

การศึกษาความเที่ยงตรงและประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของผู้สอบ
โดยใช้แบบทดสอบเทเลอรัปรีรามิดในวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6

ปริญญานิพนธ์
ของ
อศวิน คงเพชรศักดิ์

เสนอต่อมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต สาขาวิชาการวัดผลการศึกษา
พฤษภาคม 2547
ลิขสิทธิ์เป็นของมหาวิทยาลัยศรีนครินทรวิโรฒ

๓๗๔. ๗๐๔๔

๐๕๗๙ ๗

๕.๓

การศึกษาความเที่ยงตรงและประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของผู้สอบ
โดยใช้แบบทดสอบเทเลอรรูปปิรามิดในวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6

บทคัดย่อ

ของ

อศวิน คงเพชรศักดิ์

๑๔ ก.ย. ๒๕๔๗

เสนอต่อมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต สาขาวิชาการศึกษา

พฤษภาคม ๒๕๔๗

๑๒๕๑๐๓๔

อัศวิน คงเพชรศักดิ์. (2547). การศึกษาความเที่ยงตรงและประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของผู้สอบโดยใช้แบบทดสอบเทเลอรรูปปริมาตรในวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6. ปรินญานิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ.
คณะกรรมการควบคุม : อาจารย์ ดร.สุพร เข้มแข็ง, รองศาสตราจารย์นิภา ศรีไพโรจน์

การวิจัยครั้งนี้เพื่อศึกษาความเที่ยงตรงและประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของผู้สอบโดยใช้แบบทดสอบเทเลอรรูปปริมาตรในวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 โดยศึกษากับกลุ่มตัวอย่างชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2546 ของโรงเรียนสังกัดสำนักงานเขตพื้นที่การศึกษาราชบุรี เขต 1 และสังกัดเทศบาล ในเขตอำเภอเมืองราชบุรี จังหวัดราชบุรี จำนวน 364 คน ซึ่งเลือกมาโดยการสุ่มแบบกลุ่ม (Cluster Random Sampling) เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบเทเลอรรูปปริมาตรในวิชาคณิตศาสตร์ 3 รูปแบบ คือ

- 1) แบบทดสอบเทเลอรรูปปริมาตรที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น
- 2) แบบทดสอบเทเลอรรูปปริมาตรที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้น
- 3) แบบทดสอบเทเลอรรูปปริมาตรที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น

โดยมีวิธีการตรวจให้คะแนนเพื่อวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ 3 วิธี คือ

- 1) การตรวจให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบข้อที่ทำ
 - 2) การตรวจให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบเฉพาะข้อที่ตอบถูก
 - 3) การตรวจให้คะแนนตามค่าความยากตามข้อสอบที่ถัดจากข้อสุดท้ายที่ได้ตอบ
- ผลการวิจัยพบว่า

1. ค่าความเที่ยงตรงของการประมาณค่าความสามารถ ของแบบทดสอบเทเลอรรูปปริมาตร ทั้ง 3 รูปแบบ ที่ใช้วิธีการตรวจให้คะแนนทั้ง 3 วิธี มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ยกเว้นค่าความเที่ยงของการประมาณค่าความสามารถระหว่างรูปแบบที่ใช้ขนาดชั้นคงที่จำนวน 15 ชั้น กับรูปแบบที่มีหลายข้อในแต่ละชั้นจำนวน 5 ชั้น ที่ตรวจให้คะแนนโดยวิธีพิจารณาค่าความยากตามข้อสอบที่ถัดจากข้อสุดท้ายที่ได้ตอบ มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

2. ค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของแบบทดสอบเทเลอรรูปปริมาตรทั้ง 3 รูปแบบ พบว่า รูปแบบที่ใช้ขนาดชั้นคงที่จำนวน 15 ชั้น มีค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถสูงสุดในกลุ่มผู้สอบทุกระดับความสามารถ และพบว่าค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของรูปแบบที่ใช้ขนาดชั้นคงที่จำนวน 5 ชั้น สูงกว่ารูปแบบที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น ในกลุ่มผู้สอบที่มีความสามารถระดับปานกลาง และระดับต่ำ แต่ในกลุ่มผู้สอบที่มีความสามารถ

ระดับต่ำมาก ระดับสูง และระดับสูงมาก พบว่าค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่า
ความสามารถของรูปแบบที่ใช้ขนาดชั้นคงที่จำนวน 5 ชั้น มีค่าต่ำกว่ารูปแบบที่มีข้อสอบหลาย
ข้อในแต่ละชั้นจำนวน 5 ชั้น

**A STUDY OF VALIDITY AND RELATIVE EFFICIENCY OF ABILITY ESTIMATION BY
GRADE VI MATHEMATICS PYRAMIDAL TAILORED TEST**

**AN ABSTRACT
BY
ASAWIN KHONGPHETSAK**

**Presented in partial fulfillment of the requirements
for the Master of Education degree in Educational Measurement
at Srinakhainwirot University**

May 2004

Asawin Khongphetsak. (2004), *A Study of Validity and Relative Efficiency of Ability Estimation by Grade VI Mathematics Pyramidal Tailored Test.*

Master Thesis, M.Ed. (Educational Measurement). Bangkok :

Graduate School, Srinakharinwirot University.

Advisor Committee : Dr. Suwaporn Semheng, Assoc. Prof. Nipa Sripairot.

The purpose of this research were to study validity and relative efficiency of ability estimation by grade VI mathematics pyramidal tailored test. The sample were 364 grade 6 students of schools under The Office of Ratchaburi Educational Area Service 1 and Amphur muang Ratchaburi, municipality, selected by cluster random sampling technique. The instruments were three formats of Mathematics Pyramidal Tailored tests, i.e.,

- 1) 5 constant steps size pyramid,
- 2) 15 constant steps size pyramid, and
- 3) 5 steps multiple-item pyramid 5 steps.

Three methods used for scoring the tests were

- 1) scoring according to the average difficulty,
- 2) scoring according to the average difficulty of all item answers correctly of average difficulty correct, and
- 3) scoring according to the difficulty of the $N + 1^{\text{th}}$ item.

The results of the research were as follows.

1. The validity of ability estimation of all three formats of Pyramidal Tailored test using 3 methods of scoring were statistically significant different at .01 level, except the 15 constant steps size pyramid and 5 steps multiple-item pyramid using the Difficulty of the $N + 1^{\text{th}}$ item scoring method were found not statistically significant different.

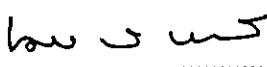
2. The 15 constant steps size pyramid had highest relative efficiency of ability estimation in every ability levels. It was also found that the relative efficiency of ability estimation of the 5 constant steps size pyramid was higher than the 5 steps multiple-item pyramid at the low and moderate ability levels students but lower than the 5 steps multiple-item pyramid at the very low, high and very high ability levels students.

ปริญญานิพนธ์
เรื่อง

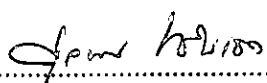
การศึกษาความเที่ยงตรงและประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของผู้สอบ
โดยใช้แบบทดสอบเทเลอ์รูปปริามิตในวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6

ของ
นายอัศวิน คงเพชรศักดิ์

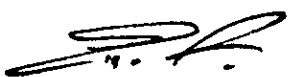
ได้รับอนุมัติจากบัณฑิตวิทยาลัยให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาการศึกษามหาบัณฑิต สาขาวิชาการวัดผลการศึกษา
ของมหาวิทยาลัยศรีนครินทรวิโรฒ

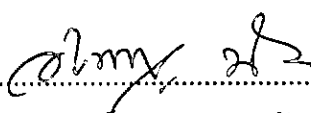
.....คณบดีบัณฑิตวิทยาลัย
(รองศาสตราจารย์ ดร.นภาพร หะวานนท์)
วันที่ 25 เดือน พฤษภาคม พ.ศ. 2547

คณะกรรมการสอบปริญญานิพนธ์

.....ประธาน
(อาจารย์ ดร. สุพร เชื้อเฮง)

.....กรรมการ
(รองศาสตราจารย์นิภา ศรีไพโรจน์)

.....กรรมการที่แต่งตั้งเพิ่มเติม
(ผู้ช่วยศาสตราจารย์ ดร.นิคม ดังกะพิภพ)

.....กรรมการที่แต่งตั้งเพิ่มเติม
(อาจารย์ ดร.องอาจ นัยพัฒน์)

ประกาศคุณูปการ

ในการทำปริญญานิพนธ์ฉบับนี้ ผู้วิจัยได้รับความกรุณาจากอาจารย์ ดร.สุวพร เข้มเฮง และรองศาสตราจารย์นิภา ศรีไพโรจน์ กรรมการที่ปรึกษาปริญญานิพนธ์ ที่กรุณาให้คำปรึกษา แนะนำ ตลอดจนข้อคิดเห็นต่าง ๆ ในการทำปริญญานิพนธ์มาโดยตลอดตั้งแต่ต้นจนสำเร็จลุล่วงไปด้วยดี ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูง

ขอกราบขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.นิคม ดังคะพิภพ และ อาจารย์ ดร.องอาจ นัยพัฒน์ กรรมการสอบปริญญานิพนธ์ ที่กรุณาให้ความรู้และคำแนะนำเพิ่มเติม อันเป็นประโยชน์อย่างยิ่งสำหรับการทำปริญญานิพนธ์ในครั้งนี้

ขอขอบพระคุณ อาจารย์ ดร.สุนิสา จัยม่วงศรี และ อาจารย์ลัษมา สุวรรณลอยล่อง ที่กรุณาให้ความรู้และคำแนะนำเพิ่มเติม ในการวิเคราะห์ข้อมูลทางสถิติ

ขอขอบพระคุณ ผู้อำนวยการ ผู้ช่วยผู้อำนวยการฝ่ายวิชาการ และคณะครูของ โรงเรียนในกลุ่มตัวอย่าง และขอขอบใจนักเรียนทุกคนที่ให้ความร่วมมือในการเก็บข้อมูลเป็นอย่างดี

สุดท้ายนี้ผู้วิจัยขอโน้มรำลึกถึงพระคุณของบิดา มารดา และครู อาจารย์ ที่ได้อบรมสั่งสอน ตลอดจนให้การสนับสนุนการศึกษาแก่ผู้วิจัยตลอดมา

อัศวิน คงเพชรศักดิ์

สารบัญ

บทที่	หน้า
1 บทนำ	1
ภูมิหลัง	1
ความมุ่งหมายของการศึกษาค้นคว้า	4
ความสำคัญของการศึกษาค้นคว้า	5
ขอบเขตของการศึกษาค้นคว้า	5
นิยามศัพท์เฉพาะ	6
สมมติฐานของการศึกษาค้นคว้า	8
2 เอกสารและงานวิจัยที่เกี่ยวข้อง	9
ทฤษฎีการตอบสนองข้อสอบ	9
การทดสอบแบบเทเลอร์	17
คุณภาพของแบบทดสอบ	35
งานวิจัยที่เกี่ยวข้อง	43
3 วิธีดำเนินการศึกษาค้นคว้า	51
ประชากรและกลุ่มตัวอย่าง	51
เครื่องมือที่ใช้ในการศึกษาค้นคว้า	52
วิธีดำเนินการสร้างแบบทดสอบ	54
วิธีดำเนินการเก็บรวบรวมข้อมูล	65
สถิติที่ใช้ในการวิเคราะห์ข้อมูล	67

สารบัญ (ต่อ)

บทที่	หน้า
4 ผลการวิเคราะห์ข้อมูล.....	71
สัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูล.....	71
การเสนอผลการวิเคราะห์ข้อมูล	72
ผลการวิเคราะห์ข้อมูล	72
5 สรุป อภิปรายผล และข้อเสนอแนะ	86
ความมุ่งหมายของการศึกษาค้นคว้า	86
กลุ่มตัวอย่าง	86
เครื่องมือที่ใช้ในการศึกษาค้นคว้า	86
วิธีดำเนินการเก็บรวบรวมข้อมูล	87
การเสนอผลการวิเคราะห์ข้อมูล	88
สรุปผลการศึกษาค้นคว้า.....	88
อภิปรายผล	91
ข้อเสนอแนะ	94
บรรณานุกรม	95
ภาคผนวก	100
ประวัติย่อของผู้วิจัย	116

บัญชีตาราง

ตาราง		หน้า
1	แสดงจำนวนนักเรียนในกลุ่มตัวอย่างที่ใช้ในการสร้างเครื่องมือ และในการวิจัยจำแนกตามสังกัด	52
2	ค่าน้ำหนักความสำคัญของจุดประสงค์เชิงพฤติกรรม วิชาคณิตศาสตร์ เรื่อง การบวก ลบ และคูณทศนิยม	55
3	ค่าความยาก (b) ของแบบทดสอบรูปปริมาตรที่มีขนาดชั้นคงที่ 5 ชั้น	59
4	ค่าความยาก (b) ของแบบทดสอบรูปปริมาตรที่มีขนาดชั้นคงที่ 15 ชั้น	61
5	ค่าเฉลี่ยความยาก (b) ของแบบทดสอบรูปปริมาตรที่มีข้อสอบหลายข้อในแต่ละชั้น	63
6	แสดงค่าความยากของข้อสอบ (b) และผลการตอบของผู้สอบ ในภาพประกอบ 12	64
7	ค่าสถิติพื้นฐานของแบบทดสอบรูปปริมาตรทั้ง 3 ฉบับ โดยวิธีการตรวจ ให้คะแนน	73
8	ค่าสถิติพื้นฐานของความสามารถของผู้สอบ (θ)	74
9	เปรียบเทียบค่าความเที่ยงตรงระหว่างแบบทดสอบปริมาตร 3 รูปแบบกับ ความสามารถของผู้สอบ โดยวิธีการตรวจให้คะแนนตามค่าความยากเฉลี่ย ของข้อสอบที่ทำ	75
10	เปรียบเทียบค่าความเที่ยงตรงระหว่างแบบทดสอบปริมาตร 3 รูปแบบกับ ความสามารถของผู้สอบ โดยวิธีการตรวจให้คะแนนตามค่าความยากเฉลี่ย ของข้อสอบเฉพาะข้อที่ทำถูก.....	75
11	เปรียบเทียบค่าความเที่ยงตรงระหว่างแบบทดสอบปริมาตร 3 รูปแบบกับ ความสามารถของผู้สอบ โดยวิธีการให้คะแนนค่าความยากตามข้อสอบ ที่ตัดจากข้อสุดท้ายที่ได้ตอบ	76
12	ผลของการทดสอบความแตกต่างของค่าความเที่ยงตรงของการประมาณค่า ความสามารถของผู้สอบ จากการทำแบบทดสอบรูปปริมาตร 3 รูปแบบ โดยวิธีการตรวจให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ทำ โดยทดสอบเป็นรายคู่.....	77

บัญชีตาราง (ต่อ)

ตาราง

หน้า

13	ผลของการทดสอบความแตกต่างของค่าความเที่ยงตรงของการประมาณ ค่าความสามารถของผู้สอบ จากการทำแบบทดสอบรูปปริมาตร 3 รูปแบบ โดยวิธีการตรวจให้คะแนนตามค่าความยากเฉลี่ยของ ข้อสอบเฉพาะข้อที่ทำถูก โดยทดสอบเป็นรายคู่	77
14	ผลของการทดสอบความแตกต่างของค่าความเที่ยงตรงของการประมาณ ค่าความสามารถของผู้สอบ จากการทำแบบทดสอบรูปปริมาตร 3 รูปแบบ โดยวิธีการตรวจให้คะแนนตามค่าความยากตามข้อสอบที่ ถัดจากข้อสุดท้ายที่ได้ตอบ โดยทดสอบเป็นรายคู่	78
15	ค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของผู้สอบ ระหว่าง แบบทดสอบรูปปริมาตร 3 รูปแบบ ที่ระดับความสามารถตั้งแต่ -4.00 ถึง 4.00 จำนวน 81 ระดับความสามารถ	79
16	ค่าไอเกน (มากกว่า 1) เปอร์เซนต์ของความแปรปรวน และเปอร์เซนต์ ของความแปรปรวนสะสม จากการนำแบบทดสอบ จำนวน 6 ฉบับ มา วิเคราะห์องค์ประกอบ.....	102
17	ค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบรูปปริมาตรแบบที่มีขนาดชั้นคงที่จำนวน 5 ชั้น (x_1) แบบทดสอบรูปปริมาตรแบบที่มีขนาดชั้นคงที่จำนวน 15 ชั้น (x_2) แบบทดสอบรูปปริมาตรแบบที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น (x_3).....	104
18	ค่าประสิทธิภาพสัมพัทธ์ (RE) ของการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบรูปปริมาตรแบบที่มีขนาดชั้นคงที่จำนวน 5 ชั้น (x_1) แบบทดสอบรูปปริมาตรแบบที่มีขนาดชั้นคงที่จำนวน 15 ชั้น (x_2) แบบทดสอบรูปปริมาตรแบบที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น (x_3).....	109

บัญชีภาพประกอบ

ภาพประกอบ

หน้า

1	โครงสร้างของการทดสอบเทเลอร์รูปปิรามิดแบบขนาดขั้นคงที่ ที่มี 5 ขั้น	23
2	โครงสร้างของการทดสอบเทเลอร์รูปปิรามิดแบบขนาดขั้นแปรผันที่มี 5 ขั้น	24
3	โครงสร้างของการทดสอบเทเลอร์รูปปิรามิดแบบรอบบิน-มอนโร (Robin-Monro)	25
4	โครงสร้างของการทดสอบเทเลอร์รูปปิรามิดแบบข้างตัด (Truncated Pyramid) สกิดการสะท้อนกลับ (Reflecting Barrier)	25
5	โครงสร้างของการทดสอบเทเลอร์รูปปิรามิดแบบข้างตัด (Truncated Pyramid) รักษาการสะท้อนกลับ (Retaining Barrier)	26
6	โครงสร้างของการทดสอบเทเลอร์รูปปิรามิด แบบมีข้อสอบหลายข้อในแต่ละ ขั้น (Multiple-Item Model)	27
7	โครงสร้างของการทดสอบเทเลอร์รูปปิรามิดแบบให้น้ำหนักแต่ตัวเลือกของ ข้อสอบเพื่อแยกทาง	28
8	การให้คะแนนแบบทดสอบปิรามิด 2 วิธี ที่ใช้ผลการตอบข้อสอบ	31
9	อินฟอร์เมชันของข้อสอบ 5 ข้อ และโค้งอินฟอร์เมชันของแบบทดสอบ	37
10	โค้งประสิทธิภาพสัมพัทธ์ของแบบทดสอบคำศัพท์ STEP เทียบกับ แบบทดสอบคำศัพท์ MAT	41
11	แผนผังแสดงขั้นตอนการดำเนินการสร้างแบบทดสอบ	54
12	โครงสร้างของแบบทดสอบรูปปิรามิดที่มีขนาดขั้นคงที่ 5 ขั้น	59
13	โครงสร้างของแบบทดสอบรูปปิรามิดที่มีขนาดขั้นคงที่ 15 ขั้น	60
14	โครงสร้างของแบบทดสอบรูปปิรามิด แบบมีข้อสอบหลายข้อในแต่ละขั้น (Multiple-Item Model)	63
15	โค้งประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของผู้สอบระหว่าง แบบทดสอบแบบทดสอบรูปปิรามิด แบบที่ใช้ขนาดขั้นคงที่ จำนวน 5 ขั้น เทียบกับแบบทดสอบรูปปิรามิด แบบที่ใช้ขนาดขั้นคงที่ จำนวน 15 ขั้น.....	82
16	โค้งประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของผู้สอบระหว่าง แบบทดสอบแบบทดสอบรูปปิรามิด แบบที่ใช้ขนาดขั้นคงที่ จำนวน 5 ขั้น เทียบกับแบบทดสอบรูปปิรามิด แบบที่มีข้อสอบหลายข้อในแต่ละขั้น จำนวน 5 ขั้น.....	83

บัญชีภาพประกอบ (ต่อ)

ภาพประกอบ	หน้า
17 โคงประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของผู้สอบระหว่าง แบบทดสอบแบบทดสอบรูปปิรามิด แบบที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น เทียบกับแบบทดสอบรูปปิรามิด แบบที่มีข้อสอบหลายข้อในแต่ละชั้น จำนวน 5 ชั้น.....	84

บทที่ 1

บทนำ

ภูมิหลัง

การวัดและประเมินผลด้านผลสัมฤทธิ์ทางการเรียนของนักเรียนในปัจจุบัน เราใช้ข้อสอบเป็นเครื่องมือวัด แล้วนำผลการวัดนั้นไปตรวจสอบกับวัตถุประสงค์ของการเรียนการสอนที่กำหนดไว้ว่านักเรียนบรรลุผลในการเรียนรู้มากน้อยเพียงใด ขั้นตอนต่าง ๆ ในการวัดผลโดยใช้ข้อสอบเป็นเครื่องมือวัดจึงมีความสำคัญเป็นอันมาก เพราะผลจากการวัดจะมีความน่าเชื่อถือมากน้อยเพียงใดขึ้นอยู่กับคุณภาพของเครื่องมือที่ใช้ในการวัดผลนั่นเอง ถ้าข้อสอบมีคุณภาพดี ผลการวัดก็จะมีมีความน่าเชื่อถือมาก แต่ถ้าข้อสอบมีความคลาดเคลื่อนผลการวัดก็จะคลาดเคลื่อนตามไปด้วย การวัดก็จะไม่ตรงตามจุดประสงค์ ธอร์นไดค์ (Thorndike, 1961 : 27) ได้กล่าวถึงความสำคัญของการทดสอบที่มีต่อการเรียนรู้ไว้ว่า ข้อสอบที่สร้างอย่างดีและนำมาใช้อย่างได้ผลสามารถเป็นแรงจูงใจในการที่จะสร้างนิสัยในการเรียนรู้ที่ดี แก่ข้อผิดพลาดได้ถูกต้องและเป็นแนวทางที่จะนำไปสู่ผลสัมฤทธิ์ในการเรียนรู้ได้ ในการทดสอบทุกครั้ง ผู้ทดสอบไม่ต้องการเพียงเพื่อให้รู้ผลคะแนนที่ผู้สอบทำได้เท่านั้น แต่ลึกไปกว่านั้น ยังต้องการรู้ด้วยว่าความสามารถจริง ๆ ของผู้สอบมีอยู่เท่าไร (สุธรรม จันทร์หอม, 2526 : 45) ดังนั้นแบบทดสอบที่ใช้วัดจึงจำเป็นต้องมีคุณภาพที่เชื่อถือได้ว่า เมื่อนำไปทำการวัดผลแล้ว คะแนนที่ได้จากการวัดจะเป็นคะแนนที่แท้จริง คือเป็นคะแนนที่มีความคลาดเคลื่อนน้อยที่สุด และสามารถวัดได้ตรงตามจุดประสงค์ที่ต้องการวัด

การวัดผลการศึกษาและจิตวิทยาเป็นการวัดความสามารถหรือคุณลักษณะแฝง (Ability or Latent Trait) ของบุคคล ซึ่งไม่สามารถวัดได้โดยตรง ต้องอาศัยการวัดโดยทางอ้อม โดยใช้แบบทดสอบเป็นเครื่องมือในการวัด แล้วนำผลการสอบหรือคะแนน (Performance or Score) ไปอธิบายหรือทำนายความสามารถที่แท้จริงของบุคคลที่ถูกวัด (Lord, 1980 : 3) ในการวัดหรือการประมาณค่าความสามารถที่แท้จริงโดยอาศัยการวัดทางอ้อมนี้มักเกิดความคลาดเคลื่อนค่อนข้างสูง ซึ่งจะทำให้ผลของการวัดขาดความถูกต้อง ในการวัดผลได้มีการสร้างทฤษฎีในการวัดขึ้นมาเพื่อที่จะได้สรุปอ้างอิงผลของการวัดซึ่งเป็นตัวแทนของความสามารถหรือคุณลักษณะแฝงของบุคคล ให้มีความถูกต้องแม่นยำและเชื่อถือได้มากที่สุด ทฤษฎีการวัดที่สำคัญ ได้แก่ ทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory : CTT) และ ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory : IRT)

ทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory : CTT) เป็นทฤษฎีดั้งเดิมที่เน้นการประมาณค่าคะแนนจริง หรือความสามารถแท้จริง โดยนำคะแนนของกลุ่มผู้สอบมาใช้ในการประมาณค่าซึ่งถูกมองว่าเป็นทฤษฎีที่ไม่เพียงพอต่อการประมาณค่าความสามารถที่แท้จริง และไม่สามารถแก้ปัญหาในการเรียนการสอนได้ ดังนั้นนักวัดผลจึงได้พยายามเสนอแนวความคิดใหม่ หรือทฤษฎีใหม่ ทฤษฎีที่พัฒนาในต่อมาและเป็นที่น่าสนใจได้แก่ ทฤษฎีประมาณคุณลักษณะภายใน (Latent Trait Theory) หรือทฤษฎีการตอบข้อคำถาม (Item Response Theory) หรือทฤษฎีโค้งลักษณะข้อสอบ (Item Characteristic Curve Theory) ซึ่งเป็นทฤษฎีที่ใช้โมเดลทางคณิตศาสตร์แสดงความสัมพันธ์ระหว่างความสามารถที่แท้จริง (θ) กับพฤติกรรมการตอบข้อสอบของแต่ละบุคคล (Performance) (ผจงจิต อินทสุวรรณ. 2528 : 54) ทฤษฎีการตอบสนองข้อสอบให้ประโยชน์อย่างมากในการสร้างแบบทดสอบ โดยเฉพาะในการตอบคำถามซึ่งไม่สามารถกระทำได้ด้วยทฤษฎีการทดสอบแบบดั้งเดิม ซึ่งปัญหาหนึ่งที่เกิดจากการวัดโดยใช้แบบทดสอบดั้งเดิมคือ ปัญหาการจัดข้อสอบให้สอดคล้องกับความสามารถของบุคคล เนื่องจากแบบทดสอบดั้งเดิมจะใช้อย่างเหมาะสมในกรณีที่ผู้สอบมีความสามารถอยู่ในระดับกลาง หรืออยู่ในช่วงค่าเฉลี่ยของกลุ่มผู้สอบ ดังนั้นเมื่อผู้ที่มีความสามารถที่สูง หรือต่ำกว่าค่าเฉลี่ยของกลุ่มผู้สอบมาทำแบบทดสอบ จะทำให้การประมาณค่าความสามารถของผู้สอบสูงหรือต่ำกว่าที่เป็นจริง ในกรณีที่ผู้สอบได้ตอบข้อสอบที่มีความยากสูงกว่าระดับความสามารถของตน ผู้สอบก็จะมีแนวโน้มในการตอบข้อสอบด้วยการเดา และถ้าผู้สอบได้ตอบข้อสอบที่มีความยากต่ำกว่าระดับความสามารถของตน ข้อสอบนั้นก็จะไม่ช่วยให้ผู้สอบได้ใช้ความสามารถในระดับสูงที่ขึ้นไป ซึ่งส่งผลให้ระดับคะแนนของผู้สอบต่ำกว่าที่ควรจะเป็น (Weiss. 1974 : 1) ปัญหานี้แก้ไขโดยใช้การทดสอบแบบเทเลอร์ (Tailored Testing) ซึ่งเป็นตัวอย่างหนึ่งของการประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบอย่างได้ผล (จิราพร ไกรสรศิวเวท. : 2529 : 2 ; Citing Rudner. 1983 : 951)

การทดสอบแบบเทเลอร์ (Tailored Testing) มีหลักการว่า ถ้าผู้สอบตอบข้อคำถามที่ยาก ได้แล้ว ข้อคำถามที่ง่ายกว่าก็น่าจะตอบถูกต้องด้วย หรือถ้าตอบข้อที่ง่าย ๆ ผิดข้อที่ยากก็น่าจะตอบผิดด้วย จึงไม่จำเป็นต้องให้ตอบข้อเหล่านั้นดังที่กรีน และคนอื่น ๆ (จิราพร ไกรสรศิวเวท. 2529 : 2 - 3 ; Citing Green and other. 1984 : 356) ได้เปรียบเทียบว่า ถ้านักกระโดดสูงทำการกระโดดสูง 72 นิ้วไม่ข้ามแล้ว ไม่จำเป็นต้องให้ลองกระโดดสูง 78 นิ้ว ก็ได้ ลักษณะที่สำคัญของการทดสอบแบบเทเลอร์ (Tailored Testing) คือ การจัดข้อคำถามให้มีความยากที่เหมาะสมกับระดับความสามารถของผู้สอบแต่ละคน โดยรูปแบบของการทดสอบเทเลอร์นั้นพอสรุปได้ 2 รูปแบบคือ ยุทธวิธีสองขั้นตอน (Two – Stage Strategies) และยุทธวิธีหลายขั้นตอน (Multi- Stage Strategies) ยุทธวิธีหลายขั้นตอนประกอบด้วย แบบจำลอง 2 รูปแบบ คือ แบบจำลองการแยกทางคงที่ (Fixed-Branching Models) และแบบจำลองการแยกทางแปรผัน (Variable Branching Models) สำหรับการจำลองการแยกทางคงที่จำแนกเป็น 3 ชนิดคือ แบบจำลองรูปปิรามิด (Pyramidal Model) แบบทดสอบเฟล็กซ์ชีเลเวล (Flexilevel Test) และแบบ

ทดสอบปรับระดับขึ้น (Stradaptive Test) ส่วนแบบจำลองการแยกทางแปรผันแบ่งออกเป็น 2 ยุทธวิธีคือ ยุทธวิธีของเบส์ (ayesion Strategies) และยุทธวิธีความเป็นไปได้สูงสุด (Maximum Likelihood Strategies)

แบบจำลองรูปปริมาตรจำแนกออกเป็น 5 แบบคือ (จิราพร ไกรสรศิวเวท. 2529 : 4 ; Citing Weiss. 1974 : 12-36) แบบจำลองที่ใช้ขนาดขั้นคงที่, แบบจำลองที่ใช้ขนาดขั้นแปรผัน, แบบจำลองรูปปริมาตรข้างตัด, แบบจำลองรูปปริมาตรที่ใช้ข้อสอบหลายข้อในแต่ละขั้น และแบบจำลองที่ใช้ในการให้ค่าน้ำหนักแก่ตัวเลือกของข้อสอบเพื่อแยกทาง ในการให้คะแนนของแบบทดสอบเทเลอรูปปริมาตรสามารถให้คะแนนได้หลายวิธี แต่ละวิธีขึ้นอยู่กับแนวทางในการแยกทางเพื่อทำข้อสอบในข้อต่อไป และขึ้นอยู่กับความยากของข้อสอบที่ผู้สอบได้ตอบ ข้อตกลงเบื้องต้นของการสร้างแบบทดสอบเทเลอรูปปริมาตรคือจะต้องเป็นมิติเดียว และมาตราวัดระดับความยากกับมาตราวัดระดับความสามารถของผู้สอบจะต้องแสดงอยู่ในมาตราเดียวกัน วิธีการให้คะแนนแบบทดสอบเทเลอรูปปริมาตรสามารถสรุปได้ 7 วิธีคือ ให้คะแนนตามค่าความยากของข้อสอบข้อสุดท้ายที่ทำ, ให้คะแนนตามค่าความยากของข้อสอบต่อจากข้อสุดท้ายที่ทำ, ให้คะแนนตามผลการตอบข้อสอบข้อสุดท้าย, ให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ทำ, ให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ตอบถูก, ให้คะแนนข้อสอบทุกข้อ และให้คะแนนตามจำนวนข้อที่ตอบถูก

จากการศึกษาของลาร์กิน และไวส์ (จิราพร ไกรสรศิวเวท. 2529 : 4 ; Citing Larkin and Weiss. 1974 : 42) พบว่า ความเชื่อมั่นคงที่ภายในของแบบทดสอบเทเลอรูปปริมาตรที่ให้คะแนนวิธีที่ 6 จะมีค่าสูงสุดถึง .99 และการให้คะแนนวิธีที่ 6 ยังมีความสัมพันธ์กับแบบทดสอบดั้งเดิมสูงกว่าวิธีการให้คะแนนโดยวิธีอื่น ๆ โดยที่การให้คะแนนวิธีที่ 2 และ 7 จะมีความสัมพันธ์กับแบบทดสอบดั้งเดิมต่ำสุด เมื่อพิจารณาความสัมพันธ์ของคะแนนที่ได้จากการให้คะแนนวิธีต่าง ๆ พบว่า การให้คะแนนโดยวิธีที่ 4 มีความสัมพันธ์กันสูงมากกับวิธีที่ 6 คะแนนที่ได้ในวิธีที่ 2 จะมีความสัมพันธ์สมบูรณ์ (1.00) กับคะแนนที่ให้โดยวิธีที่ 7 และคะแนนที่ได้โดยวิธีที่ 6 จะมีความสัมพันธ์สูงกับคะแนนที่ให้โดยวิธีอื่น ๆ และสูงมากกว่าความสัมพันธ์ของคะแนนที่ให้โดยวิธีอื่น ๆ

จะเห็นได้ว่า แบบทดสอบเทเลอรูปปริมาตรนั้นมีประโยชน์อย่างมากในการประมาณค่าความสามารถของผู้สอบ ได้มีผู้ทำวิจัยเกี่ยวกับเรื่องแบบทดสอบเทเลอรูปปริมาตรมาบ้างแล้ว แต่ก็ยังมีปัญหาเกี่ยวกับวิธีดำเนินการสอบ และวิธีการประมาณค่าความสามารถของผู้สอบ (ขรรค์ชัย คงเสน่ห์ : 2530; จิราพร ไกรสรศิวเวท : 2529; ด้าย เชิญฉิม : 2534 ; ประเวศ เวชชะ : 2530) ซึ่งดำเนินการสอบโดยใช้แบบทดสอบที่เป็นกระดาษ มีขั้นตอนที่ยุ่งยากในการตอบข้อสอบข้อที่เหมาะสมกับระดับความสามารถของผู้สอบ

ปัจจุบันคอมพิวเตอร์ได้เข้ามามีบทบาทในชีวิตประจำวันของมนุษย์อย่างมาก เนื่องจากคอมพิวเตอร์ (Computer) เป็นเครื่องมือที่ทันสมัยที่พัฒนาโดยมนุษย์เพื่อช่วยงานของมนุษย์

โดยเฉพาะงานที่มีขั้นตอนซ้ำซ้อน มีปริมาณมาก และใช้แรงงานคนมาก เมื่อนำคอมพิวเตอร์ (Computer) มาจัดการงานลักษณะดังกล่าว ก็จะสามารถทำให้งานเสร็จได้รวดเร็วขึ้นมีผลลัพธ์ที่ถูกต้องน่าเชื่อถือ ลดแรงงานของมนุษย์ลง ทำให้มนุษย์มีเวลามากขึ้นที่จะคิดสร้างสรรค์สิ่งใหม่ ๆ อันจะเกิดประโยชน์แก่มนุษยชาติต่อไป (วัชรภรณ์ สุริยาภิวัฒน์. 2535 : 1)

ในวงการศึกษาก็ได้มีการนำคอมพิวเตอร์เข้ามาช่วยในการบริหาร การเรียนการสอน การวิจัย นอกจากนี้ยังมีการนำคอมพิวเตอร์มาประยุกต์ใช้กับการทดสอบแบบเทเลอร์ (Tailored Testing) ซึ่งเรียกว่าคอมพิวเตอร์ไรซ์ อแดปติฟ เทสติง (Computerized Adaptive Testing : CAT) หรือคอมพิวเตอร์ แอด มินนิชเทอะ อแดปติฟ เทสติง (Computer-Administered Adaptive Testing CAT) มาใช้แทนวิธีดำเนินการทดสอบที่ใช้กระดาษกับดินสอ (Paper and Pencil Test) (Mcbride and Martin. 1983 : 224 ; Citing Wain. 1983 : 71 ; Citing Weiss. 1983 : 6 - 7) ด้วยการให้คอมพิวเตอร์ (Computer) ทำหน้าที่จัดข้อคำถามให้เหมาะกับระดับความสามารถของผู้สอบ ซึ่งจะทำให้สามารถแก้ปัญหาเกี่ยวกับความยุ่งยากต่าง ๆ ของแบบทดสอบเทเลอร์ ดังที่กล่าวไว้ข้างต้นได้เป็นอย่างดี (ต่าย เชียงฉี. 2534 : 4 ; Citing Weiss and Kingsbury. 1984 : 361 – 375)

จะเห็นได้ว่าแบบทดสอบเทเลอร์สามารถนำมาให้แทนแบบทดสอบแบบดั้งเดิมได้เป็นอย่างดี ในต่างประเทศได้มีการศึกษาและใช้แบบทดสอบเทเลอร์อย่างกว้างขวาง ส่วนในประเทศไทยก็มีการนำแบบทดสอบเทเลอร์มาใช้แต่ยังไม่แพร่หลาย ได้มีผู้ทำการวิจัยเกี่ยวกับแบบทดสอบเทเลอร์เช่นเดียวกันแต่เป็นการดำเนินการสอบโดยใช้กระดาษกับดินสอ (Paper and Pencil Test) ดังนั้นผู้วิจัยจึงสนใจศึกษาความเที่ยงตรง และประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของผู้สอบโดยใช้แบบทดสอบเทเลอร์รูปปิรามิดโดยใช้คอมพิวเตอร์ทำหน้าที่จัดข้อคำถามให้เหมาะสมกับระดับความสามารถของผู้สอบ และประมาณค่าความสามารถของผู้สอบโดยทำการเลือกรูปแบบ และวิธีการให้คะแนนในหลายวิธี เพื่อให้มีการพัฒนาการใช้แบบทดสอบเทเลอร์โดยใช้คอมพิวเตอร์และนำไปใช้ต่อไป

ความมุ่งหมายในการศึกษาค้นคว้า

1. เพื่อศึกษาค่าความเที่ยงตรงของการประมาณค่าความสามารถของผู้สอบระหว่างแบบทดสอบเทเลอร์รูปปิรามิดในวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ที่มีรูปแบบ และจำนวนชั้น แตกต่างกัน
2. เพื่อศึกษาค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของผู้สอบระหว่างแบบทดสอบเทเลอร์รูปปิรามิดในวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ที่มีรูปแบบจำนวนชั้นแตกต่างกัน

ความสำคัญของการศึกษาค้นคว้า

ผลการศึกษาค้นคว้านี้ทำให้ทราบถึงคุณภาพและประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของผู้สอบโดยใช้แบบทดสอบเทเลอรรูปปิรามิดในวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ที่มีรูปแบบ จำนวนชั้น แตกต่างกัน 3 รูปแบบคือ แบบขนาดชั้นคงที่มีจำนวนชั้น 5 และ 15 ชั้น และแบบมีข้อสอบหลายข้อในแต่ละชั้น และการให้คะแนน 3 แบบ คือ การให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบข้อที่ทำ การให้คะแนนตามความยากเฉลี่ยของข้อสอบเฉพาะข้อที่ตอบถูก การให้คะแนนค่าความยากตามข้อสอบที่ถัดจากข้อสุดท้ายที่ได้ตอบ ซึ่งจะเป็นประโยชน์ในการพิจารณาเลือกรูปแบบของแบบทดสอบเลือกตอบให้มีความเหมาะสมกับระดับความสามารถของผู้สอบ ทั้งยังช่วยให้สามารถตัดสินใจเลือกใช้แบบทดสอบเลือกตอบที่มีคุณภาพและประสิทธิภาพสูงในการวัดผลสัมฤทธิ์ทางการเรียน เพื่อปรับปรุงการเรียนการสอนหรือใช้ในการสอนซ่อมเสริมนักเรียนที่มีระดับความสามารถแตกต่างกัน นอกจากนี้ยังเป็นแนวทางในการนำทฤษฎีการตอบสนองข้อสอบมาประยุกต์ใช้กับการทดสอบแบบเทเลอร์ เพื่อพัฒนาแบบทดสอบ และวิธีการดำเนินการสอบให้มีประสิทธิภาพต่อไป

ขอบเขตของการศึกษาค้นคว้า

ประชากร

การศึกษานี้ใช้กับกลุ่มประชากรที่เป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ในภาคเรียนที่ 2 ปีการศึกษา 2546 ของโรงเรียนสังกัดสำนักงานเขตพื้นที่การศึกษาราชบุรี เขต 1 และสังกัดเทศบาลเมืองราชบุรี ในเขตอำเภอเมืองราชบุรี จังหวัดราชบุรี จำนวนโรงเรียน 66 โรงเรียน จำนวนห้องเรียน 109 ห้องเรียน จำนวนนักเรียน 3,165 คน

กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการศึกษานี้เป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ในภาคเรียนที่ 2 ปีการศึกษา 2546 ของโรงเรียนสังกัดสำนักงานเขตพื้นที่การศึกษาราชบุรี เขต 1 และสังกัดเทศบาลเมืองราชบุรี ในเขตอำเภอเมืองราชบุรี จังหวัดราชบุรี จำนวนโรงเรียน 10 โรงเรียน จำนวนห้องเรียน 37 ห้องเรียน จำนวนนักเรียน 1,492 คน ซึ่งเลือกมาโดยการสุ่มอย่างง่าย (Simple Random Sampling) มีโรงเรียนเป็นหน่วยการสุ่ม (Sampling Unit) ได้จำนวนโรงเรียน 10 โรงเรียน จำนวนห้องเรียน 37 ห้องเรียน จำนวนนักเรียน 1,492 คน โดยสุ่มแบ่งกลุ่มตัวอย่างออกเป็น 2 กลุ่มคือ

1. กลุ่มตัวอย่างที่ใช้ในการสร้างเครื่องมือและตรวจสอบคุณภาพ จำนวนห้องเรียน 27 ห้องเรียน จำนวนนักเรียน 1,128 คน
2. กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ จำนวนห้องเรียน 10 ห้องเรียน จำนวนนักเรียน 364 คน

ตัวแปรที่ใช้ในการศึกษา คือ

ตัวแปรต้น

1. รูปแบบของแบบทดสอบเทเลอร์รูปปิรามิด 3 รูปแบบคือ
 - 1.1 รูปแบบที่ใช้ขนาดชั้นคงที่ ชนิดเพิ่ม 1 หรือลด 1 จำนวน 5 ชั้น
 - 1.2 รูปแบบที่ใช้ขนาดชั้นคงที่ ชนิดเพิ่ม 1 หรือลด 1 จำนวน 15 ชั้น
 - 1.3 รูปแบบที่มีข้อสอบหลายข้อในแต่ละชั้น มีจำนวนชั้น 5 ชั้น
2. แบบทดสอบดั้งเดิม
3. วิธีการให้คะแนนของแบบทดสอบปิรามิด 3 วิธี คือ
 - 3.1 การให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบข้อที่ทำ
 - 3.2 การให้คะแนนตามความยากเฉลี่ยของข้อสอบเฉพาะข้อที่ตอบถูก
 - 3.3 การให้คะแนนค่าความยากตามข้อสอบที่ถัดจากข้อสุดท้ายที่ได้ตอบ

ตัวแปรตาม

1. ความเที่ยงตรงของการประมาณค่าความสามารถของผู้สอบ
2. ประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของผู้สอบ

นิยามศัพท์เฉพาะ

1. แบบทดสอบเทเลอร์รูปปิรามิด (Pyramidal Test) หมายถึง แบบทดสอบที่จัดเรียงข้อสอบตามแบบจำลองรูปปิรามิดโดยในแต่ละเนื้อหาจะเรียงลำดับข้อสอบจากง่ายไปหายากในที่นี้เป็นแบบทดสอบวิชาคณิตศาสตร์ เรื่อง การบวก ลบ และคูณทศนิยม ระดับชั้นประถมศึกษาปีที่ 6 ซึ่งยึดถือค่าความยากของข้อสอบตามโครงสร้างที่ฐานและระดับชั้นของรูปปิรามิดแบ่งออกเป็น 3 รูปแบบคือ

1.1 รูปแบบที่มีขนาดชั้นคงที่ ใช้กฎการจำแนกแนวทางการตอบข้อสอบแบบ เพิ่ม 1 หรือลด 1 แบบ 5 ชั้น มีจำนวนข้อสอบทั้งหมด 15 ข้อ นักเรียนจะต้องทำข้อสอบในแต่ละชั้นชั้นละ 1 ข้อ รวมทั้งฉบับนักเรียนจะต้องทำข้อสอบทั้งสิ้น 5 ข้อ

1.2 รูปแบบที่มีขนาดชั้นคงที่ ใช้กฎการจำแนกแนวทางการตอบข้อสอบแบบ เพิ่ม 1 หรือลด 1 แบบ 15 ชั้น มีจำนวนข้อสอบทั้งหมด 120 ข้อ นักเรียนจะต้องทำข้อสอบในแต่ละชั้นชั้นละ 1 ข้อ รวมทั้งฉบับนักเรียนจะต้องทำข้อสอบทั้งสิ้น 15 ข้อ

1.3 รูปแบบที่มีข้อสอบหลายข้อในแต่ละชั้น เป็นแบบทดสอบชนิด 5 ชั้น แบบที่มีข้อสอบ 3 ข้อในแต่ละชั้น แนวทางการตอบข้อสอบแบบ เพิ่ม 1 หรือลด 1 มีจำนวนข้อสอบทั้งหมด 45 ข้อ

2. แบบทดสอบดั้งเดิม (Conventional Test) หมายถึง แบบทดสอบที่ประกอบด้วยข้อสอบชนิดเลือกตอบ ที่มี 4 ตัวเลือก ซึ่งมีตัวเลือกที่เป็นคำตอบถูกเพียงคำตอบเดียวและมีการจัดเรียงข้อสอบตามลำดับเนื้อหาวิชา โดยในแต่ละเนื้อหาจะเรียงลำดับข้อสอบจากง่ายไปหายาก

ในที่นี้เป็นแบบทดสอบวิชาคณิตศาสตร์ เรื่อง การบวก ลบ และคูณทศนิยม ระดับชั้นประถมศึกษาปีที่ 6 จำนวน 30 ข้อที่ผู้วิจัยสร้างขึ้น

3. ความยากข้อสอบ หมายถึง ค่าความสามารถตรงจุดเปลี่ยนโค้งคุณลักษณะของ ข้อสอบแต่ละข้อ (Item Characteristic Curve) ซึ่งเป็นระดับความสามารถที่จะตอบข้อสอบนั้น ได้ถูกต้องด้วยค่าความน่าจะเป็น .50 เมื่อไม่มีการเดา (ผจงจิต อินทสุวรรณ. 2528 : 16) การ วิจัยนี้คัดเลือกข้อสอบที่มีค่าความยากระหว่าง -3.00 ถึง $+3.00$

4. อำนาจจำแนกของข้อสอบ หมายถึง ค่าที่เป็นสัดส่วนผกผันกับความชันของโค้ง หรือความชันของความน่าจะเป็นที่ผู้สอบที่มีระดับความสามารถ (θ) จะตอบข้อสอบข้อนั้นถูก ณ จุดเปลี่ยนโค้ง (ผจงจิต อินทสุวรรณ. 2528 : 16) การวิจัยครั้งนี้คัดเลือกข้อสอบข้อที่มีค่าอำนาจ จำแนกมากกว่า .30 ขึ้นไป

5. สัมประสิทธิ์การเดา หมายถึง ความน่าจะเป็นที่บุคคลหนึ่งซึ่งปราศจากความ สามารถจะตอบข้อสอบข้อนั้นได้ถูกต้อง การวิจัยครั้งนี้คัดเลือกข้อสอบข้อที่มีสัมประสิทธิ์การเดา น้อยกว่า .30

6. ฟังก์ชันอินฟอร์เมชันของข้อสอบ (Item Information Function) หมายถึง ค่าที่ เป็นสัดส่วนกับกำลังสองของช่วงแห่งความเชื่อมั่นที่เอซิมโทติก (Asymptotic confidence interval) ในการประมาณค่าความสามารถของผู้สอบจากผลการตอบ (ผจงจิต อินทสุวรรณ. 2528 : 97)

7. ฟังก์ชันอินฟอร์เมชันของแบบทดสอบ (Test Information Function) หมายถึง ผลรวมของฟังก์ชันอินฟอร์เมชันของข้อสอบ

8. ความสามารถของผู้สอบ (Examinee's Ability : θ) หมายถึง ระดับความ สามารถของผู้สอบแต่ละคน ซึ่งวัดในหน่วย θ ที่มีค่าเฉลี่ยเท่ากับ 0 และค่าเบี่ยงเบนมาตรฐาน เท่ากับ 1 สำหรับค่าความสามารถที่มีค่าเป็นไปได้ตั้งแต่ $-\infty$ ถึง $+\infty$ ในการศึกษาครั้งนี้ใช้ค่า ความสามารถในช่วง -4.00 ถึง 4.00

9. วิธีการตรวจให้คะแนน (Scoring) หมายถึง วิธีในการประมาณค่าความสามารถ ของผู้สอบจากการทดสอบเทเลอร์ ซึ่งในการวิจัยครั้งนี้ใช้ 3 วิธี คือ

9.1 วิธีการตรวจให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบข้อที่ทำ

9.2 วิธีการตรวจให้คะแนนตามความยากเฉลี่ยของข้อสอบเฉพาะข้อที่ตอบถูก

9.3 วิธีการตรวจให้คะแนนค่าความยากตามข้อสอบที่ถัดจากข้อสุดท้ายที่ได้ตอบ

10. ความเที่ยงตรง (Validity) ของการประมาณค่าความสามารถ หมายถึง ดัชนีที่บ่งชี้ถึง คุณภาพของแบบทดสอบในการประมาณค่าความสามารถของผู้สอบได้ถูกต้องตรงกับสภาพที่ เป็นจริงในปัจจุบัน ซึ่งในที่นี้เป็นการหาค่าความเที่ยงตรงเชิงสภาพ คำนวณได้จากความ สัมพันธ์ระหว่างความสามารถที่ได้จากการประมาณค่า กับการประมาณค่าความสามารถของ ผู้สอบโดยใช้แบบทดสอบดั้งเดิมที่ผู้วิจัยสร้างขึ้น

11. ประสิทธิภาพสัมพัทธ์ (Relative Efficiency : RE) ของการประมาณค่าความสามารถ หมายถึง ดัชนีที่บ่งชี้ถึงประสิทธิภาพของแบบทดสอบ ซึ่งคำนวณได้จากการหาอัตราส่วนระหว่างค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบแต่ละฉบับที่นำมาเปรียบเทียบกับ

สมมติฐานของการศึกษาค้นคว้า

ผู้วิจัยได้ตั้งสมมติฐานดังนี้

1. ค่าความเที่ยงตรงของการประมาณค่าความสามารถของผู้สอบ โดยใช้แบบทดสอบเทเลอร์รูปปริมาตรโดยในวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ทั้ง 3 รูปแบบ จำแนกตามวิธีการตรวจให้คะแนนทั้ง 3 วิธี มีค่าแตกต่างกัน
2. ค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของผู้สอบ โดยแบบทดสอบเทเลอร์รูปปริมาตร รูปแบบที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้น มีค่าสูงกว่า รูปแบบที่ใช้ขนาดชั้นคงที่จำนวน 5 ชั้น และรูปแบบที่มีข้อสอบหลายข้อในแต่ละชั้น ในทุกระดับความสามารถ
3. ค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของผู้สอบ โดยใช้แบบทดสอบเทเลอร์รูปปริมาตร รูปแบบที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้นมีค่าสูงกว่า รูปแบบที่มีข้อสอบหลายข้อในแต่ละชั้น จำนวน 5 ชั้น ในทุกระดับความสามารถ

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

เอกสารและงานวิจัยที่เกี่ยวข้องกับการศึกษาค้นคว้าครั้งนี้ ผู้วิจัยขอนำเสนอตามลำดับ
ดังนี้

1. ทฤษฎีการตอบสนองข้อสอบ
 - 1.1 หลักการของทฤษฎีการตอบสนองข้อสอบ
 - 1.2 ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ
 - 1.3 โมเดลของทฤษฎีการตอบสนองข้อสอบ
2. การทดสอบแบบเทเลอร์
 - 2.1 หลักการของแบบทดสอบเทเลอร์
 - 2.2 ประเภทของการทดสอบแบบเทเลอร์
 - 2.3 การให้คะแนนของการทดสอบแบบเทเลอร์
 - 2.4 แบบทดสอบเทเลอร์รูปปิรามิด
3. คุณภาพของแบบทดสอบ
 - 3.1 การประมาณค่าพารามิเตอร์
 - 3.2 คุณภาพของแบบทดสอบ
 - 3.3 ประสิทธิภาพของแบบทดสอบ
4. งานวิจัยที่เกี่ยวข้อง
 - 4.1 งานวิจัยภายในประเทศ
 - 4.2 งานวิจัยในต่างประเทศ

1. ทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบที่มีชื่อเรียกอื่น ๆ อีก เช่น ทฤษฎีคุณลักษณะแฝง ทฤษฎีโครงสร้างข้อสอบ หรือทฤษฎีการทดสอบแบบใหม่ (ผจงจิต อินทสุวรรณ. 2528 : 54 ; Citing Hambleton and others. 1978 : 468) เป็นต้น ทฤษฎีนี้ได้เข้ามามีบทบาทในวงการวัด ผลการศึกษาและจิตวิทยาตั้งแต่ปี ค.ศ. 1950 – 1960 โดยในระยะแรกการพัฒนาเป็นไปอย่างช้า ๆ เพราะขาดแคลนคอมพิวเตอร์ที่ใช้ในการวิเคราะห์ข้อมูล แต่ในปัจจุบันได้มีการพัฒนา โปรแกรมคอมพิวเตอร์ จึงทำให้ได้รับความสะดวก และเป็นที่สนใจของนักวัดผลมากยิ่งขึ้น สำหรับข้อดีของทฤษฎีนี้ก็คือเชื่อกันว่าสามารถแก้ข้อบกพร่องของทฤษฎีการทดสอบแบบดั้งเดิมที่เราใช้กันมาเป็นเวลานานได้ เนื่องจากทฤษฎีการทดสอบแบบดั้งเดิมมีจุดอ่อนหลาย

ประการดังนี้ (อรินทร์ น่วมถนอม. 2537 : 12 – 13 ; Citing Hambleton and Swaminathan. 1985 : 1 – 3)

1. . ค่าสถิติของข้อสอบ ซึ่งได้แก่ค่าความยากของข้อสอบ และค่าอำนาจจำแนกของข้อสอบ มีค่าไม่คงที่ จะแปรเปลี่ยนไปตามกลุ่มผู้สอบ ดังนั้นการใช้ค่าสถิติดังกล่าวมีข้อจำกัด โดยจะใช้ได้กับกลุ่มผู้สอบที่มีลักษณะเหมือน ๆ กันเท่านั้น

2. การเปรียบเทียบความสามารถของผู้สอบ จะต้องใช้แบบทดสอบฉบับเดียวกันหรือแบบทดสอบคู่ขนานกันเท่านั้น เพราะถ้าใช้แบบทดสอบคนละฉบับที่วัดความสามารถแต่ความยากง่ายแตกต่างกัน แล้วจะนำคะแนนจากแบบทดสอบคนละฉบับมาเปรียบเทียบกันไม่ได้

3. นิยามความเชื่อมั่นของแบบทดสอบในรูปแบบการวัดคู่ขนาน ซึ่งเป็นไปได้ยากที่จะสัมฤทธิ์ผลในทางปฏิบัติ เนื่องจากผู้ตอบอาจจะตอบไม่เหมือนเดิมในครั้งที่สอง

4. ไม่มีพื้นฐานในการที่จะบอกได้ว่าผู้สอบคนหนึ่งจะมีความน่าจะเป็นในการตอบข้อสอบข้อหนึ่ง ได้ถูกต้องเพียงใด ซึ่งทำให้ไม่สามารถพัฒนาแบบทดสอบให้เหมาะสมกับระดับความสามารถของผู้สอบแต่ละคนได้

5. การกำหนดข้อตกลงเบื้องต้นให้ความแปรปรวนของความคลาดเคลื่อนในการวัดของผู้สอบทุกคนมีค่าเท่ากัน ซึ่งขัดแย้งกับผลการวิจัยของ ลอร์ดและโนวิก (Lord and Novick. 1968 : 183) ที่พบว่า ความคลาดเคลื่อนในการวัดของผู้สอบแต่ละคนมีค่าไม่เท่ากัน โดยค่าคลาดเคลื่อนในการวัดสำหรับคะแนนปานกลาง จะมีค่าน้อยกว่าค่าความคลาดเคลื่อนในการวัดสำหรับคะแนนสูงหรือคะแนนต่ำ

จากปัญหาข้อบกพร่องของทฤษฎีการทดสอบแบบดั้งเดิมดังที่กล่าวมาแล้วจึงทำให้นักวัดผลการศึกษาได้ให้ความสนใจกับทฤษฎีการตอบสนองข้อสอบ ซึ่งทฤษฎีนี้มีจุดเด่นดังนี้ (อรินทร์ น่วมถนอม. 2537 : 13 ; Citing Hambleton and Swaminathan. 1985 : 11)

1. ค่าพารามิเตอร์ของข้อสอบ เป็นอิสระจากกลุ่มผู้สอบที่ใช้ในการประมาณค่า
2. ค่าความสามารถของผู้สอบ เป็นอิสระจากข้อสอบที่ใช้ในการประมาณค่า
3. สามารถหาความแม่นยำ ในการประมาณค่าความสามารถของผู้สอบได้

1.1 หลักการของทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบ เป็นทฤษฎีที่ใช้โมเดลทางคณิตศาสตร์อธิบายถึงความสัมพันธ์ระหว่างความสามารถหรือคุณลักษณะภายในของผู้สอบ กับผลการสอบหรือคะแนนของผู้สอบนั้น โดยทฤษฎีนี้เชื่อว่าจะคะแนนของผู้สอบ ซึ่งเป็นสิ่งที่วัดหรือสังเกตได้โดยตรงสามารถอธิบายหรือพยากรณ์ได้จากความสามารถของผู้สอบคนนั้น ซึ่งเป็นสิ่งที่ไม่สามารถวัดหรือสังเกตได้ (Lord and Novick. 1968 : 358 ; Hambleton and Swaminathan. 1985 : 9) และโมเดลทางคณิตศาสตร์ที่ใช้อธิบายถูกกำหนดด้วยฟังก์ชันทางเดียว (Monotonically Increasing

Function) ซึ่งเรียกว่า ฟังก์ชันลักษณะข้อสอบ หรือ โคลงลักษณะข้อสอบ (อรินทร์ น่วมทนอม. 2537 : 14 ; Citing Hambleton and Swaminathan. 1985 : 13)

สำหรับความสัมพันธ์ระหว่างความสามารถของผู้สอบกับผลการสอบ สามารถเขียนให้อยู่ในรูปความสัมพันธ์ทั่ว ๆ ไปได้ดังนี้ (สำเริง บุญเรืองรัตน์. 2525 : 2)

$$P = f(\theta)$$

เมื่อ P	แทน	ผลการสอบ
θ	แทน	ความสามารถ
f	แทน	ฟังก์ชัน

1.2 ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ

ลอร์ด และโนวิก (Lord and Novick. 1986 : 359) กล่าวว่าข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบประกอบด้วยลักษณะดังนี้ (เกศมณี พยัคฆ์. 2543 : 10)

1. ความเป็นมิติเดียว (Unidimension) หมายความว่า ความสามารถของคนที่อยู่ทั้งหมด k อย่างซึ่งความสามารถแต่ละอย่างต่างก็ส่งผลต่อการตอบข้อสอบต่าง ๆ ที่รวมกันเป็นแบบทดสอบ ถ้าผลการตอบข้อสอบหรือคะแนนของผู้สอบนั้น สามารถอธิบายได้เพียงความสามารถเดียวก็เพียงพอแล้ว เรียกเงื่อนไขข้อนี้ว่ามีความเป็นมิติเดียว ในการปฏิบัติจริงนั้นสามารถกระทำโดยการวิเคราะห์องค์ประกอบ แล้วสังเกตค่าไอเกน (Eigen Value) ค่าสูงสุดว่าแตกต่างจากค่าอื่นอย่างชัดเจนหรือไม่ นอกจากนี้ยังสามารถเลือกข้อสอบที่ความเป็นมิติเดียวได้โดยทำการเลือกข้อสอบที่มีค่าอำนาจจำแนกสูง ๆ ซึ่งไม่ควรต่ำกว่า 0.8 (Green and other. 1984 : 349)

2. ความเป็นอิสระในการตอบข้อสอบ หมายถึง การตอบข้อสอบข้อใดข้อหนึ่งจะไม่มีผลต่อการตอบข้อสอบนั้น ถ้าข้อสอบชุดนั้นมีความเป็นมิติเดียวแล้วความเป็นอิสระในการตอบข้อสอบก็จะมีเอง ซึ่งสามารถตรวจสอบได้จากการคำนวณค่าไคสแคว์ เพื่อทดสอบความเป็นอิสระของการตอบคำถาม

- ถ้าให้
- (1) U_g แทน ผลการตอบคำถามข้อที่ g ของผู้สอบคนหนึ่ง
 $g = 1, 2, 3, \dots, n$
 โดยถือหลักว่าถ้าตอบถูกให้ 1 และตอบผิดให้ 0
 - (2) P_g เป็นโอกาสที่ผู้สอบคำถามข้อ g ถูกต้อง
 - (3) $Q_g = 1 - P_g$

ข้อตกลงเกี่ยวกับ Local Independence กล่าวว่า

$$\text{Prob}[U_1 = u_1, U_2 = u_2, \dots, U_n = u_n] \\ = \prod_{g=1}^n P_g^{u_g} Q_g^{1-u_g} \quad \dots\dots\dots (2)$$

สมการสองนั้นอ่านว่าโอกาสของการตอบ (Prob ซึ่งย่อมาจาก Probability) แบบทดสอบได้ถูกต้องทั้งหมดทุกข้อคือผลคูณ ของโอกาสของการตอบคำถามแต่ละข้อได้ถูกต้อง จากข้อดังกล่าวนี้ ทำให้สามารถสร้างสมการแสดงความถี่ของคะแนนของกลุ่มผู้สอบที่มีความสามารถ θ ได้ดังนี้

$$f(x / \theta) = \sum \prod_{g=1}^n P_g^{u_g} Q_g^{1-u_g} \quad \dots\dots\dots (3)$$

เมื่อ x แทน คะแนนของผู้สอบซึ่งมีค่าตั้งแต่ 0 ถึง n

3. โค้งแสดงลักษณะของข้อสอบ (Item Characteristic Curve : ICC)

การแจกแจงความถี่ของคะแนนของกลุ่มผู้สอบที่มีระดับความสามารถ θ นั้นหาได้จากสมการ (3) ถ้าแบบทดสอบวัดความสามารถเดียวสมการที่ (3) เรียกว่าโค้งลักษณะของแบบทดสอบ (Test Characteristic Curve) แต่ถ้าวัดความสามารถหลายอย่างเรียกว่า ฟังก์ชันลักษณะข้อสอบ (Test Characteristic Function)

สำหรับคำถามแต่ละข้อนั้นก็สามารถแสดงลักษณะของโค้งผลการตอบได้เช่นกัน เรียก ว่า โค้งลักษณะข้อสอบ (Item Characteristic Curves) สำหรับคำถามที่ข้อถูกเป็น 1 และผิดเป็น 0 นั้น สมการแจกแจงความถี่สามารถแสดงได้ดังนี้

$$f_g(U_g / \theta) = P_g^{u_g} Q_g^{(1-u_g)} \quad \dots\dots\dots (4)$$

$$\text{นั่นคือ } f_g(U_g / \theta) = P_g \text{ ถ้า } U_g = 1 \\ = Q_g \text{ ถ้า } U_g = 0$$

แสดงถึงความสัมพันธ์ระหว่างโอกาสในการตอบข้อสอบข้อนั้นถูกต้องกับระดับความสามารถที่วัดได้ โดยใช้ชุดของข้อสอบ โดยที่โอกาสที่ผู้สอบตอบข้อสอบถูกจะขึ้นอยู่กับโค้งแสดงลักษณะข้อสอบ ในแต่ละโมเดลที่ใช้ ด้วยเหตุนี้โอกาสของการตอบถูกข้อสอบข้อนั้นจะไม่ขึ้นอยู่กับ การแจกแจงความสามารถของกลุ่มประชากร

1.3. โมเดลของทฤษฎีการตอบสนองข้อสอบ

โมเดลของทฤษฎีการตอบข้อคำถามเป็นโมเดลทางคณิตศาสตร์ที่กล่าวถึงกันมาก มีอยู่ 2 โมเดลคือ (Warm. 1978 : 23-25 ; Citing Hambleten and Cook. 1977 : 81 – 82)

1. โมเดลโค้งปกติสะสม (Normal Ogive Model) มี 3 รูปแบบ คือ

1.1 แบบ 1 พารามิเตอร์ (One-Parameter) เป็นรูปแบบที่มีค่าพารามิเตอร์ตัวเดียว คือค่า b ที่แสดงลักษณะของข้อสอบ เมื่อ θ คือ ความสามารถของผู้สอบ แสดงได้ด้วยฟังก์ชันดังนี้ .

$$P_g(\theta) = \int_{-\infty}^{\theta-b_g} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

1.2. แบบ 2 พารามิเตอร์ (Two-Parameter) เป็นรูปแบบที่ประกอบด้วย 2 พารามิเตอร์คือ a และ b ที่แสดงลักษณะข้อสอบ เมื่อ θ คือความสามารถของผู้สอบ มีฟังก์ชันดังนี้

$$P_g(\theta) = \int_{-\infty}^{a(\theta-b_g)} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

1.3. แบบ 3 พารามิเตอร์ (Three-Parameter) เป็นแบบที่มีค่าพารามิเตอร์ลักษณะข้อสอบ 3 ค่า คือ a , b และ c เมื่อ θ คือความสามารถของผู้สอบ มีฟังก์ชันดังนี้

$$P_g(\theta) = c_g + (1 - c_g) \int_{-\infty}^{a_g(\theta-b_g)} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

- | | | |
|-----------|-----|---|
| เมื่อ a | แทน | ค่าอำนาจจำแนกหรือสัดส่วนของความชัน ณ จุดเปลี่ยน |
| | | โค้ง |
| b | แทน | ค่าความยากของข้อสอบหรือตำแหน่งของข้อสอบบนเส้นความสามารถ |
| c | แทน | ค่าโอกาสการเดาถูก |

2. โมเดลโลจิสติก (Logistic Model) เป็นโมเดลที่พัฒนาเพื่อพัฒนาให้เหมาะสมกับสถานการณ์การทดสอบในกรณีที่ต้องตอบโดยการเดาได้ ซึ่งจะทำให้ส่วนปลายที่สุดของโค้งของ $P_i(\theta)$ ดังนั้นพารามิเตอร์ของการเดาจึงจำเป็นต้องพิจารณาด้วย นอกจากนี้หากผู้สอบทำข้อสอบด้วยความสะเพร่า โมเดลนี้จะเพิ่มความรอบคอบขึ้นอีกค่าหนึ่ง ซึ่งจะเห็นได้จากปลายสูงสุดของโค้ง ICC มีค่าน้อยกว่า 1 โมเดลโค้งปกติสะสมกับโมเดลโลจิสติก จึงมีลักษณะของโค้งใกล้เคียงกัน แต่โมเดลโลจิสติกมีความง่ายในการประมาณค่ามากกว่า จึงทำให้แพร่หลายมากกว่า รูปแบบโมเดลโลจิสติกมี 4 รูปแบบ

2.1 แบบ 1 พารามิเตอร์ (One-Parameter) หรือรูปแบบราล์ชโมเดล ราล์ช ได้

พัฒนาโมเดลนี้ โดยถือว่า ข้อสอบทุกข้อมีค่าอำนาจจำแนกเท่ากันหมด และไม่มีการเดาได้ถูกต้อง จึงมีค่าพารามิเตอร์แสดงลักษณะข้อสอบเพียงตัวเดียว คือ ค่าความยากเขียนเป็นฟังก์ชันได้ดังนี้

$$P_g(\theta) = \frac{e^{Da_g(\theta - b_g)}}{1 + e^{Da_g(\theta - b_g)}} ; g=1,2,3,\dots,n$$

หรือ

$$= \frac{e^{(\theta - b_g)}}{1 + e^{(\theta - b_g)}} = \frac{\text{EXP}(\theta - b_g)}{1 + \text{EXP}(\theta - b_g)}$$

2.2 แบบ 2 พารามิเตอร์ (Two-Parameter) หรือรูปแบบที่เบิร์ตบอม (Birnbaum) ได้พัฒนาขึ้นในปี 1968 มีค่าพารามิเตอร์แสดงลักษณะข้อสอบ 2 ค่า คือค่าอำนาจจำแนก และค่าความยาก แสดงได้โดยฟังก์ชันดังนี้

$$P_g(\theta) = \frac{e^{Da_g(\theta - b_g)}}{1 + e^{Da_g(\theta - b_g)}} ; g=1,2,3,\dots,n$$

หรือ

$$= \frac{\text{EXP}(Da_g(\theta - b_g))}{1 + \text{EXP}(Da_g(\theta - b_g))}$$

2.3 แบบ 3 พารามิเตอร์ (Three-Parameter) หรือรูปแบบที่เบิร์ตบอมได้พัฒนาโดยมีค่าพารามิเตอร์แสดงคุณลักษณะข้อสอบ 3 ค่า โดยเพิ่มค่าการเดา (C) ใน 2 พารามิเตอร์ เขียนเป็นฟังก์ชันได้ดังนี้

$$P_g(\theta) = c_g + (1 - c_g) \frac{e^{Da_g(\theta - b_g)}}{1 + e^{Da_g(\theta - b_g)}} ; g=1,2,3,\dots,n$$

หรือ

$$= c_g + (1 - c_g) \frac{\text{EXP}(Da_g(\theta - b_g))}{1 + \text{EXP}(Da_g(\theta - b_g))}$$

เมื่อ $P_g(\theta)$	แทน	ความน่าจะเป็นที่ผู้สอบคนหนึ่ง มีระดับความสามารถจะตอบข้อสอบข้อ g ได้ถูกต้อง
c_g	แทน	ค่าความน่าจะเป็นของผู้สอบที่มีความสามารถต่ำมาก มีโอกาสจะทำข้อสอบข้อนี้ได้ถูกต้อง หรือค่าการเดา
b_g	แทน	ค่าความยากของข้อสอบข้อที่ g ที่แสดงระดับความสามารถที่แท้จริงที่จุดโค้งชันที่สุดคือ θ ณ จุดความน่าจะเป็น $(1 + c) / 2$

a_g	แทน	ค่าอำนาจจำแนกข้อสอบข้อ g ที่เป็นสัดส่วนกับความชันของ $P_g(\theta)$ ณ ตำแหน่ง $\theta = b_g$
D	แทน	ค่าคงที่ของ Scaling factor มีค่าเท่ากับ 1.702
e	แทน	ค่าคงที่มีค่าเท่ากับ 2.7182818

2.4 แบบ 4 พารามิเตอร์ (Four-Parameter) โมเดลนี้เสนอโดยแม็คโคแนลต์ในปี ค.ศ. 1967 ลอร์ด และ บาร์ตัน ได้พัฒนาในปี ค.ศ. 1981 โดยเพิ่มพารามิเตอร์ค่าความสะเพร่า (Careless) ในโมเดลที่ใช้พารามิเตอร์ 3 ตัว เป็นค่าที่แสดงให้เห็นถึงกรณีที่ผู้สอบมีความสามารถสูงมักตอบข้อสอบผิด อาจเป็นเพราะความไม่รอบคอบทำให้ผู้สอบไม่เลือกคำตอบที่ถูก จะเห็นได้ว่าปลายสูงสุดของโค้ง ICC ต่ำกว่า 1 เขียนเป็นสมการได้ดังนี้

$$P_g(\theta) = c_g + (\gamma_g - c_g) \frac{e^{Da_g(\theta - b_g)}}{1 + e^{Da_g(\theta - b_g)}}$$

3. ราชส์โมเดล

ราชส์โมเดล เป็นโมเดลหนึ่งในโมเดลของทฤษฎีการตอบสนองข้อสอบที่ใช้พารามิเตอร์ตัวเดียว คือจะมีการพิจารณาค่าความยากตัวเดียว โดยจะถือว่าข้อสอบทุกข้อมีค่าอำนาจจำแนกเท่ากันหมดและไม่มีการเดาได้ถูกต้อง กล่าวคือ เมื่อคนหนึ่ง ๆ ทำข้อสอบข้อหนึ่งถ้าคนนั้นมีความสามารถมากกว่าความยากของข้อสอบข้อนั้น โอกาสที่คนจะทำข้อสอบถูกก็มีมากไปด้วย และถ้าความสามารถของคนสอบมีพอ ๆ กันกับความยากของข้อสอบโอกาสของการทำข้อสอบก็จะมีอยู่ครึ่งต่อครึ่ง คืออาจทำถูกหรืออาจทำผิดได้เช่นเดียวกัน

เมื่อกำหนดให้ θ คือความสามารถของคน และ b คือ ความยากข้อสอบ การเปรียบเทียบความสามารถกับความยากของข้อสอบผลลัพธ์ อาจเขียนในเชิงคณิตศาสตร์ได้ดังนี้

$$\text{ถ้า } \theta > b \text{ และ } (\theta - b) > 0, P > .5$$

อธิบายได้ว่า เมื่อความสามารถของคนมีมากกว่าความยากของข้อสอบ คือ b_g และ ถ้านำเอาความสามารถของคนลบด้วยความยากของข้อสอบแล้ว ผลลัพธ์จะได้มากกว่า 0 โอกาสที่คน ๆ นั้นจะทำข้อสอบข้อนั้นถูกก็จะมีค่ามากกว่า 0.50

$$\text{ถ้า } \theta < b \text{ และ } (\theta - b) < 0, P < .5$$

อธิบายได้ว่า เมื่อความสามารถของคนลบด้วยความยากของข้อสอบได้ผลลัพธ์น้อยกว่าศูนย์โอกาสที่คน ๆ นั้นจะทำข้อสอบข้อนั้นได้ถูกต้อง ก็จะมีค่าน้อยกว่า 0.50

$$\text{ถ้า } \theta = b \text{ และ } (\theta - b) = 0, P = .5$$

อธิบายได้ว่า เมื่อความสามารถของคนลบด้วยความยากของข้อสอบได้ผลลัพธ์เท่ากับ

ศูนย์พอดีแล้ว โอกาสที่คน ๆ นั้นจะทำข้อสอบข้อนั้นได้ถูกต้องก็จะมีค่าเท่ากับ 0.50 พอดี ซึ่งอาจทำข้อสอบได้ถูกหรือผิดก็ได้

จะเห็นได้ว่า เมื่อคน ๆ หนึ่งทำข้อสอบข้อหนึ่ง บุคคลนั้นต้องใช้ความสามารถต่อสู้กับความยากของข้อสอบ ผลการทำข้อสอบจึงเท่ากับความแตกต่างระหว่างความสามารถของคนกับความยากของข้อสอบ ความแตกต่างระหว่าง θ กับ b หรือ $(\theta - b)$ นั้นเอง แต่ค่า $\theta - b$ นี้อาจจะให้ผลลัพธ์ได้ตั้งแต่ $-\infty$ จนถึง $+\infty$ จึงต้องคิดแปลงให้เป็นค่าความน่าจะเป็นซึ่งจะมีค่าอยู่ระหว่าง .00 ถึง .99 หรือจาก 0 ถึง 1 เท่านั้น และเพื่อให้ค่าเปลี่ยนแปลงแต่น้อย จึงใช้ความน่าจะเป็นค่าอัตราส่วนของลอการิทึม (Natural log) ของค่าคงที่ตัวหนึ่งคือ ค่า $e^{(\theta-b)}$ และค่าอัตราส่วนนั้นก็คือ

$$P(\theta) = \frac{e^{(\theta-b)}}{1 + e^{(\theta-b)}}$$

วราลีช โมเดล เป็นโมเดลที่มีผู้นิยมนำมาประยุกต์ใช้กันมาก เนื่องจากเป็นโมเดลที่มีความแกร่ง (Robustness) ซึ่งมีรายละเอียดดังนี้

1. เป็นโมเดลที่นำมาใช้ได้อย่างง่าย ๆ ไม่ซับซ้อนเหมือนกับโมเดลอื่น ๆ ของทฤษฎีการตอบสนองข้อสอบ ถ้าข้อมูลมีไม่มากเกินไปสามารถคำนวณด้วยเครื่องคิดเลขธรรมดาได้ (Wright and Stone. 1979 : 28 - 45)

2. ความเป็นอิสระจากกลุ่มตัวอย่าง (Sample - Free) และความเป็นอิสระจากข้อสอบ (Item - Free) กล่าวคือ ความยากของข้อสอบจะไม่แปรเปลี่ยนไปตามกลุ่มตัวอย่าง และค่าความสามารถของผู้สอบจะไม่แปรเปลี่ยนไปตามค่าความยากของข้อสอบ (Wright and Panchapakesan. 1969 : 23)

3. การกระจายค่าความสามารถของกลุ่มตัวอย่าง หรือค่าความยากของข้อสอบ ไม่จำเป็นต้องเป็นโค้งปกติ เพียงแต่มีแนวโน้มว่าจะเป็นโค้งปกติเท่านั้นก็ได้ และกลุ่มตัวอย่างไม่จำเป็นที่จะได้มาด้วยการสุ่มก็ได้ (Wright and Stone. 1979 : 20)

4. ข้อสอบไม่จำเป็นต้องมีค่าอำนาจจำแนกเท่ากัน และไม่จำเป็นต้องคำนึงถึงค่าการเดา (Wright and Panchapakesan. 1969 : 25)

5. ขนาดของกลุ่มตัวอย่างที่ใช้กับวราลีช โมเดล ไม่ต้องใช้มากเหมือนกับโมเดลอื่น ๆ ที่มีจำนวนพารามิเตอร์หลายตัว ดังเช่น โลจิสติก โมเดล ที่มีพารามิเตอร์ 3 ตัว จะต้องใช้กลุ่มตัวอย่างขนาดใหญ่ประมาณ 1,000 คนขึ้นไป (Warm. 1978 : 109) แต่วราลีช โมเดล จะใช้กลุ่มตัวอย่างขนาดเล็ก ซึ่งจากการศึกษาของ แวน เดอ ไวจ์เวอร์ (Van de Vijver. 1986 : 48) พบว่า กลุ่มตัวอย่างขนาด 100 - 500 คน เมื่อใช้กับข้อสอบตั้งแต่ 25 ข้อขึ้นไป สามารถประมาณค่าพารามิเตอร์ได้อย่างมีประสิทธิภาพ

ทฤษฎีการตอบสนองข้อสอบเป็นทฤษฎีที่สามารถเข้ามาช่วยแก้ไขข้อบกพร่องของ

ทฤษฎีการทดสอบแบบดั้งเดิม ซึ่งทฤษฎีการตอบสนองข้อสอบก็มีจุดเด่นที่น่าสนใจได้แก่ ค่าพารามิเตอร์ของข้อสอบเป็นอิสระจากกลุ่มผู้สอบที่ใช้ในการประมาณค่าความสามารถของผู้สอบ เป็นอิสระจากข้อสอบที่ใช้ในการประมาณค่าสามารถหาความแม่นยำ ในการประมาณค่าความสามารถของผู้สอบได้ เป็นต้น ในการนำทฤษฎีการตอบสนองข้อสอบไปใช้ก็จะต้องปฏิบัติตามข้อตกลงเบื้องต้น คือ แบบทดสอบต้องเป็นมิติเดียว ความเป็นอิสระในการตอบข้อสอบ และความน่าจะเป็นที่ผู้สอบจะตอบข้อสอบได้ถูกต้องจะขึ้นอยู่กับโครงสร้างของข้อสอบ ส่วนโมเดลที่ใช้ขึ้นอยู่กับพารามิเตอร์ที่ใช้แสดงลักษณะข้อสอบ ได้แก่ ค่าความยาก (a_i) ค่าอำนาจจำแนก (b_i) และค่าการเดา (c_i)

สำหรับในการศึกษาครั้งนี้ ผู้วิจัยได้ใช้พารามิเตอร์ค่าความยากในการจัดข้อสอบตามค่าความยาก โดยไม่ได้ศึกษาพารามิเตอร์ที่เกี่ยวกับค่าอำนาจจำแนก และค่าการเดาของข้อสอบ ดังนั้นผู้วิจัยจึงเลือกใช้ ราล์ช โมเดล ในการวิเคราะห์ข้อมูลตามทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบได้มีการพัฒนาขึ้นมาเพื่อนำมาแก้ไขข้อบกพร่องของทฤษฎีการทดสอบแบบดั้งเดิม ซึ่งยังมีจุดอ่อนหลายประการ ได้แก่ ค่าสถิติของข้อสอบจะแปรเปลี่ยนไปตามกลุ่มผู้สอบ การใช้ค่าสถิติข้อจำกัดโดยจะใช้ได้กับกลุ่มผู้สอบที่มีลักษณะเหมือนกันเท่านั้น การเปรียบเทียบความสามารถของผู้สอบจะต้องใช้แบบทดสอบฉบับเดียวกันหรือแบบทดสอบคู่ขนานกันเท่านั้น ซึ่งเป็นไปได้ยาก

ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ ได้แก่ ความเป็นอิสระในการตอบข้อสอบ คือ การตอบข้อสอบข้อใดข้อหนึ่งได้ถูกหรือผิดจะไม่มีผลต่อการตอบข้อสอบข้ออื่น ความเป็นมิติเดียวกันของแบบทดสอบ และ โครงสร้างข้อสอบ คือ ผู้สอบมีความสามารถในระดับสูง จะมีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องมากกว่าผู้สอบที่มีความสามารถในระดับต่ำ

ในส่วนของโมเดลของทฤษฎีการตอบสนองข้อสอบแบ่งออกเป็น โมเดลที่มีการตรวจให้คะแนนเป็นสองค่า (Dichotomous) โมเดลที่มีการตรวจให้คะแนนเป็นหลายค่า (Multi - Dichotomous) และ โมเดลที่มีการตรวจให้คะแนนแบบต่อเนื่อง (Continuous)

2. การทดสอบแบบเทเลอร์

การทดสอบแบบดั้งเดิม ที่เราใช้กันอยู่ทั่ว ๆ ไปนั้น เป็นการนำแบบทดสอบซึ่งประกอบด้วยข้อสอบที่มีค่าความยากในระดับปานกลางไปทดสอบกับกลุ่มผู้สอบ โดยผู้สอบแต่ละคนจะต้องทำข้อสอบในแบบทดสอบทุกข้อ ซึ่งเป็นข้อสอบที่มีการจัดเรียงลำดับไว้แน่นอน ดังนั้นแบบทดสอบในลักษณะนี้จึงมีความเหมาะสมกับกลุ่มผู้สอบที่มีความสามารถในระดับเฉลี่ยของกลุ่มเท่านั้น แต่จะไม่เหมาะสมกับกลุ่มผู้สอบที่มีความสามารถในระดับสูงหรือต่ำ (ผจงจิต อินทสุวรรณ . 2525 : 53) แบบทดสอบที่มีประสิทธิภาพสูงจะต้องประกอบด้วยข้อสอบที่มีค่าความยากเหมาะสมกับระดับความสามารถของผู้สอบ นั่นคือ ข้อสอบต้องไม่ง่ายหรือยากเกินไป

ไปสำหรับผู้สอบแต่ละคน (Hambleton and others. 1987 : 496 – 497 ; Citing Wood. 1973) ด้วยเหตุนี้ จึงทำให้นักวัดผลได้คิดค้นวิธีการสร้างข้อสอบและวิธีดำเนินการสอบแบบใหม่ เพื่อที่จะทำให้วัดความสามารถของผู้สอบได้ครอบคลุมทุกระดับความสามารถ โดยการปรับข้อสอบให้เหมาะสมกับระดับความสามารถของผู้สอบแต่ละคน ซึ่งวิธีการทดสอบดังกล่าวนี้มีชื่อเรียกแตกต่างกันออกไป เช่น การทดสอบที่ปรับตามระดับความสามารถ การทดสอบโดยใช้คอมพิวเตอร์ การทดสอบเป็นรายบุคคล การทดสอบแบบโปรแกรม การทดสอบที่มีการจัดเรียงลำดับข้อสอบ หรือชื่อที่ เทอร์นบูล เรียกว่า การทดสอบแบบเทเลอร์ (Lord. 1980 : 150 – 151 ; Citing Turnbull. 1951) โดยในระยะแรกยังไม่ได้นำมาใช้อย่างจริงจังเนื่องจากมีปัญหาเกี่ยวกับความยุ่งยากในการดำเนินการสอบที่ใช้การเขียนตอบแบบเดิม (Urry. 1977 : 181) ต่อมาในระยะหลังได้มีการพัฒนาโดยใช้เครื่องคอมพิวเตอร์มาช่วยในการดำเนินการสอบ ซึ่งสามารถจัดข้อสอบให้มีความเหมาะสมกับระดับความสามารถของผู้สอบ พร้อมทั้งนำเอาทฤษฎีการตอบสนองข้อสอบมาประยุกต์ใช้ในการประมาณค่าความสามารถของผู้สอบ จึงทำให้การทดสอบแบบเทเลอร์มีประสิทธิภาพมากยิ่งขึ้น (Rudner. 1983 : 951)

2.1 หลักการทดสอบแบบเทเลอร์

การทดสอบแบบเทเลอร์ เป็นการทดสอบที่มีการจัดข้อสอบให้มีความเหมาะสมกับระดับความสามารถของผู้สอบแต่ละคน (Lord. 1971 : 3 ; Citing Hambleton and others. 1978 : 497) ซึ่งผู้สอบแต่ละคนไม่จำเป็นต้องทำข้อสอบเหมือนกันทุกข้อหรือจำนวนของข้อสอบก็อาจไม่เท่ากัน ทั้งนี้จะขึ้นอยู่กับรูปแบบและผลการตอบข้อสอบในแบบทดสอบนั้น ๆ (Weiss and Kingsbury. 1984 : 361) โดยทั่วไปแล้วจะให้ผู้สอบเริ่มทำข้อสอบที่มีค่าความยากปานกลางก่อน ถ้าตอบถูกข้อต่อไปจะได้ทำข้อสอบที่มีค่าความยากเพิ่มมากขึ้น แต่ถ้าตอบข้อสอบข้อนั้นผิดข้อต่อไปจะได้ทำข้อสอบที่ง่ายขึ้น และการทำข้อสอบข้อต่อ ๆ ไป จะดำเนินการในลักษณะนี้ไปจนสิ้นสุดการทดสอบ (Green and others. 1984 : 347) ตลอดช่วงระยะเวลาที่ดำเนินการสอบจะมีการประมาณค่าระดับความสามารถของผู้สอบโดยการวินิจฉัยจากผลการตอบของผู้สอบในข้อที่ตอบมาแล้วนั้น ดังนั้นผู้สอบแต่ละคนจะได้รับจำนวนข้อสอบและการจัดเรียงข้อสอบที่แตกต่างกัน (Water and Bayroff. 1971 : 125 ; Citing Spinetti and Hambleton. 1977 : 140)

2.2 ประเภทของการทดสอบแบบเทเลอร์

การทดสอบแบบเทเลอร์แบ่งออกเป็น 2 ประเภทใหญ่ ๆ ดังนี้ (Weiss. 1974 : 78 – 110 ; Citing Hambleton and Swaminathan. 1985 : 297)

2.2.1 ยุทธวิธีสองขั้นตอน (Two – Stage Strategies) เป็นยุทธวิธีของการทดสอบแบบเทเลอร์อย่างง่าย ๆ โดยผู้สอบจะต้องทำแบบทดสอบถึงสองครั้งหรือสองขั้นตอน (Two –

Stage Test) ขั้นตอนแรกเป็นแบบทดสอบแยกกลุ่ม (Routing Test) เพื่อแยกกลุ่มผู้สอบตามระดับความสามารถ ส่วนขั้นตอนที่สองเป็นแบบทดสอบวัดผล (Measurement Test) ซึ่งแบ่งออกเป็นแบบทดสอบฉบับย่อย ๆ หลายฉบับตามระดับความยาก ถ้าผู้สอบได้รับการประเมินจากแบบทดสอบแยกกลุ่มว่ามีความสามารถในระดับต่ำ ระดับปานกลาง หรือระดับสูงแล้วจะได้ทำแบบทดสอบวัดผลที่มีค่าความยากในระดับต่ำ ระดับปานกลาง หรือระดับสูงตามลำดับ

2.2.2 ยุทธวิธีหลายขั้นตอน (Multi – Stage Strategies) เป็นยุทธวิธีของการทดสอบแบบเทอร์ที่ผู้สอบจะต้องตัดสินใจแยกทางหลาย ๆ ครั้ง จากการตอบข้อสอบแต่ละข้อ ซึ่งแตกต่างกับยุทธวิธีสองขั้นตอนที่ผู้สอบจะต้องตัดสินใจแยกทางเพียงครั้งเดียวจากการตอบแบบทดสอบแยกกลุ่มไปยังแบบทดสอบวัดผล สำหรับรูปแบบของยุทธวิธีหลายขั้นตอนแบ่งออกเป็น 2 รูปแบบ ดังนี้

2.2.2.1 รูปแบบแยกทางคงที่ (Fixed – Branching Model) เป็นรูปแบบที่มีการกำหนดข้อสอบ และเส้นทางในการตอบข้อสอบไว้ล่วงหน้า เพื่อกำหนดให้ผู้สอบแต่ละคนทำข้อสอบจากขั้นหนึ่งไปยังอีกขั้นหนึ่ง ซึ่งแบ่งออกเป็น 3 รูปแบบดังนี้

1. แบบทดสอบปิรามิด (Pyramidal Model) หรือแบบทดสอบโครงสร้างต้นไม้ โดยแบ่งออกเป็นรูปแบบย่อย ๆ 5 รูปแบบ คือ รูปปิรามิดแบบขนาดขั้นคงที่ (Constant Step Size Pyramid) รูปปิรามิดแบบขนาดขั้นผันแปร (Variable Step Size Pyramid) รูปปิรามิดแบบขั้วตัด (Truncated Pyramid) รูปปิรามิดแบบมีหลายข้อในแต่ละขั้น (Multiple – Item Pyramid) และ รูปปิรามิดแบบให้น้ำหนักแก่ตัวเลือกข้อสอบเพื่อแยกทาง (Pyramid Using Differential Response Option Branching)

2. แบบทดสอบเฟล็กซิเลเวล (Flexilevel Test) เป็นแบบทดสอบที่ดัดแปลงมาจากรูปแบบปิรามิด ซึ่งแตกต่างตรงที่รูปแบบปิรามิดมีข้อสอบมากกว่าหนึ่งข้อในแต่ละระดับความยาก แต่แบบทดสอบเฟล็กซิเลเวลจะมีข้อสอบเพียงข้อเดียวในแต่ละระดับความยาก ดังนั้น จึงทำให้การแยกทางของผู้ตอบมีความแตกต่างกัน

3. แบบทดสอบปรับระดับขั้น (Stradaptive Test) เป็นแบบทดสอบที่นำข้อสอบมาจัดกลุ่ม (Stratified) ตามค่าความยากซึ่งแบ่งออกเป็นระดับขั้น (Stratum) ในแต่ละระดับขั้นจะประกอบด้วยข้อสอบที่มีค่าความยากใกล้เคียงกันและจะเรียงลำดับขั้นตามค่าความยากของข้อสอบที่สูงขึ้น สำหรับการดำเนินการสอบจะใช้เครื่องคอมพิวเตอร์

2.2.2.2 รูปแบบแยกทางแปรผัน (Variable – Branching Model) เป็นรูปแบบที่ไม่มีการกำหนดข้อสอบและเส้นทางในการตอบข้อสอบไว้ล่วงหน้า และจะดำเนินการสอบโดยใช้เครื่องคอมพิวเตอร์ ซึ่งเครื่องคอมพิวเตอร์จะคัดเลือกข้อสอบ จากกลุ่มข้อสอบที่มีค่าความยากของข้อสอบใกล้เคียงกับระดับความสามารถของผู้สอบ โดยใช้ข้อมูลจากผลการตอบในแต่ละข้อ การทดสอบจะยุติเมื่อการประมาณค่าความสามารถของผู้สอบมีค่าคงที่ และ

ให้ค่าความคลาดเคลื่อนมาตรฐานเท่ากับเกณฑ์ที่ตั้งไว้ สำหรับรูปแบบแยกทางแปรผันที่ใช้กัน ได้แก่ ยุทธวิธีความเป็นไปได้สูงสุด (Maximum Likelihood Strategies) และยุทธวิธีของเบส์ (Bayesian Strategies)

ในการศึกษาครั้งนี้ ผู้วิจัยได้นำแบบทดสอบดั้งเดิม มาทำการศึกษาร่วมกับแบบทดสอบรูปปิรามิด ซึ่งเป็นรูปแบบหนึ่งของการทดสอบแบบเทเลอร์ ดังนั้นผู้วิจัยจึงขอเสนอรายละเอียดเฉพาะแบบทดสอบรูปปิรามิดเท่านั้น

2.3 วิธีการให้คะแนนของการทดสอบแบบเทเลอร์

การทดสอบแบบเทเลอร์ ผู้สอบแต่ละคนจะได้ทำข้อสอบที่มีความแตกต่างกันทั้ง จำนวนข้อสอบ ค่าพารามิเตอร์ของข้อสอบ และเวลาในการดำเนินการสอบ การให้คะแนนเพื่อประมาณค่าความสามารถของผู้สอบโดยการนับจำนวนข้อสอบที่ตอบถูก ซึ่งเป็นวิธีการที่เราใช้กันทั่ว ๆ ไปนั้น เป็นวิธีการที่ไม่เหมาะสม ดังนั้นวิธีการให้คะแนนของการทดสอบแบบเทเลอร์ จึงต้องใช้วิธีใหม่ โดยการนำทฤษฎีการตอบสนองข้อสอบมาประยุกต์ใช้ในการประมาณค่าความสามารถของผู้สอบ เนื่องจากการให้คะแนนของการทดสอบแบบเทเลอร์มีพื้นฐานมาจากค่าความยากของข้อสอบที่ผู้สอบแต่ละคนตอบ ซึ่งสามารถจัดให้อยู่บนมาตรวัดเดียวกันได้ (Hambleton and Swaminathan. 1985 : 296) ถึงแม้ว่าผู้สอบจะตอบข้อสอบที่แตกต่างกันก็ตาม สำหรับวิธีการให้คะแนนที่สำคัญมีดังนี้ (Weiss. 1974 : 7)

2.3.1 ให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ตอบถูก (The most difficulty of item answered correctly) เป็นการให้คะแนนโดยดูจากข้อสอบในแบบทดสอบที่ผู้สอบทำมาทั้งหมด ข้อใดมีค่าความยากสูงสุดที่ทำถูก แสดงว่าเป็นความสามารถของผู้สอบ

2.3.2 ให้คะแนนตามค่าเฉลี่ยความยากของข้อสอบที่ตอบถูก (Average Difficulty of All Item Answered Correctly) เป็นการให้คะแนนโดยการนำเอาค่าความยากของข้อสอบในแบบทดสอบที่ผู้สอบตอบถูกมาหาค่าเฉลี่ย ค่าเฉลี่ยความยากที่คำนวณได้ แสดงว่าเป็นค่าความสามารถของผู้สอบ

2.3.3 ให้คะแนนตามค่าเฉลี่ยความยากของทุกข้อที่ตอบ (Average Difficulty) เป็นการให้คะแนนที่นำเอาค่าความยากของข้อสอบทุกข้อที่ผู้สอบได้ทำมาหาค่าเฉลี่ย โดยไม่คำนึงว่าผู้สอบจะตอบถูกหรือผิด ค่าเฉลี่ยความยากที่คำนวณได้ แสดงว่าเป็นค่าความสามารถของผู้สอบ

2.3.4 ให้คะแนนตามค่าความยากของข้อสุดท้ายที่ตอบ (Difficulty of the final item) ซึ่งจะไม่คำนึงว่าข้อสุดท้ายจะตอบถูกหรือผิด ค่าความยากของข้อสอบสุดท้ายที่ผู้สอบได้ทำ แสดงว่าเป็นค่าความสามารถของผู้สอบ

2.3.5 ให้คะแนนด้วยวิธีความเป็นไปได้สูงสุด (Maximum Likelihood Method)

เป็นการประมาณค่าความสามารถโดยการนำเอาคะแนนที่ผู้สอบทำได้ทั้งหมดจากแบบทดสอบ ไปประมาณค่าโดยใช้เครื่องคอมพิวเตอร์ ซึ่งมีหลายวิธีย่อย ๆ เช่น วิธีความเป็นไปได้สูงสุด แบบมีเงื่อนไข วิธีความเป็นไปได้สูงสุดแบบไม่มีเงื่อนไข เป็นต้น แต่การประมาณค่าด้วยวิธีนี้มีข้อจำกัดในกรณีที่ผู้สอบทำข้อสอบถูกหมดหรือผิดหมด จะไม่สามารถประมาณค่าได้

2.3.6 ให้คะแนนด้วยวิธีของเบส์ (Bayesian Method) เป็นการประมาณค่าความสามารถที่พัฒนาขึ้นมาเพื่อแก้ปัญหาเกี่ยวกับข้อจำกัดของการประมาณค่าด้วยวิธีความเป็นไปได้สูงสุด ซึ่งจะทำให้มีความถูกต้องแม่นยำมากกว่า แต่วิธีนี้มีการใช้สถิติค่อนข้างซับซ้อน จึงทำให้ไม่มีการนำมาใช้กันอย่างแพร่หลายมากนัก

2.4 ชนิดของแบบทดสอบรูปปิรามิด

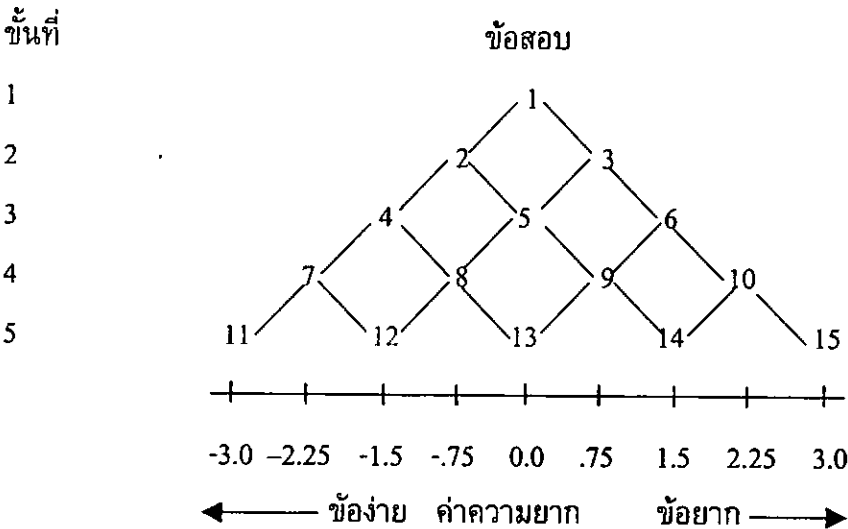
แบบทดสอบรูปปิรามิด (Pyramidal Test) สามารถจัดรูปแบบของแบบทดสอบได้ โดยอาศัยโครงสร้างของแบบจำลองรูปปิรามิด (Pyramidal Models) หรือแบบจำลองโครงสร้างต้นไม้ (Tree Structure Models) ซึ่งเป็นรูปแบบหนึ่งของแบบทดสอบเทเลอร์ (Tailored Test) ประเภทวิธีการหลายขั้นที่มีแบบจำลองการแยกทางคงที่ แบบทดสอบรูปปิรามิดเป็นแบบจำลองแบบแรกของการทดสอบเทเลอร์ที่ได้มีการศึกษาและก่อให้เกิดการวิจัยอย่างกว้างขวางในปัจจุบัน งานที่วิจัยเกี่ยวกับแบบจำลองรูปปิรามิดเริ่มต้นจากการศึกษาของ แครธว็ทท์ และ ฮายเซอร์ (Kratwohl and Huyser : 1956) ในระยะต่อมาก็มีรายงานการศึกษาของ เบย์รอฟฟ์ และคณะ (Bayroff and others. :1960,1967) แพทเทอร์สัน (Paterson : 1962) แฮนด์เซน (Hansen : 1969) และ ลอร์ด (Lord : 1970,1971) ผลการศึกษาของงานวิจัยที่ผ่านมาเหล่านี้ส่วนใหญ่สอดคล้องกับผลงานวิจัยของ มุสซิโอ (Mussio : 1972) และ ลาร์กิน และ ไวท์ (Larkin and Weiss : 1974) (ระจิต ตริพุทธรัตน์ และคนอื่น. 2535, 24 - 25.)

2.4.1 รูปแบบของแบบทดสอบรูปปิรามิด

ในปัจจุบันแบบทดสอบรูปปิรามิดได้มีการศึกษากันใน 5 รูปแบบ คือ รูปแบบที่ใช้ขนาดขั้นคงที่ รูปแบบที่ใช้ขนาดขั้นแปรผัน รูปปิรามิดข้างตัด รูปจำลองที่มีข้อสอบหลายข้อในแต่ละขั้น และรูปแบบการให้น้ำหนักตัวเลือกของข้อสอบเพื่อการแยกทาง Weiss (1974) ได้สรุปทดสอบปิรามิดทั้ง 5 ชนิดดังกล่าวไว้

1. แบบทดสอบรูปปิรามิดที่ใช้ขนาดขั้นคงที่ (Constant Step Size Pyramid)

ได้แก่ การทดสอบเทเลอร์ที่จัดเป็นรูปสามเหลี่ยม โดยมีจำนวนข้อสอบในแต่ละขั้นเท่ากับลำดับที่ของขั้น กล่าวคือ ขั้นที่หนึ่งจะมีข้อสอบหนึ่งข้อ ขั้นที่สองจะมีข้อสอบสองข้อ ในขั้นที่ห้าจะมีข้อสอบห้าข้อ ดังนั้นแบบทดสอบที่มีห้าขั้น จะมีข้อสอบทั้งหมด 15 ข้อ ดังภาพประกอบ 1

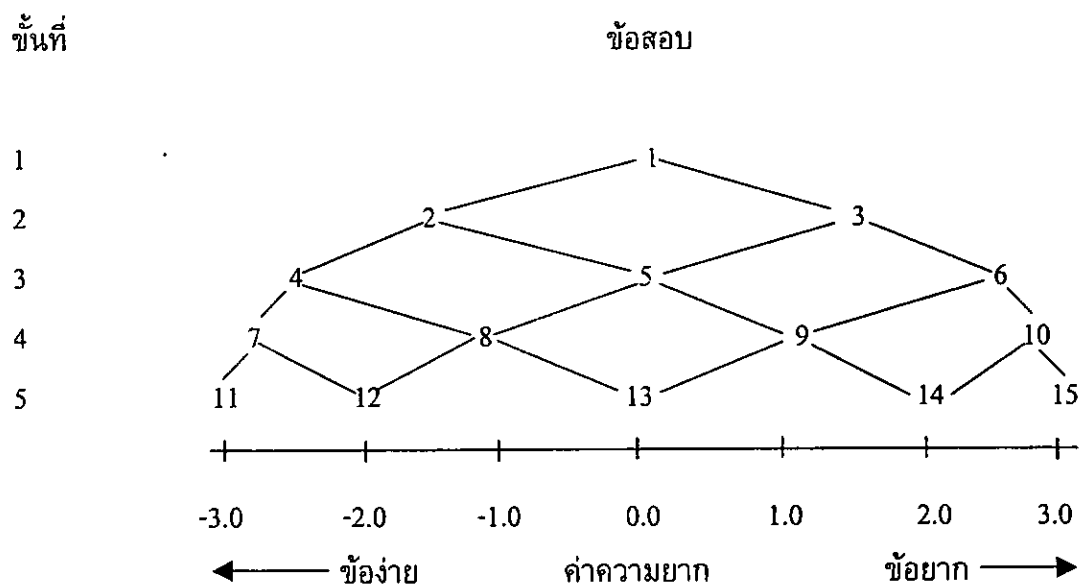


ภาพประกอบ 1 แสดงโครงสร้างของการทดสอบเทเลอรูปิรามิดแบบขนาดขั้นคงที่ ที่มี 5 ชั้น

ภาพประกอบ 1 แกนนอนแสดงถึงค่าความยากของข้อสอบ (b) ที่สัมพันธ์กับข้อสอบในการทดสอบเทเลอรูปิรามิด ซึ่งจะเห็นว่าความยากมีค่าอยู่ระหว่าง -3.0 ถึง 3.0 โดยแบ่งออกเป็น 8 ช่วงเท่า ๆ กัน ฉะนั้นค่าความยากจะห่างกันช่วงละ 0.75 ข้อสอบที่อยู่ในแนวตั้งแนวเดียวกันจะมีค่าความยากเท่ากัน และช่วงห่างค่าความยากของข้อที่อยู่ติดกันภายในชั้นมีค่าเท่ากันตลอด คือ 0.75

ในการทำแบบทดสอบ ผู้สอบจะต้องทำข้อสอบชั้นละหนึ่งข้อ โดยเริ่มทำข้อที่อยู่บนยอดสามเหลี่ยมซึ่งมีความยากปานกลางก่อน ในรูปได้แก่ข้อ 1 มีค่าความยากเท่ากับ 0 เมื่อทำในชั้นที่ 1 เสร็จจะไปทำข้อสอบในชั้นต่อไป ซึ่งผู้สอบจะทำข้อสอบชั้นละ 1 ข้อ ถ้าทำถูกชั้นต่อไปจะแยกไปทำข้อที่ยากขึ้น แต่ถ้าทำผิดชั้นต่อไปจะไปทำข้อที่ง่ายขึ้น กระบวนการในการทำข้อสอบจะเป็นเช่นนี้จนถึงขั้นสุดท้าย

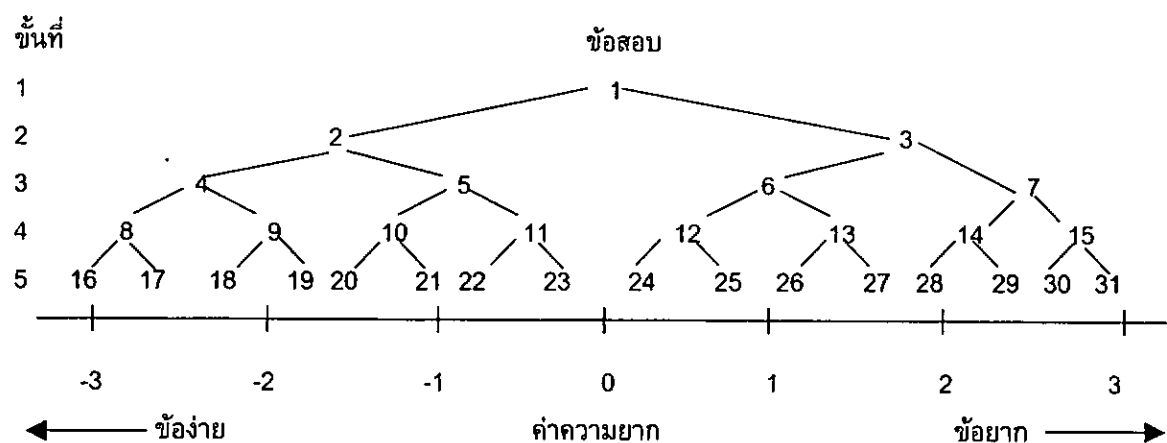
2. แบบทดสอบรูปปิรามิดที่ใช้ขนาดขั้นแปรผัน (Variable Step Size Pyramid) โดยที่แพทเตอร์สัน และลอร์ด (Weiss, 1974 : 18 ; Citing Lord, 1971) มีความเห็นว่า รูปปิรามิดแบบขนาดขั้นคงที่ ขาดประสิทธิภาพเกี่ยวกับความไวในการกำหนดเส้นทางการตอบข้อสอบ จึงได้กำหนดช่วงห่างค่าความยากของข้อที่อยู่ติดกันภายในชั้นให้มีค่าไม่เท่ากันโดยให้ข้อที่อยู่กลาง ๆ มีช่วงห่างของค่าความยากมาก แล้วค่อย ๆ ลดลงในข้อที่ค่อนข้างง่ายหรือยาก แพทเตอร์สัน ได้กำหนดการเพิ่มหรือลดนั้น จะเพิ่มหรือลดไปยังตำแหน่งกึ่งกลางระหว่าง 2 ข้อ ในชั้นที่มาก่อน หรือระหว่างข้อในชั้นที่มาก่อนกับค่าความยากสูงสุดหรือค่าต่ำสุด ดังภาพประกอบ 2



ภาพประกอบ 2 แสดงโครงสร้างของการทดสอบเทเลอร์รูปปิรามิดแบบขนาดขั้นแปดขั้นที่มี 5 ชั้น

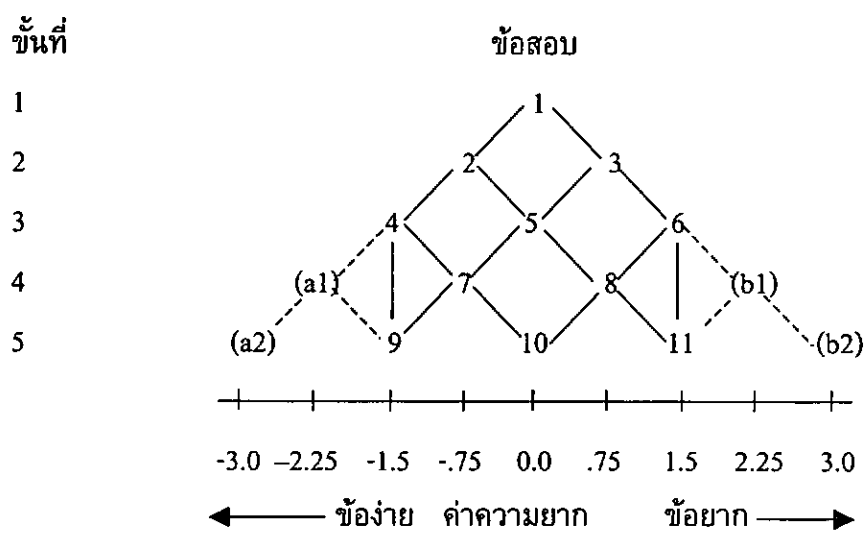
ภาพประกอบ 2 ในขั้นที่ 1 มีข้อสอบเพียงข้อเดียว คือ ข้อ 1 ซึ่งมีความยากเท่ากับ 0.00 ในขั้นที่ 2 มีข้อสอบสองข้อ คือข้อ 2 มีความยากอยู่กึ่งกลางระหว่าง -3.00 (ความยากต่ำสุด) กับ 0.00 (ความยากของข้อที่ 1) ซึ่งก็คือ -1.50 ข้อ 3 มีความยากอยู่ระหว่าง 3.00 (ความยากสูงสุด) กับ 0.00 (ความยากของข้อที่ 1) ซึ่งก็คือ 1.50 ในขั้นที่ 3 มีข้อสอบสามข้อ คือข้อ 4 มีความยากอยู่กึ่งกลางระหว่าง -3.00 (ความยากต่ำสุด) กับ -1.50 (ความยากของข้อที่ 2) ซึ่งก็คือ -2.250 ข้อ 5 มีความยากเท่ากับ 0.00 ข้อ 6 มีความยากอยู่กึ่งกลางระหว่าง 1.50 (ความยากของข้อที่ 3) กับ 3.00 (ความยากสูงสุด) ซึ่งก็คือ 2.250 ในขั้นที่ 4 มีข้อสอบสี่ข้อ คือ ข้อ 7 มีความยากอยู่กึ่งกลางระหว่าง -3.00 (ความยากต่ำสุด) กับ -2.250 (ความยากของข้อที่ 4) ซึ่งก็คือ -2.625 ข้อ 8 มีความยากอยู่กึ่งกลางระหว่าง -2.250 (ความยากของข้อที่ 4) กับ 0.00 (ความยากของข้อที่ 5) ซึ่งก็คือ -1.125 ข้อ 9 มีความยากอยู่กึ่งกลางระหว่าง 0.00 (ความยากของข้อที่ 5) กับ 2.250 (ความยากของข้อที่ 6) ซึ่งก็คือ 1.125 และข้อ 10 มีความยากอยู่กึ่งกลางระหว่าง 2.250 (ความยากของข้อที่ 6) กับ 3.00 (ความยากสูงสุด) ซึ่งก็คือ 2.625

จากรูปปิรามิดแบบขนาดขั้นแปดขั้น (Variable Step Size Pyramid) นี้ต่อมา ลอร์ด (Weiss, 1974 : 20 ; Citing Lord, 1971) ได้ใช้วิธีการเพิ่มจำนวนข้อในแต่ละขั้นให้เป็นสองเท่าของจำนวนข้อสอบในขั้นที่มาก่อน ซึ่งให้ชื่อรูปแบบใหม่นี้ว่า แบบรอบบิน-มอนโร (Robin-Monro) ดังภาพประกอบ 3



ภาพประกอบ 3 แสดงโครงสร้างของการทดสอบทีเลอร์รูปปิรามิดแบบรอบบิน-มอนโร (Robin-Monro)

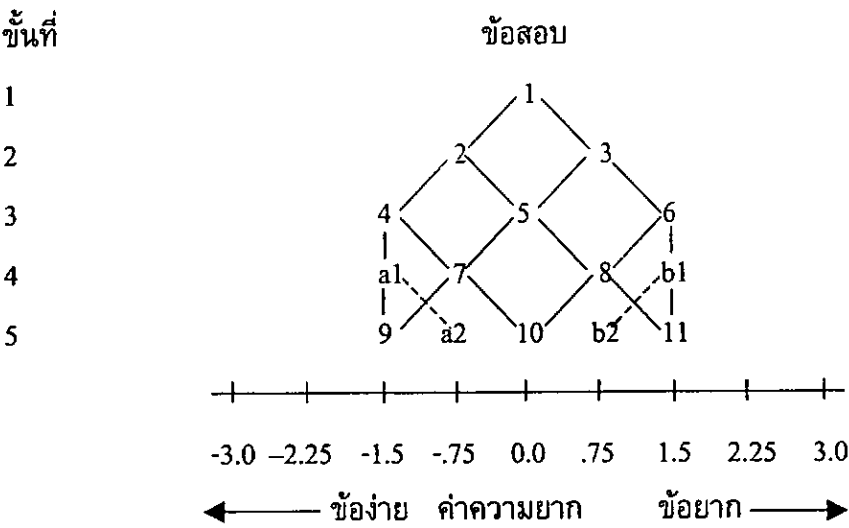
3. รูปปิรามิดแบบข้างตัด (Truncated Pyramid) เป็นรูปแบบที่ มุสลีโอ (Weiss. 1974 : 22 ; Citing Mussio. 1973) เป็นผู้เสนอรูปแบบนี้ขึ้นมา เพื่อให้ลดจำนวนข้อสอบลงจากรูปปิรามิดขนาดชั้นคงที่ จึงใช้วิธีสกัดการสะท้อนกลับ หรือวิธีรักษาการสะท้อนกลับ (Reflecting or Retaining Barrier) ดังภาพประกอบ 4



ภาพประกอบ 4 แสดงโครงสร้างของการทดสอบทีเลอร์รูปปิรามิดแบบข้างตัด (Truncated Pyramid) สกัดการสะท้อนกลับ (Reflecting Barrier)

ภาพประกอบ 4 การตอบข้อสอบใน ชั้นที่ 1 และ ชั้นที่ 2 มีวิธีการตอบเช่นเดียวกับรูปปิรามิดแบบขนาดชั้นคงที่ จนถึงจุดสกัดในชั้นที่ 3 ถ้าตอบข้อสอบข้อที่ 4 ถูก ก็จะแยกไป

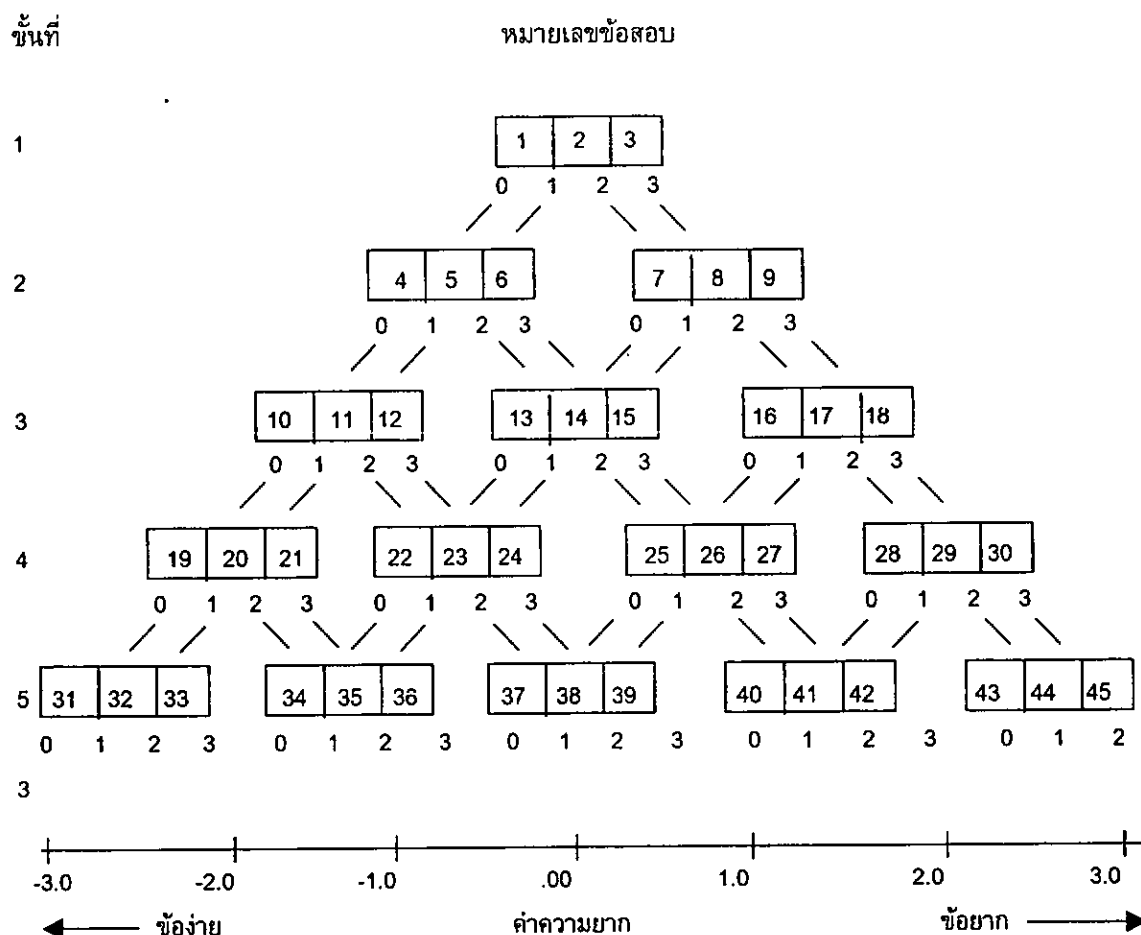
ตอบข้อ 7 แต่ถ้าตอบข้อ 4 ผิดก็จะลงไปทำข้อ 9 และในทำนองเดียวกัน ถ้าตอบข้อ 6 ผิดก็จะแยกไปตอบข้อ 8 แต่ถ้าตอบข้อ 6 ถูกก็จะไปตอบข้อ 11 เลย



ภาพประกอบ 5 แสดงโครงสร้างของการทดสอบเทเลอรูปิรามิดแบบข้างตัด (Truncated Pyramid) รักษาการสะท้อนกลับ (Retaining Barrier)

ภาพประกอบ 5 การตอบข้อสอบใน ชั้นที่ 1 และ ชั้นที่ 2 วิธีตอบเช่นเดียวกับรูปปิรามิดแบบขนาดชั้นคงที่ จนถึงจุดสกัดในชั้นที่ 3 ถ้าตอบข้อสอบข้อที่ 4 ถูกก็จะแยกไปตอบข้อ 7 แต่ถ้าตอบข้อ 4 ผิดก็จะลงไปตอบข้อ a1 ซึ่งมีความยากเท่ากับข้อ 4 และถ้าตอบข้อ a1 ผิดอีกก็จะลงไปตอบข้อ 11 แต่ถ้าตอบข้อ a1 ถูก ก็จะแยกไปตอบข้อ a2 ซึ่งมีความยากเท่ากับข้อ 7

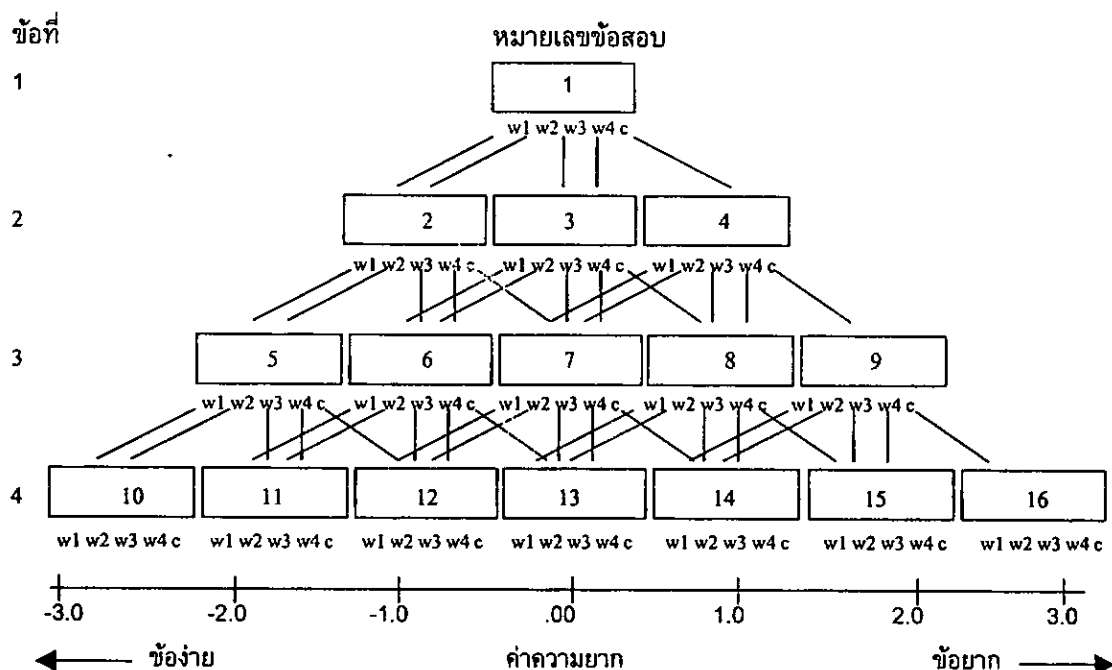
4. รูปปิรามิดแบบมีหลายข้อในแต่ละชั้น (Multiple-Item Pyramid) เป็นรูปแบบที่ แครธไวท์กับฮายเซอร์ และ ลินน์ (Weiss. 1974 : 25; Citing Krathworth and Huyser. 1956 ; Citing Linn. 1969) เป็นผู้เสนอรูปแบบนี้ขึ้นมา โดยให้มีจำนวนข้อในแต่ละชั้นมีมากกว่าหนึ่งข้อ ทั้งนี้เพื่อให้การตอบเพื่อแยกทางได้ถูกต้องมากยิ่งขึ้น ดังภาพประกอบ



ภาพประกอบ 6 แสดงโครงสร้างของการทดสอบทฤษฎีรูปร่างปริมาตร แบบมีข้อสอบหลายข้อในแต่ละ
 ชั้น (Multiple-Item Model)

ภาพประกอบ 6 ในขั้นที่ 1 ถ้าผู้สอบตอบข้อสอบถูก 0 ถึง 1 ข้อ ในขั้นที่ 2 จะแยกไปทำข้อสอบข้อที่ 4, 5 และ 6 และถ้าตอบข้อสอบถูก 2 ถึง 3 ข้อ ก็จะแยกไปทำข้อสอบข้อที่ 7, 8 และ 9 สำหรับในขั้นต่อไปทำเช่นเดียวกัน

5. รูปปริมาตรแบบให้น้ำหนักแต่ตัวเลือกของข้อสอบเพื่อแยกทาง (Differential Response Option Branching) เป็นรูปแบบที่ เบย์รอฟฟ์และซีเลย์ (Weiss. 1974 : 28 ; Citing Bayroff and Seelay. 1967) เป็นผู้เสนอขึ้น ซึ่งมีรูปแบบคล้ายกับแบบที่ใช้ข้อสอบหลาย ๆ ข้อในแต่ละชั้น กล่าวคือ แทนที่จะใช้ข้อสอบหลาย ๆ ข้อในแต่ละชั้น ก็ใช้ตัวเลือกของข้อสอบแทน รูปแบบนี้เหมาะสำหรับข้อสอบเลือกตอบ ที่สามารถกำหนดน้ำหนักคะแนนของตัวเลือกได้ ดังตัวอย่าง ภาพประกอบ



ภาพประกอบ 7 แสดงโครงสร้างของการทดสอบเทเลอรัปรีพริมาติดแบบให้น้ำหนักแต่ตัวเลือก
ของข้อสอบเพื่อแยกทาง

แบบทดสอบรูปปริมาตรูปปริมาตรแบบให้น้ำหนักแต่ตัวเลือกของข้อสอบเพื่อแยกทาง จะใช้สารสนเทศทั้งหมดจากคำตอบของผู้สอบที่ตอบข้อสอบเพื่อกำหนดการแยกทางให้ทำข้อสอบข้อต่อไปในขั้นถัดไป ซึ่งจะใช้ระดับของการตอบผิด (Degree of Wrongness) ของคำตอบที่ผิดในแต่ละข้อเป็นเกณฑ์ในการแยกทาง รูปแบบการแยกทางลักษณะนี้สามารถใช้ได้กับข้อสอบที่เป็นแบบเลือกตอบ และข้อสอบที่ให้ตอบอย่างอิสระ (Free Response Items) ที่สามารถกำหนดค่าความยากของแต่ละคำตอบได้ วิธีการนี้เสนอโดย Bayroff (Citing Weiss, 1974) ซึ่ง Bayroff ได้ศึกษาโดยใช้วิธีนี้ในขั้นที่ 1 ของแบบทดสอบชนิด 5 ขั้นเท่านั้นจากภาพประกอบ 7 แสดงรูปแบบหนึ่งของแบบทดสอบรูปปริมาตรูปปริมาตรแบบให้น้ำหนักแต่ตัวเลือกของข้อสอบเพื่อการแยกทาง ในขั้นที่ 1 จะมีข้อสอบ 1 ข้อ ขั้นที่ 2 จะมีข้อสอบ 3 ข้อ ขั้นที่ 3 มีข้อสอบ 5 ข้อ และในขั้นที่ 4 มีข้อสอบ 7 ข้อ จะเห็นได้ว่าจำนวนข้อสอบในขั้นถัดไปจะเพิ่มขึ้นจากขั้นเดิม 2 ข้อ ข้อสอบที่แสดงในภาพประกอบ 7 เป็นข้อสอบแบบเลือกตอบชนิด 5 ตัวเลือก มีคำตอบที่ถูกอยู่ 1 ตัวเลือก (C) และคำตอบที่ตอบผิดอีก 4 ตัวเลือก (W_1, W_2, W_3 และ W_4) ผิดน้อยที่สุดหรือเกือบจะถูกต้อง แทนด้วย W_4 โดยถือเอาคำตอบของกลุ่มคนปกติในการกำหนดน้ำหนักของคำตอบ

ในขั้นที่ 1 เมื่อผู้สอบตอบข้อสอบหมายเลข 1 ถูกหรือเลือกตัวเลือก C จะถูกแยกทางให้ไปตอบข้อสอบหมายเลข 4 ในขั้นที่ 2 ซึ่งเป็นข้อสอบที่มีความยากมากที่สุดในขั้นที่ 2 ถ้าตอบตัวเลือกที่มีระดับความผิด W_1 หรือ W_2 จะถูกแยกทางให้ไปตอบข้อสอบหมายเลข 2 ใน

ขั้นที่ 2 ซึ่งมีความยากน้อยกว่าข้อสอบหมายเลข 1 ในขั้นที่ 1 ส่วนผู้สอบที่เลือกตอบตัวเลือก W_3 หรือ W_4 ซึ่งมีระดับความผิดน้อยกว่า W_1 และ W_2 แต่ยังไม่ใช่คำตอบที่ถูกต้อง จะถูกแยกทางให้ไปตอบข้อสอบหมายเลข 3 ในขั้นที่ 2 ซึ่งมีความยากเท่ากับข้อสอบหมายเลข 1 ในขั้นที่ 1 ทั้งนี้เพราะเหตุผลว่า ผู้ที่จะเลือกตัวเลือก W_3 และ W_4 ส่วนใหญ่จะมีความสามารถในระดับเฉลี่ยของกลุ่มคนปกติ ส่วนผู้ที่มีระดับความสามารถที่สูงหรือต่ำกว่านี้จะถูกแยกทางให้ไปตอบข้อสอบที่มีความยากสูงหรือต่ำกว่า และจะดำเนินการเช่นนี้จนถึงขั้นสุดท้ายของแบบทดสอบลักษณะนี้

แบบทดสอบที่ให้ให้นักตัวเลือกของข้อสอบนี้ จะให้สารสนเทศของผู้สอบจากคำตอบที่เลือกได้มากกว่าแบบทดสอบรูปปิรามิดที่ให้คำตอบถูกเป็น 1 และผิดเป็น 0 ทั้งนี้เพราะการเลือกตอบในแต่ละตัวเลือกที่มีระดับความไม่ถูกต้องแตกต่างกันนี้ จะทำให้ทราบว่าผู้ตอบเข้าใจผิดในลักษณะใดในแต่ละเนื้อหาที่วัดของข้อสอบแต่ละข้อ ซึ่งจะเป็นประโยชน์มากสำหรับการสร้างแบบทดสอบเพื่อการวินิจฉัยข้อบกพร่องในการเรียน อย่างไรก็ตามแบบทดสอบรูปปิรามิดที่ให้นักตัวเลือกของข้อสอบเพื่อการแยกทางที่แสดงในภาพประกอบ 7 เป็นเพียงรูปแบบหนึ่งเท่านั้น ซึ่งถ้าจะให้การแยกทางตามให้นักตัวเลือกที่สมบูรณ์แล้วควรจะให้แยกทางไปทำข้อสอบข้อที่แตกต่างกันในแต่ละตัวเลือกที่ผิด คือถ้าผู้สอบที่ตอบตัวเลือกที่มีระดับความผิดเป็น W_1 จะได้ตอบข้อสอบที่มีความยากน้อยกว่าข้อสอบที่ผู้เลือกตอบตัวเลือก W_2 ในขั้นถัดไป วิธีการนี้เมื่อใช้กับข้อสอบแบบเลือกตอบชนิด 5 ตัวเลือก การตอบข้อสอบในขั้นหนึ่ง ๆ ก็จะมีทางเลือกในการแยกทางเพื่อทำข้อสอบในขั้นถัดไป 5 ทาง ซึ่งมีผลทำให้โครงสร้างแบบทดสอบไม่สมมาตรโดยมีทางเลือกที่มีโอกาสไปทางซ้ายมากกว่าทางขวา และจำเป็นต้องมีกลุ่มข้อสอบขนาดใหญ่ขึ้น ซึ่งเป็นข้อสอบที่ทำให้สร้างได้ยาก แต่อย่างไรก็ตามแบบทดสอบรูปปิรามิดรูปแบบนี้มีความถูกต้องในการวัดมากขึ้น จำนวนข้อสอบที่ผู้สอบแต่ละคนทำมีจำนวนคงที่ และผลการตอบข้อสอบแต่ละข้อของผู้สอบแต่ละคน จะให้สารสนเทศเกี่ยวกับผู้สอบมากในด้านการวินิจฉัยข้อบกพร่องในการเรียนรู้ แบบทดสอบรูปปิรามิดรูปแบบนี้จึงได้รับความสนใจอย่างกว้างขวาง

2.4.2 วิธีการให้คะแนนแบบทดสอบรูปปิรามิด

แบบทดสอบรูปปิรามิดทั้ง 5 รูปแบบที่กล่าวมานั้น สามารถให้คะแนนวิธีเดียวกันหรือต่างกันได้ การให้คะแนนมีอยู่หลายวิธี ผลการให้คะแนนแต่ละวิธีขึ้นอยู่กับแนวทางในการแยกทางเพื่อทำข้อสอบในขั้นถัดไปและค่าความยากของข้อสอบที่ผู้สอบได้ตอบ ดังนั้นจึงมีข้อตกลงว่ากลุ่มข้อสอบที่นำมาสร้างแบบทดสอบรูปปิรามิดจะต้องวัดในเนื้อหาเดียวกัน หรือมิติเดียว และมาตรวัดระดับความยากของข้อสอบกับมาตรวัดระดับความสามารถของผู้สอบเป็นมาตรวัดเดียวกัน การให้คะแนนแบบทดสอบรูปปิรามิดมีอยู่ 7 วิธี ดังต่อไปนี้

1. ให้คะแนนตามค่าความยากของข้อสอบข้อสุดท้ายที่ได้ตอบ (Difficulty of

the Final Item)

การให้คะแนนวิธีนี้มีข้อสมมติว่าเป็นการให้คะแนนตามความสามารถในการสอบที่ปฏิบัติได้สูงสุด (Maximum Performance) คะแนนที่ให้จะเชื่อถือได้น้อยลงเมื่อมีการตอบข้อสอบโดยการเดา และถ้าจำนวนชั้นในแบบทดสอบมีน้อยคะแนนที่ให้ก็จะน้อยเกินไป มีผลให้ค่าสหสัมพันธ์ระหว่างคะแนนแบบทดสอบรูปปรางค์ที่ให้โดยวิธีนี้กับแบบทดสอบอื่น ๆ มีค่าต่ำ กล่าวคือความตรงเชิงสัมพันธ์กับเกณฑ์มีค่าต่ำดังนั้นในทางปฏิบัติแล้วการให้คะแนนวิธีนี้จะให้ประโยชน์น้อย

2. ให้คะแนนตามค่าความยากของข้อสอบที่ถัดจากข้อสุดท้ายที่ได้ตอบ

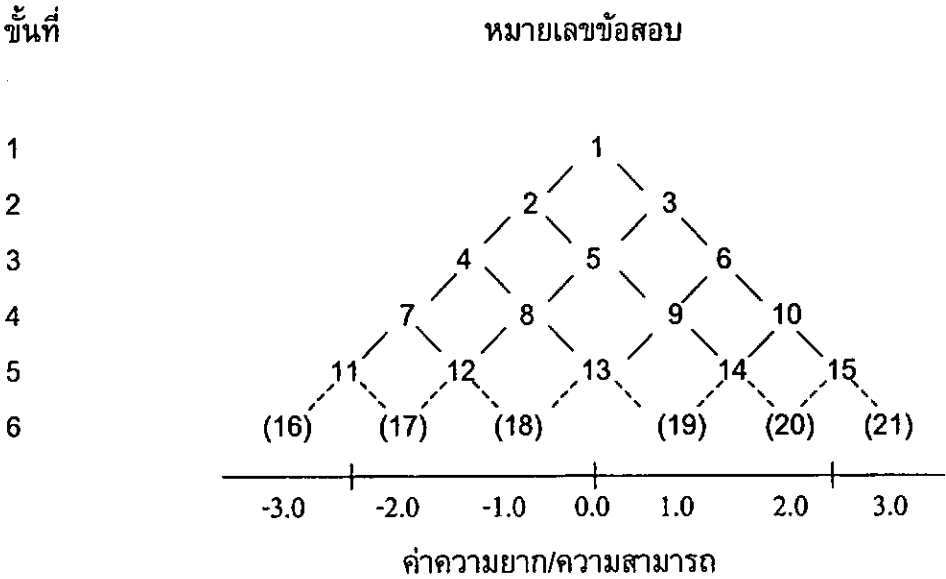
(Difficulty of The $N + 1^{\text{th}}$ Item)

วิธีการให้คะแนนตามค่าความยากของข้อสอบข้อสุดท้ายที่ได้ตอบนั้น ไม่ได้พิจารณาถึงคำตอบข้อสุดท้ายว่าถูกหรือผิด คะแนนของผู้สอบที่ทำข้อสุดท้ายข้อเดียวกันจึงเท่ากันไม่ว่าจะตอบถูกหรือผิด การให้คะแนนตามค่าความยากของข้อสอบที่ถัดจากข้อสุดท้ายที่ได้ตอบ ($N + 1^{\text{th}}$) ซึ่งสมมติว่าเป็นข้อสอบที่ควรจะได้ทำต่อไปจากผลการตอบข้อสอบข้อสุดท้าย การให้คะแนนโดยวิธีนี้จึงมีข้อสมมติว่ามีข้อสอบชั้นสุดท้ายที่ต่อจากชั้นที่ N ของแบบทดสอบจริง ดังแสดงไว้ในภาพประกอบที่ 8 ซึ่งเป็นแบบทดสอบรูปปรางค์ชนิด 5 ชั้น

ภาพประกอบที่ 8 ในชั้นที่ 5 จะมีข้อสอบหมายเลข 11 ถึง 15 และมีข้อสอบหมายเลข 16 ถึง 21 ในชั้นที่ 6 ซึ่งเป็นข้อสอบที่สมมติขึ้นสำหรับพิจารณาให้คะแนนตามผลการตอบข้อสอบข้อสุดท้ายเช่น ถ้าผู้สอบทำข้อสอบหมายเลข 12 ซึ่งมีค่าความยากเท่ากับ -1.5 เป็นข้อสุดท้าย ถ้าให้คะแนนตามค่าความยากของข้อสอบข้อสุดท้ายผู้สอบคนนี้ก็จะได้คะแนน -1.5 ไม่ว่าจะตอบข้อนี้ถูกหรือผิดก็ตาม แต่ถ้าพิจารณาให้คะแนนตามค่าความยากของผู้สอบที่ถัดจากข้อสุดท้ายที่ได้ตอบนั้น จะต้องพิจารณาว่าผลการตอบข้อสอบหมายเลข 12 นั้นถูกหรือผิด ถ้าตอบถูกก็จะได้คะแนนเท่ากับ -1.0 ซึ่งเป็นค่าความยากของข้อสอบหมายเลข 18 แต่ถ้าตอบผิดก็จะได้คะแนนเท่ากับ -2.0 ซึ่งเป็นค่าความยากของข้อสอบ 17 เป็นต้น การให้คะแนนวิธีนี้จะให้ช่วงของคะแนนแคบกว่าแบบทดสอบรูปปรางค์นั้นมีจำนวนชั้นน้อย

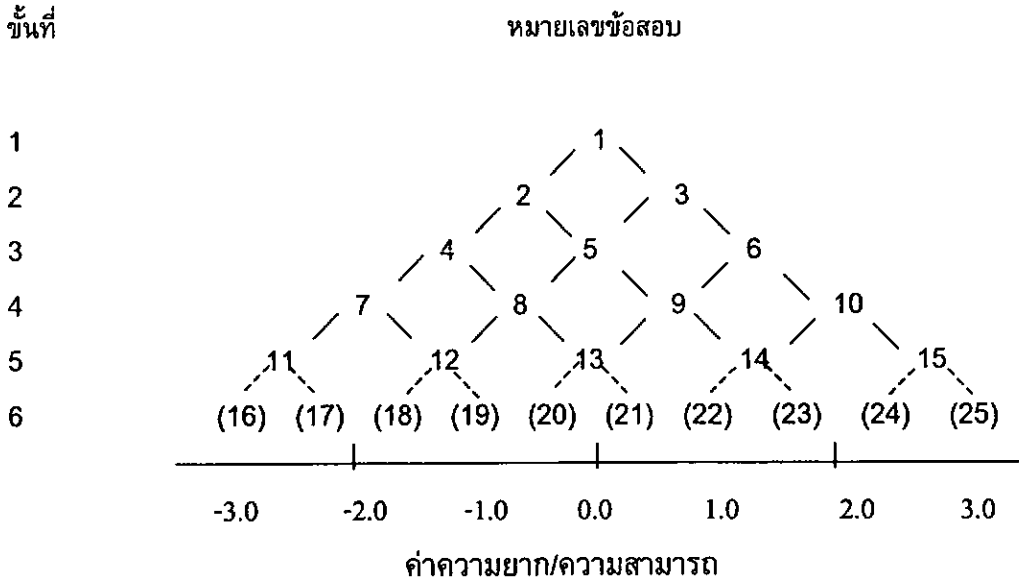
a

ให้คะแนนตามค่าความยากของข้อสอบ
ที่จัดจากข้อสุดท้ายที่ได้ตอบ



b

ให้คะแนนตามผลการตอบข้อสอบข้อสุดท้าย



ภาพประกอบ 8 แสดงการให้คะแนนแบบทดสอบรูปปิรามิด 2 วิธี ที่ใช้ผลการตอบข้อสอบข้อสุดท้าย

3. ให้คะแนนตามผลการตอบข้อสอบข้อสุดท้าย (Terminal Right-Wrong)

จากวิธีการให้คะแนนตามค่าความยากของข้อสอบข้อที่ถัดจากข้อสอบข้อสุดท้าย ที่ได้ตอบนั้น มีปัญหาอยู่ว่า ผู้สอบที่ทำข้อสอบที่ง่ายกว่าถูกกับผู้สอบที่ทำข้อสอบที่ยากกว่าในชั้นเดียวกันผิดจะได้คะแนนเท่ากันหรือตีความหมายว่ามีระดับความสามารถเท่ากัน ซึ่งวิธีการให้คะแนนตามผลการตอบข้อสุดท้ายถือว่าความสามารถของผู้สอบทั้งสองคนดังกล่าวน่าจะแตกต่างกัน จากภาพประกอบที่ 8 แสดงให้เห็นว่าข้อสอบแต่ละข้อในชั้นสุดท้ายหรือชั้นที่ 5 นั้นสามารถแยกทางไปยังข้อสอบที่สมมติขึ้นในชั้นที่ 6 อีก 12 ข้อ เช่น ข้อสอบหมายเลข 12 จะแยกทางไปยังข้อสอบหมายเลข 18 และ 19 ซึ่งความยากของข้อสอบหมายเลข 18 ($b = -1.8$) จะน้อยกว่าข้อสอบหมายเลข 12 ($b = -1.5$) และข้อสอบหมายเลข 19 ($b = -1.2$) จะมีความยากมากกว่าข้อสอบหมายเลข 12 ดังนั้นการให้คะแนนโดยวิธีนี้ ถ้าผู้สอบตอบข้อสอบหมายเลข 12 ซึ่งเป็นข้อสอบข้อสุดท้ายผิดก็จะได้คะแนนเท่ากับ -1.8 ซึ่งเป็นค่าความยากของข้อสอบหมายเลข 18 แต่ถ้าตอบข้อสอบหมายเลข 12 ถูกก็จะได้คะแนนเท่ากับ -1.2 ซึ่งเป็นค่าความยากของข้อสอบหมายเลข 19

4. ให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ได้ตอบ (Average Difficulty)

Lord (Weiss, 1974) ได้เสนอการให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ได้ตอบขึ้น วิธีการนี้จะนำค่าความยากของข้อสอบทุกข้อที่ผู้สอบได้ตอบโดยไม่คำนึงว่าจะตอบถูกหรือตอบผิดมาหาค่าเฉลี่ยเป็นคะแนน Lord ไม่ได้นำค่าความยากของข้อสอบข้อแรกมาคิดด้วยแต่จะใช้ค่าความยากของข้อสอบข้อที่ถัดจากข้อสุดท้าย ($N + 1^{\text{th}}$ Item) ที่ได้ตอบที่สมมติขึ้นมาคิดแทน วิธีการให้คะแนนโดยวิธีนี้ จะให้สารสนเทศของคะแนนเช่นเดียวกับวิธีการให้คะแนนตามจำนวนข้อสอบที่ตอบถูกในแบบทดสอบรูปปิรามิดที่ใช้ขนาดชั้นคงที่

ถึงแม้ว่าคะแนนที่ได้จากการให้ตามค่าความยากเฉลี่ยของข้อสอบที่ได้ตอบจะสูงกว่าการให้โดยวิธีอื่นบางวิธี เช่น มากกว่าคะแนนที่ให้ตามค่าความยากของข้อสอบข้อสุดท้ายที่ได้ตอบก็ตาม การให้คะแนนโดยวิธีนี้ยังมีข้อจำกัดในการวัดระดับความสามารถของผู้สอบได้ไม่กว้างมากนัก เพราะวิธีการนี้จะให้คะแนนโดยการรวมค่าความยากของข้อสอบที่มีค่าความยากใกล้ ๆ กับความสามารถเฉลี่ยของผู้สอบแล้วนำมาคิดเป็นค่าเฉลี่ยเพื่อให้เป็นคะแนน คะแนนที่ได้จึงไม่สามารถกระจายไปให้ครอบคลุมถึงระดับความสามารถที่สูงและต่ำมาก ๆ ได้ ตัวอย่างจากโครงสร้างแบบทดสอบรูปปิรามิด 5 ชั้นในภาพประกอบ 8 ถ้าผู้สอบที่มีความสามารถสูงสุดตอบข้อสอบถูกหมดทุกข้อคือได้ตอบข้อสอบหมายเลข 1, 3, 6, 10 และ 15 ถูกทั้งหมด ถ้าเป็นการให้คะแนนตามวิธีให้ตามค่าความยากของข้อที่ถัดจากข้อสุดท้ายจะได้คะแนนตามค่าความยากของข้อสอบหมายเลข 21 ในชั้นที่สมมติขึ้นซึ่งจะมีความยากสูงกว่า 3.0 อยู่ 10% หรือกล่าวได้ว่ามีความสามารถสูงกว่า 3.0 แต่การให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ได้ตอบซึ่งจะได้คะแนนเท่ากับค่าเฉลี่ยของความยากของข้อสอบหมายเลข 3, 6, 10, 15 และ 21 มีค่าเท่ากับ 1.8 ซึ่งหมายความว่ามีความสามารถต่ำกว่าระดับความสามารถ 3.0 ดังนั้นจึงสรุปได้ว่า

ถึงแม้ว่าการให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ได้ตอบจะให้ค่าคะแนนสูงกว่าการให้คะแนนโดยวิธีอื่นจะเป็นการพิจารณาจากคำตอบทั้งหมดของผู้สอบซึ่งถือว่าเป็นการประมาณความสามารถของผู้สอบหลายครั้งหรือจากข้อสอบหลายข้อ ซึ่งแตกต่างจากวิธีอื่นที่ประมาณความสามารถของผู้สอบหลายครั้งจากข้อสอบหลายข้อ ซึ่งแตกต่างจากวิธีอื่นที่ประมาณความสามารถของผู้สอบจากข้อสอบเพียงข้อเดียวก็ตาม การให้คะแนนโดยวิธีนี้ก็ยังคงให้ความหมายที่ไม่ดีไปกว่าการให้คะแนนโดยวิธีอื่น ๆ

5. ให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ตอบถูก (Average Difficulty of All Item Answered Correctly of Average difficulty correct)

การให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ตอบถูก เป็นวิธีที่จัดข้อบกพร่องของการให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ได้ตอบ วิธีนี้จะนำเอาค่าความยากของข้อสอบที่ตอบถูกเท่านั้นมาหาค่าเฉลี่ย ซึ่งคะแนนที่ได้จากค่าเฉลี่ยนี้ จะมีความหมายในการบอกระดับความสามารถของผู้สอบได้มากขึ้น คะแนนที่ได้โดยวิธีนี้จะมากพอ ๆ กับการให้คะแนนตามผลการตอบข้อสอบข้อสุดท้าย และการให้คะแนนตามค่าความยากของข้อสอบข้อสุดท้ายที่ได้ตอบ การให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ตอบถูก จะมีความคงที่ของคะแนน และให้สารสนเทศของคะแนนมากกว่าการให้คะแนนตามผลการตอบข้อสอบข้อสุดท้าย และการให้คะแนนตามค่าความยากของข้อสอบข้อสุดท้ายที่ได้ตอบ อย่างไรก็ตามการให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ตอบถูกนี้ จะมีความไวต่อการเดาตอบถูกมากกว่าการให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ได้ตอบ ซึ่งอาจมีผลให้ความเชื่อมั่นต่ำลง

6. ให้คะแนนข้อสอบทุกข้อ (All Item Score)

แฮนส์เซน (Hansen ; Citing Weiss : 1974) ได้เสนอวิธีการให้คะแนนข้อสอบทุกข้อ ซึ่งสามารถใช้ได้กับการประมาณค่าความเชื่อมั่นชนิดความคงที่ภายใน (Internal Consistency) ของแบบทดสอบรูปปรีมาตรที่ทดสอบกับผู้สอบเป็นกลุ่ม และผู้สอบทุกคนจะต้องตอบข้อสอบตามเส้นทางการตอบทุกข้อ Hansen มีข้อดกลงว่าค่าความยากของข้อสอบและระดับความสามารถของผู้สอบถือเป็นเรื่องเดียวกันวิธีการให้คะแนนตามวิธีนี้ข้อสอบที่ตอบถูกทุกข้อจะให้ 2 คะแนน ให้คะแนนข้อที่ง่ายว่าข้อที่ตอบถูกในขั้นเดียวกันอีกข้อละ 2 คะแนน ให้คะแนนข้อสอบข้อถัดไปในขั้นเดียวกันที่มีความยากมากกว่าข้อที่ตอบถูก 1 คะแนน ส่วนข้อที่มีความยากมากกว่าข้อที่ตอบถูกในขั้นเดียวกันที่เหลือจะให้ 0 คะแนน การให้คะแนนข้อที่ตอบผิดจะให้ 0 คะแนน ให้คะแนนข้อสอบที่มีความยากน้อยกว่าและอยู่ติดกับข้อสอบที่ตอบผิดที่อยู่ขั้นเดียวกัน 1 คะแนนและให้คะแนนข้อสอบที่มีความยากน้อยกว่าข้อที่ตอบผิดที่เหลือทุกข้อข้อละ 2 คะแนน Hansen ได้เสนอวิธีการให้คะแนนไว้เท่านั้นไม่ได้ศึกษาประสิทธิภาพของวิธีการนี้ ต่อมา Larkin และ Weiss ได้วิจัยเพื่อศึกษาวิธีการให้คะแนนวิธีนี้ในปี ค.ศ. 1974 ผลการวิจัยพบว่า คะแนนที่ให้โดยวิธีนี้มีความสัมพันธ์กับคะแนนที่ให้ตามค่าความยากเฉลี่ยของข้อสอบที่ได้ตอบซึ่งเสนอโดย Lord สูงถึง .99

7. ให้คะแนนตามจำนวนข้อสอบที่ตอบถูก (Number Correct)

วิธีการนี้จะให้คะแนนตามจำนวนข้อที่ตอบถูก คือให้ข้อที่ตอบถูกข้อละ 1 คะแนน การให้คะแนนวิธีนี้จะไม่มีความไวต่อการเดาตอบของผู้สอบมาก และถ้าแบบทดสอบรูปปิรามิดที่มีจำนวนชั้นน้อย คะแนนที่ได้ก็จะน้อยเกินไปทำให้ความตรงเชิงเปรียบเทียบกับเกณฑ์มีค่าต่ำ การให้คะแนนวิธีนี้จึงให้ประโยชน์ในทางปฏิบัติน้อย

แบบทดสอบรูปปิรามิดเป็นแบบทดสอบที่นักวัดผลได้คิดค้นขึ้น โดยมีจุดเด่นดังนี้คือ เป็นแบบทดสอบที่ใช้ข้อสอบน้อยข้อ ซึ่งก่อให้เกิดผลดีทั้งในการตอบแบบทดสอบและการตรวจ ให้คะแนนทำให้ช่วยประหยัดเวลา และงบประมาณ สำหรับในการตอบแบบทดสอบของผู้สอบ นั้นผู้สอบก็จะไม่เกิดความเบื่อหน่ายเพราะผู้สอบจะตอบข้อสอบเพียงไม่กี่ข้อ ใช้นาน้อย ซึ่งช่วยทำให้ผู้สอบเกิดความตั้งใจที่จะตอบข้อสอบในแบบทดสอบ และนอกจากนั้นแบบทดสอบรูปปิรามิด ในแต่ละชั้นยังมีการแยกกลุ่มผู้สอบอยู่ตลอดเวลา ทำให้ผู้สอบได้ตอบข้อสอบที่เหมาะสมกับความสามารถของเขาในทุก ๆ ชั้นของการตอบแบบทดสอบซึ่งในลักษณะเช่นนี้จะก่อให้เกิดแรงจูงใจที่จะตอบข้อสอบโดยใช้ความสามารถที่แท้จริงอย่างเต็มที่อยู่ตลอดเวลาเป็นผลให้ได้คะแนนที่เกิดจากความสามารถอย่างแท้จริง แต่แบบทดสอบรูปปิรามิดก็ยังมีจุดอ่อนบางประการ คือ การวัดไม่ตรงตามความสามารถและการเดาซึ่งจะทำให้เกิดความคลาดเคลื่อนในการวัด กล่าวคือถ้าผู้สอบคนหนึ่งมีความสามารถสูงแต่เผชิญทำข้อสอบข้อแรกซึ่งวัดไม่ตรงตามความสามารถของเขาและทำผิด ผลจากการแยกทางทำให้เขาต้องไปทำข้อสอบที่ง่ายลงเป็นผลให้คะแนนที่ได้ต่ำกว่าความสามารถที่แท้จริงของเขาส่วนอีกคนหนึ่งมีความสามารถต่ำแต่เผชิญเดาข้อแรกถูก ผลที่ได้ออกมาก็คือเขาจะได้คะแนนสูงกว่าระดับความสามารถที่แท้จริงของเขา

ในการศึกษาครั้งนี้ ผู้วิจัยได้ทำการศึกษาแบบทดสอบรูปปิรามิด 3 รูปแบบได้แก่

1. แบบทดสอบเทเลอร์รูปปิรามิดที่มีที่มีขนาดชั้นคงที่ ใช้กฎการจำแนกแนวทางการตอบข้อสอบแบบ เพิ่ม 1 หรือ ลด 1 แบบ 5 ชั้น มีจำนวนข้อสอบทั้งหมด 15 ข้อ นักเรียนจะต้องทำข้อสอบแต่ละชั้น ๆ ละ 1 ข้อ รวมทั้งสิ้น 5 ข้อ
2. แบบทดสอบเทเลอร์รูปปิรามิดที่มีที่มีขนาดชั้นคงที่ ใช้กฎการจำแนกแนวทางการตอบข้อสอบแบบ เพิ่ม 1 หรือ ลด 1 แบบ 15 ชั้น มีจำนวนข้อสอบทั้งหมด 120 ข้อ นักเรียนจะต้องทำข้อสอบแต่ละชั้น ๆ ละ 1 ข้อ รวมทั้งสิ้น 15 ข้อ
3. แบบทดสอบรูปปิรามิดที่มีจำนวนข้อหลายข้อในแต่ละชั้นใช้กฎการจำแนกแนวทางการตอบข้อสอบแบบ เพิ่ม 1 หรือ ลด 1 แบบ 5 ชั้น มีจำนวนข้อสอบในแต่ละชั้น ๆ ละ 3 ข้อ จำนวนข้อสอบทั้งหมด 45 ข้อ นักเรียนจะต้องทำข้อสอบทั้งสิ้น 15 ข้อ

เหตุผลที่ผู้วิจัยเลือกทำการศึกษาแบบทดสอบเทเลอร์ทั้ง 3 รูปแบบคือต้องการศึกษาความเที่ยงตรงและประสิทธิภาพสัมพัทธ์ของแบบทดสอบทั้ง 3 รูปแบบ ซึ่งมีจุดเด่น และจุดอ่อนที่แตกต่างกันไป

การทดสอบแบบเทเลอร์เป็นแบบทดสอบที่พัฒนาขึ้นเพื่อที่จะทำให้วัดความสามารถของผู้สอบแต่ละคน ในระยะหลังได้มีการนำคอมพิวเตอร์มาช่วยในการดำเนินการสอบ ซึ่งสามารถจัดข้อสอบให้มีความเหมาะสมกับระดับความสามารถของผู้สอบ พร้อมทั้งนำเอาทฤษฎีการตอบสนองข้อสอบมาประยุกต์ใช้ในการประมาณค่าความสามารถของผู้สอบ แบบทดสอบเทเลอร์มีหลายรูปแบบ ซึ่งแต่ละรูปแบบก็มีข้อดีและข้อเสียแตกต่างกันออกไป ในการวิจัยนี้ผู้วิจัยได้ทำการเลือกแบบทดสอบเทเลอร์รูปปิรามิดมาใช้ในการศึกษา โดยทำการศึกษา แบบทดสอบเทเลอร์รูปปิรามิดที่มีขนาดขั้นคงที่ (Constant Step Size Pyramid) จำนวน 2 รูปแบบ ได้แก่ แบบทดสอบที่มี 5 ขั้น และ 15 ขั้น และแบบทดสอบเทเลอร์รูปปิรามิดที่มีหลายข้อในแต่ละขั้น (Multiple-Item Pyramid) โดยทำการเลือกวิธีการให้คะแนนของแบบทดสอบเทเลอร์รูปปิรามิดจำนวน 3 วิธี เพื่อต้องการศึกษาความเที่ยงตรงและประสิทธิภาพสัมพัทธ์ของแบบทดสอบทั้ง 3 แบบ

3. คุณภาพของแบบทดสอบ

3.1 คุณภาพของแบบทดสอบ

การตรวจสอบคุณภาพของแบบทดสอบตามทฤษฎีการตอบสนองข้อสอบจะพิจารณาจากความถูกต้องแม่นยำในการประมาณค่าความสามารถของผู้สอบ แบบทดสอบที่ดีหรือมีคุณภาพสูงนั้นจะต้องเป็นแบบทดสอบที่ให้ผลการสอบไปประมาณค่าความสามารถของผู้สอบได้ถูกต้องแม่นยำมากที่สุด และดัชนีที่ใช้วัดความแม่นยำในการประมาณค่าความสามารถของผู้สอบก็คือ ค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบ (Test Information Function : TIF) (Birnbaum. 1968 : 418) ซึ่งได้จากผลรวมของค่าฟังก์ชันอินฟอร์เมชันของข้อสอบ (Item Information Function) โดยที่ค่าฟังก์ชันอินฟอร์เมชันของข้อสอบหรือแบบทดสอบ เป็นดัชนีผสมที่สร้างจากดัชนีบอกคุณลักษณะของข้อสอบหลาย ๆ ลักษณะด้วยกัน เช่น ค่าความยาก ค่าอำนาจจำแนก และค่าการเดาของข้อสอบ เป็นต้น โดยรวมกันเป็นดัชนีเพียงตัวเดียว ซึ่งทำให้สามารถบ่งชี้ถึงคุณภาพของข้อสอบหรือแบบทดสอบได้ และด้วยคุณสมบัติความไม่แปรเปลี่ยนของค่าพารามิเตอร์ของข้อสอบ จึงทำให้ค่าฟังก์ชันอินฟอร์เมชันมีความเหมาะสมที่จะใช้เป็นดัชนีบอกถึงคุณภาพของข้อสอบหรือแบบทดสอบ ตามแนวทฤษฎีการตอบสนองข้อสอบ ซึ่งสามารถใช้แทนการหาค่าความเชื่อมั่น และค่าความคลาดเคลื่อนมาตรฐานในการวัด ตามแนวทฤษฎีการทดสอบแบบดั้งเดิมได้ (Hambleton and Cook. 1977 : 84) เนื่องจากการหาค่าความเชื่อมั่นเป็นวิธีการที่หยาบและค่าความคลาดเคลื่อนมาตรฐานในการวัดตั้งอยู่บนสมมติฐานที่ผิด ที่กำหนดให้ความแปรปรวนของความคลาดเคลื่อนมีค่าเท่ากันสำหรับทุก ๆ คะแนน (Green and others. 1984 : 352) แต่ความเป็นจริงแล้ว ความแปรปรวนของความคลาดเคลื่อนของผู้สอบแต่ละคนมีค่าไม่เท่ากัน โดยค่าความคลาดเคลื่อนสำหรับคะแนนปานกลาง จะมีค่าน้อยกว่าค่าความคลาดเคลื่อนสำหรับคะแนนต่ำหรือคะแนนสูง (Lord and Novick. 1968 : 183)

ดังนั้นจึงเป็นจุดอ่อนในการบ่งชี้ถึงคุณภาพของข้อสอบหรือแบบทดสอบ ส่วนค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบสามารถเขียนเป็นสมการได้ดังนี้ (Lord and Novick. 1968 : 454)

$$I(\theta) = \sum_{i=1}^n I(\theta, U_i) = \sum_{i=1}^n \left\{ \frac{[P'_i(\theta)]^2}{P_i(\theta)Q_i(\theta)} \right\}$$

เมื่อ $I(\theta)$ แทน ฟังก์ชันอินฟอร์เมชันของแบบทดสอบ

$I(\theta, U_i)$ แทน ฟังก์ชันอินฟอร์เมชันของข้อสอบแต่ละข้อ

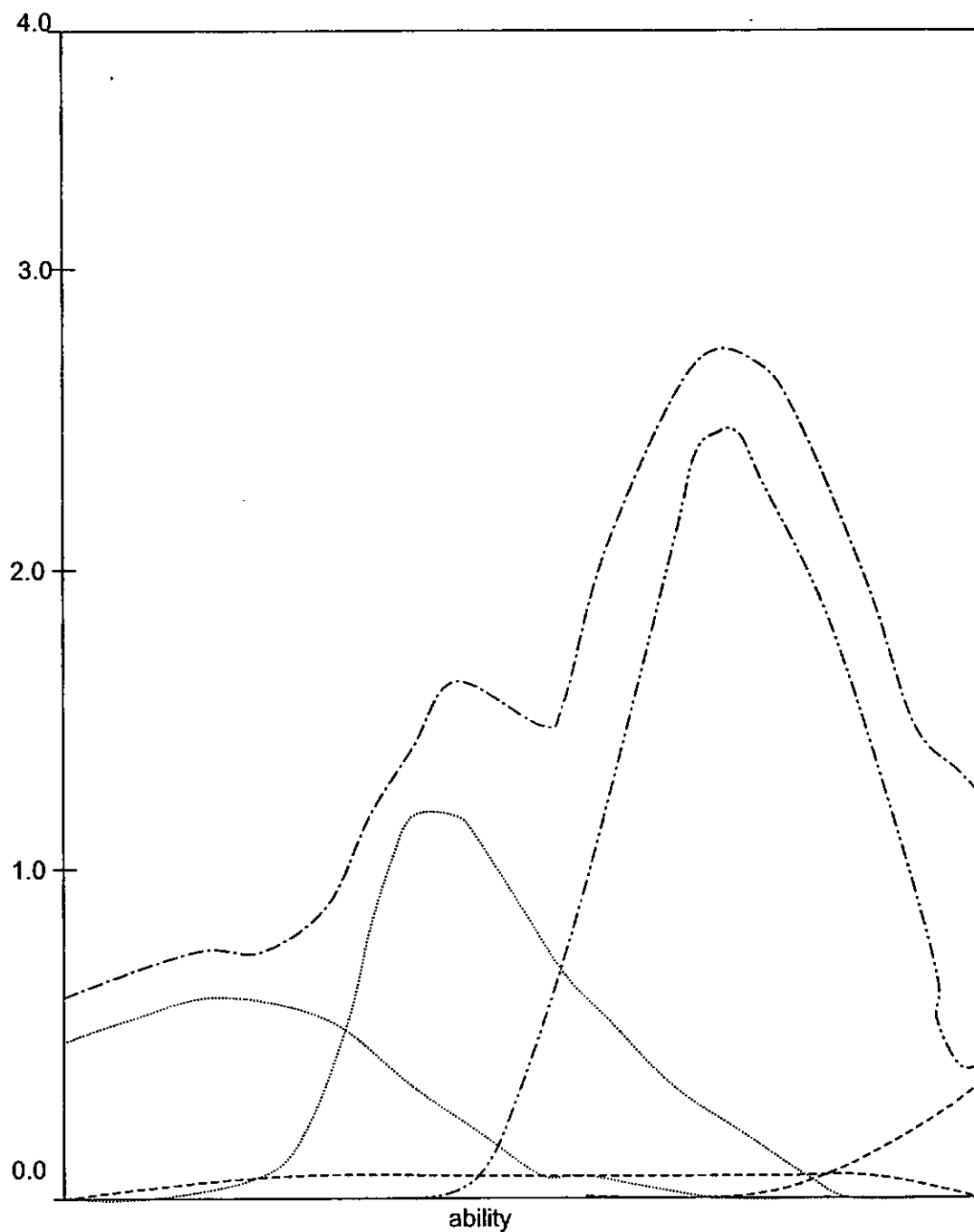
$P_i(\theta)$ แทน ความน่าจะเป็นที่ผู้สอบซึ่งมีระดับความสามารถ θ จะตอบข้อสอบข้อที่ i ได้ถูกต้อง

$Q_i(\theta)$ แทน $1 - P_i(\theta)$

$P'_i(\theta)$ แทน ความชันของโค้งลักษณะข้อสอบที่ระดับความสามารถ θ

ค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบฉบับหนึ่งๆ จะมีได้หลายค่าแตกต่างกันไปตามระดับความสามารถ ซึ่ง ณ ระดับความสามารถใดที่ฟังก์ชันอินฟอร์เมชันของแบบทดสอบมีค่าสูง แสดงว่าแบบทดสอบฉบับนั้นมีความแม่นยำในการประมาณค่าความสามารถที่ระดับนั้นจะมีค่ามาก แต่ถ้า ณ ระดับความสามารถใดที่ค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบมีค่าต่ำ แสดงว่าแบบทดสอบฉบับนั้นจะมีความแม่นยำในการประมาณค่าความสามารถที่ระดับนั้นจะมีค่าน้อย (Hambleton and Cook. 1977 : 84) และเมื่อแบบทดสอบมีจำนวนข้อสอบมากขึ้นค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบจะมีค่าเพิ่มมากขึ้นด้วย (Hambleton and Swaminathan. 1985 : 107) สำหรับประโยชน์ที่จะได้รับจากการทราบค่าฟังก์ชันอินฟอร์เมชันของข้อสอบแต่ละข้อนั้น ทำให้สามารถสร้างแบบทดสอบให้มีไค้งอินฟอร์เมชันของแบบทดสอบ ณ ระดับหนึ่งของความสามารถที่ต้องการได้ เช่น ถ้าต้องการสร้างแบบทดสอบเพื่อใช้ในการคัดเลือกนักเรียนให้ได้รับทุน ก็ต้องใช้แบบทดสอบที่มีไค้งอินฟอร์เมชันสูงสุด ณ ระดับความสามารถสูง ๆ เพื่อจะทำให้ได้แบบทดสอบที่มีประสิทธิภาพสูงสุด (ผจงจิต อินทสุวรรณ. 2525 : 66) ดังภาพประกอบ 9 (Hambleton and Cook. 1977 : 86; Citing Lord. 1968)

information



ภาพประกอบ 9 โค้งอินฟอร์เมชันของข้อสอบ 5 ข้อ และโค้งอินฟอร์เมชันของแบบทดสอบ

ภาพประกอบ 9 แสดงว่าข้อสอบทั้ง 5 ข้อนี้ มีความเหมาะสมกับกลุ่มผู้สอบที่มีความสามารถในระดับสูง นั่นคือ ถ้านำไปทดสอบกับกลุ่มผู้สอบที่มีความสามารถในระดับสูงแล้วจะทำให้การประมาณค่าความสามารถของผู้สอบมีความถูกต้องแม่นยำสูง (Hambleton and Cook. 1977 : 87)

ในการวัดหรือการประมาณค่าใด ๆ มักมีความคลาดเคลื่อนเกิดขึ้นเสมอ โดยเฉพาะอย่างยิ่งในการประมาณค่าความสามารถของผู้สอบทางด้านการศึกษา และจิตวิทยาจะมีความคลาดเคลื่อนค่อนข้างสูง ซึ่งจะส่งผลกระทบต่อความถูกต้องแม่นยำในการประมาณค่าความสามารถ โดยค่าความคลาดเคลื่อนในที่นี้จะอยู่ในรูปของความเบี่ยงเบนมาตรฐาน หรือเรียกว่า ความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ (Standard Error of Estimate : SEE (θ)) ซึ่งแตกต่างจากความคลาดเคลื่อนมาตรฐานในการวัด (Standard Error of Measurement : SE meas) ตรงที่ความคลาดเคลื่อนมาตรฐานในการประมาณค่า จะใช้ความคลาดเคลื่อนที่เกิดจากการประมาณค่าความสามารถของผู้สอบ ที่คลาดเคลื่อนไปจากความสามารถจริง ส่วนที่คลาดเคลื่อนมาตรฐานในการวัดจะใช้ความคลาดเคลื่อนที่เกิดจากคะแนนดิบหรือคะแนนที่ได้จากการวัด ที่คลาดเคลื่อนไปจากคะแนนจริง (Lord and Novick. 1968 : 66 - 67) สำหรับค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่ามีความสัมพันธ์กับค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบ ดังนี้ (Hambleton and Swaminathan. 1985 : 104 ; Citing Warm. 1978 : 77)

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

จากความสัมพันธ์ดังกล่าวแสดงว่า ค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบแปรผกผันกับค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ กล่าวคือ ถ้าแบบทดสอบฉบับใดมีค่าฟังก์ชันอินฟอร์เมชันสูง ณ ระดับความสามารถ θ แล้วความคลาดเคลื่อนมาตรฐานในการประมาณค่าจะมีค่าต่ำ แสดงว่าแบบทดสอบฉบับนั้นมีความแม่นยำในการประมาณค่าสูง ณ ระดับความสามารถ θ นั้น แต่ถ้าแบบทดสอบฉบับใดมีค่าฟังก์ชันอินฟอร์เมชันต่ำ ณ ระดับความสามารถ θ แล้วความคลาดเคลื่อนมาตรฐานในการประมาณค่าจะมีค่าสูง แสดงว่าแบบทดสอบฉบับนั้นมีความแม่นยำในการประมาณค่าต่ำ ณ ระดับความสามารถ θ นั้น

สำหรับค่าเฉลี่ยของความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ ($\overline{SEE}(\theta)$) มีความสัมพันธ์กับค่าความเชื่อมั่น (Reliability : r_{tt}) ดังนี้ (Warm. 1978 : 77)

$$r_{tt} = 1 - [\overline{SEE}(\theta)]^2$$

3.2 การประมาณค่าพารามิเตอร์

การนำทฤษฎีการตอบสนองข้อสอบมาประยุกต์ใช้ นอกจากจะต้องเลือกโมเดลให้มีความเหมาะสมแล้ว วิธีการประมาณค่าพารามิเตอร์ก็มีความสำคัญอีกประการหนึ่ง ซึ่งจะต้องเลือกใช้ให้เหมาะสมกับสภาพของการวัดแต่ละครั้ง เนื่องจากวิธีการประมาณค่าพารามิเตอร์ที่มีประสิทธิภาพจะส่งผลทำให้การสรุปอ้างอิงเกี่ยวกับข้อมูลทางสถิติได้ถูกต้อง แต่ถ้าวิธีการประมาณค่าพารามิเตอร์ขาดประสิทธิภาพแล้ว ย่อมจะทำให้การสรุปอ้างอิงมีความคลาดเคลื่อนจากความเป็นจริงได้ สำหรับวิธีการประมาณค่าพารามิเตอร์ของราล์ช โมเดล มีด้วยกันหลายวิธีแต่ละข้อเสนอวิธีที่สำคัญเพียง 3 วิธี ดังนี้ (อรินทร์ น่วมถนอม. 2537 : 25 – 26)

3.2.1 วิธีความเป็นไปได้สูงสุดแบบไม่มีเงื่อนไข (Conditional

Maximum Likelihood Procedure : CON) เป็นวิธีการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ และค่าความยากของข้อสอบได้อย่างอิสระซึ่งกันและกัน โดยการกำหนดเงื่อนไขให้การประมาณค่าพารามิเตอร์ความสามารถของผู้สอบแยกออกจากค่าความยากของข้อสอบ ซึ่งทำให้การประมาณค่าความสามารถไม่จำเป็นต้องอาศัยค่าความยาก และในทำนองเดียวกัน การประมาณค่าความยากไม่จำเป็นต้องรู้ค่าความสามารถก็ได้ สำหรับการประมาณค่าโดยวิธีนี้ ไม่มีการกำหนดข้อตกลงเบื้องต้นเกี่ยวกับการแจกแจงค่าความสามารถ และค่าความยากของข้อสอบ อย่างไรก็ตามการประมาณค่าโดยวิธีนี้จะมีความยุ่งยากในการคำนวณ และมีข้อจำกัดในกรณีที่แบบทดสอบมีจำนวนมาก ๆ โดยไม่สามารถวิเคราะห์ข้อสอบที่มีจำนวนข้อสอบมากกว่า 25 ข้อขึ้นไปได้

3.2.2 วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (Unconditional

Likelihood Procedure : UCON) เป็นวิธีการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ และค่าความยากของข้อสอบไปพร้อม ๆ กัน โดยมีการกำหนดข้อตกลงเบื้องต้นเกี่ยวกับการแจกแจงความสามารถของผู้สอบ และค่าความยากของข้อสอบมีการแจกแจงแบบปกติ สำหรับการประมาณค่า โดยวิธีนี้มีข้อดีก็คือ สามารถใช้วิเคราะห์ข้อสอบที่มีความยาวทุกขนาด ถึงแม้ว่าจะมีความลำเอียง ในการคำนวณที่เกิดจากความยาวของข้อสอบ แต่ก็สามารถปรับแก้ได้

3.2.3 วิธีพรีอ็อกซ์ (Cohen's Approximtion Procedure : PROX) เป็น

วิธีการประมาณค่าพารามิเตอร์ที่สามารถคำนวณได้อย่างง่าย ๆ ถ้าข้อมูลที่น่ามาวิเคราะห์มีไม่มากนักสามารถคำนวณได้โดยไม่ต้องใช้เครื่องคอมพิวเตอร์ก็ได้ การประมาณค่าพารามิเตอร์โดยวิธีนี้ มีข้อตกลงเบื้องต้นเกี่ยวกับค่าความสามารถของผู้สอบ และค่าความยากของข้อสอบมีการแจกแจงในลักษณะเบ้หรือมีแนวโน้มเป็นแบบปกติ สำหรับจุดเด่นของการประมาณค่าด้วยวิธีพรีอ็อกซ์ก็คือ ค่าที่ประมาณได้จะเป็นอิสระจากกลุ่มตัวอย่าง และความยาวของแบบทดสอบ โดยแปลงคะแนนให้อยู่ในรูปของคะแนนโลจิต (Logit Score)

ในปัจจุบันนี้ ได้มีการนำเอาทฤษฎีการตอบสนองข้อสอบมาใช้ในวงการวัดผลการศึกษาและจิตวิทยามากยิ่งขึ้น เนื่องจากมีการพัฒนาโปรแกรมคอมพิวเตอร์เพื่อใช้ในการวิเคราะห์ข้อมูล จึงทำให้ได้รับความสะดวกและรวดเร็วเป็นอย่างมาก ในปี ค.ศ. 1976 ลอร์ด และคนอื่น ๆ ได้พัฒนาโปรแกรมโลจิส (LOGIST) (Hambleton and Swaminathan. 1985 : 5 ; Citing Lord and others. 1976) ซึ่งต่อมาได้มีการปรับปรุงและพัฒนาเพิ่มขึ้นเรื่อย ๆ จนสามารถนำมาใช้วิเคราะห์ข้อมูลได้เป็นอย่างดี โดยสามารถประมาณค่าพารามิเตอร์แบบโลจิสติกโมเดล ที่มีพารามิเตอร์ 1, 2 หรือ 3 ตัว และในปี ค.ศ.1976 ไวท์ และ มิด ได้พัฒนาโปรแกรมไบคาล (BICAL) (Hambleton and Cook. 1977 : 90 ; Citing Wright and Mead. 1976) เพื่อใช้ประมาณค่าพารามิเตอร์ แบบราสซัส โมเดล นอกจากนี้ยังมีการพัฒนาโปรแกรมอื่น ๆ อีก เช่น โปรแกรมแอลซิลเลส (ANCILLES) พีเอ็มแอล (PML) และไบล็อก (BILOG) (Hambleton and Swaminathan. 1985 : 148) เป็นต้น

สำหรับในการศึกษาครั้งนี้ผู้วิจัยได้เลือกใช้การประมาณค่าพารามิเตอร์ ด้วยวิธีความเป็นไปได้สูงสุดแบบไม่มีเงื่อนไข โดยใช้โปรแกรมสำเร็จรูปไบคาล

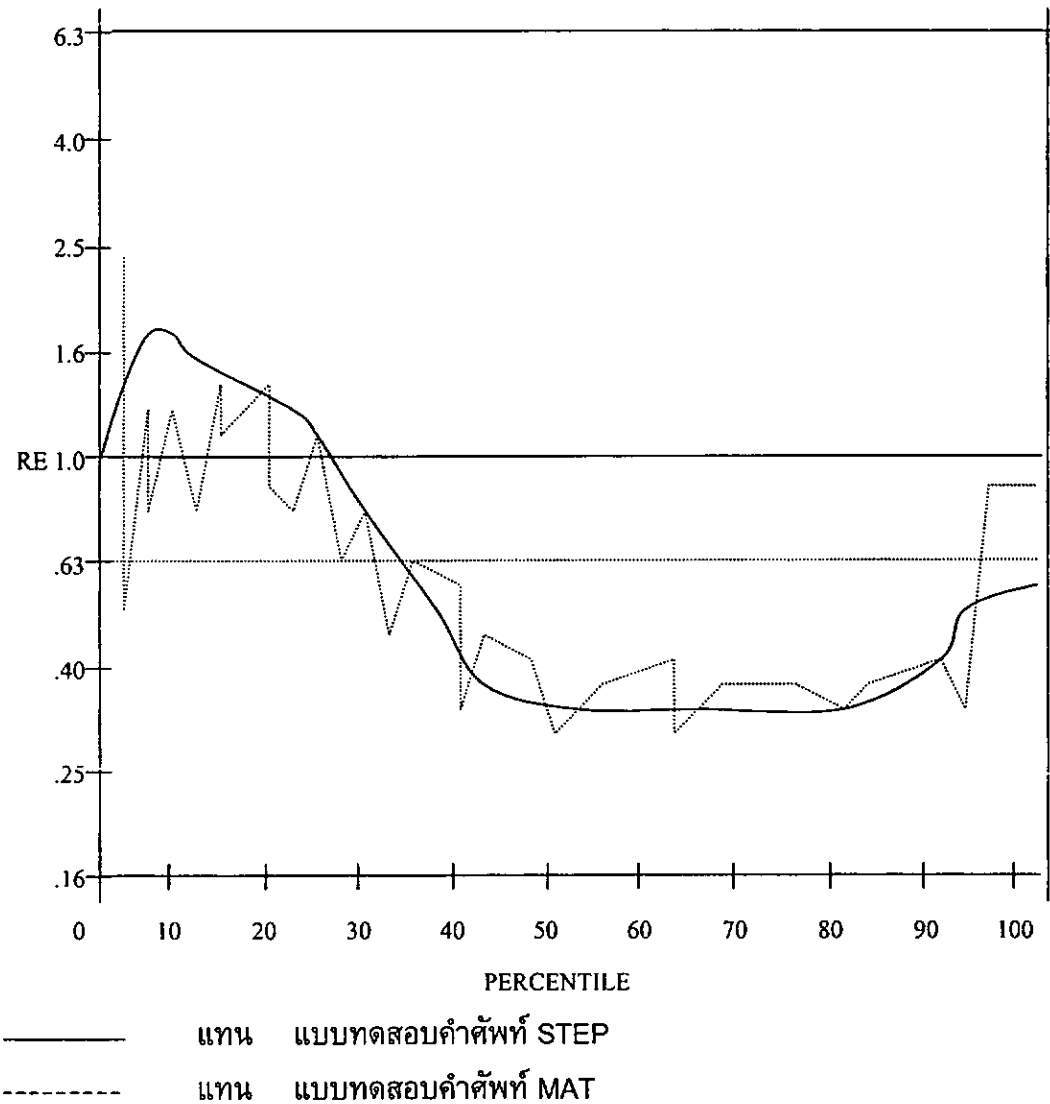
3.3 ประสิทธิภาพของแบบทดสอบ

ค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบนอกจากจะใช้พิจารณาถึงคุณภาพของแบบทดสอบแล้ว ยังสามารถนำมาใช้เปรียบเทียบหาประสิทธิภาพของแบบทดสอบได้อีกด้วย โดยการคำนวณหาอัตราส่วนของค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบ 2 ฉบับ ที่นำมาเปรียบเทียบกัน ค่าที่คำนวณได้เรียกว่า ค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่า (Relative Efficiency of Estimators : RE) ซึ่งคำนวณได้จากสูตรดังนี้ (Lord and Novick. 1968 : 471 - 472)

$$RE(\theta, X_1, X_2) = \frac{I(\theta, X_1)}{I(\theta, X_2)}$$

เมื่อ $RE(\theta, X_1, X_2)$ แทน	ค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถระหว่างแบบทดสอบรูปแบบที่ 1 (X_1) เทียบกับแบบทดสอบรูปแบบที่ 2 (X_2)
$I(\theta, X_1)$ แทน	ค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบรูปแบบที่ 1 (X_1) ณ ระดับความสามารถ θ
$I(\theta, X_2)$ แทน	ค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบรูปแบบที่ 2 (X_2) ณ ระดับความสามารถ θ

สำหรับการแปลความหมายของค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถ ระหว่างแบบทดสอบที่นำมาเปรียบเทียบกันพิจารณาดังนี้ ถ้าได้ $RE(\theta, X_1, X_2)$ มากกว่า 1 แสดงว่าแบบทดสอบรูปแบบที่ 1 (X_1) มีประสิทธิภาพสูงกว่าแบบทดสอบที่ 2 (X_2) ณ ระดับความสามารถ θ แต่ถ้า $RE(\theta, X_1, X_2)$ น้อยกว่า 1 แสดงว่าแบบทดสอบรูปแบบที่ 1 (X_1) มีประสิทธิภาพต่ำกว่าแบบทดสอบรูปแบบที่ 2 (X_2) ณ ระดับความสามารถ θ หรือถ้า $RE(\theta, X_1, X_2)$ เท่ากับ 1 แสดงว่าแบบทดสอบทั้ง 2 รูปแบบมีประสิทธิภาพพอ ๆ กัน ณ ระดับความสามารถ θ (Warm. 1978 : 75 – 76) ซึ่งสามารถแสดงได้ด้วยภาพประกอบ 4 ดังนี้ (Lord. 1980 : 97 ; Citing Lord. 1974)



ภาพประกอบ 10 ไค้งประสิทธิภาพสัมพัทธ์ของแบบทดสอบคำศัพท์ STEP เทียบกับแบบทดสอบคำศัพท์ MAT

ภาพประกอบ 10 แบบทดสอบคำศัพท์ STEP มีประสิทธิภาพสูงกว่าแบบทดสอบคำศัพท์ MAT ในกลุ่มผู้ที่มีความสามารถระดับต่ำ ๆ แต่ด้อยกว่าในกลุ่มผู้สอบที่มีความสามารถระดับสูง ๆ สำหรับเส้นประตามแนวนอนเป็นอัตราส่วนของจำนวนข้อของแบบทดสอบทั้ง 2 ฉบับ

การตรวจสอบคุณภาพของแบบทดสอบตามทฤษฎีการตอบสนองข้อสอบจะพิจารณาจากความถูกต้องแม่นยำในการประมาณค่าความสามารถของผู้สอบ แบบทดสอบที่ดีหรือมีคุณภาพสูงนั้นจะต้องเป็นแบบทดสอบที่ให้ผลการสอบไปประมาณค่าความสามารถของผู้สอบได้ถูกต้องแม่นยำมากที่สุด และดัชนีที่ใช้วัดความแม่นยำในการประมาณค่าความสามารถของผู้สอบก็คือ ค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบ (Test Information Function : TIF) ซึ่งได้จากผลรวมของค่าฟังก์ชันอินฟอร์เมชันของข้อสอบ (Item Information Function) โดยที่ค่าฟังก์ชันอินฟอร์เมชันของข้อสอบหรือแบบทดสอบ เป็นดัชนีผสมที่สร้างจากดัชนีบอกคุณลักษณะของข้อสอบหลาย ๆ ลักษณะด้วยกัน เช่น ค่าความยาก ค่าอำนาจจำแนก และค่าการเดาของข้อสอบ ซึ่งสามารถบ่งชี้ถึงคุณภาพของข้อสอบหรือแบบทดสอบได้

4. งานวิจัยที่เกี่ยวข้องกับแบบทดสอบรูปปริมาตร

4.1. งานวิจัยภายในประเทศ

จิราพร ไกรสรศิวเวท (2529 : 111) ได้ศึกษาถึงความสัมพันธ์ระหว่างรูปแบบของแบบทดสอบรูปปริมาตรกับความสามารถทางการเรียน โดยใช้แบบทดสอบรูปปริมาตรชนิด 8 ชั้น 3 แบบ คือ แบบที่ใช้ขนาดชั้นคงที่ แบบที่ใช้ขนาดชั้นแปรผัน และแบบข้างตัด และใช้วิธีการตรวจให้คะแนน 3 วิธี คือ ให้คะแนนตามความยากเฉลี่ยของข้อสอบที่ทำ ให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ตอบถูก และให้คะแนนข้อสอบทุกข้อ กับแบบทดสอบดั้งเดิมจำนวน 30 ข้อ แบบทดสอบทุกฉบับเป็นแบบเลือกตอบชนิด 5 ตัวเลือก ที่สร้างมาจากกลุ่มข้อสอบเดียวกัน จำนวน 200 ข้อ ในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 2 ทำการทดสอบกับกลุ่มตัวอย่าง 348 คน พบว่า แบบทดสอบรูปปริมาตรทั้ง 3 แบบซึ่งใช้วิธีตรวจให้คะแนนทั้ง 3 วิธี มีความสัมพันธ์ทางบวกกับความสามารถทางการเรียนอย่างมีนัยสำคัญทางสถิติ โดยแบบทดสอบรูปปริมาตรที่ใช้ขนาดชั้นแปรผันซึ่งตรวจให้คะแนนโดยวิธีให้คะแนนข้อสอบทุกข้อมีความสัมพันธ์กับความสามารถทางการเรียนสูงที่สุดเท่ากับ .55 และวิธีการตรวจให้คะแนนทั้ง 3 วิธีในแบบทดสอบรูปปริมาตรทั้ง 3 แบบมีความสัมพันธ์กันทางบวกอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 นอกจากนี้ยังได้ศึกษาเปรียบเทียบความแตกต่างของค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างแบบทดสอบรูปปริมาตรแต่ละฉบับกับความสามารถทางการเรียนและเปรียบเทียบความแตกต่างของสัมประสิทธิ์สหสัมพันธ์ภายในของวิธีการตรวจให้คะแนนทั้ง 3 วิธี ปรากฏว่า แบบทดสอบรูปปริมาตรที่ใช้ขนาดชั้นแปรผันและแบบข้างตัดแตกต่างกันอย่างมีนัยสำคัญทางสถิติในทุกวิธีการตรวจให้คะแนน สำหรับแบบที่ใช้ขนาดชั้นคงที่และแบบที่ใช้ขนาดชั้นแปรผันแตกต่างกันในการตรวจให้คะแนนตามวิธีที่ 2 และ 3 ส่วนแบบที่ใช้ขนาดชั้นคงที่กับแบบข้างตัดแตกต่างกันเฉพาะการตรวจให้คะแนนตามวิธีที่ 1 และในการตรวจให้คะแนนทั้งสามวิธีของแบบทดสอบทั้ง 3 แบบจะมีค่าสัมประสิทธิ์สหสัมพันธ์ภายในแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ยกเว้นวิธีที่ 1 และวิธีที่ 2 ของแบบทดสอบรูปปริมาตรที่ใช้ขนาดชั้นคงที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนในแบบข้างตัดไม่แตกต่างกัน และวิธีที่ 1 และกับวิธีที่ 3 ในแบบที่ใช้ขนาดชั้นแปรผันไม่แตกต่างกัน

ต่อมา ในปี พ.ศ. 2531 นันทิยา พึ่งคำ (2531) ได้ศึกษาเปรียบเทียบคุณภาพของการทดสอบแบบ ซีเอที (CAT) กับแบบดั้งเดิม (Conventional) ในการวัดความสามารถด้านคำศัพท์ภาษาอังกฤษ โดยทำการศึกษากับนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 120 คน การทดสอบแบบซีเอที ใช้กลุ่มข้อสอบ (Item Pool) จำนวน 361 ข้อ โดยมีค่าความยากอยู่ระหว่าง -.90552 ถึง 3.61860 มีค่าอำนาจจำแนกอยู่ระหว่าง 0.80334 ถึง 2.0000 และมีการเดาอยู่ระหว่าง 0.04421 ถึง 0.29886 การประมาณค่าความสามารถของผู้สอบโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (Conditional Maximum Likelihood) ส่วนการทดสอบแบบดั้งเดิมใช้ข้อสอบทั้งหมด

40 ข้อ โดยมีค่าความยากอยู่ระหว่าง -0.76488 ถึง 2.97884 มีค่าอำนาจจำแนกอยู่ระหว่าง 0.94069 ถึง 2.00000 และมีค่าการเดาอยู่ระหว่าง 0.03462 ถึง 0.28620 ผลการศึกษาพบว่า

1. การทดสอบแบบซีเอที กับแบบดั้งเดิมมีความตรงเชิงเกณฑ์สัมพันธ์ ไม่แตกต่างกันแต่การทดสอบแบบ ซีเอที สามารถลดจำนวนข้อสอบลงได้มากกว่าครึ่งหนึ่ง

2. ความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ จากการทดสอบแบบซีเอที (CAT) มีค่าต่ำกว่า ความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบจากการทดสอบแบบดั้งเดิม อย่างมีนัยสำคัญทางสถิติ (ต่าย เชิญฉวี. 2534 : 57-58)

ในปี 2535 ระจิต ตรีพุทธรัตน์ และคนอื่น ได้ศึกษาความสอดคล้องของคะแนนผลการสอบจากแบบทดสอบปิรามิดในวิชาเคมี ระดับชั้นมัธยมศึกษาปีที่ 5 ที่วิเคราะห์ข้อสอบโดยทฤษฎีคุณลักษณะแฝงกับทฤษฎีดั้งเดิม โดยทำการศึกษาความสอดคล้องของค่าความยากและคะแนนผลการสอบของข้อสอบในโครงสร้างแบบทดสอบปิรามิดที่ดำเนินการสอบด้วยคอมพิวเตอร์ วิชาเคมีชั้นมัธยมศึกษาปีที่ 5 ที่วิเคราะห์ค่าความยากโดยทฤษฎีคุณลักษณะแฝงกับทฤษฎีดั้งเดิม ใช้เครื่องมือและข้อมูลที่ไพศาล (2532) ได้ศึกษาไว้ แต่นำมาวิเคราะห์ความยากของข้อสอบโดยทฤษฎีดั้งเดิมเพิ่มขึ้น ตรวจสอบให้คะแนนเพิ่มขึ้นเป็น 5 วิธี โดยใช้การวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์แบบ Pearson Product Moment ผลการวิจัยพบว่า ค่าความยากของข้อสอบในโครงสร้างแบบทดสอบปิรามิดที่วิเคราะห์ข้อสอบโดยทฤษฎีคุณลักษณะแฝงกับทฤษฎีดั้งเดิมมีความสอดคล้องกันอย่างมีนัยสำคัญทางสถิติที่ระดับ $.001$ ($r=.95$) คะแนนผลการสอบที่ตรวจสอบให้คะแนนโดยใช้ค่าความยากที่วิเคราะห์ข้อสอบโดยทฤษฎีคุณลักษณะแฝงกับทฤษฎีดั้งเดิม มีความสอดคล้องกันอย่างมีนัยสำคัญทางสถิติที่ระดับ $.001$ คะแนนผลการสอบที่ตรวจสอบให้คะแนนโดยหาค่าความยากเฉลี่ยของข้อสอบทุกข้อที่ได้อตอบ กับวิธีหาค่าความยากเฉลี่ยของข้อสอบที่ตอบถูกสอดคล้องกันอย่างมีนัยสำคัญทางสถิติที่ระดับ $.001$ การตรวจสอบให้คะแนนโดยหาค่าความยากเฉลี่ยของข้อสอบที่ตอบถูกกับการนับจำนวนข้อที่ตอบถูกมีความสอดคล้องกันในระดับปานกลาง ส่วนการตรวจสอบให้คะแนนโดยวิธีหาค่าความยากเฉลี่ยของข้อสอบทุกข้อที่ได้อตอบกับการนับจำนวนข้อที่ตอบถูกไม่มีความสอดคล้องกัน

ปี พ.ศ. 2537 อรินทร์ น่วมถนอม ได้ศึกษาความเที่ยงตรงและประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถในวิชาคณิตศาสตร์ ระหว่างแบบทดสอบดั้งเดิม แบบทดสอบสองขั้นตอน และแบบทดสอบเฟลิกซ์เลเวล โดยศึกษาในกลุ่มตัวอย่างนักเรียนชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2536 ของโรงเรียนสังกัดกรมสามัญศึกษา กรุงเทพมหานคร จำนวน 441 คน เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ที่ผู้วิจัยสร้างขึ้นเอง มี 3 รูปแบบ ได้แก่ แบบทดสอบดั้งเดิม แบบทดสอบสองขั้นตอน และแบบทดสอบเฟลิกซ์เลเวล ผลการวิจัยพบว่า ค่าความเที่ยงตรงของการประมาณค่าความสามารถระหว่างแบบทดสอบดั้งเดิม กับแบบทดสอบสองขั้นตอน มีค่าแตกต่างกันอย่างมีนัยสำคัญทาง

สถิติที่ระดับ .01 ส่วนค่าความเที่ยงตรงของการประมาณค่าความสามารถระหว่างแบบทดสอบดั้งเดิมกับแบบทดสอบเฟลิกซิเลเวล และแบบทดสอบสองขั้นตอนกับแบบทดสอบเฟลิกซิเลเวล มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ และค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของผู้สอบระหว่างแบบทดสอบดั้งเดิมกับแบบทดสอบสองขั้นตอน และแบบทดสอบดั้งเดิมกับแบบทดสอบเฟลิกซิเลเวลมีค่าสูงกว่า 1 ในกลุ่มผู้สอบที่มีความสามารถระดับปานกลาง แต่มีค่าต่ำกว่า 1 ในกลุ่มผู้สอบที่มีความสามารถระดับต่ำลงมาและระดับสูงขึ้นไป ส่วนค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถ ระหว่างแบบทดสอบสองขั้นตอนเทียบกับแบบทดสอบเฟลิกซิเลเวลมีค่าสูงกว่า 1 ในกลุ่มผู้สอบที่มีความสามารถระดับต่ำมาก ระดับปานกลาง และระดับสูงมาก แต่มีค่าต่ำกว่า 1 ในกลุ่มผู้สอบที่มีระดับความสามารถต่ำ และระดับสูง

4.2 งานวิจัยในต่างประเทศ

แครทวอห์ และ ฮายเซอร์ (Krathwohl and Huyser : 1956) ได้ศึกษาแบบทดสอบปิรามิดชนิด 8 ขั้นที่มีข้อสอบขั้นละ 1 ข้อ และแบบทดสอบปิรามิดชนิด 4 ขั้นที่มีข้อสอบขั้นละ 2 ข้อ แบบทดสอบปิรามิดทั้ง 2 ฉบับดำเนินการสอบโดยการเขียนตอบ (Paper and Pencil) เปรียบเทียบกับแบบทดสอบดั้งเดิมที่มีข้อสอบจำนวน 60 ข้อ ข้อสอบในแบบทดสอบปิรามิดและแบบทดสอบดั้งเดิมมีลักษณะเหมือนกันคือวัดในเนื้อหาเดียวกัน จากการศึกษาแก่นักศึกษาในระดับวิทยาลัย 100 คน พบว่าแบบทดสอบปิรามิดต้องใช้คำชี้แจงประกอบมาก แต่เวลาที่ใช้ในการสอบจะน้อยกว่าแบบทดสอบดั้งเดิม ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างแบบทดสอบปิรามิดชนิด 8 ขั้นกับแบบทดสอบดั้งเดิมมีค่าเท่ากับ .78 และค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างแบบทดสอบปิรามิดชนิด 4 ขั้นกับแบบทดสอบดั้งเดิมมีค่าเท่ากับ .68

เบย์รอฟฟ์, โทมัส และ แอนเดอร์สัน (Bayroff, Thomas and Anderson ; Citing Larkin and Weiss : 1974) ได้ศึกษาในลักษณะใกล้เคียงกับวิธีการของ Krathwohl โดยสร้างแบบทดสอบปิรามิดชนิด 6 ขั้นซึ่งมีค่าความแตกต่างระหว่างค่าความยากของข้อสอบในขั้นที่อยู่ติดกันลดลงเรื่อย ๆ ข้อสอบเป็นแบบเลือกตอบชนิด 4 ตัวเลือก ได้กำหนดเกณฑ์การเลือกข้อสอบข้อที่ 2 โดยอาศัยผลการตอบข้อสอบข้อที่ 1 เป็นหลักในการพิจารณาดังนี้ ถ้าตอบข้อสอบข้อแรกถูกข้อสอบข้อที่ 2 ที่จะได้ทำจะมีความยากมากขึ้น ถ้าตอบผิดโดยเลือกตอบตัวเลือกที่มีคนเลือกตอบน้อยที่สุด จะให้ทำข้อสอบในขั้นที่สองข้อที่มีความยากลดลง ส่วนในขั้นต่อ ๆ ไปใช้รูปแบบการเพิ่ม 1 หรือ ลด 1 ซีลีย์, มอร์ดัน และ แอนเดอร์สัน (Seeley, Morton and Anderson : 1962) ได้นำแบบทดสอบดังกล่าวไปทดลองสอบกับกลุ่มตัวอย่างจำนวน 327 คน แล้วมาหาค่าสหสัมพันธ์กับแบบทดสอบชนิดดั้งเดิมได้เท่ากับ .63 และพบว่าคะแนนสอบด้วยแบบทดสอบปิรามิดมีความแปรปรวนสูง คือส่วนใหญ่จะได้คะแนนสูง ผู้วิจัยให้ข้อคิดเห็นว่าเป็นเพราะผู้ที่ได้คะแนน

ส่วนใหญ่ไม่ได้ทำข้อสอบตามคำชี้แจง ดังนั้นคะแนนของผู้ที่ได้คะแนนต่ำจึงไม่สามารถนำมาวิเคราะห์ได้

วูด (Wood : 1969) ได้ทดลองใช้แบบทดสอบปิรามิดชนิด 4, 5 และ 6 ชั้นกับตัวอย่างจำนวน 91 คน แบบทดสอบดังกล่าวใช้ขนาดของชั้นแบบคงที่ คือ $b = .05$ ข้อสอบข้อแรกมีความยากปานกลาง คือ $b = .50$ ใช้เทคนิคการแยกทางแบบเพิ่ม 1 หรือ ลด 1 และให้คะแนนโดยใช้จำนวนข้อที่ตอบถูกได้ศึกษาความเที่ยงตรงของแบบทดสอบโดยหาค่าสหสัมพันธ์ระหว่างคะแนนที่สอบได้ กับระดับผลการเรียนที่สอบโดยแบบทดสอบดั้งเดิมจำนวน 46 ข้อ ซึ่งผลได้ค่าสหสัมพันธ์เท่ากับ .35 ทั้ง 3 ฉบับ เมื่อรวมคะแนนทั้ง 3 ฉบับแล้วมาหาค่าสหสัมพันธ์กับระดับผลการเรียนพบว่ามีความเพิ่มขึ้นเป็น .51 เมื่อเปรียบเทียบกับค่าสหสัมพันธ์ระหว่างคะแนนจากแบบทดสอบดั้งเดิมกับระดับผลการเรียนซึ่งมีค่าเท่ากับ 0.68 แล้ว Wood สรุปผลการวิจัยว่าแบบทดสอบดั้งเดิมมีประสิทธิภาพดีเท่ากับแบบทดสอบปิรามิดที่มีจำนวนข้อสอบเท่ากัน

ต่อมา เบย์รอฟฟ์ และ ซีเลย์ (Bayroff and Seeley ; Citing Larkin and Weiss : 1974 : 74-3) ได้ทดลองใช้แบบทดสอบปิรามิดชนิด 8 ชั้น กับตัวอย่างจำนวน 102 คน ขนาดของชั้นของแบบทดสอบเท่ากับ .05 ($b=.05$) ข้อสอบในชั้นสุดท้ายคือชั้นที่ 8 มีความยากกระจายตั้งแต่ $b=.95$ ถึง $b=.20$ การให้คะแนนขึ้นอยู่กับผลการตอบข้อสอบข้อสุดท้าย ว่าถูกหรือผิดซึ่งช่วงคะแนนที่จะได้เท่ากับ 17 คะแนน กลุ่มตัวอย่างสอบแบบทดสอบรูปปิรามิดพร้อมกับแบบทดสอบดั้งเดิมอีก 2 ฉบับ เป็นข้อสอบทางด้านจำนวน จำนวน 40 ข้อ และเป็นข้อสอบทางด้านภาษาจำนวน 50 ข้อ ผลการวิจัยพบว่าค่าสหสัมพันธ์ระหว่างแบบทดสอบปิรามิดกับแบบทดสอบดั้งเดิม 2 ฉบับมีค่าเท่ากับ .83 และ .79 ตามลำดับซึ่งเมื่อเปรียบเทียบกับค่าสหสัมพันธ์ระหว่างคะแนนข้อสอบแบบดั้งเดิมจำนวน 8 ข้อ กับคะแนนผลการสอบจากแบบทดสอบดั้งเดิมทั้ง 2 ฉบับมีค่าเท่ากับ .75 และ .67 ตามลำดับ จึงสรุปได้ว่าแบบทดสอบรูปปิรามิดมีค่าสหสัมพันธ์กับแบบทดสอบดั้งเดิมสูงกว่าแบบทดสอบดั้งเดิมที่มีจำนวนข้อสอบเท่ากันกับแบบทดสอบดั้งเดิมฉบับเดียวกัน นอกจากนี้จากการใช้สูตรของ Spearman-Brown พบว่าจะต้องใช้จำนวนข้อสอบในแบบทดสอบดั้งเดิมเป็น 2 เท่าของจำนวนข้อสอบในแบบทดสอบปิรามิด เพื่อให้ได้ค่าสหสัมพันธ์กับแบบทดสอบที่ใช้เป็นเกณฑ์เท่ากัน

แฮนส์เซน (Hansen ; Citing Larkin and Weiss. 1974 : 74-3) ได้ทดลองดำเนินการสอบด้วยคอมพิวเตอร์กับแบบทดสอบปิรามิดจำนวน 5 ฉบับกับนักศึกษาระดับวิทยาลัยปีที่ 1 จำนวน 56 คน ข้อสอบทั้ง 5 ฉบับมีทั้งชนิด 3 และ 5 ชั้น ซึ่งนักศึกษาแต่ละคนจะได้ทำข้อสอบจำนวน 17 ข้อ ขนาดของชั้นเท่ากับ .10 ($b=.10$) การให้คะแนนใช้วิธีที่แตกต่างกัน 4 วิธี ก่อนให้ทำแบบทดสอบปิรามิด 1 สัปดาห์ให้กลุ่มตัวอย่างสอบแบบทดสอบดั้งเดิมในเนื้อหาเดียวกัน ผู้วิจัยได้ศึกษาความสัมพันธ์ระหว่างคะแนนที่ได้จากแบบทดสอบปิรามิดกับคะแนนจากแบบทดสอบดั้งเดิม คะแนนจากแบบทดสอบผลสัมฤทธิ์อีก 1 ฉบับ และระดับผลการเรียนปลายภาคการศึกษา ผลการศึกษาพบว่าแบบทดสอบดั้งเดิม แม้จะปรับให้มีความยาวเท่ากับแบบทดสอบ

ปริมาตรแล้วก็ตาม จะมีค่าความเชื่อมั่นชนิดความคงที่ภายใน (Internal Consistency) ต่ำกว่าแบบทดสอบปริมาตรทั้ง 5 ฉบับ คะแนนผลการสอบจากแบบทดสอบปริมาตรมีการกระจายเป็นปกติมากกว่าจากแบบทดสอบดั้งเดิม ซึ่งการกระจายของคะแนนมีความเบ้ทางลบ นอกจากนี้ผลการวิจัยยังพบอีกว่า โดยเฉลี่ยแล้วผู้สอบจะใช้เวลาในการสอบแบบทดสอบปริมาตรทั้ง 5 ฉบับน้อยกว่าแบบทดสอบดั้งเดิมประมาณ 5 นาที การให้คะแนนแบบทดสอบปริมาตร 2 วิธีมีค่าสหสัมพันธ์กับคะแนนแบบทดสอบผลสัมฤทธิ์ และระดับผลการเรียนปลายภาคสูงกว่าค่าสหสัมพันธ์ระหว่างแบบทดสอบดั้งเดิม กับคะแนนแบบทดสอบผลสัมฤทธิ์และระดับผลการเรียน

เบย์สัน (Bryson ; Citing Larkin and Weiss, 1974 : 74-3) ได้ศึกษาเปรียบเทียบค่าสหสัมพันธ์ระหว่างแบบทดสอบปริมาตรชนิด 5 ชั้นจำนวน 2 ฉบับและแบบทดสอบดั้งเดิมที่มีข้อสอบจำนวน 5 ข้อจำนวน 2 ฉบับ กับคะแนนผลการสอบจากแบบทดสอบฉบับหนึ่งที่มีข้อสอบ 100 ข้อ แบบทดสอบดั้งเดิมดำเนินการสอบโดยการเขียนตอบ ส่วนแบบทดสอบปริมาตรดำเนินการสอบด้วยคอมพิวเตอร์ แบบทดสอบปริมาตร ฉบับหนึ่งการเลือกข้อสอบให้ผู้สอบใช้คำอำนาจจำแนกเป็นเกณฑ์ ส่วนอีกฉบับหนึ่งนั้นใช้การพยากรณ์คะแนนสูงสุดเป็นเกณฑ์ แบบทดสอบปริมาตรทั้งสองฉบับจึงมีขนาดของชั้นแปรผัน แบบทดสอบปริมาตรทั้งสองฉบับนำไปสอบกลุ่มตัวอย่างทั้งสองกลุ่มจำนวน 263 คนและแบบทดสอบดั้งเดิมนำไปสอบกลุ่มตัวอย่างอีกกลุ่มหนึ่งจำนวน 250 คน ผลการวิจัยพบว่าแบบทดสอบดั้งเดิมฉบับหนึ่งมีค่าสหสัมพันธ์กับคะแนนรวมสูงกว่าแบบทดสอบปริมาตรทั้งสองฉบับ และแบบทดสอบปริมาตรฉบับหนึ่งมีค่าสหสัมพันธ์กับคะแนนรวมต่ำกว่าแบบทดสอบดั้งเดิมทั้งสองฉบับ

ลาร์กิน และ ไวท์ (Larkin and Weiss : 1974) ได้ศึกษาแบบทดสอบปริมาตร 3 ฉบับกับแบบทดสอบดั้งเดิม 1 ฉบับซึ่งมีข้อสอบที่ปรับค่าการเดาแล้ว 40 ข้อ มีค่าความยากตั้งแต่ -.957 ถึง 1.156 แบบทดสอบปริมาตรแต่ละฉบับเป็นแบบสมมาตรชนิด 15 ชั้น มีค่าความแตกต่างระหว่างค่าความยากของข้อสอบในชั้นที่อยู่ติดกันคงที่ ใช้กฎการแยกทางแบบ เพิ่ม 1/ลด 1 แต่ละฉบับจะมีข้อสอบ 120 ข้อ โดยจัดระดับความยากของข้อสอบออกเป็น 29 ระดับ ช่วง ระดับความยากกระจายตั้งแต่ -2.86 ถึง 2.95 ให้คะแนนแบบทดสอบปริมาตร 6 วิธี คือให้ตามจำนวนข้อที่ตอบถูก ให้ตามค่าความยากเฉลี่ยของข้อสอบที่ได้ตอบ ให้ตามค่าความยากเฉลี่ยของข้อสอบที่ตอบถูก ให้ตามค่าความยากของข้อสอบข้อสุดท้ายที่ได้ตอบ ให้ตามค่าความยากของข้อสอบข้อที่ถัดจากข้อสุดท้ายที่ได้ตอบ และให้คะแนนข้อสอบทุกข้อตามวิธีการของ Hansen ข้อสอบเป็นแบบเลือกตอบชนิด 5 ตัวเลือกเป็นคำถามที่เกี่ยวกับเนื้อหาด้านคำศัพท์ (Vocabulary) ที่อยู่ในระดับต่ำกว่าระดับวิทยาลัย มีกลุ่มข้อสอบ (Item Pool) จำนวน 369 ข้อ ใช้ดำเนินการสอบด้วยคอมพิวเตอร์ โดยแบ่งการทดสอบเป็น 2 ระยะ ดังต่อไปนี้

การทดสอบระยะที่ 1 แบ่งกลุ่มตัวอย่างออกเป็น 3 กลุ่ม กลุ่มที่ 1 และ 2 มีจำนวน 125 คน เท่ากันให้ทำแบบทดสอบปริมาตรฉบับที่ 1 และแบบทดสอบดั้งเดิมเหมือนกัน กลุ่มที่ 3 จำนวน 142 คน ให้ทำแบบทดสอบปริมาตรฉบับที่ 3 และแบบทดสอบปรับระดับชั้น

การทดสอบระยะที่ 2 ทั้งช่วงห่างจากการทดสอบระยะที่ 1 7 สัปดาห์ กลุ่มตัวอย่าง กลุ่มที่ 1 และกลุ่มที่ 2 กลุ่มละจำนวน 125 คน ได้มาโดยการสุ่มจากกลุ่มที่ 1 และกลุ่มที่ 2 ของ ระยะที่ 1 กลุ่มที่ 1 ให้ทำแบบทดสอบปริมาตรฉบับที่ 1 และแบบทดสอบดั้งเดิม กลุ่มที่ 2 ให้ทำ แบบทดสอบปริมาตรฉบับที่ 2 และแบบทดสอบดั้งเดิม ส่วนกลุ่มที่ 3 สุ่มมาจำนวน 138 คนให้ ทำแบบทดสอบปริมาตรฉบับที่ 3 และแบบทดสอบชนิดสองชั้น

ผลการศึกษาพบว่าลำดับที่ของการทำแบบทดสอบมีผลต่อคะแนนผลการสอบทั้งแบบ ทดสอบปริมาตรและแบบทดสอบดั้งเดิมอย่างไม่มีนัยสำคัญทางสถิติ คือถ้าสอบแบบทดสอบ ปริมาตรหลังจากสอบแบบทดสอบดั้งเดิมคะแนนที่ได้จะมีแนวโน้มลดลง ลักษณะการกระจายของ คะแนนผลการสอบแบบทดสอบปริมาตรฉบับที่ 1 และฉบับที่ 2 มีลักษณะใกล้เคียงกับการ กระจายแบบปกติ ส่วนแบบทดสอบปริมาตรฉบับที่ 3 มีลักษณะการกระจายแบบเบ้ไปทางบวก อย่างมีนัยสำคัญทางสถิติ การศึกษาผลของวิธีการให้คะแนนกับลักษณะการกระจายของคะแนน พบว่าการให้คะแนนทั้ง 6 วิธีส่งผลให้การกระจายของคะแนนกับลักษณะการกระจายของ คะแนน ในแบบทดสอบปริมาตรฉบับที่ 1 และฉบับที่ 2 มีความโค้งต่ำกว่าการกระจายแบบปกติ อย่างมีนัยสำคัญ และจากการใช้ร้อยละของช่วงคะแนนในการประมาณระดับความสามารถของผู้สอบ พบว่า การให้คะแนนโดยวิธีให้ตามค่าความยากเฉลี่ยของข้อสอบที่ได้ทำสามารถประมาณระดับ ความสามารถของผู้สอบได้ในช่วงกว้างเท่ากับแบบทดสอบดั้งเดิม และการให้คะแนนข้อสอบทุก ข้อตามวิธีการของ Hansen สามารถประมาณระดับความสามารถของผู้สอบได้ในช่วงกว้างมาก กว่าแบบทดสอบดั้งเดิม เล็กน้อย และการให้คะแนนทั้งสองวิธีนี้มีค่าสหสัมพันธ์สูงเท่ากับ .99 จากการเปรียบเทียบวิธีการให้คะแนนแบบทดสอบปริมาตรทั้ง 6 วิธีพบว่า การให้คะแนนโดยวิธีให้ ตามค่าความยากเฉลี่ยของข้อที่ได้ทำ และการให้คะแนนข้อสอบทุกข้อมีความคงที่ของคะแนน มากที่สุด จากการศึกษาผลของช่วงเวลาระหว่างการสอบแต่ละครั้ง ไม่มีผลต่อความคงที่ของ คะแนนทั้งในแบบทดสอบปริมาตร และแบบทดสอบดั้งเดิม ผลการศึกษาค้นพบว่าการจำไม่มีผลต่อการทำแบบทดสอบปริมาตร และแบบทดสอบดั้งเดิม จากการศึกษาผลของวิธี การให้คะแนนที่มีต่อการเพิ่มของคะแนนผลการสอบในครั้งที่ 2 พบว่าการให้คะแนนทั้ง 6 วิธีไม่ มีผลต่อการเพิ่มของคะแนนผลการสอบในครั้งที่ 2 ของแบบทดสอบปริมาตรฉบับที่ 1 และฉบับที่ 2 เพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติ แต่มีอยู่ 2 วิธีคือการให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบ ที่ได้ทำ และการให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ตอบถูกไม่มีผลต่อการเพิ่มของ คะแนนผลการสอบครั้งที่ 2 จากการศึกษาผลของการให้คะแนนที่มีต่อความเชื่อมั่นชนิดความคง ที่ภายในของแบบทดสอบพบว่า เมื่อใช้ค่าดัชนีของ Hoyt (Hoyt Index) แบบทดสอบดั้งเดิมชนิด 40 ข้อจะมีค่าความเชื่อมั่นเท่ากับ 0.89 และ 0.90 เมื่อดำเนินการสอบครั้งที่ 1 และครั้งที่ 2 ตาม ลำดับ และเมื่อเพิ่มความยาวของแบบทดสอบดั้งเดิมเป็น 120 ข้อตามวิธีการของ Spearman Brown พบว่าความเชื่อมั่นเพิ่มขึ้นเป็น 0.96 ในการทดสอบทั้งสองครั้งพบว่า ค่าความเชื่อมั่น ชนิดความคงที่ภายในของแบบทดสอบปริมาตรที่ให้คะแนนตามวิธีการให้คะแนนข้อสอบทุกข้อจะ

มีค่าสูงสุดถึง 0.99 และจากการศึกษาเปรียบเทียบวิธีการให้คะแนนทั้ง 6 วิธี พบว่า การให้คะแนนโดยวิธีการให้คะแนนข้อสอบทุกข้อในแบบทดสอบปิรามิดฉบับที่ 1 มีความสัมพันธ์กับแบบทดสอบดั้งเดิมสูงกว่าแบบทดสอบปิรามิดฉบับอื่น ๆ คือให้ค่าสหสัมพันธ์สูงถึง 0.86 คะแนนจากแบบทดสอบปิรามิดฉบับที่ 1 และ ฉบับที่ 2 ที่ให้คะแนนตามจำนวนข้อที่ตอบถูก กับ การให้คะแนนตามค่าความยากของข้อสอบข้อที่ถัดจากข้อสอบข้อสุดท้ายที่ตอบ มีความสัมพันธ์กับคะแนนจากแบบทดสอบดั้งเดิมต่ำที่สุด เมื่อพิจารณาความสัมพันธ์ของคะแนนที่ได้จากการให้คะแนนวิธีต่าง ๆ จากผลการทดสอบทั้งสองครั้งพบว่า การให้คะแนนโดยวิธีให้ตามค่าความยากเฉลี่ยของข้อสอบที่ได้ทำกับการให้คะแนนวิธีให้คะแนนข้อสอบทุกข้อมีความสัมพันธ์กันสูงมาก ส่วนการให้คะแนนวิธีให้ตามจำนวนข้อที่ตอบถูกมีความสัมพันธ์กับการให้คะแนนตามค่าความยากของข้อสอบข้อที่ถัดจากข้อสอบข้อสุดท้ายที่ตอบสูงที่สุด คือให้ค่าสหสัมพันธ์เท่ากับ 1.00 และการให้คะแนนวิธีให้คะแนนข้อสอบทุกข้อจะมีความสัมพันธ์กับคะแนนที่ให้โดยวิธีอื่น ๆ

นอกจากนี้ ลาร์กิน และ ไวท์ (Larkin and Weiss : 1975) ได้ศึกษาเปรียบเทียบแบบทดสอบชนิด 2 ชั้น กับแบบทดสอบปิรามิดที่สร้างมาจากกลุ่มข้อสอบ จำนวน 369 ข้อ เป็นข้อสอบแบบเลือกตอบชนิด 5 ตัวเลือก ซึ่งถามเนื้อหาเกี่ยวกับคำศัพท์ที่อยู่ในระดับต่ำกว่าระดับวิทยาลัย ข้อสอบได้วิเคราะห์หาค่าพารามิเตอร์ a_g (Discrimination) และ b_g (Difficulty) ตามทฤษฎีคุณลักษณะแฝง ข้อสอบแต่ละข้อมีค่าสัมประสิทธิ์สหสัมพันธ์แบบ Biserial ตั้งแต่ .30 ขึ้นไป แบบทดสอบชนิดสองชั้นที่สร้างขึ้นนี้ประกอบด้วยแบบทดสอบแยกทาง (Routing Test) และแบบทดสอบวัดผล (Measurement Tests) แบบทดสอบแยกทางประกอบด้วยข้อสอบ 10 ข้อ มีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าความยากของข้อสอบ (b_g) เท่ากับ -0.232 และ 0.050 มีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าอำนาจจำแนกของข้อสอบ (a_g) เท่ากับ 0.702 และ 0.163 ตามลำดับ ผู้สอบที่ทำแบบทดสอบแยกทางได้คะแนน 0-4, 5-6, 7-8, และ 9-10 คะแนนจะได้ตอบแบบทดสอบวัดผลฉบับที่ง่ายที่สุดหรือยากขึ้นสุดในชุดหนึ่งตามลำดับของคะแนน แบบทดสอบวัดผลซึ่งมีอยู่ 4 ฉบับ ๆ ละ 30 ข้อ มีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าความยากของข้อสอบในฉบับที่ 1 ถึง 4 เท่ากับ 1.825 และ 0.558, 0.350 และ 0.297, -0.709 และ 0.189, -1.603 และ 0.373 ตามลำดับ มีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าอำนาจจำแนกของข้อสอบในฉบับที่ 1 ถึง 4 เท่ากับ 0.530 และ 0.126, 0.684 และ 0.214, 0.611 และ 0.122, 0.693 และ 0.2132 ตามลำดับ การให้คะแนนแบบทดสอบชนิดสองชั้นใช้วิธีประมาณค่าความสามารถ (Ability Estimate) ของผู้สอบตามวิธีของ Lord

วอร์เทอร์ และ เบย์รอฟฟ์ (Waters and Bayroff : 1971) ได้ศึกษาเปรียบเทียบแบบทดสอบปิรามิดชนิด 5, 10 และ 15 ชั้นที่มีข้อสอบชั้นละ 1 ข้อ และชนิด 10 ชั้นที่มีข้อสอบชั้นละ 2 ข้อ กับแบบทดสอบดั้งเดิมที่มีจำนวนข้อสอบเท่ากัน ทั้งแบบทดสอบปิรามิดและแบบทดสอบ

ดั้งเดิมมีข้อสอบที่มีระดับความยากและค่าอำนาจจำแนกแตกต่างกันอย่างมีระบบ และมีข้อสมมติว่าการแจกแจงของความสามารถเป็นปกติ ผลการศึกษาพบว่า คะแนนผลการสอบและระดับความสามารถมีความสัมพันธ์กับการแจกแจงของความยากและค่าอำนาจจำแนกของแบบทดสอบ ซึ่งแบบทดสอบปิรามิดจะมีความสัมพันธ์สูงกว่าในแบบทดสอบดั้งเดิมโดยเฉพาะอย่างยิ่งข้อที่มีค่าอำนาจจำแนกสูง ๆ และในแบบทดสอบปิรามิดชนิดที่มีข้อสอบ 1 ข้อในแต่ละชั้นจะมีความสัมพันธ์ระหว่างคะแนนรวมกับระดับความสามารถสูงมาก

ได้มีผู้ทำการศึกษาเกี่ยวกับแบบทดสอบเทเลอร์รูปปิรามิดในหลายรูปแบบซึ่งโดยส่วนใหญ่เป็นการศึกษาเกี่ยวกับแบบทดสอบเทเลอร์รูปปิรามิดในรูปแบบขนาดชั้นคงที่ และเป็นการดำเนินการสอบโดยใช้วิธีการสอบในกระดาษคำตอบ ในการศึกษาครั้งนี้ผู้วิจัยได้ทำการศึกษาในเรื่องของความยาวของแบบทดสอบ และจำนวนข้อของข้อสอบในแต่ละชั้นว่ามีผลต่อความเที่ยงตรงของแบบทดสอบ และประสิทธิภาพสัมพันธ์ของแบบทดสอบแต่ละฉบับหรือไม่ โดยได้ทำการเลือกแบบทดสอบเทเลอร์รูปปิรามิดที่มีขนาดชั้นคงที่ แบบ 5 และ 15 ชั้นแต่ละชั้นทำข้อสอบชั้นละ 1 ข้อ และแบบทดสอบที่มีหลายข้อในแต่ละชั้น โดยในแต่ละชั้นผู้สอบจะต้องทำข้อสอบชั้นละ 3 ข้อ ซึ่งแบบทดสอบที่มีหลายข้อในแต่ละชั้นจะมีจุดเด่นในเรื่องของค่าอำนาจจำแนก โดยทำการดำเนินการสอบโดยใช้คอมพิวเตอร์มาช่วยในการดำเนินการสอบ

บทที่ 3

วิธีดำเนินการศึกษาค้นคว้า

ขอบเขตของการศึกษาค้นคว้า

ประชากร

การศึกษาค้นคว้าครั้งนี้ใช้กับกลุ่มประชากรที่เป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ในภาคเรียนที่ 2 ปีการศึกษา 2546 ของโรงเรียนสังกัดเขตพื้นที่การศึกษาราชบุรี เขต 1 และสังกัดเทศบาลเมืองราชบุรี ในเขตอำเภอเมืองราชบุรี จังหวัดราชบุรี จำนวนโรงเรียน 66 โรงเรียน จำนวนห้องเรียน 109 ห้องเรียน จำนวนนักเรียน 3,165 คน

กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้านี้เป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ในภาคเรียนที่ 2 ปีการศึกษา 2546 ของโรงเรียนสังกัดเขตพื้นที่การศึกษาราชบุรี เขต 1 และสังกัดเทศบาลเมืองราชบุรี ในเขตอำเภอเมืองราชบุรี จังหวัดราชบุรี จำนวนโรงเรียน 10 โรงเรียน จำนวนห้องเรียน 37 ห้องเรียน จำนวนนักเรียน 1,492 คน ซึ่งเลือกมาโดยการสุ่มแบบกลุ่ม (Cluster Random Sampling) มีโรงเรียนเป็นหน่วยการสุ่ม (Sampling Unit) ได้จำนวน 10 โรงเรียน จำนวน 37 ห้องเรียน จำนวนนักเรียน 1,492 คน โดยสุ่มแบ่งกลุ่มตัวอย่างออกเป็น 2 กลุ่มคือ

1. กลุ่มตัวอย่างที่ใช้ในการสร้างเครื่องมือและตรวจสอบคุณภาพ จำนวน 27 ห้องเรียน จำนวนนักเรียน 1,128 คน
2. กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ จำนวน 10 ห้องเรียน จำนวนนักเรียน 364 คน ดังแสดงในตาราง 1

ตาราง 1 แสดงจำนวนนักเรียนในกลุ่มตัวอย่างที่ใช้ในการสร้างเครื่องมือ และในการวิจัยจำแนกตามสังกัด

โรงเรียน	จำนวนนักเรียน (ห้อง) ที่ใช้ในการสอบ	
	การสร้างเครื่องมือ	การวิจัย
1. แคราย	65 (2)	-
2. ห้วยไผ่	44 (1)	-
3. วัดเขาวัง	399 (10)	-
4. อนุบาลราชบุรี	358 (8)	-
5. บางลี่	57 (1)	-
6. เทศบาล 1 (วัดสัตตนาปวิตร)	37 (1)	71 (2)
7. เทศบาล 2 (วัดช่องลม)	37 (1)	71 (2)
8. เทศบาล 3 (เทศบาลสงเคราะห์)	42 (1)	72 (2)
9. เทศบาล 4 (วัดมหาธาตุวรวิหาร)	45 (1)	77 (2)
10. เทศบาล 5 (พหลโยธินรามินทรภักดี)	44 (1)	73 (2)
รวม	1,128 (32)	364 (10)

เครื่องมือที่ใช้ในการศึกษาค้นคว้า

เครื่องมือที่ใช้ในการศึกษาค้นคว้า เป็นแบบทดสอบที่ดำเนินการสอบด้วยคอมพิวเตอร์ ซึ่งมีลักษณะดังนี้

1. แบบทดสอบรูปปริมาตร

แบบทดสอบปริมาตรเป็นแบบทดสอบที่ผู้วิจัยสร้างขึ้นตามรูปแบบจำลองรูปปริมาตรเป็นข้อสอบแบบเลือกตอบชนิด 4 ตัวเลือก ในวิชาคณิตศาสตร์ เรื่อง การบวก ลบ และคูณทศนิยม มีจำนวน 3 แบบ ดังนี้

1.1 แบบทดสอบรูปปริมาตรที่มีขนาดชั้นคงที่ ใช้กฎการจำแนกแนวทางการตอบข้อสอบแบบ เพิ่ม 1หรือลด 1 แบบ 5 ชั้น มีจำนวนข้อสอบทั้งหมด 15 ข้อ นักเรียนจะต้องทำข้อสอบในแต่ละชั้น ชั้นละ 1 ข้อ รวมนักเรียนจะต้องทำข้อสอบทั้งสิ้น 5 ข้อ

1.2 แบบทดสอบรูปปริมาตรที่มีขนาดชั้นคงที่ ใช้กฎการจำแนกแนวทางการตอบข้อสอบแบบ เพิ่ม 1หรือ ลด 1 แบบ 15 ชั้น มีจำนวนข้อสอบทั้งหมด 120 ข้อ นักเรียนจะต้องทำข้อสอบในแต่ละชั้น ชั้นละ 1 ข้อ รวมนักเรียนจะต้องทำข้อสอบทั้งสิ้น 15 ข้อ

1.3 แบบทดสอบรูปปริมาตรที่มีหลายข้อในแต่ละชั้น เป็นแบบทดสอบชนิด 5 ชั้น

แบบที่มีข้อสอบ 3 ข้อในแต่ละชั้น แนวทางการตอบข้อสอบแบบ เพิ่ม 1 หรือลด 1 มีจำนวนข้อสอบทั้งสิ้น 45 ข้อ

2. . วิธีการให้คะแนนของแบบทดสอบปิรามิด 3 วิธีคือ

2.1 การให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ทำ

2.2 การให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ตอบถูก

2.3 การให้คะแนนค่าความยากตามข้อสอบที่ถัดจากข้อสุดท้ายที่ได้ตอบ

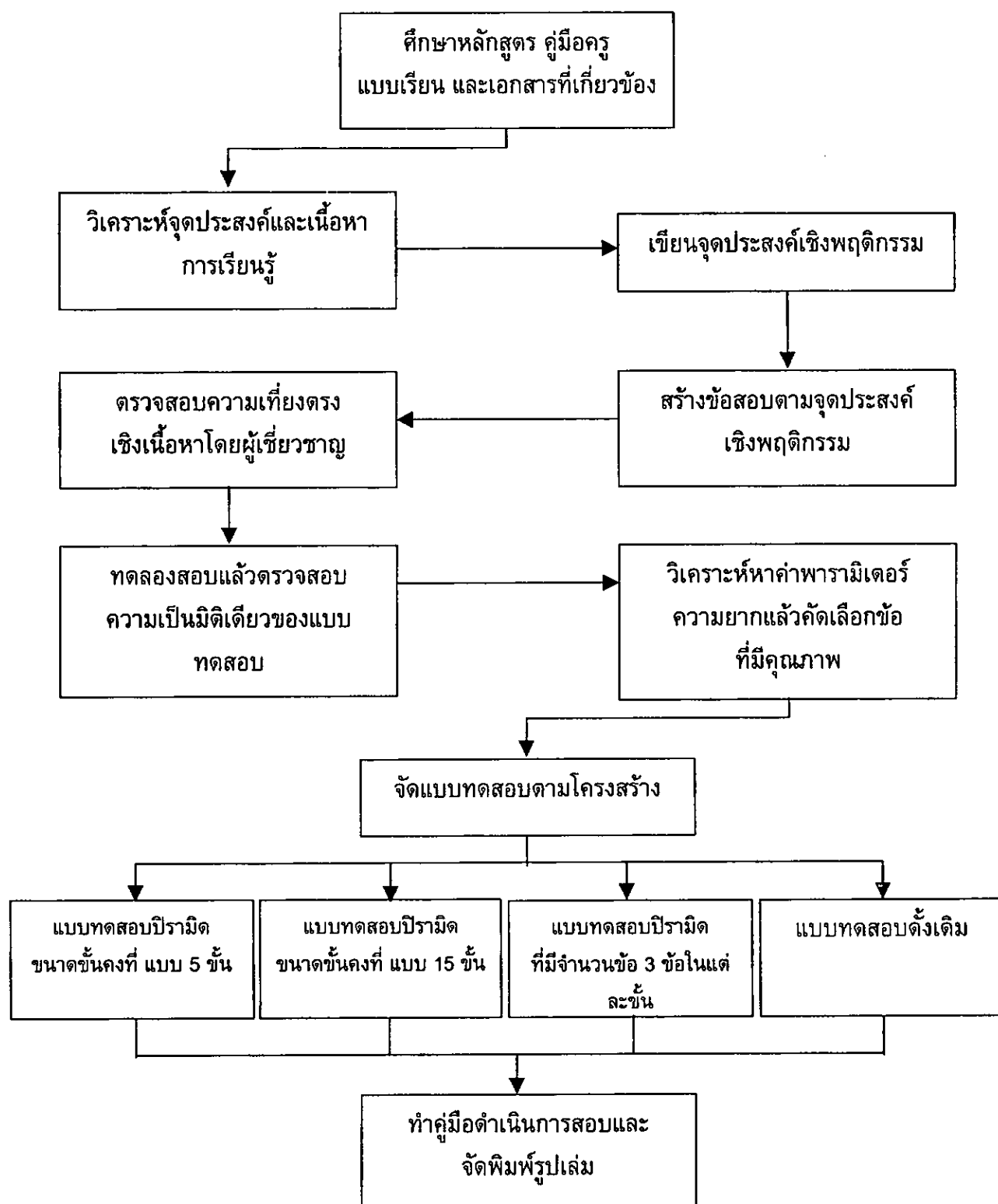
3. แบบทดสอบดั้งเดิม

แบบทดสอบดั้งเดิมเป็นแบบทดสอบที่ผู้วิจัยสร้างขึ้น เป็นข้อสอบแบบเลือกตอบชนิด

4 ตัวเลือก จำนวน 30 ข้อ วิชาคณิตศาสตร์ เรื่อง การบวก ลบ และคูณทศนิยม ระดับชั้นประถมศึกษาปีที่ 6 โดยมีช่วงระดับความยากของข้อสอบ และโค้งอินฟอร์เมชันของแบบทดสอบใกล้เคียงกับแบบทดสอบรูปปิรามิด

วิธีดำเนินการสร้างแบบทดสอบ

การสร้างแบบทดสอบดั้งเดิม และแบบทดสอบเทเลอร์รูปปิรามิด ผู้วิจัยสร้างตามแผนผัง
ดังนี้



ภาพประกอบ 11 แผนผังแสดงขั้นตอนการดำเนินการสร้างแบบทดสอบ

รายละเอียดในการสร้างแบบทดสอบ มีดังนี้

1. การสร้างกลุ่มข้อสอบ

- 1.1 ศึกษาเนื้อหาจากหลักสูตร คู่มือครู แบบเรียน และเอกสารที่เกี่ยวข้องในวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 เรื่อง การบวก ลบ และคูณทศนิยม
- 1.2 วิเคราะห์จุดประสงค์การเรียนรู้และเนื้อหา เพื่อนำมาเขียนเป็นจุดประสงค์เชิงพฤติกรรม โดยสามารถเขียนได้ 6 จุดประสงค์ ได้แก่
 - เมื่อกำหนดทศนิยมให้ สามารถอ่านและเขียนในรูปกระจายตามค่าประจำหลักได้
 - เมื่อกำหนดทศนิยมไม่เกินสามตำแหน่ง ตั้งแต่สองจำนวนขึ้นไปให้สามารถเปรียบเทียบและเรียงลำดับจากทศนิยมที่มีค่าน้อยไปหาทศนิยม ที่มีค่ามาก หรือจากทศนิยมที่มีค่ามากไปหาทศนิยมที่มีค่าน้อยได้
 - เมื่อกำหนดทศนิยมให้ สามารถเขียนในรูปเศษส่วนได้หรือเมื่อกำหนดเศษส่วนที่มีส่วนเป็น 10, 100 และ 1,000 ให้ สามารถเขียนในรูปทศนิยมได้
 - เมื่อกำหนดโจทย์ปัญหาเกี่ยวกับการบวกและการลบทศนิยมที่มีผลลัพธ์ไม่เกินสามตำแหน่งให้สามารถแสดงวิธีทำและหาคำตอบได้
 - เมื่อกำหนดจำนวนนับกับทศนิยม หรือทศนิยมกับทศนิยมที่ผลคูณเป็นทศนิยมไม่เกินสามตำแหน่งให้ สามารถหาผลคูณได้
 - เมื่อกำหนดโจทย์ปัญหาเกี่ยวกับการคูณทศนิยม ที่ผลคูณเป็นทศนิยมไม่เกินสามตำแหน่งให้สามารถแสดงวิธีทำและหาคำตอบได้

ตาราง 2 แสดงค่าน้ำหนักความสำคัญของจุดประสงค์เชิงพฤติกรรม วิชาคณิตศาสตร์ เรื่อง การบวก ลบ และคูณทศนิยม

จุดประสงค์	ค่าน้ำหนักความสำคัญ	จำนวนข้อสอบ
1	1	4
2	1	4
3	1	4
4	3	12
5	2	8
6	2	8

1.3 สร้างข้อสอบเลือกตอบ ชนิด 4 ตัวเลือก โดยสร้างตามจุดประสงค์เชิงพฤติกรรม ซึ่งได้กำหนดค่าน้ำหนักความสำคัญของแต่ละจุดประสงค์ สร้างข้อสอบได้ทั้งหมด จำนวน 240 ข้อ โดยมีข้อสอบฉบับละ 40 ข้อ จำนวน 6 ฉบับ

1.4 นำข้อสอบที่สร้างขึ้น 240 ข้อ ไปให้ผู้เชี่ยวชาญซึ่งมีความรู้ในเนื้อหาและการเขียนข้อสอบ จำนวน 5 คน เป็นผู้ตรวจสอบความเที่ยงตรงเชิงเนื้อหา (Content Validity) โดยพิจารณาความสอดคล้องระหว่างข้อสอบแต่ละข้อกับจุดประสงค์เชิงพฤติกรรม ข้อสอบที่ถือว่า วัดได้ตรงตามจุดประสงค์เชิงพฤติกรรมจะต้องมีค่า $IOC \geq .50$ จากความเห็นชอบจากผู้เชี่ยวชาญ สำหรับแบบประเมินความสอดคล้องมีลักษณะดังตัวอย่างต่อไปนี้

ตัวอย่าง แบบประเมินความสอดคล้องระหว่างข้อสอบกับจุดประสงค์เชิงพฤติกรรม
คำชี้แจง : ให้ท่านพิจารณาว่าข้อสอบต่อไปนี้วัดได้ตรงตามจุดประสงค์เชิงพฤติกรรมที่กำหนดไว้หรือไม่ โดยเขียนเครื่องหมาย ✓ ลงในช่องแสดงความคิดเห็นซึ่งพิจารณาให้คะแนนตามเกณฑ์ดังต่อไปนี้

- ให้ 1 คะแนน เมื่อท่านแน่ใจว่าข้อสอบวัดได้ตรงตามจุดประสงค์เชิงพฤติกรรม
- ให้ 0 คะแนน เมื่อท่านไม่แน่ใจว่าข้อสอบวัดได้ตรงตามจุดประสงค์เชิงพฤติกรรม
- ให้ -1 คะแนน เมื่อท่านแน่ใจว่าข้อสอบวัดได้ไม่ตรงตามจุดประสงค์เชิงพฤติกรรม

จุดประสงค์เชิงพฤติกรรม	ข้อสอบ	คะแนนการพิจารณา		
		1	0	-1
0. เมื่อกำหนดโจทย์การลบ ทศนิยมไม่เกินสอง ตำแหน่ง ให้สองจำนวน สามารถหาผลลบได้	0. 0.37 – 0.2 มีค่าตรงกับข้อใด?			
	ก. 0.35			
	ข. 0.39			
	ค. 1.70			
	ง. 0.17			
	เฉลย ง. 0.17			
	00. 30.03 – 7.57 มีค่าตรงกับข้อใด?			
	ก. 21.46			
	ข. 22.46			
	ค. 23.46			
	ง. 24.46			
	เฉลย ข. 22.46			

จากผลการตรวจสอบความเที่ยงตรงเชิงเนื้อหาของผู้เชี่ยวชาญผลปรากฏว่า ในข้อสอบ 240 ข้อ ได้รับการพิจารณาว่ามีความสอดคล้องกับจุดประสงค์เชิงพฤติกรรม จำนวน

224 ข้อ ซึ่งแสดงว่ามีข้อสอบที่จะต้องทำการปรับปรุงแก้ไข จำนวน 26 ข้อ เพื่อให้ได้ข้อสอบทั้งสิ้น 240 ข้อ

1.5 ปรับปรุงข้อสอบตามที่คุณเชี่ยวชาญได้พิจารณาไว้ แล้วนำไปให้ผู้เชี่ยวชาญพิจารณาอีกครั้งหนึ่ง แล้วจึงนำมาจัดเป็นแบบทดสอบ 6 ฉบับ ๆ ละ 40 ข้อ ซึ่งจัดโดยการสุ่มข้อสอบในแต่ละจุดประสงค์เชิงพฤติกรรม

1.6 นำแบบทดสอบทั้ง 6 ฉบับไปทดสอบกับกลุ่มตัวอย่างที่ใช้ในการวิเคราะห์หาค่าพารามิเตอร์ความยากของข้อสอบ โดยนำไปทดลองสอบในระหว่างวันที่ 6 มกราคม พ.ศ. 2547 ถึง วันที่ 30 มกราคม พ.ศ. 2547 กับกลุ่มตัวอย่างที่ใช้ในการวิเคราะห์หาค่าพารามิเตอร์ความยากของข้อสอบ

1.7 ผู้วิจัยได้สุ่มนักเรียนเพื่อทำแบบทดสอบแต่ละฉบับโดยการจัดการดำเนินการสอบให้เป็นระบบดังนี้ คือ ให้นักเรียนคนที่ 1, 2 และ 3 ทำแบบทดสอบฉบับที่ 1 และ 2, 3 และ 4, 5 และ 6 ตามลำดับ และนักเรียนคนที่ 4, 5 และ 6 ทำแบบทดสอบฉบับที่ 1 และ 2, 3 และ 4, 5 และ 6 ตามลำดับเหมือนเดิม และคนต่อ ๆ ไปก็จะได้ทำแบบทดสอบในลักษณะนี้ ซึ่งจะให้นักเรียนทำแบบทดสอบแต่ละฉบับมีจำนวนนักเรียนใกล้เคียงกัน คือ

ฉบับที่ 1	มีนักเรียนทำแบบทดสอบ จำนวน	375	คน
ฉบับที่ 2	มีนักเรียนทำแบบทดสอบ จำนวน	375	คน
ฉบับที่ 3	มีนักเรียนทำแบบทดสอบ จำนวน	376	คน
ฉบับที่ 4	มีนักเรียนทำแบบทดสอบ จำนวน	376	คน
ฉบับที่ 5	มีนักเรียนทำแบบทดสอบ จำนวน	376	คน
ฉบับที่ 6	มีนักเรียนทำแบบทดสอบ จำนวน	376	คน

1.8 ตรวจข้อสอบแต่ละข้อโดยแยกพิจารณาตามรายข้อ โดยให้ตอบถูกได้ 1 ข้อที่ตอบผิดหรือตอบเกินหนึ่งคำตอบหรือไม่ตอบได้ 0

1.9 นำผลที่ได้จากข้อ 1.8 มาวิเคราะห์องค์ประกอบด้วยโปรแกรมสำเร็จรูป SPSS เพื่อตรวจสอบว่าแบบทดสอบวัดความสามารถหรือคุณลักษณะเดียวกัน (Unidimensionality) การวิเคราะห์ตัวประกอบสำคัญ (Principal Component) แล้วหมุนแกนด้วยวิธีออบลิมิน (Oblimin)

1.10 วิเคราะห์ข้อสอบรายข้อเพื่อหาค่าความยาก ของข้อสอบตามทฤษฎีการตอบสนองข้อสอบ แบบราสซัส โมเดล ด้วยวิธีการประมาณค่าความเป็นไปได้สูงสุดแบบไม่มีเงื่อนไข โดยใช้โปรแกรมไบคาล

1.11 คัดเลือกข้อสอบที่มีคุณภาพ โดยมีค่าความยากอยู่ในช่วง -3.00 ถึง 3.00 และมีค่าความสอดคล้องกับโมเดลอยู่ในช่วง -2.00 ถึง 2.00 โดยได้ผลดังนี้

1.11.1 แบบทดสอบรูปปริมาตรที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น มีค่าความยากอยู่ระหว่าง -2.98 ถึง 2.80

1.11.2 แบบทดสอบรูปปริมาตรที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้น มีค่าความยากอยู่ระหว่าง -2.57 ถึง 2.97

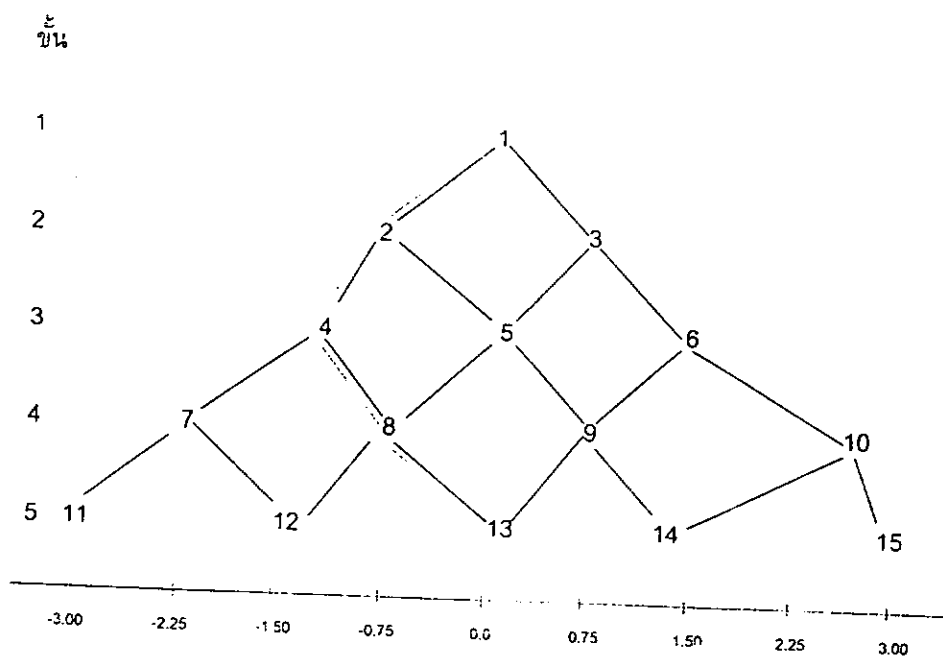
1.11.3 แบบทดสอบรูปปริมาตรที่มีข้อสอบหลายข้อในแต่ละชั้น จำนวน 5 ชั้น มีค่าความยากเฉลี่ยอยู่ระหว่าง -2.51 ถึง 2.75

2. การสร้างแบบทดสอบตามโครงสร้าง

การสร้างแบบทดสอบรูปปริมาตรทั้ง 3 แบบ โดยทำการสุ่มข้อสอบที่ได้ในข้อ 1.11 ดังนี้

2.1 การสร้างแบบทดสอบรูปปริมาตรที่มีขนาดชั้นคงที่ จำนวน 5 ชั้น จะมีจำนวนข้อสอบในแต่ละชั้นเท่ากับลำดับที่ของชั้น คือ ในชั้นที่ 1 จะมีข้อสอบ 1 ข้อ ในชั้นที่ 5 ก็จะมีข้อสอบ 5 ข้อ แกนนอนแสดงถึงค่าความยากของข้อสอบที่สัมพันธ์กับข้อสอบในแบบทดสอบปริมาตร ค่าความยากของข้อสอบอยู่ในช่วงตั้งแต่ -3.0 ถึง 3.0 ข้อสอบที่อยู่ในแนวตั้งเดียวกันจะเป็นข้อสอบที่มีระดับความยากเท่ากัน และข้อสอบที่ผู้สอบจะตอบในชั้นที่อยู่ติดกันจะมีความแตกต่างระหว่างค่าความยากเท่ากัน คือ เท่ากับ 0.75 หรือมีขนาดชั้นคงที่นั่นเอง

ในการทำแบบทดสอบ ผู้สอบจะต้องทำข้อสอบชั้นละหนึ่งข้อ โดยเริ่มทำข้อที่อยู่บนยอดสามเหลี่ยมซึ่งมีความยากปานกลางก่อน ในรูปได้แก่ข้อ 1 มีค่าความยากเท่ากับ 0 เมื่อทำในชั้นที่ 1 เสร็จจะไปทำข้อสอบในชั้นต่อไป ซึ่งผู้สอบจะทำข้อสอบชั้นละ 1 ข้อ ถ้าทำถูกชั้นต่อไปจะแยกไปทำข้อที่ยากขึ้น แต่ถ้าทำผิดชั้นต่อไปจะไปทำข้อที่ง่ายขึ้น กระบวนการในการทำข้อสอบจะเป็นเช่นนี้จนถึงชั้นสุดท้าย

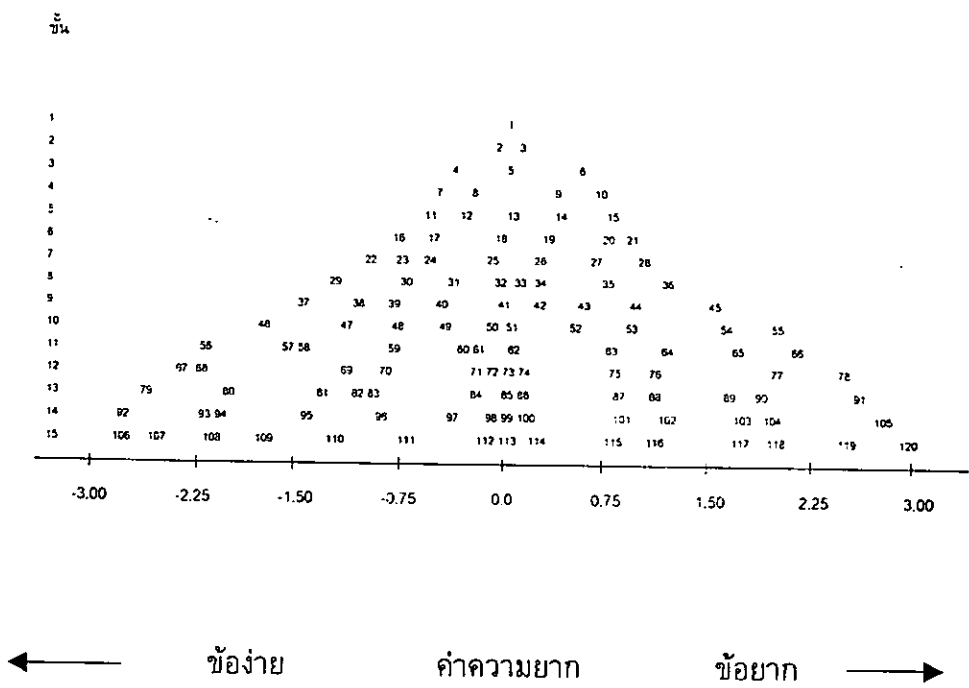


← ข้อย่าง ค่าความยาก ข้อยาก →
ภาพประกอบ 12 แสดงโครงสร้างของแบบทดสอบรูปปิรามิดที่มีขนาดชั้นคงที่ 5 ชั้น

ตาราง 3 แสดงค่าความยาก (b) ของแบบทดสอบรูปปิรามิดที่มีขนาดชั้นคงที่ 5 ชั้น

ข้อที่	b	ข้อที่	b	ข้อที่	b
1	0.03	6	1.42	11	-2.98
2	-0.79	7	-2.26	12	-1.48
3	0.75	8	-0.75	13	0.02
4	-1.31	9	0.73	14	1.38
5	0.07	10	2.48	15	2.80

2.2 การสร้างแบบทดสอบรูปปิรามิดที่มีขนาดชั้นคงที่ จำนวน 15 ชั้น จะมีรูปแบบของแบบทดสอบและวิธีการทำเหมือนกับข้อ 2.1 แต่จะมีจำนวนชั้นของแบบทดสอบจำนวน 15 ชั้น ดังภาพประกอบ 13



ภาพประกอบ 13 แสดงโครงสร้างของแบบทดสอบรูปปิรามิดที่มีขนาดชั้นคงที่ 15 ชั้น

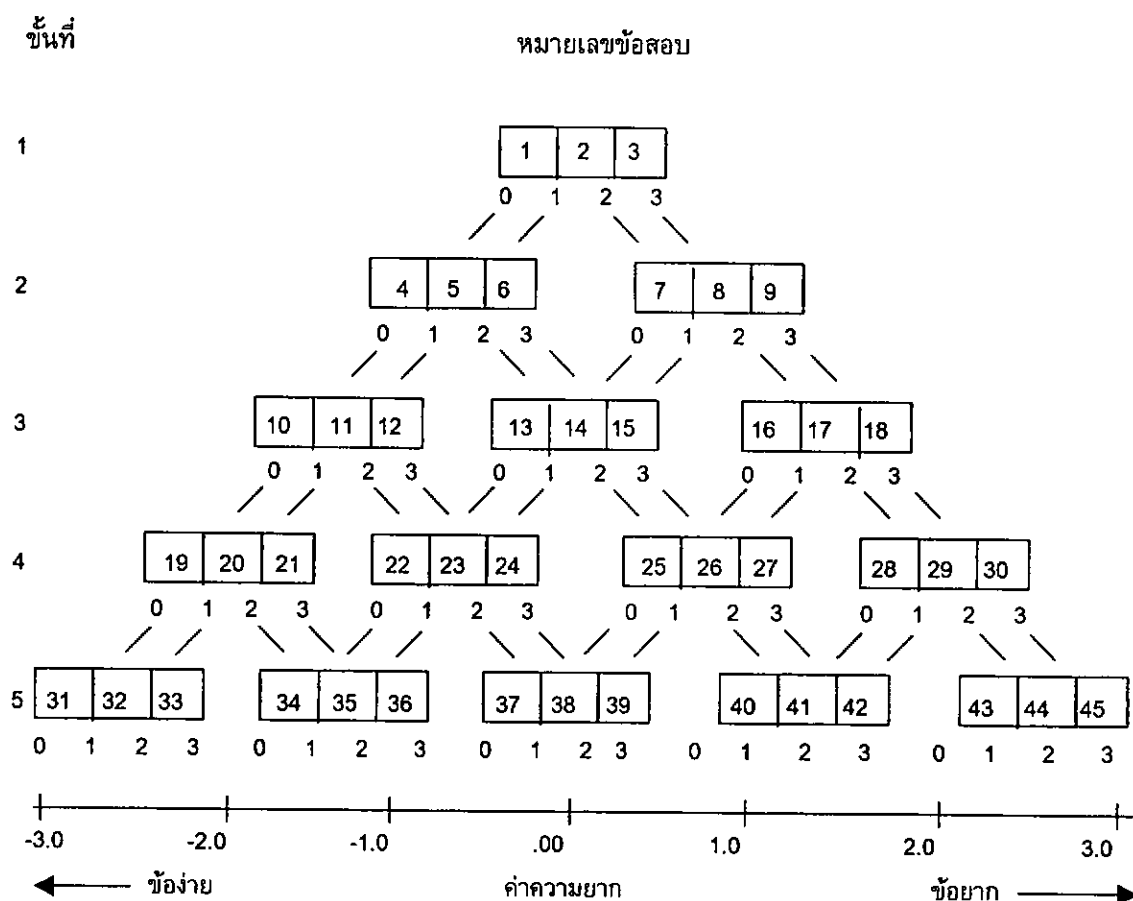
ตาราง 4 แสดงค่าความยาก (b) ของแบบทดสอบรูปปิรามิดที่มีขนาดชั้นคงที่ 15 ชั้น

ข้อที่	b	ข้อที่	b	ข้อที่	b
1	0.3	41	-0.06	81	-1.41
2	-0.02	42	0.25	82	-0.91
3	0.12	43	0.66	83	-0.78
4	-0.46	44	0.92	84	-0.24
5	-0.02	45	1.51	85	0.09
6	0.62	46	-1.76	86	0.12
7	-0.62	47	-1.38	87	0.78
8	-0.31	48	-0.80	88	1.19
9	0.39	49	-0.54	89	1.55
10	0.67	50	-0.16	90	1.86
11	-0.67	51	0.09	91	2.48
12	-0.40	52	0.54	92	-2.57
13	-0.01	53	0.84	93	-2.12
14	0.40	54	1.59	94	-1.94
15	0.77	55	2.01	95	-1.34
16	-0.79	56	-2.27	96	-0.87
17	-0.60	57	-1.64	97	-0.41
18	-0.12	58	-1.48	98	-0.08
19	0.30	59	-0.82	99	0.08
20	0.71	60	-0.21	100	0.14
21	0.91	61	-0.07	101	0.79
22	-0.87	62	0.14	102	1.33
23	-0.77	63	0.79	103	1.76
24	0.68	64	1.12	104	1.92
25	-0.07	65	1.67	105	2.51
26	0.20	66	2.12	106	-2.57
27	0.68	67	-2.37	107	-2.43
28	0.90	68	-2.23	108	-1.91
29	-1.26	69	-1.14	109	-1.54

ตาราง 4 (ต่อ)

ข้อที่	b	ข้อที่	b	ข้อที่	b
30	-0.75	70	-0.86	110	-1.36
31	-0.46	71	0.20	111	-0.68
32	-0.09	72	-0.07	112	-0.10
33	0.08	73	0.05	113	0.08
34	0.25	74	0.15	114	0.23
35	0.77	75	0.84	115	0.77
36	1.31	76	1.39	116	0.98
37	-1.47	77	2.03	117	1.72
38	-1.01	78	2.38	118	1.92
39	-0.79	79	-2.57	119	2.38
40	-0.55	80	-2.14	120	2.97

2.3 การสร้างแบบทดสอบรูปปรีามิตที่มีหลายข้อในแต่ละชั้น เป็นแบบทดสอบชนิด 5 ชั้น แบบที่มีข้อสอบ 3 ข้อในแต่ละชั้น จะใช้ผลการตอบข้อสอบทุกข้อในแต่ละชั้นเป็นเกณฑ์ในการกำหนดการแยกทางเพื่อทำข้อสอบในชั้นถัดไป



ภาพประกอบ 14 แสดงโครงสร้างของการทดสอบเทเลอรรูปปริามิต แบบมีข้อสอบหลายข้อในแต่ละชั้น (Multiple-Item Model)

ตาราง 5 แสดงค่าเฉลี่ยความยาก ($\bar{b}$) ของแบบทดสอบรูปปริมาตรที่มีข้อสอบ 3 ข้อในแต่ละชั้น

ข้อที่	$\bar{b}$	ข้อที่	$\bar{b}$	ข้อที่	$\bar{b}$
1 2 3	0.00	16 17 18	1.51	31 32 33	-2.51
4 5 6	-0.26	19 20 21	-2.17	34 35 36	-1.47
7 8 9	0.68	22 23 24	-0.77	37 38 39	0.01
10 11 12	-1.52	25 26 27	0.74	40 41 42	-1.51
13 14 15	-0.01	28 29 30	2.21	43 44 45	2.75

ภาพประกอบ 14 ในขั้นที่ 1 ถ้าผู้สอบตอบข้อสอบถูก 0 ถึง 1 ข้อ ในขั้นที่ 2 จะแยกไปทำข้อสอบข้อที่ 4, 5 และ 6 และถ้าตอบข้อสอบถูก 2 ถึง 3 ข้อ ก็จะแยกไปทำข้อสอบข้อที่ 7, 8 และ 9 สำหรับในขั้นต่อไปทำเช่นเดียวกัน

ตาราง 6 แสดงค่าความยากของข้อสอบ (b) และผลการตอบของผู้สอบในภาพประกอบ 12

ขั้นที่	หมายเลขข้อสอบที่ทำ	ค่าความยากของข้อสอบ	ผลการตอบ
1	1	0.30	ผิด
2	2	-0.79	ผิด
3	4	-1.31	ถูก
4	8	-.75	ถูก
5	13	0.2	ถูก

การประมาณค่าความสามารถของผู้สอบโดยใช้วิธีการให้คะแนน 3 รูปแบบ ดังนี้
วิธีที่ 1 ให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบข้อที่ทำ โดยวิธีนี้จะไม่นำค่าความยากของ

ข้อสอบข้อที่ 1 มาคิด แต่จะให้คะแนนแต่ละข้อใช้ค่าความยากของข้อสอบข้อที่ถัดจากข้อสุดท้าย (N + 1th Item) ที่ได้ตอบที่สมมติขึ้นมามีค่าเท่ากับ 0

ค่าความยากเฉลี่ยของข้อสอบข้อที่ทำ = $((0.30) + (-0.79) + (-1.31) + (-0.75) + 0.20) / 5$
= -0.47

ฉะนั้นความสามารถของผู้สอบมีค่าเท่ากับ -0.47

วิธีที่ 2 ให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบข้อที่ทำถูก โดยที่ผู้สอบทำข้อสอบถูกทั้งหมด 3 ข้อ ได้แก่ ข้อ 3, 4 และ 5 ดังนี้

ค่าความยากเฉลี่ยของข้อสอบข้อที่ทำถูก = $((-1.50) + (-0.75) + 0.00) / 3$
= -0.75

ฉะนั้นความสามารถของผู้สอบมีค่าเท่ากับ -0.75

วิธีที่ 3 ให้คะแนนค่าความยากตามข้อสอบที่ถัดจากข้อสุดท้ายที่ได้ตอบ ผู้สอบทำข้อสอบข้อสุดท้ายคือข้อ 13 ซึ่งมีค่าความยากเท่ากับ 0.02 ได้ถูกต้อง การคะแนนวิธีนี้จะให้คะแนนโดยสมมติให้คะแนนตามค่าความยากข้อสอบข้อถัดไปที่สมมติขึ้น ในกรณีนี้คือข้อ 19

4. นำผลที่ได้จากการสอบ มาจัดเตรียมข้อมูลสำหรับการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ แล้วนำผลการประมาณค่าไปวิเคราะห์ทางสถิติ และทดสอบสมมติฐานของการวิจัย

สถิติที่ใช้ในการวิเคราะห์ข้อมูล

1. หาค่าสถิติพื้นฐาน คือ ค่าเฉลี่ยและค่าความเบี่ยงเบนมาตรฐาน
2. การตรวจสอบความเป็นมิติเดียว (Unidimensionality) ของแบบทดสอบตามข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) โดยนำแบบทดสอบจำนวน 6 ฉบับ ๆ ละ 40 ข้อ ไปทดสอบกับนักเรียนที่ใช้ในการทดลองสอบ จำนวน 1,128 คน แล้วนำผลมาวิเคราะห์หองค์ประกอบ (Factor Analysis) โดยใช้โปรแกรม SPSS
3. ประมาณค่าพารามิเตอร์ความยากของข้อสอบและค่าความสามารถของผู้สอบ โดยนำผลการทดลองสอบจากแบบทดสอบ จำนวน 4 ฉบับ และผลการสอบในการวิจัยจากแบบทดสอบดั้งเดิม และแบบทดสอบเทเลอร์รูปปิรามิด มาวิเคราะห์หาค่าพารามิเตอร์ตามทฤษฎีการตอบสนองข้อสอบ แบบราสช์ โมเดล (Rasch Model) ด้วยวิธีการประมาณค่าความเป็นไปได้สูงสุดแบบไม่มีเงื่อนไข (Unconditional Maximum Likelihood Procedure) โดยใช้โปรแกรมไบคาล (BICAL)
4. คำนวณค่าความเที่ยงตรงเชิงสภาพ (Concurrent Validity) ของการประมาณค่าความสามารถจากแบบทดสอบที่แตกต่างกัน 3 รูปแบบ โดยทำตามลำดับขั้นดังนี้
 - 4.1 คำนวณค่าความสัมพันธ์ระหว่างความสามารถที่ได้จากการประมาณค่าด้วยการทดสอบแบบทดสอบปิรามิดแต่ละรูปแบบ กับค่าความสามารถที่ได้จากการประมาณค่าด้วยการทดสอบแบบทดสอบดั้งเดิม โดยใช้สูตรสัมประสิทธิ์สหสัมพันธ์เพียร์สัน (Pearson Product Moment Correlation Coefficient) (ล้วน สายยศ และอังคณา สายยศ. 2543 : 252)

$$r_{\theta_x, \theta_y} = \frac{N \sum \theta_x \theta_y - \sum \theta_x \sum \theta_y}{\sqrt{[N \sum \theta_x^2 - (\sum \theta_x)^2][N \sum \theta_y^2 - (\sum \theta_y)^2]}}$$

เมื่อ r_{θ_x, θ_y}	แทน	ค่าความเที่ยงตรงเชิงสภาพของการประมาณค่าความสามารถ
θ_x	แทน	ค่าความสามารถที่ได้จากการประมาณค่าด้วยแบบทดสอบปิรามิด
θ_y	แทน	ค่าความสามารถที่ได้จากการประมาณค่าด้วยแบบทดสอบดั้งเดิม
N	แทน	จำนวนนักเรียนในกลุ่มตัวอย่าง

- 4.2 เปลี่ยนค่าความเที่ยงตรงเชิงสภาพของกลุ่มตัวอย่างแต่ละโรงเรียนเป็นคะแนนมาตรฐาน (Z) ตารางของฟิชเชอร์ (Fisher's Transformation)

- 4.3 คำนวณ Fisher's Z เฉลี่ย โดยใช้สูตรดังนี้ (ล้วน สายยศและอังคณา สายยศ.

2522 : 228)

$$\bar{Z} = \frac{\sum (N_i - 3)Z_i}{(N_i - 3)}$$

เมื่อ $\bar{Z}$	แทน	คะแนนมาตรฐาน (Fisher's Z) เฉลี่ย
Z_i	แทน	คะแนนมาตรฐาน (Fisher's Z) ของกลุ่มที่ i
N_i	แทน	จำนวนนักเรียนของกลุ่มที่ i

4.4 เปลี่ยนค่าคะแนนมาตรฐานเฉลี่ย ($\bar{Z}$) เป็นค่าความเที่ยงตรงเชิงสภาพเฉลี่ยตามตารางของฟิชเชอร์

5. ทดสอบความแตกต่างของค่าความเที่ยงตรงเชิงสภาพ โดยนำค่าความเที่ยงตรงมาแปลงเป็นคะแนนมาตรฐาน (Z) ตามตารางของฟิชเชอร์ แล้วนำมาทดสอบ โดยใช้สูตร Chi Square : χ^2 ดังนี้ (Steel and Torrie. 1981 : 281)

$$\chi^2 = \sum (N-3)Z^2 - \frac{[\sum (N-3)Z]^2}{\sum (N-3)}$$

เมื่อ χ^2	แทน	ค่าจาก χ^2 - distribution
Z	แทน	คะแนนมาตรฐานที่แปลงมาจากค่าความเที่ยงตรงเชิงสภาพตามตารางของฟิชเชอร์
N	แทน	จำนวนนักเรียนในกลุ่มตัวอย่าง
K	แทน	จำนวนค่าของความเที่ยงตรงเชิงสภาพ

6. ถ้าพบค่าความเที่ยงตรงเชิงสภาพที่นำมาทดสอบในข้อ 5 มีความแตกต่างกันอย่างมีนัยสำคัญให้นำค่าความเที่ยงตรงเชิงสภาพมาทดสอบความแตกต่างเป็นรายคู่ โดยใช้สูตรดังนี้ (ล้วน สายยศ และอังคณา สายยศ. 2522 : 227)

$$Z = \frac{Z_{r1} - Z_{r2}}{\sqrt{\frac{1}{(N_1 - 3)} + \frac{1}{(N_2 - 3)}}}$$

เมื่อ Z	แทน	คะแนนมาตรฐานของโค้งปกติ
Z_{r1}, Z_{r2}	แทน	คะแนนมาตรฐานที่แปลงมาจากค่าสัมประสิทธิ์สหสัมพันธ์ตามตารางของฟิชเชอร์
N_1, N_2	แทน	จำนวนนักเรียนในกลุ่มตัวอย่างที่ทำแบบทดสอบแต่ละรูปแบบ

7. การคำนวณค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบดั้งเดิม และแบบทดสอบเทเลอร์รูปปรีรามิต โดยนำค่าพารามิเตอร์ของข้อสอบแต่ละฉบับมาคำนวณค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบ ณ ระดับความสามารถต่าง ๆ 81 ระดับ โดยใช้สูตรในการคำนวณดังนี้ (อรินทร์ น่วมถนอม. 2537 : 67 ; Citing Lord and Novick. 1968 : 454)

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

เมื่อ $I(\theta)$	แทน	ค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบ ณ ระดับความสามารถต่าง ๆ ซึ่งผู้วิจัยแบ่งออกเป็น 81 ระดับ คือ -4.00, -3.90, -3.80, ..., 3.80, 3.90, 4.00
i	แทน	ข้อสอบข้อที่ 1, 2, 3, ..., 30
$P_i(\theta)$	แทน	ความน่าจะเป็นที่ผู้สอบซึ่งมีระดับความสามารถ θ จะตอบข้อสอบข้อที่ i ได้ถูกต้อง
$Q_i(\theta)$	แทน	$1 - P_i(\theta)$
$P_i'(\theta)$	แทน	ความชันโค้งลักษณะข้อสอบที่ระดับความสามารถ θ

8. การคำนวณค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถ ระหว่างแบบทดสอบที่แตกต่างกัน 2 รูปแบบ โดยนำค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบ ณ ระดับความสามารถต่าง ๆ 81 ระดับ มาเปรียบเทียบกันทีละคู่ โดยใช้สูตรในการคำนวณ ดังนี้ (อรินทร์ น่วมถนอม. 2537 : 68 ; Citing Lord and Novick. 1968 : 472)

$$RE(\theta, X_1, X_2) = \frac{I(\theta, X_1)}{I(\theta, X_2)}$$

เมื่อ $RE(\theta, X_1, X_2)$	แทน	ค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถระหว่างแบบทดสอบรูปแบบที่ 1 (X_1) เทียบกับแบบทดสอบรูปแบบที่ 2 (X_2)
$I(\theta, X_1)$	แทน	ค่าอินฟอร์เมชันของแบบทดสอบรูปแบบที่ 1 (X_1) ณ ระดับความสามารถ θ
$I(\theta, X_2)$	แทน	ค่าอินฟอร์เมชันของแบบทดสอบรูปแบบที่ 2 (X_2) ณ ระดับความสามารถ θ

9. ค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถ ระหว่างแบบทดสอบที่นำมาเปรียบเทียบกันในข้อ 8 นำมาแปลผลดังนี้ (อรินทร์ น่วมถนอม. 2537 : 68 ; Citing Warm. 1978 : 76)

- เมื่อ $RE(\theta, X_1, X_2) > 1$ แสดงว่า แบบทดสอบรูปแบบที่ 1 (X_1) มีประสิทธิภาพสูงกว่าแบบทดสอบรูปแบบที่ 2 (X_2) ณ ระดับความสามารถ θ
- $RE(\theta, X_1, X_2) < 1$ แสดงว่า แบบทดสอบรูปแบบที่ 1 (X_1) มีประสิทธิภาพต่ำกว่าแบบทดสอบรูปแบบที่ 2 (X_2) ณ ระดับความสามารถ θ
- $RE(\theta, X_1, X_2) = 1$ แสดงว่า แบบทดสอบรูปแบบที่ 1 (X_1) มีประสิทธิภาพพอ ๆ กันกับแบบทดสอบรูปแบบที่ 2 (X_2) ณ ระดับความสามารถ θ

บทที่ 4

ผลการวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูล และการแปลความหมายของผลการวิเคราะห์ข้อมูลในการวิจัยครั้งนี้ เพื่อความสะดวกและเข้าใจตรงกัน ในการนำเสนอและแปลความหมายผลการวิเคราะห์ข้อมูล ผู้วิจัยจึงได้กำหนดสัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูลดังนี้

สัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูล

N	แทน	จำนวนนักเรียนในกลุ่มตัวอย่าง
k	แทน	จำนวนข้อของแบบทดสอบ
$\bar{X}$	แทน	คะแนนเฉลี่ย
S	แทน	ความเบี่ยงเบนมาตรฐาน
b	แทน	ความยากของข้อสอบ
θ	แทน	ความสามารถของผู้สอบ
r	แทน	สัมประสิทธิ์สหสัมพันธ์
r_{xy}	แทน	สัมประสิทธิ์สหสัมพันธ์ระหว่างแบบทดสอบรูปปิรามิดกับวิธีการให้คะแนนของแบบทดสอบ
X_1	แทน	แบบทดสอบรูปปิรามิด แบบที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น
X_2	แทน	แบบทดสอบรูปปิรามิด แบบที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้น
X_3	แทน	แบบทดสอบรูปปิรามิด แบบที่มีข้อสอบหลายข้อในแต่ละชั้น จำนวน 5 ชั้น
X_4	แทน	แบบทดสอบดั้งเดิม
Y_1	แทน	วิธีการตรวจให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ทำ
Y_2	แทน	วิธีการตรวจให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบเฉพาะข้อที่ทำถูก
Y_3	แทน	วิธีการให้คะแนนค่าความยากตามข้อสอบที่ถัดจากข้อสุดท้ายที่ได้ตอบ
χ^2	แทน	ค่าไคสแควร์
TIF	แทน	ค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบ
RE	แทน	ค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถ
SEE	แทน	ค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่า

การเสนอผลการวิเคราะห์ข้อมูล

ในการศึกษาครั้งนี้ ผู้วิจัยได้เสนอผลการวิเคราะห์ข้อมูลตามความมุ่งหมาย และ สมมติฐานของการศึกษาค้นคว้าดังนี้

1. ค่าสถิติพื้นฐาน

1.1 ค่าสถิติพื้นฐานของแบบทดสอบรูปปิรามิดทั้ง 3 รูปแบบ

1.2 ค่าสถิติพื้นฐานของความสามารถของผู้สอบ

2. ค่าความเที่ยงตรงและความแตกต่างของค่าความเที่ยงตรงของการประมาณค่า ความสามารถของผู้สอบในวิชาคณิตศาสตร์ ของแบบทดสอบรูปปิรามิดทั้ง 3 รูปแบบ จำแนก ตามวิธีการให้คะแนน 3 วิธี

3. ค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของผู้สอบในวิชา คณิตศาสตร์ ระหว่างแบบสอบรูปปิรามิดทั้ง 3 รูปแบบ

ผลการวิเคราะห์ข้อมูล

1. ค่าสถิติพื้นฐาน

1.1 ค่าสถิติพื้นฐานของแบบทดสอบรูปปิรามิดทั้ง 3 รูปแบบ

จากการนำแบบทดสอบรูปปิรามิดทั้ง 3 รูปแบบ ไปทดสอบกับกลุ่มตัวอย่าง จำนวนคน แล้วแยกผลวิเคราะห์ตามวิธีการตรวจให้คะแนน ของแบบทดสอบรูปปิรามิด ปรากฏ ว่าได้ค่าสถิติพื้นฐานดังแสดงในตาราง 7

ตาราง 7 ค่าสถิติพื้นฐานของแบบทดสอบรูปปริมาตรทั้ง 3 ฉบับ ตามวิธีการตรวจให้คะแนน

วิธีการให้ คะแนน	Y ₁ (N = 364)			Y ₂ (N = 364)			Y ₃ (N = 364)		
	คะแนน สูงสุด	$\bar{X}$	S	คะแนน สูงสุด	$\bar{X}$	S	คะแนน สูงสุด	$\bar{X}$	S
แบบทดสอบ รูปปริมาตร แบบที่ใช้ ขนาดชั้นคงที่ 5 ชั้น (K = 15)	0.90	0.04	0.77	1.63	0.45	0.82	2.13	0.51	1.20
แบบที่ใช้ ขนาดชั้นคงที่ 15 ชั้น (K = 120)	1.22	0.14	0.51	1.41	0.24	0.51	3.17	0.40	1.04
แบบที่มีหลาย ข้อในแต่ละชั้น จำนวน 5 ชั้น (K = 45)	0.80	-0.19	0.49	0.84	0.02	0.52	2.27	0.69	1.15

ตาราง 7 แสดงให้เห็นว่าเมื่อใช้วิธีการตรวจให้คะแนนทั้ง 3 วิธี ในการตรวจให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ทำ นั้นแบบทดสอบรูปปริมาตรที่ใช้ขนาดชั้นคงที่ 15 ชั้นมีค่าเฉลี่ยสูงสุด วิธีการตรวจให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบเฉพาะข้อที่ทำถูก แบบทดสอบรูปปริมาตรที่ใช้ขนาดชั้นคงที่ 5 ชั้นมีค่าเฉลี่ยสูงสุด ส่วนแบบทดสอบรูปปริมาตรที่มีหลายข้อในแต่ละชั้น มีค่าเฉลี่ยสูงสุดในการตรวจให้คะแนนตามค่าความยากตามข้อสอบที่ถัดจากข้อสุดท้ายที่ได้ตอบ นอกจากนั้นเมื่อพิจารณาค่าความเบี่ยงเบนมาตรฐานจะพบว่า เมื่อใช้วิธีการตรวจให้คะแนนทั้ง 3 วิธี แบบทดสอบรูปปริมาตรที่ใช้ขนาดชั้นคงที่ 5 ชั้นจะมีค่าความเบี่ยงเบนมาตรฐานสูงสุด แบบทดสอบรูปปริมาตรที่ใช้ขนาดชั้นคงที่ 15 ชั้น จะมีค่าความเบี่ยงเบนมาตรฐานต่ำสุด เมื่อใช้วิธีการตรวจให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบเฉพาะข้อที่ทำถูก และ วิธีการตรวจให้คะแนนค่าความยากตามข้อสอบที่ถัดจากข้อสุดท้ายที่ได้ตอบ ส่วนแบบทดสอบรูปปริมาตรที่มีหลายข้อในแต่ละชั้น จำนวน 5 ชั้น จะมีความเบี่ยงเบนมาตรฐานต่ำสุดในการตรวจให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ทำ

1.2 ค่าสถิติพื้นฐานของความสามารถของผู้สอบ

ผู้วิจัยได้นำแบบทดสอบดั้งเดิมจำนวน 30 ข้อ ไปวิเคราะห์ด้วยโปรแกรมใบคาล เพื่อหาค่าความสามารถของผู้สอบแต่ละคน ปรากฏผลค่าสถิติพื้นฐานดังแสดงในตาราง 8

ตาราง 8 ค่าสถิติพื้นฐานของความสามารถของผู้สอบ (θ)

แบบทดสอบ	N	k	θ	
			$\bar{X}$	S
แบบทดสอบดั้งเดิม	364	30	0.906	1.057

ตาราง 8 แสดงให้เห็นว่า ความสามารถของผู้สอบ ของกลุ่มตัวอย่างมีค่าเฉลี่ย เท่ากับ 0.906 และมีค่าความเบี่ยงเบนมาตรฐานเท่ากับ 1.057 ซึ่งใกล้เคียงกับค่ามาตรฐาน กล่าวคือ ความสามารถเฉลี่ยใกล้ศูนย์ และการกระจายของความสามารถใกล้ค่า 1.00 แสดง ว่า มีข้อสอบบางข้อที่ยากหรือง่ายออกจากกลุ่มข้อสอบ และมีผู้สอบบางคนความสามารถสูงหรือ ต่ำกระจายออกจากกลุ่มไปบางเล็กน้อย

2. ค่าความเที่ยงตรงและการทดสอบความแตกต่างของค่าความเที่ยงตรงของการ ประเมินค่าความสามารถในวิชาคณิตศาสตร์ ระหว่างแบบทดสอบรูปปริมาตรทั้ง 3 รูปแบบ กับ ความสามารถของผู้สอบ

ผู้วิจัยได้นำค่าประมาณค่าความสามารถของผู้สอบ ที่ได้จากการทดสอบด้วยแบบ ทดสอบรูปปริมาตร จากจำนวนนักเรียน 364 คน ไปคำนวณหาค่าความเที่ยงตรง โดยการ คำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันระหว่างค่าความสามารถของผู้สอบจากการ ทดสอบด้วยแบบทดสอบรูปปริมาตรแต่ละรูปแบบ กับความสามารถของผู้สอบจากการทดสอบ ด้วยแบบทดสอบดั้งเดิม แล้วทดสอบความแตกต่างของค่าความเที่ยงตรงระหว่างแบบทดสอบ ทั้ง 3 รูปแบบ ด้วยวิธีการแปลงค่าสัมประสิทธิ์สหสัมพันธ์ ให้อยู่ในรูปคะแนนมาตรฐานตามตาราง ฟิชเชอร์ โดยใช้ค่าไคสแควร์ ได้ผลดังตาราง 9

ตาราง 9 เปรียบเทียบค่าความเที่ยงตรงระหว่างแบบทดสอบปิรามิด 3 รูปแบบ กับความสามารถของผู้สอบ โดยวิธีการตรวจให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ทำ

วิธีการตรวจให้คะแนน	แบบทดสอบรูปปิรามิด	สัมประสิทธิ์สหสัมพันธ์	Z	df	χ^2
Y ₁	X _{1๐}	0.660	0.793	2	69.174**
	X _{2๐}	0.797	1.091	2	
	X _{3๐}	0.429	0.459	2	

ตาราง 9 แสดงให้เห็นว่าความเที่ยงตรงของการประมาณค่าความสามารถของผู้สอบจากการทดสอบด้วยแบบทดสอบรูปปิรามิดทั้ง 3 รูปแบบด้วยวิธีการตรวจให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ทำ มีค่าเท่ากับ 0.660, 0.797 และ 0.429 ตามลำดับ ซึ่งค่าความเที่ยงตรงของแบบทดสอบรูปปิรามิดแบบที่ใช้ขนาดชั้นคงที่ 15 ชั้นมีค่าสูงสุด ต่ำลงมาคือแบบทดสอบรูปปิรามิดแบบที่ใช้ขนาดชั้นคงที่ 5 ชั้น และแบบทดสอบรูปปิรามิดที่มีหลายข้อในแต่ละชั้นจำนวน 5 ชั้น มีค่าต่ำสุด และเมื่อนำมาทดสอบความแตกต่างปรากฏว่าค่าความเที่ยงตรงของการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบทั้ง 3 รูปแบบมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่ามีค่าความเที่ยงตรงต่างกันอย่างน้อย 1 คู่

ตาราง 10 เปรียบเทียบค่าความเที่ยงตรงระหว่างแบบทดสอบปิรามิด 3 รูปแบบ โดยวิธีการตรวจให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบเฉพาะข้อที่ทำถูก

วิธีการตรวจให้คะแนน	แบบทดสอบรูปปิรามิด	สัมประสิทธิ์สหสัมพันธ์	Z	df	χ^2
Y ₂	X _{1๐}	0.723	0.914	2	99.762**
	X _{2๐}	0.868	1.325	2	
	X _{3๐}	0.525	0.583	2	

ตาราง 10 แสดงให้เห็นว่าความเที่ยงตรงของการประมาณค่าความสามารถของผู้สอบจากการทดสอบด้วยแบบทดสอบรูปปิรามิดทั้ง 3 รูปแบบด้วยวิธีการตรวจให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ทำ มีค่าเท่ากับ 0.723, 0.868 และ 0.525 ตามลำดับ ซึ่งค่าความ

เที่ยงตรงของแบบทดสอบรูปปริมาตรแบบที่ใช้ขนาดชั้นคงที่ 15 ชั้นมีค่าสูงสุด ต่ำลงมาคือแบบทดสอบรูปปริมาตรแบบที่ใช้ขนาดชั้นคงที่ 5 ชั้น และแบบทดสอบรูปปริมาตรที่มีหลายข้อในแต่ละชั้นจำนวน 5 ชั้น มีค่าต่ำสุด และเมื่อนำมาทดสอบความแตกต่างปรากฏว่าค่าความเที่ยงตรงของการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบทั้ง 3 รูปแบบมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่ามีค่าความเที่ยงตรงต่างกันอย่างน้อย 1 คู่

ตาราง 11 เปรียบเทียบค่าความเที่ยงตรงระหว่างแบบทดสอบปริมาตร 3 โดยวิธีการให้คะแนนค่าความยากตามข้อสอบที่คัดจากข้อสุดท้ายที่ได้ตอบ

วิธีการตรวจให้คะแนน	แบบทดสอบรูปปริมาตร	สัมประสิทธิ์สหสัมพันธ์	Z	df	χ^2
Y ₃	X ₁₀	0.860	1.293	2	30.324**
	X ₂₀	0.930	1.658	2	
	X ₃₀	0.927	1.637	2	

ตาราง 11 แสดงให้เห็นว่าความเที่ยงตรงของการประมาณค่าความสามารถของผู้สอบจากการทดสอบด้วยแบบทดสอบรูปปริมาตรทั้ง 3 รูปแบบด้วยวิธีการตรวจให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ทำ มีค่าเท่ากับ 0.860, 0.930 และ 0.927 ตามลำดับ ซึ่งค่าความเที่ยงตรงของแบบทดสอบรูปปริมาตรแบบที่ใช้ขนาดชั้นคงที่ 15 ชั้นมีค่าสูงสุด ต่ำลงมาคือแบบทดสอบรูปปริมาตรแบบที่มีหลายข้อในแต่ละชั้น จำนวน 5 ชั้น และแบบทดสอบรูปปริมาตรที่ใช้ขนาดชั้นคงที่ 5 ชั้นมีค่าต่ำสุด และเมื่อนำมาทดสอบความแตกต่าง ปรากฏว่าค่าความเที่ยงตรงของการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบทั้ง 3 รูปแบบมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่ามีค่าความเที่ยงตรงต่างกันอย่างน้อย 1 คู่

ดังนั้นเพื่อให้ทราบค่าความเที่ยงตรงของการประมาณค่าความสามารถของผู้สอบจากแบบทดสอบคู่ใดมีความแตกต่างกัน ผู้วิจัยจึงนำค่าความเที่ยงตรงจากแบบทดสอบรูปปริมาตรทั้ง 3 รูปแบบ จำแนกตามวิธีการตรวจให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ทำ และวิธีการตรวจให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบเฉพาะข้อที่ทำถูกไปทดสอบความแตกต่างทีละคู่ได้ผลดังแสดงในตาราง 12 – 14

ตาราง 12 ผลของการทดสอบความแตกต่างของค่าความเที่ยงตรงของการประมาณค่าความสามารถของผู้สอบ จากการทำแบบทดสอบรูปปริมาตร 3 รูปแบบ โดยวิธีการตรวจให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ทำ โดยทดสอบเป็นรายคู่

วิธีการตรวจให้คะแนน	แบบทดสอบ รูปปริมาตร		X ₁₀	X ₂₀	X ₃₀
		Z	0.793	1.091	0.459
Y ₁	X ₁₀	0.793	-	4.729**	3.762**
	X ₂₀	1.091		-	8.491**
	X ₃₀	0.459			-

ตาราง 12 แสดงให้เห็นว่าความเที่ยงตรงของการประมาณค่าความสามารถของผู้สอบ จากการทำแบบทดสอบรูปปริมาตร 3 รูปแบบ โดยวิธีการตรวจให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ทำ ปรากฏว่า ระหว่างแบบทดสอบรูปปริมาตรแบบที่ใช้ขนาดชั้นคงที่จำนวน 5 ชั้น กับ แบบทดสอบรูปปริมาตร แบบที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้น และแบบทดสอบรูปปริมาตรแบบที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น กับแบบทดสอบรูปปริมาตรแบบที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น และ แบบทดสอบรูปปริมาตร แบบที่ใช้ขนาดชั้นคงที่จำนวน 15 ชั้น กับ แบบทดสอบรูปปริมาตรแบบที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

ตาราง 13 ผลของการทดสอบความแตกต่างของค่าความเที่ยงตรงของการประมาณค่าความสามารถของผู้สอบ จากการทำแบบทดสอบรูปปริมาตร 3 รูปแบบ โดยวิธีการตรวจให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบเฉพาะข้อที่ทำถูก โดยทดสอบเป็นรายคู่

วิธีการตรวจให้คะแนน	แบบทดสอบ รูปปริมาตร		X ₁₀	X ₂₀	X ₃₀
		Z	0.914	1.325	0.583
Y ₂	X ₁₀	0.914	-	5.522**	4.447**
	X ₂₀	1.325		-	6.508**
	X ₃₀	0.583			-

ตาราง 13 แสดงให้เห็นว่าความเที่ยงตรงของการประมาณค่าความสามารถของผู้สอบ จากการทำแบบทดสอบรูปปริมาตร 3 รูปแบบ โดยวิธีการตรวจให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ทำ ปรากฏว่า ระหว่างแบบทดสอบรูปปริมาตรแบบที่ใช้ขนาดชั้นคงที่จำนวน 5 ชั้น กับ แบบทดสอบรูปปริมาตร แบบที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้น และ แบบทดสอบรูปปริมาตร แบบที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น กับแบบทดสอบรูปปริมาตร แบบที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น และ แบบทดสอบรูปปริมาตรแบบที่ใช้ขนาดชั้นคงที่จำนวน 15 ชั้น กับแบบทดสอบรูปปริมาตร แบบที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

ตาราง 14 ผลของการทดสอบความแตกต่างของค่าความเที่ยงตรงของการประมาณค่าความสามารถของผู้สอบ จากการทำแบบทดสอบรูปปริมาตร 3 รูปแบบ โดยวิธีการตรวจให้คะแนนตามค่าความยากตามข้อสอบที่ถัดจากข้อสุดท้ายที่ได้ตอบ โดยทดสอบเป็นรายคู่

วิธีการตรวจให้คะแนน	แบบทดสอบรูปปริมาตร		X ₁₀	X ₂₀	X ₃₀
		Z	1.293	1.658	1.637
Y ₂	X ₁₀	1.293	-	4.904**	4.622**
	X ₂₀	1.658		-	0.282
	X ₃₀	1.637			-

ตาราง 14 แสดงให้เห็นว่าความเที่ยงตรงของการประมาณค่าความสามารถของผู้สอบ จากการทำแบบทดสอบรูปปริมาตร 3 รูปแบบ โดยวิธีการตรวจให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ทำ ปรากฏว่า ระหว่างแบบทดสอบรูปปริมาตรแบบที่ใช้ขนาดชั้นคงที่จำนวน 5 ชั้น กับ แบบทดสอบรูปปริมาตร แบบที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้น และ แบบทดสอบรูปปริมาตร แบบที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น กับแบบทดสอบรูปปริมาตร แบบที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วน แบบทดสอบรูปปริมาตรแบบที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้น กับแบบทดสอบรูปปริมาตรแบบที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

3. ผลการค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถในวิชาคณิตศาสตร์ ระหว่างแบบทดสอบรูปปริมาตรทั้ง 3 รูปแบบ

ผู้วิจัยได้คำนวณหาค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบรูปปริมาตร แบบที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้นเทียบกับแบบทดสอบรูปปริมาตรแบบที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้น แบบทดสอบรูปปริมาตร แบบที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้นเทียบกับแบบทดสอบรูปปริมาตร แบบที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น และแบบทดสอบรูปปริมาตร แบบที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้นเทียบกับแบบทดสอบรูปปริมาตรแบบที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น โดยนำค่าฟังก์ชันอินฟอर्मชันของแบบทดสอบแต่ละรูปแบบ ซึ่งมีรายละเอียดดังตาราง ประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถ ปรากฏว่าได้ผลดังตาราง 15

ตาราง 15 ค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบรูปปริมาตร 3 รูปแบบ ที่ระดับความสามารถตั้งแต่ -4.00 ถึง 4.00 จำนวน 81 ระดับความสามารถ

ระดับความสามารถ	RE		
	X_1 / X_2	X_1 / X_3	X_2 / X_3
$\theta < -3.00$	0.1568 – 0.1401	1.4150 – 1.1723	9.0268 – 8.9639
$-3.00 \leq \theta < -1.00$	0.1383 – 0.1119	1.1450 – 0.7941	8.2821 – 7.0973
$-1.00 \leq \theta \leq 1.00$	0.1112 – 0.1150	0.7859 – 0.8799	7.0688 – 7.6546
$1.00 < \theta \leq 3.00$	0.1160 – 0.1504	0.9014 – 1.7152	7.7674 – 11.4011
$\theta > 3.00$	0.1526 – 0.1702	1.7740 – 2.2726	11.6217 – 13.3500

ตาราง 15 แสดงให้เห็นช่วงการกระจายของค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถ ระหว่างแบบทดสอบที่แตกต่างกัน 3 รูปแบบ ซึ่งนำมาเปรียบเทียบกับเกณฑ์ 1 (Warm. 1978 : 76) ได้ผลดังนี้

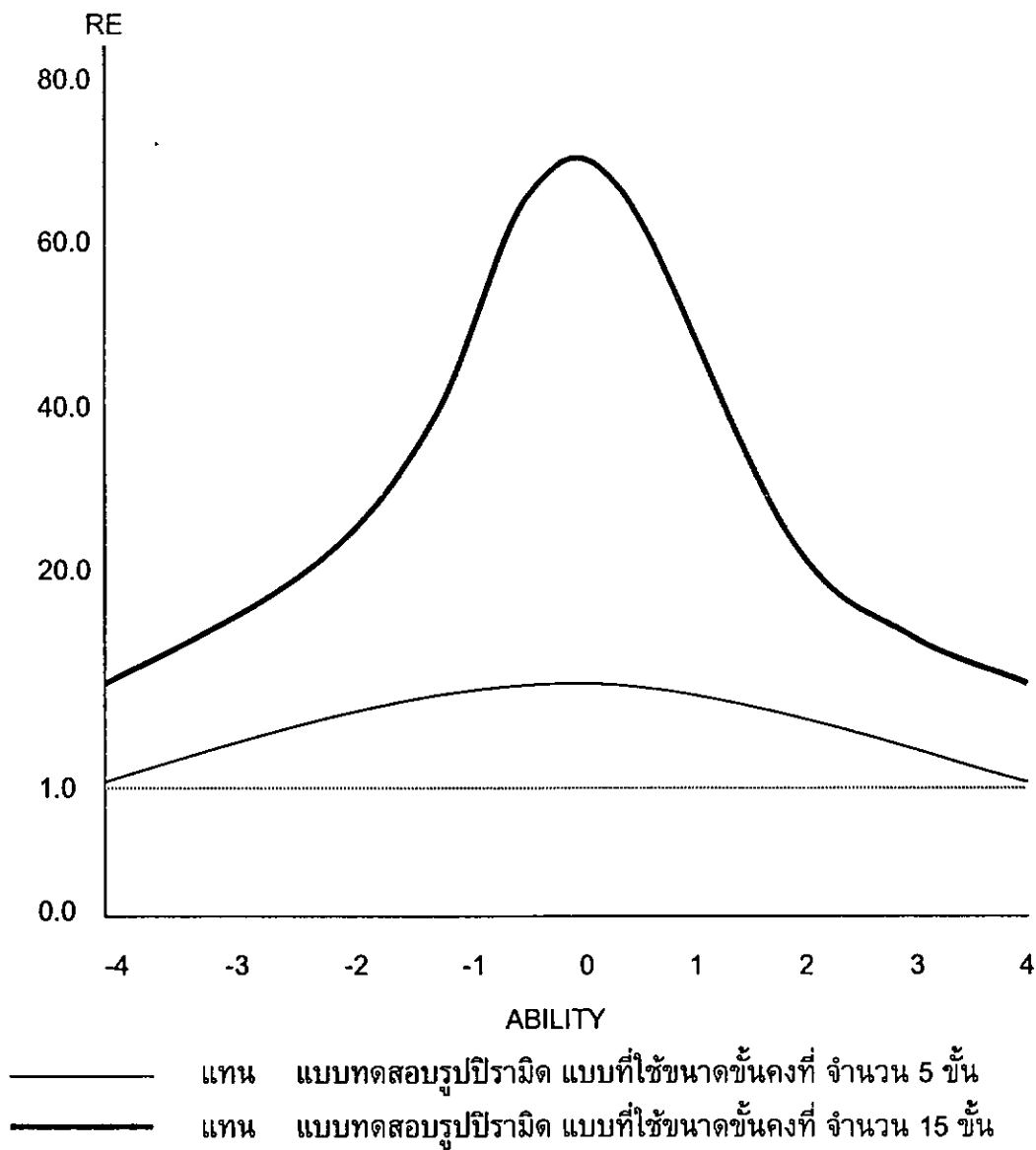
3.1 ค่าประสิทธิภาพสัมพัทธ์ ของการประมาณค่าความสามารถของผู้สอบ ในวิชาคณิตศาสตร์ ระหว่างแบบทดสอบรูปปริมาตรที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น เทียบกับแบบทดสอบรูปปริมาตรที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้น มีค่าต่ำกว่า 1 ที่ระดับความสามารถทุกระดับความสามารถ แสดงว่าแบบทดสอบรูปปริมาตรที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น มีประสิทธิภาพในการประมาณค่าความสามารถสูงกว่า แบบทดสอบรูปปริมาตรที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น

ในกลุ่มผู้สอบทุกระดับความสามารถ โดยจะมีค่าต่ำสุดในกลุ่มผู้สอบที่มีระดับความสามารถปานกลาง ($-1.00 \leq \theta \leq 1.00$) แสดงว่า แบบทดสอบรูปปริมาตรที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น มีความสามารถในการประมาณค่าความสามารถของผู้สอบได้ดีที่สุดในกลุ่มผู้สอบที่มีความสามารถระดับปานกลาง ดังนั้นจึงสรุปได้ว่า แบบทดสอบรูปปริมาตรที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้นมีประสิทธิภาพในการประมาณค่าความสามารถสูงกว่า แบบทดสอบรูปปริมาตรที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น ในกลุ่มผู้สอบที่มีความสามารถในทุกระดับความสามารถ โดยจะมีประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบในกลุ่มผู้สอบที่มีความสามารถระดับปานกลาง สูงที่สุด

3.2 ค่าประสิทธิภาพสัมพัทธ์ ของการประมาณค่าความสามารถของผู้สอบ ในวิชาคณิตศาสตร์ ระหว่าง แบบทดสอบรูปปริมาตรที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น เทียบกับ แบบทดสอบรูปปริมาตรที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น มีค่าต่ำกว่า 1 ที่ระดับความสามารถปานกลาง ($-1.00 \leq \theta \leq 1.00$) แสดงว่าแบบทดสอบรูปปริมาตรที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้น มีประสิทธิภาพในการประมาณค่าความสามารถสูงกว่า แบบทดสอบรูปปริมาตรที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น ในกลุ่มผู้สอบที่มีความสามารถระดับปานกลาง ส่วนที่มีความสามารถระดับต่ำ มีค่าประสิทธิภาพสัมพัทธ์ที่ต่ำกว่า 1 และสูงกว่า 1 โดยมีค่าสูงกว่า 1 ที่ระดับความสามารถเท่ากับ $-2.50, -2.60, -2.70, -2.80, -2.90$ และ -3.00 นอกนั้นที่ความสามารถตั้งแต่ -1.10 ถึง -2.40 จะมีค่าน้อยกว่า 1 แสดงว่า ที่ความสามารถระดับต่ำนี้ส่วนมากมีค่าน้อยกว่า 1 แสดงว่า แบบทดสอบรูปปริมาตรที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น มีประสิทธิภาพในการประมาณค่าความสามารถสูงกว่า แบบทดสอบรูปปริมาตรที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น ในกลุ่มผู้สอบที่มีความสามารถระดับต่ำ สำหรับความสามารถระดับสูง ($1.00 < \theta \leq 3.00$) มีค่าประสิทธิภาพสัมพัทธ์ที่ต่ำกว่า 1 และสูงกว่า 1 โดยมีต่ำกว่า 1 ที่ระดับความสามารถเท่ากับ $1.10, 1.20, 1.30$ และ 1.40 นอกนั้นที่ความสามารถตั้งแต่ 1.50 ถึง 3.00 จะมีค่าสูงกว่า 1 แสดงว่าที่ความสามารถระดับสูงนี้ส่วนมากมีค่ามากกว่า 1 แสดงว่าแบบทดสอบรูปปริมาตรที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น มีประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบสูงกว่า แบบทดสอบรูปปริมาตรที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น ในกลุ่มผู้สอบที่มีความสามารถระดับสูง ในกลุ่มผู้สอบที่มีความสามารถระดับต่ำมาก ($\theta < -3.00$) และระดับสูงมาก ($\theta > 3.00$) มีค่าประสิทธิภาพสัมพัทธ์สูงกว่า 1 ทุกระดับความสามารถ แสดงว่า แบบทดสอบรูปปริมาตรที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น มีประสิทธิภาพในการประมาณค่าความสามารถสูงกว่า แบบทดสอบรูปปริมาตรที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น ในกลุ่มผู้สอบที่มีความสามารถระดับต่ำมาก และระดับสูงมาก จึงสรุปได้ว่า แบบทดสอบรูปปริมาตรที่ใช้ขนาดชั้นคงที่จำนวน 5 ชั้น มีประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบสูงกว่า แบบทดสอบรูปปริมาตรที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น ในกลุ่มผู้สอบที่มีความสามารถระดับต่ำมาก

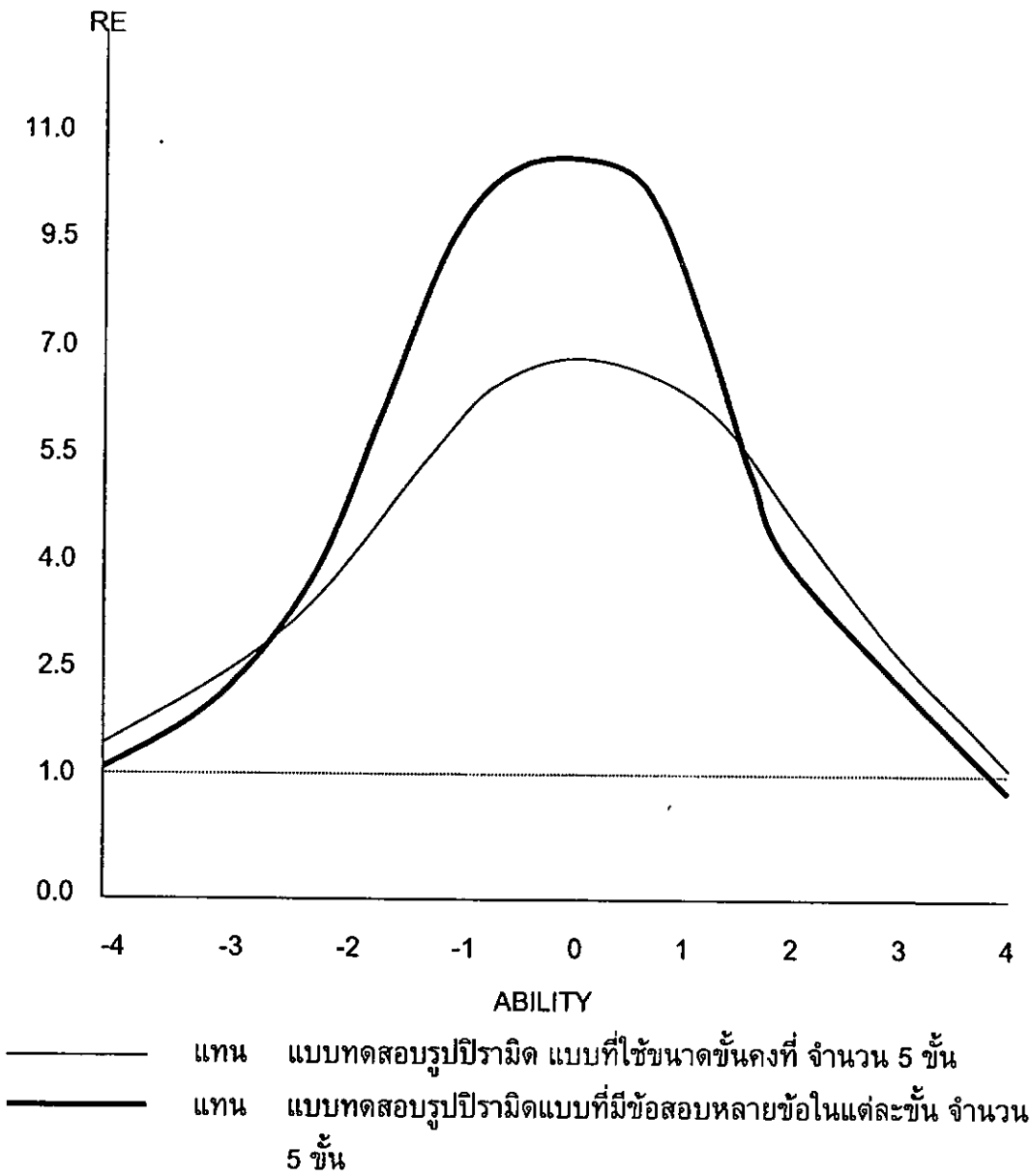
ระดับสูง และระดับสูงมาก แต่จะมีประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบน้อยกว่าในระดับต่ำ และระดับปานกลาง

3.3 ค่าประสิทธิภาพสัมพัทธ์ ของการประมาณค่าความสามารถของผู้สอบ ในวิชาคณิตศาสตร์ ระหว่าง แบบทดสอบรูปปริมาตรที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้น เทียบกับแบบทดสอบรูปปริมาตรที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น มีค่าสูงกว่า 1 ในระดับความสามารถทุกระดับความสามารถ แสดงว่า แบบทดสอบรูปปริมาตรที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้น มีประสิทธิภาพในการประมาณค่าความสามารถสูงกว่าแบบทดสอบรูปปริมาตรที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น ในกลุ่มผู้สอบทุกระดับความสามารถ โดยจะมีค่าต่ำสุดในกลุ่มผู้สอบที่มีระดับความสามารถปานกลาง ($-1.00 \leq \theta \leq 1.00$) แสดงว่า แบบทดสอบรูปปริมาตรที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้น มีความสามารถในการประมาณค่าความสามารถของผู้สอบได้ดีที่สุดในกลุ่มผู้สอบที่มีความสามารถระดับปานกลาง ดังนั้นจึงสรุปได้ว่า แบบทดสอบรูปปริมาตรที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้น มีประสิทธิภาพในการประมาณค่าความสามารถสูงกว่า แบบทดสอบรูปปริมาตรมีข้อสอบหลายข้อในแต่ละชั้น จำนวน 5 ชั้น ในกลุ่มผู้สอบที่มีความสามารถในทุกระดับความสามารถ โดยจะมีประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบในกลุ่มผู้สอบที่มีความสามารถระดับปานกลาง สูงที่สุด



ภาพประกอบ 15 โค้งประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของผู้สอบ
ระหว่างแบบทดสอบรูปปิรามิด แบบที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น เทียบกับ
แบบทดสอบรูปปิรามิด แบบที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้น

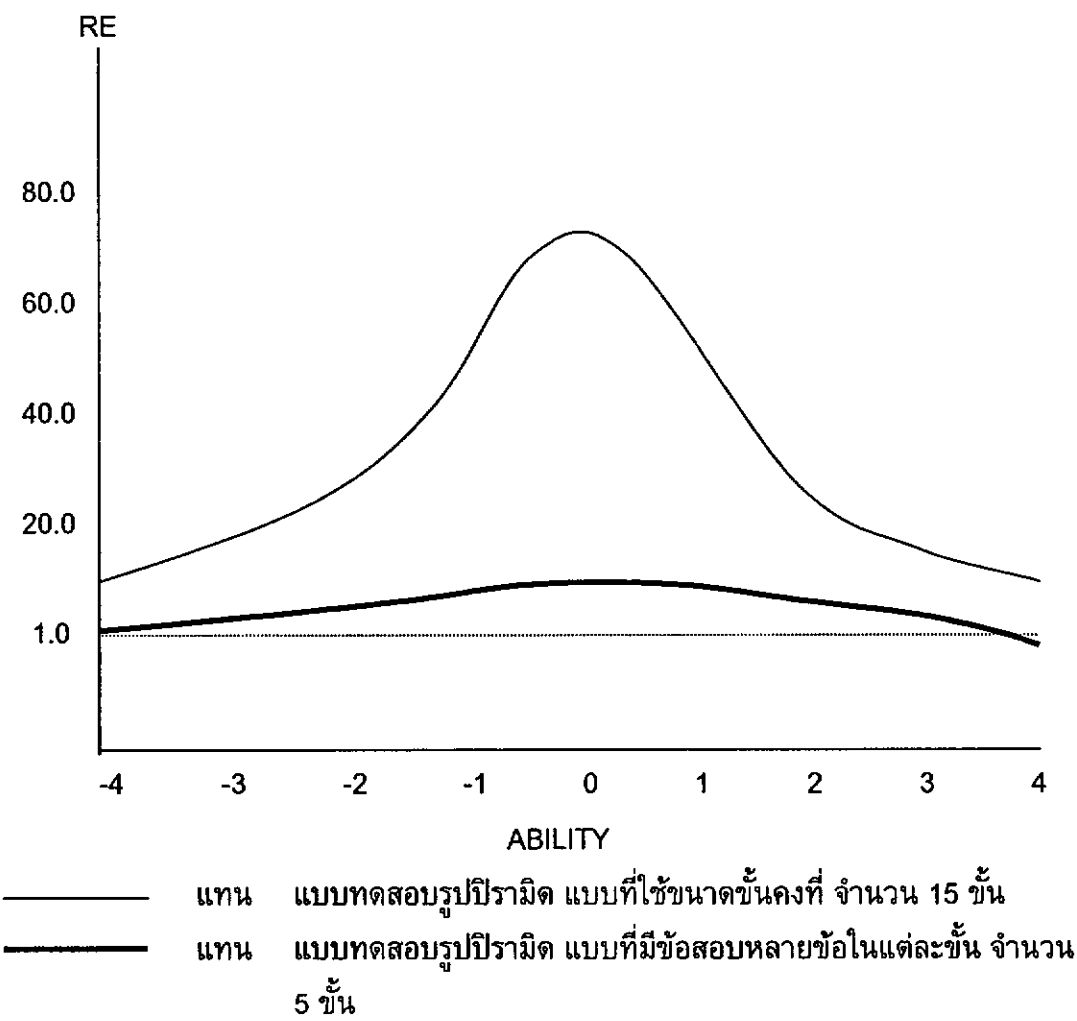
ภาพประกอบ 15 แสดงให้เห็นว่า ค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของผู้สอบ ระหว่าง แบบทดสอบรูปปิรามิด แบบที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น เทียบกับแบบทดสอบรูปปิรามิด แบบที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้น มีค่าต่ำกว่า 1 ในช่วงระดับความสามารถทุกระดับ นั่นคือ แบบทดสอบรูปปิรามิด แบบที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น มีประสิทธิภาพในการประมาณค่าต่ำกว่าแบบทดสอบรูปปิรามิด แบบที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้น ในกลุ่มผู้สอบที่ระดับความสามารถทุกระดับ



ภาพประกอบ 16 โค้งประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบรูปปิรามิด แบบที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น เทียบกับแบบทดสอบรูปปิรามิด แบบที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น

ภาพประกอบ 16 แสดงให้เห็นว่า ค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของผู้สอบ ระหว่าง แบบทดสอบแบบทดสอบรูปปิรามิด แบบที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น เทียบกับแบบทดสอบรูปปิรามิด แบบที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น มีค่าสูงกว่า 1 ในช่วงระดับความสามารถตั้งแต่ -4.00 ถึง -2.50 และ 1.50 ถึง 4.00 แต่มีค่าต่ำกว่า 1 ในช่วงระดับความสามารถตั้งแต่ -2.40 ถึง 1.40 นั่นคือ แบบทดสอบแบบทดสอบรูปปิรามิด

แบบที่ใช้ขนาดขั้นคงที่ จำนวน 5 ขั้น จะมีประสิทธิภาพในการประมาณค่าความสามารถสูงกว่าแบบทดสอบรูปปิรามิด แบบที่มีข้อสอบหลายข้อในแต่ละขั้นจำนวน 5 ขั้น ในกลุ่มผู้สอบที่มีความสามารถระดับต่ำมาก ($\theta \leq -2.50$) ระดับสูง และระดับสูงมาก ($\theta \geq 1.50$) แต่จะมีประสิทธิภาพในการประมาณค่าความสามารถต่ำกว่าแบบทดสอบรูปปิรามิด แบบที่มีข้อสอบหลายข้อในแต่ละขั้นจำนวน 5 ขั้น ในกลุ่มผู้สอบที่มีความสามารถระดับต่ำ และระดับปานกลาง ($-2.40 \leq \theta \leq 1.40$)



ภาพประกอบ 17 โค้งประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบรูปปิรามิด แบบที่ใช้ขนาดขั้นคงที่ จำนวน 15 ขั้น เทียบกับแบบทดสอบรูปปิรามิด แบบที่มีข้อสอบหลายข้อในแต่ละขั้นจำนวน 5 ขั้น

ภาพประกอบ 17 แสดงให้เห็นว่า ค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบรูปปิรามิด แบบที่ใช้ขนาดขั้นคงที่ จำนวน 15 ขั้น เทียบกับแบบทดสอบรูปปิรามิด แบบที่มีข้อสอบหลายข้อในแต่ละขั้นจำนวน 5 ขั้น มีค่าสูงกว่า

1 ในช่วงระดับความสามารถทุกระดับ นั่นคือ แบบทดสอบรูปปิรามิด แบบที่ใช้ขนาดขั้นคงที่จำนวน 15 ขั้น มีประสิทธิภาพในการประมาณค่าสูงกว่าแบบทดสอบรูปปิรามิด แบบที่มีข้อสอบหลายข้อในแต่ละขั้นจำนวน 5 ขั้นในกลุ่มผู้สอบที่ระดับความสามารถทุกระดับ

บทที่ 5

สรุป อภิปรายผล และข้อเสนอแนะ

ความมุ่งหมายของการศึกษาค้นคว้า

1. เพื่อศึกษาค่าความเที่ยงตรงของการประมาณค่าความสามารถของผู้สอบระหว่างแบบทดสอบเทเลอรรูปปริามิตในวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ที่มีรูปแบบ จำนวนชั้น แตกต่างกัน
2. เพื่อศึกษาค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของผู้สอบระหว่างแบบทดสอบเทเลอรรูปปริามิตในวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ที่มีรูปแบบ จำนวนชั้น แตกต่างกัน

กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้านี้เป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ในภาคเรียนที่ 2 ปีการศึกษา 2546 ของโรงเรียนสังกัดสำนักงานเขตพื้นที่การศึกษาราชบุรี เขต 1 และสังกัดเทศบาลเมืองราชบุรี ในเขตอำเภอเมืองราชบุรี จังหวัดราชบุรี จำนวนโรงเรียน 10 โรงเรียน จำนวนห้องเรียน 37 ห้องเรียน จำนวนนักเรียน 1,492 คน ซึ่งเลือกมาโดยการสุ่มอย่างง่าย(Simple Random Sampling) มีโรงเรียนเป็นหน่วยการสุ่ม (Sampling Unit) ได้จำนวน โรงเรียน 10 โรงเรียน จำนวนห้องเรียน 37 ห้องเรียน จำนวนนักเรียน 1,492 คน โดยสุ่มแบ่งกลุ่มตัวอย่างออกเป็น 2 กลุ่มคือ

1. กลุ่มตัวอย่างที่ใช้ในการสร้างเครื่องมือและตรวจสอบคุณภาพ จำนวนห้องเรียน 27 ห้องเรียน จำนวนนักเรียน 1,128 คน
2. กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ จำนวนห้องเรียน 10 ห้องเรียน จำนวนนักเรียน 364 คน

เครื่องมือที่ใช้ในการศึกษาค้นคว้า

เครื่องมือที่ใช้ในการศึกษาค้นคว้า เป็นแบบทดสอบที่ดำเนินการสอบด้วยคอมพิวเตอร์ ซึ่งมีลักษณะดังนี้

1. แบบทดสอบรูปปริามิต

แบบทดสอบปริามิตเป็นแบบทดสอบที่ผู้วิจัยสร้างขึ้นตามรูปแบบจำลองรูปปริามิตเป็นข้อสอบแบบเลือกตอบชนิด 4 ตัวเลือก ในวิชาคณิตศาสตร์ เรื่อง การบวก ลบ และคูณทศนิยม มีจำนวน 3 แบบ ดังนี้

1.1 แบบทดสอบรูปปริามิตที่มีขนาดขั้นคงที่ ใช้กฎการจำแนกแนวทางการตอบ

ข้อสอบแบบ เพิ่ม 1 หรือลด 1 แบบ 5 ชั้น มีจำนวนข้อสอบทั้งหมด 15 ข้อ นักเรียนจะต้องทำข้อสอบในแต่ละชั้น ชั้นละ 1 ข้อ รวมนักเรียนจะต้องทำข้อสอบทั้งสิ้น 5 ข้อ

1.2 แบบทดสอบรูปปริมาตรที่มีขนาดชั้นคงที่ ใช้กฎการจำแนกแนวทางการตอบข้อสอบแบบ เพิ่ม 1 หรือ ลด 1 แบบ 15 ชั้น มีจำนวนข้อสอบทั้งหมด 120 ข้อ นักเรียนจะต้องทำข้อสอบในแต่ละชั้น ชั้นละ 1 ข้อ รวมนักเรียนจะต้องทำข้อสอบทั้งสิ้น 15 ข้อ

1.3 แบบทดสอบรูปปริมาตรที่มีหลายข้อในแต่ละชั้น เป็นแบบทดสอบชนิด 5 ชั้น แบบที่มีข้อสอบ 3 ข้อในแต่ละชั้น แนวทางการตอบข้อสอบแบบ เพิ่ม 1 หรือลด 1 มีจำนวนข้อสอบทั้งสิ้น 45 ข้อ

2. วิธีการให้คะแนนของแบบทดสอบปริมาตร 3 วิธีคือ

2.1 การให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ทำ

2.2 การให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ตอบถูก

2.3 การให้คะแนนค่าความยากตามข้อสอบที่ถัดจากข้อสุดท้ายที่ได้ตอบ

3. แบบทดสอบดั้งเดิม

แบบทดสอบดั้งเดิมเป็นแบบทดสอบที่ผู้วิจัยสร้างขึ้น เป็นข้อสอบแบบเลือกตอบชนิด 4 ตัวเลือก จำนวน 30 ข้อ วิชาคณิตศาสตร์ เรื่อง การบวก ลบ และคูณทศนิยม ระดับชั้นประถมศึกษาปีที่ 6 โดยมีช่วงระดับความยากของข้อสอบ และโค้งอินฟอर्मชันของแบบทดสอบใกล้เคียงกับแบบทดสอบรูปปริมาตร

วิธีดำเนินการเก็บรวบรวมข้อมูล

1. วางแผนในการดำเนินการสอบ โดยการติดต่อขอความร่วมมือในการนำแบบทดสอบไปทดสอบกับนักเรียนจากผู้อำนวยการโรงเรียนที่ใช้เป็นกลุ่มตัวอย่าง เพื่อกำหนดวันเวลาในการดำเนินการสอบ

2. จัดเตรียมแบบทดสอบ 4 ฉบับ ได้แก่ แบบทดสอบดั้งเดิม และแบบทดสอบรูปปริมาตร 3 ฉบับ สำหรับนำไปทดสอบกับนักเรียนกลุ่มตัวอย่าง

3. ผู้วิจัยนำแบบทดสอบทั้ง 4 ฉบับไปทดสอบกับกลุ่มตัวอย่างจำนวน 364 คนด้วยตนเองในระหว่างวันที่ 5 กุมภาพันธ์ พ.ศ. 2547 ถึงวันที่ 10 มีนาคม พ.ศ. 2547 นักเรียนแต่ละคนต้องทำแบบทดสอบทั้ง 4 ฉบับ โดยจะทิ้งช่วงเวลาการทดสอบแต่ละฉบับเป็นเวลา 1 สัปดาห์ ในการดำเนินการสอบผู้วิจัยได้อธิบายให้นักเรียนกลุ่มตัวอย่างเข้าใจในวัตถุประสงค์การสอบ และขอความร่วมมือในการสอบ เพื่อให้ได้ผลตรงตามความเป็นจริง

4. นำผลที่ได้จากการสอบ มาจัดเตรียมข้อมูลสำหรับการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ แล้วนำผลการประมาณค่าไปวิเคราะห์ทางสถิติ และทดสอบสมมติฐานของการวิจัย

การเสนอผลการวิเคราะห์ข้อมูล

ในการศึกษาค้นคว้าครั้งนี้ได้เสนอผลการวิเคราะห์ข้อมูลตามลำดับ ดังนี้

1. ค่าสถิติพื้นฐาน
 - 1.1 ค่าสถิติพื้นฐานของแบบทดสอบรูปปริมาตรทั้ง 3 รูปแบบ
 - 1.2 ค่าสถิติพื้นฐานของความสามารถของผู้สอบ
2. ค่าความเที่ยงตรงและความแตกต่างของค่าความเที่ยงตรงของการประมาณค่าความสามารถของผู้สอบในวิชาคณิตศาสตร์ ของแบบทดสอบรูปปริมาตรทั้ง 3 รูปแบบ
3. ค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของผู้สอบในวิชาคณิตศาสตร์ ระหว่างแบบสอบรูปปริมาตรทั้ง 3 รูปแบบ

สรุปผลการศึกษาค้นคว้า

1. ค่าความเที่ยงตรงและความแตกต่างของค่าความเที่ยงตรงของการประมาณค่าความสามารถของผู้สอบในวิชาคณิตศาสตร์ ของแบบทดสอบรูปปริมาตรทั้ง 3 รูปแบบ จำแนกตามวิธีการให้คะแนน 3 วิธี

1.1 จำแนกตามวิธีการตรวจให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ทำ

1.1.1 ค่าความเที่ยงตรงของการประมาณค่าความสามารถของผู้สอบในวิชาคณิตศาสตร์ ระหว่าง แบบทดสอบรูปปริมาตรแบบที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น กับแบบทดสอบรูปปริมาตรแบบที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้น มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยแบบทดสอบรูปปริมาตรแบบที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้นมีความเที่ยงตรงของการประมาณค่าความสามารถสูงกว่าแบบทดสอบรูปปริมาตรแบบที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น

1.1.2 ค่าความเที่ยงตรงของการประมาณค่าความสามารถของผู้สอบในวิชาคณิตศาสตร์ ระหว่าง แบบทดสอบรูปปริมาตร แบบที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น กับแบบทดสอบรูปปริมาตร แบบที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยแบบทดสอบรูปปริมาตรที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น มีความเที่ยงตรงสูงกว่าแบบทดสอบรูปปริมาตรที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น

1.1.3 ค่าความเที่ยงตรงของการประมาณค่าความสามารถของผู้สอบในวิชาคณิตศาสตร์ ระหว่าง แบบทดสอบรูปปริมาตร แบบที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้น กับแบบทดสอบรูปปริมาตร แบบที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยแบบทดสอบรูปปริมาตรที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้น มีความเที่ยงตรงสูงกว่าแบบทดสอบรูปปริมาตรที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น

คณิตศาสตร์ ระหว่าง แบบทดสอบรูปปริมาตร แบบที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้น กับแบบทดสอบรูปปริมาตร แบบที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

2. ค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของผู้สอบในวิชาคณิตศาสตร์

2.1 ค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบรูปปริมาตร แบบที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น เทียบกับแบบทดสอบรูปปริมาตรแบบที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้น มีค่าต่ำกว่า 1 ในช่วงระดับความสามารถทุกระดับ นั่นคือแบบทดสอบรูปปริมาตร แบบที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น มีประสิทธิภาพในการประมาณค่าต่ำกว่าแบบทดสอบรูปปริมาตร แบบที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้น ในกลุ่มผู้สอบที่ระดับความสามารถทุกระดับ

2.2 ค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบแบบทดสอบรูปปริมาตร แบบที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น เทียบกับแบบทดสอบรูปปริมาตร แบบที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น มีค่าสูงกว่า 1 ในช่วงระดับความสามารถตั้งแต่ -4.00 ถึง -2.50 และ 1.50 ถึง 4.00 แต่มีค่าต่ำกว่า 1 ในช่วงระดับความสามารถตั้งแต่ -2.40 ถึง 1.40 นั่นคือ แบบทดสอบแบบทดสอบรูปปริมาตร แบบที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น จะมีประสิทธิภาพในการประมาณค่าความสามารถสูงกว่า แบบทดสอบรูปปริมาตร แบบที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น ในกลุ่มผู้สอบที่มีความสามารถระดับต่ำมาก ($0 \leq -2.50$) ระดับสูง และระดับสูงมาก ($0 \geq 1.50$) แต่จะมีประสิทธิภาพในการประมาณค่าความสามารถต่ำกว่า แบบทดสอบรูปปริมาตร แบบที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น ในกลุ่มผู้สอบที่มีความสามารถระดับต่ำ และระดับปานกลาง ($-2.40 \leq 0 \leq 1.40$)

2.3 ค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบรูปปริมาตร แบบที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้น เทียบกับแบบทดสอบรูปปริมาตรแบบที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น มีค่าสูงกว่า 1 ในช่วงระดับความสามารถทุกระดับ นั่นคือ แบบทดสอบรูปปริมาตร แบบที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้น มีประสิทธิภาพในการประมาณค่าสูงกว่าแบบทดสอบรูปปริมาตร แบบที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น ในกลุ่มผู้สอบที่ระดับความสามารถทุกระดับ

อภิปรายผล

1. ค่าความเที่ยงตรงของการประมาณค่าความสามารถของผู้สอบในวิชาคณิตศาสตร์ โดยใช้แบบทดสอบรูปปริมาตร จำแนกตาม วิธีการตรวจให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ทำ โดยทำการเปรียบเทียบแบบทดสอบรูปปริมาตรทั้ง 3 รูปแบบ มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ซึ่งเป็นไปตามสมมติฐานที่ตั้งไว้ในข้อ 1 โดยรูปแบบที่ใช้ขนาดชั้น

คงที่ จำนวน 15 ชั้น มีค่าความเที่ยงตรงสูงสุด รองลงมาคือรูปแบบที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น และ รูปแบบที่มีข้อสอบหลายข้อในแต่ละชั้นตามลำดับ วิธีการตรวจให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบเฉพาะข้อที่ทำถูก โดยทำการเปรียบเทียบแบบทดสอบรูปปิรามิดทั้ง 3 รูปแบบ มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ซึ่งเป็นไปตามสมมติฐานที่ตั้งไว้ในข้อ 1 โดยรูปแบบที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้น มีค่าความเที่ยงตรงสูงสุด รองลงมาคือรูปแบบที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น และ รูปแบบที่มีข้อสอบหลายข้อในแต่ละชั้นตามลำดับ วิธีการตรวจให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบเฉพาะข้อที่ทำถูก พบว่า รูปแบบที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น กับ รูปแบบที่มีหลายข้อในแต่ละชั้น จำนวน 5 ชั้น และรูปแบบที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้นกับ รูปแบบที่มีหลายข้อในแต่ละชั้น จำนวน 5 ชั้น มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ซึ่งเป็นไปตามสมมติฐานที่ตั้งไว้ในข้อ 1 ส่วนรูปแบบที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้นกับรูปแบบที่มีหลายข้อในแต่ละชั้น จำนวน 5 ชั้น มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ซึ่งไม่เป็นไปตามสมมติฐานที่ตั้งไว้ในข้อ 1 โดย แบบทดสอบรูปปิรามิดแบบที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้น มีค่าความเที่ยงตรงสูงกว่าแบบทดสอบรูปปิรามิดแบบที่มีหลายข้อในแต่ละชั้น จำนวน 5 ชั้น อย่างไม่มีนัยสำคัญทางสถิติ

ทั้งนี้ในการพิจารณาค่าความเที่ยงตรงของแบบทดสอบรูปปิรามิดทั้ง 3 รูปแบบ โดยการจำแนกด้วยวิธีการตรวจให้คะแนนทั้ง 3 รูปแบบ พบว่า แบบทดสอบรูปปิรามิดแบบที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้น มีค่าความเที่ยงตรงสูงที่สุด ส่วนแบบทดสอบรูปปิรามิดแบบที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น จะมีค่าความเที่ยงตรงรองลงมา ในวิธีการตรวจให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ทำ และวิธีการตรวจให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบเฉพาะข้อที่ทำถูก ส่วนวิธีการตรวจให้คะแนนค่าความยากตามข้อสอบที่ถัดจากข้อสุดท้ายที่ได้ตอบนั้น แบบทดสอบรูปปิรามิดแบบที่มีหลายข้อในแต่ละชั้น จำนวน 5 จะมีค่าความเที่ยงตรงรองลงมาเป็นอันดับที่สอง ทั้งนี้อาจเนื่องมาจาก จำนวนของข้อสอบมีผลต่อค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ซึ่งไม่ควรน้อยเกินไป สอดคล้องกับการศึกษาของ ต่าย เชียงฉี (ต่าย เชียงฉี. 2534 : 159 - 160) ที่ใช้แบบทดสอบรูปปิรามิดที่มีจำนวนชั้น 4 รูปแบบ ได้แก่ 7 ชั้น, 10 ชั้น, 13 ชั้น และ 16 ชั้น พบว่า แบบทดสอบรูปปิรามิดที่มีขนาด 16 ชั้นจะมีค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบสูงที่สุด ส่วนค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบต่ำสุดได้แก่แบบทดสอบรูปปิรามิดที่มีขนาด 7 ชั้น ในส่วนแบบทดสอบรูปปิรามิดที่มีขนาดชั้นคงที่ จำนวน 5 ชั้น มีค่าความเที่ยงตรงรองลงมาเป็นอันดับสอง สอดคล้องกับผลการศึกษาของ แครทวอลล์ และฮายเซอร์ (Larkin and Weiss. 1974 : 4 citing Krathwohl and Huyser. 1956) ได้ศึกษาแบบทดสอบรูปปิรามิดชนิด 8 ชั้น ๑ ละ 1 ข้อ และแบบทดสอบรูปปิรามิดชนิด 4 ชั้น ๑ ละ 2 ข้อ เปรียบเทียบกับแบบทดสอบดั้งเดิมซึ่งมี 60 ข้อ ที่มีคุณลักษณะเช่นเดียวกับข้อสอบที่สร้างแบบทดสอบรูปปิรามิด พบว่า แบบทดสอบรูปปิรามิดต้องใช้คำชี้แจงมาก แต่ใช้เวลาในการสอบน้อยกว่าแบบดั้งเดิม คำสัมประสิทธิ์

สหสัมพันธ์ระหว่างแบบทดสอบรูปปริมาตรชนิด 8 ชั้น ๆ ละ 1 ข้อ กับแบบทดสอบดั้งเดิมมีค่า .78 มากกว่า แบบทดสอบรูปปริมาตรชนิด 4 ชั้น ๆ ละ 2 ข้อ กับแบบทดสอบดั้งเดิมที่มีค่า .68

2. ค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของผู้สอบในวิชาคณิตศาสตร์ระหว่าง แบบสอบรูปปริมาตรทั้ง 3 ฉบับ พบว่า ผลการวิจัยนี้ขัดแย้งและ สอดคล้องกับสมมติฐานที่ตั้งไว้ ซึ่งพิจารณาได้ดังนี้

ผลการวิจัยที่สอดคล้องกับสมมติฐาน มีดังนี้

2.1 แบบทดสอบรูปปริมาตร แบบที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น มีประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบสูงกว่า แบบทดสอบรูปปริมาตร แบบที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น ในกลุ่มผู้สอบที่มีความสามารถต่ำมาก ระดับสูงขึ้นไป

2.2 แบบทดสอบรูปปริมาตร แบบที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้น มีประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบสูงกว่า แบบทดสอบรูปปริมาตร แบบที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น ในทุกระดับความสามารถ

2.3 แบบทดสอบรูปปริมาตร แบบที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้น มีประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบสูงกว่า แบบทดสอบรูปปริมาตรที่มีข้อสอบหลายข้อในแต่ละชั้น จำนวน 5 ชั้น ในทุกระดับความสามารถ

ผลจากการวิจัยทั้ง 3 ข้อ การหาค่าประสิทธิภาพสัมพัทธ์ของแบบทดสอบ ซึ่งแสดงถึงการเปรียบเทียบคุณภาพของแบบทดสอบ โดยคำนวณจากค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบ โดยค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบได้มาจากผลรวมของฟังก์ชันของข้อสอบ ค่าฟังก์ชันอินฟอร์เมชัน เป็นฟังก์ชันของค่าพารามิเตอร์ของข้อสอบ เป็นดัชนีกำหนดคุณภาพของข้อสอบแต่ละข้อในแบบทดสอบ โดยในการวิจัยครั้งนี้ใช้ แบบ 1 พารามิเตอร์ คือจะใช้ค่าความยากของข้อสอบเป็นดัชนีชี้วัด นอกจากนี้จำนวนข้อของแบบทดสอบก็ยังมีผลต่อค่าฟังก์ชันของแบบทดสอบ โดยจำนวนข้อสอบที่มีมากขึ้นก็ส่งผลต่อค่าฟังก์ชันของแบบทดสอบทำให้ค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของผู้สอบของแบบทดสอบรูปปริมาตร แบบที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้น มีค่าสูงกว่าแบบทดสอบรูปปริมาตรทั้งสองฉบับ ซึ่งสอดคล้องกับผลการวิจัยของ พรศรี ฉิมแก้ว (พรศรี ฉิมแก้ว : 2538) ที่ศึกษาเกี่ยวกับประสิทธิภาพสัมพัทธ์ของแบบทดสอบที่มีจำนวนข้อ และจำนวนตัวเลือกแตกต่างกัน ในส่วนแบบทดสอบรูปปริมาตรที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น มีค่าประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบสูงกว่า แบบทดสอบรูปปริมาตรที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น ในกลุ่มผู้สอบที่มีความสามารถระดับต่ำ และระดับสูงขึ้นไปนั้น อาจเนื่องมาจากค่าความยากของแบบทดสอบรูปปริมาตร แบบที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น เป็นค่าการจัดข้อสอบแบบเฉลี่ยความยากในแต่ละกลุ่มข้อสอบ ทำให้ค่าความยากอยู่ในช่วงระดับปานกลาง ส่วนแบบทดสอบรูปปริมาตรที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น มีค่าความยากของข้อสอบกระจายกว่า ทำให้เมื่อผู้สอบทำข้อสอบสามารถประมาณค่าความสามารถของผู้สอบได้ดีกว่า

ผลการวิจัยที่ขัดแย้งกับสมมติฐาน มีดังนี้

แบบทดสอบรูปปิรามิดที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้น มีประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบต่ำกว่าแบบทดสอบรูปปิรามิดที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น ในกลุ่มผู้สอบที่มีความสามารถปานกลาง และระดับต่ำ ผลจากการวิจัยที่ขัดแย้งกับสมมติฐานดังกล่าว เมื่อพิจารณาค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบทั้ง 3 แบบ พบว่า ค่าสูงสุดของค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบรูปปิรามิด แบบที่ใช้ขนาดชั้นคงที่ จำนวน 5 ชั้นอยู่ที่ระดับความสามารถ 0.0 ส่วนแบบทดสอบรูปปิรามิด แบบที่ใช้ขนาดชั้นคงที่ จำนวน 15 ชั้น และแบบทดสอบรูปปิรามิด แบบที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น อยู่ที่ระดับความสามารถ -0.10 ตามลำดับ กล่าวคือถ้าวัดค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบมาก ค่าความเคลื่อนไหวก็จะน้อยลง (Hambleton and Cook 1977 : 86 – 87) ถ้าวัดค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบมีค่าสูงสุด ณ ระดับความสามารถใด แสดงว่าแบบทดสอบนั้นมีประสิทธิภาพในการวัดสูงสุดที่ระดับความสามารถนั้น ก็จะทำให้ความคลาดเคลื่อนมาตรฐานในการประมาณค่า ณ ระดับความสามารถนั้นลดลง ทำให้มีความแม่นยำในการประมาณค่า ณ ระดับความสามารถนั้นสูงตามไปด้วย เมื่อพิจารณาคูณภาพของแบบทดสอบในกลุ่มผู้สอบระดับความสามารถปานกลาง และกลุ่มผู้สอบระดับความสามารถต่ำนั้น อาจจะทำให้มีการแกว่ง (Fluctuation) ของคุณลักษณะภายในหรือความสามารถในการวัดนั้นไม่แจ่มแจ้งแบบปกติ จึงน่าจะทำให้มีผลการวิจัยที่ขัดแย้งกับสมมติฐาน จึงไม่สามารถสรุปได้ชัดเจน ในกลุ่มผู้สอบที่มีความสามารถในระดับปานกลาง และระดับต่ำ

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำไปใช้

ในการเลือกใช้แบบทดสอบเพื่อวัดผลสัมฤทธิ์ทางการเรียนให้มีความถูกต้องแม่นยำสูงควรเลือกใช้แบบทดสอบที่มีความเหมาะสมกับระดับความสามารถของผู้สอบ โดยในกลุ่มผู้สอบที่มีความสามารถระดับปานกลาง และระดับต่ำ ควรเลือกใช้แบบทดสอบรูปปิรามิดที่มีจำนวนข้อสอบหลายข้อในแต่ละชั้น จำนวน 5 ชั้น ในส่วนของผู้ที่มีความสามารถระดับต่ำมาก ระดับสูง และระดับสูงขึ้นไป ควรเลือกใช้แบบทดสอบรูปปิรามิดที่ใช้ขนาดชั้นคงที่ 15 ชั้น หรือ 5 ชั้น

2. ข้อเสนอแนะในการวิจัยต่อไป

2.1 ควรมีการศึกษาในลักษณะเดียวกัน โดยใช้รูปแบบของการทดสอบแบบทেলอร์ชนิดอื่นที่แตกต่างกันออกไป และควรนำคอมพิวเตอร์มาใช้ในการดำเนินการสอบเพื่อความสะดวกและรวดเร็วในการดำเนินการสอบ ตรวจคำตอบ และประมาณค่าความสามารถของผู้สอบ

2.2 ควรมีการศึกษาเกี่ยวกับแบบทดสอบรูปปิรามิดในรูปแบบอื่น ๆ วิธีการให้คะแนน จำนวนชั้น จำนวนข้อของแบบทดสอบ

2.3 ควรศึกษาเพิ่มเติมเกี่ยวกับการนำค่าพารามิเตอร์ของแบบทดสอบที่แตกต่างออกไปได้แก่ 2 พารามิเตอร์ 3 พารามิเตอร์ ศึกษาเกี่ยวกับ ค่าอำนาจจำแนก ค่าสัมประสิทธิ์การเดา ที่จะนำมาคัดเลือกข้อสอบเพื่อสร้างเป็นแบบทดสอบรูปปิรามิดต่อไป

บรรณานุกรม

บรรณานุกรม

- เกศมณี พยัคฆ์. (2543). การเปรียบเทียบความคลาดเคลื่อนมาตรฐานในการประมาณค่าพารามิเตอร์ของแบบทดสอบความเข้าใจในการอ่านที่มีจำนวนข้อและจำนวนกลุ่มตัวอย่างต่างกันที่วิเคราะห์ด้วยราล์ชโมเดล. ปรินญานิพนธ์ กศ.ม. (การวัดผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- ขรรค์ชัย คงเสน่ห์. (2530). การสร้างแบบทดสอบสองขั้นตอน โดยใช้ทฤษฎีการตอบข้อคำถาม. ปรินญานิพนธ์ กศ.ม. (การวัดผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร
- จิราพร ไกรสรทิวเวา. (2529). ความสัมพันธ์ระหว่างรูปแบบบางประการของแบบทดสอบรูปปริมาตรกับความสมารถทางการเรียน ในวิชาคณิตศาสตร์. ปรินญานิพนธ์ กศ.ม. (การวัดผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- ด้าย เชียงฉ. (2534). การศึกษาเปรียบเทียบประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบจากการทดสอบเทเลอร์รูปปริมาตรที่มี รูปแบบ จำนวนชั้น และวิธีการให้คะแนนที่แตกต่างกัน โดยใช้วิธีมอนติคาร์โล. กศ.ด. (การวิจัยและพัฒนาหลักสูตร) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- นันทิยา พึ่งคำ. (2531). การเปรียบเทียบคุณภาพการทดสอบแบบซี เอ ที และแบบประเพณีนิยมในการวัดความสามารถด้านคำศัพท์ภาษาอังกฤษ ของนักเรียนชั้นมัธยมศึกษาปีที่ 3. ปรินญานิพนธ์ ค.ม. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย.
- ประสาร สุขสอน. (2536). การสร้างโปรแกรมคอมพิวเตอร์เพื่อเป็นรูปแบบของการทดสอบโดยยุทธวิธีสองขั้นตอน. วิทยานิพนธ์ กศ.ม. (การวัดผลการศึกษา) พิษณุโลก : บัณฑิตวิทยาลัย มหาวิทยาลัยนเรศวร. ถ่ายเอกสาร.
- ประเวช เวชชะ. (2530). การศึกษาความเที่ยงตรงในการจำแนกความรอบรู้และเปรียบเทียบผล การทดสอบบางประการจากการใช้แบบทดสอบอิงเกณฑ์และแบบทดสอบเทเลอร์ (Tailored Test) บางรูปแบบ. วิทยานิพนธ์ ศศ.ม. เชียงใหม่ : มหาวิทยาลัยเชียงใหม่.
- ผจงจิต อินทสุวรรณ. (2525, มกราคม - เมษายน). "Latent Trait Theory" วารสารการวัดผลการศึกษา. 3(3) : 51 – 69.
- _____. (2528). ทฤษฎีการตอบข้อคำถาม. กรุงเทพฯ : สถาบันวิจัยพฤติกรรมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.
- พรศรี ฉิมแก้ว. (2538). การศึกษาประสิทธิภาพสัมพัทธ์ และคุณภาพของแบบทดสอบแบบเลือกตอบที่มีจำนวนข้อ และจำนวนตัวเลือกต่างกัน ด้วยโมเดลโลจิสติกสามพารามิเตอร์. วิทยานิพนธ์ กศ.ม. (การวัดผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย

มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

ระจิต ตรีพุทธรัตน์ และคนอื่น. (2535). ความสอดคล้องของคะแนนผลการสอบจากแบบทดสอบ
ปริมาตรในวิชาเคมีระดับชั้นมัธยมศึกษาปีที่ 5 ที่วิเคราะห์ข้อสอบโดยทฤษฎีคุณลักษณะ
แฝงกับทฤษฎีดั้งเดิม. คณะศึกษาศาสตร์ขอนแก่น : มหาวิทยาลัยขอนแก่น.

วัชรารมณ์ สุริยาภิวัณ. (2534). คอมพิวเตอร์เบื้องต้นและเทคนิคการเขียนโปรแกรม.
พิมพ์ครั้งที่ 8. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย.

ล้วน สายยศ และอังคณา สายยศ. (2522). สถิติวิทยาทางการศึกษา. กรุงเทพฯ : วัฒนาพานิช.

_____. (2543). เทคนิคการวัดผลการเรียนรู้. กรุงเทพฯ :

ภาควิชาการวัดผลและวิจัยการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.

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

สำเร็จ บุญเรืองรัตน์. (2525, 1 – 3 กันยายน – ธันวาคม). การพัฒนาทฤษฎีลาแทนท์เพื่อ
วิเคราะห์ข้อสอบ. วารสารวัดผลการศึกษา 4(2).

สุธรรม จันทร์หอม. (2526, มกราคม – เมษายน). วิเคราะห์ผลการสอบแบบบราซซ์โมเดล.
วารสารวัดผลการศึกษา 4(3) : 45 – 61.

เสรี ปรมขลิตโรจน์. (2531). การศึกษาความเที่ยงตรงในการจำแนกความรอบรู้และการ
ประมาณค่าความสามารถ ในวิชาคณิตศาสตร์ของแบบทดสอบเพลิกซีเลเวลและแบบ
ทดสอบรูปปริมาตรที่ใช้ขนาดขั้นแปรผัน. ปริญญาโท กศ.ม. (การวัดผลการศึกษา)
กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

อรินทร์ น่วมถนอม. (2537). การศึกษาความเที่ยงตรงและประสิทธิภาพสัมพัทธ์ของการ
ประมาณค่าความสามารถในวิชาคณิตศาสตร์ ระหว่างแบบทดสอบดั้งเดิม แบบทดสอบ
สองขั้นตอน และแบบทดสอบเพลิกซีเลเวล. ปริญญาโท กศ.ม. (การวัดผลการ
ศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

Betz, N.E. and D.J. Weiss. (1973). *An Empirical Study of Computer-Administered Two-
Stage Ability Testing*. Research Report 73 – 4. Minneapolis : University of
Minnesota, Psychometric Methods Program, Department of Psychology.

Birnbaum, Allan. (1968). *Some Latent Trait Model and Their Use in Inferring in
Examinee's Ability*, In F.M. Novick. *Statistical Theories of Mental Test Scores*.
Reading, Massachusetts : Addison – Wesley.

Green, Bert F. and others. (1984, Winter). Technical Guidelines for Assessing
Computerized Adaptive Tests. *Journal of Educational Measurement*. 21(4) : 347
– 360.

- Hambleton, R.K. and Linda L. Cook. (1977, Summer). Latent Trait Models and Their Use in the Analysis of Educational Test Data. *Journal of Education Measurement*. 14(2) : 75 – 96.
- Hambleton, R.K. and Hariharan Swaminathan. (1985). *Item Response Theory*. Boston : Kluwer – Nijhoff Publishing.
- Hambleton, R.K. and others. (1987) "Developments in Latent Trait Theory : Model, Technical Issues, and Applications," *Review of Educational Research*. 48 (4) : 467 – 510.
- Krathwohl, D.R. and Huyser, R.J. (1956). The sequential item test (SIT). *American Psychologist*. 2(1) : 419.
- Larkin, K.C. and Weiss, D.J. (1974). *An Empirical Investigation of Computer-Administered Pyramidal Ability Testing*. Research Report 74-3 Psychometric Methods Program, Department of Psychology, University of Minnesota.
- Larkin, K.C. and Weiss, D.J. (1975). *An Empirical Comparison of Two-Stage and Pyramidal Adaptive Ability Testing*, Research Report 75 –1 . Minneapolis : University of Minnesota, Department of Psychology, Psychometric Methods Program.
- Linn, R.L., D.A. Rock. And T.A. Cleary. (1969). The Development and Evaluation of Several Programmed Testing Methods. *Educational and Psychological Measurement*. 29(1) : 129 – 146.
- Lord, F.M. and Novick, M.R. (1968). *Statistical Theories of Mental Test Scores*. Massachusetts, Addison-Wesley Publishing Company Inc., 568 p.
- Lord, F.M. (1971, Summer). A Theoretical Study of the Measurement Effectiveness of Flexilevel Tests. *Educational and Psychological Measurement*. 31(4) : 805 – 813.
- _____. (1971, Summer). Robbins – Monro Procedures for Tailored Testing. *Educational and Psychological Measurement*. 31(1) : 3 – 31
- _____. (1980). *Applications of Item Response Theory to Practical Testing Problem*. Hillsdale, New Jersey : Lawrence Erlbaum Associates.
- McBride, James R. and John T. Martin. (1983). "Reliability and Validity of Adaptive Ability Tests in a Military Setting," in *NEW HORIZONS IN TESTING : Latent Trait Test Theory and Computerized Adaptive Testing*. 223 – 236. New York : Academic Press, Inc.
- Rudner, Lawrence. (1983). *A Closer Look At Latent Trait parameter Invariance*,

- Educational and Psychological Measurement*. 43 : 951 – 955.
- Steel, Robert G.D. and James H. Torrie. (1981). *Principles and Procedures of Statistics*. 2nd ed. Singapore : McGraw – Hill, Inc.,
- Thorndike, Robert L. (1961). *Measurement and Evaluation in Psychology and Education*. New Wiley : 575 p.
- Urry, V.W. (1977, Winter). Tailored Testing : A Successful Application of Latent Trait Theory. *Journal of Educational Measurement*. 14(2) : 181 – 196.
- Vijver, Fons J.R. Van De. (1986). "The Robustness of Rasch Estimates," *Applied Psychological Measurement*. 10(1) : 45 – 57.
- Warm, Thomas (1978). *A. A Primer of Item Response Theory*. Technical Report 941278. Oklahoma : U.S. Coast Guard Institute.
- Waters, C.W. and Bayroff, A.G. (1971). A comparison of computer-simulated conventional and Braching tests. *Educational and Psychological Measurement*. 31(1) : 125 – 136.
- Weiss, D.J. (1974). *Strategies of Adaptive Ability Measurement*. Research Report 74 – 5. Minneapolis : University of Minnesota, Psychometric Methods Program, epartment of Psychology.
- Weiss, J.J. and G. Gage Kingsbury. (1984). "Application of Computerized Adaptive Tesinting to Educational Problems," *Journal of Educational Measurement*. 21 (4) : 361 – 375.
- Wood, R. (1969). The ifficiency of tailored testing. *Educational Research*. 11(1) : 219-222.
- Wright, Benjamin D. and N. Panchapakesan. (1969). "A Procedure for Sample Free Item Analysis," *Educational and Psychological measurement*. 29 : 23 – 48.
- Wright, Benjamin D. and Mark H. Stone. (1979). *Best Test Design*. Chicago : Mesa Press.

ภาคผนวก

ภาคผนวก ก

ผลการวิเคราะห์องค์ประกอบที่มีค่าไอเกนมากกว่า 1
เพื่อตรวจสอบความเป็นมิติเดียวของแบบทดสอบที่นำไปทดลองสอบ
จำนวน 6 ฉบับ

ตาราง 16 ค่าไอเกน (มากกว่า 1) เพอร์เซ็นต์ของความแปรปรวน และเพอร์เซ็นต์ของความแปรปรวนสะสม จากการนำแบบทดสอบ จำนวน 6 ฉบับ มาวิเคราะห์องค์ประกอบ

แบบทดสอบ ฉบับที่	องค์ประกอบที่	ค่าไอเกน	เพอร์เซ็นต์ของ ความแปรปรวน	เพอร์เซ็นต์ของ ความแปรปรวน สะสม
1	1	2.863	47.709	47.709
	2	1.032	17.204	64.913
2	1	3.021	50.350	50.350
3	1	3.726	62.100	62.100
4	1	3.227	53.785	53.785
5	1	2.766	46.093	46.093
6	1	2.148	35.802	35.802
	2	1.660	27.675	63.477

ภาคผนวก ข

ค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบรูปปิรามิด 3 รูปแบบ

ตาราง 17 ค่าฟังก์ชันอินฟอร์เมชัน (TIF) ของแบบทดสอบรูปปิรามิดแบบที่มีขนาดชั้นคงที่จำนวน 5 ชั้น (x_1) แบบทดสอบรูปปิรามิดแบบที่มีขนาดชั้นคงที่จำนวน 15 ชั้น (x_2) แบบทดสอบรูปปิรามิดแบบที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น (x_3)

ระดับความสามารถ	TIF		
	x_1	x_2	x_3
-4.00	1.7434	11.1219	1.2321
-3.90	1.8725	12.0766	1.3472
-3.80	2.0074	13.0951	1.4717
-3.70	2.1480	14.1788	1.6060
-3.60	2.2939	15.3287	1.7505
-3.50	2.4450	16.5455	1.9056
-3.40	2.6009	17.8293	2.0718
-3.30	2.7613	19.1797	2.2494
-3.20	2.9258	20.5960	2.4385
-3.10	3.0940	22.0766	2.6393
-3.00	3.2655	23.6195	2.8519
-2.90	3.4398	25.2224	3.0761
-2.80	3.6167	26.8819	3.3118
-2.70	3.7955	28.5945	3.5587
-2.60	3.9758	30.3559	3.8162
-2.50	4.1573	32.1614	4.0837
-2.40	4.3393	34.0055	4.3604
-2.30	4.5214	35.8824	4.6454
-2.20	4.7030	37.7855	4.9375
-2.10	4.8836	39.7077	5.2355
-2.00	5.0625	41.6414	5.5380
-1.90	5.2392	43.5782	5.8434
-1.80	5.4129	45.5091	6.1501

ตาราง 17 (ต่อ)

ระดับความสามารถ	TIF		
	x_1	x_2	x_3
-1.70	5.5831	47.4246	6.4563
-1.60	5.7489	49.3145	6.7601
-1.50	5.9097	51.1679	7.0595
-1.40	6.0648	52.9736	7.3525
-1.30	6.2133	54.7199	7.6367
-1.20	6.3546	56.3947	7.9101
-1.10	6.4880	57.9858	8.1701
-1.00	6.6128	59.4809	8.4146
-0.90	6.7283	60.8678	8.6410
-0.80	6.8338	62.1348	8.8471
-0.70	6.9288	63.2706	9.0304
-0.60	7.0128	64.2649	9.1887
-0.50	7.0852	65.1081	9.3198
-0.40	7.1456	65.7920	9.4219
-0.30	7.1936	66.3098	9.4933
-0.20	7.2289	66.6560	9.5325
-0.10	7.2512	66.8272	9.5386
0.00	7.2603	66.8213	9.5108
0.10	7.2562	66.6383	9.4491
0.20	7.2388	66.2801	9.3537
0.30	7.2083	65.7501	9.2254
0.40	7.1647	65.0538	9.0653
0.50	7.1084	64.1979	8.8752

ตาราง 17 (ต่อ)

ระดับความสามารถ	TIF		
	x_1	x_2	x_3
0.60	7.0397	63.1908	8.6571
0.70	6.9591	62.0421	8.4135
0.80	6.8670	60.7623	8.1472
0.90	6.7640	59.3629	7.8611
1.00	6.6507	57.8559	7.5584
1.10	6.5280	56.2536	7.2423
1.20	6.3964	54.5687	6.9161
1.30	6.2568	52.8136	6.5831
1.40	6.1100	51.0005	6.2463
1.50	5.9567	49.1415	5.9085
1.60	5.7976	47.2481	5.5727
1.70	5.6335	45.3312	5.2411
1.80	5.4652	43.4011	4.9160
1.90	5.2933	41.4676	4.5992
2.00	5.1183	39.5395	4.2923
2.10	4.9410	37.6252	3.9967
2.20	4.7618	35.7323	3.7134
2.30	4.5813	33.8679	3.4431
2.40	4.3999	32.0381	3.1863
2.50	4.2181	30.2486	2.9434
2.60	4.0363	28.5045	2.7145
2.70	3.8551	26.8103	2.4994
2.80	3.6748	25.1697	2.2981

ตาราง 17 (ต่อ)

ระดับความสามารถ	TIF		
	x_1	x_2	x_3
2.90	3.4959	23.5860	2.1101
3.00	3.3190	22.0620	1.9351
3.10	3.1444	20.5998	1.7725
3.20	2.9728	19.2011	1.6219
3.30	2.8045	17.8670	1.4826
3.40	2.6402	16.5980	1.3541
3.50	2.4801	15.3945	1.2357
3.60	2.3249	14.2560	1.1268
3.70	2.1749	13.1819	1.0268
3.80	2.0304	12.1712	0.9350
3.90	1.8917	11.2225	0.8510
4.00	1.7592	10.3341	0.7741

ภาคผนวก ค

ค่าประสิทธิภาพสัมพัทธ์ (RE) ของการประมาณค่าความสามารถของผู้สอบระหว่าง
แบบทดสอบรูปปิรามิด แบบที่มีขนาดชั้นคงที่จำนวน 5 ชั้น (x_1)
แบบทดสอบรูปปิรามิด แบบที่มีขนาดชั้นคงที่จำนวน 15 ชั้น (x_2)
และแบบทดสอบรูปปิรามิดแบบที่มี ข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น (x_3)

ตาราง 18 ค่าประสิทธิภาพสัมพัทธ์ (RE) ของการประมาณค่าความสามารถของผู้สอบระหว่าง
 แบบทดสอบรูปปิรามิดแบบที่มีขนาดชั้นคงที่จำนวน 5 ชั้น (x_1) แบบทดสอบรูปปิรามิด
 แบบที่มีขนาดชั้นคงที่จำนวน 15 ชั้น (x_2) แบบทดสอบรูปปิรามิดแบบที่มี
 ข้อสอบหลายข้อในแต่ละชั้นจำนวน 5 ชั้น (x_3)

ระดับความสามารถ	RE		
	x_1/x_2	x_1/x_3	x_2/x_3
-4.00	0.1568	1.4150	9.0268
-3.90	0.1550	1.3899	8.9639
-3.80	0.1533	1.3640	8.8979
-3.70	0.1515	1.3375	8.8289
-3.60	0.1497	1.3105	8.7570
-3.50	0.1478	1.2831	8.6825
-3.40	0.1459	1.2554	8.6056
-3.30	0.1440	1.2276	8.5268
-3.20	0.1421	1.1999	8.4463
-3.10	0.1401	1.1723	8.3646
-3.00	0.1383	1.1450	8.2821
-2.90	0.1364	1.1182	8.1994
-2.80	0.1345	1.0920	8.1169
-2.70	0.1327	1.0665	8.0351
-2.60	0.1310	1.0418	7.9545
-2.50	0.1293	1.0180	7.8755
-2.40	0.1276	0.9952	7.7987
-2.30	0.1260	0.9733	7.7243
-2.20	0.1245	0.9525	7.6528

ตาราง 18 (ต่อ)

ระดับความสามารถ	RE		
	x_1/x_2	x_1/x_3	x_2/x_3
-2.10	0.1230	0.9328	7.5843
-2.00	0.1216	0.9141	7.5192
-1.90	0.1202	0.8966	7.4577
-1.80	0.1189	0.8801	7.3997
-1.70	0.1177	0.8647	7.3455
-1.60	0.1166	0.8504	7.2949
-1.50	0.1155	0.8371	7.2480
-1.40	0.1145	0.8249	7.2049
-1.30	0.1135	0.8136	7.1654
-1.20	0.1127	0.8034	7.1295
-1.10	0.1119	0.7941	7.0973
-1.00	0.1112	0.7859	7.0688
-0.90	0.1105	0.7786	7.0440
-0.80	0.1100	0.7724	7.0232
-0.70	0.1095	0.7673	7.0064
-0.60	0.1091	0.7632	6.9939
-0.50	0.1088	0.7602	6.9860
-0.40	0.1086	0.7584	6.9828
-0.30	0.1085	0.7578	6.9849
-0.20	0.1085	0.7583	6.9925
-0.10	0.1085	0.7602	7.0060

ตาราง 18 (ต่อ)

ระดับความสามารถ	RE		
	x_1/x_2	x_1/x_3	x_2/x_3
0.00	0.1087	0.7634	7.0258
0.10	0.1089	0.7679	7.0523
0.20	0.1092	0.7739	7.0859
0.30	0.1096	0.7814	7.1271
0.40	0.1101	0.7903	7.1761
0.50	0.1107	0.8009	7.2334
0.60	0.1114	0.8132	7.2993
0.70	0.1122	0.8271	7.3741
0.80	0.1130	0.8429	7.4581
0.90	0.1139	0.8604	7.5515
1.00	0.1150	0.8799	7.6546
1.10	0.1160	0.9014	7.7674
1.20	0.1172	0.9249	7.8901
1.30	0.1185	0.9504	8.0226
1.40	0.1198	0.9782	8.1650
1.50	0.1212	1.0081	8.3170
1.60	0.1227	1.0404	8.4785
1.70	0.1243	1.0749	8.6492
1.80	0.1259	1.1117	8.8286
1.90	0.1276	1.1509	9.0163
2.00	0.1294	1.1924	9.2117

ตาราง 18 (ต่อ)

ระดับความสามารถ	RE		
	x_1/x_2	x_1/x_3	x_2/x_3
2.10	0.1313	1.2363	9.4140
2.20	0.1333	1.2823	9.6225
2.30	0.1353	1.3306	9.8365
2.40	0.1373	1.3809	10.0548
2.50	0.1394	1.4331	10.2766
2.60	0.1416	1.4870	10.5009
2.70	0.1438	1.5424	10.7265
2.80	0.1460	1.5991	10.9525
2.90	0.1482	1.6568	11.1777
3.00	0.1504	1.7152	11.4011
3.10	0.1526	1.7740	11.6217
3.20	0.1548	1.8329	11.8385
3.30	0.1570	1.8916	12.0507
3.40	0.1591	1.9497	12.2575
3.50	0.1611	2.0071	12.4581
3.60	0.1631	2.0633	12.6518
3.70	0.1650	2.1182	12.8383
3.80	0.1668	2.1715	13.0170
3.90	0.1686	2.2230	13.1876
4.00	0.1702	2.2726	13.3500

รายนามผู้เชี่ยวชาญ

รายนามผู้เชี่ยวชาญ

- | | |
|------------------------------------|---|
| 1. รองศาสตราจารย์ฉวีวรรณ เศวตมัลย์ | สาขาวิชาการมัธยมศึกษา
คณะศึกษาศาสตร์
มหาวิทยาลัยศรีนครินทรวิโรฒ |
| 2 อาจารย์ ดร. ละเอียต รักษ์เฝ้า | ภาควิชาการวัดผลและวิจัยการศึกษา
คณะศึกษาศาสตร์
มหาวิทยาลัยศรีนครินทรวิโรฒ |
| 3. อาจารย์สุชาดา จันทรแพง | อาจารย์ 2 ระดับ 7
โรงเรียนเทศบาล 4
(วัดมหาธาตุวรวิหาร) |
| 4. อาจารย์ชนธิชา เศตะพราหมณ์ | อาจารย์ 1 ระดับ 5
โรงเรียนบ้านเก่าวิทยา |
| 5. อาจารย์สุภาพร สุขมูล | อาจารย์ 1 ระดับ 4
โรงเรียนเทศบาล 5
(พหลโยธินรามินทรภักดี) |

ประวัติย่อของผู้ทำวิจัย

ประวัติย่อของผู้วิจัย

ชื่อ นายอัศวิน ชื่อสกุล คงเพชรศักดิ์
เกิดวันที่ 29 เดือนสิงหาคม พุทธศักราช 2519
สถานที่เกิด อำเภอเมืองอ่างทอง จังหวัดอ่างทอง
สถานที่อยู่ปัจจุบัน บ้านเลขที่ 43/3 หมู่ 2 ตำบลโคกหม้อ
 อำเภอเมืองราชบุรี จังหวัดราชบุรี โทรศัพท์ 032-316677
ตำแหน่งหน้าที่การงาน อาจารย์ 1 ระดับ 4
สถานที่ทำงานปัจจุบัน โรงเรียนเทศบาล 4 (วัดมหาธาตุวรวิหาร)
 เทศบาลเมืองราชบุรี
 จังหวัดราชบุรี โทรศัพท์ 032-337465

ประวัติการศึกษา

พ.ศ. 2539	ประกาศนียบัตรวิชาชีพชั้นสูง (ปวส.) จากวิทยาลัยเทคนิคราชบุรี (คณะบริหารธุรกิจ)
พ.ศ. 2541	บ.ธ.บ. (ระบบสารสนเทศ) จาก คณะบริหารธุรกิจ ศูนย์กลางสถาบันเทคโนโลยีราชมงคล
พ.ศ. 2547	กศ.ม. (การวัดผลการศึกษา) จากมหาวิทยาลัยศรีนครินทรวิโรฒ