรือปัจจุบันโดยอาศัยสิ่งที่น่าสนใจของตัวเลือก เช่น คะแนนมาตรฐานโดยเฉลี่ยของแต่ละตัวเลือกที่ผู้สอบเลือก ความสัมพันธ์ระหว่างตัวเลือกที่ผู้สอบเลือกกับคะแนนรวมทั้งหมด เป็นต้น

2. การเปลี่ยนแปลงโครงสร้างของข้อสอบ (Changing the Item Structure)

วิธีนี้มีรูปแบบของโครงสร้างข้อสอบ และ/หรือ คำอธิบายที่แตกต่างจากแบบเลือกตอบธรรมดาตัวอย่างที่สำคัญ มีดังนี้

2.1 ข้อสอบที่มี 2 ตัวเลือก คือ ถูก และ ผิด ผู้ตอบจะต้องเลือกคำตอบอย่างใดอย่างหนึ่ง คะแนนที่ได้จากผลรวมของจำนวนคำตอบที่ถูกต้อง เนื่องจากการตอบแต่ละข้อใช้เวลาน้อย จึงต้องเพิ่มจำนวนข้อให้มากขึ้นกว่าข้อสอบแบบเลือกตอบธรรมดา

2.2 ข้อสอบที่มีตัวถูกหลายๆ ตัวเลือก (มากกว่า 1) โดยต้องมีคำชี้แจงอธิบายให้แก่ผู้สอบซึ่งอาจจะระบุจำนวนตัวเลือกที่ถูกหรือไม่ระบุก็ได้ คะแนนที่ได้เท่ากับผลรวมของการระบุตัวเลือกถูกได้ถูกต้อง ในบางกรณีมีการหักคะแนนการตอบที่ตอบตัวเลือกผิด

2.3 ข้อสอบที่มีกลุ่มของข้อสอบ และกลุ่มของคำตอบ ที่เรียกว่าข้อสอบแบบจับคู่ โดยผู้สอบต้องเลือกคำตอบที่ถูกคู่กับคำถามที่กำหนดให้ ถ้าจำนวนคำถามเท่ากับจำนวนคำตอบ

เรียกว่า การจับคู่อย่างง่าย (simple matching) (Simon. et al. 1997; อ้างอิงจาก Horst. 1966, Wesman. 1971) แต่ถ้าจำนวนคำตอบมากกว่าจำนวนคำถาม เรียกว่า การจับคู่แบบพหุ (multiple matching) (Simon et al. 1997; อ้างอิงจาก Gulli & en. 1986) ความแตกต่างของ 2 แบบ คือ การจับคู่แบบพหุ ช่วยลดโอกาสในการเดามากกว่า เนื่องจากมีจำนวนตัวเลือกมากกว่า แต่ข้อจำกัดที่สำคัญของวิธีนี้ คือ ความยากในการสร้างข้อสอบให้มีความเหมาะสม

3. การเปลี่ยนแปลงวิธีการตอบ (Changing the Response Method)

วิธีนี้แตกต่างจากวิธีก่อนที่กล่าวมาทั้งหมด คือ ใช้น้ำหนักที่ให้โดยผู้สอบเอง ซึ่งสะท้อนถึงความรู้ที่มีอยู่ในตัวผู้สอบ การประเมินด้วยตนเองนี้เป็นวิธีที่ดีที่สุดในการได้มาซึ่งสารสนเทศเกี่ยวกับความรู้ของผู้สอบในแต่ละข้อ ผู้สอบจะให้สารสนเทศอย่างสมบูรณ์เกี่ยวกับน้ำหนัก วิธีการให้คะแนน และการควบคุมระดับความมั่นใจในการให้ ซึ่งจะช่วยลดการเดาและความคลาดเคลื่อนในการวัด วิธีนี้ใช้รูปแบบข้อสอบเหมือนแบบทดสอบชนิดเลือกตอบทั่วไปเพียงแต่คำสั่ง วิธีการตอบและการให้คะแนนแตกต่างจากแบบทดสอบชนิดเลือกตอบธรรมดา ซึ่งแต่ละวิธีใช้วิธีการที่ไม่ยุ่งยาก โดยแบ่งเป็นชนิดต่างๆได้ 6 วิธี คือ

3.1 วิธีการแก้การเดา (Correction for guessing)

วิธีการนี้ให้ผู้สอบทำข้อสอบเหมือนกับแบบทดสอบชนิดเลือกตอบธรรมดา แต่ต่างกันที่การให้คะแนน วิธีการแก้การเดา มีวิธีที่น่าสนใจ 2 วิธี

3.1.1 วิธีการแก้การเดาโดยการลงโทษ เป็นวิธีที่มีข้อตกลงเบื้องต้นว่าข้อที่ผู้สอบตอบผิดเป็นผลมาจากการเดา ดังนั้นข้อที่ผู้สอบตอบผิดแทนที่จะได้ 0 แต่จะให้คะแนนติดลบ $-1/(K-1)$ K คือ จำนวนตัวเลือกของแบบสอบ (Simon et al. 1997; อ้างอิงจาก Lord. 1975) โดยใช้สูตรการให้คะแนนเพื่อแก้การเดาคำตอบ สูตรการให้คะแนนที่ใช้อยู่คือ

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

X_c คือ คะแนนที่ได้หลังจากการแก้การเดา

R คือ จำนวนข้อสอบที่ผู้สอบตอบถูก

W คือ จำนวนข้อสอบที่ผู้สอบตอบผิด

K คือ จำนวนตัวเลือกของข้อสอบแต่ละข้อ

วิธีการแก้การเดาสุ่มเหมาะกับแบบทดสอบความเร็วที่ผู้ตอบตอบไม่ทัน แต่ถ้าเป็นแบบทดสอบที่ไม่จำกัดเวลา ผู้ตอบอาจไม่ได้ตอบอย่างเดาสุ่มเพียงอย่างเดียว ผู้ตอบอาจมีความรู้บางส่วนในการตัดตัวลวงบางตัวออกได้ ดังนั้นการใช้สูตรแก้การเดาอาจไม่เหมาะสม เพราะควรที่จะพิจารณาถึงการตอบโดยมีความรู้บางส่วนของผู้ตอบด้วย

3.1.2 วิธีการให้รางวัลแก่ข้อที่ไม่ตอบ เป็นวิธีที่มีข้อตกลงเบื้องต้นว่า ข้อที่ผู้ตอบเว้นไว้นั้นถ้าเขาตอบจะมีโอกาสตอบถูกเป็น $1/K$ เมื่อ K คือ จำนวนตัวเลือกของแบบทดสอบ ซึ่งเขียนเป็นสูตรได้ดังนี้ (Simon. Et al. 1997; อ้างอิงจาก Ebel. 1965)

$$X_c = R + O/K$$

X_c คือ คะแนนหลังจากการแก้ไข

R คือ จำนวนข้อสอบที่ผู้สอบตอบถูก

O คือ จำนวนข้อสอบที่ผู้สอบไม่ตอบ

3.2 วิธีการให้คะแนนจากการตัดตัวลวง (Elimination Scoring)

วิธีนี้เป็นการให้ผู้สอบตัดตัวลวงออกไปได้บ้างอย่างถูกต้อง แม้จะไม่ทราบว่าคำตอบที่ถูกต้องคืออะไร มีวิธีการและงานวิจัยที่สนับสนุนวิธีนี้ดังนี้

3.2.1 วิธีการตอบและให้คะแนนแบบคูมบ์ (the Coombs response/scoring method) Coombs ได้จำแนกระดับความรู้ของผู้สอบออกเป็น 5 ประเภท คือ (Simon et al.1997; อ้างอิงจากCoombs, Milholland และ Womer 1956)

1) มีความรู้เต็มที่ (Full knowledge) คือ ผู้ที่สามารถตัดตัวลวงทั้งหมดออกได้

2) มีความรู้บางส่วน (Partial knowledge) คือ ผู้ที่สามารถตัดตัวลวงบางตัวออกได้แต่ไม่ทั้งหมด

3) มีความรู้ที่ผิดบางส่วน (Partial misinformation) คือ ผู้ที่สามารถตัดตัวลวงบางตัวออกได้บ้างและตัดตัวถูกออกด้วย

4) มีความรู้ที่ผิดเต็มที่ (Full misinformation) คือ ผู้ที่ตัดตัวถูกออกเพียงตัวเดียว

5) ไม่มีความรู้ (Absence of knowledge) คือ ผู้ที่ไม่ตอบ หรือตัดตัวเลือกทุกตัวออกหมด

วิธีการตอบและการให้คะแนนแบบคูมบ์ ใช้รูปแบบของแบบทดสอบเลือกตอบที่มีคำตอบถูกเพียงคำตอบเดียว โดยให้ผู้สอบตัดตัวลวงที่ไม่ใช่คำตอบถูกที่เขาแน่ใจออกไปมากที่สุด การให้คะแนนผู้ตอบจะได้รับ 1 คะแนน ในแต่ละตัวลวงที่ตัดออกอย่างถูกต้อง และในกรณีที่ผู้สอบมีความรู้ที่ผิดไปตัดตัวถูกออกโดยเข้าใจว่าเป็นตัวลวงจะถูกตัดคะแนน $K-1$ โดยที่ K คือ จำนวนตัวเลือก ดังนั้นช่วงคะแนนในแต่ละข้ออยู่ระหว่าง $-(K-1)$ ถึง $K-1$

3.2.2 วิธีการตอบและการให้คะแนนแบบอาร์โนลด์ (The Arnold Scoring/Response Method) (Simon et al. 1997; อ้างอิงจาก Arnold และ Arnold 1970) อาร์โนลด์เสนอวิธีที่พัฒนาจากทฤษฎีเกมเบื้องต้น (Game theory) มีการใช้ข้อสอบแบบเลือกตอบ

เช่นเดียวกับแบบของคูมบ์ โดยให้ผู้ตอบตัดตัวเลือกที่แน่ใจว่าเป็นตัวลวงหรือคำตอบผิดออกไปมากที่สุดที่เขารู้ แต่วิธีนี้มีระบบการให้คะแนนที่แตกต่างจากของคูมบ์ คือจะให้คะแนน $1/(K-1)$ คะแนนสำหรับการตัดตัวลวงถูกต้อง 1 ตัว แต่ถ้าผู้สอบตัดตัวเลือกถูกออกโดยเข้าใจว่าเป็นตัวลวงจะถูกหัก $-1/(K-1)$ คะแนน เมื่อ K คือจำนวนตัวเลือกแต่ละข้อ ระบบการให้คะแนนของอาร์โนลด์วางอยู่บนพื้นฐานคะแนนที่ถูกคาดหวัง (expected item score) เท่ากับ 0 เมื่อ ผู้สอบเดาสุ่ม ทำให้สามารถพัฒนาสูตรการให้คะแนนที่ยุติธรรมได้ดังนี้

$$C_d = (p)\left(d/(K-d)\right)$$

โดยที่ C_d คือ คะแนนเมื่อสามารถตัดตัวลวงบางตัวออกไปได้ และคะแนนที่เพิ่มขึ้นจากการเดาเป็น 0

d คือ จำนวนตัวลวงที่ตัดออกได้ถูกต้อง

K คือ จำนวนตัวเลือก

p คือ $1/(K-1)$

ส่วนในกรณีที่ผู้สอบตัดตัวเลือกที่ถูกเข้าไว้เป็นตัวลวงจะได้คะแนนลงโทษคือ $-p$ ตัวอย่างในกรณีของ 4 ตัวเลือก คะแนนที่ได้เป็นผลจากการตอบของผู้สอบ ดังนี้

คะแนน C_0 คือ ไม่สามารถตัดตัวลวงใดๆ ได้เลย จะได้คะแนนจากการแทนค่าในสูตรเท่ากับ 0

คะแนน C_1 คือ ตัดตัวลวงออกได้ 1 ตัว จะได้คะแนนจากการแทนค่าในสูตรเท่ากับ1/9

คะแนน C_2 คือ ตัดตัวลวงออกได้ 2 ตัว จะได้คะแนนจากการแทนค่าในสูตรเท่ากับ 1/3

คะแนน C_3 คือ ตัดตัวลวงออกได้ 3 ตัว จะได้คะแนนจากการแทนค่าในสูตรเท่ากับ 1

คะแนน $-P$ คือ ตัดตัวถูกออกเพราะเข้าใจว่าเป็นตัวลวงจะถูกหักออก -1/3 คะแนน

3.2.3 วิธีการตอบและให้คะแนนแบบครอส (The Cross Response/Scoring Method.

CRS) วิธีการนี้ให้ผู้สอบตัดตัวลวงออกเช่นเดียวกับวิธีการของคูมบ์ และ อาร์โนลด์ แต่การให้คะแนนแตกต่างกัน คือ วิธีการให้คะแนนของครอสจะให้คะแนนรายข้อเฉลี่ยต่ำสุด เมื่อผู้ตอบตัดตัวเลือกที่เป็นคำตอบโดยเข้าใจว่าเป็นตัวลวงและตัดตัวลวงอีก $K-2$ ตัว โดยเข้าใจว่าเป็นชุดของตัวเลือกที่ผิด ซึ่งผิดกับการให้คะแนนแบบคูมบ์ที่คะแนนรายข้อเฉลี่ยต่ำสุดจะเกิดจากการตัดตัวเลือกที่ถูกว่าเป็นตัวลวงเพียงคำตอบเดียว การให้คะแนนแบบครอสผู้สอบจะได้ 2 คะแนนในแต่ละตัวลวงที่เขาตัด

ออกได้อย่างถูกต้อง และจะได้ 1 คะแนนสำหรับแต่ละตัวเลือกที่ไม่ได้ตัดว่าเป็นตัวลวง เมื่อคำตอบถูกยังไม่ได้ถูกตัด แต่เมื่อใดที่ผู้ตอบตัดตัวเลือกที่ถูกด้วย misinformation เขาจะได้คะแนนเฉพาะตัวเลือกที่ไม่ได้ตัดตัวเลือกละ 1 คะแนน สำหรับในกรณีที่ข้อสอบมี 4 ตัวเลือก คะแนนที่ยอมให้เกิดขึ้นได้มีค่าตั้งแต่ 1 คะแนน ถึง 7 คะแนน

3.2.4 วิธีการเลือกชุดย่อยของคำตอบ (The Subset Selection Method. SST) เสนอโดยDressel และ Schmidt (1953) เป็นวิธีที่ตรงข้ามกับวิธีให้คะแนนแบบการตัดตัวลวง (ET) คือ วิธีนี้ให้ผู้สอบเลือกชุดย่อยของคำตอบที่เขาเห็นว่าน่าจะมีตัว เลือกที่เป็นคำตอบถูกรวมอยู่ด้วย การให้คะแนนขึ้นอยู่กับจำนวนตัวเลือกที่มีอยู่ในชุดย่อยของคำตอบ ยิ่งน้อยเท่าใดคะแนนจะได้เพิ่มขึ้น ในกรณีที่ตัวเลือกที่ถูกต้องรวมอยู่ในชุดย่อยของคำตอบนั้นด้วย แต่ถ้าชุดย่อยนั้นไม่มีคำตอบที่ถูกต้องอยู่ด้วย คะแนนจะถูกหักออกไปเท่ากับขนาดของจำนวนตัวเลือกที่ผู้ตอบเลือก ในกรณีที่ผู้ตอบไม่เลือกตัวเลือกใดเลยจะได้ 0 คะแนน

3.2.5 วิธีที่ประยุกต์จากวิธีของคูมบ์ (Modified Coombs Approach) หลักการ/แนวคิด

4. หลีกเลี่ยงการให้คะแนนลงโทษแก่ผู้สอบ ในกรณีที่ผู้สอบไม่มีความรู้หรือความรู้ผิด(misinformation) โดยที่ผู้สอบตัดตัวเลือกที่เป็นคำตอบถูกออกไป ถึงแม้ผู้สอบจะตัดตัวเลือกผิดออกไปได้บ้าง ซึ่งจะถือว่าตัวเลือกผิดที่เขาตัดออกไป ตัดออกไปโดยไม่มีความรู้บางส่วนที่แท้จริง คะแนนที่ได้ในกรณีนี้จึงควรเป็น 0 คะแนน

5. เพื่อแก้ปัญหาคะแนนที่ได้ในแต่ละข้อ ไม่ให้ไปมีผลกระทบต่อคะแนนในข้ออื่น หรือเกิดการชดเชยคะแนนกัน ในกรณีที่ข้อสอบทั้งฉบับไม่ได้มีความเป็นเอกพันธ์ ดังนั้นการให้คะแนนมีค่าติดลบตามวิธีของคูมบ์ จึงมีผลต่อคะแนนในข้ออื่น เมื่อนำไปหาคะแนนรวม จะเกิดการชดเชยกัน

6. เพื่อให้เกิดความยุติธรรมแก่ผู้สอบที่มีความรู้เต็ม (full knowledge) โดยสามารถตัดตัวเลือกที่เป็นคำตอบผิดออกไปได้ทุกตัว ซึ่งจะได้คะแนนสูงสุดในข้อนั้น หรือผู้ที่มีความรู้บางส่วนที่แท้จริง (partial knowledge) ซึ่งสามารถตัดตัวเลือกผิดออกไปได้บ้าง จะได้คะแนนตามจำนวนตัวเลือกที่เป็นคำตอบผิดที่ตัดออกไปได้จะเป็นตัวบอกถึงระดับความรู้ที่เขามีอยู่อย่างแท้จริง

7. เพื่อแก้ปัญหาการให้คะแนนติดลบ ที่ไม่เสมอภาคกันระหว่างผู้สอบที่ตัดจำนวนตัวเลือกออกไม่เท่ากัน โดยที่ตัดตัวเลือกออกมากแล้วตัดตัวเลือกถูกออกไปด้วย น่าจะได้คะแนนติดลบมาก

วิธีการตอบทำเช่นเดียวกับวิธีของคูมบ์ คือให้ผู้ตอบทำเครื่องหมาย (×) ตรงตัวเลือกที่เห็นว่าผิดทุกตัวเลือก ถ้าไม่แน่ใจให้เว้นข้ามไป ไม่ต้องทำ ส่วนวิธีการตรวจให้คะแนนแสดงให้เห็นดังสูตรต่อไปนี้

$$X = \frac{1}{2}\left[n + (n)c\right]$$

เมื่อ X คือ คะแนนที่ผู้สอบจะได้จากการตอบข้อสอบข้อหนึ่ง

n คือ จำนวนตัวเลือกที่ผู้สอบเลือกในการตอบข้อสอบแต่ละข้อ

c คือ ค่ากำหนดเงื่อนไข คือ $c=1$ เมื่อผู้ตอบตัดตัวเลือกที่เป็นคำตอบผิดออกได้ถูกต้อง

สำหรับสูตรการให้คะแนน ได้ใช้แนวคิดการปรับสเกลของคะแนนเพื่อไม่ให้มีค่าติดลบ และให้เกิดความยุติธรรมสำหรับผู้สอบที่เลือกจำนวนตัวเลือกในการตอบในแต่ละข้อไม่เท่ากัน ซึ่งมีผลต่อความน่าจะเป็นในการได้คะแนนมากน้อยต่างกัน (สุพจน์ เกิดสุวรรณ์. 2545: 61-64)

วิธีที่ประยุกต์จากวิธีของเดรสเซลและชมิด (Modified Dressel and Schmidt Approach)

หลักการ/แนวคิด

หลักในการพัฒนาวิธีที่ประยุกต์จากวิธีของเดรสเซลและชมิด

1. หลีกเลี่ยงการให้คะแนนลงโทษผู้สอบ ในกรณีผู้สอบตอบผิดและเป็นคะแนนติดลบ ซึ่งไม่ให้ความหมายในการบอกระดับความรู้ของผู้สอบ

2. เพื่อแก้ปัญหาคะแนนที่ได้ในแต่ละข้อ ไม่ให้ไปมีผลกระทบต่อคะแนนในข้ออื่น หรือเกิดการชดเชยคะแนนกัน ในกรณีที่ข้อสอบทั้งฉบับไม่ได้มีความเป็นเอกพันธ์ ถ้ามีการให้คะแนนติดลบซึ่งมีผลต่อคะแนนในข้ออื่น เมื่อนำไปหาคะแนนรวมจะเกิดการชดเชยคะแนนกัน

3. แสดงให้เห็นถึงความน่าจะเป็นที่ผู้สอบจะได้คะแนนในกรณีที่ผู้สอบเลือกหลายตัวเลือก และนำค่าความน่าจะเป็นที่ผู้สอบจะได้คะแนนลบออกจากคะแนนเต็ม ในข้อสอบข้อนั้น เพื่อให้เกิดความยุติธรรมของผู้สอบทุกคน

เพื่อให้เกิดความยุติธรรมแก่ผู้สอบที่มีความรู้เต็ม (full knowledge) โดยการเลือกคำตอบที่ถูกที่สุดเพียงคำตอบเดียว ซึ่งจะได้คะแนนสูงสุดในข้อนั้น หรือผู้ที่มีความรู้บางส่วนที่แท้จริง (partial knowledge) จะได้คะแนนจากการตอบข้อสอบนั้น โดยการเลือกตัวเลือกที่น่าจะเป็นคำตอบที่ถูก ซึ่งจะได้คะแนนตามจำนวนตัวเลือกที่เหลืออยู่ ถ้าตัวเลือกที่เลือกนั้นมีคำตอบถูกรวมอยู่ด้วย

วิธีการตอบทำเช่นเดียวกับวิธีของเดรสเซลและชมิด คือ ให้ผู้ตอบทำเครื่องหมาย (×) ตรงตัวเลือกที่เห็นว่าเป็นคำตอบที่ถูก ถ้าไม่แน่ใจเห็นว่ามีตัวเลือกอื่นอีกที่น่าจะเป็นคำตอบที่ถูก ก็สามารถที่จะทำเครื่องหมาย (×) ที่ตัวเลือกนั้นอีกกี่ตัวก็ได้ ส่วนวิธีการตรวจให้คะแนนแสดงให้เห็นดังสูตรต่อไปนี้

$$X=\left[k-k\frac{n}{k}\right]c \quad \text{หรือ} \quad X=[k-n](c)$$

เมื่อ X คือ คะแนนที่ผู้สอบจะได้จากการตอบข้อสอบข้อหนึ่ง

n คือ จำนวนตัวเลือกที่ผู้ตอบเลือก

k	คือ	จำนวนตัวเลือกทั้งหมดในข้อสอบ
c	คือ	ค่าความสอดคล้องในชุดของตัวเลือกที่ผู้สอบเลือกมี 2 ค่า คือ $c=1$ เมื่อมีตัวเลือกถูกอยู่ในชุดของตัวเลือกที่ผู้ตอบเลือก $c=0$ เมื่อไม่มีตัวเลือกถูกอยู่ในชุดของตัวเลือกที่ผู้ตอบเลือกคะแนนรายข้อที่ได้จะอยู่ในช่วง 0 ถึง $(k-n)$ เมื่อ k คือ จำนวนตัวเลือก ในข้อสอบ n คือ จำนวนตัวเลือกที่ผู้สอบเลือก (สุพจน์ เกิดสุวรรณ์. 2545: 66-69)

วิธีที่ประยุกต์จากวิธีของอาร์โนลด์ (Modified Arnold Approach)

หลักการ/แนวคิด

1. เพื่อให้เกิดความยุติธรรมแก่ผู้สอบทุกคนในกรณีที่อาร์โนลด์กำหนดให้มีคะแนนลงโทษเท่ากันทุกคน ในเมื่อผู้สอบตอบผิด โดยตัดตัวเลือกที่เป็นคำตอบถูกออกไปด้วย ซึ่งจะได้คะแนนติดลบ จึงได้ปรับค่า p (คะแนนลงโทษ) ของอาร์โนลด์ให้เป็น 0 เพื่อจะให้คะแนนผู้ที่ตอบผิดในข้อนั้นเป็น 0

2. เพื่อให้เกิดความยุติธรรมแก่ผู้สอบที่มีความรู้เต็ม (full knowledge) ที่สามารถตัดตัวเลือกที่เป็นคำตอบผิดออกไปได้ทุกตัว ซึ่งจะได้คะแนนสูงสุดในข้อนั้น และผู้ที่มีความรู้บางส่วน (partial knowledge) สามารถตัดตัวลวงออกไปได้บ้าง จะได้คะแนนบางส่วน คะแนนที่ได้จะแสดงถึงระดับความสามารถที่แท้จริงของผู้สอบ

3. เพื่อแก้ปัญหาคะแนนแต่ละข้อ มีผลกระทบต่อกัน หรือเกิดการชดเชยคะแนนกันในกรณีที่ข้อสอบทั้งฉบับไม่ได้มีความเป็นเอกพันธ์ เมื่อมีการให้คะแนนติดลบ เมื่อนำไปหาคะแนนรวมจะมีการชดเชยคะแนนกัน

วิธีการตอบทำเช่นเดียวกับวิธีของอาร์โนลด์ คือ ให้ผู้ตอบทำเครื่องหมาย (×) ตรงตัวเลือกที่เห็นว่าผิดทุกตัวเลือก ถ้าไม่แน่ใจให้เว้นข้ามไป ไม่ต้องทำ ส่วนวิธีการตรวจให้คะแนนแสดงให้เห็นดังสูตรต่อไปนี้

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

เมื่อ	C_d	คือ	คะแนนที่ผู้สอบได้เมื่อสามารถตัดตัวเลือกที่เป็นคำตอบที่ผิดออกไปได้
	d	คือ	จำนวนตัวเลือกที่เป็นคำตอบผิดที่ผู้สอบตัดออกได้ถูกต้อง
	k	คือ	จำนวนตัวเลือกในข้อสอบ
และ	p	คือ	คะแนนลงโทษเมื่อผู้สอบตัดตัวเลือกที่เป็นคำตอบถูกออกไปด้วย ในที่นี้จะกำหนดค่า p เป็น 2 ค่า คือ

$p = 1$ เมื่อผู้ตอบตัดตัวเลือกผิดออกไปได้ถูกต้อง

$p = 0$ เมื่อผู้ตอบตัดตัวเลือกที่เป็นคำตอบถูกออกไปด้วย

สูตรนี้ได้ปรับค่า p ซึ่งอาร์โนลด์ ได้กำหนดให้เป็นคะแนนการลงโทษในกรณีที่ผู้สอบตัดตัวเลือกที่เป็นคำตอบที่ถูกออกไป แต่ค่า p ที่อาร์โนลด์กำหนดไม่ได้ตั้งบนพื้นฐานของความน่าจะเป็นในการตอบข้อสอบถูก และความยุติธรรมของผู้สอบเพราะแต่ละคนจะได้คะแนนลงโทษเท่ากันหมด คือ $-1/(k-1)$ เมื่อ k คือจำนวนตัวเลือกทั้งหมดของข้อสอบ ในการปรับค่า p เพื่อที่จะให้คะแนนอยู่ในสเกลของค่าบวก และให้เท่ากับ 1 เพื่อให้คะแนนที่ตอบถูกแต่ละตัวเลือกเพิ่มทีละ 1 ช่วงสเกลทั้งนี้เพื่อความยุติธรรมสำหรับผู้สอบทุกคน(สุพจน์ เกิดสุวรรณ์. 2545: 72-75)

1.1 วิธีการของการทดสอบความน่าจะเป็น (Probability Testing)

วิธีนี้มีความยืดหยุ่น และเปิดโอกาสให้ผู้สอบแสดงความรู้บางส่วนมากที่สุด โดยรายงานความน่าจะเป็นในตัวเลือกที่ผู้สอบคิดว่าเป็นคำตอบที่ถูก โดยมีตำแหน่งคะแนนที่เป็นไปได้ 101 ตำแหน่ง (ช่วงจาก 0-1.00) โดยแบ่งระดับความรู้เป็น 5 ประเภท ดังนี้ (Wallsten, Budescu; & Zwick. 1993)

1. มีความรู้เต็มที่ (Full knowledge) คือ ให้ความน่าจะเป็นเท่ากับ 1 แก่คำตอบที่ถูกต้อง
2. มีความรู้บางส่วน (Partial knowledge) คือ ให้ความน่าจะเป็นมากกว่า0 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 of misinformation) คือ ให้ความน่าจะเป็นแก่ทุกตัวเลือกเท่ากันหรือไม่ตอบ

1.2 วิธีการแสดงความมั่นใจ (Confidence Marking)

วิธีนี้คิดค้นโดย Dressel และ Schmidt (1953) โดยที่ผู้สอบถูกถามให้แสดงความมั่นใจในคำตอบที่คิดว่าถูก โดยใช้ C-point scale (โดยทั่วไป $3 \le C \le 5$ หรืออาจเป็นการบอกความมั่นใจเป็นภาษา คือ ไม่แน่ใจ ค่อนข้างแน่ใจ แน่ใจมาก ซึ่งมีการแบ่งระดับความรู้ออกเป็น 5 ระดับ คือ

1. Full knowledge คือ การเลือกคำตอบได้ถูก ด้วยความมั่นใจสูงสุด
2. Partial knowledge คือ การเลือกคำตอบได้ถูกต้อง ด้วยความมั่นใจระดับต่ำ
3. Partial misinformation คือ การเลือกคำตอบผิด ด้วยความมั่นใจระดับต่ำ
4. Full misinformation คือ การเลือกคำตอบผิด ด้วยความมั่นใจระดับสูงสุด
5. Absence of knowledge คือ การไม่เลือกคำตอบเว้นว่างไว้

1.3 วิธีการจัดอันดับอย่างสมบูรณ์ (Complete Ordering)

วิธีการนี้เป็นวิธีเฉพาะของวิธีการของการทดสอบความน่าจะเป็น (PT) คือ แทนที่จะให้แสดงความน่าจะเป็นในแต่ละตัวเลือก ผู้สอบจะต้องจัดอันดับของตัวเลือก ซึ่งทำให้ง่ายขึ้นต่อการนำไปใช้ ทำให้สามารถจำแนกระดับความรู้เป็น 3 ระดับ คือ

1. Full knowledge คือ การจัดอันดับให้คำตอบถูกอยู่ในตำแหน่งที่สูงที่สุด
2. Partial knowledge คือ การจัดอันดับให้คำตอบถูกอยู่ในตำแหน่งกลาง
3. Absence of knowledge คือ การจัดอันดับให้คำตอบถูกอยู่ในตำแหน่งต่ำที่สุด

1.4 วิธีการจัดอันดับเป็นบางส่วน (Partial Ordering)

วิธีนี้เป็นวิธีที่ผสมระหว่างวิธีให้คะแนนแบบการตัดตัวลวง (ET) และ วิธีการจัดอันดับอย่างสัมบูรณ์ (CO) คือ ให้ผู้สอบจัดอันดับตัวเลือกที่ไม่สามารถ ตัดออกไปได้ วิธีนี้เสนอโดย De Finetti (1965) และถูกนำไปใช้โดย Diamond (1975)วิธีการจัดอันดับเป็นบางส่วนนี้จำแนกระดับความรู้เป็น 5 ระดับ คือ

1. Full knowledge คือ การจัดอันดับแก่คำตอบถูกในตำแหน่งสูงที่สุด
2. Partial knowledge คือ การจัดอันดับแก่หลายตัวเลือก โดยคำตอบถูกถูกจัดอันดับในตำแหน่งสูงที่สุด
3. Partial misinformation คือ การจัดอันดับแก่คำตอบถูกในตำแหน่งที่ต่ำกว่า ตัวลวง
4. Full misinformation คือ การจัดตำแหน่งแก่ตัวลวง โดยตัดคำตอบถูกออก
5. Absence of knowledge คือ การจัดอันดับแก่ตัวถูกในตำแหน่งที่ต่ำที่สุดหรือไม่ตอบ (เอมอร จังศิริพรปกรณ์. 2545.)

นอกจาก Partial Credit Model (PCM) แล้วยังมีโมเดลที่ถูกพัฒนาขึ้น ทำให้โมเดล PCM มีลักษณะเป็นโมเดลทั่วไป นั่นคือ Generalized Partial Credit Model (G-PCM)

ศิริชัย กาญจนวาสี (2555: 101) ได้กล่าวถึง Generalized Partial Credit Model ว่า มูรากิ (Muraki. 1992, 1993) ได้พัฒนา Generalized Partial Credit Model (G-PCM) ทำให้โมเดล PCM มีลักษณะเป็นโมเดลทั่วไป โดยยอมให้ข้อคำถามแต่ละข้อสามารถมีค่าพารามิเตอร์ความชันแตกต่างกันได้ G-PCM มีลักษณะเป็นโมเดลทั่วไปของ PCM และใช้หลักการคำนวณความน่าจะเป็นของการตอบแต่ละระดับขั้นการตอบโดยตรงแบบขั้นตอนเดียว (Direct IRT Method)

โมเดล

ใน G-PCM ลักษณะคำถามแต่ละข้อ (i) อธิบายด้วยค่าพารามิเตอร์ระดับความยากของขั้นการตอบ (δ_{ij}) และค่าพารามิเตอร์ความชันของข้อคำถาม (α_i) ฟังก์ชันทางคณิตศาสตร์ของโมเดลนี้จึงเขียนได้โดยแทนค่าพารามิเตอร์ความชันลงในสูตรของโมเดล PCM ดังนี้

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

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

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

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

α_i = ค่าพารามิเตอร์ความชันของข้อคำถามที่ i

ค่าระดับความยากของขั้นการตอบ (δ_{ij}) ซึ่งเป็นค่าสเกล θ ตรงตำแหน่งที่ตัดกันของโค้งรายการคำตอบ มีความหมายเหมือนกับโมเดล PCM ส่วนค่าความชัน (α_i) มีความหมายต่างจากโมเดลการตอบสนองข้อสอบแบบ 2 ค่า เพราะว่าในโมเดลการตอบสนองข้อสอบแบบมากกว่า 2 ค่า ค่าอำนาจจำแนกของข้อคำถามขึ้นอยู่กับผลร่วมกันระหว่างพารามิเตอร์ความชัน และการกระจายของพารามิเตอร์ Threshold ของรายการคำตอบ (สำหรับโมเดล GRM และ M-GRM) หรือตำแหน่งจุดตัดระหว่างรายการคำตอบ (สำหรับโมเดล PCM และ G-PCM) ในโมเดล G-PCM ค่าพารามิเตอร์ความชันแสดงถึงระดับความผันแปรของรายการคำตอบระหว่างข้อ เมื่อ θ ของผู้ตอบเปลี่ยนไป ถ้า α_i มีค่ามากกว่า 1.00 แสดงว่าโค้งรายการคำตอบของข้อนั้น มีความสูงชันมากกว่าโค้งรายการคำตอบในโมเดล PCM

จากการศึกษาเอกสารและงานวิจัย พบว่างานวิจัยส่วนใหญ่จะเปรียบเทียบวิธีการให้คะแนนความรู้บางส่วนหลายๆ วิธี และวิธีที่จะถูกนำมาเปรียบเทียบส่วนใหญ่จะเป็นวิธีการเปลี่ยนแปลงวิธีการตอบเนื่องจากเป็นวิธีที่สะท้อนถึงความรู้ที่มีอยู่ในตัวผู้สอบมากที่สุด ซึ่งการตรวจให้คะแนนความรู้บางส่วนสามารถให้สารสนเทศได้มากกว่าวิธีแบบดั้งเดิม จึงเป็นเรื่องที่น่าสนใจและทำการศึกษาต่อไป

3. การทำหน้าที่ต่างกันของข้อสอบ

3.1 ความหมายของการทำหน้าที่ต่างกันของข้อสอบและการทำหน้าที่ต่างกันของตัวลวง

3.1.1 ความหมายของการทำหน้าที่ต่างกันของข้อสอบ

ศิริชัย กาญจนวาสี. (2555: 117) ได้ให้ความหมายของการทำหน้าที่ต่างกันของข้อสอบ (DIF) หมายถึง การที่ข้อสอบทำให้ผู้สอบจากต่างกลุ่มกันที่มีความสามารถหรือคุณลักษณะที่มุ่งวัดเท่ากัน มีโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกัน หรือมีฟังก์ชันการตอบสนองข้อสอบแตกต่างกัน การทำหน้าที่ต่างกันของข้อสอบเกิดขึ้นเมื่อนำข้อสอบไปทดสอบกับผู้สอบกลุ่มย่อยต่างกันที่มีความสามารถหลัก (Primary ability) ระดับเดียวกันหรือมีคุณลักษณะแฝง (Latent trait) ที่ต้องการวัดเท่ากัน แต่มีความสามารถรอง (Secondary ability) แตกต่างกัน ทำให้ผู้สอบต่างกลุ่มที่นำมาจับคู่เปรียบเทียบมีโอกาสตอบข้อสอบถูกแตกต่างกัน

Holland; & Wainer. (1993) กล่าวถึงการทำหน้าที่ต่างกันของข้อสอบ หมายถึง สารสนเทศทางสถิติของข้อสอบที่ได้จากผลการตอบของผู้สอบต่างกลุ่มกัน และมีความสามารถเท่ากัน แต่มีโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกัน

Popham. (1981: 78) ได้ให้ความหมายของการทำหน้าที่ต่างกันของข้อสอบ หมายถึง ความชอบหรือความโอนเอียงที่ทำให้การพิจารณาตัดสินเป็นไปอย่างไม่ยุติธรรม

Allen; & Yen. (1979: 129) กล่าวเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบว่า เป็นข้อสอบที่วัดความสามารถหรือคุณลักษณะทางจิตวิทยาของผู้สอบแต่ละกลุ่มไม่ตรงกัน

Camilli; & Shepard. (1994: 8) ได้ให้ความหมายของการทำหน้าที่ต่างกันของข้อสอบว่า เป็นความคลาดเคลื่อนอย่างเป็นระบบ (Systematic Error) หรือความไม่เที่ยงตรง (Invalidity) ของการวัดชนิดหนึ่ง ทำให้สมาชิกบางกลุ่มได้ประโยชน์หรือเสียประโยชน์ เช่น ข้อสอบที่เข้าข้างกลุ่มนักเรียนชายมากกว่านักเรียนหญิง หรือคนผิวขาวกับคนผิวดำ หรือระหว่างคนที่อยู่ต่างภูมิภาคกัน เป็นต้น

ดังนั้นจึงสรุปได้ว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ข้อสอบที่นำไปวัดผู้สอบที่มีความสามารถหรือคุณลักษณะเท่ากันจากต่างกลุ่มกัน แต่มีโอกาสในการตอบถูกแตกต่างกัน เนื่องด้วยมีลักษณะบางอย่างที่แตกต่างกัน ทำให้ได้ผลการตอบข้อสอบที่ถูกต้องไม่เท่ากันทำให้เกิดการได้เปรียบเสียเปรียบจากการตอบข้อสอบเดียวกัน

3.1.2 การทำหน้าที่ต่างกันของตัวลวง

เรืองเดช ศิริกิจ. (2554: 21) ได้กล่าวถึงการทำหน้าที่ต่างกันของตัวลวงว่า การทำหน้าที่ต่างกันของตัวลวง (Differential Distractor Functioning) หมายถึง การที่สมาชิกในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีความสามารถหรือคุณลักษณะที่มุ่งวัดเท่ากันมีโอกาสในการเลือกตัวลวงแทนที่จะเลือกตัวถูกได้แตกต่างกัน

Kato, Moen; & Thurlow. (2009) ได้ให้ความหมายของว่าการทำหน้าที่ต่างกันของตัวลวง หมายถึง การวิเคราะห์ความแตกต่างในการเลือกตอบตัวเลือกที่เป็นตัวลวงของข้อสอบ โดยกลุ่มผู้สอบที่มีความแตกต่างกัน จะมีโอกาสในการเลือกตัวลวงได้แตกต่างกัน

Penfield. (2010) กล่าวเกี่ยวกับการทำหน้าที่ต่างกันของตัวลวง หมายถึง การที่สมาชิกในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีความสามารถเท่ากัน มีโอกาสในการเลือกตัวลวงแทนที่จะเลือกตัวถูกได้แตกต่างกัน

Green, Crone & Folk. (1989) ให้ความหมายของการทำหน้าที่ต่างกันของตัวลวงว่า คือ การวิเคราะห์การตอบข้อสอบโดยเป็นการตรวจสอบความแตกต่างของการทำหน้าที่ของข้อสอบผ่านตัวเลือกที่เป็นตัวลวงเมื่อผู้สอบต่างกลุ่มกัน แต่มีความสามารถเท่ากันเลือกตัวเลือกที่เป็นตัวลวง

ดังนั้นจึงสรุปได้ว่า การทำหน้าที่ต่างกันของตัวลวง หมายถึง การตรวจสอบความแตกต่างในการตอบตัวเลือกที่เป็นตัวลวงเมื่อผู้สอบต่างกลุ่มกัน ซึ่งมีความสามารถเท่ากัน จะมีโอกาสในการเลือกตัวลวงได้แตกต่างกัน ทำให้ได้ผลการตอบข้อสอบที่ถูกต้องไม่เท่ากันทำให้เกิดการได้เปรียบเสียเปรียบจากการตอบข้อสอบเดียวกัน

3.2 หลักการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

การตรวจการทำหน้าที่ต่างกันของข้อสอบ (DIF detection) เป็นการเปรียบเทียบผลการตอบสนองของข้อสอบเป็นรายข้อระหว่างกลุ่มผู้สอบอย่างน้อย 2 กลุ่ม ที่มีความสามารถหลัก (primary ability) ที่มุ่งวัดเท่ากัน แต่คาดว่าจะไม่มีความได้เปรียบหรือเสียเปรียบกัน โดยกลุ่มหนึ่งถือเป็น **กลุ่มอ้างอิง** (Reference group) ซึ่งคาดว่าน่าจะได้เปรียบในการตอบข้อสอบข้อนั้น หรือมีโอกาสตอบข้อสอบได้ถูกต้องมากกว่า ส่วนอีกกลุ่มคือ **กลุ่มเปรียบเทียบ** (Focal group) ซึ่งเป็นกลุ่มสนใจศึกษาและคาดว่าน่าจะได้เป็นกลุ่มที่เสียเปรียบ

ในการเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบจำเป็นต้องจับคู่ (matching) ผู้สอบตามความสามารถซึ่งเป็นเงื่อนไขสำคัญของการตรวจการทำหน้าที่ต่างกันของข้อสอบ เกณฑ์การจับคู่ (matching criteria) ที่นิยมใช้กันมี 2 วิธี (ศิริชัย กาญจนวาสี. 2555: 120–123) ดังนี้

1. ***เกณฑ์ภายนอก*** (External Criterion)

การวิเคราะห์การทำหน้าที่ต่างกัน โดยใช้เกณฑ์ภายนอกนี้สามารถนำไปใช้ได้ทั้งข้อสอบรายข้อ แบบทดสอบทั้งฉบับ โดยการใช้คะแนนจากแบบทดสอบอื่นเป็นเกณฑ์ภายนอกแล้วใช้เทคนิคการถดถอย (regression analysis) เพื่อทำการเปรียบเทียบเส้นกราฟความสัมพันธ์ระหว่างตัวแปรเกณฑ์กับตัวแปรทำนายระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ

หลักการนี้มีจุดมุ่งหมายเพื่อสร้างสมการทำนายตัวแปรเกณฑ์ ซึ่งเป็นคะแนนของแบบทดสอบอื่นจากตัวแปรทำนายซึ่งเป็นคะแนนรายข้อ หรือคะแนนแบบทดสอบระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ จะใช้คะแนนรายข้อเป็น

ตัวแปรทำนาย แต่ถ้าเป็นการวิเคราะห์การทำหน้าที่ต่างกันของแบบทดสอบจะใช้คะแนนรวมของแบบทดสอบทั้งสองฉบับเป็นตัวแปรทำนาย สำหรับตัวแปรเกณฑ์ที่ใช้เป็นเกณฑ์ภายนอก อาจใช้คะแนนรวมทั้งฉบับ หรือเกรดเฉลี่ย หรือผลสัมฤทธิ์ในงานที่เกี่ยวข้องของผู้สอบ (ศิริชัย กาญจนวาสี. 2555: 120 – 123; อ้างอิงจาก Cronbach. 1970)

สมการทำนาย สำหรับกลุ่มอ้างอิงและกลุ่มเปรียบเทียบแสดงผลได้ดังนี้

กลุ่มอ้างอิง (R) $Y_i = A_R + B_R X_i$

กลุ่มเปรียบเทียบ (F) $Y_F = A_F + B_F X_i$

เมื่อ Y_i = คะแนนของตัวแปรเกณฑ์ภายนอก

X_i = คะแนนของตัวแปรทำนาย

A = ค่าคงที่หรือค่าตัดแกน (intercept)

B = ค่าความชัน (Slope)

การใช้เกณฑ์ภายนอกมีข้อดี คือเกณฑ์ที่ใช้มีความเป็นอิสระจากข้อสอบ และแบบทดสอบที่ดีต้องมีการตรวจสอบ แต่มีจุดอ่อนตรงที่ความเหมาะสมของเกณฑ์ที่จะนำมาใช้ในทางปฏิบัติเป็นการยากที่จะหาตัวแปรเกณฑ์ภายนอกจากแบบทดสอบฉบับอื่นที่มีความตรงเชิงทำนายและมีความยุติธรรมสำหรับกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ถ้าตัวแปรเกณฑ์ภายนอกขาดคุณสมบัติดังกล่าวจะทำให้ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบหรือแบบทดสอบขาดความแม่นยำและความสมบูรณ์

2. **เกณฑ์ภายใน** (Internal Criterion)

การวิเคราะห์การทำหน้าที่ต่างกัน โดยใช้เกณฑ์ภายในเป็นการนำวิธีทางสถิติมาตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ หรือแบบทดสอบ โดยเน้นการพิจารณาจากโครงสร้างภายในของแบบทดสอบเป็นหลัก ด้วยการวิเคราะห์ผลการตอบข้อสอบและความสามารถหรือคะแนนจริงของผู้สอบจากแบบทดสอบฉบับนั้น เพื่อนำมาเปรียบเทียบระหว่างผู้สอบจากกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ที่มีความสามารถหรือคะแนนจริงเท่ากัน ว่าจะมีผลการตอบหรือโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกันหรือไม่ เพื่อบ่งชี้ถึงการทำหน้าที่ต่างกันของข้อสอบ การวิเคราะห์ในลักษณะนี้นิยมใช้ค่าสถิติต่าง ๆ เป็นตัวบ่งชี้ถึงการทำหน้าที่ต่างกันของข้อสอบ ค่าสถิติทดสอบที่นิยมนำมาใช้พอสรุปได้ดังนี้

2.1 การทดสอบปฏิสัมพันธ์ (Interaction)

ในระยะเริ่มแรกของการศึกษาความลำเอียงของข้อสอบ มีการใช้สถิติทดสอบเอฟ(F – test) จากการวิเคราะห์ความแปรปรวน (ANOVA) เพื่อทดสอบปฏิสัมพันธ์ระหว่างกลุ่มผู้สอบกับข้อสอบ ถ้าการทดสอบมีนัยสำคัญเป็นสัญญาณของการทำหน้าที่ต่างกันของข้อสอบ

(Cleary & Hiltion. 1968; Jensen. 1974) จากนั้นจึงทำการวิเคราะห์ต่อด้วยวิธีการ Post Hoc เพื่อระบุข้อสอบที่มีผลต่อการปฏิสัมพันธ์ ซึ่งเป็นข้อที่ทำหน้าที่ต่างกัน

วิธีการนี้มีข้อดีที่สามารถศึกษาผู้ตอบหลาย ๆ กลุ่มได้สะดวก แต่มีจุดอ่อนในเรื่องการควบคุมกลุ่มต่าง ๆ ให้มีความสามารถที่ทัดเทียมกัน ขนาดกลุ่มตัวอย่างของกลุ่มต่าง ๆและอัตราการคลาดเคลื่อนประเภทที่ 1 จะสูงขึ้น ถ้าจำนวนข้อสอบเพิ่มมากขึ้น

2.2 การวัดความเบี่ยงเบนสัมพันธ์ (Relative Deviation)

การคำนวณค่าความยากของข้อสอบ เช่น p, b เป็นต้น เมื่อคำนวณแยกระหว่างกลุ่มและแปลงให้เป็นค่าความยากมาตรฐาน สามารถนำมาพล๊อตเปรียบเทียบรายข้อ ถ้าข้อใดเบี่ยงเบนไปจากแกนหลักที่คาดหมาย หรือเบี่ยงเบนเกินจากความคลาดเคลื่อนมาตรฐานของ

ค่าความยากที่กำหนด ย่อมแสดงถึงการทำหน้าที่ต่างกันของข้อสอบ (Cleary & Hilton. 1968; Angoff & Ford. 1973) รวมทั้งสามารถคำนวณค่าสหสัมพันธ์ระหว่างค่าความยากง่ายรายข้อระหว่างกลุ่ม เพื่อแสดงถึงการทำหน้าที่ต่างกันของข้อสอบ ถ้าสหสัมพันธ์เข้าใกล้ 1.00 แสดงว่าค่าความยากสัมพัทธ์ของข้อสอบมีค่าใกล้เคียงกันระหว่างกลุ่มดังนั้นแบบทดสอบวัดคุณลักษณะคล้ายกันระหว่างกลุ่ม วิธีการนี้มีข้อดีและข้อเสียคล้ายการทดสอบปฏิสัมพันธ์ นอกจากนี้ค่าความยากง่ายของข้อสอบ (p) มิใช่ตัวแทนของค่าความยากจริงของข้อสอบ และได้รับอิทธิพลจากค่าแทรกซ้อนอื่น ได้แก่ ค่าอำนาจจำแนกและความสามารถของผู้สอบ

2.3 การเปรียบเทียบน้ำหนักตัวประกอบ (Factor Analysis)

เป็นเทคนิคทางสถิติที่นิยมใช้ในการตรวจสอบความตรงเชิงทฤษฎีหรือโครงสร้าง (Construct Validity) เมื่อนำมาวิเคราะห์ตัวประกอบมาใช้ในการวิเคราะห์โครงสร้างแบบทดสอบแยกตามกลุ่มผู้สอบ ความไม่สอดคล้องกันระหว่างน้ำหนักตัวประกอบบนคุณลักษณะสำคัญที่มุ่งวัด หรือความแตกต่างของค่าเฉลี่ยคะแนนตัวประกอบ(factor Score) ระหว่างกลุ่มผู้สอบ ย่อมสะท้อนการทำหน้าที่ต่างกันของข้อสอบและแบบทดสอบ

การใช้เทคนิคการวิเคราะห์ตัวประกอบเชิงสำรวจ (Exporator Factor Analysis; EFA) สำหรับศึกษาการทำหน้าที่ต่างกัน จะมีจุดอ่อนในเรื่องความไม่สอดคล้องระหว่างน้ำหนักตัวประกอบอาจเกิดจากความแตกต่างของความสามารถกลุ่มก็ได้ แนวทางที่เหมาะสมควรใช้เทคนิคการวิเคราะห์ตัวประกอบเชิงยืนยัน (Confirmator Factor Analysis ; CFA) นอกจากนี้ยังสมารถใช้ CFA สำหรับตรวจสอบความแตกต่างระหว่างกลุ่ม ในด้านคุณลักษณะหรือความสามารถรองได้อีกด้วย (Camili & Shepard. 1994)

2.4 การเปรียบเทียบโอกาสตอบข้อสอบถูก

การวิเคราะห์โอกาสตอบข้อสอบถูกของผู้สอบจากกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีความสามารถเท่ากัน เป็นแนวทางสำคัญที่นิยมใช้กันและเป็นที่ยอมรับกันในปัจจุบัน สำหรับบ่งชี้การทำหน้าที่ต่างกันของข้อสอบ มีการคำนวณค่าสถิติ 2 แนวทางดังนี้

- เปรียบเทียบค่าสัดส่วนหรือความน่าจะเป็นในการตอบข้อสอบถูกของผู้สอบต่างกลุ่มที่มีระดับความสามารถเท่ากัน เช่นวิธี แมนเทล – แฮนส์เซล (MH) เป็นต้น

- เปรียบเทียบค่าฟังก์ชันการตอบสนองข้อสอบ หรือโค้งลักษณะข้อสอบระหว่างกลุ่มที่มีระดับความสามารถเท่ากัน เป็นวิธีที่อยู่บนพื้นฐานของทฤษฎี IRT เช่น วิธีวัดความแตกต่างของพื้นที่ วิธีวัดความแตกต่างของค่าพารามิเตอร์ความยาก วิธีทดสอบไค – สแควร์ของลอร์ด (Lords' χ^2 - test) เป็นต้น

วิธีการนี้มีข้อดีที่สำคัญได้แก่ การคำนวณค่าทางสถิติของข้อสอบมีความน่าเชื่อถือ มีกลไกการควบคุมความสามารถเท่ากัน จึงเป็นวิธีการยอมรับกันทั่วไป แต่มีข้อจำกัดในด้านความความสลับซับซ้อนของแนวคิดพื้นฐาน และการวิเคราะห์มีความจำเป็นต้องใช้โปรแกรมคอมพิวเตอร์โดยเฉพาะ

3.3 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF detection) จำแนกตามลักษณะการตรวจให้คะแนนได้เป็น 2 ประเภท คือ ข้อสอบที่มีการให้คะแนนแบบทวิภาค หรือสองค่า (Dichotomous Scoring) และข้อสอบที่มีการให้คะแนนแบบพหุวิภาค หรือหลายค่า (Polytomous Scoring) วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแต่ละประเภท ยังสามารถจำแนกได้อีก 2 มิติ ได้แก่ **มิติลักษณะของตัวแปรเกณฑ์** ซึ่งแบ่งเป็นกลุ่มวิธีที่ใช้คะแนนสังเกตได้ (Observed Score) และกลุ่มวิธีที่ใช้คะแนนสังเกตไม่ได้หรือคะแนนของตัวแปรแฝง (Latent variable) และ**มิติลักษณะของสถิติวิเคราะห์** ซึ่งแบ่งเป็นกลุ่มวิธีที่ใช้สถิติพาราเมตริก (Parametric approach) และกลุ่มวิธีที่ใช้สถิตินันพาราเมตริก (Nonparametric approach) (ศิริชัย กาญจนวาสี. 2555: 124 – 126) ดังแสดงในตาราง 1

ตาราง 1 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่มีการตรวจให้คะแนนแบบทวิภาค (Dichotomous DIF) และพหุวิภาค (Polytomous DIF)

ประเภทและตัวแปรเกณฑ์	พาราเมตริก	นันพาราเมตริก
1. DIF แบบทวิภาค **1.1 คะแนนที่สังเกตได้** **(observed score)**	**ANOVA** **Logistic Regression**	**TID** **MH** **STND**
1.2 คุณลักษณะ/ตัวแปรแฝง **(latent variable)**	**IRT – D^2** **Lord's χ^2** **General IRTLR** **Loglinear IRTLR**	**SIBTEST**
2. DIF แบบพหุวิภาค **2.1 คะแนนที่สังเกตได้** **(observed score)** **2.2 คุณลักษณะ/ตัวแปรแฝง** **(latent variable)**	**ANOVA** **Polytomous Logistic Regression** **General IRTLR** **PCM**	**Polytomous STND** **GMH** **PolytomousSIBTEST** **GPCM**

1. วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนแบบทวิภาค

1.1 กลุ่มวิธีที่ใช้คะแนนที่สังเกตได้

วิธีในกลุ่มนี้มักวิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิม (CTT) หรือกลุ่มที่ไม่ใช้ทฤษฎีการตอบสนองข้อสอบ (Non – IRT approach) โดยใช้คะแนนรวมของผู้สอบเป็นเกณฑ์การจับคู่ของกลุ่มผู้สอบ วิธีการตรวจสอบที่สำคัญในกลุ่มนี้ ได้แก่

- การวิเคราะห์ความแปรปรวน (ANOVA)
- วิธีการวิเคราะห์การถดถอยโลจิสติก (Logistic Regression ; LR)
- วิธีแปลงค่าความยากง่ายของข้อสอบ (Transformed Item Difficulty ; TID)
- วิธีแมนเทล – แฮนส์เซล (Mantel – Haenzel ; MH)
- วิธีวัดดัชนีมาตรฐาน (Standardization ; STND) การปรับให้เป็นมาตรฐานด้วยน้ำหนักตัวประกอบ

1.2 กลุ่มที่ใช้คุณลักษณะแฝง

วิธีในกลุ่มนี้ใช้คุณลักษณะหรือตัวแปรแฝง ซึ่งวิเคราะห์บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ (IRT) สำหรับใช้เกณฑ์การจับคู่กลุ่มผู้สอบ วิธีการตรวจสอบที่สำคัญในกลุ่มนี้ได้แก่

- วิธีวัดพื้นที่ความแตกต่างระหว่างโค้งการตอบสนองข้อสอบ (IRT - D^2)
- วิธีไค – สแควร์ของลอร์ด (Lord's χ^2)
- วิธีอัตราส่วนไลค์ลิฮูดทั่วไป (Gerneral IRT Likelihood Ratio)
- วิธีอัตราส่วนไลค์ลิฮูด ลอกลิเนียร์ (Loglinear IRT Likelihood Ratio)
- วิธีซิปเทสท์ (SIBTEST)

2. วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนแบบพหุวิภาค

2.1 กลุ่มวิธีใช้คะแนนที่สังเกตได้

- วิธีการวิเคราะห์ความแปรปรวน (ANOVA)
- วิธีการวิเคราะห์การถดถอยโลจิสติกพหุวิภาค (Polytomous Logistic Regression)
- วิธีดัชนีมาตรฐานพหุวิภาค (Polytomous Standardization)
- วิธีแมนเทล – แฮนส์เซลทั่วไป (General Mantel – Haenzel ; GMH)

2.2 กลุ่มวิธีที่ใช้คุณลักษณะแฝง

- วิธีอัตราส่วนไลค์ลิฮูดในรูปทั่วไป (Gerneral IRT Likelihood Ratio)
- วิธีการให้คะแนนบางส่วน (Partial Credit Model ; PCM)
- วิธีซิปเทสท์พหุวิภาค (Polytomous SIBTEST)
- วิธีการให้คะแนนบางส่วนทั่วไป (Generalized partial Credit Model; GPCM)

3.4 เกณฑ์เปรียบเทียบคุณภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (Criteria for comparison between DIF Detecting Methods)

คุณภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบนิยมพิจารณาจากดัชนีบ่งชี้สำคัญ 2 ตัว ได้แก่ อำนาจการทดสอบ หรืออัตราความถูกต้องของการตรวจพบข้อที่ทำหน้าที่ต่างกัน (Power rate) และ ความคลาดเคลื่อนของการตรวจสอบ หรืออัตราตามความคลาดเคลื่อนประเภทที่ 1 (Type I error rate) ซึ่งเป็นโอกาสของการเกิดความคลาดเคลื่อนในลักษณะที่ตรวจพบว่า ข้อสอบทำหน้าที่ต่างกันทั้งๆที่ความเป็นจริงข้อสอบไม่ได้ทำหน้าที่ต่างกัน (ศิริชัย กาญจนวาสี. 2555: 151 – 152)

ในการคำนวณค่าสถิติตามวิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ มีจุดมุ่งหมายเพื่อทดสอบนัยสำคัญของผลการตรวจสอบ โดยมีสมมติฐานศูนย์ของการทดสอบ คือ ข้อสอบไม่ได้ทำหน้าที่ต่างกัน หรือ Ho: No DIF ผลการทดสอบสมมุติฐานของวิธีการตรวจสอบ DIF วิธีต่างๆนำสู่

การตัดสินใจว่า ยอมรับสมมุติฐานศูนย์ (Accept Ho) หรือปฏิเสธสมมุติฐานศูนย์ (Reject Ho) ผลของการตัดสินใจมีโอกาสเกิดเหตุการณ์ 4 ลักษณะ ดังตาราง 2

ตาราง 2 คุณภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

การตัดสินใจตามผลการตรวจสอบ	Ho: No DIF	
	ความเป็นจริง	
	Ho ถูก	Ho ผิด
Accept Ho	• ตัดสินถูก (True negative) ระดับความเชื่อมั่น (1-α)	• ตัดสินผิด (Type II Error, β) • False negative
Reject Ho	• ตัดสินผิด (Type I Error, α) • False positive	• ตัดสินถูก (True positive) • อำนาจการทดสอบ (1-β)

จากค่าสถิติที่คำนวณตามวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ จะนำไปสู่การตัดสินใจสรุปผลการตรวจสอบดังนี้

1. ตัดสินใจถูก มีโอกาสเกิดขึ้น 2 ลักษณะคือ การสรุปถูกว่า
   1.1 ข้อสอบ No DIF ตามความเป็นจริง (True negative)
   1.2 ข้อสอบ DIF ตามความเป็นจริง (True positive)
2. ตัดสินผิด มีโอกาสเกิดขึ้น 2 ลักษณะคือ การสรุปผิดว่า
   2.1 ข้อสอบ DIF ทั้งๆที่ความเป็นจริงข้อสอบ No DIF (False positive)
   2.2 ข้อสอบ No DIF ทั้งๆที่ความจริงข้อสอบ DIF (False negative)

เนื่องจากอำนาจการทดสอบ (1-β) กับ β เป็นค่าดัชนีมีสเกลที่ผกผันกันและความคลาดเคลื่อนประเภทที่ 1 (α) กับ 1-α ต่างเป็นดัชนีมีสเกลที่ผกผันเช่นเดียวกัน ดังนั้น

การพิจารณาดัชนีบ่งชี้คุณภาพ 2 ตัว คือ อัตราความถูกต้องของการตรวจพบ DIF (Power rate) และอัตราความคลาดเคลื่อนประเภทที่ 1 (Type I error rate) ก็เพียงพอที่จะให้สารสนเทศครบทั้ง 4 เหตุการณ์

สำหรับเกณฑ์ในการพิจารณาการทำหน้าที่ต่างกันของข้อคำถาม พิจารณาได้จากค่า Z(LOR) และค่า LOR โดยมีขั้นตอนการแปลผลดังนี้ (นิอร ไชยพรพัฒนา. 2554: 150; อ้างอิงจาก Penfield. 2012)

1. พิจารณาค่า Z(LOR)

1.1 ค่า Z(LOR) > 2 หรือค่า Z(LOR) < -2 แสดงว่าข้อคำถามข้อนั้นมีการทำหน้าที่ต่างกันระหว่างกลุ่มอย่างมีนัยสำคัญ ให้พิจารณาค่า LOR ในขั้นตอนที่ 2

1.2 -2 ≤ ค่า Z(LOR) ≤ 2 แสดงว่าข้อคำถามนั้นทำหน้าที่ไม่ต่างกันระหว่างกลุ่ม

2. พิจารณาค่า LOR สำหรับข้อคำถามที่มีการทำหน้าที่ต่างกันระหว่างกลุ่ม

2.1 ค่า LOR มีค่าเป็นบวก แสดงว่าข้อคำถามจะเข้าข้างกลุ่มอ้างอิง (Reference group)

2.2 ค่า LOR มีค่าเป็นลบ แสดงว่าข้อคำถามจะเข้าข้างกลุ่มเปรียบเทียบ (Focal group)

งานวิจัยนี้ผู้วิจัยตรวจสอบลักษณะที่แตกต่างกัน คือ ตัวแปรเพศ เพื่อใช้เป็นเกณฑ์ในการคัดเลือกข้อสอบ โดยที่ข้อที่คัดเลือกมานั้น ต้องไม่เกิดการได้เปรียบเสียเปรียบระหว่างเพศ ซึ่งใช้ค่า Z(LOR) และค่า LOR ในการพิจารณา

4. การแก้ปัญหาทางคณิตศาสตร์

การแก้ปัญหา เป็นกระบวนการที่ผู้เรียนควรจะเรียนรู้ ฝึกฝน และพัฒนาให้เกิดทักษะขึ้นในตัวนักเรียน การเรียนการแก้ปัญหาทางคณิตศาสตร์จะช่วยให้ผู้เรียนมีแนวทางการคิดที่หลากหลาย มีนิสัยกระตือรือร้น ไม่ย่อท้อ และมีความมั่นใจในการแก้ปัญหาที่เผชิญอยู่ทั้งภายในและภายนอกห้องเรียน ตลอดจนเป็นทักษะพื้นฐานที่ผู้เรียนสามาถนำติดตัวไปใช้แก้ปัญหาในชีวิตประจำวันได้นานตลอดชีวิต

ปัญหา หมายถึง สถานการณ์ที่เผชิญอยู่และต้องการค้นหาคำตอบ โดยที่ยังไม่รู้วิธีการหรือขั้นตอนที่จะได้คำตอบของสถานการณ์นั้นในทันที แล้วสถานการณ์นั้นก็ไม่ใช่ปัญหาอีกต่อไป อย่างไรก็ตามปัญหาสำหรับคนหนึ่งอาจไม่ใช่ปัญหาสำหรับอีกคนหนึ่งก็ได้

ปัญหาทางคณิตศาสตร์ หมายถึง สถานการณ์ที่เกี่ยวกับคณิตศาสตร์ ซึ่งเผชิญอยู่และต้องการค้นหาคำตอบ โดยที่ยังไม่รู้วิธีการหรือขั้นตอนที่จะได้คำตอบของสถานการณ์นั้นในทันที 0
การแก้ปัญหาทางคณิตศาสตร์ หมายถึง กระบวนการในการประยุกต์ความรู้ทางคณิตศาสตร์ ขั้นตอน/กระบวนการแก้ปัญหา ยุทธวิธีแก้ปัญหา และประสบการณ์ที่มีอยู่ไปใช้ในการค้นหาคำตอบของปัญหาทางคณิตศาสตร์

(สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี กระทรวงศึกษาธิการ. 2555: 6-7)

กระบวนการแก้ปัญหา

กระบวนการแก้ปัญหาที่ยอมรับและนำมาใช้กันอย่างแพร่หลาย คือ กระบวนการแก้ปัญหาตามแนวคิดของโพยา (Polya) ซึ่งประกอบด้วยขั้นตอนสำคัญ 4 ขั้นตอน ดังนี้

ขั้นที่ 1 ทำความเข้าใจปัญหา

ขั้นที่ 2 วางแผนแก้ปัญหา

ขั้นที่ 3 ดำเนินการตามแผน

ขั้นที่ 4 ตรวจสอบผล

(สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี กระทรวงศึกษาธิการ. 2555: 8)

ดิวอี้ ได้เสนอกระบวนการคิดแก้ปัญหาไว้ 5 ขั้นตอน ดังนี้

1, ขั้นเตรียมการ

2, ขั้นวิเคราะห์ปัญหา

3. ขั้นเสนอแนวทางในการแก้ปัญหา

4. ขั้นตรวจสอบผล

5. ขั้นนำไปประยุกต์ใช้

(ชมนาด เชื้อสุวรรณทวี. 2544: 124-125)

ทอลตัน ได้ศึกษาและรวบรวมขั้นตอนในการแก้โจทย์ปัญหาจากตำราคณิตศาสตร์ที่ผลิตในช่วง ค.ศ.1960-1979 พบว่า ผู้เขียนตำราหลายท่านในสมัยนั้น แนะนำให้นักเรียนแก้โจทย์ปัญหาโดยใช้ขั้นตอน ดังนี้

1. อ่านโจทย์

2. กำหนดว่าโจทย์ถามอะไรบ้าง

3. กำหนดว่าโจทย์ให้อะไรบ้าง

4. เลือกวิธีการ

5. ลงมือแก้ปัญหา

ส่วนทอลตัน ได้ศึกษาผลการวิจัยของโพลยา และคนอื่นๆ ได้ข้อสรุปว่า ในการแก้ปัญหามีความส่งเสริมให้นักเรียนเพิ่มทักษะการคิดวางแผนแก้ปัญหาโดยใช้ขั้นตอนดังนี้

1. อ่านโจทย์

2. วางแผนการแก้ปัญหา ซึ่งอาจสร้างตาราง คิดย้อนกลับ ใช้ตรรกศาสตร์ เขียนแผนภาพ แปลงปัญหาให้ง่ายขึ้น เลือกวิธีการเดาและทดสอบ หรืออื่นๆ

3. ดำเนินการแก้ปัญหา

4. ตรวจสอบกระบวนการแก้ปัญหา

(ประพนธ์ จ่ายเจริญ. 2547: 118-119)

จากที่กล่าวมาสรุปได้ว่า กระบวนการแก้ปัญหาประกอบด้วยขั้นตอนสำคัญ 4 ขั้นตอน คือ

1. การเตรียมการและทำความเข้าใจกับปัญหา
2. วิเคราะห์วางแผนการแก้ปัญหา
3. เลือกวิธีและดำเนินการแก้ปัญหา
4. ตรวจสอบผล

ยุทธวิธีแก้ปัญหา

สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี กระทรวงศึกษาธิการ (2555: 11-36) ได้กล่าวเกี่ยวกับยุทธวิธีแก้ปัญหาว่า ยุทธวิธีแก้ปัญหาเป็นเครื่องมือสำคัญและสามารถนำมาใช้ในการแก้ปัญหาได้ดี ที่พบบ่อยในคณิตศาสตร์ มีดังนี้

1. การค้นหาแบบรูป

การค้นหาแบบรูป เป็นการวิเคราะห์ปัญหาและค้นหาความสัมพันธ์ของข้อมูลที่มีลักษณะเป็นระบบหรือเป็นแบบรูปในสถานการณ์ปัญหานั้นๆ แล้วคาดเดาคำตอบ ซึ่งคำตอบที่ได้จะยอมรับว่าเป็นคำตอบที่ถูกต้องเมื่อผ่านการตรวจสอบยืนยัน ยุทธวิธีนี้มักจะใช้ในการแก้ปัญหาที่เกี่ยวกับเรื่องจำนวนและเรขาคณิต การฝึกฝนการหาแบบรูปในเรื่องดังกล่าวเป็นประจำ จะช่วยนักเรียนในการพัฒนาความรู้สึกเชิงจำนวนและทักษะการสื่อสาร ซึ่งเป็นทักษะที่ช่วยให้นักเรียนสามารถประมาณและคาดคะเน จำนวนที่พิจารณาโดยยังไม่ต้องคิดคำนวณก่อน ตลอดจนสามารถสะท้อนความรู้ความเข้าใจในแนวคิดทางคณิตศาสตร์และกระบวนการคิดของตนได้

2. การสร้างตาราง

การสร้างตาราง เป็นการจัดระบบข้อมูลใส่ในตาราง ตารางสร้างขึ้นจะช่วยในการวิเคราะห์หาความสัมพันธ์ อันจะนำไปสู่การค้นพบแบบรูปหรือข้อชี้แนะอื่นๆ ตลอดจนช่วยให้ไม่หลงลืมหรือสับสนในกรณีใดกรณีหนึ่ง เมื่อต้องแสดงกรณีที่เป็นไปได้ทั้งหมดของปัญหา

3. การเขียนภาพหรือแผนภาพ

การเขียนภาพหรือแผนภาพ เป็นการอธิบายสถานการณ์และแสดงความสัมพันธ์ของข้อมูลต่างๆ ของปัญหาด้วยภาพหรือแผนภาพ ซึ่งการเขียนภาพหรือแผนภาพจะช่วยให้เข้าใจปัญหาได้ง่ายขึ้น และบางครั้งก็สามารถหาคำตอบของปัญหาได้โดยตรงจากภาพหรือแผนภาพนั้น

4. การแจงกรณีที่เป็นไปได้ทั้งหมด

การแจงกรณีที่เป็นไปได้ทั้งหมด เป็นการจัดระบบข้อมูล โดยแยกเป็นกรณีๆ ที่เกิดขึ้นทั้งหมด ในการแจงกรณีที่เป็นไปได้ทั้งหมด นักเรียนอาจขจัดกรณีที่ไม่ใช่ออกก่อน แล้วค่อยค้นหาระบบหรือแบบรูปของกรณีที่เหลืออยู่ ซึ่งถ้าไม่มีระบบในการแจงกรณีที่เหมาะสม ยุทธวิธีนี้ก็จะไม่มีประสิทธิภาพ ยุทธวิธีนี้จะใช้ได้ดีถ้าปัญหานั้นมีจำนวนกรณีที่เป็นไปได้แน่นอน ซึ่งบางครั้งเราอาจใช้การค้นหาแบบรูปและการสร้างตารางมาช่วยในการแจงกรณีด้วยก็ได้

5. การคาดเดาและตรวจสอบ

การคาดเดาและตรวจสอบ เป็นการพิจารณาข้อมูลและเงื่อนไขต่างๆ ที่เป็นปัญหากำหนดผสมผสานกับประสบการณ์เดิมที่เกี่ยวข้อง มาสร้างข้อความคาดการณ์ แล้วตรวจสอบความถูกต้องของข้อความคาดการณ์นั้น ถ้าการคาดเดาไม่ถูกต้องก็คาดเดาใหม่โดยอาศัยประโยชน์จากความไม่ถูกต้องของการคาดเดาในครั้งแรกๆ เป็นกรอบในการคาดเดาคำตอบของปัญหาครั้งต่อไป นักเรียนควรคาดเดาอย่างมีเหตุผลและทิศทาง เพื่อให้สิ่งที่คาดเดานั้นเข้าใกล้คำตอบที่ต้องการมากที่สุด

6. การคิดแบบย้อนกลับ

การคิดแบบย้อนกลับ เป็นการวิเคราะห์ปัญหาที่พิจารณาจากผลย้อนกลับไปสู่เหตุ โดยเริ่มจากข้อมูลที่ได้ในขั้นตอนสุดท้าย แล้วคิดย้อนขั้นตอนกลับมาสู่ข้อมูลที่ได้ในขั้นตอนเริ่มต้น การคิดแบบย้อนกลับใช้ได้ดีกับการแก้ปัญหาที่ต้องการอธิบายถึงขั้นตอนการได้มาซึ่งคำตอบ

7. การเขียนสมการ

การเขียนสมการ เป็นการแสดงความสัมพันธ์ของข้อมูลที่กำหนดของปัญหาในรูปของสมการ ซึ่งบางครั้งอาจเป็นอสมการก็ได้ ในการแก้สมการนักเรียนต้องวิเคราะห์สถานการณ์ปัญหาเพื่อหาว่า ข้อมูลและเงื่อนไขที่กำหนดมามีอะไรบ้าง และสิ่งที่ต้องการหาคืออะไร หลังจากนั้นกำหนดตัวแปรแทนสิ่งที่ต้องการหาหรือแทนสิ่งที่เกี่ยวข้องกับข้อมูลที่กำหนดมาให้ แล้วเขียนสมการหรืออสมการแสดงความสัมพันธ์ของข้อมูลเหล่านั้น ในการหาคำตอบของสมการ มักใช้สมบัติของการเท่ากันมาช่วยในการแก้สมการ ซึ่งได้แก่ สมบัติสมมาตร สมบัติถ่ายทอด สมบัติการบวกและสมบัติการคูณ และเมื่อใช้สมบัติของการเท่ากันมาช่วยแล้ว ต้องมีการตรวจสอบคำตอบของสมการตามเงื่อนไขของปัญหา ถ้าเป็นไปตามเงื่อนไขของปัญหาถือว่าคำตอบที่ได้เป็นคำตอบที่ถูกต้องของปัญหานี้ ยุทธวิธีนี้มักใช้บ่อยในปัญหาทางพีชคณิต

8. การเปลี่ยนมุมมอง

การเปลี่ยนมุมมอง เป็นการเปลี่ยนการคิดหรือมุมมองให้แตกต่างไปจากที่คุ้นเคยหรือที่ต้องทำตามขั้นตอนทีละขั้น ทั้งนี้เพื่อให้แก้ปัญหาได้ง่ายขึ้น ยุทธวิธีนี้มักใช้ในกรณีที่แก้ปัญหาด้วยยุทธวิธีอื่นไม่ได้ผล สิ่งสำคัญของยุทธวิธีนี้ก็คือ การเปลี่ยนมุมมองที่แตกต่างไปจากเดิม

9. การแบ่งเป็นปัญหาย่อย

การแบ่งเป็นปัญหาย่อย เป็นการแบ่งปัญหาใหญ่หรือปัญหาที่มีความซับซ้อนหลายขั้นตอนออกเป็นปัญหาย่อยหรือเป็นส่วนๆ ซึ่งในการแบ่งเป็นปัญหาย่อยนั้นนักเรียนอาจลดจำนวนของข้อมูลลง หรือเปลี่ยนข้อมูลให้อยู่ในรูปที่คุ้นเคยและไม่ซับซ้อน หรือเปลี่ยนให้เป็นปัญหาที่คุ้นเคยหรือเคยแก้ปัญหามาก่อนหน้านี้

10. การให้เหตุผลทางตรรกศาสตร์

การให้เหตุผลทางตรรกศาสตร์ เป็นการอธิบายข้อความหรือข้อมูลที่ปรากฏอยู่ในปัญหานั้นว่าเป็นจริง โดยใช้เหตุผลทางตรรกศาสตร์มาช่วยในการแก้ปัญหา บางปัญหาเราใช้การให้เหตุผลทางตรรกศาสตร์ ร่วมกับการคาดเดาและตรวจสอบ หรือการเขียนภาพและแผนภาพ จนทำให้บางครั้งเราไม่สามารถแยกการให้เหตุผลทางตรรกศาสตร์ออกจากยุทธวิธีอื่นได้อย่างเด่นชัด ยุทธวิธีนี้มักใช้บ่อยในปัญหาทางเรขาคณิตและพีชคณิต

11. การให้เหตุผลทางอ้อม

การให้เหตุผลทางอ้อม เป็นการแสดงหรืออธิบายข้อความหรือข้อมูลที่ปรากฏอยู่ในปัญหานั้นว่าเป็นจริง โดยการสมมติว่าข้อความที่ต้องการแสดงนั้นเป็นเท็จ แล้วหาข้อขัดแย้ง ยุทธวิธีนี้มักใช้กับการแก้ปัญหาที่ยากแก่การแก้ปัญหาโดยตรง และง่ายที่จะหาข้อขัดแย้งเมื่อกำหนดให้ข้อความที่จะแสดงเป็นเท็จ

จากเอกสารและงานวิจัยที่เกี่ยวข้อง พบว่า กระบวนการแก้ปัญหาทางคณิตศาสตร์จำเป็นจะต้องมีขั้นตอนในการดำเนินการชัดเจนและยังมีวิธีการที่หลายหลาย ซึ่งการแก้ปัญหานี้ได้สอดแทรกอยู่ในทุกๆเนื้อหาของวิชาคณิตศาสตร์ไม่ว่าเรื่องใดก็ตาม สังเกตได้จากยุทธวิธีในการแก้ปัญหาซึ่งครอบคลุมเนื้อหาโดยรวมทั้งหมด

5. การศึกษาเกี่ยวกับแบบทดสอบ

5.1 แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน

ชวาล แพรัตกุล (2552: 74) ให้ความหมายว่า แบบทดสอบประเภทความสัมฤทธิ์ หมายถึงแบบทดสอบที่วัดความรู้ ทักษะ และสมรรถภาพสมองด้านต่างๆที่เด็กได้รับจากประสบการณ์ทั้งปวงทั้งจากทางโรงเรียนและทางบ้าน, ยกเว้น การวัดทางร่างกาย, ความถนัด, และทางบุคคล-สังคม อันได้แก่ อารมณ์ และการปรับตน เป็นต้น

ศิริชัย กาญจนวาสี (2556: 166) ได้กล่าวถึง ผลสัมฤทธิ์ว่า เป็นผลการเรียนรู้ตามแผนที่กำหนดไว้ล่วงหน้าอันเกิดจากกระบวนการเรียนการสอนในช่วงเวลาใดเวลาหนึ่งที่ผ่านมา แบบทดสอบผลสัมฤทธิ์จึงเป็นแบบทดสอบที่ใช้วัดผลการเรียนรู้ที่เกิดขึ้น (What person has learned) จากกิจกรรมการเรียนการสอนที่ผู้สอนได้จัดขึ้นเพื่อการเรียนรู้นั้น สิ่งที่มุ่งวัดจึงเป็นสิ่งที่ผู้เรียนได้เรียนรู้ภายใต้สถานการณ์ที่กำหนดขึ้น ซึ่งอาจเป็นความรู้หรือทักษะบางอย่าง อันบ่งบอกถึง สถานภาพของการเรียนรู้ที่ผ่านมา หรือสภาพการเรียนรู้ที่บุคคลนั้นได้รับ

มหาวิทยาลัยสุโขทัยธรรมาธิราช (2545: 219) ให้ความหมายว่า แบบทดสอบผลสัมฤทธิ์ทางการเรียน หมายถึง ชุดของคำถามที่มุ่งวัดความรู้ความสามารถ ทักษะและสมรรถภาพทางสมองด้านต่างๆ ของผู้เรียนหลังที่เกิดขึ้นการเรียนรู้

ข้อตกลงเบื้องต้นของการวัดผลสัมฤทธิ์

ศิริชัย กาญจนวาสี (2556: 167) ได้กล่าวว่า แบบทดสอบผลสัมฤทธิ์เป็นเครื่องมืออย่างหนึ่งที่ออกแบบไว้สำหรับวัดความรู้ หรือทักษะที่เกิดขึ้นกับผู้เรียนในช่วงเวลาหนึ่ง แต่เนื่องจากข้อจำกัดหลายประการ ทำให้ผู้สอนไม่สามารถทำการวัดความรู้และทักษะที่ได้สอนไปแล้วทั้งหมดได้ แบบทดสอบที่ใช้จึงเป็นเพียงกลุ่มตัวอย่างของเนื้อหาสาระส่วนหนึ่งที่ใช้เป็นตัวแทนมวลเนื้อหาหรือประสบการณ์ที่เป็นไปได้ทั้งหมด ดังนั้นการสร้างแบบทดสอบเพื่อใช้วัดผลสัมฤทธิ์ทางการเรียนของผู้เรียน จึงตั้งอยู่บนข้อตกลงเบื้องต้น 3 ประการ ดังนี้

1. มวลเนื้อเรื่อง หรือทักษะที่เป็นจุดมุ่งหมายของการเรียนการสอน สามารถนำมาระบุได้อย่างชัดเจน (Well defined) ในเชิงพฤติกรรมหรือกลุ่มพฤติกรรม

2. แบบทดสอบที่สร้างขึ้นได้ทำหน้าที่วัดพฤติกรรมสำคัญอันเป็นเป้าหมายของการเรียนรู้ นั่นคือ แบบทดสอบได้สร้างขึ้นมาอย่างมีความเที่ยงตรงตามเนื้อเรื่อง (Content Validity)

3. ผู้สอบได้มีโอกาสเรียนรู้ครอบคลุมสิ่งที่แบบทดสอบมุ่งวัด การวัดผลสัมฤทธิ์ของผู้เรียนจะนำมาแปลผลได้อย่างมีความหมาย เมื่อผู้เรียนได้มีโอกาสเรียนรู้ (Opportunity to learn) ครอบคลุมสิ่งที่แบบทดสอบมุ่งวัด

จากที่กล่าวมาสรุปได้ว่า แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน หมายถึง แบบทดสอบที่วัดความรู้ความสามารถทางการเรียน ประสบการณ์ที่มีอยู่ในสภาพปัจจุบันของแต่ละบุคคล และทักษะต่างๆ จากกิจกรรมการเรียนการสอนที่เกิดขึ้น ซึ่งในการสร้างก็มีข้อจำกัดหลายประการ ดังนั้นในการสร้างจึงต้องมีข้อตกลงบางประการเพื่อเป็นที่เข้าใจตรงกันและสร้างแบบทดสอบไปในทิศทางเดียวกัน

5.2 แบบทดสอบแบบเลือกตอบ

แบบทดสอบแบบเลือกตอบใช้วัดได้ครอบคลุมทั้งด้านความรู้ ความคิด หลักการ ทฤษฎี การตัดสินใจ การประเมินตัวแปร การแปลความหมายข้อมูล การแสดงความเข้าใจในธรรมชาติของคณิตศาสตร์ ตลอดจนความสามารถด้านทักษะกระบวนการทางคณิตศาสตร์ แบบทดสอบแบบเลือกตอบมีส่วนประกอบสำคัญ 2 ส่วน คือ 1) ส่วนของคำถาม และ 2) ส่วนของคำตอบ เรียกว่าตัวเลือก ซึ่งมีทั้งตัวเลือกที่เป็นคำตอบถูกและตัวเลือกที่เป็นคำตอบผิดเรียกว่าตัวลวง

ลักษณะของข้อสอบแบบเลือกตอบ

ดิลก ดิลกานนท์(2557: 77) กล่าวเกี่ยวกับลักษณะของข้อสอบแบบเลือกตอบ (Multiple choice item) ว่าแต่ละข้อประกอบด้วยสองส่วนที่สำคัญ คือ

1. ส่วนที่เป็นข้อคำถาม (Stem) ซึ่งเป็นส่วนที่ถามตามจุดประสงค์ที่ต้องการจะตรวจสอบว่าผู้สอบเกิดความรู้-ความเข้าใจในประเด็นที่ถามนั้นหรือไม่ ซึ่งข้อคำถามที่ดีจะต้องชัดเจน ไม่คลุมเครือทำให้ใครๆอ่านคำถามแล้วต้องเข้าใจประเด็นในการถามตรงกัน

2. ส่วนที่เป็นตัวเลือก (Choices or alternatives) หรือคำตอบ (Response) ของข้อคำถามข้างต้น โดยทั่วไปคำถามแต่ละข้อจะประกอบด้วยตัวคำตอบ 3-5 ตัวเลือก ให้ผู้สอบเลือกหาคำตอบที่ถูกต้องตรงกับประเด็นคำถามเพียงคำตอบเดียวเท่านั้น ดังนั้นในแต่ละชุดของคำตอบข้อหนึ่งๆนั้นจึงมีคำตอบที่ถูกต้องจริงๆ เพียงหนึ่งคำตอบเท่านั้น ส่วนที่เหลือจะเป็นตัวลวง ตัวลวงที่ดีในคำถามข้อหนึ่งๆนั้น ต้องมีความเป็นเอกพันธ์กับตัวคำตอบที่ถูก คือ ต้องไม่เด่นชัดจนเกินไปจนผู้สอบสามารถเดาได้

รูปแบบของข้อสอบเลือกตอบ

รูปแบบที่สำคัญๆ ของข้อสอบแบบเลือกตอบมี 4 รูปแบบ ดังนี้ (ดิลก ดิลกานนท์. 2557: 77-79)

1. ข้อสอบแบบคำถามโดด (Single question) เป็นข้อสอบที่แต่ละข้อจะถามอย่างเป็นอิสระในตัวเองไม่เกี่ยวข้องกับข้อคำถามข้ออื่น ซึ่งประกอบด้วยข้อคำถาม และตัวคำตอบหรือตัวเลือกข้อละ 1 ชุด ไม่ซ้ำกัน ข้อสอบแบบคำถามโดดนี้จัดเป็นรูปแบบพื้นฐานของข้อสอบแบบเลือกตอบ

2. ข้อสอบแบบตัวเลือกคงที่ (Constant choice) เป็นข้อสอบที่มีคำถามมากกว่าหนึ่งข้อใช้ตัวเลือกชุดเดียวกัน (เฉลยของแต่ละข้ออาจแตกต่างกัน)

3. ข้อสอบแบบสถานการณ์ (Situational item) ข้อสอบแบบนี้จะกำหนดสถานการณ์หรือเหตุการณ์ให้ก่อน แล้วจะถามในสาระหรือประเด็นต่างๆ จากสถานการณ์ที่กำหนดให้นั้น ซึ่งผู้ตอบจะต้องตอบโดยยึดสถานการณ์หรือเหตุการณ์ที่กำหนดนั้น

4. ข้อสอบแบบเลือกตอบสองชั้น (Double multiple choice) เป็นข้อสอบที่ดัดแปลงจากข้อสอบแบบเลือกตอบที่คำถามแต่ละข้อมีคำตอบถูกหรือเฉลยมากกว่าหนึ่งคำตอบ โดยจะให้ผู้ตอบพิจารณาว่า จากตัวเลือกที่กำหนดให้นั้นมีตัวเลือกใดบ้างที่ถูกต้อง

การสร้างแบบทดสอบแบบเลือกตอบให้มีคุณภาพ มีหลักการดังนี้

1. การสร้างคำถาม คำถามที่ดีควรมีลักษณะดังต่อไปนี้

1.1 สั้น ชัดเจน และใช้ภาษาที่เข้าใจง่าย

1.2 เขียนเป็นประโยคบอกเล่า ถ้าจำเป็นต้องใช้ประโยคปฏิเสธก็ควรเน้นข้อความหรือขีดเส้นใต้ข้อความที่แสดงการปฏิเสธ

1.3 คำถามแต่ละข้อต้องเป็นอิสระแก่กัน โดยไม่ให้การตอบคำถามของข้อหนึ่งชี้นำหรือขึ้นอยู่กับอีกข้อหนึ่ง

1.4 หลีกเลี่ยงการใช้ภาษาที่ชี้นำหรือสื่อความไปถึงคำตอบถูกหรือคำตอบผิด

1.5 แต่ละคำถามต้องมีคำตอบที่ถูกต้องเพียงคำตอบเดียว

2. การสร้างตัวเลือก ตัวเลือกที่ดีควรมีลักษณะดังต่อไปนี้

2.1 ตัวเลือกควรเป็นเรื่องหรือประเด็นเดียวกัน มีความยาวใกล้เคียงกัน

2.2 ต้องกระจายคำตอบถูกของแบบทดสอบทั้งฉบับ ให้มีสัดส่วนของแต่ละตัวเลือกใกล้เคียงกัน

2.3 ใช้คำให้สั้นที่สุดเท่าที่จะทำได้และหลีกเลี่ยงการใช้คำศัพท์หรือข้อความที่เข้าใจยาก

2.4 ไม่ควรใช้ตัวเลือก "ถูกทุกข้อ" หรือ "ไม่มีข้อใดถูก" (อาจเป็นการสื่อความหมายไม่แน่ใจในคำถามหรือการเลือกตอบด้วยความไม่มั่นใจก็ได้)

เกณฑ์การให้คะแนนแบบทดสอบแบบเลือกตอบ

การให้คะแนนแบบทดสอบแบบเลือกตอบ โดยส่วนใหญ่พิจารณาจากความถูกผิดของการเลือกคำตอบเป็นส่วนใหญ่ เช่น ตอบถูกได้ 1 คะแนน และตอบผิดได้ 0 คะแนน

จุดเด่น-จุดด้อยของข้อสอบแบบเลือกตอบ

ข้อสอบแบบเลือกตอบจัดเป็นเครื่องมือวัดผลชนิดหนึ่งที่ค่อนข้างนิยมใช้กันมาก เพราะมีจุดเด่นหรือข้อได้เปรียบกว่าเครื่องมือชนิดอื่นในหลาย ๆประเด็น แต่ก็ต้องไม่ลืมว่าข้อสอบแบบเลือกตอบก็มีข้อจำกัดอยู่หลายประการเช่นกัน โดยเฉพาะอย่างยิ่งถ้ามีการสร้างและใช้อย่างขาดความระมัดระวังแล้วอาจทำให้ผลที่ได้มีความคลาดเคลื่อนสูงกว่าเครื่องมือวัดชนิดอื่นก็เป็นได้

จุดเด่นของข้อสอบแบบเลือกตอบ ได้แก่

1. สามารถถามได้มากข้อในเวลาจำกัด ทำให้ครอบคลุมจุดประสงค์และเนื้อหาที่ต้องการวัดได้

2. ข้อสอบแบบเลือกตอบมักมีความเป็นปรนัยสูง ถ้าสร้างอย่างระมัดระวัง โดยเฉพาะการตรวจให้คะแนน ซึ่งสามารถตรวจตามเฉลยที่กำหนดได้อย่างมีความเชื่อมั่นและยุติธรรม

3. ใช้เวลาในการตรวจให้คะแนนไม่มาก และสามารถใช้สอบวัดกับคนจำนวนมาก ๆในเวลาเดียวกันได้

4. สามารถนำมาวิเคราะห์หาคุณภาพเชิงประจักษ์ได้ง่าย จึงทำให้เป็นข้อสอบมาตรฐานได้ดีกว่าข้อสอบแบบอื่น

จุดด้อยของข้อสอบแบบเลือกตอบ ได้แก่

1. ต้องใช้เวลาในการสร้างมากและสร้างข้อคำถามที่ดี ๆ ได้ยาก ข้อสอบแบบเลือกตอบโดยทั่วไปถ้าไม่ระมัดระวังในการสร้างข้อคำถามแล้วมักจะได้ข้อคำถามที่มีคุณภาพต่ำ คือ คำถามมักถามในพฤติกรรมความจำและตัวเลือกมีโอกาสเดาคำตอบถูกได้สูง

2. มีข้อจำกัดในการถามวัดพฤติกรรมระดับสูง เช่น การวิเคราะห์ สังเคราะห์ และประเมินค่า ทั้งนี้เพราะพฤติกรรมดังกล่าวผู้สอบต้องเป็นผู้สร้างคำตอบขึ้นเอง ไม่ควรมีคำตอบให้เลือกซึ่งเป็นการชี้นำและสะกัดกั้นความคิดของผู้เข้าสอบ

จากที่ได้กล่าวมาสรุปได้ว่า ส่วนใหญ่แบบทดสอบแบบเลือกตอบจะถูกนำมาใช้ในการทดสอบทางการศึกษา และจะมีวิธีการให้คะแนนแบบ 2 ค่า คือตอบถูกได้ 1 คะแนน ตอบผิดได้ 0 คะแนน ซึ่งวิธีการให้คะแนนแบบนี้ ยังไม่สามารถให้สารสนเทศที่ชัดเจนได้ ถ้าไม่ระวังเกี่ยวกับการสร้างข้อคำถามแล้วมักจะทำให้ข้อคำถามนั้นมีคุณภาพต่ำได้ และยังเปิดโอกาสให้มีการเดาที่สูงอีกด้วย

6. งานวิจัยที่เกี่ยวข้อง

6.1 งานวิจัยในประเทศ

รัตนา ไชยตรี (2546: 116-128) ได้ศึกษาเปรียบเทียบคุณภาพของวิธีการตอบและการตรวจให้คะแนนความรู้บางส่วนที่ประยุกต์วิธีของคูมบ์ ประยุกต์วิธีของเดรสเซลและชมิดท์ และวิธีการตอบโดยบอกระดับความมั่นใจ มีวัตถุประสงค์เพื่อวิเคราะห์และเปรียบเทียบคุณภาพ ด้านความเที่ยงตรง ความเชื่อมั่น ฟังก์ชันสารสนเทศและประสิทธิภาพสัมพัทธ์เฉลี่ยของแบบทดสอบ และเพื่อเปรียบเทียบอัตราความคลาดเคลื่อนในการกำหนดเกรดแบบอิงกลุ่มและอิงเกณฑ์ ของคะแนนที่ได้จากวิธีการตอบและการตรวจให้คะแนนความรู้บางส่วนที่ประยุกต์วิธีของคูมบ์ ประยุกต์วิธีของเดรสเซลและชมิดท์ และวิธีการตอบโดยบอกระดับความมั่นใจ กลุ่มตัวอย่าง ได้แก่ นักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2548 ของโรงเรียนมัธยมศึกษา สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ในกรุงเทพมหานคร จำนวน 8 โรงเรียน รวมทั้งสิ้น 1,005 คน เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์พื้นฐานชั้นมัธยมศึกษาปีที่ 3 โดยนำข้อมูลมาวิเคราะห์ความเที่ยงตรงตามสภาพ ความเที่ยงตรงแบบความสอดคล้องภายใน ค่าอำนาจจำแนก ค่าความยากง่าย ฟังก์ชันสารสนเทศ ประสิทธิภาพสัมพัทธ์เฉลี่ยของแบบทดสอบ และอัตราความคลาดเคลื่อนของการกำหนดเกรด ใช้โปรแกรมวิเคราะห์ ได้แก่ PASCALE และ SPSS ผลการวิจัยพบว่า วิธีการตอบและตรวจให้คะแนนความรู้บางส่วนตามวิธีการตอบโดยบอกระดับความมั่นใจ ส่วนใหญ่จะให้ค่าฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบทดสอบ และอัตราส่วนสารสนเทศสูงสุด รองลงมาคือ วิธีการตอบและตรวจให้คะแนนวิธีที่ประยุกต์การให้คะแนนของเดรสเซลและชมิดท์ และวิธีที่ประยุกต์การให้คะแนนของคูมบ์ตามลำดับ

สุพจน์ เกิดสุวรรณ์ (2545: 147-161) ได้ทำการเปรียบเทียบคุณภาพวิธีการตอบและตรวจให้คะแนนตามวิธีของคูมบ์ วิธีของอาร์โนลด์ วิธีของเดรสเซลและชมิท วิธีที่ประยุกต์จากวิธีของคูมบ์ วิธีที่ประยุกต์จากวิธีของอาร์โนลด์ วิธีที่ประยุกต์จากวิธีของเดรสเซลและ ชมิท และวิธีการตรวจให้คะแนนแบบประเพณีนิยม มีวัตถุประสงค์เพื่อพัฒนาวิธีการวัดความรู้บางส่วนของผู้ตอบแบบทดสอบเลือกตอบ และเพื่อเปรียบเทียบคุณภาพวิธีการตอบและตรวจให้คะแนนทั้ง 7 วิธีข้างต้น

ในด้านความเที่ยงตรงตามเกณฑ์ ความเที่ยงตรงเชิงโครงสร้าง ความเชื่อมั่นแบบความสอดคล้องภายใน ค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบทดสอบ และค่าประสิทธิภาพสัมพัทธ์เฉลี่ยของแบบทดสอบ กลุ่มตัวอย่างได้แก่ นักเรียนระดับมัธยมศึกษาปีที่ 3 จากโรงเรียนมัธยมศึกษาในจังหวัดสระบุรี ปีการศึกษา 2544 จำนวน 15 โรงเรียน รวมทั้งสิ้น 946 คน เครื่องมือที่ใช้ในการวิจัยเป็นแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ขั้นพื้นฐานระดับมัธยมศึกษาตอนต้น ข้อมูลที่ได้นำมาตรวจสอบความเที่ยงตรงเชิงโครงสร้าง โดยการวิเคราะห์เมตริกพหุลักษณะพหุวิธี และวิเคราะห์ตัวประกอบเชิงยืนยันด้วยโปรแกรม LISREL ตรวจสอบความเที่ยงตรงตามเกณฑ์สัมพันธ์จากการวิเคราะห์สหสัมพันธ์ของคะแนนกับเกณฑ์ วิเคราะห์ค่าความเชื่อมั่นแบบความสอดคล้องภายใน และวิเคราะห์ค่าฟังก์ชันสารสนเทศของแบบทดสอบ ด้วยโปรแกรม BILOG และ PARSCALE ผลการวิจัยพบว่า วิธีที่ประยุกต์จากวิธีของคูมบ์ ให้ค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบทดสอบและประสิทธิภาพสัมพัทธ์เฉลี่ยของแบบทดสอบสูงกว่าวิธีอื่นๆ รองลงมาคือวิธีของ อาร์โนลด์ และวิธีที่ประยุกต์จากวิธีของเดรสเซลและชมิด ตามลำดับ ส่วนวิธีประเพณีนิยมจะให้ค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบทดสอบและประสิทธิภาพสัมพัทธ์เฉลี่ยของแบบทดสอบต่ำกว่าทุกวิธี

เอมอร จังศิริพรปกรณ์ (2545: 77-85) ได้ทำการเปรียบเทียบคุณภาพของแบบทดสอบเลือกตอบเมื่อตรวจด้วยวิธีการให้คะแนนความรู้บางส่วนกับวิธีประเพณีนิยม มีวัตถุประสงค์เพื่อเปรียบเทียบคุณภาพของแบบทดสอบในด้านความเที่ยงตรงตามสภาพ อำนาจจำแนก ความยาก ฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบทดสอบและอัตราส่วนสารสนเทศเฉลี่ยโดยใช้ทฤษฎีการตอบสนองข้อสอบเมื่อแบบทดสอบตรวจให้คะแนนแบบประเพณีนิยมกับวิธีให้คะแนนความรู้บางส่วน ซึ่งประกอบด้วยวิธีประยุกต์การให้คะแนนของคูมบ์ และวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท กลุ่มตัวอย่างได้แก่ นิสิตคณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ชั้นปีที่ 3 ที่ลงทะเบียนเรียนวิชาการประเมินผลการเรียนการสอน ในภาคการศึกษาต้น ปีที่ศึกษา 2544 จำนวน 297 คน เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบที่ใช้เป็นแบบวัดผลสัมฤทธิ์ทางการเรียนวิชาการประเมินผลการเรียนการสอน ที่มีวิธีการตรวจให้คะแนน 3 วิธีดังกล่าวจากแบบทดสอบชุดเดียวกัน วิเคราะห์ข้อมูลโดยใช้โปรแกรม BILOG สำหรับการให้คะแนนแบบประเพณีนิยม และใช้โปรแกรม PARSCALE สำหรับการให้คะแนนด้วยวิธีประยุกต์การให้คะแนนของคูมบ์และวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท ผลการวิจัยเมื่อเปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบทดสอบ และอัตราส่วนสารสนเทศระหว่างวิธีการตรวจให้คะแนนที่แตกต่างกัน พบว่าวิธีประยุกต์การให้คะแนนของคูมบ์ มีค่าฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบทดสอบ และอัตราส่วนสารสนเทศเฉลี่ยสูงที่สุด รองลงมาคือวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท ส่วนวิธีประเพณีนิยมมีค่าต่ำที่สุด

6.2 งานวิจัยต่างประเทศ

David A. Bradbard, Darrell F. Parker and Gary L. Stone (2004: 11-26) ได้ทำการเปรียบเทียบวิธีการตรวจให้คะแนนระหว่างวิธีการตัดตัวเลือกแบบคูมบ์กับวิธีเลือกตอบแบบธรรมดาทั่วไป เครื่องมือที่ใช้เป็นแบบทดสอบปรนัยวิชาเศรษฐศาสตร์มหภาค ใช้ระยะเวลาในการเก็บข้อมูล 1 ภาคเรียน ซึ่งมีการทดสอบทั้งหมด 4 ครั้ง กลุ่มตัวอย่างเป็นนักศึกษาจำนวน 44 คน แต่มีเพียง 35 คน ที่ทำแบบทดสอบครบทั้ง 4 ครั้ง ผลการวิจัยพบว่า เมื่อทำการทดลองครบทั้ง 4 ครั้ง คะแนนของนักศึกษาเมื่อคิดเป็นเปอร์เซ็นต์แล้ววิธีการตรวจให้คะแนนด้วยวิธีการตัดตัวเลือกแบบคูมบ์สูงกว่าวิธีการเลือกตอบแบบธรรมดาทั่วไป

Hee-Sun Lee, Ou Lydia Liu and Marcia C. Linn (2011: 115-136) ได้ทำเปรียบเทียบการตอบสนองต่อข้อสอบ โดยใช้ข้อสอบที่มีทั้งปรนัยและอัตนัยอยู่ในฉบับเดียวกันในสาขาวิชาวิทยาศาสตร์ซึ่งจะออกครอบคลุมทั้ง 6 รายวิชา คือ 1. ชีววิทยา ม.ต้น 2. ฟิสิกส์ ม.ต้น 3. โลกศาสตร์ ม.ต้น 4. ชีววิทยา ม.ปลาย 5. ฟิสิกส์ ม.ปลาย และ 6. เคมี ม.ปลาย ในการทดลองครั้งนี้ใช้ Rasch Partial Credit Model ในปีแรกทำกับนักเรียนที่เรียนในโรงเรียน ม.ต้น และ ม.ปลาย จำนวน 3,716 ที่สอนโดยครู 51 คน จาก 16 โรงเรียนในรัฐอริโซน่านอร์ทคาโรไลน่า เวอจิเนียร์ เมซซาซูเซส และแคลิฟอเนีย ในปีที่สอง ได้ทำกับนักเรียน 4,507 คน สอนโดยครู 50 คน ที่มาจาก 13 โรงเรียนโดยสลับข้ามกันมาจาก 5 รัฐ การให้คะแนนจะให้ตามแบบของ Wilson's (2005) ส่วนอัตนัยใช้เป็นรูบิค

ผลการวิจัยพบว่า

1. มีเพียง 33% ของผู้ตอบข้อสอบแบบปรนัยถูกมีระดับความสามารถอยู่สูงกว่าระดับการบูรณาการความรู้ ซึ่งอีกประมาณเกือบ 75% ไม่สามารถบูรณาการได้
2. ข้อสอบอัตนัยสามารถแยกความสามารถของนักเรียนในการบูรณาการความรู้ได้มีประสิทธิภาพกว่า
3. ข้อสอบอัตนัยสามารถวัดช่วงความกว้างของระดับการบูรณาการความรู้ได้ดีกว่าแบบปรนัย
4. ข้อสอบอัตนัยกระตุ้นการบูรณาการความรู้ได้ดีกว่าแบบปรนัย

Yuan H. Li and Robert W. Lissitz (2004: 85-117) ได้ทำการเปรียบเทียบวิธีการหาความคลาดเคลื่อน (SE) ของ maximum likelihood ระหว่างวิธี $AEASE_s$ กับ $EMBSE_s$ ผลการวิจัยพบว่า ค่าความคลาดเคลื่อน (SE) ทั้งสองวิธีให้ค่าใกล้เคียงกัน ยิ่งถ้าหากใช้กับ G-PCM ค่าที่ได้จะยิ่งใกล้เคียงกันมากเข้าไปอีก และวิธี $EMBSE_s$ สามารถประมาณค่าความคลาดเคลื่อน (SE) ได้แม่นยำมากขึ้น เมื่อคำตอบเพิ่มมากขึ้น แต่มีข้อเสียคือใช้เวลานานในการคำนวณ

Tsung-Han Ho and Barbara G. Dodd (2012: 305-326)ได้ทำการเปรียบเทียบกระบวนการเลือกข้อสอบ 5 แบบ คือ 1. MI 2. MPWI 3. MPWKL 4. MEI 5. MEPWKL โดยใช้ 3

วิธีในการเลือกคือ 1. MLE 2. WLE 3. EAP ซึ่งเป็นพื้นฐานของ G-PCM ในการวิจัยครั้งนี้ได้ทำการจำลองข้อมูลเพื่อใช้ในการประมาณค่าความสามารถ และใช้โปรแกรมคอมพิวเตอร์ในการวิเคราะห์ข้อมูลคือ IRTGEN ผลการวิจัยพบว่า ข้อสอบทั้ง 5 แบบ ใช้ได้เหมือนกันหมด แต่การใช้วิธีการเลือกแบบ MLE ดีที่สุด และ WLE สามารถใช้ได้ดีเมื่อมีการกระจายของประชากร และEAP จะใช้ได้ดีเมื่อมีข้อมูลที่เจาะจง

จากเอกสารและงานวิจัยที่เกี่ยวข้องกับการตอบและตรวจให้คะแนนความรู้บางส่วนทั้งในและต่างประเทศ ทำให้ทราบว่ายังไม่สามารถสรุปได้ว่าวิธีการใดเป็นวิธีการที่ดีที่สุดและมีคุณภาพที่สุด ซึ่งทุกวิธีมีส่วนดีและไม่ดี และยังมีการศึกษาเพื่อพัฒนาวิธีการตอบและตรวจให้คะแนนความรู้บางส่วนอย่างต่อเนื่องเพื่อประโยชน์แก่การนำไปใช้ ดังนั้นจำเป็นจะต้องเลือกใช้ให้เหมาะสมกับลักษณะของผู้สอบและวิชาที่สอบนั้นๆ

บทที่ 3
วิธีการดำเนินงานวิจัย

ในการวิจัยครั้งนี้ ผู้วิจัยได้ดำเนินการตามขั้นตอนดังนี้

1. การกำหนดประชากรและการสุ่มกลุ่มตัวอย่าง
2. การสร้างเครื่องมือที่ใช้ในการวิจัย
3. การเก็บรวบรวมข้อมูล
4. การจัดกระทำและการวิเคราะห์ข้อมูล

1. การกำหนดประชากรและการสุ่มกลุ่มตัวอย่าง

ประชากร

ประชากรที่ใช้ในการวิจัยครั้งนี้ เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2559 สำนักงานเขตพื้นที่การศึกษามัธยมศึกษาเขต 2 จังหวัดกรุงเทพมหานคร จำนวน 52 โรงเรียน จำนวน 548 ห้องเรียน นักเรียน 23,056 คน (สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน. 2559)

การเลือกกลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2559 สำนักงานเขตพื้นที่การศึกษามัธยมศึกษาเขต 2 จังหวัดกรุงเทพมหานคร จำนวน 1,200 คน ซึ่งแต่ละวิธีการจะมีนักเรียนทำแบบทดสอบในจำนวนที่เท่าๆกัน นั่นคือ วิธีการละ 600 คน ได้มาโดยการสุ่มแบบหลายขั้นตอน (Multi-stage random sampling) เนื่องจากในการวิเคราะห์ข้อมูลตามทฤษฎีตอบสนองข้อสอบ (Item Response Model) ควรกำหนดขนาดกลุ่มตัวอย่างให้มีขนาดใหญ่พอที่จะทำให้ความคลาดเคลื่อนมาตรฐานของการประมาณค่าพารามิเตอร์มีขนาดเล็กถึงระดับที่ยอมรับได้ตามเป้าหมายของการนำผลไปใช้ในทางปฏิบัติ ซึ่งควรใช้ขนาดกลุ่มตัวอย่างไม่ต่ำกว่า 500 คน (ศิริชัย กาญจนวาสี. 2555: 113)

การสุ่มกลุ่มตัวอย่างมีขั้นตอน ดังนี้

1. สำรวจข้อมูลหน่วยสมาชิกของประชากรชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2559 จากสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 2 จังหวัดกรุงเทพมหานคร

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

2.1 จำแนกโรงเรียนในเขตพื้นที่การศึกษามัธยมศึกษาเขต 2 เป็น 3 ขนาด (ขนาดใหญ่พิเศษ, ขนาดใหญ่ และขนาดกลาง) จากการสำรวจพบว่า

2.1.1 ขนาดใหญ่พิเศษ พิจารณาจากจำนวนนักเรียนทั้งหมดมากกว่า 2,500 คน จำนวน 25 โรงเรียน จำนวนนักเรียน 14,311 คน

2.1.2 ขนาดใหญ่ พิจารณาจากจำนวนนักเรียนทั้งหมด 1,501 -2,500 คน จำนวน 20 โรงเรียน จำนวนนักเรียน 7,588 คน

2.1.3 ขนาดกลาง พิจารณาจากจำนวนนักเรียนทั้งหมด 501 -1,500 คน จำนวน 6 โรงเรียน จำนวนนักเรียน 1,094 คน

2.2 สุ่มโรงเรียนตามสัดส่วนของโรงเรียนแต่ละขนาด ได้โรงเรียนขนาดใหญ่พิเศษ 5 โรงเรียน ได้แก่ บดินทรเดชา(สิงห์ สิงหเสนี), เทพศิรินทร์ร่มเกล้า, บางกะปิ, พรตพิทยพยัต และมัธยมวัดหนองจอก จำนวน 68 ห้องเรียน จำนวน 3,070 คน โรงเรียนขนาดใหญ่ 3 โรงเรียน ได้แก่ รัตนโกสินทร์สมโภชบางเขน, บดินทรเดชา (สิงห์ สิงหเสนี) 4 และลาดปลาเค้าพิทยาคม จำนวน 28 ห้องเรียน จำนวน 1,200 คน โรงเรียนขนาดกลาง 2 โรงเรียน ได้แก่ สุขุมนวพันธ์อุปถัมภ์ และจันทร์หุ่นบำเพ็ญ จำนวน 12 ห้องเรียน จำนวน 432 คน

3. สุ่มแบบแบ่งกลุ่ม โดยแบ่งห้องเรียนของโรงเรียนแต่ละขนาด ดังนี้ โรงเรียนขนาดใหญ่พิเศษ เป็น 68 ห้องเรียน สุ่มได้มา 16 ห้องเรียน โรงเรียนขนาดใหญ่ จำนวน 28 ห้องเรียน สุ่มได้มา 6 ห้องเรียน โรงเรียนขนาดกลาง จำนวน 12 ห้องเรียน สุ่มได้มา 4 ห้องเรียน รวมทั้งสิ้น จำนวน 26 ห้องเรียน จำนวน 1,200 คน

ตาราง 3 ข้อมูลการสุ่มห้องเรียนแต่ละโรงเรียนของประชากร ชั้นมัธยมศึกษาปีที่ 1

โรงเรียน	จำนวนห้องเรียน	จำนวนนักเรียน
โรงเรียนขนาดใหญ่พิเศษ		
1. บดินทรเดชา(สิงห์ สิงหเสนี)	3	153
2. เทพศิรินทร์ร่มเกล้า	4	201
3. บางกะปิ	3	151
4. พรตพิทยพยัต	3	148
5. มัธยมวัดหนองจอก	3	145
รวม	16	798
โรงเรียนขนาดใหญ่		
1. รัตนโกสินทร์สมโภชบางเขน	2	90
2. บดินทรเดชา(สิงห์ สิงหเสนี) 4	2	87
3. ลาดปลาเค้าพิทยาคม	2	86
รวม	6	263
โรงเรียนขนาดกลาง		
1. สุขุมนวพันธ์อุปถัมภ์	2	76
2. จันทร์หุ่นบำเพ็ญ	2	63
รวม	4	139
รวมทั้งสิ้น	26	1,200

2. การสร้างเครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยครั้งนี้ คือ แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ เรื่อง ตัวหารร่วมมากและตัวคูณร่วมน้อย ชั้นมัธยมศึกษาปีที่ 1 ที่ผู้วิจัยสร้างขึ้น มีวิธีการตรวจให้คะแนน 2 วิธี คือ วิธีประยุกต์การให้คะแนนของคูมบ์ (Modified Coombs Approach) และวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ (The Arnold Scoring/Response Method) มีลักษณะเป็นข้อสอบเลือกตอบ 4 ตัวเลือก จำนวน 25 ข้อ

ตัวอย่างแบบทดสอบเมื่อตรวจด้วย

วิธีประยุกต์การให้คะแนนของคูมบ์และวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์

คำชี้แจง:

1. แบบทดสอบฉบับนี้มีคำถามทั้งหมด 25 ข้อ ให้เวลาทำ 60 นาที
2. แบบทดสอบฉบับนี้เป็นแบบเลือกตอบ ชนิด 4 ตัวเลือก มีตัวเลือกถูกเพียง 1 ตัวเลือก

ให้นักเรียนเลือกตัวเลือกที่ผิด โดยทำเครื่องหมาย x ลงในกระดาษคำตอบที่แนบมาพร้อมนี้ ดังตัวอย่างข้อ (0)

ร้านขายของแห่งหนึ่งรับนมจากผู้ผลิตเป็นงวดๆ ดังนี้

รับนมรสช็อกโกแลตทุก 2 วัน รับนมรสสตอเบอรี่ทุก 3 วัน และรับนมรสวานิลาทุก 4 วัน โดยผู้ผลิตจะมาส่งนมใหม่จะรับนมเก่าที่เหลือกลับ แอ๋วไปซื้อนมที่ร้านนี้ในวันที่ 1 สิงหาคม ซึ่งตรงกับวันที่ร้านรับนมทั้งสามรสพร้อมกันพอดี อยากทราบว่าแอ๋วควรไปซื้อนมที่ร้านนี้ครั้งต่อไปเมื่อใดจึงจะได้นมที่มาส่งใหม่ทั้งสามรส

ก. วันที่ 6

ข. วันที่ 8

ค. วันที่ 12

ง. วันที่ 13

วิธีการตอบและการตรวจให้คะแนน มีรายละเอียดดังนี้

1. วิธีประยุกต์การให้คะแนนของคูมบ์ (Modified Coombs Approach)

วิธีการตอบ คือ ให้ผู้ตอบทำเครื่องหมาย (×) ตรงตัวเลือกที่เห็นว่าผิดทุกตัวเลือกถ้าไม่แน่ใจให้เว้นข้ามไปไม่ต้องทำ

วิธีการตรวจให้คะแนน อธิบายได้ด้วยสูตรดังนี้ (สุพจน์ เกิดสุวรรณ์. 2545: 62)

$$X=\frac{1}{2}\left[n+(n)c\right]$$

เมื่อ X คือคะแนนที่ผู้สอบจะได้จากการตอบข้อสอบข้อหนึ่ง

n คือจำนวนตัวเลือกที่ผู้สอบเลือกในการตอบข้อสอบแต่ละข้อ

c คือค่ากำหนดเงื่อนไข มี 2 ค่า คือ

$c=1$ เมื่อผู้ตอบตัดตัวเลือกที่เป็นคำตอบผิดออกได้ถูกต้อง

$c=-1$ เมื่อผู้ตอบตัดตัวเลือกที่เป็นคำตอบถูกออกไป

ตาราง 4 การตรวจให้คะแนนโดยใช้วิธีประยุกต์การให้คะแนนของคูมบ์โดยจำนวน ตัวเลือกมี 4 ตัวเลือก

จำนวนคำตอบที่ผู้สอบเลือก	ตัวเลือกที่ถูกตัด	จำนวนตัวลวงที่ถูกเลือก	คะแนน
3	FFF	ตัดตัวลวงได้ทั้งหมด	3
2	FF	ตัดตัวลวงได้ 2 ตัว	2
1	F	ตัดตัวลวงได้ 1 ตัว	1
1	T	ตัดตัวลวงไม่ตรงกับตัวลวงข้อนั้น	0
2	TF	ตัดตัวลวงได้ 1 ตัว	0
3	TFF	ตัดตัวลวงได้ 2 ตัว	0
4	TFFF	ตัดตัวเลือกทั้งหมด	0

หมายเหตุ: T คือ ตัวเลือกถูก และ F คือ ตัวลวง

2. วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ (The Arnold Scoring/Response Method)

วิธีการตอบ คือ ให้ผู้ตอบทำเครื่องหมาย (×) ตรงตัวเลือกที่เห็นว่าผิดทุกตัวเลือกถ้าไม่แน่ใจให้เว้นข้ามไปไม่ต้องทำ

วิธีการตรวจให้คะแนน อธิบายได้ด้วยสูตรดังนี้ (สุพจน์ เกิดสุวรรณ์. 2545: 69-70)

$$C_d = (p)\left[d/(k-d)\right]$$

เมื่อ C_d คือคะแนนที่ผู้สอบได้เมื่อสามารถตัดตัวเลือกที่เป็นคำตอบที่ผิดออกไปได้โดยไม่มีคะแนนที่เพิ่มจากการเดา

d คือจำนวนตัวเลือกที่เป็นคำตอบผิดที่ผู้สอบตัดออกไปได้ถูกต้อง

k คือจำนวนตัวเลือกในข้อสอบ

p คือ $1/(k-1)$

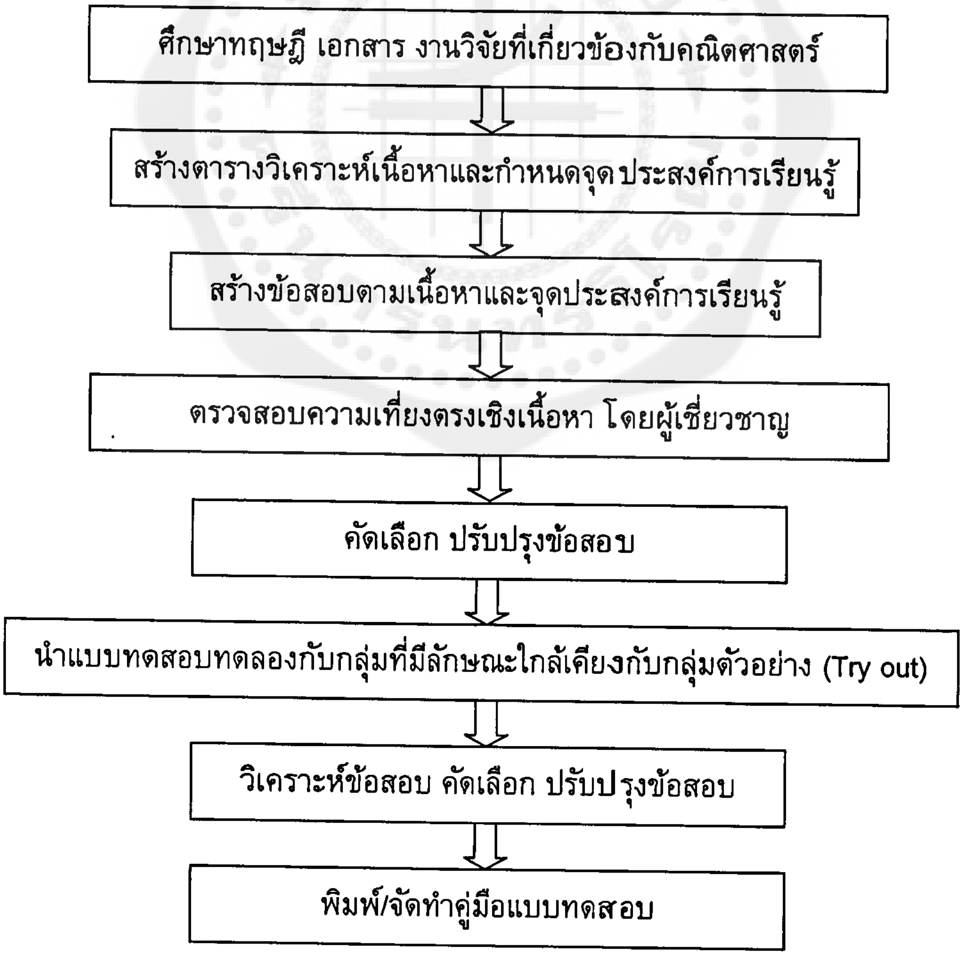
ส่วนในกรณีที่ผู้สอบตัดตัวเลือกที่ถูกเข้าไว้เป็นตัวลวงจะได้คะแนนลงโทษคือ $-p$ ซึ่งไม่ต้องแทนค่าในสูตรข้างต้นอีกไม่ว่าผู้สอบจะตัดตัวเลือกออกไปกี่ตัวก็ตาม

ตาราง 5 การตรวจให้คะแนนโดยใช้วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ โดยจำนวนตัวเลือกมี 4 ตัวเลือก

จำนวนคำตอบที่ผู้สอบเลือก	ตัวเลือกที่ถูกตัด	จำนวนตัวลวงที่ถูกเลือก	คะแนน
3	FFF	ตัดตัวลวงได้ทั้งหมด	1
2	FF	ตัดตัวลวงได้ 2 ตัว	1/3
1	F	ตัดตัวลวงได้ 1 ตัว	1/9
1	T	ตัดตัวลวงไม่ตรงกับตัวลวงข้อนั้น	-1/3
2	TF	ตัดตัวลวงได้ 1 ตัว	-1/3
3	TFF	ตัดตัวลวงได้ 2 ตัว	-1/3

หมายเหตุ: T คือ ตัวเลือกถูก และ F คือ ตัวลวง

การดำเนินการสร้างเครื่องมือวัดแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1 มีขั้นตอนการสร้างดังนี้



ภาพประกอบ 5 ขั้นตอนในการสร้างแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์

ในการสร้างแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 อาศัยเอกสารหลักสูตรและการวิเคราะห์หลักสูตร เพื่อที่จะระบุจุดมุ่งหมายในการชี้แนวทางการเลือกตัวแทนของเนื้อหาที่จะออกข้อสอบว่าออกข้อสอบอย่างไรจึงจะเป็นตัวแทนที่ดีของเนื้อหา วิธีการนี้จะเป็นหลักประกันได้ว่าแบบทดสอบนั้นจะมีความเที่ยงตรงด้านเนื้อหา (Content validity) มีการดำเนินการสร้างแบบทดสอบดังนี้

1. ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับวิชาคณิตศาสตร์และการสร้างแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์

2. สร้างตารางวิเคราะห์เนื้อหา และจุดประสงค์การเรียนรู้เกี่ยวกับเรื่อง ตัวหารร่วมมากและตัวคูณร่วมน้อย ชั้นมัธยมศึกษาปีที่ 1

3. สร้างแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ เรื่อง ตัวหารร่วมมากและตัวคูณร่วมน้อย ชนิดเลือกตอบ 4 ตัวเลือก จำนวน 50 ข้อ

4. ตรวจสอบคุณภาพแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนเรื่อง ตัวหารร่วมมากและตัวคูณร่วมน้อยของนักเรียนชั้นมัธยมศึกษาปีที่ 1 ด้านความเที่ยงตรงเชิงเนื้อหา โดยนำแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน ที่ผู้วิจัยสร้างขึ้น ให้ผู้เชี่ยวชาญที่มีความรู้ทางการวัดผลและคณิตศาสตร์ จำนวน 5 ท่าน พิจารณาแบบทดสอบ และนำแบบทดสอบไปตรวจสอบหาความเที่ยงตรงของแบบทดสอบโดยใช้สูตร IOC และคัดเลือกข้อสอบจำนวน 40 ข้อ ผลปรากฏว่า ค่า IOC ที่ได้อยู่ในช่วง 0.6 ถึง 1

ดัชนีความสอดคล้องระหว่างข้อสอบกับจุดประสงค์รายข้อของผู้เชี่ยวชาญ (ล้วน สายยศและอังคณา สายยศ. 2543: 249)

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

เมื่อ IOC แทน ดัชนีความสอดคล้องระหว่างข้อสอบกับจุดประสงค์

$\sum R$ แทน ผลรวมคะแนนความคิดเห็นของผู้เชี่ยวชาญทั้งหมด

N แทน จำนวนผู้เชี่ยวชาญทั้งหมด

คะแนนค่าดัชนีความสอดคล้องระหว่างข้อสอบกับจุดประสงค์รายข้อของผู้เชี่ยวชาญ ดังนี้

+1 หมายถึง แน่ใจว่าข้อสอบวัดได้ตรงกับจุดประสงค์ที่กำหนด

0 หมายถึง ไม่แน่ใจว่าข้อสอบวัดได้ตรงกับจุดประสงค์ที่กำหนด

-1 หมายถึง แน่ใจว่าข้อสอบวัดได้ไม่ตรงกับจุดประสงค์ที่กำหนด

5. นำแบบทดสอบที่ปรับปรุงไปทดลองใช้กับนักเรียนชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2559 สำนักงานเขตพื้นที่การศึกษามัธยมศึกษาเขต 2 จังหวัดกรุงเทพมหานคร ที่มี

ลักษณะใกล้เคียงกับกลุ่มตัวอย่าง จำนวน 200 คน โดยแบ่งเป็นวิธีประยุกต์การให้คะแนนของคูมบ์ 100 คน เป็น ชาย 50 คน หญิง 50 คน และวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ 100 คน เป็น ชาย 50 คน หญิง 50 คน

6. นำแบบทดสอบมาตรวจให้คะแนนตามวิธีประยุกต์การให้คะแนนของคูมบ์ และวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์

7. นำผลการทดสอบมาวิเคราะห์ หาคุณภาพข้อสอบโดยใช้ทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory: CTT) เพื่อหาค่าความยากง่าย ค่าอำนาจจำแนกของข้อสอบ และค่าความเชื่อมั่น จากสัมประสิทธิ์แอลฟาของครอนบาค ได้ค่าความเชื่อมั่นของวิธีประยุกต์การให้คะแนนของคูมบ์ เท่ากับ 0.8177 ส่วนวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ได้ค่าความเชื่อมั่นเท่ากับ 0.8272

8. นำผลการทดสอบมาวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบทั้งตัวถูกและตัวลวงในเรื่องเพศ เพื่อตรวจสอบให้แน่ใจว่าข้อสอบที่คัดเลือกมามีคุณภาพเพียงพอและจะไม่เกิดการทำหน้าที่ต่างกันของข้อสอบทั้งตัวถูกและตัวลวง

9. ปรับปรุงข้อสอบและคัดเลือกข้อสอบที่มีคุณภาพจำนวน 25 ข้อ โดยวิธีประยุกต์การให้คะแนนของคูมบ์ ค่าความยากง่ายอยู่ระหว่าง 0.27 – 0.92 ค่าอำนาจจำแนกอยู่ระหว่าง 0.04 – 0.45 วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ ค่าความยากง่ายอยู่ระหว่าง 0.02 – 0.89 ค่าอำนาจจำแนกอยู่ระหว่าง 0.07 – 0.63 ส่วนการทำหน้าที่ต่างกันของข้อสอบทั้งตัวถูกและตัวลวง ผู้วิจัยมีหลักเกณฑ์ในการพิจารณาแต่ละข้อเรียงตามลำดับดังนี้

9.1 ข้อสอบต้องไม่เกิดการทำหน้าที่ต่างกันของข้อสอบทั้งตัวถูกและตัวลวง (DIFและ DDF)

9.2 ข้อสอบต้องไม่เกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF)

9.3 ข้อสอบต้องไม่เกิดการทำหน้าที่ต่างกันของตัวลวง (DDF)

ทั้งนี้ผู้วิจัยพบข้อสอบที่เกิดการทำหน้าที่ต่างกันของข้อสอบทั้งตัวถูกและตัวลวง แต่การคัดเลือกผู้วิจัยคำนึงถึงโครงสร้างข้อสอบเป็นหลักในการพิจารณา

10. จัดพิมพ์และนำไปใช้กับนักเรียนกลุ่มตัวอย่าง จำนวน 1,200 คน

3. การเก็บรวบรวมข้อมูล

ระยะเวลาในการเก็บรวบรวมข้อมูล

การดำเนินการเก็บรวบรวมข้อมูลการวิจัย ผู้วิจัยใช้เวลาในการรวบรวมข้อมูลระหว่าง วันที่ 1 มกราคม 2560 – 28 กุมภาพันธ์ 2560

ขั้นตอนการเก็บรวบรวมข้อมูล

1. ติดต่อโรงเรียนที่ใช้เป็นกลุ่มตัวอย่าง จำนวน 10 โรงเรียน จำนวน 26 ห้องเรียน แบบทดสอบจำนวน 1,277 ฉบับ เพื่อขอความอนุเคราะห์ในการเก็บข้อมูล และกำหนด วัน เวลา ในการสอบ

2. เตรียมแบบทดสอบให้เพียงพอกับจำนวนนักเรียนที่สอบในแต่ละครั้ง และวางแผนในการดำเนินการสอบล่วงหน้า

3. ดำเนินการสอบกับนักเรียนกลุ่มตัวอย่าง

4. ผู้วิจัยหรือตัวแทนผู้วิจัย อธิบายให้นักเรียนเข้าใจจุดประสงค์ของการทำแบบทดสอบ

5. แบบทดสอบฉบับนี้มี 25 ข้อ นักเรียนทุกคนได้รับฉบับเดียวกัน แต่แตกต่างกันที่วิธีการตรวจให้คะแนน โดยที่โรงเรียนแต่ละโรงเรียนจะได้รับแบบทดสอบที่มีวิธีการตรวจทั้งวิธีประยุกต์การให้คะแนนของคูมบ์และวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ ซึ่งใน 1 ห้องเรียนจะได้รับแบบทดสอบที่มีวิธีการตรวจแบบเดียวกัน ให้นักเรียนทำแบบทดสอบโดยที่นักเรียนสามารถตัดตัวลวงที่เป็นคำตอบที่ผิดได้ 1 หรือ 2 หรือ 3 ตัวเลือก ก็ได้ โดยไม่จำเป็นต้องตัดตัวเลือกทั้ง 3 ตัวในแต่ละข้อ แต่พยายามตัดออกตามความรู้ที่สามารถตัดออกได้เท่านั้น ถ้าไม่สามารถตัดออกได้ก็ไม่ต้องตัดออก

6. คัดเลือกกระดาษคำตอบที่นักเรียนทำแบบทดสอบได้อย่างสมบูรณ์ วิธีการละ 600 คน รวมจำนวนทั้งหมด 1,200 คน ทั้งนี้เพื่อจะได้ไม่มีความแตกต่างกันทางด้านจำนวนข้อมูลที่นำมาวิเคราะห์ของทั้ง 2 วิธี

4. การจัดกระทำข้อมูลและการวิเคราะห์ข้อมูล

4.1 การจัดกระทำข้อมูล

1. ในการทดลอง (Try out) เครื่องมือ มีขั้นตอน ดังนี้

1.1 ตรวจสอบความสมบูรณ์ของการตอบแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์จากกระดาษคำตอบของนักเรียนก่อน

1.2 นำแบบทดสอบมาตรวจให้คะแนนตามเกณฑ์ที่กำหนดไว้

1.3 วิเคราะห์หาค่าความยากง่าย ค่าอำนาจจำแนก ค่าความเชื่อมั่น จากสัมประสิทธิ์แอลฟาของครอนบาค โดยใช้โปรแกรมคอมพิวเตอร์

1.4 วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบทั้งตัวถูกและตัวลวง โดยใช้โปรแกรมคอมพิวเตอร์

2. การวิเคราะห์ข้อมูลเพื่อตอบคำถามการวิจัย มีขั้นตอน ดังนี้

2.1 วิเคราะห์ข้อมูลเบื้องต้น (preliminary data analysis) มีขั้นตอน ดังนี้

2.1.1 นำข้อมูลที่รวบรวมได้มาตรวจสอบความสมบูรณ์ของการตอบจากกระดาษคำตอบของนักเรียนว่ามีสภาพเป็นอย่างไร

2.1.2 ตรวจให้คะแนนตามเกณฑ์ที่กำหนด

2.1.3 วิเคราะห์ค่าสถิติพื้นฐานโดยใช้โปรแกรมคอมพิวเตอร์เพื่อหาค่าสูงสุด ค่าต่ำสุด ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน สัมประสิทธิ์ของการแปรผัน ความเบ้ และความโด่ง ของข้อมูล

2.2 ตรวจสอบการเป็นมิติเดียวของแบบทดสอบด้วยวิธีการวิเคราะห์องค์ประกอบ (Factor Analysis) โดยใช้โปรแกรมคอมพิวเตอร์

2.3 นำข้อมูลที่ผ่านการวิเคราะห์ขั้นต้นและการตรวจสอบการเป็นมิติเดียวมาทำการหาคุณภาพข้อสอบตามทฤษฎีตอบสนองข้อสอบ (IRT) โดยวิเคราะห์เพื่อหาค่าพารามิเตอร์ของข้อสอบ ค่าฟังก์ชันสารสนเทศของข้อสอบ ค่าฟังก์ชันสารสนเทศของแบบทดสอบ และค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่า วิเคราะห์ข้อมูลโดยใช้ Generalized Partial Credit Model (G-PCM) ด้วยโปรแกรมคอมพิวเตอร์และนำค่าที่ได้มาคำนวณค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบทดสอบ และค่าประสิทธิภาพสัมพัทธ์เฉลี่ยของแบบทดสอบ

4.2 สถิติที่ใช้ในการวิเคราะห์ข้อมูล

สถิติที่ใช้ในการหาคุณภาพเครื่องมือ

1. ค่าความเที่ยงตรงเชิงเนื้อหา (Content Validity) โดยพิจารณาหาค่าดัชนีความสอดคล้อง (Index of Congruency: IOC) (ล้วน สายยศและอังคณา สายยศ. 2543: 249)

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

เมื่อ IOC แทน ดัชนีความสอดคล้องระหว่างแบบทดสอบกับนิยาม

$\sum R$ แทน ผลรวมคะแนนผู้เชี่ยวชาญทั้งหมด

N แทน จำนวนผู้เชี่ยวชาญทั้งหมด

2. ค่าความยากง่าย (Difficulty) โดยใช้การวิเคราะห์ข้อสอบรายข้อ (Item Analysis) (ศิริชัย กาญจนวาสี. 2556: 243) ดังนี้

$$p = \frac{P_H + P_L}{2}$$

เมื่อ $\sum H$ แทน คะแนนรวมรายข้อของทุกคนในกลุ่มสูง

$\sum L$ แทน คะแนนรวมรายข้อของทุกคนในกลุ่มต่ำ

$\sum T_H$ แทน คะแนนเต็มรวมรายข้อของทุกคนในกลุ่มสูง

$\sum T_L$ แทน คะแนนเต็มรวมรายข้อของทุกคนในกลุ่มต่ำ

P_H แทน $\frac{\sum H}{\sum T_H}$, P_L แทน $\frac{\sum L}{\sum T_L}$

3. ค่าอำนาจจำแนก (r) คำนวณจากสูตร(ศิริชัย กาญจนวาสี. 2556: 243)

$$r = P_H - P_L$$

เมื่อ $\sum H$ แทน คะแนนรวมรายข้อข องทุกคนในกลุ่มสูง
$\sum L$ แทน คะแนนรวมรายข้อข องทุกคนในกลุ่มต่ำ
$\sum T_H$ แทน คะแนนเต็มรวมรายข้อของทุกคนในกลุ่มสูง
$\sum T_L$ แทน คะแนนเต็มรวมรายข้อของทุกคนในกลุ่มต่ำ
P_H แทน $\frac{\sum H}{\sum T_H}$, P_L แทน $\frac{\sum L}{\sum T_L}$

4. ค่าความเชื่อมั่น คำนวณได้จากสัมประสิทธิ์แอลฟาของครอนบาค (Cronbach's Alpha) (ศิริชัย กาญจนวาสี. 2556: 71) ดังนี้

$$\alpha = \left[\frac{k}{k-1}\right]\left[1 - \frac{\sum S_i^2}{S_X^2}\right]$$

เมื่อ α แทน สัมประสิทธิ์ความเชื่อมั่นของแบบทดสอบ
S_i^2 แทน ความแปรปรวนของคะแนนส่วนที่ i (หรือข้อที่ i)
S_X^2 แทน ความแปรปรวนของคะแนนรวม x
k แทน จำนวนส่วน (components) ที่นำมารวมเป็น x (หรือจำนวนแบบทดสอบ)

5. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) และการทำหน้าที่ ต่างกันของตัวลวง (DDF) (เรืองเดช ศิริกิจ. 2554: 126)

$$\hat{\alpha}_{MH_i} = \frac{\sum_{s=1}^{S} R_{1s}F_{0s}/n_s}{\sum_{s=1}^{S} R_{0s}F_{1s}/n_s}$$

4.3 การวิเคราะห์ข้อมูลในการวิจัย

ข้อมูลที่ได้รับจากการสอบของนักเรียนนำมาตรวจให้คะแนนตามเกณฑ์ที่กำหนดแล้วนำคะแนนที่ได้ในแบบทดสอบมาวิเคราะห์ข้อมูล ดังนี้

1. วิเคราะห์ข้อมูลเบื้องต้นจากการหาค่าสูงสุด ค่าต่ำสุด ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน สัมประสิทธิ์ของการแปรผัน ความเบ้ และความโด่ง ของข้อมูล

2. ตรวจสอบการเป็นมิติเดียวของแบบทดสอบด้วยวิธีการวิเคราะห์องค์ประกอบ (Factor Analysis)

3. วิเคราะห์เพื่อหาค่าพารามิเตอร์ของข้อสอบและวัดคุณภาพของข้อสอบและแบบทดสอบจากฟังก์ชันสารสนเทศของข้อสอบและฟังก์ชันสารสนเทศของแบบทดสอบ และค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่า โดยวิเคราะห์ข้อสอบรายข้อตามทฤษฎีการตอบสนองข้อสอบ (IRT) แบบพหุวิภาค วิเคราะห์ข้อมูลโดยใช้ Generalized Partial Credit Model (G-PCM) ด้วยโปรแกรมคอมพิวเตอร์และนำค่าที่ได้มาคำนวณค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบทดสอบและค่าประสิทธิภาพสัมพัทธ์เฉลี่ยของแบบทดสอบ

3.1 ค่าฟังก์ชันสารสนเทศของข้อสอบ (Item Information Function: IIF) (ศิริชัย กาญจนวาสี. 2555: 110)

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

เมื่อ $I_i(\theta)$ แทน ฟังก์ชันสารสนเทศของข้อสอบข้อที่ i

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

3.2 ค่าฟังก์ชันสารสนเทศของแบบทดสอบ (Test Information Function: TIF) (ศิริชัย กาญจนวาสี. 2555: 110)

จาก IIF สามารถนำมาคำนวณ TIF

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

3.3 ค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่า (Standard Error of Estimation: $SE(\theta)$) (ศิริชัย กาญจนวาสี. 2555: 110)

จาก TIF สามารถนำมาคำนวณ $SE(\theta)$

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

เมื่อ $SE(\theta)$ แทน ความคลาดเคลื่อนมาตรฐานของการประมาณค่าสำหรับผู้ตอบที่มีความสามารถ θ

$I(\theta)$ แทน สารสนเทศที่ได้จากแบบทดสอบ สำหรับผู้ตอบที่มีความสามารถ θ

3.4 ค่าประสิทธิภาพสัมพัทธ์เฉลี่ย (ratio of average information: RAI) (ศิริชัย กาญจนวาสี. 2555: 66-67)

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

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

$AI(\theta, Y)$ แทน ค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบทดสอบฉบับ Y ณ ทุกตำแหน่ง θ

บทที่ 4
ผลการวิเคราะห์ข้อมูล

สัญลักษณ์ที่ใช้ในการวิเคราะห์ข้อมูล

การเสนอผลการวิเคราะห์ข้อมูลในการวิจัยครั้งนี้ ผู้วิจัยได้กำหนดสัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูลเพื่อสื่อความหมายในการเสนอผลงานวิจัยให้เข้าใจตรงกัน ดังนี้

Min แทน คะแนนต่ำสุด

Max แทน คะแนนสูงสุด

$\overline{x}$ แทน คะแนนเฉลี่ย

SD แทน ค่าส่วนเบี่ยงเบนมาตรฐาน

C.V. แทน สัมประสิทธิ์ของการแปรผัน

θ แทน ความสามารถ

IIF แทน ค่าฟังก์ชันสารสนเทศของข้อสอบ

TIF แทน ค่าฟังก์ชันสารสนเทศของแบบทดสอบ

AI แทน ค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบทดสอบ

RAI แทน ค่าประสิทธิภาพสัมพัทธ์เฉลี่ยของแบบทดสอบ

SE(θ) แทน ความคลาดเคลื่อนมาตรฐานของการประมาณค่าสำหรับผู้ตอบที่มีระดับความสามารถ θ

การเสนอผลการวิเคราะห์ข้อมูล

การวิจัยครั้งนี้มีความมุ่งหมายเพื่อศึกษาฟังก์ชันสารสนเทศของข้อสอบ (Item Information Function) และฟังก์ชันสารสนเทศของแบบทดสอบ (Test Information Function) จากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 ที่ตรวจให้คะแนนความรู้บางส่วนด้วยวิธีประยุกต์การให้คะแนนของคูมบ์ (Modified Coombs Approach) และวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ (The Arnold Scoring/Response Method) ผู้วิจัยได้เสนอผลการวิเคราะห์ข้อมูลเรียงตามลำดับขั้นตอน ดังนี้

ตอนที่ 1 ผลการวิเคราะห์ค่าสถิติพื้นฐาน

ตอนที่ 2 ผลการวิเคราะห์องค์ประกอบเพื่อตรวจสอบความเป็นเอกมิติของข้อมูล

ตอนที่ 3 ผลการวิเคราะห์ค่าพารามิเตอร์ของข้อสอบ

ตอนที่ 4 ผลการวิเคราะห์ค่าฟังก์ชันสารสนเทศของข้อสอบ ค่าฟังก์ชันสารสนเทศของแบบทดสอบ ค่าประสิทธิภาพสัมพัทธ์เฉลี่ย และค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่า

ตอนที่ 1 ผลการวิเคราะห์ค่าสถิติพื้นฐาน

การวิเคราะห์ค่าสถิติพื้นฐานในตอนนี้ เป็นการวิเคราะห์เพื่อหาค่าสูงสุด ค่าต่ำสุด ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน สัมประสิทธิ์ของการแปรผัน ค่าความเบ้ และค่าความโด่ง ของข้อมูล จากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ จำนวน 25 ข้อ ที่มีวิธีการตรวจแตกต่างกัน 2 วิธี จากแบบทดสอบชุดเดียวกัน คือ วิธีประยุกต์การให้คะแนนของคูมบ์ และวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ ซึ่งวิธีประยุกต์การให้คะแนนของคูมบ์ มีคะแนนสูงสุด เท่ากับ 3 ต่ำสุดเท่ากับ 0 ในแต่ละข้อ รวมคะแนนเต็มเท่ากับ 75 คะแนน วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ มีคะแนนสูงสุด เท่ากับ 1 ต่ำสุดเท่ากับ -1/3 ในแต่ละข้อ รวมคะแนนเต็มเท่ากับ 25 คะแนน ดังตาราง 6

ตาราง 6 ค่าสถิติพื้นฐานของคะแนนสอบจำแนกตามวิธีการตรวจให้คะแนน

วิธีการตรวจให้คะแนน	ประยุกต์การให้คะแนนของคูมบ์	การตอบและการตรวจให้คะแนนแบบอาร์โนลด์
จำนวนผู้สอบ	600	600
คะแนนเต็ม	75	25
Min	2	-5.16
Max	72	23.67
$\overline{x}$	33.88	7.56
SD	13.53	6.53
C.V.(%)	39.94	86.38
Skewness	0.72	0.43
Kurtosis	0.18	-0.62

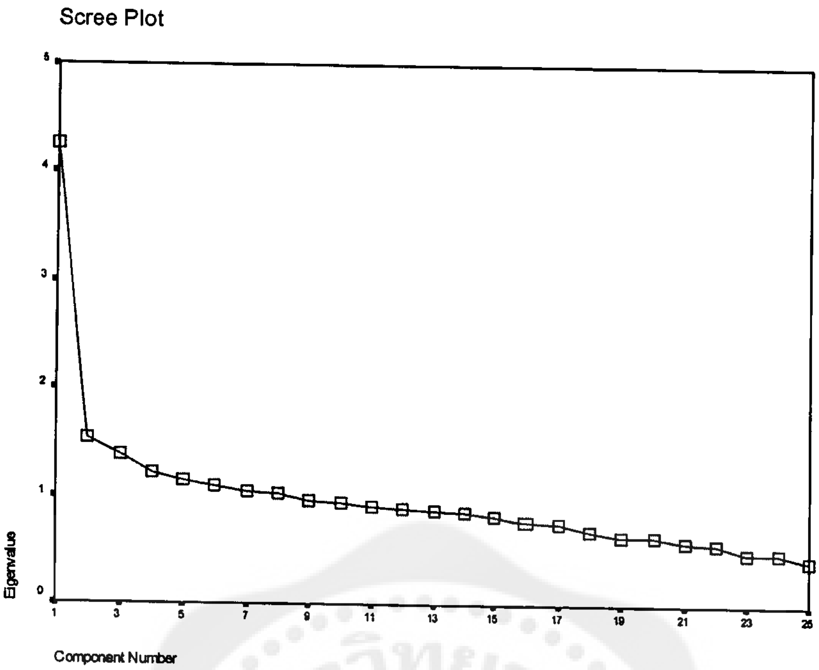
ตาราง 6 พบว่า วิธีประยุกต์การให้คะแนนของคูมบ์มีคะแนนเฉลี่ยเท่ากับ 33.88 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 13.53 วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์มีคะแนนเฉลี่ยเท่ากับ 7.56 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 6.53 ซึ่งวิธีการตอบและการตรวจให้คะแนนแบบ อาร์โนลด์มีการกระจายของข้อมูลมากกว่าวิธีประยุกต์การให้คะแนนของคูมบ์สังเกตได้จากสัมประสิทธิ์ของการแปรผัน (%) ต่างกันอยู่ถึง 46.44% เมื่อพิจารณาถึงความเบ้ของข้อมูลทั้ง 2 วิธี พบว่าข้อมูลเบ้ขวาทั้ง 2 วิธี หมายถึง คนส่วนใหญ่ที่ตอบแบบทดสอบมีความสามารถค่อนข้างต่ำ ซึ่งคะแนนส่วนใหญ่จะต่ำกว่าค่าเฉลี่ย และเมื่อพิจารณาถึงความโด่งของข้อมูล พบว่า ข้อมูลของทั้ง 2 วิธีเป็นโค้งที่แบนกว่าปกติ หมายความว่า ข้อมูลมีการกระจายมาก

ตอนที่ 2 ผลการวิเคราะห์องค์ประกอบเพื่อตรวจสอบความเป็นเอกมิติของข้อมูล

การตรวจสอบความเป็นมิติเดียวของแบบทดสอบ เป็นข้อตกลงพื้นฐานที่สำคัญของทฤษฎีการตอบสนองข้อสอบ (IRT) เป็นการตรวจสอบว่าแบบทดสอบวัดคุณลักษณะแฝงเพียงคุณลักษณะเดียว ผู้วิจัยได้นำคะแนนที่ได้จากการตอบแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ เรื่อง ตัวหารร่วมมากและตัวคูณร่วมน้อย จำนวน 25 ข้อ มาตรวจสอบความเป็นมิติเดียว โดยใช้การวิเคราะห์องค์ประกอบ (Factor Analysis) ด้วยวิธีการวิเคราะห์องค์ประกอบหลัก (Principle Componant Analysis) หาค่าไอเกน (Eigen Value) และร้อยละของความแปรปรวน ของวิธีประยุกต์การให้คะแนนของคูมบ์ และวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ ผลการวิเคราะห์แสดงรายละเอียดในตาราง 7 และ 8

ตาราง 7 ค่าไอเกน และร้อยละของความแปรปรวนขององค์ประกอบของแบบทดสอบจำนวน 25 ข้อ เมื่อตรวจด้วยวิธีประยุกต์การให้คะแนนของคูมบ์

องค์ประกอบ	ค่าไอเกน	ร้อยละของความแปรปรวน
1	4.259	17.036
2	1.531	6.125
3	1.379	5.518
4	1.217	4.867
5	1.137	4.547
6	1.089	4.355
7	1.051	4.203
8	1.020	4.080

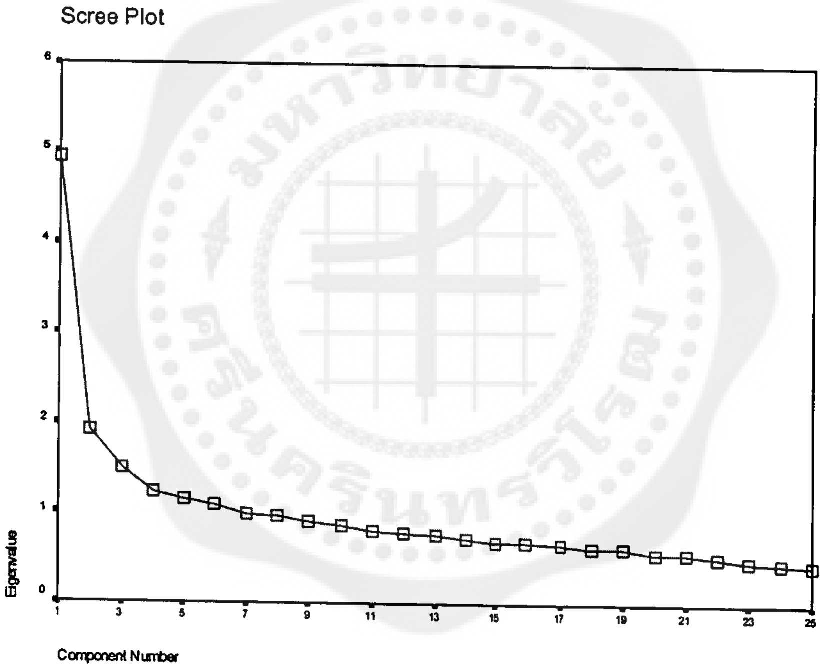


ภาพประกอบ 6 ผลการตรวจสอบความเป็นมิติเดียวกันของข้อสอบเมื่อตรวจด้วยวิธีประยุกต์ การให้คะแนนของคูมบ์

ตาราง 7 พบว่า แบบทดสอบจำนวน 25 ข้อ มีค่า Eigen Ration ระหว่างองค์ประกอบที่ 1 และองค์ประกอบที่ 2 ประมาณ 2.78 เท่า จะเห็นได้ว่าองค์ประกอบที่ 1 มีค่าไอเกนแตกต่างจากไอเกนขององค์ประกอบอื่น ส่วนค่าไอเกนขององค์ประกอบอื่นๆ มีค่าใกล้เคียงกัน และเมื่อพิจารณาถึงร้อยละของความแปรปรวน พบว่า องค์ประกอบที่ 1 มีค่าร้อยละของความแปรปรวนเท่ากับ 17.036 ซึ่งน้อยกว่าเกณฑ์ที่กำหนดเล็กน้อย เกณฑ์ในที่นี้คือ ค่าร้อยละของความแปรปรวนควรมีค่าไม่น้อยกว่าร้อยละ 20 หรือค่าไอเกนขององค์ประกอบที่ 1 แตกต่างจากองค์ประกอบอื่นอย่างเด่นชัด จึงจะถือว่าแบบทดสอบมีความเป็นมิติเดียว (อุทัยวรรณ สายพัฒนะ. 2547: 64-65; อ้างอิงจาก Raju. 1993: 304) ดังนั้นจึงถือได้ว่าแบบทดสอบฉบับนี้เป็นมิติเดียว สามารถนำข้อมูลไปวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบได้

ตาราง 8 ค่าไอเกน และร้อยละของความแปรปรวนขององค์ประกอบของแบบทดสอบจำนวน 25 ข้อ เมื่อตรวจด้วยวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์

องค์ประกอบ	ค่าไอเกน	ร้อยละของความแปรปรวน
1	4.950	19.800
2	1.918	7.673
3	1.487	5.947
4	1.241	4.963
5	1.157	4.630
6	1.097	4.387



ภาพประกอบ 7 ผลการตรวจสอบความเป็นมิติเดียวกันของข้อสอบเมื่อตรวจด้วยวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์

ตาราง 8 พบว่า แบบทดสอบจำนวน 25 ข้อ มีค่า Eigen Ration ระหว่างองค์ประกอบที่ 1 และองค์ประกอบที่ 2 ประมาณ 2.58 เท่า จะเห็นได้ว่าองค์ประกอบที่ 1 มีค่าไอเกนแตกต่างจากไอเกนขององค์ประกอบอื่น ส่วนค่าไอเกนขององค์ประกอบอื่น ๆ มีค่าใกล้เคียงกัน และเมื่อพิจารณาถึงร้อยละของความแปรปรวน พบว่า องค์ประกอบที่ 1 มีค่าร้อยละของความแปรปรวนเท่ากับ 19.800 ซึ่งน้อยกว่าเกณฑ์ที่กำหนดเล็กน้อย เกณฑ์ในที่นี้คือ ค่าร้อยละของความแปรปรวนควรมีค่าไม่น้อยกว่าร้อยละ 20 หรือค่าไอเกนขององค์ประกอบที่ 1 แตกต่างจากองค์ประกอบอื่นอย่างเด่นชัด

จึงจะถือว่าแบบทดสอบมีความเป็นมิติเดียว (อุทัยวรรณ สายพัฒนะ. 2547: 64-65; อ้างอิงจาก Raju. 1993: 304) ดังนั้นจึงถือได้ว่าแบบทดสอบฉบับนี้เป็นมิติเดียว สามารถนำข้อมูลไปวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบได้

ตอนที่ 3 ผลการวิเคราะห์ค่าพารามิเตอร์ของข้อสอบ

ค่าพารามิเตอร์ของข้อสอบจะนำเสนอ 3 ค่า คือ ค่าพารามิเตอร์อำนาจจำแนก (a) ค่าพารามิเตอร์ความยากง่าย (b) และค่าพารามิเตอร์ระดับความยากของขั้นการตอบ(δ_{ij}) ของวิธีประยุกต์การให้คะแนนของคูมบ์ และวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ ผลการวิเคราะห์แสดงรายละเอียดในตาราง 9 , 10 และ 11

ตาราง 9 ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ (a) ที่วิเคราะห์จากคะแนนที่ได้จากวิธีการตอบและตรวจให้คะแนนทั้ง 2 วิธี โดยวิเคราะห์ตามโมเดล GPCM

ข้อที่	ค่าพารามิเตอร์อำนาจจำแนก (a)	
	วิธีประยุกต์การให้คะแนนของคูมบ์	วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์
1	0.33	0.43
2	0.38	0.52
3	0.44	0.36
4	0.39	0.35
5	0.49	0.59
6	0.37	-0.06
7	0.24	0.20
8	0.17	0.16
9	0.12	0.37
10	0.28	0.22
11	0.25	0.53
12	0.30	0.47
13	-0.01	-0.08
14	0.16	0.20
15	0.16	0.12
16	0.22	0.48
17	0.39	0.57
18	0.30	0.28

ตาราง 10 ค่าพารามิเตอร์ความยากง่ายของข้อสอบ (*b*) ที่วิเคราะห์จากคะแนนที่ได้จากวิธีการตอบและตรวจให้คะแนนทั้ง 2 วิธี โดยวิเคราะห์ตามโมเดล GPCM

ข้อที่	ค่าพารามิเตอร์ความยากง่าย (*b*)	
	วิธีประยุกต์การให้คะแนนของคูมบ์	วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์
1	-0.76	-0.43
2	0.13	-0.40
3	-0.61	-0.66
4	-0.57	-0.81
5	-0.55	-1.00
6	-1.38	1.20
7	0.64	0.54
8	1.84	0.96
9	0.98	0.14
10	0.90	0.64
11	0.21	-0.05
12	0.46	0.18
13	-63.40	-5.95
14	2.47	1.78
15	1.87	1.81
16	0.52	0.19
17	0.68	-0.22
18	-0.25	-0.21
19	0.53	0.16
20	0.38	5.78
21	0.64	0.41
22	0.72	0.04
23	0.14	0.04
24	0.34	0.22
25	0.05	-0.41

จากตาราง 10 พบว่า ค่าความยากง่ายของวิธีประยุกต์การให้คะแนนของคูมบ์มีค่าตั้งแต่ -63.4 ถึง 2.47 ส่วนค่าความยากง่ายของวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์มีค่าตั้งแต่

-5.95 ถึง 5.78 โดยภาพรวมทั้งวิธีประยุกต์การให้คะแนนของคูมบ์ และวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์มีค่าความยากง่ายของข้อสอบค่อนข้างต่ำ ซึ่งหมายถึงผู้ตอบที่มีความสามารถค่อนข้างต่ำจะตอบข้อสอบได้ แต่เมื่อพิจารณาแต่ละวิธีปรากฏว่า วิธีประยุกต์การให้คะแนนของคูมบ์ ข้อที่ 13 มีค่าต่ำมาก หมายถึง ว่าคนที่สามารถตอบข้อนี้ได้ มีความสามารถอยู่ในระดับที่ต่ำมาก ข้อนี้เป็นข้อที่ง่ายเกินไป ไม่ค่อยดี ส่วน วิธีการตอบและการตรวจให้คะแนนแบบ อาร์โนลด์ ข้อ 13 มีค่าต่ำมากเช่นกัน เหมาะกับผู้ที่มีความสามารถที่ต่ำมาก ในทางกลับกัน ข้อ 20 มีค่าที่สูงมาก หมายถึงว่า คนที่สามารถตอบข้อนี้ได้ ต้องมีความสามารถอยู่ในระดับที่สูงมาก แปลว่าข้อนี้ยากเกินไป ทั้งนี้ในทางปฏิบัตินิยมใช้ข้อสอบที่มีค่าความยากง่ายอยู่ระหว่าง -2.50 ถึง 2.50 (ศิริชัย กาญจนวาสี. 2555: 55)

ตาราง 11 ค่าพารามิเตอร์ระดับความยากของขั้นการตอบ(δ_{ij}) ที่วิเคราะห์จากคะแนนที่ได้จากวิธีการตอบและตรวจให้คะแนนทั้ง 2 วิธี โดยวิเคราะห์ตามโมเดล GPCM

ข้อที่	ค่าพารามิเตอร์ระดับความยากของขั้นการตอบ(δ_{ij})					
	วิธีประยุกต์การให้คะแนนของคูมบ์			วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์		
	δ_1	δ_2	δ_3	δ_1	δ_2	δ_3
1	0.25	-4.29	-0.73	0.19	-4.24	-1.26
2	-0.05	-3.47	-0.56	0.21	-3.69	-0.71
3	0.27	-2.59	-0.01	0.24	-3.03	-0.55
4	0.22	-2.65	-0.37	0.28	-2.95	-0.38
5	0.27	-2.47	0.08	0.59	-2.51	-0.02
6	0.50	-2.64	-0.41	0.08	-3.14	-0.76
7	-0.15	-2.53	-0.66	-0.10	-2.41	-0.58
8	-0.31	-2.65	-0.19	-0.15	-2.93	-0.55
9	-0.12	-2.64	-0.41	-0.05	-2.77	-0.29
10	-0.25	-2.39	-0.38	-0.14	-2.74	-0.41
11	-0.05	-2.44	-0.46	0.03	-2.38	-0.81
12	-0.14	-2.08	-0.49	-0.08	-2.18	-0.40
13	-0.42	-2.28	-0.49	-0.48	-2.56	-0.79
14	-0.40	-2.14	-0.47	-0.35	-2.73	-0.50
15	-0.30	-2.30	-0.55	-0.21	-2.14	-0.48
16	-0.11	-2.18	-0.56	-0.09	-2.51	-0.68

ตาราง 11 (ต่อ)

ข้อที่	ค่าพารามิเตอร์ระดับความยากของขั้นการตอบ(δ_{ij})					
	วิธีประยุกต์การให้คะแนนของคูมบ์			วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์		
	δ_1	δ_2	δ_3	δ_1	δ_2	δ_3
17	-0.27	-2.34	-0.52	0.12	-2.19	-0.37
18	0.08	-2.82	-0.35	0.06	-3.27	-0.48
19	-0.25	-2.56	-0.53	-0.06	-2.75	-0.53
20	-0.18	-2.80	-0.61	-0.20	-2.41	-0.75
21	-0.15	-1.79	-0.43	-0.14	-2.43	-0.76
22	-0.26	-2.18	-0.32	-0.02	-2.32	-0.12
23	-0.06	-1.96	-0.27	-0.02	-2.04	-0.43
24	-0.11	-1.95	-0.25	-0.09	-2.11	-0.25
25	-0.02	-2.03	-0.39	0.23	-2.59	0.09

แบบทดสอบฉบับนี้มีระดับความยากของขั้นการตอบทั้งหมด 3 ขั้น จากตาราง 11 พบว่า ขั้นความยากของการตอบของรายการคำตอบในแต่ละข้อไม่ได้ถูกจัดเรียงอย่างเป็นลำดับจากการตอบขั้นที่ 1 (ง่ายสุด) ไปยังการตอบขั้นที่ 3 (ยากที่สุด) แสดงว่า ส่วนใหญ่ข้อคำถามไม่มีรายการคำตอบที่มีการเรียงลำดับของการตัดสินใจ

ตอนที่ 4 ผลการวิเคราะห์ค่าฟังก์ชันสารสนเทศของข้อสอบ ค่าฟังก์ชันสารสนเทศของแบบทดสอบ ค่าประสิทธิภาพสัมพัทธ์เฉลี่ย และค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่า

ผลการวิเคราะห์ค่าฟังก์ชันสารสนเทศของข้อสอบ ค่าฟังก์ชันสารสนเทศของแบบทดสอบ ค่าประสิทธิภาพสัมพัทธ์เฉลี่ย และค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่า โดยวิเคราะห์จากคะแนนที่ได้จากวิธีการตอบและตรวจให้คะแนนด้วยวิธีประยุกต์การให้คะแนนของคูมบ์ และวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ ซึ่งในการวิจัยครั้งนี้ วิธีการตอบและตรวจให้คะแนนความรู้บางส่วนเป็นแบบพหุวิภาค กล่าวคือ 1) วิธีประยุกต์การให้คะแนนของคูมบ์ ให้คะแนนในแต่ละลำดับขั้นตั้งแต่ต่ำสุดถึงสูงสุด เป็น 0, 1, 2, 3 ซึ่งมี 3 ลำดับขั้น 2) วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ ให้คะแนนในแต่ละลำดับขั้นตั้งแต่ต่ำสุดถึงสูงสุด เป็น -1/3, 1/9, 1/3, 1 เมื่อปรับสเกลคะแนนจะได้ 0, 1, 2, 3 เพื่อให้โปรแกรมสามารถคำนวณได้ ซึ่งมี 3 ลำดับขั้น (สุพจน์

เกิดสุวรรณ์. 2545: 134) ในการวิเคราะห์ทั้ง 2 วิธี วิเคราะห์ข้อมูลโดยใช้ Generalized Partial Credit Model (G-PCM) ด้วยโปรแกรมคอมพิวเตอร์ ซึ่งนำเสนอดังต่อไปนี้

4.1 การเปรียบเทียบค่าฟังก์ชันสารสนเทศของข้อสอบตามวิธีประยุกต์การให้คะแนนของคูมบ์ และวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ จำแนกตามรายข้อ แสดงผลดังตาราง 12

ตาราง 12 ค่าฟังก์ชันสารสนเทศสูงสุดของข้อสอบ จำแนกตามรายข้อและวิธีการตรวจให้คะแนน

ข้อ	ฟังก์ชันสารสนเทศของข้อสอบ (IIF)	
	วิธีประยุกต์การให้คะแนนของคูมบ์	วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์
1	0.24	0.40
2	0.30	0.59
3	0.40	0.27
4	0.30	0.25
5	0.48	0.69
6	0.27	0.01
7	0.11	0.08
8	0.06	0.05
9	0.03	0.28
10	0.16	0.10
11	0.13	0.56
12	0.18	0.43
13	0.00	0.01
14	0.05	0.08
15	0.05	0.03
16	0.09	0.45
17	0.31	0.64
18	0.18	0.17
19	0.47	0.25
20	0.44	0.00
21	0.11	0.23
22	0.25	0.49

ตาราง 12 (ต่อ)

ข้อ	ฟังก์ชันสารสนเทศของข้อสอบ (IIF)	
	วิธีประยุกต์การให้คะแนนของคูมบ์	วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์
23	0.29	0.53
24	0.23	0.35
25	0.26	0.67

จากตาราง 12 พบว่า ค่าฟังก์ชันสารสนเทศสูงสุดของข้อสอบเป็นดังนี้ วิธีประยุกต์การให้คะแนนของคูมบ์ให้ค่าฟังก์ชันสารสนเทศของข้อสอบสูงกว่าวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ทั้งหมด 10 ข้อ คิดเป็น 40% ของแบบทดสอบทั้งฉบับ และวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ให้ค่าฟังก์ชันสารสนเทศของข้อสอบสูงกว่าวิธีประยุกต์การให้คะแนนของคูมบ์ทั้งหมด 15 ข้อ คิดเป็น 60% ของแบบทดสอบทั้งฉบับ ซึ่งโดยภาพรวมแล้ว พบว่า วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ให้ค่าฟังก์ชันสารสนเทศสูงสุดของข้อสอบสูงกว่าวิธีประยุกต์การให้คะแนนของคูมบ์ แต่เมื่อพิจารณาค่าฟังก์ชันสารสนเทศสูงสุดของข้อสอบแต่ละข้อแล้ว วิธีประยุกต์การให้คะแนนของคูมบ์และวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์จะให้ค่าที่ใกล้เคียงกันมาก

4.2 ค่าฟังก์ชันสารสนเทศของแบบทดสอบที่มีวิธีการตอบและตรวจให้คะแนน ทั้ง 2 วิธี ผลการวิเคราะห์แสดงรายละเอียดในตาราง 13

ตาราง 13 ค่าฟังก์ชันสารสนเทศของแบบทดสอบที่วิเคราะห์จากคะแนนที่ได้จากวิธีการตอบและตรวจให้คะแนนทั้ง 2 แบบ

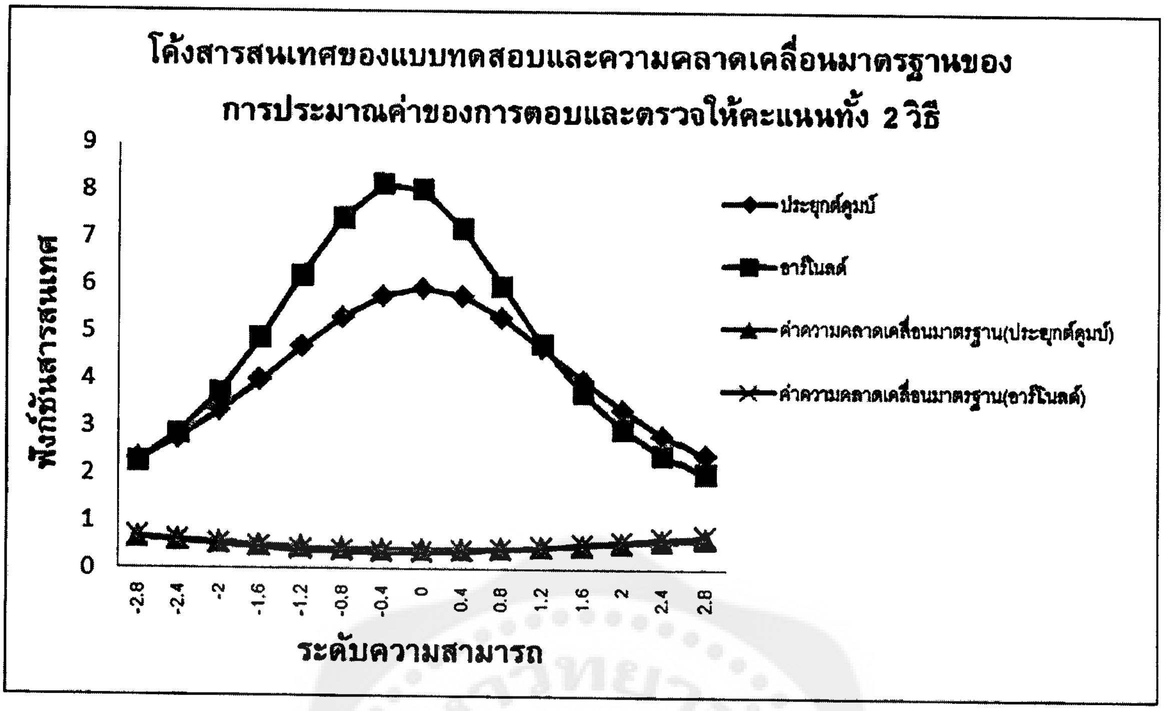
ระดับความสามารถ (θ)	ค่าฟังก์ชันสารสนเทศของแบบทดสอบ (TIF)	
	วิธีประยุกต์การให้คะแนนของคูมบ์	วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์
-2.8	2.30	2.25
-2.4	2.75	2.85
-2.0	3.32	3.71
-1.6	3.98	4.87
-1.2	4.68	6.21
-0.8	5.32	7.43

ตาราง 13 (ต่อ)

ระดับความสามารถ (θ)	ค่าฟังก์ชันสารสนเทศของแบบทดสอบ (TIF)	
	วิธีประยุกต์การให้คะแนนของคูมบ์	วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์
-0.4	5.77	8.14
0	5.94	8.04
0.4	5.77	7.21
0.8	5.32	5.99
1.2	4.68	4.76
1.6	3.99	3.73
2.0	3.36	2.96
2.4	2.83	2.41
2.8	2.41	2.03

จากตาราง 13 พบว่า วิธีประยุกต์การให้คะแนนของคูมบ์ให้ค่าฟังก์ชันสารสนเทศของแบบทดสอบสูงกว่าวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ที่ระดับความสามารถ(θ) เท่ากับ -2.8 และช่วงตั้งแต่ 1.6 ถึง 2.8 แสดงว่า วิธีประยุกต์การให้คะแนนของคูมบ์ เหมาะสำหรับผู้ที่มีความสามารถค่อนข้างต่ำและผู้ที่มีความสามารถค่อนข้างสูง ส่วนวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ให้ค่าฟังก์ชันสารสนเทศของแบบทดสอบสูงกว่าวิธีประยุกต์การให้คะแนนของคูมบ์ที่ระดับความสามารถ(θ) ในช่วง -2.4 ถึง 1.2 แสดงว่า วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์เหมาะสำหรับผู้ที่มีความสามารถปานกลาง

สำหรับโค้งฟังก์ชันสารสนเทศของแบบทดสอบและความคลาดเคลื่อนมาตรฐานของการประมาณค่าที่มีวิธีการตอบและตรวจให้คะแนนทั้ง 2 วิธี ได้แสดงไว้ในภาพประกอบ 8 จากภาพจะเห็นได้ว่า วิธีประยุกต์การให้คะแนนของคูมบ์ และวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์จะให้โค้งฟังก์ชันสารสนเทศคล้ายคลึงกัน ซึ่งวิธีประยุกต์การให้คะแนนของคูมบ์ที่ระดับความสามารถ (θ) เท่ากับ 0 เส้นกราฟจะมีความโด่งมากที่สุด ส่วนวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ที่ระดับความสามารถ (θ) เท่ากับ -0.4 เส้นกราฟจะมีความโด่งมากที่สุด ซึ่งในช่วงระดับความสามารถ (θ) เท่ากับ -2.4 ถึง 1.2 กราฟของวิธีตอบและการตรวจให้คะแนนแบบอาร์โนลด์ จะเป็นกราฟที่โด่งกว่าวิธีประยุกต์การให้คะแนนของคูมบ์ ส่วนกราฟความคลาดเคลื่อนมาตรฐานของการประมาณค่าของทั้ง 2 วิธี แตกต่างกันอย่างไม่ชัดเจน



ภาพประกอบ 8 เปรียบเทียบโค้งฟังก์ชันสารสนเทศของแบบทดสอบที่มีวิธีการตอบและตรวจให้คะแนนทั้ง 2 วิธี

4.3 ค่าสารสนเทศเฉลี่ยของแบบทดสอบที่วิเคราะห์จากคะแนนที่ได้จากวิธีการตอบและตรวจให้คะแนนทั้ง 2 วิธี แสดงรายละเอียดในตาราง 14

ตาราง 14 ค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบทดสอบ

วิธีการตอบและตรวจให้คะแนน	ค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบทดสอบ (AI)
วิธีประยุกต์การให้คะแนนของคูมบ์	4.16
วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์	4.84

จากตาราง 14 พบว่า วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์มีค่าเท่ากับ 4.84 ซึ่งให้ค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบทดสอบสูงกว่าวิธีประยุกต์การให้คะแนนของคูมบ์ที่มีค่าเท่ากับ 4.16

4.4 ค่าประสิทธิภาพสัมพัทธ์เฉลี่ยของแบบทดสอบที่วิเคราะห์จากคะแนนที่ได้จากวิธีการตอบและตรวจให้คะแนนทั้ง 2 วิธี แสดงรายละเอียดในตาราง 15

ตาราง 15 ประสิทธิภาพสัมพัทธ์เฉลี่ยของแบบทดสอบ

วิธีการตอบและตรวจให้คะแนน	ประสิทธิภาพสัมพัทธ์เฉลี่ย (RAI)	
	วิธีประยุกต์การให้คะแนนของคูมบ์	วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์
วิธีประยุกต์การให้คะแนนของคูมบ์	1	0.86
วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์	1.16	1

หมายเหตุ ตัวเลขในเมตริกซ์เป็นค่าประสิทธิภาพสัมพัทธ์เฉลี่ยที่ได้จากการนำค่าฟังก์ชันสารสนเทศ เฉลี่ยของแบบทดสอบของแต่ละวิธีมาเทียบอัตราส่วนระหว่างแต่ละคู่ โดย

$$RAI = \frac{AI_{row}}{AI_{column}}$$

จากตาราง 15 พบว่า สำหรับค่าประสิทธิภาพสัมพัทธ์เฉลี่ย เมื่อเทียบสัดส่วนกันแล้วถ้ามีค่าอัตราส่วนสารสนเทศมากกว่า 1 แสดงว่า วิธีแรกมีประสิทธิภาพสัมพัทธ์เฉลี่ยสูงกว่าวิธีหลัง ในทางตรงกันข้าม ถ้ามีค่าอัตราส่วนสารสนเทศน้อยกว่า 1 แสดงว่า วิธีหลังมีประสิทธิภาพสัมพัทธ์เฉลี่ยสูงกว่าวิธีแรก ซึ่งจากผลการวิเคราะห์ เมื่อเปรียบเทียบระหว่างวิธีประยุกต์การให้คะแนนของคูมบ์และวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ ปรากฏว่าค่าที่ได้เท่ากับ 0.86 แสดงว่า วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์มีประสิทธิภาพสัมพัทธ์เฉลี่ยสูงกว่าวิธีประยุกต์การให้คะแนนของคูมบ์

4.5 ค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าของแบบทดสอบที่วิเคราะห์จากคะแนนที่ได้จากวิธีการตอบและตรวจให้คะแนนทั้ง 2 วิธี แสดงรายละเอียดในตาราง 16

ตาราง 16 ค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าของแบบทดสอบ

ระดับความสามารถ(θ)	วิธีการตรวจให้คะแนน	ค่าฟังก์ชันสารสนเทศของแบบทดสอบ (TIF)	SE (θ)
-2.8	วิธีประยุกต์การให้คะแนนของคูมบ์	2.30	0.66
	วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์	2.25	0.67
-2.4	วิธีประยุกต์การให้คะแนนของคูมบ์	2.75	0.60
	วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์	2.85	0.59
-2.0	วิธีประยุกต์การให้คะแนนของคูมบ์	3.32	0.55
	วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์	3.71	0.52
-1.6	วิธีประยุกต์การให้คะแนนของคูมบ์	3.98	0.50
	วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์	4.87	0.45
-1.2	วิธีประยุกต์การให้คะแนนของคูมบ์	4.68	0.46
	วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์	6.21	0.40
-0.8	วิธีประยุกต์การให้คะแนนของคูมบ์	5.32	0.43
	วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์	7.43	0.37
-0.4	วิธีประยุกต์การให้คะแนนของคูมบ์	5.77	0.42
	วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์	8.14	0.35
0	วิธีประยุกต์การให้คะแนนของคูมบ์	5.94	0.41
	วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์	8.04	0.35
0.4	วิธีประยุกต์การให้คะแนนของคูมบ์	5.77	0.42
	วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์	7.21	0.37
0.8	วิธีประยุกต์การให้คะแนนของคูมบ์	5.32	0.43
	วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์	5.99	0.41

ตาราง 16 (ต่อ)

ระดับความสามารถ(θ)	วิธีการตรวจให้คะแนน	ค่าฟังก์ชันสารสนเทศของแบบทดสอบ (TIF)	SE (θ)
1.2	วิธีประยุกต์การให้คะแนนของคูมบ์	4.68	0.46
	วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์	4.76	0.46
1.6	วิธีประยุกต์การให้คะแนนของคูมบ์	3.99	0.50
	วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์	3.73	0.52
2.0	วิธีประยุกต์การให้คะแนนของคูมบ์	3.36	0.55
	วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์	2.96	0.58
2.4	วิธีประยุกต์การให้คะแนนของคูมบ์	2.83	0.59
	วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์	2.41	0.64
2.8	วิธีประยุกต์การให้คะแนนของคูมบ์	2.41	0.64
	วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์	2.03	0.70

จากตาราง 16 พบว่า ค่าความคลาดเคลื่อนมาตรฐานที่ระดับความสามารถ (θ) เท่ากับ -2.8 และช่วง 1.6 ถึง 2.8 วิธีประยุกต์การให้คะแนนของคูมบ์ให้ค่าความคลาดเคลื่อนมาตรฐานต่ำกว่าวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ ที่ระดับความสามารถ (θ) เท่ากับ -2.4 ถึง 0.8 วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ให้ค่าความคลาดเคลื่อนมาตรฐานต่ำกว่าวิธีประยุกต์การให้คะแนนของคูมบ์ และที่ระดับความสามารถ (θ) เท่ากับ 1.2 ค่าความคลาดเคลื่อนมาตรฐานของวิธีประยุกต์การให้คะแนนของคูมบ์และวิธีการตอบและการตรวจให้คะแนนแบบ อาร์โนลด์ มีค่าเท่ากัน

บทที่ 5
สรุปผล อภิปราย และข้อเสนอแนะ

การวิจัยครั้งนี้ ต้องการศึกษาวิธีการตอบและตรวจให้คะแนนความรู้บางส่วน จากการตรวจให้คะแนนด้วยวิธีประยุกต์การให้คะแนนของคูมบ์ และวิธีการตอบและการตรวจให้คะแนนแบบ อาร์โนลด์ เพื่อศึกษาฟังก์ชันสารสนเทศของข้อสอบ (Item Information Function) และฟังก์ชันสารสนเทศของแบบทดสอบ (Test Information Function) จากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์แบบเลือกตอบ กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2559 สำนักงานเขตพื้นที่การศึกษามัธยมศึกษาเขต 2 จังหวัดกรุงเทพมหานคร จำนวน 10 โรงเรียน จำนวน 26 ห้องเรียน จำนวน 1,200 คน ได้มาโดยการสุ่มแบบหลายขั้นตอน (Multi-stage random sampling) มีขั้นตอนคือ สุ่มโรงเรียน โดยวิธีการสุ่มแบบแบ่งชั้น (Stratified Random Sampling) โดยมีขนาดโรงเรียนเป็นชั้น (Strata) มีโรงเรียนเป็นหน่วยสุ่ม จำแนกโรงเรียนในเขตพื้นที่การศึกษามัธยมศึกษาเขต 2 เป็น 3 ขนาด ได้แก่ ขนาดใหญ่พิเศษ, ขนาดใหญ่ และขนาดกลาง สุ่มโรงเรียนตามสัดส่วนของโรงเรียนแต่ละขนาด ได้โรงเรียนขนาดใหญ่พิเศษ 5 โรงเรียน ได้แก่ บดินทรเดชา(สิงห์ สิงหเสนี), เทพศิรินทร์ร่มเกล้า, บางกะปิ, พรตพิทยพยัต และมัธยมวัดหนองจอก จำนวน 68 ห้องเรียน จำนวน 3,070 คน โรงเรียนขนาดใหญ่ 3 โรงเรียน ได้แก่ รัตนโกสินทร์สมโภช บางเขน, บดินทรเดชา (สิงห์ สิงหเสนี) 4 และลาดปลาเค้าพิทยาคม จำนวน 28 ห้องเรียน จำนวน 1,200 คน โรงเรียนขนาดกลาง 2 โรงเรียน ได้แก่ สุขุมนวพันธ์อุปถัมภ์ และ จันทร์หุ่นบำเพ็ญ จำนวน 12 ห้องเรียน จำนวน 432 คน หลังจากนั้นได้ทำการสุ่มแบบแบ่งกลุ่ม โดยแบ่งห้องเรียนของโรงเรียนแต่ละขนาดเป็นดังนี้ โรงเรียนขนาดใหญ่พิเศษ เป็น 68 ห้องเรียน สุ่มได้มา 16 ห้องเรียน โรงเรียนขนาดใหญ่ จำนวน 28 ห้องเรียน สุ่มได้มา 6 ห้องเรียน โรงเรียนขนาดกลาง จำนวน 12 ห้องเรียน สุ่มได้มา 4 ห้องเรียน รวมทั้งสิ้น จำนวน 26 ห้องเรียน จำนวน 1,200 คน

เครื่องมือที่ใช้ในการวิจัย คือ แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ เรื่อง ตัวหารร่วมมากและตัวคูณร่วมน้อย ชั้นมัธยมศึกษาปีที่ 1 มีวิธีการตรวจให้คะแนน 2 วิธี คือวิธีประยุกต์การให้คะแนนของคูมบ์ (Modified Coombs Approach) และวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ (The Arnold Scoring/Response Method) ซึ่งผู้วิจัยเป็นผู้สร้างขึ้นเป็นแบบทดสอบเลือกตอบชนิด 4 ตัวเลือก โดยมีตัวเลือกที่เป็นคำตอบเพียงตัวเลือกเดียว จำนวน 25 ข้อ โดยวิธีประยุกต์การให้คะแนนของคูมบ์ ค่าความยากง่ายอยู่ระหว่าง 0.27 – 0.92 ค่าอำนาจจำแนกอยู่ระหว่าง 0.04 – 0.45 วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ ค่าความยากง่ายอยู่ระหว่าง 0.02 – 0.89 ค่าอำนาจจำแนกอยู่ระหว่าง 0.07 – 0.63 ตลอดจนพิจารณาการทำหน้าที่ต่างกันของข้อสอบทั้งตัวถูกและตัวลวงร่วมด้วย สำหรับการทดลองนั้นนักเรียนทุกคนได้รับฉบับเดียวกัน แต่แตกต่างกันที่วิธีการตรวจให้คะแนน โดยที่โรงเรียนแต่ละโรงเรียนจะได้รับแบบทดสอบที่มีวิธีการตรวจทั้งวิธีประยุกต์การให้คะแนนของคูมบ์และวิธีการตอบและการตรวจให้คะแนนแบบอาร์

โนลด์ ซึ่งใน 1 ห้องเรียนจะได้รับแบบทดสอบที่มีวิธีการตรวจแบบเดียวกัน ให้นักเรียนทำแบบทดสอบโดยที่นักเรียนสามารถตัดตัวลวงที่เป็นคำตอบที่ผิดได้ 1 หรือ 2 หรือ 3 ตัวเลือก ก็ได้ โดยไม่จำเป็นต้องตัดตัวเลือกทั้ง 3 ตัวในแต่ละข้อ แต่พยายามตัดออกตามความรู้ที่สามารถตัดออกได้เท่านั้น ถ้าไม่สามารถตัดออกได้ก็ไม่ต้องตัดออก

ผู้วิจัยดำเนินการในการทดลองและเก็บรวบรวมข้อมูลดังนี้คือ

ขั้นที่ 1 การดำเนินการสร้างเครื่องมือเป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1

ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับวิชาคณิตศาสตร์และการสร้างแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ สร้างตารางวิเคราะห์เนื้อหา และจุดประสงค์การเรียนรู้เกี่ยวกับเรื่อง ตัวหารร่วมมากและตัวคูณร่วมน้อย ชั้นมัธยมศึกษาปีที่ 1 สร้างแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ เรื่อง ตัวหารร่วมมากและตัวคูณร่วมน้อย ชนิดเลือกตอบ 4 ตัวเลือก จำนวน 50 ข้อ

ขั้นที่ 2 หาคุณภาพของแบบทดสอบ

ตรวจสอบคุณภาพแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนเรื่อง ตัวหารร่วมมากและตัวคูณร่วมน้อยของนักเรียนชั้นมัธยมศึกษาปีที่ 1 ด้านความเที่ยงตรงเชิงเนื้อหา โดยนำแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน ที่ผู้วิจัยสร้างขึ้น ให้ผู้เชี่ยวชาญที่มีความรู้ทางการวัดผลและคณิตศาสตร์ จำนวน 5 ท่าน พิจารณาแบบทดสอบแล้วนำไปตรวจสอบหาความเที่ยงตรงของแบบทดสอบโดยใช้สูตร IOC และคัดเลือกข้อสอบจำนวน 40 ข้อ ผลปรากฏว่า ค่า IOC ที่ได้อยู่ในช่วง 0.6 ถึง 1 นำแบบทดสอบที่ปรับปรุงไปทดลองใช้กับนักเรียนชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2559 จำนวน 200 คน ซึ่งมีลักษณะใกล้เคียงกับกลุ่มตัวอย่าง และนำแบบทดสอบมาตรวจให้คะแนนตามวิธีการให้คะแนนทั้ง 2 วิธี แล้วนำผลการทดสอบมาวิเคราะห์ หาคุณภาพข้อสอบโดยใช้ทฤษฎีการวัดแบบดั้งเดิม (Classical Test Theory: CTT) เพื่อหาค่าความยากง่าย ค่าอำนาจจำแนกของข้อสอบ และค่าความเชื่อมั่น จาก สัมประสิทธิ์แอลฟาของครอนบาค ได้ค่าความเชื่อมั่นของวิธีประยุกต์การให้คะแนนของคูมบ์ เท่ากับ 0.8177 ส่วนวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ได้ค่าความเชื่อมั่น เท่ากับ 0.8272 นำผลการทดสอบมาวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบทั้งตัวถูกและตัวลวงในเรื่องเพศ ปรับปรุงข้อสอบและคัดเลือกข้อสอบที่มีคุณภาพจำนวน 25 ข้อ โดยวิธีประยุกต์การให้คะแนนของคูมบ์ ค่าความยากง่ายอยู่ระหว่าง 0.27 – 0.92 ค่าอำนาจจำแนกอยู่ระหว่าง 0.04 – 0.45 วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ ค่าความยากง่ายอยู่ระหว่าง 0.02 – 0.89 ค่าอำนาจจำแนกอยู่ระหว่าง 0.07 – 0.63 และการทำหน้าที่ต่างกันของข้อสอบทั้งตัวถูกและตัวลวง โดยผู้วิจัยมีหลักเกณฑ์ในการพิจารณาแต่ละข้อเรียงตามลำดับดังนี้ 1) ข้อสอบต้องไม่เกิดการทำหน้าที่ต่างกันของข้อสอบทั้งตัวถูกและตัวลวง (DIFและDDF) 2) ข้อสอบต้องไม่เกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF) 3) ข้อสอบต้องไม่เกิดการทำหน้าที่ต่างกันของตัวลวง (DDF) ทั้งนี้ผู้วิจัยพบข้อสอบที่เกิดการทำหน้าที่ต่างกันของข้อสอบทั้งตัวถูกและตัวลวง แต่การ

คัดเลือก ผู้วิจัยคำนึงถึงโครงสร้างข้อสอบเป็นหลักในการพิจารณา จัดพิมพ์และนำไปใช้กับนักเรียนกลุ่มตัวอย่าง จำนวน 1,200 คน

ขั้นที่ 3 การเก็บรวบรวมข้อมูล

ดำเนินการเก็บรวบรวมข้อมูลการวิจัย ผู้วิจัยหรือตัวแทนผู้วิจัยใช้เวลาในการรวบรวมข้อมูลระหว่าง วันที่ 1 มกราคม 2560 – 28 กุมภาพันธ์ 2560 โดยใช้แบบทดสอบชุดเดียว จำนวน 25 ข้อ

การวิเคราะห์ข้อมูลที่ได้รับจากการสอบของนักเรียนกลุ่มตัวอย่าง ชั้นมัธยมศึกษาปีที่ 1 นำมาตรวจให้คะแนนตามวิธีการตรวจให้คะแนนทั้ง 2 วิธี แล้วนำคะแนนที่ได้ในแบบทดสอบมาทำการวิเคราะห์ข้อมูล ดังนี้

1. วิเคราะห์ค่าสถิติพื้นฐานโดยใช้โปรแกรมคอมพิวเตอร์เพื่อหาค่าสูงสุด ค่าต่ำสุด ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน สัมประสิทธิ์ของการแปรผัน ความเบ้ และความโด่ง ของวิธีการตอบและตรวจให้คะแนนทั้ง 2 วิธี

2. ตรวจสอบการเป็นมิติเดียวของแบบทดสอบด้วยวิธีการวิเคราะห์องค์ประกอบ (Factor Analysis) โดยใช้โปรแกรมคอมพิวเตอร์

3. วิเคราะห์เพื่อหาค่าพารามิเตอร์ของข้อสอบ ค่าฟังก์ชันสารสนเทศของข้อสอบ ค่าฟังก์ชันสารสนเทศของแบบทดสอบ และค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่า วิเคราะห์ข้อมูลโดยใช้ Generalized Partial Credit Model (G-PCM) ด้วยโปรแกรมคอมพิวเตอร์ และนำค่าที่ได้มาคำนวณค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบทดสอบ และค่าประสิทธิภาพสัมพัทธ์เฉลี่ยของแบบทดสอบ

3. เปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบทดสอบ ค่าประสิทธิภาพสัมพัทธ์เฉลี่ยของแบบทดสอบ และค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าของวิธีการตอบและตรวจให้คะแนนทั้ง 2 วิธี คือ วิธีประยุกต์การให้คะแนนของคูมบ์ และวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์

สรุปผลการวิจัย

ผลการวิเคราะห์ค่าฟังก์ชันสารสนเทศของข้อสอบ พบว่า ค่าฟังก์ชันสารสนเทศสูงสุดของข้อสอบเป็นดังนี้ วิธีประยุกต์การให้คะแนนของคูมบ์ให้ค่าฟังก์ชันสารสนเทศของข้อสอบสูงกว่าวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ทั้งหมด 10 ข้อ คิดเป็น 40% ของแบบทดสอบทั้งฉบับ และวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ให้ค่าฟังก์ชันสารสนเทศของข้อสอบสูงกว่าวิธีประยุกต์การให้คะแนนของคูมบ์ทั้งหมด 15 ข้อ คิดเป็น 60% ของแบบทดสอบทั้งฉบับ ซึ่งโดยภาพรวมแล้ว พบว่า วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ให้ค่าฟังก์ชันสารสนเทศสูงสุดของข้อสอบสูงกว่าวิธีประยุกต์การให้คะแนนของคูมบ์ แต่เมื่อพิจารณาค่าฟังก์ชันสารสนเทศ

สูงสุดของข้อสอบแต่ละข้อแล้ว วิธีประยุกต์การให้คะแนนของคูมบ์และวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์จะให้ค่าที่ใกล้เคียงกันมาก

ผลการวิเคราะห์ค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบทดสอบ พบว่า วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์มีค่าเท่ากับ 4.84 ซึ่งให้ค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบทดสอบสูงกว่าวิธีประยุกต์การให้คะแนนของคูมบ์ที่มีค่าเท่ากับ 4.16

เมื่อพิจารณาถึงค่าฟังก์ชันสารสนเทศของแบบทดสอบจำแนกตามระดับความสามารถของผู้สอบ (θ) พบว่า วิธีประยุกต์การให้คะแนนของคูมบ์ให้ค่าฟังก์ชันสารสนเทศของแบบทดสอบสูงกว่าวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ที่ระดับความสามารถ(θ) เท่ากับ -2.8 และช่วงตั้งแต่ 1.6 ถึง 2.8 แสดงว่า วิธีประยุกต์การให้คะแนนของคูมบ์ เหมาะสำหรับผู้ที่มีความสามารถค่อนข้างต่ำและผู้ที่มีความสามารถค่อนข้างสูง ส่วนวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ให้ค่าฟังก์ชันสารสนเทศของแบบทดสอบสูงกว่าวิธีประยุกต์การให้คะแนนของคูมบ์ที่ระดับความสามารถ(θ) ในช่วง -2.4 ถึง 1.2 แสดงว่า วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์เหมาะสำหรับผู้ที่มีความสามารถปานกลาง

ผลการวิเคราะห์ค่าประสิทธิภาพสัมพัทธ์เฉลี่ย พบว่า วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์มีประสิทธิภาพสัมพัทธ์เฉลี่ยสูงกว่าวิธีประยุกต์การให้คะแนนของคูมบ์

ผลการวิเคราะห์ค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าของแบบทดสอบ พบว่า ค่าความคลาดเคลื่อนมาต่รฐานที่ระดับความสามารถ (θ) เท่ากับ -2.8 และช่วง 1.6 ถึง 2.8 วิธีประยุกต์การให้คะแนนของคูมบ์ให้ค่าความคลาดเคลื่อนมาตรฐานต่ำกว่าวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ ที่ระดับความสามารถ (θ) เท่ากับ -2.4 ถึง 0.8 วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ให้ค่าความคลาดเคลื่อนมาตรฐานต่ำกว่าวิธีประยุกต์การให้คะแนนของคูมบ์ และที่ระดับความสามารถ (θ) เท่ากับ 1.2 ค่าความคลาดเคลื่อนมาตรฐานของวิธีประยุกต์การให้คะแนนของคูมบ์และวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ มีค่าเท่ากัน

อภิปรายผลการวิจัย

คุณภาพของวิธีการตอบและตรวจให้คะแนนเกี่ยวกับฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบทดสอบ ค่าประสิทธิภาพสัมพัทธ์เฉลี่ยของแบบทดสอบ และค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่า ผลการวิจัยพบว่า ให้หลักฐานที่ยอมรับสมมติฐานการวิจัยที่กำหนดให้ค่าฟังก์ชันสารสนเทศของข้อสอบ และฟังก์ชันสารสนเทศของแบบทดสอบ ระหว่างการตรวจให้คะแนนความรู้บางส่วนด้วยวิธีประยุกต์การให้คะแนนของคูมบ์และวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์แตกต่างกัน โดยผลปรากฏว่า วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ มีค่าฟังก์ชันสารสนเทศของข้อสอบ ค่าฟังก์ชันสารสนเทศของแบบทดสอบ ค่าประสิทธิภาพสัมพัทธ์เฉลี่ยของแบบทดสอบสูงกว่าวิธีประยุกต์การให้คะแนนของคูมบ์ และวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์มีค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าของ

แบบทดสอบต่ำกว่าวิธีประยุกต์การให้คะแนนของคูมบ์ ซึ่งจากการวิเคราะห์นั้นไม่สอดคล้องกับข้อสรุปของงานวิจัยของ สุพจน์ เกิดสุวรรณ์. (2545: 151) ที่ได้ข้อสรุปว่า วิธีประยุกต์การให้คะแนนของคูมบ์ ให้ค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบทดสอบสูงที่สุด รองลงมาคือวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ และวิธีที่ประยุกต์จากวิธีของเดรสเซลและชมิด และวิธีที่ให้ค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบทดสอบต่ำสุด คือ วิธีการตรวจให้คะแนนแบบประเพณีนิยม และจากงานวิจัยของ เอมอร จังศิริพรปกรณ์ (2545: 83) ได้ข้อสรุปว่า วิธีประยุกต์การให้คะแนนของคูมบ์ มีค่าฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบทดสอบ และอัตราส่วนสารสนเทศเฉลี่ยสูงที่สุด รองลงมา คือ วิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท ส่วนวิธีประเพณีนิยมมีค่าต่ำที่สุด และจากงานวิจัยของ David A. Bradbard, Darrell F. Parker and Gary L. Stone (2004: 11-26) ได้ข้อสรุปว่า วิธีการตรวจให้คะแนนด้วยวิธีการตัดตัวเลือกแบบคูมบ์สูงกว่าวิธีการเลือกตอบแบบธรรมดาทั่วไปหรือวิธีการแบบประเพณีนิยม จะเห็นได้ว่าจากงานวิจัยอื่นๆนั้นวิธีประยุกต์การให้คะแนนของคูมบ์ จะให้ค่าฟังก์ชันสารสนเทศของข้อสอบและแบบทดสอบสูงที่สุด ซึ่งตรงกันข้ามกับงานวิจัยที่ผู้วิจัยได้ทำการวิเคราะห์คือวิธีประยุกต์การให้คะแนนของคูมบ์จะให้ค่าฟังก์ชันสารสนเทศของข้อสอบและแบบทดสอบต่ำซึ่งต่ำกว่าวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ แต่ข้อค้นพบในครั้งนี้ได้มีความสอดคล้องกับงานวิจัยของ รัตนา ไชยตรี (2546: 125) ที่ได้ข้อสรุปว่า วิธีการตอบและตรวจให้คะแนนความรู้บางส่วนตามวิธีการตอบโดยบอกระดับความมั่นใจส่วนใหญ่จะให้ค่าฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบทดสอบ และอัตราส่วนสารสนเทศสูงสุด รองลงมาคือ วิธีการตอบและตรวจให้คะแนนวิธีที่ประยุกต์การให้คะแนนของเดรสเซลและชมิดท์ และวิธีที่ประยุกต์การให้คะแนนของคูมบ์ตามลำดับ ทำให้พบว่าน่าจะมีเหตุผลบางประการที่ทำให้วิธีที่ประยุกต์การให้คะแนนของคูมบ์ให้ค่าฟังก์ชันสารสนเทศของข้อสอบและแบบทดสอบต่ำ อาจจะเป็นผลสืบเนื่องมาจากขั้นตอนในการคัดเลือกข้อสอบที่นำมาใช้ยังมีความบกพร่อง เพราะมีข้อสอบที่ค่าอำนาจจำแนก และค่าความยากง่ายของข้อสอบไม่เป็นไปตามเกณฑ์ที่กำหนด ที่ได้กำหนดไว้ว่า ค่าอำนาจจำแนกควรอยู่ระหว่าง 0.50 ถึง 2.50 และค่าความยากง่ายควรอยู่ระหว่าง -2.50 ถึง 2.50 (ศิริชัย กาญจนวาสี. 2555: 55) ซึ่งในงานวิจัยฉบับนี้มีค่าอำนาจจำแนกและค่าความยากง่าย ค่อนข้างต่ำ และสิ่งที่เป็นข้อสังเกตเพิ่มเติมในครั้งนี้คือ รายวิชาที่นำมาสร้างเป็นแบบทดสอบน่าจะมีผลต่อฟังก์ชันสารสนเทศของข้อสอบและแบบทดสอบด้วยส่วนหนึ่ง ซึ่งจากงานวิจัยครั้งนี้ผู้วิจัยได้สร้างแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 เช่นเดียวกับงานของ รัตนา ไชยตรี (2546: 65) ที่ได้สร้างแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 3 จึงส่งผลให้ผลการวิเคราะห์ค่าฟังก์ชันสารสนเทศของข้อสอบ และฟังก์ชันสารสนเทศของแบบทดสอบเป็นไปในทิศทางเดียวกันคือวิธีประยุกต์การให้คะแนนของคูมบ์ให้ค่าต่ำกว่าวิธีอื่นๆที่ได้ทำการศึกษา

ในการพิจารณาค่าฟังก์ชันสารสนเทศของข้อสอบ และฟังก์ชันสารสนเทศของแบบทดสอบหากพิจารณาถึงช่วงของระดับความสามารถนั้น ปรากฏว่า วิธีประยุกต์การให้คะแนนของคูมบ์ให้ค่าฟังก์ชันสารสนเทศของแบบทดสอบสูงกว่าวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ที่ระดับ

ความสามารถ(θ) เท่ากับ -2.8 และช่วงตั้งแต่ 1.6 ถึง 2.8 แสดงว่า วิธีประยุกต์การให้คะแนนของคูมบ์ เหมาะสำหรับผู้ที่มีความสามารถค่อนข้างต่ำและผู้ที่มีความสามารถค่อนข้างสูง ส่วนวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ให้ค่าฟังก์ชันสารสนเทศของแบบทดสอบสูงกว่าวิธีประยุกต์การให้คะแนนของคูมบ์ที่ระดับความสามารถ(θ) ในช่วง -2.4 ถึง 1.2 แสดงว่า วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์เหมาะสำหรับผู้ที่มีความสามารถปานกลาง ดังนั้นจากการวิเคราะห์จึงสามารถอธิบายด้วยเหตุผลที่ว่า วิธีการตอบและตรวจให้คะแนนทั้ง 2 วิธีนั้น มีความเหมาะสมกับระดับความสามารถ(θ)ของผู้ตอบในลักษณะที่แตกต่างกัน โดยกลุ่มผู้ที่มีความสามารถค่อนข้างต่ำและสูงวิธีการตรวจให้คะแนนที่เหมาะสม คือ วิธีประยุกต์การให้คะแนนของคูมบ์ กลุ่มผู้ที่มีความสามารถปานกลาง วิธีการตรวจให้คะแนนที่เหมาะสม คือ วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ จะเห็นได้ว่าวิธีการตรวจให้คะแนนความรู้บางส่วนนั้นมีความเหมาะสมกับกลุ่มผู้ที่มีความสามารถที่แตกต่างกันไป เพราะฉะนั้นในการเลือกวิธีการให้คะแนนความรู้บางส่วนจึงควรให้ความสำคัญกับระดับความสามารถของกลุ่มผู้ตอบเพื่อให้ได้สารสนเทศที่เป็นประโยชน์และเกิดประโยชน์กับผู้ตอบมากที่สุด

ข้อเสนอแนะ

ข้อเสนอแนะสำหรับการวิจัยครั้งต่อไป

1. ควรศึกษาและทำการทดลองกับกลุ่มตัวอย่างในระดับชั้นอื่น ในรายวิชาคณิตศาสตร์
2. ใช้แบบทดสอบแบบเลือกตอบที่มากกว่า 4 ตัวเลือก เพื่อทำให้ผู้ตอบได้ใช้ความรู้บางส่วนมากยิ่งขึ้น
3. ในการเลือกวิธีการให้คะแนนความรู้บางส่วนเพื่อนำมาทำการทดลองนั้น ควรพิจารณาถึงระดับชั้นหรือช่วงอายุของผู้ตอบด้วย เพื่อทำให้ได้ผลที่ตรงกับความสามารถที่แท้จริงของผู้ตอบมากที่สุด

ข้อเสนอแนะสำหรับการพัฒนาการศึกษา

1. ครูผู้สอนควรนำแบบทดสอบไปใช้ให้เหมาะสมกับระดับความสามารถของนักเรียน
2. การนำแบบทดสอบวิธีการตัดตัวลวงไปใช้ต้องระวังเกี่ยวกับคำชี้แจงในการทำแบบทดสอบ เนื่องจากเป็นวิธีการทำที่นักเรียนไม่คุ้นเคย อาจเกิดข้อผิดพลาดได้ง่าย

บรรณานุกรม

บรรณานุกรม

กระทรวงศึกษาธิการ. (2551). *หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551.* กรุงเทพฯ: สำนักพิมพ์ชุมนุมสหกรณ์การเกษตรแห่งประเทศไทย.

ชมนาด เชื้อสุวรรณทวี. (2544). *การสอนคณิตศาสตร์.* มหาวิทยาลัยศรีนครินทรวิโรฒ.

ชวาล แพรัตกุล. (2552). *เทคนิคการวัดผล.* พิมพ์ครั้งที่ 7. กรุงเทพฯ: วิฑูรย์การปก.

ดิลก ดิลกานนท์. (2557). *เอกสารประกอบการอบรม การวัดและประเมินทางการศึกษา.* กรุงเทพฯ: สำนักทดสอบทางการศึกษาและจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ.

นิอร ไชยพรพัฒนา. (2554). *การพัฒนามาตรวัดแบบการคิดพหุมิติตามทฤษฎีจิตในการปกครองตนเอง สำหรับนักเรียนชั้นประถมศึกษาปีที่ 6.* วิทยานิพนธ์ ค.ด. (วิจัยพัฒนาหลักสูตร). กรุงเทพฯ: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.

บุญเชิด ภิญโญอนันตพงษ์. (2545). *คุณภาพของเครื่องมือวัด หน่วยที่ 3 ประมวลสาระชุดวิชาการพัฒนาเครื่องมือสำหรับการประเมินการศึกษา.* นนทบุรี: มหาวิทยาลัยสุโขทัยธรรมาธิราช.

ประพนธ์ จ่ายเจริญ. (2547). *เอกสารคำสอนวิชา ปถ 522 ปัญหาและกลวิธีการสอนคณิตศาสตร์ในโรงเรียนประถมศึกษา.* กรุงเทพฯ: ภาคหลักสูตรและการสอน คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.

พรทิพย์ ไชยโส. (2533). การพัฒนาสูตรการให้คะแนนแบบเลือกตอบสำหรับความรู้บางส่วนของผู้ตอบ: การประยุกต์ใช้วิธีการอาร์โนลด์และวิธีการของแฮมดาน. วิทยานิพนธ์ ค.ด. (วิจัยและพัฒนาหลักสูตร). กรุงเทพฯ: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.

พวงรัตน์ ทวีรัตน์. (2540). *การสร้างและการพัฒนาแบบทดสอบวัดผลสัมฤทธิ์.* กรุงเทพฯ: สำนักทดสอบทางการศึกษาและจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ.

เยาวดี รางชัยกุล วิบูลย์ศรี. (2553). *การวัดผลและการสร้างแบบสอบผลสัมฤทธิ์.* พิมพ์ครั้งที่ 9. กรุงเทพฯ: สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.

มหาวิทยาลัยสุโขทัยธรรมาธิราช. (2545). *การพัฒนาเครื่องมือสำหรับการประเมินการศึกษา.* นนทบุรี: สำนักพิมพ์มหาวิทยาลัยสุโขทัยธรรมาธิราช.

------------. (2549). *การพัฒนาแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน.* พิมพ์ครั้งที่ 7. นนทบุรี: สำนักพิมพ์มหาวิทยาลัยสุโขทัยธรรมาธิราช.

รัตนา ไชยตรี. (2546). *การเปรียบเทียบคุณภาพและอัตราความคลาดเคลื่อนของการกำหนดเกรดแบบอิงกลุ่ม และอิงเกณฑ์ เมื่อใช้แบบสอบเลือกตอบที่มีการตอบ และการตรวจให้คะแนนความรู้บางส่วนด้วยวิธีแตกต่างกัน.* วิทยานิพนธ์ ค.ม. (วิจัยและพัฒนาหลักสูตร). กรุงเทพฯ: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.

เรืองเดช ศิริกิจ. (2554). *การวิเคราะห์เปรียบเทียบโมเดลการประเมินคุณภาพการจัดการศึกษาวิชาคณิตศาสตร์: การประยุกต์ใช้โมเดลมูลค่าเพิ่มที่มีการวิเคราะห์การทำ หน้าที่ต่างกันของข้อสอบและการวิเคราะห์การทำหน้าที่ต่างกันของตัวลวง.* วิทยานิพนธ์ ค.ด. (วิจัยและพัฒนาหลักสูตร). กรุงเทพฯ: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

ล้วน สายยศ; และอังคณา สายยศ. (2543). *เทคนิคการวัดผลการเรียนรู้.* กรุงเทพฯ: ชมรมเด็ก.

ศิริชัย กาญจนวาสี. (2555). *ทฤษฎีการทดสอบแนวใหม่.* กรุงเทพฯ: สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.

------------. (2556). *ทฤษฎีการทดสอบแบบดั้งเดิม.* กรุงเทพฯ: สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.

สุพจน์ เกิดสุวรรณ์. (2545). *การพัฒนาการวัดความรู้บางส่วนของผู้ตอบแบบสอบเลือกตอบ.* กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.

สุเทพ สันติวรานนท์. (2533). *การกำหนดลักษณะเฉพาะของข้อสอบ.* สงขลา: มหาวิทยาลัยสงขลานครินทร์.

------------. (2534). *ปัญหาการสร้างแบบทดสอบของครูที่สอนชั้นมัธยมศึกษา เขตการศึกษา 2.* สงขลา: มหาวิทยาลัยสงขลานครินทร์.

สุรศักดิ์ อมรรัตนศักดิ์. (2544). *ทฤษฎีทางการทดสอบ.* กรุงเทพฯ: คณะศึกษาศาสตร์มหาวิทยาลัยรามคำแหง.

สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี กระทรวงศึกษาธิการ. (2555). *ทักษะและกระบวนการทางคณิตศาสตร์.* กรุงเทพ ฯ: 3-คิว มีเดีย.

สมจวน มานะก่อ. (2551). *การศึกษาเปรียบเทียบคุณภาพของแบบทดสอบเลือกตอบเมื่อ ตรวจด้วยวิธีการให้คะแนนความรู้บางส่วนของแบบทดสอบวัดผลสัมฤทธิ์คณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 3.* ปริญญานิพนธ์ กศ.ม. (การมัธยมศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัยมหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน. (2559). *ข้อมูลพื้นฐานโรงเรียนสังกัดสพม. เขต 2.* สืบค้นเมื่อ 6 ธันวาคม 2559. จากระบบสารสนเทศเพื่อบริหารการศึกษา: http://data.bopp-obec.info/emis/school.php?Area_CODE=101702.

อัมพร ม้าคนอง. (2553). *ทักษะและกระบานการทางคณิตศาสตร์.* กรุงเทพฯ: ศูนย์ตำราและเอกสารทางวิชาการ คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

อุทัยวรรณ สายพัฒนะ. (2547). *การเปรียบเทียบประสิทธิภาพของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่าระหว่างวิธี GMH และวิธี Polytomous SIBTES.* ปริญญานิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

เอมอร จังศิริพรปกรณ์. (2545ก). รายงานการวิจัย การเปรียบเทียบคุณภาพของแบบสอบเลือกตอบเมื่อตรวจด้วยวิธีการให้คะแนนความรู้บางส่วนกับวิธีประเพณีนิยม. กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.

------------. (2545ข). การให้คะแนนความรู้บางส่วนทำได้อย่างไร. กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.

Allen, M.J.; & Yen, W.M. (1979). Introduction to Measurement Theory. Monterey: Califormia Brooks/Cole.

Camilli, Gregory; & Lorrie A Shepard. (1994). Methodes of Identifying Biased Test Items. California: Sage Publications.

David A. Bradbard, Darrell F. Parker; & Gary L. Stone. (2004). An Alternate Multiple-Choice Scoring Procedure in a Macroeconomics Course. *Journal of Innovative Education* .11-26.

Frary, R.B. (1980). The effect of misinformation, partial information, and guessing on expected multiple-choice test item scores. *Applied Psychological Measurement*. 4: 79-90.

Green B.F., Crone C R.; & Folk V.G. (1989). A Method for studying differential distractor functioning. *Journal of Educational Measurement*. 26(2): 147-160.

Hee-Sun Lee, Ou Lydia Liu; & Marcia C. Linn. (2011). Validating Measurement of Knowledge Integration in Science Using Multiple-Choice and Explanation Items. *Applied Measurement in Education*. 24: 115-136.

Kato K., Moen R.E.; & Thurlow M.L. (2009). Differentials of a state reading assessment: item functioning, distractor functioning, and omission frequency for disability categories. *Educational Measurement: Issues and Practice*. 28(2): 28-40

Penfield R.D. (2010). *DDFS 1.0, Differential Distractor Functioning Software User's Manual.* Unpublished manuscript

Popham, W.J. (1981). *Modern educational Measurment*. Englewood Cliffs, New Jersey: Prentice-Hall.

Tsung-Han Ho; & Barbara G. Dodd. (2012). Item Selection and Ability Estimation Procedures for a Mixed-Format Adaptive Test. *Applied Measurement in Education*. 25: 305-326.

Yuan H. Li and Robert W. Lissitz. (2004). Applications of the Analytically Derived Asymptotic Standard Errors of Item Response Theory Item Parameter Estimates. *Journal of Educational Measurement*. 85-117.

ภาคผนวก

ภาคผนวก ก

คุณภาพของเครื่องมือที่ใช้ในการวิจัย

ตาราง 17 ผลการตรวจสอบความเที่ยงตรงด้านเนื้อหา(Content Validity) โดยผู้เชี่ยวชาญ

ข้อ	ผู้เชี่ยวชาญคนที่					IOC	การคัดเลือก
	1	2	3	4	5		
1	1	1	1	0	1	0.8	คัดเลือก
2	1	1	1	1	1	1	คัดเลือก
3	1	1	1	0	1	0.8	คัดเลือก
4	1	0	1	1	1	0.8	คัดเลือก
5	1	1	1	0	1	0.8	คัดเลือก
6	1	1	1	0	1	0.8	คัดเลือก
7	1	0	0	1	1	0.6	คัดเลือก
8	1	1	1	1	1	1	ไม่ใช้
9	1	1	1	1	1	1	คัดเลือก
10	1	1	1	1	1	1	คัดเลือก
11	1	1	1	1	1	1	คัดเลือก
12	1	1	1	0	0	0.6	คัดเลือก
13	1	1	1	1	1	1	คัดเลือก
14	1	1	1	0	0	0.6	ไม่ใช้
15	1	1	1	1	1	1	คัดเลือก
16	1	1	1	1	1	1	คัดเลือก
17	1	1	1	1	1	1	คัดเลือก
18	1	1	-1	0	1	0.4	ไม่ใช้
19	1	1	0	0	1	0.6	ไม่ใช้
20	1	1	0	0	1	0.6	ไม่ใช้
21	1	1	1	1	1	1	คัดเลือก
22	1	1	1	0	1	0.8	ไม่ใช้
23	1	1	1	1	1	1	ไม่ใช้
24	1	1	1	1	1	1	คัดเลือก
25	1	1	1	1	1	1	คัดเลือก

ตาราง 17 (ต่อ)

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

ตาราง 18 ค่าความยากง่าย (p) ค่าอำนาจจำแนกรายข้อ (r) การทำหน้าที่ต่างกันของข้อสอบทั้งตัวถูกและตัวลวง

ข้อ	วิธีประยุกต์การให้คะแนนของคูมบ์		วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์		การทำหน้าที่ต่างกันของข้อสอบ		ผลการคัดเลือก
	p	r	p	r	ตัวถูก	ตัวลวง	
1	0.92	0.09	0.88	0.14	—	1	ไม่ใช้
2	0.92	0.04	0.89	0.07	—	—	คัดเลือก
3	0.58	0.45	0.43	0.61	—	—	คัดเลือก
4	0.89	0.11	0.85	0.16	—	—	คัดเลือก
5	0.74	0.32	0.65	0.44	—	—	คัดเลือก
6	0.74	0.29	0.65	0.41	—	—	คัดเลือก
7	0.43	0.19	0.22	0.23	√	—	ไม่ใช้
8	0.91	0.10	0.87	0.16	—	—	คัดเลือก
9	0.64	0.29	0.51	0.40	—	—	คัดเลือก
10	0.47	0.31	0.28	0.44	—	—	คัดเลือก
11	0.42	0.15	0.22	0.28	—	—	คัดเลือก
12	0.51	0.37	0.34	0.51	—	—	คัดเลือก
13	0.77	0.18	0.66	0.28	√	—	ไม่ใช้
14	0.44	0.35	0.25	0.53	√	1	ไม่ใช้
15	0.57	0.25	0.44	0.33	—	—	คัดเลือก
16	0.61	0.35	0.47	0.48	—	—	คัดเลือก
17	0.54	0.39	0.38	0.58	—	1	ไม่ใช้
18	0.47	0.41	0.29	0.59	—	—	คัดเลือก
19	0.28	0.14	0.04	0.20	—	—	คัดเลือก
20	0.27	0.05	0.02	0.08	—	—	คัดเลือก
21	0.70	0.28	0.58	0.44	√	—	ไม่ใช้
22	0.19	0.33	-0.08	0.44	—	—	ไม่ใช้
23	0.26	0.20	0.01	0.29	—	—	ไม่ใช้

ตาราง 18 (ต่อ)

ข้อ	วิธีประยุกต์การให้คะแนนของคูมบ์		วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์		การทำหน้าที่ต่างกันของข้อสอบ		ผลการคัดเลือก
	p	r	p	r	ตัวถูก	ตัวลวง	
24	0.37	0.17	0.16	0.20	—	—	ไม่ใช้
25	0.46	0.27	0.27	0.35	—	—	คัดเลือก
26	0.35	0.28	0.13	0.42	—	—	คัดเลือก
27	0.77	0.23	0.69	0.31	√	—	คัดเลือก
28	0.61	0.45	0.48	0.63	—	—	คัดเลือก
29	0.71	0.37	0.61	0.50	—	—	คัดเลือก
30	0.48	0.39	0.28	0.52	—	—	คัดเลือก
31	0.47	0.19	0.28	0.28	—	—	ไม่ใช้
32	0.48	0.31	0.30	0.46	—	—	คัดเลือก
33	0.21	0.03	-0.07	0.06	√	2	ไม่ใช้
34	0.60	0.25	0.45	0.30	—	—	คัดเลือก
35	0.37	0.19	0.14	0.28	—	—	ไม่ใช้
36	0.47	-0.04	0.28	-0.04	√	—	ไม่ใช้
37	0.46	0.18	0.27	0.23	—	—	คัดเลือก
38	0.42	0.31	0.21	0.43	—	—	ไม่ใช้
39	0.33	-0.11	0.09	-0.09	—	—	ไม่ใช้
40	0.59	0.23	0.43	0.35	—	—	คัดเลือก

Reliability

วิธีประยุกต์การให้คะแนนของคูมบ์

****** Method 1 (space saver) will be used for this analysis ******

R E L I A B I L I T Y   A N A L Y S I S   -   S C A L E   (A L P H A)

Reliability Coefficients

N of Cases =   100.0   N of Items = 40

Alpha =   .8177

วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์

****** Method 1 (space saver) will be used for this analysis ******

R E L I A B I L I T Y   A N A L Y S I S   -   S C A L E   (A L P H A)

Reliability Coefficients

N of Cases =   100.0   N of Items = 40

Alpha =   .8272

ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบทั้งตัวถูกและตัวลวง โดยใช้โปรแกรม DDFS 1.0

Opened the text file: d:\DDFs\ct1.txt

Number of Cases: 101

Number of Variables: 41

Differential Distractor Functioning Analysis

Stratifying variable: Sum of correct responses

Stratum size: 1

Number of strata: 18

Number of reference group members: 50

Number of focal group members: 50

Grouping variable: Var1

Reference Value = 1, Focal Value = 2

Variable: 2 Correct Option: 4

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	-0.7802	0.5122	-1.5232	4.129	.	.
Correct	.	.	.	.	30	36
1	.	.	.	.	1	2
2	0.7073	0.8445	0.8375	.	3	6
3	-2.3877	1.0405	-2.2948	.	16	6

Variable: 3 Correct Option: 4

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	-0.4118	0.8095	-0.5087	0.006	.	.
Correct	.	.	.	.	43	45
1	-1.2078	1.5595	-0.7745	.	3	1
2	0.5261	1.669	0.3152	.	2	2
3	-0.2068	1.1125	-0.1859	.	2	2

Variable: 4 Correct Option: 3

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	-0.1981	0.5019	-0.3947	0.116	.	.
Correct	.	.	.	.	20	26
1	-0.6433	0.6546	-0.9827	.	17	13
2	-0.1448	0.5788	-0.2502	.	11	11
4	.	.	.	.	2	0

Variable: 5 Correct Option: 4

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	-0.2451	0.6726	-0.3644	2.454	.	.
Correct	.	.	.	.	42	43
1	.	.	.	.	1	4
2	.	.	.	.	6	2
3	.	.	.	.	1	1

Variable: 6 Correct Option: 3

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	-0.6156	0.517	-1.1907	0.86	.	.
Correct	.	.	.	.	32	37
1	-1.4486	0.7988	-1.8135	.	10	5
2	-0.5662	0.8126	-0.6968	.	4	4
4	0.8089	1.1412	0.7088	.	4	4

Variable: 7 Correct Option: 2

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	0.1781	0.4694	0.3794	0.659	.	.
Correct	.	.	.	.	33	30
1	0.2525	0.6054	0.4171	.	9	12
3	-0.2797	0.6811	-0.4107	.	8	5
4	.	.	.	.	0	3

Variable: 8 Correct Option: 1

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	1.7768	0.5597	3.1746	2.37	.	.
Correct	.	.	.	.	36	20
2	1.0013	0.7905	1.2667	.	6	8
3	.	.	.	.	3	13
4	1.2521	0.7226	1.7328	.	5	9

Variable: 9 Correct Option: 2

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	-1.4157	1.236	-1.1454	0.509	.	.
Correct	.	.	.	.	44	49
1	.	.	.	.	1	0
3	.	.	.	.	2	1
4	.	.	.	.	3	0

Variable: 10 Correct Option: 4

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	0.8385	0.4994	1.679	0.103	.	.
Correct	.	.	.	.	23	16
1	0.4055	0.7065	0.574	.	7	9
2	1.2976	0.8168	1.5886	.	10	12
3	0.8999	0.649	1.3866	.	10	13

Variable: 11 Correct Option: 3

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	-0.4372	0.471	-0.9282	0.411	.	.
Correct	.	.	.	.	16	20
1	-0.3627	0.5569	-0.6513	.	19	21
2	1.3132	0.9382	1.3997	.	2	9
4	.	.	.	.	13	0

Variable: 12 Correct Option: 2

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC

DIF	0.5775	0.4563	1.2656	0.013	.	.
Correct	.	.	.	.	25	15
1	0.7849	0.5051	1.5539	.	17	25
3	-0.0541	0.8067	-0.0671	.	7	6
4	1.4441	1.3057	1.106	.	1	4

Variable: 13 Correct Option: 3

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	0.4461	0.5531	0.8065	0.093	.	.
Correct	.	.	.	.	14	9
1	0.5615	0.6187	0.9075	.	24	25
2	0.5889	0.6811	0.8646	.	10	16
4	.	.	.	.	2	0

Variable: 14 Correct Option: 3

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	-1.4194	0.532	-2.668	0.002	.	.
Correct	.	.	.	.	24	34
1	-1.0342	0.5573	-1.8557	.	18	15
2	.	.	.	.	7	1
4	.	.	.	.	1	0

Variable: 15 Correct Option: 4

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC

DIF	-1.3638	0.6627	-2.0579	0.011	.	.
Correct	.	.	.	.	9	13
1	-1.3652	1.1251	-1.2134	.	7	6
2	-1.5349	0.7353	-2.0874	.	26	21
3	-1.0589	1.0852	-0.9758	.	8	10

Variable: 16 Correct Option: 2

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	0.0422	0.4832	0.0873	0.454	.	.
Correct	.	.	.	.	19	17
1	-1.3264	0.8443	-1.571	.	11	5
3	0.8318	0.6576	1.2649	.	8	16
4	0.04	0.7849	0.051	.	12	12

Variable: 17 Correct Option: 1

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	1.0407	0.5813	1.7903	0.469	.	.
Correct	.	.	.	.	17	8
2	0.9706	0.6672	1.4547	.	12	17
3	0.5438	0.766	0.7099	.	16	12
4	1.9021	0.9623	1.9766	.	5	13

Variable: 18 Correct Option: 2

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	0.7766	0.5407	1.4363	0.305	.	.

Correct	.	.	.	.	16	11
1	1.354	0.8601	1.5742	.	11	15
3	1.6404	0.771	2.1276	.	10	22
4	-1.8192	1.0755	-1.6915	.	13	2

Variable: 19 Correct Option: 1

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	0.0259	0.6039	0.0429	0.228	.	.
Correct	.	.	.	.	9	8
2	-0.574	0.7515	-0.7638	.	17	16
3	0.4888	0.7174	0.6813	.	10	16
4	0.7557	0.8387	0.901	.	14	10

Variable: 20 Correct Option: 4

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	0.5859	0.6205	0.9442	0.185	.	.
Correct	.	.	.	.	11	10
1	0.5024	0.6772	0.7419	.	18	21
2	1.5732	1.1828	1.3301	.	8	12
3	0.1809	0.9187	0.1969	.	13	7

Variable: 21 Correct Option: 3

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	-0.6911	0.5437	-1.2711	0.001	.	.
Correct	.	.	.	.	12	17

1	-1.2274	0.7678	-1.5986	.	10	6
2	-0.8281	0.5961	-1.3892	.	27	22
4	1.0986	1.2958	0.8478	.	1	5

Variable: 22 Correct Option: 4

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	-1.4505	0.5092	-2.8486	0.233	.	.
Correct	.	.	.	.	14	28
1	-1.2707	0.6894	-1.8432	.	11	6
2	-1.1991	0.663	-1.8086	.	15	13
3	-1.5658	0.7919	-1.9773	.	10	3

Variable: 23 Correct Option: 1

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	1.7085	0.8875	1.9251	0.022	.	.
Correct	.	.	.	.	11	5
2	2.3979	1.4845	1.6153	.	11	15
3	1.0116	1.4213	0.7117	.	12	8
4	1.7188	0.9584	1.7934	.	16	22

Variable: 24 Correct Option: 2

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	-0.0433	0.5219	-0.083	0.036	.	.
Correct	.	.	.	.	17	17
1	0.7299	0.6809	1.072	.	14	19

3	0.0541	0.8038	0.0673	.	9	10
4	-0.954	0.7638	-1.249	.	10	4

Variable: 25 Correct Option: 3

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	-0.4094	0.6045	-0.6773	0.563	.	.
Correct	.	.	.	.	9	14
1	-1.3942	0.8708	-1.6011	.	9	4
2	0.0273	0.6636	0.0411	.	24	28
4	.	.	.	.	8	4

Variable: 26 Correct Option: 3

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	0.2316	0.4228	0.5478	3.775	.	.
Correct	.	.	.	.	22	17
1	0.352	0.625	0.5632	.	7	7
2	0.5897	0.5094	1.1576	.	14	18
4	0.2647	0.6695	0.3954	.	7	8

Variable: 27 Correct Option: 2

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	-0.3994	0.477	-0.8373	3.65	.	.
Correct	.	.	.	.	12	18
1	-0.5921	0.7295	-0.8117	.	10	7
3	-0.4778	0.5772	-0.8278	.	18	16

4	-0.5066	0.7842	-0.646	.	10	9

Variable: 28 Correct Option: 3

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	-1.0316	0.4941	-2.0878	0.014	.	.
Correct	.	.	.	.	18	34
1	-1.4505	0.8466	-1.7133	.	10	4
2	-0.5567	0.5737	-0.9704	.	13	10
4	.	.	.	.	9	2

Variable: 29 Correct Option: 4

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	-0.6933	0.4751	-1.4593	8.502	.	.
Correct	.	.	.	.	13	22
1	-0.7932	0.6333	-1.2525	.	11	8
2	-0.4692	0.567	-0.8275	.	13	15
3	-0.9167	0.7542	-1.2155	.	13	5

Variable: 30 Correct Option: 1

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	-0.9625	0.5096	-1.8887	0.085	.	.
Correct	.	.	.	.	17	29
2	-1.7583	1.0149	-1.7325	.	13	2
3	-0.4426	0.7867	-0.5626	.	7	5
4	-0.6299	0.5903	-1.0671	.	13	14

Variable: 31 Correct Option: 3

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	-0.2706	0.4574	-0.5916	0.766	.	.
Correct	.	.	.	.	18	23
1	-0.0495	0.6669	-0.0742	.	8	8
2	-0.009	0.5638	-0.016	.	12	14
4	-0.8438	0.7626	-1.1065	.	12	5

Variable: 32 Correct Option: 4

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	0.8617	0.6423	1.3416	0.227	.	.
Correct	.	.	.	.	13	6
1	1.128	0.6808	1.6569	.	18	30
2	-0.5208	0.9514	-0.5474	.	11	6
3	0.4055	0.9154	0.443	.	8	8

Variable: 33 Correct Option: 3

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	-0.5578	0.5704	-0.9779	0.028	.	.
Correct	.	.	.	.	10	13
1	-0.7646	0.8855	-0.8635	.	13	7
2	-0.5551	0.6176	-0.8988	.	22	26
4	-0.7569	0.9031	-0.8381	.	5	4

Variable: 34 Correct Option: 4

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	1.6915	0.7099	2.3827	0.955	.	.
Correct	.	.	.	.	13	6
1	0.619	1.025	0.6039	.	11	6
2	1.8412	0.7623	2.4153	.	11	19
3	2.032	0.915	2.2208	.	15	19

Variable: 35 Correct Option: 3

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	-0.2483	0.4873	-0.5095	0.538	.	.
Correct	.	.	.	.	20	21
1	-0.5609	0.614	-0.9135	.	15	12
2	0.0078	0.6833	0.0114	.	8	10
4	0.4488	0.7732	0.5804	.	7	7

Variable: 36 Correct Option: 1

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	0.6906	0.4822	1.4322	1.1	.	.
Correct	.	.	.	.	19	17
2	0.6786	0.6029	1.1256	.	16	13
3	0.2323	0.5992	0.3877	.	13	15
4	.	.	.	.	2	5

Variable: 37 Correct Option: 2

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	1.0686	0.4776	2.2374	5.011	.	.
Correct	.	.	.	.	28	16
1	1.0354	0.8359	1.2387	.	10	12
3	1.1327	0.5965	1.8989	.	8	14
4	1.0602	0.7555	1.4033	.	4	8

Variable: 38 Correct Option: 2

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	0.6776	0.4957	1.367	0.245	.	.
Correct	.	.	.	.	18	11
1	1.1278	0.6155	1.8323	.	10	22
3	0.6273	0.742	0.8454	.	9	7
4	0.5023	0.6036	0.8322	.	13	10

Variable: 39 Correct Option: 4

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	0.3187	0.5451	0.5847	0.586	.	.
Correct	.	.	.	.	15	8
1	0.7179	0.6884	1.0429	.	11	18
2	0.1736	0.6899	0.2516	.	9	11
3	-0.9946	1.089	-0.9133	.	15	13

Variable: 40 Correct Option: 4

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	0.4233	0.5816	0.7278	0.207	.	.
Correct	.	.	.	.	12	8
1	0.9163	1.037	0.8836	.	7	7
2	0.6638	0.7167	0.9262	.	17	19
3	0.4239	0.6783	0.6249	.	14	16

Variable: 41 Correct Option: 3

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	-0.1509	0.4525	-0.3335	3.777	.	.
Correct	.	.	.	.	26	23
1	-0.2823	0.6125	-0.4609	.	12	9
2	-0.9555	0.801	-1.1929	.	6	4
4	0.6122	0.6694	0.9146	.	6	14

3	0.0541	0.8038	0.0673	.	9	10
4	-0.954	0.7638	-1.249	.	10	4

Variable: 25 Correct Option: 3

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	-0.4094	0.6045	-0.6773	0.563	.	.
Correct	.	.	.	.	9	14
1	-1.3942	0.8708	-1.6011	.	9	4
2	0.0273	0.6636	0.0411	.	24	28
4	.	.	.	.	8	4

Variable: 26 Correct Option: 3

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	0.2316	0.4228	0.5478	3.775	.	.
Correct	.	.	.	.	22	17
1	0.352	0.625	0.5632	.	7	7
2	0.5897	0.5094	1.1576	.	14	18
4	0.2647	0.6695	0.3954	.	7	8

Variable: 27 Correct Option: 2

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	-0.3994	0.477	-0.8373	3.65	.	.
Correct	.	.	.	.	12	18
1	-0.5921	0.7295	-0.8117	.	10	7
3	-0.4778	0.5772	-0.8278	.	18	16

4	-0.5066	0.7842	-0.646	.	10	9

Variable: 28 Correct Option: 3

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	-1.0316	0.4941	-2.0878	0.014	.	.
Correct	.	.	.	.	18	34
1	-1.4505	0.8466	-1.7133	.	10	4
2	-0.5567	0.5737	-0.9704	.	13	10
4	.	.	.	.	9	2

Variable: 29 Correct Option: 4

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	-0.6933	0.4751	-1.4593	8.502	.	.
Correct	.	.	.	.	13	22
1	-0.7932	0.6333	-1.2525	.	11	8
2	-0.4692	0.567	-0.8275	.	13	15
3	-0.9167	0.7542	-1.2155	.	13	5

Variable: 30 Correct Option: 1

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	-0.9625	0.5096	-1.8887	0.085	.	.
Correct	.	.	.	.	17	29
2	-1.7583	1.0149	-1.7325	.	13	2
3	-0.4426	0.7867	-0.5626	.	7	5
4	-0.6299	0.5903	-1.0671	.	13	14

Variable: 31 Correct Option: 3

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	-0.2706	0.4574	-0.5916	0.766	.	.
Correct	.	.	.	.	18	23
1	-0.0495	0.6669	-0.0742	.	8	8
2	-0.009	0.5638	-0.016	.	12	14
4	-0.8438	0.7626	-1.1065	.	12	5

Variable: 32 Correct Option: 4

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	0.8617	0.6423	1.3416	0.227	.	.
Correct	.	.	.	.	13	6
1	1.128	0.6808	1.6569	.	18	30
2	-0.5208	0.9514	-0.5474	.	11	6
3	0.4055	0.9154	0.443	.	8	8

Variable: 33 Correct Option: 3

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	-0.5578	0.5704	-0.9779	0.028	.	.
Correct	.	.	.	.	10	13
1	-0.7646	0.8855	-0.8635	.	13	7
2	-0.5551	0.6176	-0.8988	.	22	26
4	-0.7569	0.9031	-0.8381	.	5	4

Variable: 34 Correct Option: 4

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	1.6915	0.7099	2.3827	0.955	.	.
Correct	.	.	.	.	13	6
1	0.619	1.025	0.6039	.	11	6
2	1.8412	0.7623	2.4153	.	11	19
3	2.032	0.915	2.2208	.	15	19

Variable: 35 Correct Option: 3

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	-0.2483	0.4873	-0.5095	0.538	.	.
Correct	.	.	.	.	20	21
1	-0.5609	0.614	-0.9135	.	15	12
2	0.0078	0.6833	0.0114	.	8	10
4	0.4488	0.7732	0.5804	.	7	7

Variable: 36 Correct Option: 1

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	0.6906	0.4822	1.4322	1.1	.	.
Correct	.	.	.	.	19	17
2	0.6786	0.6029	1.1256	.	16	13
3	0.2323	0.5992	0.3877	.	13	15
4	.	.	.	.	2	5

Variable: 37 Correct Option: 2

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	1.0686	0.4776	2.2374	5.011	.	.
Correct	.	.	.	.	28	16
1	1.0354	0.8359	1.2387	.	10	12
3	1.1327	0.5965	1.8989	.	8	14
4	1.0602	0.7555	1.4033	.	4	8

Variable: 38 Correct Option: 2

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	0.6776	0.4957	1.367	0.245	.	.
Correct	.	.	.	.	18	11
1	1.1278	0.6155	1.8323	.	10	22
3	0.6273	0.742	0.8454	.	9	7
4	0.5023	0.6036	0.8322	.	13	10

Variable: 39 Correct Option: 4

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	0.3187	0.5451	0.5847	0.586	.	.
Correct	.	.	.	.	15	8
1	0.7179	0.6884	1.0429	.	11	18
2	0.1736	0.6899	0.2516	.	9	11
3	-0.9946	1.089	-0.9133	.	15	13

Variable: 40 Correct Option: 4

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	0.4233	0.5816	0.7278	0.207	.	.
Correct	.	.	.	.	12	8
1	0.9163	1.037	0.8836	.	7	7
2	0.6638	0.7167	0.9262	.	17	19
3	0.4239	0.6783	0.6249	.	14	16

Variable: 41 Correct Option: 3

Option	LOR	SE(LOR)	Z(LOR)	BD	N-REF	N-FOC
DIF	-0.1509	0.4525	-0.3335	3.777	.	.
Correct	.	.	.	.	26	23
1	-0.2823	0.6125	-0.4609	.	12	9
2	-0.9555	0.801	-1.1929	.	6	4
4	0.6122	0.6694	0.9146	.	6	14

ภาคผนวก ข

ตารางวิเคราะห์เนื้อหาและตัวอย่างเครื่องมือที่ใช้ในการวิจัย

ตาราง 19 ตารางวิเคราะห์เนื้อหาวิชาคณิตศาสตร์ เรื่องตัวหารร่วมมากและตัวคูณร่วมน้อย

เรื่อง	เนื้อหาและจุดประสงค์การเรียนรู้	จำนวนข้อสอบ	ความรู้ความจำ	ความเข้าใจ	การนำไปใช้	การวิเคราะห์
1.	การแยกตัวประกอบของจำนวนนับและจำนวนเฉพาะ					
	1.1 จำแนกจำนวนซึ่งเป็นจำนวนเฉพาะและไม่เป็นจำนวนเฉพาะ	2		2		
	1.2 หาตัวประกอบ ตัวประกอบเฉพาะ และตัวประกอบร่วมของจำนวนใดๆ	7		6		1
2.	การหาตัวหารร่วมมากของจำนวนนับและการแก้ปัญหา					
	2.1 บอกความหมายของตัวหารร่วมมาก	1	1			
	2.2 หาตัวหารร่วมมากโดยวิธีพิจารณาจากตัวประกอบร่วม วิธีตั้งหาร และวิธีแยกตัวประกอบ	1				1
	2.3 หาตัวหารร่วมมากในกรณีเศษเหลือ	2		2		
	2.4 แก้โจทย์ปัญหาเกี่ยวกับตัวหารร่วมมาก	3			3	
3.	การหาตัวคูณร่วมน้อยของจำนวนนับและการแก้ปัญหา					
	3.1 บอกความหมายของตัวคูณร่วมน้อย	1	1			
	3.2 หาตัวคูณร่วมน้อยโดยพิจารณาจากพหุคูณร่วม วิธีตั้งหาร และวิธีแยกตัวประกอบ	4		4		
	3.3 แก้โจทย์ปัญหาเกี่ยวกับตัวคูณร่วมน้อย	2			2	
4.	การประยุกต์การหาตัวหารร่วมมากและการหาตัวคูณร่วมน้อยของจำนวนนับ					
	4.1 หาตัวหารร่วมมากหรือตัวคูณร่วมน้อยโดยใช้ความสัมพันธ์ของตัวหารร่วมมากและตัวคูณร่วมน้อย	2	1	1		
รวม		25	3	15	5	2

ตาราง 20 ตารางแสดงจำนวนข้อสอบในแบบทดสอบวิชาคณิตศาสตร์ เรื่องตัวหารร่วมมากและตัวคูณร่วมน้อย

เรื่อง	เนื้อหา	จุดประสงค์การเรียนรู้	จำนวนข้อสอบ				
			จำนวนที่สร้าง		จำนวนที่ใช้จริง	Try Out (40ข้อ)	ฉบับจริง(25ข้อ)
1.	การแยกตัวประกอบของจำนวนนับและจำนวนเฉพาะ	1.1 จำแนกจำนวนซึ่งเป็นจำนวนเฉพาะและไม่เป็นจำนวนเฉพาะ	12	2	9	2	2
		1.2 หาตัวประกอบ ตัวประกอบเฉพาะและตัวประกอบร่วมของจำนวนใดๆ		10		9	7
2.	การหาตัวหารร่วมมากของจำนวนนับและการแก้ปัญหา	2.1 บอกความหมายของตัวหารร่วมมาก	22	2	7	1	1
		2.2 หาตัวหารร่วมมากโดยวิธีพิจารณาจากตัวประกอบร่วม วิธีตั้งหาร และวิธีแยกตัวประกอบ		8		3	1
		2.3 หาตัวหารร่วมมากในกรณีเศษเหลือ		3		3	2
		2.4 แก้โจทย์ปัญหาเกี่ยวกับตัวหารร่วมมาก		9		7	3
3.	การหาตัวคูณร่วมน้อยของจำนวนนับและการแก้ปัญหา	3.1 บอกความหมายของตัวคูณร่วมน้อย	9	1	7	1	1
		3.2 หาตัวคูณร่วมน้อยโดยพิจารณาจากพหุคูณร่วม วิธีตั้งหาร และวิธีแยกตัวประกอบ		5		5	4
		3.3 แก้โจทย์ปัญหาเกี่ยวกับตัวคูณร่วมน้อย		3		3	2
4.	การประยุกต์การหาตัวหารร่วมมาก ตัวณร่วมน้อย การแยกตัวประกอบ และจำนวนเฉพาะ	4.1 หาตัวหารร่วมมากหรือตัวคูณร่วมน้อยโดยใช้ความสัมพันธ์ของตัวหารร่วมมากและตัวคูณร่วมน้อย	7	7	2	6	2

เครื่องมือที่ใช้ในการวิจัย

ตัวอย่างคำชี้แจงในการตอบและตรวจให้คะแนนแบบทดสอบ

1. ตัวอย่างคำชี้แจงวิธีประยุกต์การให้คะแนนของคูมบ์

แบบทดสอบ

เรื่อง ตัวหารร่วมมากและตัวคูณร่วมน้อย

ชั้นมัธยมศึกษาปีที่ 1

คำอธิบายวิธีทำแบบทดสอบ

1. แบบทดสอบฉบับนี้มีคำถามทั้งหมด 25 ข้อ ให้เวลาทำ 60 นาที

2. คำถามในแบบทดสอบฉบับนี้เป็นแบบเลือกตอบ ชนิด 4 ตัวเลือก มีตัวเลือกถูกเพียง 1 ตัวเลือก ให้นักเรียนเลือก<u>ตัวเลือกที่ผิด</u> โดยทำเครื่องหมาย x ลงในกระดาษคำตอบที่แนบมาพร้อมนี้

3. ตัวอย่างข้อสอบ วิธีการตอบ และการตรวจให้คะแนน

ตัวอย่างข้อ (0) ห.ร.ม. ของ 36, 52 คือข้อใด

ก. 2 (ข.) 4

ค. 12 ง. 468

จากข้อ (0) มีวิธีทำดังนี้

2	36	52
2	18	26
	9	13

ดังนั้น ห.ร.ม. คือ 2 x 2 = 4

ตอบ (ข.) 4

3.1 นักเรียนได้ **3 คะแนน**เมื่อเลือกคำตอบที่ผิด ครบ 3 ตัวเลือก ดังตัวอย่าง

ข้อ	ก	ข	ค	ง
(0)	x		x	x

3.2 นักเรียนได้ **2 คะแนน**เมื่อเลือกคำตอบที่ผิด **2** ตัวเลือก ดังตัวอย่าง

ข้อ	ก	ข	ค	ง
(0)	x		x	

3.3 นักเรียนได้ **1 คะแนน**เมื่อเลือกคำตอบที่ผิด **1** ตัวเลือก ดังตัวอย่าง

ข้อ	ก	ข	ค	ง
(0)	x			

3.4 นักเรียนได้ **0 คะแนน**เมื่อเลือกคำตอบที่ถูก **หรือ ไม่เลือกคำตอบใดเลย** ดังตัวอย่าง

ข้อ	ก	ข	ค	ง
(0)		x	x	

***(เลือกคำตอบที่ถูก)**

ข้อ	ก	ข	ค	ง
(0)				

****(ไม่เลือกคำตอบใดเลย)**

5. **ห้ามขีดเขียนหรือทำเครื่องหมายใด ๆในแบบทดสอบ ฉบับนี้ ถ้าต้องการทดให้ทดลงในกระดาษที่กำหนดให้**

6. ห้ามใช้เครื่องคิดเลขหรืออุปกรณ์ในการคำนวณทุกชนิด

2. ตัวอย่างคำชี้แจงวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์

แบบทดสอบ

เรื่อง ตัวหารร่วมมากและตัวคูณร่วมน้อย

ชั้นมัธยมศึกษาปีที่ 1

คำอธิบายวิธีทำแบบทดสอบ

1. แบบทดสอบฉบับนี้มีคำถามทั้งหมด 25 ข้อ ให้เวลาทำ 60 นาที

2. คำถามในแบบทดสอบฉบับนี้เป็นแบบเลือกตอบ ชนิด 4 ตัวเลือก มีตัวเลือกถูกเพียง 1 ตัวเลือก ให้นักเรียนเลือก<u>ตัวเลือกที่ผิด</u> โดยทำเครื่องหมาย x ลงในกระดาษคำตอบที่แนบมาพร้อมนี้

3. ตัวอย่างข้อสอบ วิธีการตอบ และการตรวจให้คะแนน

ตัวอย่างข้อ (0) ห.ร.ม. ของ 36, 52 คือข้อใด

ก. 2 (ข.) 4

ค. 12 ง. 468

จากข้อ (0) มีวิธีทำดังนี้

2) 36 52

2) 18 26

9 13

ดังนั้น ห.ร.ม. คือ 2 x 2 = 4

ตอบ (ข.) 4

3.1 นักเรียนได้ 1 คะแนนเมื่อเลือกคำตอบที่ผิด ครบ 3 ตัวเลือก ดังตัวอย่าง

ข้อ	ก	ข	ค	ง
(0)	x		x	x

3.2 นักเรียนได้ 1/3 คะแนนเมื่อเลือกคำตอบที่ผิด 2 ตัวเลือก ดังตัวอย่าง

ข้อ	ก	ข	ค	ง
(0)	x		x	

3.3 นักเรียนได้ 1/9 คะแนนเมื่อเลือกคำตอบที่ผิด 1 ตัวเลือก ดังตัวอย่าง

ข้อ	ก	ข	ค	ง
(0)	x			

3.4 นักเรียนได้ -1/3 คะแนนเมื่อเลือกคำตอบที่ถูก หรือ ไม่เลือกคำตอบใดเลย ดังตัวอย่าง

ข้อ	ก	ข	ค	ง
(0)		x	x	

*(เลือกคำตอบที่ถูก)

ข้อ	ก	ข	ค	ง
(0)				

**(ไม่เลือกคำตอบใดเลย)

5. **ห้ามขีดเขียนหรือทำเครื่องหมายใด ๆในแบบทดสอบฉบับนี้ ถ้าต้องการทดให้ทดลงในกระดาษที่กำหนดให้**

6. ห้ามใช้เครื่องคิดเลขหรืออุปกรณ์ในการคำนวณทุกชนิด

ตัวอย่างแบบทดสอบที่ใช้ในการวิจัย

แบบทดสอบ
เรื่อง ตัวหารร่วมมากและตัวคูณร่วมน้อย
ชั้นมัธยมศึกษาปีที่ 1

1. ข้อใด**ไม่ใช่**ตัวประกอบของ 18

ก. 1, 9, 18 ข. 1, 3, 6 ค. 2, 3, 6 ง. 4, 8, 10

2. ตัวประกอบของ 24 มีทั้งหมดกี่จำนวน

ก. 6 จำนวน ข. 7 จำนวน ค. 8 จำนวน ง. 9 จำนวน

3. 150 แยกตัวประกอบได้ดังข้อใด

ก. 3 x 5 x 10 ข. 5 x 5 x 6 ค. 2 x 5 x 15 ง. 2 x 3 x 5 x 5

4. จำนวนนับที่**น้อยที่สุด**ที่มี 1, 2, 3, 4, 6 เป็นตัวประกอบคือข้อใด

ก. 6 ข. 8 ค. 12 ง. 24

5. ข้อใดเป็นจำนวนนับที่**น้อยที่สุด** ที่มี 1, 2, 3, 4, 6, 8, 12 เป็นตัวประกอบ

ก. 12 ข. 24 ค. 48 ง. 72

6. ข้อใดแยกตัวประกอบได้ถูกต้อง

ก. $18 = 3 \times 6$ ข. $24 = 2 \times 2 \times 2 \times 3$ ค. $28 = 4 \times 7$ ง. $36 = 3 \times 3 \times 4$

7. จำนวนใด**ไม่เป็น**จำนวนเฉพาะ

ก. 37 ข. 43 ค. 59 ง. 91

8. จำนวนที่กำหนดให้ในข้อใดต่อไปนี้เป็นจำนวนเฉพาะทั้งชุด

ก. 1, 2, 5, 7 ข. 3, 19, 21, 31 ค. 23, 41, 53, 97 ง. 1, 29, 67, 83

9. พิจารณาข้อความต่อไปนี้

 1. ตัวประกอบร่วมของ 24 และ 36 คือ 1, 2, 3, 4, 6, 12
 2. จำนวนเฉพาะของ 38 คือ 1, 2, 19

ข้อใดต่อไปนี้สรุปถูกต้อง

ก. 1 และ 2 ถูก ข. 1 ถูก และ 2 ผิด ค. 1 ผิด และ 2 ถูก ง. 1 และ 2 ผิด

10. ข้อใดให้ความหมายของ " 3 เป็น ห.ร.ม. ของ 15 และ 24" ได้ถูกต้อง

ก. 3 เป็นตัวหารร่วมของ 15 และ 24

ข. 3 เป็นตัวประกอบร่วมที่มีค่าน้อยที่สุดของ 15 และ 24

ค. 3 เป็นจำนวนที่มากที่สุดที่หาร 15 และ 24 แล้วลงตัว

ง. 3 เป็นจำนวนที่มากที่สุดที่มี 15 และ 24 เป็นตัวประกอบ

11. ข้อใดเป็นจำนวนที่มากที่สุดที่นำไปหาร 254 แล้วเหลือเศษ 2 และเมื่อนำไปหาร247 แล้วเหลือเศษ 4

ก. 3 ข. 9 ค. 27 ง. 28

12. ถ้า B เป็นจำนวนที่มากที่สุดที่หาร 29, 41 และ 69 แล้วเหลือเศษ 1 ทุกจำนวน ข้อใดต่อไปนี้ถูกต้อง

ก. $B + 3 = 7$ ข. $4 - B = 2$ ค. $4 \times B = 8$ ง. $4 + B = 6$

13. กำหนดให้

$$A = 2 \times 3 \times 3 \times a$$
$$B = 2 \times b \times 5 \times 7$$
$$C = c \times 3 \times 5 \times 11$$

ถ้า ห.ร.ม. ของ A, B และ C เท่ากับ 30 แล้ว $(b+c)-a$ มีค่าเท่ากับเท่าใด

ก. 0 ข. 2 ค. 3 ง. 5

14. มีมังคุด 52 ผล เงาะ 164 ผล แบ่งเป็นกองๆ ละเท่าๆ กัน ให้ได้ผลไม้มากที่สุด แต่ละกองจะมีผลไม้กี่ผล

ก. 4 ผล ข. 13 ผล ค. 41 ผล ง. 54 ผล

15. มีลูกอมรสกาแฟ 35 เม็ด รสช็อกโกแลต 42 เม็ด และรสวานิลา 49 เม็ด ต้องการแบ่งใส่ถุง โดยให้แต่ละถุงมีจำนวนเท่ากันและรสเดียวกันทั้งหมด ต้องใช้ถุง<u>**น้อยที่สุด**</u>กี่ถุง

ก. 5 ถุง ข. 7 ถุง ค. 18 ถุง ง. 210 ถุง

16. ต้องการล้อมรั้วบ้านเป็นรูปสี่เหลี่ยมที่มีความกว้าง 100 เมตร และความยาว 120 เมตร ต้องปักเสาให้มีระยะห่างกันเท่าใดจึงจะใช้เสาจำนวน**น้อยที่สุด**และใช้เสาทั้งหมดกี่ต้น

ก. 5 เมตร, 88 ต้น ข. 6 เมตร, 77 ต้น ค. 20 เมตร, 22 ต้น ง.600 เมตร,4 ต้น

17. ข้อใดให้ความหมายของ " 24 เป็น ค.ร.น. ของ 8 และ 12" ได้ถูกต้อง

ก. 24 เป็นจำนวนที่มากที่สุดที่ 8 และ 12 หารลงตัว

ข. 24 เป็นจำนวนที่น้อยที่สุดที่ 8 และ 12 หารลงตัว

ค. 24 เป็นจำนวนที่ 8 และ 12 หารลงตัว

ง. 24 เป็นตัวคูณร่วมของ 8 และ 12

18. จำนวนใด**ไม่เป็น**พหุคูณของ 8

ก. 24 ข. 40 ค. 58 ง. 96

19. ข้อใดคือ ค.ร.น. ของ 27 และ 54

ก. 3 ข. 9 ค. 27 ง. 54

20. ข้อใดคือ ค.ร.น. ของ 8, 40 และ 56

ก. 280 ข. 56 ค. 40 ง. 8

21. 600 เป็นตัวคูณร่วมน้อยของจำนวนชุดใด

ก. 100 และ 150 ข. 135 และ 200 ค. 120 และ 200 ง. 100 และ 135

22. หนังสือจำนวนหนึ่งสามารถนำมาใส่กล่องครั้งละ 48 เล่ม หรือ 54 เล่ม หรือ 60 เล่ม ได้โดยไม่เหลือเศษ มีหนังสืออย่างน้อยกี่เล่ม

ก. 60 ข. 162 ค. 2,160 ง. 155,520

23. สวนสนุกแห่งหนึ่งเปิดให้คนเล่นเครื่องเล่น 3 ชนิดพร้อมกัน เวลา 10.00 น เครื่องเล่นชนิดที่ 1, 2 และ 3 ใช้เวลาต่อ 1 รอบเป็น 6, 9 และ 15 นาที ตามลำดับ อยากทราบว่าเครื่องเล่นทั้ง 3 ชนิดนี้จะเปิดพร้อมกันอีกในเวลาเท่าใด

ก. 10.30 น. ข. 11.00 น. ค. 11.30 น. ง. 12.00 น.

24. ค่า ค.ร.น. และ ห.ร.ม. ของ 15, 21 และ 36 **ต่างกัน**อยู่เท่าใด

ก. 420 ข. 1,257 ค. 1,263 ง. 3,780

25. ข้อใดถูกต้อง

ก. 1 เป็นตัวคูณร่วมน้อยของทุกจำนวน

ข. 1 เป็นตัวหารร่วมมากของทุกจำนวน

ค. 1 เป็นตัวประกอบของทุกจำนวน

ง. 1 เป็นตัวประกอบเฉพาะของทุกจำนวน

ภาคผนวก ค

ผลการวิเคราะห์ข้อมูล

ผลการวิเคราะห์ข้อมูลโดยใช้โปรแกรม SPSS
การวิเคราะห์ค่าสถิติพื้นฐาน

วิธีประยุกต์การให้คะแนนของคูมบ์

ตาราง 21 ค่าสถิติพื้นฐานของข้อมูลเมื่อตรวจด้วยวิธีประยุกต์การให้คะแนนของคูมบ์

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std.	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
SUM	599	2	72	33.88	13.527	.717	.100	.184	.199
Valid N (listwise)	599								

วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์

ตาราง 22 ค่าสถิติพื้นฐานของข้อมูลเมื่อตรวจด้วยวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std.	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
SUM	599	-5.16	23.67	7.5610	6.52864	.425	.100	-.617	.199
Valid N (listwise)	599								

ผลการวิเคราะห์ข้อมูลโดยใช้โปรแกรม SPSS
การวิเคราะห์องค์ประกอบเพื่อตรวจสอบความเป็นเอกมิติของข้อมูล

วิธีประยุกต์การให้คะแนนของคูมบ์

ตาราง 23 ค่าไอเกน และร้อยละของความแปรปรวนขององค์ประกอบของแบบทดสอบเมื่อตรวจด้วยวิธีประยุกต์การให้คะแนนของคูมบ์

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.259	17.036	17.036	4.259	17.036	17.036	2.121	8.483	8.483
2	1.531	6.125	23.161	1.531	6.125	23.161	1.774	7.097	15.579
3	1.379	5.518	28.679	1.379	5.518	28.679	1.727	6.907	22.486
4	1.217	4.867	33.546	1.217	4.867	33.546	1.653	6.614	29.100
5	1.137	4.547	38.093	1.137	4.547	38.093	1.628	6.511	35.611
6	1.089	4.355	42.448	1.089	4.355	42.448	1.356	5.424	41.035
7	1.051	4.203	46.650	1.051	4.203	46.650	1.261	5.044	46.078
8	1.020	4.080	50.730	1.020	4.080	50.730	1.163	4.651	50.730
9	.960	3.841	54.570						
10	.939	3.754	58.324						
11	.903	3.610	61.934						
12	.889	3.555	65.490						
13	.877	3.507	68.997						
14	.854	3.415	72.411						
15	.828	3.311	75.722						
16	.775	3.098	78.820						
17	.758	3.031	81.851						
18	.690	2.758	84.609						
19	.639	2.556	87.165						
20	.634	2.537	89.702						
21	.593	2.371	92.073						
22	.575	2.301	94.374						
23	.494	1.977	96.351						
24	.491	1.964	98.315						
25	.421	1.685	100.000						

Extraction Method: Principal Component Analysis.

วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์

ตาราง 24 ค่าไอเกน และร้อยละของความแปรปรวนขององค์ประกอบของแบบทดสอบเมื่อตรวจด้วยวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.950	19.800	19.800	4.950	19.800	19.800	3.297	13.188	13.188
2	1.918	7.673	27.474	1.918	7.673	27.474	2.284	9.137	22.325
3	1.487	5.947	33.420	1.487	5.947	33.420	2.143	8.573	30.898
4	1.241	4.963	38.383	1.241	4.963	38.383	1.486	5.944	36.842
5	1.157	4.630	43.013	1.157	4.630	43.013	1.452	5.807	42.648
6	1.097	4.387	47.400	1.097	4.387	47.400	1.188	4.752	47.400
7	.982	3.929	51.329						
8	.972	3.890	55.219						
9	.912	3.649	58.867						
10	.861	3.443	62.310						
11	.813	3.251	65.561						
12	.780	3.118	68.679						
13	.764	3.058	71.737						
14	.735	2.941	74.678						
15	.694	2.776	77.454						
16	.677	2.709	80.163						
17	.666	2.663	82.827						
18	.630	2.521	85.348						
19	.621	2.486	87.833						
20	.569	2.276	90.109						
21	.556	2.226	92.335						
22	.529	2.117	94.452						
23	.481	1.926	96.378						
24	.459	1.834	98.212						
25	.447	1.788	100.000						

Extraction Method: Principal Component Analysis.

ผลการวิเคราะห์ข้อมูลโดยใช้โปรแกรม IRTPRO Version 4.0
การวิเคราะห์ค่าพารามิเตอร์ของข้อสอบ
วิธีประยุกต์การให้คะแนนของคูมบ์

IRTPRO Version 4.0

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

ตาราง 25 GPCM Model Item Parameter Estimates, logit: $a[k(\theta - b) + \Sigma d_k]$

Item	a	s.e.	b	s.e.
1	0.33	0.05	-0.76	0.13
2	0.38	0.05	0.13	0.09
3	0.44	0.06	-0.61	0.1
4	0.39	0.05	-0.57	0.1
5	0.49	0.07	-0.55	0.09
6	0.37	0.06	-1.38	0.19
7	0.24	0.04	0.64	0.17
8	0.17	0.04	1.84	0.42
9	0.12	0.03	0.98	0.36
10	0.28	0.04	0.9	0.16
11	0.25	0.04	0.21	0.13
12	0.3	0.04	0.46	0.13
13	-0.01	0.04	-63.4	357.73
14	0.16	0.04	2.47	0.58
15	0.16	0.04	1.87	0.44
16	0.22	0.04	0.52	0.17
17	0.39	0.05	0.68	0.11
18	0.3	0.04	-0.25	0.11
19	0.48	0.06	0.53	0.09
20	0.46	0.06	0.38	0.09
21	0.24	0.04	0.64	0.17
22	0.36	0.05	0.72	0.13
23	0.39	0.05	0.14	0.09
24	0.34	0.05	0.34	0.11
25	0.36	0.05	0.05	0.1

ตาราง 26 ค่าพารามิเตอร์ระดับความยากของขั้นการตอบ (δ_{ij}) ที่วิเคราะห์จากคะแนนที่ได้จากวิธีประยุกต์การให้คะแนนของคูมบ์ โดยวิเคราะห์ตามโมเดล GPCM

Item	δ_1	s.e.	δ_2	s.e.	δ_3	s.e.
1	0.25	0.04	-4.29	0.37	-0.73	0.37
2	-0.05	0.03	-3.47	0.25	-0.56	0.24
3	0.27	0.04	-2.59	0.18	-0.01	0.17
4	0.22	0.04	-2.65	0.18	-0.37	0.17
5	0.27	0.04	-2.47	0.17	0.08	0.16
6	0.5	0.05	-2.64	0.2	-0.41	0.19
7	-0.15	0.03	-2.53	0.17	-0.66	0.16
8	-0.31	0.03	-2.65	0.17	-0.19	0.17
9	-0.12	0.03	-2.64	0.17	-0.41	0.16
10	-0.25	0.03	-2.39	0.16	-0.38	0.15
11	-0.05	0.03	-2.44	0.16	-0.46	0.15
12	-0.14	0.03	-2.08	0.15	-0.49	0.14
13	-0.42	0.03	-2.28	0.16	-0.49	0.15
14	-0.4	0.04	-2.14	0.15	-0.47	0.14
15	-0.3	0.03	-2.3	0.16	-0.55	0.15
16	-0.11	0.03	-2.18	0.15	-0.56	0.14
17	-0.27	0.04	-2.34	0.17	-0.52	0.16
18	0.08	0.03	-2.82	0.19	-0.35	0.18
19	-0.25	0.04	-2.56	0.18	-0.53	0.17
20	-0.18	0.04	-2.8	0.2	-0.61	0.19
21	-0.15	0.03	-1.79	0.13	-0.43	0.12
22	-0.26	0.04	-2.18	0.15	-0.32	0.14
23	-0.06	0.04	-1.96	0.14	-0.27	0.13
24	-0.11	0.03	-1.95	0.14	-0.25	0.13
25	-0.02	0.03	-2.03	0.14	-0.39	0.13

วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์

IRTPRO Version 4.0

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

ตาราง 27 GPCM Model Item Parameter Estimates, logit: $a[k(\theta - b) + \sum d_k]$

Item	*a*	s.e.	*b*	s.e.
1	0.43	0.05	-0.43	0.08
2	0.52	0.06	-0.4	0.07
3	0.36	0.05	-0.66	0.11
4	0.35	0.05	-0.81	0.12
5	0.59	0.08	-1	0.1
6	-0.06	0.03	1.2	0.73
7	0.2	0.04	0.54	0.18
8	0.16	0.03	0.96	0.27
9	0.37	0.05	0.14	0.09
10	0.22	0.04	0.64	0.17
11	0.53	0.06	-0.05	0.07
12	0.47	0.05	0.18	0.08
13	-0.08	0.04	-5.95	2.95
14	0.2	0.04	1.78	0.34
15	0.12	0.03	1.81	0.58
16	0.48	0.05	0.19	0.08
17	0.57	0.07	-0.22	0.07
18	0.28	0.04	-0.21	0.11
19	0.35	0.04	0.16	0.1
20	0.03	0.03	5.78	5.59
21	0.34	0.04	0.41	0.11
22	0.49	0.06	0.04	0.08
23	0.52	0.06	0.04	0.08
24	0.42	0.05	0.22	0.09
25	0.57	0.07	-0.41	0.07

ตาราง 28 ค่าพารามิเตอร์ระดับความยากของขั้นการตอบ (δ_{ij}) ที่วิเคราะห์จากคะแนนที่ได้จากวิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์ โดยวิเคราะห์ตามโมเดล GPCM

Item	δ_1	s.e.	δ_2	s.e.	δ_3	s.e.
1	0.19	0.04	-4.24	0.43	-1.26	0.43
2	0.21	0.04	-3.69	0.3	-0.71	0.29
3	0.24	0.04	-3.03	0.22	-0.55	0.21
4	0.28	0.04	-2.95	0.21	-0.38	0.2
5	0.59	0.06	-2.51	0.19	-0.02	0.18
6	0.08	0.03	-3.14	0.22	-0.76	0.22
7	-0.1	0.03	-2.41	0.16	-0.58	0.15
8	-0.15	0.03	-2.93	0.19	-0.55	0.19
9	-0.05	0.03	-2.77	0.18	-0.29	0.18
10	-0.14	0.03	-2.74	0.18	-0.41	0.17
11	0.03	0.04	-2.38	0.18	-0.81	0.17
12	-0.08	0.04	-2.18	0.16	-0.4	0.14
13	-0.48	0.04	-2.56	0.19	-0.79	0.18
14	-0.35	0.03	-2.73	0.19	-0.5	0.18
15	-0.21	0.03	-2.14	0.15	-0.48	0.14
16	-0.09	0.04	-2.51	0.18	-0.68	0.17
17	0.12	0.04	-2.19	0.16	-0.37	0.15
18	0.06	0.03	-3.27	0.22	-0.48	0.22
19	-0.06	0.03	-2.75	0.19	-0.53	0.18
20	-0.2	0.03	-2.41	0.17	-0.75	0.16
21	-0.14	0.03	-2.43	0.17	-0.76	0.17
22	-0.02	0.04	-2.32	0.16	-0.12	0.15
23	-0.02	0.04	-2.04	0.15	-0.43	0.14
24	-0.09	0.04	-2.11	0.15	-0.25	0.14
25	0.23	0.04	-2.59	0.18	0.09	0.17

ผลการวิเคราะห์ข้อมูลโดยใช้โปรแกรม IRTPRO Version 4.0
การวิเคราะห์ค่าฟังก์ชันสารสนเทศ
วิธีประยุกต์การให้คะแนนของคูมบ์

ตาราง 29 Item Information Function Values for Group **1** at 15 Values of θ from -2.8 to 2.8

Item	θ: -2.8	-2.4	-2	-1.6	-1.2	-0.8	-0.4	0	0.4	0.8	1.2	1.6	2	2.4	2.8
1	0.1	0.14	0.17	0.21	0.23	0.24	0.23	0.21	0.18	0.14	0.11	0.08	0.06	0.04	0.03
2	0.05	0.07	0.1	0.14	0.19	0.24	0.28	0.3	0.3	0.26	0.22	0.17	0.12	0.09	0.06
3	0.1	0.14	0.21	0.28	0.35	0.4	0.39	0.35	0.28	0.2	0.14	0.09	0.06	0.04	0.03
4	0.1	0.13	0.18	0.23	0.28	0.3	0.3	0.28	0.23	0.18	0.13	0.1	0.07	0.05	0.03
5	0.09	0.14	0.21	0.3	0.4	0.47	0.48	0.42	0.33	0.23	0.15	0.1	0.06	0.04	0.02
6	0.17	0.21	0.25	0.27	0.27	0.25	0.21	0.17	0.13	0.09	0.07	0.05	0.03	0.02	0.02
7	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.11	0.11	0.11	0.11	0.1	0.09	0.08	0.07
8	0.02	0.02	0.03	0.03	0.04	0.04	0.04	0.05	0.05	0.05	0.06	0.06	0.06	0.06	0.05
9	0.02	0.02	0.02	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
10	0.03	0.04	0.06	0.07	0.09	0.11	0.13	0.15	0.16	0.16	0.16	0.15	0.13	0.11	0.1
11	0.05	0.06	0.08	0.09	0.11	0.12	0.13	0.13	0.13	0.12	0.11	0.1	0.09	0.07	0.06
12	0.05	0.06	0.08	0.1	0.12	0.14	0.16	0.17	0.18	0.17	0.16	0.14	0.12	0.1	0.08
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0.02	0.02	0.02	0.03	0.03	0.03	0.04	0.04	0.04	0.05	0.05	0.05	0.05	0.05	0.05
15	0.02	0.02	0.03	0.03	0.03	0.04	0.04	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.05
16	0.04	0.05	0.06	0.07	0.07	0.08	0.09	0.09	0.09	0.09	0.09	0.08	0.07	0.07	0.06
17	0.03	0.05	0.07	0.1	0.13	0.18	0.23	0.27	0.3	0.31	0.28	0.24	0.19	0.14	0.1
18	0.07	0.09	0.11	0.14	0.16	0.18	0.18	0.18	0.17	0.15	0.13	0.1	0.08	0.06	0.05
19	0.03	0.04	0.07	0.1	0.16	0.24	0.34	0.42	0.47	0.45	0.38	0.28	0.2	0.13	0.08
20	0.03	0.05	0.08	0.12	0.18	0.27	0.35	0.42	0.44	0.41	0.33	0.24	0.17	0.11	0.07
21	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.1	0.11	0.11	0.1	0.09	0.09	0.08	0.07
22	0.03	0.05	0.07	0.09	0.12	0.16	0.19	0.23	0.25	0.25	0.24	0.21	0.17	0.14	0.1
23	0.05	0.08	0.11	0.14	0.19	0.23	0.27	0.29	0.28	0.25	0.21	0.17	0.13	0.09	0.07
24	0.05	0.07	0.09	0.12	0.15	0.18	0.21	0.22	0.23	0.22	0.19	0.16	0.13	0.1	0.08
25	0.06	0.08	0.11	0.15	0.19	0.22	0.25	0.26	0.25	0.22	0.19	0.15	0.11	0.08	0.06
Test Information:	2.30	2.75	3.32	3.98	4.68	5.32	5.77	5.94	5.77	5.32	4.68	3.99	3.36	2.83	2.41
Expected s.e.:	0.66	0.60	0.55	0.50	0.46	0.43	0.42	0.41	0.42	0.43	0.46	0.50	0.55	0.59	0.64

Marginal Reliability for Response Pattern Scores: 0.80

วิธีการตอบและการตรวจให้คะแนนแบบอาร์โนลด์

ตาราง 30 Item Information Function Values for Group 1 at 15 Values of θ from -2.8 to 2.8

Item	θ: -2.8	-2.4	-2	-1.6	-1.2	-0.8	-0.4	0	0.4	0.8	1.2	1.6	2	2.4	2.8
1	0.08	0.12	0.18	0.25	0.33	0.38	0.4	0.37	0.3	0.23	0.16	0.11	0.07	0.04	0.03
2	0.06	0.11	0.18	0.29	0.42	0.54	0.59	0.54	0.41	0.28	0.17	0.1	0.06	0.03	0.02
3	0.1	0.14	0.18	0.22	0.26	0.27	0.27	0.24	0.2	0.16	0.12	0.09	0.06	0.04	0.03
4	0.11	0.15	0.18	0.22	0.24	0.25	0.24	0.21	0.18	0.14	0.11	0.08	0.06	0.04	0.03
5	0.14	0.24	0.38	0.56	0.69	0.69	0.56	0.39	0.24	0.14	0.08	0.04	0.03	0.02	0.01
6	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
7	0.04	0.05	0.05	0.06	0.06	0.07	0.07	0.08	0.08	0.08	0.07	0.07	0.06	0.06	0.05
8	0.03	0.03	0.04	0.04	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.04
9	0.05	0.07	0.1	0.14	0.18	0.23	0.26	0.28	0.28	0.25	0.21	0.16	0.12	0.09	0.06
10	0.04	0.05	0.06	0.07	0.08	0.08	0.09	0.1	0.1	0.1	0.1	0.09	0.08	0.08	0.07
11	0.05	0.08	0.13	0.21	0.32	0.44	0.54	0.56	0.49	0.38	0.26	0.17	0.1	0.06	0.04
12	0.04	0.07	0.1	0.16	0.23	0.31	0.39	0.43	0.42	0.37	0.29	0.21	0.14	0.09	0.06
13	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
14	0.02	0.03	0.03	0.04	0.04	0.05	0.06	0.06	0.07	0.07	0.08	0.08	0.08	0.08	0.07
15	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.03	0.03	0.03	0.03
16	0.04	0.06	0.1	0.15	0.23	0.32	0.4	0.45	0.44	0.38	0.29	0.21	0.14	0.09	0.06
17	0.05	0.09	0.15	0.25	0.39	0.54	0.64	0.63	0.51	0.36	0.23	0.14	0.08	0.05	0.03
18	0.07	0.09	0.11	0.13	0.15	0.16	0.17	0.17	0.16	0.14	0.12	0.1	0.08	0.06	0.05
19	0.05	0.07	0.1	0.13	0.17	0.2	0.23	0.25	0.24	0.22	0.19	0.15	0.12	0.09	0.06
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0.05	0.06	0.09	0.11	0.14	0.18	0.2	0.22	0.23	0.22	0.19	0.16	0.13	0.1	0.08
22	0.04	0.07	0.11	0.18	0.26	0.36	0.45	0.49	0.46	0.37	0.27	0.19	0.12	0.08	0.05
23	0.05	0.07	0.12	0.19	0.28	0.39	0.49	0.53	0.49	0.39	0.28	0.19	0.12	0.08	0.05
24	0.05	0.07	0.1	0.14	0.2	0.26	0.31	0.35	0.34	0.31	0.26	0.2	0.14	0.1	0.07
25	0.06	0.11	0.18	0.3	0.46	0.6	0.67	0.6	0.45	0.3	0.18	0.11	0.06	0.03	0.02
Test Information:	2.25	2.85	3.71	4.87	6.21	7.43	8.14	8.04	7.21	5.99	4.76	3.73	2.96	2.41	2.03
Expected s.e.:	0.67	0.59	0.52	0.45	0.4	0.37	0.35	0.35	0.37	0.41	0.46	0.52	0.58	0.64	0.7

Marginal Reliability for Response Pattern Scores: 0.83

ภาคผนวก ง
รายนามผู้เชี่ยวชาญ

รายนามผู้เชี่ยวชาญ

ดร. มนตา ตุลย์เมธาการ	อาจารย์ประจำภาควิชาการวัดผลและวิจัยการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ
ดร. นวรินทร์ ตาก้อนทอง	อาจารย์ประจำสำนักทดสอบทางการศึกษาและจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ
อาจารย์ปานวาสน์ มหาลวเลิศ	อาจารย์ประจำสำนักทดสอบทางการศึกษาและจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ
อาจารย์พรปวีณ์ ศรียะพันธุ์	ครูชำนาญการพิเศษ สาระการเรียนรู้คณิตศาสตร์ โรงเรียนเทพศิรินทร์ร่มเกล้า
อาจารย์ชัยพร นาควิเชียร	ครูชำนาญการพิเศษ สาระการเรียนรู้คณิตศาสตร์ โรงเรียนวัดราชโอรส

ประวัติย่อผู้วิจัย

ประวัติย่อผู้วิจัย

ชื่อ ชื่อสกุล	นางสาววรารัตน์ แสงสุข
วันเดือนปีเกิด	29 สิงหาคม 2531
สถานที่เกิด	เขตหนองจอก จังหวัดกรุงเทพมหานคร
สถานที่อยู่ปัจจุบัน	41 ถนนอยู่วิทยา แขวงกระทุ่มราย เขตหนองจอก จังหวัดกรุงเทพมหานคร 10530
ตำแหน่งหน้าที่การงานในปัจจุบัน	ครู ค.ศ. 1
สถานที่ทำงานปัจจุบัน	โรงเรียนเทพศิรินทร์ร่มเกล้า แขวงคลองสามประเวศ เขตลาดกระบัง จังหวัดกรุงเทพมหานคร

ประวัติการศึกษา

พ.ศ. 2550	มัธยมศึกษาปีที่ 6 จาก โรงเรียนมัธยมวัดหนองจอก เขตหนองจอก จังหวัดกรุงเทพมหานคร
พ.ศ. 2555	ปริญญาการศึกษาบัณฑิต (กศ.บ.) สาขาวิชาคณิตศาสตร์ จาก มหาวิทยาลัยศรีนครินทรวิโรฒ
พ.ศ. 2560	ปริญญาการศึกษามหาบัณฑิต (กศ.ม.) สาขาวิชาการวิจัยและพัฒนาศักยภาพมนุษย์ (แขนงการทดสอบและวัดผลการศึกษา) จาก มหาวิทยาลัยศรีนครินทรวิโรฒ