

ผลของการให้คะแนนที่มีต่อประสิทธิภาพการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

ภาณุวัชร ปุระณะศิริ

คุณฉันทิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปรัชญาดุษฎีบัณฑิต

สาขาวิชาวิจัย วัฒนผลและสถิติการศึกษา

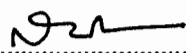
คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา

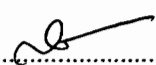
สิงหาคม 2558

ลิขสิทธิ์เป็นของมหาวิทยาลัยบูรพา

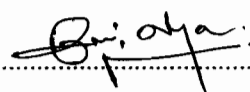
คณะกรรมการควบคุมคุษฎีนิพนธ์ และคณะกรรมการสอบคุษฎีนิพนธ์ ได้พิจารณา
คุษฎีนิพนธ์ของ ภาณุวัชร ปุระณะศิริ ฉบับนี้แล้ว เห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตาม
หลักสูตรปรัชญาคุษฎีบัณฑิต สาขาวิชาวิจัย วัฒนผลและสติดิการศึกษา ของมหาวิทยาลัยบูรพาได้

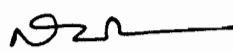
คณะกรรมการควบคุมคุษฎีนิพนธ์



..... อาจารย์ที่ปรึกษาหลัก
(ดร.สมศักดิ์ ลิลา)

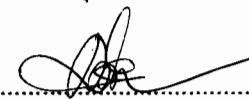

..... อาจารย์ที่ปรึกษาร่วม
(ดร.สมพงษ์ ปั่นห่วน)

คณะกรรมการสอบคุษฎีนิพนธ์


..... ประธาน
(รองศาสตราจารย์ ดร.บุญชม ศรีสะอาด)


..... กรรมการ
(ดร.สมศักดิ์ ลิลา)


..... กรรมการ
(ดร.สมพงษ์ ปั่นห่วน)


..... กรรมการ
(รองศาสตราจารย์ ดร.ไพรัตน์ วงษ์นาม)

คณะศึกษาศาสตร์ อนุมัติให้รับคุษฎีนิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตาม
หลักสูตรปรัชญาคุษฎีบัณฑิต สาขาวิชาวิจัย วัฒนผลและสติดิการศึกษา ของมหาวิทยาลัยบูรพา


..... คณบดีคณะศึกษาศาสตร์
(รองศาสตราจารย์ ดร.วิชิต สุรัตน์เรืองชัย)

วันที่.....17.....เดือน.....สิงหาคม.....พ.ศ. 2558

กิตติกรรมประกาศ

คุณฐิณิพนธ์ฉบับนี้สำเร็จลงได้ด้วยได้รับความเมตตา กรุณาจาก ดร.สมศักดิ์ ธิลา อาจารย์ที่ปรึกษาหลัก ดร.สมพงษ์ ปั่นหุ่น อาจารย์ที่ปรึกษาร่วม และรองศาสตราจารย์ ดร.ไพรัตน์ วงษ์นาม ประธานหลักสูตรปรัชญาดุษฎีบัณฑิต สาขาวิชาวิจัย วัฒนผลและสถิตการศึกษ ผู้ช่วยศาสตราจารย์ ดร.สุริพร อนุศาสนนันท์ หัวหน้าภาควิชาวิจัยและจิตวิทยาประยุกต์ คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา รองศาสตราจารย์ ดร.บุญชม ศรีสะอาด คณะศึกษาศาสตร์ มหาวิทยาลัยมหาสารคาม ที่กรุณาให้คำปรึกษา แนะนำแนวทางที่ถูกต้อง ด้วยจิตวิญญาณของความเป็นครูที่เต็มเปี่ยม ถ่ายทอดความรู้ และศาสตร์ต่าง ๆ ทางด้านการวิจัย วัฒนผลและสถิตการศึกษ ตลอดจนแก้ไขข้อบกพร่องต่าง ๆ ด้วยความละเอียดถี่ถ้วนและเอาใจใส่ด้วยดีเสมอมา ผู้วิจัยรู้สึกซาบซึ้งและขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

ขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.สมศักดิ์ จีวัฒนา อาจารย์ประจำภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์ อาจารย์ ดร.กระพัน ศรีงาน คณบดี คณะครุศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์ อาจารย์ ดร.ชัยยศ ชาวระนอง อาจารย์ประจำ มหาวิทยาลัยราชภัฏสวนดุสิต อาจารย์ ดร.ชนาวุฒิ ประกอบผล อาจารย์ประจำวิทยาลัยการบริหาร และการจัดการ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ที่กรุณาให้ความรู้ คำปรึกษา ตรวจสอบแก้ไข และวิจารณ์ผลงาน ทำให้งานวิจัยมีความสมบูรณ์ยิ่งขึ้น ผู้เชี่ยวชาญทุกท่าน ที่ให้ความอนุเคราะห์ในการตรวจสอบ รวมทั้งให้คำแนะนำ แก้ไขเครื่องมือที่ใช้ในการวิจัย ให้มีคุณภาพ นอกจากนี้ ยังได้รับความอนุเคราะห์จากผู้อำนวยการสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 32 ผู้อำนวยการ โรงเรียนบุรีรัมย์พิทยาคม รองผู้อำนวยการ โรงเรียนบุรีรัมย์พิทยาคม ตลอดจนเพื่อนครู โรงเรียนมัธยมศึกษา จังหวัดบุรีรัมย์ ที่ให้ความร่วมมือเป็นอย่างดีในการเก็บรวบรวมข้อมูลที่ใช้ในการวิจัย ทำให้คุณฐิณิพนธ์ฉบับนี้สำเร็จด้วยดี

ขอขอบพระคุณคณาจารย์ภาควิชาวิจัยและจิตวิทยาประยุกต์ ทุกท่าน เพื่อน พี่ น้อง นิสิต สาขาวิชาวิจัย วัฒนผลและสถิตการศึกษ ทุกคนที่ให้กำลังใจ และสนับสนุนผู้วิจัยเสมอมา

คุณค่าและประโยชน์ของคุณฐิณิพนธ์ฉบับนี้ ผู้วิจัยขอมอบเป็นกตัญญูกตเวทิตาแด่บุพการี บูรพาจารย์ และผู้มีพระคุณทุกท่านทั้งในอดีตและปัจจุบัน ที่ทำให้ข้าพเจ้าเป็นผู้มีการศึกษา และประสบความสำเร็จมาจนตราบเท่าทุกวันนี้

ภาณุวัชร ปุระณะศิริ

50810838: สาขาวิชา: วิจัย วัดผลและสถิติการศึกษา; ปร.ด. (วิจัย วัดผลและสถิติการศึกษา)

คำสำคัญ: ผลของการให้คะแนน/ การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

ภาณุวัชร ปุระณะศิริ: ผลของการให้คะแนนที่มีต่อประสิทธิภาพการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (THE EFFECTS OF SCORING METHODS ON EFFICIENCY OF COMPUTERIZED ADAPTIVE TESTING) คณะกรรมการควบคุมคุรุวิชาชีพ: สมศักดิ์ ธิลา, กศ.ด., สมพงษ์ ปิ่นหุ่น, ค.ด. 173 หน้า. ปี พ.ศ. 2558.

การวิจัยครั้งนี้เป็นการวิจัยเชิงทดลอง มีวัตถุประสงค์เพื่อศึกษาผลของการให้คะแนนที่มีต่อคุณภาพการวัดในด้านความตรง ความเที่ยง ความคลาดเคลื่อนมาตรฐานในการวัด และเปรียบเทียบผลของการให้คะแนนที่มีต่อประสิทธิภาพการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ในด้านจำนวนข้อสอบ และค่าฟังก์ชันสารสนเทศของชุดข้อสอบ เมื่อพิจารณาตามองค์ประกอบของการทดสอบ ได้แก่ การคัดเลือกข้อสอบ การประมาณค่าความสามารถผู้สอบ การยุติการทดสอบ และความสามารถของผู้สอบที่แตกต่างกัน การดำเนินการวิจัยแบ่งเป็น 2 ระยะ โดยระยะแรกดำเนินการสร้างคลังข้อสอบวิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษาตอนปลาย ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 กลุ่มตัวอย่าง คือ นักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2557 จำนวน 3,330 คน วิเคราะห์คุณภาพตามทฤษฎีการตอบสนองข้อสอบ ได้ข้อสอบทั้งหมด 230 ข้อ ที่ยังไม่ได้ปรับสเกล มีค่าอำนาจจำแนกอยู่ระหว่าง 0.739-0.957 และค่าเบี่ยงเบนมาตรฐานอยู่ระหว่าง 0.060-0.238 อยู่ในเกณฑ์ดี ค่าความยากง่ายอยู่ระหว่าง 0.798-1.095 และค่าเบี่ยงเบนมาตรฐานอยู่ระหว่าง 0.295-1.270 อยู่ในเกณฑ์ค่อนข้างไปทางยาก และค่าการเดาอยู่ระหว่าง 0.1120-0.1640 และค่าเบี่ยงเบนมาตรฐานอยู่ระหว่าง 0.024-0.047 อยู่ในเกณฑ์ดี ทำการตรวจสอบความเป็นมิติเดียวของข้อสอบ โดยการวิเคราะห์องค์ประกอบ พบว่า ค่าไอเกน ของตัวประกอบที่ 1 สูงกว่าค่าไอเกนของตัวประกอบอื่น ๆ ที่เหลือ และตัวประกอบที่เหลือดังกล่าวก็มีค่าไอเกนที่ใกล้เคียงกัน ซึ่งลักษณะดังกล่าวนี้อนุมานได้ว่า ข้อสอบที่ได้มีความเป็นมิติเดียว ตรงตามข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ ระยะที่สองเป็นการศึกษาผลของการให้คะแนนตามตัวแปรอิสระ 5 ตัว ได้แก่ การคัดเลือกข้อสอบ การประมาณค่าความสามารถผู้สอบ การยุติการทดสอบ และความสามารถของผู้สอบ ที่มีต่อการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ในด้านจำนวนข้อสอบ และค่าฟังก์ชันสารสนเทศของชุดข้อสอบ กลุ่มตัวอย่าง คือ นักเรียนชั้นมัธยมศึกษาปีที่ 6 โรงเรียนบุรีรัมย์พิทยาคม ปีการศึกษา 2557 จำนวน 540 คน เครื่องมือที่ใช้ในการวิจัย คือ โปรแกรมคอมพิวเตอร์และข้อสอบวิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษา ตอนปลาย ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 สำหรับ

การศึกษาประสิทธิภาพการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ วิเคราะห์ข้อมูลโดยใช้ค่าเฉลี่ยเลขคณิต ค่าเบี่ยงเบนมาตรฐาน

ผลการวิจัย พบว่า

1. คุณภาพของการวัดจากผลการให้คะแนนแบบ MR แบบ MTF และแบบ OMTF มีค่าความตรง คือ 0.7202, 0.7233, 0.7239 ตามลำดับ ค่าความเที่ยง คือ 0.7716, 0.7750, 0.7757 ตามลำดับ และความคลาดเคลื่อนมาตรฐานในการวัด คือ 0.2326, 0.4609, 0.2305 ตามลำดับ
2. ผลของการให้คะแนนแบบ MR แบบ MTF และแบบ OMTF เมื่อพิจารณาตามเกณฑ์การคัดเลือกข้อสอบ การประมาณค่าความสามารถผู้สอบ การยุติการทดสอบ และความสามารถของผู้สอบมีผลต่อจำนวนข้อสอบและค่าฟังก์ชันสารสนเทศของชุดข้อสอบ

50810838: MAJOR: EDUCATIONAL RESEARCH, MEASUREMENT AND STATISTICS;
Ph.D. (EDUCATIONAL RESEARCH, MEASUREMENT AND STATISTICS)

KEYWORDS: EFFECTS OF SCORING METHODS/ COMPUTERIZED ADAPTIVE
TESTING

PANUWAT PURANASIRI: THE EFFECTS OF SCORING METHODS ON
EFFICIENCY OF COMPUTERIZED ADAPTIVE TESTING. ADVISORY COMMITTEE:
SOMSAK LILA, Ed.D., SOMPONG PANHOON, Ph.D. 173 P. 2015.

The purposes of this research were to study effects of scoring methods on the quality measure of validity, reliability, and standard error of measurement and to compare results of effects of scoring methods on efficiency of computerized adaptive testing to investigate on number of items and test functional information by considering the composition of the main testing, item selection criteria, ability estimation procedure, termination criteria, examinee's different abilities. The process of this study consisted of two phases. The first phase was developing test item bank on Mathematics for upper secondary school level based on the Basic Education Core Curriculum B.E. 2551 (A.D. 2008). The participants were 3,330 Grade 12 students in the academic year of 2014, the test was analyzed the qualities of items by using Item Response Theory. A total of 230 test items were created that have not adapted the scale, having the discrimination index between 0.739 and 0.957. The standard deviation is between 0.060 and 0.238, which is considered good. The difficulty of test items was between 0.798 and 1.095 and the standard deviation is between 0.295 and 1.270. The guessing values were between 0.1120 and 0.1640, the standard deviation ranged from 0.024 and 0.047, which is considered good. Checking the unidimensional of the test by using factor analysis revealed that the first eigen value was higher than other values, with similar values all together. It is assumed that the test is unidimensional, meeting basic terms of the theory. The second phase was studying effects of scoring methods by the five independent variables including selection criteria, ability estimation procedure, termination criteria and examinee's ability on efficiency of computerized adaptive testing to investigate on number of items and test functional information. The participants were 540 Grade 12 student in the academic year of 2014 at Burirampitthayakhom School. The instruments used in this phase were test item pools on fraction for upper secondary school level

based on the Basic Education Core Curriculum B.E. 2551 (A.D. 2008). The computer program for computerized adaptive testing were employed to analyze means, standard deviation, and Analysis of Variance.

The results indicated that

1. The quality of the measurement results scoring of MR, MTF and OMTF revealed validity of 0.7202, 0.7233, 0.7239 respectively. The reliability values were 0.7716, 0.7750, 0.7757 respectively. The standard error of measurement values were 0.2326, 0.4609, 0.2305 respectively.

2. Effects of MR, MTF and OMTF scoring after item considering selection criteria, ability estimation procedure, termination criteria examinee's ability, number of items, and functional test information.

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ญ
สารบัญภาพ.....	ฎ
บทที่	
1 บทนำ.....	1
ความเป็นมาและความสำคัญของปัญหา.....	1
วัตถุประสงค์ของการวิจัย.....	11
คำถามการวิจัย.....	11
สมมติฐานของการวิจัย.....	12
ความสำคัญของการวิจัย.....	14
กรอบแนวคิดในการวิจัย.....	15
ประโยชน์ที่ได้รับจากการวิจัย.....	15
ขอบเขตของการวิจัย.....	16
ข้อตกลงเบื้องต้นของการวิจัย.....	18
นิยามศัพท์เฉพาะ.....	18
2 เอกสารและงานวิจัยที่เกี่ยวข้อง.....	21
การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์.....	21
การให้คะแนนข้อสอบเลือกตอบแบบถูกผิด.....	36
ทฤษฎีการตอบสนองข้อสอบ.....	46
ประสิทธิภาพของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์.....	57
งานวิจัยที่เกี่ยวข้อง.....	64
3 วิธีการดำเนินการวิจัย.....	75
การดำเนินการวิจัย ระยะแรก.....	75
การดำเนินการวิจัย ระยะที่สอง.....	77
เครื่องมือที่ใช้ในการวิจัย.....	81

สารบัญ (ต่อ)

บทที่	หน้า
การสร้างเครื่องมือที่ใช้ในการวิจัย.....	84
การเก็บรวบรวมข้อมูล.....	90
การวิเคราะห์ข้อมูลการวิจัย.....	90
สถิติที่ใช้ในการวิจัย.....	92
4 ผลการวิจัย.....	98
สัญลักษณ์ที่ใช้ในการแสดงผลการวิเคราะห์ข้อมูล.....	98
ผลการวิเคราะห์ข้อมูล.....	99
5 สรุป อภิปรายผลและข้อเสนอแนะ.....	108
สรุปผลการวิจัย.....	109
อภิปรายผล.....	111
ข้อเสนอแนะ.....	114
บรรณานุกรม.....	115
ภาคผนวก.....	122
ภาคผนวก ก.....	123
ภาคผนวก ข.....	153
ภาคผนวก ค.....	155
ภาคผนวก ง.....	166
ประวัติย่อของผู้วิจัย.....	173

สารบัญตาราง

ตารางที่	หน้า
1 ความแตกต่างระหว่างการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ กับการทดสอบที่ใช้กระดาษเขียนตอบ.....	26
2 โครงสร้างของการจัดแบบทดสอบเฟล็กซิเบิล.....	34
3 ฟังก์ชันทางคณิตศาสตร์ของโมเดลการตอบสนองข้อสอบ.....	52
4 ตัวอย่างค่าพารามิเตอร์ของข้อสอบ 3 ข้อ และ $P_i(\theta)$ ที่ระดับ θ ต่าง ๆ.....	54
5 จำนวนประชากรและกลุ่มตัวอย่างที่ใช้ในการหาคุณภาพของการวัดผล จำแนกตามโรงเรียนและชุดข้อสอบ.....	77
6 เจ็อนไขของโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์.....	78
7 โครงสร้างของข้อสอบวิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษาตอนปลาย จำแนกตามมาตรฐานและตัวชี้วัด.....	81
8 จำนวนข้อสอบ ผู้สอบ ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของค่าพารามิเตอร์ ของข้อสอบที่ใช้วัดความรู้วิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ยังไม่ได้ปรับสเกล.....	100
9 ค่าไอเกนที่ได้จากการวิเคราะห์ตัวประกอบของแบบทดสอบวิชาคณิตศาสตร์ ฉบับที่ 1-9 จำแนกเป็นรายตัวประกอบ (Eigen value >1.00).....	101
10 ผลการให้คะแนนที่แตกต่างกัน 3 แบบ จำแนกตามตัวแปรความตรง ความเที่ยง และความคลาดเคลื่อนมาตรฐานในการวัด.....	103
11 ค่าเฉลี่ย และค่าเบี่ยงเบนมาตรฐานของผลการให้คะแนน 3 แบบ ที่ใช้ในการทดสอบ แบบปรับเหมาะด้วยคอมพิวเตอร์ จำแนกตามเกณฑ์การคัดเลือกข้อสอบ.....	103
12 ค่าเฉลี่ย และค่าเบี่ยงเบนมาตรฐานของผลการให้คะแนน 3 แบบ ที่ใช้ในการทดสอบ แบบปรับเหมาะด้วยคอมพิวเตอร์ จำแนกตามวิธีการประมาณค่าความสามารถ ของผู้สอบ.....	105
13 ค่าเฉลี่ย และค่าเบี่ยงเบนมาตรฐานของผลการให้คะแนน 3 วิธี ที่ใช้ในการทดสอบ แบบปรับเหมาะด้วยคอมพิวเตอร์ จำแนกตามเกณฑ์ยุติการทดสอบ.....	106
14 ค่าเฉลี่ย และค่าเบี่ยงเบนมาตรฐานของผลการให้คะแนน 3 วิธี ที่ใช้ในการทดสอบ แบบปรับเหมาะด้วยคอมพิวเตอร์ จำแนกตามความสามารถทางคณิตศาสตร์ ของผู้เข้าสอบ.....	107

สารบัญญภาพ

ภาพที่	หน้า
1 กรอบแนวคิดในการวิจัย.....	15
2 โครงสร้างของการทดสอบเฉพาะบุคคลแบบสองขั้นตอน.....	29
3 โครงสร้างของการทดสอบเฉพาะบุคคลรูปแบบพีระมิดขนาดขั้นคงที่.....	30
4 โครงสร้างของแบบทดสอบรูปพีระมิดขนาดขั้นแปรผัน.....	31
5 โครงสร้างของแบบทดสอบรูปพีระมิดแบบโรบินส์มอนโร.....	31
6 โครงสร้างของแบบทดสอบรูปพีระมิดแบบสกัดการสะท้อนกลับ.....	32
7 โครงสร้างของแบบทดสอบรูปพีระมิดแบบรักษาการสะท้อนกลับ.....	32
8 โครงสร้างของแบบทดสอบรูปพีระมิดแบบมีหลายข้อในแต่ละขั้น.....	33
9 โครงสร้างของแบบทดสอบรูปพีระมิดแบบให้น้ำหนักแก่ตัวเลือกของข้อสอบ เพื่อแยกทาง.....	33
10 โครงสร้างของแบบทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบแบบแบ่งขั้น....	35
11 เส้นโค้งคุณลักษณะข้อสอบของโมเดลแบบ 3-พารามิเตอร์.....	50
12 โค้งคุณลักษณะของข้อสอบตามโมเดลการตอบสนองข้อสอบ 3 โมเดล.....	53
13 ตัวอย่าง โค้งคุณลักษณะข้อสอบ (ICCs) ของข้อสอบ 3 ข้อ.....	54
14 ขั้นตอนดำเนินการวิจัย.....	80
15 แสดงขั้นตอนการทดสอบ.....	89
16 ค่าไอเกนที่ได้จากการวิเคราะห์ตัวประกอบของแบบทดสอบวิชาคณิตศาสตร์.....	102

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

ปรัชญาของการวัดผลการศึกษาตามแนวคิดหลักของ ศาสตราจารย์ ดร.ชวาล แพร์ตกุล ผู้ได้รับยกย่องว่าเป็นบิดาแห่งการวัดผลของไทย กล่าวว่า “ทดสอบเพื่อค้นและพัฒนาสมรรถภาพของมนุษย์” (Test to discover and develop human talent) (ชวาล แพร์ตกุล, 2552, หน้า 199) แสดงให้เห็นถึงความสำคัญของการทดสอบทางการศึกษาว่า เป็นกระบวนการวัดที่สำคัญในการค้นหาสมรรถภาพของบุคคลว่า มีอะไร ไม่มีอะไร และมากน้อยแค่ไหน สิ่งใดที่มีอยู่ สิ่งนั้นต้องวัดได้ เพื่อพัฒนา ส่งเสริมสิ่งที่มีอยู่ให้มากขึ้น และถ้าไม่มีก็ปลูกฝังสิ่งใหม่ขึ้นมา ซึ่งการทดสอบที่มีมาในอดีตเน้นการใช้แบบสอบพร้อมฉบับเดียวและตอบคำถามโดยใช้กระดาษและดินสอหรือปากกา ถูกนำมาใช้เป็นระยะเวลาที่ยาวนาน หลังจากปี ค.ศ. 1980 เป็นต้นมาการใช้คอมพิวเตอร์ส่วนตัว (PC) ได้ขยายตัวอย่างรวดเร็ว การทดสอบจึงเริ่มพลิกโฉมหน้าใหม่ด้วยการใช้คอมพิวเตอร์สำหรับการทดสอบกันมากขึ้น คอมพิวเตอร์ช่วยให้การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบมีความเป็นไปได้ มีประสิทธิภาพ และถูกต้องแม่นยำยิ่งขึ้น การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ โดยใช้คอมพิวเตอร์ (Computerized Adaptive Testing: CAT) มีจุดเด่นที่สำคัญ คือ ทำให้สามารถจัดการสอบได้สะดวกตามความต้องการของผู้สอบ ในการเลือกข้อสอบให้เหมาะสมกับความสามารถของผู้สอบ ทำให้การทดสอบมีสภาพสอดคล้องกับธรรมชาติของผู้สอบมากขึ้น ความวิตกกังวลในการสอบจะลดลง นอกจากนี้ยังสามารถประมวลผลได้อย่างรวดเร็ว รายงานผลได้ทันที และมีความถูกต้องแม่นยำมากกว่า การทดสอบแบบประเพณีนิยม การทดสอบลักษณะนี้จึงเหมาะกับสังคมในยุคโลกาภิวัตน์ องค์การการศึกษาหลายแห่งจึงหันมาใช้ในการทดสอบแบบนี้มากขึ้นตามลำดับ (ศิริชัย กาญจนวาสี, 2550, หน้า 220) การนำเทคโนโลยีมาใช้บนพื้นฐานของทฤษฎีการวัดแนวใหม่ จะช่วยเสริมประสิทธิภาพของการสร้างข้อสอบ พัฒนาค้างข้อสอบ การจัดแบบสอบ การตรวจข้อสอบ การวิเคราะห์ข้อสอบ การแปลผลการสอบ และการรายงานผล ทำให้ระบบการทดสอบมีความแม่นยำ กระจับ รัดกุม และยืดหยุ่น แนวโน้มของการพัฒนาเครื่องมือทางเทคโนโลยีคอมพิวเตอร์มีบทบาทต่อสังคมมากขึ้น เนื่องมาจากว่า คอมพิวเตอร์เป็นเครื่องมือที่สามารถหาได้ง่าย ราคาถูก มีประสิทธิภาพ และยังจะเป็นการช่วยเพิ่มคุณภาพและประสิทธิภาพของการวัดผลการเรียนรู้ของบุคคลได้อย่างรวดเร็ว ถูกต้อง ทำให้กระบวนการวัดผลมีความคล่องตัว (สมศักดิ์ ลิลา, 2539,

หน้า 52-53) ในยุคแรก กระบวนการวัดผลโดยการทดสอบแบบนี้จะเรียกว่า การทดสอบแบบเทลอร์ (Tailored testing) คำนี้ใช้ครั้งแรก โดย Tumbull (1951 อ้างถึงใน Lord, 1980) ต่อมา ไวส์ (Weiss, 1974) ได้เสนอให้ใช้คำว่า Adaptive test ซึ่งเป็นกระบวนการประมาณค่าความสามารถของผู้สอบ และจัดข้อสอบที่เหมาะสมกับระดับความสามารถของผู้สอบเป็นรายบุคคลด้วยคอมพิวเตอร์ และเรียกว่า Computerized adaptive testing ใช้ตัวย่อว่า CAT (บุญชม ศรีสะอาด, 2540, หน้า 73)

การวัดคุณลักษณะของผู้สอบโดยการทดสอบทำได้หลายวิธี เช่น การสอบปากเปล่า การสอบปฏิบัติ การสอบข้อเขียน เหล่านี้ก็เพื่อค้นหาระดับความสามารถหรือทักษะของผู้สอบ ให้ได้ถูกต้อง แม่นยำมากที่สุด หรือเกิดความคลาดเคลื่อนน้อยที่สุด กระบวนการทดสอบดังกล่าว เป็นการวัดเพียงตัวอย่างของพฤติกรรม แล้วแปลงพฤติกรรมที่วัดได้ให้ออกมาในรูปของปริมาณ (Lord, 1980, p. 12) การวัดคุณลักษณะหรือความสามารถของบุคคลไม่สามารถวัดได้โดยตรง จะต้องสร้างเครื่องมือขึ้นเพื่อวัดสิ่งที่แฝงอยู่ในตัวบุคคล โดยจะอาศัยผลของการตอบสนอง ต่อสิ่งเร้าในลักษณะใดลักษณะหนึ่งตามเครื่องมือวัดผลที่ใช้ ซึ่งมีอยู่หลายชนิดด้วยกัน แต่เครื่องมือ ชนิดที่ได้รับความนิยมมากที่สุดคือ แบบทดสอบ โดยเฉพาะแบบทดสอบเลือกตอบ เนื่องจากสามารถ วัดความรู้ความสามารถและวัดผลสัมฤทธิ์ได้อย่างกว้างขวาง ประกอบกับมีความเป็นปรนัย ในการตรวจให้คะแนน ใช้เวลารวดเร็วในการตรวจให้คะแนน ผลการวิเคราะห์ข้อสอบสามารถ นำมาวินิจฉัยปัญหาการเรียนรู้อันได้ และสามารถใช้อุปกรณ์คอมพิวเตอร์ในการตรวจให้คะแนนได้อีกด้วย แต่วิธีการตอบและการให้คะแนนที่เปิดโอกาสให้ผู้สอบมีการตอบถูกโดยการเดานั้น จะส่งผลต่อ ความคลาดเคลื่อนในการวัดผล และการให้คะแนนบางวิธียังไม่สามารถให้สารสนเทศที่เพียงพอ ที่จะจำแนกผู้สอบในระดับต่าง ๆ ของความรู้ได้ ระหว่างผู้ที่มีความรู้สมบูรณ์ (Full knowledge) ผู้ที่มีความรู้บางส่วน (Partial knowledge) และผู้ที่ไม่มีความรู้ (Absence of knowledge) จากจุดอ่อน ดังกล่าวจึงได้มีการศึกษาวิธีการแก้ปัญหา เพื่อลดโอกาสในการเดาสุ่ม และเพิ่มสารสนเทศเกี่ยวกับการตอบข้อสอบให้ได้มากที่สุด โดยปรับปรุงวิธีการให้คะแนน ซึ่งเป็นกระบวนการสำคัญ ในการวัดผลการให้คะแนนที่เหมาะสม สามารถบอกถึงตำแหน่งจริงของความรู้ของผู้สอบ จะทำให้คุณภาพของการวัดผลในประเด็นที่เกี่ยวกับความเที่ยงและความตรงสูงขึ้น ทำให้คะแนน มีความเชื่อมั่นมากขึ้น (Frany, 1980, pp. 79-90)

คะแนนการสอบเป็นแหล่งอ้างอิงสำคัญของการวัดผลจิตวิทยา ผลของการให้คะแนน ที่มีคุณภาพจะทำให้ได้ผลของการวัดมีคุณภาพด้วย ซึ่งสามารถแสดงได้จากค่าความตรง ค่าความเชื่อมั่น และความคลาดเคลื่อนมาตรฐานในการวัดของคะแนนที่เป็นผลจากการวัด ในสิ่งเดียวกัน การวัดที่ถูกต้องเที่ยงตรงที่สุดนั้น นอกจากจะต้องพิจารณาถึงผลของการให้คะแนน ที่เหมาะสมแล้ว การวิเคราะห์คุณภาพของข้อสอบ การวิเคราะห์เพื่อประมาณค่าความสามารถ

ของผู้สอบก็เป็นกระบวนการหนึ่งที่จะทำให้สารสนเทศเกี่ยวกับบุคคลและข้อสอบมีความถูกต้องและแม่นยำ การพัฒนาวิธีการตอบสนองข้อสอบ (IRT) เพื่อแก้ปัญหาการวิเคราะห์คุณภาพตามแนวทฤษฎีแบบดั้งเดิม (CTT) โดยอธิบายกระบวนการวัดในรูปคุณลักษณะแฝง (Latent trait) หรืออาจกล่าวได้ว่า โอกาสที่ผู้สอบจะตอบข้อสอบ ได้ถูกนั้น ขึ้นอยู่กับความสามารถของผู้สอบ และคุณลักษณะของข้อสอบ ซึ่งอาจอธิบายด้วยพารามิเตอร์ของข้อสอบ (Item parameter) สามค่า คือ อำนาจจำแนก (a) ความยาก (b) และการเดาตอบ (c) ฐานความเชื่อในทฤษฎีนี้เชื่อว่า สามารถประมาณค่าคุณลักษณะของผู้สอบโดยอิสระจากชุดของข้อคำถามหรือข้อความที่เลือกใช้ (Test free) เพราะค่าพารามิเตอร์ของผู้สอบ (Examinee parameter) หรือคุณลักษณะ (Trait) หรือความสามารถ (Ability) ที่แท้จริงของผู้สอบก็เป็นคุณลักษณะที่มีอยู่ในตัวผู้สอบนั้นจริง จึงไม่ควรแปรเปลี่ยนไปตามชุดของข้อสอบที่เลือกใช้ การประมาณค่าพารามิเตอร์นี้เป็นแนวคิดที่สำคัญของทฤษฎีการตอบสนองข้อสอบ โดยมีดัชนีความแม่นยำของการประมาณค่าคือ ฟังก์ชันสารสนเทศของข้อสอบ (Item Information Function: IIF) และค่าฟังก์ชันสารสนเทศของชุดข้อสอบ (Test Information Function: TIF) นั่นเอง (Birnbaum, 1968) โดย ณ ระดับความสามารถใดที่ค่าฟังก์ชันสารสนเทศมีค่าสูง แสดงว่า การวัดนั้นมีคุณภาพดีมากในการประมาณค่าความสามารถที่ระดับนั้น การวัดความสามารถในทางจิตวิทยาที่มุ่งอธิบายคุณลักษณะภายในของบุคคลนั้น เป็นเพราะว่า การวัดคุณลักษณะภายในจะทำให้เข้าใจการเกิดพฤติกรรมภายนอกของบุคคลที่สามารถสังเกตได้โดยตรง อันจะนำไปสู่การทำนาย ควบคุม และพัฒนาพฤติกรรมบุคคล การวัดความสามารถ (Ability) ของบุคคลทางด้านความรู้ ความคิด เป็นการวัดที่ไม่สามารถสังเกตหรือวัดได้โดยตรง ต้องอาศัยการวัดทางอ้อม ซึ่งเป็นการวัดทางจิตวิทยา จะใช้วิธีการตอบสนองต่อสิ่งเร้าที่จัดขึ้น แล้วนำผลการตอบสนองที่ได้ สรุปล้างอิงไปยังคุณลักษณะภายในของบุคคล ปัจจุบันใช้วิธีการทดสอบ (Testing) โดยใช้ข้อสอบ (Item) กระตุ้นการตอบสนองต่อบุคคลแล้วนับหน่วยความสามารถเป็นคะแนนรายข้อ เพื่ออ้างอิงถึงคุณลักษณะภายในของบุคคล แต่เนื่องจากวิธีการดังกล่าวเป็นการวัดทางอ้อม จะมีความคลาดเคลื่อนในการวัด (Measurement error) รวมอยู่ด้วย ฉะนั้น การวัดและประเมินผลที่สำคัญต้องประกอบด้วย ข้อสอบที่มีคุณภาพ กระบวนการทดสอบที่มีประสิทธิภาพ และวิธีการคิดคะแนนความสามารถที่เหมาะสม สอดคล้องกับโมเดลการคิดของบุคคล จึงจะทำให้การลงข้อสรุปเกี่ยวกับความสามารถของบุคคลถูกต้อง แม่นยำมากที่สุด หรือมีความคลาดเคลื่อนต่ำ คะแนนที่วัดได้จึงจะสอดคล้องกับปริมาณความสามารถที่แท้จริงมากที่สุด (Lord, 1980, p. 3) อย่างไรก็ตาม คะแนนที่ได้จากการสอบเป็นข้อมูลเชิงปริมาณ ที่รวมเอาความคลาดเคลื่อนจากแหล่งต่าง ๆ ไม่ว่าจะเป็นความคลาดเคลื่อนจากตัวผู้สอบเอง จากข้อสอบ สถานการณ์ในการสอบ และความคลาดเคลื่อนจากแหล่งอื่น ๆ ทั้งที่เป็นระบบและไม่เป็นระบบ

การให้คะแนนจึงเป็นองค์ประกอบที่สำคัญของการวัดผลที่จะส่งผลให้กระบวนการทดสอบมีประสิทธิภาพ สนองตอบต่อผู้สอบในเชิงจิตวิทยา ผู้สอบที่มีความสามารถสูงควรจะได้รับข้อสอบที่ยาก ส่วนผู้สอบที่มีความสามารถต่ำก็ควรจะได้รับข้อสอบที่ง่าย นั่นคือ การทดสอบมีความเหมาะสมกับความสามารถของผู้สอบ (Lord & Novick, 1968, p. 359) ด้วยเหตุผลดังกล่าว การให้คะแนนที่เหมาะสมแก่ผู้สอบ จึงส่งเสริมให้คุณภาพของการวัดผลดีขึ้น การพิจารณาประสิทธิภาพของแบบทดสอบจึงสามารถอ้างอิงได้จากความตรง ความเชื่อมั่น และค่าสารสนเทศที่ได้จากคะแนนการสอบในตัวแปรที่มุ่งวัด และจากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง ในชุดข้อสอบที่มีการตรวจให้คะแนนความรู้บางส่วน จะให้ค่าความเที่ยงสูง (Simon, Budescu, & Nevo, 1997) และมีความแม่นยำ (Precision) ในการประมาณค่าพารามิเตอร์ของบุคคลและข้อสอบมากกว่าการตรวจให้คะแนนแบบประเพณีนิยม (Thissen, 1976; Muraki, 1993; Samejima, 1996) จากความพยายามของนักวัดผลทางการศึกษาที่จะแก้ไขปัญหาที่เกิดขึ้น เพื่อที่จะให้คะแนนผลการสอบเป็นตัวแทนระดับความรู้ความสามารถที่แท้จริงของบุคคล และเพิ่มสารสนเทศเกี่ยวกับการตอบข้อสอบในชุดข้อสอบให้มากที่สุด จึงมีวิธีการต่าง ๆ ที่ตอบสนองวัตถุประสงค์ดังกล่าว (Lord & Novick, n.d. cited in Budescu et al., 1997) สรุปได้ดังนี้

1. การให้น้ำหนักคะแนนข้อสอบที่แตกต่างกัน (Differential item weighting) วิธีนี้ถึงแม้ว่าจะมีการกำหนดน้ำหนักคะแนนรายข้อที่เหมาะสม แต่ก็ยังวางอยู่บนพื้นฐานของการตอบและให้คะแนนเฉพาะข้อที่ตอบถูกแบบเดิม คือ ตอบถูกได้ 1 คะแนน ตอบผิดได้ 0 คะแนน ซึ่งให้สารสนเทศเกี่ยวกับความรู้ความสามารถของผู้สอบน้อยมาก ไม่สามารถวัดความรู้ในส่วนที่ผู้สอบมีอยู่เพียงบางส่วนออกมาได้ คะแนนที่ได้จึงไม่สามารถที่จะใช้แทนการวัดระดับความรู้ความสามารถที่แท้จริงของผู้สอบได้
2. การให้น้ำหนักตัวเลือกที่แตกต่างกัน (Differential option weighting) หลักการนี้ความรู้บางส่วนสามารถวัดได้จากการให้น้ำหนักคะแนนแก่ตัวเลือก ซึ่งการให้น้ำหนักคะแนนตัวเลือกทำได้ 2 วิธี วิธีแรกใช้วิจารณ์ญาณของผู้เชี่ยวชาญ ส่วนวิธีที่สองเป็นการให้น้ำหนักจากข้อมูลเชิงประจักษ์ที่ผ่านมาในอดีตหรือปัจจุบัน โดยอาศัยสิ่งที่น่าสนใจของตัวเลือก เช่น คะแนนมาตรฐานเฉลี่ยของผู้สอบที่เลือกแต่ละตัวเลือก ความสัมพันธ์ระหว่างตัวเลือกที่ผู้สอบเลือกกับคะแนนรวมทั้งหมด จากผลการศึกษาพบว่า วิธีนี้มีประโยชน์เล็กน้อยในแง่เพิ่มความเที่ยงแบบคงที่ภายใน (Internal consistency reliability) นอกจากนี้ก็ยังไม่เป็นที่นิยม เพราะต้องใช้ทรัพยากรมาก การคำนวณการให้คะแนนมีวิธีการที่ยุ่งยาก และยากในการอธิบายและตัดสินใจการให้คะแนนแก่ผู้สอบ
3. การเปลี่ยนแปลงโครงสร้างข้อสอบ (Changing the item structure) วิธีนี้มีข้อจำกัดคือสร้างข้อสอบให้มีความเหมาะสมได้ยาก เมื่อได้วิเคราะห์วิธีการเปลี่ยนแปลงรูปแบบข้อสอบ พบว่า

มีข้อจำกัดหลายประการ เช่น ในเรื่องของเวลาที่ใช้ในการทำข้อสอบ มีความยุ่งยากในการสร้างข้อสอบให้เหมาะสม และมีความแปรเปลี่ยนในด้านความเที่ยงและความตรงของผลการวัด

4. การเปลี่ยนแปลงวิธีตอบข้อสอบ (Changing the response method) วิธีการนี้จะแตกต่างจากวิธีที่กล่าวมาทั้งหมด คือ ใช้น้ำหนักคะแนนที่ให้โดยผู้สอบเอง ซึ่งสะท้อนถึงความรู้ที่มีอยู่ในตัวผู้สอบ เป็นการประเมินตนเอง เป็นวิธีที่ดีที่สุดในการได้มาซึ่งสารสนเทศเกี่ยวกับความรู้ของผู้สอบในแต่ละข้อ (Budesu et al., 1997) ผู้สอบจะให้สารสนเทศอย่างสมบูรณ์เกี่ยวกับน้ำหนักคะแนนที่ให้ การให้คะแนน และระดับความมั่นใจในการตอบ ซึ่งจะเป็นการช่วยลดการเดาคำตอบและความคลาดเคลื่อนในการวัด วิธีการนี้จะมีการเปลี่ยนแปลงวิธีการตอบ และการให้คะแนนแตกต่างจากชุดข้อสอบชนิดเลือกตอบธรรมดา ซึ่งมีอยู่หลายวิธี ได้แก่ การแก้การเดา การให้คะแนนจากการตัดตัวลวง การทดสอบความน่าจะเป็น การแสดงความมั่นใจ การจัดอันดับอย่างสมบูรณ์ และการจัดอันดับเป็นบางส่วน ซึ่งในแต่ละวิธีนั้น จะพิจารณาการให้คะแนนแก่ผู้สอบแตกต่างกัน

จากการศึกษาการให้คะแนนแบบเปลี่ยนแปลงวิธีการตอบข้อสอบเพื่อให้สามารถวัดความรู้ของผู้สอบ พบว่า แต่ละวิธีมีแนวโน้มที่จะให้สารสนเทศเกี่ยวกับความรู้ของผู้สอบได้อย่างน่าเชื่อถือ แต่บางวิธีก็มีข้อจำกัด เช่น มีความยุ่งยากในการปฏิบัติ ใช้ทรัพยากรจำนวนมาก การสอบต้องใช้เวลามาก แนวคิดเชิงทฤษฎีในการวัดความรู้ที่แท้จริงของผู้สอบ ฮัทชินสัน (Hutchinson, 1982) ได้จำแนกเป็น 2 แนวทาง คือ แนวทางแรก การแสดงออกซึ่งความรู้ของบุคคลที่มีต่อข้อสอบจะมีเพียง 2 ลักษณะ คือ รู้คำตอบกับไม่รู้คำตอบ ซึ่งแนวทางนี้คือ ทฤษฎีการให้คะแนนแบบจำกัด (Finite state score theory) ซึ่งได้นำมาเป็นกรอบแนวคิดในวิธีการตอบและการให้คะแนนในชุดข้อสอบเลือกตอบที่มีตัวเลือกถูก 1 ตัว นอกจากนั้นเป็นตัวลวง การตอบจะให้ผู้สอบเลือกตัวเลือกที่ถูกต้องเพียงตัวเลือกเดียว เมื่อตอบถูกจะได้ 1 คะแนน ถ้าตอบผิดจะได้ 0 คะแนน และแนวทางที่สอง การแสดงออกซึ่งความรู้ของผู้สอบควรจะเป็นลักษณะต่อเนื่อง ซึ่งความรู้ของผู้สอบเป็นตัวแปรต่อเนื่อง ผู้สอบจะมีความรู้ในช่วงตั้งแต่รู้คำตอบไปจนถึงไม่รู้คำตอบ เป็นลักษณะต่อเนื่องกัน ซึ่งเป็นแนวของทฤษฎีการกระจายแบบต่อเนื่อง (Continuous distribution theory) ซึ่งการตอบข้อสอบของผู้สอบที่ไม่ใช่การเลือกคำตอบที่ถูกต้องเพียงคำตอบเดียว แต่จะต้องแสดงได้ว่าเขาเป็นผู้ที่รู้จริง รู้บางส่วน หรือไม่รู้เลย การให้คะแนนจะต้องกำหนดวิธีให้คะแนนที่แสดงถึงระดับความรู้ของบุคคลได้ ชอปปีน (Choppin, 1971 cited in Leclercq, 1983, pp. 161-287) ได้เสนอว่า กระบวนการวัดผลที่ให้ผู้สอบได้แสดงความรู้บางส่วนเป็นกระบวนการที่สอดคล้องกับโมเดลในการคิดของผู้สอบ ซึ่งกระบวนการสอบและการให้คะแนนมีความสัมพันธ์ใกล้ชิดกับโมเดลทางความรู้ (Epistemological model) ทั้งนี้เพราะสิ่งที่วัดก็คือ ความรู้ ซึ่งชอปปีนได้เสนอโมเดลที่เกี่ยวข้องกับกระบวนการคิดในการทำแบบสอบและการให้คะแนนแบบสอบเลือกตอบ

ไว้ 3 โมเดล ได้แก่ โมเดลแรก ได้เสนอว่า “เมื่อผู้ตอบรู้คำตอบ เขาก็จะเลือกตัวเลือกที่เป็นคำตอบ ที่ถูกต้องในข้อนั้น แต่เมื่อผู้ตอบไม่รู้คำตอบที่ถูกต้อง เขาก็จะเดาสุ่มในระหว่างตัวเลือกที่ให้มา ทั้งหมด” โมเดลที่สอง เสนอว่า “เมื่อผู้ตอบรู้คำตอบ เขาจะเลือกตัวเลือกที่เป็นคำตอบที่ถูกต้อง ในข้อนั้น แต่เมื่อผู้ตอบไม่รู้คำตอบ เขาจะตัดตัวเลือกที่เขารู้ว่าไม่ถูกต้องออกบ้าง และเขาจะเลือก อย่างสุ่มในระหว่างตัวเลือกที่เหลือ” โมเดลที่สาม เสนอว่า “ขั้นความรู้ของผู้ตอบไม่สามารถแบ่ง ได้ เป็นรู้คำตอบกับไม่รู้คำตอบเท่านั้น แต่มีความต่อเนื่องของขั้นความรู้จากขั้นที่ไม่รู้อย่างสมบูรณ์ ไปสู่ขั้นที่รู้อย่างสมบูรณ์ ซึ่งในช่วงระหว่าง 2 ขั้น ของความรู้นี้จะเรียกกันว่า ความรู้บางส่วน ซึ่งเมื่อผู้ตอบเผชิญกับปัญหาที่เป็นตัวเลือกต่าง ๆ เขาจะตอบสนองต่อตัวเลือกนั้นแต่ละตัว ตามค่าความน่าจะเป็นของการที่จะเป็นคำตอบที่ถูกต้อง แต่ด้วยคำชี้แจงในการตอบที่เขาเลือก เพียงตัวเลือกเดียว เขาจึงต้องเลือกตัวเลือกที่มีความน่าจะเป็นว่าจะถูกต้องมากที่สุด” เมื่อวิเคราะห์ ถึงโมเดลที่แสดงกระบวนการคิดเพื่อตอบแบบสอบเลือกตอบที่น่าเสนอ โดยชอบป็นทั้ง 3 โมเดล พบว่า โมเดลที่สองและโมเดลที่สามได้สนับสนุนแนวคิดในเรื่องการแสดงออกซึ่งความรู้ของผู้ตอบ แบบสอบเลือกตอบ เป็นลักษณะต่อเนื่อง ซึ่งสอดคล้องกับแนวคิดของฮัทชินสัน ดังนั้น การตอบ แบบสอบเลือกตอบ ผู้ตอบจะต้องแสดงได้ว่าเป็นผู้รู้จริง รู้บางส่วน หรือ ไม่รู้เลย ซึ่งได้มีนักวัดผล ทางการศึกษาได้นำแนวคิดตามทฤษฎีนี้ไปใช้ในการออกแบบการวัดความรู้ที่ไม่สมบูรณ์ของผู้ตอบ แบบสอบเลือกตอบ อีเบล (Ebel, 1972, pp. 102-104) และธอร์นไคค์ (Thomdike, 1969, p. 58) ให้ความเห็นตรงกันว่า ข้อสอบที่ดีที่สุด คือ ข้อสอบแบบเลือกตอบ (Multiple choices test) เพราะเป็นข้อสอบชนิดที่ได้รับความนิยมมากที่สุด สามารถสร้างเป็นข้อสอบแบบมาตรฐานได้ง่าย วัดได้ครอบคลุมทั้งเนื้อหา จุดประสงค์ และสมรรถภาพทางสมองได้ทุกระดับ ทั้งยังสามารถ ตรวจให้คะแนนได้ง่าย ให้ความเป็นธรรมสูง หรือใช้คอมพิวเตอร์ช่วยตรวจได้ แต่ข้อสอบ แบบเลือกตอบที่มีคำตอบถูกต้องที่สุดเพียงตัวเลือกเดียวมีจุดอ่อนคือ การเปิดโอกาสให้ผู้สอบเดาได้ง่าย นักวัดผลจึงพยายามปรับปรุงแก้ไขเพื่อให้ผลที่ได้จากการวัดด้วยข้อสอบเลือกตอบมีประสิทธิภาพ ยิ่งขึ้น (Coombs, Milholland, & Womer, 1956, pp. 13-17) หลากหลายวิธี แต่การเปลี่ยนแปลง วิธีตอบข้อสอบ โดยใช้ข้อสอบเลือกตอบแบบถูกผิด (Multiple True-False: MTF) มีจุดเด่น ผู้สอบจะต้องตอบให้ได้ว่าตัวเลือกใดบ้างเป็นตัวเลือกที่ถูก และตัวเลือกใดบ้างเป็นตัวเลือกที่ผิด นั่นคือ จะต้องพิจารณาตัดสินคำตอบทุกตัวเลือก ซึ่งการให้ตอบลักษณะนี้เป็นการวัดที่คล้ายกับ การวัดความคิดรวบยอด โดยผู้สอบจะต้องสามารถแยกแยะสิ่งที่เป็นความคิดรวบยอดกับสิ่งที่ไม่เป็น ความคิดรวบยอดออกจากกันให้ได้ จึงจะถือว่ารู้จริง (Roid & Haladyna, 1982, p. 150) และข้อสอบ เลือกตอบแบบถูกผิดน่าจะเป็นเครื่องมือที่สามารถวัดส่วนของความรู้ (Partial knowledge) ของผู้สอบ ได้ใกล้เคียงกับความเป็นจริงมากยิ่งขึ้น ในการวัดผลทางการศึกษานั้น ต้องอยู่บนพื้นฐาน

แนวคิดที่ว่า พฤติกรรมแสดงออกกำหนดมาจากลักษณะภายในตัวบุคคล ซึ่งสังเกตไม่ได้โดยตรง ต้องอาศัยสิ่งเร้ามากระตุ้นให้บุคคลแสดงพฤติกรรมตอบสนอง เครื่องมือการวัดผลสัมฤทธิ์ทางการศึกษาที่ดีนั้น ต้องสามารถใช้เป็นตัวแทนเนื้อหาการเรียนรู้ของผู้เรียน เปรียบเสมือนสิ่งเร้า นำไปกระตุ้นผู้เรียนให้แสดงพฤติกรรมตอบสนอง ผลการตอบหาคือ สาระสนเทศเกี่ยวกับผลสัมฤทธิ์การศึกษาตามหลักสูตรกำหนด หรือความสามารถทางการเรียนรู้ด้านต่าง ๆ ที่ต้องการให้เกิดขึ้นภายในตัวผู้เรียน ถ้าได้รับสารสนเทศมากเพียงพอ ระดับความสามารถภายในตัวผู้เรียน จะประมาณได้อย่างแม่นยำและลดการตัดสินใจผิดพลาด ประกอบกับการให้คะแนนที่เหมาะสมกับกระบวนการทดสอบโดยการเลือกข้อสอบที่เหมาะสมกับระดับความสามารถหรือคุณลักษณะของผู้สอบ และสนองต่อการวัดผลความสามารถของผู้สอบในเชิงจิตวิทยาจะเพิ่มคุณภาพการวัดให้มีความแม่นยำ นอกจากนี้ยังได้รับการยอมรับว่า การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Computerized Adaptive Testing: CAT) สามารถลดระยะเวลาของการทดสอบลงได้ และมีประสิทธิภาพมากกว่าการทดสอบแบบดั้งเดิม (Segall, 2005, p. 429; ศิริชัย กาญจนวาสี, 2550, หน้า 197; Frey & Seitz, 2009, p. 90) ทั้งนี้ โดยทั่วไปในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ สามารถที่จะลดจำนวนข้อสอบลงได้จากการทดสอบแบบดั้งเดิมครั้งหนึ่ง โดยไม่สูญเสียความแม่นยำในการวัด จึงช่วยให้ลดภาระงาน เวลา และค่าใช้จ่ายลงเป็นอย่างมาก (Frey & Seitz, 2009, p. 89) ในกระบวนการทดสอบโดยใช้ข้อสอบแบบเลือกตอบยังคงเป็นเครื่องมือที่มีความสำคัญและจำเป็นสำหรับใช้ในการประเมินคุณลักษณะ ความสามารถในตัวผู้เรียน โดยเฉพาะคุณลักษณะทางด้านความรู้หรือผลสัมฤทธิ์ทางการเรียน เพราะข้อสอบแบบเลือกตอบยังเป็นเครื่องมือที่สามารถใช้วัดได้ครอบคลุมเนื้อหาคุณลักษณะทางด้านสติปัญญาและวัดผลผู้เรียนได้ครั้งละหลาย ๆ คนพร้อมกัน ประหยัดทั้งเวลา ค่าใช้จ่าย สะดวกในการแปลความหมาย และให้ผลเป็นที่น่าเชื่อถือ ข้อสอบแบบเลือกตอบที่ให้ผลการตรวจสอบทางการศึกษามีความแม่นยำขึ้นได้ และแก้ไขจุดบกพร่องในข้อสอบที่ว่า การตอบคำถามในข้อสอบนั้น ผู้สอบได้อ่านทุกส่วนของแต่ละข้อคำถาม โดยละเอียดหรือไม่ หรืออ่านเฉพาะตัวคำถามและตัวเลือก จนกระทั่งถึงคำตอบที่พอใจเท่านั้น ในข้อนี้ ถ้าผู้สอบไม่ต้องอ่านคำถามและตัวเลือกในแต่ละข้อกระทั่งจนจบ ก็สามารถตอบคำถามได้ถูกต้อง ก็ย่อมแสดงว่า ข้อสอบแบบเลือกตอบที่สร้างขึ้นนั้น เป็นข้อสอบที่ไม่มีคุณภาพ เพราะตัวलगต่าง ๆ ไม่สามารถนำมาใช้ประโยชน์ได้ และอีกประเด็นหนึ่ง ในกรณีที่เนื้อหาวิชา มีความสำคัญมาก การใช้ข้อสอบแบบเลือกตอบจะสามารถกำหนดคำถามได้ครอบคลุมเนื้อหาวิชานั้นหรือไม่ เพราะในแต่ละข้อคำถามของข้อสอบแบบเลือกตอบนั้น มีคำตอบถูกให้เลือกตอบเพียงตัวเดียว จากปัญหานี้ กรอนดัลด์ (Gronlund, 1982, p. 44) ได้สนับสนุนว่า ในการกล่าวเนื้อหาวิชาบางอย่าง บางครั้งก็ไม่เหมาะสมที่จะนำมาสร้างเป็นข้อสอบแบบเลือกตอบที่มีคำตอบถูก

ตัวเดียว ควรจะสร้างให้มีจำนวนตัวเลือกที่ถูกหรือผิดได้มากกว่า 1 ตามความเหมาะสมของเนื้อหาวิชา ได้กว้างขวางขึ้น ซึ่งจากกรณีทั้งสองที่ยกมานี้ จะเห็นว่า ถ้าผู้สอบทราบที่ตัวเลือกมีเพียงตัวเลือกเดียว ผู้สอบก็มีหน้าที่อย่างเดียวกัน คือ หาตัวเลือกที่ถูกเท่านั้น ซึ่งการทำเช่นนั้นไม่ตรงกับจุดประสงค์ของผู้ออกข้อสอบ เพราะถ้าเป็นเช่นนั้นจริง ตัวลวงย่อมไม่มีคุณค่า การเขียนข้อสอบที่ดีต้องสามารถช่วยแก้ปัญหาในเรื่องนี้ได้ โดยในการสร้างข้อสอบต้องทำให้ผู้สอบอ่านคำถามและตัวเลือกทุกตัว เพื่อพิจารณาว่าตัวเลือกใดถูกหรือผิดบ้าง นอกจากนี้ ยังสามารถสร้างข้อสอบได้ครอบคลุมเนื้อหาได้มากกว่าข้อสอบแบบเลือกตอบแบบที่มีคำตอบถูกที่สุดเพียงคำตอบเดียว เพราะไม่ได้กำหนดแน่นอนว่าในแต่ละข้อคำถามจะต้องมีจำนวนตัวเลือกถูกหรือผิดเท่าใด ซึ่งเป็นการเปิดกว้างในการเขียนรูปแบบคำถามและการจัดตัวเลือกให้เหมาะสม นอกจากนี้ อีเบล (Ebel, 1972, pp. 37-44) ได้กล่าวว่า ข้อสอบลักษณะนี้ คำถามไม่จำเป็นต้องเป็นการถามความคิดรวบยอด คำนิยม ความสัมพันธ์ หรือขบวนการ แต่อาจจะเป็น “ข้อความใดถูกหรือข้อความใดผิด” ซึ่งข้อความที่กำหนดให้อาจเป็นรายละเอียดของเรื่องใดเรื่องหนึ่ง หรือเหตุการณ์ที่สำคัญทางประวัติศาสตร์ ข้อสอบที่มีคุณภาพจะสามารถจำแนกผู้เรียน และให้ความยุติธรรมแก่ผู้เรียนได้เป็นอย่างดี เพราะข้อสอบมีความชัดเจนหรือมีความเป็นปรนัยในการให้คะแนนสูง ในขณะที่เครื่องมือวัดผลชนิดอื่น ๆ มีความชัดเจนในการให้คะแนนน้อยกว่า ความแม่นยำในการประมาณค่าความสามารถต่ำ จึงอาจเกิดความคลาดเคลื่อน ในการให้คะแนนได้มากกว่า โดยเฉพาะถ้าผู้ให้คะแนนมีความลำเอียง หรือไม่มีความรอบคอบในการให้คะแนนยังมีโอกาสทำให้ขาดความยุติธรรมต่อตัวผู้สอบมากยิ่งขึ้น วิธีการให้คะแนนที่เหมาะสมจะทำให้คุณภาพของการวัดในด้านความตรงและความเที่ยงสูงขึ้น การให้คะแนนบางส่วนของผู้สอบจะแสดงถึงตำแหน่งจริงของความรู้ของผู้สอบ ซึ่งมีผลช่วยลดความคลาดเคลื่อนจากแหล่งการเดาสุ่มของผู้สอบลง ทำให้คะแนนมีความเที่ยงมากขึ้น ในขณะเดียวกัน ก็ส่งผลต่อความตรงของกระบวนการวัดผลด้วย (พรทิพย์ ไชยโส, 2534)

กระบวนการวัดผลทางการศึกษาต้องอาศัยการทดสอบที่มีประสิทธิภาพ และเครื่องมือที่มีคุณภาพ จึงจะทำให้ได้ข้อมูลหรือคะแนนที่ถูกต้อง เชื่อถือได้ ข้อสอบเลือกตอบจัดว่าเป็นเครื่องมือที่มีบทบาทมากกว่าเครื่องมืออื่น ๆ ที่ใช้ในการวัดผลการศึกษา ผลการวิจัยเกี่ยวกับการทดสอบที่มีประสิทธิภาพพบในหลายประเด็นที่น่าสนใจ เช่น เรื่อง โมเดลของการตอบสนอง ข้อสอบ พบว่า โมเดลการตอบสนองข้อสอบตามทฤษฎีการตอบสนองข้อสอบแบบ โลจิสติก แบบสามพารามิเตอร์นั้น มีความเหมาะสมกว่าโมเดลอื่น ๆ (Urry, 1970; 1977; Koch & Reckase, 1978; McKinley & Reckase, 1980) สำหรับในส่วนที่เกี่ยวกับคลังข้อสอบนั้น พบว่า ขนาดของคลังข้อสอบส่งผลต่อการประมาณค่าความสามารถของผู้สอบน้อยมาก (Ho, 1989, pp. 421-A) และขนาดของคลังข้อสอบนั้นควรมีประมาณ 100-200 ข้อ และคลังข้อสอบที่มีข้อสอบ 116-150 ข้อ

จะให้ผลดีที่สุด (Weiss, 1988, pp. 374-375) ซึ่งสอดคล้องกับยูริ (Urry, 1970; 1977) ที่พบว่า คลังข้อสอบที่ใช้ควรมีอย่างน้อย 100 ข้อ ส่วนข้อสอบในคลังข้อสอบนั้น ควรมีค่าอำนาจจำแนกมากกว่า 0.80 ค่าความยากง่ายอยู่ระหว่าง -2.00 ถึง 2.00 และค่าสัมประสิทธิ์การเดาน้อยกว่า 0.30 นอกจากนี้ โฮ (Ho, 1989, pp. 421-A) ยังพบว่า ชนิดของคลังข้อสอบ (มีข้อสอบแบบสุ่ม ข้อสอบที่มีค่าอำนาจจำแนกสูง และข้อสอบที่ง่าย ๆ) นั้น มีปฏิสัมพันธ์กับวิธีการประมาณค่าความสามารถของผู้สอบ สำหรับระบบคลังข้อสอบทางการศึกษา สมศักดิ์ ลีลา (2539, หน้า 38) กล่าวว่า ระบบคลังข้อสอบที่ดี จะทำให้ครูมีข้อสอบที่ดีจำนวนมากไว้ใช้ สามารถเลือกข้อสอบมาสร้างเป็นแบบทดสอบได้ตรงกับจุดประสงค์ที่จะวัดในโอกาสต่าง ๆ ทั้งในการสอบก่อนเรียน ระหว่างเรียน และหลังการเรียน ได้อย่างรวดเร็ว สามารถสร้างแบบทดสอบได้หลาย ๆ ฉบับ ทำให้สะดวกในการสอบซ้ำ เช่น ในกิจกรรมการสอนซ่อมเสริม สะดวกในการสอบนักเรียนโดยใช้แบบทดสอบคู่ขนานกัน และสามารถจัดข้อสอบให้มีความยากเหมาะสมกับผู้สอบได้ ซึ่งจะมีประโยชน์ต่อการพัฒนาและการจัดการเรียนการสอนอีกด้วย สำหรับในประเด็นที่เกี่ยวกับระดับความยากง่ายของข้อสอบที่ใช้เมื่อเริ่มต้นในการสอบนั้น ลอร์ด (Lord, 1977; 1980, p. 153) พบว่า ค่าความยากง่ายของข้อสอบข้อแรก que ที่เลือกให้แก่ผู้สอบนั้น ไม่มีผลต่อความถูกต้อง ในการประมาณค่าความสามารถของผู้สอบ แต่ โฮ (Ho, 1989, pp. 421-A) พบว่า ถ้าข้อสอบข้อแรกมีความยากง่ายเท่ากับหรือต่ำกว่าความสามารถของผู้สอบแล้ว จะทำให้การประมาณค่าความสามารถของผู้สอบมีความถูกต้องมากยิ่งขึ้นในด้านวิธีการประมาณค่าความสามารถของผู้สอบนั้น ไวส์ และแม็คไบรด์ (Weiss & McBride, 1984, p. 274) พบว่า วิธีการประมาณค่าความสามารถ ตามวิธีของเบส์ที่ปรับปรุงใหม่ (Bayesian updating) เป็นวิธีการประมาณค่าที่ไม่ลำเอียง ค่าย เชียงฉี (2534) พบว่า วิธีการประมาณค่าความสามารถของผู้สอบตามวิธีของเบส์ที่ปรับปรุงใหม่ (Bayesian updating) และวิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (Conditional maximum likelihood) ให้ผล ไม่ต่างกัน ซึ่งให้ผลสอดคล้องกับข้อค้นพบของสแกจส์ และสตีเวนสัน (Skaggs & Stevenson, 1989, pp. 391-402) ที่ได้ระบุว่า ถ้าผู้สอบมากกว่า 2,000 คนแล้ว ผลที่ได้รับจากการประมาณค่าความสามารถทั้ง 2 วิธี ไม่แตกต่างกัน แต่ถ้ามีผู้สอบน้อยกว่า 500 คนแล้ว วิธีการของเบส์จะมีความคงที่ และถูกต้องมากกว่า และถ้าข้อสอบที่ใช้ในการสอบน้อยกว่า 15 ข้อแล้ว วิธีความเป็นไปได้สูงสุด จะมีความถูกต้องมากกว่า จากที่กล่าวมา มีบางส่วนสอดคล้องกับข้อค้นพบของกิฟฟอร์ด และสวามินาธาน (Gifford & Swaminathan, 1990, p. 33) พบว่า ถ้ากลุ่มผู้สอบมีขนาดเล็กและข้อสอบที่ใช้มีจำนวนน้อยข้อแล้ว การประมาณค่าความสามารถด้วยวิธีของเบส์แบบร่วมกัน (Joint Bayesian) จะมีความถูกต้องมากกว่า นอกจากนี้ การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์นั้น ยังพบว่า ผู้สอบมีเจตคติที่ดีต่อการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์มากกว่าการทดสอบแบบปรับเหมาะแบบเขียน

ตอบและการทดสอบแบบดั้งเดิม (Wisniewski, 1985; Garrison & Baumgarten, 1986; Lee, 1987; Blackmore, 1987; Carroll, 1990; Vispoel, 1993; งามนิตย์ ชาติทอง และปรีชา เครือวัลย์, 2527) สำหรับทางด้านประสิทธิภาพของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ พบว่า สูงกว่า การทดสอบแบบประเพณีนิยม (Weiss, 1982, pp. 473-492; Morino et al., 1984, pp. 155-163; Garrison & Baumgarten, 1986) และเมื่อนำการทดสอบแบบปรับเหมาะแบบเขียนตอบ (Paper and pencil adaptive testing) มาทำการเปรียบเทียบกับ การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Computerized Adaptive Testing; CAT) ก็พบว่า การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ มีประสิทธิภาพสูงกว่าการทดสอบแบบปรับเหมาะแบบเขียนตอบ (Urry, 1977, p. 181)

จากการศึกษาการให้คะแนนเพื่อประมาณค่าความรู้ที่ผ่านมาส่วนใหญ่ให้ค่าสารสนเทศของคะแนนได้ไม่ครอบคลุมความสามารถที่แท้จริงของผู้สอบ ประกอบกับวิธีการให้คะแนนข้อสอบแบบเลือกตอบในปัจจุบันยังไม่อยู่บนพื้นฐานหลักของความน่าจะเป็น คะแนนต้องให้ความยุติธรรมแก่ผู้สอบที่มีความรู้เต็ม (Full knowledge) ที่ควรได้คะแนนสูงสุด โดยให้มีค่าเท่ากับ 1 หรือผู้ที่ไม่มีความรู้ (Null knowledge) ที่ควรได้คะแนนต่ำสุด โดยให้มีค่าเท่ากับ 0 ซึ่งอยู่บนหลักของความน่าจะเป็นเกี่ยวกับการประมาณค่าความสามารถของผู้สอบในการตอบข้อสอบถูกในข้อนั้น และสำหรับผู้ที่มีความรู้เพียงบางส่วน (Partial knowledge) ที่ควรได้คะแนนระหว่าง 0 ถึง 1 เป็นค่าที่บ่งบอกถึงระดับความรู้ที่เขามีอยู่บางส่วนอย่างแท้จริง การให้คะแนนต้องหลีกเลี่ยงการลงโทษหรือให้คะแนนคิดลบ ในกรณีที่ผู้สอบไม่มีความรู้หรือมีความรู้ผิด (Misinformation) หรือผู้สอบแสดงคำตอบที่ไม่เป็นไปตามเงื่อนไขของคำชี้แจง แม้ผู้สอบจะตอบเป็นไปตามเงื่อนไขของคำชี้แจงการสอบไปได้บ้าง ซึ่งจะถือว่าคำตอบที่ตอบนั้น มาจากการที่เขาไม่มีความรู้ที่แท้จริง คะแนนที่ได้กรณีนี้ควรเป็น 0 คะแนน และยังสามารถป้องกันปัญหาคะแนนที่ได้ในแต่ละข้อไม่ให้ไปมีผลกระทบต่อคะแนนในข้ออื่น หรือเกิดการชดเชยคะแนนกันเมื่อนำไปหาคะแนนรวม

จากการศึกษางานวิจัย พบว่า การทดสอบที่มีประสิทธิภาพต้องเป็นกระบวนการบูรณาการศาสตร์แห่งการวัดผลที่มีการให้คะแนนที่เหมาะสมกับความนำสมัยของเทคโนโลยีการทดสอบเข้าด้วยกัน จะทำให้การทดสอบมีความคล่องตัว สะดวก มีประสิทธิภาพสูงขึ้น ยังผลให้ได้สารสนเทศที่ตอบสนองต่อความต้องการในการพัฒนามนุษย์อย่างเสมอภาคและเที่ยงธรรม นับตั้งแต่เริ่มมีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์เป็นต้นมา ผู้ที่เกี่ยวข้องในวงการวัดผลการศึกษาที่มีความสนใจวิธีดำเนินการสอบลักษณะนี้ต่างก็ทำการศึกษา ค้นคว้า และพัฒนารูปแบบของการทดสอบแบบปรับเหมาะมากขึ้น สำหรับในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ จะมีองค์ประกอบที่สำคัญ ได้แก่ การคัดเลือกข้อสอบจากคลังข้อสอบ (Item selection rule) การประมาณค่าความสามารถของผู้สอบ (Ability estimation procedure) และการยุติการทดสอบ

(Termination criteria) และการวิจัยครั้งนี้ผู้วิจัยมีความสนใจศึกษาผลของการให้คะแนนที่มีต่อประสิทธิภาพการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ทั้งนี้เพราะว่าในปัจจุบัน พบว่าการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ส่วนใหญ่ประมาณค่าความสามารถของผู้สอบจากผลการให้คะแนนแบบสองค่า (0-1) หรือถูก 1 ผิด 0 ที่อาจเกิดความคลาดเคลื่อนในการวัดได้ง่าย จึงน่าจะมีการศึกษาเปรียบเทียบถึงการให้คะแนนดังกล่าวกับการให้คะแนนแบบอื่นว่าจะส่งผลต่อประสิทธิภาพของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ในด้านจำนวนข้อสอบและค่าฟังก์ชันสารสนเทศของชุดข้อสอบหรือไม่ มากหรือน้อยเพียงไร และเพื่อเป็นการนำเสนอทางเลือกใหม่ของการทดสอบ ที่มุ่งศึกษาถึงการประมาณความสามารถของบุคคลในการอธิบายคะแนนการวัดผลที่ใกล้เคียงกับความสามารถจริงของบุคคลให้มากที่สุดจากผลการทดสอบที่มีประสิทธิภาพ และเป็นการต่อยอดแนวคิดของนักวัดผลอื่น ๆ ที่ได้เคยคิดไว้ อันจะทำให้เกิดความเชื่อมั่นในการนำผลของการให้คะแนนไปใช้กับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ต่อไป

วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาผลของการให้คะแนนที่มีต่อคุณภาพการวัดผลด้านความตรง ความเที่ยง และความคลาดเคลื่อนมาตรฐานในการวัด
2. เพื่อเปรียบเทียบผลของการให้คะแนนที่มีต่อประสิทธิภาพการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ด้านจำนวนข้อสอบ และค่าฟังก์ชันสารสนเทศของชุดข้อสอบ เมื่อกำหนดให้การคัดเลือกข้อสอบ การประมาณค่าความสามารถผู้สอบ การยุติการทดสอบ และความสามารถของผู้สอบต่างกัน

คำถามการวิจัย

ผลของการให้คะแนนที่มีต่อประสิทธิภาพการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ มีคำถามที่น่าสนใจศึกษา ดังต่อไปนี้

1. ผลของการให้คะแนนที่ต่างกันจะส่งผลต่อคุณภาพการวัดในด้านความตรง (Validity) ความเที่ยง (Reliability) และความคลาดเคลื่อนมาตรฐานในการวัด (Standard error of measurement) อย่างไร
2. ผลของการให้คะแนนที่ต่างกันจะส่งผลต่อประสิทธิภาพของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ในด้านจำนวนข้อสอบ (Length testing) และค่าฟังก์ชันสารสนเทศของชุดข้อสอบ (Test information function) อย่างไร

สมมติฐานของการวิจัย

การให้คะแนนที่เหมาะสมจะทำให้คุณสมบัติของการวัดในประเด็นที่เกี่ยวข้องกับความเที่ยงและความตรงสูงขึ้น กระบวนการวัดที่ให้ผู้สอบได้แสดงความรู้ส่วนย่อยเป็นกระบวนการที่สอดคล้องกับโมเดลในการคิดของผู้สอบ (Choppin, 1971 cited in Leclercq, 1983, pp. 161-287) ทำให้คะแนนที่ได้ของผู้สอบแสดงถึงตำแหน่งจริงของความรู้ของผู้สอบ ซึ่งมีส่วนช่วยลดความคลาดเคลื่อนจากแหล่งการเดาสุ่มของผู้สอบลง ทำให้คะแนนมีความเที่ยงเพิ่มขึ้น ในขณะเดียวกันก็ส่งผลต่อความตรงของกระบวนการวัดนั้นด้วย (Fraty, 1980) ผลของการให้คะแนนที่มีการพิจารณาให้คะแนนส่วนย่อยของความรู้แก่ผู้สอบจึงส่งเสริมให้คุณภาพของกระบวนการวัดดีขึ้น การพิจารณาคุณภาพของทดสอบจึงสามารถอ้างอิงได้จากความเที่ยงและความตรงของผลการวัดที่ได้จากคะแนนที่มีการวัดส่วนย่อยของความรู้ของผู้สอบในการวิจัย และจากการศึกษางานวิจัยที่เกี่ยวข้องกับการวัดความรู้ส่วนย่อยในชุดข้อสอบแบบเลือกตอบ ส่วนใหญ่จะให้ค่าความเที่ยงสูง (Simon et al., 1997) และมีความแม่นยำ (Precision) ในการประมาณค่าพารามิเตอร์ของบุคคลและข้อสอบมากกว่า การตรวจให้คะแนนแบบวัดความรู้สมบูรณ์หรือแบบประเพณีนิยม (Samejima, 1976; Thissen, 1976; Muraki, 1993)

ดังนั้น เพื่อศึกษาประสิทธิภาพของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ที่มีการให้คะแนนแตกต่างกัน ผู้วิจัยจึงได้ตั้งสมมติฐานของการวิจัยและเหตุผลของการตั้งสมมติฐานแยกเป็นประเด็นต่าง ๆ ไว้ โดยสอดคล้องกับวัตถุประสงค์ในการวิจัย และเพื่อให้มองเห็นภาพของผลที่อาจเกิดขึ้นต่อตัวแปรตามได้ชัดเจนยิ่งขึ้น ผู้วิจัยจึงได้กล่าวถึงผลของตัวแปรตาม 2 ตัว อันเนื่องมาจากตัวแปรอิสระ 1 ตัว ไว้ในสมมติฐานข้อเดียวกัน ดังนี้

1. การคัดเลือกข้อสอบตามเกณฑ์ที่ต่างกัน ได้แก่

1.1 ค่าความยากง่ายของข้อสอบ (b) กับค่าระดับความสามารถที่ประมาณได้ของผู้สอบ ($\hat{\theta}$) [Match b_i to $\hat{\theta}$]

1.2 ค่าความสามารถที่ทำให้ข้อสอบมีสารสนเทศสูงสุด (θ_{max}) กับค่าความสามารถที่ประมาณได้ของผู้สอบ ($\hat{\theta}$) [Match Q_{max} to $\hat{\theta}$]

1.3 ข้อสอบที่มีค่าสารสนเทศสูงสุด (maximum item information, I_{max})

การศึกษาผลของการให้คะแนนที่ต่างกัน จะส่งผลต่อจำนวนข้อสอบ และค่าฟังก์ชันสารสนเทศของชุดข้อสอบนั้นน่าจะต่างกัน ซึ่งการคัดเลือกข้อสอบที่มีค่าสารสนเทศสูงสุดน่าจะใช้ข้อสอบน้อยกว่าการคัดเลือกข้อสอบด้วยวิธีการอื่น ทั้งนี้เนื่องจากข้อสอบที่ถูกเลือกขึ้นมาใช้ในแต่ละวิธีนั้น จะมีค่าสารสนเทศของข้อสอบแตกต่างกันไป ซึ่งค่าสารสนเทศดังกล่าวจะส่งผลต่อค่าความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ ดังนั้น จึงทำให้วิธีการคัดเลือก

ข้อสอบแต่ละวิธีผู้เข้าหาค่าความคลื่อนในการประมาณค่าที่กำหนดที่เป็นเกณฑ์ยุติการสอบต่างกัน ซึ่งการใช้จำนวนข้อสอบที่ต่างกันจะส่งผลให้การคัดเลือกข้อสอบที่มีค่าสารสนเทศสูงสุด น่าจะทำให้ค่าฟังก์ชันสารสนเทศของชุดข้อสอบมีค่าสูงด้วย

2. การประมาณค่าความสามารถของผู้สอบที่ต่างกัน ได้แก่

2.1 วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (Conditional maximum likelihood)

2.2 วิธีของเบส์ที่ปรับใหม่ (Bayesian updating)

การศึกษาผลของการให้คะแนนที่ต่างกัน จะส่งผลต่อจำนวนข้อสอบ และค่าฟังก์ชันสารสนเทศของชุดข้อสอบนั้นน่าจะต่างกัน วิธีของเบส์ที่ปรับใหม่ น่าจะใช้ข้อสอบน้อยกว่าวิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข เนื่องจากมีข้อค้นพบว่า ถ้ากลุ่มตัวอย่างมีขนาดใหญ่ (มากกว่า 2,000 คน) และทดสอบโดยใช้ข้อสอบมากกว่า 35 ข้อแล้ว การประมาณค่าความสามารถตามวิธีการของเบส์ และวิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข จะให้ผลการประมาณค่าไม่ต่างกัน และมีความถูกต้องสูง และถึงแม้ว่ากลุ่มผู้สอบมีขนาดเล็ก (น้อยกว่า 500 คน) วิธีการของเบส์จะให้การประมาณที่มีความคงเส้นคงวา (Consistency) และมีความถูกต้องมากกว่าวิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไขก็ตาม แต่ก็ไม่ทำให้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าต่างกัน เพราะการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ครั้งนี้ ได้กำหนดให้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถเป็นเกณฑ์ยุติการทดสอบ ซึ่งจะมีค่าเท่ากันสำหรับผู้สอบทุกคน แต่จะส่งผลให้จำนวนข้อสอบต่างกัน เพราะที่ระดับความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถเดียวกันนั้น วิธีการประมาณค่าความสามารถตามวิธีการของเบส์จะใช้ข้อสอบน้อยกว่าวิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (ต่าย เชียงฉิน, 2534, หน้า 122)

3. การยุติการทดสอบโดยใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า

(Standard error of estimation) ที่ต่างกัน ได้แก่ $SEE \leq 0.30$, $SEE \leq 0.40$, $SEE \leq 0.50$

การศึกษาผลของการให้คะแนนที่ต่างกัน จะส่งผลต่อจำนวนข้อสอบ และค่าฟังก์ชันสารสนเทศของชุดข้อสอบนั้นน่าจะต่างกัน เพราะถ้าใช้ค่าความคลาดเคลื่อนมาตรฐานในการวัดความสามารถที่มีค่าต่ำ ๆ เป็นเกณฑ์ยุติการทดสอบ จะใช้จำนวนข้อสอบมากกว่าใช้ค่าความคลาดเคลื่อนมาตรฐานในการวัดความสามารถที่มีค่าสูง ๆ ดังนั้น เมื่อกำหนดค่าความคลาดเคลื่อนในการประมาณค่าความสามารถแตกต่างกันเป็นเกณฑ์ยุติการทดสอบ จะทำให้เมื่อจำนวนข้อสอบที่ใช้เพิ่มขึ้น ค่าความคลาดเคลื่อนในการประมาณค่าก็จะเกิดการเปลี่ยนแปลง ดังนั้น หากกำหนดค่าความคลาดเคลื่อนมาตรฐานให้มีค่าต่างกัน จำนวนข้อสอบที่ถูกนำมาใช้ในการทดสอบ และค่าฟังก์ชันสารสนเทศของชุดข้อสอบก็จะต่างกัน

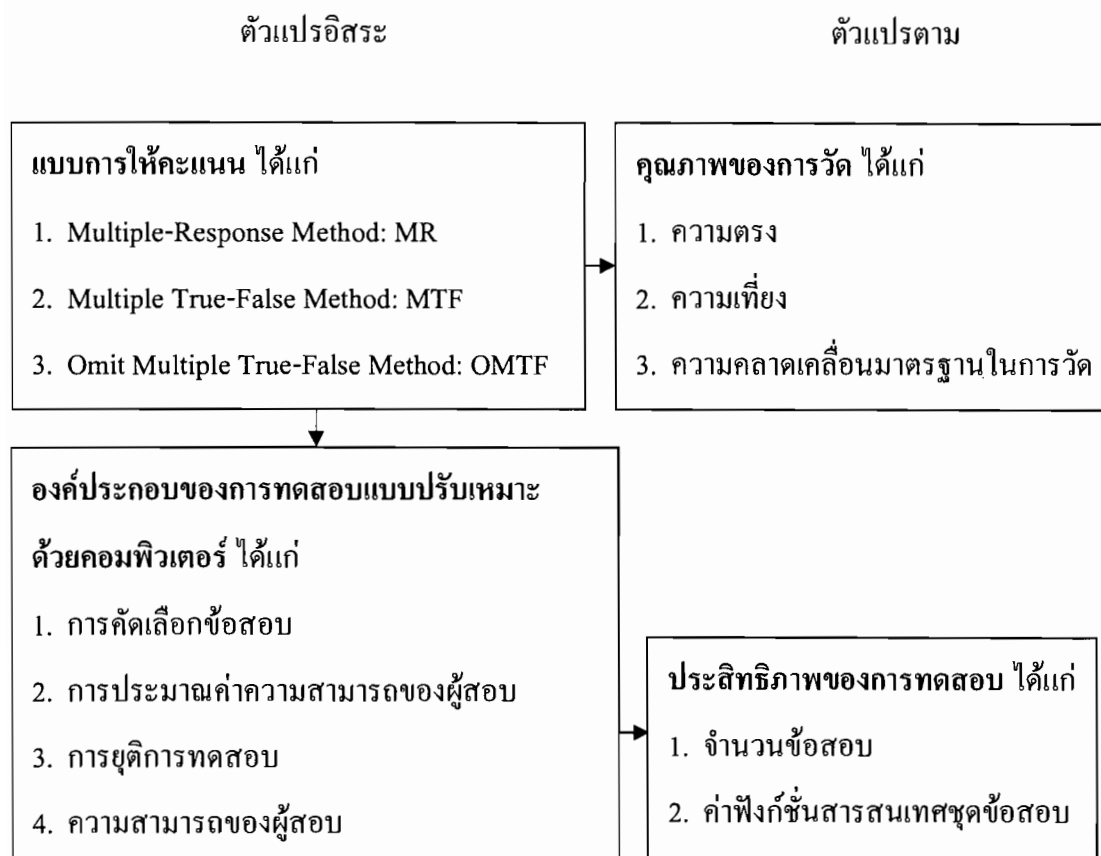
4. ความสามารถของผู้สอบที่ต่างกัน ได้แก่ ผู้สอบที่มีความสามารถทางคณิตศาสตร์สูงปานกลาง และต่ำ ซึ่งการศึกษาผลของการให้คะแนนที่ต่างกัน จะส่งผลต่อจำนวนข้อสอบและค่าฟังก์ชันสารสนเทศของชุดข้อสอบนั้นน่าจะต่างกัน ทั้งนี้เพราะว่าผู้สอบที่มีความสามารถสูงน่าจะใช้จำนวนข้อสอบน้อยกว่าผู้ที่มีความสามารถต่ำ และค่าฟังก์ชันสารสนเทศของชุดข้อสอบน่าจะสูงกว่าผู้ที่มีความสามารถต่ำ ทั้งนี้เนื่องจากการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์นั้นมีวัตถุประสงค์เพื่อวัดคุณลักษณะของผู้สอบ โดยใช้ข้อสอบที่มีความเหมาะสมกับความสามารถของผู้สอบ โดยผู้สอบที่มีความสามารถสูงก็จะได้รับข้อสอบที่ยาก สำหรับผู้สอบที่มีความสามารถต่ำก็จะได้รับข้อสอบที่ง่าย ดังนั้น จึงทำให้เกิดจำนวนข้อสอบ และค่าฟังก์ชันสารสนเทศของชุดข้อสอบจะต่างกัน เพราะผู้สอบที่มีความสามารถต่ำนั้นน่าจะตอบข้อสอบด้วยการเดา ถึงแม้ว่าจะเป็นข้อสอบที่มีความเหมาะสมกับความสามารถของผู้สอบแล้วก็ตาม เมื่อการทดสอบเป็นลักษณะนี้จะทำให้ผู้สอบดังกล่าวต้องตอบข้อสอบมากขึ้น กว่าจะประมาณค่าความสามารถของผู้สอบให้มีความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถเท่ากับที่กำหนดไว้

ความสำคัญของการวิจัย

การศึกษาประสิทธิภาพการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ โดยการนำผลของการให้คะแนนผู้สอบข้อสอบแบบเลือกตอบที่แตกต่างกันมาใช้ในการประมาณค่าความสามารถของผู้สอบ เนื่องจากคะแนนเป็นองค์ประกอบที่สำคัญของกระบวนการวัดผลและการให้คะแนนที่เหมาะสมจะทำให้คุณภาพของการวัดสูงขึ้น กระบวนการวัดที่เปิดโอกาสให้ผู้สอบได้แสดงความรู้ส่วนย่อยเป็นกระบวนการที่สอดคล้องกับ โมเดลในการคิดของผู้สอบ ทำให้คะแนนที่ให้แกผู้สอบแสดงถึงตำแหน่งจริงของความรู้ของผู้สอบ ซึ่งมีส่วนช่วยลดความคลาดเคลื่อนจากแหล่งการเดาสุ่มของผู้สอบลงได้ จึงทำให้คะแนนมีความเที่ยงเพิ่มขึ้น ในขณะเดียวกันก็ส่งผลต่อความตรงของกระบวนการวัดนั้นด้วย นอกจากนี้ ผลการวิจัยยังนำไปใช้เป็นพื้นฐานสำหรับประยุกต์ลักษณะการทดสอบที่มีการนำเอาเทคโนโลยีคอมพิวเตอร์เข้ามาร่วมด้วย ทำให้การจัดคลังข้อสอบมีขนาดเล็กกลง ช่วยในการลดค่าใช้จ่ายในการสร้างข้อสอบ และการบำรุงรักษาคลังข้อสอบทำได้ง่าย จำนวนข้อสอบอาจสามารถปรับให้มีขนาดสั้นลง ทำให้สามารถประหยัดข้อสอบในการสอบ การรักษาความปลอดภัยของข้อสอบยังคงอยู่ด้วยคุณสมบัติของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ สามารถคาดการณ์ วิจัย ตัดสินลักษณะคลังข้อสอบและกำหนดรูปแบบต่าง ๆ โดยยังรักษาประสิทธิภาพสูงสุดในการประมาณค่าความสามารถของผู้สอบ เพื่อใช้ในการจัดการสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ต่อไป ข้อสังเกตที่ได้รับระหว่างการสอบจะนำไปใช้ศึกษาเพื่อเชื่อมโยงความรู้เข้าด้วยกัน สร้างเป็นชุดความรู้ใหม่ สำหรับการทดสอบแบบปรับเหมาะ

ด้วยคอมพิวเตอร์ที่มีการนำเอาผลของการให้คะแนนข้อสอบที่แตกต่างกัน สามารถนำไปประยุกต์ใช้ในสถานการณ์จริง และการวิจัยครั้งนี้ยังสามารถใช้เป็นแนวทางในการศึกษาการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ที่จะเกิดขึ้นในโอกาสต่อไป

กรอบแนวคิดในการวิจัย



ภาพที่ 1 กรอบแนวคิดในการวิจัย

ประโยชน์ที่ได้รับจากการวิจัย

จากการศึกษาวิจัยในครั้งนี้ ทำให้ทราบข้อมูลและได้รับประโยชน์จากผลการวิจัย ดังนี้

1. เป็นการเสนอขอบเขตความรู้เกี่ยวกับการทดสอบที่มีการพิจารณาจากผลของการให้คะแนนที่คำนวณได้จากค่าสังเกตที่วัดได้ เพื่อประมาณระดับความสามารถจริงของผู้สอบได้ใกล้เคียง มีประสิทธิภาพ เป็นแบบที่ไม่ยุ่งยากซับซ้อน และนำไปปฏิบัติได้ในสภาพปกติ

2. ได้ข้อความรู้เกี่ยวกับการให้คะแนนที่เหมาะสมกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ในมิติด้านจำนวนข้อสอบ ค่าฟังก์ชันสารสนเทศของชุดข้อสอบ
3. ได้แนวทางที่ช่วยทำให้ผู้ใช้เครื่องมือวัดผลทางการศึกษาได้ตระหนักถึงการให้คะแนนที่มีคุณภาพในกระบวนการวัด เพื่อใช้ในการตัดสินใจได้อย่างแม่นยำมากขึ้น
4. ใช้เป็นแนวทางการพัฒนาวิธีการให้คะแนนเพื่อประมาณค่าความสามารถในการสอบประเภทอื่น ๆ ต่อไป และยังสามารถนำไปประยุกต์ใช้ร่วมกับการประเมินตามสภาพจริงได้

ขอบเขตของการวิจัย

1. ประชากร

เป็นนักเรียนของโรงเรียนมัธยมศึกษา จังหวัดบุรีรัมย์ ระดับชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2557 จำนวน 9,789 คน (สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 32, 2557)

2. กลุ่มตัวอย่าง

กลุ่มตัวอย่างสำหรับการสร้างคลังข้อสอบเป็นนักเรียนของโรงเรียนมัธยมศึกษา จังหวัดบุรีรัมย์ ระดับชั้นมัธยมศึกษาปีที่ 6 ได้มาด้วยการสุ่มแบบหลายขั้นตอน (Multi stage random sampling) จำนวน 4,096 คน

กลุ่มตัวอย่างสำหรับหาค่าคุณภาพของการวัดผลเป็นนักเรียนของโรงเรียนมัธยมศึกษา จังหวัดบุรีรัมย์ ระดับชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2557 จำนวน 3,330 คน

กลุ่มตัวอย่างหาค่าประสิทธิภาพการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ เป็นนักเรียนโรงเรียนบุรีรัมย์พิทยาคม จังหวัดบุรีรัมย์ ระดับชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2557 จำนวน 540 คน

3. เครื่องมือที่ใช้ในการวิจัย

เครื่องมือการวิจัยครั้งนี้เป็นข้อสอบวัดผลสัมฤทธิ์ทางการเรียนวิชา คณิตศาสตร์ ตามมาตรฐานการเรียนรู้หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ระดับชั้นมัธยมศึกษาตอนปลาย เป็นชุดข้อสอบเลือกตอบที่ผู้วิจัยสร้างขึ้น ซึ่งเป็นข้อสอบที่มีคุณภาพสร้างขึ้นตามหลักวิชาการสร้างข้อสอบ และตรวจสอบคุณภาพวิเคราะห์ค่าพารามิเตอร์ตามแนวทฤษฎีการตอบสนองข้อสอบ

4. ตัวแปรที่เกี่ยวข้อง

การศึกษาประสิทธิภาพการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ในครั้งนี้ มีตัวแปรที่เกี่ยวข้อง คือ ผลคะแนนจากการทดสอบระดับชาติขั้นพื้นฐาน (O-NET) ของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2557 ซึ่งเป็นตัวแปรที่ผู้วิจัยเลือกเพื่อศึกษาความตรงเชิงทฤษฎี

ของการให้คะแนน ตัวแปรที่เลือกต้องมีลักษณะ (Trait) ที่ต้องการวัดมากกว่า 1 ลักษณะ หรือตัวแปรมากกว่า 1 ตัว สำหรับความตรงตามเกณฑ์ของการให้คะแนนที่ผู้วิจัยนำมาศึกษา เพื่อหาความสัมพันธ์กับตัวแปรทำนาย คือ คะแนนจากการสอบ โดยใช้การให้คะแนนแบบ MR แบบ MTF และแบบ OMTF ซึ่งเป็นวิธีการที่มีความสะดวกในการใช้ สามารถที่จะพัฒนาวิธีการตอบข้อสอบและการให้คะแนนในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ได้อย่างยุติธรรม และยังสามารถนำไปตรวจสอบคุณภาพด้านความเที่ยงและความตรงของผลการวัดที่ให้คะแนน ตามวิธีการให้คะแนนที่ผู้วิจัยนำมาศึกษา ส่วนตัวแปรเกณฑ์ คือ คะแนนการสอบ O-NET รายวิชาคณิตศาสตร์ของนักเรียนที่เป็นกลุ่มตัวอย่าง ทั้งนี้ ในการเลือกคะแนน O-NET รายวิชาคณิตศาสตร์ มาเป็นตัวแปรเกณฑ์นั้น เพราะว่า ตัวแปรดังกล่าวมีความคล้ายคลึงกันในด้าน โครงสร้างทางสติปัญญา (Cognitive structure) ทั้งนี้เนื่องมาจากหลักสูตรในรายวิชาคณิตศาสตร์ ซึ่งพัฒนาขึ้นตาม โครงสร้างหลักสูตรแกนกลาง การศึกษาขั้นพื้นฐาน พุทธศักราช 2551 มีมาตรฐานและตัวชี้วัดที่สอดคล้องกับทักษะทางคณิตศาสตร์ เข้าไปในกระบวนการจัดการเรียนการสอน และในรายละเอียดของเนื้อหาวิชา เพื่อต้องการให้ผู้เรียนเป็นผู้มีความสามารถในการแสวงหาความรู้ โดยใช้กระบวนการทางคณิตศาสตร์ต่อไป

5. ตัวแปรที่ใช้ในการวิจัย

5.1 ตัวแปรอิสระ (Independent variables) ประกอบด้วย

5.1.1 การคัดเลือกข้อสอบ ใช้ 3 เกณฑ์ ได้แก่

5.1.1.1 ค่าความยากง่ายของข้อสอบ (b) กับค่าระดับความสามารถที่ประมาณได้ของผู้สอบ ($\hat{\theta}$) [Match b_j to $\hat{\theta}$]

5.1.1.2 ค่าความสามารถที่ทำให้ข้อสอบมีสารสนเทศสูงสุด (θ_{max}) กับค่าความสามารถที่ประมาณได้ของผู้สอบ ($\hat{\theta}$) [Match Q_{max} to $\hat{\theta}$]

5.1.1.3 การใช้ค่าสารสนเทศสูงสุด (Maximum item information, I_{max}) ของข้อสอบ

5.1.2 การประมาณค่าความสามารถของผู้สอบ

5.1.2.1 วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (Conditional maximum likelihood)

5.1.2.2 วิธีการของเบย์ที่ปรับใหม่ (Bayesian updating)

5.1.3 การยุติการทดสอบ ใช้เกณฑ์ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า (Standard Error of Estimation; SEE) 3 เกณฑ์ ได้แก่

5.1.3.1 $SEE \leq 0.30$

5.1.3.2 $SEE \leq 0.40$

5.1.3.3 $SEE \leq 0.50$

5.1.4 ความสามารถทางคณิตศาสตร์ของผู้เข้าสอบ

5.1.4.1 ระดับสูง

5.1.4.2 ระดับปานกลาง

5.1.4.3 ระดับต่ำ

5.2 ตัวแปรตาม (Dependent variables) คือ ประสิทธิภาพการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ได้แก่

5.2.1 จำนวนข้อสอบ (Length testing)

5.2.2 ค่าฟังก์ชันสารสนเทศชุดข้อสอบ (Test information functions)

ข้อตกลงเบื้องต้นของการวิจัย

คะแนนถือว่าเป็นปริมาณที่แสดงความสามารถของผู้สอบเกี่ยวกับเรื่องใดเรื่องหนึ่ง ทั้งที่สมบูรณ์แล้วและที่ยังไม่สมบูรณ์ในการตอบข้อสอบ สำหรับที่ยังไม่สมบูรณ์จะเป็นปริมาณความรู้เพียงบางส่วนที่ผู้สอบสามารถเลือกตัวเลือกถูกและผิด โดยไม่ทราบว่าคำตอบที่ถูกต้องคืออะไร ถือว่าตัวเลือกผิดทุกตัวเกี่ยวข้องกับคำตอบถูก การเลือกตัวเลือกผิดในแต่ละข้อได้ถือว่าเป็นการใช้ความรู้ส่วนย่อยเท่านั้น และในกรณีที่ผู้สอบเลือกตัวเลือกที่เป็นตัวลงได้ไม่ถูกต้อง โดยระบุว่าเป็นตัวเลือกถูก หรือในกรณีที่ผู้สอบเลือกตัวเลือกที่เป็นตัวถูกได้ไม่ถูกต้อง โดยระบุว่าเป็นตัวเลือกผิด จะถือว่าผู้สอบมีความรู้ไม่ถูกต้อง (Misinformation) จะไม่ได้คะแนน และคะแนนของผู้สอบในข้อนี้จะเป็น 0 ซึ่งคะแนนความรู้ทั้งหมดของผู้สอบเป็นผลจากการรวมคะแนนรายข้อที่ได้จากการให้คะแนนแบบ MR แบบ MTF และแบบ OMTF ซึ่งเป็นผลจากการตรวจให้คะแนนของข้อสอบทั้งชุดเข้าด้วยกัน

นิยามศัพท์เฉพาะ

1. การทดสอบปรับเหมาะด้วยคอมพิวเตอร์ (Computerized adaptive testing) หมายถึง การดำเนินการทดสอบที่ใช้ข้อสอบแตกต่างกัน จัดให้กับผู้สอบแต่ละคน โดยการพิจารณาเลือกข้อสอบที่มีค่าความยากพอเหมาะกับความสามารถภายในที่ต้องการวัดผู้สอบ ด้วยโปรแกรมคอมพิวเตอร์ที่ผู้วิจัยพัฒนาขึ้น

2. ข้อสอบเลือกตอบแบบถูกผิด (Multiple true-false item) หมายถึง ข้อสอบที่ให้เลือกรูปแบบหนึ่ง ประกอบด้วยส่วนที่เป็นคำถามและส่วนที่เป็นตัวเลือก ซึ่งการตอบข้อสอบ

ผู้สอบจะต้องแสดงความสามารถสูงสุด ระบุว่าตัวเลือกแต่ละตัวถูก (True) หรือผิด (False) ตามข้อความ สถานการณ์ ปรัชญาการณ ที่ข้อคำถามได้กล่าวถึงไว้ สำหรับการวิจัยนี้มีการตอบดังนี้

- 2.1 กรณีที่รู้คำตอบว่าผิด เป็นการเลือกคำตอบผิดบางตัวได้อย่างถูกต้อง โดยทำเครื่องหมายกากบาท (X) ตรงตัวเลือกที่เป็นตัวผิด
- 2.2 กรณีที่รู้คำตอบว่าถูก เป็นการเลือกตัวเลือกบางตัวที่ถูกได้อย่างถูกต้อง โดยทำเครื่องหมายถูก (✓) ตรงตัวเลือกที่เป็นตัวถูก
- 2.3 กรณีที่ไม่แน่ใจว่าเป็นคำตอบที่ถูกหรือคำตอบที่ผิด จะทำเครื่องหมายไม่แน่ใจ (?) ตรงตัวเลือกนั้น

3. คะแนน (Scoring) หมายถึง การกำหนดค่าคะแนนความรู้แก่ผู้สอบข้อสอบเลือกตอบแบบถูกผิด วิชาคณิตศาสตร์ สำหรับการวิจัยนี้มีแบบการให้คะแนนดังนี้

3.1 แบบ MR (Multiple-Response Method) หรือ MTF-A (Multiple True-False Item with Absolute Knowledge) หรือ 0-1 หมายถึง การให้คะแนนแก่ความรู้ที่ถูกต้องสมบูรณ์ กล่าวคือ ผู้สอบจะต้องตอบได้ตรงกับสภาพความเป็นจริงหรือตรงตามเฉลยทุกตัวเลือก จึงจะได้ 1 คะแนน แต่ถ้าไม่ตรงกับสภาพความเป็นจริงแม้เพียงตัวเลือกเดียวก็จะได้ 0 คะแนน วิธีนี้จึงมีจุดมุ่งหมายเพื่อบ่งชี้ว่าผู้สอบมีความรู้ ความสามารถในเรื่องนั้น ๆ ถูกต้องสมบูรณ์หรือไม่ ถ้าได้ 0 คะแนน ก็แสดงว่า ยังมีบางประเด็นที่บกพร่องอยู่ ดังนั้น คะแนนที่เป็นไปได้ของข้อสอบที่ตรวจให้คะแนนแบบนี้จึงมีค่า 2 ค่า ได้แก่ 0 และ 1 คะแนน

3.2 แบบ MTF (Multiple True-False Method) หมายถึง การตรวจให้คะแนนทุกตัวเลือก ถ้าตัวเลือกใดนักเรียนกาได้ถูกต้องจะได้ 1 คะแนน ถ้ากาผิดจะได้คะแนน 0 ดังนั้น ถ้าข้อสอบมี 4 ตัวเลือก ก็จะได้คะแนนอยู่ระหว่าง 0-4 คะแนน

3.3 แบบ OMTF (Omit Multiple True-False Method) หมายถึง การตรวจให้คะแนนทุกตัวเลือก โดยคำนึงถึงตัวเลือกที่ผู้สอบเว้นไว้จะไม่คิดคะแนน เพื่อให้เกิดความยุติธรรมแก่ผู้สอบที่มีความรู้เต็ม (Full knowledge) ที่ควรได้คะแนนสูงสุด ในที่นี้ ผู้วิจัยให้มีค่าเท่ากับ 4 หรือผู้ที่ไม่มีความรู้ (Null knowledge) ที่ควรได้คะแนนต่ำสุด ในที่นี้ ผู้วิจัยให้มีค่าเท่ากับ 0 ดังนั้น ถ้าข้อสอบมี 4 ตัวเลือก ก็จะได้คะแนนอยู่ระหว่าง 0-4 คะแนน

4. คุณภาพของการวัด (Measurement efficiency) หมายถึง ผลของการให้คะแนนกับผู้สอบข้อสอบเลือกตอบแบบถูกผิด วิชาคณิตศาสตร์ พิจารณาจาก

4.1 ความตรง (Validity) หมายถึง ค่าสหสัมพันธ์ระหว่างผลคะแนนจากการทดสอบระดับชาติขั้นพื้นฐาน (O-NET) ของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2557

กับคะแนนการสอบวิชาคณิตศาสตร์ระดับชั้นมัธยมศึกษาตอนปลาย ตามมาตรฐานการเรียนรู้
หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551

4.2 ความเที่ยง (Reliability) หมายถึง ค่าความเชื่อมั่นแบบสอดคล้องภายใน
ซึ่งเป็นความสอดคล้องกันระหว่างคะแนนรายข้อหรือความเป็นเอกพันธ์ของเนื้อหารายข้อ
อันเป็นตัวแทนของคุณลักษณะเด่นเดียวกับที่ต้องการวัด ประเมินค่าได้จากการคำนวณ
ค่าสัมประสิทธิ์แอลฟา (α) ของครอนบาค (Cronbach's alpha coefficient)

4.3 ความคลาดเคลื่อนมาตรฐานในการวัด (Standard error of measurement)
หมายถึง ค่าเบี่ยงเบนมาตรฐานของผลต่างระหว่างค่าความสามารถจริง (θ) กับค่าที่ได้จาก
การวัดความสามารถ ($\hat{\theta}$) ของผู้เข้าสอบ

5. ประสิทธิภาพของการทดสอบ (Test efficiency) หมายถึง ผลของการให้คะแนน
แบบ MR แบบ MTF และแบบ OMTF สำหรับข้อสอบเลือกตอบแบบถูกผิด พิจารณาจาก

5.1 จำนวนข้อสอบ (Length testing) หมายถึง จำนวนข้อสอบที่ผู้สอบใช้ในการทดสอบ
แบบปรับเหมาะด้วยคอมพิวเตอร์ เมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ
มีค่าเท่ากับหรือน้อยกว่าเกณฑ์ที่กำหนด

5.2 ค่าฟังก์ชันสารสนเทศชุดข้อสอบ (Test information function) หมายถึง ผลรวม
ของค่าสารสนเทศของข้อสอบทุกข้อในชุดข้อสอบที่มีการตรวจให้คะแนนอย่างเดียวกันทั้งฉบับ
โดยแสดงค่าฟังก์ชันสารสนเทศตามช่วงพิสัยของระดับความสามารถของผู้สอบ (θ) ที่แตกต่างกัน
ถ้าสูงที่ระดับ θ ใดก็แสดงว่ามีความแม่นยำสูงในการประมาณค่าพารามิเตอร์ของผู้สอบ ณ ระดับ
 θ นั้น ๆ

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

การวิจัยครั้งนี้ มีวัตถุประสงค์เพื่อเปรียบเทียบผลของการให้คะแนนที่มีต่อประสิทธิภาพการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ผู้วิจัยได้ศึกษาแนวคิด ทฤษฎี หลักการ เอกสาร และงานวิจัยที่เกี่ยวข้อง เพื่อนำความรู้ที่ได้มาเป็นกรอบการวิจัยและนำไปประยุกต์ใช้ในการวิจัยครั้งนี้ โดยเสนอเนื้อหาตามลำดับดังนี้

1. การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Computerized adaptive testing)
2. การให้คะแนนข้อสอบเลือกตอบแบบถูกผิด (Scoring method of multiple true false)
3. ทฤษฎีการตอบสนองข้อสอบ (Item response theory)
4. ประสิทธิภาพของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Efficiency of computerized adaptive testing)
5. งานวิจัยที่เกี่ยวข้อง

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Computerized adaptive testing)

ลักษณะทั่วไปของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ เป็นการจัดการสอบโดยใช้คอมพิวเตอร์ คัดเลือกข้อสอบที่เหมาะสมกับความสามารถของผู้สอบจากคลังข้อสอบ ซึ่งเรียกว่า Tailored test คำนี้ใช้เป็นครั้งแรกโดย Tumbull (1951 cited in Lord, 1980) ซึ่งคลังข้อสอบจะมีข้อสอบที่เหมาะสมกับระดับความสามารถของผู้สอบในระดับต่าง ๆ และมีเนื้อหาที่เกี่ยวข้องกับการทดสอบในแต่ละครั้งนั้น ๆ และข้อสอบแต่ละข้อจะกระจายออกไปตามเนื้อหาและระดับความยากของข้อสอบ โดยข้อสอบจะถูกตรวจสอบความยากง่ายโดยใช้ทฤษฎีการตอบสนองข้อสอบ และมีความรวดเร็วในการระบุความสามารถของผู้สอบที่มีความแม่นยำด้วยข้อสอบเพียงไม่กี่ข้อ (<http://www.ascp.org>) การทดสอบเทเลอร์นั้น เป็นการทดสอบที่จัดข้อสอบให้มีความยากให้เหมาะสมกับระดับความสามารถของผู้สอบแต่ละคน (Lord, 1975, p. 43) และผู้สอบแต่ละคนก็ไม่จำเป็นต้องทำข้อสอบเหมือนกัน ทุกข้อ จำนวนข้อสอบก็ไม่จำเป็นต้องเท่ากัน ทั้งนี้ ขึ้นอยู่กับรูปแบบและผลการตอบข้อสอบในแต่ละการทดสอบเทเลอร์นั้น ๆ (Weiss & Kingsbury, 1984, p. 361) โดยทั่วไปแล้ว จะให้ผู้สอบเริ่มทำข้อสอบข้อที่มีความยากปานกลางก่อน ถ้าผลการตอบข้อสอบถูก ข้อต่อไปจะมีความยากมากขึ้น หากตอบข้อสอบข้อนั้นผิด ข้อต่อไปที่ได้รับก็จะง่ายลง การทดข้อสอบขั้นต่อไปจะดำเนินการ

เช่นนี้ไปจนถึงสิ้นสุดการทดสอบ (Green et al., 1984, p. 347) ต่อมา ไวส์ ได้เสนอให้ใช้คำว่า Adaptive test (Weiss, 1974 อ้างถึงใน ศิริชัย กาญจนวาสี, 2550, หน้า 193)

การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ (Adaptive testing) มีเป้าหมายเพื่อที่จะดำเนินการสอบด้วยการคัดเลือกข้อสอบให้เหมาะสม (Tailoring) กับความสามารถของผู้สอบแต่ละคน โดยอาศัยทฤษฎีการตอบสนองข้อสอบ (Item Response Theory; IRT) มาใช้เป็นพื้นฐานในการคำนวณค่าสารสนเทศของข้อสอบ ซึ่งจะมีการคัดเลือกข้อสอบที่สามารถให้สารสนเทศอันเป็นประโยชน์เกี่ยวกับความสามารถของผู้สอบแต่ละคน ทำให้การทดสอบเป็นไปอย่างมีประสิทธิภาพ ใช้จำนวนข้อสอบที่น้อยลง ประหยัดเวลาของการสอบ มีความคล่องตัวผลที่ได้มีความคลาดเคลื่อนต่ำ และสามารถนำผลมาเปรียบเทียบกันได้อย่างมีความหมาย (Weiss, 1973; ศิริชัย กาญจนวาสี, 2538) ในระยะเริ่มต้นของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ มีการดำเนินการสอบโดยให้ผู้สอบเขียนตอบลงในกระดาษคำตอบ ซึ่งการคัดเลือกข้อสอบข้อต่อไปให้เหมาะกับความสามารถของผู้สอบเป็นเรื่องยุ่งยาก ไม่สะดวก ต่อมาวิทยาการทางคอมพิวเตอร์พัฒนาก้าวหน้ามากขึ้น ดังนั้น การดำเนินการสอบแบบปรับเหมาะจึงได้นำคอมพิวเตอร์เข้ามาใช้ เนื่องจากคอมพิวเตอร์รักษาความลับของข้อสอบได้เป็นอย่างดี การบันทึกคำตอบ การแสดงผลการทดสอบสามารถทำได้ง่าย รวดเร็วกว่าการใช้กระดาษคำตอบและการตรวจข้อสอบแบบเดิม การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบจะต้องมีคลังข้อสอบขนาดใหญ่ ข้อสอบจะต้องมีค่าพารามิเตอร์ประจำตัวของข้อสอบแต่ละข้อ ได้แก่ ค่าอำนาจจำแนก (a) ค่าความยากง่าย (b) ค่าการเดา (c) และค่าสารสนเทศของข้อสอบ (Item information function) เป็นพารามิเตอร์แสดงถึงคุณภาพของข้อสอบไว้ เพื่อความสะดวกในการเลือกใช้ข้อสอบต่อไป (Hambleton & Swaminathan, 1985 อ้างถึงใน รังสรรค์ เล็กมณี, 2540)

การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ (Computerized Adaptive Testing; CAT) เป็นการรวมพัฒนาการของศาสตร์ 2 ด้าน คือ พัฒนาการทางทฤษฎีการตอบสนองข้อสอบ (Item response theory) และพัฒนาการทางเทคโนโลยีซิลิคอนชิป (Silicon chip technology) เข้าด้วยกัน เวียนเนอร์ กล่าวว่า การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์เป็นการสร้างแบบทดสอบโดยการนำคอมพิวเตอร์เข้ามาช่วยในการดำเนินการสอบ เพื่อให้ได้แบบทดสอบที่เหมาะสมกับระดับความสามารถของผู้เข้าสอบ ซึ่งเป็นความพยายามของผู้สร้างแบบทดสอบที่จะเลียนแบบเทคนิคการสร้างแบบทดสอบโดยผู้เชี่ยวชาญด้านการสอบ กล่าวคือ ถ้าผู้เชี่ยวชาญด้านการสอบถามคำถามยากเกินไปสำหรับผู้สอบในคำถามแรก คำถามข้อถัดไปผู้เชี่ยวชาญจะพิจารณาคำถามที่ง่ายขึ้น (Wainer, 1990 cited in Straetmans & Eggen, 1998)

การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์เป็นการทดสอบที่มีวัตถุประสงค์เพื่อปรับข้อสอบให้เหมาะกับความสามารถของผู้สอบ โดยผู้สอบจะใช้เป็นพิมพ์ของคอมพิวเตอร์เพื่อตอบคำถามที่แสดงอยู่บนหน้าจอคอมพิวเตอร์ หลังจากนั้นคอมพิวเตอร์จะประมาณความสามารถของผู้สอบและทำการปรับข้อสอบให้เหมาะสมกับระดับความสามารถหรือผลสัมฤทธิ์ของผู้เข้าสอบแต่ละคน โดยหลักการแล้ว โปรแกรมคอมพิวเตอร์จะเลือกข้อสอบที่ยากขึ้นหลังจากผู้สอบตอบคำถามถูกต้อง หรือจะเลือกคำถามง่ายขึ้นถ้าผู้สอบตอบคำถามผิด การทดสอบแบบปรับเหมาะนี้ สามารถใช้วัดผลทางจิตวิทยาได้อย่างมีประสิทธิภาพ มีความแม่นยำในการประมาณค่าความสามารถของผู้สอบ และรายงานผลการสอบได้ทันที โดยใช้ข้อสอบจำนวนน้อยลงและใช้เวลาในการสอบลดลงประมาณครึ่งหนึ่งของการทดสอบแบบมาตรฐานเดิม แต่บรรลุผลในระดับเดียวกัน ทั้งนี้ เนื่องมาจากลักษณะของการทดสอบแบบปรับเหมาะจะเลือกเฉพาะข้อสอบที่มีความสอดคล้องเหมาะสมกับความสามารถของผู้สอบ และจะตัดข้อสอบส่วนใหญ่ที่มีความยากง่ายไม่เหมาะสมกับผู้สอบออกไป (Meijer & Nering, 1999; ศิริชัย กาญจนวาสี, 2538)

ลักษณะการทำงานของ การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

สตรีทแมน และเอจเจน (Straetmans & Eggen, 1998) กล่าวว่า การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ (CAT) เป็นการสร้างแบบทดสอบให้เหมาะสมกับระดับความสามารถของผู้เข้าสอบ ซึ่งเป็นความพยายามที่จะพัฒนาเทคนิคการทดสอบที่ชาญฉลาด ซึ่งในช่วงปี ค.ศ. 1980-1989 ได้มีการรวมพัฒนาการศาสตร์ 2 ด้าน นั่นคือ ทฤษฎีการตอบสนองข้อสอบ (Item response theory) และเทคโนโลยี Silicon chip (Silicon chip technology) เข้าด้วยกัน เพื่อใช้ประโยชน์ในการทดสอบการนำคอมพิวเตอร์มาใช้ในการทดสอบ ให้ประโยชน์ดังนี้

1. สามารถนำเสนอข้อสอบที่ตรงกับความสามารถของผู้สอบได้อย่างถูกต้องรวดเร็ว
2. สามารถนำเสนอวิธีการทดสอบที่รวมเอาภาพกราฟฟิก เสียงวีดิโอเคลื่อนไหว และเนื้อหาสาระเข้าไว้ด้วยกันจะคล้ายคลึงกับวิถีชีวิตจริง
3. สามารถแสดงผลคะแนนการสอบของผู้เข้าสอบและรายงานผลขั้นสุดท้ายทันทีที่สิ้นสุดการทดสอบ จะประหยัดทั้งเวลาและแรงงาน
4. การจัดชุดแบบทดสอบให้กับผู้เข้าสอบจึงเป็นการแบ่งเบาภาระแก่ครูผู้สอน
5. สามารถประมวลผลที่มีความรวดเร็วถูกต้องและมีประสิทธิภาพมาก
6. การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์สามารถประเมินผลคุณลักษณะทางจิตวิทยาได้เที่ยงตรงน่าเชื่อถือ โดยใช้จำนวนข้อสอบและเวลาตอบน้อยกว่าเมื่อเทียบกับการทดสอบแบบมาตรฐานเดิม

ส่วนขั้นตอนของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ (CAT) มีดังนี้

1. ทำการคัดเลือกข้อสอบจากธนาคารข้อสอบ
2. นำเสนอข้อสอบให้นักเรียนตอบบนหน้าจอคอมพิวเตอร์
3. นักเรียนตอบคำถามโดยการพิมพ์หรือเลือกคำตอบบนหน้าจอคอมพิวเตอร์
4. ทำการประเมินคำตอบว่าตอบถูกหรือตอบผิด
5. ถ้าตอบผิดจะคัดเลือกข้อสอบข้อต่อไปง่ายขึ้นถ้าตอบถูกข้อสอบก็จะยากขึ้น
6. การทดสอบจะสิ้นสุดลงเมื่อเป็นไปตามกฎเกณฑ์ของการยุติการสอบ

กระบวนการของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จะต้องมีองค์ประกอบสำคัญดังนี้ (Straetmans & Eggen, 1998)

1. ธนาคารข้อสอบ (A calibrated item bank)

การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ (CAT) มีความจำเป็นอย่างยิ่งที่จะต้องมีธนาคารข้อสอบ ซึ่งเป็นระบบการจัดข้อสอบที่มีคุณภาพ ธนาคารข้อสอบจะเป็นที่รวมของข้อสอบที่สามารถค้นหาได้ง่าย (Accessible) ข้อสอบจะถูกแยกประเภทหรือจัดระบบตามลักษณะที่จะนำไปใช้งานได้ง่ายในการทดสอบแต่ละครั้ง ข้อสอบในธนาคารข้อสอบโดยปกติแล้วจะจัดแยกประเภทตามเนื้อหา ลักษณะข้อคำถาม ลักษณะการปฏิบัติการ เชื่อมโยงกับข้อสอบข้ออื่น ๆ หรือจัดแยกตามวัสดุสื่อผู้เขียนประวัติการทดสอบ และลักษณะทางจิตวิทยา รวมทั้งระดับความยากของข้อสอบ สำหรับข้อสอบที่รวบรวมไว้จะมีสารสนเทศเกี่ยวกับระดับความยากของข้อสอบซึ่งเป็นเรื่องที่สำคัญมากในทฤษฎีมาตรฐานเดิม (Classical test theory) ค่าดัชนีที่ใช้อธิบายความยากของข้อสอบคือค่า p ซึ่งเป็นสัดส่วนของกลุ่มผู้เข้าสอบกับผู้ที่ตอบข้อสอบนี้ถูก ค่า p ต่ำ หมายถึงข้อสอบยาก และค่า p สูง หมายถึง ข้อสอบง่าย อย่างไรก็ตาม ค่า p จะขึ้นอยู่กับความสามารถของกลุ่มที่สอบข้อสอบนั้น ค่า p ของกลุ่มข้อสอบที่ใช้กับนักเรียนกลุ่มที่มีความสามารถสูงจะแตกต่างอย่างมากกับค่า p ของข้อสอบชุดเดียวกันที่ใช้สอบกับนักเรียนกลุ่มที่มีความสามารถต่ำ ด้วยเหตุนี้เองทำให้ค่า p ตามทฤษฎีมาตรฐานเดิมไม่เหมาะสมสำหรับการอธิบายระดับความยากของข้อสอบในธนาคารข้อสอบ ในขณะที่เดียวกันก็ไม่สามารถรับรองได้ว่าระดับความสามารถของกลุ่มที่ทำแบบทดสอบแต่ละชุดนั้นจะสอดคล้องกับระดับความสามารถของกลุ่มตัวอย่างที่น่าจะนำมาใช้คำนวณค่า p ดังนั้น ทฤษฎีการตอบสนองข้อสอบ (Item response theory) สามารถแก้ปัญหานี้ได้ โดยการประมาณค่าพารามิเตอร์ความยากซึ่งไม่ขึ้นอยู่กับระดับความสามารถของกลุ่มที่นำมาพัฒนา (Weiss et al., 1990) ในทฤษฎีการตอบสนองข้อสอบ ค่าพารามิเตอร์ความยากของข้อสอบแต่ละชุดนั้นเป็นอิสระในแต่ละกลุ่ม โอกาสของการตอบถูก

แตกต่างกันไปตามระดับความสามารถของแต่ละบุคคล นั่นคือ โอกาสของผู้ที่มีความสามารถสูงในการตอบข้อสอบได้ถูกต้องจะสูงกว่าผู้ที่มีความสามารถต่ำ

2. ขั้นตอนในการทดสอบ (Testing algorithm)

กระบวนการในการทดสอบเป็นหัวใจสำคัญของการทดสอบแบบปรับเหมาะ ซึ่งจะต้องทราบว่า จะเริ่มต้นดำเนินการสอบและสิ้นสุดการทดสอบอย่างไร

2.1 การเริ่มต้นการทดสอบ (Starting)

การทดสอบแบบปรับเหมาะจะให้ความสำคัญกับค่าความยากของข้อสอบข้อต่อไปที่จะนำมาเสนอแก่ผู้สอบ การพิจารณาความยากของข้อสอบจะพิจารณาจากการประมาณค่าความสามารถของผู้เข้าสอบในขณะนั้น เมื่อผู้เข้าสอบตอบข้อสอบในแต่ละครั้ง ความสามารถของผู้สอบจะถูกประมาณค่าใหม่ทุกครั้งเสมอ เมื่อเริ่มต้นการทดสอบมักจะไม่มีข้อมูลเกี่ยวกับความสามารถของผู้เข้าสอบ ดังนั้น การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ จึงคัดเลือกข้อสอบที่มีความยากปานกลางเสนอแก่ผู้เข้าสอบเป็นข้อแรกในการเริ่มต้นการสอบ หรืออาจคัดเลือกข้อสอบที่ง่ายเป็นข้อสอบในการเริ่มต้นสอบเพื่อลดความวิตกกังวลและสร้างแรงจูงใจในการสอบของผู้เข้าสอบ

2.2 การดำเนินการสอบ (Continuing)

การดำเนินการสอบด้วยการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ ด้วยคอมพิวเตอร์จะต้องคำนึงถึงประสิทธิภาพของการทดสอบ กล่าวคือ การทดสอบจะต้องใช้ข้อสอบน้อยข้อและลดเวลาที่ใช้ในการทดสอบลงเพื่อลดความเครียดและความเหนื่อยล้าจากการทดสอบ แต่ผลการทดสอบสามารถประมาณค่าความสามารถของผู้สอบได้ตรงกับระดับความสามารถที่แท้จริง ดังนั้น ในขั้นตอนของการดำเนินการสอบจะต้องกระทำอย่างรัดกุม กระบวนการคัดเลือกข้อสอบจะต้องสามารถคัดเลือกข้อสอบที่มีสารสนเทศ โดยเฉพาะค่าความยากจะต้องสอดคล้องกับค่าความสามารถของผู้สอบที่ถูกประมาณค่าครั้งสุดท้ายมากที่สุด การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์สามารถคัดเลือกข้อสอบเสนอแก่ผู้สอบได้ถูกต้องรวดเร็ว การทดสอบแบบปรับเหมาะจึงเป็นกระบวนการทดสอบที่มีประสิทธิภาพสูง

2.3 การสิ้นสุดการทดสอบ (Stopping)

การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบจะสิ้นสุดลงเมื่อผลการสอบที่ผ่านมาเป็นไปตามเกณฑ์ยุติการทดสอบ ซึ่งเกณฑ์การยุติการทดสอบนั้น อาจจะถูกกำหนดโดยใช้เกณฑ์จำนวนข้อสอบ เกณฑ์เวลาที่ใช้ในการสอบ เกณฑ์คะแนนจุดตัด (Cut off score) เพื่อแยกระดับความสามารถของผู้สอบออกเป็นกลุ่มรอบรู้และกลุ่มไม่รอบรู้ และเกณฑ์ความคลาดเคลื่อน

มาตรฐานในการประมาณค่าความสามารถของผู้สอบเป็นไปตามเกณฑ์ยุติการทดสอบที่ตั้งไว้ การทดสอบของแต่ละคนก็จะสิ้นสุดลง

ความแตกต่างระหว่างการทดสอบแบบเดิมที่ใช้กระดาษเขียนตอบและการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์

สตรีทแมน และเอจเจน (Straetmans & Eggen, 1998) กล่าวว่าแบบทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แตกต่างจากการทดสอบที่ใช้กระดาษเขียนตอบพอสรุปได้ดังตารางที่ 1

ตารางที่ 1 ความแตกต่างระหว่างการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์กับการทดสอบที่ใช้กระดาษเขียนตอบ

ประเด็น	การทดสอบที่ใช้กระดาษเขียนตอบ	การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์
ความยาก	ปรับความสามารถของผู้เข้าสอบให้อยู่ในระดับปานกลาง	ปรับความสามารถของผู้เข้าสอบเป็นรายบุคคล
จำนวนข้อสอบ	กำหนดและข้อสอบมีจำนวนมากข้อ	ไม่กำหนดและข้อสอบจำนวนน้อยข้อ
ความปลอดภัย	เมื่อผู้เข้าสอบไม่ได้สอบในเวลาเดียวกัน อาจจะมีปัญหาเกี่ยวกับความปลอดภัยของข้อสอบ	ป้องกันความปลอดภัยของข้อสอบได้ดีกว่า เพราะผู้เข้าสอบแต่ละคนจะใช้ข้อสอบที่ต่างกัน
ความสมดุลของเนื้อหา	ไม่มีปัญหา	ไม่ได้ขึ้นอยู่กับเนื้อหา
การควบคุมการบริหารจัดการ	ผู้เข้าสอบอยู่ในการควบคุม เป็นภาระการทำงานหนัก	คอมพิวเตอร์อยู่ในการควบคุม สามารถกระจายการทำงานได้

1. ความยาก (Difficulty)

การทดสอบที่ใช้กระดาษเขียนตอบ โดยเฉพาะข้อสอบแบบเลือกตอบ มีการใช้กันอย่างแพร่หลายในปัจจุบัน นักเรียนทุกคนจะต้องทำการทดสอบในเวลาเดียวกันด้วยข้อสอบชุดเดียวกัน ระดับความยากของแบบทดสอบจะถูกปรับตามค่าเฉลี่ยความสามารถของกลุ่มผู้เข้าสอบซึ่งมีผลทำให้ผู้เข้าสอบที่มีความสามารถสูงหรือต่ำกว่าค่าเฉลี่ยเกิดความสะเพร่าหรือมีความเครียดเป็นอย่างมาก ผู้เข้าสอบจะเผชิญกับข้อสอบที่ยากหรือง่ายเกินไป ตามหลักจิตวิทยาแล้ว ความถูกต้อง

ของการวัดผลจากการสอบจะลดลง ส่วนการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ผู้เข้าสอบแต่ละคนจะได้รับแบบทดสอบที่แตกต่างกัน ซึ่งเป็นแบบทดสอบที่ไม่ยากหรือง่ายเกินไป ดังนั้นแบบทดสอบจึงทำลายความรู้สึกของผู้เข้าสอบ ผู้เข้าสอบไม่วิตกกังวล ความถูกต้องของการวัดของผู้เข้าสอบแต่ละคนจะมีค่าสูง

2. จำนวนข้อสอบ (Test length)

การทดสอบที่ใช้กระดาษเขียนตอบจะกำหนดจำนวนข้อสอบของแบบทดสอบตายตัว จำนวนข้อสอบของแบบทดสอบนั้นจะถูกกำหนดจากกลุ่มผู้เข้าสอบทั้งหมด ผู้เข้าสอบทุกคนจะต้องใช้ข้อสอบชุดเดียวกัน ถ้าความสามารถของกลุ่มผู้เข้าสอบกระจายเป็น โค้งปกติแล้วแบบทดสอบที่ดีจะต้องมีความยากกระจายไปให้ครอบคลุมความสามารถของกลุ่มผู้เข้าสอบด้วย ดังนั้น จึงมีข้อสอบที่ไม่เหมาะกับความสามารถของผู้เข้าสอบแต่ละคนรวมอยู่ด้วย ส่วนการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ผู้เข้าสอบแต่ละคนจะได้รับแบบทดสอบที่แตกต่างกัน ซึ่งจะถูปรับตามความสามารถของผู้เข้าสอบ หลังจากทำข้อสอบในแต่ละข้อแล้ว ความสามารถของผู้เข้าสอบแต่ละคนจะถูกประมาณค่าใหม่ทุกครั้ง และเมื่อค่าความสามารถของผู้สอบจากการประมาณครั้งหลังสุดเป็นไปตามเกณฑ์ยุติการสอบ การทดสอบก็จะสิ้นสุดลง

3. ความปลอดภัย (Security)

การทดสอบที่ใช้กระดาษเขียนตอบผู้เข้าสอบจะต้องทำแบบทดสอบชุดเดียวกันในเวลาเดียวกันด้วย จึงจะทำให้ข้อสอบไม่รั่ว ถ้าทำการทดสอบในเวลาที่แตกต่างกัน ความเชื่อมั่นของการทดสอบจะลดลง แต่การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สามารถแก้ปัญหานี้ได้ โดยการให้ผู้เข้าสอบแต่ละคนทดสอบโดยใช้ข้อสอบที่ต่างกัน ในเวลาที่ต่างกันได้ด้วย แต่ผลการสอบยังคงเปรียบเทียบกันได้

4. ความสมดุลของเนื้อหา (Balancing of content)

การสร้างแบบทดสอบที่ใช้กระดาษเขียนตอบจะสร้างตามแผนการออกข้อสอบ ทั้งนี้ เพื่อให้เกิดความมั่นใจว่าข้อสอบที่เขียนออกมานั้นเป็นตัวแทนของความรู้และทักษะที่กำหนด ในจุดประสงค์ของหลักสูตรแต่ข้อสอบ ในแบบทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์จะถูกคัดเลือกโดยยุทธศาสตร์การคัดเลือกจากสารสนเทศสูงสุด เนื้อหาในแบบทดสอบจึงถูกกำหนดโดยการพิจารณาตามหลักการทางจิตวิทยา ดังนั้น ผู้เข้าสอบจะได้รับข้อสอบที่มีความสมดุล ถึงแม้ว่าการทดสอบจะสิ้นสุดลงหลังจากที่ทำแบบทดสอบที่มีจำนวนข้อสอบเพียงไม่กี่ข้อก็ตาม

5. การควบคุม (Locus of control)

การทดสอบที่ใช้กระดาษเขียนตอบผู้ใช้ข้อสอบจะต้องตัดสินใจว่าคำตอบที่ได้จากแบบทดสอบจะนำไปทำอะไร มีความเป็นไปได้ที่จะมีคำตอบที่แตกต่างกันในคำถามเดียวกัน จึงอาจมีการทดสอบโดยให้เปิดหนังสือตอบและมีการทบทวนคำตอบต่าง ๆ ซึ่งปกติแล้ว ผู้บริหารการสอบจะเป็นผู้อนุญาตให้ทำในลักษณะใดลักษณะหนึ่ง ส่วนการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ คอมพิวเตอร์จะเป็นผู้ตัดสินใจว่าคำถามข้อต่อไปควรจะเป็นอะไร การควบคุมกระบวนการทดสอบจะรัดกุมมากยิ่งขึ้น

6. การบริหารจัดการ (Logistics)

การทดสอบที่ใช้กระดาษเขียนตอบจะต้องมีครูผู้ควบคุมการสอบหรือผู้ที่เกี่ยวข้องกับการสอบจำนวนมาก แบบทดสอบจะนำมาจากธนาคารข้อสอบ และถ่ายทำสำเนาต้องมีการใช้ห้องสอบและจ้างผู้ควบคุมการสอบ หลังจากทำการทดสอบจะมีการเก็บรวบรวมกระดาษคำตอบและข้อสอบพร้อมทั้งทำเครื่องหมายต่าง ๆ ไว้ ส่วนการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ คอมพิวเตอร์จะดึงงานต่าง ๆ เหล่านี้มาจากครู และลดอุปสรรคในการทดสอบลง

ประเภทของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ

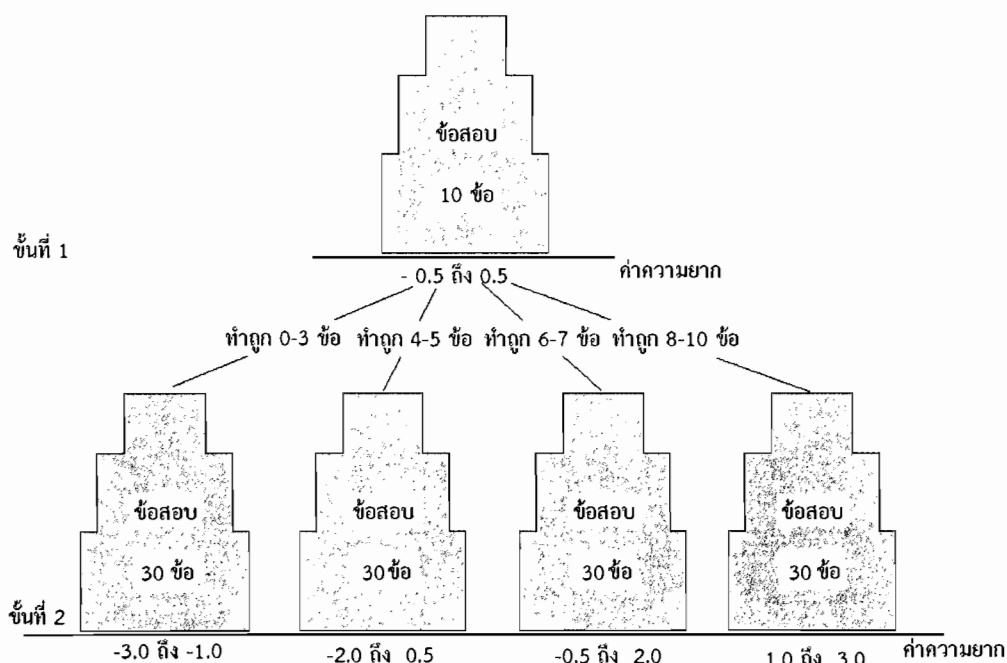
ไวส์ แฮมเบิลตัน และสวามินาทาน (Weiss, 1974; Hambleton & Swaminathan, 1985)

จำแนกการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบออกเป็น 2 ประเภทใหญ่ ๆ คือ

1. การทดสอบสองขั้นตอน (Two-stage strategies)
2. การทดสอบหลายขั้นตอน (Multi-stage strategies)
 - 2.1 แบบแยกทางคงที่ (Fixed branching model)
 - 2.1.1 รูปพีระมิดขนาดขั้นคงที่ (Constant step size pyramid)
 - 2.1.2 รูปพีระมิดขนาดขั้นแปรผัน (Variable step size pyramid)
 - 2.1.3 รูปพีระมิดข้างตัด (Truncated pyramid)
 - 2.1.4 รูปพีระมิดที่มีหลายข้อในแต่ละขั้น (Multiple item pyramid)
 - 2.1.5 รูปพีระมิดที่ให้น้ำหนักแก่ตัวเลือก (Differential response option branching)
 - 2.1.6 แบบทดสอบเฟล็กซิเลเวล (Flexilevel test)
 - 2.2 แบบแยกทางแปรผัน (Variable branching model)
 1. การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบสองขั้นตอน

การทดสอบสองขั้นตอนเป็นการทดสอบโดยใช้ข้อสอบ 2 ชุด การดำเนินการสอบเริ่มด้วยให้ผู้สอบทำแบบทดสอบเพื่อแยกทาง (Routing test) ผลจากการทดสอบนี้จะเห็นตัวชี้วัดว่าผู้สอบจะได้รับแบบทดสอบชุดใดในขั้นที่ 2 ซึ่งเรียกแบบทดสอบในขั้นนี้ว่า แบบทดสอบเพื่อวัดผล

(Measurement test) โดยหลักการแล้ว ผู้สอบที่ได้คะแนนน้อยจากการทดสอบแยกทางถือว่าเป็นผู้ที่มีความสามารถต่ำ จึงจะได้รับข้อสอบเพื่อวัดผลชุดที่ง่าย ส่วนผู้ที่ได้คะแนนจากการทดสอบเพื่อแยกทางมากหรือมีความสามารถสูงจะได้รับข้อสอบเพื่อวัดผลชุดที่ยากขึ้น ดังภาพที่ 2



ภาพที่ 2 โครงสร้างของการทดสอบเฉพาะบุคคลแบบสองขั้นตอน

จากภาพที่ 2 จะเห็นว่า มีข้อสอบเพื่อแยกทางในชั้นที่ 1 จำนวน 10 ข้อ ถ้าผู้สอบตอบคำถามถูก 0-3 ข้อ จะได้ทำข้อสอบเพื่อวัดผลในชั้นที่ 2 ในชุดที่ง่ายที่สุด จำนวน 30 ข้อ ส่วนผู้ที่มีความสามารถสูงขึ้นไปก็จะได้รับข้อสอบในชุดที่ยากขึ้นตามลำดับ

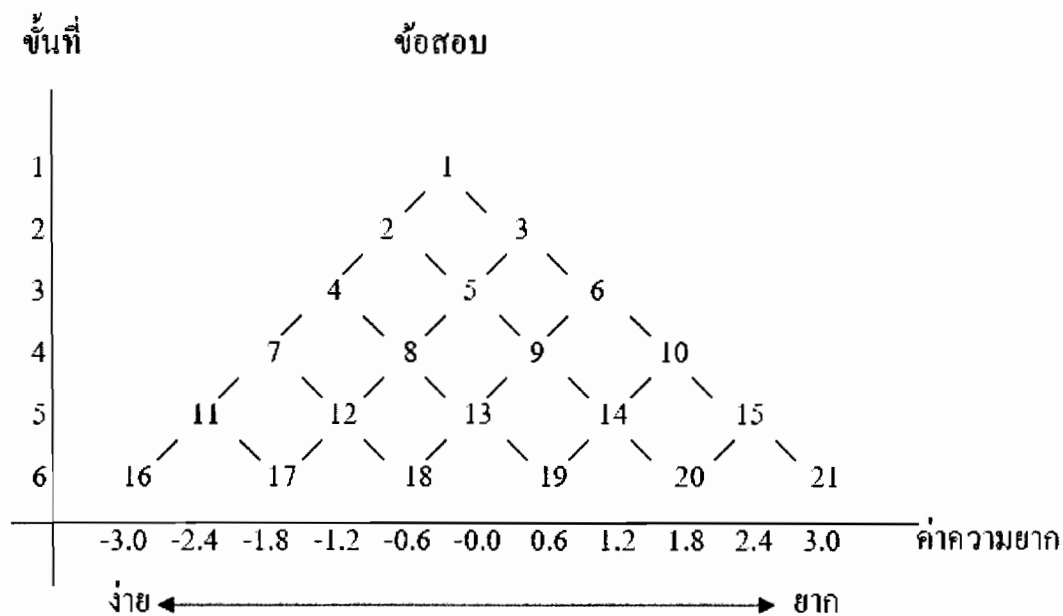
2. การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบหลายขั้นตอน

การทดสอบหลายขั้นตอนเป็นการทดสอบที่มีโครงสร้างของแบบทดสอบเป็นรูปพีระมิด โดยใช้ค่าความยากง่ายของข้อสอบเป็นเกณฑ์ในการจัดเรียงข้อสอบที่มีความยากง่ายปานกลาง จะอยู่จุดยอดของพีระมิด การดำเนินการสอบจะเริ่มด้วยให้ผู้สอบตอบข้อสอบข้อแรกที่มีความยากง่ายปานกลาง ถ้าผู้สอบตอบคำถามถูก ข้อสอบข้อต่อไปจะยากขึ้น และถ้าผู้สอบตอบผิด ข้อสอบข้อต่อไปก็จะง่ายลง การทดสอบจะดำเนินการเช่นนี้ไปเรื่อย ๆ จนกระทั่งถึงเกณฑ์ยุติการสอบตามที่กำหนดไว้ การสอบจึงสิ้นสุดลง การทดสอบหลายขั้นตอนสามารถแยกเป็นประเภทย่อย ๆ ได้ดังนี้

2.1 แบบแยกคงที่

2.1.1 รูปพีระมิดขนาดชั้นคงที่

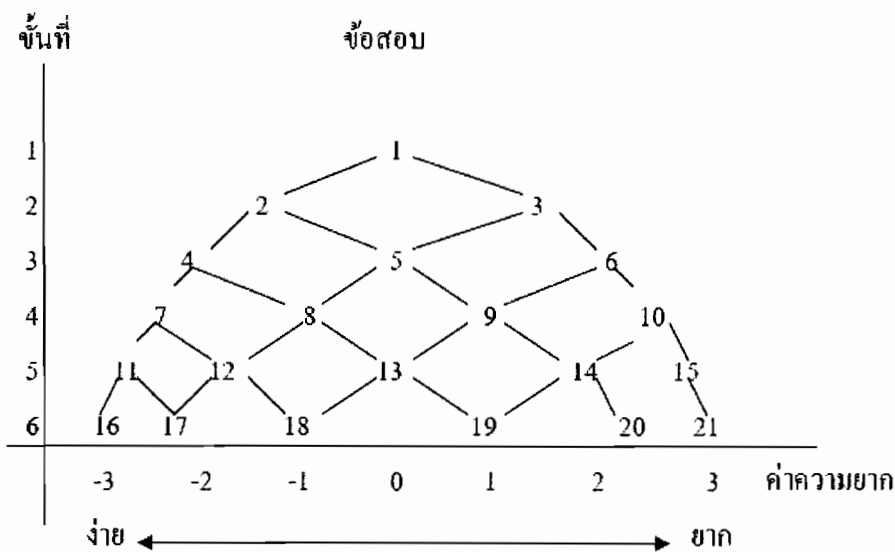
เป็นการจัดชุดแบบทดสอบเป็นรูปสามเหลี่ยม โดยมีจำนวนข้อสอบในแต่ละชั้นเท่ากับลำดับที่ของชั้น ผู้สอบจะต้องตอบข้อสอบชั้นละ 1 ข้อ โดยเริ่มตอบข้อ 1 ก่อน ถ้าตอบถูกก็จะไปตอบข้อที่ 3 ซึ่งยากกว่าข้อ 1 แต่ถ้าตอบผิดก็จะไปตอบข้อ 2 ซึ่งง่ายกว่าข้อ 1 ดังภาพที่ 3



ภาพที่ 3 โครงสร้างของการทดสอบเฉพาะบุคคลรูปแบบพีระมิดขนาดชั้นคงที่

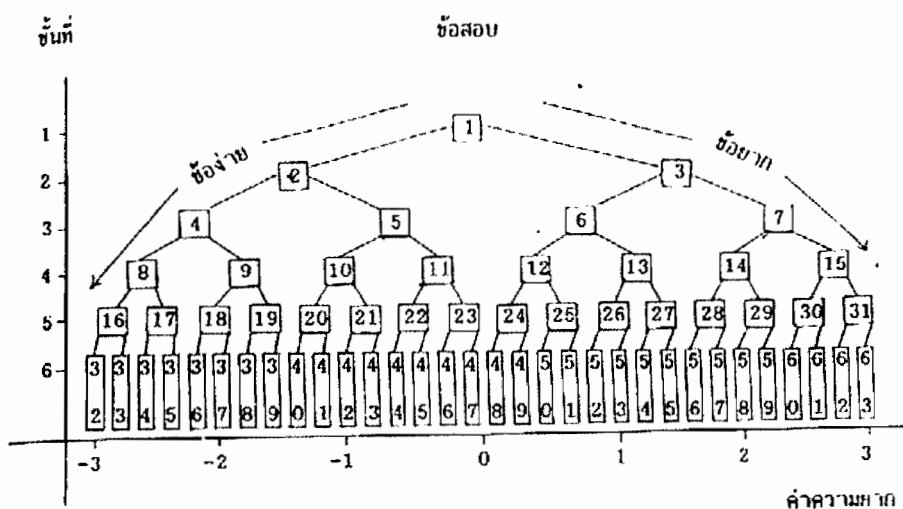
2.1.2 รูปพีระมิดขนาดชั้นแปรผัน

การจัดชุดแบบทดสอบลักษณะนี้ได้กำหนดช่วงห่างของค่าความยากง่ายของข้อสอบในชั้นเดียวกันไม่เท่ากัน โดยข้อสอบที่มีความยากง่ายปานกลางจะมีช่วงห่างของค่าความยากง่ายมากกว่าข้อสอบที่ยากหรือข้อสอบที่ง่าย ดังภาพที่ 4



ภาพที่ 4 โครงสร้างของแบบทดสอบรูปพีระมิดขนาดชั้นแปด

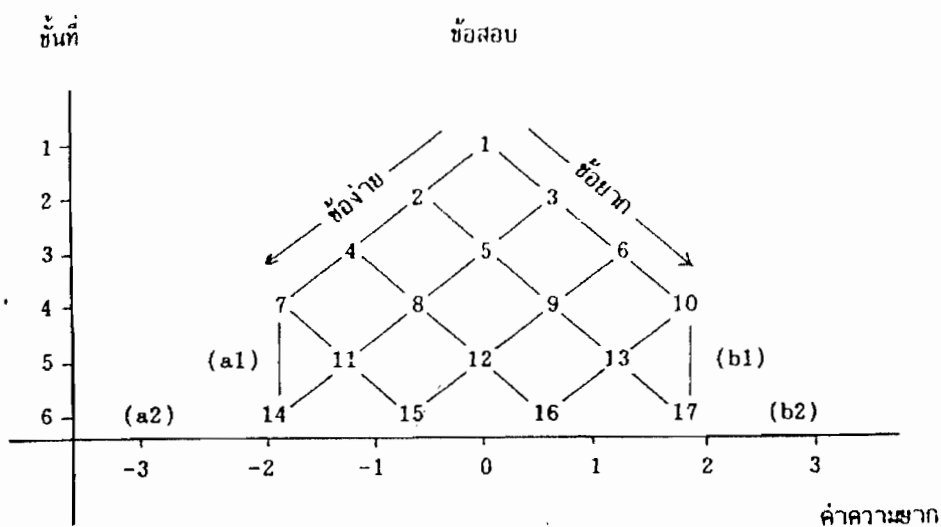
ต่อมาได้มีการพัฒนาโครงสร้างของแบบทดสอบรูปพีระมิดขนาดชั้นแปดใหม่ โดยมีการเพิ่มจำนวนข้อสอบในแต่ละชั้น ให้มีจำนวนเป็น 2 เท่าของจำนวนข้อสอบในชั้นที่ผ่านมา รูปแบบใหม่นี้มีชื่อว่า โรบินส์มอนโร (Robbins monro) ซึ่งมีโครงสร้างดังภาพที่ 5



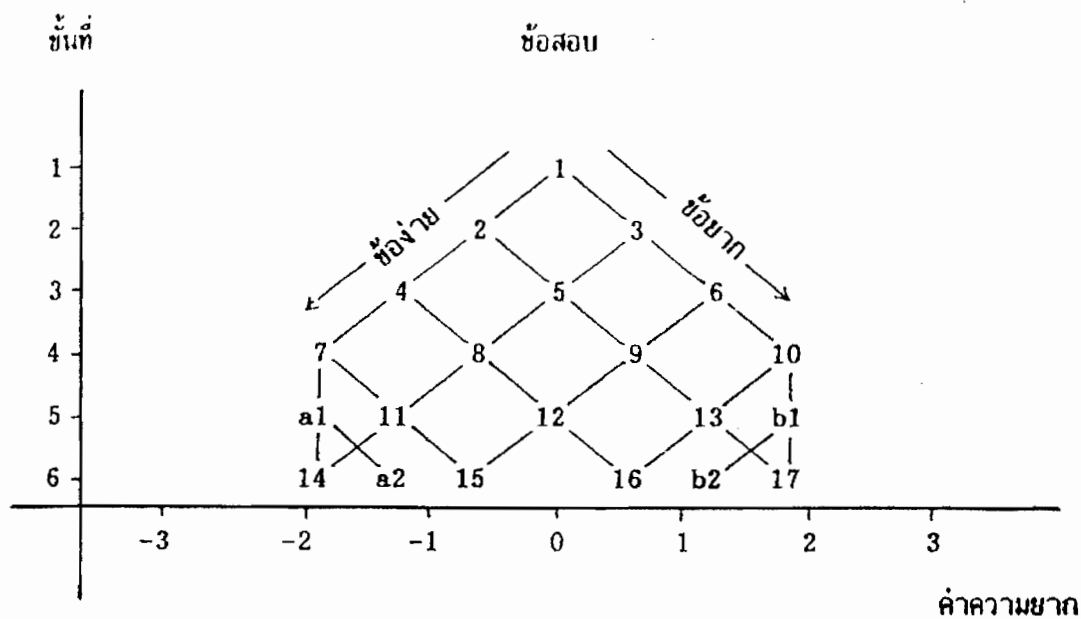
ภาพที่ 5 โครงสร้างของแบบทดสอบรูปพีระมิดแบบโรบินส์มอนโร

2.1.3 รูปพีระมิดแบบข้างตัด

เป็นรูปแบบที่มูสลีโอ (Mussio, 1973 cited in Weiss, 1974) ที่สร้างขึ้น เพื่อลดจำนวนข้อสอบจากโครงสร้างแบบทดสอบรูปพีระมิดแบบขนาดชั้นคงที่ ซึ่งจำแนกเป็น แบบสกัดการสะท้อนกลับ (Reflecting barrier) และรักษาการสะท้อนกลับ (Retaining barrier) ดังภาพที่ 6 และ 7 ตามลำดับ



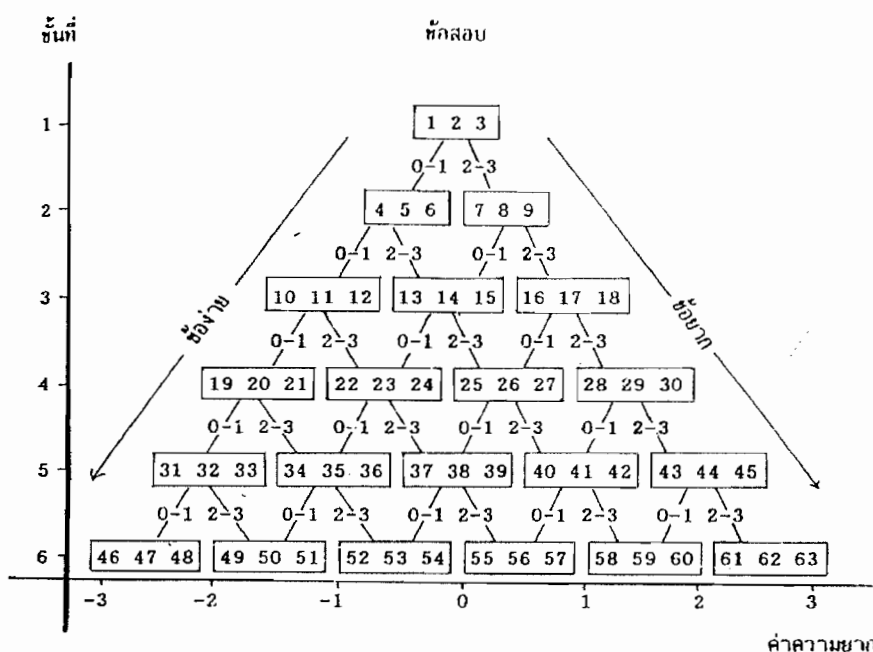
ภาพที่ 6 โครงสร้างของแบบทดสอบรูปพีระมิดแบบสกัดการสะท้อนกลับ



ภาพที่ 7 โครงสร้างของแบบทดสอบรูปพีระมิดแบบรักษาการสะท้อนกลับ

2.1.4 รูปพีระมิดแบบมีหลายข้อในแต่ละชั้น

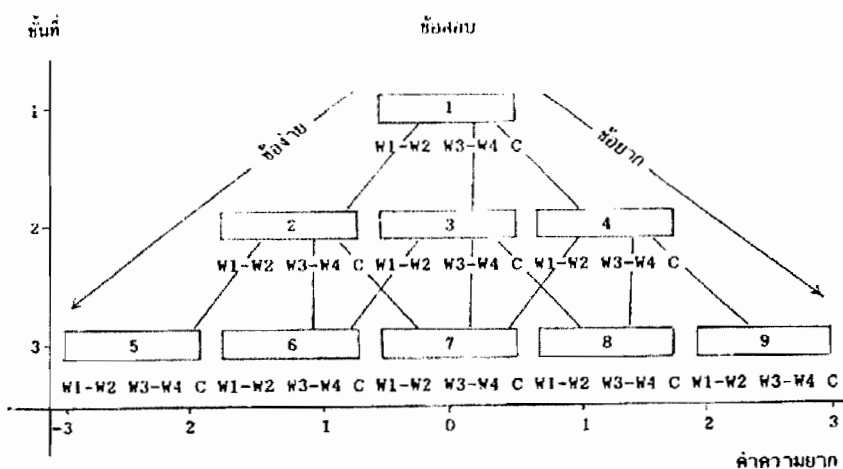
รูปแบบนี้กำหนดให้มีข้อสอบหลายข้อในแต่ละชั้น ทั้งนี้ เพื่อให้การแยกทางมีความถูกต้องมากยิ่งขึ้น ซึ่งมีโครงสร้างของการจัดข้อสอบดังภาพที่ 8



ภาพที่ 8 โครงสร้างของแบบทดสอบรูปพีระมิดแบบมีหลายข้อในแต่ละชั้น

2.1.5 รูปพีระมิดแบบให้น้ำหนักแก่ตัวเลือก

รูปแบบนี้คล้ายกับรูปแบบที่มีหลายข้อในแต่ละชั้นต่างกันตรงที่จะมีการกำหนดน้ำหนักให้กับตัวเลือกแต่ละข้อ ดังภาพที่ 9



ภาพที่ 9 โครงสร้างของแบบทดสอบรูปพีระมิดแบบให้น้ำหนักแก่ตัวเลือกของข้อสอบเพื่อแยกทาง

2.1.6 การทดสอบเฟลิกซ์เลเวล

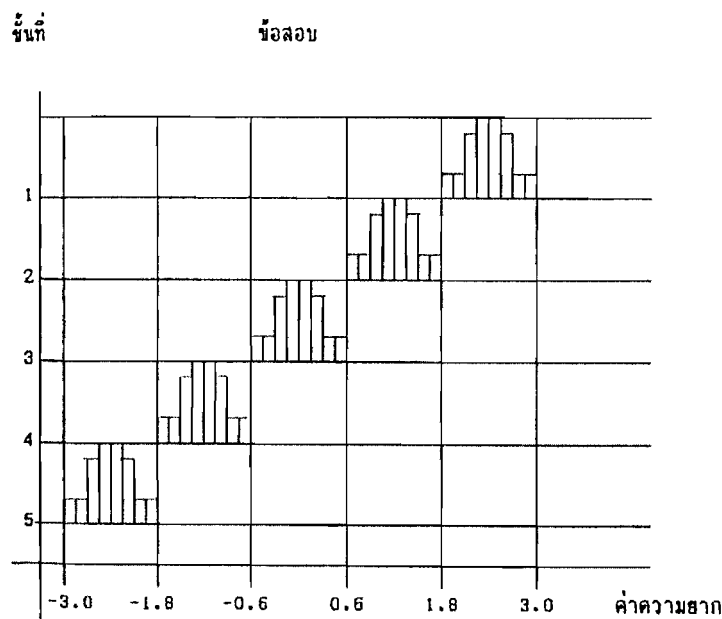
ลอร์ด (Lord, 1980, pp. 114-117) เป็นผู้นำเสนอการทดสอบลักษณะนี้ขึ้นมา การจัดเรียงข้อสอบจะเรียงจากง่ายไปหายาก ข้อที่มีความยากง่ายปานกลางจะถูกนำเสนอแก่ผู้สอบก่อน เช่น ถ้าข้อสอบมีทั้งหมด 75 ข้อ โครงสร้างของการจัดข้อสอบลักษณะนี้สามารถแสดงได้ดังตารางที่ 2

ตารางที่ 2 โครงสร้างของการจัดแบบทดสอบเฟลิกซ์เลเวล

ข้อ 0. (มีความยากอยู่กึ่งกลางระหว่างข้อสอบกลุ่มง่ายกับกลุ่มยาก)	
(ข้อสอบกลุ่มง่าย)	(ข้อสอบกลุ่มยาก)
(พิมพ์สีแดง)	(พิมพ์สีน้ำเงิน)
ข้อ 1. (ข้อที่ยากที่สุดในกลุ่มนี้)	ข้อ 1. (ข้อที่ง่ายที่สุดในกลุ่มนี้)
ข้อ 2.	ข้อ 2.
ข้อ 3.	ข้อ 3.
:	:
ข้อ 37. (ข้อที่ง่ายที่สุดในกลุ่มนี้)	ข้อ 37. (ข้อที่ยากที่สุดในกลุ่มนี้)

กระดาษคำตอบของการทดสอบลักษณะนี้จะทำให้ผู้สอบทราบได้ทันทีหลังจากการตอบแต่ละข้อว่า ถ้าตอบผิดจุดแดงจะขึ้นมา (ข้อสอบง่ายกว่าเดิม) ถ้าตอบถูกต้อง จุดน้ำเงินจะขึ้นมา (ข้อสอบยากกว่าเดิม) ซึ่งอาจใช้วิธีการเจาะหรือแกะกระดาษที่ทับตัวเลือกแต่ละข้อ ผู้สอบจะต้องตอบข้อสอบครั้งหนึ่งของข้อสอบทั้งหมดการทดสอบจึงจะยุติ วิธีการง่าย ๆ เพื่อให้ผู้สอบสามารถยุติการสอบได้ด้วยตนเองก็คือ จัดเรียงข้อสอบด้านยากใหม่ โดยกลับหัวจากง่ายลงไปหายากที่สุด เป็นจากยากที่สุดลงไปหาง่าย (เหมือนการจัดข้อสอบด้านง่าย) การตอบก็จะมีลักษณะเหมือนเดิม คือตอบข้อสอบที่ยากง่ายปานกลางก่อน ถ้าตอบผิดก็จะไปตอบข้อที่ง่ายกว่าเดิม ถ้าตอบถูกก็จะไปตอบข้อที่ยากกว่าเดิม ตอบไปจนกระทั่งพบว่าในแต่ละแถวของข้อสอบ (มี 2 ข้อ) ถูกตอบครบทุกแถว แถวละ 1 ข้อ จึงยุติการสอบ

ต่อมาได้มีการปรับปรุงโครงสร้างของการทดสอบลักษณะนี้ โดยแบ่งความยากง่ายของข้อสอบออกเป็นช่วง ๆ ในแต่ละช่วงความยากง่ายจะมีข้อสอบหลายข้อ การจัดแบบทดสอบลักษณะนี้เรียกว่า แบบทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบแบบแบ่งชั้น (Stradaptive test) ซึ่งมีโครงสร้างของการจัดข้อสอบดังภาพที่ 10



ภาพที่ 10 โครงสร้างของแบบทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบแบบแบ่งชั้น

2.2 การทดสอบแบบหลายชั้นตอนรูปแบบแยกทางแปรผัน

การทดสอบแบบแยกทางแปรผันนี้ ต่างจากการทดสอบแบบแยกทางคงที่

ตรงที่ไม่สามารถกำหนดเส้นทางของการตอบข้อสอบได้แน่นอนตายตัว ข้อสอบที่จะตอบข้อต่อไปนั้นขึ้นอยู่กับผลการตอบข้อสอบข้อที่ผ่านมา ซึ่งต้องมีการประมาณค่าความสามารถของผู้สอบทุกครั้ง หลังจากตอบข้อสอบแต่ละข้อ จึงเป็นเรื่องที่ยุ่ยากที่จะจัดแบบทดสอบเป็นแบบเขียนตอบ (Paper and pencil test) ดังนั้น จึงมีการนำไมโครคอมพิวเตอร์เข้ามาใช้เพื่อความสะดวกในการดำเนินการสอบ และบันทึกผลการสอบ การทดสอบจะเริ่มต้น โดยให้ผู้สอบนั่งที่หน้าจอคอมพิวเตอร์ให้ข้อมูลของตนเองตามที่ผู้ดำเนินการสอบต้องการ เช่น เลขที่ ชื่อ ชั้นเรียน ห้องเรียน เป็นต้น มีการแนะนำวิธีการตอบข้อสอบหลังจากที่ผู้สอบเข้าใจวิธีการตอบดีแล้ว ข้อสอบข้อแรกที่จะถูกนำเสนอแก่ผู้สอบ ซึ่งตามปกติจะใช้ข้อสอบที่มีความยากง่ายปานกลางหรือใช้ข้อสอบที่ยากง่ายเหมาะกับผู้สอบ หากทราบสภาพภูมิหลังของผู้สอบมาก่อน ข้อสอบถูกนำเสนอที่หน้าจอคอมพิวเตอร์ครั้งละ 1 ข้อ หลังจากการตอบข้อสอบแต่ละข้อผ่านไปข้อสอบข้อใหม่ก็จะปรากฏขึ้นมาที่หน้าจอตามเกณฑ์การคัดเลือกข้อสอบที่กำหนดไว้ การทดสอบจะดำเนินไปเรื่อย ๆ จนกระทั่งถึงเกณฑ์ยุติการสอบ การทดสอบก็จะสิ้นสุดลง

ประโยชน์ในการสร้างแบบทดสอบปรับเหมาะ

ในการทดสอบตามแบบประเพณีนิยม (Conventional testing) นั้น แบบทดสอบจะประกอบด้วยข้อทดสอบจำนวนมาก เพื่อให้การทดสอบมีความแม่นยำและความเชื่อมั่นสูง ข้อทดสอบเหล่านี้จะถูกเรียงลำดับไว้ก่อนแล้ว ตามเนื้อหาที่ต้องการทดสอบ และมักจะประกอบด้วยข้อทดสอบที่มีระดับความยากในระดับปานกลาง เพื่อให้เหมาะสมกับระดับความสามารถทั่วไปของผู้สอบ ซึ่งการกระทำดังกล่าวนี้ทำให้เกิดความคลาดเคลื่อนในการวัดสำหรับผู้สอบที่มีความรู้ในระดับเก่งหรืออ่อน เพราะข้อทดสอบที่ง่ายเกินไปจะไม่ท้าทายความสามารถในการสอบของผู้สอบที่เก่ง ทำให้ผู้สอบเกิดความเบื่อหน่ายในการตอบอย่างไม่เต็มความสามารถ ส่วนข้อทดสอบที่ยากเกินไปสำหรับผู้สอบที่อ่อนก็จะทำให้ผู้สอบรู้สึกท้อถอยและใช้วิธีเดาสิ่งต่าง ๆ เหล่านี้ ทำให้ความถูกต้องและแม่นยำในการทดสอบลดลงและทำให้การประมาณค่าความสามารถของผู้สอบคลาดเคลื่อน ดังนั้น การทดสอบแบบประเพณีนิยมจึงใช้ได้ผลดีเฉพาะกรณีที่ผู้สอบมีความสามารถในระดับปานกลางเท่านั้น และความคลาดเคลื่อนจะมีมากขึ้นเมื่อความสามารถของผู้สอบเบี่ยงเบนจากความสามารถปานกลางมากขึ้น (สุพัฒน์ สุขมลสันต์, 2538, หน้า 197)

การให้คะแนนข้อสอบเลือกตอบแบบถูกผิด (Scoring method of multiple true false)

หลักทั่วไปของข้อสอบเลือกตอบแบบถูกผิด

ข้อสอบเลือกตอบแบบถูกผิด (Multiple true false item) ได้พัฒนาขึ้นและใช้กันมานานแล้ว ตั้งแต่ก่อนปี ค.ศ. 1939 เพราะในบทความของ Cronbach ในปี ค.ศ. 1939 ได้กล่าวถึงข้อสอบชนิดนี้ว่ามีการใช้กันมาแล้ว แต่เรียกชื่อต่าง ๆ กัน เช่น Plural multiple-answer หรือ Plural choice (Cronbach, 1939, p. 628 อ้างถึงใน ยศสินี วิพชานานนท์, 2538, หน้า 25) อย่างไรก็ตาม แบบทดสอบเลือกตอบแบบถูกผิดนี้ ไม่ได้ได้รับความสนใจมากนักในวงการวัดผลการศึกษา แต่กลับมีการนำไปใช้กันอย่างแพร่หลายในการทดสอบในวงการแพทยศึกษาในชื่อที่เรียกว่า Indeterminate type และเรียกชื่อว่า ข้อสอบแบบ Multi-facet ในการสอบระดับประโยคมัธยม (Wood, 1977, p. 23 อ้างถึงใน พิชรี มีวรรณ, 2540, หน้า 4)

ในการสร้างข้อสอบเลือกตอบแบบถูกผิด (จินดา โตอนันต์, 2526, หน้า 14-15)

ใช้แนวความคิดที่พัฒนามาจากหลักการสร้างข้อสอบเลือกตอบ และข้อสอบถูก-ผิด โดยมีส่วนตอนนำของแต่ละข้อเป็นคำถาม (Stem) และมีตัวเลือก (Alternative or choice) หลายตัวเลือก ซึ่งแต่ละตัวเลือกจะต้องเป็นข้อความหรือประโยคที่สามารถตอบได้ว่าถูกหรือผิด แต่มิได้กำหนดว่าต้องมีตัวเลือกถูกหรือผิดในข้อนั้นจำนวนเท่าใด ทั้งนี้ ขึ้นอยู่กับความเหมาะสมของเนื้อหา

ในข้อคำถามหนึ่ง ๆ อาจประกอบด้วยตัวเลือกที่เป็นตัวถูกทั้งหมด ผิดทั้งหมด ถูกและผิดเท่า ๆ กัน หรือถูกผิด ไม่เท่ากันก็ได้ ซึ่งหลักการเขียนตัวเลือกก็ใช้หลักการเดียวกับข้อสอบเลือกตอบนั่นเอง

ข้อสอบเลือกตอบแบบถูกผิด (MTF) เป็นข้อสอบปรนัยชนิดหนึ่งที่กำลังพัฒนา เพื่อให้ผลการตรวจสอบทางการศึกษาแม่นยำยิ่งขึ้น และเพื่อแก้ไขจุดบกพร่องในข้อสอบเลือกตอบที่ว่า การตอบคำถามในข้อสอบนั้น ผู้สอบได้อ่านทุกส่วนของแต่ละข้อคำถาม โดยละเอียดหรือไม่ หรืออ่านเฉพาะตัวคำถามและตัวเลือก จนกระทั่งถึงคำตอบที่พอใจเท่านั้น ในข้อนี้ ถ้าผู้สอบไม่ต้องอ่านคำถามและตัวเลือกในแต่ละข้อกระทั่งจบ ก็สามารถตอบคำถามได้ถูกต้อง ก็ย่อมแสดงว่า ข้อสอบเลือกตอบที่สร้างขึ้นนั้น เป็นข้อสอบที่มีประสิทธิภาพ เพราะตัวลวงต่าง ๆ ไม่สามารถนำมาใช้ประโยชน์ได้ และอีกประเด็นหนึ่ง ในกรณีที่เนื้อหาวิชามีความสำคัญมาก การใช้ข้อสอบเลือกตอบจะสามารถถามได้ครอบคลุมเนื้อหาวิชานั้นหรือไม่ เพราะในแต่ละคำถามของข้อสอบเลือกตอบนั้น มีคำตอบถูกให้เลือกตอบเพียงตัวเลือกเดียว (กมลทิพย์ อธิการยานันท์, 2530, หน้า 13) จากปัญหานี้ กรอนดึนส์ (Gronlund, 1982, p. 44) ได้สนับสนุนว่า ในการกล่าวเนื้อหาวิชาบางอย่าง บางครั้งก็ไม่เหมาะสมที่จะนำมาสร้างเป็นแบบทดสอบเลือกตอบที่มีคำตอบถูกตัวเดียว ควรจะสร้างให้มีจำนวนข้อความ (ตัวเลือก) ที่ถูกหรือผิดได้มากกว่า 1 ตามความเหมาะสมของเนื้อหาวิชาได้กว้างขวางขึ้น ซึ่งจากกรณีทั้งสองที่ยกมานี้ จะเห็นว่า ถ้าผู้สอบทราบว่าตัวเลือกมีเพียงตัวเลือกเดียว ผู้สอบก็มีหน้าที่อย่างเดียว คือ หาตัวเลือกที่ถูกเท่านั้น ซึ่งการทำเช่นนั้นไม่ตรงกับจุดประสงค์ของผู้ออกข้อสอบ เพราะถ้าเป็นเช่นนั้นจริง ตัวลวงย่อมไม่มีคุณค่าการเขียนข้อสอบเลือกตอบแบบถูกผิด สามารถช่วยแก้ปัญหาในเรื่องนี้ได้ เพราะในการสร้างข้อสอบเลือกตอบแบบถูกผิด ผู้สอบต้องอ่านคำถาม และตัวเลือกทุกตัวเพื่อพิจารณาว่าตัวเลือกใดถูกหรือผิดบ้าง นอกจากนี้ ข้อสอบเลือกตอบแบบถูกผิดยังสามารถออกได้ครอบคลุมเนื้อหาได้มากกว่าข้อสอบเลือกตอบธรรมดา เพราะไม่ได้กำหนดแน่นอนว่าในแต่ละข้อคำถามจะต้องมีจำนวนตัวเลือกถูกหรือผิดเท่าใด ซึ่งเป็นการเปิดกว้างในการเขียนรูปแบบคำถามและการจัดตัวเลือกให้เหมาะสม นอกจากนี้ อีเบล (Ebel, 1972, pp. 37-44) ได้กล่าวว่า ข้อสอบเลือกตอบแบบถูกผิดนี้ คำถามไม่จำเป็นต้องเป็นการถามความคิดรวบยอด ค่านิยม ความสัมพันธ์หรือขบวนการ แต่อาจจะเป็น “ข้อความใดถูกหรือข้อความใดผิด” ซึ่งข้อความที่กำหนดให้อาจเป็นรายละเอียดของเรื่องใดเรื่องหนึ่ง หรือเหตุการณ์ที่สำคัญทางประวัติศาสตร์

หลักการสำหรับข้อสอบเลือกตอบแบบถูกผิด

ฮาလာไดนา (Haladyna, 1994, p. 83 อ้างถึงใน พัทรี มีวรรณ, 2540, หน้า 20) กล่าวถึงหลักการสำหรับข้อสอบเลือกตอบแบบถูกผิด ดังนี้

1. จำนวนตัวเลือกในแต่ละคำถามของข้อสอบในแต่ละฉบับไม่เท่ากัน

2. สามารถนำหลักการของข้อสอบเลือกตอบที่ถูกเพียงตัวเดียว (Multiple-Choice: MC) ข้อสอบเลือกตอบแบบตัวเลือกซ้อน (Complex Multiple-Choice: CMC) และข้อสอบแบบถูกผิด (True-False: TF) มาพัฒนาเป็นข้อสอบแบบนี้ได้เป็นอย่างดี

3. จำนวนข้อสอบในชุดข้อสอบขึ้นอยู่กับ ความเหมาะสมของหน้ากระดาษ อาจมีประมาณ 30-35 ข้อ

4. ไม่จำกัดเกี่ยวกับจำนวนข้อถูก และผิดในแต่ละข้อคำถาม

ชวาล แพร์ตกุล (ม.ป.ป., หน้า 136-138) กล่าวถึงหลักในการสร้างข้อสอบเลือกตอบแบบถูกผิด ดังนี้

1. ข้อสอบชนิดนี้ เหมาะที่จะใช้กับเรื่องราวที่มีคำตอบถูกต้องอยู่หลายประการ โดยแต่ละประการเหล่านั้นต่างก็เป็นความจริงหรือมีความสำคัญทัดเทียมกัน และเราก็ต้องการวัดดูว่า เด็กรู้คำตอบเหล่านั้นครบถ้วนหรือไม่

2. ตัวเลือกที่ใช้กับข้อสอบชนิดนี้ จะต้องมีลักษณะเป็นประเภทถูกจริงหรือผิดจริง

3. ตัวเลือกชนิดนี้จะใช้ตัวเลือกปลายเปิดหรือปลายปิดไม่ได้ เพราะทุกข้อมีทั้งตัวถูก และตัวผิดปนกันอยู่ในนั้นหลายตัว

4. ตัวเลือกชนิดนี้เขียนได้ง่ายมาก เพราะตัวเลือกจะเป็นข้อความธรรมดา ๆ แบบประโยคบอกเล่า ที่พูดถึงเนื้อหาของวิชาเหล่านั้น สิ่งที่เราควรระวังเป็นพิเศษคือ อย่าลอกถ้อยคำ จากตำรามาตั้งเป็นตัวเลือกหรือให้เป็นตัวเลือกกันตรง ๆ ตามนั้น กับอย่าเขียนคำถาม ให้ตีความหมายได้เป็นสองแง่หรือกำวม

5. ข้อสอบเลือกตอบแบบถูกผิด สามารถใช้ได้กับทุกวิชา

6. ข้อสอบชนิดนี้เหมาะที่จะใช้กับการทดสอบย่อยในระหว่างการเรียนมากกว่าที่จะใช้เป็นข้อสอบรวบยอดปลายเทอมหรือปลายปีที่เป็นพิธีการ

7. ข้อสอบที่ใช้คำถามชนิดนี้ จะต้องใช้เวลาทำแก้เด็กมากกว่าข้อสอบเลือกตอบ สัก 3-4 เท่าตัวจึงจะพอ เพราะเขาจะต้องคิดตอบทุกตัวเลือก

กระบวนการปฏิบัติในการใช้ข้อสอบเลือกตอบแบบถูกผิด

ฮาร์ดเดน (Harden, 1979 อ้างถึงใน วัชรารักษ์ จิตรมาศ, 2538, หน้า 13) ยังได้เสนอแนะ กระบวนการปฏิบัติ ในการใช้ข้อสอบเลือกตอบแบบถูกผิด ดังนี้

1. ประชุมคณะกรรมการจัดทำข้อสอบก่อนการเขียนข้อสอบเพื่อวัตถุประสงค์ดังนี้

1.1 เลือกชนิดของข้อสอบว่าจะให้มีจำนวนตัวเลือกเท่าใด จำนวนข้อเท่าใด

และเวลาที่เหมาะสมในการทำข้อคำถามทั้งหมด โดยเสนอแนะว่าข้อสอบ 60 ข้อที่มีจำนวนตัวเลือก

5 ตัวเลือกควรใช้เวลาประมาณ 60 ถึง 90 นาที ทั้งนี้ ขึ้นกับค่าความยาก ความยาวของปัญหา หรือคำถาม

- 1.2 วางแผน วิเคราะห์เนื้อหา กำหนดระดับความยากที่เหมาะสม
- 1.3 เลือกคำถาม อาจมาจากธนาคารข้อสอบ หรือร่วมกันเขียน
- 1.4 วางแผนเวลาในการทำงานโดยมีการเขียนคำถามไว้ก่อน แล้วหาคำตอบที่ถูกมาให้เพียงพอดตามที่ต้องการ และเฉลยด้วย
2. เมื่อได้ข้อสอบมาแล้วควรมีการวิเคราะห์ หรือพิจารณาร่วมกันอีกครั้งหนึ่งเพื่อ
 - 2.1 ข้อสอบที่เขียนขึ้น วัดในสิ่งที่ต้องการวัดหรือไม่ วัดตามจุดประสงค์หรือไม่
 - 2.2 ข้อสอบวัดในส่วนที่สำคัญของเนื้อหาหรือไม่
 - 2.3 ข้อความที่เป็นตัวเลือกที่ถูกหรือผิดมีค่าความยากใกล้เคียงกันหรือไม่
3. คำชี้แจงวิธีการตอบ การตรวจให้คะแนน ควรจัดทำไว้ให้ชัดเจน พร้อมสำหรับผู้ดำเนินการสอบ
 4. มีการตรวจทานการสะกด การันต์อย่างละเอียดถี่ถ้วนก่อนจัดทำไปทดสอบ
 5. นำผลการสอบมาวิเคราะห์อีกครั้งหนึ่ง เพื่อเป็นการปรับปรุงข้อสอบ และความรู้ความสามารถของนักเรียน

การให้คะแนน (Scoring method)

คณิต ไช่มุกต์ (2533, หน้า 1) กล่าวว่า การสอบทางการศึกษาส่วนใหญ่เป็นการตรวจสอบความสามารถของผู้เรียนอันเป็นผลจากการเรียนรู้ที่เกิดขึ้น เพื่ออธิบายความสามารถในปัจจุบัน หรือทำนายความสามารถ ความพร้อมของการเรียนรู้ในอนาคต ความสามารถของบุคคล เป็นคุณลักษณะภายในที่ต้องอาศัยวิธีการวัดทางอ้อม โดยการตอบสนองต่อสิ่งเร้าที่จัดขึ้น แล้วนำผลการตอบสนองที่ได้ ทำการสรุปอ้างอิงไปยังคุณลักษณะภายในสิ่งที่มีงวัด

ลอร์ด และ โนวิก (Lord & Novick, 1968) กล่าวว่า คะแนนที่วัดได้ทางการศึกษา เป็นปริมาณที่รวมเอาความคลาดเคลื่อนที่เกิดจากแหล่งต่าง ๆ ทั้งข้อสอบที่ใช้ ลักษณะของผู้สอบ สถานการณ์การสอบ เข้าไว้ด้วยทุกครั้งทีวัด คะแนนที่ได้จึงไม่ใช่ค่าที่แท้จริงของสิ่งที่มีงวัด

เมสสิก (Messick, n.d. cited in Linn, 1981) กล่าวว่า คะแนนจึงเป็นแหล่งสำคัญของการอ้างอิงในการวัดทางการศึกษา เพราะคะแนนเป็นผลการตอบข้อสอบซึ่งไม่เพียงแต่เป็นฟังก์ชันที่เกี่ยวกับข้อสอบ งานที่กำหนดหรือเงื่อนไขของสิ่งเร้าต่าง ๆ เท่านั้น แต่ยังเกี่ยวกับวิธีการตอบของผู้สอบและเกี่ยวกับบริบทของการวัดด้วย การอ้างอิงถึงความเที่ยงตรงและความเชื่อมั่นของการวัดใด ๆ จึงต้องอ้างอิงจากคะแนนนี้

ปัจจุบันได้มีนักทฤษฎีวัดผลอ้างอิงถึงความสัมพันธ์ระหว่างความสามารถที่แท้จริงกับ ลักษณะการตอบสนองข้อสอบ เพื่อทำนายลักษณะที่ซ่อนอยู่ในตัว เรียกว่า ทฤษฎีตอบสนองข้อสอบ (Item Response Theory: IRT) เป็นทฤษฎีที่มีพื้นฐานความเชื่อว่าพฤติกรรมกรตอบสนองต่อข้อสอบ ของผู้สอบ เป็นสิ่งสังเกตได้โดยตรงว่าถูกหรือผิด จะถูกกำหนด โดยคุณลักษณะภายใน (Trait) หรือ ความสามารถ (Ability) ที่อยู่ในตัวบุคคลซึ่ง เป็นสิ่งที่ไม่สามารถสังเกตได้โดยตรง สามารถอธิบาย ความสัมพันธ์ของคุณลักษณะภายในหรือความสามารถที่มีอยู่ในตัวบุคคลกับพฤติกรรมกรตอบ ข้อสอบ โดยใช้โค้งลักษณะข้อสอบ (Item Characteristic Curve: ICC) ซึ่งกำหนดลักษณะของข้อสอบ ด้วยพารามิเตอร์ความยาก (b) ค่าอำนาจจำแนก (a) และ โอกาสการเดา (c) มีลักษณะเป็นฟังก์ชัน ทางคณิตศาสตร์ เรียกว่า ฟังก์ชัน โลจิสติก (Logistic function) หรือใกล้เคียงกับฟังก์ชันปกติสะสม (Normal ogive function) (ศิริชัย กาญจนวาสี, 2545, หน้า 45-46)

การนำทฤษฎีตอบสนองข้อสอบ (Item Response Theory: IRT) ไปใช้ในการแก้ปัญหา การวัดที่สำคัญ ๆ โดยเฉพาะการแก้ปัญหาที่ไม่สามารถแก้ไขด้วยทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory: CTT) เนื่องจากทฤษฎีการตอบสนองข้อสอบสามารถประมาณค่า ความสามารถหรือคุณลักษณะของผู้ตอบเป็นอิสระจากชุดของข้อคำถาม ไม่ว่าจะเลือก กลุ่มตัวอย่างย่อยกลุ่มใดมาเพื่อใช้ในการประมาณค่าพารามิเตอร์ ข้อความค่าพารามิเตอร์นั้นจะคงที่ คุณสมบัตินี้เป็นประโยชน์อย่างมากสำหรับการเลือกโมเดลเพื่อนำไปใช้ในการสร้าง และพัฒนาข้อสอบในปัจจุบัน นอกจากการเลือกโมเดลที่เหมาะสมแล้ว วิธีการประมาณค่าพารามิเตอร์ ของข้อคำถามหรือข้อความ และคุณลักษณะภายในตัวบุคคลที่ต้องการวัด เป็นอีกกระบวนการหนึ่ง ที่ต้องเลือกให้เหมาะสมกับสภาพการวัด

ไรท์ และมาสเตอร์ (Wright & Masters, 1982) ได้พัฒนาและขยายแนวคิดการประมาณค่า พารามิเตอร์ที่ได้จากการตรวจให้คะแนนแบบสองค่า (Dichotomous) มาเป็นการประมาณค่า พารามิเตอร์ที่ได้จากการตรวจให้คะแนน คะแนนแบบหลายค่า หรือการตรวจให้คะแนนมากกว่า 2 ค่า (Polytomous item response theory) รวมถึงการตรวจให้คะแนนความรู้บางส่วน (Partial Credit Model: PCM) ที่พัฒนาโดยมาสเตอร์ (Masters, 1982) ในการวิเคราะห์ข้อสอบหรือข้อคำถาม ที่มีกระบวนการตอบหลายลำดับขั้น (Ordered response categories) ซึ่งจำเป็นต้องมีการตรวจ ให้คะแนนการตอบถูกต้องหรือตอบถูกบางส่วนในแต่ละลำดับขั้นของกระบวนการตอบ (Ordered performance levels)

Partial Credit Model = PCM พัฒนาโดยมาสเตอร์ให้สามารถใช้กับแบบสอบที่มี หลายตัวเลือก และการให้คะแนนตัวเลือกแต่ละตัวแตกต่างกัน โมเดลนี้ได้รับการพัฒนาให้ดีขึ้นโดย ไรท์ และมาสเตอร์ (Wright & Masters, 1984) นักวัดผลการศึกษาทั้งสองยังได้ประมวลโมเดล

โลจิสติกหนึ่งพารามิเตอร์ หรือ Rasch model แบบต่าง ๆ ที่มีการพัฒนาขึ้นรวมเรียกว่า โมเดลการวัดขั้นพื้นฐาน (Fundamental measurement model) ซึ่งมีคุณสมบัติสำคัญ คือ การแยกพารามิเตอร์ผู้สอบและพารามิเตอร์ข้อคำถาม และมีการใช้สถิติที่พอเพียงในการประมาณค่าพารามิเตอร์ของโมเดล การพัฒนาล่าสุด คือ Muraki (1993) ได้พัฒนาโมเดลรวมค่าพารามิเตอร์ความชัน (Slope) ของโค้งลักษณะข้อคำถามแต่ละข้อ ไว้ในโมเดล เรียกว่า Generalized Partial Credit Model: GPCM (นงลักษณ์ วิรัชชัย, 2545, หน้า 16)

วิธีการตรวจให้คะแนนแบบ 0-1 มีข้อบกพร่องหลายประการ (อนันต์ ศรีโสภากา, 2532, หน้า 11) เช่น เป็นการช่วยหรือส่งเสริมให้นักเรียนที่ไม่มีความรู้ความสามารถเดามากขึ้น แม้ว่าจะมีสูตรแก้การเดาก็มิได้ทำให้คะแนนหลังจากแก้การเดาแล้ว แทนความรู้ความสามารถที่แท้จริงของนักเรียนได้ การที่นักเรียนทำถูกได้ 1 คะแนน และทำผิดหรือไม่ทำเลยได้ 0 คะแนน ความจริงนักเรียนบางคนอาจจะมีความรู้ความสามารถไม่สมบูรณ์แต่ก็ไม่ถึงกับไม่รู้เลย คือ มีความรู้บางส่วน วิธีการแบบนี้ไม่สามารถวัดความรู้ความสามารถบางส่วนเหล่านั้นออกมาได้ อนึ่ง ข้อสอบแต่ละข้อระดับความยากไม่เท่ากัน คือ บางข้อยากมาก บางข้อยากน้อย แต่ก็ให้น้ำหนักของคะแนนเท่ากันหมดทุกข้อ คือ นักเรียนตอบถูก ไม่ว่าข้อนั้นจะยากหรือง่ายก็ได้ 1 คะแนนเท่ากัน คะแนนรวมที่ได้จึงไม่สอดคล้องกับความรู้ความสามารถที่แท้จริงของนักเรียน จากข้อบกพร่องของวิธีการตรวจให้คะแนนแบบ 0-1 ดังกล่าว จึงได้มีนักวัดผลการศึกษาหลายท่านพยายามคิดหาวิธีเพื่อพัฒนาวิธีการตรวจให้คะแนนข้อสอบเลือกตอบให้มีประสิทธิภาพยิ่งขึ้น

การให้คะแนนความรู้บางส่วนของผู้ตอบแสดงถึงตำแหน่งความรู้จริงของผู้ตอบซึ่งมีส่วนช่วยลดความคลาดเคลื่อนจากแหล่งการเดาสุ่มของผู้ตอบทำให้คะแนนมีความเที่ยงตรงมากขึ้น ขณะเดียวกันก็ส่งผลต่อความเชื่อมั่นของกระบวนการวัด (Frany, 1980) คะแนนความรู้บางส่วนแบ่งเป็นประเภทใหญ่ ดังนี้ (Simon et al., 1997, pp. 65-88)

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

โดยใช้น้ำหนักที่ให้ผู้สอบประเมินตนเอง เป็นวิธีที่ดีที่สุดในการได้มาซึ่งสารสนเทศเกี่ยวกับความรู้ของผู้สอบในแต่ละข้อ ผู้สอบจะให้สารสนเทศอย่างสมบูรณ์เกี่ยวกับน้ำหนัก วิธีการให้คะแนนและการควบคุมระดับความมั่นใจ จะช่วยลดการเดาและความคลาดเคลื่อนในการวัด วิธีนี้ใช้รูปแบบข้อสอบเหมือนแบบทดสอบชนิดเลือกตอบทั่วไป เพียงแต่คำสั่ง วิธีการตอบ และการให้คะแนนแตกต่างจากแบบทดสอบชนิดเลือกตอบธรรมดา แต่ละวิธีการไม่ยุ่งยาก จำแนกได้ดังนี้

- 4.1 วิธีการแก้การเดา (Correction for guessing)
- 4.2 วิธีการให้คะแนนจากการตัดตัวลวง (Elimination scoring)
- 4.3 วิธีการของการทดสอบความน่าจะเป็น (Probability testing)
- 4.4 วิธีการแสดงความมั่นใจ (Confidence marking)
- 4.5 วิธีการจัดอันดับอย่างสมบูรณ์ (Complete ordering)
- 4.6 วิธีการจัดอันดับเป็นบางส่วน (Partial ordering)

การที่จะทำให้ทราบว่าคะแนนที่ได้จากการสอบเกิดจากความรู้ที่แท้จริง วิธีการตรวจให้คะแนนจึงมีบทบาทสำคัญ เพื่อให้คะแนนที่ได้จากแบบทดสอบน่าเชื่อถือมากกว่านี้ การตรวจให้คะแนนแบบสองค่า (Binary) เช่น การตรวจให้คะแนนแบบ 0 หรือ 1 ถูกหรือผิด เห็นด้วยหรือไม่เห็นด้วย เป็นวิธีการตรวจให้คะแนนกันอย่างแพร่หลายในปัจจุบัน เพราะสามารถวัดความรู้ความสามารถ และวัดผลสัมฤทธิ์ได้อย่างกว้างขวาง

เนื่องจากมีความเป็นปรนัยในการตรวจให้คะแนนใช้เวลาตรวจน้อย แต่ยังมีข้อจำกัดที่สำคัญ คือ เปิดโอกาสให้มีการตอบถูกโดยการเดา ปัญหาการเดาสุ่ม จึงเป็นแหล่งความคลาดเคลื่อนในการวัด นอกจากนี้ ผลการวัดที่ได้ไม่สามารถให้สารสนเทศที่เพียงพอที่จะจำแนกผู้ตอบในระดับต่าง ๆ ของความรู้ระหว่างผู้รู้จริง (Full knowledge) ผู้ที่มีความรู้บางส่วน (Partial knowledge) ผู้ที่ไม่รู้ (Absence of knowledge)

การนำวิธีการให้คะแนนไปใช้เป็นวิธีการแก้ปัญหาเพื่อลดโอกาสในการเดาและเพิ่มสารสนเทศเกี่ยวกับการตอบข้อสอบเลือกตอบให้มากที่สุด โดยใช้วิธีการให้คะแนนที่เป็นองค์ประกอบที่สำคัญของกระบวนการวัดสำหรับวิธีการให้คะแนนที่เหมาะสมจะทำให้คุณสมบัติของการวัดในประเด็นที่เกี่ยวข้องกับความเที่ยงตรงและความเชื่อมั่นสูงขึ้น การให้คะแนนความรู้บางส่วน (Partial credit) ของผู้ตอบแสดงถึงตำแหน่งจริงของความรู้ผู้ตอบ ซึ่งมีส่วนช่วยลดความคลาดเคลื่อนจากแหล่งการเดาของผู้ตอบ ทำให้คะแนนมีความเที่ยงตรงมากขึ้น ในขณะที่เดียวกันก็ส่งผลต่อความเชื่อมั่นของกระบวนการวัดด้วย (Frany, 1980)

จากการศึกษาทฤษฎีเกี่ยวกับ โมเดลของการให้คะแนนที่ผ่านมา เห็นว่า ทฤษฎีการตอบสนองข้อสอบที่มีการตรวจให้คะแนนมากกว่า 2 ค่า (Polytomous item response model) โดยเฉพาะการให้คะแนนความรู้บางส่วน (Partial Credit Model: PCM) ที่เป็นโมเดลที่แสดงถึงการให้คะแนนที่ใกล้เคียงกับความรู้ที่แท้จริงของผู้ตอบ สามารถอธิบายถึงความสัมพันธ์ระหว่างความสามารถของผู้สอบและโอกาสของการตอบได้ถูก ด้วยฟังก์ชันทางคณิตศาสตร์ที่นำเสนอในรูปแบบโค้งลักษณะข้อสอบ จากวิธีการประมาณค่าพารามิเตอร์ความสามารถของผู้ตอบ ตลอดจน

มีความเหมาะสมในการวัดความสามารถทางสติปัญญา เช่น วัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ ที่เชื่อว่า ผลการตอบสนองข้อสอบขึ้นอยู่กับลักษณะของข้อสอบและความสามารถของผู้สอบ

การประเมินผู้เรียนโดยใช้แบบทดสอบยังมีความจำเป็นและใช้กันอย่างแพร่หลาย ในปัจจุบัน ถึงแม้ว่านักการศึกษา ครูอาจารย์นักวิชาการ ตลอดจนนักวัดผล ได้มีความพยายาม ในการพัฒนารูปแบบและเครื่องมือการประเมินผู้เรียนให้มีความหลากหลายยิ่งขึ้น เพื่อให้สามารถ วัดคุณลักษณะของผู้เรียนได้เที่ยงตรง ครบคลุม มีความน่าเชื่อถือ และได้ข้อมูลที่มีความหมายต่อ ตัวผู้เรียน เครื่องมือวัดผลต่าง ๆ ที่ได้พัฒนาขึ้นนั้น ยังไม่มีเครื่องมือชนิดใดใช้แทนแบบทดสอบ ได้ทั้งหมด เพราะการวัดและการประเมินผลการเรียนรู้ที่เน้นผู้เรียนเป็นสำคัญต้องวัดและประเมิน ให้ครอบคลุมทุกด้าน ทั้งในส่วนของกระบวนการและผลงาน ทั้งด้านความรู้ ด้านความรู้สึก และทักษะการแสดงออกทุกด้าน และการประเมินตามสภาพจริง (ส.วาสนา ประवालพฤษย์, 2545)

วิธีการตรวจให้คะแนนข้อสอบเลือกตอบแบบถูกผิด

วิธีการให้คะแนนสำหรับข้อสอบนั้นมีอยู่หลายวิธี การจะนำวิธีใดไปใช้ก็ขึ้นอยู่กับ ผู้ตรวจให้คะแนนว่าต้องการใช้วิธีการให้คะแนนแบบใดที่ผู้ตรวจให้คะแนนคิดว่าเหมาะสม จากการศึกษาพบว่า มีวิธีการให้คะแนนหลายวิธี ดังนี้

1. วิธี 0-1 เป็นวิธีการให้คะแนนที่นิยมปฏิบัติกัน ข้อสอบเลือกตอบ ที่หมายถึงวิธีการ ให้นักเรียนเลือกตอบเฉพาะตัวเลือกที่ถูก จะได้คะแนน 0 คะแนน ต่อข้อสอบ 1 ข้อ ถ้าผู้ตอบ เลือกตัวเลือกที่ผิดหรือไม่เลือกเลย และให้ 1 คะแนน ถ้าผู้ตอบเลือกตัวเลือกที่ถูก เมื่อนำมาใช้ ในแบบทดสอบถูกผิดแบบหลายตัวเลือกจะหมายถึง การให้ผู้ตอบเลือกตอบตัวเลือกที่เห็นว่าถูก และตัวเลือกที่เห็นว่าผิด คะแนนของข้อสอบแต่ละข้อจะเป็น 1 คะแนน เมื่อผู้ตอบเลือกตัวเลือก ที่ตรงกับสภาพจริงของตัวเลือกนั้น ๆ ได้ทุกตัวเลือก และได้คะแนน 0 คะแนน เมื่อผู้ตอบ เลือกตัวเลือกที่ไม่ตรงกับสภาพจริงของตัวเลือกนั้น ๆ ได้ทุกตัวเลือก และได้คะแนน 0 คะแนน เมื่อผู้ตอบเลือกตัวเลือกที่ไม่ตรงกับสภาพจริงของตัวเลือกนั้น ๆ ตั้งแต่ 1 ตัวเลือก ขึ้นไป

2. วิธีการให้คะแนนแบบให้คะแนนทุกตัวเลือก (MTF) วิธีให้คะแนนวิธีนี้ อัลบานีส, โทมัส และดักลาส (Albanese, Thomas, & Douglas, 1979, pp. 948-950) ได้เสนอไว้ในการให้ผู้ตอบ ตอบแบบทดสอบเลือกตอบชนิดถูกผิดหลายตัวเลือก โดยในข้อสอบหนึ่งข้อจะมีตัวเลือกที่ถูกและผิด มากกว่าหนึ่งตัวเลือก แล้วให้ผู้ตอบพิจารณาตัวเลือกที่ผิดและที่ถูกซึ่งวิธีการตรวจจะให้ 1 คะแนน ต่อตัวเลือก เมื่อผู้ตอบตอบได้ตรงกับสภาพจริงของตัวเลือกนั้น และให้ 0 คะแนน ต่อตัวเลือก ที่ผู้ตอบตอบไม่ตรงกับสภาพจริงของตัวเลือกนั้น

3. วิธีของอนันต์ ศรีโสภา (อนันต์ ศรีโสภา, 2516, หน้า 13-19) ได้เสนอวิธีการตอบ โดยให้ตอบทุกตัวเลือก การตอบวิธีนี้จะต้องพิจารณาตอบทุกตัวเลือกว่าตัวเลือกใดถูก ตัวเลือกใดผิด

ซึ่งจะทำให้เป็นการวัดความรู้ในข้อคำถามนั้น ๆ ได้สมบูรณ์ยิ่งขึ้นกว่าการเลือกคำตอบที่ถูกเพียงคำตอบเดียว และยังสามารถทราบความรู้ที่ผิดของนักเรียนได้ดีกว่าวิธีตอบแบบธรรมดาอีกด้วย การให้คะแนนก็ให้เป็นรายตัวเลือก ถ้านักเรียนทำเครื่องหมายได้ตรงกับสภาพความเป็นจริงว่าตัวเลือกนั้นถูกหรือผิด จะได้คะแนนตัวเลือกละ 1 คะแนน ถ้าทำเครื่องหมายตรงข้ามกับความเป็นจริง จะได้คะแนนตัวเลือกละ -1 คะแนน ซึ่งวิธีนี้จะทำให้คะแนนที่ได้จากการวัดถูกต้องและเหมาะสมกับความรู้ของนักเรียนมากขึ้น

4. วิธีให้คะแนนแบบนับสอง (Count 2) วิธีการให้คะแนนแบบนี้ ฟูจู ไต และซุน (Fu-Ju Tsai & Suen, 1993, pp. 399-404) ได้เสนอไว้ โดยใช้กับการตอบแบบทดสอบเลือกตอบชนิดถูกผิดหลายตัวเลือก มีวิธีการให้คะแนนดังนี้คือ จะได้คะแนนในข้อสอบแต่ละข้อ เมื่อผู้ตอบเลือกตัวเลือกได้ตรงกับสภาพจริงของตัวเลือกนั้น ๆ ตั้งแต่ 1 ใน 2 ของจำนวนตัวเลือกทั้งหมดในข้อนั้นขึ้นไป ดังเงื่อนไขดังนี้ คือ

$$f(i) = 0 \quad \text{เมื่อ } (x) i = 0$$

$$f(i) = \frac{i-1}{k-1} \quad \text{เมื่อ } (y) \text{ นอกเหนือจาก } (x)$$

โดยที่ i แทน จำนวนตัวเลือกที่ผู้ตอบตอบได้ตรงกับสภาพจริงของตัวเลือกนั้น ๆ
 k แทน จำนวนตัวเลือกทั้งหมดของข้อสอบแต่ละข้อ

ตัวอย่างเช่น ถ้าจำนวนตัวเลือกทั้งหมด (k เป็น 5 ตัวเลือก ผู้ตอบเลือกตัวเลือกได้ตรงกับสภาพจริงของตัวเลือกนั้น 2 ตัวเลือก จะได้คะแนนในข้อนั้นเท่ากับ 1/4

5. วิธีการให้คะแนนแบบนับ 3 (Count 3) วิธีการให้คะแนนแบบนี้ ฟูจู ไต และซุน (Fu-Ju Tsai & Suen, 1993, pp. 399-404) ได้เสนอไว้ โดยใช้กับการตอบแบบทดสอบเลือกตอบชนิดถูกผิดหลายตัวเลือก มีวิธีการให้คะแนนดังนี้ คือจะได้คะแนนในข้อสอบแต่ละข้อ เมื่อผู้ตอบเลือกตัวเลือกได้ตรงกับสภาพจริงของตัวเลือกนั้นตั้งแต่ 3 ใน 4 ของตัวเลือกทั้งหมดของข้อนั้น ดังเงื่อนไขดังนี้ คือ

$$f(i) = 0 \quad \text{เมื่อ } (x) i = 0$$

$$f(i) = \frac{i-2}{k-2} \quad \text{เมื่อ } (y) \text{ นอกเหนือจาก } (x)$$

โดยที่ i แทน จำนวนตัวเลือกที่ผู้ตอบตอบได้ตรงกับสภาพจริงของตัวเลือกนั้น ๆ
 k แทน จำนวนตัวเลือกทั้งหมดของข้อสอบแต่ละข้อ

ตัวอย่างเช่น ถ้าตัวเลือกทั้งหมดเป็น 5 เมื่อผู้ตอบเลือกตัวเลือกได้ตรงกับสภาพจริง
 3 ตัวเลือก จะได้คะแนนในข้อนั้นเท่ากับ $1/3$

6. วิธีการให้คะแนนแบบแก้การเดา (Correction for Guessing: CFG) วิธีการให้คะแนน
 แบบนี้ ฟุจู่ ไต และซุน (Fu-Ju Tsai & Suen, 1993, pp. 399-404) ได้เสนอไว้ เป็นวิธีการตรวจ
 ให้คะแนนที่เป็นการแก้การเดา มีเงื่อนไขการให้คะแนนดังนี้ คือ

$$f(i) = \frac{i - (k - i)}{k} \quad \text{เมื่อ } k > i > 0$$

โดยที่ i แทน จำนวนตัวเลือกที่ผู้ตอบตอบได้ตรงกับสภาพจริงของตัวเลือกนั้น ๆ
 k แทน จำนวนตัวเลือกทั้งหมดของข้อสอบแต่ละข้อ

ตัวอย่างเช่น เมื่อจำนวนตัวเลือกทั้งหมดเป็น 5 ผู้ตอบเลือกตัวเลือกได้ตรงกับสภาพจริง
 ของตัวเลือกนั้น ๆ 3 ตัวเลือก ผู้ตอบจะได้คะแนนในข้อนั้นเป็น $1/5$

7. วิธีการให้คะแนนแบบการเพิ่มคะแนนตัวเลือกที่เว้นไว้ (Let omit) วิธีการให้คะแนน
 แบบนี้ ฟุจู่ ไต และซุน (Fu-Ju Tsai & Suen, 1993, pp. 399-404) ได้เสนอไว้ เป็นวิธีการให้คะแนน
 แบบที่เพิ่มคะแนนให้กับข้อที่เว้นไว้ โดยถือว่าข้อที่ผู้ตอบเว้นไว้ หมายถึง ผู้ตอบไม่มีความรู้
 หรือทำไม่ทัน แต่ถ้าผู้ตอบตอบตัวเลือกนั้น ๆ ผิดไปจากสภาพจริง อาจเป็นเพราะผู้ตอบเดา
 จึงหักคะแนนข้อที่ตอบผิดด้วย เงื่อนไขการให้คะแนนมีดังนี้ คือ

$$f(i) = \frac{i - (k - i - j)}{k} \quad \text{เมื่อ } k > i, j > 0$$

โดยที่ i แทน จำนวนตัวเลือกที่ผู้ตอบตอบได้ตรงกับสภาพจริงของตัวเลือกนั้น ๆ
 j แทน จำนวนตัวเลือกที่ผู้ตอบเว้นไว้
 k แทน จำนวนตัวเลือกทั้งหมดของข้อสอบแต่ละข้อ

ตัวอย่างเช่น ถ้าจำนวนตัวเลือกทั้งหมดเป็น 5 ผู้ตอบเลือกตัวเลือกได้ตรงกับสภาพจริง
 ของตัวเลือกนั้น ๆ 3 ตัวเลือกและมีตัวเลือกที่เว้นไว้ 1 ตัวเลือกจะได้คะแนนในข้อนั้นเป็น $2/5$

จะเห็นได้ว่า การให้คะแนนนั้นมีอยู่หลากหลายแบบ แต่ละแบบต่างก็มีทั้งข้อดีและข้อเสียแตกต่างกันไป การจะนำแบบการให้คะแนนใดไปใช้ก็ต้องพิจารณาถึงความเหมาะสมด้วย สำหรับในการวิจัยครั้งนี้ ผู้วิจัยเลือกใช้การตรวจให้คะแนนแบบวิธี MR (0-1) ซึ่งเป็นการตรวจให้คะแนนแก่ความรู้ที่ถูกต้องสมบูรณ์ และการตรวจให้คะแนนทุกตัวเลือก (MTF) ของอัลบานีส และคณะ นอกจากนี้ ผู้วิจัยยังได้เพิ่มเติมแบบการให้คะแนนทุกตัวเลือกที่ยอมให้ผู้สอบเว้นแสดงความไม่แน่ใจได้ เรียกว่า แบบ OMTF ซึ่งเป็นการตรวจให้คะแนนแก่ความรู้ส่วนย่อย สอดคล้องกับ โมเดลการคิดของบุคคล จึงจะทำให้การลงข้อสรุปเกี่ยวกับความสามารถของบุคคล ถูกต้อง แม่นยำมากที่สุด หรือมีความคลาดเคลื่อนต่ำ คะแนนที่วัดได้จึงจะสอดคล้องกับปริมาณความสามารถที่แท้จริงมากที่สุด (Lord, 1980, หน้า 3)

ทฤษฎีการตอบสนองข้อสอบ (Item response theory)

ทฤษฎีการตอบสนองข้อสอบ เป็นทฤษฎีการวัดที่อธิบายความสัมพันธ์ระหว่างคุณลักษณะภายในหรือความสามารถที่มีอยู่ภายในตัวบุคคลกับพฤติกรรมการตอบสนองข้อสอบของบุคคลนั้น ว่ามีโอกาสตอบข้อสอบถูกมากน้อยเพียงไร ทฤษฎีนี้มีพื้นฐานความเชื่อว่า พฤติกรรมการตอบสนองต่อข้อสอบของผู้สอบ ซึ่งเป็นสิ่งสังเกตได้โดยตรงว่าถูกหรือผิด จะถูกกำหนดโดยคุณลักษณะภายใน (Trait) หรือความสามารถ (Ability) ที่อยู่ภายในตัวบุคคล ซึ่งเป็นสิ่งที่ไม่สามารถสังเกตได้โดยตรง ทฤษฎีนี้ได้อธิบายความสัมพันธ์ดังกล่าวในรูปของฟังก์ชันคณิตศาสตร์หรือโมเดลที่แสดงความสัมพันธ์ระหว่างระดับความสามารถ คุณลักษณะของข้อสอบ และโอกาสของการตอบข้อสอบได้ถูก ที่เรียกว่า “ฟังก์ชันการตอบสนองข้อสอบ” (Item response function) ซึ่งมีหลายรูปแบบ เช่น ฟังก์ชันปกติสะสม (Normal ogive function) ฟังก์ชันโลจิส (Logistic function) เป็นต้น ซึ่งขึ้นอยู่กับข้อตกลงเบื้องต้น ธรรมชาติของการทดสอบ และข้อมูลจากการทดสอบ

จากฟังก์ชันการตอบสนองข้อสอบสามารถนำมาใช้ศึกษาความสัมพันธ์ระหว่างความน่าจะเป็นในการตอบข้อสอบแต่ละข้อได้ถูกต้อง $[P_i(\theta)]$ กับระดับความสามารถของผู้สอบที่วัดได้โดยแบบสอบฉบับนั้น (θ) เมื่อนำมาเขียนเป็นกราฟได้ “โค้งลักษณะข้อสอบ” (Item Characteristic Curve; ICC) โค้งลักษณะข้อสอบมีได้หลายลักษณะ ขึ้นอยู่กับโมเดล (Model) หรือแบบจำลองที่ใช้อธิบายความสัมพันธ์ดังกล่าว เช่น โมเดลแบบหนึ่งพารามิเตอร์ สองพารามิเตอร์ หรือสามพารามิเตอร์ เป็นต้น

หลักการของการวัดความสามารถผู้สอบ

การทดสอบทางการศึกษาส่วนใหญ่เป็นการตรวจสอบความสามารถของผู้เรียน อันเป็นผลจากการเรียนรู้ที่เกิดขึ้น เพื่ออธิบายความสามารถในปัจจุบัน หรือทำนายความสามารถ/ความพร้อมของการเรียนรู้ในอนาคต ความสามารถของบุคคลเป็นคุณลักษณะภายในที่ต้องอาศัย

วิธีการวัดทางอ้อมโดยการตอบสนองต่อสิ่งเร้าที่จัดขึ้น แล้วนำผลการตอบสนองที่ได้ทำการสรุปอ้างอิงไปยังคุณลักษณะภายในที่มุ่งวัดนั้น

1. ความสัมพันธ์ระหว่างความสามารถของผู้สอบ ลักษณะของข้อสอบ และโอกาสของการตอบถูก

การแสดงออกทางพฤติกรรมของมนุษย์เป็นผลอันเนื่องมาจากอิทธิพลของคุณลักษณะเฉพาะบางอย่างภายใน (Latent trait) หรืออาจเรียกว่า คุณลักษณะทางจิตวิทยา (Psychological trait) พฤติกรรมการตอบสนองต่อข้อสอบ (Item) ของผู้สอบก็เช่นกัน ย่อมขึ้นอยู่กับคุณลักษณะเฉพาะภายในของผู้สอบแต่ละคน ในที่นี้ ขอเรียกคุณลักษณะเฉพาะของผู้สอบนั้น ด้วยคำที่มีความหมายกว้าง ๆ ว่า “ความสามารถ (Ability; θ)” อันเกิดจากการเรียนรู้ซึ่งอาจเป็นผลสัมฤทธิ์ทางการเรียน ความสามารถเฉพาะด้านต่าง ๆ เช่น ความสามารถทางคณิตศาสตร์ วิทยาศาสตร์ ภาษา ศิลปะ เป็นต้น

ลอร์ด และ โนวิก (Lord & Novick, 1968) ได้กล่าวไว้ว่า ถ้าเราสามารถนิยามความหมายของ “ความสามารถ” ได้ เราจะสามารถประมาณค่าความสามารถของผู้สอบและละคนได้ ซึ่งคะแนนความสามารถนั้นย่อมนำมาใช้ทำนายหรืออธิบายพฤติกรรมการตอบสนองข้อสอบแต่ละข้อของผู้สอบแต่ละคนได้ จากแนวคิดลักษณะดังกล่าวได้นำไปสู่การสร้างและพัฒนาทฤษฎีการตอบสนองข้อสอบ (Item Response Theory; IRT) ซึ่งประกอบด้วยรูปแบบต่างกันของโมเดลการตอบสนองข้อสอบ (Item response models) โดยนักทฤษฎีวัดผลที่มีชื่อเสียงหลายท่าน เช่น Lowley (1943) Lord (1953; 1980) Birnbaum (1958; 1968) Rasch (1960) Wright (1977; 1979) Lord and Novick (1968) Wright and Stone (1979) Wingersley et al. (1982) Mislevy and Bock (1984) เป็นต้น

โมเดลการตอบสนองข้อสอบเป็นรูปแบบจำลองที่แสดงความสัมพันธ์ระหว่างผลการตอบข้อสอบของผู้สอบ ซึ่งสามารถสังเกตได้ (Observable) กับความสามารถที่แท้จริงของผู้สอบ ซึ่งสังเกตไม่ได้โดยตรง (Unobservable) ความสัมพันธ์ระหว่างความสามารถ (θ) กับโอกาสของการตอบข้อสอบแต่ละข้อได้ถูกต้อง $[P_i(\theta)]$ ดังกล่าว สามารถเขียนได้ในรูปของฟังก์ชันคณิตศาสตร์ที่เป็นฟังก์ชันแบบเพิ่มตลอด (Monotonically increasing function)

สมมติว่า ผู้สอบคนหนึ่งมีระดับความสามารถ θ_k ทำการตอบข้อสอบที่ i ได้ผลการตอบเป็น U_i (ถ้าตอบถูก $U_i = 1$, แต่ถ้าตอบผิด $U_i = \theta$) โอกาส (Probability) ของการตอบข้อสอบที่ i ได้ถูกของผู้นั้นเป็น P_i หรือ $P_i(\theta)$ สามารถอธิบายได้ด้วยฟังก์ชันทางคณิตศาสตร์ $f_i(U_i / \theta_k, B_i)$

เมื่อ B_i เป็นพารามิเตอร์ของข้อสอบที่ i ซึ่ง B_i ประกอบด้วยพารามิเตอร์ a_i (ค่าอำนาจจำแนกของข้อสอบ) b_i (ค่าความยากของข้อสอบ) และ c_i (ค่าโอกาสของการเดาข้อสอบได้ถูก)

$$\text{นั่นคือ} \quad P_i = P_i(\theta) = \text{Prob.}(U_i = 1 / \theta_k, B_i)$$

เมื่อ P_i = โอกาสที่ผู้สอบซึ่งมีความสามารถ θ_k จะตอบข้อสอบที่ i ซึ่งมีพารามิเตอร์ของข้อสอบเป็น B_i ได้อย่างถูกต้อง ($U_i = 1$)

ดังนั้น โอกาสของการตอบข้อสอบแต่ละข้อได้ถูกต้องจึงขึ้นอยู่กับความสามารถของผู้ตอบ และพารามิเตอร์ของข้อสอบข้อนั้น

2. การแจกแจงคะแนนรายข้อและคะแนนรวม

การแจกแจงความถี่ของคะแนนรายข้อ (U_i) ของผู้สอบที่มีความสามารถ θ_k มีลักษณะของการแจกแจงแบบทวินาม (Binomial Distribution) ดังนี้

$$f(U_i / \theta_k, B_i) = P_i Q_i$$

$$\text{เมื่อ} \quad Q_i = 1 - P_i$$

เนื่องจากข้อสอบแต่ละข้อเป็นอิสระจากกัน (ตามข้อตกลงเบื้องต้นของ IRT) การแจกแจงความถี่ของคะแนนรวมทั้งฉบับเมื่อมีข้อสอบ n ข้อ ของผู้สอบที่มีความสามารถ θ_k จะเป็นดังนี้

$$f(U / \theta_k, B) = \prod_{i=1}^n P_i Q_i$$

เมื่อ $U = \{U_i\}$ = เวกเตอร์ของคะแนนสอบรายข้อของผู้สอบคนหนึ่งที่มีความสามารถ θ_k
และ $B = \{B_i\}$ = เมตริกส์ของพารามิเตอร์ของข้อสอบทุกข้อ

สมมติว่า มีผู้เข้าสอบทั้งหมด N คน ซึ่งได้มาอย่างสุ่มจากประชากรของผู้สอบทั้งหมด จึงทำให้เวกเตอร์ของคะแนนสอบรายข้อ U ของผู้สอบจำนวน N เวกเตอร์มีการแจกแจงเป็นอิสระจากกัน ดังนั้น เวกเตอร์ของคะแนนสอบรายข้อทั้งหมด U จำนวน $n \times N$ เวกเตอร์ มีการแจกแจงดังนี้

$$L(U/\theta, B) = \prod_{ij} Q_{ij}$$

เมื่อ $\theta = \{\theta_j\}$ = เวกเตอร์ของความสามารถของผู้สอบจำนวน N คน

$U = \{U\}$ = เวกเตอร์ของคะแนนสอบรายข้อทั้งหมด

3. การประมาณค่าพารามิเตอร์

การแจกแจงร่วมของคะแนนสอบรายข้อทั้งหมด (U) ขึ้นอยู่กับพารามิเตอร์ความสามารถของผู้สอบ ($\theta_1, \theta_2, \dots, \theta_n$) จำนวน N ตัว และพารามิเตอร์ของข้อสอบ (B_1, B_2, \dots, B_n) จำนวน n ตัว เมื่อ r เป็นจำนวนพารามิเตอร์ข้อสอบในโมเดลของทฤษฎีการตอบสนองข้อสอบ ซึ่งอาจเป็น 1, 2, หรือ 3 ตัว ส่วน n เป็นจำนวนข้อสอบในแบบสอบฉบับนั้น

โดยทั่วไปเมื่อมีกลุ่มผู้สอบขนาดใหญ่ (N) ทำแบบสอบที่มีจำนวนข้อสอบมากพอ (n) มากพอ (n) การประมาณค่าพารามิเตอร์ของผู้สอบและข้อสอบจำนวน $N + n$ ตัว มักไม่มีใครมีปัญหาวิธีที่ใช้ในการประมาณค่าพารามิเตอร์มีอยู่หลายวิธี ส่วนใหญ่ที่ใช้กัน ได้แก่ วิธีแมกซิมัมไลค์ลิฮูด (Maximum likelihood) แต่ถ้าทราบการแจกแจงความถี่ของ θ และ B ล่วงหน้า ควรใช้วิธีของเบส์ (Bayesian procedure) เพราะจะทำให้ค่าประมาณของพารามิเตอร์มีความคงที่สูงกว่า (Lord, 1980)

โมเดลการตอบสนองข้อสอบ U (Item response model)

1. ประเภทของโมเดลการตอบสนองข้อสอบ

จากแนวคิดตามทฤษฎีการตอบสนองข้อสอบ ได้มีการพัฒนาโมเดลหรือแบบจำลองขึ้นมาหลายรูปแบบด้วยกัน โดยแต่ละโมเดลจะมีฟังก์ชันทางคณิตศาสตร์และจำนวนพารามิเตอร์ในฟังก์ชันที่แตกต่างกัน ซึ่งสามารถจำแนกเป็น 3 ประเภทใหญ่ ๆ (Hambleton & Swaminathan, 1985) ดังนี้

1.1 โมเดลการตอบสนองข้อสอบที่ใช้กับคะแนนรายข้อแบบ 2 ค่า (Dichotomous) เป็นข้อสอบที่ตรวจให้คะแนน 0, 1 (ตอบผิดได้ 0 คะแนน, ตอบถูกได้ 1 คะแนน) โมเดลประเภทนี้ในระยะเริ่มแรก (ค.ศ. 1943-1968) เช่น Guttman perfect scale, Latent distance model, Linear model เป็นต้น ในระยะต่อมา (ค.ศ. 1952-1982) ได้มีการพัฒนาโมเดลประเภทนี้เพิ่มขึ้นมา เช่น

One-, Two-, Three-parameter normal ogive model และ One-, Two-, Three-, Four-parameter logistic model เป็นต้น

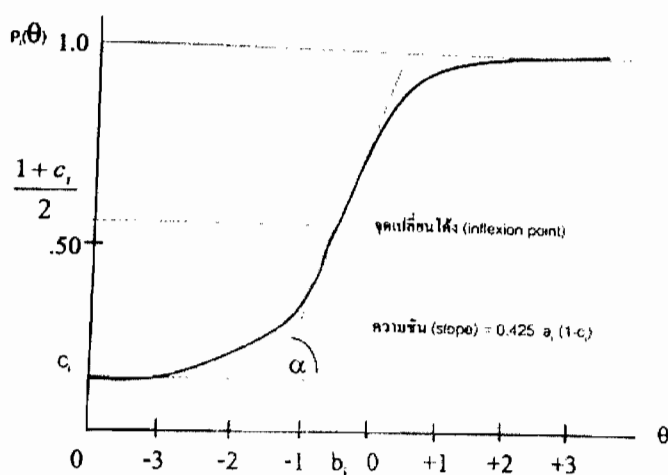
1.2 โมเดลการตอบสนองข้อสอบที่ใช้กับคะแนนรายข้อแบบมากกว่า 2 ค่า (Multichotomous) เช่น Norminal response model, Grade response model, Partial credit model (Samejima, 1972) เป็นต้น

1.3 โมเดลการตอบสนองข้อสอบที่ใช้กับคะแนนรายข้อแบบต่อเนื่อง (Continuous) เช่น Continuous response model (Samejima, 1972) เป็นต้น

ในการศึกษานี้ ผู้ศึกษาขอเสนอรายละเอียดเฉพาะ โมเดลการตอบสนองข้อสอบที่ใช้กับคะแนนรายข้อแบบ 2 ค่าเท่านั้น เพราะเป็นที่นิยมใช้กันทั่วไป โมเดลดังกล่าวเป็นที่รู้จักกัน ในปัจจุบัน ได้แก่ โมเดลปกติสะสม (Normal ogive model) และ โมเดลโลจิส (Logistic model)

โมเดลปกติสะสมใช้ฟังก์ชันปกติสะสม (Normal ogive function) แสดงความสัมพันธ์ระหว่างผลการตอบข้อสอบกับความสามารถของผู้สอบ ส่วน โมเดลโลจิสใช้ฟังก์ชันโลจิส (Logistic function) แสดงความสัมพันธ์ระหว่างผลการตอบกับความสามารถดังกล่าว ซึ่งเป็นฟังก์ชันทั้งสอง ให้ผลลัพธ์ของการประมาณค่าใกล้เคียงกันมาก แต่ฟังก์ชันโลจิสมีลักษณะของสูตรทางคณิตศาสตร์ และวิธีคำนวณง่ายและสะดวกกว่า นอกจากนี้ โมเดลโลจิสยังมีความทนทานต่อความคลาดเคลื่อน ที่เกิดขึ้นกับผู้สอบที่มีความสามารถสูงจะตอบข้อสอบได้ดีกว่า จึงทำให้ โมเดลโลจิสเป็นที่นิยมกันมากในการนำไปใช้จริง (Lord, 1980)

ตัวอย่างของ โค้งคุณลักษณะข้อสอบ (Item Characteristic Curve; ICC) ของ โมเดลโลจิสแบบสามพารามิเตอร์สามารถแสดงได้ดังภาพประกอบที่ 11 (ศิริชัย กาญจนวาสี, 2550, หน้า 55)



ภาพที่ 11 เส้น โค้งคุณลักษณะข้อสอบของ โมเดลแบบ 3-พารามิเตอร์

2. พารามิเตอร์ของโมเดลการตอบสนองข้อสอบ

โมเดลการตอบสนองข้อสอบประกอบด้วยพารามิเตอร์และค่าคงที่ดังนี้

2.1 พารามิเตอร์ของผู้สอบ (Examinee's parameter)

$(\theta)_p$ = ระดับความสามารถของผู้สอบคนที่ P ซึ่งประมาณได้จากโมเดลตามทฤษฎีการตอบสนองข้อสอบ และปรับให้เป็นคะแนนมาตรฐานที่มีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 ค่า θ มีพิสัยอยู่ระหว่าง $-\infty$ ถึง $+\infty$ แต่ในทางปฏิบัตินิยมใช้ในช่วง -3 ถึง +3

2.2 พารามิเตอร์ของข้อสอบ (Item parameters)

b_i = ค่าความยากของข้อสอบข้อที่ i (Item difficulty) ซึ่งเป็นการวัดตำแหน่งของโค้งคุณลักษณะของข้อสอบ (ICC) ตามแกนนอนบนสเกลของ θ ณ จุดที่โค้งมีความชันมากที่สุด (จุดเปลี่ยนโค้ง) หรือที่ตำแหน่งต่อไปนี้

$$b_i = \theta \text{ ที่ } P_i(\theta) = 0.50 \text{ (สำหรับ 1-และ 2-Parameter model)}$$

$$b_i = \theta \text{ ที่ } P_i(\theta) = \frac{1+C}{2} \text{ (สำหรับ 3-Parameter model)}$$

ค่า b มีพิสัยอยู่ระหว่าง $-\infty$ ถึง $+\infty$ แต่ในทางปฏิบัตินิยมใช้ข้อสอบที่ค่า b_i อยู่ระหว่าง -2.5 ถึง +2.5 ค่า b_i ที่อยู่ใกล้ -2.5 แสดงว่าเป็นข้อสอบที่ง่าย ส่วน b_i ที่อยู่ใกล้ +2.5 แสดงว่าเป็นข้อสอบที่ยาก

a_i = ค่าอำนาจจำแนกของข้อสอบที่ i (Item discrimination) ซึ่งเป็นความชันของโค้ง ICC ณ จุดเปลี่ยนโค้ง หรือที่จุด $\theta = b_i$ ค่า a_i มีพิสัยอยู่ระหว่าง $-\infty$ ถึง $+\infty$ ค่า a_i ที่เป็นลบเป็นสิ่งที่ไม่พึงปรารถนา ในทางปฏิบัติจึงนิยมใช้ข้อสอบที่ค่า a_i อยู่ระหว่าง 0.5 ถึง 2.5 ค่า a_i ที่สูงแสดงว่าข้อสอบข้อนั้นมี Slope ที่ชัน จึงจำแนกผู้สอบที่มีความสามารถแตกต่างกันได้ดี

การแปลความหมายของค่าอำนาจจำแนก (Baker, 2001, p. 35)

0.00	จำแนกไม่ได้
0.01-0.34	จำแนกได้ต่ำมาก (Very low)
0.35-0.64	จำแนกได้ต่ำ (Low)
0.65-1.34	จำแนกได้ปานกลาง (Moderate)
1.35-1.69	จำแนกได้สูง (High)
สูงกว่า 1.70	จำแนกได้สูงมาก (Very high)

+infinity จำแนกได้สมบูรณ์

c_i = ความน่าจะเป็นของการเดาถูก (Guessing parameter หรือ Pseudo-chance score level) ซึ่งเป็นความน่าจะเป็นที่ผู้สอบที่มีความสามารถต่ำมาก ๆ จะทำข้อสอบข้อที่ i ได้ถูก c_i เป็นค่ากำกับต่ำที่สุด (Lower asymptote ของ ICC) ค่า c_i มีพิสัยอยู่ระหว่าง 0 ถึง 1 โดยทั่วไปนิยมใช้ข้อสอบมีค่า c_i อยู่ระหว่าง 0 ถึง 0.3

2.3 ค่าคงที่ (Constant)

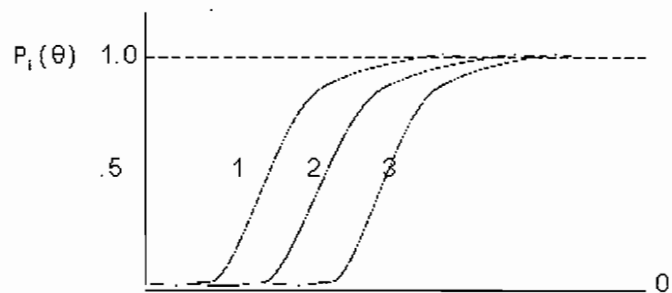
e = ค่าคงที่ ซึ่งมีค่าประมาณ 2.71828

D = ค่าการปรับสเกล (Scaling factor) ซึ่งมีค่าเท่ากับ 1.70

ตารางที่ 3 ฟังก์ชันทางคณิตศาสตร์ของโมเดลการตอบสนองข้อสอบ

Models	Normal ogive function	Logistic function
1. Parameter	$P_i(\theta) = \int_{-\infty}^{\theta-b_i} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$	$P_i(\theta) = \frac{1}{1 + e^{-(\theta-b_i)}}$
2. Parameter	$P_i(\theta) = \int_{-\infty}^{a_i(\theta-b_i)} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$	$P_i(\theta) = \frac{1}{1 + e^{-Da_i(\theta-b_i)}}$
3. Parameter	$P_i(\theta) = c_i + (1 - c_i) \int_{-\infty}^{a_i(\theta-b_i)} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$	$P_i(\theta) = c_i + \frac{1}{1 + e^{-Da_i(\theta-b_i)}}$

ก) 1 - Parameter model

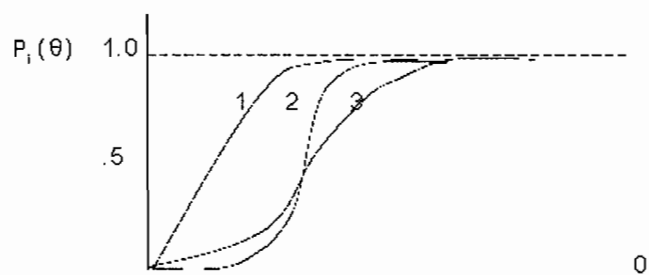


$b_i = \text{vary}$

$a_i = \text{constant (1.0, 0.4 etc.)}$

$c_i = 0$

ข) 2 - Parameter model

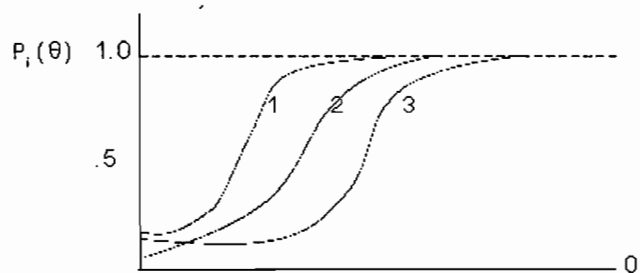


$b_i = \text{vary}$

$a_i = \text{vary}$

$c_i = 0$

ค) 3 - Parameter model



$b_i = \text{vary}$

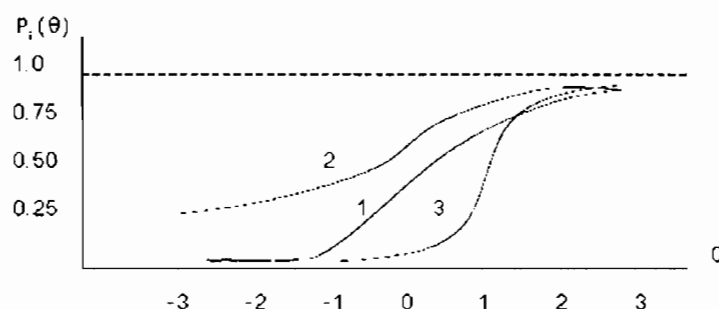
$a_i = \text{vary}$

$c_i = \text{vary}$

ภาพที่ 12 โค้งคุณลักษณะของข้อสอบตามโมเดลการตอบสนองข้อสอบ 3 โมเดล

ตารางที่ 4 ตัวอย่างค่าพารามิเตอร์ของข้อสอบ 3 ข้อ และ $P_i(\theta)$ ที่ระดับ θ ต่าง ๆ

Item	Item Parameters			$P_i(\theta)$						
	a_i	b_i	c_i	-3.0	-2.0	-1.0	0	1.0	2.0	3.0
1.	1.00	0.50	0	0	.01	.07	.30	.70	.93	.99
2.	0.59	0	2.50	.29	.34	.45	.63	.79	.91	.96
3.	1.50	1.00	0	0	0	.01	.07	.50	.93	.99
4.	0.50	-2.50	0.15	.20	.82	.92	.96	.97	.99	.99



ภาพที่ 13 ตัวอย่างโค้งคุณลักษณะข้อสอบ (ICCs) ของข้อสอบ 3 ข้อ

3. ฟังก์ชันสารสนเทศ (Information function)

การวิเคราะห์ตามทฤษฎี IRT จะใช้แบบแผนการตอบสนองแบบสอบเป็นรายข้อ ในการประมาณค่าความสามารถของผู้สอบ ดังนั้น การประเมินคุณภาพของแบบสอบจึงสามารถพิจารณาจากความถูกต้องแม่นยำในการประมาณค่าความสามารถของผู้สอบ ซึ่งมีดัชนีตัวหนึ่งสามารถใช้ชี้ถึงความถูกต้องแม่นยำดังกล่าว เรียกว่า “สารสนเทศของแบบสอบ” (Test information) ซึ่งเป็นค่าฟังก์ชันสารสนเทศของแบบสอบ เกิดจากผลรวมเชิงพีชคณิตของค่าฟังก์ชันสารสนเทศของข้อสอบ (Item information function) แต่ละข้อรวมเข้าด้วยกัน โดยค่าฟังก์ชันสารสนเทศของข้อสอบเป็นดัชนีผสม (Composite index) ที่สร้างจากดัชนีคุณลักษณะของข้อสอบหลายลักษณะ ได้แก่ ค่าความยาก ค่าอำนาจจำแนก และค่าความแปรปรวนของคะแนนรายข้อ เพื่อบ่งชี้คุณภาพของข้อสอบ (Birnbaum, 1968) เนื่องจากค่าสารสนเทศมีความสัมพันธ์ผกผันกับความคลาดเคลื่อนมาตรฐานของการประมาณค่า (θ) ดังนั้น ถ้าค่าความสามารถของผู้สอบในช่วง (θ) นั้น ๆ โดยมีความคลาดเคลื่อนมาตรฐานของการประมาณค่าต่ำ (ศิริชัย กาญจนวาสี, 2545)

ด้วยคุณสมบัติด้านความไม่แปรเปลี่ยนตามกลุ่มตัวอย่าง (Invariance) ของค่าพารามิเตอร์ของข้อสอบจากการวิเคราะห์ ตามทฤษฎี IRT จึงทำให้ค่าฟังก์ชันสารสนเทศเหมาะสมที่จะใช้เป็นดัชนีบ่งบอกคุณภาพของข้อสอบและแบบสอบแทนการหาค่าความเที่ยงและความคลาดเคลื่อนมาตรฐานของการวัดตามทฤษฎีแบบสอบดั้งเดิม (Hambleton & Cook, 1977)

เมื่อ $a_j^{(k)}$ และ $a_j^{(k+1)}$ = ค่าประมาณอำนาจจำแนกของข้อสอบข้อที่ j ที่ควรจะเป็นจากการประมาณครั้งที่ k และ $k+1$ ตามลำดับ

$b_j^{(k)}$ และ $b_j^{(k+1)}$ = ค่าประมาณความยากของข้อสอบข้อที่ j ที่ควรจะเป็นจากการประมาณครั้งที่ k และ $k+1$ ตามลำดับ

$c_j^{(k)}$ และ $c_j^{(k+1)}$ = ค่าประมาณ โอกาสการเดาข้อสอบข้อที่ j ที่ควรจะเป็นจากการประมาณครั้งที่ k และ $k+1$ ตามลำดับ

$$g [a_j^{(k)}] = \text{อนุพันธ์อันดับที่ 1 ของฟังก์ชันการแจกแจงภายหลัง}$$

$$= d \ln f(\theta; b, a, c/u) d a_j$$

$$h [a_j^{(k)}] = \text{อนุพันธ์อันดับที่ 2 ของฟังก์ชันการแจกแจงภายหลัง}$$

$$= d^2 \ln f(\theta; b, a, c/u) d a_j^2$$

$g [b_j^{(k)}]$ และ $h [b_j^{(k)}]$ = อนุพันธ์อันดับที่ 1 และ 2 ของฟังก์ชันการแจกแจงภายหลังของ b_j

$g [c_j^{(k)}]$ และ $h [c_j^{(k)}]$ = อนุพันธ์อันดับที่ 1 และ 2 ของฟังก์ชันการแจกแจงภายหลังของ c_j ค่าอนุพันธ์อันดับที่ 1 และที่ 2 ประกอบไปด้วยส่วนของฟังก์ชัน ไลค์ลิฮูด และส่วนของการแจกแจงเริ่มแรก

ทำการประมาณซ้ำจนกว่าค่าประมาณ a_j , b_j และ c_j จะลู่อู่เข้าสู่ค่าคงที่ค่าใดค่าหนึ่งของค่าพารามิเตอร์นั้น ๆ

4. ประมาณค่าพารามิเตอร์ซ้ำตามขั้น 2 และ 3 จนกว่าจะได้ค่าประมาณพารามิเตอร์ θ_j , b_j , a_j และ c_j ที่ทำให้ค่าของฟังก์ชันการแจกแจงร่วมภายหลัง (Joint posterior distribution) $(\theta; b, a, c/u)$ มีค่าสูงสุด

การวิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบ ตามทฤษฎีนี้การวิเคราะห์ข้อสอบเป็นรายข้อจะวิเคราะห์ค่าความยาก (b) ค่าอำนาจจำแนก (a) ค่าความน่าจะเป็นของการเดาถูก (c) ค่าฟังก์ชันสารสนเทศของข้อสอบ (Item information function) ส่วนการวิเคราะห์ข้อสอบทั้งฉบับจะวิเคราะห์ค่าฟังก์ชันสารสนเทศของแบบสอบ (Test information function) และความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถ ซึ่งถ้าความคลาดเคลื่อนมาตรฐานของการประมาณค่า มีค่าต่ำ แสดงว่า การประมาณค่าความสามารถของผู้เรียนมีความแม่นยำสูง สำหรับการวิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบเหมาะที่จะใช้วิเคราะห์แบบทดสอบที่ได้พัฒนาแล้ว (บุญชม ศรีสะอาด, 2541, หน้า 173)

ทฤษฎีการตอบสนองข้อสอบเป็นทฤษฎีการวัดที่อธิบายให้เห็นความสัมพันธ์ระหว่างความสามารถของผู้สอบกับพฤติกรรมคำตอบข้อสอบ โดยแสดงความสัมพันธ์ดังกล่าวด้วยฟังก์ชันในเชิงคณิตศาสตร์ ซึ่งปัจจุบันได้พัฒนาเป็น โมเดล โลจิสติก จำแนกเป็น โมเดล หนึ่ง สอง และสาม พารามิเตอร์

ค่าความยากของข้อสอบ ตามทฤษฎีนี้ใช้สัญลักษณ์ “b” จะมีค่าอยู่ระหว่าง -4 ถึง +4 ค่าความยากที่ติดลบหมายถึงข้อสอบนั้นง่าย ข้อสอบที่มีค่าความยากเท่ากับศูนย์ หมายถึง ข้อสอบนั้นมีความยากปานกลาง กรณีที่ข้อสอบมีค่าความยากเท่ากับ 3 แสดงว่าข้อสอบนั้นยาก

ค่าอำนาจจำแนกของข้อสอบ ตามทฤษฎีนี้ใช้สัญลักษณ์ “a” มีค่าตั้งแต่ -1 ถึง 1 ค่าอำนาจจำแนกตั้งแต่ 0.20 ขึ้นไป ถือว่าใช้ได้ ค่าอำนาจจำแนกติดลบ หมายถึง จำแนกไม่ได้เช่นกัน และค่าการเดาของข้อสอบ ตามทฤษฎีนี้ใช้สัญลักษณ์ “c” แต่ละข้อไม่ควรเกิน 0.30

การคำนวณค่าพารามิเตอร์ความยาก อำนาจจำแนก และการเดา ต้องคำนวณจากโปรแกรมการคำนวณ โดยเฉพาะ ซึ่งคลังข้อสอบแบบง่ายหากต้องการคำนวณค่าพารามิเตอร์นี้ จะใช้โปรแกรม BILOG หรือ โปรแกรม LOGIST ในการคำนวณได้ การนำทฤษฎีการตอบสนองข้อสอบมาใช้ต้องคำนึงถึงข้อตกลงเบื้องต้นของทฤษฎี ดังนี้

1. แบบทดสอบนั้นต้องวัดความสามารถเดียว
2. ความน่าจะเป็นที่ผู้เข้าสอบจะตอบข้อสอบข้อหนึ่งถูก ต้องไม่ได้รับผลกระทบจากการตอบข้ออื่น ๆ
3. ข้อสอบที่นำมาวิเคราะห์ต้องไม่เป็นข้อสอบวัดความเร็ว

ทฤษฎีการตอบสนองข้อสอบสามารถใช้หาคะแนนจริงของผู้สอบได้ โดยโอกาสที่ผู้มีความสามารถสูงจะตอบข้อสอบข้อที่ยากได้มากกว่าผู้ที่มีความสามารถต่ำ การวิเคราะห์ข้อสอบโดยใช้ทฤษฎีนี้ไม่มีวัตถุประสงค์ในการใช้เพื่อปรับปรุงข้อสอบเช่นเดียวกับการวิเคราะห์

ตามทฤษฎีการทดสอบแบบมาตรฐานเดิม แต่การวิเคราะห์ตามทฤษฎีนี้จะทำให้ได้ค่าพารามิเตอร์ของข้อสอบ ได้แก่ ค่าความยาก ค่าอำนาจจำแนก และการเดาที่เป็นอิสระ ไม่ขึ้นกับความสามารถของผู้สอบ ดังนั้น เมื่อได้ค่าพารามิเตอร์ประจำข้อแล้ว สามารถนำข้อสอบต่าง ๆ มารวบรวมเก็บไว้ในคลังข้อสอบต่อไป

ประสิทธิภาพของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Efficiency of computerized adaptive testing)

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Computerized Adaptive Testing: CAT) เป็นการทดสอบที่ใช้ข้อสอบแตกต่างกันจัดให้กับผู้สอบแต่ละคน โดยพิจารณาจากความสามารถภายในของผู้สอบที่ต้องการวัดและดำเนินการทดสอบด้วยเครื่องคอมพิวเตอร์ ดังนั้น วัตถุประสงค์ของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์จึงเป็นการสร้างแบบทดสอบให้เหมาะสมกับผู้สอบแต่ละคน (Meijer & Nering, 1999) และอาศัยคอมพิวเตอร์ช่วยดำเนินการขั้นตอนการทดสอบต่าง ๆ ปัจจุบันการศึกษาประสิทธิภาพการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์มีความก้าวหน้าอย่างต่อเนื่อง องค์ประกอบสำคัญมีรายละเอียดดังนี้

1. รูปแบบการตอบข้อสอบ (Item response model)

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ใช้หลักการตามทฤษฎีการตอบข้อสอบรูปแบบ 1 2 และ 3 พารามิเตอร์เป็นไปตามข้อตกลงของความเป็นมิติเดียว (Unidimension) ความเป็นอิสระในการตอบข้อสอบ (Local independence) และ โค้งลักษณะข้อสอบ (Item characteristic curve) ก่อนพิจารณากำหนดรูปแบบการตอบข้อสอบ ความเป็นมิติเดียวของแบบทดสอบใด ๆ มีความหมายเหมือนจำนวนของลักษณะแฝง (Latent trait) หรือมิติ (Dimension) ของลักษณะภายในหรือความสามารถที่ต้องการทดสอบ ในขณะเดียวกัน ต้องการความเป็นอิสระในการตอบข้อสอบเพื่อให้การตอบข้อสอบข้อหนึ่งไม่มีความเกี่ยวข้องกับข้อสอบข้ออื่น รักษาพารามิเตอร์ของข้อสอบให้คงที่ไม่ว่าข้อสอบนั้นอยู่ตำแหน่งใดบนแบบทดสอบ ผลที่ได้รับคือการประมาณค่าความสามารถของผู้สอบเป็นอิสระจากชุดข้อสอบ แม้ว่าผู้สอบแต่ละคนจะได้รับชุดข้อสอบแตกต่างกัน ความยากต่างกัน แต่สามารถเปรียบเทียบค่าประมาณความสามารถของผู้สอบเหล่านี้ได้ (Chalhoub-Deville, Alcaya, & Lozier, 1996) และถ้าความเป็นอิสระในการตอบข้อสอบอยู่บนเงื่อนไขบนความเป็นมิติเดียว จะสามารถอธิบายความแตกต่างของรูปแบบการตอบข้อสอบระหว่างผู้สอบได้ (Lord & Novick, 1968, pp. 358-394) การกำหนดรูปแบบการตอบข้อสอบว่าจะเลือกใช้รูปแบบโลจิสติก 1 2 และ 3 พารามิเตอร์ใดนั้น มีคำแนะนำโดยกรีนและคณะ (Green et al., 1984) แนะนำว่า การใช้รูปแบบโลจิสติก 3 พารามิเตอร์ถือว่า

มีความเหมาะสมกับการทดสอบแบบปรับเหมาะมากกว่าใช้รูปแบบโลจิสติก 1 หรือ 2 พารามิเตอร์ เพราะรูปแบบจะไม่สนิทรูป (Fit) กับข้อสอบหลายตัวเลือกถ้าค่าการเดาเท่ากับศูนย์ ไวนเนอร์ และ มีสวาลี (Wainer & Mislevy, 1990, p. 72) กล่าวว่า ในทางปฏิบัติถึงแม้ว่าค่าการเดาอาจไม่จำเป็นเท่าไรนักในระหว่างการดำเนินการทดสอบ เพราะเป็นไปได้ยากที่ผู้สอบจะได้รับข้อสอบความยากไม่เหมาะสมกับความสามารถของตนเอง แต่การใช้ค่าการเดาร่วมด้วยเป็นสิ่งจำเป็น ด้วยเหตุผลว่าในระยะเริ่มแรกของการทดสอบ สารสนเทศของผู้สอบอาจไม่เพียงพอที่จะทำให้การประมาณค่าความสามารถถูกต้องได้ โอกาสเกิดการเดาจากการจัดข้อสอบไม่เหมาะสมในระยะเริ่มแรกจึงเป็นไปได้ และค่าการเดาจำเป็นต้องใช้เพื่อประสิทธิภาพในการคำนวณ อย่างไรก็ตาม การใช้รูปแบบโลจิสติก 3 พารามิเตอร์ก็พบปัญหาได้ คือ พารามิเตอร์ค่าการเดาประมาณค่าได้ยาก เพื่อแก้ปัญหานี้ ชาลฮอบ-ดีวาลี และคณะ (Chalhoub-Deville et al., 1996) แนะนำให้กำหนดค่าการเดาคงที่เท่ากับ 1 และหารด้วยจำนวนตัวเลือก เช่น ข้อสอบมีตัวเลือก 5 ตัวเลือก กำหนดค่าการเดาจะถูกกำหนดไว้ที่ 0.20 อย่างไรก็ตาม วิธีนี้ช่วยแก้ปัญหาลเฉพาะหน้าได้ในระดับหนึ่ง แต่ยังไม่สามารถตอบปัญหาในการระบุค่าการเดาแน่นอนได้

สรุปว่า การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามทฤษฎีการตอบข้อสอบรูปแบบการตอบข้อสอบ 1 2 และ 3 พารามิเตอร์ จะต้องเป็นไปตามข้อตกลง ความเป็นมิติเดียว ความเป็นอิสระในการตอบข้อสอบ และ คุ้งลักษณะข้อสอบก่อน ส่วนการเลือกใช้รูปแบบโลจิสติก 1 2 และ 3 พารามิเตอร์รูปแบบใดนั้น ควรพิจารณาถึงลักษณะของข้อมูลและภาวะสนิทรูป (Fit) ระหว่างข้อมูลการตอบข้อสอบและรูปแบบที่เลือกใช้

2. คลังข้อสอบ (Item pool)

คลังข้อสอบ (Item pool หรือ Question banks, Item collection, Item reservoirs, Test item libraries) เป็นแหล่งรวบรวมข้อสอบหรือคำถามแบบทดสอบ การทดสอบด้วยแบบทดสอบประเพณีนิยม ผู้สอบทั้งหมดจะได้รับข้อสอบชุดเดียวกัน แต่สำหรับการทดสอบแบบปรับเหมาะ ผู้สอบจะได้รับข้อสอบเหมาะสมกับความสามารถของผู้สอบแต่ละคน คลังข้อสอบจึงเป็นเสมือนแหล่งรวบรวมชุดแบบทดสอบจำนวนมากและมีความเป็นคู่ขนานกัน ข้อสอบบรรจุในคลังข้อสอบมีลักษณะค่าความยากกระจายเต็มระดับความสามารถของประชากรผู้สอบ ค่าอำนาจจำแนกสูง เพื่อให้ประสิทธิภาพการทดสอบสูงสุด และค่าการเดาเข้าใกล้หรือเท่ากับศูนย์ ข้อสอบควรมีจำนวนมากเพียงพอและกระจายทุกระดับความสามารถของผู้สอบ ดังนั้น ต้องใช้ข้อสอบจำนวนมากสำหรับสร้างคลังข้อสอบ ทำให้คลังข้อสอบสำหรับการทดสอบแบบปรับเหมาะมีขนาดใหญ่กว่าการทดสอบประเพณีนิยม (Flaugher, 1990, pp. 41-62)

การสร้างข้อสอบจำนวนมากใช้สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์
ขั้นตอนดังนี้ ได้แก่

- 1.1 สร้างข้อสอบใหม่ให้มีจำนวนเพียงพอในแต่ละเนื้อหา เน้นที่ครอบคลุม
เนื้อหาหลักและกฎเบื้องต้นในการเขียนข้อสอบ
- 1.2 ตรวจสอบคุณภาพของข้อสอบเพื่อให้ข้อสอบมีคุณภาพสูงสุดเท่าที่เป็นไปได้
- 1.3 ทำการทดสอบเบื้องต้นเพื่อตรวจสอบข้อบกพร่องของข้อสอบใหม่
- 1.4 ทำการเลือกชุดย่อยของข้อสอบใหม่ ตอบข้อสอบ และวิเคราะห์ผลสอบ
ตามทฤษฎีการทดสอบแบบมาตรฐานเดิมและทฤษฎีการตอบข้อสอบ เพื่อจัดข้อสอบไม่เหมาะสม
ออกไปหรือทำการปรับปรุงข้อสอบใหม่
- 1.5 เปรียบเทียบสมมูลเนื้อหาของคลังข้อสอบด้วยแบบทดสอบตามทฤษฎี
การทดสอบแบบมาตรฐานเดิม เพื่ออ้างอิงสมมูลเนื้อหาของคลังข้อสอบในการทดสอบ
แบบปรับเหมาะ ประเมินระบบการทดสอบด้วยการจำลองพฤติกรรมกรรมการตอบข้อสอบของผู้สอบ
ระดับความสามารถต่าง ๆ กัน คะแนนทั้ง 2 การทดสอบนี้ต้องเปรียบเทียบกันได้
- 1.6 นำแสดงข้อสอบตามแนวทางการจัดการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์
ต่อไป (Flaugher, 1990, p. 43)

สรุป คลังข้อสอบเป็นแหล่งรวบรวมข้อสอบ จำนวนข้อสอบต้องมีจำนวนมากเพียงพอ
กับทุกระดับความสามารถของผู้สอบ ข้อสอบในคลังข้อสอบควรมีค่าความยากกระจายเต็มระดับ
ความสามารถของประชากรผู้สอบ ค่าอำนาจจำแนกสูง และค่าการเดาเข้าใกล้หรือเท่ากับศูนย์
แบบทดสอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับผู้สอบแต่ละคนต้องม
ีความเป็นคู่ขนานเปรียบเทียบกันได้ คลังข้อสอบควรมีความเป็นมิติเดียว ถ้าคลังข้อสอบมีหลายมิติ
อาจแบ่งคลังข้อสอบใหญ่เป็นคลังข้อสอบย่อยหลาย ๆ คลังข้อสอบแทนแต่ละมิติ คลังข้อสอบ
ขนาดใหญ่จะให้สารสนเทศแบบทดสอบได้ดีกว่าคลังข้อสอบขนาดเล็ก การกำหนดขนาดคลังข้อสอบ
ควรพิจารณาปัจจัยต่าง ๆ นำมาพิจารณาออกแบบคลังข้อสอบ

3. จำนวนข้อสอบ (Test length)

ทฤษฎีการตอบข้อสอบให้ความสำคัญกับสารสนเทศข้อสอบ สารสนเทศแบบทดสอบ
บนระดับความสามารถหนึ่ง ได้รับมาจากผลรวมสารสนเทศของข้อสอบทุกข้อในแบบทดสอบ
บนระดับ 41 ความสามารถนั้น ๆ ดังนั้น เมื่อความยาวแบบทดสอบเพิ่มขึ้นสารสนเทศแบบทดสอบ
จะเพิ่มขึ้นตามไปด้วย หลังจากการตอบข้อสอบแต่ละข้อ ความคลาดเคลื่อนของการประมาณค่า
ความสามารถของผู้สอบจะลดลง จนค่าประมาณความสามารถเข้าใกล้ค่าความสามารถจริงของ
ผู้สอบมากที่สุด (Hambleton & Swaminathan, 1985, pp. 101-124) ไวน์เนอร์ และมีสวาลี (Wainer &

Mislevy, 1990, p. 77) เสนอว่า ในทางปฏิบัติการทดสอบแบบปรับเหมาะ ความถูกต้องของการประมาณค่าความสามารถของผู้สอบทำได้โดยการเพิ่มจำนวนข้อสอบ และเพิ่มข้อสอบความยากเหมาะสมเข้าไป ข้อเสนอแนะนี้ตรงกับการศึกษาของ ลุง ชาง และฮัว (Leung Chang & Hau, 2003) พบว่า เมื่อจำนวนข้อสอบเพิ่มขึ้นผลตามมาก็คือ ความสัมพันธ์ของค่าประมาณความสามารถและค่าความสามารถจริงของผู้สอบสูงขึ้นตามกัน ความคลาดเคลื่อนของการประมาณค่าความสามารถของผู้สอบค่อย ๆ ลดลง ดังนั้น การเพิ่มจำนวนข้อสอบจะเพิ่มประสิทธิภาพการประมาณค่าความสามารถของผู้สอบสูงตามไปด้วย การศึกษาของ ลุง ชาง และฮัว (Leung Chang & Hau, 2003) รายงานผลข้างเคียงของการเพิ่มจำนวนข้อสอบว่า ทำให้การแสดงผลข้อสอบมากเกินไป อัตราการแสดงผลข้อสอบกำหนด และให้ค่าความลำเอียงเพิ่มขึ้นแต่ไม่มากนัก ในขณะที่ทิสเซน (Thissen, 1990, p. 169) รายงานผลความยาวแบบทดสอบที่ไม่เหมาะสมอาจให้สารสนเทศแบบทดสอบมีลักษณะเป็นโค้งแบน จำนวนข้อสอบจะแสดงความยาวแบบทดสอบจึงเป็นสิ่งสำคัญ มีผลต่อประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ และสมดุลการใช้ข้อสอบในคลังข้อสอบ การออกแบบความยาวแบบทดสอบควรพิจารณาปัจจัยเกี่ยวข้องต่าง ๆ ในการทดสอบ อาจกำหนดความยาวแบบทดสอบเบื้องต้นจากข้อเสนอของซีเกลให้คลังข้อสอบมีขนาดประมาณ 6-8 เท่า ของความยาวแบบทดสอบ (Segall, 2003)

สรุป การเพิ่มจำนวนข้อสอบที่แสดงถึงความยาวแบบทดสอบสามารถเพิ่มประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ แต่ความยาวแบบทดสอบมากเกินไปอาจให้สารสนเทศแบบทดสอบลดลง และการแสดงผลข้อสอบมากเกินไป อัตราการแสดงผลข้อสอบกำหนด ขาดสมดุลการใช้ข้อสอบในคลังข้อสอบ การกำหนดความยาวแบบทดสอบจึงเป็นสิ่งสำคัญ ต้องพิจารณาถึงปัจจัยประกอบต่าง ๆ ที่เกี่ยวข้องในการออกแบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ร่วมไปด้วย

นักพัฒนาแบบทดสอบ ที่ผ่านมาใช้การศึกษาในสถานการณ์จำลองเป็นส่วนใหญ่ เพื่อปรับวิธีกำหนดและทดสอบคลังข้อสอบก่อนทำการทดสอบในสถานการณ์จริงโดยกำหนดค่าประมาณพารามิเตอร์ข้อสอบ บรรจุข้อสอบในคลังข้อสอบ ทำการทดสอบกับกลุ่มผู้สอบจำลองระดับความสามารถต่าง ๆ ผลการศึกษาจะให้ค่าเฉลี่ยและความแปรปรวนของการประมาณค่าความสามารถของผู้สอบอย่างมีเงื่อนไข คอค, ดิอียาลา และน็อซ (Dodd, De Ayala, & Koch, 1995) กล่าวถึงข้อดีของการใช้คลังข้อสอบจำลองว่า การกำหนดค่าพารามิเตอร์สำหรับจัดการออกแบบระบบ ทำให้สามารถติดตามตัวแปรพื้นฐานที่สนใจได้ การศึกษาในสถานการณ์จำลองนี้จะยอมให้ผลการทดสอบมาจากการเปลี่ยนแปลงต่าง ๆ เช่น การควบคุมเนื้อหา ขนาดคลังข้อสอบกฎการยุติ อัตราการแสดงผลข้อสอบ เป็นต้น ซึ่งจะนำมาทดสอบและเปรียบเทียบ ผลการศึกษาใช้เป็นพื้นฐาน

สำหรับการกำหนดลักษณะคลังข้อสอบและวิธีกำหนดที่เหมาะสมต่อไป (Segall, 2003) การศึกษาเปรียบเทียบการทดสอบแบบปรับเหมาะในสถานการณ์จำลองและในสถานการณ์การทดสอบปฏิบัติจริงเมื่อใช้รูปแบบตามทฤษฎีการตอบข้อสอบพบว่า การทดสอบทั้งสองให้ผลสอดคล้องกันไม่แตกต่างกันมากนัก (Johnson & Weiss, 1980 cited in Weiss, 1985) เช่น การศึกษาการเลือกข้อสอบงานวิจัยวิธีตามระดับขั้นของค่าอำนาจจำแนก (Chang & Ying, 1999; Leung et al., 2002) การศึกษาในสถานการณ์จำลองเป็นเพียงการออกแบบการทดลองเพื่อจำลองสถานการณ์เสมือนจริงเมื่อไม่สามารถได้ข้อมูลจากการวิจัยอื่นหรือยากจะได้รับ การทดลองดำเนินการเป็นระบบภายใต้เงื่อนไขกำหนด ผลการศึกษานำไปอ้างอิงได้กับผลการทดลองเชิงประจักษ์ ขั้นตอนเริ่มด้วยการกำหนดปัญหาการทดสอบ การออกแบบการศึกษา การใช้โปรแกรมคอมพิวเตอร์และ/หรือ การเขียนคำสั่งใหม่ และการวิเคราะห์ผลการศึกษา การทดสอบแบบปรับเหมาะเมื่อศึกษาในสถานการณ์จำลองจะให้ความสำคัญกับลักษณะคลังข้อสอบ วิธีกำหนด และผลจากปฏิสัมพันธ์เชิงซ้อนของการทดสอบซึ่งส่งผลต่อประสิทธิภาพการทดสอบ เมื่อเงื่อนไขอย่างใดอย่างหนึ่งเปลี่ยนไป จะต้องทำการศึกษาในสถานการณ์จำลองก่อนเพื่อใช้เป็นข้อมูลพื้นฐานสำหรับการกำหนดลักษณะคลังข้อสอบและวิธีกำหนดอย่างเหมาะสมในการทดสอบจริงต่อไป

การศึกษาประสิทธิภาพของทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ จำเป็นต้องศึกษาระบบคลังข้อสอบ (Item pool หรือ Question banks, Item collection, Item reservoirs, Test item libraries) ซึ่งเป็นแหล่งรวบรวมข้อสอบหรือคำถามแบบทดสอบ การทดสอบด้วยแบบทดสอบประเพณีนิยม ผู้สอบทั้งหมดจะได้รับข้อสอบชุดเดียวกัน แต่สำหรับการทดสอบแบบปรับเหมาะ ผู้สอบจะได้รับข้อสอบเหมาะสมกับความสามารถของผู้สอบแต่ละคน คลังข้อสอบจึงเป็นเสมือนแหล่งรวบรวมชุดแบบทดสอบจำนวนมากและมีความเป็นคู่ขนานกัน ข้อสอบบรรจุในคลังข้อสอบมีลักษณะค่าความยากกระจายเต็มระดับความสามารถของประชากรผู้สอบ ค่าอำนาจจำแนกสูง เพื่อให้ประสิทธิภาพการทดสอบสูงสุด และค่าการเดาเข้าใกล้หรือเท่ากับศูนย์ ข้อสอบควรมีจำนวนมากเพียงพอและกระจายทุกระดับความสามารถของผู้สอบ ดังนั้น ต้องใช้ข้อสอบจำนวนมากสำหรับสร้างคลังข้อสอบ ทำให้คลังข้อสอบสำหรับการทดสอบแบบปรับเหมาะมีขนาดใหญ่กว่าการทดสอบประเพณีนิยม (Flaughner, 1990, pp. 41-62)

การสร้างข้อสอบจำนวนมากใช้สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ขั้นตอนดังนี้ ได้แก่

1. สร้างข้อสอบใหม่ให้มีจำนวนเพียงพอในแต่ละเนื้อหาเน้นที่ครอบคลุมเนื้อหาหลัก และกฎเบื้องต้นในการเขียนข้อสอบ
2. ตรวจสอบคุณภาพของข้อสอบเพื่อให้ข้อสอบมีคุณภาพสูงสุดเท่าที่เป็นไปได้

3. ทำการทดสอบเบื้องต้นเพื่อตรวจสอบข้อบกพร่องของข้อสอบใหม่
4. ทำการเลือกชุดย่อยของข้อสอบใหม่ ตอบข้อสอบ และวิเคราะห์ผลสอบตามทฤษฎีการทดสอบแบบมาตรฐานเดิมและทฤษฎีการตอบข้อสอบ เพื่อจัดข้อสอบไม่เหมาะสมออกไปหรือทำการปรับปรุงข้อสอบใหม่
5. เปรียบเทียบสมมูลเนื้อหาของคลังข้อสอบด้วยแบบทดสอบตามทฤษฎีการทดสอบแบบมาตรฐานเดิม เพื่ออ้างอิงไปยังสมมูลเนื้อหาของคลังข้อสอบในการทดสอบแบบปรับเหมาะประเมินระบบการทดสอบด้วยการจำลองพฤติกรรมกรตอบข้อสอบของผู้สอบระดับความสามารถต่าง ๆ กัน คะแนนทั้ง 2 การทดสอบนี้ต้องเปรียบเทียบกันได้ และ
6. นำแสดงข้อสอบตามแนวทางการจัดการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ต่อไป (Flaugher, 1990, p. 43)

วิธีเลือกข้อสอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ให้สามารถเปรียบเทียบได้กับการทดสอบใช้กระดาษและดินสอ มีจำเป็นต้องจัดสมมูลเนื้อหาเพื่อใช้เป็นตัวแทนของแบบทดสอบ ถ้าการทดสอบใช้กระดาษและดินสอเป็นพหุมิติ แต่การทดสอบแบบปรับเหมาะข้อตกลงกำหนดว่า คลังข้อสอบต้องเป็นมิติเดียว การแก้ปัญหานี้อาจใช้วิธีการแบ่งคลังข้อสอบใหญ่เป็นคลังข้อสอบย่อย คลังข้อสอบย่อยแต่ละคลังแทนแต่ละมิติ คะแนนแต่ละส่วนมาจากแต่ละมิติหรือแต่ละเนื้อหา ผลการทดสอบหาโดยคำนวณรวมน้ำหนักของคะแนนต่าง ๆ ในแต่ละมิติ แต่วิธีนี้อาจยุ่งยากในการทดสอบและการรายงานผลตามมา (Flaugher, 1990, p. 60) กรีน และคณะ (Green et al., 1984) ให้เหตุผลสำหรับใช้คลังข้อสอบมิติเดียวว่า รูปแบบการทดสอบแบบปรับเหมาะที่เหมาะสมเกิดได้ก็ต่อเมื่อข้อสอบมีมิติเด่นมิติเดียว (Dominant dimension) ข้อสอบมิติเด่น 2 มิติหรือข้อสอบพหุมิติมีแนวโน้มให้ค่าอำนาจจำแนกต่ำ และการเลือกข้อสอบอำนาจจำแนกสูงมีแนวโน้มให้ความเป็นมิติเดียวโดยตัวของมันเองอยู่แล้ว

การสร้างคลังข้อสอบมากกว่าหนึ่งคลังข้อสอบหรือแบ่งคลังข้อสอบใหญ่เป็นคลังข้อสอบย่อยสามารถสุ่มเลือกคลังข้อสอบและหมุนเวียนการใช้ชุดแบบทดสอบไปตามคลังข้อสอบต่าง ๆ ได้ (Stocking, 1988; 1994 cited in Wang & Kolen, 2001) การสร้างคลังข้อสอบจำนวนมากต้องทำการศึกษาเปรียบเทียบระหว่างคลังข้อสอบก่อนนำไปใช้ พิจารณาปัจจัยจากต่าง ๆ ดังนี้

1. ขนาดของคลังข้อสอบ
2. จำนวนคลังข้อสอบเมื่อใช้พร้อม ๆ กัน
3. การทับซ้อนของคลังข้อสอบ
4. วิธีเข้าสู่แต่ละคลังข้อสอบ

5. การหมุนเวียนใช้คลังข้อสอบและเกณฑ์การออกจากคลังข้อสอบ
6. กฎการนำข้อสอบกลับมาใช้ใหม่
7. การทดสอบข้อสอบก่อนนำมาใช้ วิธีการคำนวณข้อสอบ และวิธีการควบคุมช่วงความสามารถบนเส้นต่อเนื่อง และ
8. วิธีการเลือกข้อสอบเพื่อให้ครอบคลุมแต่ละเนื้อหาวิชา ควบคุมการแสดงข้อสอบ อัตราการทับซ้อนและอื่น ๆ (Wang & Kolen, 2001)

ปฏิบัติการรวมแบบทดสอบในคลังข้อสอบย่อยเข้าเป็นแบบทดสอบใหญ่ อาจประยุกต์จากผลการศึกษาของเวนเคอลินเดน (Van der Linden, 1998) ศึกษาเปรียบเทียบวิธีรวมข้อสอบเข้าแบบทดสอบ 4 วิธี (1. Heuristic-base test 2. 0-1 Linear programming 3. Network-float programming และ 4. Optimal design approach) ขณะที่สต็อกกิง และสเวนสัน (Stocking & Swanson, 1998) ประยุกต์วิธีการออกแบบเหมาะสม (Optimal design) เพื่อใช้สำหรับรวบรวมแบบทดสอบแบบปรับเหมาะเข้าไว้ด้วยกัน การศึกษาเหล่านี้อยู่ในระยะเริ่มต้น อาจนำไปใช้กับการรวมคลังข้อสอบสำหรับการทดสอบแบบปรับเหมาะในโอกาสต่อไป

ขนาดคลังข้อสอบมีส่วนช่วยเพิ่มประสิทธิภาพการทดสอบ ไวส์ (Weiss, 1985) แนะนำว่าการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์จะให้ประสิทธิภาพเมื่อคลังข้อสอบมีจำนวนข้อสอบมากเพียงพอสำหรับจำแนกผู้สอบ การศึกษาของซิง และแฮมเบิลตัน (Xing & Hambleton, 2004) พบว่า คลังข้อสอบขนาดใหญ่ทำให้สารสนเทศแบบทดสอบเพิ่มขึ้นตามไปด้วย สอดคล้องกับความคิดเห็นของลอร์ด (Lord, 1980, pp. 150-161) ซึ่งอ้างถึงการศึกษานในสถานการณ์จำลอง เมื่อใช้จำนวนข้อสอบ 25 ข้อ ขนาดคลังข้อสอบ 363 ข้อ และ 183 ข้อ พบว่า ขนาดคลังข้อสอบ 363 ข้อ สามารถให้สารสนเทศได้มากกว่า เป็นเช่นนี้เพราะคลังข้อสอบขนาดใหญ่มีโอกาสเลือกข้อสอบที่ดีที่สุดจัดให้ผู้สอบได้มากกว่าคลังข้อสอบขนาดเล็ก

การกำหนดขนาดคลังข้อสอบพิจารณาปัจจัยหลายอย่าง เชาฮูบ-คิว และคณะ (Chalhoub-Deville et al., 1996) กล่าวถึงปัจจัยในการออกแบบคลังข้อสอบที่ส่งผลต่อขนาดคลังข้อสอบ เช่น วิธีการเลือกข้อสอบ การควบคุมเนื้อหามาตรฐานทางจิตวิทยา กฎการยุติ การทับซ้อนข้อสอบ การให้คะแนนแบบทดสอบ ความเป็นคู่ขนานกับแบบทดสอบใช้กระดาษและดินสอ เป็นต้น เบิร์กสตรอม และลูนซ์ (Bergstrom & Lunz, 1999, pp. 67-91) อ้างถึงลักษณะแบบทดสอบสามารถนำไปประมาณขนาดคลังข้อสอบได้ว่า

1. จำนวนข้อสอบต่อขนาดคลังข้อสอบเท่ากับ 1 ต่อ 6-8
2. ใช้ข้อสอบเพียง 10-15 เปอร์เซ็นต์ของประชากรผู้สอบ
3. อัตราการทับซ้อนข้อสอบที่จัดให้แก่ผู้สอบแต่ละคู่ต้องไม่มากเกินไป 15-20 เปอร์เซ็นต์

4. เปอร์เซ็นต์การทับซ้อนข้อสอบสูงสุดของแบบทดสอบแบบปรับเหมาะ 2 ชุดไม่เกิน 40 เปอร์เซ็นต์ ระหว่างผู้สอบความสามารถระดับเดียวกัน

เบิร์กสตรอม และ ลูนซ์ อ้างถึงจำนวนผู้สอบถ้าน้อยกว่า 1,000 คน ขนาดคลังข้อสอบอย่างน้อยที่สุดควรประมาณ 500 ข้อ ขนาดคลังข้อสอบโดยทั่วไปประมาณ 600 ถึง 800 ข้อ และต้องกระจายครอบคลุมเนื้อหาข้อสอบ (Bergstrom & Lunz, 1999, pp. 67-91) ซีเกิล (Segall, 2003) เสนอว่า ขนาดคลังข้อสอบควรมีขนาดประมาณ 6-8 เท่าของจำนวนข้อสอบของแบบทดสอบ

โดยสรุป คลังข้อสอบเป็นแหล่งรวบรวมข้อสอบ จำนวนข้อสอบต้องมีจำนวนมากเพียงพอกับทุกระดับความสามารถของผู้สอบ ข้อสอบในคลังข้อสอบควรมีค่าความยากกระจายเต็มระดับความสามารถของประชากรผู้สอบ ค่าอำนาจจำแนกสูง และค่าการเดาเข้าใกล้หรือเท่ากับศูนย์ แบบทดสอบของการทดสอบแบบปรับเหมาะสำหรับผู้สอบแต่ละคนต้องมีความเป็นคู่ขนานเปรียบเทียบกันได้ คลังข้อสอบควรมีความเป็นมิติเดียว ถ้าคลังข้อสอบมีหลายมิติ อาจแบ่งคลังข้อสอบใหญ่เป็นคลังข้อสอบย่อยหลาย ๆ คลังข้อสอบแทนแต่ละมิติ คลังข้อสอบขนาดใหญ่จะให้สารสนเทศแบบทดสอบได้ดีกว่าคลังข้อสอบขนาดเล็ก การกำหนดขนาดคลังข้อสอบ ควรพิจารณาปัจจัยต่าง ๆ นำมาพิจารณาออกแบบคลังข้อสอบ

งานวิจัยที่เกี่ยวข้อง

สมจวน มานะก่อ (2551) ได้ศึกษาเปรียบเทียบคุณภาพของแบบทดสอบเลือกตอบ เมื่อตรวจด้วยวิธีการให้คะแนนความรู้บางส่วนของแบบทดสอบวัดผลสัมฤทธิ์คณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ผลการวิจัยพบว่า

1. ความเที่ยงตรงตามสภาพ (Concurrent validity) ของแบบทดสอบวัดผลสัมฤทธิ์คณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 เมื่อตรวจด้วยวิธีการให้คะแนนความรู้บางส่วน โดยพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ (r) ระหว่างคะแนนสอบกลางภาคกับคะแนนที่ได้มาจากการตรวจด้วยวิธีของการทดสอบความน่าจะเป็นกับวิธีการตรวจให้คะแนนตามความยากประจำชั้น มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ใช้สถิติไควสแควร์ (χ^2) ทดสอบความแตกต่าง โดยใช้คะแนนพิชเชอร์ซีเปรียบเทียบในภาพรวม พบว่า ความเที่ยงตรงตามสภาพไม่ขึ้นอยู่กับวิธีการตรวจให้คะแนน
2. ฟังก์ชันสารสนเทศของข้อสอบ (Item information) วิธีการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็น ให้ค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงที่สุดในระดับความสามารถ -1 มีค่าเท่ากับ 0.097 โดยเฉลี่ยสูงกว่าวิธีการตรวจให้คะแนนตามความยากประจำชั้นที่มีค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงที่สุดในระดับความสามารถ -1 มีค่าเท่ากับ

0.092 และค่าประสิทธิภาพสัมพัทธ์ (Relative Efficiency: $RE(\theta)$) ที่ระดับความสามารถ (θ)-1.4 ถึง 0.6 วิธีการตรวจให้คะแนนด้วยวิธีการของการทดสอบความน่าจะเป็นมีประสิทธิภาพสูงกว่าวิธีการตรวจให้คะแนนด้วยวิธีการให้คะแนนตามความยากประจำชั้น

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

สุนิสา จุยม่วงศรี (2546) ได้ศึกษาผลการเทียบคะแนนข้อสอบที่ตรวจให้คะแนนแบบหลายค่า โดยมีจุดมุ่งหมายเพื่อศึกษาผลการเทียบคะแนนที่ข้อสอบตรวจให้คะแนนหลายค่าตามวิธีการเทียบคะแนนในทฤษฎีการตอบสนองข้อสอบและทฤษฎีการทดสอบแบบมาตรฐานเดิม ผลการเทียบคะแนนพิจารณาความคล้ายคลึงกันของการแจกแจงคะแนนสะสม ดัชนีคะแนนความแตกต่างและดัชนีความแตกต่าง (C) จากกลุ่มสอบทาน ผลการศึกษาพบว่า

การสร้างตารางเปรียบเทียบคะแนนแบบทดสอบ จากการเทียบคะแนนทั้ง 5 วิธี 8 รูปแบบ จำแนกเป็นการเทียบคะแนนตามทฤษฎีการทดสอบมาตรฐานเดิม ได้แก่ วิธีการเทียบคะแนนเชิงเส้นตรงและอิกวิเปอร์เซน ไทล์ เป็นการเทียบคะแนนดิบจากแบบทดสอบฉบับ Y ปรับสู่สเกลแบบทดสอบฉบับ X และการเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบจำแนกตามวิธีการวิเคราะห์ค่าพารามิเตอร์ ซึ่งมีทั้งวิเคราะห์แยกฉบับและวิเคราะห์พร้อมกันทั้ง 3 วิธี แต่ละวิธีเทียบคะแนนโดยใช้คะแนนจริงและคะแนนสังเกตจากแบบทดสอบฉบับ Y มีค่าสูงกว่าคะแนนของแบบทดสอบฉบับ X ในช่วงคะแนนที่แตกต่างกันในแต่ละวิธีการเทียบคะแนน

สุพจน์ เกิดสุวรรณ (2545) ได้พัฒนาวิธีการวัดความรู้บางส่วนของผู้ตอบแบบสอบถาม การดำเนินการวิจัยประกอบด้วย 2 ขั้นตอน ขั้นตอนแรกเป็นการพัฒนาวิธีการตอบและตรวจให้คะแนน โดยประยุกต์วิธีของคูมบ์ ประยุกต์วิธีของอาร์โนลด์ และประยุกต์วิธีของเดรสเชลและชมิค ขั้นตอนที่สอง วิเคราะห์คุณภาพวิธีการตอบและตรวจให้คะแนนที่พัฒนาขึ้นเพื่อเปรียบเทียบคุณภาพกับวิธีการตอบและตรวจให้คะแนนตามวิธีของคูมบ์ วิธีของอาร์โนลด์วิธีของเดรสเชลและชมิค และวิธีการตรวจให้คะแนนแบบประเพณีนิยม ในด้านความตรงตามเกณฑ์ความตรงเชิงโครงสร้าง ความเที่ยงแบบความสอดคล้องภายใน ค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบสอบ และค่าประสิทธิภาพสัมพัทธ์เฉลี่ยของแบบสอบ เครื่องมือที่ใช้ในการวิจัยเป็นแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ขั้นพื้นฐาน ระดับมัธยมศึกษาตอนต้น โดยการสุ่มกลุ่มตัวอย่างนักเรียนระดับมัธยมศึกษาปีที่ 3 จากโรงเรียนมัธยมศึกษาในจังหวัดสระบุรี ปีการศึกษา 2544 จำนวน 15 โรงเรียน รวมทั้งสิ้น 946 คน ข้อมูลที่ได้นำมาตรวจสอบความตรงเชิงโครงสร้าง โดยการวิเคราะห์เมตริกพหุลักษณะ-พหุวิธี และวิเคราะห์ตัวประกอบเชิงยืนยันด้วย โปรแกรม

LISREL ตรวจสอบความตรงตามเกณฑ์สัมพัทธ์จากการวิเคราะห์ สหสัมพันธ์ของคะแนนกับ เกณฑ์วิเคราะห์ค่าความเที่ยงแบบความสอดคล้องภายใน และวิเคราะห์ค่าฟังก์ชันสารสนเทศ ของแบบสอบ ด้วยโปรแกรม BILOG และ PARSCALE ผลการวิจัยสรุปได้ดังนี้

1. วิธีที่ประยุกต์จากวิธีของคูมบ์ ให้ค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบสอบ และประสิทธิภาพสัมพัทธ์เฉลี่ยของแบบสอบสูงกว่าวิธีอื่น ๆ รองลงมาคือ วิธีของอาร์โนลด์ และวิธีที่ประยุกต์จากวิธีของเครสเซลและชมิค ตามลำดับ ส่วนวิธีประเพณีนิยม จะให้ค่าฟังก์ชัน สารสนเทศเฉลี่ยของแบบแบบสอบ และประสิทธิภาพสัมพัทธ์เฉลี่ยของแบบสอบต่ำกว่าทุกวิธี
2. วิธีของอาร์โนลด์มีความตรงเชิง โครงสร้างสูงกว่าวิธีอื่น ๆ รองลงมาคือ วิธีที่ประยุกต์ จากวิธีของคูมบ์และวิธีของคูมบ์ ตามลำดับ ส่วนวิธีประเพณีนิยมมีความตรงเชิง โครงสร้างต่ำกว่า ทุกวิธี
3. วิธีของคูมบ์ และวิธีที่ประยุกต์จากวิธีของคูมบ์มีค่าความตรงตามเกณฑ์สัมพัทธ์ สูงกว่าวิธีที่ประยุกต์จากวิธีของอาร์โนลด์
4. วิธีที่ประยุกต์จากวิธีของอาร์โนลด์ มีค่าความเที่ยงสูงกว่าวิธีอื่น ๆ รองลงมาคือ วิธีของอาร์โนลด์ และวิธีที่ประยุกต์จากวิธีของเครสเซลและชมิค ตามลำดับ
5. คุณภาพของวิธีการตอบและการตรวจให้คะแนน เมื่อพิจารณาตามเกณฑ์โดยรวม ทุกด้าน พบว่า วิธีที่ประยุกต์จากวิธีของคูมบ์ มีคุณภาพดีกว่าวิธีอื่น ๆ รองลงมาคือ วิธีของอาร์โนลด์ และวิธีที่ประยุกต์จากวิธีของเครสเซลและชมิค ตามลำดับ

ฉวีวรรณ บุญมั่ง (2540) ได้ศึกษาการเปรียบเทียบความสัมพันธ์ระหว่างคะแนน ความสามารถจากการวิเคราะห์ โดยทฤษฎีการตอบสนองข้อสอบกับคะแนนที่ได้จากการตอบ และตรวจให้คะแนน 4 วิธี ผลการศึกษาพบว่า

ผลการเปรียบเทียบค่าจากการวิเคราะห์โดยทฤษฎีการตอบสนองข้อสอบกับสัดส่วน การเดาที่ได้จากการตอบและตรวจให้คะแนน 4 วิธี พบว่า ค่าการเดาแตกต่างกับสัดส่วนการเดา ที่ได้จากวิธีของ 0-1 วิธีของคูมบ์ วิธีของกิบบอนส์และคณะ และวิธีของอนันต์ ศรีโสภา และสัดส่วนการเดาที่ได้จากวิธีของคูมบ์แตกต่างกับวิธีของ 0-1 และวิธีของกิบบอนส์ และคณะ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนสัดส่วนการเดาที่ได้จากการตอบและตรวจให้คะแนน คู่อื่น ๆ ไม่แตกต่างกัน

ผลการเปรียบเทียบค่าการเดาจากการวิเคราะห์โดยทฤษฎีการตอบสนองข้อสอบกับ สัดส่วนการเดาที่ได้จากการตอบและตรวจให้คะแนนวิธีของคูมบ์ วิธีของกิบบอนส์ และคณะ และวิธีของอนันต์ ศรีโสภา พบว่า ค่าการเดาแตกต่างกับสัดส่วนการเดาที่ได้จากวิธีของคูมบ์

และวิธีของกิบบอนส์ และคณะ สัดส่วนการเคาะที่ได้จากวิธีของคูมบ์แตกต่างกับวิธีของกิบบอนส์ และคณะ และวิธีของอนันต์ ศรีโสภา อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

พรทิพย์ ไชยโส (2534) ได้พัฒนาสูตรการให้คะแนนแบบทดสอบเลือกตอบสำหรับการให้คะแนนความรู้บางส่วนของผู้ตอบ การประยุกต์ใช้วิธีการของอาโนลด์และวิธีการของแฮมคาน จากการศึกษาพบว่า ไม่สามารถให้ข้อสรุปได้ชัดเจนถึงคุณภาพของสูตรการให้คะแนนที่พัฒนาขึ้น ถึงแม้ว่าคุณภาพเหนือกว่าสูตรการให้คะแนนอื่นทั้ง 3 เกณฑ์ คือ ความตรงเชิงทฤษฎี ความเที่ยงตรงตามเกณฑ์ และความเชื่อมั่นแบบสอดคล้องภายใน เนื่องจากสูตรการให้คะแนนแบบประเพณีนิยมมีความเที่ยงตรงเชิงทฤษฎี ความเที่ยงตรงตามเกณฑ์ และความเชื่อมั่นแบบสอดคล้องภายในไม่ยิ่งหย่อนกว่าสูตรการให้คะแนนอื่น ๆ มากนัก กลับมีความเที่ยงตรงเชิงทฤษฎีและความเที่ยงตรงตามเกณฑ์ไม่แตกต่างจากสูตรการให้คะแนนที่ประยุกต์จากวิธีของแฮมคาน

พินิจ อุไรรักษ์ (2533) ศึกษาผลของวิธีการให้คะแนนที่มีต่อคะแนนสอบ ค่าความเชื่อมั่น และค่าความเที่ยงตรงจากวิธีการให้คะแนน 4 วิธี คือ วิธีให้น้ำหนักคะแนนรายข้อเท่ากัน วิธีการให้น้ำหนักคะแนนรายข้อต่างกันตามระดับความมั่นใจในการตอบ วิธีการให้คะแนนรายข้อต่างกันตามค่าพารามิเตอร์ตามสูตรของลอร์ด และวิธีการให้คะแนนรายข้อต่างกันตามค่าความสามารถของผู้สอบ ผลการวิจัยพบว่า

1. ความสัมพันธ์ของคะแนนสอบระหว่างวิธีให้คะแนนทั้ง 4 วิธี มีค่าตั้งแต่ 0.941-0.983
2. ค่าความเชื่อมั่นของแบบทดสอบ พบว่า วิธีการให้น้ำหนักคะแนนรายข้อเท่ากัน วิธีการให้น้ำหนักคะแนนรายข้อต่างกันตามระดับความมั่นใจในการตอบ วิธีการให้น้ำหนักคะแนนรายข้อต่างกันตามค่าพารามิเตอร์ตามสูตรของลอร์ด และวิธีการให้คะแนนรายข้อต่างกันตามค่าความสามารถของผู้สอบ มีค่าเท่ากับ 0.075

3. วิธีการให้น้ำหนักคะแนนรายข้อต่างกันตามระดับความมั่นใจในการตอบวิธีการให้น้ำหนักคะแนนรายข้อต่างกันตามค่าพารามิเตอร์ตามสูตรของลอร์ด และวิธีการให้คะแนนรายข้อต่างกันตามค่าความสามารถของผู้สอบ สูงกว่าวิธีการให้น้ำหนักคะแนนรายข้อเท่ากัน

กาญจนา ศรีวัฒนพงษ์ (2520) ได้ศึกษาเปรียบเทียบวิธีการตอบและตรวจให้คะแนนแบบทดสอบปรนัยเลือกตอบที่มีลักษณะต่างกัน คือ แบบทดสอบที่มีตัวเลือกถูกต้องเดียว แบบตัวเลือกถูกต้องที่สุดเพียงตัวเดียว และแบบตัวเลือกถูกต้องหลายตัว ที่มีลักษณะต่างกัน โดยใช้วิธีการตรวจให้คะแนนต่างกัน 3 วิธี คือ วิธี 0-1 วิธีของคูมบ์ส วิธีของอนันต์ ผลการศึกษาพบว่า

1. เมื่อตอบและตรวจด้วยวิธี 0-1 และวิธีของคูมบ์ส ค่าความเชื่อมั่น ความเที่ยงตรง และค่าอำนาจจำแนกไม่แตกต่างกัน แต่เมื่อตอบด้วยวิธีอนันต์ ค่าความเชื่อมั่นของแบบทดสอบ

ที่มีตัวเลือกถูกเพียงตัวเดียวสูงกว่าตัวเลือกถูกที่สุดและตัวเลือกถูกหลายตัว ค่าความเที่ยงตรง และค่าอำนาจจำแนกไม่แตกต่างกัน

2. แบบทดสอบที่มีตัวเลือกถูกเพียงตัวเดียวและตัวเลือกถูกที่สุดเพียงตัวเดียว เมื่อตอบ และตรวจให้คะแนน โดยวิธีของอนันต์ ให้ค่าความเชื่อมั่นและค่าความเที่ยงตรงสูงกว่าวิธีการตรวจ ให้คะแนนของคูมบัส แต่ค่าความเที่ยงตรงและค่าอำนาจจำแนกไม่แตกต่างกัน ส่วนแบบทดสอบ ที่มีตัวเลือกถูกหลายตัวค่าความเที่ยงตรงและค่าอำนาจจำแนกไม่แตกต่างกัน

3. เมื่อตรวจให้คะแนนทั้ง 3 วิธี แบบทดสอบปรนัยเลือกตอบที่มีตัวเลือกถูกเพียงตัวเดียว และตัวเลือกถูกหลายตัวมีการเดาน้อยกว่าตัวเลือกถูกที่สุดเพียงตัวเดียว ส่วนแบบทดสอบที่มี ตัวเลือกถูกเพียงตัวเดียวมีค่าการเดาน้อยกว่าตัวเลือกถูกหลายตัว และเมื่อตรวจให้คะแนน โดยวิธีของ อนันต์และของคูมบัส ไม่พบว่าแตกต่างกัน

4. แบบทดสอบปรนัยเลือกตอบทั้ง 3 ลักษณะเมื่อตอบและตรวจให้คะแนนด้วยวิธีของ อนันต์ มีการเดาน้อยกว่าเมื่อตอบและตรวจให้คะแนนด้วย วิธี 0-1 และวิธีของคูมบัส แต่ไม่พบ ความแตกต่างในแบบทดสอบที่มีตัวเลือกถูกที่สุดเพียงตัวเดียว

ปรมินทร์ อริเดช (2547) ได้ศึกษาการเปรียบเทียบฟังก์ชันสารสนเทศในการใช้โมเดล โลจิสติก จีอาร์เอ็ม และจีพีซีเอ็มของมาตรวัดเจตคติแบบลิเคิร์ตและมาตรวัดแบบตัวเลือกบังคับตอบ ที่มีวิธีการให้คะแนนแบบสองค่าและแบบหลายค่า กลุ่มตัวอย่างเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 จังหวัดเชียงราย

ผลการวิจัยพบว่า ในมาตรวัดเจตคติแบบลิเคิร์ตการตรวจให้คะแนนแบบสองค่า ที่วิเคราะห์ด้วยโมเดล โลจิสติกให้ค่าเฉลี่ยฟังก์ชันสารสนเทศสูงกว่าการตรวจให้คะแนนแบบหลายค่า ที่วิเคราะห์ตาม GRM GCPM และ GPCM ที่มีการยุบรวมลำดับชั้นคะแนนใหม่ แต่ถ้าพิจารณา โคลังฟังก์ชันสารสนเทศจะพบว่า การวิเคราะห์ด้วย GPCM ที่มีการลำดับชั้นคะแนนใหม่ จะให้ค่าฟังก์ชันสูงสุดและพิจารณาจากมาตรวัดแบบลิเคิร์ตแบบซึ่งมีค่าตรงกลางและไม่มีค่า ตรงกลาง การตรวจให้คะแนนแบบไม่มีค่าตรงกลางจะให้ค่าเฉลี่ยฟังก์ชันสารสนเทศสูงกว่า การตรวจให้คะแนนแบบมีค่าตรงกลาง ในมาตรวัดความมีระเบียบวินัยแบบตัวเลือกบังคับตอบ การให้คะแนนแบบหลายค่าที่วิเคราะห์ด้วย GRM ให้ค่าเฉลี่ยฟังก์ชันสารสนเทศสูงสุด รองลงมา คือ GPCM ที่มีการยุบรวมลำดับชั้นคะแนนใหม่ GPCM ซึ่งสูงกว่าการให้คะแนนแบบสองค่า ที่วิเคราะห์ด้วยโมเดล โลจิสติก 3, 2 และ 1 พารามิเตอร์ ตามลำดับ เมื่อพิจารณามาตรวัด ความมีระเบียบวินัยแบบตัวเลือกบังคับตอบที่มีการให้คะแนนแบบสองค่า ที่มีวิธีกำหนดน้ำหนัก คะแนนแตกต่างกัน 3 แบบ พบว่า การวิเคราะห์ด้วยโมเดล โลจิสติก ด้วยวิธีการให้คะแนนแบบที่ 3 ให้คะแนนเฉลี่ยฟังก์ชันสารสนเทศสูงสุดรองลงมา คือ 2 และ 1 พารามิเตอร์

เอมอร์ จังศิริพรปกรณ์ (2545) ได้ศึกษาการเปรียบเทียบคุณภาพของแบบสอบเลือกตอบ เมื่อตรวจด้วยวิธีการให้คะแนนความรู้บางส่วน กับวิธีประเพณีนิยมกลุ่มตัวอย่างที่ใช้ในการศึกษา เป็นนิติตถณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยชั้นปีที่ 3 ที่ลงทะเบียนเรียนวิชาการประเมินผลการเรียนการสอน ในภาคต้น ปีการศึกษา 2544 จำนวน 297 คน ผลการวิจัยพบว่า

1. ผลการเปรียบเทียบความตรงตามสภาพ ระหว่างวิธีการตรวจให้คะแนนที่แตกต่างกัน จากการศึกษาความตรงตามสภาพตามวิธีประยุกต์การให้คะแนนของคูมบ์ วิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท และวิธีประเพณีนิยม แล้วนำมาเปรียบเทียบความแตกต่างโดยใช้คะแนน ฟิชเซอร์ซี เปรียบเทียบในภาพรวมโดยใช้สถิติทดสอบไคว์สแควร์ พบว่า ค่าความตรงตามสภาพ ไม่ขึ้นอยู่กับวิธีการตรวจให้คะแนนในแต่ละวิธี และเมื่อนำมาเปรียบเทียบความแตกต่างกัน เป็นรายคู่ ด้วยสถิติทดสอบซี (Z) พบว่า ไม่มีความแตกต่างกันของความตรงสภาพในแต่ละคู่ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

2. ผลการเปรียบเทียบความยากและอำนาจจำแนกระหว่างวิธีการตรวจให้คะแนน ที่แตกต่างกัน ผลการเปรียบเทียบความแตกต่างของค่าเฉลี่ยของค่าอำนาจจำแนกและค่าเฉลี่ยของ ค่าความยาก ด้วยสถิติ F พบว่า ค่าเฉลี่ยของค่าอำนาจจำแนกตามวิธีการตรวจให้คะแนน มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ($F = 85.369, p\text{-value} < .000$) และเมื่อนำมาเปรียบเทียบความแตกต่างรายคู่ด้วยวิธีของ Scheffe พบว่า วิธีประยุกต์การให้คะแนน ของคูมบ์ มีค่าอำนาจจำแนกเฉลี่ยสูงกว่าวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ในขณะที่เดียวกัน วิธีประเพณีนิยมมีค่าอำนาจจำแนกเฉลี่ย สูงกว่าวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิทอย่างมีนัยสำคัญทางสถิติ เช่นเดียวกัน ส่วนค่าเฉลี่ยของความยากตามวิธีการตรวจให้คะแนนมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ที่ระดับ .05 ($F = 20.592, p\text{-value} < .000$) และเมื่อนำมาเปรียบเทียบความแตกต่างรายคู่ด้วยวิธีของ Scheffe พบว่า วิธีประยุกต์การให้คะแนนของคูมบ์ มีค่าความยากเฉลี่ยต่ำกว่าวิธีประยุกต์ การให้คะแนนของเดรสเซลและสมิท และวิธีประเพณีนิยมอย่างมีนัยสำคัญทางสถิติ ในขณะที่เดียวกัน วิธีประเพณีนิยมมีค่าความยากเฉลี่ยไม่แตกต่างจากวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท อย่างมีนัยสำคัญทางสถิติ

3. ผลการเปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบสอบ และอัตราส่วนสารสนเทศ ระหว่างวิธีการตรวจให้คะแนนที่แตกต่างกันวิธีประยุกต์การให้คะแนน ของคูมบ์มีค่าฟังก์ชันสารสนเทศของข้อสอบ โดยเฉลี่ยสูงกว่าวิธีประยุกต์การให้คะแนนของ เดรสเซลและสมิท และวิธีประเพณีนิยมอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนวิธีประยุกต์ การให้คะแนนของเดรสเซลและสมิท มีค่าฟังก์ชันสารสนเทศของข้อสอบ โดยเฉลี่ยสูงกว่า

วิธีประเพณีนิยมอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เมื่อเปรียบเทียบค่าฟังก์ชันสารสนเทศของข้อสอบ จำแนกตามค่าความสามารถของผู้สอบตามวิธีการตรวจให้คะแนน พบว่า วิธีประยุกต์การให้คะแนนของคும்บ์ ให้ค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงที่สุดในระดับความสามารถ -1 และมีค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยเท่ากับ 0.673 ซึ่งสูงกว่าวิธีอื่น ๆ ในขณะที่วิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท ให้ค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงที่สุดในระดับความสามารถ 0 มีค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยเท่ากับ 0.395 ส่วนวิธีประเพณีนิยมให้ค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงที่สุดในระดับความสามารถ 0 มีค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยเท่ากับ 0.187

วิธีประยุกต์การให้คะแนนของคும்บ์ ให้ค่าฟังก์ชันสารสนเทศของแบบสอบสูงกว่าวิธีอื่น ๆ รองลงมาคือ วิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท ส่วนวิธีประเพณีนิยมให้ค่าฟังก์ชันสารสนเทศของแบบสอบต่ำที่สุด ส่วนอัตราส่วนสารสนเทศเฉลี่ยระหว่างวิธีประยุกต์การให้คะแนนของคும்บ์กับวิธีประเพณีนิยม มีค่าสูงสุด รองลงมาคือ อัตราส่วนสารสนเทศเฉลี่ยระหว่างวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิทกับวิธีประเพณีนิยม ซึ่งแสดงให้เห็นว่าวิธีประยุกต์การให้คะแนนของคும்บ์มีประสิทธิภาพสัมพัทธ์สูงที่สุด รองลงมาคือ วิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท ส่วนวิธีประเพณีนิยมมีประสิทธิภาพสัมพัทธ์ต่ำที่สุด

สุรัชย์ มีชาญ (2539) เปรียบเทียบความยากประจำชั้นและประสิทธิภาพในการประมาณค่าเจตคติของมาตรวัดเจตคติแบบลิเคิร์ตที่มีรูปแบบของการตอบ จำนวนของลำดับขั้นและทิศทางของการเรียงลำดับแตกต่างกันด้วยพาเซี่ยลคริตโมเดลของมาสเตอร์ กลุ่มตัวอย่างที่ใช้เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2538 ของโรงเรียนสังกัดกรมสามัญศึกษาจังหวัดลพบุรีและสิงห์บุรี จำนวน 3,226 คน

ผลการวิจัยพบว่า มาตรวัดเจตคติมีจำนวนของลำดับขั้นเพียง 4 ระดับ มีแนวโน้มว่าหากมาตรวัดนั้นใช้รูปแบบของการตอบและทิศทางของการเรียงลำดับขั้นแตกต่างกัน ค่าเฉลี่ยความยากประจำชั้นที่ 1 และชั้นที่ 3 ของมาตรวัด จะมีความแตกต่างกัน แต่ค่าเฉลี่ยความยากประจำชั้นที่ 2 จะไม่แตกต่างกัน และเมื่อมาตรวัดเหล่านั้นมีจำนวนของลำดับขั้นเพิ่มขึ้นเป็น 5 ระดับ ค่าเฉลี่ยความยากประจำชั้นของการเปรียบเทียบค่าสัดส่วนของผู้ตอบที่มีความเหมาะสมกับโมเดลระหว่างกลุ่มผู้ตอบมาตรวัดเจตคติต่างฉบับกัน พบว่า ค่าสัดส่วนของผู้ตอบที่เหมาะสมกับโมเดลสูงสุด คือ ฉบับที่ 3 รองลงมา คือฉบับที่ 7 ฉบับที่ 4 ฉบับที่ 8 ฉบับที่ 1 ฉบับที่ 2 ฉบับที่ 6 และฉบับที่ 5 ตามลำดับขั้นในมาตรวัดเหล่านั้นจะไม่แตกต่างกัน ในการศึกษาความสัมพันธ์ของค่าความยากประจำชั้นของมาตรวัดเจตคติแต่ละฉบับ พบว่า จำนวนลำดับขั้น 4 ระดับ มาตรวัดฉบับที่ 1, 3 และ 4 มีค่าความยากประจำชั้นที่ 1 มีความสัมพันธ์กันทางบวกทุกคู่ ในกรณีมาตรวัด

มีจำนวนของลำดับชั้น 5 ระดับ ฉบับที่ 3 กับฉบับที่ 7 และฉบับที่ 4 กับฉบับที่ 7 มีค่าความยากประจำชั้นที่ 1 ไม่สัมพันธ์กัน การเปรียบเทียบค่าฟังก์ชันอินฟอร์เมชันของมาตรวัด แต่ละฉบับพบว่า มาตรวัดที่มีจำนวนลำดับชั้น 5 ระดับ มีค่าอินฟอร์เมชันสูงกว่ามาตรวัดที่มีจำนวนลำดับชั้น 4 ระดับ ทุกคู่ของการเปรียบเทียบตลอดช่วงเจตคติ มาตรวัดที่มีจำนวนลำดับชั้นเท่ากันมีค่าเฉลี่ยของ RMSE_nใกล้เคียงกัน โดยมาตรวัดที่มีจำนวนลำดับ 4 ระดับ มีค่าเฉลี่ย RMSE_n ไม่แตกต่างกันทุกคู่ ยกเว้นฉบับที่ 1 และฉบับที่ 5 ขณะที่มาตรวัดที่มีจำนวนลำดับชั้น 5 ระดับ มีค่าเฉลี่ย RMSE_n ไม่แตกต่างกันทุกคู่ และมีค่าสูงกว่ามาตรวัดที่มีจำนวนลำดับชั้น 4 ระดับ

ธนวัฒน์ แสนสุข (2538) ได้ศึกษาการใช้ จีอาร์เอ็ม จีพีซีเอ็ม และ โมเดลโลจิสติก ในการเปรียบเทียบฟังก์ชันสารสนเทศของแบบวัดที่มีวิธีการให้คะแนนต่างกัน โดยที่แบบวัดมีการตรวจให้คะแนนแบบทวิภาค (0, 1) และแบบพหุภาค (1, 2, 3, 4) ทำการตรวจสอบความสมบูรณ์ของข้อมูลโดยวิเคราะห์ด้วยโปรแกรม BIGSTEPS และวิเคราะห์หาค่าฟังก์ชันสารสนเทศด้วยโปรแกรม MULTILOG และ PARSCALE ผลการวิจัย พบว่า การตรวจให้คะแนนแบบพหุภาค เมื่อวิเคราะห์ตาม GRM ให้ค่าฟังก์ชันสารสนเทศสูงกว่าการให้คะแนนแบบทวิภาค สำหรับการตรวจให้คะแนนแลตพหุภาควิเคราะห์ตาม GPCM และแบบทวิภาคยังไม่สามารถสรุปได้ว่าวิธีใดให้ค่าฟังก์ชันสารสนเทศสูงกว่ากัน

องอาจ นัยวัฒน์ (2535) ได้ศึกษาประสิทธิภาพของผลการประมาณค่าพารามิเตอร์ของ พาเชียด เครดิต โมเดล (CREDIT) ระหว่างวิธีพรีออกซ์ (PROX) และวิธียูคอน (UCON) ในแบบทดสอบวัดเจตคติแบบลิเคิร์ตสเกล จำนวน 50 รายการ กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ปีการศึกษา 2534 เขตพระ โขนง กรุงเทพมหานคร จำนวน 500 คน ผลการวิจัยพบว่า ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยที่เป็นผลมาจากการประมาณค่าเจตคติของผู้ตอบด้วยวิธียูคอนมีค่าต่ำกว่าวิธีพรีออกซ์ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยที่เป็นผลมาจากการประมาณค่าความยากประจำชั้นของข้อความระหว่างวิธีทั้งสองมีค่าไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ค่าความเที่ยงตรงเชิงสภาพของการประมาณค่าพารามิเตอร์เจตคติของผู้ตอบที่ประมาณค่าด้วยวิธีพรีออกซ์มีค่าสูงกว่าวิธียูคอนอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบที่ประมาณค่าพารามิเตอร์ด้วยวิธีพรีออกซ์สูงกว่าวิธียูคอนในช่วงของเจตคติดระดับปานกลาง ส่วนในช่วงเจตคติดระดับต่ำและสูง พบว่า วิธียูคอนสูงกว่า ด้านความสัมพันธ์ระหว่างเจตคติที่เป็นผลมาจากการประมาณค่าพารามิเตอร์ด้วยวิธีพรีออกซ์และวิธียูคอน มีความสัมพันธ์ในทางบวกซึ่งกันและกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ทำนองเดียวกับความยากประจำชั้นของข้อความที่ประมาณค่าด้วยวิธีทั้งสองก็ พบว่า ความสัมพันธ์ในทางบวกระดับสูงมาก

อนันต์ ศรี โสกา (2532) ได้ศึกษาเปรียบเทียบประสิทธิภาพของแบบทดสอบที่ใช้วิธีการตอบและการตรวจให้คะแนนแบบใหม่ แบบของคูมบ์ส และแบบ 0-1 โดยใช้แบบทดสอบวิชาวิจัยเบื้องต้น กลุ่มตัวอย่างที่ใช้เป็นนักศึกษาวิทยาลัยพยาบาลกรุงเทพ ชั้นปีที่ 4 จำนวน 144 คน แบ่งออกเป็น 3 กลุ่ม กลุ่มละ 48 คน แต่ละกลุ่มให้ตอบและตรวจให้คะแนนด้วยวิธีของอนันต์แบบของคูมบ์ส และแบบ 0-1 ตามลำดับ ผลการศึกษาพบว่า วิธีการตอบและตรวจให้คะแนนแบบใหม่มีค่าการเดาน้อยกว่าวิธี 0-1 อย่างมีนัยสำคัญ เมื่อเปรียบเทียบกับวิธีของคูมบ์ส จำนวนข้อถูกและข้อผิดของวิธีการตรวจให้คะแนนแบบใหม่แตกต่างจากวิธี 0-1 อย่างมีนัยสำคัญ แต่ไม่แตกต่างจากวิธีของคูมบ์ส การค้นพบนี้สรุปได้ว่า คะแนนการเดา คะแนนข้อถูก คะแนนข้อผิด และคะแนนถูกบางส่วนของวิธีการใหม่ไม่แตกต่างจากวิธีของคูมบ์ส แต่แตกต่างจากวิธีการตรวจให้คะแนนแบบ 0-1 อย่างมีนัยสำคัญ ค่าความเชื่อมั่นของการตรวจให้คะแนนแบบใหม่สูงกว่าวิธีของคูมบ์สอย่างมีนัยสำคัญ ส่วนค่าความเที่ยงตรงของแบบทดสอบที่มีตรวจให้คะแนนแบบใหม่วิธีของคูมบ์ส ไม่แตกต่างกันอย่างมีนัยสำคัญ วิธีการตรวจให้คะแนนแบบใหม่สามารถปรับปรุงให้การสอบ โดยใช้แบบทดสอบเลือกตอบมีประสิทธิภาพสูงขึ้นและมีคุณค่าแก่การศึกษา

ไซมอน และคณะ (Simon et al., 1997) ได้เปรียบเทียบวิธีการตอบสนองข้อสอบเพื่อให้คะแนนความรู้บางส่วน 6 วิธี กับวิธีเลือกตอบแบบประเพณีนิยม เก็บข้อมูลกับนักเรียน 2,989 คน ที่สอบคัดเลือกเข้ามหาวิทยาลัยในอิสราเอล ระหว่างเดือนเมษายน 1985 กรกฎาคม 1986 โดยใช้แบบทดสอบ 4 วิชา คือ General knowledge 30 ข้อ วิชา General reasoning 22 ข้อ วิชา Figural reasoning 18 ข้อ และวิชา Mathematical reasoning 21 ข้อ แต่ละวิชามีวิธีการตอบ 7 วิธี ได้แก่

1. วิธีแก้การเดา (Correction for Guessing: CG)
2. วิธีการตัดตัวลวง (Elimination Scoring: ET)
3. วิธีการทดสอบความน่าจะเป็น (Probability Testing: PT)
4. วิธีการแสดงความมั่นใจ (Confidence Marking: CM)
5. วิธีการจัดอันดับอย่างสมบูรณ์ (Complete Ordering: CO)
6. วิธีการจัดอันดับเป็นบางส่วน (Partial Ordering: PO) และ
7. วิธีประเพณีนิยม

อัลเบิร์ต (Albert, 1970) ได้ศึกษาวิธีการตรวจให้คะแนนความรู้บางส่วนวิชาคณิตศาสตร์ โดยใช้โจทย์ปัญหาคณิตศาสตร์ จำนวน 25 ข้อ กับกลุ่มตัวอย่างที่เป็นนักเรียนเกรด 11 จำนวน 2 กลุ่ม กลุ่มละ 146 คน มีวิธีการตรวจให้คะแนน 2 ลักษณะ คือ 1) แบบทดสอบเลือกตอบที่มีตัวเลือกถูกเพียงตัวเดียว มีการให้คะแนน 3 วิธี คือ วิธี 0-1 วิธีกำหนดน้ำหนักตัวเลือก โดยดูจาก

โอกาสผิดพลาดมากหรือน้อย และวิธีกำหนดน้ำหนักตัวเลือกโดยครูเป็นผู้กำหนด 2) แบบขยายคำตอบ เป็นแบบที่มีการแก้ปัญหาหลายขั้นตอน มีการตรวจให้คะแนน 2 วิธี คือ วิธีการให้คะแนนขั้นละ 1 คะแนน ในการตอบที่ถูกต้อง และวิธีให้คะแนนขั้นละ 1 คะแนน ถ้านักเรียนเริ่มต้นจากข้อมูลที่ถูกต้องถึงคำตอบที่ถูกต้อง ผลการศึกษาพบว่า

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

หาความคลาดเคลื่อนที่เกิดจากการเดาได้

จากการศึกษาจะเห็นว่า แบบขยายคำตอบซึ่งเป็นตรวจให้คะแนนความรู้บางส่วน ของนักเรียนในการตอบจะทำให้ค่าอำนาจจำแนกและค่าความเชื่อมั่นสูง เปอร์เซนต์การเดาลดลง และยังให้ค่าความเที่ยงตรงตามสภาพสูงด้วย

เดรสเซล และชมิค (Dressel & Schmid, 1953) ได้ศึกษาวิธีการตอบและวิธีการตรวจให้คะแนนที่ต่างกัน เพื่อปรับปรุงอำนาจจำแนกของแบบทดสอบเลือกตอบ โดยมีวิธีการตอบ 4 วิธี คือ

1. แบบเลือกตอบเสรี เป็นแบบทดสอบเลือกตอบที่มี 5 ตัวเลือก ผู้สอบจะเลือกกี่ตัวเลือกก็ได้
2. แบบบอกระดับความมั่นใจในการตอบ เป็นแบบที่ผู้เข้าสอบจะต้องแสดงระดับความมั่นใจในการตอบ ซึ่งมี 4 ระดับ แต่คำถามแต่ละข้อให้ตอบเพียงคำตอบเดียว
3. แบบคำตอบถูกมากกว่าหนึ่ง แบบทดสอบชนิดนี้จะมีคำตอบถูกอยู่หลายคำตอบ ผู้เข้าสอบจะต้องทำเครื่องหมายทุกข้อที่ถูก
4. แบบมีคำตอบถูกสองคำตอบ ผู้เข้าสอบจะเลือกตอบได้เพียงสองคำตอบเท่านั้น

ผลการศึกษาพบว่า การข้อสอบของนักเรียนเก่ง (พิจารณาจากคะแนนสอบ) จะแตกต่างกับนักเรียนปานกลางและอ่อน อย่างมีนัยสำคัญทางสถิติ เมื่อใช้แบบทดสอบเลือกตอบชนิดตอบเสรี คือ กลุ่มเก่งจะทำเครื่องหมายน้อยกว่ากลุ่มอ่อน ทั้งในข้อสอบที่มีระดับความยากสูง ปานกลาง และต่ำ ในทางตรงกันข้าม แบบทดสอบชนิดทดสอบความมั่นใจไม่ได้จำแนกนักเรียนเก่ง ปานกลาง อ่อน ได้ดีกว่าวิธีแรก เพราะเมื่อพิจารณาการตอบข้อสอบแล้ว พบว่า ผู้เข้าสอบตอบข้อสอบที่มีความยากปานกลาง และยากมากด้วยระดับความมั่นใจเดียวกัน

เดรสเซล และชมิค ได้สรุปว่า องค์ประกอบของความมั่นใจที่วัด โดยการให้คำตอบแบบเสรีนั้น แตกต่างกับที่วัด โดยผู้เข้าสอบบอกระดับความมั่นใจ

สมิท (Smith, 1987) ได้ศึกษาเปรียบเทียบความตรงสภาพและความเชื่อมั่น โดยใช้สัมประสิทธิ์แอลฟาระหว่างแบบทดสอบที่มีการให้คะแนนแบบทวิภาค (Dichotomous, 0-1) และแบบทดสอบที่มีการให้คะแนนแบบพหุวิภาค (Polytomous) โดยเก็บข้อมูลจากนักศึกษาที่มาสอบความถนัดทางภาษา จำนวน 1,399 คน ให้ทำแบบทดสอบทางภาษาที่ใช้โดยทั่วไป เพื่อนำไปเป็นเกณฑ์ในการหาความตรงตามสภาพกับแบบทดสอบภาษาที่ใช้ในการศึกษาครั้งนี้ มี 2 ชุด ชุดแรกมีการให้คะแนนแบบทวิภาค (Dichotomous) และชุดที่ 2 ให้คะแนนแบบพหุวิภาค (Polytomous) โดยที่แต่ละตัวเลือกรับคะแนนแต่ละข้อไม่เท่ากัน โดยมีหลักการให้คะแนนความรู้บางส่วนจากทฤษฎีโครงสร้างความรู้ และมีการแบ่งกลุ่มตัวอย่างออกเป็นสองกลุ่ม กลุ่มที่ 1 จำนวน 400 คน ทดสอบตามปกติ กลุ่มที่ 2 จำนวน 999 คน ทดสอบความตรงข้ามกลุ่ม โดยแบ่งเป็นกลุ่มที่มีความสามารถสูงและกลุ่มที่มีความสามารถต่ำ ผลการวิจัยพบว่า การให้คะแนนแบบพหุวิภาค (Polytomous) ค่าความตรงสภาพและความเชื่อมั่นสูงกว่าการให้คะแนนแบบทวิภาค (Dichotomous) อย่างมีนัยสำคัญทางสถิติทั้งกลุ่มที่มีความสามารถสูงและกลุ่มที่มีความสามารถต่ำ

บทที่ 3

วิธีการดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการวิจัยเชิงทดลอง เพื่อศึกษาประสิทธิภาพการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ด้านจำนวนข้อสอบและฟังก์ชันสารสนเทศของชุดข้อสอบที่เกิดจากผลของการให้คะแนนที่ต่างกัน เมื่อพิจารณาตามองค์ประกอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ได้แก่ การคัดเลือกข้อสอบ การประมาณค่าความสามารถของผู้สอบ การยุติการทดสอบ และความสามารถของผู้สอบ ซึ่งผู้วิจัยแบ่งการดำเนินการวิจัยออกเป็น 2 ระยะ ได้แก่ ระยะแรกดำเนินการสร้างคลังข้อสอบวิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษาตอนปลาย ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551

ระยะที่สองเป็นการศึกษาคุณภาพของการวัดและประสิทธิภาพของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ในด้านจำนวนข้อสอบ และค่าฟังก์ชันสารสนเทศของชุดข้อสอบที่เกิดจากผลของการให้คะแนนตามตัวแปรอิสระ 5 ตัว ได้แก่ การคัดเลือกข้อสอบ การประมาณค่าความสามารถผู้สอบ การยุติการทดสอบ และความสามารถของผู้สอบ โดยมีวิธีการดำเนินการวิจัยในแต่ละระยะ ดังนี้

การดำเนินการวิจัย ระยะแรก

ประชากรและกลุ่มตัวอย่าง

ประชากร เป็นนักเรียนของ โรงเรียนมัธยมศึกษา จังหวัดบุรีรัมย์ ระดับชั้นมัธยมศึกษาปีที่ 6 จำนวน 9,789 คน (สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 32, 2557)

กลุ่มตัวอย่างที่ใช้ในการสร้างคลังข้อสอบเป็นนักเรียนของ โรงเรียนมัธยมศึกษา จังหวัดบุรีรัมย์ ระดับชั้นมัธยมศึกษาปีที่ 6 ได้มาด้วยการสุ่มแบบหลายขั้นตอน (Multi stage random sampling) จำนวน 4,096 คน โดยมีขั้นตอนในการสุ่มดังนี้

ขั้นตอนแรก สุ่มโรงเรียนมัธยมศึกษา จำนวน 66 โรงเรียน ด้วยวิธีการสุ่มแบบแบ่งชั้น (Stratified random sampling) โดยใช้ขนาดของโรงเรียนเป็นชั้นของการสุ่ม ได้แก่ ขนาดใหญ่ (นักเรียนตั้งแต่ 1,500 คนขึ้นไป) ขนาดกลาง (นักเรียนตั้งแต่ 500 คนไม่เกิน 1,500 คน) และขนาดเล็ก (นักเรียนไม่เกิน 500 คน) สุ่มโรงเรียนได้ทั้งหมด 30 โรงเรียน แบ่งเป็นโรงเรียนขนาดใหญ่ 4 โรงเรียน ขนาดกลาง 12 โรงเรียน และขนาดเล็ก 14 โรงเรียน

ขั้นตอนที่สอง สุ่มห้องเรียนจากโรงเรียนที่สุ่มได้ในขั้นตอนแรก โดยใช้วิธีการสุ่มอย่างง่าย (Simple random sampling) ซึ่งสุ่มห้องเรียนจากโรงเรียนขนาดใหญ่ จำนวน 4 โรงเรียน ขนาดกลาง จำนวน 25 โรงเรียน และขนาดเล็ก จำนวน 28 โรงเรียน โรงเรียนละ 5 ห้อง 3 ห้อง และ 1 ห้อง ตามลำดับ ได้ห้องเรียนทั้งหมด 513 ห้อง และผู้วิจัยใช้นักเรียนทุกคนที่สุ่มได้ในแต่ละห้องเป็นกลุ่มตัวอย่าง

วิธีดำเนินการวิจัย

การพัฒนาข้อสอบ โดยการสร้างข้อสอบเลือกตอบและให้ผู้เชี่ยวชาญพิจารณาความถูกต้องเหมาะสมของข้อสอบ จากนั้นนำไปทดสอบกับนักเรียนกลุ่มตัวอย่างดำเนินการดังนี้

1. กำหนดกรอบแนวคิดของการตรวจให้คะแนนสำหรับรูปแบบข้อสอบเลือกตอบแบบถูกผิด โดยดำเนินการดังนี้
 - 1.1 เลือกตัวแปรและเนื้อหาที่เกี่ยวข้องกับผลการวัด
 - 1.2 สร้างเครื่องมือเพื่อใช้ในการวัด
 - 1.3 กำหนดวิธีการตอบและคำชี้แจงในการตอบข้อสอบ
 - 1.4 เลือกประชากรและกำหนดขนาดกลุ่มตัวอย่าง
 - 1.5 ดำเนินการสอบกับนักเรียนกลุ่มตัวอย่าง
2. การกำหนดการให้คะแนนรายข้อ 3 แบบ ได้แก่ แบบ Multiple-Response Method (MR) แบบ Multiple True-False Method (MTF) และแบบที่ผู้วิจัยเสนอคือ Omit Multiple True-False Method (OMTF)
3. เก็บรวบรวมข้อมูลกับกลุ่มตัวอย่าง โดยนำข้อสอบไปใช้ร่วมกับโครงการเตรียมความพร้อมสำหรับการทดสอบระดับชาติขั้นพื้นฐาน (Pre-O-NET) ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2557 โดยสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 32

เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้สำหรับสร้างคลังข้อสอบ คือ ข้อสอบวัดผลสัมฤทธิ์ทางการเรียน วิชาคณิตศาสตร์ ตามมาตรฐานการเรียนรู้หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ระดับชั้นมัธยมศึกษาตอนปลาย เป็นชุดข้อสอบเลือกตอบที่ผู้วิจัยสร้างขึ้น จำนวน 190 ข้อ จัดพิมพ์เป็นชุดข้อสอบ จำนวน 9 ฉบับ ฉบับละ 30 ข้อ (แต่ละฉบับมีข้อสอบต่างจากฉบับอื่นอยู่ 20 ข้อ และมีข้อสอบที่เหมือนกันอยู่ 10 ข้อในทุกฉบับซึ่งใช้เป็นข้อสอบร่วม) หากค่าคุณภาพตามทฤษฎีการทดสอบแบบดั้งเดิม ได้แก่ ค่าความยาก มีค่าระหว่าง 0.20-0.80 และค่าอำนาจจำแนก ตั้งแต่ 0.20 ขึ้นไป พิจารณาข้อสอบที่มีคุณภาพเก็บไว้สำหรับใช้ในการวิจัยระยะที่สองต่อไป

2. ขั้นตอนที่ 2 พัฒนาโปรแกรมการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ เป็นการออกแบบโปรแกรมคอมพิวเตอร์ และเขียนโปรแกรมด้วยภาษา ASP.NET โดยนำข้อสอบจากคลังข้อสอบในขั้นตอนที่ 1 มาเป็นฐานข้อมูลสำหรับใช้ในการพัฒนาโปรแกรม จากนั้นให้ผู้เชี่ยวชาญพิจารณาตรวจสอบความถูกต้องเหมาะสมของโปรแกรม ปรับปรุงตามคำแนะนำของผู้เชี่ยวชาญแล้วทำการทดลองใช้โปรแกรมการทดสอบ โดยนำโปรแกรม ไปให้นักเรียนทดลองใช้งานปรับปรุงตามคำแนะนำของนักเรียนและจัดทำเป็นชุดโปรแกรมสำหรับการทดสอบต่อไป

ตารางที่ 6 เงื่อนไขของโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

โปรแกรมย่อย	การคัดเลือกข้อสอบ	การประมาณค่า
1	[Match b_i to $\hat{\theta}$]	Conditional maximum likelihood (MLE)
2	[Match b_i to $\hat{\theta}$]	Bayesian Updating (Bay)
3	[Match Q_{max} to $\hat{\theta}$]	Conditional maximum likelihood (MLE)
4	[Match Q_{max} to $\hat{\theta}$]	Bayesian Updating (Bay)
5	I_{max}	Conditional maximum likelihood (MLE)
6	I_{max}	Bayesian Updating (Bay)

3. ขั้นตอนที่ 3 ดำเนินการทดสอบกับนักเรียนกลุ่มตัวอย่างโดยใช้โปรแกรมคอมพิวเตอร์เพื่อศึกษาประสิทธิภาพการทดสอบ ในด้านจำนวนข้อสอบและค่าฟังก์ชันสารสนเทศของชุดข้อสอบแล้วเปรียบเทียบตามแนวคิดการทดสอบแบบปรับเหมาะ ดังนี้

- เกณฑ์การคัดเลือกข้อสอบ (Item selection rule)
- การประมาณค่าความสามารถของผู้สอบ (Ability estimation procedure)
- เกณฑ์ยุติการทดสอบ (Termination criteria)
- ความสามารถของผู้สอบ (Examinee's ability)

ประชากรและกลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการหาประสิทธิภาพของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ได้แก่ โรงเรียนบุรีรัมย์พิทยาคม จังหวัดบุรีรัมย์ ซึ่งเป็นโรงเรียนมัธยมศึกษาขนาดใหญ่ ประจำจังหวัดบุรีรัมย์ ที่มีการสอบคัดเลือกนักเรียนเข้ามาเรียนจากทั้งจังหวัดบุรีรัมย์ เนื่องจากผู้วิจัยต้องการให้คะแนนที่ได้จากการสอบครั้งนี้ ได้จากกลุ่มนักเรียนที่ได้รับการสอนด้วยวิธีที่ใกล้เคียงกัน

หลักสูตรเดียวกัน และจัดการสอบได้พร้อมกัน ดังนั้นผู้วิจัยจึงได้เลือกนักเรียนมาวิจัยเพียง 1 โรงเรียนเท่านั้น คือ นักเรียนระดับชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2557 จำนวน 16 ห้องเรียน จำนวน 650 คน สุ่มนักเรียนโดยแบ่งตามความสามารถทางคณิตศาสตร์ได้แก่ ความสามารถสูง ปานกลาง และต่ำ ผู้วิจัยต้องการกลุ่มตัวอย่าง จำนวน 540 คน เพื่อใช้เป็นกลุ่มผู้เข้าสอบ โดยใช้โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ จำนวน 6 โปรแกรม ๆ ละ 90 คน ซึ่งได้จากการจัดหมู่ตัวแปรอิสระ 2 ตัว ได้แก่ การคัดเลือกข้อสอบจากคลัง 3 เกณฑ์ (ค่าความยากง่าย ค่าความสามารถที่ทำให้ข้อสอบมีสารสนเทศสูงสุด ค่าสารสนเทศของข้อสอบสูงสุด) การประมาณค่าความสามารถของผู้สอบ 2 วิธี (ความเป็นไปได้สูงสุดและเบส) ส่วนการยุติการทดสอบ 3 เกณฑ์ (ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า เท่ากับหรือน้อยกว่า 0.30, 0.40, 0.50) และความสามารถของผู้สอบ 3 ระดับ (สูง กลาง ต่ำ) นั้นไม่ได้นำมารวมในการจัดหมู่ เพราะโปรแกรมสามารถใช้ร่วมกันได้

รายละเอียดการดำเนินการวิจัย ผู้วิจัยนำเสนอเป็นแผนภาพได้ดังนี้



ภาพที่ 14 ขั้นตอนดำเนินการวิจัย

เครื่องมือที่ใช้ในการวิจัย

แบบทดสอบวิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษาตอนปลาย ตามหลักสูตรแกนกลาง การศึกษาขั้นพื้นฐาน พุทธศักราช 2551 เป็นข้อสอบที่หาค่าคุณภาพจากระยะแรก และคัดเลือกไว้ รวมจำนวน 150 ข้อ นำมาจัดพิมพ์ข้อสอบเป็น 9 ฉบับ ฉบับละ 30 ข้อ (แต่ละฉบับมีจำนวนข้อสอบ ต่างจากฉบับอื่นอยู่ 15 ข้อ และมีข้อสอบที่เหมือนกันอยู่ 15 ข้อทุกฉบับ ซึ่งใช้เป็นข้อสอบร่วม) ผู้วิจัยสร้างข้อสอบดังกล่าวให้เป็นข้อสอบคู่ขนาน โดยสร้างจากมาตรฐานและตัวชี้วัด วิชาคณิตศาสตร์พื้นฐาน จำนวน 30 ตัวชี้วัด สาขาที่ต้องมีการสร้างแบบทดสอบหลายฉบับ ก็เนื่องจากผู้วิจัยต้องการให้มีข้อสอบไว้ในคลังข้อสอบจำนวนมาก มีค่าความยากง่ายของข้อสอบ เหมาะกับผู้สอบที่มีความสามารถทุกระดับตั้งแต่ความสามารถต่ำไปจนถึงความสามารถสูง เนื่องจากกลุ่มตัวอย่างที่ใช้ในการสร้างคลังข้อสอบเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 6 จึงมีศักยภาพ ในการตอบข้อสอบจำนวนมากข้อ ในแต่ละครั้ง ผู้วิจัยจึงได้แยกข้อสอบออกเป็นฉบับย่อย ๆ 9 ฉบับ ซึ่งเมื่อนำค่าพารามิเตอร์ของข้อสอบแต่ละฉบับมาปรับให้อยู่บนสเกลเดียวกัน ข้อสอบทั้งหมด เป็นแบบเลือกตอบ 4 ตัวเลือก มีความเป็นมิตติเดียว ข้อสอบที่มีคุณภาพจะถูกคัดเลือกเก็บไว้ ในลักษณะของคลังข้อสอบ เพื่อใช้สำหรับทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ ด้วยคอมพิวเตอร์ ซึ่งสัดส่วนของจำนวนข้อสอบ จำแนกเป็นมาตรฐานและตัวชี้วัด ดังตารางที่ 7

ตารางที่ 7 โครงสร้างของข้อสอบวิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษาตอนปลาย จำแนกตามมาตรฐานและตัวชี้วัด

ที่	รหัส	มาตรฐาน/ ตัวชี้วัด	จำนวน (ข้อ)	ข้อสอบ
1	ค 1.1/1	แสดงความสัมพันธ์ของจำนวนต่าง ๆ ในระบบจำนวนจริง	5	23, 50, 70, 111, 117
2	ค 1.1/2	มีความคิดรวบยอดเกี่ยวกับค่าสัมบูรณ์ของจำนวนจริง	7	24, 25, 51, 71, 72, 97, 118
3	ค 1.1/3	มีความคิดรวบยอดเกี่ยวกับจำนวนจริงที่อยู่ในรูปเลขยกกำลังที่มีเลขชี้กำลัง เป็นจำนวนตรรกยะ และจำนวนจริงที่อยู่ในรูปกรณฑ์	9	1, 2, 3, 26, 52, 73, 102, 104, 119

ตารางที่ 7 (ต่อ)

ที่	รหัส	มาตรฐาน/ ตัวชี้วัด	จำนวน (ข้อ)	ข้อสอบ
4	ค 1.2/1	เข้าใจความหมายและหาผลลัพธ์ที่เกิดจากการบวก การลบ การคูณ การหารจำนวนจริง จำนวนจริงที่อยู่ในรูปเลขยกกำลังที่มีเลขชี้กำลังเป็นจำนวนตรรกยะ และจำนวนจริงที่อยู่ในรูปกรณฑ์	8	4, 5, 6, 27, 53, 74, 101, 120
5	ค 1.3/1	หาค่าประมาณของจำนวนจริงที่อยู่ในรูปกรณฑ์ และจำนวนจริงที่อยู่ในรูปเลขยกกำลังโดยใช้วิธีการคำนวณที่เหมาะสม	4	28, 54, 75, 121
6	ค 1.4/1	เข้าใจสมบัติของจำนวนจริงเกี่ยวกับการบวก การคูณ การเท่ากัน การไม่เท่ากัน และนำไปใช้ได้	4	29, 55, 76, 122
7	ค 1.4/1, ค 4.2/3	เข้าใจสมบัติของจำนวนจริงเกี่ยวกับการบวก การคูณ การเท่ากัน การไม่เท่ากัน และนำไปใช้ได้, แก้วสมการและอสมการตัวแปรเดียวดีกรีไม่เกินสอง	5	116, 144, 145, 149, 150
8	ค 2.1/1	ใช้ความรู้เรื่อง อัตราส่วนตรีโกณมิติของมุมในการคาดคะเนระยะทาง/ ความสูง	6	30, 56, 77, 109, 112, 123
9	ค 2.1/2	อัตราส่วนตรีโกณมิติและการนำไปใช้	4	78, 146, 147, 148
10	ค 2.2/1	แก้โจทย์ปัญหาเกี่ยวกับระยะทางและความสูง โดยใช้อัตราส่วนตรีโกณมิติ	4	31, 57, 110, 124
11	ค 4.1/1	มีความคิดรวบยอดในเรื่องเซตและการดำเนินการของเซต	4	32, 58, 79, 114
12	ค 4.1/2	เข้าใจและสามารถให้การให้เหตุผลแบบอุปนัยและนิรนัย	5	7, 33, 59, 80

ตารางที่ 7 (ต่อ)

ที่	รหัส	มาตรฐาน/ ตัวชี้วัด	จำนวน (ข้อ)	ข้อสอบ
13	ค 4.1/3	มีความคิดรวบยอดเกี่ยวกับความสัมพันธ์ และฟังก์ชัน เขียนแสดงความสัมพันธ์ และฟังก์ชันในรูปแบบต่าง ๆ เช่น ตาราง กราฟ และสมการ	5	12, 34, 60, 81, 113
14	ค 4.1/4	เข้าใจความหมายของลำดับและหาพจน์ ทั่วไปของลำดับจำกัด	5	9, 35, 61, 82, 125
15	ค 4.1/5	เข้าใจความหมายของลำดับเลขคณิต และลำดับเรขาคณิต หาพจน์ต่าง ๆ ของลำดับเลขคณิตและลำดับเรขาคณิต และนำไปใช้	6	10, 36, 62, 83, 108, 126
16	ค 4.2/1	เขียนแผนภาพเวเนน-ออยเลอร์แสดงเซต และนำไปใช้แก้ปัญหา	5	38, 63, 84, 99, 127
17	ค 4.2/1, ค 4.1/2	เขียนแผนภาพเวเนน-ออยเลอร์แสดงเซต และนำไปใช้แก้ปัญหา, เข้าใจและสามารถ ใช้การให้เหตุผลแบบอุปนัยและนิรนัย	2	20, 21
18	ค 4.2/2	ตรวจสอบความสมเหตุสมผลของการให้ เหตุผล โดยใช้แผนภาพเวเนน-ออยเลอร์	4	39, 64, 115, 128
19	ค 4.2/2, ค 4.1/2	ตรวจสอบความสมเหตุสมผลของการให้ เหตุผล โดยใช้แผนภาพเวเนน-ออยเลอร์, เข้าใจและสามารถใช้การให้เหตุผล แบบอุปนัยและนิรนัย	4	22, 141, 142, 143
20	ค 4.2/3	แก้สมการและอสมการตัวแปรเดียวดีกรี ไม่เกินสอง	3	40, 86, 129
21	ค 4.2/4	สร้างความสัมพันธ์หรือฟังก์ชันจาก สถานการณ์ หรือปัญหาและนำไปใช้ ในการแก้ปัญหา	4	13, 41, 87, 130
22	ค 4.2/5	ใช้กราฟของสมการ อสมการ ฟังก์ชัน ในการแก้ปัญหา	6	14, 37, 43, 88, 89, 131

ตารางที่ 7 (ต่อ)

ที่	รหัส	มาตรฐาน/ ตัวชี้วัด	จำนวน (ข้อ)	ข้อสอบ
23	ค 4.2/6	เข้าใจความหมายของผลบวก n พจน์แรก ของอนุกรมเลขคณิตและอนุกรมเรขาคณิต หาผลบวก n พจน์แรกของอนุกรมเลขคณิต และอนุกรมเรขาคณิต โดยใช้สูตรและ นำไปใช้	5	8, 11, 65, 98, 132, 98
24	ค 5.1/1	เข้าใจวิธีการสำรวจความคิดเห็นอย่างง่าย	4	15, 66, 90, 133, 90, 133
25	ค 5.1/2	หาค่าเฉลี่ยเลขคณิต มัธยฐาน ฐานนิยม ส่วนเบี่ยงเบนมาตรฐาน และเปอร์เซ็นต์ไทล์ ของข้อมูล	9	16, 42, 44, 67, 91, 103, 106, 107, 134
26	ค 5.1/3	เลือกใช้ค่ากลางที่เหมาะสมกับข้อมูล และวัดดูประสงค์	6	17, 45, 68, 92, 100, 135
27	ค 5.2/1	นำผลที่ได้จากการสำรวจความคิดเห็น ไปใช้ คาดการณ์ในสถานการณ์ที่กำหนดให้	4	18, 69, 93, 136, 18, 69
28	ค 5.2/2	อธิบายการทดลองสุ่มเหตุการณ์ ความน่าจะเป็น ของเหตุการณ์ และนำผลที่ได้ไปใช้คาดการณ์ ในสถานการณ์ที่กำหนดให้	4	46, 47, 94, 137, 46
29	ค 5.3/1	ใช้ข้อมูลข่าวสารและคำสถิติช่วย ในการตัดสินใจ	5	19, 48, 95, 105, 139
30	ค 5.3/2	ใช้ความรู้เกี่ยวกับความน่าจะเป็นช่วย ในการตัดสินใจและแก้ปัญหา	4	49, 96, 138, 140
รวม			150	

การสร้างเครื่องมือที่ใช้ในการวิจัย

ข้อสอบแบบเลือกตอบแบบถูกผิด

การสร้างข้อสอบวัดผลสัมฤทธิ์วิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษาตอนปลาย
มีขั้นตอนในการสร้าง ดังนี้

1. ศึกษาจุดประสงค์ เนื้อหาและเวลาการเรียนการสอนวิชา คณิตศาสตร์ ระดับชั้นมัธยมศึกษาตอนปลาย ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551
2. ศึกษาสมรรถภาพที่ใช้วัดความรู้ทางคณิตศาสตร์ มี 30 ตัวชี้วัด ผู้วิจัยเลือกวัดสมรรถภาพในด้านทักษะและกระบวนการทางคณิตศาสตร์ เนื่องจาก เป็นสมรรถภาพและเนื้อหาพื้นฐานทางคณิตศาสตร์ที่นักเรียนทุกคนต้องผ่าน และเป็นเนื้อหาสำคัญที่นักเรียนต้องนำไปใช้ในการทดสอบระดับชาติขั้นพื้นฐาน (O-NET)
3. กำหนดลักษณะของเครื่องมือวัด ผู้วิจัยเลือกใช้รูปแบบของข้อสอบแบบเลือกตอบแบบถูกผิด 4 ตัวเลือก
4. กำหนดวัตถุประสงค์ย่อย สำหรับใช้วัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ตามมาตรฐานและตัวชี้วัดวิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษาตอนปลาย
5. กำหนดโครงสร้างของข้อสอบวิชาคณิตศาสตร์เพื่อใช้เป็นกรอบในการสร้างข้อสอบ และนำไปใช้เขียนข้อคำถามและตัวเลือก
6. สร้างข้อสอบเพื่อวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ จำนวน 190 ข้อ โดยแยกออกเป็น 9 ฉบับ ฉบับละ 30 ข้อ แต่ละฉบับจะมีลักษณะใกล้เคียงกันตามโครงสร้างของข้อสอบที่เป็นชุดคำถามเดียวกัน ซึ่งแต่ละฉบับมีข้อสอบที่ต่างจากฉบับอื่นอยู่ 10 ข้อ และเหมือนกันทุกฉบับอยู่ 20 ข้อ ข้อสอบที่เหมือนกันทุกฉบับนี้จะถูกใช้เป็นข้อสอบร่วม (Anchor item) สำหรับเทียบคะแนนของข้อสอบแต่ละฉบับให้อยู่บนสเกลเดียวกัน ซึ่งจำนวนข้อสอบร่วมที่ใช้ในทุกฉบับนั้น ผู้วิจัยได้ใช้จำนวนข้อสอบร่วมตามข้อเสนอของลอร์ด (Lord) ที่แนะนำว่าควรใช้ข้อสอบร่วมอย่างน้อยร้อยละ 20 ของข้อสอบทั้งหมดในฉบับ
7. นำข้อสอบไปให้ผู้เชี่ยวชาญด้านหลักสูตร การสอน การวัดและประเมินผลการศึกษา คณิตศาสตร์ ตรวจสอบความเที่ยงตรงตามเนื้อหาและเที่ยงตรงเชิงโครงสร้าง โดยใช้เทคนิคของโรวินเนลลี และแฮมเบิลตัน (Rovinelli & Hambleton) กำหนดเกณฑ์ในการพิจารณา 3 เกณฑ์ คือ ตรง (+1) ไม่ตรง (-1) และไม่แน่ใจ (0) ตรงตามมาตรฐานและตัวชี้วัดหรือไม่
8. ปรับปรุงข้อสอบตามคำแนะนำของผู้เชี่ยวชาญ ซึ่งส่วนใหญ่แล้วจะปรับปรุงเกี่ยวกับภาษาที่ใช้ และรูปภาพ เพื่อให้สื่อความหมายได้ตรงกับสิ่งที่ต้องการวัด
9. นำข้อสอบไปทดสอบกับกลุ่มตัวอย่างระยะแรก แล้วหาค่าคุณภาพของข้อสอบตามทฤษฎีการทดสอบแบบดั้งเดิม เพื่อนำไปใช้ในระยะที่สองต่อไป

โปรแกรมคอมพิวเตอร์สำหรับการทดสอบ

การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์

(Computerized Adaptive Testing: CAT) นั้นดำเนินการสอบโดยให้ผู้สอบนั่งหน้าจอคอมพิวเตอร์ และตอบข้อสอบครั้งละ 1 ข้อ โดยคอมพิวเตอร์เป็นผู้เลือกข้อสอบจากคลังข้อสอบขึ้นมาให้ผู้สอบได้ตอบ ข้อสอบที่ถูกเลือกขึ้นมาจะมีความเหมาะสมกับความสามารถของผู้สอบแต่ละคน ในการดำเนินการสอบลักษณะนี้ จำเป็นต้องมีโปรแกรมสั่งให้คอมพิวเตอร์จัดกระทำกับข้อมูล ในลักษณะต่าง ๆ ตามเงื่อนไขของการทดสอบ ซึ่งผู้วิจัยได้สร้างโปรแกรมคอมพิวเตอร์ดังกล่าว ตามขั้นตอน ดังนี้

1. ออกแบบโปรแกรม ได้แก่ หน้าหลัก การลงทะเบียนผู้สอบ การทดสอบ รายงานผล หน้าหลัก เป็นการให้ผู้สอบเลือกชื่อ โรงเรียนที่นักเรียนกำลังศึกษาอยู่ แล้วพิมพ์

Username: [เลขประจำตัวนักเรียน 5 หลัก] Password: [เลขประจำตัวประชาชน 13 หลัก]

การลงทะเบียนผู้สอบ เป็นการป้อนข้อมูลผู้เข้าสอบก่อนที่จะเริ่มทดสอบ เกี่ยวกับ รหัสนักเรียน 5 หลัก เช่น 29009

ชื่อ-สกุล เช่น สมชาย รักเรียน

ระดับชั้น เช่น 6/1

ผลการเรียนเฉลี่ยวิชาคณิตศาสตร์พื้นฐาน เช่น 3.50

โรงเรียนของผู้เข้าสอบ เช่น บุรีรัมย์พิทยาคม

Email เช่น 29001@brp.ac.th

การทดสอบ การทดสอบจะเริ่มเลือกข้อสอบตามความสามารถของผู้สอบบนพื้นฐาน คะแนนเฉลี่ยวิชาคณิตศาสตร์พื้นฐาน แบ่งออกเป็น 3 กลุ่ม ได้แก่

กลุ่มสูง คะแนนระหว่าง 80-100 คะแนน กำหนดค่าความยากให้ $b > 0.50$

กลุ่มกลาง คะแนนระหว่าง 65-79 คะแนน กำหนดค่าความยากให้ $-0.50 \leq b \leq 0.50$

กลุ่มต่ำ คะแนนระหว่าง 0-64 คะแนน กำหนดค่าความยากให้ $b < -0.50$

และโปรแกรมจะเลือกข้อสอบตามความสามารถเริ่มต้นที่ประมาณได้จากคะแนนวิชาคณิตศาสตร์ พื้นฐานซึ่งเป็นเป็น 3 ระดับ และดำเนินการสอบไปจน $SEE \leq 0.50$, $SEE \leq 0.40$, $SEE \leq 0.30$

โดยพิจารณาข้อสอบและเลือกตอบตามเงื่อนไขของโปรแกรมการทดสอบ

โดยการทดสอบครั้งที่ 1 จะเป็นรูปแบบ Multiple Response Method (MR) ผู้สอบ จะต้องเลือกตัวเลือกที่ถูกต้องที่สุดเพียงตัวเลือกเดียว ต่อมาอีก 2 สัปดาห์จะเป็นการสอบครั้งที่ 2 โดยใช้รูปแบบ Multiple True-False Method (MTF) ผู้สอบจะต้องตอบทุกตัวเลือกโดยพิจารณา คำตอบว่าถ้าผิดจะเป็นการเลือกตัวลวงบางตัวได้อย่างถูกต้อง โดยทำเครื่องหมายกากบาท (X)

ตรงตัวเลือกที่เป็นตัวลวง จะกำหนดให้ค่าข้อมูลเท่ากับ 2 กรณีที่รู้คำตอบว่าถูกจะเป็นการเลือกตัวเลือกบางตัวที่ถูกได้อย่างถูกต้อง โดยทำเครื่องหมายถูก (✓) ตรงตัวเลือกที่เป็นตัวถูกจะกำหนดให้ค่าข้อมูลเท่ากับ 1 และต่อมาอีก 2 สัปดาห์จะเป็นการสอบครั้งที่ 3 โดยใช้รูปแบบ Omit Multiple True-False Method (OMTF) ผู้สอบจะต้องตอบทุกตัวเลือกโดยพิจารณาคำตอบว่าถ้าผิดจะเป็นการเลือกตัวลวงบางตัวได้อย่างถูกต้อง โดยทำเครื่องหมายกากบาท (×) ตรงตัวเลือกที่เป็นตัวลวง จะกำหนดให้ค่าข้อมูลเท่ากับ 2 กรณีที่รู้คำตอบว่าถูกจะเป็นการเลือกตัวเลือกบางตัวที่ถูกได้อย่างถูกต้อง โดยทำเครื่องหมายถูก (✓) ตรงตัวเลือกที่เป็นตัวถูกจะกำหนดให้ค่าข้อมูลเท่ากับ 1 กรณีที่ไม่แน่ใจว่าเป็นตัวเลือกที่ถูกหรือตัวเลือกที่ผิด จะทำเครื่องหมายคำถาม (?) ในตัวเลือกนั้น จะกำหนดให้ค่าข้อมูลเท่ากับ 0

การรายงานผล เป็นการรายงานข้อมูลผู้สอบ ผลการวิเคราะห์ความสามารถ และสรุปผลการประเมินความสามารถ ได้แก่

ข้อมูลผู้สอบ

ชื่อ-สกุลผู้สอบ:

ชั้น:ผลการเรียนเฉลี่ย :โรงเรียน :

Email: Report No :

ผลการวิเคราะห์ความสามารถ

มาตรฐาน/ ตัวชี้วัด.....

ค่าความสามารถ

ความหมาย

ผลการวิเคราะห์.....

สรุปผลการประเมินความสามารถ

เวลาในการสอบ.....นาที

ค่าฟังก์ชันสารสนเทศของชุดข้อสอบ

จำนวนข้อสอบ.....ข้อ

ความคลาดเคลื่อนมาตรฐานในการวัด.....

ความเชื่อมั่นในการประมาณค่า.....

ความสามารถโดยรวม..... ระดับ.....

ผลการประเมิน.....

เกณฑ์การวิเคราะห์ความสามารถของผู้สอบ ตามมาตรฐานและตัวชี้วัด 30 ตัวชี้วัด

ค่าความสามารถ	ความหมาย	ผลวิเคราะห์
$\hat{\theta} > 0.50$	ดีมาก	ระดับสูง
$0.25 < \hat{\theta} \leq 0.50$	ดี	ระดับปานกลาง
$-0.25 < \hat{\theta} \leq 0.25$	ค่อนข้างดี	ระดับปานกลาง
$-0.50 < \hat{\theta} \leq -0.25$	ปกติ	ระดับปานกลาง
$\hat{\theta} < -0.50$	น้อยกว่าปกติ	ระดับต่ำ

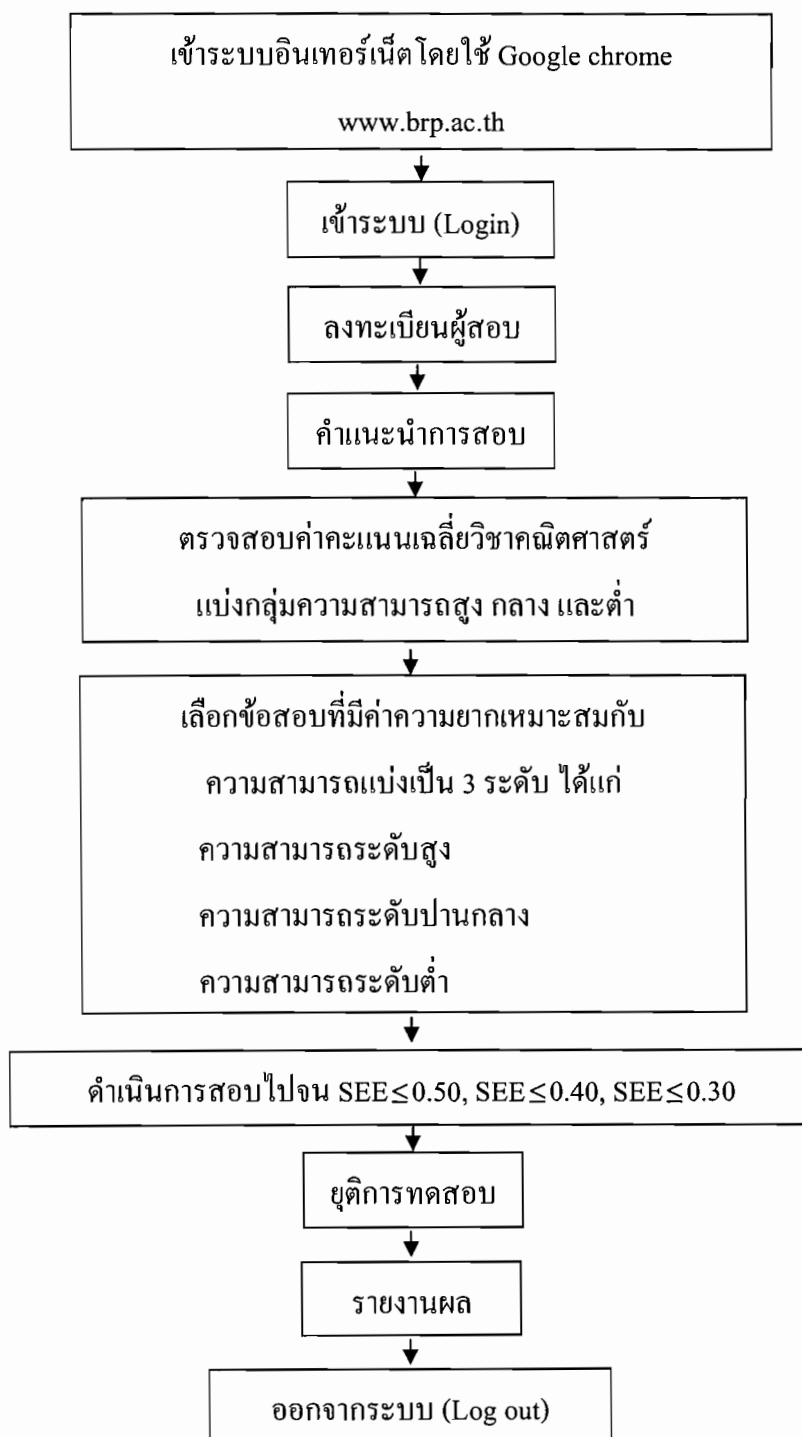
เกณฑ์การประเมินความสามารถของผู้สอบ

ค่าเฉลี่ยความสามารถ	ความหมาย	ผลประเมิน
$\hat{\theta} > 0.50$	ระดับสูง	ผ่าน
$-0.50 \leq \hat{\theta} \leq 0.50$	ระดับปานกลาง	ผ่าน
$\hat{\theta} < -0.50$	ระดับต่ำ	ไม่ผ่าน

2. ศึกษาและเปรียบเทียบโปรแกรมภาษาคอมพิวเตอร์ที่จะนำมาใช้ในการพัฒนา

โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ผู้วิจัยคำนึงถึงความสะดวกในการนำไปใช้ จึงเลือก โปรแกรมภาษา ASP.NET ทั้งนี้เนื่องจาก ASP.NET เป็นภาษาสคริปต์บนเซิร์ฟเวอร์ไซด์ (Serverside scripting language) ซึ่งการประมวลผลจะเกิดขึ้นบนเครื่องแม่ข่าย (Server) ที่เป็นระบบปฏิบัติการของ Windows แล้วจึงสร้างผลลัพธ์เป็นภาษา HTML ส่งให้กับเครื่องลูกข่าย (Client) เพื่อแสดงผล ซึ่งลดภาระการส่งถ่ายข้อมูลจำนวนมากการเขียนสามารถทำได้โดยการเขียน โค้ด ASP.NET แทรกลงไปใน โค้ด HTML ได้อีกทั้ง ASP.NET นั้นเป็นภาษาสคริปต์ยุคใหม่ที่ได้รับความนิยมในปัจจุบัน

3. ดำเนินการพัฒนาโปรแกรมเพื่อใช้ในการทดสอบหาประสิทธิภาพด้านจำนวนข้อสอบ และฟังก์ชันสารสนเทศของชุดข้อสอบ ดังนี้



ภาพที่ 15 แสดงขั้นตอนการทดสอบ

4. ให้ผู้เชี่ยวชาญตรวจสอบการทำงานของโปรแกรมและปรับปรุงแก้ไขความถูกต้องให้พร้อมสำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

การเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูลเพื่อสร้างคลังข้อสอบวิชาคณิตศาสตร์ เรื่อง ทักษะ และกระบวนการทางคณิตศาสตร์ ได้ดำเนินการ ดังนี้

1. นำหนังสือจากมหาวิทยาลัยบูรพา ไปให้สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 32 เพื่อขอความร่วมมือในการวิจัย โดยขอเข้าร่วม โครงการทดสอบความรู้ขั้นพื้นฐาน ของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 6 (Pre-O-NET) จำนวน 67 โรงเรียน
2. นำข้อสอบมอบให้สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 32 เพื่อนำไป จัดเป็นชุดข้อสอบทำการทดสอบกับกลุ่มตัวอย่าง (พฤศจิกายน-ธันวาคม 2557)
3. ทดสอบนักเรียนระดับชั้นมัธยมศึกษาปีที่ 6 ในโรงเรียนกลุ่มตัวอย่าง ซึ่งใช้วิธีการแจก แบบทดสอบให้แก่ นักเรียนในห้องสอบแบบเรียงสลับอย่างเป็นระบบ กล่าวคือ คนที่ 1-9 จะได้รับ แบบทดสอบฉบับที่ 1-9 ตามลำดับ คนที่ 10-18 จะได้รับแบบทดสอบฉบับที่ 1-9 ตามลำดับ การแจกแบบทดสอบจะกระทำเช่นนี้ไปเรื่อย ๆ จนครบนักเรียนทุกคนที่มีอยู่ในห้องสอบ

การเก็บรวบรวมข้อมูลเพื่อหาประสิทธิภาพการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ผู้วิจัยเลือกกลุ่มตัวอย่าง คือ โรงเรียนบุรีรัมย์พิทยาคม เนื่องจากผู้วิจัยต้องการให้คะแนนที่ได้ จากการสอบครั้งนี้ ได้จากกลุ่มนักเรียนที่ได้รับการสอนด้วยวิธีที่ใกล้เคียงกัน หลักสูตรเดียวกัน และจัดการสอบได้พร้อมกัน ดังนั้นผู้วิจัยจึงได้เลือกนักเรียนระดับชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2557 จำนวน 540 คน นัดหมายการสอบที่ห้องคอมพิวเตอร์ ดำเนินการสอบ 3 ครั้ง โดยให้ กลุ่มตัวอย่างตอบข้อสอบ โดยใช้รูปแบบการตอบทุกตัวเลือกแบบถูกผิดและให้คะแนนแบบ 2 ค่า (Multiple Response Method: MR) ครั้งที่ 2 ให้กลุ่มตัวอย่างตอบข้อสอบ โดยใช้รูปแบบ การตอบทุกตัวเลือกแบบถูกผิด และให้คะแนนแบบทุกตัวเลือก (Multiple True-False Method: MTF) และครั้งที่ 3 ให้กลุ่มตัวอย่างตอบข้อสอบ โดยใช้รูปแบบการตอบทุกตัวเลือกแบบถูกผิด และไม่แน่ใจ (Omit Multiple True-False Method: OMTF) แต่ละครั้งเว้นระยะห่างกัน 2-3 สัปดาห์

การวิเคราะห์ข้อมูลการวิจัย

ผู้วิจัยทำการวิเคราะห์ข้อมูล มีขั้นตอนดังนี้

1. หาค่าสถิติพื้นฐานของข้อสอบ ดังนี้
 - 1.1 ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของค่าพารามิเตอร์ข้อสอบ การประมาณ ค่าพารามิเตอร์ของข้อสอบ โดยวิเคราะห์ข้อสอบรายข้อตามรูปแบบ โลจิสติก 3 พารามิเตอร์ เพื่อให้ได้ค่าอำนาจจำแนก (a) ค่าความยาก (b) ค่าการเดา (c) ด้วยโปรแกรมสำเร็จรูปไปล็อก

(BILOG Version 3.04) ซึ่งความสัมพันธ์เชิงคณิตศาสตร์ของพารามิเตอร์ตามรูปแบบ โมเดล โลจิสติก 3 พารามิเตอร์ สามารถเขียนได้ดังนี้ (Hambleton, Swaminathan & Roger, 1991, p. 17)

$$P_i(\theta) = c_i + (1 - c_i) \frac{e^{D_i(\theta - b_i)}}{1 + e^{D_i(\theta - b_i)}} \quad ; \quad i = 1, 2, 3, \dots, n$$

- เมื่อ $P_i(\theta)$ แทน โอกาสการตอบที่ผู้มีความสามารถ (θ) จะตอบข้อสอบข้อ i
- θ แทน ระดับความสามารถของผู้สอบ
- a_i แทน ค่าอำนาจจำแนกของข้อสอบข้อ i
- b_i แทน ค่าความยากของข้อสอบข้อ i
- c_i แทน ค่าการเดาของข้อสอบข้อ i
- D แทน Scaling factor มีค่าเท่ากับ 1.7
- e แทน ค่าคงที่มีค่าเท่ากับ 2.71828...

1.2 วิเคราะห์องค์ประกอบเพื่อตรวจสอบความเป็นมิติเดียว การทดสอบความเป็นคุณลักษณะเดียวของแบบทดสอบ โดยการวิเคราะห์องค์ประกอบ (Factor analysis) ใช้โปรแกรมคอมพิวเตอร์ การวิเคราะห์นี้ใช้การวิเคราะห์ตัวประกอบสำคัญ (Principal component) แล้วหมุนแกนด้วยวิธีเวรีแมกซ์ (Varimax) สำหรับการตรวจสอบพิจารณาค่าไอเกน (Eigen value) ถ้าหากค่าไอเกนของตัวประกอบที่ 1 ของแบบทดสอบมีค่าสูงกว่าค่าไอเกนของตัวประกอบที่ 2 อย่างมาก และค่าไอเกนของตัวประกอบที่ 2 สูงกว่าค่าไอเกนของตัวประกอบถัดไปเพียงเล็กน้อย จึงจะกล่าวได้ว่าแบบทดสอบทั้งฉบับมีคุณสมบัติการวัดเพียงมิติเดียว (Lord, 1980, p. 21)

2. หาค่าคุณภาพของข้อสอบแบบเลือกตอบแบบถูกผิด ตามตัวแปรดังนี้

2.1 ความตรง (Validity)

2.2 ความเที่ยง (Reliability)

2.3 ความคลาดเคลื่อนมาตรฐานในการวัด (Standard error of measurement)

3. หาประสิทธิภาพของการทดสอบตามตัวแปรจำนวนข้อสอบ (Length testing)

และฟังก์ชันสารสนเทศของชุดข้อสอบ (Test information function) ภายใต้เงื่อนไขของวิธีการให้คะแนนที่ต่างกัน 3 วิธี ได้แก่ วิธี Multiple-Response Method (MR) วิธี Multiple True-False Method

(MTF) และวิธีที่ผู้วิจัยเสนอคือวิธี Omit Multiple True-False Method (OMTF) เปรียบเทียบตามแนวคิดของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ดังนี้

- 3.1 เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ (Item selection rule)
- 3.2 การประมาณค่าความสามารถของผู้สอบ (Ability estimation procedure)
- 3.3 เกณฑ์ยุติการทดสอบ (Termination criteria)
- 3.4 ความสามารถของผู้สอบ (Ability's examinee)

สถิติที่ใช้ในการวิจัย

การวิเคราะห์ข้อมูลนั้นได้ใช้สถิติสำหรับคำนวณค่าของตัวแปรตามเพื่อศึกษาผลที่เกิดขึ้นอันเนื่องมาจากการให้คะแนนที่แตกต่างกัน 3 แบบ แล้วเปรียบเทียบตามองค์ประกอบสำคัญของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ เพื่อหาประสิทธิภาพของการทดสอบ

1. สถิติที่ใช้ในการสร้างและพัฒนาเครื่องมือ

1.1 ตรวจสอบความตรงตามเนื้อหา (Content validity) โดยพิจารณาค่าดัชนีความสอดคล้อง ของข้อสอบกับจุดประสงค์การเรียนรู้ โดยใช้สูตรของโรวินลลี และแฮมเบิลตัน (Rovinelli & Hambleton, 1937 อ้างถึงใน บุญเชิด ภิญโญนนตพงษ์, 2527, หน้า 68)

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

เมื่อ IOC หมายถึง ดัชนีความสอดคล้องระหว่างข้อสอบกับจุดประสงค์ การเรียนรู้

$\sum R$ หมายถึง ผลรวมของคะแนนความคิดเห็นของผู้เชี่ยวชาญ

N หมายถึง จำนวนผู้เชี่ยวชาญ

1.2 วิเคราะห์ข้อสอบเป็นรายข้อ โดยตรวจให้คะแนนด้วยวิธี 0-1 ตามทฤษฎีการทดสอบ แบบดั้งเดิม (Classical Test Theory: CTT) เพื่อหาค่าความยาก (P) และค่าอำนาจจำแนก (D) ของข้อสอบ

1.2.1 ค่าความยาก (P) หมายถึง สัดส่วนของจำนวนผู้ที่ตอบข้อนั้น ๆ ได้ถูกต้อง ต่อจำนวนผู้ตอบข้อนั้น ๆ ทั้งหมด (สุพัฒน์ สุขมลสันต์, 2538, หน้า 18) ใช้สูตรดังนี้

$$P = \frac{N_c}{N_t}$$

- P หมายถึง คำนีค่าความยาก
 N_c หมายถึง จำนวนผู้เข้าสอบทั้งหมดที่ตอบข้อสอบข้อนั้นถูก
 N_t หมายถึง จำนวนผู้เข้าสอบทั้งหมดที่ตอบข้อสอบข้อนั้น

1.2.2 ค่าอำนาจจำแนก (D) หมายถึง คำนีที่ชี้ว่าข้อทดสอบแต่ละข้อมีความสามารถ ในการจำแนกผู้สอบที่เก่งออกจากผู้ที่ไม่เก่งได้มากน้อยเพียงใด ใช้สูตรคำนวณคำนีค่าอำนาจจำแนก ดังนี้ (ล้วน สายยศ และอังคณา สายยศ, 2539, หน้า 186)

$$D = \frac{U}{n_u} - \frac{L}{n_L}$$

- เมื่อ D หมายถึง คำนีค่าอำนาจจำแนกของข้อสอบ
 U หมายถึง จำนวนนักเรียนที่ตอบถูกในกลุ่มคะแนนสูง
 L หมายถึง จำนวนนักเรียนที่ตอบถูกในกลุ่มคะแนนต่ำ
 n_u หมายถึง จำนวนนักเรียนทั้งหมดที่ตอบถูกในกลุ่มคะแนนสูง
 n_L หมายถึง จำนวนนักเรียนทั้งหมดที่ตอบถูกในกลุ่มคะแนนต่ำ

ในการวิเคราะห์ข้อสอบเป็นรายตัวเลือก โดยตรวจให้คะแนนด้วยวิธีให้คะแนนทุกตัวเลือก (MTF) ตามทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory: CTT) เพื่อหาค่าความยาก (P) และคำนีค่าอำนาจจำแนก (D) ของตัวเลือก จะใช้สูตรเดียวกันคือ สูตรในข้อ 1.2.1 และ 1.2.2 ตามลำดับ เนื่องจากวิธีการตรวจให้คะแนนทุกตัวเลือกนี้ แต่ละตัวเลือก ทำหน้าที่เหมือนข้อสอบ 1 ข้อย่อย

1.3 การตรวจสอบความเป็นมิติเดียวของข้อสอบ ใช้โปรแกรม SPSS for Windows วิเคราะห์ ตัวประกอบสำคัญ (Principal component) แล้วหมุนแกนด้วยวิธีแวนิเม็กซ์ (Varimax) สำหรับการตรวจสอบพิจารณาค่าไอเกน (Eigen value) ถ้าหากค่าไอเกนของตัวประกอบที่ 1 ของแบบทดสอบมีค่าสูงกว่า ค่าไอเกนของตัวประกอบที่ 2 อย่างมาก และค่าไอเกนของตัวประกอบที่ 2

มีค่าสูงกว่าค่าไอเกนของตัวประกอบ ถัดไปเพียงเล็กน้อย จึงจะกล่าวได้ว่าแบบทดสอบทั้งฉบับนี้มีคุณสมบัติเพียงมิติเดียว (Lord, 1980, p. 21)

1.4 ประมาณค่าพารามิเตอร์ของข้อสอบ โดยการนำคะแนนผลการตอบแบบทดสอบแต่ละฉบับมาวิเคราะห์รายข้อตามรูปแบบโลจิสติก 3 พารามิเตอร์โดยวิธีความเป็นไปได้สูงสุด (Maximum likelihood) หาค่าอำนาจจำแนก (a) ค่าความยาก (b) และค่าการเดา (c) โดยใช้โปรแกรมสำเร็จรูป BILOG Version 3.04 ซึ่งความสัมพันธ์เชิงคณิตศาสตร์ของค่าพารามิเตอร์ตามรูปแบบโลจิสติก 3 พารามิเตอร์ สามารถเขียนได้ ดังนี้ (Hambleton et al., 1991, p. 17)

$$P_i(\theta) = c_i + \frac{1}{1 + e^{-D_i(\theta - b_i)}} \quad ; i=1,2,3,\dots$$

เมื่อ $P_i(\theta)$ หมายถึง ความน่าจะเป็นที่ผู้ตอบที่มีระดับความสามารถ θ ตอบข้อสอบข้อที่ i ถูก

θ หมายถึง ระดับความสามารถที่แท้จริงของผู้สอบที่ประมาณได้จากคะแนนรวมของการตอบแบบทดสอบโดยการปรับให้เป็นคะแนนที่มีค่าเฉลี่ยเท่ากับ 0 และมีส่วนเบี่ยงเบนมาตรฐาน เท่ากับ 1

a_i หมายถึง ค่าอำนาจจำแนกของข้อสอบข้อที่ i ซึ่งมีค่าเป็น สัดส่วน โดยตรงกับความชันของโค้งลักษณะข้อสอบ ณ จุด เปลี่ยน โคง (ที่จุด $\theta = b_i$) โดยทั่วไป มักจะเลือกใช้ช่วงของ ค่าอำนาจจำแนกของข้อสอบอยู่ระหว่าง 0.30 ถึง 2.00

b_i หมายถึง ค่าความยากของข้อสอบข้อที่ i ซึ่งมีค่าเท่ากับ ระดับความสามารถ ของผู้สอบ ณ จุดเปลี่ยน โคง โดยทั่วไปมักจะ เลือกใช้ ช่วงของค่าความยากของข้อสอบระหว่าง -2 ถึง +2 ซึ่งค่าที่ โกล้ -2 หมายถึงข้อสอบที่ง่าย มากและค่าที่ โกล้ +2 หมายถึง ข้อสอบ ที่ยากมาก

c_i หมายถึง ค่าการเดาของข้อสอบข้อที่ i ซึ่งเป็นค่าความน่าจะเป็นที่ ผู้สอบที่มี ความสามารถต่ำมากจะ ตอบข้อสอบ ข้อที่ i ได้ถูก หรือเป็นค่ากำกับ โคงที่ต่ำที่สุด (Lower asymptote) ของ โคง ลักษณะข้อสอบ มีค่า 0 ถึง 1 โดยทั่วไปมักจะเลือก ใช้ค่าการเดา ของข้อสอบต่ำกว่า 0.3

D หมายถึง Scaling Factor มีค่าเท่ากับ 1.7

e หมายถึง ค่าคงที่ มีค่าเท่ากับ 2.71828...

n หมายถึง จำนวนข้อสอบในแบบทดสอบ

2. สถิติที่ใช้ในการวิเคราะห์ข้อมูล

2.1 หาค่าความตรง จากสูตร (Glass & Stanley, 1970, p. 114)

$$r_{xy} = \frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}}$$

เมื่อ r_{xy} แทน ความสัมพันธ์ระหว่างคะแนนจากการตอบแบบทดสอบวิชา
คณิตศาสตร์ และคะแนนเฉลี่ยวิชาคณิตศาสตร์พื้นฐาน

n แทน จำนวนผู้สอบ

x, y แทน คะแนนการตอบข้อสอบวิชาคณิตศาสตร์

2.2 การหาความเที่ยงโดยวิธีของคูเดอร์-ริชาร์ดสัน (Kuder-Richardson: KR) วิธีการนี้จะแตกต่างจากวิธีการหาความเที่ยงแบบต่างๆ ที่กล่าวมา จะไม่ได้ใช้การหาค่าสหสัมพันธ์เพื่อทดสอบความเที่ยง แต่จะใช้วิธีหาความสัมพันธ์ระหว่างองค์ประกอบภายใน ได้แก่ ความสัมพันธ์ระหว่างข้อสอบในฉบับเดียวกัน และการคำนวณหาความสัมพันธ์คะแนนของข้อสอบแต่ละข้อจะต้องแปลงให้เป็นคะแนน 2 ค่าเท่านั้น ได้แก่ ถ้าถูกจะได้ค่า 1 และถ้าผิดจะได้ค่า 0 สูตรในการหาความเชื่อมั่นแบบคูเดอร์-ริชาร์ดสัน (KR-20) เป็นสูตรในการหาค่าความเที่ยงที่เหมาะสมสำหรับแบบทดสอบที่มีค่าความยากง่ายในลักษณะกระจาย สูตรที่ใช้ในการหา มีรูปแบบดังนี้

$$r_t = \frac{n}{n-1} \left\{ 1 - \frac{\sum pq}{S_t^2} \right\}$$

$$S_t^2 = \frac{N \sum X^2 - (\sum X)^2}{N^2}$$

เมื่อ r_t	คือ สัมประสิทธิ์ของความเชื่อมั่นของแบบทดสอบทั้งฉบับ
n	คือ จำนวนข้อของแบบทดสอบ
p	คือ สัดส่วนของผู้เรียนที่ทำข้อสอบข้อนั้นถูกต้องกับผู้เรียนทั้งหมด
q	คือ สัดส่วนของผู้เรียนที่ทำข้อสอบข้อนั้นผิดกับผู้เรียนทั้งหมด
S_t^2	คือ ความแปรปรวนของคะแนนสอบทั้งฉบับ
N	คือ จำนวนผู้เรียน

2.3 หาค่าความคลาดเคลื่อนมาตรฐานในการวัด (Standard error of measurement)

$$SEM = \sigma_0 \sqrt{1 - r_{tt}}$$

เมื่อ σ_0	แทน ค่าเบี่ยงเบนมาตรฐานของคะแนนการสอบ
r_{tt}	แทน ค่าสัมประสิทธิ์แห่งความเที่ยงของแบบสอบ

2.4 หาค่าเฉลี่ยของจำนวนข้อสอบ/ฟังก์ชันสารสนเทศของชุดข้อสอบ คำนวณ

จากสูตร

$$\bar{X} = \frac{\sum x}{n}$$

เมื่อ \bar{X}	แทน จำนวนข้อสอบเฉลี่ยที่ใช้ในการสอบ
$\sum x$	แทน ผลรวมของข้อสอบที่ใช้ในการสอบแต่ละคน
n	แทน จำนวนผู้สอบ

2.5 ฟังก์ชันสารสนเทศของข้อสอบ (Item information function) หาค่าโดยการนำ

ค่าพารามิเตอร์ของข้อสอบในชุดข้อสอบ มาคำนวณ ณ ระดับความสามารถ (θ) ซึ่งแบ่งระดับความสามารถออกเป็น 81 ระดับ คือ -4.0, -3.9, -3.8, ..., 3.8, 3.9 และ 4.0 ใช้สูตรในการคำนวณดังนี้ (Hambleton et al., 1991, p. 91) ดังนี้

$$I_i(\theta) = \frac{2.89a_i^2(1-c_i)}{[c_i + e^{D_i(\theta-b_i)}][1 + c_i + e^{-D_i(\theta-b_i)}]^2} ; i = 1, 2, 3, \dots, n$$

เมื่อ $I_i(\theta)$ แทน ค่าฟังก์ชันสารสนเทศของข้อสอบข้อที่ i ณ ระดับความสามารถ (θ) ใดๆ

θ แทน ระดับความสามารถของผู้สอบ

a_i แทน ค่าอำนาจจำแนกของข้อสอบข้อ i

b_i แทน ค่าความยากของข้อสอบข้อ i

c_i แทน ค่าการเดาของข้อสอบข้อ i

D แทน Scaling factor มีค่าเท่ากับ 1.7

e แทน ค่าคงที่มีค่าเท่ากับ 2.71828...

2.6 ค่าฟังก์ชันสารสนเทศของชุดข้อสอบ (Test information function) คำนวณได้จากสูตร ดังนี้ (Hambleton et al., 1991, p. 94)

$$I(\theta) = \sum_{j=1}^m I_j(\theta) ; i = 1, 2, 3, \dots, n$$

เมื่อ $I(\theta)$ แทน ค่าฟังก์ชันสารสนเทศของแบบทดสอบ ณ ระดับความสามารถ (θ) ใดๆ

m แทน จำนวนข้อสอบ

บทที่ 4

ผลการวิจัย

การเสนอผลการวิเคราะห์ข้อมูลเพื่อศึกษาผลของการให้คะแนนที่มีต่อประสิทธิภาพการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ในด้านจำนวนข้อสอบ และฟังก์ชันสารสนเทศของชุดข้อสอบ จำแนกตามตัวแปรอิสระ 4 ตัวแปร คือ เกณฑ์การคัดเลือกข้อสอบ การประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบ และความสามารถของผู้สอบ ซึ่งมีรายละเอียดของผลการวิเคราะห์ข้อมูล ดังนี้

สัญลักษณ์ที่ใช้ในการแสดงผลการวิเคราะห์ข้อมูล

ผลการวิเคราะห์ข้อมูล แบ่งเป็น 3 ตอน ได้แก่

ตอนที่ 1 ค่าพารามิเตอร์ของข้อสอบในคลังข้อสอบ

ตอนที่ 2 คุณภาพของข้อสอบ

ตอนที่ 3 ประสิทธิภาพของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

สัญลักษณ์ที่ใช้ในการแสดงผลการวิเคราะห์ข้อมูล

เพื่อความเข้าใจตรงกัน ผู้วิจัยได้กำหนดคสัญลักษณ์และอักษรย่อที่ใช้ในการเสนอผลการวิเคราะห์ข้อมูล ดังนี้

Item	แทน	ข้อสอบ
N	แทน	จำนวนผู้สอบ
$\hat{\theta}$	แทน	ค่าความสามารถโดยประมาณของผู้สอบ
θ_{\max}	แทน	ค่าความสามารถที่ทำให้ข้อสอบมีสารสนเทศสูงสุด
I_{\max}	แทน	ค่าสารสนเทศสูงสุดของข้อสอบ
a	แทน	ค่าอำนาจจำแนก (Item discriminating)
b	แทน	ค่าความยากง่าย (Item difficulty)
c	แทน	ค่าการเดา (Item guessing)
\bar{X}	แทน	ค่าเฉลี่ยเลขคณิต (Arithmetic mean)
SD	แทน	ค่าเบี่ยงเบนมาตรฐาน (Standard deviation)
MR	แทน	การให้คะแนนแบบ Multiple-Response Method

MTF แทน การให้คะแนนแบบ Multiple True-False Method

OMTF แทน การให้คะแนนแบบ Omit Multiple True-False Method

ผลการวิเคราะห์ข้อมูล

ตอนที่ 1 ค่าพารามิเตอร์ของข้อสอบในคลังข้อสอบ

ข้อสอบที่นำมาสร้างเป็นคลังข้อสอบนั้น เป็นข้อสอบวัดผลสัมฤทธิ์ทางการเรียน วิชาคณิตศาสตร์ ที่สร้างตามมาตรฐานการเรียนรู้หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ระดับชั้นมัธยมศึกษาตอนปลาย ซึ่งมุ่งวัดทักษะทางสติปัญญาของบุคคล ในการแสวงหาความรู้หรือแก้ปัญหาทางคณิตศาสตร์ เป็นชุดข้อสอบแบบเลือกตอบที่ผู้วิจัย สร้างขึ้นจำนวน 190 ข้อ สร้างขึ้นตามหลักวิชาการสร้างข้อสอบ และตรวจสอบคุณภาพข้อสอบ ตามทฤษฎีการทดสอบแบบดั้งเดิม วิเคราะห์และหาค่าคุณภาพของข้อสอบได้ข้อสอบที่มีคุณภาพ จำนวน 150 ข้อ นำไปจัดเป็นชุดข้อสอบเพื่อตรวจสอบคุณภาพวิเคราะห์ค่าพารามิเตอร์ตามแนว ทฤษฎีการตอบสนองข้อสอบ โดยจัดทำเป็นแบบทดสอบคู่ขนาน จำนวน 9 ฉบับ ฉบับละ 30 ข้อ โดยแต่ละฉบับจะมีลักษณะใกล้เคียงกันตามโครงสร้างของข้อสอบที่เป็นชุดคำถามเดียวกัน ซึ่งแต่ละฉบับจะมีข้อสอบที่เหมือนกันอยู่ 15 ข้อ และต่างจากฉบับอื่นอยู่ 15 ข้อ ข้อสอบ ที่เหมือนกันทุกฉบับนี้จะถูกใช้เป็นส่วนข้อสอบร่วม (Anchor item) สำหรับเทียบคะแนนของข้อสอบ แต่ละฉบับให้อยู่บนสเกลเดียวกัน ซึ่งจำนวนข้อสอบร่วมที่ใช้ในทุกฉบับนั้น ผู้วิจัยได้ใช้ จำนวนข้อสอบร่วมตามข้อเสนอของลอร์ด (Lord, 1980) ที่แนะนำว่าควรใช้ข้อสอบร่วมอย่างน้อย ร้อยละ 20 ของข้อสอบทั้งหมดในฉบับเพื่อนำไปใช้กับกลุ่มตัวอย่าง จากผลการทดลอง ได้ค่าเฉลี่ย และค่าเบี่ยงเบนมาตรฐาน ของค่าอำนาจจำแนก ค่าความยากง่าย และค่าการเดา ของแต่ละฉบับ ดังตารางที่ 8

ตารางที่ 8 จำนวนข้อสอบ ผู้สอบ ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของค่าพารามิเตอร์ของข้อสอบที่ใช้วัดความรู้วิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ยังไม่ได้ปรับสเกล

ฉบับที่	จำนวน		ค่าอำนาจจำแนก (a)		ค่าความยากง่าย (b)		ค่าการเดา (c)	
	ข้อสอบ	ผู้สอบ	\bar{X}	SD	\bar{X}	SD	\bar{X}	SD
1	29	847	0.957	0.174	0.840	0.295	0.126	0.024
2	28	847	0.833	0.097	0.978	0.566	0.125	0.047
3	25	847	0.739	0.060	1.095	1.178	0.142	0.039
4	23	847	0.917	0.184	0.877	1.279	0.152	0.060
5	25	847	0.828	0.105	0.851	0.625	0.135	0.020
6	24	847	0.879	0.173	0.798	1.209	0.164	0.033
7	25	847	0.935	0.238	0.839	0.640	0.112	0.031
8	24	847	0.830	0.076	0.851	0.345	0.152	0.031
9	27	847	0.842	0.120	0.829	0.297	0.135	0.035

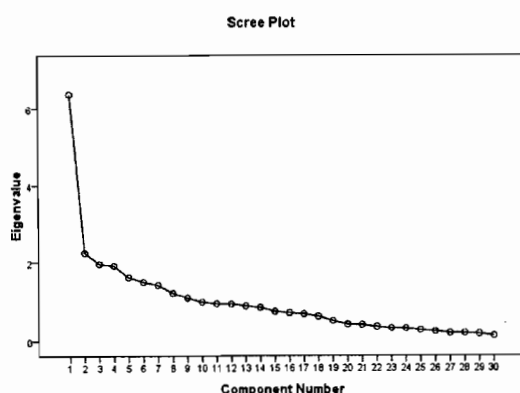
จากตารางที่ 8 จะเห็นว่าแบบทดสอบที่ใช้วัดความรู้วิชาคณิตศาสตร์ของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 6 จำนวน 270 ข้อ ได้ข้อสอบจำนวน 230 ข้อที่ยังไม่ได้ปรับสเกล การวัดนั้นมีลักษณะ ใกล้เคียงกัน ซึ่งสังเกตได้จากค่าเฉลี่ยของค่าพารามิเตอร์ของข้อสอบที่วิเคราะห์ได้พบว่า ค่าอำนาจจำแนกอยู่ระหว่าง 0.739-0.957 และค่าเบี่ยงเบนมาตรฐานอยู่ระหว่าง 0.060-0.238 ซึ่งค่าอำนาจจำแนกอยู่ในเกณฑ์ดี ค่าความยากง่ายอยู่ระหว่าง 0.798-1.095 และค่าเบี่ยงเบนมาตรฐานอยู่ระหว่าง 0.295-1.270 ซึ่งอยู่ในเกณฑ์ค่อนข้างไปทางยาก และค่าการเดามีค่าอยู่ระหว่าง 0.1120-0.1640 และค่าเบี่ยงเบนมาตรฐานอยู่ระหว่าง 0.024-0.047 ซึ่งอยู่ในเกณฑ์ดี

เมื่อนำผลของการตอบข้อสอบวิชาคณิตศาสตร์ จำนวน 9 ฉบับ ไปวิเคราะห์องค์ประกอบเพื่อตรวจสอบความเป็นมิติเดียวของข้อสอบ ซึ่งผลของการวิเคราะห์ดังกล่าว แสดงได้ตารางที่ 9

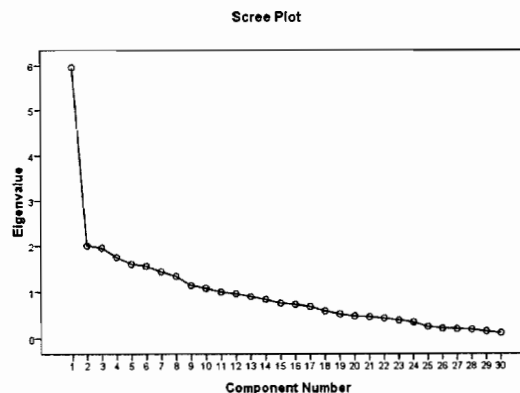
ตารางที่ 9 ค่าไอเกนที่ได้จากการวิเคราะห์ตัวประกอบของแบบทดสอบวิชาคณิตศาสตร์
ฉบับที่ 1-9 จำแนกเป็นรายตัวประกอบ (Eigen value >1.00)

ตัวประกอบ	ฉบับที่								
	1	2	3	4	5	6	7	8	9
1	6.3808	5.9627	6.0972	3.9175	4.7498	4.8785	4.5194	5.1459	6.4254
2	2.2416	2.0103	1.9895	1.9281	2.4960	2.1582	2.2296	2.7698	2.1656
3	1.9564	1.9607	1.8524	1.8484	1.7343	2.0165	1.8787	2.0894	2.0203
4	1.9057	1.7550	1.7204	1.7310	1.7109	1.8193	1.7236	1.8257	1.9806
5	1.6069	1.6067	1.5811	1.6303	1.5788	1.6550	1.6069	1.6947	1.6444
6	1.4850	1.5594	1.4794	1.5651	1.5225	1.4353	1.5287	1.5318	1.5533
7	1.4006	1.4366	1.3858	1.4140	1.4575	1.4035	1.4164	1.3854	1.3663
8	1.2009	1.3332	1.3513	1.3775	1.3537	1.2877	1.2862	1.2694	1.3133
9	1.0836	1.1313	1.2145	1.3231	1.2975	1.2064	1.1759	1.2140	1.2173
10		1.0712	1.1758	1.2496	1.1817	1.1143	1.1140	1.1300	1.1084
11			1.0160	1.1282	1.1226	1.0680	1.0467	1.0409	1.0534
12				1.0713		1.0180			1.0419

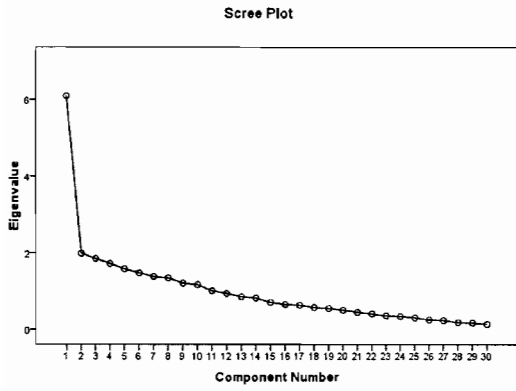
จากตารางที่ 9 จะเห็นว่า แบบทดสอบทั้ง 9 ฉบับ มีค่าไอเกนของตัวประกอบที่ 1 สูงกว่าค่าไอเกนของตัวประกอบอื่น ๆ ที่เหลือ และตัวประกอบที่เหลือดังกล่าวก็มีค่าไอเกนที่ใกล้เคียงกัน ซึ่งลักษณะดังกล่าวนี้พออนุมานได้ว่า ข้อสอบที่ได้มีความเป็นมิติเดียวตรงตามข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ



ฉบับที่ 1

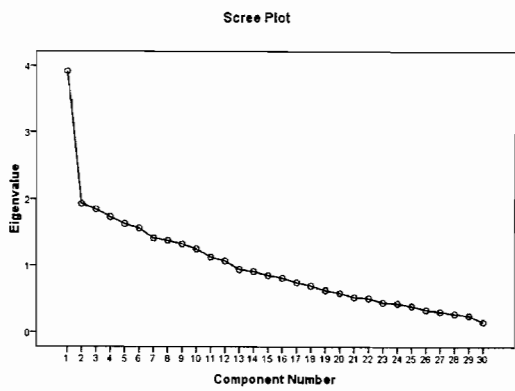


ฉบับที่ 2



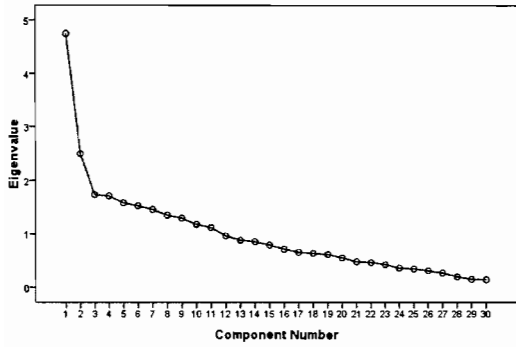
ฉบับที่ 3

Scree Plot



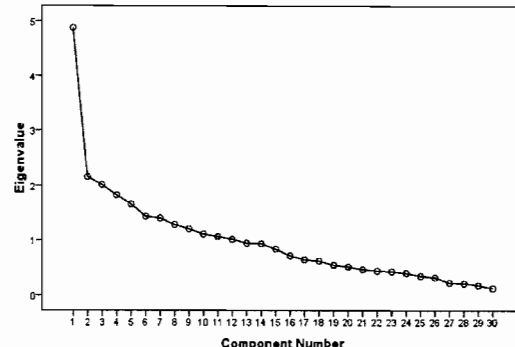
ฉบับที่ 4

Scree Plot



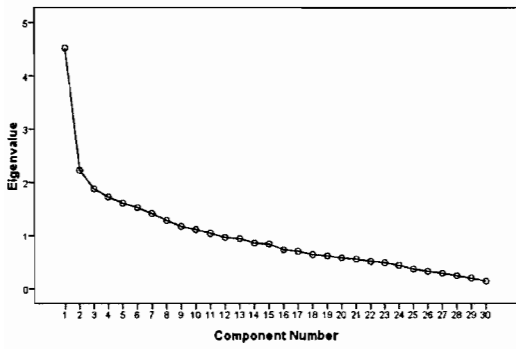
ฉบับที่ 5

Scree Plot



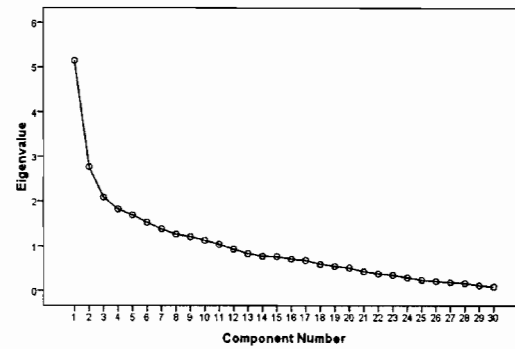
ฉบับที่ 6

Scree Plot

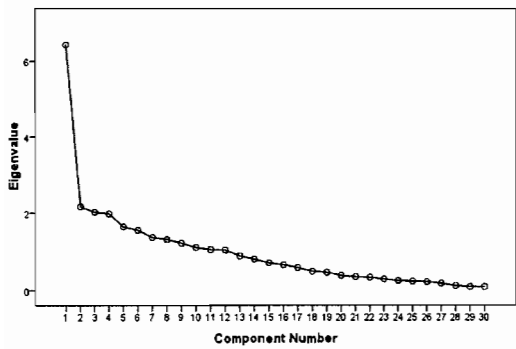


ฉบับที่ 7

Scree Plot



ฉบับที่ 8



ฉบับที่ 9

ภาพที่ 16 ค่าไอเกนที่ได้จากการวิเคราะห์ตัวประกอบของแบบทดสอบวิชาคณิตศาสตร์

ตอนที่ 2 ผลการวิเคราะห์คุณภาพของข้อสอบ ด้านความตรง ความเที่ยง และความคลาดเคลื่อนมาตรฐานในการวัด

ตารางที่ 10 ผลการให้คะแนนที่แตกต่างกัน 3 แบบ จำแนกตามตัวแปรความตรง ความเที่ยง และความคลาดเคลื่อนมาตรฐานในการวัด

การให้คะแนน	ความตรง	ความเที่ยง	ความคลาดเคลื่อนมาตรฐานในการวัด
MR	0.7202	0.7716	0.2326
MTF	0.7233	0.7750	0.4609
OMTF	0.7239	0.7757	0.2305

จากตารางที่ 10 พบว่า ผลการให้คะแนนแบบ MR แบบ MTF และแบบ OMTF มีค่าความตรง คือ 0.7202, 0.7233, 0.7239 ตามลำดับ ค่าความเที่ยง คือ 0.7716, 0.7750, 0.7757 ตามลำดับ และความคลาดเคลื่อนมาตรฐานในการวัด คือ 0.2326, 0.4609, 0.2305 ตามลำดับ

ตอนที่ 3 ผลการวิเคราะห์ประสิทธิภาพของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ จากผลของการให้คะแนน 3 แบบ ด้านจำนวนข้อสอบ และค่าฟังก์ชันสารสนเทศของชุดข้อสอบ

ตารางที่ 11 ค่าเฉลี่ย และค่าเบี่ยงเบนมาตรฐานของผลการให้คะแนน 3 แบบ ที่ใช้ในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ จำแนกตามเกณฑ์การคัดเลือกข้อสอบ

การให้คะแนน	จำนวนข้อสอบ		ฟังก์ชันสารสนเทศ		
	\bar{X}	SD	\bar{X}	SD	
MR	b	17.59	5.78	11.462	0.292
	Q_{max}	14.45	5.23	11.818	0.531
	I_{max}	12.45	3.76	11.870	0.600
MTF	b	15.44	8.78	11.757	0.648
	Q_{max}	14.40	7.23	11.752	0.619
	I_{max}	11.34	4.76	11.804	0.616

ตารางที่ 11 (ต่อ)

การให้คะแนน	จำนวนข้อสอบ		ฟังก์ชันสารสนเทศ	
	\bar{X}	SD	\bar{X}	SD
OMTF b	16.34	6.78	11.740	0.598
Q_{max}	14.50	6.23	11.801	0.655
I_{max}	10.34	4.76	12.045	0.807

จากตารางที่ 11 พบว่า จากผลของการให้คะแนนทั้ง 3 แบบ เมื่อใช้เกณฑ์การคัดเลือกข้อสอบข้อต่อ ไปจากคลังข้อสอบด้วยวิธีการแตกต่างกัน ดังนี้

จำนวนข้อสอบที่ใช้วิธีให้คะแนนแบบ Multiple Response Method (MR) ที่ใช้เกณฑ์การคัดเลือกข้อสอบข้อต่อ ไปจากคลังข้อสอบด้วยวิธีการจับคู่ระหว่างค่าความยากง่ายของข้อสอบ (b) สอดคล้องกับค่าความสามารถที่ประมาณได้ของผู้สอบ ($\hat{\theta}$) มีค่าเฉลี่ยของการใช้ข้อสอบสูงสุดคือ 17.59 ค่าเบี่ยงเบนมาตรฐาน 5.78 และวิธีให้คะแนนแบบ Omit Multiple True-False Method (OMTF) ที่ใช้เกณฑ์การคัดเลือกข้อสอบข้อต่อ ไปจากคลังข้อสอบด้วยวิธีการใช้ข้อสอบที่มีค่าสารสนเทศสูงสุด (Maximum item information, I_{max}) มีค่าเฉลี่ยการใช้จำนวนข้อสอบต่ำที่สุดคือ 10.34 และค่าเบี่ยงเบนมาตรฐาน 4.76

ค่าฟังก์ชันสารสนเทศของชุดข้อสอบที่ใช้การให้คะแนนแบบ Omit Multiple True-False Method (OMTF) ที่ใช้เกณฑ์การคัดเลือกข้อสอบข้อต่อ ไปจากคลังข้อสอบด้วยวิธีการใช้ข้อสอบที่มีค่าสารสนเทศสูงสุด (Maximum item information, I_{max}) มีค่าเฉลี่ยของฟังก์ชันสารสนเทศสูงที่สุด คือ 12.045 ค่าเบี่ยงเบนมาตรฐาน 0.807 และวิธีให้คะแนนแบบ Multiple Response Method (MR) ที่ใช้เกณฑ์การคัดเลือกข้อสอบข้อต่อ ไปจากคลังข้อสอบด้วยวิธีการจับคู่ระหว่างค่าความยากง่ายของข้อสอบ (b) สอดคล้องกับค่าความสามารถที่ประมาณได้ของผู้สอบ ($\hat{\theta}$) มีค่าเฉลี่ยฟังก์ชันสารสนเทศต่ำที่สุด คือ 11.462 ค่าเบี่ยงเบนมาตรฐาน 0.292

เป็นที่น่าสังเกตว่า การให้คะแนนแบบ Omit Multiple True-False Method (OMTF) ที่ใช้เกณฑ์การคัดเลือกข้อสอบที่มีค่าสารสนเทศสูงสุด (maximum item information, I_{max}) จะใช้จำนวนข้อสอบและให้ค่าฟังก์ชันสารสนเทศของชุดข้อสอบสูงที่สุด

ตารางที่ 12 ค่าเฉลี่ย และค่าเบี่ยงเบนมาตรฐานของผลการให้คะแนน 3 แบบ ที่ใช้ในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ จำแนกตามวิธีการประมาณค่าความสามารถของผู้สอบ

การให้คะแนน		จำนวนข้อสอบ		ฟังก์ชันสารสนเทศ	
		\bar{X}	SD	\bar{X}	SD
MR	MLE	12.53	6.2388	11.5214	0.3361
	Bay	16.32	8.693	11.4223	0.2469
MTF	MLE	12.35	5.9557	11.4419	0.2817
	Bay	14.49	7.0149	11.4428	0.2758
OMTF	MLE	13.28	7.6622	11.4806	0.3074
	Bay	14.30	6.1118	11.8507	0.5037

จากตารางที่ 12 พบว่า จากผลของการให้คะแนนทั้ง 3 แบบ เมื่อใช้วิธีการประมาณค่าความสามารถของผู้สอบ ดังนี้

จำนวนข้อสอบที่ใช้การให้คะแนนแบบ Multiple Response Method (MR) ที่ใช้วิธีการประมาณค่าความสามารถของผู้สอบด้วยวิธีของเบส์ที่ปรับใหม่ (Bayesian updating) มีค่าเฉลี่ยจำนวนข้อสอบที่ใช้สอบสูงสุด คือ 16.32 ค่าเบี่ยงเบนมาตรฐาน 8.693 และการให้คะแนนแบบ Multiple True-False Method (MTF) ที่ใช้วิธีการประมาณค่าความสามารถของผู้สอบด้วยวิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (Conditional maximum likelihood) คือ 12.35 ค่าเบี่ยงเบนมาตรฐาน 5.9557

ค่าฟังก์ชันสารสนเทศของชุดข้อสอบที่ใช้การให้คะแนนแบบ Omit Multiple True-False Method (OMTF) ที่ใช้วิธีการประมาณค่าความสามารถของผู้สอบด้วยวิธีของเบส์ที่ปรับใหม่ (Bayesian Updating) มีค่าฟังก์ชันสารสนเทศเฉลี่ยของชุดข้อสอบสูงที่สุด คือ 11.8507 ค่าเบี่ยงเบนมาตรฐาน 0.5037

ตารางที่ 13 ค่าเฉลี่ย และค่าเบี่ยงเบนมาตรฐานของผลการให้คะแนน 3 วิธี ที่ใช้ในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ จำแนกตามเกณฑ์ยุติการทดสอบ

การให้คะแนน	จำนวนข้อสอบ		ฟังก์ชันสารสนเทศ		
	\bar{X}	SD	\bar{X}	SD	
MR	SEE ≤ .30	16.22	8.4682	11.3701	0.2370
	SEE ≤ .40	13.31	5.8653	11.9572	0.7488
	SEE ≤ .50	12.15	6.7749	11.9910	0.5106
MTF	SEE ≤ .30	16.54	8.5861	11.4705	0.2501
	SEE ≤ .40	13.27	6.2098	11.4790	0.3259
	SEE ≤ .50	15.04	7.7980	12.0302	0.8752
OMTF	SEE ≤ .30	16.40	5.8998	12.0607	0.8185
	SEE ≤ .40	13.37	5.1596	12.0446	0.7257
	SEE ≤ .50	12.07	7.3632	12.1379	0.8579

จากตารางที่ 13 พบว่า ผลของการให้คะแนนทั้ง 3 แบบ เมื่อใช้เกณฑ์ยุติการทดสอบ โดยใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า (Standard Error of Estimation: *SEE*) ดังนี้ จำนวนข้อสอบที่ใช้การให้คะแนนแบบ Multiple True-False Method (MTF) ที่ใช้เกณฑ์ยุติการทดสอบ โดยใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า 0.30 มีค่าเฉลี่ยของจำนวนข้อสอบสูงที่สุด คือ 16.54 ค่าเบี่ยงเบนมาตรฐาน 8.5861 และการให้คะแนนแบบ Omit Multiple True-False Method (OMTF) ที่ใช้เกณฑ์ยุติการทดสอบ โดยใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า 0.50 มีค่าเฉลี่ยของจำนวนข้อสอบต่ำที่สุด คือ 12.07 ค่าเบี่ยงเบนมาตรฐาน 7.3632

ค่าฟังก์ชันสารสนเทศของชุดข้อสอบที่ใช้การให้คะแนนแบบ Omit Multiple True-False Method (OMTF) ที่ใช้เกณฑ์ยุติการทดสอบ โดยใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า 0.50 มีค่าฟังก์ชันสารสนเทศของชุดข้อสอบเฉลี่ยสูงที่สุด คือ 12.1379 ค่าเบี่ยงเบนมาตรฐาน 0.8579 และการให้คะแนนแบบ Multiple Response Method (MR) ที่ใช้เกณฑ์ยุติการทดสอบ โดยใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า 0.30 มีค่าฟังก์ชันสารสนเทศของชุดข้อสอบต่ำที่สุด คือ 11.3701 ค่าเบี่ยงเบนมาตรฐาน 0.2370

ตารางที่ 14 ค่าเฉลี่ย และค่าเบี่ยงเบนมาตรฐานของผลการให้คะแนน 3 วิธี ที่ใช้ในการทดสอบ
แบบปรับเหมาะด้วยคอมพิวเตอร์ จำแนกตามความสามารถทางคณิตศาสตร์ของผู้เข้าสอบ

การให้คะแนน		จำนวนข้อสอบ		ฟังก์ชันสารสนเทศ	
		\bar{X}	SD	\bar{X}	SD
MR	สูง	11.19	6.6125	11.8622	0.6737
	กลาง	13.25	5.5005	11.6335	0.4493
	ต่ำ	18.02	9.3765	11.6698	0.4608
MTF	สูง	11.21	6.5828	11.9197	0.5879
	กลาง	13.21	7.6256	11.8273	0.5469
	ต่ำ	14.10	5.3671	11.8801	0.7003
OMTF	สูง	11.09	5.7075	12.1515	0.8617
	กลาง	11.46	5.1748	11.9338	0.6208
	ต่ำ	13.22	6.3245	11.8080	0.5179

จากตารางที่ 14 พบว่า จากผลของการให้คะแนนทั้ง 3 แบบ เมื่อผู้สอบมีความสามารถทางคณิตศาสตร์สูง ปานกลาง และต่ำ ดังนี้

จำนวนข้อสอบที่ใช้การให้คะแนนแบบ Multiple Response Method (MR) ของผู้สอบที่มีความสามารถทางคณิตศาสตร์ต่ำ มีค่าเฉลี่ยของการใช้ข้อสอบสูงที่สุด คือ 18.02 ค่าเบี่ยงเบนมาตรฐาน 9.3765 และการให้คะแนนแบบ Omit Multiple True-False Method (OMTF) ของผู้สอบที่มีความสามารถทางคณิตศาสตร์สูง มีค่าเฉลี่ยของการใช้ข้อสอบต่ำที่สุด คือ 11.09 ค่าเบี่ยงเบนมาตรฐาน 5.7075

ค่าฟังก์ชันสารสนเทศของชุดข้อสอบ ที่ใช้การให้คะแนนแบบ Omit Multiple True-False Method (OMTF) ของผู้สอบที่มีความสามารถทางคณิตศาสตร์สูง มีค่าฟังก์ชันสารสนเทศเฉลี่ยสูงที่สุด คือ 12.1515 ค่าเบี่ยงเบนมาตรฐาน 0.8617 และการให้คะแนนแบบ Multiple Response Method (MR) ของผู้สอบที่มีความสามารถทางคณิตศาสตร์ปานกลาง มีค่าฟังก์ชันสารสนเทศของชุดข้อสอบเฉลี่ยต่ำที่สุด คือ 11.6335 ค่าเบี่ยงเบนมาตรฐาน 0.4493

บทที่ 5

สรุป อภิปรายผลและข้อเสนอแนะ

การวิจัยเพื่อศึกษาคุณภาพของการวัดในด้านความตรง ความเที่ยง ความคลาดเคลื่อนมาตรฐานในการวัด และเปรียบเทียบประสิทธิภาพการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ในด้านจำนวนข้อสอบ และค่าฟังก์ชันสารสนเทศของชุดข้อสอบ จากผลของการให้คะแนนที่แตกต่างกัน 3 แบบ เมื่อพิจารณาตามองค์ประกอบสำคัญของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ได้แก่ เกณฑ์การคัดเลือกข้อสอบ การประมาณค่าความสามารถของผู้สอบ การยุติการทดสอบและความสามารถของผู้สอบ โดยดำเนินการวิจัยเป็น 2 ระยะ ได้แก่ ระยะแรกคือการพัฒนาคลังข้อสอบวิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษาตอนปลาย และระยะที่สองเป็นการศึกษาคุณภาพของการวัดและเปรียบเทียบประสิทธิภาพการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ แบ่งเป็น 3 ขั้นตอน ดังนี้

ขั้นตอนที่ 1 การตรวจสอบคุณภาพของการวัดในด้านความตรง ความเที่ยง ความคลาดเคลื่อนมาตรฐานในการวัด และวิเคราะห์หาค่าพารามิเตอร์ของข้อสอบ เพื่อคัดเลือกข้อสอบเข้าคลังข้อสอบไว้ใช้สำหรับการทดสอบปรับเหมาะด้วยคอมพิวเตอร์

ขั้นตอนที่ 2 การพัฒนาโปรแกรมการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ เป็นการออกแบบโปรแกรมคอมพิวเตอร์ และเขียนโปรแกรมด้วยภาษา ASP.NET โดยนำข้อสอบจากคลังข้อสอบในขั้นตอนที่ 1 มาเป็นฐานข้อมูลสำหรับใช้ในการพัฒนาโปรแกรม จากนั้นให้ผู้เชี่ยวชาญพิจารณาตรวจสอบความถูกต้องเหมาะสมของโปรแกรม ปรับปรุงตามคำแนะนำของผู้เชี่ยวชาญ และจัดทำเป็นชุดโปรแกรมเพื่อนำไปใช้ในการศึกษาวิจัย

ขั้นตอนที่ 3 การใช้โปรแกรมการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ โดยนำโปรแกรมจากขั้นตอนที่ 2 ไปให้นักเรียนทดลองใช้งาน รวมทั้งให้ข้อเสนอแนะเกี่ยวกับการใช้งานโปรแกรม แล้วนำมาปรับปรุงตามคำแนะนำของนักเรียน เพื่อศึกษาประสิทธิภาพการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ในด้านจำนวนข้อสอบ และค่าฟังก์ชันสารสนเทศของชุดข้อสอบ จากผลการให้คะแนนที่แตกต่างกัน 3 แบบ

สรุปผลการวิจัย

จากผลการวิเคราะห์ห้ข้อมูล พบว่า ผลการให้คะแนนแบบ OMTF มีค่าความตรงและความเที่ยงสูงสุด คือ 0.7239, 0.7757 รองลงมาได้แก่ การให้คะแนนแบบ MTF เท่ากับ 0.7202, 0.7750 และการให้คะแนนแบบ MR เท่ากับ 0.7233, 0.7716 และความคลาดเคลื่อนมาตรฐานในการวัดผลการให้คะแนนแบบ OMTF มีค่าต่ำสุด เท่ากับ 0.2305 รองลงมาได้แก่ การให้คะแนนแบบ MR เท่ากับ 0.2326 และการให้คะแนนแบบ MTF เท่ากับ 0.4609 สำหรับประสิทธิภาพของการทดสอบเมื่อพิจารณาตามองค์ประกอบสำคัญของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ จากผลการให้คะแนนทั้ง 3 แบบ พบว่า

1. การคัดเลือกข้อสอบจากคลังข้อสอบในการทดสอบแบบเหมาะด้วยคอมพิวเตอร์ นั้นปรากฏว่า การคัดเลือกข้อสอบจากคลังข้อสอบด้วยการใช้ค่าสารสนเทศสูงสุดของข้อสอบ จะใช้ข้อสอบน้อยกว่าการคัดเลือกข้อสอบด้วยวิธีอื่น ๆ และให้ค่าฟังก์ชันสารสนเทศสูงสุด จากผลของการให้คะแนนทั้ง 3 แบบ ซึ่งจำนวนข้อสอบที่ใช้การให้คะแนนแบบ Multiple Response Method (MR) ที่ใช้เกณฑ์การคัดเลือกข้อสอบข้อต่อไปจากคลังข้อสอบด้วยวิธีการจับคู่ระหว่างค่าความยากง่ายของข้อสอบ (b) สอดคล้องกับค่าความสามารถที่ประมาณได้ของผู้สอบ ($\hat{\theta}$) มีค่าเฉลี่ยของการใช้ข้อสอบสูงสุด คือ 17.59 ค่าเบี่ยงเบนมาตรฐาน 5.78 และวิธีให้คะแนนแบบ Omit Multiple True-False Method (OMTF) ที่ใช้เกณฑ์การคัดเลือกข้อสอบข้อต่อไปจากคลังข้อสอบด้วยวิธีการใช้ข้อสอบที่มีค่าสารสนเทศสูงสุด (Maximum item information, I_{max}) มีค่าเฉลี่ยการใช้จำนวนข้อสอบต่ำที่สุด คือ 10.34 และค่าเบี่ยงเบนมาตรฐาน 4.76 สำหรับค่าฟังก์ชันสารสนเทศของชุดข้อสอบที่ใช้การให้คะแนนแบบ Omit Multiple True-False Method (OMTF) ที่ใช้เกณฑ์การคัดเลือกข้อสอบข้อต่อไปจากคลังข้อสอบด้วยวิธีการใช้ข้อสอบที่มีค่าสารสนเทศสูงสุด (Maximum item information, I_{max}) มีค่าเฉลี่ยของฟังก์ชันสารสนเทศสูงที่สุด คือ 12.045 ค่าเบี่ยงเบนมาตรฐาน 0.807 และการให้คะแนนแบบ Multiple Response Method (MR) ที่ใช้เกณฑ์การคัดเลือกข้อสอบข้อต่อไปจากคลังข้อสอบด้วยวิธีการจับคู่ระหว่างค่าความยากง่ายของข้อสอบ (b) สอดคล้องกับค่าความสามารถที่ประมาณได้ของผู้สอบ ($\hat{\theta}$) มีค่าเฉลี่ยฟังก์ชันสารสนเทศต่ำที่สุด คือ 11.462 ค่าเบี่ยงเบนมาตรฐาน 0.292 และค่าฟังก์ชันสารสนเทศของชุดข้อสอบที่ใช้การให้คะแนนแบบ Omit Multiple True-False Method (OMTF) ที่ใช้เกณฑ์การคัดเลือกข้อสอบข้อต่อไปจากคลังข้อสอบด้วยวิธีการใช้ข้อสอบที่มีค่าสารสนเทศสูงสุด (maximum item information, I_{max}) มีค่าเฉลี่ยของฟังก์ชันสารสนเทศสูงที่สุด คือ 12.045 ค่าเบี่ยงเบนมาตรฐาน 0.807 และการให้คะแนนแบบ Multiple Response Method (MR) ที่ใช้เกณฑ์การคัดเลือกข้อสอบข้อต่อไป

จากคลังข้อสอบด้วยวิธีการจับคู่ระหว่างค่าความยากง่ายของข้อสอบ (b) สอดคล้องกับค่าความสามารถที่ประมาณได้ของผู้สอบ ($\hat{\theta}$) มีค่าเฉลี่ยฟังก์ชันสารสนเทศต่ำที่สุด คือ 11.462 ค่าเบี่ยงเบนมาตรฐาน 0.292

2. การประมาณค่าความสามารถของผู้สอบในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ทำให้จำนวนข้อสอบที่ใช้ในการทดสอบต่างกัน วิธีของเบส์ใช้ข้อสอบมากกว่าวิธีความเป็นไปได้สูงสุด และให้ค่าฟังก์ชันสารสนเทศใกล้เคียงกัน ซึ่งจำนวนข้อสอบที่ใช้การให้คะแนนแบบ Multiple Response Method (MR) ที่ใช้วิธีการประมาณค่าความสามารถของผู้สอบด้วยวิธีของเบส์ที่ปรับใหม่ (Bayesian Updating) มีค่าเฉลี่ยจำนวนข้อสอบที่ใช้สอบสูงสุด คือ 16.32 ค่าเบี่ยงเบนมาตรฐาน 8.693 และการให้คะแนนแบบ Multiple True-False Method (MTF) ที่ใช้วิธีการประมาณค่าความสามารถของผู้สอบด้วยวิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (Conditional maximum likelihood) คือ 12.35 ค่าเบี่ยงเบนมาตรฐาน 5.9557 สำหรับค่าฟังก์ชันสารสนเทศของชุดข้อสอบที่ใช้การให้คะแนนแบบ Omit Multiple True-False Method (OMTF) ที่ใช้วิธีการประมาณค่าความสามารถของผู้สอบด้วยวิธีของเบส์ที่ปรับใหม่ (Bayesian Updating) มีค่าฟังก์ชันสารสนเทศเฉลี่ยของชุดข้อสอบสูงสุด คือ 11.8507 ค่าเบี่ยงเบนมาตรฐาน 0.5037

3. การยุติการทดสอบในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ทำให้จำนวนข้อสอบที่ใช้ในการทดสอบต่างกัน ค่าความคลาดเคลื่อนมาตรฐานในการวัดเท่ากับหรือน้อยกว่า 0.30 ใช้ข้อสอบมากกว่าเกณฑ์อื่น ซึ่งจำนวนข้อสอบที่ใช้การให้คะแนนแบบ Multiple True-False Method (MTF) ที่ใช้เกณฑ์ยุติการทดสอบ โดยใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า 0.30 มีค่าเฉลี่ยของจำนวนข้อสอบสูงสุด คือ 16.54 ค่าเบี่ยงเบนมาตรฐาน 8.5861 และการให้คะแนนแบบ Omit Multiple True-False Method (OMTF) ที่ใช้เกณฑ์ยุติการทดสอบ โดยใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า 0.50 มีค่าเฉลี่ยของจำนวนข้อสอบต่ำที่สุด คือ 12.07 ค่าเบี่ยงเบนมาตรฐาน 7.3632 สำหรับค่าฟังก์ชันสารสนเทศของชุดข้อสอบที่ใช้การให้คะแนนแบบ Omit Multiple True-False Method (OMTF) ที่ใช้เกณฑ์ยุติการทดสอบ โดยใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า 0.50 มีค่าฟังก์ชันสารสนเทศของชุดข้อสอบเฉลี่ยสูงสุด คือ 12.1379 ค่าเบี่ยงเบนมาตรฐาน 0.8579 และการให้คะแนนแบบ Multiple Response Method (MR) ที่ใช้เกณฑ์ยุติการทดสอบ โดยใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า 0.30 มีค่าฟังก์ชันสารสนเทศของชุดข้อสอบต่ำที่สุด คือ 11.3701 ค่าเบี่ยงเบนมาตรฐาน 0.2370

4. ความสามารถทางคณิตศาสตร์ของผู้สอบในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ทำให้จำนวนข้อสอบที่ใช้ในการทดสอบต่างกัน ผู้สอบที่มีความสามารถสูงจะใช้

ข้อสอบน้อยกว่าผู้ที่มีความสามารถต่ำ ซึ่งจำนวนข้อสอบที่ใช้การให้คะแนนแบบ Multiple Response Method (MR) ของผู้สอบที่มีความสามารถทางคณิตศาสตร์ต่ำ มีค่าเฉลี่ยของการใช้ข้อสอบสูงที่สุด คือ 18.02 ค่าเบี่ยงเบนมาตรฐาน 9.3765 และการให้คะแนนแบบ Omit Multiple True-False Method (OMTF) ของผู้สอบที่มีความสามารถทางคณิตศาสตร์สูง มีค่าเฉลี่ยของการใช้ข้อสอบต่ำที่สุด คือ 11.09 ค่าเบี่ยงเบนมาตรฐาน 5.7075 สำหรับค่าฟังก์ชันสารสนเทศของชุดข้อสอบที่ใช้การให้คะแนนแบบ Omit Multiple True-False Method (OMTF) ของผู้สอบที่มีความสามารถทางคณิตศาสตร์สูง มีค่าฟังก์ชันสารสนเทศเฉลี่ยสูงที่สุด คือ 12.1515 ค่าเบี่ยงเบนมาตรฐาน 0.8617 และการให้คะแนนแบบ Multiple Response Method (MR) ของผู้สอบที่มีความสามารถทางคณิตศาสตร์ปานกลาง มีค่าฟังก์ชันสารสนเทศของชุดข้อสอบเฉลี่ยต่ำที่สุด คือ 11.6335 ค่าเบี่ยงเบนมาตรฐาน 0.4493

อภิปรายผล

ข้อค้นพบที่ได้จากการวิจัยในครั้งนี้ บางประเด็นก็ตรงตามสมมติฐานที่ผู้วิจัยกำหนดไว้บางประเด็นก็ไม่ตรงตามสมมติฐานที่กำหนดไว้ ซึ่งผู้วิจัยได้ตั้งข้อสังเกตไว้ในการอภิปรายผลการวิจัยแยกเป็นประเด็นสำคัญ ๆ สอดคล้องกับวัตถุประสงค์ของการวิจัย ดังนี้

1. การคัดเลือกข้อสอบจากคลังข้อสอบในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ปรากฏว่า การให้คะแนนแบบ Omit Multiple True-False Method (OMTF) ที่ใช้เกณฑ์การคัดเลือกข้อสอบที่มีค่าสารสนเทศสูงสุด (Maximum item information, I_{\max}) จะใช้จำนวนข้อสอบและให้ค่าฟังก์ชันสารสนเทศของชุดข้อสอบสูงที่สุด ทั้งนี้เพราะว่า การคัดเลือกข้อสอบจากคลังข้อสอบต่างกันในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ นั้นจำนวนข้อสอบและค่าฟังก์ชันสารสนเทศของชุดข้อสอบก็ต่างกัน ผลการวิจัยนี้ตรงตามสมมติฐานที่กำหนดไว้ และสอดคล้องกับผลการวิจัยของ รังสรรค์ มณีเล็ก (2540) ที่ว่าข้อสอบที่ถูกเลือกขึ้นมาใช้แต่ละวิธีนี้ จะมีสารสนเทศของข้อสอบแตกต่างกันออกไป ซึ่งค่าสารสนเทศของข้อสอบดังกล่าวจะส่งผลต่อค่าความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ ตามสูตร $SEE = 1/\sqrt{|\{\theta\}|}$ ซึ่ง $|\{\theta\}|$ (สารสนเทศของชุดข้อสอบ) ก็เป็นผลรวมของ $|\{\theta, u\}|$ (สารสนเทศของข้อสอบ) ดังนั้นจึงทำให้วิธีการคัดเลือกข้อสอบแต่ละวิธีผู้เข้าหาความคลาดเคลื่อนในการประมาณค่าที่กำหนดเป็นเกณฑ์ยุติการทดสอบเร็วช้าต่างกัน ซึ่งการเร็วช้าต่างกันนี้ก็คือ การใช้ข้อสอบมากน้อยต่างกันในการทดสอบนั่นเอง

2. การประมาณค่าความสามารถของผู้สอบในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ซึ่งสอดคล้องกับผลการวิจัยของกีฟฟอร์ด และสวามินาทาน (Gifford & Swaminathan, 1990, p. 33) ที่พบว่า ถ้ากลุ่มตัวอย่างขนาดเล็กและข้อสอบน้อยข้อ การประมาณค่าความสามารถของวิธี Joint Bayesian estimate จะมีความถูกต้องมากกว่า Join maximum likelihood และข้อค้นพบของโฮ (Ho, 1989) ที่พบว่า วิธีการประมาณค่าความสามารถวิธีของเบส มีความเชื่อมั่นและมีประสิทธิภาพสูงกว่าวิธีความเป็นไปได้สูงสุด ซึ่งมักไม่ค่อยคงเส้นคงวานัก เมื่อใช้วิธีการประมาณค่าความสามารถของผู้สอบต่างกัน ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ ปรากฏว่า จำนวนข้อสอบ และค่าฟังก์ชันสารสนเทศของชุดข้อสอบต่างกัน ผลการวิจัยนี้สอดคล้องกับสมมติฐานที่กำหนดไว้ นั่นคือ เมื่อใช้วิธีการประมาณค่าความสามารถของผู้สอบต่างกันแล้ว จำนวนข้อสอบ และ ค่าฟังก์ชันสารสนเทศก็จะต่างกัน ไปด้วย ซึ่งวิธีของเบสจะใช้ข้อสอบมากกว่าวิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข ซึ่งต่างจากข้อค้นพบของต่าย เชียงฉี ที่พบว่า วิธีการประมาณค่าความสามารถของเบสจะใช้ข้อสอบน้อยกว่าวิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (ต่าย เชียงฉี, 2530, หน้า 122)

3. การยุติการทดสอบในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ จำนวนข้อสอบที่ใช้การให้คะแนนแบบ Multiple True-False Method (MTF) ที่ใช้เกณฑ์ยุติการทดสอบโดยใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า 0.30 มีค่าเฉลี่ยของจำนวนข้อสอบสูงที่สุด คือ 16.54 ค่าเบี่ยงเบนมาตรฐาน 8.5861 และการให้คะแนนแบบ Omit Multiple True-False Method (OMTF) ที่ใช้เกณฑ์ยุติการทดสอบ โดยใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า 0.50 มีค่าเฉลี่ยของจำนวนข้อสอบต่ำที่สุด คือ 12.07 ค่าเบี่ยงเบนมาตรฐาน 7.3632 และค่าฟังก์ชันสารสนเทศของชุดข้อสอบที่ใช้การให้คะแนนแบบ Omit Multiple True-False Method (OMTF) ที่ใช้เกณฑ์ยุติการทดสอบโดยใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า 0.50 มีค่าฟังก์ชันสารสนเทศของชุดข้อสอบเฉลี่ยสูงที่สุด คือ 12.1379 ค่าเบี่ยงเบนมาตรฐาน 0.8579 และการให้คะแนนแบบ Multiple Response Method (MR) ที่ใช้เกณฑ์ยุติการทดสอบโดยใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า 0.30 มีค่าฟังก์ชันสารสนเทศของชุดข้อสอบต่ำที่สุด คือ 11.3701 ค่าเบี่ยงเบนมาตรฐาน 0.2370 ดังนั้นการยุติการทดสอบ ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ ปรากฏว่า จำนวนข้อสอบ และค่าฟังก์ชันสารสนเทศของชุดข้อสอบต่างกัน ผลการวิจัยนี้สอดคล้องกับสมมติฐานที่กำหนดไว้ ทั้งนี้ เนื่องมาจากเมื่อจำนวนข้อสอบ และค่าฟังก์ชันสารสนเทศของชุดข้อสอบ เปลี่ยนไป ค่าความถูกต้องในการประมาณค่าความสามารถของผู้สอบ ซึ่งก็รวมไปถึงค่าความคลาดเคลื่อนในการประมาณค่า จะเกิดการเปลี่ยนแปลง กล่าวคือ เมื่อใช้ข้อสอบมากขึ้น ความเชื่อมั่นก็จะมีค่าสูงขึ้น ก็คือ

ความคลาดเคลื่อนมาตรฐานก็จะมีค่าลดลงดังนั้น หากกำหนดค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าต่างกัน จำนวนข้อสอบ และค่าฟังก์ชันสารสนเทศของชุดข้อสอบก็จะต่างกันไป

4. ความสามารถทางคณิตศาสตร์ของผู้สอบในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ จากผลการให้คะแนนที่แตกต่างกัน ซึ่งผู้วิจัยกำหนดไว้ว่า เมื่อผู้สอบมีความสามารถทางคณิตศาสตร์ต่างกัน จะไม่ทำให้ค่าความตรงของการทดสอบต่างกัน ทั้งอาจเป็นเพราะค่าความสามารถของผู้สอบนั้นจะส่งผลต่อความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง และค่าความน่าจะเป็นดังกล่าวจะส่งผลต่อค่าสารสนเทศของข้อสอบ และค่าสารสนเทศของชุดข้อสอบสูงตามไปด้วย กรณีเช่นนี้จะส่งผลทำให้ค่าความคลาดเคลื่อนมาตรฐานในการวัดความสามารถของผู้สอบมีค่าต่ำลง นั่นคือ มีความถูกต้องในการวัดสูง ในทางกลับกันถ้าความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องมีค่าต่ำก็จะส่งผลให้ค่าสารสนเทศของข้อสอบ และสารสนเทศของชุดข้อสอบมีค่าต่ำ และจะส่งผลให้ค่าความคลาดเคลื่อนมาตรฐานในการวัดมีค่าสูง นั่นคือ มีความถูกต้องในการวัดต่ำนั่นเอง และค่าความถูกต้องในการวัดนั้น เกี่ยวข้องโดยตรงกับค่าความตรงของชุดข้อสอบ ดังนั้น จึงเป็นไปได้ที่เมื่อทดสอบกับผู้สอบที่มีความสามารถสูงจะมีความตรงสูงกว่าเมื่อทดสอบกับผู้ที่มีความสามารถต่ำกว่า อีกประการหนึ่งที่น่าจะเป็นไปได้ก็คือ ข้อสอบในคลังข้อสอบนั้น หากพิจารณาโครงสร้างสารสนเทศของชุดข้อสอบแล้ว จะพบว่า เป็นข้อสอบที่เหมาะสมกับผู้สอบที่มีความสามารถปานกลางค่อนข้างสูง ซึ่งหากกำหนดข้อสอบให้มีความเหมาะสมหรือตรงกับความสามารถของผู้สอบแล้ว ก็จะทำให้ค่าของความคลาดเคลื่อนมาตรฐานในการวัดต่ำลง (Haladyna & Roid, 1983) เป็นที่น่าสังเกตว่า ความสามารถของผู้สอบนั้นจะไม่สูงมากนัก ทั้งนี้อาจเป็นเพราะเมื่อแบ่งกลุ่มผู้สอบออกไป 3 กลุ่ม คือ ความสามารถสูง กลาง และต่ำ ทำให้พิสัยของค่าความสามารถของผู้สอบที่จะนำมาหาค่าสหสัมพันธ์แคบลง พิสัยค่าความสามารถที่แคบลงนี้จะทำให้ความแปรปรวนของค่าความสามารถมีค่าต่ำ ผลการวิจัยนี้สอดคล้องกับสมมติฐานที่กำหนดไว้ นั่นคือ ผู้สอบที่มีความสามารถทางคณิตศาสตร์ต่างกัน จะใช้ข้อสอบในการทดสอบไม่เท่ากัน ผู้สอบที่มีความสามารถสูง จะใช้ข้อสอบน้อยกว่าผู้สอบที่มีความสามารถปานกลาง และต่ำ กล่าวคือ ผู้สอบที่มีความสามารถสูงก็จะมีค่าความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง ซึ่งจะส่งผลทำให้ค่าสารสนเทศของข้อสอบและค่าสารสนเทศของชุดข้อสอบสูงตามไปด้วย แต่จะส่งผลทำให้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่ามีค่าต่ำลง และในการวิจัยครั้งนี้ผู้วิจัยได้ใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าเป็นเกณฑ์ยุติการทดสอบ เมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าจากผู้สอบที่มีความสามารถสูงมีค่าต่ำลงเร็วกว่าผู้สอบที่มีความสามารถต่ำ นั่นแสดงว่า ผู้สอบที่มีความสามารถสูงใช้ข้อสอบน้อยข้อกว่าผู้สอบที่มีความสามารถต่ำ

ข้อเสนอแนะ

จากผลการวิจัยที่ได้ ผู้วิจัยมีข้อเสนอแนะดังนี้

1. การนำผลการวิจัยไปใช้ สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์นี้ เป็นการทดสอบรายบุคคลที่ค่อนข้างมีความถูกต้อง และมีความตรงเชิงสภาพที่แท้จริงของนักเรียน ตามโดเมนที่ถูกวัด การทดสอบแต่ละครั้งไม่จำเป็นต้องทำพร้อมกันทุกคน ดังนั้น ปริมาณของอุปกรณ์ทางคอมพิวเตอร์ จึงไม่ใช่ปัญหาสำหรับการทดสอบลักษณะนี้ เมื่อผู้สอบต้องการที่จะทดสอบเมื่อไรก็ตามดำเนินการเข้าระบบ ลงทะเบียน และทดสอบด้วยตนเองผลการทดสอบแต่ละครั้งจะถูกเก็บไว้ในแฟ้มฐานข้อมูลของนักเรียนแต่ละคน ซึ่งจะเป็นแหล่งข้อมูลที่จะช่วยให้ครูผู้สอนสามารถติดตามความก้าวหน้าทางการเรียน นอกจากนั้นแล้วยังสามารถนำข้อมูลไปใช้ในการประเมินผลนักเรียนตามสภาพและเหตุการณ์ที่แท้จริงได้อย่างมีประสิทธิภาพอีกด้วย
2. การวิจัยในครั้งต่อไป ควรมีการศึกษาเกี่ยวกับการข้ามข้อสอบเป็นบางข้อ การทบทวนข้อสอบที่ได้ทำไปแล้ว การเลือกข้อสอบในคลังข้อสอบได้ด้วยตนเอง การสร้างความพอดีของเนื้อหา ในรูปแบบต่าง ๆ และนอกจากนี้ควรจะได้มีการศึกษาเกี่ยวกับรูปแบบของคลังข้อสอบ เช่น การแจกแจงของค่าพารามิเตอร์ของข้อสอบในคลังข้อสอบควรมีลักษณะเป็นโค้งปกติ หรือ มีจำนวนข้อสอบเท่ากันในแต่ละระดับความสามารถ หรือ ข้อสอบควรแจกแจงเป็นโค้งปกติ ในแต่ละระดับความสามารถ

บรรณานุกรม

- กมลทิพย์ อธิการยานันท์. (2530). การเปรียบเทียบคุณภาพของแบบสอบเลือกตอบชนิดคำตอบ ถูกตัวเดียว คำตอบถูกไม่จำกัด และคำตอบรวม. วิทยานิพนธ์ครุศาสตรมหาบัณฑิต, สาขาวิชาวัดและประเมินผลทางการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.
- กาญจนา ศิริวัฒนพงษ์. (2520). การศึกษาเปรียบเทียบการตอบและการตรวจให้คะแนน แบบทดสอบปรนัยชนิดเลือกตอบที่มีลักษณะแตกต่างกัน. ปริญญาานิพนธ์การศึกษามหาบัณฑิต, มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร.
- กณิศ ไช้มุกต์. (2533). การพัฒนาเกณฑ์ในการคัดเลือกข้อสอบที่เหมาะสมกับระดับความสามารถของผู้สอบ. วิทยานิพนธ์ดุขฎิบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.
- งามนิตย์ ธาตุทอง และปรีชา เกรือวัลย์. (2527). การให้คะแนนตนเองในการสอบแบบทดสอบชนิดเฟลกซิลเวล. ขอนแก่น: มหาวิทยาลัยขอนแก่น.
- จินดา โตอนันต์. (2526). การเปรียบเทียบคุณลักษณะของแบบทดสอบสัมฤทธิ์ผลแบบเลือกตอบ ถูกผิดทุกตัวเลือกกับแบบเลือกตอบตัวเลือกเดียว. วิทยานิพนธ์ครุศาสตรมหาบัณฑิต, สาขาวิชาวิจัยการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.
- ฉวีวรรณ บุญมั่ง. (2541). การเปรียบเทียบความสัมพันธ์ระหว่างคะแนนความสามารถจากการวิเคราะห์ โดยทฤษฎีการตอบสนองข้อสอบกับคะแนนที่ได้จากการตอบและตรวจให้คะแนน 4 วิธี. วิทยานิพนธ์การศึกษามหาบัณฑิต, สาขาวิชาวัดผลการศึกษา, บัณฑิตวิทยาลัย, มหาวิทยาลัยมหาสารคาม.
- ชวาล แพรัตกุล. (2552). เทคนิคการวัดผล (พิมพ์ครั้งที่ 7). กรุงเทพฯ: มหาวิทยาลัยศรีนครินทรวิโรฒ.
- ชวาล แพรัตกุล. (ม.ป.ป.). เทคนิคการเขียนคำถามเลือกตอบ. กรุงเทพฯ: กิ่งจันทร์การพิมพ์.
- ต่าย เชียงฉิ. (2530). การวิเคราะห์ข้อสอบ. ใน เอกสารการฝึกอบรมเชิงปฏิบัติการ เรื่องการวัดและประเมินผลการศึกษา. เชียงใหม่: ภาควิชาประเมินผลและวิจัยการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยเชียงใหม่.
- ต่าย เชียงฉิ. (2534). การศึกษาเปรียบเทียบประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบจากการทดสอบเทเลอร์รูปปริมาตรที่มีรูปแบบจำนวนชั้นและวิธีการให้คะแนนที่แตกต่างกันโดยใช้วิธีมอนติคาร์โล. ปริญญาานิพนธ์การศึกษาดุขฎิบัณฑิต, สาขาวิชาการวิจัยและพัฒนาหลักสูตร, บัณฑิตวิทยาลัย, มหาวิทยาลัยศรีนครินทรวิโรฒ.

- ธนวัฒน์ แสนสุข. (2538). *การใช้ไออาร์เอ็ม จีพีซีเอ็มและโมเดลโลจิสติกในการเปรียบเทียบฟังก์ชัน
สารสนเทศของแบบวัดที่มีการให้คะแนนต่างกัน*. วิทยานิพนธ์ครุศาสตรมหาบัณฑิต
สาขาวิชาการวัดและประเมินผลการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.
- บุญชม ศรีสะอาด. (2540). *การวิจัยทางการวัดผลและประเมินผล*. กรุงเทพฯ: สุวีริยาสาส์น.
- บุญชม ศรีสะอาด. (2541). *นวัตกรรมทางการศึกษา (พิมพ์ครั้งที่ 2)*. กรุงเทพฯ: ชมรมเด็ก.
- บุญชม ศรีสะอาด. (2543). *การวิจัยทางการวัดผลและประเมินผล*. กรุงเทพฯ: สุวีริยาสาส์น.
- บุญชม ศรีสะอาด. (2553). *การวิจัยเบื้องต้น (พิมพ์ครั้งที่ 8)*. กรุงเทพฯ: สุวีริยาสาส์น.
- บุญเชิด ภิญโญอนันตพงษ์. (2527). *การทดสอบอิงเกณฑ์: แนวคิดและวิธีการ*. กรุงเทพฯ:
ไอเดียนสโตร์.
- ปรมิินทร์ อริเดช. (2547). *การเปรียบเทียบฟังก์ชันสารสนเทศในการใช้โมเดลโลจิสติก จีอาร์เอ็ม
และจีพีซีเอ็มของมาตรวัดเจตคติแบบลิเคิร์ตและมาตรวัดแบบตัวเลือกบังคับตอบ
ที่มีการให้คะแนนแบบสองค่าและแบบหลายค่า*. วิทยานิพนธ์การศึกษาดุขฎีบัณฑิต,
สาขาวิชาการทดสอบและวัดผลการศึกษา, บัณฑิตวิทยาลัย, มหาวิทยาลัยศรีนครินทร
วิโรฒ.
- พรทิพย์ ไชยโส. (2534). *การพัฒนาสูตรการให้คะแนนแบบสอบเลือกตอบสำหรับความรู้
บางส่วนของผู้ตอบ: การประยุกต์ใช้วิธีการของอาร์โนลด์ และวิธีการของแฮมดาน*.
วิทยานิพนธ์ครุศาสตรดุขฎีบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา,
บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.
- พินิจ อุไรรักษ์. (2533). *ผลของวิธีการให้คะแนนที่มีต่อคะแนนสอบ ค่าสัมประสิทธิ์ความเที่ยง
และค่าสัมประสิทธิ์ความเที่ยง*. วิทยานิพนธ์ครุศาสตรมหาบัณฑิต, การวัดและ
การประเมินผลการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.
- รังสรรค์ เล็กมณี. (2540). *ผลของตัวแปรบางตัวต่อความเที่ยงตรงเชิงสภาพและจำนวนข้อสอบ
ที่ใช้ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์*.
วิทยานิพนธ์ดุขฎีบัณฑิต, สาขาวิชาการทดสอบและวัดผลการศึกษา, บัณฑิตวิทยาลัย,
มหาวิทยาลัยศรีนครินทรวิโรฒ.
- ล้วน สายยศ และอังคณา สายยศ. (2539). *เทคนิคการวัดการเรียนรู้*. กรุงเทพฯ: สุวีริยาสาส์น.
- ศิริชัย กาญจนวาสี. (2538). *การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ
(Adaptive Testing)*. กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี. (2550). *ทฤษฎีการทดสอบแนวใหม่: MODERN TEST THEORY*. กรุงเทพฯ:
โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

- ส.วาสนา ประवालพุดักษ์. (2545). *หลักการและเทคนิคการประเมินทางการศึกษา*. กรุงเทพฯ: มหาวิทยาลัยศรีนครินทรวิโรฒ.
- สมจวน มานะก่อ. (2551). *การศึกษาเปรียบเทียบคุณภาพของแบบทดสอบเลือกตอบ เมื่อตรวจด้วยวิธีการให้คะแนนความรู้บางส่วนของแบบทดสอบวัดผลสัมฤทธิ์คณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3*. ปรียญานิพนธ์การศึกษามหาบัณฑิต, สาขาวิชาการวัดผลการศึกษา, บัณฑิตวิทยาลัย, มหาวิทยาลัยศรีนครินทรวิโรฒ.
- สมศักดิ์ ธิลา. (2539). *การพัฒนาระบบคลังข้อสอบเพื่อการเรียนการสอนด้วยไมโครคอมพิวเตอร์*. ปรียญานิพนธ์การศึกษาคณะศึกษาศาสตร์, สาขาวิชาการวิจัยและพัฒนาหลักสูตร, บัณฑิตวิทยาลัย, มหาวิทยาลัยศรีนครินทรวิโรฒ.
- สุนิสา จุ้ยม่วงศรี. (2546). *การศึกษาผลของการเทียบคะแนนข้อสอบที่ตรวจให้คะแนนแบบหลายค่า*. ปรียญานิพนธ์การศึกษาคณะศึกษาศาสตร์, สาขาวิชาการทดสอบและวัดผลการศึกษา, บัณฑิตวิทยาลัย, มหาวิทยาลัยศรีนครินทรวิโรฒ.
- สุพจน์ เกิดสุวรรณ. (2545). *การพัฒนาการวัดความรู้บางส่วนของผู้ตอบแบบสอบเลือกตอบ*. วิทยานิพนธ์ครุศาสตรศึกษาคณะศึกษาศาสตร์, สาขาวิชาการวัดและประเมินผลการศึกษา, คณะครุศาสตร์, จุฬาลงกรณ์มหาวิทยาลัย.
- สุพัฒน์ สุกมลสันต์. (2538). *การวิเคราะห์ข้อทดสอบแนวใหม่ด้วยคอมพิวเตอร์*. กรุงเทพฯ: อักษรการพิมพ์.
- สุรัชย์ มีชาญ. (2539). *การเปรียบเทียบความยากประจำชั้นและประสิทธิภาพในการประมาณค่าเจตคติของมาตรวัดเจตคติแบบลิเคอร์ท์ที่มีรูปแบบของการตอบจำนวนของลำดับชั้น และทิศทางของการเรียงลำดับชั้นแตกต่างกันด้วยพาเซี่ยลเครดิท โมเดลของมาสเตอร์*. ปรียญานิพนธ์การศึกษาคณะศึกษาศาสตร์, การทดสอบและวัดผลการศึกษา, บัณฑิตวิทยาลัย, มหาวิทยาลัยศรีนครินทรวิโรฒ.
- สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 32. (2557). *ข้อมูลสารสนเทศปีการศึกษา 2557 โรงเรียนมัธยมศึกษาจังหวัดบุรีรัมย์*. บุรีรัมย์: สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 32.
- อนันต์ ศรีโสภณ. (2532). *การทดลองรูปแบบใหม่ของข้อสอบแบบเลือกตอบ*. กรุงเทพฯ: มหาวิทยาลัยศรีนครินทรวิโรฒ.
- อนันต์ ศรีโสภณ. (2524). *การวัดและประเมินผลการศึกษา (พิมพ์ครั้งที่ 2)*. กรุงเทพฯ: ไทยวัฒนาพานิช.

- องอาจ นัยวัฒน์. (2535). *การศึกษาประสิทธิภาพของผลการประมาณค่าพารามิเตอร์ของพหุเชิงเส้น เครดิต โมเดล (CREDIT) ระหว่างวิธีพรีออกซ์ (PROX) และวิธียูคอน (UCON) ในแบบทดสอบวัดเจตคติแบบลิเคิร์ตสเกล*. ปรินซิพส์การศึกษามหาบัณฑิต, สาขาวิชาการวัดผลการศึกษา, บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ.
- อุทุมพร จามรมาน. (2535). *ข้อสอบ: การสร้างและการพัฒนา*. กรุงเทพฯ: ฟีนีซ์พับลิชชิ่ง.
- อุทุมพร จามรมาน. (2537). *ทฤษฎีการวัดทางจิตวิทยา*. กรุงเทพฯ: ฟีนีซ์พับลิชชิ่ง.
- เอมอร จังศิริพรปกรณ์. (2545). *การเปรียบเทียบคุณภาพของแบบสอบเลือกตอบระหว่างวิธีการ ให้คะแนนความรู้บางส่วนกับวิธีประเพณีนิยม*. กรุงเทพฯ: ภาควิชาวิจัยการศึกษาคณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.
- Albert, Wilfred Georjn. (1970). Scoring for partial knowledge in mathematics testing; A study of modification and extension of multiple – choice item applied to the testing of achievement in mathematics. *Dissertation Abstracts International*, 31(4), 1619A-1620A.
- Albanese, M. A., Thomas, H. K., & Douglas, R. W. (1979). Cluing in the multiple-choice test items with combinations of correct response. *Journal of Medical Education*, 54, 948-950.
- Albanese, M. A., & Sabers, D. L. (1988). Multiple true-false item: A study of inter-item correlations, scoring alternatives, and reliability estimation. *Journal of Educational Measurement*, 25, 111-123.
- Birnbaum, A. (1968). *Some latent trait models and their use in inferring an examinee's ability*, *Contributed chapters in Lord, F.M. and Novick, M.R., Statistical theories of mental test scores*. Chapters 17-20, Reading, MA: Addison Wesley.
- Blackmore, L. M. (1987). Computerized adaptive and pencil-and-paper test administrations: A comparative study in a high school setting. *Dissertation Abstracts International*, 47(07), 2554-A
- Chalhoub-Deville, Micheline, Alcaya, Cheryl & Lozier Vashti McCollum. (1996). *An operational ramework for constructing a computer-adaptive test of L2 reading ability: Theoretical and practical issues*. Retrieved from <http://www.carla.umn.edu/resources/working-papers/samples/CAT.pdf>.

- Coombs, C. H., Milholland, J. E., & Womer, F. B. (1956). The assessment of partial knowledge. *Education and Psychological Measurement, 16*, 13-37.
- Coombs, C. H. (1953). On the use of objective elimination. *Educational and Psychological Measurement, 13*, 308-310.
- Dodd, B. G., De Ayala, R. J., & Koch, W. R. (1995). Computerized adaptive testing with polytomous items. *Applied Psychological Measurement, 19*(1), 5-22.
- Dressel, P. L., & Schmid, J. (1953). Some modification of the multiple-choice item. *Education and Psychological Measurement, 13*, 574-595.
- Ebel, R. L. (1972). *Essentials of education measurement* (2nd ed.). New Jersey: Prentice Hall.
- Flaugher, R. (1990). Item pools. In *Computerized adaptive testing: A primer*. By Howard Wainer & et al. (pp. 41-63). New Jersey: Lawrence Erlbaum.
- Frey, A., & Seitz, N. N. (2009). Multidimensional adaptive testing in educational and psychological measurement: Current state and future challenges. *Studies in Educational Evaluation, 35*, 89-94.
- 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.
- Gifford, J. A., & Swaminathan, H. (1990). Bias and the effect of priors in bayesian estimation of parameters of item response model. *Applied Psychological Measurement, 14*(1), 33-43.
- Glass, G. V., & Stanley, J. C. (1970). *Statistical methods in education and psychology*. Englewood Cliffs, NJ: Prentice-Hall.
- Green, B. F. (1984). Tech guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement, 21*(4), 347-360.
- Haladyna, T. M., & Roid, G. H. (1983). Two alternative methods for criterion-referenced instructional program assessment. *Educational Technology, 28*, 35-38.
- Hambleton, R. K., & Cook, L. L. (1977). Latent trait models and their use in the analysis of educational test data. *Journal of Education Measurement, 14*, 75-96.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer Nijhoff.

- Hambleton, R. K., Swaminathan H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. London.
- Ho, R. G. (1989). Using micro CAT in computerized adaptive testing: A comparison of three adaptive testing strategies. *Dissertation Abstracts International*, 50(2), 421-A.
- Lord, F. M. (1975). Formula scoring and number right scoring. *Journal of Educational Measurement*, 12, 7-11.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading Massachusetts: Addison-Wesley.
- Meijer, R. R., & Michael, L. N. (1999). Computerized adaptive testing: Overview and introduction. *Applied Psychological Measurement*, 23(3), 187-194.
- Muraki, E. (1993). Information function of generalized partial credit model. *Applied Psychological Measurement*, 17(4), 351-363.
- Riod, G. H., & Haladyna, T. M. (1982). *A technology for test item-writing*. New York: Academic Press.
- Samejima, F. (1996). Graded response model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 85-100). New York: Springer.
- Samejima, F. (1972). A general model for free-response data. *Psychometrical Monograph*, 18.
- Segall, D. O. (2002). Principles of multidimensional adaptive testing. In W. J. Van Der Linden & C. A. W. Glas (Eds.), *Computerized adaptive testing: Theory and practice* (pp. 53-74). Netherlands: Kluwer.
- Segall, D. O. (2005). Computerized adaptive testing. In K. Kempf-Leonard (Ed.), *Encyclopedia of Social Measurement* (pp. 429-438). New York: Academic Press.
- Simon, A. B., Budescu, D. V., & Nevo, H. B. (1997). A comparative study of measurement procedure. *Psychometrika*.
- Smith, R. M. (1987). Assessing partial knowledge in vocabulary. *Journal of Educational Measurement*, 24, 217-231.

- Stocking, M. L., & Swanson, L. (1998). Optimal design of item banks for computerized adaptive tests. *Applied Psychological Measurement, 22*(3), 271-279.
- Straetmans, Gerard, J. J. M., & Eggen, Theo, J. H. M. (1998). Computerized adaptive testing: What it is and how it works. *Educational Technology, 38*(1), 45-52.
- Thorndike, R. L. (1969). *Measurement and evaluation in psychology and education*. New York: Wiley.
- Urry, V. W. (1977). Tailored testing: A successful application of latent trait theory. *Journal of Educational Measurement, 14*(2), 181-196.
- van der Linden, Wim, J. (1998). Optimal assembly of psychological and educational tests. *Applied Psychological Measurement, 22*(3), 195-211.
- Wainer, H. (1990). Introduction and History. In *Computerized adaptive testing: A primer*. by Howard Wainer, & et al. (pp. 1-21). New Jersey: Lawrence Erlbaum.
- Wainer, H., & Mislevy, R. J. (1990). Item response theory, item calibration and proficiency estimation. In *Computerized adaptive testing: A primer*. by Howard Wainer, & et al. (pp. 65-102). New Jersey: Lawrence Erlbaum.
- Weiss, D. J. (1974). *Strategies of adaptive ability measurement*. Minnesota: Department of Psychology University of Minnesota.

ภาคผนวก

ภาคผนวก ก

ข้อสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ตามมาตรฐานการเรียนรู้
และตัวชี้วัดระดับชั้นมัธยมศึกษาตอนปลาย หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน

พุทธศักราช 2551

**ข้อสอบวิชาคณิตศาสตร์ตามมาตรฐานการเรียนรู้ระดับชั้นมัธยมศึกษาตอนปลาย
หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551**

คำชี้แจง

1. แบบทดสอบนี้เป็นชุดข้อสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ สำหรับนักเรียนระดับชั้นมัธยมศึกษาตอนปลาย
2. ข้อสอบในแบบทดสอบฉบับนี้เป็น ข้อสอบเลือกตอบแบบถูกผิด ชนิด 4 ตัวเลือก
3. ข้อสอบในแต่ละข้ออาจจะมีตัวเลือกถูกหรือผิดมากกว่า 1 ตัวเลือก
4. ในการทำข้อสอบให้ผู้สอบพิจารณาคำตอบทุกตัวเลือก โดยทำเครื่องหมายดังนี้
 - 4.1 กรณีที่รู้คำตอบว่าผิด เป็นการเลือกคำตอบผิดบางตัวได้อย่างถูกต้อง โดยทำเครื่องหมายกากบาท (X) ตรงตัวเลือกที่เป็นตัวผิด
 - 4.2 กรณีที่รู้คำตอบว่าถูก เป็นการเลือกตัวเลือกบางตัวที่ถูกต้องได้อย่างถูกต้อง โดยทำเครื่องหมายถูก (✓) ตรงตัวเลือกที่เป็นตัวถูก
 - 4.3 กรณีที่ไม่แน่ใจว่าเป็นคำตอบที่ถูกหรือคำตอบที่ผิด จะทำเครื่องหมาย (?) แสดงความไม่แน่ใจ ตรงตัวเลือกนั้น ผู้สอบไม่ควรเว้นว่างไว้ เพราะจะถือว่าไม่ได้พิจารณาตัวเลือกดังกล่าว และจะทำให้ไม่มีคะแนนในข้อสอบข้อนี้

ตัวอย่างการตอบข้อสอบ

00. ถ้า N เป็นเลขคู่ ตัวเลือกใดเป็นเลขคี่
- (1) N+2 (ผิด)
 - (2) N+3 (ถูก)
 - (3) N+6 (ผิด)
 - (4) N+9 (ถูก)

ผู้สอบคนที่ 1

ข้อที่	ตัวเลือก	เฉลย	ตอบ	ตรวจ	MR	MTF	OMTF
00.	(1)	X	✓	0	0	1	0
	(2)	✓	X	0			
	(3)	X	✓	0			
	(4)	✓	✓	1			

ผู้สอบคนที่ 2

ข้อที่	ตัวเลือก	เฉลย	ตอบ	ตรวจ	MR	MTF	OMTF
00.	(1)	×	✓	0	0	2	0
	(2)	✓	×	0			
	(3)	×	×	1			
	(4)	✓	✓	1			

ผู้สอบคนที่ 3

ข้อที่	ตัวเลือก	เฉลย	ตอบ	ตรวจ	MR	MTF	OMTF
00.	(1)	×	✓	0	0	3	0
	(2)	✓	✓	1			
	(3)	×	×	1			
	(4)	✓	✓	1			

ผู้สอบคนที่ 4

ข้อที่	ตัวเลือก	เฉลย	ตอบ	ตรวจ	MR	MTF	OMTF
00.	(1)	×	×	1	0	4	4
	(2)	✓	✓	1			
	(3)	×	×	1			
	(4)	✓	✓	1			

ผู้สอบคนที่ 5

ข้อที่	ตัวเลือก	เฉลย	ตอบ	ตรวจ	MR	MTF	OMTF
00.	(1)	×	?		0	3	3
	(2)	✓	✓	1			
	(3)	×	×	1			
	(4)	✓	✓	1			

1. ถ้าผลบวกของ 4 และสามเท่าของจำนวนจริง x มีค่าไม่น้อยกว่า 20 แล้ว x มีค่าเท่าใด
- (1) x ต้องมีค่าน้อยกว่า $\frac{16}{3}$ (2) x ต้องมีค่ามากกว่า $\frac{16}{3}$
 (3) x ต้องมีค่าน้อยกว่าหรือเท่ากับ $\frac{16}{3}$ (4) x ต้องมีค่ามากกว่าหรือเท่ากับ $\frac{16}{3}$
2. ถ้า $4^{x-y} = 128$ และ $3^{2x+y} = 81$ แล้ว y มีค่าเท่าใด
- (1) -2 (2) -1 (3) 1 (4) 2
3. จงหาค่าของ $\sqrt{(2)^2} + \left[\frac{8^{1/2} + 2\sqrt{2}}{\sqrt{32}} \right]$ ว่ามีค่าเท่าใด
- (1) -1 (2) 1 (3) 3 (4) 5
4. จงหาค่าของ $\left(\sqrt[4]{81} + 64^{1/2} - \sqrt[3]{-32} \right)^{-2}$ ว่ามีค่าเท่าใด
- (1) $\frac{1}{13}$ (2) $-\frac{1}{13}$ (3) $\frac{1}{169}$ (4) $-\frac{1}{169}$
5. กำหนดให้ S เป็นเซตคำตอบของอสมการ $\frac{x^4 - 13x^2 + 36}{x^2 + 5x + 6} \geq 0$ ถ้า a เป็นจำนวนที่มีค่าน้อยที่สุดในเซต $S \cap (2, \infty)$ และ b เป็นจำนวนลบที่มีค่ามากที่สุด ซึ่ง $b \notin S$ แล้ว $a^2 - b^2$ เท่ากับข้อใด
- (1) -9 (2) -5 (3) 5 (4) 9
6. ถ้า x, y, z เป็นจำนวนจริงซึ่งสอดคล้องกับระบบสมการเชิงเส้น
- $$\begin{aligned} 2x - 2y - z &= 1 \\ x - 3y + z &= 7 \\ -x + y - z &= -5 \end{aligned}$$
- แล้ว $\frac{1}{x} + \frac{2}{y} + \frac{3}{z}$ มีค่าเท่ากับข้อใดต่อไปนี้
- (1) 0 (2) 2 (3) 5 (4) 8
7. เหตุ 1) ไม่มีศัลยแพทย์คนใดเป็นผู้หญิง
 2) นางสาวไทยทุกคนเป็นผู้หญิง
- ผลในข้อใดต่อไปนี้เป็นการสรุปผล จาก เหตุ ข้างต้นที่เป็นไปอย่างสมเหตุสมผล
- (1) นางสาวไทยบางคนเป็นผู้หญิง
 (2) ไม่มีศัลยแพทย์คนใดเป็นนางสาวไทย
 (3) ไม่มีศัลยแพทย์คนใดเป็นผู้หญิง
 (4) นางสาวไทยบางคนเป็นศัลยแพทย์

8. ถ้าผลบวกของ n พจน์แรกของอนุกรมหนึ่ง คือ $S_n = 5n^2 + 4$ แล้ว พจน์ที่ 10 ของอนุกรมนี้มีค่าเท่ากับข้อใดต่อไปนี้
- (1) 104 (2) 204 (3) 500 (4) 504
9. ข้อใดต่อไปนี้เป็นพจน์ทั่วไปของลำดับ $\frac{1}{2}, \frac{1}{4}, \frac{1}{6}, \frac{1}{8}, \frac{1}{10}, \dots$
- (1) $a_n = \frac{1}{n}$ (2) $a_n = \frac{1}{2^n}$
 (3) $a_n = \frac{1}{2n}$ (4) $a_n = 2 + n$
10. ถ้าลำดับเรขาคณิต 4, 20, 100, ... แล้วพจน์ที่ 7 เท่ากับข้อใด
- (1) 1200 (2) 15000 (3) 15625 (4) 62500
11. ขวัญตาตั้งใจว่าจะออมเงินไว้เพื่อซื้อคอมพิวเตอร์ โดยวันแรกจะออมไว้ 15 บาท วันที่สอง 30 บาท วันที่สาม 60 บาท เช่นนี้เรื่อยไปจนครบ 10 วัน เมื่อครบ 10 วัน ขวัญตาจะมีเงินออมเท่าใด
- (1) 15,360 บาท (2) 15,345 บาท (3) 12,485 บาท (4) 11,960 บาท
12. ความสัมพันธ์ในข้อใดไม่เป็นฟังก์ชัน
- (1) $r_1 = \{(a,1), (b,2), (c,3), (d,4)\}$ (2) $r_2 = \{(a,1), (b,1), (c,1), (d,1)\}$
 (3) $r_3 = \{(1,a), (2,b), (3,v), (1,d)\}$ (4) $r_4 = \{(1,1), (2,2), (3,3), (1,1)\}$
13. กำหนดให้ $A = \{2, 3, 4\}$ และ $B = \{a, b\}$ ฟังก์ชันในข้อใดต่อไปนี้เป็นฟังก์ชันจาก B ไป A
- (1) $f_1 = \{(a,2), (2,a), (4,a)\}$ (2) $f_2 = \{(a,3), (b,3)\}$
 (3) $f_3 = \{(2,b), (3,b), (4,b)\}$ (4) $f_4 = \{(2,a), (3,b), (4,b)\}$
14. กำหนดให้ $f(x) = -x^2 - 6x - 8$ ข้อความในข้อใดต่อไปนี้ไม่ถูกต้อง
- (1) กราฟของ f จะคว่ำลง (2) กราฟของ f มีจุดวกกลับที่จุด $(-3, 1)$
 (3) f มีค่าสูงสุดเท่ากับ 1 (4) f ไม่มีค่าสูงสุด

15. ในการสำรวจความคิดเห็นของนักเรียนต่อราคาสินค้าในโรงอาหารของโรงเรียนแห่งหนึ่ง โดยสำรวจกับนักเรียน ม. 4 – ม. 6 ตามกลุ่มตัวอย่างดังนี้

ชั้น	ห้อง	จำนวน
ม.4	4/1	20
	4/2	15
ม. 5	5/1	10
	5/2	15
ม. 6	6/1	30
	6/2	10

จงพิจารณาข้อความต่อไปนี้ว่าถูกหรือผิด

(1) ข้อมูลปฐมภูมิ เลือกตัวอย่างแบบหลายชั้น

(2) ข้อมูลปฐมภูมิ เลือกตัวอย่างแบบชั้นภูมิ

(3) ข้อมูลทุติยภูมิ เลือกตัวอย่างแบบหลายชั้น

(4) ข้อมูลทุติยภูมิ เลือกตัวอย่างแบบชั้นภูมิ

16. แผนภาพต้นไม้ แสดงน้ำหนักสัมภาระของผู้โดยสารบนเครื่องบินกลุ่มหนึ่ง เป็นดังนี้

0	6	7						
1	3	4	6	6	9			
2	0	0	1	2	2	2	4	7
3	1	1	2	3	3	5	8	
4	1	1	2					

ข้อใดต่อไปนี้ถูก

- (1) ค่าเฉลี่ยเลขคณิตน้อยกว่าฐานนิยม
- (2) ฐานนิยม น้อยกว่ามัธยฐาน
- (3) มัธยฐาน น้อยกว่าค่าเฉลี่ยเลขคณิต
- (4) ค่าเฉลี่ยเลขคณิตเท่ากับมัธยฐาน และฐานนิยม

17. มัธยฐาน เป็นค่ากลางที่เหมาะสมกับข้อมูลแบบใดต่อไปนี้

- (1) ขนาดของรองเท้า
- (2) รายได้ของประชากรทั่วประเทศ
- (3) คะแนนสอบของนักเรียนที่มีความสามารถใกล้เคียงกัน
- (4) ความคิดเห็นของคนกรุงเทพฯกับการแก้ไขรัฐธรรมนูญ

18. จากการสำรวจความคิดเห็นของประชาชนเกี่ยวกับการประกันสุขภาพถ้วนหน้า พ.ศ. 2546 ผลการสำรวจแสดงให้เห็นว่า ผู้ไม่มีสิทธิเข้าร่วมโครงการทั่วประเทศมีถึงร้อยละ 23.2 ต้องการให้บัตรประกันสุขภาพถ้วนหน้าและสวัสดิการที่มีอยู่ร้อยละ 8.2 ในจำนวนผู้มีสิทธิเข้าร่วมโครงการมีถึงร้อยละ 74.9 ที่เคยใช้บริการ จากข้อมูลข้างต้นข้อใดบ้างที่รัฐบาลควรทำ
- (1) ดำเนินโครงการต่อไป เพราะโครงการนี้เกิดประโยชน์กับคนส่วนใหญ่
 - (2) ดำเนินโครงการต่อไป เพราะสร้างความเชื่อมั่นในการเข้ารับการรักษา
 - (3) ยุติโครงการ เพราะ ประชาชนไม่มั่นใจในมาตรฐานการรักษา
 - (4) ยุติโครงการ เพราะ ประชาชนไม่คุ้มค่าในการรักษา
19. สินค้าชิ้นแรกมีส่วนเบี่ยงเบนมาตรฐาน 425 บาท ค่าเฉลี่ยเลขคณิต 17 บาท สินค้าชิ้นที่สองมีส่วนเบี่ยงเบนมาตรฐาน 483 บาท ค่าเฉลี่ยเลขคณิต 23 บาท นักเรียนจะเลือกซื้อสินค้าจากร้านค้าใด
- (1) ร้านแรก เพราะ มีการกระจายของราคาสินค้ามากกว่าร้านที่สอง
 - (2) ร้านที่สอง เพราะ มีการกระจายของราคาสินค้าน้อยกว่าร้านแรก
 - (3) ร้านแรก เพราะ มีค่าเฉลี่ยของราคาสินค้าน้อยกว่าร้านที่สอง
 - (4) ร้านใดก็ได้ เพราะมีการกระจายของราคาสินค้าเท่ากัน
20. ร้านค้าแห่งหนึ่งได้ทำการสำรวจความนิยมของลูกค้าเกี่ยวกับการใช้ผงซักฟอก พบว่า 40% ใช้ผงซักฟอกยี่ห้อบรีส 25% ใช้ผงซักฟอกยี่ห้อโอโม และ 10% ใช้ผงซักฟอกทั้งสองยี่ห้อ อยากทราบว่าลูกค้าที่ไม่ใช้ผงซักฟอกทั้งสองยี่ห้อ มีกี่คน ถ้าสำรวจความนิยมของลูกค้าจำนวน 500 คน
- (1) 225 คน
 - (2) 75 คน
 - (3) 150 คน
 - (4) 50 คน
21. ถ้า $A - B = \{a, b, c\}$, $B - A = \{x, y, z\}$ และ $A \cup B = \{0, 1, 2, a, b, c, x, y, z\}$ และ $(A \cap B)'$ เป็นเซตในข้อใดต่อไปนี้
- (1) $\{a, b, c, x, y, z\}$
 - (2) $\{0, 1, 2, x, y, z\}$
 - (3) $\{3, y, z\}$
 - (4) $\{a, y, z\}$
22. เหตุ 1) จำนวนเต็มที่หารด้วย 2 ลงตัวทุกจำนวนเป็นจำนวนคู่
2) 11 หารด้วย 2 ลงตัว
- ผลในข้อใดต่อไปนี้เป็นการสรุปผลจาก เหตุ ข้างต้นที่เป็นไปอย่างสมเหตุสมผล
- (1) 11 ไม่ใช่จำนวนคู่
 - (2) 11 เป็นจำนวนคี่
 - (3) 11 เป็นจำนวนคู่
 - (4) สรุปผลไม่ได้

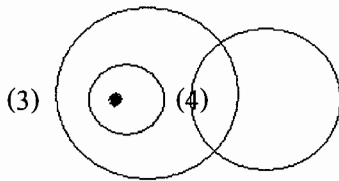
32. จากการสุ่มตัวอย่างนักเรียนชั้นมัธยมศึกษาปีที่ 6 จำนวน 1000 คน พบว่า มีผู้ต้องการศึกษาต่อจำนวน 370 คน มีผู้ต้องการทำงานจำนวน 550 คน และมีนักเรียนที่ต้องการศึกษาต่อหรือต้องการทำงานจำนวน 850 คน จะมีผู้ที่ต้องการศึกษาต่อและต้องการทำงานด้วยทั้งหมดกี่คน
- (1) 850 คน (2) 280 คน (3) 170 คน (4) 70 คน
33. เหตุ 1) คนที่ไม่มีหนี้สินและมีเงินฝากในธนาคารมากกว่า 100 ล้านบาทเป็นเศรษฐี
บุญเกิดไม่มีหนี้สินและมีเงินฝากในธนาคาร 180 ล้านบาท
- ผล กือข้อใด.....
- (1) บุญเกิดรวยกว่าเศรษฐี (2) บุญเกิดมีเงินมากกว่าเศรษฐี
(3) บุญเกิดเป็นเศรษฐี (4) บุญเกิดดีกว่าเศรษฐี
34. จงตรวจสอบและสรุปต่อไปนี้เป็นสมเหตุสมผลหรือไม่ โดยใช้แผนภาพ
- เหตุ 1) ถ้ามีสิ่งมีชีวิตบนดาวเคราะห์แล้ว ดาวเคราะห์ต้องมีก๊าซออกซิเจน
มีสิ่งมีชีวิตบนดาวเคราะห์ A
- ผล ดาวเคราะห์ A มีก๊าซออกซิเจน
- (1) ผลสรุปที่กล่าวว่า ดาวเคราะห์ A มีออกซิเจนสมเหตุสมผล
(2) ผลสรุปที่กล่าวว่าดาวเคราะห์ A อาจจะไม่มีการออกซิเจน
(3) ผลสรุปที่กล่าวว่าดาวเคราะห์ A มีก๊าซออกซิเจน ไม่สมเหตุสมผล
(4) ผลสรุปที่กล่าวว่า ดาวเคราะห์ A ไม่สมเหตุสมผล เพราะไม่ทราบว่าดาวเคราะห์ A มีก๊าซออกซิเจนหรือไม่
35. ลำดับในข้อใดต่อไปนี้ เป็นลำดับเรขาคณิต
- (1) $a_n = 3^n \cdot 2^{2^n}$ (2) $a_n = 3^n + 2^n$
(3) $a_n = 2^{n^2}$ (4) $a_n = (3n)^n$
36. ลำดับเรขาคณิต $\frac{1}{625}, \frac{1}{125}, \frac{1}{25}, \dots$ แล้วพจน์ที่ 9 ของลำดับนี้เท่ากับข้อใด
- (1) 5 (2) 25 (3) 125 (4) 625
37. กำหนดให้ $s = \{ 11, 12, 13, \dots, 99 \}$ ถ้า a เท่ากับ ผลบวกของจำนวนที่ทั้งหมดใน S และ b เท่ากับผลบวกของจำนวนคู่ทั้งหมดใน S แล้ว $b - a$ มีค่าเท่ากับเท่าไรต่อไปนี้
- (1) -45 (2) -50 (3) -55 (4) -60
38. กำหนดให้ A และ B เป็นเซต ซึ่ง $n(A \cup B) = 75$ และ $[n(A - B) \cup (B - A)] = 20$ ถ้า $n(A) = 50$ แล้ว $n(B)$ เท่ากับข้อใดต่อไปนี้
- (1) 70 (2) 76 (3) 78 (4) 80

39. จงพิจารณาข้อความต่อไปนี้

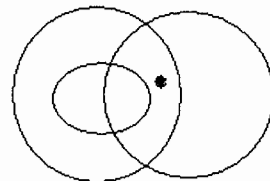
- 1) นักเรียนทุกคนตั้งใจเรียน
- 2) คนที่ตั้งใจเรียนบางคนเกรด 4
- 3) วรเทพเป็นนักเรียน และได้เกรด 4

แผนภาพในข้อใดต่อไปนี้ มีความเป็นไปได้ที่จะสอดคล้องกับข้อความทั้งสามข้อข้างต้น
เมื่อจุดแทนวรเทพ

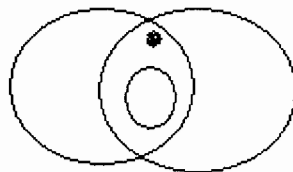
(1)



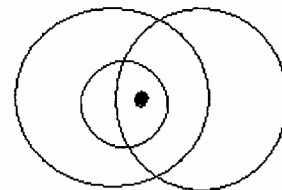
(2)



(3)



(4)



40. ถ้า $x = \frac{-1}{3}$ เป็นรากของสมการ $ax^2 - 7x - 3 = 0$ แล้วรากอีกรากหนึ่งของสมการนี้

มีค่าเท่ากับเท่าไร

(1) $-\frac{3}{2}$

(2) $-\frac{2}{3}$

(3) $\frac{2}{3}$

(4) $\frac{3}{2}$

41. ข้อใดเป็นฟังก์ชันจาก $B \rightarrow A$ โดยที่ $A = \{1,2,3\}$, $B = \{a,b,c\}$

(1) $\{(1,a), (2,a), (3,b)\}$

(2) $\{(a,1), (b,2), (c,3)\}$

(3) $\{(1,1), (2,2), (3,3)\}$

(4) $\{(1,a), (2,a), (3,c)\}$

42. ผลการสอบวิชาคณิตศาสตร์ของนาย ก ในชั้นมัธยมศึกษาปีที่ 1 เป็นดังนี้

รหัสวิชา	ค21101	ค21201	ค21102	ค21202
หน่วยกิต	1.0	2.0	1.0	2.0
เกรด	4	3.5	3	3.5

เกรดเฉลี่ยวิชาคณิตศาสตร์ของนาย ก ในชั้นมัธยมศึกษาปีที่ 1 เป็นเท่าใด

- (1) 2.48 (2) 3.5 (3) 3.63 (4) 3.98

43. กราฟของฟังก์ชันในข้อใดต่อไปนี้ตัดแกน x มากกว่า 1 จุด

- (1) $y = x^2$ (2) $y = |x| - 3$ (3) $y = |x + 1|$ (4) $y = 2^x$

44. คะแนนสอบของนักเรียนกลุ่มหนึ่งเป็นดังนี้ 10, 12, 9, 18, 16, 13, 10, 12, 11, 14, 8, 16, 18, 17, 16, 12 เปอร์เซ็นไทล์ที่ 80 มากกว่าเปอร์เซ็นไทล์ที่ 20 อยู่เท่าใด

- (1) 6.6 (2) 7.6 (3) 8.6 (4) 9.6

45. นักเรียนคนหนึ่งได้เงินมาโรงเรียนในรอบสัปดาห์ดังนี้

วัน	จันทร์	อังคาร	พุธ	พฤหัสบดี	ศุกร์
เงิน(บาท)	30	30	60	50	90

ข้อใดต่อไปนี้ถูกต้อง

- (1) ค่าเฉลี่ยเลขคณิตเท่ากับมัธยฐาน
 (2) ฐานนิยมมากกว่าค่าเฉลี่ยเลขคณิต
 (3) มัธยฐานเท่ากับฐานนิยม
 (4) ค่าเฉลี่ยเลขคณิตมากกว่าฐานนิยม

46. ความน่าจะเป็นที่ นาย ก ถูกรางวัลเลขท้าย 2 ตัว ของสลากกินแบ่งรัฐบาลคือข้อใด

- (1) $\frac{1}{10}$ (2) $\frac{2}{10}$ (3) $\frac{1}{100}$ (4) $\frac{1}{99}$

47. โยนลูกเต๋า 3 ลูก ความน่าจะเป็นที่ลูกเต๋ายกจะขึ้นแต้มคู่เหมือนกันทุกลูกคือข้อใด

- (1) $\frac{1}{216}$ (2) $\frac{6}{216}$ (3) $\frac{1}{72}$ (4) $\frac{1}{12}$

48. กำหนดข้อมูลดังนี้ 1 2 5 8 10 25 ควรใช้ค่ากลางใด

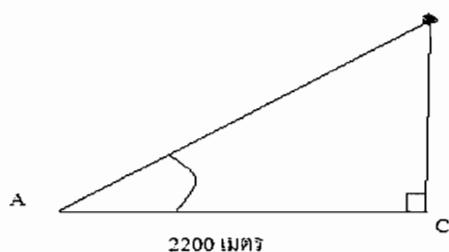
- (1) \bar{x} (2) มัธยฐาน (3) ฐานนิยม (4) ค่ากึ่งกลางพิสัย

49. ความน่าจะเป็นที่รางวัลเลขท้าย 3 ตัว (หมุน 1 ครั้ง) ของสลากกินแบ่งรัฐบาลที่จะออกเลขสามหลักเป็นเลขเดียวกัน เท่ากับข้อใด

- (1) $\frac{1}{10}$ (2) $\frac{1}{100}$ (3) $\frac{1}{1000}$ (4) $\frac{1}{10,000}$

50. $\left(\frac{\sqrt{3}}{2\sqrt{5}} - \frac{\sqrt{5}}{2\sqrt{3}}\right)^2$ มีค่าเท่ากับเท่าไร
- (1) $\frac{1}{15}$ (2) $\frac{2}{15}$ (3) $\sqrt{3} - 5$ (4) $3 - \sqrt{5}$
51. $\left[\sqrt{\left(\frac{2}{5}\right)^3}\right]^3 = \left(\frac{8}{125}\right)^{\frac{2}{x}}$ มีค่าเท่ากับเท่าไร
- (1) $\frac{2}{3}$ (2) $\frac{4}{3}$ (3) $\frac{3}{2}$ (4) $\frac{3}{4}$
52. $(\sqrt{125} + 3\sqrt{-27} - 5\sqrt{625})$ มีค่าเท่ากับเท่าไร
- (1) $-34 + 5\sqrt{5}$ (2) $-34 - 5\sqrt{5}$
 (3) $34 + 5\sqrt{5}$ (4) $34 - 5\sqrt{5}$
53. $\frac{(\sqrt{32} - \sqrt{243}) + (\sqrt{72} + \sqrt{27})}{(\sqrt{12} + 3\sqrt{8}) - (\sqrt{75} - \sqrt{48})}$ มีค่าเท่ากับจำนวนในข้อใดต่อไปนี้
- (1) $\frac{2}{3}(\sqrt{3} - \sqrt{2})$ (2) $\frac{2\sqrt{3}}{3}(\sqrt{3} - \sqrt{2})$
 (3) $\frac{2}{3}(\sqrt{2} - \sqrt{6})$ (4) $\frac{2}{3}(\sqrt{6} - \sqrt{2})$
54. จากสมการ $(x+1)^2 = \sqrt{x^3 + ax^2 + 31x + 19} + (x+b)(x-2)$ จงหาจำนวนของค่า a และ b ที่ทำให้ $x = 1, 2, 3$
- (1) a มีหนึ่งค่า , b มีค่าบวกหนึ่งค่า ลบหนึ่งค่า
 (2) a มีหนึ่งค่า , b มีค่าบวกสองค่า
 (3) a มีหนึ่งค่า , b มีค่าสามค่า
 (4) a มีหนึ่งค่า , b มีหนึ่งค่า
55. ให้ x, y, z เป็นจำนวนจริงใด ๆ ข้อใดต่อไปนี้ถูก
- (1) ถ้า $x < y$ แล้ว $xz < yz$ หรือ $xz > yz$
 (2) ถ้า $1 < x \leq y$ และ n เป็นจำนวนจริงเต็ม แล้ว $(x-1)^n \leq (y-1)^n$
 (3) $2xy \leq \frac{(x+y)^2 + (x-y)^2}{2} \leq (x+y)^2 - 2xy$
 (4) ถ้า $|x-1| < 2$ แล้ว $\frac{1}{2} < 2^x < 2^3 - 2^{\frac{1}{3}}$

56.



ผู้สังเกตคนหนึ่งอยู่บนพื้นดินที่ A ทำมุมเงยของบอลูนที่จุด B เป็น 37° จุด C เป็นจุดหนึ่งบนพื้นดินอยู่ตรงลงไปใต้บอลูนระยะทางจาก AC เป็น 2,200 เมตร อยากทราบว่าบอลูนอยู่สูงจากพื้นดินกี่เมตร ($\cot 53^\circ = 0.754$)

- (1) 1324 เมตร (2) 1658 เมตร
(3) 1757 เมตร (4) 2919 เมตร

57. ในการเล่นฟุตบอลชายหาด ผู้เล่นคนหนึ่งอยู่ที่จุด F พยายามที่ยิงประตู ผู้ยิงอยู่ห่างจากเสาประตู G 6 เมตร และอยู่ห่างจากเสาประตู H 8 เมตร จากจุดที่เสากำลึงยืนอยู่ มุมที่สามารถยิงประตูได้คือ 19° ระยะทางระหว่างเสาประตู G และ H ใกล้เคียงที่สุดคือข้อใด

- (1) 2.3 เมตร (2) 2.6 เมตร (3) 3.0 เมตร (4) 5.3 เมตร

58. กำหนดให้ $A = \{ 1, 2, 3, 4 \}$

$$B = \{ 1, 2, 3, 4, \dots, 14, 15 \}$$

$$S = \{ (a, b) \in A \times B \mid b = a^2 - 2a \}$$

แล้วจำนวนสมาชิกของ S เท่ากับข้อใด

- (1) 1 (2) 2 (3) 3 (4) 4

59. โดยการให้เหตุผลแบบอุปนัย $(a + 2b) - c$ มีค่าเท่ากับเท่าไรต่อไปนี้

- (1) 256 (2) 128 (3) 64 (4) 512

60. เซตของคู่อันดับในข้อใดเป็นฟังก์ชัน

(1) $\{ (1, 2), (2, 3), (3, 4) \}$

(2) $\{ (1, 2), (1, 3), (2, 5) \}$

(3) $\{ (2, 4), (3, 4), (2, 3) \}$

(4) $\{ (-1, 1), (0, 2), (0, 3) \}$

61. จงหาพจน์ทั่วไปของ $\frac{1}{2}, \frac{1}{8}, \frac{1}{18}, \frac{1}{32}, \dots$

(1) $\frac{1}{2n+2}$

(2) $\frac{1}{2n+1}$

(3) $\frac{1}{2n^n}$

(4) $\frac{1}{2n^2}$

62. ลำดับเรขาคณิตลำดับหนึ่ง มีพจน์แรกเท่ากับ 2 และพจน์ที่สองเท่ากับ 1 พจน์ที่ห้าของลำดับเรขาคณิตเป็นเท่าใด

(1) $\frac{1}{4}$

(2) $\frac{1}{8}$

(3) $\frac{1}{16}$

(4) $\frac{1}{32}$

63. จากการสอบถามนักเรียน 40 คน ไม่มีนักเรียนคนใดไม่ชอบโก๊ะตี้และไม่ชอบสายัณฑ์ มีนักเรียน 20 คน ชอบโก๊ะตี้ มีนักเรียน 10 คน ชอบโก๊ะตี้และสายัณฑ์ มีนักเรียนกี่คนที่ชอบสายัณฑ์และไม่ชอบโก๊ะตี้
- (1) 15 (2) 18 (3) 20 (4) 22
64. กำหนด เหตุ 1) จำนวนเป็นนักกีฬาว่ายน้ำสมัครเล่น
2) นักกีฬาว่ายน้ำสมัครเล่นบางคนได้เหรียญทอง
ผล จำนวนได้เหรียญทอง
- จากเหตุที่กำหนด เมื่อต้องการตรวจสอบผลสรุปว่าเป็นการสรุปสมเหตุสมผลหรือไม่ จะเขียนแผนภาพตรวจสอบได้ทั้งหมดกี่แบบ
- (1) 1 (2) 2 (3) 3 (4) 4
65. ผลบวกของจำนวนเต็มที่อยู่ระหว่าง 59 และ 229 ที่หารด้วย 3 ลงตัว ตรงกับข้อใด
- (1) 8129 (2) 8202 (3) 8205 (4) 8208
66. การสำรวจความคิดเห็นของกลุ่มบุคคลเกี่ยวกับเรื่องใดเรื่องหนึ่งเป็นการเก็บรวบรวมข้อมูลโดยใช้วิธีใด
- (1) การสังเกต (2) การทดลอง (3) การสอบถาม (4) การสัมภาษณ์
67. ในการสอบสัมภาษณ์นักเรียน 3 คน ปรากฏว่า ค่าเฉลี่ยของคะแนนมีค่าเท่ากับ 53 มัธยฐานเท่ากับ 50 และพิสัยเท่ากับ 21 ส่วนเบี่ยงเบนมาตรฐานของคะแนนในการสอบสัมภาษณ์ครั้งนี้เท่ากับข้อใด
- (1) 13.72 (2) 17.32 (3) 18.41 (4) 19.62
68. อายุของเด็กกลุ่มหนึ่งจำนวน 7 คน เป็นดังนี้ 10 , 5 , 8 , 6 , 7 , 5 , 8 ปี ข้อความที่กล่าวถึงค่ากลางของอายุของเด็กกลุ่มนี้ ข้อใดถูกต้อง
- (1) ค่าฐานนิยมมากกว่าค่ามัธยฐาน (2) ค่าฐานนิยมมากกว่าค่าเฉลี่ยเลขคณิต
(3) ค่าฐานนิยมเท่ากับค่าเฉลี่ยเลขคณิต (4) มัธยฐานเท่ากับค่าเฉลี่ยเลขคณิต
69. นักเรียนคนหนึ่งมีถุงเท้านักเรียน 2 คู่ ที่ต่างกันและรองเท้านักเรียน 3 คู่ ที่ต่างกัน นักเรียนคนนี้จะมียุวิธีใส่ถุงเท้า 1 คู่ และรองเท้า 1 คู่ เพื่อไปโรงเรียนได้กี่วิธี
- (1) 3 วิธี (2) 4 วิธี (3) 5 วิธี (4) 6 วิธี
70. พิจารณาข้อความต่อไปนี้ ข้อใดถูกหรือผิด
- (1) มีจำนวนเต็มที่มีมากที่สุดที่น้อยกว่า 9
(2) มีจำนวนตรรกยะที่มีมากที่สุดที่น้อยกว่า 9
(3) มีจำนวนอตรรกยะที่น้อยที่สุดที่มากกว่า 1
(4) มีจำนวนตรรกยะที่น้อยที่สุดที่มากกว่า 1

71. พิจารณาข้อความต่อไปนี้ ว่าข้อใดถูกหรือผิด

- (1) มีจำนวนจริงซึ่ง $|x| = -x$
- (2) มีจำนวนจริง ซึ่ง $|x| > x$
- (3) มีจำนวนจริง ซึ่ง $-|x| < x$
- (4) มีจำนวนจริง ซึ่ง $|x| < x$

72. พิจารณาข้อความต่อไปนี้ ว่าข้อใดถูกหรือผิด

- (1) $|-a - b| \leq |-a| - |-b|$
- (2) $|a| < a$ ก็ต่อเมื่อ a เป็นจำนวนจริงลบ
- (3) $|a| \leq b$ ก็ต่อเมื่อ $-a \leq b \leq a$
- (4) $|b - a| \geq ||a| - |b||$

73. ข้อใดกล่าวถูกต้อง

- (1) $a^m + a^n = a^{m+n}$
- (2) $a^{-n} = -\frac{1}{a^n}$ เมื่อ $a \neq 0$
- (3) $(a^m)^n = a^{m+n}$
- (4) $\frac{1}{a^{m-n}} = a^{n-m}$ เมื่อ $a \neq 0$

74. $\left(\frac{2^{3n+1} + 2^{2n+1}}{2^{2n+1} + 2^{n+1}}\right)^{\frac{1}{n}}$ มีค่าเท่ากับเท่าไรต่อไปนี้

- (1) 0
- (2) 1
- (3) 2
- (4) 3

75. กำหนดให้ $\sqrt{2} \approx 1.41421$, $\sqrt{3} \approx 1.73205$ จงหาค่าประมาณของ $\frac{5\sqrt{3}-3}{5\sqrt{3}+3}$

ให้ถูกต้องถึงทศนิยมตำแหน่งที่ 4

- (1) 0.5580
- (2) 0.5620
- (3) 0.5704
- (4) 0.5860

76. พิจารณาข้อความต่อไปนี้ ข้อสรุปใดต่อไปนี้ถูกต้อง

- (1) $75 + 50 = 25(3+2)$ เป็นจริงตามสมบัติการแจกแจง
- (2) $2.45 \times \left(\frac{1}{3} \times \sqrt{2}\right) = \left(2.45 \times \frac{1}{3}\right) \times \sqrt{2}$ เป็นจริงตามสมบัติการสลับที่ของการคูณ
- (3) $\frac{5}{3} \times 1 = \frac{5}{3}$ เป็นจริงตามสมบัติการมีเอกลักษณ์ของการคูณ
- (4) $7 + (-7) = 0$ เป็นจริงตามสมบัติตัวผกผันการบวก (อินเวิร์สการบวก)

77. ข้อใดต่อไปนี้ เป็นจริง

- (1) $\sin 30^\circ \cos 60^\circ > 1$
- (2) $\cos 30^\circ \sin 60^\circ > 1$
- (3) $\cot 60^\circ \cot 45^\circ > 1$
- (4) $\tan 60^\circ \tan 45^\circ > 1$

78. นักเรียนคนหนึ่งยืนมองเสาตรงเป็นมุมเงย 60° จากระดับสายตา ถ้านักเรียนคนนี้อยู่ห่างจากเสาตรงเป็นระยะทาง 25 เมตร และนักเรียนมีความสูง 160 เซนติเมตร เสาตรงสูงประมาณกี่เมตร

- (1) 37 (2) 45 (3) 47 (4) 55

79. กำหนดให้ $A = \{1, 2, \{1, 2\}, \{1, 2, 3\}\}$ ข้อใดต่อไปนี้ผิด

- (1) $\{1, 2\} \in A$
 (2) $\{1, 2, 3\} \in A$
 (3) $\{1, 2\} \subset A$
 (4) $\{1, 2, 3\} \subset A$

80. พิจารณาผลต่างระหว่างพจน์ของลำดับ 2, 5, 10, 17, 26, ... โดยการให้เหตุผลแบบอุปนัย พจน์ที่ 8 ของลำดับคือข้อใดต่อไปนี้

- (1) 50 (2) 65 (3) 82 (4) 101

81. กำหนดให้ $A = \{1, 3, 5\}$ และ $B = \{a, b\}$ คู่อันดับในข้อใดต่อไปนี้ เป็นสมาชิกของผลคูณคาร์ทีเซียน $A \times B$

- (1) (1,3) (2) (b,a) (3) (a,5) (4) (3,b)

82. ข้อใดเป็นพจน์ทั่วไปของ $1, \frac{4}{3}, \frac{3}{2}, \frac{8}{5}, \dots$ และพจน์ที่ 10 เท่ากับเท่าไร

- (1) $a_n = \frac{n+1}{n+1} + 1, a_{10} = 2$
 (2) $a_n = \frac{n+2}{n+1} - 1, a_{10} = \frac{1}{11}$
 (3) $a_n = \frac{n-1}{n+1} + 1, a_{10} = \frac{20}{11}$
 (4) $a_n = \frac{1}{n+1} + n - 1, a_{10} = \frac{100}{11}$

83. กำหนดให้ลำดับเรขาคณิตเป็น 8, 16, 32, ... และลำดับเลขคณิตเป็น 3, 6, 9, ... พจน์ที่เท่าใดของทั้งสองลำดับทำให้ลำดับเรขาคณิตมากกว่าลำดับเลขคณิตอยู่ 491

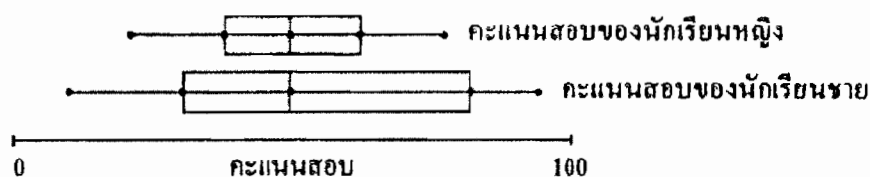
- (1) 4 (2) 5 (3) 6 (4) 7

84. จากการสอบถามนักเรียน 150 คน ว่าชอบกีฬาประเภทใดต่อไปนี้ พบว่า นักเรียนตอบว่า ชอบฟุตบอล 68 คน ชอบบาสเกตบอล 44 คน และชอบวอลเลย์บอล 37 คน ถ้ามักเรียนชอบทั้งบาสเกตบอลและวอลเลย์บอล 10 คน ชอบฟุตบอลและบาสเกตบอลแต่ไม่ชอบวอลเลย์บอล 14 คน ชอบฟุตบอลอย่างเดียว 45 คน จะมีนักเรียนที่ไม่ชอบกีฬาทั้งสามชนิดกี่คน

- (1) 14 (2) 24 (3) 34 (4) 44

85. คุณครูณัฐพร พูดว่า : วันนี้เด็กชายวงศธรเข้าเรียนสายและไม่ได้กินอาหารเช้า
คุณครูนิตยา พูดว่า : วันนี้ถ้าเด็กชายสิริวิชญ์เข้าเรียนสายแล้วเด็กชายสิริวิชญ์ได้ทานอาหารเช้า
ผู้อำนวยการโรงเรียนทราบภายหลังจากนั้นว่า : คุณครูณัฐพรพูดจริงแต่คุณครูนิตยาพูดเท็จ
ข้อสรุปใดเป็นเท็จ สำหรับเหตุการณ์ในวันนี้
- (1) ถ้าเด็กชายวงศธรเข้าเรียนสายแล้วเด็กชายสิริวิชญ์ไม่ได้ทานอาหารเช้า
 - (2) เด็กชายสิริวิชญ์ได้ทานอาหารเช้าก็ต่อเมื่อเด็กชายวงศธรทานอาหารเช้า
 - (3) เด็กชายวงศธรได้ทานอาหารเช้าหรือเด็กชายสิริวิชญ์ไม่ได้เข้าเรียนสาย
 - (4) เด็กชายสิริวิชญ์ไม่ได้ทานอาหารเช้าและเด็กชายวงศธรเข้าเรียนสาย
86. ต้องการล้อมรั้วรอบที่ดินรูปสี่เหลี่ยมผืนผ้า ซึ่งมีพื้นที่ 230 ตารางวา โดยด้านยาวของที่ดินยาวกว่าสองเท่าของด้านกว้างอยู่ 3 วา จะต้องใช้รั้วที่มีความยาวเท่ากับข้อใด
- (1) 36 วา
 - (2) 33 วา
 - (3) 56 วา
 - (4) 66 วา
87. ให้ $A = \{2, 88\}$ ความสัมพันธ์ใน A ในข้อใดไม่เป็นฟังก์ชัน
- (1) เท่ากับ
 - (2) ไม่เท่ากับ
 - (3) หารลงตัว
 - (4) หารไม่ลงตัว
88. จากความสัมพันธ์ m ข้อใดต่อไปนี้เป็นฟังก์ชัน
- (1) m เป็นฟังก์ชันเพราะ $(2,1), (3,2)$ และ $(4,3)$ อยู่ในแนวเส้นตรงเดียวกัน
 - (2) m เป็นฟังก์ชันเพราะมีจำนวนจุดเป็นจำนวนจำกัด
 - (3) m ไม่เป็นฟังก์ชันเพราะมีจุด $(4,3)$ และ $(4,-2)$ อยู่บนกราฟ
 - (4) m ไม่เป็นฟังก์ชันเพราะมีจุด $(-2,-2)$ และ $(4,-2)$ อยู่บนกราฟ
89. กำหนดให้อนุกรมเลขคณิต $2+5+8+\dots$ และอนุกรมเรขาคณิต $1+3+9+\dots$
จงหาว่าผลบวกของจำนวนที่ n พจน์ที่ทำให้ผลต่างของอนุกรมเรขาคณิตกับอนุกรมเลขคณิตมีค่าไม่น้อยกว่า 29369 โดยที่พจน์ที่ n มีค่าต่ำสุดเป็นเท่าใด
- (1) 8
 - (2) 9
 - (3) 10
 - (4) 11
90. ข้อใดต่อไปนี้เป็นข้อมูลเชิงปริมาณ
- (1) บ้านเลขที่
 - (2) อายุ
 - (3) หมายเลขโทรศัพท์
 - (4) เพศ
91. พิจารณาข้อมูลต่อไปนี้ 9, 4, 5, 8, 10, 14, 6, 16 ค่าของ D_8 ใกล้เคียงกับข้อใดต่อไปนี้มากที่สุด
- (1) 14.1
 - (2) 14.3
 - (3) 14.7
 - (4) 15.0

92. ข้อมูลต่อไปนี้แสดงส่วนสูงในหน่วยเซนติเมตรของนักเรียนกลุ่มหนึ่ง 145, 149, 150, 162, 155, 156, 165, 190, 155 ค่ากลางในข้อใดเป็นค่าที่ไม่เหมาะสมที่จะเป็นตัวแทนของข้อมูลชุดนี้
- (1) มัธยฐาน (2) ฐานนิยม
(3) ค่าเฉลี่ยเลขคณิต (4) ค่าเฉลี่ยของค่าสูงสุดและค่าต่ำสุด
93. ถ้าข้อมูลชุดหนึ่งประกอบด้วย 10, 12, 15, 13 และ 10 ข้อใดเป็นเท็จ
- (1) มัธยฐานเท่ากับ 12
(2) ฐานนิยมน้อยกว่า 12
(3) ฐานนิยมน้อยกว่าค่าเฉลี่ยเลขคณิต
(4) ค่าเฉลี่ยเลขคณิตมากกว่า 12
94. เกมต้องการสร้างหมายเลขที่มี 2 หลัก จากเลขโดด 0, 1, 2 และ 3 สามารถใช้หมายเลขซ้ำกันได้ จงพิจารณาข้อความต่อไปนี้ว่าถูกหรือผิด
- (1) หมายเลขที่สร้างขึ้นมีจำนวนคู่มากกว่าจำนวนคี่
(2) หมายเลขที่ผลบวกทั้ง 2 หลักเป็น 6 มีจำนวนมากที่สุด
(3) จำนวนหมายเลขที่ผลบวกทั้ง 2 หลัก เป็น 3 เท่ากับจำนวนหมายเลขที่ผลบวกทั้ง 2 หลักเป็น 4
(4) จำนวนหมายเลขที่ผลบวกทั้ง 2 หลัก เป็น 1 มากกว่าจำนวนหมายเลขที่ผลบวกทั้ง 2 หลัก เป็น 5
95. จากแผนภาพกล่องของคะแนนสอบวิชาคณิตศาสตร์ของนักเรียน จำแนกตามเพศเป็นดังนี้



ข้อสรุปในข้อใดต่อไปนี้ ถูกหรือผิด

- (1) คะแนนสอบเฉลี่ยวิชาคณิตศาสตร์ของนักเรียนชายสูงกว่าคะแนนสอบเฉลี่ยวิชาคณิตศาสตร์ของนักเรียนหญิง
- (2) คะแนนสอบวิชาคณิตศาสตร์ของนักเรียนชายมีการกระจายเบ้ลาดขวา
- (3) คะแนนสอบวิชาคณิตศาสตร์ของนักเรียนหญิงมีการกระจายมากกว่าคะแนนสอบวิชาคณิตศาสตร์ของนักเรียนชาย
- (4) คะแนนสอบวิชาคณิตศาสตร์ของนักเรียนหญิงมีการกระจายเบ้ลาดขวา

96. เจนนี่ต้องการซื้อถุงน่องแดง 1 ถุง ซึ่งทางร้านมีถุงน่องอยู่ 2 กล่อง กล่องแรกมีถุงน่องแดง 2 ถุง และถุงน่องเขียว 4 ถุง และกล่องที่สองมีถุงน่องแดง 10 ถุง และถุงน่องเขียว 30 ถุง จงพิจารณาข้อความต่อไปนี้ว่าเป็นอย่างไร

- (1) เจนนี่ ต้องซื้อถุงน่องจากกล่องที่ 1 จึงจะมีโอกาสได้ถุงน่องแดงมากกว่ากล่องที่ 2
- (2) เจนนี่ ต้องซื้อถุงน่องจากกล่องที่ 2 เพราะมีจำนวนถุงน่องแดงมากกว่ากล่องที่ 2
- (3) เจนนี่ ต้องซื้อถุงน่องจากกล่องที่ 2 จึงจะมีโอกาสได้ถุงน่องแดงมากกว่ากล่องที่ 1
- (4) เจนนี่ ซื้อถุงน่องจากกล่องใดก็มีโอกาสได้ถุงน่องแดงเท่ากัน

97. $(\sqrt{-125})^2 + (\sqrt[3]{-125})^3$ มีค่าเท่ากับเท่าใด

- (1) 0
- (2) 125
- (3) -125
- (4) 250

98. กำหนดให้ $\frac{3}{2}, 1, \frac{1}{2}, \dots$ เป็นลำดับเลขคณิตผลบวก 9 พจน์แรกเท่ากับข้อใด

- (1) $\frac{7}{2}$
- (2) $\frac{9}{2}$
- (3) $-\frac{7}{2}$
- (4) $-\frac{9}{2}$

99. จากการสัมภาษณ์ผู้เข้าชมรายการโทรทัศน์จำนวน 220 คน พบว่า

มี 140 คน ชอบดูรายการ “เกมส์พันหน้า”

มี 110 คน ชอบดูรายการ “ชิงร้อยชิงล้าน”

มี 105 คน ชอบดูรายการ “ตีสิบ”

มี 45 คน ชอบดูทั้ง เกมส์พันหน้าและตีสิบ

มี 40 คน ชอบดูทั้ง เกมส์พันหน้าและชิงร้อยชิงล้าน

มี 15 คน ชอบดูทั้งสามรายการ

ถ้าไม่มีผู้ชมคนใดที่ไม่ชอบดูทั้งสามรายการ จงหาจำนวนผู้ชมรายการโทรทัศน์ที่ชอบดูรายการดังกล่าวอย่างน้อยสองรายการ

- (1) 80 คน
- (2) 110 คน
- (3) 120 คน
- (4) 130 คน

100. จากแผนภาพต้นไม้ แสดงข้อมูลซึ่งเป็นความสูง (เซนติเมตร) ของนักเรียนกลุ่มหนึ่ง ดังนี้

13	8	9	7	7			
14	3	4	5	5	6		
15	5	8	1	3	2	11	4
16	1	3	2				

ข้อสรุปเป็นอย่างไรบ้าง

(1) จำนวนนักเรียนกลุ่มนี้ 20 คน

(2) ฐานนิยมเท่ากับ มัธยฐาน

(3) ค่าเฉลี่ยเลขคณิตเท่ากับ 149.25

(4) ค่าเฉลี่ยเลขคณิตมากกว่าฐานนิยม

101. ค่าของ $\sqrt{(-3)^2} - \left(\frac{1}{27^{\frac{1}{2}} + \sqrt{12}} \right)$ เท่ากับข้อใด

(1) $\frac{7}{4}$

(2) $\frac{5}{4}$

(3) 3

(4) 2

102. กำหนด $\sqrt{2} \approx 1.414, \sqrt{3} \approx 1.732$ แล้วค่าของ

$$\left(|3\sqrt{2} - 4\sqrt{5}| - |4\sqrt{5} - 5\sqrt{3}| - |3\sqrt{2} - 5\sqrt{3}| \right)^2$$

ตรงกับข้อใด

(1) 243

(2) 3

(3) 0

(4) -2.3

103. ค่าเฉลี่ยของน้ำหนักของนักเรียนในโรงเรียนแห่งหนึ่งเท่ากับ 48.01 กิโลกรัม ถ้าโรงเรียนนี้มีนักเรียนชาย 43 คน และนักเรียนหญิง 57 คน และค่าเฉลี่ยเลขคณิตของน้ำหนักของนักเรียนหญิงเท่ากับ 45 กิโลกรัม แล้วน้ำหนักของนักเรียนรวมทั้งหมดเท่าใด

(1) 2,365 กิโลกรัม

(2) 2,322 กิโลกรัม

(3) 2,277 กิโลกรัม

(4) 2,236 กิโลกรัม

104. ข้อใดต่างจากข้ออื่น

(1) $(-1)^{0.3}$

(2) $(-1)^{0.4}$

(3) $(-1)^{0.6}$

(4) $(-1)^{0.7}$

105. ข้อมูลชุดหนึ่งมีบางส่วนถูกนำเสนอในตาราง

อันตรภาคชั้น	ความถี่	ความถี่สัมพัทธ์
2-6	2	
3-11	8	0.4
12-16		x
17-21	6	

จงหาค่า x มีค่าเท่ากับเท่าไร

(1) 0.1

(2) 0.2

(3) 0.3

(4) 0.4

106. ปริมาณการส่งออกข้าวของประเทศไทยไปยัง 10 ประเทศ หน่วยเป็น (พันตัน) ดังนี้

ประเทศ	ปริมาณข้าวส่งออก (พันตัน)
กัมพูชา	42
จีน	255
ญี่ปุ่น	156
ไต้หวัน	22
ฟิลิปปินส์	389
มาเลเซีย	316
สิงคโปร์	270
อินโดนีเซีย	764
บรูไน	20
เกาหลีเหนือ	95

อยากทราบว่าเปอร์เซ็นต์ไทล์ที่ 75 ของปริมาณการส่งออกข้าวของประเทศไทยเป็นเท่าไร

- (1) 250.00 พันตัน (2) 275.25 พันตัน
(3) 334.25 พันตัน (4) 389.00 พันตัน

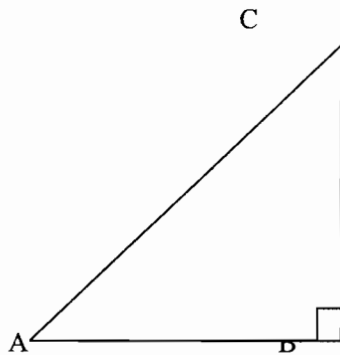
107. ค่าเฉลี่ยเลขคณิตของน้ำหนักของนักเรียนสามคนคือ 38 กิโลกรัม และนักเรียนหนึ่งคนในกลุ่มนั้นหนัก 46 กิโลกรัม ส่วนอีกสองคนที่เหลือหนักเท่ากัน จงหาว่านักเรียนสองคนที่เหลือหนักคนละกี่กิโลกรัม

- (1) 32 กิโลกรัม (2) 34 กิโลกรัม
(3) 36 กิโลกรัม (4) 38 กิโลกรัม

108. กำหนดให้ $\frac{3}{2}, 1, \frac{1}{2}, \dots$ เป็นลำดับเลขคณิต จงหาค่าของ $a_{40} - a_{30}$

- (1) -18 (2) -13 (3) -8 (4) -5

109. จากรูปสามเหลี่ยม ABC มีมุม $B = 90^\circ$ และ มุม $A = 35^\circ$ จงพิจารณาว่าข้อใดถูกหรือผิด



(1) $\sin 35^\circ = \cos 35^\circ$

(2) $\cos 35^\circ = \sec 35^\circ$

(3) $\tan 55^\circ = \cot 55^\circ$

(4) $\sec 55^\circ = \operatorname{cosec} 35^\circ$

110. กล้องวงจรปิดซึ่งถูกติดตั้งอยู่สูงจากพื้นถนน 3.0 เมตร สามารถจับภาพได้ต่ำสุดที่มุมก้ม 45° และจับภาพได้สูงที่สุดที่มุมก้ม 30° ระยะทางบนพื้นถนนในแนวที่กล้องนี้สามารถจับภาพได้ มีค่าตรงกับข้อใด ($\sqrt{3} \approx 1.732$)

(1) 2.196 เมตร

(2) 4.196 เมตร

(3) 1.46 เมตร

(4) 0.372 เมตร

111. พิจารณาข้อความต่อไปนี้ ว่าข้อใดบ้างถูกหรือผิด

(1) จำนวนที่เป็นทศนิยมไม่รู้จบทุกจำนวน เป็นจำนวนอตรรกยะ

(2) จำนวนที่เป็นทศนิยมไม่รู้จบทุกจำนวน เป็นจำนวนตรรกยะ

(3) จำนวนอตรรกยะทุกจำนวน เป็นจำนวนจริง

(4) เศษส่วนที่มีส่วนเป็นศูนย์เป็น จำนวนตรรกยะ

112. ข้อใดต่อไปนี้ถูกต้อง

(1) $\sin 30^\circ < \sin 45^\circ$

(2) $\cos 30^\circ < \cos 45^\circ$

(3) $\tan 45^\circ < \cot 45^\circ$

(4) $\tan 60^\circ < \cot 60^\circ$

113. ความสัมพันธ์ในข้อใดไม่เป็นฟังก์ชัน

(1) $\{(1,2), (2,3), (3,4), (4,5)\}$

(2) $\{(1,1), (2,2), (3,3), (4,4)\}$

(3) $\{(1,3), (2,4), (3,1), (4,2)\}$

(4) $\{(1,3), (1,2), (1,1), (1,4)\}$

114. กำหนดให้ $U = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$

$$A = \{2, 4, 7, 8, 9\}$$

$$B = \{x / x^2 - 10x + 21 = 0\}$$

$$C = \{x / 2 < x \leq 6\}$$

แล้ว $A \cap (B' - C)$ ตรงกับข้อใด

(1) $\{2\}$

(2) $\{2, 8\}$

(3) $\{2, 8, 9\}$

(4) $\{2, 4, 8, 9\}$

115. พิจารณาการให้เหตุผลแต่ละข้อต่อไปนี้

1. เหตุ 1. ถ้าเป็นนักบาสเกตบอลแล้วต้องเป็นนักวิ่ง

2. สมชายเป็นนักบาสเกตบอล

ผล สมชายเป็นนักวิ่ง

2. เหตุ 1. ถ้าสายตาเอียงแล้วต้องใส่แว่น

2. รุ่งระวีไม่ใส่แว่น

ผล รุ่งระวีไม่ได้สายตาเอียง

การให้เหตุผลข้อใดสมเหตุสมผล

(1) ข้อ 1 เท่านั้น

(2) ข้อ 2 เท่านั้น

(3) ข้อ 1 และ ข้อ 2

(4) ไม่ใช่ทั้งข้อ 1 และข้อ 2

116. ถ้า -2 เป็นผลเฉลยหนึ่งของสมการ $4x^2 + bx - 6 = 0$ เมื่อ b เป็นจำนวนจริงแล้ว
อีกผลเฉลยหนึ่งของสมการนี้ตรงกับข้อใด

(1) 2

(2) $\frac{1}{2}$

(3) $\frac{3}{4}$

(4) $-\frac{1}{2}$

117. ให้ a, b, c เป็นจำนวนเต็ม ข้อสรุปใดต่อไปนี้จริง

(1) ถ้า $a|b$ และ $b|c$ แล้ว $ab|c$

(2) ถ้า $a|c$ และ $b|c$ แล้ว $c|ab$

(3) ถ้า $a|b$ และ $b|c$ แล้ว $a|c$

(4) ถ้า $a|c$ และ $b|c$ แล้ว $a|b$

118. กำหนดให้ a และ b เป็นจำนวนจริงลบทั้งคู่ โดยที่ $a < x < b$ แล้วข้อใดต่อไปนี้ เป็นจริง

(1) $|x| + a > 0$

(2) $|x| + b < 0$

(3) $\frac{1}{|x|} < \frac{1}{|b|}$

(4) $\frac{1}{|x|} < \frac{1}{|a|}$

119. ค่าของ $(\sqrt{5} - 2)^{-2}$ เป็นจริงตามข้อใด

(1) เป็นจำนวนอตรรกยะที่น้อยกว่า 17.9 (2) เป็นจำนวนอตรรกยะที่มากกว่า 17.9

(3) เป็นจำนวนตรรกยะที่น้อยกว่า 17.9 (4) เป็นจำนวนตรรกยะที่มากกว่า 17.9

120. ถ้า $m = \frac{\sqrt{3} - \sqrt{5}}{\sqrt{3} + \sqrt{5}}$ และ $n = \frac{\sqrt{3} + \sqrt{5}}{\sqrt{3} - \sqrt{5}}$ แล้ว $m^2 + 8mn + n^2$ เท่ากับเท่าใด

(1) 70

(2) 80

(3) 90

(4) 100

121. ค่า $4(3)^{\frac{1}{3}} - (24)^{\frac{1}{3}} - (576)^{\frac{1}{6}} + 6\left(\frac{1}{81}\right)^{\frac{1}{6}}$ เท่ากับข้อใดต่อไปนี้

(1) $\sqrt[3]{2}$

(2) $2\sqrt[3]{3}$

(3) $4\sqrt[3]{5}$

(4) 2

122. พิจารณาข้อความต่อไปนี้ ว่าข้อใดถูกหรือผิด

(1) $\frac{2xy + 3x\sqrt{x} - 2y\sqrt{y} - 3\sqrt{xy}}{x - \sqrt{y}} = 2y + 3\sqrt{x}$

(2) $\frac{2a\sqrt{ab} + a^2\sqrt{b} + 2b\sqrt{ab} + ab\sqrt{b}}{2\sqrt{ab} + a\sqrt{b}} = a + b$

(3) ถ้า $4 \times 4^{-1} = 4^{1+(-1)} = 4^0 = 1$ เป็นสมบัติการมีเอกลักษณ์การคูณของจำนวนจริง

(4) ถ้า $a \in \mathbb{R}$, $b \in \mathbb{R}$ และ $c \in \mathbb{R}$ แล้ว $a + (b + c) = (a + b) + c$ เป็นสมบัติการสลับที่การบวก

123. อยู่บนยอดเสากระโดงเรือลำหนึ่งซึ่งสูง 160 ฟุต มองเห็นเรืออีกลำหนึ่งลอยอยู่ในทะเล เป็นมุมก้ม 30° แล้ว ระยะห่างระหว่างเรือสองลำเท่ากับเท่าใด

(1) 234.63 ฟุต

(2) 245.16 ฟุต

(3) 268.31 ฟุต

(4) 277.12 ฟุต

124. ชายคนหนึ่งเห็นยอดเจดีย์เป็นมุมเมย α เมื่อเดินเข้าไปใกล้เจดีย์ 20 ฟุต มุมเมยเปลี่ยนเป็น 48° และเดินต่อไปอีก 15 ฟุต มุมเมยเป็นมุมประกอบหนึ่งมุมฉากของมุมเมยครั้งแรก แล้วความสูงของเจดีย์เท่ากับข้อใด

(1) 54 ฟุต

(2) 56 ฟุต

(3) 58 ฟุต

(4) 60 ฟุต

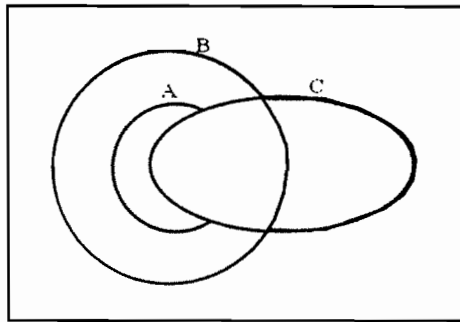
125. กำหนด $\frac{3}{2}, 1, \frac{1}{2}, \dots$ เป็นลำดับเลขคณิต ผลบวกของพจน์ที่ 40 และพจน์ที่ 42 เท่าใด

- (1) 44 (2) 43
(3) -37 (4) -38

126. จำนวนที่อยู่ระหว่าง 200 ถึง 900 ที่หารด้วย 2 ลงตัว แต่หารด้วยสามไม่ลงตัวมีกี่จำนวน

- (1) 233 จำนวน (2) 234 จำนวน
(3) 235 จำนวน (4) 236 จำนวน

127. แผนภาพที่แรเงาแทนเซตในข้อใด



- (1) $(C - A) \cup (C - B)$
(2) $(A' \cap C) \cup (A' \cap B)$
(3) $(A \cap B) \cup (A \cup B)'$
(4) $(A - C) \cup (C - B)$

128. พิจารณาการอ้างเหตุผลต่อไปนี้

เหตุ 1. เฉชะ สอบ O-NET ผ่าน หรือ เฉชะ จะสอบ PIZA ผ่าน

2. เฉชะ สอบ O-NET ไม่ผ่าน

ผล เฉชะ สอบ PIZA ไม่ผ่าน

เหตุ 1. ถ้าฝนไม่ตก แล้ว อมรรัตน์ไปโรงเรียน

2. ฝนไม่ตก

ผล อมรรัตน์ไม่ไปโรงเรียน

ข้อใดต่อไปนี้ถูกหรือผิด

- (1) ข้อ 1 สมเหตุสมผล และ ข้อ 2 สมเหตุสมผล
(2) ข้อ 1 สมเหตุสมผล และ ข้อ 2 ไม่สมเหตุสมผล
(3) ข้อ 1 ไม่สมเหตุสมผล และ ข้อ 2 สมเหตุสมผล
(4) ข้อ 1 และ ข้อ 2 ไม่สมเหตุสมผล

129. กำหนด $2^{x-3} = 8^{-(x-5)}$ แล้ว ผลบวกของคำตอบมีค่าเท่าใด

- (1) 6 (2) 3
(3) -3 (4) -6

130. กำหนด $f(x) = x + 5$, $g(x + 5) = x - 7$ แล้ว $f(x) \cdot g(x)$ มีค่าเท่าใด
- (1) $x^2 + 7x - 60$ (2) $x^2 - 7x - 60$
 (3) $x^2 - 7x + 60$ (4) $x^2 + 7x + 60$
131. กราฟใดต่อไปนี้ไม่เป็นฟังก์ชัน
- (1) $y = 3x^2 - 5$ (2) $y = 2x - 5$
 (3) $x^2 + y^2 = 9$ (4) $y = |x + 1|$
132. ลำดับ 2, 4, 8, ..., 1024 มีจำนวน n พจน์ แล้วผลบวก n พจน์แรกของอนุกรม $3 + 6 + 9 + \dots$ เท่ากับเท่าใด
- (1) 165 (2) 166 (3) 167 (4) 168
133. วิธีการสำคัญในการสำรวจความคิดเห็น ประกอบด้วยข้อใดบ้างต่อไปนี้
- (1) กำหนดขอบเขตของการสำรวจ
 (2) สร้างแบบสำรวจความคิดเห็น
 (3) วิธีเลือกตัวอย่าง
 (4) การประมวลผลและวิเคราะห์ความคิดเห็น
134. การสุ่มชั่งน้ำหนัก (กิโลกรัม) ของมันสำปะหลังจำนวน 15 กระสอบ ซึ่งมีน้ำหนักดังนี้
 49 52 47 53 55 48 50 50 54 52 51 52 49 50 53
 จงหาส่วนเบี่ยงเบนมาตรฐานของน้ำหนักมันสำปะหลัง
- (1) 2.15 (2) 2.20
 (3) 2.27 (4) 2.37
135. แผนภาพต้น-ใบ ของข้อมูลชุดหนึ่งเป็นดังนี้
- 1 0 0 3 5 8
 2 1 4 4 6 7
 3 3 3 5 7
 4 1 2 2 2
 5 3 5
- พิจารณาข้อความต่อไปนี้
- (1) ข้อมูลชุดนี้ไม่มีฐานนิยม (2) มัธยฐานของข้อมูลชุดนี้เท่ากับ 42
 (3) ค่าเฉลี่ยเลขคณิตเท่ากับ 31.7 (4) ค่าพิสัยของข้อมูล เท่ากับ 27

136. สำรวจความนิยมที่ชอบดารา จากคน 20 คน ดังนี้

ชื่อที่	ระดับความนิยม					รวม
	5	4	3	2	1	
1. ศรราม	5	5	3	3	4	20
2. ณเดช	6	5	3	3	3	20

เกณฑ์การแปรผล (\bar{x})

4.00-5.00 มากที่สุด

3.00-3.99 มาก

2.00-1.99 ปานกลาง

1.00-1.99 น้อย

0.00-0.99 น้อยที่สุด

ข้อใดแปรค่าได้ถูกต้อง

- (1) ศรราม แปรผล มาก และ ณเดช แปรผล มาก
- (2) ศรราม แปรผล มาก และ ณเดช แปรผล น้อย
- (3) ศรราม แปรผล น้อย และ ณเดช แปรผล น้อย
- (4) ศรราม แปรผล น้อย และ ณเดช แปรผล มาก

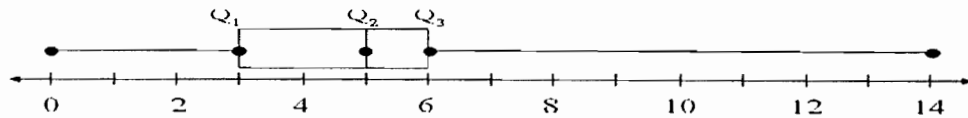
137. หมายเลขประจำตัวประกอบด้วยตัวเลข 7 ตัว ข้อใดต่อไปนี้เป็นจำนวนหมายเลขที่ขึ้นต้นด้วย 427

- (1) 10,000
- (2) 5,040
- (3) 840
- (4) 210

138. กำหนด $P(A) = \frac{2}{3}$, $P(A \cup B) = \frac{3}{4}$, $P(A' \cup B') = \frac{1}{2}$ แล้ว $P(B')$ มีค่าเท่าใด

- (1) $\frac{4}{12}$
- (2) $\frac{5}{12}$
- (3) $\frac{6}{12}$
- (4) $\frac{7}{12}$

139. จากการตรวจสอบปริมาณน้ำตาล (กรัม) ต่อปริมาณขนม 100 กรัม ของขนม 31 ชนิด
ในร้านค้า ได้ข้อมูลดังนี้



จงพิจารณาข้อความต่อไปนี้

- (1) ข้อมูลมีค่าอยู่ระหว่าง 6–14 คือ ประมาณ 35 %
 - (2) ข้อมูลมีค่าอยู่ระหว่าง 3–6 คือ ประมาณ 60 %
 - (3) ข้อมูลมีมัธยฐานเท่ากับ Q_2
 - (4) ข้อมูลที่อยู่ระหว่าง Q_1 และ Q_2 มีการกระจายน้อยกว่าข้อมูลที่อยู่ระหว่าง Q_2 และ Q_3
140. ในการสอบวิชาคณิตศาสตร์ครั้งหนึ่งมีนักเรียนเข้าสอบจำนวน 20 คน ความน่าจะเป็น
ที่นักเรียนคนหนึ่งจะได้คะแนนไม่ต่ำกว่า 16 คะแนน คือข้อใด
- ผลสอบดังตาราง

คะแนน	ความถี่
5-10	3
11-15	13
16-20	4

- (1) $\frac{1}{3}$
 - (2) $\frac{1}{4}$
 - (3) $\frac{1}{5}$
 - (4) $\frac{1}{13}$
141. จงพิจารณาการให้เหตุผลต่อไปนี้
- (1) เหตุ 1. สัตว์เลี้ยงทุกตัวเป็นสัตว์ไม่ดุร้าย
2. แมวทุกตัวเป็นสัตว์เลี้ยง
ผล แมวทุกตัวเป็นสัตว์ไม่ดุร้าย
 - (2) เหตุ 1. นักเรียน ม.6 ทุกคนแต่งกายถูกระเบียบ
2. สมชายเป็นนักเรียนชั้น ม.6
ผล สมชายเป็นนักเรียน

(3) เหตุ 1. วันที่มีฝนตกทั้งวันจะมีท้องฟ้ามีดครึ้มทุกวัน

2. วันนี้ท้องฟ้ามีดครึ้ม

ผล วันนี้ฝนตกทุกวัน

(4) เหตุ 1. สัตว์เลี้ยงลูกด้วยนมเป็นสัตว์เลี้ยงลูกด้วยนม

2. สุนัขเลี้ยงลูกด้วยนม

ผล สุนัขเป็นสัตว์เลี้ยงลูกด้วยนม

142. เหตุ 1. หมูอวกาศทุกตัวบินได้

2. โน้ตบินได้

ผลจงพิจารณาการให้เหตุผลต่อไปนี้

(1) โน้ตเป็นหมูอวกาศ (2) โน้ตและหมูอวกาศบินได้

(3) หมูอวกาศบินได้ (4) โน้ตบินได้

143. พิจารณาข้อใดต่อไปนี้

นิดออกจากบ้านเพื่อไปโรงเรียน นิดพบว่า

ถ้า นิดออกจากบ้านเวลา 6.30 น. นิดจะถึงโรงเรียนเวลา 7.30 น.

ถ้า นิดออกจากบ้านเวลา 6.50 น. นิดจะถึงโรงเรียนเวลา 7.55 น.

ถ้า นิดออกจากบ้านเวลา 6.00 น. นิดจะถึงโรงเรียนเวลา 7.05 น.

ถ้า นิดอยากไปถึงโรงเรียน ก่อนเวลา 7.00 น. นิดควรออกจากบ้านเวลาใด

สรุปสถานการณ์ที่ใช้การให้เหตุผลแบบอุปนัย

(1) ออกจากบ้านเวลา 5.50 น.

(2) ออกจากบ้านเวลา 6.00 น.

(3) ออกจากบ้านเวลา 6.15 น.

(4) ออกจากบ้านเวลา 6.45 น.

144. จงพิจารณาข้อใดต่อไปนี้ใช้สมบัติการเปลี่ยนกลุ่มการบวกและการสลับที่ของการบวก

(1) $(1+2)+3 = 3+(2+1)$ (2) $(1+2)+3 = 3+(1+2)$

(3) $(1+2)+3 = (2+3)+1$ (4) $(1+2)+3 = 1+(2+3)$

145. ถ้า a และ b เป็นจำนวนจริงแล้ว จงพิจารณาข้อความต่อไปนี้

(1) $a(-b) = -ab$ (2) $\frac{a}{b} = a(b^{-1})$

(3) ถ้า $ab = 0$ แล้ว $a = 0$ หรือ $b = 0$ (4) $(-a)(-b) = ab$

146. ถ้า $\tan A = \frac{8}{15}$ แล้ว $\cos A$ มีค่าเท่ากับเท่าไร

(1) $\cos A = \frac{5}{17}$ (2) $\cos A = \frac{8}{17}$

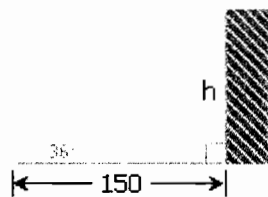
(3) $\cos A = \frac{15}{17}$ (4) $\cos A = 1$

147. ถ้า $\sin A = \frac{1}{2}$ แล้ว $\tan A + \cos A$ มีค่าเท่ากับเท่าไร

(1) $\frac{2\sqrt{3}}{5}$ (2) $\frac{2}{5\sqrt{3}}$

(3) $\frac{5\sqrt{3}}{2}$ (4) $\frac{5\sqrt{3}}{6}$

148. นักเรียนยืนห่างจากตึกแห่งหนึ่ง 150 เมตร เมื่อมองขึ้นไปบนยอดตึกเป็นมุมเงยขนาด 36 องศา อยากทราบว่า ตึกนี้สูงประมาณกี่เมตร โดยไม่คิดความสูงของนักเรียน



(1) 106 เมตร (2) 107 เมตร

(3) 108 เมตร (4) 109 เมตร

149. อินเวอร์สการคูณของ $1 - \sqrt{5}$ มีค่าเท่ากับเท่าไร

(1) $1 + \sqrt{5}$ (2) $1 - \sqrt{5}$

(3) $\frac{1 + \sqrt{7}}{4}$ (4) $\frac{1 + \sqrt{7}}{-4}$

150. จงพิจารณาข้อความต่อไปนี้ ว่าถูกหรือผิด

(1) เมื่อ a, b เป็นจำนวนจริงใด ๆ $a - b = a + (-b)$

(2) $\left(\frac{a}{b}\right)^{-1} = \frac{b}{a}$

(3) $\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}$

(4) $\left(\frac{a}{b}\right)\left(\frac{c}{d}\right) = \frac{ac}{bd}$

ภาคผนวก ข

รายชื่อผู้เชี่ยวชาญตรวจสอบเครื่องมือการวิจัย

รายชื่อผู้เชี่ยวชาญตรวจสอบเครื่องมือการวิจัย

ด้านโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

1. ผู้ช่วยศาสตราจารย์ ดร.สมศักดิ์ จีวัฒนา การศึกษาระดับการศึกษาปริญญาเอก
ผู้อำนวยการสำนักวิทยบริการและเทคโนโลยีสารสนเทศสารสนเทศ
อาจารย์ประจำสาขาวิชาวิทยาการคอมพิวเตอร์คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์
2. ดร.ชนาวุฒิ ประกอบผล การศึกษาระดับการศึกษาปริญญาเอก
อาจารย์ประจำวิทยาลัยการบริหารและการจัดการ
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
3. นายพีระพันธ์ เพชรสุวรรณ การศึกษาระดับการศึกษาปริญญาโท
ครูวิทยฐานะชำนาญการพิเศษ หัวหน้างานวิทยาการคอมพิวเตอร์ โรงเรียนบุรีรัมย์พิทยาคม
อำเภอเมือง จังหวัดบุรีรัมย์

ด้านข้อสอบวิชาคณิตศาสตร์

1. ดร.กระพัน ศรีงาน ระดับการศึกษาปริญญาเอก
คณบดีคณะครุศาสตร์ และอาจารย์ประจำสาขาวิชาทดสอบและวิจัยการศึกษา
มหาวิทยาลัยราชภัฏบุรีรัมย์
2. ดร.โกวิท วัชรินทรางกูร ระดับการศึกษาปริญญาเอก
อาจารย์ประจำ สาขาวิชาหลักสูตรและการสอน คณะครุศาสตร์
มหาวิทยาลัยราชภัฏบุรีรัมย์
3. นางสาวพวงเพชร ราชประโคน ระดับการศึกษาปริญญาโท (การสอนคณิตศาสตร์)
สาขาวิชาคณิตศาสตร์ ภาควิชาคณิตศาสตร์และคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี
มหาวิทยาลัยราชภัฏบุรีรัมย์
4. นายพีระวัฒน์ เสวตรพัชร ระดับการศึกษาปริญญาโท (หลักสูตรและการสอน)
ศึกษานิเทศก์ชำนาญการพิเศษ หัวหน้ากลุ่มงานหลักสูตร
สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 32
5. นายอัครณัฐ บุญมะยา ระดับการศึกษาปริญญาโท (คณิตศาสตร์)
ครูวิทยฐานะชำนาญการพิเศษ โรงเรียนบุรีรัมย์พิทยาคม อำเภอเมือง จังหวัดบุรีรัมย์

ภาคผนวก ก

ค่าพารามิเตอร์ของข้อสอบในคลังข้อสอบ

ค่าพารามิเตอร์ของข้อสอบในคลังข้อสอบ

Test No.	Item Bank No	Standard / Indicator	Item Test No	Keys	IRT-a	IRT-b	IRT-c
001	23	ค 1.1/1	1	1	0.8060	2.3910	0.1440
001	24	ค 1.1/2	2	1	0.4050	1.2550	0.2050
001	1	ค 1.1/3	3	4	2.6910	0.6030	0.1890
001	4	ค 1.2/1	4	3	0.7650	2.2810	0.1980
001	28	ค 1.3/1	5	3	0.9900	2.2800	0.1060
001	29	ค 1.4/1	6	2	0.7230	2.3190	0.1120
001	116	ค 1.4/1 ,ค 4.2/3	7	3	0.9350	2.4530	0.1040
001	30	ค 2.1/1	8	3	0.9690	2.2580	0.1170
001	78	ค 2.1/2	9	2	0.7730	1.7820	0.1410
001	31	ค 2.2/1	10	1	0.9290	2.0020	0.1240
001	32	ค 4.1/1	11	4	0.7120	3.3170	0.1210
001	7	ค 4.1/2	12	2	0.8470	0.4860	0.1800
001	12	ค 4.1/3	13	3	0.0760	4.4420	0.0060
001	9	ค 4.1/4	14	3	0.8720	1.2820	0.1340
001	10	ค 4.1/5	15	4	0.3200	0.5440	0.1230
001	38	ค 4.2/1	16	1	0.3220	0.5060	0.1710
001	20	ค 4.2/1 , ค 4.1/2	17	4	0.6880	0.8510	0.2240
001	39	ค 4.2/2	18	4	0.7320	1.4440	0.0930
001	40	ค 4.2/3	20	4	0.4850	1.2120	0.2420
001	13	ค 4.2/4	21	2	1.5780	1.8730	0.0220
001	14	ค 4.2/5	22	4	0.2370	0.8530	0.2450
001	8	ค 4.2/6	23	4	0.8500	0.8960	0.2240
001	15	ค 5.1/1	24	1	0.3380	1.4210	0.2260
001	16	ค 5.1/2	25	3	0.6700	4.2460	0.0360

ค่าพารามิเตอร์ของข้อสอบในคลังข้อสอบ

Test No.	Item Bank No	Standard / Indicator	Item Test No	Keys	IRT-a	IRT-b	IRT-c
001	17	ค 5.1/3	26	2	0.1710	1.2710	0.2620
001	18	ค 5.2/1	27	4	0.9090	2.6270	0.0260
001	46	ค 5.2/2	28	3	0.8530	1.0860	0.1700
001	19	ค 5.3/1	29	2	0.3060	1.1590	0.1130
001	49	ค 5.3/2	30	2	0.8950	0.5990	0.2400
002	50	ค 1.1/1	1	1	0.7680	2.9670	0.1330
002	25	ค 1.1/2	2	4	0.9510	2.1070	0.1100
002	2	ค 1.1/3	3	2	0.8920	2.2850	0.1530
002	5	ค 1.2/1	4	3	0.9630	1.5520	0.1660
002	54	ค 1.3/1	5	4	0.8620	2.4650	0.1480
002	55	ค 1.4/1	6	3	0.7250	0.1230	0.2070
002	116	ค 1.4/1 ,ค 4.2/3	7	3	0.9350	2.4530	0.1040
002	56	ค 2.1/1	8	2	0.8710	3.2050	0.0640
002	78	ค 2.1/2	9	2	0.7730	1.7820	0.1410
002	57	ค 2.2/1	10	3	0.9410	2.2370	0.1180
002	58	ค 4.1/1	11	2	0.8290	2.2280	0.1710
002	33	ค 4.1/2	12	3	1.1950	1.6000	0.1160
002	35	ค 4.1/4	14	2	0.2190	1.1550	0.1070
002	36	ค 4.1/5	15	4	1.0160	0.6340	0.1640
002	63	ค 4.2/1	16	3	0.5490	1.3140	0.2220
002	21	ค 4.2/1 , ค 4.1/2	17	4	0.2160	1.0870	0.2300
002	64	ค 4.2/2	18	2	0.4520	0.9580	0.2530
002	86	ค 4.2/3	20	4	0.5470	1.4400	0.0980
002	41	ค 4.2/4	21	1	0.3020	1.0530	0.1360

ค่าพารามิเตอร์ของข้อสอบในคลังข้อสอบ

Test No.	Item Bank No	Standard / Indicator	Item Test No	Keys	IRT-a	IRT-b	IRT-c
002	37	ค 4.2/5	22	2	0.6350	0.7120	0.1900
002	11	ค 4.2/6	23	2	3.1470	1.2300	0.0160
002	66	ค 5.1/1	24	3	0.2600	0.9880	0.2160
002	42	ค 5.1/2	25	2	0.1540	1.2440	0.2240
002	45	ค 5.1/3	26	4	0.1900	1.2770	0.2450
002	69	ค 5.2/1	27	4	0.1510	1.3610	0.0750
002	47	ค 5.2/2	28	3	0.2960	0.5180	0.2010
002	48	ค 5.3/1	29	2	0.3950	0.5390	0.0940
002	96	ค 5.3/2	30	4	0.5750	1.0120	0.2500
003	70	ค 1.1/1	1	4	0.8340	2.3720	0.1600
003	51	ค 1.1/2	2	3	0.7200	4.3780	0.1150
003	3	ค 1.1/3	3	3	1.8950	0.6300	0.0770
003	6	ค 1.2/1	4	4	0.7290	2.1910	0.2000
003	75	ค 1.3/1	5	2	1.2570	1.8020	0.0960
003	76	ค 1.4/1	6	1	0.7610	2.0340	0.0820
003	116	ค 1.4/1 ,ค 4.2/3	7	3	0.9350	2.4530	0.1040
003	77	ค 2.1/1	8	4	0.2600	0.9010	0.2510
003	78	ค 2.1/2	9	2	0.7730	1.7820	0.1410
003	110	ค 2.2/1	10	4	1.0280	2.2250	0.1140
003	79	ค 4.1/1	11	3	0.8000	2.8330	0.1550
003	59	ค 4.1/2	12	2	0.7610	2.9970	0.1050
003	60	ค 4.1/3	13	4	0.1860	1.2940	0.1950
003	61	ค 4.1/4	14	3	0.1990	1.3420	0.0750
003	62	ค 4.1/5	15	4	0.4830	1.4710	0.0740

ค่าพารามิเตอร์ของข้อสอบในคลังข้อสอบ

Test No.	Item Bank No	Standard / Indicator	Item Test No	Keys	IRT-a	IRT-b	IRT-c
003	84	ท 4.2/1	16	3	0.6650	4.6880	0.0210
003	20	ท 4.2/1 , ท 4.1/2	17	4	0.6880	0.8510	0.1350
003	115	ท 4.2/2	18	4	2.4200	0.6180	0.1130
003	87	ท 4.2/4	21	2	0.4740	0.5500	0.0950
003	43	ท 4.2/5	22	3	0.2990	1.3460	0.2550
003	65	ท 4.2/6	23	4	0.6340	0.6660	0.1140
003	90	ท 5.1/1	24	4	0.2670	1.0850	0.1110
003	44	ท 5.1/2	25	2	0.3050	0.5240	0.1130
003	68	ท 5.1/3	26	4	0.1910	1.2450	0.2330
003	95	ท 5.3/1	29	2	0.3690	1.0580	0.1570
004	111	ท 1.1/1	1	4	0.6690	2.0990	0.1910
004	71	ท 1.1/2	2	3	0.7010	3.6710	0.0910
004	26	ท 1.1/3	3	3	0.1880	1.0610	0.1260
004	27	ท 1.2/1	4	4	0.7080	1.7020	0.2630
004	121	ท 1.3/1	5	2	0.8960	2.4770	0.1570
004	122	ท 1.4/1	6	1	0.9050	2.5120	0.1280
004	116	ท 1.4/1 , ท 4.2/3	7	3	0.9350	2.4530	0.1040
004	109	ท 2.1/1	8	4	0.8570	2.5420	0.1370
004	78	ท 2.1/2	9	2	0.7730	1.7820	0.1410
004	124	ท 2.2/1	10	4	1.1730	0.8710	0.1370
004	114	ท 4.1/1	11	3	0.9650	2.6120	0.0930
004	80	ท 4.1/2	12	2	1.1900	0.7710	0.2440
004	81	ท 4.1/3	13	4	0.4560	0.9100	0.1350
004	82	ท 4.1/4	14	3	0.1370	1.3670	0.1040

ค่าพารามิเตอร์ของข้อสอบในคลังข้อสอบ

Test No.	Item Bank No	Standard / Indicator	Item Test No	Keys	IRT-a	IRT-b	IRT-c
004	83	ค 4.1/5	15	4	0.1320	0.9310	0.1120
004	99	ค 4.2/1	16	3	0.2320	1.3780	0.1860
004	21	ค 4.2/1 , ค 4.1/2	17	4	0.2160	1.0870	0.0860
004	40	ค 4.2/3	20	4	0.4850	1.2120	0.0940
004	88	ค 4.2/5	22	3	0.2340	0.5370	0.2140
004	98	ค 4.2/6	23	4	0.1720	1.1070	0.1820
004	67	ค 5.1/2	25	2	0.2740	1.0610	0.1930
004	92	ค 5.1/3	26	4	0.1300	0.8860	0.1310
004	105	ค 5.3/1	29	2	1.8450	0.5570	0.2170
005	117	ค 1.1/1	1	3	0.8390	2.3570	0.1600
005	72	ค 1.1/2	2	4	0.8660	1.7620	0.1680
005	52	ค 1.1/3	3	1	0.9430	2.2510	0.1180
005	53	ค 1.2/1	4	2	0.6980	3.1340	0.1280
005	28	ค 1.3/1	5	3	0.9900	2.2800	0.1060
005	29	ค 1.4/1	6	2	0.7230	2.3190	0.1120
005	116	ค 1.4/1 , ค 4.2/3	7	3	0.9350	2.4530	0.1040
005	112	ค 2.1/1	8	1	0.6120	0.8090	0.1710
005	78	ค 2.1/2	9	2	0.7730	1.7820	0.1410
005	31	ค 2.2/1	10	1	0.9290	2.0020	0.1240
005	32	ค 4.1/1	11	4	0.7120	3.3170	0.1210
005	85	ค 4.1/2	12	3	0.8350	2.3530	0.1550
005	113	ค 4.1/3	13	4	0.6910	1.4700	0.1480
005	108	ค 4.1/5	15	4	1.4370	0.8410	0.2130
005	20	ค 4.2/1 , ค 4.1/2	17	4	0.6880	0.8510	0.0720

ค่าพารามิเตอร์ของข้อสอบในคลังข้อสอบ

Test No.	Item Bank No	Standard / Indicator	Item Test No	Keys	IRT-a	IRT-b	IRT-c
005	39	ค 4.2/2	18	4	0.7320	1.4440	0.1260
005	86	ค 4.2/3	20	4	0.5470	1.4400	0.2380
005	13	ค 4.2/4	21	2	1.5780	1.8730	0.0220
005	89	ค 4.2/5	22	3	0.5180	1.3100	0.2130
005	90	ค 5.1/1	24	2	0.2670	1.0850	0.2130
005	91	ค 5.1/2	25	2	0.5480	1.2170	0.0860
005	100	ค 5.1/3	26	4	0.2530	1.0520	0.0960
005	18	ค 5.2/1	27	4	0.9090	2.6270	0.0260
005	46	ค 5.2/2	28	3	0.8530	1.0860	0.1720
005	49	ค 5.3/2	30	2	0.8950	0.5990	0.1930
006	23	ค 1.1/1	1	1	0.8060	2.3910	0.1440
006	97	ค 1.1/2	2	1	0.1790	0.7100	0.0870
006	73	ค 1.1/3	3	4	0.2450	0.7960	0.2420
006	74	ค 1.2/1	4	3	0.7590	1.2830	0.1420
006	54	ค 1.3/1	5	4	0.8620	2.4650	0.1480
006	55	ค 1.4/1	6	3	0.7250	0.1230	0.2070
006	116	ค 1.4/1 ,ค 4.2/3	7	3	0.9350	2.4530	0.1040
006	123	ค 2.1/1	8	4	1.1930	0.0380	0.1760
006	78	ค 2.1/2	9	2	0.7730	1.7820	0.1410
006	57	ค 2.2/1	10	3	0.9410	2.2370	0.1180
006	58	ค 4.1/1	11	2	0.8290	2.2280	0.1710
006	7	ค 4.1/2	12	2	0.8470	0.4860	0.1800
006	12	ค 4.1/3	13	3	0.0760	4.4420	0.0060
006	9	ค 4.1/4	14	3	0.8720	1.2820	0.0760

ค่าพารามิเตอร์ของข้อสอบในคลังข้อสอบ

Test No.	Item Bank No	Standard / Indicator	Item Test No	Keys	IRT-a	IRT-b	IRT-c
006	38	ค 4.2/1	16	4	0.3220	0.5060	0.2180
006	21	ค 4.2/1 , ค 4.1/2	17	4	0.2160	1.0870	0.0940
006	64	ค 4.2/2	18	2	0.4520	0.9580	0.2590
006	41	ค 4.2/4	21	1	0.3020	1.0530	0.1690
006	98	ค 4.2/6	23	4	0.1720	1.1070	0.1160
006	103	ค 5.1/2	25	4	0.8730	1.3490	0.1680
006	69	ค 5.2/1	27	4	0.1510	1.3610	0.0880
006	47	ค 5.2/2	28	3	0.2960	0.5180	0.1470
006	19	ค 5.3/1	29	2	0.3060	1.1590	0.1300
006	96	ค 5.3/2	30	4	0.5750	1.0120	0.0730
007	50	ค 1.1/1	1	1	0.7680	2.9670	0.1330
007	118	ค 1.1/2	2	3	0.5180	1.3710	0.1800
007	102	ค 1.1/3	3	3	0.5090	1.3340	0.2520
007	101	ค 1.2/1	4	1	0.6340	3.4430	0.1610
007	75	ค 1.3/1	5	2	1.2570	1.8020	0.0960
007	76	ค 1.4/1	6	2	0.7610	2.0340	0.0820
007	116	ค 1.4/1 ,ค 4.2/3	7	3	0.9350	2.4530	0.1040
007	30	ค 2.1/1	8	3	0.9690	2.2580	0.1170
007	78	ค 2.1/2	9	2	0.7730	1.7820	0.1410
007	110	ค 2.2/1	10	1	1.0280	2.2250	0.1140
007	79	ค 4.1/1	11	4	0.8000	2.8330	0.1550
007	33	ค 4.1/2	12	3	1.1950	1.6000	0.1160
007	35	ค 4.1/4	14	2	0.2190	1.1550	0.2300
007	10	ค 4.1/5	15	4	0.3200	0.5440	0.0840

ค่าพารามิเตอร์ของข้อสอบในคลังข้อสอบ

Test No.	Item Bank No	Standard / Indicator	Item Test No	Keys	IRT-a	IRT-b	IRT-c
007	63	ค 4.2/1	16	3	0.5490	1.3140	0.2250
007	20	ค 4.2/1 , ค 4.1/2	17	4	0.6880	0.8510	0.1230
007	115	ค 4.2/2	18	3	2.4200	0.6180	0.2500
007	40	ค 4.2/3	20	4	0.4850	1.2120	0.1740
007	87	ค 4.2/4	21	3	0.4740	0.5500	0.0820
007	14	ค 4.2/5	22	4	0.2370	0.8530	0.1010
007	8	ค 4.2/6	23	4	0.8500	0.8960	0.2250
007	15	ค 5.1/1	24	1	0.3380	1.4210	0.1340
007	106	ค 5.1/2	25	3	0.3280	0.8130	0.2440
007	17	ค 5.1/3	26	2	0.1710	1.2710	0.1960
007	48	ค 5.3/1	29	2	0.3950	0.5390	0.2180
008	70	ค 1.1/1	1	1	0.8340	2.3720	0.1600
008	24	ค 1.1/2	2	1	0.4050	1.2550	0.1520
008	104	ค 1.1/3	3	2	0.8480	2.1280	0.1580
008	120	ค 1.2/1	4	1	0.9340	1.6680	0.1650
008	121	ค 1.3/1	5	2	0.8960	2.4770	0.1570
008	122	ค 1.4/1	6	1	0.9050	2.5120	0.1280
008	116	ค 1.4/1 , ค 4.2/3	7	3	0.9350	2.4530	0.1040
008	56	ค 2.1/1	8	2	0.8710	3.2050	0.0640
008	78	ค 2.1/2	9	2	0.7730	1.7820	0.1410
008	124	ค 2.2/1	10	4	1.1730	0.8710	0.1370
008	114	ค 4.1/1	11	3	0.9650	2.6120	0.0930
008	59	ค 4.1/2	12	3	0.7610	2.9970	0.1050
008	60	ค 4.1/3	13	1	0.1860	1.2940	0.0800

ค่าพารามิเตอร์ของข้อสอบในคลังข้อสอบ

Test No.	Item Bank No	Standard / Indicator	Item Test No	Keys	IRT-a	IRT-b	IRT-c
008	61	ค 4.1/4	14	4	0.1990	1.3420	0.1100
008	36	ค 4.1/5	15	4	1.0160	0.6340	0.1860
008	84	ค 4.2/1	16	3	0.6650	4.6880	0.0210
008	21	ค 4.2/1 , ค 4.1/2	17	4	0.2160	1.0870	0.2230
008	86	ค 4.2/3	20	4	0.5470	1.4400	0.2570
008	37	ค 4.2/5	22	2	0.6350	0.7120	0.0830
008	11	ค 4.2/6	23	2	3.1470	1.2300	0.0160
008	66	ค 5.1/1	24	3	0.2600	0.9880	0.2520
008	107	ค 5.1/2	25	2	0.2390	0.9990	0.2340
008	45	ค 5.1/3	26	4	0.1900	1.2770	0.1990
008	95	ค 5.3/1	29	2	0.3690	1.0580	0.1630
009	111	ค 1.1/1	1	4	0.6690	2.0990	0.1910
009	25	ค 1.1/2	2	4	0.9510	2.1070	0.1100
009	119	ค 1.1/3	3	1	0.7440	2.9050	0.1410
009	4	ค 1.2/1	4	3	0.7650	2.2810	0.1980
009	28	ค 1.3/1	5	3	0.9900	2.2800	0.1060
009	29	ค 1.4/1	6	2	0.7230	2.3190	0.1120
009	116	ค 1.4/1 ,ค 4.2/3	7	3	0.9350	2.4530	0.1040
009	77	ค 2.1/1	8	4	0.2600	0.9010	0.0850
009	78	ค 2.1/2	9	2	0.7730	1.7820	0.1410
009	124	ค 2.2/1	10	4	1.1730	0.8710	0.1370
009	114	ค 4.1/1	11	3	0.9650	2.6120	0.0930
009	80	ค 4.1/2	12	2	1.1900	0.7710	0.1120
009	81	ค 4.1/3	13	4	0.4560	0.9100	0.0890

ค่าพารามิเตอร์ของข้อสอบในคลังข้อสอบ

Test No.	Item Bank No	Standard / Indicator	Item Test No	Keys	IRT-a	IRT-b	IRT-c
009	82	ค 4.1/4	14	3	0.1370	1.3670	0.2060
009	62	ค 4.1/5	15	2	0.4830	1.4710	0.2210
009	99	ค 4.2/1	16	3	0.2320	1.3780	0.1170
009	20	ค 4.2/1 , ค 4.1/2	17	4	0.6880	0.8510	0.2170
009	39	ค 4.2/2	18	4	0.7320	1.4440	0.2230
009	13	ค 4.2/4	21	2	1.5780	1.8730	0.0220
009	43	ค 4.2/5	22	2	0.2990	1.3460	0.2000
009	65	ค 4.2/6	23	4	0.6340	0.6660	0.1010
009	90	ค 5.1/1	24	2	0.2670	1.0850	0.1980
009	68	ค 5.1/3	26	4	0.1910	1.2450	0.2180
009	69	ค 5.2/1	27	4	0.1510	1.3610	0.1050
009	47	ค 5.2/2	28	3	0.2960	0.5180	0.1880
009	105	ค 5.3/1	29	2	1.8450	0.5570	0.1250
009	96	ค 5.3/2	30	4	0.5750	1.0120	0.2520

ภาคผนวก ง

การใช้งานโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

การใช้งานโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Unit of Computerized Adaptive Testing: UCAT)

บทนำ:

โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ เพื่อประมาณความสามารถทางคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 จำนวน 32 ตัวชี้วัด ที่มีการให้คะแนนโดยวิธีการวัดความรู้บางส่วน เอกสารฉบับนี้จัดทำขึ้น เพื่อช่วยให้ผู้ใช้เข้าใจถึงขั้นตอนการใช้งาน โปรแกรม ซึ่งโปรแกรมนี้เป็นส่วนหนึ่งของการทำวิทยานิพนธ์ ระดับปริญญาเอก สาขาวิจัย วัดผลและสถิติการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา เรื่อง ผลการให้คะแนนความรู้บางส่วนที่มีต่อประสิทธิภาพการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (The partial knowledge scoring effecting to efficiency of computerized adaptive testing) แบ่งเป็น 4 ขั้นตอน คือ

ขั้นตอนที่ 1 การเข้าโปรแกรม

ขั้นตอนที่ 2 การลงทะเบียนข้อมูลผู้สอบ

ขั้นตอนที่ 3 การทดสอบ

ขั้นตอนที่ 4 การรายงานผล

ผู้วิจัยหวังเป็นอย่างยิ่งว่า เอกสารฉบับนี้จะช่วยให้ผู้ใช้งาน มีความรู้ความเข้าใจเกี่ยวกับขั้นตอนต่าง ๆ และสามารถนำไปใช้ได้สะดวก ก่อให้เกิดประโยชน์ต่อการเรียนการสอนดังนี้

1. ครุนำไปใช้ตรวจสอบความรู้พื้นฐานของนักเรียนในเนื้อหาคณิตศาสตร์ระดับชั้นมัธยมศึกษาตอนปลาย ทั้งก่อนเรียน ระหว่างเรียน และหลังเรียน หรือทดสอบนักเรียนก่อนที่จะเรียนเนื้อหาในระดับชั้นที่สูงขึ้นไป
2. นักเรียนนำไปใช้ตรวจสอบความรู้พื้นฐานของตนเองในเนื้อหาคณิตศาสตร์ระดับชั้นมัธยมศึกษาตอนปลาย เพื่อนำไปปรับปรุง พัฒนาตนเอง หรือเตรียมความพร้อมที่จะเรียนเนื้อหาในระดับชั้นที่สูงขึ้นไป
3. ผู้ปกครองนำไปใช้ตรวจสอบความรู้พื้นฐานของนักเรียนที่อยู่ในความดูแลของตนเอง เพื่อนำไปเป็นแนวทางในการให้คำแนะนำส่งเสริม หรือเตรียมความพร้อมให้กับนักเรียนที่อยู่ในความดูแลต่อไป

ขั้นตอนที่ 1 การเข้าโปรแกรม



1. เข้าระบบอินเทอร์เน็ตโดยใช้ Google Chrome
2. พิมพ์ url ในช่อง address bar : <http://www.brp.ac.th>
3. คลิกที่ Banner : UCAT ศูนย์การสอบออนไลน์

โรงเรียนบุรีรัมย์พิทยาคม
BURIRAM PITTHAYAKHOM SCHOOL

ภาษาไทย Choose language

Home Curriculum Personnel Calendar Data Center Downloads Webboard FAQ E-mail

Announcement :

UCAT
ศูนย์การสอบออนไลน์

เข้าสู่ระบบ

USERNAME 3450100361413
PASSWORD ***** LOGIN

จำนวนไป หนังสือคำสั่ง ฉบับวิธีใช้ ฉบับเนื้อหา

นักเรียนปัจจุบัน ครูและบุคลากร ผู้ปกครองนักเรียน คณิตเก่า เอกสารวิชาการ

สถิติการรับสมัครนักเรียน
ปีการศึกษา 2558

พิธีมอบใบประกาศนียบัตร
ประจำปีการศึกษา 2557
โรงเรียนบุรีรัมย์พิทยาคม
วันที่ 22 ตุลาคม 2558 ณ อาคารอเนกประสงค์
โดย นายศรี ศรีศรีใจ พ.จ.บุรีรัมย์ เป็นประธาน

ปฏิทินวิชาการ

ธ.	จ.	อ.	พ.	พฤ.	ศ.	ส.
31	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	1	2	3	4
5	6	7	8	9	10	11

4. เลือกชื่อโรงเรียนที่นักเรียนกำลังศึกษาอยู่ แล้วพิมพ์ Username : [เลขประจำตัวนักเรียน 5 หลัก]

Password: [เลขประจำตัวประชาชน 13 หลัก]

โรงเรียน :	บุรีรัมย์พิทยาคม	Log In
Username :	32117	
Password :	*****	

ขั้นตอนที่ 2 การลงทะเบียนข้อมูลผู้สอบ เป็นการป้อนข้อมูลผู้เข้าสอบก่อนที่จะเริ่มทดสอบ เกี่ยวกับ

รหัสนักเรียน 5 หลัก เช่น 29009

ชื่อ-สกุล เช่น สมชาย รักเรียน

ระดับชั้น เช่น 6/1

คะแนนเฉลี่ยวิชาคณิตศาสตร์ เช่น 64

โรงเรียนของผู้เข้าสอบ เช่น บุรีรัมย์พิทยาคม

Email เช่น 29001@brp.ac.th



ขั้นตอนที่ 3 การทดสอบ ผู้สอบคลิกปุ่ม

เริ่มทดสอบ

การทดสอบจะเริ่มนับเวลาในการทำข้อสอบรายข้อ ผู้สอบจะทำข้อสอบตั้งแต่ข้อที่ 1 ถึงข้อ 30 โดยพิจารณาข้อสอบ และเลือกตอบทุกตัวเลือก

ตัวเลือกที่ 1	ตัวเลือกที่ 2	ตัวเลือกที่ 3	ตัวเลือกที่ 4
<input type="radio"/> ถูก	<input type="radio"/> ถูก	<input type="radio"/> ถูก	<input type="radio"/> ถูก
<input type="radio"/> ผิด	<input type="radio"/> ผิด	<input type="radio"/> ผิด	<input type="radio"/> ผิด
<input type="radio"/> ไม่แน่ใจ	<input type="radio"/> ไม่แน่ใจ	<input type="radio"/> ไม่แน่ใจ	<input type="radio"/> ไม่แน่ใจ

เมื่อมั่นใจในการตอบเรียบร้อยแล้วให้คลิกปุ่ม

ยืนยันคำตอบ

00:00:43

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

ข้อที่ 1/30

พิจารณาข้อความต่อไปนี้

ก) มีจำนวนเต็มทีมากที่สุดที่น้อยกว่า 9

ข) มีจำนวนตรรกยะทีมากที่สุดที่น้อยกว่า 9

ค) มีจำนวนอตรรกยะทีน้อยทีสุดทีมากกว่า 1

ง) มีจำนวนตรรกยะทีน้อยทีสุดทีมากกว่า 1

ข้อสรุปใดต่อไปนีถูกต้อง

ก. มีข้อถูก 1 ข้อ

ข. มีข้อถูก 2 ข้อ

ค. มีข้อถูก 3 ข้อ

ง. มีข้อถูก 4 ข้อ

ตัวเลือกที่ 1

ถูก

ผิด

ไม่แน่ใจ

ตัวเลือกที่ 2

ถูก

ผิด

ไม่แน่ใจ

ตัวเลือกที่ 3

ถูก

ผิด

ไม่แน่ใจ

ตัวเลือกที่ 4

ถูก

ผิด

ไม่แน่ใจ

ข้อปัจจุบัน
 ข้อที่หาคำแล้ว
ข้อที่ยังไม่ได้เปิด

ขั้นตอนที่ 4 การรายงานผล

- ผลการวิเคราะห์ความสามารถของผู้สอบตามมาตรฐานและตัวชี้วัด 32 ตัวชี้วัด

ค่าความสามารถ	ความหมาย	ผลวิเคราะห์
$\hat{\theta} > 0.50$	ดีมาก	ระดับสูง
$0.25 < \hat{\theta} \leq 0.50$	ดี	ระดับปานกลาง
$-0.25 < \hat{\theta} \leq 0.25$	ค่อนข้างดี	ระดับปานกลาง
$-0.50 < \hat{\theta} \leq -0.25$	ปกติ	ระดับปานกลาง
$\hat{\theta} < -0.50$	น้อยกว่าปกติ	ระดับต่ำ

- สรุปผลการประเมินความสามารถของผู้สอบ

ค่าเฉลี่ยความสามารถ	ความหมาย	ผลประเมิน
$\hat{\theta} > 0.50$	ระดับสูง	ผ่าน
$-0.50 \leq \hat{\theta} \leq 0.50$	ระดับปานกลาง	ผ่าน
$\hat{\theta} < -0.50$	ระดับต่ำ	ไม่ผ่าน

รายงานผลการสอบแบบปรับเหมาะด้วยคอมพิวเตอร์
วิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษาตอนปลาย

ชื่อ-สกุลผู้สอบ : ชั้น :

คะแนนเฉลี่ยคณิตศาสตร์พื้นฐาน : โรงเรียน :

Email : Report No :

ผลการวิเคราะห์ความสามารถ

มาตรฐาน/ตัวชี้วัด	ค่าความสามารถ	ความหมาย	ผลการวิเคราะห์
ค 1.1/1.....
ค 1.1/2.....
ค 1.1/3.....
.....
.....
.....
.....
.....
ค 5.3/1.....
ค 5.3/2.....

สรุปผลการประเมินความสามารถ

เวลาในการสอบ..... นาที ค่าฟังก์ชันสารสนเทศของแบบสอบ

จำนวนข้อสอบ..... ข้อ ความคลาดเคลื่อนมาตรฐานในการวัด.....

ความเชื่อมั่นในการประมาณค่า.....

ความสามารถโดยรวม..... ระดับ..... ผลการประเมิน.....

(ลงชื่อ).....นักเรียน (ลงชื่อ).....ผู้ปกครองนักเรียน

(ลงชื่อ).....ครูที่ปรึกษา (ลงชื่อ).....ครูผู้สอน

...../...../.....

...../...../.....