

321, 260 B

๓๓๔๔

๙.๓

การศึกษาเปรียบเทียบประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบ จากการทดสอบ
เทเลอร์รูปปิรามิดที่มี รูปแบบ จำนวนชั้น และวิธีการให้คะแนน ที่แตกต่างกัน
โดยใช้วิธีมอนติ คาร์โล

ปริณิธานิพนธ์

ของ

คำย เชียงฉี

10 ก.ย. 2534

เสนอต่อมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร เพื่อเป็นส่วนหนึ่งของการศึกษา

ตามหลักสูตรปริญญาการศึกษาคุณวุฒิ สาขาการวิจัยและพัฒนาหลักสูตร

กุมภาพันธ์ 2534

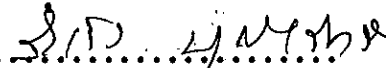
ลิขสิทธิ์เป็นของมหาวิทยาลัยศรีนครินทรวิโรฒ

ปรีชญานินพนธ์นี้ได้รับพระราชทานทุน "ภูมิพล" ผู้วิจัยรู้สึกสำนึกในพระมหากรุณาธิคุณเป็น
ล้นเกล้าล้นกระหม่อมหาที่สุดมิได้

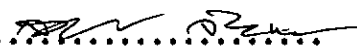
คำย เชียงฉี

คณะกรรมการควบคุมและคณะกรรมการสอบได้พิจารณาปฏิญานพนธ์ฉบับนี้แล้ว
เห็นสมควรเป็นส่วนหนึ่งของการศึกษาค้นคว้าหลักสูตรปริญญาการศึกษาคุณวุฒิบัณฑิต สาขา
การวิจัยและพัฒนาหลักสูตร ของมหาวิทยาลัยศรีนครินทรวิโรฒได้

คณะกรรมการควบคุม

.......... ประธาน

(ศ. ดร. สำเริง บุญเรืองรัตน์)

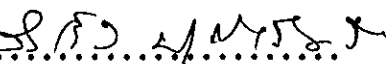
.......... กรรมการ

(รศ. ดร. สธรรม จันทน์หอม)

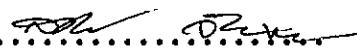
.......... กรรมการ

(อ. ดร. ชุติศักดิ์ ชัมภลชิต)

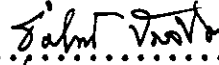
คณะกรรมการสอบ

.......... ประธาน

(ศ. ดร. สำเริง บุญเรืองรัตน์)

.......... กรรมการ

(รศ. ดร. สธรรม จันทน์หอม)

.......... กรรมการ

(อ. ดร. ชุติศักดิ์ ชัมภลชิต)

.......... กรรมการที่แต่งตั้งเพิ่มเติม

(อ. ดร. สงบ ลักษณะ)

บัณฑิตวิทยาลัยอนุมัติให้รับปฏิญานพนธ์ฉบับนี้ เป็นส่วนหนึ่งของการศึกษาค้นคว้าหลักสูตร
ปริญญาการศึกษาคุณวุฒิบัณฑิต สาขาการวิจัยและพัฒนาหลักสูตร ของมหาวิทยาลัยศรีนครินทรวิโรฒ

.......... คณบดีบัณฑิตวิทยาลัย

(ศ. ดร. สมพร บัวทอง)

วันที่ ...1... เดือน ...มิถุนายน... พ.ศ. ๒๕๓๔.

ประกาศคุณประการ

ในการทําปริญญานิพนธ์นี้ ผู้วิจัยได้รับความกรุณาจากศาสตราจารย์ ดร. สำเริง บุญเรืองรัตน์ รองศาสตราจารย์ ดร. สุธรรม จันทน์หอม และอาจารย์ ดร. ชูศักดิ์ ชัมภลชีต ที่ได้ให้คำปรึกษา แนะนำในการทําปริญญานิพนธ์นี้ตั้งแต่ต้นจนสำเร็จลงไปด้วยดี ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูง ณ โอกาสนี้ และขอขอบพระคุณอาจารย์ ดร. สงบ ลักษณะ ที่ได้ช่วยเหลือในการตรวจสอบ กลั่นกรอง และให้คำแนะนำอันเป็นประโยชน์ยิ่งสำหรับการทําปริญญานิพนธ์นี้

ขอขอบพระคุณรองศาสตราจารย์ ดร. สมบูรณ์ ชิตพงศ์ ที่ได้ให้คำแนะนำและให้กำลังใจในการทําปริญญานิพนธ์นี้

นอกจากนี้ขอขอบคุณผู้อำนวยการสำนักทดสอบทางการศึกษา กรมวิชาการ กระทรวงศึกษาธิการ ที่อนุญาตให้ใช้ข้อมูลการตรวจสอบคุณภาพทางการศึกษา อาจารย์กนก อินทรพฤกษ์ และอาจารย์ประพันธ์ พุทธิรักษา ที่ได้ให้ความช่วยเหลือในการสัมภาษณ์เพื่อนำมาใช้ในการวิจัยครั้งนี้

สุดท้ายนี้ขอขอบคุณผู้อำนวยการสำนักคอมพิวเตอร์ อาจารย์และเจ้าหน้าที่ในสำนักคอมพิวเตอร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร ที่ได้อนุญาต และให้ความสะดวกในการใช้ไมโครคอมพิวเตอร์ ในการจำลองข้อมูล

คำย เชียงฉี

สารบัญ

บทที่	หน้า
1 บทนำ	1
ความเป็นมาและความสำคัญของปัญหา	1
กรอบความคิด	7
วัตถุประสงค์ของการวิจัย	15
ตัวแปรที่ศึกษา	16
สมมุติฐานในการวิจัย	17
ประโยชน์ที่คาดว่าจะได้รับ	18
ข้อตกลงเบื้องต้น	18
ขอบเขตในการวิจัย	18
นิยามศัพท์เฉพาะ	19
2 เอกสารและผลงานวิจัยที่เกี่ยวข้อง	20
ทฤษฎีการตอบข้อสอบ (Item Response Theory: IRT)	20
หลักการของทฤษฎีการตอบข้อสอบ	20
เงื่อนไขเบื้องต้นเกี่ยวกับการใช้ทฤษฎีการตอบข้อสอบ (IRT)	22
โมเดลของทฤษฎีการตอบข้อสอบ (Item Response Model)	24
นอร์มัล โอจีฟ โมเดล (Normal Ogive Model)	24
โลจิสติก โมเดล (Logistic Model)	25
ราซค์ โมเดล (Rasch Model)	30
การนำเอาทฤษฎีการตอบข้อสอบมาประยุกต์ใช้	32
การทดสอบเทเลอร์ (Tailored Testing)	35
ความหมายและหลักการในการทดสอบเทเลอร์	35
ประเภทของการทดสอบเทเลอร์	36

การทดสอบทีเลอร์ โดยใช้วิธีสองขั้นตอน	37
การทดสอบทีเลอร์ โดยใช้วิธีหลายขั้นตอนแบบแยกทางคงที่	38
การทดสอบทีเลอร์ โดยใช้วิธีหลายขั้นตอนแบบแยกทางแปรผัน	49
วิธีการให้คะแนนการทดสอบทีเลอร์ (Scoring)	49
ผลงานวิจัยที่เกี่ยวข้องกับการทดสอบทีเลอร์	55
ผลงานวิจัยที่เกี่ยวข้องภายในประเทศ	55
ผลงานวิจัยต่างประเทศ	58
การศึกษาโดยใช้เทคนิควิธีมอนติ คาร์โล (Monte Carlo Method)	62
ความเป็นมาของวิธีมอนติ คาร์โล	62
ขั้นตอนของระเบียบวิธีมอนติ คาร์โล	63
จุดเด่นของการใช้เทคนิควิธีมอนติ คาร์โล	68
การนำเอาเทคนิควิธีมอนติ คาร์โล มาศึกษาเกี่ยวกับการทดสอบทีเลอร์	68
3 วิธีดำเนินการวิจัย	70
ประชากรและกลุ่มตัวอย่าง	70
ขั้นตอนในการทดลอง	70
การวิเคราะห์ข้อมูล	98
ขั้นตอนในการวิเคราะห์ข้อมูล	98
4 ผลการวิจัย	101
ผลการศึกษาและ เปรียบเทียบค่าความคลาดเคลื่อนในการประมาณค่าความ	
สามารถของผู้สอบ (E) จากการทดสอบทีเลอร์รูปปริมาตรที่มี รูปแบบ	
จำนวนชั้น และ วิธีการให้คะแนน ที่แตกต่างกัน	102

ผลการศึกษาและ เปรียบเทียบค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า

ความสามารถของผู้สอบ (SEE) จากการทดสอบเทเลอร์รูปปิรามิดที่มี

รูปแบบ จำนวนชั้น และวิธีการให้คะแนน ที่แตกต่างกัน 117

ผลการศึกษาและ เปรียบเทียบค่าความเที่ยงตรงในการประมาณค่า

ความสามารถของผู้สอบ (re θ) จากการทดสอบเทเลอร์รูปปิรามิดที่มี

รูปแบบ จำนวนชั้น และวิธีการให้คะแนน ที่แตกต่างกัน 131

5 สรุป อภิปรายผล และข้อเสนอแนะ 147

วัตถุประสงค์ของการวิจัย 147

ตัวแปรที่ศึกษา 147

สมมุติฐานในการวิจัย 148

ประชากรและกลุ่มตัวอย่าง 149

รูปแบบการทดลอง 149

ขั้นตอนในการทดลอง 149

การวิเคราะห์ข้อมูล 150

สรุปผลการวิจัย 150

อภิปรายผลการวิจัย 155

ข้อเสนอแนะ 161

บรรณานุกรม 164

ภาคผนวก 172

ประวัติย่อของผู้วิจัย 179

บัญชีตาราง

ตาราง

หน้า

1	แสดงค่าต่ำสุด สูงสุด ค่าเฉลี่ย และค่าส่วนเบี่ยงเบนมาตรฐาน ของค่าความ สามารถจริง (θ_1) ของกลุ่มตัวอย่าง	71
2	แสดงค่าความยากเป็นรายข้อ (β_j) ของข้อสอบการทดสอบเทเลอ์รูปปิรามิด แบบขนาดชั้นคงที่ ที่มี 7 ชั้น	75
3	แสดงค่าความยากเป็นรายข้อ (β_j) ของข้อสอบการทดสอบเทเลอ์รูปปิรามิด แบบขนาดชั้นแปรผัน ที่มี 7 ชั้น	81
4	แสดงค่าความยากเป็นรายข้อ (β_j) ของข้อสอบการทดสอบเทเลอ์รูปปิรามิด แบบขนาดชั้นแปรผัน ที่มี 10 ชั้น	81
5	แสดงค่าความยากเป็นรายข้อ (β_j) ของข้อสอบการทดสอบเทเลอ์รูปปิรามิด แบบขนาดชั้นแปรผัน ที่มี 13 ชั้น	82
6	แสดงค่าความยากเป็นรายข้อ (β_j) ของข้อสอบการทดสอบเทเลอ์รูปปิรามิด แบบขนาดชั้นแปรผัน ที่มี 16 ชั้น	83
7	แสดงค่าโอกาสในการตอบข้อสอบถูก ของผู้สอบแต่ละคนในข้อสอบ แต่ละข้อ (P_{ij})	85
8	แสดงผลการตอบข้อสอบ (เมื่อตอบถูกให้ $u_{ij}=1$; ตอบผิดให้ $u_{ij}=0$) ของผู้สอบแต่ละคนในข้อสอบแต่ละข้อ	86
9	แสดงตัวอย่าง สมมุติว่าผลการตอบของผู้สอบ (3 คนแรก ในข้อสอบ 55 ข้อ)	87
10	แสดงค่าความยากของข้อสอบ (β_j) และผลการตอบของผู้สอบคนที่ 1	90
11	แสดงผลการประมาณค่าความสามารถของผู้สอบ โดยใช้วิธีความเป็นไปได้สูงสุด แบบมีเงื่อนไข (Conditional Maximum Likelihood) ของผู้สอบคนที่ 1..	96

12	แสดงผลการทดสอบความแตกต่างระหว่างความสามารถจริงกับความสามารถที่ได้ จากการประมาณค่า ในแต่ละกลุ่มการทดลอง โดยทำการทดลองซ้ำ 40 ครั้ง ...	103
13	แสดงค่าเฉลี่ยของค่าเฉลี่ย ($\bar{E}$) และส่วนเบี่ยงเบนมาตรฐานของค่าเฉลี่ย ($SD\bar{E}$) ความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ ในแต่ละกลุ่มการทดลอง โดยทำการทดลองซ้ำ 40 ครั้ง	105
14	แสดงผลการวิเคราะห์ความแปรปรวนของค่าเฉลี่ยความคลาดเคลื่อนในการ ประมาณค่าความสามารถของผู้สอบ ($\bar{E}$) ในแต่ละกลุ่มการทดลอง โดยทำการ ทดลองซ้ำ 40 ครั้ง	110
15	แสดงผลการทดสอบความแตกต่างของค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่า ความสามารถของผู้สอบ เป็นรายคู่ โดยใช้วิธีของ เชฟเฟ (Scheffe')	113
16	แสดงค่าเฉลี่ย ($\overline{SEE}$) และส่วนเบี่ยงเบนมาตรฐาน (SD_{SEE}) ของค่า ความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ในแต่ละกลุ่มการทดลอง โดยทำการทดลองซ้ำ 40 ครั้ง	118
17	แสดงผลการวิเคราะห์ความแปรปรวนของค่าความคลาดเคลื่อนมาตรฐานในการ ประมาณค่าความสามารถของผู้สอบ (SEE) ในแต่ละกลุ่มการทดลอง โดยทำ การทดลองซ้ำ 40 ครั้ง	123
18	แสดงผลการทดสอบความแตกต่างของค่าเฉลี่ยความคลาดเคลื่อนมาตรฐาน ในการประมาณค่าความสามารถของผู้สอบ เป็นรายคู่ โดยใช้วิธีของ เชฟเฟ (Scheffe')	126
19	แสดงช่วงของค่า $re\hat{\theta}$, นัยสำคัญของค่า $re\hat{\theta}$, $\bar{Z}_r$, SD_{Z_r} , และค่า $\overline{re\hat{\theta}}$ ในแต่ละกลุ่มการทดลอง โดยทำการทดลองซ้ำ 40 ครั้ง	131

- 20 แสดงผลการวิเคราะห์ความแปรปรวนของค่าความเที่ยงตรงในการประมาณค่า
 ความความสามารถของผู้สอบ (ในรูปของ Fisher's $Z: Z_r$) ในแต่ละกลุ่ม
 การทดลองโดยการทดลองซ้ำ 40 ครั้ง 137
- 21 แสดงผลการทดสอบความแตกต่างของค่าเฉลี่ยความเที่ยงตรงในการประมาณค่า
 ความความสามารถของผู้สอบ ในรูปของ พิชเชอร์ซี ($\bar{Z}_r$) เป็นรายคู่ โดยใช้
 วิธีของ เชฟเฟ่ (Scheffe') 140

บัญชีภาพประกอบ

ภาพประกอบ

หน้า

1	โค้งแสดงลักษณะข้อสอบ (ICC) แบบ 3-พารามิเตอร์ (Three-Parameters Logistic Model)	29
2	โค้งแสดงลักษณะข้อสอบ (ICC) แบบราซค์ โมเดล	31
3	แสดงโครงสร้างของการทดสอบเทเลอร์แบบสองชั้นตอน	38
4	แสดงโครงสร้างของการทดสอบเทเลอร์รูปปिरามิดแบบขนาดชั้นคงที่ ที่มี 6 ชั้น	39
5	แสดงโครงสร้างของการทดสอบเทเลอร์รูปปिरามิดแบบขนาดชั้นแปรผันที่มี 6 ชั้น ...	40
6	แสดงโครงสร้างของการทดสอบเทเลอร์รูปปिरามิดแบบروبิน-มอนโร (Robin-Monro)	42
7	แสดงโครงสร้างของการทดสอบเทเลอร์รูปปिरามิดแบบข้างตัด (Truncated Pyramid) สกิดการสะท้อนกลับ (Reflecting Barrier)	43
8	แสดงโครงสร้างของการทดสอบเทเลอร์รูปปिरามิดแบบข้างตัด (Truncated Pyramid) รักษาการสะท้อนกลับ (Retaining Barrier)	44
9	แสดงโครงสร้างของการทดสอบเทเลอร์รูปปिरามิด แบบมีข้อสอบหลายข้อในแต่ละชั้น (Multiple-Item Model)	45
10	แสดงโครงสร้างของการทดสอบเทเลอร์รูปปिरามิดแบบให้หน้านักแก้ตัวเลือกของข้อสอบ เพื่อแยกทาง (Differential Response Option Branching) ..	46
11	แสดงตัวอย่างข้อสอบแบบเฟล็กซ์เลเวล (Flexilevel Test)	47
12	แสดงโครงสร้างของการทดสอบเทเลอร์แบบสเตรดิคเด้ฟัลฟ์ เทสคิง (Stradaptive Testing)	48
13	แสดงโครงสร้างและค่าความยากของข้อสอบการทดสอบเทเลอร์รูปปिरามิดแบบขนาดชั้นคงที่ ที่มี 7 ชั้น	74

14	แสดงโครงสร้างและค่าความยากของข้อสอบการทดสอบเทเลอร์ รูปปิรามิดแบบขนาดชั้นคงที่ ที่มี 10 ชั้น	76
15	แสดงโครงสร้างและค่าความยากของข้อสอบการทดสอบเทเลอร์ รูปปิรามิดแบบขนาดชั้นคงที่ ที่มี 13 ชั้น	77
16	แสดงโครงสร้างและค่าความยากของข้อสอบการทดสอบเทเลอร์ รูปปิรามิดแบบขนาดชั้นคงที่ ที่มี 16 ชั้น	78
17	แสดงโครงสร้างของการทดสอบเทเลอร์รูปปิรามิดแบบขนาดชั้นแปรผัน ที่มี 7 ชั้น	79
18	แสดงพื้นที่ของค่าโอกาสในการตอบข้อสอบถูก (P_{ij})	86
19	แสดงเส้นทางในการตอบ การทดสอบเทเลอร์รูปปิรามิดแบบขนาดชั้นคงที่ ที่มี 10 ชั้น ของผู้สอบในตัวอย่าง (3 คนแรก)	89
20	แสดงรูปแบบการทดลอง	98
21	กราฟเปรียบเทียบค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่า ความสามารถของผู้สอบ ($\bar{E}$) จากการทดสอบเทเลอร์รูปปิรามิด ที่มี รูปแบบต่างกัน	107
22	กราฟเปรียบเทียบค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่า ความสามารถของผู้สอบ ($\bar{E}$) จากการทดสอบเทเลอร์รูปปิรามิด ที่มี จำนวนชั้นต่างกัน	108
23	กราฟเปรียบเทียบค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่า ความสามารถของผู้สอบ ($\bar{E}$) จากการทดสอบเทเลอร์รูปปิรามิด ที่มี วิธีการให้คะแนนที่ต่างกัน	109

- 24 กราฟเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานในการประมาณค่า
ความสามารถของผู้สอบ ($\overline{SEE}$) จากการทดสอบเทเลอร์รูปปิรามิด ที่มี
รูปแบบต่างกัน 120
- 25 กราฟเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานในการประมาณค่า
ความสามารถของผู้สอบ ($\overline{SEE}$) จากการทดสอบเทเลอร์รูปปิรามิด ที่มี
จำนวนชั้นต่างกัน 121
- 26 กราฟเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานในการประมาณค่า
ความสามารถของผู้สอบ ($\overline{SEE}$) จากการทดสอบเทเลอร์รูปปิรามิด ที่มี
วิธีการให้คะแนนที่ต่างกัน..... 122
- 27 กราฟเปรียบเทียบค่าเฉลี่ยความเที่ยงตรงในการประมาณค่าความสามารถของ
ผู้สอบ ($\overline{r_{95}}$) จากการทดสอบเทเลอร์รูปปิรามิด ที่มีรูปแบบต่างกัน 134
- 28 กราฟเปรียบเทียบค่าเฉลี่ยความเที่ยงตรงในการประมาณค่าความสามารถของ
ผู้สอบ ($\overline{r_{95}}$) จากการทดสอบเทเลอร์รูปปิรามิด ที่มีจำนวนชั้นต่างกัน 135
- 29 กราฟเปรียบเทียบค่าเฉลี่ยความเที่ยงตรงในการประมาณค่าความสามารถของ
ผู้สอบ ($\overline{r_{95}}$) จากการทดสอบเทเลอร์รูปปิรามิด ที่มีวิธีการให้คะแนนที่ต่างกัน.. 136

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

หัวใจของการจัดการศึกษา คือ การจัดการเรียน การสอน และการสอบ ทั้งสามอย่างนี้ จะต้องสอดคล้องและดำเนินไปในทิศทางเดียวกัน ในสมัยก่อนจำนวนนักเรียนยังมีน้อย การสอนใช้วิธีสอนแบบตัวต่อตัว การสอบก็ใช้วิธีสอบปากเปล่าเป็นรายบุคคล ต่อมาจำนวนนักเรียนเพิ่มขึ้น การสอนก็ใช้วิธีสอนเป็นกลุ่ม หรือ เป็นชั้นเรียนเล็ก ๆ การสอบก็ใช้วิธีเขียนตอบแบบความเรียง (Essay Test) สอบเป็นกลุ่มหรือเป็นชั้นเรียน และต่อมาจำนวนนักเรียนได้เพิ่มมากขึ้นอีก ห้องเรียนได้ขยายใหญ่ขึ้น สำหรับการสอนครูคนหนึ่งต้องสอนหลาย ๆ ห้องเรียน การสอบก็ได้พัฒนามาเป็นการใช้ข้อสอบแบบปรนัย (Objective Test) ทั้งนี้เพื่อให้การตรวจให้คะแนนทำได้ง่ายขึ้น ข้อสอบมีความเป็นปรนัยมากขึ้น และให้ค่าความเชื่อมั่นสูงขึ้น (ชาวล แพร่ศกุล. 2509: 2) มาในปัจจุบัน การเรียนการสอนกลับมาเน้นความแตกต่างเป็นรายบุคคลอีก โดยเฉพาะอย่างยิ่งจากการเสนอแนวคิดเกี่ยวกับการเรียนเพื่อรอบรู้ (Mastery learning) ของ บลูม (Bloom. 1976: 111) และวิธีสอนที่มีประสิทธิภาพสูงสุด คือ การสอนแบบตัวต่อตัว (One-to-One Tutoring) (Bloom. 1984: 4-15) โดยมีความเชื่อว่า ถ้าให้เวลาแก่ผู้เรียนอย่างเพียงพอแล้ว นักเรียนทุกคนมีโอกาที่จะเรียนรู้ได้เท่าเทียมกัน แต่จะใช้เวลาในการเรียนแตกต่างกัน ฉะนั้นการเรียนการสอนจึงได้หันมาเน้นในเรื่องของ จุดประสงค์การเรียนรู้เป็นรายบทเรียน โดยมีเงื่อนไขว่า ถ้าผู้เรียนคนใดผ่านจุดประสงค์การเรียนรู้ในบทเรียนนั้นแล้วก็สามารถที่จะ ไปเรียนในบทต่อไปได้ แต่ถ้ายังไม่ผ่านก็จะมี การสอนซ่อมเสริม และวิธีสอนก็ได้พัฒนาเครื่องช่วยสอนต่าง ๆ มาใช้ เช่น ชุดการสอน บทเรียนสำเร็จรูป คอมพิวเตอร์ช่วยสอน (Computer Assisted Instruction: CAI) เป็นต้น สำหรับการสอบก็ได้มีการหาวิธี ที่จะจัดข้อสอบให้เหมาะสมกับระดับความสามารถของผู้สอบเป็นรายบุคคล นั่นคือ การทดสอบเพอร์ (Tailored Testing)

แนวคิดเกี่ยวกับการทดสอบเทเลอร์ (Tailored Testing) ได้มีมานานแล้ว เทอร์นบูล (Lord. 1980: 151; citing Turnbull. 1951) เป็นผู้ตั้งชื่อนี้ ในระยะแรก ๆ รู้จักกันเฉพาะในวงการนักวิจัย และนักวัดผลการศึกษาเท่านั้น ยังไม่ได้นำมาใช้อย่างจริงจัง (Urry. 1977: 181) ทั้งนี้เพราะยังมีปัญหาเกี่ยวกับความยุ่งยากในการดำเนินการสอบ เมื่อใช้กับการทดสอบที่ใช้กระดาษกับดินสอ (Paper and Pencil Test) ต่อมาในระยะหลังนี้ ได้มีการนำเอาคอมพิวเตอร์ มาช่วยในการดำเนินการสอบ การจัดข้อสอบให้เหมาะสมกับระดับความสามารถของผู้สอบเป็นรายบุคคล พร้อมทั้งประยุกต์เอาทฤษฎีการตอบข้อสอบ (Item Response Theory: IRT) มาใช้ในการประมาณค่าความสามารถของผู้สอบ จึงทำให้การทดสอบเทเลอร์ (Tailored Testing) สามารถนำมาใช้ได้อย่างมีประสิทธิภาพ (Rudner. 1983: 951; Weiss and Kingsbury. 1984: 362) ซึ่งนับเป็นก้าวใหม่อีกก้าวหนึ่งในวงการวัดผลการศึกษา (Urry. 1977: 181; Dodd and others. 1989: 129)

การทดสอบเทเลอร์ (Tailored Testing) หรือการทดสอบ โดยการจัดข้อสอบให้เหมาะสมกับระดับความสามารถของผู้สอบเป็นรายบุคคล (Adaptive Testing) นั้น มีหลักการอยู่ว่า ถ้าผู้สอบทำข้อสอบข้อที่ยาก ๆ ได้ถูกต้องแล้ว ข้อที่ง่าย ๆ ก็น่าจะทำได้ถูกต้องด้วย จึงไม่จำเป็นต้องให้ทำข้อเหล่านั้นก็ได้ หรือถ้าทำข้อที่ง่าย ๆ ผิดแล้ว ข้อที่ยาก ๆ ก็น่าจะทำได้ด้วย จึงไม่จำเป็นต้องให้ทำข้อเหล่านั้นก็ได้ ดังที่ กรีนและคนอื่น ๆ (Green and others. 1984: 356) ได้เปรียบเทียบว่า ถ้านักกระโดดสูงที่กระโดดสูง 72 นิ้วไม่ข้ามแล้ว ไม่จำเป็นต้องให้ลองกระโดดสูง 78 นิ้วก็ได้ นอกจากนั้น ข้อสอบที่ง่ายเกินไปหรือยากเกินไป จะให้รายละเอียดเกี่ยวกับตัวผู้สอบค่า (Vale and Weiss. 1975: 6) และมีแนวโน้มว่า จะทำให้ค่าความเชื่อมั่นของคะแนนต่ำกว่าที่ควรจะเป็นด้วย ทั้งนี้เพราะเด็กอ่อนจะตอบข้อสอบข้อที่ยากมาก ๆ โดยการเดา และข้อง่าย ๆ ก็จะไม่ยั่วให้เด็กเก่งให้ตั้งใจทำ (Weiss. 1974: 1)

การประมาณค่าความสามารถของผู้สอบด้วยการทดสอบเทเลอร์ โดยปกติจะเริ่มให้ผู้สอบทำข้อสอบข้อที่มีความยากปานกลางก่อน ถ้าทำถูกข้อต่อไปก็จะให้ทำข้อที่ยากขึ้น แต่ถ้าทำผิดข้อต่อไปก็จะให้ทำข้อที่ง่ายขึ้น การทำข้อสอบในข้อต่อ ๆ ไป ก็จะปฏิบัติ เช่น เดียวกันนี้ จนถึงที่สุดการทดสอบเทเลอร์ สำหรับการคัดเลือกข้อสอบที่จะให้ผู้สอบทำในแต่ละขั้นนั้น จะเป็นไปตามรูปแบบ

และเงื่อนไขของการทดสอบเทเลอร์นั้น ๆ (Lord. 1971: 3-4)

รูปแบบของการทดสอบเทเลอร์ พอสรุปได้เป็น 2 รูปแบบใหญ่ ๆ และในแต่ละรูปแบบยังแยกย่อยได้เป็น ดังนี้

1. ยุทธวิธีสองขั้นตอน (Two-Stage Strategies) หมายถึง การทดสอบเทเลอร์ที่ประกอบด้วยข้อสอบ 2 ชุด ชุดแรกมีไว้สำหรับแยกกลุ่มผู้สอบตามระดับความสามารถ (Routing Test) ส่วนชุดที่สองเป็นข้อสอบหลัก หรือ ข้อสอบวัดจริง (Main or Measurement Test) ข้อสอบหลัก หรือข้อสอบวัดจริงนี้ ได้แยกเป็นข้อสอบชุดย่อย ๆ (Subtest) หลายฉบับตามระดับความยาก ถ้าผู้สอบได้รับการประเมินจากข้อสอบชุดแรก (Routing Test) ว่า มีความสามารถต่ำก็จะได้รับข้อสอบหลักที่ง่าย มีความสามารถปานกลางก็จะได้รับข้อสอบหลักที่ยากปานกลาง และถ้ามีความสามารถสูงก็จะได้รับข้อสอบหลักที่ยาก

2. ยุทธวิธีหลายขั้นตอน (Multi-Stage Strategies) ซึ่งยังแยกออกได้เป็น 2 รูปแบบ คือ

2.1 แบบแยกทางคงที่ (Fixed-Branching Model) ได้แก่ รูปแบบที่มีการกำหนดข้อสอบ และเส้นทางในการตอบข้อสอบที่แน่นอน ซึ่งยังแยกออกได้อีกหลายรูปแบบ ดังนี้

2.1.1 รูปปริมาตรแบบขนาดขั้นคงที่ (Constant Step Size Pyramid) ได้แก่การทดสอบเทเลอร์ที่จัดเป็นรูปสามเหลี่ยม โดยมีจุดยอดเป็นจุดเริ่มต้นในการสอบซึ่งมีข้อสอบเพียงข้อเดียวและมีค่าความยากปานกลาง ขั้นตอนต่อไปจะมีจำนวนข้อสอบเพิ่มขึ้นเรื่อย ๆ โดยมีช่วงห่างค่าความยากของข้อที่อยู่ติดกันภายในชั้น มีค่าเท่ากันตลอด

2.1.2 รูปปริมาตรแบบขนาดขั้นแปรผัน (Variable Step Size Pyramid) เป็นรูปแบบที่เสนอโดย แพทเตอร์สัน และ ลอร์ด (Weiss. 1974: 18; citing Paterson. 1962; Lord. 1971) ซึ่งเขามีความเห็นว่าเป็นว่า รูปปริมาตรแบบขนาดขั้นคงที่ ขาดประสิทธิภาพเกี่ยวกับความไวในการกำหนดเส้นทางในการตอบข้อสอบ จึงได้กำหนดช่วงห่างค่าความยากของข้อที่อยู่ติดกันภายในชั้นมีค่าไม่เท่ากัน กล่าวคือ ในช่วงกลาง ๆ จะมีค่ามาก แล้วค่อย ๆ ลดลงในข้อที่ค่อนข้างง่ายหรือยากและต่อมา ลอร์ด (Weiss. 1974: 20; citing

Lord. 1971) ได้เพิ่มจำนวนข้อในแต่ละชั้นให้เป็นสองเท่าของจำนวนข้อในชั้นที่มาก่อน ซึ่งให้ชื่อเรียกรูปแบบใหม่นี้ว่า แบบรอนบิน-มอนโร (Robin-Monro)

2.1.3 รูปปริมาตรแบบข้างตัด (Truncated Pyramid) ซึ่งเป็นรูปแบบที่เสนอโดย มุสลีโอ (Weiss. 1974: 22; citing Mussio. 1973) เพื่อให้ลดจำนวนข้อสอบลงจากรูปปริมาตรแบบขนาดชั้นคงที่ จึงใช้วิธีสกัดการสะท้อนกลับ หรือ วิธีรักษาการสะท้อนกลับ (Reflecting or Retaining Barrier)

2.1.4 รูปปริมาตรแบบมีหลายข้อในแต่ละชั้น (Multiple-Item model) เป็นรูปแบบที่ แครธวอร์ท กับฮายเซอร์ และลินน์ (Weiss. 1974: 25; citing Krathwohl and Huyser. 1956; Linn. 1969) เป็นผู้เสนอรูปแบบนี้ขึ้นมา โดยให้มีจำนวนข้อในแต่ละชั้นมีมากกว่าหนึ่งข้อ

2.1.5 รูปปริมาตรแบบให้น้ำหนักแก่ตัวเลือกของข้อสอบเพื่อแยกทาง (Differential Response Option Branching) เป็นรูปแบบที่เสนอโดย เบย์รอฟฟ์และซีลีย์ (Weiss. 1974: 28; citing Bayroff and Seeley. 1967) รูปแบบนี้เหมาะสมสำหรับข้อสอบเลือกตอบที่มีการกำหนดน้ำหนักคะแนนของตัวเลือกของข้อสอบ

2.2 แบบแยกทางแปรผัน (Variable-Branching Model) ได้แก่รูปแบบที่ไม่มีการกำหนดข้อสอบและเส้นทางในการตอบข้อสอบไว้แน่นอน โดยมากจะใช้คอมพิวเตอร์ เป็นผู้เลือกข้อสอบให้มีค่าความยากที่เหมาะสมกับระดับความสามารถของผู้สอบ

ในปัจจุบันนี้ได้มีการนำเอาคอมพิวเตอร์มาประยุกต์ใช้กับการทดสอบเทเลอร์ ซึ่งเรียกว่า คอมพิวเตอร์ไรซ์ อแดปทีฟ เทสติ้ง (Computerized Adaptive Testing: CAT) โดยให้คอมพิวเตอร์เป็นผู้ประมาณค่าความสามารถของผู้สอบ และจัดข้อสอบให้เหมาะสมกับระดับความสามารถของผู้สอบ (Weiss. 1982: 473-492; Weiss and Kingsbury. 1984: 361-375; Moreno and others. 1984: 155-163; Ayala. 1989: 789) นอกจากนี้ยังได้นำเอาทฤษฎีการตอบข้อสอบ (Item Response Theory: IRT) มาใช้ในการประมาณค่าความสามารถของผู้สอบ จึงทำให้การประมาณค่าความสามารถของผู้สอบได้เที่ยงตรง (Validity)

และเชื่อถือได้ (Reliability) ยิ่งขึ้น โดยใช้จำนวนข้อสอบน้อยข้อกว่าเดิม และใช้เวลาในการสอบน้อยกว่าเดิมด้วย ดังผลการวิจัยของ ไวส์ (Weiss. 1982: 473-492) และ โมรีโน และคนอื่น ๆ (Moreno and others. 1984: 155-163) พบว่า การทดสอบโดยใช้คอมพิวเตอร์ไรซ์ อแคคทิฟ เทสติ้ง (CAT) ให้ค่าความเที่ยงตรง และความเชื่อมั่นเท่ากับ หรือมากกว่า การทดสอบแบบดั้งเดิม (Conventional Test) ทั้ง ๆ ที่ใช้จำนวนข้อของข้อสอบน้อยกว่าเดิมกว่าครึ่งหนึ่ง ด้วยเหตุนี้การใช้คอมพิวเตอร์ไรซ์ อแคคทิฟ เทสติ้ง (CAT) จึงเริ่มมีบทบาทเข้ามาแทนที่การสอบโดยการเขียนตอบแบบเดิม (Green and others. 1984: 347-360; Weiss and Kingsbury. 1984: 361-375; Lee. 1987: 3742-A; Blackmore. 1987: 2554-A) โดยเฉพาะอย่างยิ่งข้อสอบมาตรฐานที่มีชื่อเสียงหลายฉบับในสหรัฐอเมริกา ได้มีการปรับปรุงจากการสอบแบบเดิมที่ใช้กระดาษคินสอ มาเป็นการสอบแบบคอมพิวเตอร์ไรซ์ อแคคทิฟ เทสติ้ง (CAT) เช่น แบบทดสอบ เอเอ็สวีเอบี (Armed Services Vocational Aptitude Battery: ASVAB) โมรีโน, เวทเซล, เม็กซ์ไบรด์และไวส์ และคูเด็ค (Henly and others. 1989: 363; citing Moreno, Wetzel, McBride and Weiss. 1984; Cudeck. 1985) ได้พบว่า การทดสอบแบบคอมพิวเตอร์ไรซ์ อแคคทิฟ เทสติ้ง สามารถใช้สอบแทนการทดสอบแบบดั้งเดิมได้เกือบทุกฉบับย่อย (Arithmetic Reasoning, Word Knowledge, General Science, Mathematics Knowledge and Paragraph Comprehension) นอกจากสองฉบับย่อยที่สอบโดยอาศัยความเร็ว (Speed Test) (Numerical Operations and Coding Speed) ซึ่งไม่เหมาะสมที่จะสอบโดยใช้คอมพิวเตอร์ไรซ์ อแคคทิฟ เทสติ้ง (CAT) ซึ่งก็สอดคล้องกับผลการวิจัยของ เฮนลี่และคนอื่น ๆ (Henly and others. 1989: 363-371) ที่ศึกษาโดยเปรียบเทียบข้อสอบ ดีเอที (Differential Aptitude Test: DAT) ฉบับที่เป็นคอมพิวเตอร์ไรซ์ อแคคทิฟ เทสติ้ง กับ ฉบับดั้งเดิม (Paper and Pencil Test) ได้พบว่า การทดสอบแบบคอมพิวเตอร์ไรซ์ อแคคทิฟ เทสติ้ง สามารถใช้สอบแทนการทดสอบแบบดั้งเดิมได้เกือบทุกฉบับย่อย (Verbal Reasoning, Numerical Ability, Abstract Reasoning, Mechanical Reasoning, Space Relations, Spelling and Language Usage) เช่นนั้น นอกจากสองฉบับย่อยที่สอบโดยอาศัยความเร็ว (Speed Test) (Clerical Speed

and Accuracy) ซึ่งไม่เหมาะสมที่จะสอบโดยใช้ คอมพิวเตอร์ไรซ์ อแดปทีฟ เทสติ้ง (CAT) สำหรับเมืองไทยในอนาคตอันใกล้นี้โอกาสที่จะนำเอาคอมพิวเตอร์ไรซ์ อแดปทีฟ เทสติ้ง (Computerized Adaptive Testing) มาใช้สอบแทนการเขียนตอบแบบเดิม (Paper and Pencil Test) นั้น เป็นไปได้สูงมาก โดยเฉพาะอย่างยิ่งการวัดผลสัมฤทธิ์ระหว่างเรียน เพื่อการปรับปรุงการเรียนการสอน และการสอนซ่อมเสริม ทั้งนี้เพราะการใช้คอมพิวเตอร์ไรซ์ อแดปทีฟ เทสติ้ง (CAT) ให้ประสิทธิภาพสูงกว่าการทดสอบโดยการเขียนตอบแบบเดิม หลายประการ ดังนี้

(1) ใช้เวลาในการสอบน้อย (Urry. 1977: 194) เนื่องจากการทดสอบเทเลอ์ (Tailored Testing) ใช้จำนวนข้อสอบน้อยกว่าการทดสอบแบบดั้งเดิม ดังผลการวิจัยของ นันทิยา พิงศา (2531: 92) พบว่าการทดสอบโดยใช้คอมพิวเตอร์ไรซ์ อแดปทีฟ เทสติ้ง (CAT) สามารถลดจำนวนข้อสอบลงได้เกินกว่าครึ่งหนึ่ง โดยยังคงให้ค่าความเที่ยงตรงเท่าเดิม ซึ่งก็สอดคล้องกับผลการวิจัยของ ไวส์ (Weiss. 1982: 473-492) และโมรีโนและคณะอื่น ๆ (Moreno and others. 1984: 155-163)

(2) สามารถรายงานผลการสอบได้ทันที ผู้สอบเพียงแต่นั่งหน้าจอคอมพิวเตอร์ เมื่อคำถามขึ้นที่จอคอมพิวเตอร์แล้ว ผู้สอบก็ตอบคำถามลงไปทางแป้นพิมพ์ เครื่องคอมพิวเตอร์ก็จะตรวจสอบว่า คำตอบนั้นถูกต้องหรือไม่ แล้วก็จะประมาณค่าความสามารถของผู้สอบ พร้อมทั้งเลือกข้อสอบข้อต่อไปที่เหมาะสมกับระดับความสามารถของผู้สอบ ออกมาทางจอคอมพิวเตอร์ เพื่อให้ผู้สอบทำต่อไป การทดสอบนี้จะสิ้นสุดลง เมื่อมีความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบถึงระดับที่กำหนดไว้ (Urry. 1977: 194; Linden and Zwarte. 1989: 182; จักรณา พรหมมา. 2525: 46) เครื่องคอมพิวเตอร์จะใช้เวลาน้อยมากในการตรวจคำตอบ การประมาณค่าความสามารถของผู้สอบ และการค้นหาข้อสอบข้อที่เหมาะสมกับระดับความสามารถของผู้สอบ หลังจากสิ้นสุดการทดสอบแล้ว เครื่องคอมพิวเตอร์ยังสามารถที่จะรายงานผลการสอบได้ทันทีอีกด้วย

(3) เครื่องคอมพิวเตอร์สามารถบันทึกผลการสอบได้ โดยในขณะที่ผู้สอบทำข้อสอบอยู่นั้น เครื่องคอมพิวเตอร์สามารถที่จะเก็บบันทึกผลการสอบไว้ให้ได้ด้วย

(4) ผู้สอบที่มีความสามารถสูง ไม่เกิดความเบื่อหน่าย เพราะ ไม่ต้องตอบข้อสอบที่ง่าย ๆ และผู้สอบที่มีความสามารถต่ำก็ไม่เกิดความเครียด เพราะไม่ต้องพบกับข้อสอบที่ยาก ๆ จึงทำให้ผู้สอบมีทัศนคติที่ดีต่อการทดสอบเทเลอร์ ซึ่งจะเห็นได้จากผลการวิจัยของ งานนิตย์ ธาตุทอง และปรีชา เครือวรรณ (2527) และ ลี (Lee, 1987: 3742-A) ที่พบว่า ผู้สอบส่วนใหญ่มีทัศนคติที่ดีต่อการทดสอบเทเลอร์

(5) เป็นการสอบที่ย้ำให้ผู้ตอบคิด เพราะมีการปรับความยากของข้อสอบให้เหมาะสมกับระดับความสามารถของผู้สอบอยู่ตลอดเวลา

(6) สามารถสอบกับนักเรียนเป็นรายบุคคลได้โดยที่ครูไม่จำเป็นต้องมาดำเนินการสอบ นักเรียนพร้อมที่จะสอบเมื่อไร ก็ไปทดสอบกับเครื่องคอมพิวเตอร์ได้

จากที่กล่าวมาจะเห็นว่าการทดสอบเทเลอร์ (Tailored Testing) สามารถนำมาใช้สอบแทนการทดสอบแบบเดิมที่ใช้เขียนตอบได้เป็นอย่างดี และยังมีประสิทธิภาพสูงกว่าการทดสอบแบบเดิมหลายประการ ดังกล่าวแล้ว แต่การทดสอบเทเลอร์ยังมีหลายรูปแบบ และมีวิธีการให้คะแนนได้หลายวิธีด้วยกัน การที่จะเลือกใช้การทดสอบเทเลอร์รูปแบบใด และจะให้คะแนนโดยวิธีใดนั้นสิ่งที่ควรคำนึงถึงก็คือ ความสะดวกสบายไม่ยุ่งยากซับซ้อนในการใช้ และมีความถูกต้องแม่นยำในการประมาณค่าความสามารถของผู้สอบ นั่นคือ มีประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบ จากผลงานวิจัยที่เกี่ยวข้องยังไม่พบว่ามีผู้ใดศึกษาในเรื่องนี้มาก่อน ด้วยเหตุนี้ผู้วิจัยจึงได้ ศึกษาเปรียบเทียบประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบ จากการทดสอบเทเลอร์รูปแบบที่มี รูปแบบ จำนวนชั้น และวิธีการให้คะแนนที่แตกต่างกัน โดยใช้วิธีมอนติ คาร์โล (Monte Carlo Method) ในการศึกษา

กรอบความคิด

ในการศึกษาครั้งนี้มีกรอบความคิด ดังนี้

1. ตัวบ่งชี้ถึงประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบที่สำคัญ ที่จะศึกษาในครั้งนี้ 3 ประการ คือ

1.1 ความถูกต้องแม่นยำในการประมาณค่าความสามารถของผู้สอบ (Accuracy)

โดยปกติไม่ว่าจะเป็นการวัดหรือการประมาณค่าใด ๆ ก็ตาม จะต้องมีความคลาดเคลื่อนอยู่เสมอ โดยเฉพาะอย่างยิ่งในการวัด หรือการประมาณค่าความสามารถในทางการศึกษาและจิตวิทยา จะมีความคลาดเคลื่อนค่อนข้างสูง ถ้าความคลาดเคลื่อนในการประมาณค่า มีค่าสูง ความถูกต้องแม่นยำในการประมาณค่าก็จะต่ำ ซึ่งสามารถเขียนเป็นสมการ แสดงความสัมพันธ์ระหว่างความสามารถจริง (θ) ความสามารถที่ได้จากการประมาณค่า ($\hat{\theta}$) และค่าความคลาดเคลื่อน (E) ได้ดังนี้ (Lord and Novick. 1968: 66)

$$E = \hat{\theta} - \theta$$

จากสมการนี้จะเห็นว่า ถ้าค่าความคลาดเคลื่อน (E) มีค่าน้อยหรือเข้าใกล้ศูนย์ ความสามารถที่ได้จากการประมาณค่า ($\hat{\theta}$) ก็จะเข้าใกล้ความสามารถจริง (θ) นั่นคือ มีความถูกต้องแม่นยำในการประมาณค่าความสามารถของผู้สอบสูง

1.2 ความคลาดเคลื่อนมาตรฐานในการประมาณค่า (Standard Error of Estimate: SEE) ก็คือ ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนในการประมาณค่า ซึ่งแตกต่างจากความคลาดเคลื่อนมาตรฐานในการวัด (Standard Error of Measurement: SE_{meas}) (Lord and Novick. 1968: 66-67; Urry. 1977: 183) ตรงที่ความคลาดเคลื่อนมาตรฐานในการประมาณค่า (SEE) ใช้ความคลาดเคลื่อนที่เกิดจากการประมาณค่าความสามารถของผู้สอบ ที่คลาดเคลื่อนไปจากความสามารถจริง ($E = \hat{\theta} - \theta$) ส่วนความคลาดเคลื่อนมาตรฐานในการวัด (SE_{meas}) จะใช้ความคลาดเคลื่อนที่เกิดจากคะแนนดิบ หรือคะแนนที่ได้จากการวัด ที่คลาดเคลื่อนไปจากคะแนนจริง ($E = X - T$)

ความคลาดเคลื่อนมาตรฐานในการประมาณค่า (SEE) มีความสัมพันธ์กับค่าอินฟอร์เมชัน (Information: I) ดังนี้ $SEE = 1/\sqrt{I}$ (Hambleton and Cook. 1977: 84; Hambleton and Swaminathan. 1985: 104, 232) ค่าอินฟอร์เมชัน (I) จะเป็นตัวบ่งชี้ถึงคุณภาพของการประมาณค่า ซึ่งยังสามารถใช้แทนค่าความเชื่อมั่น (Reliability: re)

หรือค่าความคลาดเคลื่อนมาตรฐานในการวัด (Standard Error of Measurement: SE_{meas}) ในทฤษฎีการทดสอบดั้งเดิม (Classical Theory) ได้ด้วย (Hambleton and Cook, 1977: 84) กรีนและคนอื่น ๆ (Green and others, 1984: 352) ได้ชี้ให้เห็นว่า ความคลาดเคลื่อนมาตรฐานในการวัด (SE_{meas}) มีข้อดกลงเบื้องตันที่ไม่น่าจะถูกต้อง ที่ว่าความแปรปรวนของคะแนนความคลาดเคลื่อนมีค่าเท่ากันสำหรับทุก ๆ คะแนน ฉะนั้น จะเน้นการทดสอบเทเลอร์ (Tailored Testing) จึงควรใช้ความคลาดเคลื่อนมาตรฐานในการประมาณค่า (SEE) แทน โดยเฉพาะอย่างยิ่ง การทดสอบเทเลอร์ ที่ใช้การประมาณค่าความสามารถของผู้สอบโดยวิธีของเบย์ส์ (Bayesian Updating) สามารถที่จะกำหนดค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า (SEE) ค่าสุดท้ายในการทดสอบเทเลอร์แต่ละครั้ง ได้ด้วย

ส่วนค่าความเชื่อมั่นนั้น วอม (Warm, 1978: 77) ได้แสดงความสัมพันธ์ระหว่างความคลาดเคลื่อนมาตรฐานในการประมาณค่า (SEE) กับค่าความเชื่อมั่น (r_{tt}) ไว้ ดังนี้ $r_{tt} = 1 - (\overline{SEE})^2$ เมื่อ $\overline{SEE}$ คือ ค่าเฉลี่ยของ SEE ได้มีนักวัดผลการศึกษาหลายท่าน (Samejima, 1977: 243; Green and others, 1984: 352) มีความเห็นว่า ไม่ควรนำเอาการหาค่าความเชื่อมั่น (Reliability) มาใช้กับการทดสอบเทเลอร์ แต่ ไควจิ (Divgi, 1989: 146) มีความเห็นว่าค่าความเชื่อมั่นยังมีประโยชน์สำหรับการทดสอบเทเลอร์ (Tailored Testing) อยู่

ความคลาดเคลื่อนมาตรฐานในการประมาณค่า (SEE) ในการวิจัยครั้งนี้หาได้จาก สูตรดังนี้

$$SEE = \sqrt{\frac{N\sum E^2 - (\sum E)^2}{N(N-1)}}$$

เมื่อ SEE คือ ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ

E คือ ค่าความคลาดเคลื่อนในการประมาณค่า ($E = \hat{\theta} - \theta$)

N คือ จำนวนผู้สอบ

1.3 ความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ (Validity)

ความเที่ยงตรงก็เป็นดัชนี (Index) อีกตัวหนึ่ง ที่จะชี้ถึงคุณภาพของการประมาณค่าความสามารถของผู้สอบ กรีนและคนอื่น ๆ (Green and others. 1984: 354) ได้ให้ความเห็นว่า ค่าความเที่ยงตรง (Validity) มีความสำคัญสำหรับการทดสอบเทเลอร์ (Tailored Testing) ถ้าใช้ข้อสอบที่มีค่าอำนาจจำแนก (a) สูง จะทำให้เกิดความเที่ยงตรงสูงขึ้น ความเที่ยงตรงที่จะศึกษาในครั้งนี้ จะเป็นความเที่ยงตรงตามโครงสร้าง (Construct Validity) ซึ่งสามารถหาได้จาก ความสัมพันธ์ระหว่างความสามารถจริง (o) กับความสามารถที่ได้จากการประมาณค่า (ô) (Urry. 1977: 183; citing Cronbach and Meehl. 1957) โดยใช้สูตร ดังนี้

$$r_{o\hat{o}} = \frac{N\sum o\hat{o} - \sum o\sum \hat{o}}{\sqrt{[N\sum o^2 - (\sum o)^2][N\sum \hat{o}^2 - (\sum \hat{o})^2]}}$$

เมื่อ $r_{o\hat{o}}$ คือ ค่าความเที่ยงตรง (Construct Validity)

o คือ ความสามารถจริงของผู้สอบ

$\hat{o}$ คือ ความสามารถของผู้สอบที่ได้จากการประมาณค่า

N คือ จำนวนผู้สอบ

2. องค์ประกอบที่ส่งผลต่อประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบ

ในการศึกษาครั้งนี้มุ่งเน้นที่ องค์ประกอบทางด้านตัวข้อสอบของการทดสอบเทเลอร์เป็นสำคัญ และองค์ประกอบนั้นจะต้องสามารถจัดกระทำ (Manipulate) ได้ง่าย จากการศึกษาทฤษฎีและผลงานวิจัยที่เกี่ยวข้อง พบว่า องค์ประกอบทางด้านตัวข้อสอบของการทดสอบเทเลอร์ ที่ส่งผลโดยตรงต่อประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบ และสามารถที่จะจัดกระทำ (Manipulate) ได้ง่าย ที่สำคัญมี 3 ประการ คือ (1) รูปแบบของการทดสอบเทเลอร์ (2) จำนวนขั้นหรือจำนวนข้อในการทดสอบเทเลอร์ และ (3) วิธีการให้คะแนน

2.1 รูปแบบของการทดสอบเทเลอร์ (Form) รูปแบบของการทดสอบเทเลอร์

จะต้องเป็นรูปแบบที่ง่ายต่อการปฏิบัติหรือการนำไปใช้ จากการศึกษาของธอร์นไคค์ (Thordike. 1982: 317) พบว่า การทดสอบเทเลอรรูปปิรามิด (Pyramid Model) เป็นรูปแบบที่ง่ายต่อการปฏิบัติ เข้าใจง่าย และยังสามารถที่จะดัดแปลงให้ใช้สื่อทั้งกระดาษและดินสอ (Paper and Pencil Test) ได้ (จิราพร ไกรสรศิวาท. 2529; เสรี ประมวลศิริโรจน์. 2531) หรือจะใช้คอมพิวเตอร์ช่วยในการดำเนินการสอบ (CAT) ก็ได้ การทดสอบเทเลอรรูปปิรามิดยังมีหลายแบบดังได้กล่าวมาแล้ว ผู้วิจัยได้เลือกศึกษารูปปิรามิดแบบขนาดขั้นคงที่ (Constant Step Size Pyramid) เพราะเป็นรูปแบบที่มีคนนิยมใช้กันมาก และเป็นแม่แบบของการทดสอบเทเลอรรูปปิรามิดแบบอื่น ๆ กับอีกรูปแบบหนึ่งคือรูปปิรามิดแบบขนาดขั้นแปรผัน (Variable Step Size Pyramid) ซึ่งเป็นรูปแบบที่ แพทเตอร์สัน ได้พัฒนาขึ้นมาโดยอาศัยโครงสร้างของรูปปิรามิดแบบขนาดขั้นคงที่ แต่มีประสิทธิภาพเกี่ยวกับความไวในการกำหนดเส้นทางในการตอบข้อสอบมากกว่าแบบขนาดขั้นคงที่ (Wiess. 1974: 18; citing Paterson. 1962)

2.2 จำนวนขั้น (Stage) ในการทดสอบเทเลอรรูปปิรามิด จำนวนขั้นในการทดสอบเทเลอรรูป จะมีความสัมพันธ์โดยตรงกับจำนวนข้อที่ผู้สอบจะต้องทำในแบบทดสอบ และจำนวนข้อในการสอบ จะมีความสัมพันธ์โดยตรงกับค่าความเที่ยงตรง (Validity) และค่าความเชื่อมั่น (Reliability) กล่าวคือ โดยทฤษฎีแล้วจำนวนข้อสอบยิ่งมากขึ้น ค่าความเที่ยงตรง และความเชื่อมั่นก็จะยิ่งสูงขึ้น (Gulliksen. 1967: 74-105; Lord and Novick. 1968: 112-118) แต่ถ้าใช้จำนวนข้อมากเกินไป จะมีผลเสียเช่นกัน กล่าวคือ ต้องใช้จำนวนข้อสอบมากขึ้น เสียเวลาในการสอบมากขึ้น เสียค่าใช้จ่ายมากขึ้น ทำให้ผู้สอบเกิดความเบื่อหน่ายในการสอบ มีการทำข้อสอบโดยการเดามากขึ้น ซึ่งจะส่งผลต่อค่าความเที่ยงตรงและความเชื่อมั่นในทางลบก็ได้ ฉะนั้นในทางปฏิบัติการทดสอบเทเลอรรูปปิรามิดที่ปรารถนา คือ ใช้จำนวนข้อสอบ หรือจำนวนขั้นในการทดสอบน้อยที่สุด แต่ยังคงให้ค่าความถูกต้องแม่นยำ และความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบได้ไม่แตกต่างจากเดิม จากการวิจัยของ จิราพร ไกรสรศิวาท (2529) โดยใช้การทดสอบเทเลอรรูปปิรามิดที่มีขนาด 8 ขั้น ในวิชาคณิตศาสตร์ กับนักเรียนมัธยมศึกษาปีที่ 2 จำนวน 348 คน พบว่า คะแนนจากการทดสอบเทเลอรรูปมีความสัมพันธ์กับความสามารถทางการเรียน มีค่าเป็นบวกอย่างมีนัยสำคัญทางสถิติ และต่อมาในปี พ.ศ. 2531 วัฒนา พงศา

(2531: 93) ได้ใช้ ซีเอที (Computerized Adaptive Testing: CAT) ทดสอบความสามารถทางด้านศัพท์ภาษาอังกฤษ กับนักเรียนชั้นมัธยมศึกษาปีที่ 3 ได้พบว่า ผู้สอบจะได้รับข้อสอบโดยเฉลี่ยประมาณคนละ 17 ข้อ ในขณะที่การทดสอบแบบดั้งเดิมใช้ข้อสอบถึง 40 ข้อ ซึ่งให้ค่าความตรงเชิงเกณฑ์สัมพันธ์ (Criterion-Related Validity) ไม่แตกต่างกัน แต่มีค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ค่าว่าการทดสอบแบบดั้งเดิม ส่วน อะยาลา (Ayala. 1989: 796) ได้พบว่า การทดสอบ ซีเอที (CAT) ผู้สอบส่วนใหญ่ (Mode) ใช้ข้อสอบ 16 ข้อ สำหรับ ธอร์นไดค์ (Thorndike. 1982: 321) ได้ชี้ให้เห็นว่า ความคลาดเคลื่อนมาตรฐานในการประมาณค่า (Standard Error of Estimate) จะมาก หรือ น้อย นอกจากจะขึ้นอยู่กับจำนวนขั้นในการทดสอบเทเลอรัปปริมาตรแล้ว ยังขึ้นอยู่กับค่าอำนาจจำแนก (α) ของข้อสอบด้วย กล่าวคือ การทดสอบเทเลอรัปปริมาตรที่มีจำนวนขั้น (Stage) มากขึ้น หรือข้อสอบมีค่าอำนาจจำแนก (α) สูงขึ้น ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าก็จะลดลง โดยปกติแล้วถ้าข้อสอบที่มีค่าอำนาจจำแนกโดยเฉลี่ย 0.7 ควรจะใช้ข้อสอบประมาณ 10 ถึง 15 ขั้น

จากที่กล่าวมาจะเห็นว่า การทดสอบเทเลอรัปปริมาตรส่วนใหญ่จะใช้จำนวนข้อสอบ หรือจำนวนขั้นอยู่ระหว่าง 7 ถึง 16 ขั้น ฉะนั้นการวิจัยครั้งนี้จึงได้เลือกศึกษาการทดสอบเทเลอรัปปริมาตรที่มีจำนวน 7, 10, 13 และ 16 ขั้น

2.3 วิธีการให้คะแนน (Scoring) วิธีการให้คะแนน จะมีผลโดยตรงต่อการประมาณค่าความสามารถของผู้สอบ สำหรับการวิจัยครั้งนี้ ได้เลือกศึกษาวิธีการให้คะแนน 4 วิธี คือ

(1) ให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำได้ ซึ่งเป็นวิธีการให้คะแนน โดยยึดหลักการเดียวกับการกระโดดสูง ที่เป็นแนวคิดเบื้องต้นของการทดสอบเทเลอรัป (Green and others. 1984: 356) กล่าวคือ ความสูงสุดที่สามารถกระโดดข้ามได้ คือ ความสามารถในการกระโดดสูงของผู้นั้น การให้คะแนนโดยวิธีนี้ส่วนใหญ่จะประมาณค่าความสามารถของผู้สอบ สูงกว่าความสามารถจริง (Over Estimate) (ค่าย เชียงฉ. 2533: 64) แต่เป็นวิธีการให้คะแนนที่ง่าย ไม่จำเป็นต้องใช้เครื่องคอมพิวเตอร์ในการประมาณค่าความสามารถของผู้สอบ จึง

สามารถนำมาใช้กับการทดสอบเทเลอร์ ที่สอบโดยใช้กระดาษคินสอ (Paper and Pencil Test) ได้

(2) ให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทำถูก เป็นวิธีการให้คะแนนที่แก้จุดอ่อนในการให้คะแนนตามวิธีแรกที่ประมาณค่าความสามารถของผู้สอบสูงกว่าความสามารถจริง (Over Estimate) ฉะนั้นการให้คะแนนโดยวิธีนี้จึงใช้ค่ากลาง ๆ นั่นคือ ค่าเฉลี่ย (Mean) แทน การให้คะแนนโดยวิธีนี้ทำได้ง่ายเช่นเดียวกับวิธีแรก สามารถนำมาใช้กับการทดสอบเทเลอร์ที่สอบโดยใช้กระดาษคินสอได้

(3) ให้คะแนนโดยใช้วิธีของเบส์ (Bayesian Updating) เป็นวิธีการที่ โอเวน (Owen. 1969; 1975: 351-356) ได้พัฒนาขึ้นมาเพื่อใช้ในการประมาณค่าความสามารถของผู้สอบในการทดสอบเทเลอร์โดยเฉพาะ ซึ่งจะทำให้ความถูกต้องแม่นยำ (Accuracy) มากกว่าใน 2 วิธีแรกที่กล่าวมา แต่ก็มีการใช้สถิติที่ค่อนข้างซับซ้อน จึงเหมาะสำหรับการทดสอบเทเลอร์ที่ใช้คอมพิวเตอร์ช่วยในการดำเนินการสอบ (Computerized Adaptive Testing: CAT) ไวส์ และแมกซ์ไบรด์ (Wiess and McBride. 1984: 273-285) ได้ศึกษาความลำเอียง (Bias) ในการประมาณค่าความสามารถของผู้สอบ เมื่อใช้วิธีการของเบส์ (Bayesian Adaptive Testing) ได้พบว่า การประมาณค่าไม่มีความลำเอียง (Unbias) นอกจากนี้ ยูรี (Wiess and McBride. 1984: 274; citing Urry. 1971) ได้พบว่า ความสัมพันธ์ระหว่างความสามารถจริง กับความสามารถที่ได้จากการประมาณค่าโดยวิธีของเบส์ มีค่าอยู่ระหว่าง .919 ถึง .936 และเจนเซมา (Weiss and McBride. 1984: 274; citing Jensema. 1972) ก็ได้ศึกษาในทางองเดียวกันนี้ ได้พบว่า ความสัมพันธ์ระหว่างความสามารถจริง กับความสามารถที่ได้จากการประมาณค่าโดยวิธีของเบส์ มีค่าเท่ากับ .92 ซึ่งอยู่ในเกณฑ์ที่สูงมาก

(4) ให้คะแนนโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (Conditional Maximum Likelihood) เป็นวิธีการประมาณค่าความสามารถของผู้สอบที่ใช้กันอย่างแพร่หลาย และให้ความถูกต้องแม่นยำ (Accuracy) ค่อนข้างสูง แต่มีการใช้สถิติที่ค่อนข้างซับซ้อน จึงเหมาะสำหรับการทดสอบเทเลอร์ ที่ใช้คอมพิวเตอร์ช่วยในการดำเนินการสอบ (Computerized Adaptive Testing: CAT) เช่นเดียวกับวิธีของเบส์ แต่เป็นวิธีการประมาณค่าความสามารถ

ของผู้สอบ ที่มีข้อจำกัดในกรณีที่ผู้สอบทำข้อสอบถูกต้องหรือผิดหมด จะไม่สามารถประมาณค่าได้

สกแกส์และสตีเวนสัน (Skaggs and Stevenson. 1989: 391-402) ได้ศึกษาเปรียบเทียบผลการประมาณค่าพารามิเตอร์ต่าง ๆ โดยใช้วิธีการของเบส์ (Pseudo-Bayesian) กับวิธีความเป็นไปได้สูงสุด (Joint Maximum Likelihood Procedure) โดยใช้ 3-พารามิเตอร์ ผลปรากฏว่า ถ้าใช้กลุ่มตัวอย่างขนาดใหญ่มากตั้งแต่ 2,000 คนขึ้นไป และข้อสอบตั้งแต่ 35 ข้อขึ้นไป การประมาณค่าโดยวิธีทั้ง 2 นี้ จะมีความถูกต้องสูงและไม่แตกต่างกัน แต่ถ้ากลุ่มตัวอย่างลดลง ($N < 500$) การประมาณค่าโดยวิธีของเบส์ (Pseudo-Bayesian) จะให้ค่าความคงที่ (Consistency) และความถูกต้องแม่นยำ (Accuracy) มากกว่า เมื่อเทียบกับการใช้วิธีความเป็นไปได้สูงสุด (Joint Maximum Likelihood Procedure) และถ้าจำนวนข้อของข้อสอบลดลง ($k < 15$) การประมาณค่าโดยวิธีความเป็นไปได้สูงสุด จะให้ความถูกต้องแม่นยำมากกว่า เมื่อเทียบกับการใช้วิธีของเบส์ และใน ปี ค.ศ. 1990 กิฟฟอร์ดและสวามินาธาน (Gifford and Swaminathan. 1990: 33) ก็ได้ศึกษาในทำนองเดียวกันนี้ ได้พบว่า ถ้ากลุ่มตัวอย่างขนาดเล็ก และข้อสอบมีจำนวนน้อยข้อ การประมาณค่าพารามิเตอร์ต่าง ๆ โดยใช้วิธีของเบส์ (Joint Bayesian Estimate) จะให้ความถูกต้องมากกว่า เมื่อเทียบกับการใช้วิธีความเป็นไปได้สูงสุด (Joint Maximum Likelihood Estimate)

จากที่กล่าวมาจึงสรุปได้ว่า การวิจัยในครั้งนี้ได้เลือกวิธีการให้คะแนน 4 วิธี เพราะวิธีการให้คะแนนใน 2 วิธีแรก เป็นวิธีที่ง่ายต่อการปฏิบัติ ไม่มีการใช้วิธีการทางสถิติที่ซับซ้อน จึงสามารถนำมาใช้กับการทดสอบเทเลอร์ ที่ดำเนินการสอบโดยใช้ กระดาษดินสอ (Paper and Pencil Test) ได้ สำหรับวิธีการให้คะแนนใน 2 วิธีหลังนั้น เป็นวิธีการให้คะแนนที่ให้ความถูกต้องแม่นยำมากกว่าใน 2 วิธีแรก แต่มีการใช้สถิติที่ค่อนข้างซับซ้อน จึงเหมาะสำหรับการทดสอบเทเลอร์ที่ใช้คอมพิวเตอร์ช่วยในการดำเนินการสอบ (Computerized Adaptive Testing: CAT) และวิธีการให้คะแนนทั้ง 4 วิธีนี้ วิธีที่มีประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบสูงสุด ควรจะเป็นวิธีของเบส์ (Bayesian Updating) รองลงมา คือ วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (Conditional Maximum Likelihood) วิธีให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทำได้ และให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำได้ ตามลำดับ

จากเอกสารและผลงานวิจัยที่เกี่ยวข้องที่กล่าวมา จะเห็นว่าการศึกษเกี่ยวกับรูปแบบของการทดสอบเทเลอรัูปิรามิด จำนวนชั้น และวิธีการให้คะแนน ส่วนใหญ่เป็นการศึกษาในรายองค์ประกอบย่อยที่ละองค์ประกอบ ยังไม่มีการศึกษาถึงปฏิสัมพันธ์ (Interaction) ระหว่างองค์ประกอบทั้ง 3 ดังกล่าว ด้วยเหตุนี้ ในการวิจัยครั้งนี้จึงมุ่งเน้นที่จะศึกษาปฏิสัมพันธ์ระหว่างองค์ประกอบทั้ง 3 ที่จะร่วมกันส่งผลต่อความถูกต้องแม่นยำ และความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ทั้งนี้เพราะว่าในทางการปฏิบัติจริงนั้น การทดสอบเทเลอรัูปิรามิดทุกครั้งจะต้องประกอบด้วยรูปแบบ จำนวนชั้น และวิธีการให้คะแนน ทั้ง 3 องค์ประกอบนี้พร้อม ๆ กันไปเสมอ ถ้าขาดองค์ประกอบใดองค์ประกอบหนึ่งแล้ว การทดสอบเทเลอรัูปิรามิดก็จะไม่สมบูรณ์ และองค์ประกอบทั้ง 3 นี้ มีแนวโน้มว่าจะมีปฏิสัมพันธ์ (Interaction) กัน โดยเฉพาะอย่างยิ่ง จำนวนชั้น หรือจำนวนข้อที่ผู้สอบจะต้องทำ กับ วิธีการให้คะแนนโดยใช้อัตราของเบส์กับวิถีความเป็นไปได้สูงสุด ซึ่งจะเห็นได้จากผลงานวิจัยของ สแควกส์และสตีเวนสัน (Skaggs and Stevenson, 1989: 391-402)

3. วิธีการศึกษา การศึกษาครั้งนี้ได้เลือกศึกษาโดยวิธีมอนติ คาร์โล (Monte Carlo Method) ทั้งนี้เพราะ การศึกษาครั้งนี้เป็นการศึกษาในเชิงตรวจสอบทางทฤษฎี เพื่อให้สามารถศึกษาจากความสามารถจริง (True Ability: θ) ของผู้สอบได้ และยังสามารถกระทำได้หลาย ๆ ครั้ง โดยที่ไม่มีตัวแปรแทรกซ้อนอื่น ๆ เข้ามาเกี่ยวข้อง เช่น ความเหนื่อยหน่าย การจำคำตอบได้ การขาดหายไปของตัวอย่าง เป็นต้น นอกจากนี้ยังได้นำเอาทฤษฎีการตอบข้อสอบ (Item Response Theory: IRT) แบบ 1-พารามิเตอร์ โลจิสติก โมเดล (One-Parameter Logistic Model) มาประยุกต์ใช้กับการศึกษาครั้งนี้ด้วย ทั้งนี้เพื่อให้การประมาณค่าความสามารถของผู้สอบมีความถูกต้องและสามารถนำมาเปรียบเทียบกันได้

วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาเปรียบเทียบ ค่าความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ (E) จากการทดสอบเทเลอรัูปิรามิดที่มี รูปแบบ จำนวนชั้น และวิธีการให้คะแนน ที่แตกต่างกัน

2. เพื่อศึกษาเปรียบเทียบ ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ (SEE) จากการทดสอบเทเลอรรูปปิรามิดที่มี รูปแบบ จำนวนชั้น และวิธีการให้คะแนน ที่แตกต่างกัน

3. เพื่อศึกษาเปรียบเทียบ ค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ (reliability) จากการทดสอบเทเลอรรูปปิรามิดที่มี รูปแบบ จำนวนชั้น และวิธีการให้คะแนน ที่แตกต่างกัน

ตัวแปรที่ศึกษา

1. ตัวแปรอิสระ ได้แก่

1.1 รูปแบบของการทดสอบเทเลอรรูปปิรามิด มี 2 รูปแบบ คือ

- (1) แบบขนาดขั้นคงที่ (Constant Step Size Pyramid)
- (2) แบบขนาดขั้นแปรผัน (Variable Step Size Pyramid)

1.2 จำนวนชั้นของการทดสอบเทเลอรรูปปิรามิด มี 4 แบบ คือ

- (1) แบบที่มีจำนวน 7 ชั้น
- (2) แบบที่มีจำนวน 10 ชั้น
- (3) แบบที่มีจำนวน 13 ชั้น
- (4) แบบที่มีจำนวน 16 ชั้น

1.3 วิธีการให้คะแนนในการทดสอบเทเลอรรูปปิรามิด มี 4 วิธี คือ

- (1) ให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำถูก ✓
- (2) ให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทำถูก ✓
- (3) ให้คะแนนโดยใช้วิธีของเบย์ (Bayesian Updating)
- (4) ให้คะแนนโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข

(Conditional Maximum Likelihood)

2. ตัวแปรตาม ได้แก่

- (1) ความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ (E)

- (2) ความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ (SEE)
- (3) ความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ (reô)

สมมติฐานในการวิจัย

การวิจัยครั้งนี้ ผู้วิจัยได้กำหนด ประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบสูง หมายถึง (1) มีความคลาดเคลื่อนในการประมาณค่าต่ำ (2) มีความคลาดเคลื่อนมาตรฐานในการประมาณค่าต่ำ และ (3) มีความเที่ยงตรงในการประมาณค่าสูง

จากกรอบความคิด และผลงานวิจัยที่เกี่ยวข้องที่กล่าวมา จะเห็นว่า รูปแบบ การทดสอบ เทเลอร์รูปปริมาตรแบบขนาดชั้นแปรผัน มีประสิทธิภาพเกี่ยวกับความไวในการกำหนดเส้นทาง ในการตอบข้อสอบสูงกว่าแบบขนาดชั้นคงที่ (Wiess. 1974: 18; citing Paterson. 1962) จำนวนชั้น หรือ จำนวนข้อที่ผู้สอบจะต้องทำ ถ้าผู้สอบยิ่งทำข้อสอบมากข้อ ความถูกต้องแม่นยำ (Accuracy) ค่าความเชื่อมั่น (Reliability) และความเที่ยงตรง (Validity) ในการประมาณค่าความสามารถของผู้สอบ ก็จะมีมากกว่าเมื่อใช้จำนวนข้อสอบน้อยกว่า (Gulliksen. 1967:74-105; Lord and Novick. 1968: 112-118) สำหรับ วิธีการให้คะแนน นั้น วิธีที่มีประสิทธิภาพสูงสุดควรจะเป็น การให้คะแนนโดยใช้วิธีของ เบย์ (Bayesian Updating) รองลงมาคือ วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (Conditional Maximum Likelihood) ให้ตามค่าเฉลี่ยความยากของข้อที่ทำถูก และให้ตามค่าความยากของข้อที่ยากที่สุดที่ทำถูก ฉะนั้น การวิจัยครั้งนี้จึง ได้ตั้งสมมติฐานว่า

การทดสอบเทเลอร์รูปปริมาตรแบบขนาดชั้นแปรผัน ที่มี 16 ชั้น ให้คะแนนโดยใช้วิธีของ เบย์ จะมีประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบสูงสุด และการทดสอบเทเลอร์รูปปริมาตรแบบขนาดชั้นคงที่ ที่มี 7 ชั้น ให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำถูก จะมีประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบต่ำสุด

ประโยชน์ที่คาดว่าจะได้รับ

1. เป็นการขยายความรู้ในเชิงทฤษฎีเกี่ยวกับการทดสอบเทเลอรัปรีรามิด
2. เพื่อให้นักพัฒนการศึกษา ครู และอาจารย์ได้ข้อมูลในการตัดสินใจเลือกใช้ รูปแบบ จำนวนชั้น และวิธีการให้คะแนน ในการทดสอบเทเลอรัปรีรามิดได้ถูกต้อง

ข้อตกลงเบื้องต้น

เนื่องจากผู้สอบที่ทำข้อสอบถูกต้อง หรือผิดหมด การประมาณค่าความสามารถของผู้สอบ โดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (Conditional Maximum Likelihood) ไม่สามารถที่จะประมาณค่าได้ จึงต้องตัดผู้สอบเหล่านี้ออกไป รวมทั้งในการตีพิมพ์ก็มีค่าความเป็นไปได้สูงสุดมีหลายค่าก็ต้องตัดผู้สอบเหล่านี้ออกไปด้วย เช่นกัน

ขอบเขตในการวิจัย

ในการวิจัยครั้งนี้ได้กำหนดขอบเขตในการศึกษาไว้ ดังนี้

1. การหาค่าโอกาสในการตอบข้อสอบถูกต้องของกลุ่มตัวอย่างจะใช้ 1-พารามิเตอร์ แบบ โลจิสติก โมเดล (One-Parameter Logistic Model)
2. รูปแบบ (Form) ของการทดสอบเทเลอรัปรีรามิด ที่จะศึกษาในครั้งนี้มี 2 แบบ คือ แบบขนาดขั้นที่ (Constant Step Size Pyramid) และแบบขนาดขั้นแปรผัน (Variable Step Size Pyramid)
3. จำนวนชั้น (Stage) ของการทดสอบเทเลอรัปรีรามิด ที่จะศึกษาในครั้งนี้มี 4 แบบ คือ แบบที่มีจำนวน 7, 10, 13 และ 16 ชั้น
4. วิธีการให้คะแนน (Scoring) ในการทดสอบเทเลอรัปรีรามิด ที่จะศึกษาในครั้งนี้มี 4 วิธี คือ ให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำถูก ให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทำถูก ให้คะแนนโดยใช้วิธีของเบย์ (Bayesian Updating) และให้คะแนนโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (Conditional Maximum Likelihood)
5. ค่าความยากของข้อสอบ (p_j) ในการทดสอบเทเลอรัปรีรามิด จะกำหนดให้มีความยากอยู่ในช่วง -3.000 ถึง $+3.000$

6. ขอบเขตกลุ่มตัวอย่างในการทดลอง 1000 คน
7. จำนวนครั้งที่ทำการทดลองซ้ำ 40 ครั้ง

6. ขนาดกลุ่มตัวอย่างในแต่ละกลุ่มใช้ 1,000 คน
7. จำนวนครั้งในการทดลองซ้ำ จะทดลองซ้ำ 40 ครั้ง

นิยามศัพท์เฉพาะ

1. การทดสอบเทเลอร์ (Tailored Testing) หมายถึง การทดสอบโดยใช้ชุดของข้อสอบ ที่มีการปรับความยากให้เหมาะสมกับระดับความสามารถของผู้สอบเป็นรายบุคคล
2. ประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบ หมายถึง ความถูกต้องแม่นยำในการประมาณค่า ซึ่งได้จาก ค่าความคลาดเคลื่อนในการประมาณค่า (E) ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า (SEE) และค่าความเที่ยงตรงในการประมาณค่า (Validity: re $\hat{e}$)
3. รูปแบบ (Form) หมายถึง การจัดโครงสร้างภายในของการทดสอบเทเลอร์รูปปิรามิด ในที่นี้หมายถึงใน 2 รูปแบบ คือแบบขนาดขั้นคงที่ (Constant Step Size Pyramid) และแบบขนาดขั้นแปรผัน (Variable Step Size Pyramid)
4. ขนาดขั้นคงที่ (Constant Step Size Pyramid) หมายถึง การทดสอบเทเลอร์รูปปิรามิด ที่มีช่วงห่างค่าความยากของข้อที่อยู่ติดกันภายในชั้น มีค่าเท่ากันตลอด
5. ขนาดขั้นแปรผัน (Variable Step Size Pyramid) หมายถึง การทดสอบเทเลอร์รูปปิรามิด ที่มีช่วงห่างค่าความยากของข้อที่อยู่ติดกันภายในชั้น มีค่าไม่เท่ากัน คือ ในช่วงกลาง ๆ จะมีค่ามากแล้วค่อย ๆ ลดลงในข้อที่ค่อนข้างง่ายหรือยาก สำหรับการเพิ่มหรือลดค่าความยากนั้น จะเพิ่มหรือลดไปยังตำแหน่งกึ่งกลางระหว่าง 2 ข้อในชั้นที่มาก่อน หรือ ระหว่างข้อในชั้นที่มาก่อนกับค่าความยากสูงสุด หรือต่ำสุด
6. วิธีการให้คะแนน (Scoring) หมายถึง วิธีการในการประมาณค่าความสามารถของผู้สอบจากการทดสอบเทเลอร์ ซึ่งในการวิจัยครั้งนี้ จะใช้ทั้งหมด 4 วิธี คือ ให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำถูก ให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทำถูก ให้คะแนนโดยใช้วิธีของเบย์ (Bayesian Updating) และ ให้คะแนนโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (Conditional Maximum Likelihood)

บทที่ 2

เอกสารและผลงานวิจัยที่เกี่ยวข้อง

การศึกษาเปรียบเทียบประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบ จากการทดสอบเทเลอร์รูปปารามิคที่มี รูปแบบ จำนวนชั้น และวิธีการให้คะแนน ที่แตกต่างกัน ในการศึกษาครั้งนี้ได้นำเอาทฤษฎีการตอบข้อสอบ (Item Response Theory: IRT) และเทคนิคมอนติคาร์โล (Monte Carlo Method) มาประยุกต์ใช้ในการทดสอบเทเลอร์ ในการประมาณค่าความสามารถของผู้สอบ การนำเสนอเอกสารและผลงานวิจัยที่เกี่ยวข้องครั้งนี้ จึงได้แบ่งเนื้อหาในการนำเสนอออกเป็น 3 หัวข้อ ดังนี้

- (1) ทฤษฎีการตอบข้อสอบ (Item Response Theory: IRT)
- (2) การทดสอบเทเลอร์ (Tailored Testing)
- (3) การศึกษาโดยใช้เทคนิคมอนติคาร์โล (Monte Carlo Method)

1. ทฤษฎีการตอบข้อสอบ (Item Response Theory: IRT)

1.1 หลักการของทฤษฎีการตอบข้อสอบ

ทฤษฎีการทดสอบดั้งเดิม (Classical Test Theory) ที่เราใช้กันมานานนั้นยังมีข้อบกพร่อง ในเรื่องของค่าพารามิเตอร์ (Parameter) ของข้อสอบ เช่น ค่าความยาก และค่าอำนาจจำแนกของข้อสอบไม่คงที่ แต่จะแปรเปลี่ยนไปตามกลุ่มตัวอย่าง (Hambleton and Swaminathan, 1985: 1-2; สำเร็จ บุญเรืองรัตน์, 2527: 97) และ คะแนนของผู้สอบก็แปรเปลี่ยนไปตามค่าความยากของข้อสอบ จากข้อบกพร่องอันนี้ทำให้เกิดปัญหาเกี่ยวกับการเปรียบเทียบความสามารถของผู้สอบ ซึ่งความสามารถของผู้สอบจะเปรียบเทียบกันได้นั้น ต้องใช้แบบทดสอบที่มีข้อคำถามเหมือนกัน มีค่าความยากเท่ากัน หรือเป็นคู่ขนานกันเท่านั้น ถ้าใช้ข้อสอบที่มีความยาก หรือง่ายแตกต่างกันแล้ว ไม่สามารถที่จะเปรียบเทียบกันได้ (สำเร็จ บุญเรืองรัตน์, 2525: 1-2) ฉะนั้นทฤษฎีการทดสอบดั้งเดิม (Classical Test Theory) จึงไม่เหมาะสมที่จะนำมาใช้ในการทดสอบเทเลอร์ (Tailored Testing) ที่มีการจัดข้อสอบให้เหมาะสมกับ

ระดับความสามารถของผู้สอบ โดยที่ผู้สอบที่ต่างความสามารถกันจะได้รับชุดของข้อสอบที่แตกต่างกัน และมีค่าความยาก หรือง่ายไม่เท่ากัน ทฤษฎีที่เหมาะสมและสามารถนำมาใช้กับการทดสอบ เทเลอร์ (Tailored Testing) ได้อย่างมีประสิทธิภาพ ก็คือ ทฤษฎีการตอบข้อสอบ (IRT) (Urry. 1977: 181; Green and others. 1984. 348; Hambleton. 1989: 123)

ทฤษฎีการตอบข้อสอบ (Item Response Theory: IRT) หรือทฤษฎีคุณลักษณะแฝง (Latent Trait Theory) มีความเชื่อว่าค่าพารามิเตอร์ (Parameter) ต่าง ๆ ของข้อสอบ ไม่ว่าจะเป็นค่าความยาก (b) อำนาจจำแนก (a) หรือค่าการเดา (c) ของข้อสอบแต่ละข้อ เป็นคุณลักษณะที่มีอยู่ประจำ และคงที่พอสมควรในตัวข้อสอบนั้นจริง ฉะนั้นค่าพารามิเตอร์ (Parameter) เหล่านี้จึงไม่ควรแปรเปลี่ยนไปตามกลุ่มตัวอย่าง (Sample-free) และในทางองเดียว กับความสามารถของผู้สอบ (Ability) ก็เป็นคุณลักษณะที่มีอยู่ในตัวผู้สอบนั้นจริง จึงไม่ควรจะเปลี่ยนแปลงไปตามค่าความยากของข้อสอบ (Test-free) ซึ่งเป็นคุณลักษณะภายนอก แต่เนื่องจากว่า ความสามารถของผู้สอบ (Ability) เป็นคุณลักษณะแฝง (Latent Trait) ซึ่งเราไม่สามารถที่จะวัดหรือสังเกตได้โดยตรง (Unobservable) แต่จะเป็นตัวพยากรณ์ (Predict) หรือ อธิบาย (Explain) ผลการสอบ (Test Performance) หรือ คะแนน (Score) ซึ่งเป็นสิ่งที่เราสามารถสังเกตหรือวัดได้ (Observable) (Lord and Novick. 1968: 358; Hambleton and Cook. 1977: 75; Hambleton and Swaminathan. 1985: 9; สานิจ บุษุเรืองรัตน์. 2527: 98) ฉะนั้นนักวัดผลการศึกษาจึงได้พยายามหาความสัมพันธ์ระหว่างปริมาณ การกระทำข้อสอบ หรือคะแนน (Test Performance or Score) กับ ปริมาณความสามารถ (Ability) เพื่อเขียนออกมาเป็นโมเดลทางคณิตศาสตร์ (Mathematical Model) (Hambleton and Cook. 1977: 75; Hambleton and Swaminathan. 1985: 9; Wright and Stone. 1979: 13)

ความสัมพันธ์ระหว่างปริมาณของการกระทำข้อสอบหรือคะแนนของผู้สอบ (Test Performance) กับปริมาณความสามารถของผู้สอบ (Ability) สามารถเขียนในรูปของความสัมพันธ์ทั่ว ๆ ไปได้ ดังนี้ (Lord. 1982: 141)

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

เมื่อ P แทน ผลการสอบ (Performance)

f แทน ฟังก์ชัน (function)

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

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

β_j แทน ค่าพารามิเตอร์ของข้อสอบข้อที่ j

จากความสัมพันธ์ดังกล่าวนี้เป็นฟังก์ชันความสัมพันธ์ในลักษณะทั่ว ๆ ไป นักวัดผลการศึกษาจะต้องหา โมเดลทางคณิตศาสตร์ (Mathematical Model) ที่เหมาะสม เพื่อใช้แทนฟังก์ชัน (Function) ความสัมพันธ์ดังกล่าวโดยอาศัยเงื่อนไขเบื้องต้นต่าง ๆ ดังนี้

1.2 เงื่อนไขเบื้องต้นเกี่ยวกับการใช้ทฤษฎีการตอบข้อสอบ (IRT)

การที่จะใช้ทฤษฎีการตอบข้อสอบ (IRT) ได้ จะต้องเป็นไปตามเงื่อนไขเบื้องต้น ดังนี้

(1) ความเป็นมิติเดียว (Unidimension) สมมุติว่าความสามารถของคน (Ability or Trait) มีอยู่ทั้งหมด k อย่าง ซึ่งความสามารถแต่ละอย่างนี้ ต่างก็ส่งผลต่อการตอบข้อสอบ (Test Performance) ต่าง ๆ ที่รวมกันเป็นแบบทดสอบ ถ้าผลการตอบข้อสอบ หรือคะแนนของผู้สอบนั้น สามารถอธิบายได้เพียงความสามารถเดียวก็เพียงพอแล้ว เรียกเงื่อนไขข้อนี้ว่ามีความเป็นมิติเดียว (Hambleton and Swaminathan. 1985: 16; สำเร็จ บุญเรืองรัตน์. 2527: 99) สำหรับในทางปฏิบัติ ก็คือ ข้อสอบแต่ละข้อในแบบทดสอบจะต้องวัดในคุณลักษณะ (Trait) เดียว แต่อย่างไรก็ตาม แฮมเบิลตัน และสวามินาธาน (Hambleton and Swaminathan. 1985: 16-17) ได้กล่าวว่าข้อตกลงเบื้องต้นข้อนี้ไม่ได้เข้มงวดนัก ขอให้มึลักษณะเด่น (Dominant) ที่จะวัดในองค์ประกอบใดองค์ประกอบหนึ่ง ก็ใช้ได้ ซึ่งก็สอดคล้องกับผลการวิจัยของ คราสโกว์และพาร์สันส์ และเรคเคส (McCauley and Mendoza. 1985: 390; citing Drasgow and Parsons. 1983; Reckase. 1979) ที่พบว่า การประมาณค่าพารามิเตอร์

ต่าง ๆ ยังสามารถกระทำได้ เมื่อการวัดนั้นอาจจะมีการบ่ายเบนไปในทางหลายมิติบ้างเล็กน้อย

วิธีการตรวจสอบว่าแบบทดสอบนั้นวัดในมิติเดียวหรือไม่นั้น ทำได้หลายวิธี เช่น โดยการวิเคราะห์องค์ประกอบ (Factor analysis) แล้วสังเกตค่าไอเก้น (Eigen value) ค่าสูงสุด ว่าแตกต่างจากค่าอื่น ๆ อย่างชัดเจนหรือไม่ (Hambleton and Thumb. 1973: 195-211; Reckase. 1979: 207-230) สำหรับ ลอร์ดและ โนวิค (Lord and Novick. 1968: 382; Lord. 1980: 20) ยังได้แนะนำว่าการวิเคราะห์องค์ประกอบควรจะใช้ค่าสหสัมพันธ์แบบเททระคอร์ริก (Tetrachoric correlation) ไม่ควรใช้การคำนวณค่าสหสัมพันธ์แบบเพียร์สันโปรดักท์โมเมนต์ (Pearson Product Moment) เพราะว่าข้อมูลที่เหมาะสม (Fit) กับโมเดลของทฤษฎีการตอบข้อสอบ (IRT) ไม่ได้สัมพันธ์กันโดยตรง (Hambleton and Swaminathan. 1985: 306) สำหรับการทดสอบเทเลอร์ กรีนและคนอื่น ๆ (Green and others. 1984:349) ได้เสนอแนะว่าการที่จะทำให้เกิดความเป็นมิติเดียว (Unidimension) ได้ง่ายวิธีหนึ่งคือ การเลือกข้อสอบที่มีค่าอำนาจจำแนก (a) สูง ควรมีค่าอำนาจจำแนกไม่ต่ำกว่า 0.8 (Urry. 1977: 184) แต่ก็ควรระวังถ้าคัดเลือกแต่เฉพาะข้อสอบที่มีค่าอำนาจจำแนกสูง ๆ อาจจะทำให้ขาดคุณสมบัติที่ดีของข้อสอบในส่วนอื่นก็ได้

(2) ความเป็นอิสระในการตอบข้อสอบ (Local Independence) หมายถึง โอกาสในการตอบข้อสอบแต่ละข้อได้ถูกต้องเป็นอิสระจากกัน นั่นคือ การตอบข้อสอบข้อใดข้อหนึ่งได้ถูกหรือผิดจะ ไม่มีผลต่อการตอบข้ออื่น ๆ ด้วย (Lord. 1980: 19; Hambleton and Swaminathan. 1985: 23) หรือ จะกล่าวในเชิงคณิตศาสตร์ได้ว่า ความเป็นอิสระในการตอบข้อสอบ (Local Independent) หมายถึง โอกาสที่จะตอบข้อสอบถูกต้องทั้งหมด จะเท่ากับผลคูณของโอกาสการตอบข้อสอบถูกเป็นรายข้อ (Lord. 1980: 19; Hambleton and Swaminathan. 1985: 23) แมกซ์โดนัลด์ (McDonald. 1982: 379-396) และ แฮมเบิลตันกับสวามินาธาน (Hambleton and Swaminathan. 1985: 22) มีความเห็นตรงกันว่า ถ้ามีความเป็นมิติเดียว (Unidimension) แล้ว ความเป็นอิสระในการตอบข้อสอบ (Local Independent) ก็จะมีเอง ความเป็นอิสระในการตอบข้อสอบ (Local Independence) มีความสำคัญต่อการประมาณค่าพารามิเตอร์ต่าง ๆ มาก ถ้าขาดความเป็นอิสระในการตอบข้อสอบแล้ว ค่าพารามิเตอร์ต่าง ๆ

ที่ประมาณค่าได้ จะมีความคลาดเคลื่อนจากความเป็นจริง

(3) โค้งแสดงลักษณะข้อสอบ (Item Characteristic Curve: ICC) โค้งแสดงลักษณะข้อสอบ (ICC) เป็นฟังก์ชันทางคณิตศาสตร์ ที่แสดงถึงความสัมพันธ์ระหว่างโอกาสในการตอบข้อสอบข้อนั้นได้ถูกต้อง กับระดับความสามารถที่วัดได้โดยใช้ชุดของข้อสอบ หรือแบบทดสอบนั้น (Hambleton and Swaminathan. 1985: 25) ซึ่งจะเห็นว่า โอกาสที่ผู้สอบตอบข้อสอบถูกต้อง (Probability) จะขึ้นอยู่กับโค้งแสดงลักษณะข้อสอบ (ICC) ในแต่ละ โมเดลที่ใช้ โดยที่รูปร่าง (Shape) ของโค้งแสดงลักษณะข้อสอบ (ICC) ในแต่ละข้อมีคุณสมบัติไม่แปรเปลี่ยน (Invariant) ไปตามกลุ่มตัวอย่างหรือตัวผู้สอบ (Hambleton and Cook. 1977: 80) ฉะนั้นจึงทำให้โอกาสในการตอบข้อสอบถูกต้องในแต่ละข้อไม่แปรเปลี่ยน (Invariant) ด้วย

เนื่องจากมีความเชื่อหลายความเชื่อ เกี่ยวกับความสัมพันธ์ระหว่างโอกาสในการตอบข้อสอบได้ถูกต้อง กับระดับความสามารถที่วัดได้โดยใช้ชุดของข้อสอบ หรือแบบทดสอบนั้น จึงทำให้เกิดมีโมเดลของโค้งแสดงลักษณะข้อสอบ (Item Response Model) ขึ้นมาหลายโมเดล ดังจะกล่าวต่อไปนี้

1.3 โมเดลของทฤษฎีการตอบข้อสอบ (Item Response Model)

จากการใช้เงื่อนไขเบื้องต้นดังกล่าว ทำให้เกิดมีโมเดลทางคณิตศาสตร์ (Mathematical Model) เพื่อแสดงความสัมพันธ์ระหว่างผลการสอบ (Test Performance) กับความสามารถของผู้สอบ (Ability or Trait) หลายโมเดลด้วยกัน แต่ที่สำคัญที่จะนำมากล่าวในครั้งนี้มี 3 โมเดล คือ

- (1) นอร์มัล โอจีฟ โมเดล (Normal Ogive Model)
- (2) โลจิสติก โมเดล (Logistic Model)
- (3) ราซค์ โมเดล (Rasch Model)

1.3.1 นอร์มัล โอจีฟ โมเดล (Normal Ogive Model)

โมเดลนี้ได้รับการพัฒนาขึ้นมาตั้งแต่ปี ค.ศ. 1943-1944 โดย ลอร์ดและ (Lord and

Novick. 1968: 369; citing Lowley. 1943, 1944) ต่อมาราว ค.ศ. 1952 ลอร์ด (Hambleton and Swaminathan. 1985: 4; citing Lord. 1952) ได้ศึกษาและพัฒนาวิธีการประมาณค่าพารามิเตอร์ต่าง ๆ และการนำมาประยุกต์ใช้ ซึ่งในครั้งแรก ลอร์ด ได้ใช้ 2-พารามิเตอร์ โดยมีโมเดลทางคณิตศาสตร์ ดังนี้ (Lord and Novick. 1968: 366)

$$P_j(\theta) = \int_{-\infty}^{a_j(\theta - b_j)} \phi(t) dt$$

เมื่อ $P_j(\theta)$ แทน โอกาสที่ผู้สอบซึ่งมีระดับความสามารถ (θ) ตอบข้อสอบข้อที่ j ได้ถูกต้อง

$\phi(t)$ แทน นอร์มัล ฟรีควเอนซี ฟังก์ชัน (Normal frequency function)

a_j แทน ค่าอำนาจจำแนกของข้อสอบข้อที่ j

b_j แทน ค่าความยากของข้อสอบข้อที่ j

θ แทน ระดับความสามารถของผู้สอบ

นอร์มัล โอจีฟ โมเดล (Normal Ogive Model) เป็นโมเดลที่ ลอร์ด ได้ใช้เวลาอันยาวนานในการพัฒนาโมเดลนี้ ในระยะแรก ๆ เป็นไปอย่างเชื่องช้ามากเนื่องจากความสลับซับซ้อนทางคณิตศาสตร์ที่เป็นพื้นฐานของทฤษฎี ยากแก่การปฏิบัติ หรือนำมาใช้จริง คนทั่วไปจึงไม่นิยมนำโมเดลนี้มาใช้ แม้แต่ ลอร์ด เอง ในระยะหลังนี้ก็ได้หันมาใช้ โลจิสติก โมเดล (Logistic Model) แทน เพราะเป็นโมเดลที่มีลักษณะคล้ายกันมากแต่มีวิธีการคำนวณค่าพารามิเตอร์ต่าง ๆ ได้ง่ายกว่า (Lord and Novick. 1968: 366) แต่อย่างไรก็ตาม โมเดลนี้ก็พบว่ามิประโยชน์อย่างมาก สำหรับการพัฒนาโลจิสติก โมเดล (Logistic Model) ต่อมา

1.3.2 โลจิสติก โมเดล (Logistic Model)

ผู้ที่มีบทบาทสำคัญในการพัฒนาโมเดลนี้ คือ เบิร์นบอม (Hambleton and Swaminathan. 1985: 36; citing Birnbaum. 1957, 1958a, 1958b, 1968) โดยในครั้งแรก

เบิร์นบอม ได้พัฒนาโลจิสติก โมเดล แบบ 2-พารามิเตอร์ คือ พารามิเตอร์ค่าความยาก (b) และค่าอำนาจจำแนก (a) ซึ่งเป็นโมเดลที่มีโค้งแสดงลักษณะข้อสอบ (Item Characteristic Curve: ICC) คล้ายกับโค้งแสดงลักษณะข้อสอบ (ICC) แบบนอร์มัล โอจีฟ โมเดล (Normal Ogive Model) แบบ 2-พารามิเตอร์ ที่ ลอร์ด ได้เสนอไว้เมื่อปี ค.ศ. 1952 มาก โดยโค้งแสดงลักษณะข้อสอบทั้งสองมีความแตกต่างกันไม่เกิน .01 สำหรับทุก ๆ ค่าของความสามารถ (Birnbau. 1968: 399; Lord. 1980: 14; Hambleton and Swaminathan. 1985: 37; citing Haley. 1952)

เนื่องจากโลจิสติก โมเดล (Logistic Model) ได้ใช้โมเดลทางคณิตศาสตร์ที่สามารถนำไปประยุกต์ใช้ได้ง่ายกว่าแบบนอร์มัล โอจีฟ โมเดล (Birnbau. 1968: 399) จึงทำให้มีคนนำโมเดลนี้ไปใช้กันมาก ต่อมา เบิร์นบอม ได้เพิ่มพารามิเตอร์ค่าการเดา (c) เข้าไปอีกตัวหนึ่ง จึงกลายเป็นแบบ 3-พารามิเตอร์ แยกจากนั้นยังได้ตัดพารามิเตอร์ค่าอำนาจจำแนก (a) ออกจากแบบ 2-พารามิเตอร์ จึงเหลือเพียงพารามิเตอร์เดียว คือ ค่าความยาก (b) ซึ่งเรียกว่า 1-พารามิเตอร์ สำหรับโมเดลทางคณิตศาสตร์ (Mathematical Model) ของแต่ละโมเดล ย่อย เป็นดังนี้

(1) โลจิสติก โมเดล แบบ 1-พารามิเตอร์ (One-Parameter Logistic Model) เป็นโมเดลที่ถือว่าข้อสอบทุกข้อมีค่าอำนาจจำแนกเท่ากันหมดซึ่งเท่ากับ $\bar{a}$ และมีค่าการเดา (c_j) เป็นศูนย์ โดยมีโมเดลทางคณิตศาสตร์ ดังนี้ (Hambleton and Swaminathan. 1985: 47)

$$P_j(\theta) = \frac{e^{D\bar{a}(\theta - b_j)}}{1 + e^{D\bar{a}(\theta - b_j)}}$$

(2) โลจิสติก โมเดล แบบ 2-พารามิเตอร์ (Two-Parameter Logistic Model) มีโมเดลทางคณิตศาสตร์ ดังนี้ (Hambleton and Swaminathan. 1985: 36)

$$P_j(\theta) = \frac{e^{Da_j(\theta - b_j)}}{1 + e^{Da_j(\theta - b_j)}}$$

(3) โลจิสติก โมเดล แบบ 3-พารามิเตอร์ (Three-Parameter Logistic Model)
มีโมเดลทางคณิตศาสตร์ ดังนี้ (Hambleton and Swaminathan, 1985: 37)

$$P_j(\theta) = c_j + (1 - c_j) \cdot \frac{e^{Da_j(\theta - b_j)}}{1 + e^{Da_j(\theta - b_j)}}$$

เมื่อ $P_j(\theta)$ คือ โอกาสที่ผู้สอบซึ่งมีระดับความสามารถ (θ) ตอบข้อสอบข้อที่ j ได้ถูกต้อง

a_j คือ ค่าอำนาจจำแนกของข้อสอบข้อที่ j โดยทฤษฎีแล้ว a จะมีค่าอยู่ระหว่าง $-\infty$ ถึง $+\infty$ ข้อสอบที่ใช้ได้ควรมีค่า a เป็นบวก โดยปกติจะมีค่าไม่เกิน 2.00

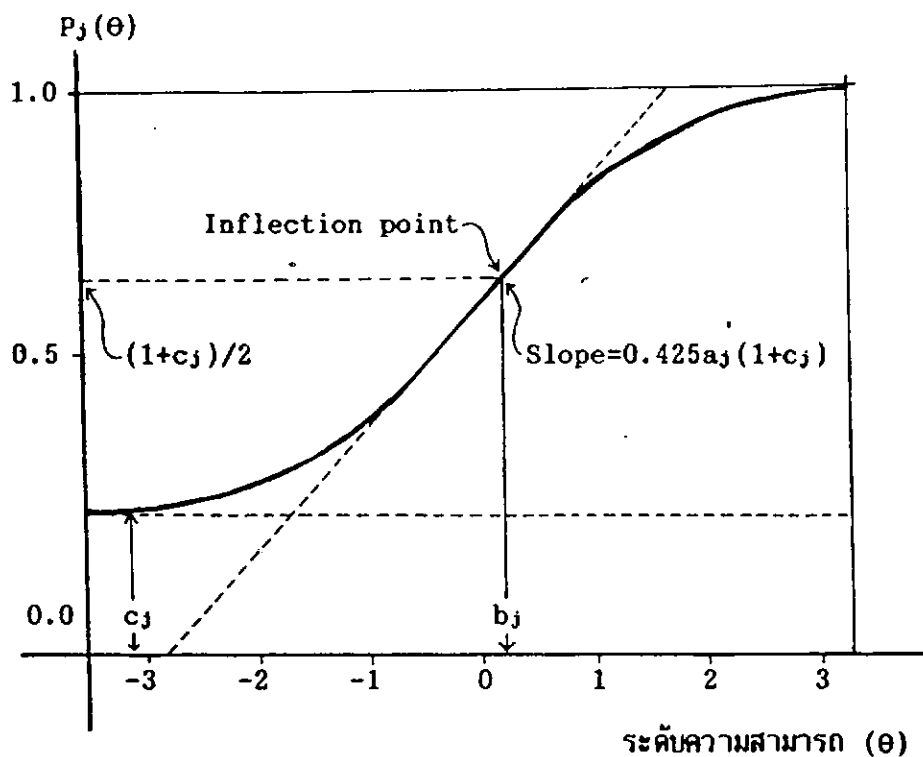
b_j คือ ค่าความยากของข้อสอบข้อที่ j โดยทฤษฎีแล้ว b จะมีค่าอยู่ระหว่าง $-\infty$ ถึง $+\infty$ แต่ในทางปฏิบัติ b จะมีค่าอยู่ระหว่าง -2.00 ถึง +2.00 ข้อที่มีค่า b ใกล้ -2.00 เป็นข้อที่ง่ายมาก และข้อที่มีค่า b ใกล้ +2.00 เป็นข้อที่ยากมาก (Hambleton and Cook, 1977: 81)

c_j คือ ค่าการเดาของข้อสอบข้อที่ j หมายถึง ค่าความน่าจะเป็นที่ผู้สอบที่มีความสามารถต่ำสุด ที่ข้อสอบนี้ได้ถูกต้อง ข้อสอบที่มีค่าการเดาสูง เป็นข้อสอบที่ไม่ดี ควรมีค่าการเดาไม่เกิน 0.30 (Urry, 1977: 184)

- o คือ ระดับความสามารถของผู้สอบ โดยทฤษฎีแล้วจะมีค่าอยู่ระหว่าง $-\infty$ ถึง $+\infty$ ในทางปฏิบัติจะมีค่าอยู่ระหว่าง -3.00 ถึง $+3.00$ (Hambleton and Cook. 1977: 81)
- D คือ ค่าคงที่ (Scaling Factor) เพื่อปรับโค้งแสดงลักษณะข้อสอบ (ICC) ให้มีลักษณะใกล้เคียงกับโค้งแสดงลักษณะข้อสอบของนอร์มัล โอจีฟ โมเดล (Normal Ogive Model) มากที่สุด ซึ่ง D จะมีค่าเท่ากับ 1.7
- e คือ ค่าเอ็กซ์โปเนนเชียล (Exponential) เป็นค่าคงที่ตัวหนึ่งซึ่งมีค่าเท่ากับ 2.7182818

จากโมเดลทางคณิตศาสตร์ของ 3-พารามิเตอร์นี้ เมื่อนำมาเขียนเป็นกราฟโค้งแสดงลักษณะข้อสอบ (Item Characteristic Curve: ICC) ได้ ดังนี้ (Lord. 1980: 12-14; Hambleton and Swaminathan. 1985: 38-39)

โอกาสที่จะตอบข้อสอบได้ถูกต้อง



ภาพประกอบ 1 ได้แสดงลักษณะข้อสอบ (ICC) แบบ 3-พารามิเตอร์ (Three-Parameters Logistic Model)

จากภาพประกอบ 1 เป็นได้แสดงลักษณะข้อสอบ (ICC) แบบ 3-พารามิเตอร์ จะเห็นว่า

(1) ค่าความชัน (slope) a_j จะเป็นสัดส่วนกับความชัน (Slope) ของโค้ง ณ จุดเปลี่ยนโค้ง หรือจุดที่มีความชันมากที่สุด [ความชัน ณ จุดนี้มีค่าเท่ากับ $0.425a_j(1+c_j)$]

(2) ค่าความยาก b_j จะอยู่บนแกนความสามารถ θ ตรงจุดตัดกับเส้นที่ลากมาจากจุดเปลี่ยนโค้งมาตั้งฉากกับแกนความสามารถ

(3) ค่าการเดา c_j เป็นส่วนตัดบนแกนตั้งของโค้งแสดงลักษณะข้อสอบ (ICC)

1.3.3 ราชค์ โมเดล (Rasch Model)

ราชค์โมเดล (Rasch Model) ก็เป็นอีกโมเดลหนึ่งในตระกูลของทฤษฎีคุณลักษณะแฝง (Latent Trait) ที่ได้พัฒนาขึ้นมาโดยนักคณิตศาสตร์ชาวเดนมาร์ก (Danish) ชื่อจอร์จ ราชค์ (Georg Rasch) ในราวปี ค.ศ.1951 ถึง 1960 ราชค์ ได้ให้ชื่อโมเดลนี้ว่า พรอบเบบิลิสติก โมเดล (Probabilistic Model) ในครั้งแรก ราชค์ ได้อธิบายโมเดลนี้ด้วยทฤษฎีความน่าจะเป็นที่มีการแจกแจงแบบปัวซอง (Poisson) (Rasch. 1980: 185) ต่อมา ไรท์กับสโตน (Wright and Stone. 1979) ได้นำเอา พรอบเบบิลิสติก โมเดล มาอธิบายใหม่ในรูปของ โลจิสติก (Logistic) โดยให้ชื่อว่าราชค์ โมเดล ทำให้ง่ายต่อการเข้าใจและการประมาณค่าพารามิเตอร์ต่าง ๆ โมเดลนี้มีความเชื่อว่า ผู้สอบมีความสามารถจริง (Ability: β_i) แฝงอยู่ และข้อสอบก็มีความยาก (Difficulty: d_j) แฝงอยู่ประจำในแต่ละข้อ พฤติกรรมการตอบข้อสอบจึงขึ้นอยู่กับผลต่างระหว่างความสามารถของผู้สอบ (β_i) กับ ความยากของข้อสอบ (d_j) ถ้าความสามารถของผู้สอบมีมากกว่าความยากของข้อสอบ ($\beta_i - d_j > 0$) โอกาสที่ผู้สอบจะตอบข้อสอบข้อนั้นได้ถูกต้องก็มีมากกว่า 0.5 แต่ถ้าความสามารถของผู้สอบมีน้อยกว่าความยากของข้อสอบ ($\beta_i - d_j < 0$) โอกาสที่ผู้สอบจะตอบข้อสอบข้อนั้นได้ถูกต้องก็มีน้อยกว่า 0.5 หรือถ้าผู้สอบมีความสามารถพอ ๆ กับ ความยากของข้อสอบ ($\beta_i - d_j = 0$) โอกาสที่ผู้สอบจะตอบข้อสอบข้อนั้นได้ถูกต้องก็มีครึ่งต่อครึ่ง (Wright and Stone. 1979: 9-12) สำหรับ ราชค์ โมเดลนี้ มีโมเดลทางคณิตศาสตร์ (Wright and Stone. 1979: 15) ดังนี้

$$P_j(\beta_i) = \frac{e^{(\beta_i - d_j)}}{1 + e^{(\beta_i - d_j)}}$$

เมื่อ $P_j(\beta_i)$ คือ โอกาสที่ผู้สอบซึ่งมีระดับความสามารถ β_i ตอบข้อสอบข้อที่ j ได้ถูกต้อง

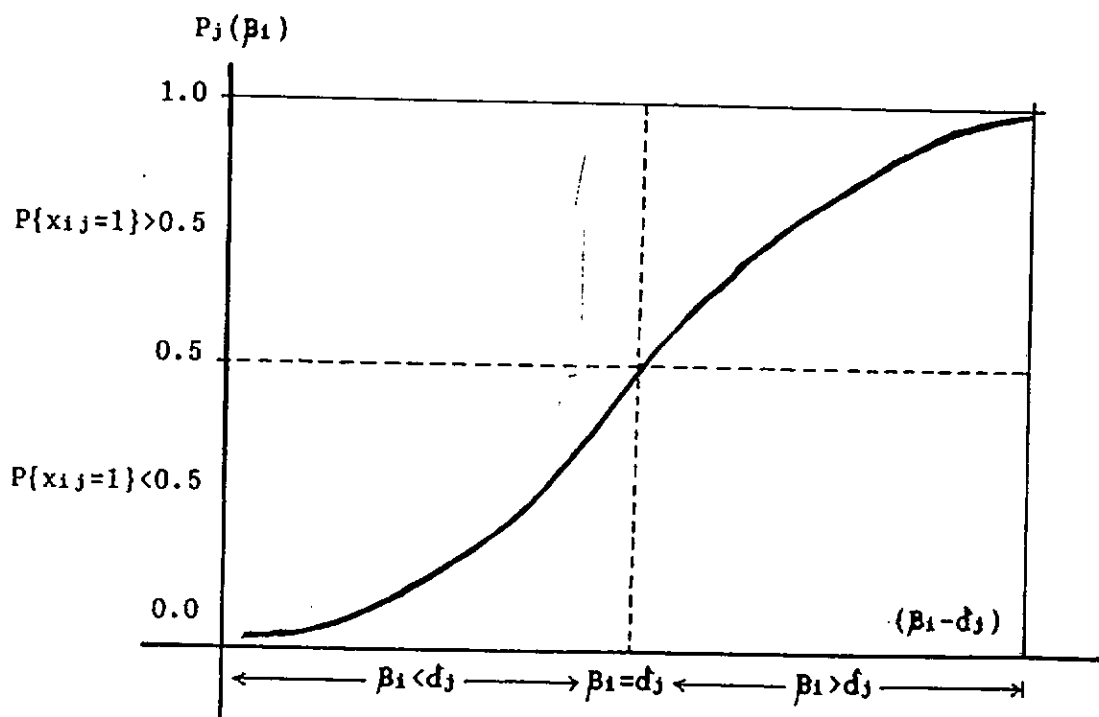
β_i คือ ค่าระดับความสามารถของผู้สอบคนที่ i

d_j คือ ค่าความยากของข้อสอบข้อที่ j

จะเห็นว่าราชค์ โมเดล มีลักษณะคล้าย กับ 1-พารามิเตอร์ แบบโลจิสติก โมเดลมาก
ทั้ง ๆ ที่โมเดลทั้งสองนี้ต่างคนต่างพัฒนามา ใน 1-พารามิเตอร์ ค่า a และ D เป็นค่าคงที่
ฉะนั้น ไรท์ (Hambleton and Swaminathan. 1985: 47; citing Wright. 1977a)
จึงให้ $\beta_i = Da\alpha_i$ และ $d_j = Dab_j$ ซึ่งก็จะเป็นโมเดลที่เหมือนกับ แบบ 1-พารามิเตอร์

จากโมเดลทางคณิตศาสตร์ของราชค์โมเดล เมื่อนำมาเขียนเป็นกราฟโค้งแสดงลักษณะ
ข้อสอบ (ICC) จะได้ ดังนี้ (Wright and Stone. 1979: 14)

โอกาสที่จะตอบข้อสอบได้ถูกต้อง



ภาพประกอบ 2 โค้งแสดงลักษณะข้อสอบ (ICC) แบบราชค์ โมเดล

จากภาพประกอบ 2 จะเห็นว่าค่าความยากของข้อสอบ (d_j) จะอยู่บนแกนระดับความสามารถ (β_i) ตรงจุดตัดของเส้นที่ลากมาจากจุดเปลี่ยนโค้ง หรือมีค่าโอกาสในการตอบข้อสอบได้ถูกต้องเท่ากับ 0.5 มาตั้งฉากกับแกนระดับความสามารถ

ราซค์ โมเดล (Rasch Model) เป็นโมเดลที่มีคนนิยมใช้กันมากเพราะเป็นโมเดล ที่ใช้งานง่ายไม่ซับซ้อน ถ้าข้อมูลไม่มากเกินไปก็สามารถคำนวณด้วยเครื่องคิดเลขธรรมดาได้ (Wright and Stone, 1979: 28-45) และต่อมาได้มีการพัฒนาโมเดลนี้เพื่อให้ใช้ได้กับลักษณะของข้อมูลแบบต่าง ๆ (Wright and Masters, 1982: 38-59; citing Andrich, 1978c, 1978d, 1979; Rasch, 1972; Master, 1980; Masters and Wright, 1981)

การวิจัยครั้งนี้การหาค่าโอกาสในการตอบข้อสอบทุก $[P_j(\theta)]$ ในการทดลองโดยใช้วิธีมอนติ คาร์โล (Monte Carlo Method) นั้น จะเลือกใช้ 1-พารามิเตอร์ แบบโลจิสติก โมเดล (One-Parameter Logistic Model) ทั้งนี้เพราะพารามิเตอร์ค่าอำนาจจำแนก (a) และค่าการเดา (c) เป็นตัวแปรที่ผู้วิจัยไม่ประสงค์จะศึกษา จึงควบคุมโดยให้มีค่าคงที่ นั่นคือ ให้มีค่าอำนาจจำแนกทุกข้อเป็นหนึ่ง และค่าการเดาทุกข้อเป็นศูนย์ และมีผลงานวิจัยหลายชิ้นด้วยกันที่สนับสนุนว่า การวิเคราะห์ค่าพารามิเตอร์ต่าง ๆ ของข้อสอบและของผู้สอบโดยใช้ราซค์ โมเดล หรือ แบบ 1-พารามิเตอร์ มีคุณภาพดีเท่าเทียมกับการวิเคราะห์ แบบ 3-พารามิเตอร์ (Henly and others, 1989: 367; citing McBride, Cope and Wing, 1987)

1.4 การนำเอาทฤษฎีการตอบข้อสอบมาประยุกต์ใช้

การนำเอาทฤษฎีการตอบข้อสอบ (IRT) มาประยุกต์ใช้จริงในระยะแรก ๆ ยังมีปัญหาหลายประการ (Hambleton and Cook, 1977: 75-76) คือ

- (1) ตัวทฤษฎีอธิบายด้วยโมเดลทางคณิตศาสตร์ที่ซับซ้อนยากแก่การที่จะเข้าใจ
- (2) ผู้เสนอทฤษฎีส่วนใหญ่เน้นในทางทฤษฎีมากกว่าที่จะนำมาประยุกต์ใช้
- (3) ยังขาดโปรแกรมคอมพิวเตอร์ที่มีประสิทธิภาพ ที่ใช้สะดวกและรวดเร็วในการ

ประมาณค่าความสามารถของผู้สอบและค่าพารามิเตอร์ของข้อสอบ

(4) ยังมีนักวิจัยหลายท่าน ตั้งข้อสงสัยถึงประโยชน์ที่จะได้รับการวิจัย และพัฒนาในทฤษฎีนี้

(5) เป็นทฤษฎีที่มีข้อตกลงเบื้องต้นที่ค่อนข้างเข้มงวด (Restrictive) ทำให้มีข้อจำกัดในการนำมาใช้

การนำเอาทฤษฎีการตอบข้อสอบ, (IRT) มาใช้กันอย่างจริงจังภายหลังจากที่มีการพัฒนาโปรแกรมคอมพิวเตอร์มาใช้ในการประมาณค่าความสามารถของผู้สอบ และค่าพารามิเตอร์ต่าง ๆ ของข้อสอบ ไรท์และเมด (Wright and Mead, 1969) ได้พัฒนาโปรแกรมคอมพิวเตอร์ไบคาล (BICAL) สำหรับประมาณค่าพารามิเตอร์ แบบบราซค์ โมเดล หรือ 1-พารามิเตอร์ แบบโลจิสติก โมเดล ต่อมาในปี ค.ศ. 1976 ลอร์ดและคนอื่น ๆ (Lord and others, 1976) ได้พัฒนาโปรแกรมโลจิส (LOGIST) สำหรับการประมาณค่า 2-พารามิเตอร์ และ 3-พารามิเตอร์ แบบโลจิสติก โมเดล และ ต่อมาในปี ค.ศ. 1982 ลอร์ดกับคณะ แห่งสถาบันบริการการทดสอบทางการศึกษา (Educational Testing Service: ETS) ได้ปรับปรุงโปรแกรมโลจิส (LOGIST) ให้ใช้ได้ง่ายขึ้น สามารถทำงานได้เร็วขึ้น และเปิดโอกาสให้เลือกแสดงผลได้มากขึ้น (Hambleton and Swaminathan, 1985: 5) นอกจากนี้ยังมีโปรแกรมดาต้าเจ็น (DATAGEN) ของ แฮมเบิลตันและคูก (Hambleton and Cook, 1977: 89-90) โปรแกรม พีเอ็มแอล (PML) ของ กัสตฟัสสัน และ บล็อก (BILOG) ของ มิสเลวิกกับบล็อค (Hambleton and Swaminathan, 1985: 148; citing Gustafsson, 1980a; Mislevy and Block, 1982) จากโปรแกรมคอมพิวเตอร์เหล่านี้ ประกอบกับในระยะหลังนี้ ได้มีเครื่องคอมพิวเตอร์ที่มีขนาดเล็ก แต่มีประสิทธิภาพสูงใช้กันอย่างแพร่หลาย และที่สำคัญที่สุด คือ คุณสมบัติที่สำคัญของทฤษฎีการตอบข้อสอบ (IRT) ในเรื่องค่าพารามิเตอร์ของข้อสอบไม่ขึ้นอยู่กับความสามารถของผู้สอบ (Sample free) และความสามารถของผู้สอบก็ไม่ขึ้นอยู่กับความยาก หรือง่ายของข้อสอบ (Test-free) ดังได้ กล่าวแล้ว จึงสามารถนำเอาคุณสมบัติเด่นเหล่านี้มาประยุกต์ใช้ในทางการวัดผลการศึกษาได้อย่างมีประสิทธิภาพในหลาย ๆ เรื่อง ดังนี้

(1) การสร้างคลังข้อสอบ (Item Bank) โดยการนำเอาทฤษฎีการตอบข้อสอบ

(IRT) มาใช้ในการวิเคราะห์และสร้างคลังข้อสอบ ทำให้เกิดผลดีที่เห็นเด่นชัดมีสองประการ คือ ประการแรก ด้านค่าพารามิเตอร์ (Parameter) ของข้อสอบ ซึ่งได้แก่ค่าความยาก (b) ค่าอำนาจจำแนก (a) และค่าการคาด (c) ของข้อสอบ ที่วิเคราะห์โดยใช้ทฤษฎีการตอบข้อสอบ (IRT) มีคุณสมบัติที่ไม่แปรเปลี่ยนตามกลุ่มตัวอย่าง (Invariant) ประการที่สอง ด้านการรายงานคุณภาพของข้อสอบ ในทฤษฎีการตอบข้อสอบ (IRT) จะรายงานคุณภาพของข้อสอบในรูปของอินฟอร์เมชัน (Information) ซึ่งสามารถรายงานได้ทั้งเป็นรายข้อ (Item Information) และ ทั้งฉบับ (Test Information) ค่าอินฟอร์เมชันนี้ จะเป็นตัวบ่งชี้ถึงความถูกต้องแม่นยำ (Accuracy) ในการประมาณค่า ซึ่งสามารถใช้แทนค่าความเชื่อมั่น (Reliability) และค่าความคลาดเคลื่อนมาตรฐานในการวัด (Standard Error of Measurement) ในทฤษฎีดั้งเดิมได้ (Hambleton and Swaminathan. 1985: 236) ซึ่งค่าความเชื่อมั่นยังมีจุดอ่อน คือ นอกจากจะขึ้นอยู่กับคุณภาพของข้อสอบแล้ว ยังขึ้นอยู่กับความแตกต่างของกลุ่มตัวอย่าง (Heterogeneous) ด้วย แซมิจิม่า (Samejima. 1977: 243) ได้กล่าวว่า การหาค่าความเชื่อมั่นของข้อสอบเป็นความคิดรวบยอดที่ผิด (Dead Concept) และค่าความคลาดเคลื่อนมาตรฐานในการวัด (Standard Error of Measurement) ก็ตั้งอยู่บนสมมุติฐานที่ผิด (False Assumption) (Green and others. 1984: 352) ที่กำหนดให้ความแปรปรวนของความคลาดเคลื่อนมีค่าเท่ากันสำหรับทุก ๆ คะแนน

จากการใช้ทฤษฎีการตอบข้อสอบ (IRT) ทำให้ข้อสอบแต่ละข้อมีค่าสถิติต่าง ๆ เป็นของประจำข้อไม่ขึ้นอยู่กับข้ออื่นและกลุ่มตัวอย่าง จึงทำให้เกิดความสะดวกในการคัดเลือกข้อสอบมาใช้ และจัดชุดข้อสอบให้ได้คุณภาพตามเกณฑ์ที่ต้องการ (Hambleton and Swaminathan. 1985: 261)

(2) การหาความลำเอียงของข้อสอบ (Item Bias) ข้อสอบที่ใช้กันอยู่ โดยปกติจะมีบางข้อที่เหมาะสมเฉพาะคนบางกลุ่ม แต่จะไม่เหมาะสมกับคนอีกบางกลุ่ม โดยมากจะเกี่ยวข้องกับเพศ ศาสนา วัฒนธรรม ความเชื่อต่าง ๆ เป็นต้น ซึ่งจะส่งผลให้เกิดการได้เปรียบและเสียเปรียบขึ้นในการทดสอบ เราสามารถใช้โค้งแสดงลักษณะข้อสอบ (ICC) มาตรวจสอบความลำเอียงของข้อคำถาม โดยปกติถ้าข้อคำถามนั้นไม่มีความลำเอียง โค้งแสดงลักษณะข้อสอบ (ICC)

ไม่ว่าจะได้มาจากผู้สอบกลุ่มใดก็ตาม จะมีรูปร่าง (Shape) เหมือนกัน แต่ถ้าข้อคำถามนั้นมีความลำเอียง ได้ังแสดงลักษณะข้อสอบก็จะมีรูปร่างแตกต่างกัน (Hambleton and Swaminathan. 1985: 285-289)

(3) การเทียบมาตรฐาน (Test Equating) หมายถึง กระบวนการทางสถิติเพื่อปรับคะแนนที่ได้จากการทดสอบด้วยข้อสอบต่างชุดกัน ที่มีโครงสร้างเดียวกัน ให้สามารถเทียบกันได้ (เรวดี อินทสระ. 2530: 7) แต่ถ้าการทดสอบนั้นวัดความสามารถ (Ability or Trait) ที่ต่างกันแล้วจะนำมาเทียบกันไม่ได้ (Hambleton and Swaminathan. 1985: 199; citing Lord. 1977b; 1980b: 195-198) เนื่องจากความสามารถของผู้สอบที่ประมาณค่ามาจากทฤษฎีการตอบข้อสอบ (IRT) จะมีหน่วยของการวัดเดียวกัน และไม่ขึ้นอยู่กับชุดของแบบทดสอบ จึงสามารถนำมาเปรียบเทียบกันได้ (Hambleton and Swaminathan. 1985: 202)

(4) การทดสอบเทเลอร์ (Tailored Testing) หมายถึง การทดสอบที่มีการจัดข้อสอบให้มีความยากที่เหมาะสมกับระดับความสามารถของผู้สอบเป็นรายบุคคล นับเป็นการนำเอาทฤษฎีการตอบข้อสอบ (IRT) มาประยุกต์ใช้ได้อย่างมีประสิทธิภาพ (Urry. 1977: 181) โดยการนำเอาคุณสมบัติความไม่แปรเปลี่ยน (Invariant) ของค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบมาใช้ การทดสอบเทเลอร์นั้นผู้สอบแต่ละคนจะได้รับชุดของข้อสอบที่แตกต่างกันตามระดับความสามารถของผู้สอบ บางคนอาจจะได้รับข้อสอบชุดที่ยาก และอีกบางคนอาจจะได้รับข้อสอบชุดที่ง่าย แต่ก็ยังสามารถนำเอาความสามารถที่ประมาณค่าได้ จากการทดสอบเทเลอร์มาเปรียบเทียบกันได้ ทั้งนี้เพราะ ความสามารถที่ประมาณค่าได้ ยังอยู่ในมาตรวัดเดียวกัน (Common Ability Scale) (Hambleton and Cook. 1977: 90-91; Hambleton and Swaminathan. 1985: 296)

2. การทดสอบเทเลอร์ (Tailored Testing)

2.1 ความหมายและหลักการในการทดสอบเทเลอร์

การทดสอบเทเลอร์ (Tailored Testing) เป็นการทดสอบที่จัดข้อสอบให้มีความยากให้เหมาะสมกับระดับความสามารถของผู้สอบแต่ละคน (Lord. 1971: 43) และผู้สอบแต่ละคนก็

ไม่จำเป็นต้องทำข้อสอบเหมือน ๆ กันทุกข้อ จำนวนข้อก็ไม่จำเป็นต้องเท่ากัน ทั้งนี้ขึ้นอยู่กับรูปแบบและผลการตอบข้อสอบในการทดสอบเทเลอร์นั้น ๆ (Weiss and Kingsbury. 1984: 361) โดยทั่วไปแล้ว จะให้ผู้สอบเริ่มทำข้อสอบข้อที่มีความยากปานกลางก่อน ถ้าผลการตอบข้อสอบทุกข้อต่อไปจะมีความยากมากขึ้น หากตอบข้อสอบข้อนั้นผิดข้อต่อไปที่ได้รับก็จะง่ายขึ้น การทำข้อสอบข้อต่อไปจะดำเนินการเช่นนี้ไปจนสิ้นสุดการทดสอบ (Green and others. 1984: 347)

2.2 ประเภทของการทดสอบเทเลอร์

การทดสอบเทเลอร์พอจัดได้เป็น 2 ประเภทใหญ่ ๆ (Weiss. 1974: 78-110; Hambleton and Swaminathan. 1985: 297) ในแต่ละประเภทยังแบ่งย่อยออกได้อีก ดังนี้

(1) ยุทธวิธีสองขั้นตอน (Two-Stage Strategies)

(2) ยุทธวิธีหลายขั้นตอน (Multi-Stage Strategies)

(2.1) แบบแยกทางคงที่ (Fixed-Branching Model)

(2.1.1) รูปปิรามิดแบบขนาดขั้นคงที่ (Constant Step Size Pyramid)

(2.1.2) รูปปิรามิดแบบขนาดขั้นแปรผัน (Variable Step Size Pyramid)

(2.1.3) รูปปิรามิดแบบข้างตัด (Truncated Pyramid)

(2.1.4) รูปปิรามิดแบบมีหลายข้อในแต่ละขั้น (Multiple-Item Pyramid)

(2.1.5) รูปปิรามิดแบบให้หน้าหนักแก่ตัวเลือกของข้อสอบ เพื่อแยกทาง (Differential Response Option Branching)

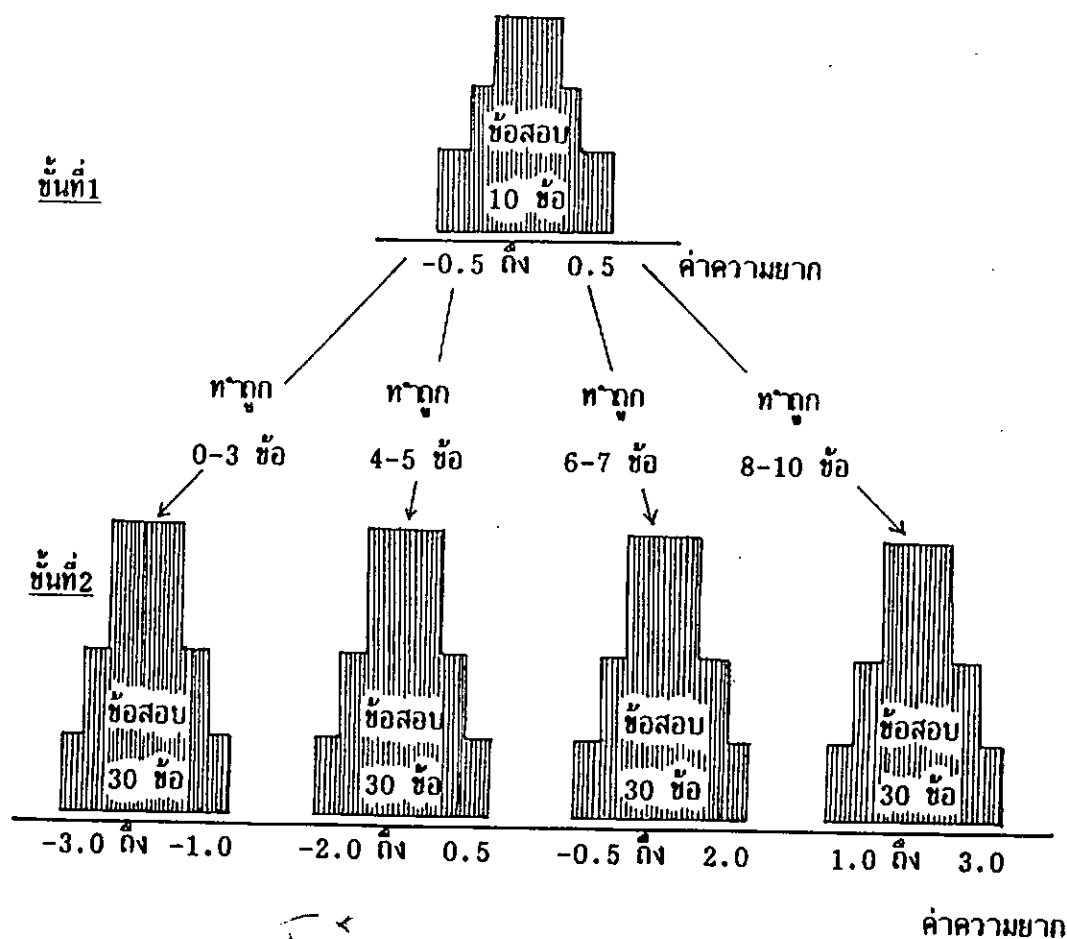
(2.1.6) แบบเฟล็กซ์เลเวล (Flexilevel Test)

(2.2) แบบแยกทางแปรผัน (Variable-Branching Model)

2.2.1 การทดสอบเทเลอร์โดยใช้อัตราวิธีสองชั้นตอน

การทดสอบเทเลอร์ โดยใช้อัตราวิธีสองชั้นตอน จะใช้ข้อสอบ 2 ชุด ข้อสอบชุดแรก (Routing Test) ซึ่งผู้สอบทุกคนต้องสอบเหมือนกัน สอบเพื่อแยกระดับความสามารถเพื่อที่จะไปรับข้อสอบในชุดที่สอง โดยปกติข้อสอบชุดแรกจะมีประมาณ 10 ข้อ ส่วนข้อสอบชุดที่สอง (Main Test) จะแบ่งเป็นข้อสอบชุดย่อย ๆ หลายชุดแยกตามระดับความยาก โดยปกติจะมีชุดละประมาณ 20-30 ข้อ ผู้สอบแต่ละคนจะได้รับข้อสอบชุดที่สองไม่เหมือนกัน ผู้สอบที่ได้รับการประเมินจากข้อสอบชุดแรกว่ามีความสามารถสูงจะได้รับข้อสอบชุดที่สองที่ยาก ส่วนผู้สอบที่มีความสามารถปานกลางจะได้รับข้อสอบชุดที่สองที่มีความยากปานกลาง และผู้สอบที่มีความสามารถต่ำจะได้รับข้อสอบชุดที่สองที่ง่าย (Lord. 1971: 3-31; Betz and Weiss. 1973)

ลอร์ด (Lord. 1980: 140-141) ได้แนะนำว่าข้อสอบชุดย่อย (Subtest) แต่ละฉบับในชั้นที่สอง (Main Test) ควรมีข้อสอบบางส่วนที่คาบเกี่ยวกันในเรื่องของความยากของข้อสอบรายข้อ (Overlap) ถ้าให้ n คือ จำนวนข้อสอบที่คาบเกี่ยวกัน L คือ จำนวนฉบับของข้อสอบชุดย่อยในชั้นที่สอง n คือ จำนวนข้อสอบทั้งหมด n_1 คือ จำนวนข้อสอบในชุดย่อยแต่ละฉบับในชั้นที่สอง จะได้ $n = (n_1 L - n) / (L - 1)$ โครงสร้างของการทดสอบเทเลอร์แบบสองชั้นตอน ดังตัวอย่าง ภาพประกอบ 3



ภาพประกอบ 3 แสดง โครงสร้างของการทดสอบเทเลอร์แบบสองชั้นคอน

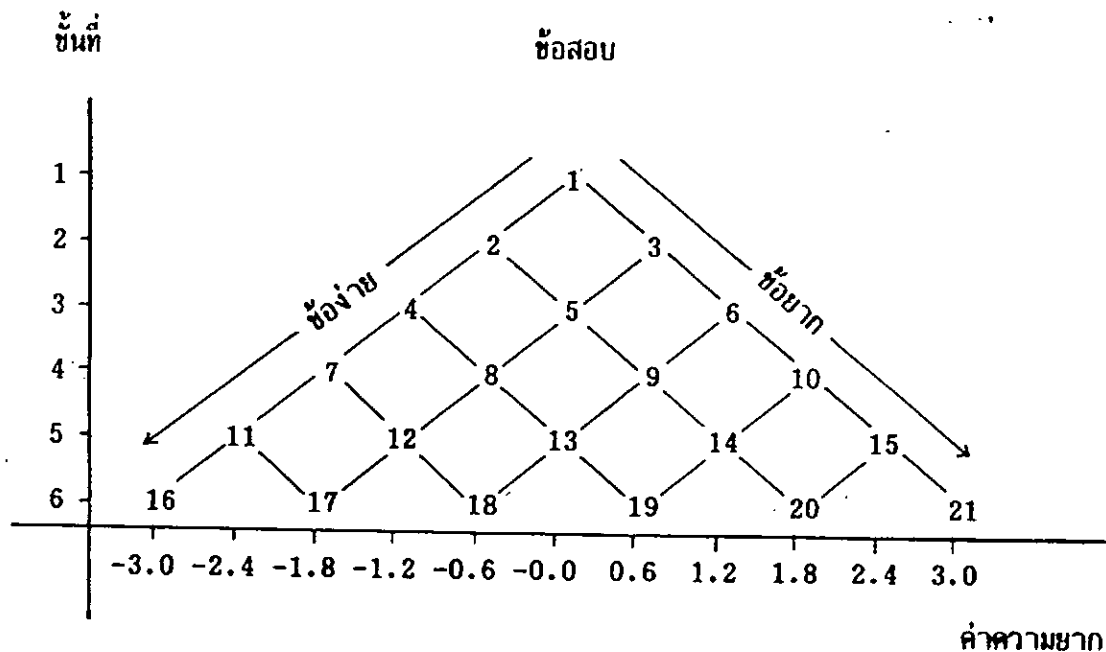
จากภาพประกอบ 3 ผู้สอบทุกคนต้องทำข้อสอบใน ชั้นที่ 1 ซึ่งมีอยู่ 10 ข้อ ผู้สอบที่ทำข้อสอบถูก 0-3 ข้อ จะไปทำข้อสอบ ชั้นที่ 2 ที่ง่ายที่สุด ส่วนผู้สอบที่ทำข้อสอบ ชั้นที่ 1 ถูก 4-5, 6-7, หรือ 8-10 ข้อ จะแยกไปทำข้อสอบใน ชั้นที่ 2 ที่ยากขึ้นตามลำดับ

2.2.2 การทดสอบเทเลอร์โดยใช้วิธีหลายชั้นคอนแบบแยกทางคงที่

การทดสอบเทเลอร์โดยใช้วิธีหลายชั้นคอน (Multi-Stage Strategies) แบบแยกทางคงที่ (Fixed-Branching Model) หมายถึง การทดสอบเทเลอร์ที่มีหลายชั้นคอน ในแต่ละชั้นคอนอาจจะมีข้อสอบเพียงข้อเดียวหรือหลายข้อก็ได้ โดยมีรูปแบบ หรือเส้นทางในการตอบ

ข้อสอบที่แน่นอน ซึ่งได้แก่การทดสอบเทเลอรัปราคูปริมาตรูปแบบต่าง ๆ

(1) รูปปริมาตรูปแบบขนาดขั้นคงที่ (Constant Step Size Pyramid) ได้แก่ การทดสอบเทเลอรัปราคูปริมาตรูปแบบเหลี่ยม โดยมีจำนวนข้อสอบในแต่ละชั้นเท่ากับลำดับที่ของชั้น กล่าวคือ ชั้นที่หนึ่งจะมีข้อสอบหนึ่งข้อ ชั้นที่สองจะมีข้อสอบสองข้อ ในชั้นที่หกจะมีข้อสอบหกข้อ ดังนั้นการทดสอบเทเลอรัปราคูปริมาตรูปแบบที่มีหกชั้น จะมีข้อสอบทั้งหมดยี่สิบเอ็ดข้อ ดังภาพประกอบ 4



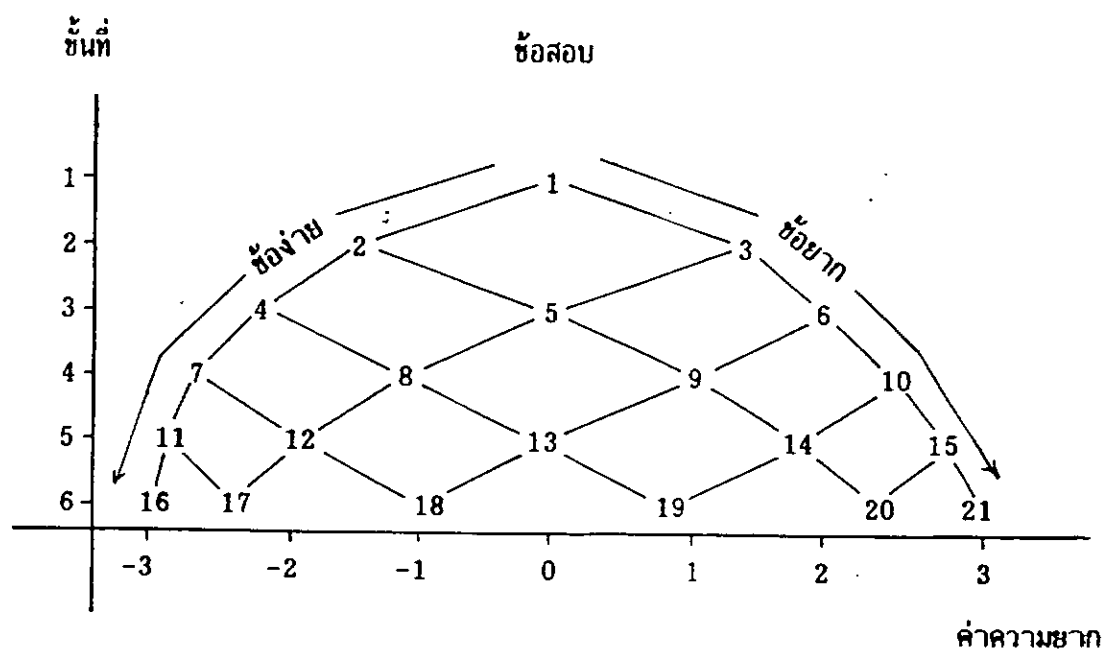
ภาพประกอบ 4 แสดงโครงสร้างของการทดสอบเทเลอรัปราคูปริมาตรูปแบบขนาดขั้นคงที่ ที่มี 6 ชั้น

จากภาพประกอบ 4 แขนงบนแสดงถึงค่าความยากของข้อสอบ (b) ที่สัมพันธ์กับข้อสอบในการทดสอบเทเลอรัปราคูปริมาตรูปแบบ ซึ่งจะเห็นว่าความยากมีค่าอยู่ระหว่าง -3.0 ถึง 3.0 โดยแบ่งออกเป็น 10 ช่วงเท่า ๆ กัน ฉะนั้นค่าความยากจะห่างกันช่วงละ 0.6 ข้อสอบที่อยู่ในแนวตั้งแนวเดียวกัน จะมีค่าความยากเท่ากัน และช่วงห่างค่าความยากของข้อที่อยู่ติดกันภายในชั้นมีค่าเท่ากันตลอด คือ 1.2

ในการทำแบบทดสอบ ผู้สอบจะต้องทำข้อสอบชั้นละหนึ่งข้อ โดยเริ่มทำข้อที่อยู่บนยอดสาม

เหลี่ยมซึ่งมีความยากปานกลางก่อน ถ้าทำถูกขั้นต่อไปจะยากขึ้น แต่ถ้าทำผิดขั้นต่อไปจะ ไปทำข้อที่ง่ายขึ้น กระบวนการทดสอบจะเป็นเช่นนี้จนถึงขั้นสุดท้าย

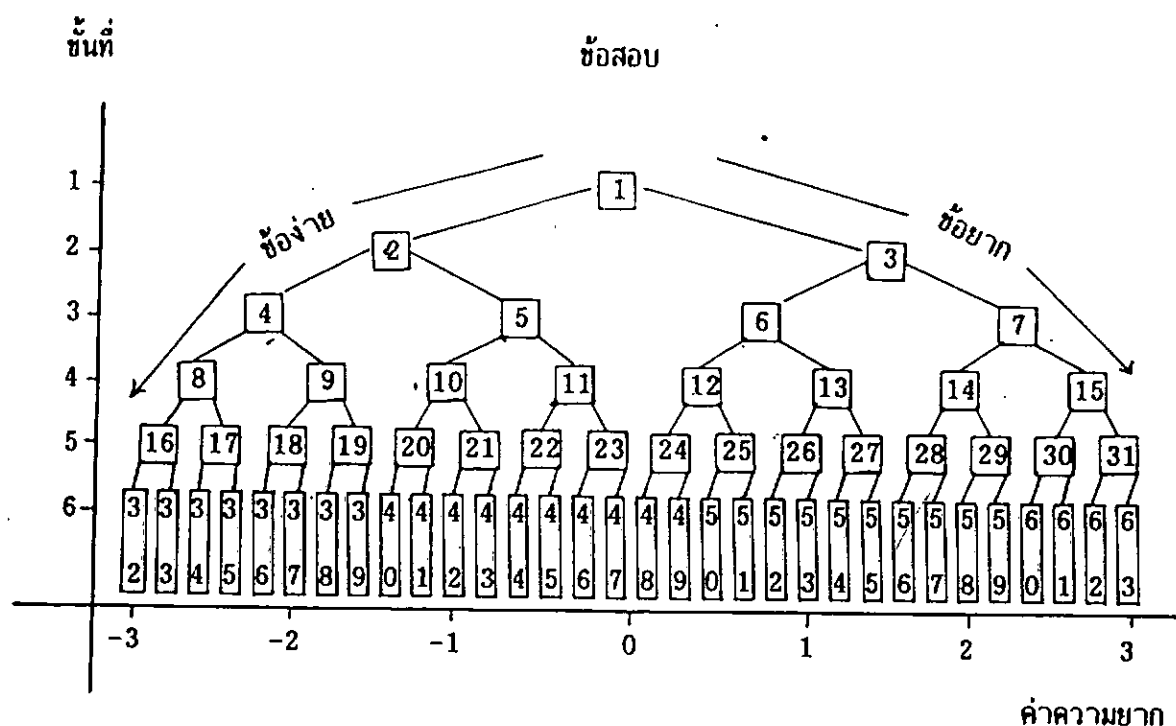
(2) รูปปิรามิดแบบขนาดขั้นแปรผัน (Variable Step Size Pyramid) โดยที่ แพทเตอร์สัน และลอร์ด (Wiess. 1974: 18; citing Paterson. 1962; Lord. 1971) มีความเห็นว่า รูปปิรามิดแบบขนาดขั้นคงที่ ขาดประสิทธิภาพเกี่ยวกับความไวในการกำหนดเส้นทางการตอบข้อสอบ จึงได้กำหนดช่วงห่างค่าความยากของข้อที่อยู่ติดกันภายในชั้นให้มีค่าไม่เท่ากัน โดยให้ข้อที่อยู่กลาง ๆ มีช่วงห่างของค่าความยากมาก แล้วค่อย ๆ ลดลงในข้อที่ค่อนข้างง่ายหรือยาก แพทเตอร์สัน ได้กำหนดการเพิ่มหรือลดนั้น จะเพิ่มหรือลดไปยังตำแหน่งกึ่งกลางระหว่าง 2 ข้อในชั้นที่มาก่อน หรือ ระหว่างข้อในชั้นที่มาก่อนกับค่าความยากสูงสุดหรือต่ำสุด ดังตัวอย่าง ภาพประกอบ 5



ภาพประกอบ 5 แสดงโครงสร้างของการทดสอบเทเลอรัูปปิรามิดแบบขนาดขั้นแปรผันที่มี 6 ชั้น

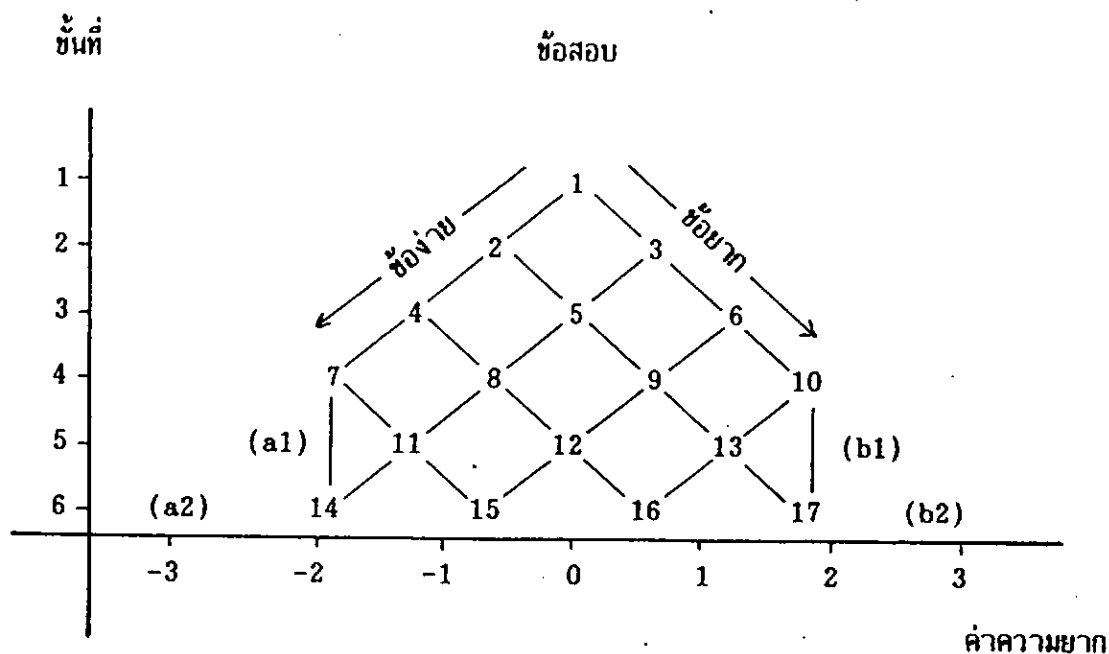
จากภาพประกอบ 5 ในขั้นที่ 1 มีข้อสอบเพียงข้อเดียว คือ ข้อ 1 ซึ่งมีความยากเท่ากับ 0.000 ในขั้นที่ 2 มีข้อสอบสองข้อ คือ ข้อ 2 มีความยากอยู่กึ่งกลางระหว่าง -3.000 (ความยากต่ำสุด) กับ 0.000 (ความยากของข้อที่ 1) ซึ่งก็คือ -1.500 ข้อ 3 มีความยากอยู่กึ่งกลางระหว่าง 3.000 (ความยากสูงสุด) กับ 0.000 (ความยากของข้อที่ 1) ซึ่งก็คือ 1.500 ในขั้นที่ 3 มีข้อสอบสามข้อ คือ ข้อ 4 มีความยากอยู่กึ่งกลางระหว่าง -3.000 (ความยากต่ำสุด) กับ -1.500 (ความยากของข้อที่ 2) ซึ่งก็คือ -2.250 ข้อ 5 มีความยากเท่ากับ 0.000 ข้อ 6 มีความยากอยู่กึ่งกลางระหว่าง 1.500 (ความยากของข้อที่ 3) กับ 3.000 (ความยากสูงสุด) ซึ่งก็คือ 2.250 ในขั้นที่ 4 มีข้อสอบสี่ข้อ คือ ข้อ 7 มีความยากอยู่กึ่งกลางระหว่าง -3.000 (ความยากต่ำสุด) กับ -2.250 (ความยากของข้อที่ 4) ซึ่งก็คือ -2.625 ข้อ 8 มีความยากอยู่กึ่งกลางระหว่าง -2.250 (ความยากของข้อที่ 4) กับ 0.000 (ความยากของข้อที่ 5) ซึ่งก็คือ -1.125 ข้อ 9 มีความยากอยู่กึ่งกลางระหว่าง 0.000 (ความยากของข้อที่ 5) กับ 2.250 (ความยากของข้อที่ 6) ซึ่งก็คือ 1.125 และข้อ 10 มีความยากอยู่กึ่งกลางระหว่าง 2.250 (ความยากของข้อที่ 6) กับ 3.000 (ความยากสูงสุด) ซึ่งก็คือ 2.625 สำหรับในขั้นอื่น ๆ ก็ในทำนองเดียวกัน

จากรูปปริมิตแบบขนาดขั้นแปรผัน (Variable Step Size Pyramid) นี้ คอมาลอร์ด (Weiss, 1974: 20; citing Lord, 1971) ได้ใช้วิธีการเพิ่มจำนวนข้อในแต่ละชั้นให้เป็นสองเท่าของจำนวนข้อสอบในขั้นที่มาก่อน ซึ่งให้ชื่อรูปแบบใหม่นี้ว่า แบบรอนบิน-มอนโร (Robin-Monro) ดังตัวอย่าง ภาพประกอบ 6



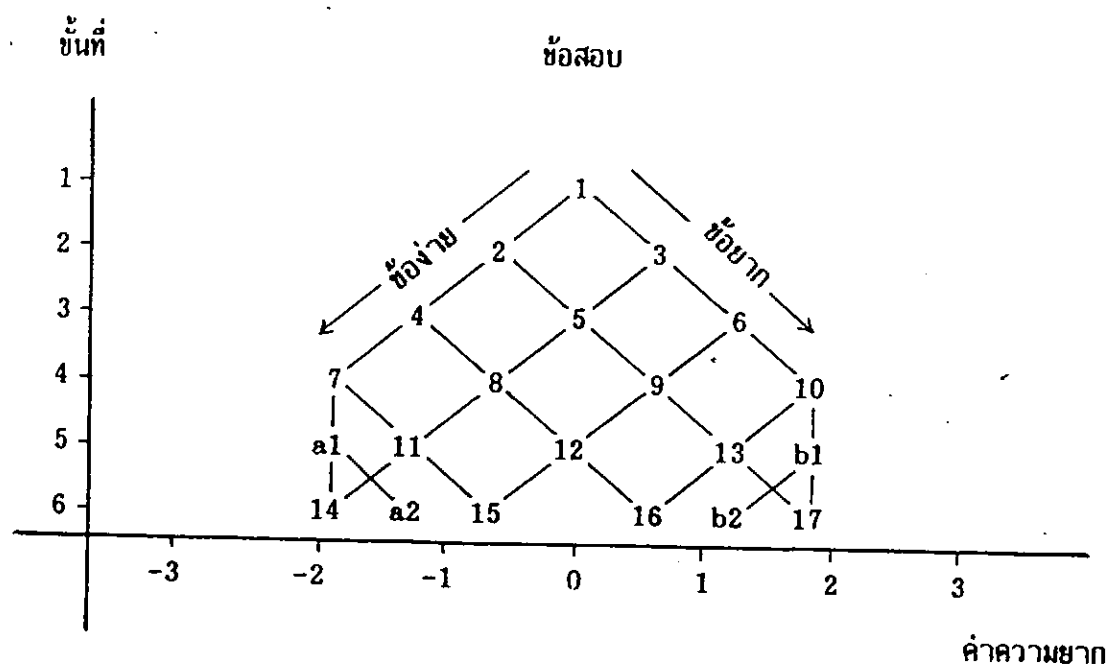
ภาพประกอบ 6 แสดงโครงสร้างของการทดสอบเทเลอรัปปีรามิดแบบรอนบิน-มอนโร (Robin-Monro)

(3) รูปปิรามิดแบบข้างตัด (Truncated Pyramid) เป็นรูปแบบที่ มุสซิโอ (Weiss, 1974: 22; citing Mussio, 1973) เป็นผู้เสนอรูปแบบนี้ขึ้นมา เพื่อให้ลดจำนวนข้อสอบลงจากรูปปิรามิดแบบขนาดชั้นคงที่ จึงใช้วิธีสกัดการสะท้อนกลับ หรือวิธีรักษาการสะท้อนกลับ (Reflecting or Retaining Barrier) ดังตัวอย่าง ภาพประกอบ 7 และ 8



ภาพประกอบ 7 แสดงโครงสร้างของการทดสอบเทเลอรัปป์ปิรามิดแบบข้างคัต (Truncated Pyramid) สกัการสะท้อนกลับ (Reflecting Barrier)

จากภาพประกอบ 7 การตอบข้อสอบใน ชั้นที่ 1 ถึง ชั้นที่ 3 มีวิธีตอบเช่นเดียวกับรูปปิรามิดแบบขนาดชั้นคงที่ จนถึงจุดสกัดในชั้นที่ 4 ถ้าตอบข้อสอบข้อที่ 7 ถูก ก็จะแยกไปตอบข้อ 11 แต่ถ้าตอบข้อ 7 ผิด ก็จะลงไปตอบข้อ 14 และในทางนองเดียวกัน ถ้าตอบข้อ 10 ผิด ก็จะแยกไปตอบข้อ 13 แต่ถ้าตอบข้อ 10 ถูก ก็จะลงไปตอบข้อ 17 เลย



ภาพประกอบ 8 แสดงโครงสร้างของการทดสอบเทเลอร์รูปปิรามิดแบบข้างตัด (Truncated Pyramid) รักษาการสะท้อนกลับ (Retaining Barrier)

จากภาพประกอบ 8 การตอบข้อสอบใน ชั้นที่ 1 ถึง ชั้นที่ 3 วิธีตอบ เช่นเดียวกับรูปปิรามิดแบบขนาดชั้นคงที่ จนถึงจุดสกัดในชั้นที่ 4 ถ้าตอบข้อสอบข้อที่ 7 ถูก ก็จะแยกไปตอบข้อ 11 แต่ถ้าตอบข้อ 7 ผิด ก็จะลงไปตอบข้อ a1 ซึ่งมีความยากเท่ากับข้อ 7 และถ้าตอบข้อ a1 ผิดอีก ก็จะลงไปตอบข้อ 14 แต่ถ้าตอบข้อ a1 ถูก ก็จะแยกไปตอบข้อ a2 ซึ่งมีความยากเท่ากับข้อ 11

(4) รูปปิรามิดแบบมีหลายข้อในแต่ละชั้น (Multiple-Item Pyramid) เป็นรูปแบบที่ แครธไวท์โฮซายเซอร์ และ ลินน์ (Weiss. 1974: 25; citing Krathworth and Huyser. 1956; Linn. 1969) เป็นผู้เสนอรูปแบบนี้ขึ้นมา โดยให้มีจำนวนข้อในแต่ละชั้นมีมากกว่าหนึ่งข้อ ทั้งนี้เพื่อให้การตอบเพื่อแยกทางได้ถูกต้องมากยิ่งขึ้น ดังตัวอย่าง ภาพประกอบ 9

เป็นข้อเริ่มต้นสอบสำหรับผู้สอบทุกคน ภายในข้อสอบกลุ่มง่ายจะ เรียงลำดับข้อสอบจากข้อที่ยากที่สุด ภายในกลุ่มนี้ไปหาข้อง่าย ข้อสอบกลุ่มง่ายนี้จะพิมพ์ด้วยหมึกสีแดง ส่วนข้อสอบในกลุ่มยากจะ เรียงลำดับจากข้อที่ง่ายที่สุดภายในกลุ่มนี้ไปหาข้อที่ยาก และจะพิมพ์ด้วยหมึกสีน้ำเงิน ดังตัวอย่าง ภาพประกอบ 11 (Lord. 1971: 805; Lord. 1980: 114)

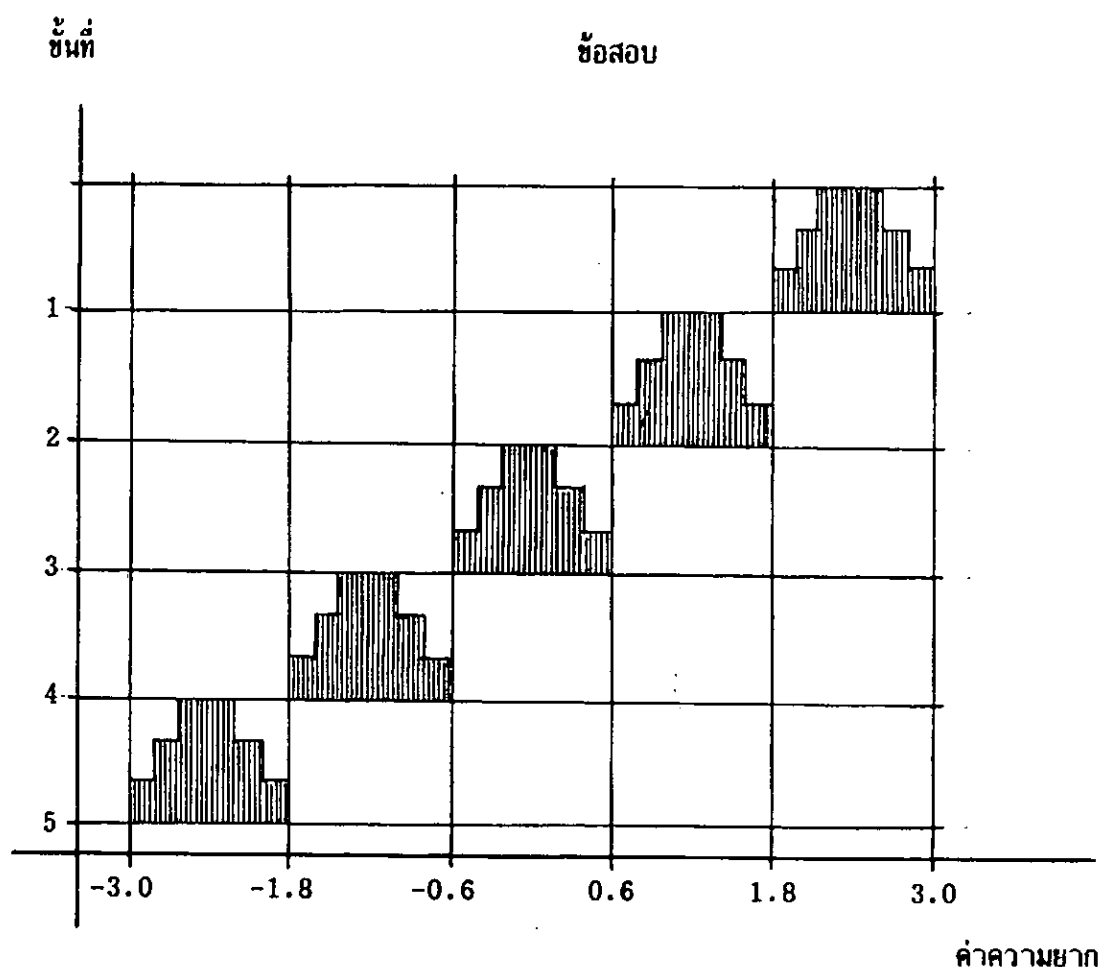
ข้อ 0. [มีความยากอยู่กึ่งกลางระหว่างข้อสอบกลุ่มง่ายกับกลุ่มยาก]	
[ข้อสอบกลุ่มง่าย] [พิมพ์สีแดง]	[ข้อสอบกลุ่มยาก] [พิมพ์สีน้ำเงิน]
ข้อ 1. [ข้อที่ยากที่สุดในกลุ่มนี้]	ข้อ 1. [ข้อที่ง่ายที่สุดในกลุ่มนี้]
ข้อ 2.	ข้อ 2.
ข้อ 3.	ข้อ 3.
.	.
.	.
ข้อ 37. [ข้อที่ง่ายที่สุดในกลุ่มนี้]	ข้อ 37. [ข้อที่ยากที่สุดในกลุ่มนี้]

ภาพประกอบ 11 แสดงตัวอย่างข้อสอบแบบเฟล็กซ์เลเวล (Flexilevel Test)

สำหรับกระดาษตอบก็ได้รับการออกแบบมาพิเศษ คือ เมื่อนักเรียนตอบถูก จุดสีน้ำเงิน ก็จะปรากฏขึ้น ข้อต่อไปผู้สอบจะต้อง ไปเลือกทำข้อสอบในกลุ่มสีน้ำเงิน (กลุ่มยาก) ในข้อที่มีหมายเลขต่ำสุดที่ยังไม่ได้ทำ และถ้าตอบผิดจุดสีแดงก็จะปรากฏขึ้น ผู้สอบจะต้อง ไปเลือกทำข้อสอบในกลุ่มสีแดง (กลุ่มง่าย) ในข้อที่มีหมายเลขต่ำสุดที่ยังไม่ได้ทำ ถ้าให้ k เป็นจำนวนข้อสอบเดิม ผู้สอบในข้อสอบเฟล็กซ์เลเวลจะต้องทำข้อสอบทั้งหมด $(k+1)/2$ ข้อ (Lord. 1971: 805)

ข้อสอบเพ็ล็กซ์เลเวล ก็คือ ข้อสอบเทเลอร์รูปปิรามิดที่มีข้อสอบเพียงข้อเดียวในแต่ละระดับความยากนั่นเอง (Hambleton and Swaminathan. 1985: 300; citing Betz and Weiss. 1975) แต่มีข้อแตกต่างกันที่การใช้เส้นทางในการเลือกตอบข้อสอบ

ต่อมาได้มีการปรับปรุงการทดสอบแบบนี้ โดยการแบ่งความยากของข้อสอบออกเป็นช่วง ๆ (Stratified) ในแต่ละช่วงความยาก (Stratum) จะมีข้อสอบหลายข้อ โดยให้ชื่อการทดสอบแบบนี้ว่า สเตร็คเค็พคัพ เทสคิง (Stradaptive=Stratified+Adaptive) , (Hambleton and Swaminathan. 1985: 300; citing Weiss. 1973; Waters. 1977) ดังตัวอย่างภาพประกอบ 12



ภาพประกอบ 12 แสดงโครงสร้างของการทดสอบเทเลอร์แบบสเตร็คเค็พคัพ เทสคิง
(Stradaptive Testing)

2.2.3 การทดสอบเทเลอร์โดยใช้วิธีหลายขั้นตอนแบบแยกทางแปรผัน

การทดสอบเทเลอร์โดยใช้วิธีหลายขั้นตอน (Multi-Stage Strategies) แบบแยกทางแปรผัน (Variable Branching Models) หมายถึง รูปแบบการตอบข้อสอบหลายขั้นตอนที่ไม่ได้กำหนดข้อสอบ และเส้นทางในการตอบข้อสอบไว้แน่นอน โดยมากจะใช้คอมพิวเตอร์เป็นผู้เลือกข้อสอบจากกลุ่มข้อสอบ (Item Pool) โดยให้มีค่าความยากของข้อสอบใกล้เคียงกับความสามารถของผู้สอบ สำหรับการประมาณค่าความสามารถของผู้สอบ หลังจากตอบข้อสอบไปแล้วในแต่ละข้อ หรือหลังจากสิ้นสุดการทดสอบแล้ว ที่นิยมใช้กันมากมี 2 วิธี คือ

(1) ยุทธวิธีความเป็นไปได้สูงสุด (Maximum Likelihood) สำหรับการประมาณค่าโดยวิธีนี้ยังมีหลายแบบ เช่น แบบ คอนดิชันนัล แมกซิมัม โลลิฮูด (Conditional Maximum Likelihood) (Hambleton and Swaminathan. 1985: 81-91) แบบ จอยท์ แมกซิมัม โลลิฮูด (Joint Maximum Likelihood) (Lord. 1986: 157; citing Wingersky. 1983) แบบ มาจินอล แมกซิมัม โลลิฮูด (Marginal Maximum Likelihood) (Lord. 1986: 157; citing Wingersky. 1983)

(2) ยุทธวิธีของเบย์ (Bayesian Strategies) การประมาณค่าโดยวิธีนี้ยังมีหลายวิธี คือ วิธีของ โอเวน (Owen. 1969, 1975: 351-356) วิธีของ แมกซ์ไบรด์, สวามินาธาน, และ วูด (Hambleton and Swaminathan. 1985: 300; citing McBride. 1977; Swaminathan, in press; Wood. 1976: 196)

2.3 วิธีการให้คะแนนการทดสอบเทเลอร์ (Scoring)

ได้มีผู้เสนอวิธีการให้คะแนนการทดสอบเทเลอร์ ไว้หลายวิธี พอสรุปได้ดังนี้

(1) ให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ถูกต้อง (The most difficulty of Item answered correctly) เป็นการให้คะแนน โดยดูจากข้อสอบการทดสอบเทเลอร์ที่ผู้สอบทำมาทั้งหมด ข้อใดที่มีค่าความยากสูงสุดที่ผู้สอบทำถูกต้อง ถือว่าเป็นค่าความสามารถของผู้สอบคนนั้น ค่าความยากที่ว่านี้ ควรจะเป็นค่าความยากที่วิเคราะห์มาจากทฤษฎีการตอบข้อสอบ (IRT)

(2) ให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ถูกต้อง (Average difficulty of

all Item answered correctly) เป็นการให้คะแนนโดยการเอาค่าความยากของข้อสอบเทเลอร์ทุกข้อที่ผู้สอบทำถูกมาหาค่าเฉลี่ย ค่าเฉลี่ยความยากนี้ ถือเป็นค่าความสามารถของผู้สอบคนนั้น

(3) ให้คะแนนตามค่าเฉลี่ยความยากของทุกข้อที่ทำ (Average difficulty) วิธีการให้คะแนนแบบนี้เสนอโดย ลอร์ด (Weiss. 1974: 33; citing Lord. 1970) วิธีการคือ นำค่าความยากของข้อสอบเทเลอร์ทุกข้อที่ผู้สอบได้ทำ โดยไม่คำนึงว่าจะตอบถูกหรือไม่ มาหาค่าเฉลี่ย ค่าเฉลี่ยนี้จะถือเป็นความสามารถของผู้สอบผู้นั้น

(4) ให้คะแนนตามค่าความยากของข้อสุดท้ายที่ทำ (Difficulty of the final term) นั่นคือ ไม่คำนึงถึงว่าข้อสุดท้ายจะตอบถูกหรือไม่ก็ตาม จะถือว่าค่าความยากของการทดสอบเทเลอร์ข้อสุดท้ายที่ผู้สอบได้ทำคือค่าความสามารถของผู้สอบผู้นั้น

(5) ให้คะแนนตามค่าความยากของข้อสอบที่ต่อจากข้อสุดท้ายที่ทำ [Difficulty of (N+1)th Item] ซึ่งมีลักษณะคล้ายกับวิธีที่ 4 แต่เป็นการตรวจสอบต่อไปอีกชั้นหนึ่ง กล่าวคือ ถ้าข้อสุดท้ายที่ทำถูก คะแนนในชั้นต่อไปก็จะเพิ่มขึ้น และถ้าข้อสุดท้ายที่ทำผิด ชั้นต่อไปคะแนนก็จะลดลง

(6) ให้คะแนนโดยใช้หลักการของเบย์ส (Bayesian Strategies) การใช้หลักการของเบย์สนี้ยังมีวิธีการย่อยหลายวิธี วิธีที่นิยมใช้กันมาก คือ วิธีที่นักสถิติ โรเจอร์ เจ โอเวน (Owen. 1969; 1975: 351-356; Thorndike. 1982: 303; citing Owen. 1975) เป็นผู้เสนอขึ้น เพื่อใช้ในการประมาณค่าความสามารถของผู้สอบในการทดสอบเทเลอร์โดยเฉพาะ ทำให้ชื่อว่า เบย์เซียน อัปเดตติง (Bayesian Updating) ซึ่งสามารถประมาณค่าความสามารถของผู้สอบได้ค่อนข้างคงที่ โดยมีสูตร ดังนี้

ในการนี้ที่ตอบข้อสอบถูก

$$\theta_{n+1} = \theta_n + (1 - c_g) \left[\frac{\frac{\sigma_m^2}{\frac{1}{2} + \frac{\sigma_m^2}{a_g}}}{\sqrt{\frac{1}{2} + \frac{\sigma_m^2}{a_g}}} \right] \left[\frac{\phi(D)}{c_g + (1 - c_g)\Phi(-D)} \right]$$

และ

$$\sigma_{m+1}^2 = \sigma_m^2 \left\{ 1 - \left[\frac{1 - c_g}{1 + \frac{1}{\frac{2}{a_g} \sigma_m^2}} \right] \left[\frac{\phi(D)}{A} \right] \left[\frac{(1 - c_g)\phi(D)}{A} - D \right] \right\}$$

ในการที่หาค่าข้อสอบผิด

$$\theta_{m+1} = \theta_m - \left\{ \frac{\sigma_m^2}{\sqrt{\frac{1 + \sigma_m^2}{\frac{2}{a_g}}}} \right\} \left[\frac{\phi(D)}{\Phi(D)} \right]$$

และ

$$\sigma_{m+1}^2 = \sigma_m^2 \left\{ 1 - \left[\frac{\phi(D)}{1 + \frac{1}{\frac{2}{a_g} \sigma_m^2}} \right] \left[\frac{\phi(D)}{\Phi(D)} + D \right] / \Phi(D) \right\}$$

เมื่อ $\phi(D)$ คือ ค่าออร์ดิเนต (Ordinate) ของโค้งปกติ ณ จุด D

$\Phi(D)$ คือ พื้นที่ใต้โค้งปกติจากค่า D ต่ำสุดจนถึงจุด D

$$D = \frac{b_g - \theta_m}{\sqrt{\frac{1 + \sigma_m^2}{\frac{2}{a_g}}}}$$

$$A = c_g + (1 - c_g)\Phi(-D)$$

θ_m คือ ความสามารถของผู้สอบที่ประมาณค่าได้ก่อนตอบข้อสอบข้อที่ $m+1$ โดยปกติถ้าเป็นการประมาณค่าครั้งแรก ที่ยังไม่ทราบความสามารถของผู้สอบ จะให้ $\theta_m = 0.000$

σ_m^2 คือ ความแปรปรวนในการประมาณค่าความสามารถของผู้สอบ ก่อนตอบ

ข้อสอบข้อที่ $m+1$ โดยปกติถ้าเป็นการประมาณค่าครั้งแรกที่ยังไม่ทราบ

ค่าความแปรปรวน จะให้ $\sigma_m^2 = 1.000$

θ_{m+1} คือ ความสามารถของผู้สอบที่ประมาณค่าได้เมื่อตอบข้อสอบข้อที่ $m+1$ แล้ว

σ_{m+1}^2 คือ ความแปรปรวนในการประมาณค่าความสามารถของผู้สอบ เมื่อตอบ
ข้อสอบข้อที่ $m+1$ แล้ว

a_g คือ ค่าอำนาจจำแนกของข้อสอบข้อที่ $m+1$

b_g คือ ค่าความยากของข้อสอบข้อที่ $m+1$

c_g คือ ค่าการเดาของข้อสอบข้อที่ $m+1$

(7) ให้คะแนนโดยใช้หลักการความเป็นไปได้สูงสุด (Maximum Likelihood) การให้คะแนนโดยใช้หลักการนี้ยังมีวิธีย่อยหลายวิธี วิธีที่นิยมใช้มาก คือ วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (Conditional Maximum Likelihood) การประมาณค่าโดยวิธีนี้มีข้อจำกัดในกรณีที่ผู้สอบตอบข้อสอบทุกหมัดหรือผิดหมด จะไม่สามารถประมาณค่าได้ นอกจากนี้ แฮมมิลตัน (Hambleton and Swaminathan, 1985: 87; citing Samejima, 1973a) ยังพบว่าในกรณีที่ใช้ 3-พารามิเตอร์บางครั้งอาจจะมีค่าสูงสุดหลายค่า ในกรณีนี้ ลอร์ด ได้ให้ข้อสังเกตว่าปัญหานี้จะไม่เกิดขึ้นถ้าใช้ข้อสอบมากข้อ ($k > 20$) ซึ่งมีขั้นตอนในการประมาณค่า ดังนี้ (Hambleton and Swaminathan, 1985: 81-91)

ขั้นที่ 1 ประมาณค่าความสามารถเริ่มต้น (θ_m ; $m=0$) โดยใช้สูตร ดังนี้

$$\theta_0 = \ln[r_a / (k - r_a)]$$

เมื่อ $r_a = \sum a_j U_j$; ($U_j=1$ เมื่อตอบข้อสอบถูก, $U_j=0$ เมื่อตอบข้อสอบผิด,

a_j คือ ค่าอำนาจจำแนกของข้อสอบข้อที่ j)

k คือ จำนวนข้อสอบทั้งหมดที่ผู้สอบได้ทำ ในการทดสอบเทเลอร์

ขั้นที่ 2 หาค่า $P_j(\theta_m)$ และ $Q_j(\theta_m)$ โดยใช้สูตร ดังนี้

$$P_j(\theta_m) = \frac{c_j + (1 - c_j) \cdot \frac{e^{Da_j(\theta_m - b_j)}}{1 + e^{Da_j(\theta_m - b_j)}}}{1 + e^{Da_j(\theta_m - b_j)}}$$

$$Q_j(\theta_m) = 1 - P_j(\theta_m) = \frac{1 - c_j}{1 + e^{Da_j(\theta_m - b_j)}}$$

เมื่อ θ_m คือ ความสามารถของผู้สอบที่ประมาณค่าได้ในครั้งที่ m

a_j คือ ค่าอำนาจจำแนกของข้อสอบข้อที่ j

b_j คือ ค่าความยากของข้อสอบข้อที่ j

c_j คือ ค่าการเดาของข้อสอบข้อที่ j

D คือ ค่าคงที่ ซึ่งมีค่าเท่ากับ 1.7

ขั้นที่ 3 หาค่าตัวปรับแก้ (h_m) โดยใช้สูตร

$$h_m = \frac{D[r_a - \sum P_j(\theta_m)]}{-D^2 \sum P_j(\theta_m) Q_j(\theta_m)}$$

ขั้นที่ 4 ประมาณค่าความสามารถของผู้สอบใหม่ (θ_{m+1}) โดยใช้สูตร

$$\theta_{m+1} = \theta_m - h_m$$

ขั้นที่ 5 ทำการคำนวณซ้ำในขั้นที่ 2, 3, และ 4 จนกระทั่ง h_n เข้าใกล้ศูนย์ ($h_n < 0.001$)

สำหรับการวิจัยครั้งนี้ได้เลือกวิธีให้คะแนน 4 วิธี ด้วยกัน คือ

- (1) ให้คะแนนตามค่าความยาก (β_j) ของข้อที่ยากที่สุดที่ทำได้
- (2) ให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทำได้
- (3) ให้คะแนนโดยใช้วิธีของเบส์ (Bayesian Updating)

สำหรับการให้คะแนนโดยวิธีนี้ เนื่องจากการวิจัยครั้งนี้ได้ใช้ทฤษฎีการตอบข้อสอบ แบบ 1-พารามิเตอร์ ฉะนั้นค่าอำนาจจำแนก (a_g) จะเป็น 1.000 และค่าการเดา (c_g) จะเป็น 0.000 สูตรที่ใช้จึงเป็น ดังนี้

ในการให้คะแนนข้อสอบถูก

$$\theta_{n+1} = \theta_n + \{\sigma_n^2 / \sqrt{(1+\sigma_n^2)}\} \{\phi(D)/\Phi(-D)\}$$

และ

$$\sigma_{n+1}^2 = \sigma_n^2 [1 - \{1/(1+1/\sigma_n^2)\} \{\phi(D)/\Phi(-D)\} \{\phi(D)/\Phi(-D) - D\}]$$

ในการให้คะแนนข้อสอบผิด

$$\theta_{n+1} = \theta_n - \{\sigma_n^2 / \sqrt{(1+\sigma_n^2)}\} \{\phi(D)/\Phi(D)\}$$

และ

$$\sigma_{n+1}^2 = \sigma_n^2 [1 - \{\phi(D)/(1+1/\sigma_n^2)\} \{\phi(D)/\Phi(D) + D\} / \Phi(D)]$$

เมื่อ

$$D = (b_g - \theta_n) / \sqrt{(1+\sigma_n^2)}$$

- (4) ให้คะแนนโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (Conditional Maximum Likelihood)

สำหรับการให้คะแนนโดยวิธีนี้ เนื่องจากการวิจัยครั้งนี้ได้ใช้ทฤษฎีการตอบข้อสอบ แบบ 1-พารามิเตอร์ ฉะนั้นค่าอำนาจจำแนก (a_g) จะเป็น 1.000 และค่าการเดา (c_g) จะเป็น 0.000 สูตรที่ใช้จึงเป็น ดังนี้

ขั้นที่ 1 ประมวลค่าความสามารถเริ่มต้นของผู้สอบ (θ_m ; $m=0$) จากสูตร

$$\theta_m = \ln[r/(k - r)] \quad ; \quad \text{เมื่อ } r \text{ คือ จำนวนข้อที่ตอบถูก}$$

ขั้นที่ 2 หาค่า $P_j(\theta_m)$ และ $Q_j(\theta_m)$ จากสูตร *

$$P_j(\theta_m) = \exp D(\theta_m - b_j) / [1 + \exp D(\theta_m - b_j)]$$

$$Q_j(\theta_m) = 1 - P_j(\theta_m) = 1 / [1 + \exp D(\theta_m - b_j)]$$

ขั้นที่ 3 หาค่าตัวปรับแก้ (h_m) โดยใช้สูตร

$$h_m = D[r - \sum P_j(\theta_m)] / [-D^2 \sum P_j(\theta_m) Q_j(\theta_m)]$$

ขั้นที่ 4 ประมวลค่าความสามารถของผู้สอบใหม่ (θ_{m+1}) โดยใช้สูตร

$$\theta_{m+1} = \theta_m - h_m$$

ขั้นที่ 5 ทำการคำนวณซ้ำในขั้นที่ 2, 3, และ 4 จนกระทั่ง h_m เข้าใกล้ศูนย์ ($h_m < 0.001$)

2.2.4 ผลงานวิจัยที่เกี่ยวข้องกับการทดสอบเทเลอร์

2.2.4.1 ผลงานวิจัยที่เกี่ยวข้องภายในประเทศ ได้มีผลงานวิจัยเกี่ยวกับการทดสอบเทเลอร์ ภายในประเทศหลายเรื่อง บางเรื่องที่น่าสนใจ ดังจะนำเสนอต่อไปนี้

งามนิศย์ ธาตุทอง และปรีชา เครือวรรณ (2527: 89-92) ได้สร้างและประเมินผล การใช้การทดสอบเทเลอร์แบบแฟล็กซีเลเวล ในวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 2 โดยการใช้การทดสอบเทเลอร์แบบแฟล็กซีเลเวล 1 ฉบับ ที่มี 18 ข้อ มีข้อสอบจำนวน 35 ข้อ และแบบทดสอบดั้งเดิม 1 ฉบับ ที่มีจำนวนข้อสอบ 35 ข้อ จากการศึกษาพบว่า

ความสัมพันธ์ระหว่างคะแนนจากการทดสอบแบบแฟล็กซีเลเวล กับ คะแนนผลสัมฤทธิ์ทางการเรียน มีค่าเท่ากับ .7238 และ ความสัมพันธ์ระหว่างคะแนนจากการทดสอบแบบดั้งเดิมกับ

ผลสัมฤทธิ์ทางการเรียน มีค่าเท่ากับ .6109 เมื่อทดสอบความแตกต่างระหว่างค่าสัมประสิทธิ์สหสัมพันธ์ดังกล่าว ปรากฏว่า ไม่แตกต่างกัน แต่ผู้ถูกทดสอบส่วนใหญ่มีทัศนคติที่ดีต่อการสอบด้วยข้อสอบแบบแฟล็กชีลเวล

จิราพร ไกรสรศิริเวท (2529) ได้ศึกษาความสัมพันธ์ระหว่างรูปแบบบางประการของการทดสอบเทเลอรัปปีรามีคกับความสามารถทางการเรียน โดยใช้การทดสอบเทเลอรัปปีรามีคที่มี 8 ชั้น 3 แบบ คือ แบบที่ใช้ขนาดชั้นคงที่ ขนาดชั้นแปรผัน และแบบข้างตัด และใช้วิธีการให้คะแนน 3 วิธี คือ ให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ได้ทำ ให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทำถูก และให้คะแนนทุกข้อ โดยทำการศึกษากับนักเรียนชั้นมัธยมศึกษาปีที่ 2 จำนวน 348 คน ผลการศึกษาพบว่า

ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนที่ได้จากการทดสอบเทเลอรัปปีรามีคทั้ง 3 แบบ โดยให้คะแนนทั้ง 3 วิธี กับคะแนนจากการทดสอบแบบดั้งเดิม มีความสัมพันธ์กันในทางบวกอย่างมีนัยสำคัญทางสถิติทุกค่า

เมื่อพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ภายใน ระหว่างคะแนนที่ได้จากการทดสอบเทเลอรัปปีรามีคทั้ง 3 แบบ โดยให้คะแนนทั้ง 3 วิธี ปรากฏว่า มีค่าสหสัมพันธ์เป็นบวกอย่างมีนัยสำคัญทางสถิติทุกค่า

ต่อมาในปี พ.ศ. 2532 เกรียงศักดิ์ สุวรรณภาค (2532) ได้ศึกษาความสัมพันธ์ระหว่างรูปแบบของแบบทดสอบแยกกลุ่ม กับ ความสามารถทางการเรียน ในวิชาคณิตศาสตร์ โดยทำการศึกษานักเรียน 2 ตัว คือ ตัวแปรตัวแรก รูปแบบของแบบทดสอบแยกกลุ่มซึ่งมี 3 แบบ คือ แบบที่มีค่าความยากของข้อสอบใกล้เคียงกัน แบบที่มีค่าความยากของข้อสอบกระจายจากน้อยที่สุดถึงมากที่สุด และแบบที่มีการแยกกลุ่มสองครั้ง และ ตัวแปรที่สอง คือ วิธีการให้คะแนน ซึ่งมี 2 วิธี คือ ให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำถูก และ ให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทำถูก โดยทำการศึกษากับนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 2,455 คน ผลการศึกษาพบว่า

ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างผลการสอบที่ใช้แบบทดสอบแยกกลุ่มแต่ละฉบับ ซึ่งตรวจให้คะแนนในแต่ละวิธี กับความสามารถทางการเรียน มีค่าเป็นบวกอย่างมีนัยสำคัญทางสถิติทุกค่า และผลการทดสอบความแตกต่างของค่าสหสัมพันธ์ดังกล่าว เป็นรายคู่ พบว่า แตกต่างกันอย่างมีนัยสำคัญทางสถิติทุกคู่

ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนที่ได้จากการตรวจให้คะแนนทั้ง 2 วิธี ดังกล่าว โดยแยกวิเคราะห์ตามแบบทดสอบแยกกลุ่มแต่ละแบบ มีค่าเป็นบวกอย่างมีนัยสำคัญทางสถิติทุกค่า และผลการทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ดังกล่าว พบว่า เมื่อพิจารณาในแบบทดสอบแยกกลุ่มแต่ละแบบ ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนที่ได้จากการตรวจให้คะแนนทั้ง 2 วิธี มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติทุกคู่

ส่วน เสรี ประมวลลิขโรจน์ (2531) ได้ศึกษาความเที่ยงตรงในการจำแนกความรอบรู้ และการประมาณค่าความสามารถ โดยใช้แบบทดสอบเพลิกซีเลเวลจำนวน 2 ฉบับ และแบบทดสอบ รูปปริมาตรขนาดชั้นแปรผันจำนวน 2 ฉบับ โดยทำการศึกษากับนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 5,987 คน ผลการศึกษาพบว่า

ความเที่ยงตรงในการจำแนกความรอบรู้ เมื่อใช้ข้อสอบที่สร้างขึ้นเป็นเกณฑ์ ปรากฏว่า ข้อสอบแบบเพลิกซีเลเวล และรูปปริมาตรขนาดชั้นแปรผันทุกฉบับ มีความเที่ยงตรงในการจำแนกความรอบรู้ได้อย่างมีนัยสำคัญทางสถิติ และเมื่อเปรียบเทียบค่าความเที่ยงตรงแต่ละแบบ พบว่า แบบทดสอบแบบเพลิกซีเลเวลทั้งสองฉบับ มีความเที่ยงตรงในการจำแนกความรอบรู้ได้สูงกว่าแบบทดสอบรูปปริมาตรอยู่หนึ่งฉบับ อย่างมีนัยสำคัญทางสถิติ

ในปี พ.ศ. 2531 นันทิยา พึ่งคำ (2531) ได้ศึกษาเปรียบเทียบคุณภาพของการทดสอบแบบ ซีเอที (CAT) กับแบบดั้งเดิม (Conventional) ในการวัดความสามารถด้านคำศัพท์ภาษาอังกฤษ โดยทำการศึกษากับนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 120 คน การทดสอบแบบ ซีเอที ใช้กลุ่มข้อสอบ (Item Pool) จำนวน 361 ข้อ โดยมีค่าความยากอยู่ระหว่าง $-.90552$ ถึง 3.61860 มีค่าอำนาจจำแนกอยู่ระหว่าง 0.80334 ถึง 2.0000 และมีค่าพารามิเตอร์

ระหว่าง 0.04421 ถึง 0.29886 การประมาณค่าความสามารถของผู้สอบโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (Conditional Maximum Likelihood) ส่วนการทดสอบแบบดั้งเดิมใช้ข้อสอบทั้งหมด 40 ข้อ โดยมีค่าความยากอยู่ระหว่าง -0.76488 ถึง 2.97884 มีค่าอำนาจจำแนกอยู่ระหว่าง 0.94069 ถึง 2.0000 และมีค่าการเดาอยู่ระหว่าง 0.03462 ถึง 0.28620 ผลการศึกษาพบว่า

(1) การทดสอบแบบซีเอที กับแบบดั้งเดิม มีความตรงเชิงเกสส์สัมพันธ์ ไม่แตกต่างกัน แต่การทดสอบแบบ ซีเอที สามารถลดจำนวนข้อสอบลงได้มากกว่าครึ่งหนึ่ง

(2) ความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ จากการทดสอบแบบ ซีเอที (CAT) มีค่าต่ำกว่า ความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบจากการทดสอบแบบดั้งเดิม อย่างมีนัยสำคัญทางสถิติ

2.2.4.2 ผลงานวิจัยที่เกี่ยวข้องต่างประเทศ

วูด (Larkin and Weiss. 1975; citing Wood. 1969) ได้ศึกษาการทดสอบเทเลอ์รูปบิรามิดที่มี 4, 5, และ 6 ชั้น โดยใช้กฎการแยกทางแบบที่ถูกเพิ่มหนึ่งชั้นและทวนลดหนึ่งชั้น ได้ใช้ขนาดชั้นคงที่ ซึ่งเท่ากับ 0.05 และ ให้คะแนนตามจำนวนข้อที่ถูก ทว่าการทดสอบกับกลุ่มตัวอย่าง จำนวน 91 คน ผลการศึกษาพบว่า

ความสัมพันธ์ระหว่างคะแนนรวมทั้ง 3 ฉบับที่ได้จากการทดสอบเทเลอ์รูปบิรามิด กับเกรดเฉลี่ยในชั้นเรียนเท่ากับ .51 ในขณะที่เมื่อใช้การทดสอบแบบเดิมจำนวน 46 ข้อ คะแนนมีความสัมพันธ์กับเกรดเฉลี่ยเท่ากับ .68

ลอร์ด (Lord. 1971: 805-813) ได้ใช้คอมพิวเตอร์จำลองสถานการณ์ เพื่อศึกษาประเมินคุณภาพการทดสอบเทเลอ์แบบเพิลิกซีเลเวล โดยใช้ข้อสอบเพิลิกซีเลเวลจำนวน 119 ข้อ ซึ่งผู้สอบแต่ละคนจะต้องตอบข้อสอบแบบเพิลิกซีเลเวล จำนวน 60 ข้อ ผลการศึกษาพบว่า

เมื่อขนาดของช่วงห่างความยาก (d) ของข้อสอบมากขึ้นจะทำให้ค่าอินฟอร์เมชันลดลงทุกระดับความสามารถ และจะมีประสิทธิภาพสูงสุดเมื่อมีค่าอยู่ระหว่าง 0.033 ถึง 0.067 แบบ

ทดสอบจะ ไม่มีประสิทธิภาพเมื่อใช้ช่วงห่างความยาก (d) ใกล้เคียงกับ 0.1

เมื่อเปรียบเทียบกับ การทดสอบแบบเดิม การทดสอบแบบเพลิกซีเลเวล จะ ให้ค่าอินฟอร์-
เมชันสูงกว่าการทดสอบแบบเดิม เมื่อผู้สอบมีความสามารถสูงหรือต่ำ ส่วนผู้สอบที่มีความสามารถ
กลาง ๆ การทดสอบแบบเดิมจะให้ค่าอินฟอร์เมชันสูงกว่าการทดสอบแบบเพลิกซีเลเวล

แบล็คมอร์ (Blackmore. 1987: 2554-A) ได้ศึกษาเปรียบเทียบการดำเนินการสอบ
3 วิธี คือใช้คอมพิวเตอร์ดำเนินการสอบ (Computerized) ใช้คอมพิวเตอร์ปรับความยากให้
เหมาะสมกับระดับความสามารถของผู้สอบ (Computerized Adaptive Testing) และการ
ดำเนินการสอบแบบดั้งเดิม คือ ใช้กระดาษกับดินสอ (Paper and Pencil) ได้ทำการศึกษ
กับนักเรียนเกรดสิบสองที่กำลังจะจบการศึกษาในปี ค.ศ. 1986 ของโรงเรียนประจำตำบล
การ์เดนโกรฟ (Garden Grove High School) ในรัฐแคลิฟอร์เนีย จำนวน 24 คน โดยแบ่ง
ออกเป็น 3 กลุ่ม ๆ ละ 8 คน เพื่อดำเนินการทดสอบด้วยวิธีการทดสอบ 3 วิธีดังกล่าว วิชาที่
ใช้ในการทดสอบคือความถนัดทางการเรียน จำนวน 6 ฉบับย่อย คือ การให้เหตุผลทางภาษา
(Verbal Reasoning) ความสามารถทางตัวเลข (Numerical Ability) การให้เหตุผล
เชิงนามธรรม (Abstract Reasoning) การให้เหตุผลเชิงกล (Mechanical Reasoning)
มิติสัมพันธ์ (Space Relations) และการใช้ภาษา (Language Usage) ผลการศึกษาพบว่า
คะแนนจากข้อสอบย่อยทุกฉบับจากการดำเนินการสอบทั้ง 3 วิธี ไม่แตกต่างกัน ส่วน
ทัศนคติและความคิดเห็นต่าง ๆ เกี่ยวกับการดำเนินการสอบทั้ง 3 วิธี จากการใช้แบบสอบถาม
และสัมภาษณ์ปรากฏว่า แตกต่างกันอย่างมีนัยสำคัญทางสถิติ

แฮนคินส์ (Hankins. 1987: 3013-A) ได้ศึกษาการทดสอบเทเลอร์โดยใช้ยุทธวิธี
ของ เบย์ส์ (Bayesian Adaptive Test) 2 วิธี คือ แบบกำหนดความสามารถเริ่มต้นของ
ผู้สอบคงที่ (Fixed Entry) กับแบบกำหนดความสามารถเริ่มต้นของผู้สอบแปรผัน (Variable
Entry) ผลการศึกษาพบว่าทั้ง 2 วิธี ไม่มีความลำเอียง (Bias) เมื่อประมาณค่าความสามารถ
ของผู้สอบที่มีความสามารถปานกลาง แต่จะมีความลำเอียง เมื่อผู้สอบมีความสามารถค่อนข้าง

เก่ง หรืออ่อนมาก ๆ ทั้ง 2 วิธี ให้ค่าอินฟอร์เมชัน (Information) มีการแจกแจงเป็นรูปสมมาตร (Symmetry) ที่มีโค้งค่อนข้างโค้ง และการทดสอบแบบกำหนดความสามารถเริ่มต้นของผู้สอบแปรผัน (Variable Entry) จะต้องใช้จำนวนข้อสอบมากกว่าแบบกำหนดความสามารถเริ่มต้นของผู้สอบคงที่ (Fixed Entry)

ไฮ (Ho. 1989: 421-A) ได้ใช้ไมโครคอมพิวเตอร์ศึกษาเปรียบเทียบยุทธวิธีการใช้อัดแดปทีฟ เทสต์ติ้ง (Adaptive Testing) 3 วิธี คือ ยุทธวิธีของเบส์ (Bayesian) โมดัลเบส์ (Modal Bayesian) และวิธีแมกซ์ลิคไลฮูด (Maximum Likelihood) ยุทธวิธีทั้ง 3 นี้ได้เปรียบเทียบโดยใช้ขนาดของคลังข้อสอบ (Bank Sizes) 3 ขนาด คือ ขนาด 86 ข้อ, 71 ข้อ และ 56 ข้อ ชนิดของคลังข้อสอบ (Bank Type) 3 แบบ คือ แบบที่มีข้อสอบแบบสุ่ม, มีเฉพาะข้อสอบที่มีค่าอำนาจจำแนกสูง และแบบที่มีเฉพาะข้อสอบที่ง่าย ๆ ส่วน ระดับความสามารถของผู้สอบ มี 3 ระดับ คือ สูง, ปานกลาง และต่ำ การวิเคราะห์ข้อมูลโดยใช้การวิเคราะห์ความแปรปรวน 3 ทาง (Three ways ANOVA with repeated measures) ผลการศึกษาพบว่า

(1) ยุทธวิธีของโมดัลเบส์ (Modal Bayesian) จะมีประสิทธิภาพ (Efficient) สูงสุด ยุทธวิธีของเบส์ (Bayesian) จะให้ค่าความเชื่อมั่น (Reliable) สูงสุด และยุทธวิธีแมกซ์ลิคไลฮูด (Maximum Likelihood) จะให้ค่าไม่คงที่ (Inconsistent) ภายใต้การทดสอบในสถานการณ์ต่าง ๆ

(2) การเลือกขนาดคลังข้อสอบ ดูเหมือนว่าจะส่งผลน้อยมาก ยกเว้นในกลุ่มผู้สอบที่มีความสามารถต่ำ เมื่อใช้ข้อสอบเริ่มต้นที่มีความยากสูงกว่าความสามารถที่แท้จริงของผู้สอบ

(3) ถ้าใช้ข้อสอบเริ่มต้นที่มีความยากเท่ากับ หรือน้อยกว่าความสามารถที่แท้จริงของผู้สอบแล้ว ความถูกต้องในการประมาณค่าจะมีมากขึ้น

(4) มีปฏิสัมพันธ์ (Interaction) ระหว่างชนิดของคลังข้อสอบ กับ ยุทธวิธีของอัดแดปทีฟ เทสต์ติ้ง และระหว่างชนิดของคลังข้อสอบกับระดับความสามารถของผู้สอบ

(5) ความแตกต่างระหว่างคุณลักษณะทางสถิติของการเลือกข้อสอบแบบไม่เลือกข้อสอบ มีความสัมพันธ์กันค่า

วิสโนวสกี (Wisniewski. 1986: 159-A) ได้ตรวจสอบความถูกต้องแม่นยำในการประมาณค่าความสามารถของผู้สอบ จากการประยุกต์ใช้รหัสโมเดลกับการทดสอบแบบคิฟเทสตั้งแต่แบบ ไบนารีเซิร์ชเมธอด (Binary Search Method: BSM) การวิเคราะห์ข้อมูล ใช้การวิเคราะห์การถดถอยพหุคูณ (Multiple Regression) โดยใช้คะแนนจากการทดสอบบีเอสเอ็ม (BSM) เป็นตัวทำนายคะแนนการทดสอบแบบดั้งเดิมที่ใช้กระดาษดินสอ (Paper-Pencil Testing) ผลการศึกษาพบว่า คะแนนจากการทดสอบบีเอสเอ็ม (BSM) สามารถทำนายคะแนนจากการทดสอบแบบดั้งเดิมที่ใช้กระดาษ-ดินสอได้ดี นอกจากนี้การทดสอบแบบบีเอสเอ็ม (BSM) ยังสามารถที่จะควบคุมเวลาในการตอบเป็นรายข้อ และสามารถควบคุมสภาพแวดล้อมที่เป็นตัวแปรแทรกซ้อนได้ดีกว่าการทดสอบแบบดั้งเดิม

ลี (Lee. 1987: 3742-A) ได้พัฒนาการทดสอบแบบ เอ็มเอเอที (Microcomputerized Adaptive Achievement Test: MAAT) โดยใช้ยุทธวิธี 2 แบบ คือ แบบ เอสพีอาร์ที ของวอลท์ (Wald's Sequential Probability Ratio Test: SPRT) และยุทธวิธีของลอร์ด แบบ เฟล็กซ์เลเวล (Lord's Flexilevel Test) โดยใช้ข้อสอบที่มีอยู่ในคลังข้อสอบพีอีเอ (The Pitt Educational Testing Aids: PETA) ได้ทดลองศึกษากับนักศึกษาระดับบัณฑิตศึกษา จำนวน 3 ห้องเรียน ในรายวิชาการวิจัยการศึกษาเบื้องต้น ผลการศึกษาพบว่า

(1) มีความเป็นไปได้สูงในอนาคตที่จะนำเอาการทดสอบแบบเอ็มเอเอที (MAAT) มาใช้ในการทดสอบผลสัมฤทธิ์ทางการเรียน ในระดับบัณฑิตศึกษา ผู้ดำเนินการสอบไม่จำเป็นต้องมีความรู้ทางด้านจิตวิทยาหรือคอมพิวเตอร์มากนักก็สามารถดำเนินการสอบได้ เพียงแต่ดำเนินการตามขั้นตอนต่าง ๆ ตามคู่มือครูที่เขียนไว้อย่างละเอียด

(2) คะแนนจากการทดสอบแบบเอ็มเอเอที (MAAT) มีความสัมพันธ์ทางบวกกับคะแนนผลการสอบกลาง เทอม

(3) นักศึกษามีทัศนคติที่ดีต่อการทดสอบแบบเอ็มเอเอที (MAAT) ในทางบวก

จากผลงานวิจัยที่เกี่ยวข้องที่กล่าวมาแล้ว ทั้งในประเทศและต่างประเทศ ส่วนใหญ่พบว่า

การทดสอบเทเลอร์มีคุณภาพ ไม่แตกต่างจากการทดสอบแบบดั้งเดิม แต่ใช้จำนวนข้อสอบน้อยกว่า และผู้สอบส่วนใหญ่มีความพอใจกับการทดสอบเทเลอร์ และมีความเป็นไปได้ที่จะนำเอาการทดสอบเทเลอร์ หรือ ซีเอที (CAT) มาใช้แทนการทดสอบแบบดั้งเดิมได้

3. การศึกษาโดยใช้เทคนิควิธีมอนติ คาร์โล (Monte Carlo Method)

3.1 ความเป็นมาของวิธีมอนติ คาร์โล

เทคนิควิธี มอนติ คาร์โล เป็นสาขาหนึ่งของคณิตศาสตร์ ที่ใช้คอมพิวเตอร์ช่วยในการจำลองสถานการณ์ (Simulation) โดยอาศัยตัวเลขสุ่ม (Random Number) มาสร้างตัวแปรให้เหมือนกับสถานการณ์จริง และมีการทดลองซ้ำหลาย ๆ ครั้ง เพื่อให้ได้ค่าที่แน่นอนที่จะใช้เป็นข้อสรุปหรืออธิบายปรากฏการณ์ต่าง ๆ ในสถานการณ์จริง (Kerlinger. 1988: 192; Krobe and Forge. 1980: 242; Taha. 1988: 13) หรือ ช่วยหาคำตอบในเรื่องราวต่าง ๆ ที่ยังไม่แน่ใจในผลที่จะเกิดขึ้น (Hammersley and Handscomb. 1964: 2) เทคนิควิธี มอนติ คาร์โล มีการนำมาใช้นานแล้ว โดยในราวคริสต์ศตวรรษที่ 17 มีการนำเอาวิธีมอนติ คาร์โล ซึ่งในสมัยนั้นยังไม่ได้เรียกว่ามอนติ คาร์โล มาพัฒนาทฤษฎีความน่าจะเป็น (Probability Theory) (Barton. 1970: 137) ในราวปี ค.ศ. 1733 จอร์ส หลุยส์ เลคเลอว์ และบุฟฟง (Georges Louis Leclerc and Comte de Buffon) ได้ทำการทดลองหาค่าพาย (π) โดยการโยนเข็มที่มีความยาวขนาด k หน่วย อย่างสุ่มลงมาบนพื้นราบที่มีเส้นขนานอยู่ โดยให้ระยะห่างระหว่างเส้นขนานแต่ละเส้นห่างกัน d หน่วย และกำหนดให้ $d > k$ จะได้ความน่าจะเป็น (Probability: P) ที่เข็มจะตัดกับเส้นขนาน $P = 2k/\pi d$ ซึ่งถ้าความน่าจะเป็น (P) เป็นค่าสุ่ม ก็จะหาค่า π ได้ (Mihram. 1972: 186; Rubinstein. 1981: 11; บัณฑิต พนเจริญสวัสดิ์. 2527: 700; เลิศลักษณ์ กลิ่นหอม. 2532: 78) ในราว ปี ค.ศ. 1908 กอสเซต (W.S. Gosset) ได้ศึกษา การแจกแจงความถี่ของความสูงของนักโทษอาชญากรรมจำนวน 3,000 คน โดยเทียบกับการแจกแจงความถี่ของกลุ่มตัวอย่างที่สุ่มมาครั้งละ 4 คน จำนวน 750 กลุ่มตัวอย่าง ผลการศึกษาพบว่า การแจกแจงความถี่ทั้งสองมีลักษณะเหมือนกัน กอสเซต ได้ตั้งชื่อการแจกแจงความถี่ที่ค้นพบนี้ว่า การแจกแจงค่าที (t-distribution) (Meyer. 1956: 6-7; Spence.

1983: 406) ทั้งหมดนี้ถือว่าเป็นจุดเริ่มต้นของการศึกษาที่ใช้เทคนิควิธีมอนติ คาร์โล (Monte Carlo Method)

สำหรับการพัฒนาระเบียบวิธีการ มอนติ คาร์โล อย่างจริงจังนั้น ในราวปี ค.ศ. 1944 ในช่วงระหว่างสงครามโลกครั้งที่สอง อุลาม และ วอน นิวแมน (Ulam and Von Neumann) เป็นผู้ตั้งชื่อ มอนติ คาร์โล ซึ่งเป็นชื่อรหัสลับของงานที่ทำใน ลอส อลามอส (Los Alamos) ซึ่งงานนี้เป็นงานที่เกี่ยวข้องกับการสร้างระเบิดปรมาณูในโครงการแมนฮัตตัน (Manhattan Project) อุลาม (Ulam) ได้นำเทคนิควิธีมอนติ คาร์โล มาหาผลของการแพร่อย่างสุ่มของนิวตรอน (Neutron diffusion) ในวัสดุเชื้อเพลิง ซึ่งเป็นการทดลองทางคณิตศาสตร์ เพื่อหาผลของคำตอบก่อนที่จะทำการทดลองจริง ซึ่งเป็นการลดอันตราย และช่วยประหยัดค่าใช้จ่ายก่อนการทดลองจริง (Spence.1983: 405; สมชาย ยืนนาน. 2528: 29) หลังจากนั้นก็ได้มีการนำเอาวิธีมอนติ คาร์โล มาใช้อย่างกว้างขวางทั้งทางด้านฟิสิกส์ คณิตศาสตร์ สถิติ และการวิจัยเทคนิควิธีมอนติ คาร์โล (Monte Carlo Method) นับว่ามีประโยชน์อย่างมากในการขยายความรู้ในเชิงทฤษฎี เช่น การนำมาศึกษาค่าความคลาดเคลื่อนทางสถิติ เปรียบเทียบประสิทธิภาพของการทดสอบต่าง ๆ เป็นต้น

3.2 ขั้นตอนของระเบียบวิธีมอนติ คาร์โล

หลักการสำคัญของวิธีมอนติ คาร์โล (Monte Carlo Method) ก็คือ การนำเอาตัวเลขสุ่ม (Random Number) มาประยุกต์ใช้ในการแก้ปัญหาต่าง ๆ ซึ่งมีขั้นตอนที่สำคัญ ดังนี้

3.2.1 สร้างตัวเลขสุ่ม (Generate random number) ในระยะแรก ๆ กระทำโดยการอาศัยเครื่องมือทางกายภาพ เช่น ล้อรูเล็ต ลูกเต๋า ไพ่ กระดาษเขียนเบอร์ เป็นต้น เครื่องมือดังกล่าวใช้ได้เมื่อต้องการใช้ตัวเลขสุ่มจำนวนไม่มากนัก ต่อมาเมื่อมีความต้องการใช้ตัวเลขสุ่มจำนวนมาก ๆ ก็มีการหันมาใช้เครื่องมืออิเล็กทรอนิกส์ ที่ใช้แพร่หลายที่สุดคือ เครื่องสร้างตัวเลขสุ่มที่สร้างขึ้นโดยบริษัทแรนด์ (Rand) ที่สร้างตัวเลขสุ่มจากเครื่องกำเนิดพัลส์ (Pulse) เครื่องดังกล่าวสามารถสร้างตัวเลขสุ่มได้เป็นล้านตัว

การสร้างหรือเลือกใช้ตัวเลขสุ่มดังกล่าวกับเครื่องคอมพิวเตอร์ยังมีปัญหา 2 ประการ

คือ เป็นการยากที่จะทำให้คอมพิวเตอร์สามารถเรียกใช้ได้เมื่อมีความต้องการ และเป็นการยากที่จะทำให้เครื่องมืองดกล่าว สร้างตัวเลขสุ่มชุดเดิม เมื่อต้องการใช้เปรียบเทียบวิธีการต่าง ๆ ภายใต้งานของระบบเลขสุ่มชุดเดียวกัน หรือถ้าจะเก็บเลขสุ่มเหล่านี้ไว้ในหน่วยความจำ หรือจานแม่เหล็ก ก็จะทำให้สูญเสียหน่วยความจำหรือเสียเวลาในการค้นหา ฉะนั้นการสร้างตัวเลขสุ่มในคอมพิวเตอร์ จึงนิยมสร้างตัวเลขสุ่มเทียม (Pseudo Random number) โดยอาศัยสูตรทางคณิตศาสตร์ วิธีที่ใช้กันมากมี 2 วิธี คือ

วิธีที่ 1 โดยวิธีส่วนกลางกำลังสอง (Midsquare Method) ซึ่งเป็นวิธีที่ใช้ในยุคแรก ๆ ของการสร้างตัวเลขสุ่มเทียม โดยมีขั้นตอนในการสร้าง ดังนี้ (ศิริจันทร์ ทองประเสริฐ. 2532: 87-88)

- (1) เลือกตัวเลขขึ้นมาสี่หลัก
- (2) ยกกำลังสองของตัวเลขนั้น ถ้าตัวเลขที่ได้ไม่ครบแปดหลัก ให้เติมศูนย์ข้างหน้าให้ครบแปดหลัก
- (3) ใช้เลขสี่หลักกลางที่ได้ในขั้นที่ (2) เป็นตัวเลขแบบสุ่ม
- (4) ยกกำลังสองของตัวเลขในข้อ (3)
- (5) ทำซ้ำในขั้นที่ (3) และ (4) จนได้จำนวนตัวเลขสุ่มตามต้องการ

ตัวอย่าง สมมุติว่าเลือกตัวเลขสุ่มตัวแรกเป็น $X_0 = 1376$

$$X_0 = 1376 \text{ จะได้ } X_0^2 = 01893376$$

$$X_1 = 8933 \text{ จะได้ } X_1^2 = 79798489$$

$$X_2 = 7984 \text{ จะได้ } X_2^2 = 63744256$$

$$X_3 = 7442 \text{ จะได้ } X_3^2 = 55383364$$

$$X_4 = 3833 \text{ จะได้ } X_4^2 = 14691889$$

$$X_5 = 6918 \text{ etc}$$

ฉะนั้นตัวเลขสุ่มที่ได้ คือ 1376, 8933, 7984, 7442, 3833, 6918 ...

การสร้างตัวเลขสุ่มโดยวิธีนี้ยังมีปัญหาสองประการ คือ ไม่ทราบช่วงของตัวเลขสุ่มที่จะหมุนกลับมาซ้ำเดิมอีก และถ้ากำหนดตัวเลขสุ่มตัวแรกไม่เหมาะสม จะทำให้ได้ตัวเลขสุ่มที่ไม่เป็น

แบบสุ่ม เช่น ถ้ากำหนดตัวเลขสุ่มตัวแรกเป็น $X_0 = 4500$

$$X_0 = 4500 \text{ จะได้ } X_0^2 = 20250000$$

$$X_1 = 2500 \text{ จะได้ } X_1^2 = 06250000$$

$$X_2 = 2500 \text{ จะได้ } X_2^2 = 06250000$$

$$X_3 = 2500 \text{ จะได้ } X_3^2 = 06250000$$

$$X_4 = 2500 \text{ etc.}$$

ฉะนั้นตัวเลขสุ่มที่ได้ คือ 4500, 2500, 2500, 2500, ... ซึ่งไม่เป็นตัวเลขแบบสุ่ม

วิธีที่ 2 โดยวิธีเศษเหลือ (Congruent Method) วิธีเศษเหลือที่นิยมใช้กันมากที่สุด คือ เศษเหลือของผลคูณ (Multiplication Congruential Method) ซึ่งมีสูตร ดังนี้ (ศิริจันทร์ ทองประเสริฐ. 2529: 88)

$$X_{i+1} = aX_i \pmod{m}$$

ตัวอย่าง สมมติว่าเลือก $a=2$, $m=10$ และ $X_0 = 1$

$$X_0 = 1$$

$$X_1 = 2 \times 1 \pmod{10} = 2$$

$$X_2 = 2 \times 2 \pmod{10} = 4$$

$$X_3 = 2 \times 4 \pmod{10} = 8$$

$$X_4 = 2 \times 8 \pmod{10} = 6$$

$$X_5 = 2 \times 6 \pmod{10} = 2$$

ฉะนั้นตัวเลขสุ่มที่ได้ คือ 1, 2, 4, 8, 6, 2, ... ซึ่งจะเห็นว่าเลขสุ่มเริ่มวนกลับมาเลขเดิมแล้ว เพื่อไม่ให้เลขสุ่มวนกลับมาเลขเดิมเร็วเกินไป ในทางปฏิบัติเมื่อใช้คอมพิวเตอร์สร้างตัวเลขสุ่มโดยวิธีนี้ จึงนิยมกระทำดังนี้ (ศิริจันทร์ ทองประเสริฐ. 2929: 89)

(1) เลือกตัวเลขสุ่มใด ๆ ที่น้อยกว่า 9 หลัก ใช้เป็น X_0 โดยปกติถ้าเป็นคอมพิวเตอร์

ฐานสอง x_0 จะให้เป็นเลขคี่ที่เป็นบวก และคอมพิวเตอร์ฐานสิบ x_0 จะให้เป็นเลขบวกที่หารด้วย 2 และ 5 ไม่ลงตัว

(2) เลขในข้อ (1) ด้วยค่า a ซึ่ง a นี้ ควรมีค่าไม่น้อยกว่า 5 หลัก โดยปกติค่า a จะหาได้จาก

$a = 8T \pm 3$ สำหรับคอมพิวเตอร์ฐานสอง

$a = 200T \pm Q$ สำหรับคอมพิวเตอร์ฐานสิบ

เมื่อ T คือ เลขใด ๆ ที่เป็นบวก

Q คือ ค่าหนึ่งค่าใดต่อไปนี้ $\pm(3, 11, 13, 19, 21, 27, 29, 37, 53, 59, 61, 67, 69, 77, 83, \text{ หรือ } 91)$

(3) คูณผลคูณในข้อ 2 ด้วย $1/m$ โดยปกติค่า m จะหาได้จาก

$m = 2^b$ สำหรับคอมพิวเตอร์ฐานสอง

$m = 10^d$ สำหรับคอมพิวเตอร์ฐานสิบ

เมื่อ b คือ จำนวน bites ใน 1 word

d คือ จำนวนหลักใน 1 word

(4) เลือกใช้ตัวเลขหลังทศนิยมของเลขที่ได้ในข้อ (3) เป็นตัวเลขสุ่มที่ต้องการ

(5) คัดจุดทศนิยมออกจากตัวเลขในข้อ (4) แล้วดำเนินการตามขั้นตอนที่ (2)

(6) ทำซ้ำในขั้นที่ (2) ถึง (5) จนกว่าจะได้จำนวนตัวเลขสุ่มมากเท่าที่ต้องการ

ในปัจจุบันนี้ได้มีโปรแกรมสำเร็จรูปสำหรับสร้างตัวเลขสุ่มในเครื่องคอมพิวเตอร์ เช่น ในภาษาเบสิก (BASIC) มีคำสั่งที่ใช้เรียกตัวเลขสุ่มได้ คือ RANDOMIZED และ RND ส่วนในภาษาฟอร์แทรน (FORTRAN) ก็มีคำสั่งเรียกตัวเลขสุ่ม คือ RANDOM เป็นต้น

ในปี ค.ศ. 1987 เจโนสกี (Janosky. 1987: 4367-A) ได้ศึกษาการใช้ไมโครคอมพิวเตอร์ในการสร้างตัวเลขสุ่มเทียม (Pseudo-Random Number) ที่มี การแจกแจงแบบยูนิฟอร์ม (Uniform distribution) 3 วิธี คือ วิธีของไอบีเอ็ม-แรนดุม (IBM-RANDU) วิธีของมอร์แกน (The Morgan. 1984) และวิธีของเลแมนกับไบเลย์ (The Leman and Bailey.

1968) และที่มี การแจกแจงแบบปกติ (Normal Distribution) 3 วิธี คือใช้ทฤษฎีเซ็นทรัล-ลิมิต (Central Limit Theorem) วิธีของบ็อกซ์และมุลเลอร์ (The Box and Muler method. 1958) และวิธีโพลาร์ (The Polar method, Marsaglia and Bray. 1964) พร้อมทั้งศึกษาปฏิสัมพันธ์ระหว่างการแจกแจงแบบยูนิฟอร์มกับการแจกแจงแบบปกติ ผลการศึกษาพบว่า

การสร้างตัวเลขสุ่มเทียมที่มีการแจกแจงแบบยูนิฟอร์มโดยใช้วิธีของ เลแมนกับไบเลย์ (The Leman and Bailey. 1968) ให้ผลดีที่สุด ส่วนการแจกแจงปกติวิธีที่ดีที่สุดเป็นการผสมกันระหว่างการใช้อยู่ทฤษฎีเซ็นทรัล ลิมิต (Central Limit Theory) กับ วิธีของเลแมนและไบเลย์ (The Leman and Bailey. 1968)

คุณสมบัติของตัวเลขสุ่มที่ดี

- (1) ตัวเลขสุ่มที่ได้จะต้องมีลักษณะการกระจายความน่าจะเป็นแบบสม่ำเสมอ (Uniform Distribution)
- (2) ตัวเลขสุ่มที่ได้ต้องเป็นอิสระแก่กัน
- (3) อนุกรมของตัวเลขสุ่มที่ได้ต้องสามารถสร้างซ้ำเดิมได้ (Reproducible)
- (4) อนุกรมของตัวเลขสุ่มที่ได้ต้องไม่ซ้ำเดิมในช่วงที่ต้องการใช้ตัวเลขสุ่ม
- (5) ต้องใช้เวลาน้อยในการสร้างตัวเลขสุ่ม
- (6) ต้องใช้หน่วยความจำในคอมพิวเตอร์น้อย

3.2.2 นวัตกรรมตัวเลขสุ่มมาประยุกต์ใช้กับปัญหาต่าง ๆ เป็นขั้นตอนของการนำตัวเลขสุ่มไปสร้างตัวแปรตามลักษณะการแจกแจงของปัญหาที่จะศึกษา เพื่อเป็นข้อมูลของปัญหานั้น (Swarup and Mohan. 1982: 666-674) เช่น สร้างตัวเลขสุ่มขึ้นมาจำนวนหนึ่ง แล้วนำเลขสุ่มนั้นไปสร้างเป็นคะแนนผลการสอบของผู้เรียน บางครั้งตัวแปรของปัญหาที่จะศึกษา ไม่ได้สร้างจากตัวเลขสุ่มโดยตรง แต่มีขั้นตอนที่ต้องอาศัยตัวเลขสุ่มเป็นพื้นฐานก็ได้

3.2.3 ท้าการทดลองซ้ำหลาย ๆ ครั้ง ต้องมีการการทดลองซ้ำหลาย ๆ ครั้ง (Krober, 1980: 242) เพื่อลดความคลาดเคลื่อนของคำตอบที่จะได้ (Hammersley and Handscomb, 1964: 5, 21) และ สามารถสรุปเป็นความน่าจะเป็นของการเกิดเหตุการณ์ในปัญหานั้น ๆ

3.3 จุดเด่นของการใช้เทคนิควิธีมอนติ คาร์โล

เนื่องจากเทคนิควิธี มอนติ คาร์โล จะใช้ตัวเลขสุ่มเป็นพื้นฐานในการสร้างตัวแปรของปัญหาโดยอาศัยทฤษฎี สูตร หรือกฎเกณฑ์ต่าง ๆ ที่มีอยู่ และมีการทดลองซ้ำหลาย ๆ ครั้ง เพื่อลดความคลาดเคลื่อนต่าง ๆ ซึ่งนับว่ามีประโยชน์ที่สาคัญ ดังนี้

(1) เทคนิควิธี มอนติ คาร์โล สามารถควบคุมตัวแปรแทรกซ้อนและสามารถสังเกตได้อย่างสมบูรณ์ (Completely Controlled and Completely Observed) (Payne, 1988: 3)

(2) นอกจากนี้ยังสามารถท้าทายได้สภาพแวดล้อม (Context) เดิมหลาย ๆ ครั้ง ได้ ส่วนในการทดลองจริงนั้นทำไม่ได้เพราะ ไม่สามารถรักษาสภาพแวดล้อมให้เหมือนเดิมทุกอย่างได้ เมื่อเวลาได้เปลี่ยนไป

(3) เทคนิควิธีมอนติ คาร์โล ถ้ามีทฤษฎี สูตร หรือกฎเกณฑ์ต่าง ๆ ที่ถูกต้องรองรับในการสร้างตัวแปรของปัญหาในการทดลองแล้ว จะให้ผลที่ได้ ถูกต้องแม่นยำกว่า เมื่อใช้ทดลองในสถานการณ์จริง ทั้งนี้เพราะสามารถลดตัวแปรแทรกซ้อนที่เป็นในเชิงจิตวิทยาได้

(4) ทำให้สิ้นเปลืองเวลา แรงงาน และค่าใช้จ่ายน้อยกว่าเมื่อเทียบกับการทดลองในสถานการณ์จริง

3.4 การนำเอาเทคนิควิธีมอนติ คาร์โล มาศึกษาเกี่ยวกับการทดสอบเทเลอร์

การทดสอบเทเลอร์ได้นำเอาทฤษฎีการตอบข้อสอบ (IRT) มาประยุกต์ใช้ ซึ่งจุดเด่นของทฤษฎีการตอบข้อสอบ คือ มีโมเดลทางคณิตศาสตร์ (Mathematical Model) ที่แน่นอนสำหรับแสดงความสัมพันธ์ระหว่างค่าพารามิเตอร์ต่าง ๆ ของข้อสอบและผู้สอบ ประกอบกับในระยะหลังนี้มีคอมพิวเตอร์ขนาดเล็กที่มีประสิทธิภาพสูงใช้กันอย่างแพร่หลาย จึงทำให้นักวิจัยนำเอา

เทคนิคอีมอนติ คาร์โล มาศึกษาเกี่ยวกับการทดสอบเทเลอร์ก็นมากขึ้น ทั้งนี้เพราะการศึกษาโดยใช้เทคนิคอีมอนติ คาร์โล ประหยัดแรงงาน ค่าใช้จ่าย และสามารถทดลองหลาย ๆ ครั้ง ได้ จึงจะเห็นได้จากงานวิจัยที่ใช้เทคนิคอีมอนติ คาร์โล กับการทดสอบเทเลอร์ ดังนี้

งานวิจัยที่ศึกษาเกี่ยวกับประสิทธิภาพในการใช้วิธีของเบส์ที่ โอเวน (Owen. 1969, 1975) ได้พัฒนาขึ้นมาเพื่อใช้กับการทดสอบเทเลอร์ ศึกษาโดยใช้อีมอนติ คาร์โล มีผลงานวิจัยหลายชิ้นด้วยกัน ดังนี้ เออร์รี่, เจนเซมา, แมกซ์เบิร์ก, และกอร์แมน (Weiss and McBride. 1984:274-275; citing Urry. 1971; Jensema: 1972; McBride. 1977; Gorman. 1980)

ไวส์และแมกซ์เบิร์ก (Weiss and McBride. 1984: 273-285) ได้ใช้อีมอนติ คาร์โล ศึกษาความคลาดเคลื่อน (Bias) และค่าอินฟอร์เมชัน (Information) ของการทดสอบโดยใช้ เบส์เซียนอแดปติฟ เทสติ้ง (Bayesian Adaptive Testing)

คิงส์เบอรีและไวส์ (Wiess and Kingsbury. 1984: 365; citing Kingsbury and Weiss. 1983) ได้ใช้อีมอนติ คาร์โล ศึกษาการทดสอบแบบปรับระดับการรอบรู้ (Adaptive Mastery Testing) เพื่อเปรียบเทียบกับวิธีการทดสอบความรอบรู้แบบดั้งเดิม (Conventional Mastery Testing)

ดอดและคนอื่น ๆ (Dodd and others: 1989: 129-143) ได้ใช้เทคนิค มอนติ คาร์โล ศึกษาลักษณะบางประการของขบวนการจัดการทดสอบเทเลอร์ โดยใช้เกรด เรสพอน โมเดล (Graded Response Model)

วีชเวอร์ (Vijver. 1986: 45-57) ได้ใช้อีมอนติ คาร์โล ศึกษาความแกร่ง (Robustness) ของการใช้รัศมีโมเดลในการประมาณค่าพารามิเตอร์ต่าง ๆ

จากตัวอย่างงานวิจัยที่ยกมาจะเห็นว่า เราสามารถนำเอาเทคนิคอีมอนติ คาร์โล มาตรวจสอบการทดสอบเทเลอร์ได้ในหลาย ๆ กรณี โดยเฉพาะอย่างยิ่งในบางกรณีที่เรามิสามารถใช้อ้อมลูในสถานการณ์จริงได้ แต่สามารถใช้เทคนิคอีมอนติ คาร์โล ตรวจสอบได้

บทที่ 3

วิธีดำเนินการวิจัย

การศึกษาเปรียบเทียบประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบ จากการทดสอบเทเลอร์รูปปราคที่มี รูปแบบ จำนวนชั้น และวิธีการให้คะแนน ที่แตกต่างกัน ในการวิจัยครั้งนี้ เป็นการวิจัยเชิงทดลอง โดยใช้เทคนิคมอนติ คาร์โล (Monte Carlo Method) คือ ใช้เครื่องคอมพิวเตอร์ในการสุ่มตัวอย่าง จำลองผลการตอบข้อสอบ และประมาณค่าความสามารถของผู้สอบจากการทดสอบเทเลอร์ ซึ่งมีขั้นตอนในการวิจัย ดังนี้

ประชากรและกลุ่มตัวอย่าง

ประชากรที่ใช้ในการทดลองครั้งนี้ ได้ใช้คะแนนจากการตรวจสอบคุณภาพทางการศึกษา วิชาภาษาไทย ปีการศึกษา 2531 ชั้นประถมศึกษาปีที่ 6 ทั่วประเทศ จำนวน 106,637 คน กลุ่มตัวอย่าง ได้มาจากการสุ่มจากประชากร มาจำนวน 32 กลุ่ม ๆ ละ 1,000 คน

ขั้นตอนในการทดลอง

การวิจัยครั้งนี้มีลำดับขั้นตอนในการทดลอง ดังนี้

1. จำลองกลุ่มตัวอย่างและความสามารถจริง (True-Ability: θ_1) ของตัวอย่าง
2. จำลองโครงสร้างของการทดสอบเทเลอร์รูปปราคตาม รูปแบบ และ จำนวนชั้น ความที่จะศึกษา พร้อมทั้งหาค่าความยากของข้อสอบแต่ละข้อ
3. หาค่าโอกาส (Probability) ในการตอบข้อสอบถูกต้องของกลุ่มตัวอย่างในข้อ 1. กับข้อสอบในข้อ 2.
4. หาผลการตอบข้อสอบ โดยนำเอาโอกาสในการตอบข้อสอบถูกต้องที่ได้ในข้อ 3. มาหาผลการตอบข้อสอบ คือ ตอบถูก=1 ตอบผิด=0
5. นำผลการตอบข้อสอบในข้อ 4. มาหาเส้นทางในการตอบข้อสอบการทดสอบเทเลอร์ ตามที่จำลองขึ้นในข้อ 2. และให้คะแนนตามวิธีการที่กำหนด ก็จะได้ค่าความสามารถของผู้สอบที่ประมาณค่าจากการทดสอบเทเลอร์ ($\hat{\theta}_1$)

รายละเอียดในแต่ละขั้นตอนของการทดลอง มีดังนี้

1. จัดกลุ่มตัวอย่างและความสามารถจริง (True-Ability: θ_1) ของตัวอย่าง
กระทำ ดังนี้

นำคะแนนจากการตรวจสอบคุณภาพทางการศึกษาทั่วประเทศ ปีการศึกษา 2531 วิชา
ภาษาไทย ชั้นประถมศึกษาปีที่ 6 จำนวน 106,637 คน มาแปลงเป็นคะแนนมาตรฐานซี (Z-Score)
แล้วสุ่มคะแนนมาตรฐานซี มาเป็นกลุ่มตัวอย่างจำนวน 32 กลุ่ม ๆ ละ 1,000 คน โดยใช้คะแนน
มาตรฐานซี (Z-Score) แทนความสามารถจริงของผู้สอบ (True-Ability: θ_1) ซึ่งปรากฏว่า
ในแต่ละกลุ่มตัวอย่างมีค่าความสามารถจริง ค่าต่ำสุด, สูงสุด, ค่าเฉลี่ย, และค่าส่วนเบี่ยงเบน
มาตรฐาน ดังตาราง 1

ตาราง 1 แสดงค่าต่ำสุด สูงสุด ค่าเฉลี่ย และค่าส่วนเบี่ยงเบนมาตรฐาน ของค่าความสามารถ
จริง (θ_1) ของกลุ่มตัวอย่าง

กลุ่มตัวอย่าง	ความสามารถจริง (θ_1)			
	ค่าต่ำสุด	ค่าสูงสุด	ค่าเฉลี่ย	ค่าส่วนเบี่ยงเบนมาตรฐาน
c7s1	-2.9252	2.6782	0.0799	0.8233
c10s1	-2.8170	2.8628	0.0481	0.9558
c13s1	-2.8170	2.7853	0.0484	0.9559
c16s1	-2.8170	2.7853	-0.0052	0.9835
c7s2	-2.9130	2.6782	0.0235	0.9229
c10s2	-2.8170	2.6782	-0.0368	0.8894
c13s2	-2.8170	2.6782	-0.0461	0.9739
c16s2	-2.8170	2.6782	-0.0086	0.9203
c7s3	-2.9130	2.7853	-0.0836	0.9434
c10s3	-2.8170	2.6782	0.0521	0.9374
c13s3	-2.8170	2.7853	-0.0464	0.9551

ตาราง 1 (ต่อ)

กลุ่มตัวอย่าง	ความสามารถจริง (θ_i)			
	ค่าต่ำสุด	ค่าสูงสุด	ค่าเฉลี่ย	ค่าส่วนเบี่ยงเบนมาตรฐาน
c16s3	-2.9130	2.6782	-0.0205	0.9802
c7s4	-2.8170	2.6782	-0.0115	0.9795
c10s4	-2.9130	2.6782	0.0084	0.9504
c13s4	-2.8170	2.6782	0.0766	0.9758
c16s4	-2.9130	2.6782	0.0685	0.8281
v7s1	-2.8170	2.6782	0.0864	0.9764
v10s1	-2.9130	2.6782	-0.0388	0.9881
v13s1	-2.8170	2.6782	0.0658	0.9604
v16s1	-2.9130	2.6782	-0.0766	1.0022
v7s2	-2.9723	2.7853	0.0282	0.9310
v10s2	-2.9130	2.6782	0.0139	0.9383
v13s2	-2.8170	2.8834	0.0158	0.9057
v16s2	-2.8170	2.6782	0.0316	0.8836
v7s3	-2.9130	2.8628	-0.0531	1.0210
v10s3	-2.9130	2.7853	-0.0382	0.9283
v13s3	-2.8170	2.7853	-0.0473	0.8739
v16s3	-2.8170	2.6782	-0.0139	0.9566
v7s4	-2.9130	2.6782	-0.0200	0.9315
v10s4	-2.8170	2.9394	0.0947	0.9333
v13s4	-2.9130	2.7853	-0.0717	0.9423
v16s4	-2.8170	2.7853	-0.0153	0.9517

2. จําลองโครงสร้างของการทดสอบเทเลอร์รูปปรีramidตาม รูปแบบ และจํานวนชั้น ความที่จะศึกษา พร้อมทั้งหาค่าความยากของข้อสอบแต่ละข้อ กระทั่งดังนี้

2.1 จําลองโครงสร้างของการทดสอบเทเลอร์รูปปรีramid ตาม รูปแบบ และจํานวนชั้น ความที่จะศึกษา

(1) รูปแบบของการทดสอบเทเลอร์รูปปรีramidที่จะศึกษาในครั้งนี้มี 2 รูปแบบ คือ แบบขนาดชั้นคงที่ (Constant Step Size Pyramid) และแบบขนาดชั้นแปรผัน (Variable Step Size Pyramid)

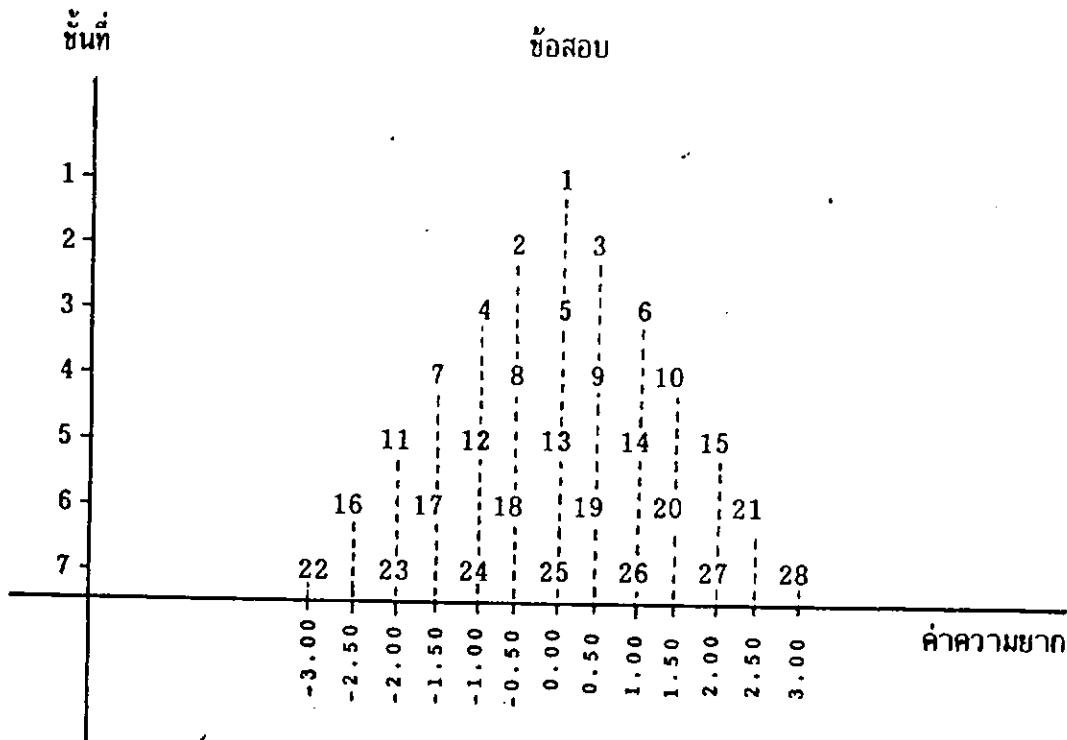
(2) จํานวนชั้นของรูปปรีramidที่จะศึกษาในครั้งนี้มี 4 แบบ คือ แบบ 7, 10, 13 และ 16 ชั้น

ฉะนั้นโครงสร้างของการทดสอบเทเลอร์ที่จําลองขึ้นจะมีทั้งหมด 8 แบบ ดังนี้

- (1) รูปปรีramidแบบขนาดชั้นคงที่ ที่มี 7 ชั้น (c7)
- (2) รูปปรีramidแบบขนาดชั้นคงที่ ที่มี 10 ชั้น (c10)
- (3) รูปปรีramidแบบขนาดชั้นคงที่ ที่มี 13 ชั้น (c13)
- (4) รูปปรีramidแบบขนาดชั้นคงที่ ที่มี 16 ชั้น (c16)
- (5) รูปปรีramidแบบขนาดชั้นแปรผัน ที่มี 7 ชั้น (v7)
- (6) รูปปรีramidแบบขนาดชั้นแปรผัน ที่มี 10 ชั้น (v10)
- (7) รูปปรีramidแบบขนาดชั้นแปรผัน ที่มี 13 ชั้น (v13)
- (8) รูปปรีramidแบบขนาดชั้นแปรผัน ที่มี 16 ชั้น (v16)

2.2 หาค่าความยากของข้อสอบแต่ละข้อ (β_j) ในการวิจัยครั้งนี้ ได้กำหนดค่าความยากของข้อสอบ ค่าสุด-สูงสุด ให้อยู่ในช่วง -3.000 ถึง $+3.000$ สำหรับการกำหนดค่าความยากของข้อสอบแต่ละข้อ จะเป็นไปตามรูปแบบ และจํานวนชั้นของการทดสอบเทเลอร์นั้น ๆ ดังตัวอย่าง การหาค่าความยากของข้อสอบการทดสอบเทเลอร์รูปปรีramidแบบขนาดชั้นคงที่ ที่มี 7 ดังแสดงใน ภาพประกอบ 13 กับ ตาราง 2 ส่วนค่าความยากของข้อสอบการทดสอบเทเลอร์รูปปรีramidแบบขนาดชั้นคงที่ ที่มี 10, 13 และ 16 ชั้นนั้น แสดงไว้ใน ภาพประกอบ 14 ถึง 16 และ ตัวอย่าง การหาค่าความยากของข้อสอบการทดสอบเทเลอร์รูปปรีramidแบบขนาดแปรผัน ที่มี

7 ชั้น ดังแสดงใน ภาพประกอบ 17 กับ ตาราง 3 ส่วนค่าความยากของข้อสอบการทดสอบ
 เทลอร์รูปปิรามิดแบบขนาดชั้นแปรรัน ที่มี 10, 13 และ 16 ชั้นนั้น ได้แสดงไว้ใน ตาราง 4
 ถึง 6



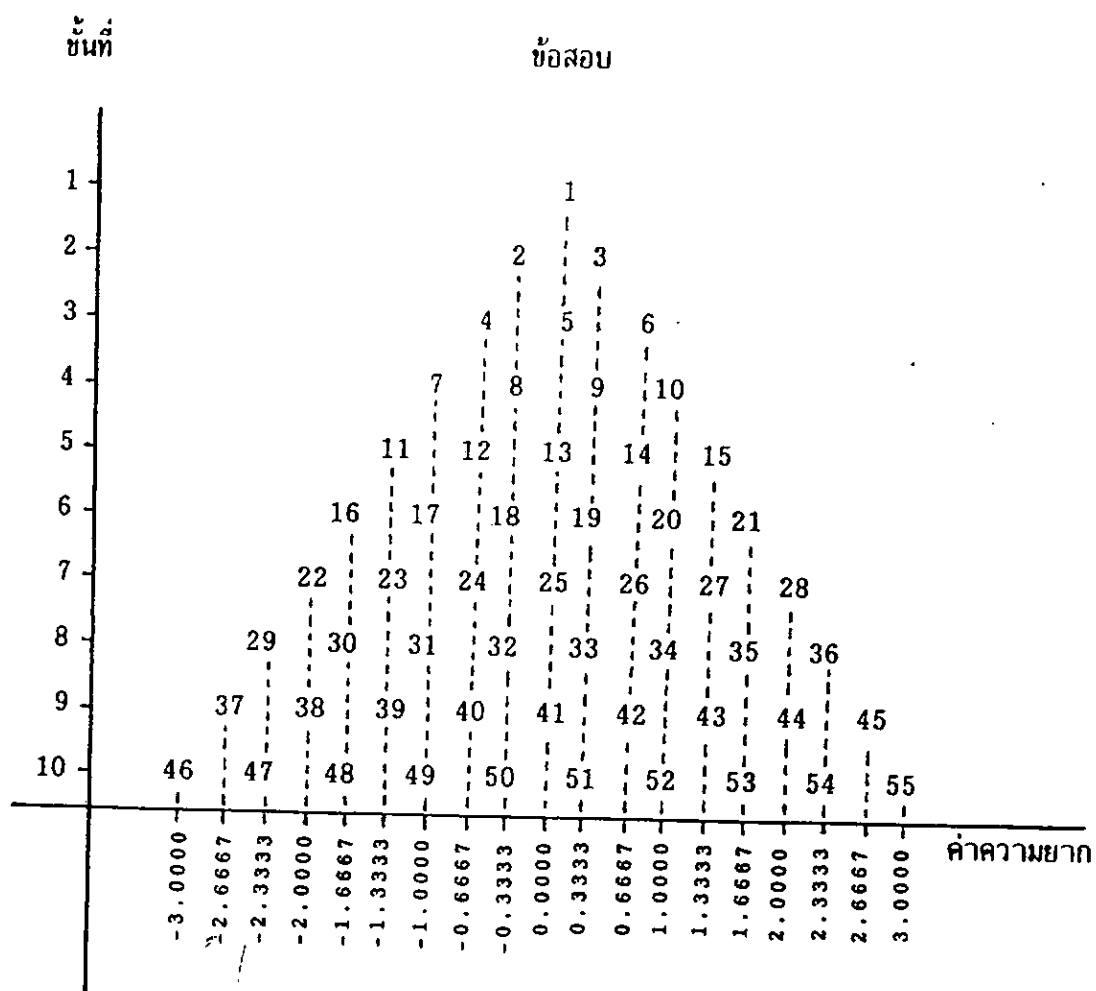
ภาพประกอบ 13 แสดงโครงสร้างและค่าความยากของข้อสอบการทดสอบเทลอร์รูปปิรามิด
 แบบขนาดชั้นคงที่ ที่มี 7 ชั้น

จากภาพประกอบ 13 แกนนอนแสดง ค่าความยากของข้อสอบที่สัมพันธ์กับข้อสอบในการ
 ทดสอบเทลอร์รูปปิรามิดแบบขนาดชั้นคงที่ ที่มี 7 ชั้น ซึ่งจะเห็นว่า ค่าความยากต่ำสุดกับสูงสุด
 อยู่ระหว่าง -3.00 ถึง +3.00 โดยแบ่งออกเป็น 12 ช่วงเท่า ๆ กัน จะได้ค่าความยากห่าง
 กันช่วงละ 0.50 และข้อสอบที่อยู่ในแนวดังเดียวกันจะมีค่าความยากเท่ากัน ฉะนั้นจะได้ค่าความ
 ยากของข้อสอบแต่ละข้อ ดังตาราง 2

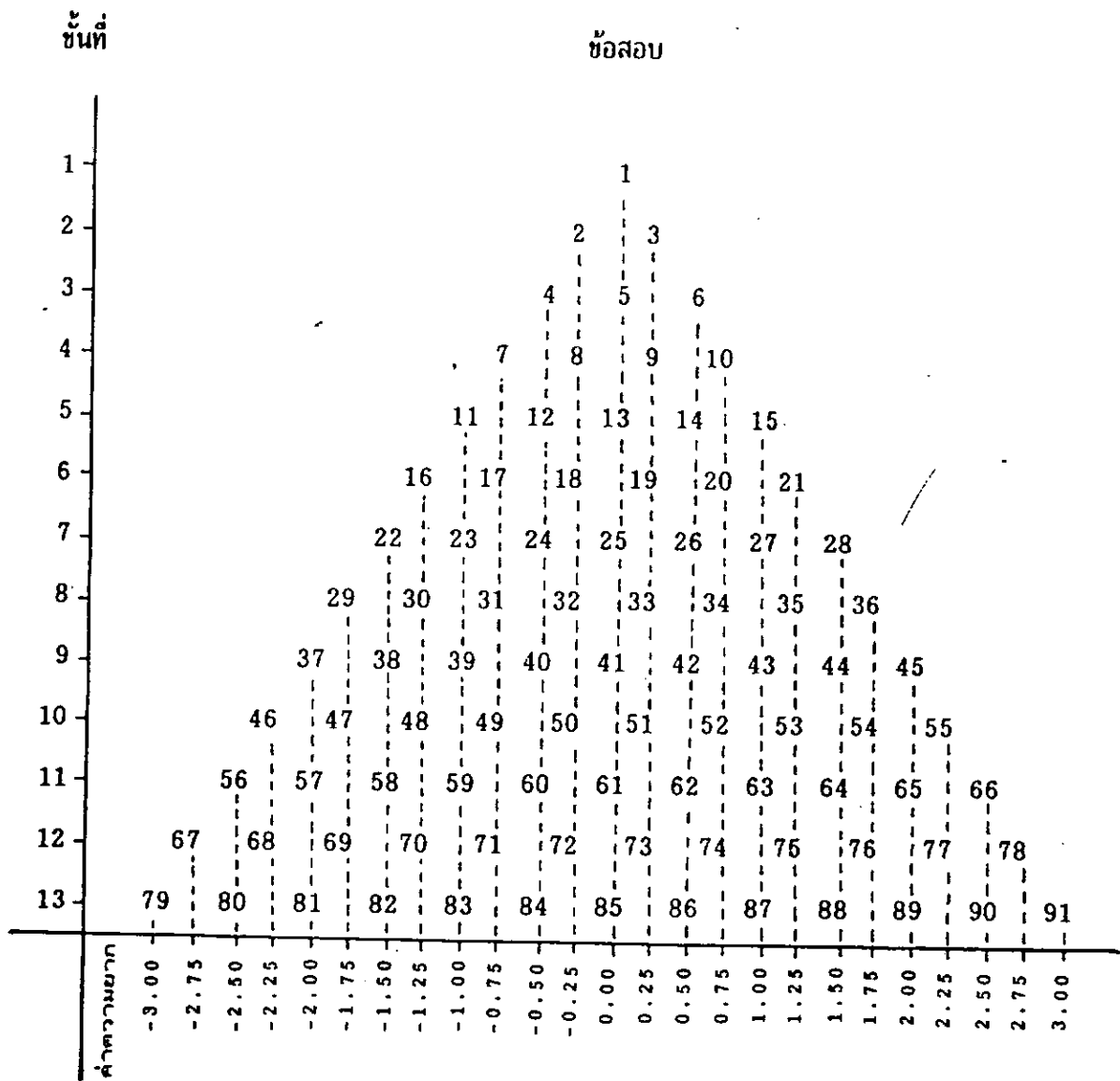
สำหรับค่าความยากของข้อสอบการทดสอบเทลอร์รูปปิรามิดแบบขนาดชั้นคงที่ ที่มี 10,
 13 และ 16 ชั้น ดังได้แสดงไว้ใน แผนภาพประกอบ 14 ถึง 16

ตาราง 2 แสดงค่าความยากเป็นรายข้อ (β_j) ของข้อสอบการทดสอบเทเลอ์รูปปิรามิดแบบ
ขนาดชั้นคงที่ ที่มี 7 ชั้น

ค่าความยาก (β_j)	จำนวนข้อ	หมายเลขข้อสอบ
-3.00	1	22
-2.50	1	16
-2.00	2	11, 23
-1.50	2	7, 17
-1.00	3	4, 12, 24
-0.50	3	2, 8, 18
0.00	4	1, 5, 13, 25
0.50	3	3, 9, 19
1.00	3	6, 14, 26
1.50	2	10, 20
2.00	2	15, 27
2.50	1	21
3.00	1	28



ภาพประกอบ 14 แสดง โครงสร้างและค่าความยากของข้อสอบการทดสอบเทเลอรัปปีรามิด
แบบขนาดชั้นคงที่ ที่มี 10 ชั้น



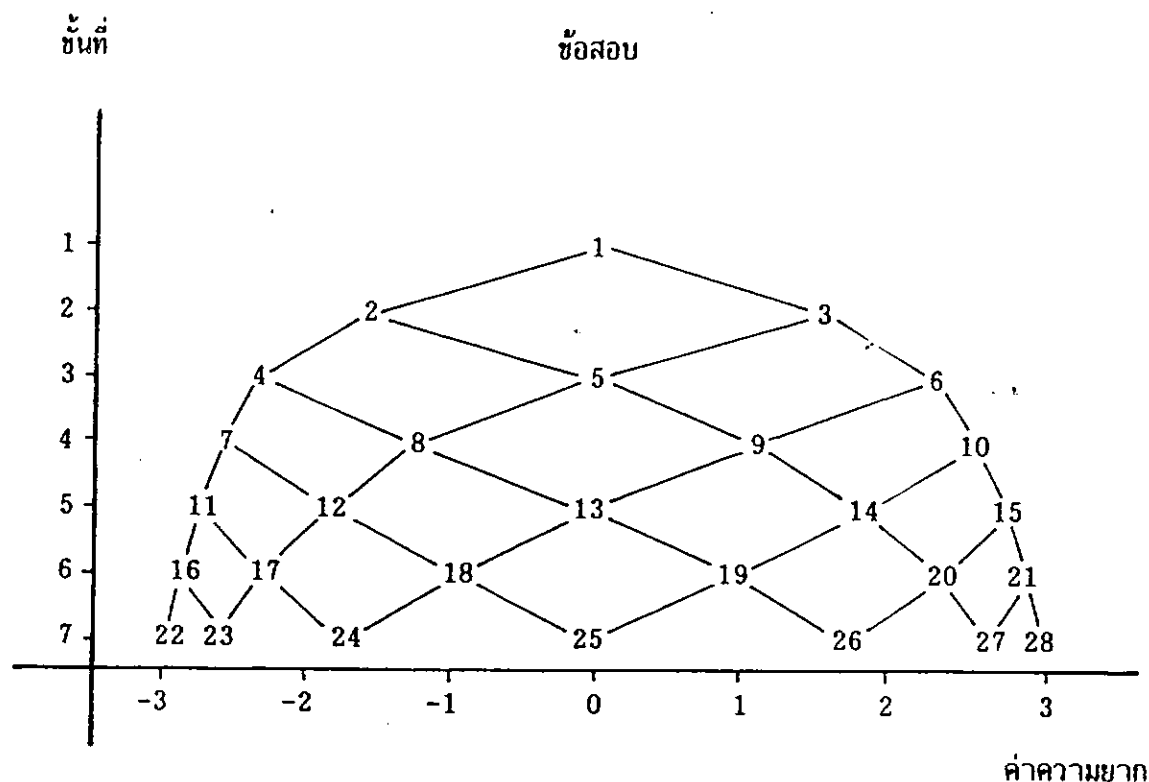
ภาพประกอบ 15 แสดงโครงสร้างและค่าความยากของข้อสอบการทดสอบเทเลอรัปบีรามีคแบบ
ขนาดชั้นคงที่ ที่มี 13 ชั้น

ข้อสอบ

ชั้นที่

ชั้นที่	ข้อสอบ	ค่าความยาก
1	1	-3.00
2	2	-2.80
3	3	-2.60
4	4	-2.40
5	5	-2.20
6	6	-2.00
7	7	-1.80
8	8	-1.60
9	9	-1.40
10	10	-1.20
11	11	-1.00
12	12	-0.80
13	13	-0.60
14	14	-0.40
15	15	-0.20
16	16	0.00
	17	0.20
	18	0.40
	19	0.60
	20	0.80
	21	1.00
	22	1.20
	23	1.40
	24	1.60
	25	1.80
	26	2.00
	27	2.20
	28	2.40
	29	2.60
	30	2.80
	31	3.00
	32	3.20
	33	3.40
	34	3.60
	35	3.80
	36	4.00
	37	4.20
	38	4.40
	39	4.60
	40	4.80
	41	5.00
	42	5.20
	43	5.40
	44	5.60
	45	5.80
	46	6.00
	47	6.20
	48	6.40
	49	6.60
	50	6.80
	51	7.00
	52	7.20
	53	7.40
	54	7.60
	55	7.80
	56	8.00
	57	8.20
	58	8.40
	59	8.60
	60	8.80
	61	9.00
	62	9.20
	63	9.40
	64	9.60
	65	9.80
	66	10.00
	67	10.20
	68	10.40
	69	10.60
	70	10.80
	71	11.00
	72	11.20
	73	11.40
	74	11.60
	75	11.80
	76	12.00
	77	12.20
	78	12.40
	79	12.60
	80	12.80
	81	13.00
	82	13.20
	83	13.40
	84	13.60
	85	13.80
	86	14.00
	87	14.20
	88	14.40
	89	14.60
	90	14.80
	91	15.00
	92	15.20
	93	15.40
	94	15.60
	95	15.80
	96	16.00
	97	16.20
	98	16.40
	99	16.60
	100	16.80
	101	17.00
	102	17.20
	103	17.40
	104	17.60
	105	17.80
	106	18.00
	107	18.20
	108	18.40
	109	18.60
	110	18.80
	111	19.00
	112	19.20
	113	19.40
	114	19.60
	115	19.80
	116	20.00
	117	20.20
	118	20.40
	119	20.60
	120	20.80
	121	21.00
	122	21.20
	123	21.40
	124	21.60
	125	21.80
	126	22.00
	127	22.20
	128	22.40
	129	22.60
	130	22.80
	131	23.00
	132	23.20
	133	23.40
	134	23.60
	135	23.80
	136	24.00

ภาพประกอบ 16 แสดงโครงสร้างและค่าความยากของข้อสอบการทดสอบเทคโนโลยีการบำบัดน้ำเสีย ที่มี 16 ข้อ



ภาพประกอบ 17 แสดงโครงสร้างของการทดสอบเทเลอรูปิรามิดแบบขนาดชั้นแปดที่มี 7 ชั้น

จากภาพประกอบ 17 จะเห็นว่าการทดสอบเทเลอรูปิรามิดแบบขนาดชั้นแปด จะมีช่วงห่างค่าความยากของข้อที่อยู่ติดกันภายในชั้นไม่เท่ากัน ในชั้นกลาง ๆ จะมีค่ามากแล้วค่อย ๆ ลดลงในข้อที่ค่อนข้างง่ายหรือยาก สำหรับการเพิ่มหรือลดนั้น จะเพิ่มหรือลดไปยังตำแหน่งกึ่งกลางระหว่าง 2 ข้อในชั้นที่มาก่อน หรือระหว่างข้อในชั้นที่มาก่อนกับค่าความยากสูงสุดหรือค่าสุดคั่งตัวอย่างใน ภาพประกอบ 17

ชั้นที่ 1 มีข้อสอบเพียงข้อเดียวคือข้อที่ 1 ซึ่งมีความยาก 0.0000

ชั้นที่ 2 มีข้อสอบสองข้อ คือข้อ 2 และ 3

ข้อ 2 มีความยากอยู่กึ่งกลางระหว่าง -3.0000 (ความยากค่าสุด) กับ 0.0000 (ความยากของข้อที่ 1) ซึ่งก็คือ -1.5000

ข้อ 3 มีความยากอยู่กึ่งกลางระหว่าง 3.0000 (ความยากสูงสุด) กับ 0.0000
(ความยากของข้อที่ 1) ซึ่งก็คือ 1.5000

ขั้นที่ 3 มีข้อสอบสามข้อ คือข้อ 4, 5 และ 6.

ข้อ 4 มีความยากอยู่กึ่งกลางระหว่าง -3.0000 (ความยากต่ำสุด) กับ -1.5000
(ความยากของข้อที่ 2) ซึ่งก็คือ -2.2500

ข้อ 5 มีความยากเท่ากับ 0.0000

ข้อ 6 มีความยากอยู่กึ่งกลางระหว่าง 1.5000 (ความยากของข้อที่ 3)
กับ 3.0000 (ความยากสูงสุด) ซึ่งก็คือ 2.2500

ขั้นที่ 4 มีข้อสอบสี่ข้อ คือข้อ 7, 8, 9 และ 10.

ข้อ 7 มีความยากอยู่กึ่งกลางระหว่าง -3.0000 (ความยากต่ำสุด) กับ -2.2500
(ความยากของข้อที่ 4) ซึ่งก็คือ -2.6250

ข้อ 8 มีความยากอยู่กึ่งกลางระหว่าง -2.2500 (ความยากของข้อที่ 4)
กับ 0.0000 (ความยากของข้อที่ 5) ซึ่งก็คือ -1.1250

ข้อ 9 มีความยากอยู่กึ่งกลางระหว่าง 0.0000 (ความยากของข้อที่ 5) กับ 2.2500
(ความยากของข้อที่ 6) ซึ่งก็คือ 1.1250

ข้อ 10 มีความยากอยู่กึ่งกลางระหว่าง 2.2500 (ความยากของข้อที่ 6)
กับ 3.0000 (ความยากสูงสุด) ซึ่งก็คือ 2.6250

สำหรับค่าความยากของข้อสอบในขั้นอื่น ๆ ก็สามารถหาได้ในตนเองเดียวกัน ในที่สุดก็
จะได้ค่าความยากของข้อสอบเป็นรายข้อ ดังตาราง 3

อนึ่ง สำหรับค่าความยากของข้อสอบการทดสอบเทเลอรรูปปริามิคแบบขนาดชั้นแปรรัน ที่
มี 10, 13 และ 16 ชั้น ก็หาได้ในตนเองเดียวกัน ได้แสดงค่าความยากเป็นรายข้อ ดังตาราง
4 ถึง 6

ตาราง 3 แสดงค่าความยากเป็นรายข้อ (β_j) ของข้อสอบการทดสอบเทเลอรรูปปิรามิดแบบ
ขนาดชั้นแปดชั้น ที่มี 7 ชั้น

ข้อที่	ค่าความยาก	ข้อที่	ค่าความยาก	ข้อที่	ค่าความยาก	ข้อที่	ค่าความยาก
1	0.0000	8	-1.1250	15	2.8125	22	-3.0000
2	-1.5000	9	1.1250	16	-2.9063	23	-2.6251
3	1.5000	10	2.6250	17	-2.3438	24	-1.6407
4	-2.2500	11	-2.8125	18	-0.9375	25	0.0000
5	0.0000	12	-1.8750	19	0.9375	26	1.6407
6	2.2500	13	0.0000	20	2.3438	27	2.6251
7	-2.6250	14	1.8750	21	2.9063	28	3.0000

ตาราง 4 แสดงค่าความยากเป็นรายข้อ (β_j) ของข้อสอบการทดสอบเทเลอรรูปปิรามิดแบบ
ขนาดชั้นแปดชั้น ที่มี 10 ชั้น

ข้อที่	ค่าความยาก	ข้อที่	ค่าความยาก	ข้อที่	ค่าความยาก	ข้อที่	ค่าความยาก
1	0.0000	15	2.8125	29	-2.9766	43	2.4611
2	-1.5000	16	-2.9063	30	-2.7892	44	2.8829
3	1.5000	17	-2.3438	31	-2.1329	45	2.9883
4	-2.2500	18	-0.9375	32	-0.8204	46	-3.0000
5	0.0000	19	0.9375	33	0.8204	47	-2.9356
6	2.2500	20	2.3438	34	2.1329	48	-2.6720
7	-2.6250	21	2.9063	35	2.7892	49	-1.9689
8	-1.1250	22	-2.9532	36	2.9766	50	-0.7384
9	1.1250	23	-2.6251	37	-2.9883	51	0.7384
10	2.6250	24	-1.6407	38	-2.8829	52	1.9689
11	-2.8125	25	0.0000	39	-2.4611	53	2.6720
12	-1.8750	26	1.6407	40	-1.4767	54	2.9356
13	0.0000	27	2.6251	41	0.0000	55	3.0000
14	1.8750	28	2.9532	42	1.4767		

ตาราง 5 แสดงค่าความยากเป็นรายข้อ (p_j) ของข้อสอบการทดสอบเทเลอ์รูปปิรามิดแบบ
ขนาดชั้นแปดชั้น ที่มี 13 ชั้น

ข้อที่	ค่าความยาก	ข้อที่	ค่าความยาก	ข้อที่	ค่าความยาก	ข้อที่	ค่าความยาก
1	0.0000	24	-1.6407	47	-2.9356	70	-2.5622
2	-1.5000	25	0.0000	48	-2.6720	71	-1.8371
3	1.5000	26	1.6407	49	-1.9689	72	-0.6769
4	-2.2500	27	2.6251	50	-0.7384	73	0.6769
5	0.0000	28	2.9532	51	0.7384	74	1.8371
6	2.2500	29	-2.9766	52	1.9689	75	2.5622
7	-2.6250	30	-2.7892	53	2.6720	76	2.8844
8	-1.1250	31	-2.1329	54	2.9356	77	2.9810
9	1.1250	32	-0.8204	55	2.9942	78	2.9986
10	2.6250	33	0.8204	56	-2.9971	79	-3.0000
11	-2.8125	34	2.1329	57	-2.9649	80	-2.9898
12	-1.8750	35	2.7892	58	-2.8038	81	-2.9327
13	0.0000	36	2.9766	59	-2.3205	82	-2.7233
14	1.8750	37	-2.9883	60	-1.3537	83	-2.1997
15	2.8125	38	-2.8829	61	0.0000	84	-1.2570
16	-2.9063	39	-2.4611	62	1.3537	85	0.0000
17	-2.3438	40	-1.4767	63	2.3205	86	1.2570
18	-0.9375	41	0.0000	64	2.8038	87	2.1997
19	0.9375	42	1.4767	65	2.9649	88	2.7233
20	2.3438	43	2.4611	66	2.9971	89	2.9327
21	2.9063	44	2.8829	67	-2.9986	90	2.9898
22	-2.9532	45	2.9883	68	-2.9810	91	3.0000
23	-2.6251	46	-2.9942	69	-2.8844		

ตาราง 6 แสดงค่าความยากเป็นรายข้อ (β_j) ของข้อสอบการทดสอบเทเลรูปปริมาตรแบบ
ขนาดชั้นแปดที่มี 16 ชั้น

ข้อที่	ค่าความยาก	ข้อที่	ค่าความยาก	ข้อที่	ค่าความยาก	ข้อที่	ค่าความยาก
1	0.0000	35	2.7892	69	-2.8844	103	2.9613
2	-1.5000	36	2.9766	70	-2.5622	104	2.9946
3	1.5000	37	-2.9883	71	-1.8371	105	2.9997
4	-2.2500	38	-2.8829	72	-0.6769	106	-2.9999
5	0.0000	39	-2.4611	73	0.6769	107	-2.9972
6	2.2500	40	-1.4767	74	1.8371	108	-2.9780
7	-2.6250	41	0.0000	75	2.5622	109	-2.8947
8	-1.1250	42	1.4767	76	2.8844	110	-2.6448
9	1.1250	43	2.4611	77	2.9810	111	-2.0950
10	2.6250	44	2.8829	78	2.9986	112	-1.1785
11	-2.8125	45	2.9883	79	-2.9993	113	0.0000
12	-1.8750	46	-2.9942	80	-2.9898	114	1.1785
13	0.0000	47	-2.9356	81	-2.9327	115	2.0950
14	1.8750	48	-2.6720	82	-2.7233	116	2.6448
15	2.8125	49	-1.9689	83	-2.1997	117	2.8947
16	-2.9063	50	-0.7384	84	-1.2570	118	2.9780
17	-2.3438	51	0.7384	85	0.0000	119	2.9972
18	-0.9375	52	1.9689	86	1.2570	120	2.9999
19	0.9375	53	2.6720	87	2.1997	121	-3.0000
20	2.3438	54	2.9356	88	2.7233	122	-2.9986
21	2.9063	55	2.9942	89	2.9327	123	-2.9876
22	-2.9532	56	-2.9971	90	2.9898	124	-2.9364
23	-2.6251	57	-2.9649	91	2.9993	125	-2.7698

ตาราง 6 (ต่อ)

ข้อที่	ค่าความยาก	ข้อที่	ค่าความยาก	ข้อที่	ค่าความยาก	ข้อที่	ค่าความยาก
24	-1.6407	58	-2.8038	92	-2.9997	126	-2.3699
25	0.0000	59	-2.3205	93	-2.9946	127	-1.6368
26	1.6407	60	-1.3537	94	-2.9613	128	-0.5893
27	2.6251	61	0.0000	95	-2.8280	129	0.5893
28	2.9532	62	1.3537	96	-2.4615	130	1.6368
29	-2.9766	63	2.3205	97	-1.7284	131	2.3699
30	-2.7892	64	2.8038	98	-0.6285	132	2.7698
31	-2.1329	65	2.9649	99	0.6285	133	2.9364
32	-0.8204	66	2.9971	100	1.7284	134	2.9876
33	0.8204	67	-2.9986	101	2.4615	135	2.9986
34	2.1329	68	-2.9810	102	2.8280	136	3.0000

3. หาค่าโอกาส (Probability) ในการตอบข้อสอบถูกของกลุ่มตัวอย่างที่จำลองขึ้น
ในข้อ 1. กับข้อสอบที่จำลองขึ้นในข้อ 2. โดยใช้สูตร แบบ 1-พารามิเตอร์ (One-Parameter
Logistic Model) (Hambleton and Swaminathan. 1985: 47) ดังนี้

$$P_{ij} = \frac{e^{1.7(\theta_i - \beta_j)}}{1 + e^{1.7(\theta_i - \beta_j)}}$$

เมื่อ P_{ij} คือ โอกาสในการตอบข้อสอบถูกของคนที่ i ในข้อสอบข้อที่ j

θ_i คือ ค่าความสามารถจริงของผู้สอบคนที่ i

β_j คือ ค่าความยากของข้อสอบข้อที่ j

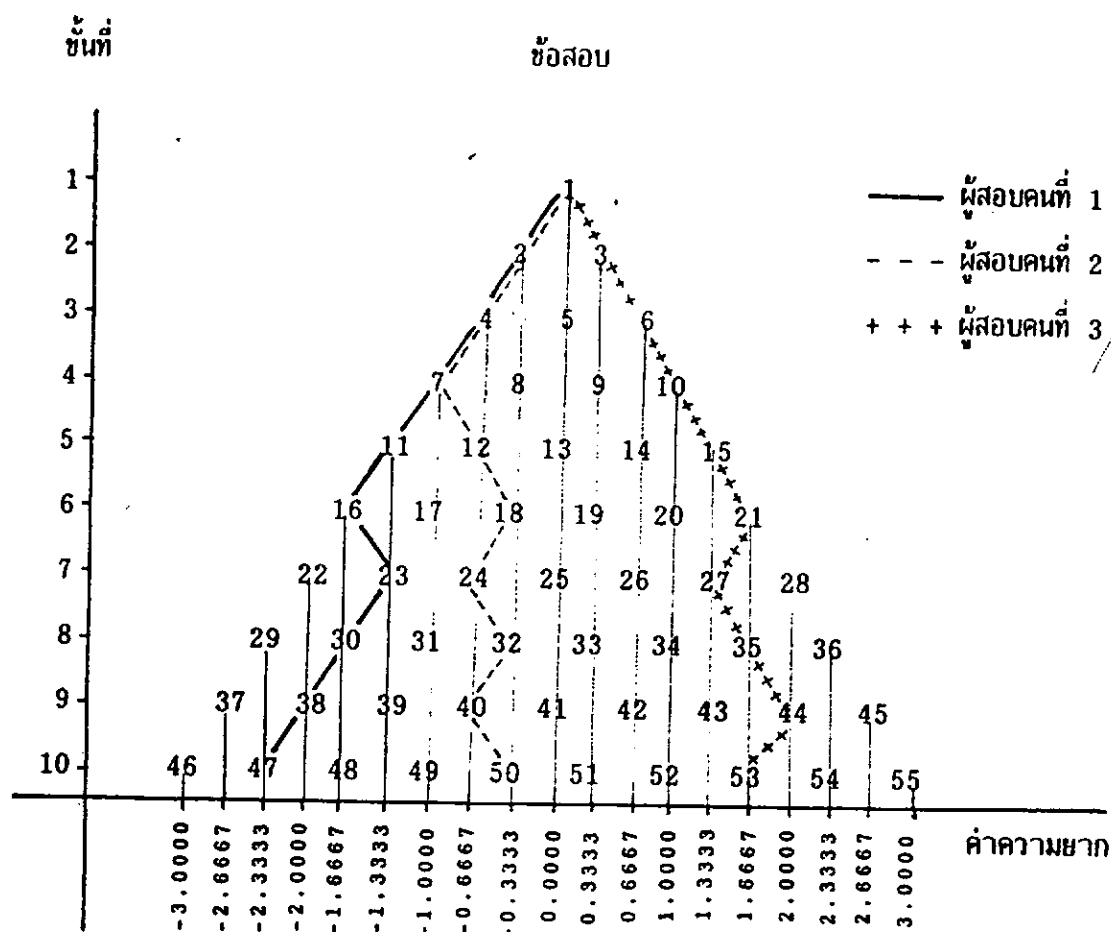
จากการใช้สูตรนี้หาค่าโอกาสในการตอบข้อสอบถูก (P_{1j}) ของกลุ่มตัวอย่างที่จำลองขึ้น
ในข้อ 1. กับข้อสอบที่จำลองขึ้นในข้อ 2. ในแต่ละครั้งการทดลองจะได้ผล ดังตาราง 7

ตาราง 7 แสดงค่าโอกาสในการตอบข้อสอบถูก ของผู้สอบแต่ละคนในข้อสอบแต่ละข้อ (P_{1j})
(เมื่อใช้สูตรในข้อ 3)

คนที่	ข้อที่									
	1	2	3	-	-	-	-	-	-	k
1	$P_{1,1}$	$P_{1,2}$	$P_{1,3}$	-	-	-	-	-	-	$P_{1,k}$
2	$P_{2,1}$	$P_{2,2}$	$P_{2,3}$	-	-	-	-	-	-	$P_{2,k}$
3	$P_{3,1}$	$P_{3,2}$	$P_{3,3}$	-	-	-	-	-	-	$P_{3,k}$
-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-
n	$P_{n,1}$	$P_{n,2}$	$P_{n,3}$	-	-	-	-	-	-	$P_{n,k}$

4. หาผลการตอบข้อสอบของผู้สอบ จากโอกาสในการตอบข้อสอบถูก (P_{1j}) ที่ได้ใน
ข้อ 3. นามหาผลการตอบ คือ ตอบถูก=1 ตอบผิด=0 ของผู้สอบแต่ละคนในข้อสอบแต่ละข้อ โดย
การเรียกเลขสุ่มที่มีการแจกแจงเป็นแบบยูนิฟอร์ม (Uniform) ที่มีค่า ตั้งแต่ 0 ถึง 1 (X_{1j})
จากเครื่องคอมพิวเตอร์ แล้วนำค่า X_{1j} ไปเปรียบเทียบกับค่า P_{1j} ถ้า $X_{1j} < P_{1j}$ แสดงว่า
ตกอยู่ในพื้นที่ที่ยอมรับว่าตอบข้อสอบถูก (ดูภาพประกอบ 18) จึงให้ $U_{1j}=1$ และถ้า $X_{1j} > P_{1j}$
ก็จะปฏิเสธการตอบถูก นั่นคือ ตอบผิด ก็จะให้ $U_{1j}=0$ (Weiss and McBride. 1984: 276;
Vijver. 1986: 46) ตัวอย่างนี้กับผู้สอบทุกคนในข้อสอบทุกข้อ ก็จะได้ผลการตอบ (0, 1)
ดังตาราง 11 โดยจะทาสีทั้งหมด 40 ครั้ง ซึ่งเปรียบเหมือนผู้สอบทำข้อสอบชุดเดิมซ้ำ 40 ครั้ง

จากผลการตอบถูกหรือผิด (1,0) ของผู้สอบ (3 คนแรก) นี้ เมื่อนำมาหาเส้นทางในการตอบ การทดสอบทฤษฎีรูปปิรามิดแบบขนาดชั้นคงที่ ที่มี 10 ชั้น จะได้ดัง ภาพประกอบ 19



ภาพประกอบ 19 แสดงเส้นทางในการตอบ การทดสอบทฤษฎีรูปปิรามิดแบบขนาดชั้นคงที่ ที่มี 10 ชั้น ของผู้สอบในตัวอย่าง (3 คนแรก)

จากภาพประกอบ 19 จะเห็นว่าผู้สอบคนที่ 1 เริ่มทำข้อสอบข้อที่ 1 ซึ่งมีความยากเท่ากับ 0.0000 ปรากฏว่าทำผิดจึงไปทำข้อที่ง่ายขึ้นคือข้อ 2 ซึ่งมีความยากเท่ากับ -0.3333 ปรากฏว่าทำผิดอีกจึงไปทำข้อที่ง่ายขึ้นอีกคือข้อ 4 ซึ่งมีความยากเท่ากับ -0.6667 ปรากฏว่าทำผิดอีกจึงไปทำข้อที่ง่ายขึ้นอีกคือข้อ 7 ซึ่งมีความยากเท่ากับ -1.0000 ปรากฏว่าทำผิดอีกจึงไปทำ

ข้อที่ง่ายขึ้นอีกคือข้อ 11 ซึ่งมีความยากเท่ากับ -1.3333 ปรากฏว่าทำผิดอีกจึง ไปทำข้อที่ง่ายขึ้น
 อีกคือข้อ 16 ซึ่งมีความยากเท่ากับ -1.6667 ปรากฏว่าทำถูกจึง ไปทำข้อที่ยากขึ้นคือข้อ 23 ซึ่งม
 ความยากเท่ากับ -1.3333 ปรากฏว่าทำผิดจึงไปทำข้อที่ง่ายขึ้นอีกคือข้อ 30 ซึ่งมีความยาก
 เท่ากับ -1.6667 ปรากฏว่าทำผิดอีกจึง ไปทำข้อที่ง่ายขึ้นอีกคือข้อ 38 ซึ่งมีความยากเท่ากับ
 -2.0000 ปรากฏว่าทำผิดอีก จึงไปทำข้อที่ง่ายขึ้นอีกคือข้อ 47 ซึ่งเป็นขั้นสุดท้ายมีความยาก
 เท่ากับ -2.3333 ปรากฏว่าทำถูก ฉะนั้นผลการทดสอบเทเลอร์ทั้ง 10 ขั้น ของผู้สอบคนที่ 1
 ปรากฏผล ดังตาราง 10

ตาราง 10 แสดงค่าความยากของข้อสอบ (β_j) และผลการตอบของผู้สอบคนที่ 1

ขั้นที่	หมายเลขข้อสอบที่ทำ	ค่าความยากของข้อสอบ (β_j)	ผลการตอบ
1	1	0.0000	ผิด
2	2	-0.3333	ผิด
3	4	-0.6667	ผิด
4	7	-1.0000	ผิด
5	11	-1.3333	ผิด
6	16	-1.6667	ถูก
7	23	-1.3333	ผิด
8	30	-1.6667	ผิด
9	38	-2.0000	ผิด
10	47	-2.3333	ถูก

การประมาณค่าความสามารถของผู้สอบจากการทดสอบเทเลอร์ ($\hat{\theta}_1$)

วิธีที่ 1 ให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำได้ จะเห็นว่าข้อที่ยากที่สุดที่ทำได้ คือ ข้อที่ 16 ซึ่งมีค่าความยากเท่ากับ -1.6667 ฉะนั้นความสามารถของผู้สอบที่ประมาณค่าจากการทดสอบเทเลอร์ ($\hat{\theta}_1$) มีค่าเท่ากับ -1.6667

วิธีที่ 2 ให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทำได้ ดังนี้

$$\text{ค่าความยากเฉลี่ยของข้อที่ทำได้} = (-1.6667 - 2.3333) / 2 = -2.0000$$

ฉะนั้นความสามารถของผู้สอบที่ประมาณค่าจากการทดสอบเทเลอร์ ($\hat{\theta}_1$) มีค่าเท่ากับ -2.0000

วิธีที่ 3 ให้คะแนนโดยใช้วิธีของเบส์ (Bayesian Updating) ซึ่งจะใช้สูตร ในการคำนวณ ดังนี้

กรณีที่ตอบข้อสอบถูก

$$\theta_{n+1} = \theta_n + \{\sigma_n^2 / \sqrt{(1+\sigma_n^2)}\} \{\phi(D) / \Phi(-D)\}$$

และ

$$\sigma_{n+1}^2 = \sigma_n^2 [1 - \{1/(1+1/\sigma_n^2)\} \{\phi(D) / \Phi(-D)\} \{\phi(D) / \Phi(-D) - D\}]$$

กรณีที่ตอบข้อสอบผิด

$$\theta_{n+1} = \theta_n - \{\sigma_n^2 / \sqrt{(1+\sigma_n^2)}\} \{\phi(D) / \Phi(D)\}$$

และ

$$\sigma_{n+1}^2 = \sigma_n^2 [1 - \{\phi(D) / (1+1/\sigma_n^2)\} \{\phi(D) / \Phi(D) + D\} / \Phi(D)]$$

เมื่อ $\phi(D)$ คือ ค่าออร์ดิเนต (Ordinate) ของโค้งปกติ ณ จุด D

$\Phi(D)$ คือ พื้นที่ใต้โค้งปกติจากค่า D ต่ำสุดจนถึงจุด D

$$D = (b_g - \theta_n) / \sqrt{(1+\sigma_n^2)}$$

θ_n คือ ความสามารถของผู้สอบที่ประมาณค่าได้ก่อนตอบข้อสอบข้อที่ $n+1$ โดยปกติ ถ้าเป็นการประมาณค่าครั้งแรก ที่ยังไม่ทราบค่าความสามารถของผู้สอบ จะให้ $\theta_n = 0.0000$

σ_n^2 คือ ความแปรปรวนในการประมาณค่าความสามารถของผู้สอบก่อนตอบข้อสอบ
ข้อที่ $n+1$ โดยปกติถ้าเป็นการประมาณค่าครั้งแรก ที่ยังไม่ทราบค่าความ
แปรปรวน จะให้ $\sigma_n^2 = 1.0000$

θ_{n+1} คือ ความสามารถของผู้สอบที่ประมาณค่าได้เมื่อตอบข้อสอบข้อที่ $n+1$ แล้ว

σ_{n+1}^2 คือ ความแปรปรวนในการประมาณค่าความสามารถของผู้สอบเมื่อตอบข้อสอบ
ข้อที่ $n+1$ แล้ว

b_n คือ ค่าความยากของข้อสอบข้อที่ $n+1$

ผลการประมาณค่าความสามารถของผู้สอบคนที่ 1

จุดเริ่มต้นกำหนดให้ $n=0$, $\theta_n = 0.0000$, $\sigma_n^2 = 1.0000$

ขั้นที่ 1 ทาข้อสอบข้อที่ 1 ซึ่งมีความยากเท่ากับ 0.0000 ผลปรากฏว่าทาผิด

$$D_1 = (0.0000 - 0.0000) / \sqrt{(1 + 1.0000)} = 0.0000$$

จากตารางพื้นที่ใต้โค้งปกติ จะได้ $\phi(D) = .3989$; $\Phi(D) = .5000$

$$\begin{aligned}\theta_1 &= 0.0000 - \{1.0000 / \sqrt{(1 + 1.0000)}\} \{ .3989 / .5000 \} \\ &= -0.5641\end{aligned}$$

$$\begin{aligned}\sigma_1^2 &= 1.0000 [1 - \{ .3989 / (1 + 1/1.0000) \} \{ .3989 / .5000 + .0000 \} / .5000] \\ &= 0.6818\end{aligned}$$

ขั้นที่ 2 ทาข้อสอบข้อที่ 2 ซึ่งมีความยากเท่ากับ -0.3333 ผลปรากฏว่าทาผิด

$$D_2 = (-0.3333 + 0.5641) / \sqrt{(1 + .6818)} = 0.1780$$

จากตารางพื้นที่ใต้โค้งปกติ จะได้ $\phi(D) = .3925$; $\Phi(D) = .5714$

$$\begin{aligned}\theta_2 &= -0.5641 - \{ .6818 / \sqrt{(1 + .6818)} \} \{ .3925 / .5714 \} \\ &= -0.9252\end{aligned}$$

$$\begin{aligned}\sigma_2^2 &= 0.6818 [1 - \{ .3925 / (1 + 1/.6818) \} \{ .3925 / .5714 + .1780 \} / .5714] \\ &= 0.5176\end{aligned}$$

ขั้นที่ 3 ทาข้อสอบข้อที่ 4 ซึ่งมีความยากเท่ากับ -0.6667 ผลปรากฏว่าทาผิด

$$D_3 = (-0.6667 + 0.9252) / \sqrt{(1 + .5176)} = 0.2098$$

จากตารางพื้นที่ใต้โค้งปกติ จะได้ $\phi(D) = .3902$; $\Phi(D) = .5832$

$$\begin{aligned} \theta_3 &= -0.9252 - \{ .5176 / \sqrt{(1 + .5176)} \} \{ .3902 / .5832 \} \\ &= -1.2063 \end{aligned}$$

$$\begin{aligned} \phi_3^2 &= 0.5176 [1 - \{ .3902 / (1 + 1 / .5176) \} \{ .3902 / .5832 + .2098 \} / .5832] \\ &= 0.4138 \end{aligned}$$

ขั้นที่ 4 ทดสอบข้อที่ 7 ซึ่งมีความยากเท่ากับ -1.0000 ผลปรากฏว่าผ่าน

$$D_4 = (-1.0000 + 1.2063) / \sqrt{(1 + .4138)} = 0.1735$$

จากตารางพื้นที่ใต้โค้งปกติ จะได้ $\phi(D) = .3932$; $\Phi(D) = .5675$

$$\begin{aligned} \theta_4 &= -1.2063 - \{ .4138 / \sqrt{(1 + .4138)} \} \{ .3932 / .5675 \} \\ &= -1.4474 \end{aligned}$$

$$\begin{aligned} \phi_4^2 &= 0.4138 [1 - \{ .3932 / (1 + 1 / .4138) \} \{ .3932 / .5675 + .1735 \} / .5675] \\ &= 0.3411 \end{aligned}$$

ขั้นที่ 5 ทดสอบข้อที่ 11 ซึ่งมีความยากเท่ากับ -1.3333 ผลปรากฏว่าผ่าน

$$D_5 = (-1.3333 + 1.4474) / \sqrt{(1 + .3411)} = 0.0985$$

จากตารางพื้นที่ใต้โค้งปกติ จะได้ $\phi(D) = .3970$; $\Phi(D) = .5398$

$$\begin{aligned} \theta_5 &= -1.4474 - \{ .3411 / \sqrt{(1 + .3411)} \} \{ .3970 / .5398 \} \\ &= -1.6640 \end{aligned}$$

$$\begin{aligned} \phi_5^2 &= 0.3411 [1 - \{ .3970 / (1 + 1 / .3411) \} \{ .3970 / .5398 + .0985 \} / .4602] \\ &= 0.2879 \end{aligned}$$

ขั้นที่ 6 ทดสอบข้อที่ 16 ซึ่งมีความยากเท่ากับ -1.6667 ผลปรากฏว่าถูก

$$D_6 = (-1.6667 + 1.6640) / \sqrt{(1 + .2879)} = -0.0024$$

จากตารางพื้นที่ใต้โค้งปกติ จะได้ $\phi(D) = .3989$; $\Phi(-D) = .5000$

$$\begin{aligned} \theta_6 &= -1.6640 + \{ .2879 / \sqrt{(1 + .2879)} \} \{ .3989 / .5000 \} \\ &= -1.4616 \end{aligned}$$

$$\begin{aligned} \delta_6^2 &= 0.2879[1 - \{1/(1+1/.2879)\}\{.3989/.5000\}\{.3989/.5000+.0024\}] \\ &= 0.2468 \end{aligned}$$

ขั้นที่ 7 ทําสอบข้อที่ 23 ซึ่งมีความยากเท่ากับ -1.3333 ผลปรากฏว่าทําคิด

$$D_7 = (-1.3333+1.4616)/\sqrt{(1+.2468)} = 0.1149$$

จากตารางพื้นที่ใต้โค้งปกติ จะได้ $\phi(D)=.3961$; $\Phi(D)=.5478$

$$\begin{aligned} \theta_7 &= -1.4616 - \{.2468/\sqrt{(1+.2468)}\}\{.3961/.5478\} \\ &= -1.6214 \end{aligned}$$

$$\begin{aligned} \delta_7^2 &= 0.2468[1 - \{.3961/(1+1/.2468)\}\{.3961/.5478+.1149\}/.5478] \\ &= 0.2172 \end{aligned}$$

ขั้นที่ 8 ทําสอบข้อที่ 30 ซึ่งมีความยากเท่ากับ -1.6667 ผลปรากฏว่าทําคิด

$$D_8 = (-1.6667+1.6214)/\sqrt{(1+.2172)} = -0.0411$$

จากตารางพื้นที่ใต้โค้งปกติ จะได้ $\phi(D)=.3986$; $\Phi(D)=.4840$

$$\begin{aligned} \theta_8 &= -1.6214 - \{.2172/\sqrt{(1+.2172)}\}\{.3986/.4840\} \\ &= -1.7835 \end{aligned}$$

$$\begin{aligned} \delta_8^2 &= 0.2172[1 - \{.3986/(1+1/.2172)\}\{.3986/.4840-.0411\}/.4840] \\ &= 0.1922 \end{aligned}$$

ขั้นที่ 9 ทําสอบข้อที่ 38 ซึ่งมีความยากเท่ากับ -2.0000 ผลปรากฏว่าทําคิด

$$D_9 = (-2.0000+1.7835)/\sqrt{(1+.1922)} = -0.1983$$

จากตารางพื้นที่ใต้โค้งปกติ จะได้ $\phi(D)=.3910$; $\Phi(D)=.4207$

$$\begin{aligned} \theta_9 &= -1.7835 - \{.1922/\sqrt{(1+.1922)}\}\{.3910/.4207\} \\ &= -1.9471 \end{aligned}$$

$$\begin{aligned} \delta_9^2 &= 0.1922[1 - \{.3910/(1+1/.1922)\}\{.3910/.4207-.1983\}/.4207] \\ &= 0.1712 \end{aligned}$$

ขั้นที่ 10 ทําสอบข้อที่ 47 ซึ่งมีความยากเท่ากับ -2.3333 ผลปรากฏว่าทําคิด

$$D_{10} = (-2.3333+1.9471)/\sqrt{(1+.1712)} = -0.3569$$

จากตารางพื้นที่ใต้โค้งปกติ จะได้ $\phi(D)=.3739$; $\Phi(-D)=.6406$

$$\theta_{10} = -1.9471 + \{.1712 / \sqrt{(1+.1712)}\} \{ .3739 / .6406 \}$$

$$= -1.8548$$

$$\delta_{10}^2 = 0.1712 [1 - \{1 / (1+.1712)\}] \{ .3739 / .6406 \} \{ .3739 / .6406 + .3569 \}$$

$$= 0.1575$$

ฉะนั้นความสามารถของผู้สอบที่ประมาณค่าจากการทดสอบเทเลอรัมีค่าเท่ากับ -1.8548

วิธีที่ 4 ให้คะแนนโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (Conditional Maximum Likelihood) โดยมีขั้นตอนในการคำนวณ ดังนี้

(1) ประมาณค่าความสามารถเริ่มต้นของผู้สอบ (θ_m ; $m=0$) จากสูตร

$$\theta_m = \ln[r / (k - r)]$$

(2) หาค่า $p_j(\theta_m)$ และ $q_j(\theta_m)$ จากสูตร

$$p_j(\theta_m) = \exp D(\theta_m - \beta_j) / [1 + \exp D(\theta_m - \beta_j)]$$

$$q_j(\theta_m) = 1 - p_j(\theta_m) = 1 / [1 + \exp D(\theta_m - \beta_j)]$$

(3) หาค่าตัวปรับแก้ h_m โดยใช้สูตร

$$h_m = D[r - \sum p_j(\theta_m)] / [-D^2 \sum p_j(\theta_m) q_j(\theta_m)]$$

(4) ประมาณค่าความสามารถของผู้สอบใหม่ (θ_{m+1}) โดยใช้สูตร

$$\theta_{m+1} = \theta_m - h_m$$

(5) ทำการคำนวณซ้ำในข้อ (2), (3) และ (4) จนกระทั่ง h_m เข้าใกล้ศูนย์

$$(h_m < .001)$$

เมื่อ $p_j(\theta_m)$ คือ โอกาสที่ผู้สอบที่มีความสามารถ θ_m ตอบข้อสอบข้อที่ j ถูก

$q_j(\theta_m)$ คือ โอกาสที่ผู้สอบที่มีความสามารถ θ_m ตอบข้อสอบข้อที่ j ผิด

θ_m คือ ความสามารถของผู้สอบที่ประมาณค่าได้ในครั้งที่ m

β_j คือ ค่าความยากของข้อสอบข้อที่ j

r คือ จำนวนข้อที่ตอบถูก

k คือ จำนวนข้อทั้งหมดที่ทำในการทดสอบเทเลอรั

ผลการประมาณค่าความสามารถของผู้สอบคนที่ 1

$$\theta_0 = \ln[2/(10 - 2)] = -1.3863$$

ตาราง 11 แสดงผลการประมาณค่าความสามารถของผู้สอบ โดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (Conditional Maximum Likelihood) ของผู้สอบ คนที่ 1

ข้อสอบ ข้อที่	ค่าความยาก (β_j)	$m=0$ $P_j(\theta_0)$	$m=1$ $P_j(\theta_1)$	$m=2$ $P_j(\theta_2)$	$m=3$ $P_j(\theta_3)$
1	0.0000	0.0865	0.0241	0.0207	0.0206
2	-0.3333	0.1431	0.0417	0.0360	0.0358
4	-0.6667	0.2274	0.0713	0.0617	0.0614
7	-1.0000	0.3415	0.1192	0.1039	0.1034
11	-1.3333	0.4775	0.1925	0.1696	0.1689
16	-1.6667	0.6170	0.2959	0.2647	0.2637
23	-1.3333	0.4775	0.1925	0.1696	0.1689
30	-1.6667	0.6170	0.2959	0.2647	0.2637
38	-2.0000	0.7395	0.4255	0.3882	0.3869
47	-2.3333	0.8334	0.5662	0.5279	0.5266
$\Sigma P_j(\theta_m)$		4.5602	2.2247	2.0069	1.9999
$\Sigma P_j(\theta_m)Q_j(\theta_m)$		1.9053	1.4523	1.3636	1.3606
h_m		0.7904	0.0910	0.0030	0.0000
$\theta_{m+1} = \theta_m - h_m$		-2.1767	-2.2677	-2.2707	-2.2707

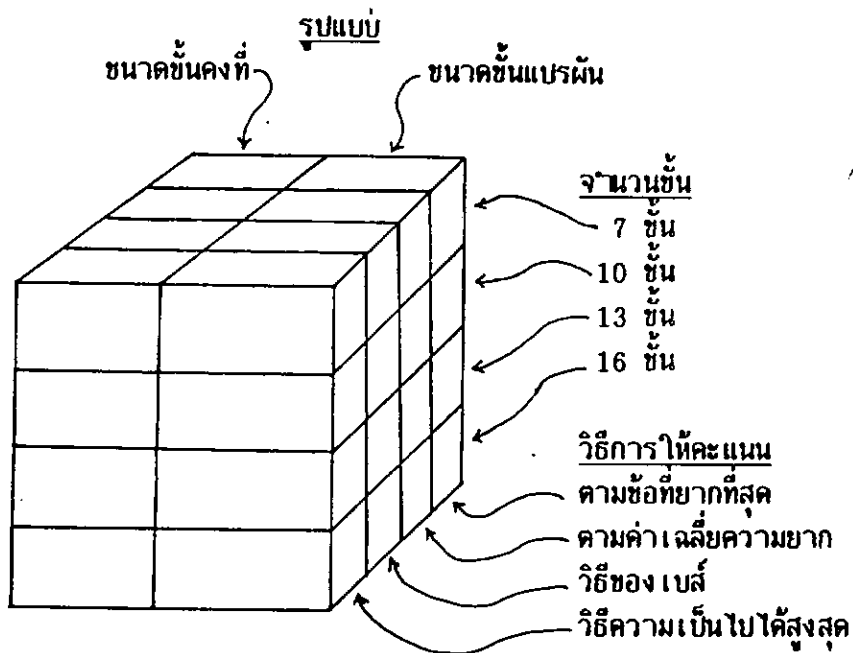
ฉะนั้นความสามารถของผู้สอบที่ประมาณค่าจากการทดสอบเทเลอร์มีค่าเท่ากับ -2.2707

ผู้สอบคนที่ 2 เริ่มทบทวนข้อที่ 1 ซึ่งมีความยากเท่ากับ 0.0000 ปรากฏว่าทบทวนข้อที่ง่ายขึ้นคือข้อ 2 ซึ่งมีความยากเท่ากับ -0.3333 ปรากฏว่าทบทวนข้อที่ง่ายขึ้นอีกคือข้อ 4 ซึ่งมีความยากเท่ากับ -0.6667 ปรากฏว่าทบทวนข้อที่ง่ายขึ้นอีกคือข้อ 7 ซึ่งมีความยากเท่ากับ -1.0000 ปรากฏว่าทบทวนข้อที่ยากขึ้นคือข้อ 12 ซึ่งมีความยากเท่ากับ -0.6667 ปรากฏว่าทบทวนข้อที่ยากขึ้นอีกคือข้อ 18 ซึ่งมีความยากเท่ากับ -0.3333 ปรากฏว่าทบทวนข้อที่ง่ายขึ้นคือข้อ 24 ซึ่งมีความยากเท่ากับ -0.6667 ปรากฏว่าทบทวนข้อที่ยากขึ้นคือข้อ 32 ซึ่งมีความยากเท่ากับ -0.3333 ปรากฏว่าทบทวนข้อที่ง่ายขึ้นคือข้อ 40 ซึ่งมีความยากเท่ากับ -0.6667 ปรากฏว่าทบทวนข้อ 50 ซึ่งเป็นขั้นสุดท้ายมีความยากเท่ากับ -0.3333 ปรากฏว่าทบทวน จะเห็นว่า ผู้สอบคนที่ 2 ทบทวนทดสอบมาแล้วทั้งหมด ข้อที่ยากที่สุดที่ทบทวนคือข้อ 40 ซึ่งมีความยากเท่ากับ -0.6667 ฉะนั้นถ้าให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทบทวน ผู้สอบคนที่ 2 จะมีความสามารถที่ประมาณค่าได้จากการทดสอบทดสอบ ($\hat{\theta}_1$) เท่ากับ -0.6667

ผู้สอบคนที่ 3 เริ่มทบทวนข้อที่ 1 ซึ่งมีความยากเท่ากับ 0.0000 ปรากฏว่าทบทวนข้อที่ยากขึ้นคือข้อ 3 ซึ่งมีความยากเท่ากับ 0.3333 ปรากฏว่าทบทวนข้อที่ยากขึ้นอีกคือข้อ 6 ซึ่งมีความยากเท่ากับ 0.6667 ปรากฏว่าทบทวนข้อที่ยากขึ้นอีกคือข้อ 10 ซึ่งมีความยากเท่ากับ 1.0000 ปรากฏว่าทบทวนข้อที่ยากขึ้นอีกคือข้อ 15 ซึ่งมีความยากเท่ากับ 1.3333 ปรากฏว่าทบทวนข้อที่ยากขึ้นอีกคือข้อ 21 ซึ่งมีความยากเท่ากับ 1.6667 ปรากฏว่าทบทวนข้อที่ง่ายขึ้นคือข้อ 27 ซึ่งมีความยากเท่ากับ 1.3333 ปรากฏว่าทบทวนข้อที่ยากขึ้นอีกคือข้อ 35 ซึ่งมีความยากเท่ากับ 1.6667 ปรากฏว่าทบทวนข้อที่ยากขึ้นอีกคือข้อ 44 ซึ่งมีความยากเท่ากับ 2.0000 ปรากฏว่าทบทวนข้อ 53 ซึ่งเป็นขั้นสุดท้ายมีความยากเท่ากับ 1.6667 ปรากฏว่าทบทวน จะเห็นว่าผู้สอบคนที่ 3 ทบทวนทดสอบมาแล้วทั้งหมด ข้อที่ยากที่สุดที่ทบทวนคือข้อ 53 ซึ่งมีความยากเท่ากับ 1.6667 ฉะนั้นถ้าให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทบทวน ผู้สอบคนที่ 3 จะมีความสามารถที่ประมาณค่าได้จากการทดสอบทดสอบ ($\hat{\theta}_1$) เท่ากับ 1.6667

การวิเคราะห์ข้อมูล

การวิจัยครั้งนี้ได้ออกแบบการวิจัยแบบ คอมพลีทลี่ แรนดอมไมซ์ แฟคทอเรียล ดีไซน์ (Completely Randomized Factorial Design: CRF_{2x4x4}) (Kirk. 1982: 429-455) โดยใช้ขนาดกลุ่มตัวอย่างในแต่ละเซลล์ 1,000 คน ทำการทดลองซ้ำ 40 ครั้ง และจะใช้ในการทดลองซ้ำแต่ละครั้ง เป็นหน่วยในการวิเคราะห์เปรียบเทียบ



ภาพประกอบ 20 แสดงรูปแบบการทดลอง

ขั้นตอนในการวิเคราะห์ข้อมูล

1. ทดสอบความแตกต่างระหว่างความสามารถจริง (θ_1) กับความสามารถที่ได้จากการประมาณค่า ($\hat{\theta}_1$) จากการทดสอบทีเลอร์ โดยใช้ t-test
2. หาค่าความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ (E_1) จากการทดสอบทีเลอร์รูปปริมาตร โดยใช้สูตร

$$E_i = \hat{\theta}_i - \theta_i$$

3. ในแต่ละครั้งการทดลองจะหาค่า ดังนี้

3.1 หาค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ โดยไม่คิดเครื่องหมาย ($\bar{E}$) โดยใช้สูตร

$$\bar{E} = \Sigma |E_i| / N$$

3.2 หาค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ (SEE) โดยใช้สูตร

$$SEE = \sqrt{\frac{N \Sigma E_i^2 - (\Sigma E_i)^2}{N(N-1)}}$$

3.3 หาค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ (Validity: $r_{\theta\hat{\theta}}$) ซึ่งหาได้จากความสัมพันธ์ระหว่างความสามารถจริง (θ_i) กับความสามารถที่ได้จากการประมาณค่า ($\hat{\theta}_i$) โดยใช้สูตร ดังนี้

$$r_{\theta\hat{\theta}} = \frac{N \Sigma \theta_i \hat{\theta}_i - \Sigma \theta_i \Sigma \hat{\theta}_i}{\sqrt{[N \Sigma \theta_i^2 - (\Sigma \theta_i)^2][N \Sigma \hat{\theta}_i^2 - (\Sigma \hat{\theta}_i)^2]}}$$

3.4 เปลี่ยนค่าความเที่ยงตรง ($r_{\theta\hat{\theta}}$) ให้เป็นค่า Fisher's Z (Z_r) และเปลี่ยนค่า Fisher's Z ให้เป็น $r_{\theta\hat{\theta}}$ โดยใช้สูตร ดังนี้ (Charter and Larsen. 1983: 41-42)

$$Z_r = 0.5[\log_e\{(1+re\hat{o})/(1-re\hat{o})\}]$$

$$re\hat{o} = (e^{2Z_r} - 1)/(e^{2Z_r} + 1)$$

4. จากการทดลองซ้ำจำนวน 40 ครั้ง ในแต่ละเซลล์การทดลอง (คุณภาพประกอบ 20) ก็จะมีค่า $\bar{E}$, SEE และ $re\hat{o}$ หรือ Z_r อย่างละ 40 ค่า
 - 4.1 เขียนกราฟเปรียบเทียบค่า $\bar{E}$, $\overline{SEE}$ และ $\overline{re\hat{o}}$
 - 4.2 เปรียบเทียบค่า $\bar{E}$, SEE และ Z_r โดยการวิเคราะห์ความแปรปรวนสามทาง (CRF_{2x4x4} ANOVA)
 - 4.3 ถ้าผลการวิเคราะห์ในข้อ 4.2 มีนัยสำคัญทางสถิติ ก็จะเปรียบเทียบเป็นรายคู่ โดยใช้วิธีของเชฟเฟ่ (Scheffe')

บทที่ 4

ผลการวิจัย

เพื่อความสะดวกในการนำเสนอผลการวิจัย จึงได้กำหนดสัญลักษณ์และตัวย่อต่าง ๆ
เพื่อใช้ในการนำเสนอผลการวิจัยในครั้งนี้ ดังนี้

- c7 หมายถึง การทดสอบเทเลอร์รูปปริมาตรแบบขนาดชั้นคงที่ ที่มี 7 ชั้น
- c10 หมายถึง การทดสอบเทเลอร์รูปปริมาตรแบบขนาดชั้นคงที่ ที่มี 10 ชั้น
- c13 หมายถึง การทดสอบเทเลอร์รูปปริมาตรแบบขนาดชั้นคงที่ ที่มี 13 ชั้น
- c16 หมายถึง การทดสอบเทเลอร์รูปปริมาตรแบบขนาดชั้นคงที่ ที่มี 16 ชั้น
- v7 หมายถึง การทดสอบเทเลอร์รูปปริมาตรแบบขนาดชั้นแปรผัน ที่มี 7 ชั้น
- v10 หมายถึง การทดสอบเทเลอร์รูปปริมาตรแบบขนาดชั้นแปรผัน ที่มี 10 ชั้น
- v13 หมายถึง การทดสอบเทเลอร์รูปปริมาตรแบบขนาดชั้นแปรผัน ที่มี 13 ชั้น
- v16 หมายถึง การทดสอบเทเลอร์รูปปริมาตรแบบขนาดชั้นแปรผัน ที่มี 16 ชั้น
- s1 หมายถึง การให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำได้
- s2 หมายถึง การให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทำได้
- s3 หมายถึง การให้คะแนนโดยใช้วิธีของเบส์ (Bayesian Updating)
- s4 หมายถึง การให้คะแนนโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข
(Conditional Maximum Likelihood)
- β_j หมายถึง ค่าความยากของข้อสอบ ข้อที่ j
- θ_i หมายถึง ค่าความสามารถจริงของผู้สอบ คนที่ i
- $\hat{\theta}_i$ หมายถึง ค่าความสามารถที่ได้จากการประมาณค่าจากการทดสอบเทเลอร์
รูปปริมาตร ของผู้สอบคนที่ i
- E_i หมายถึง ค่าความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ คนที่ i
($E_i = \hat{\theta}_i - \theta_i$)
- $\bar{E}$ หมายถึง ค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ

โดยไม่คิดเครื่องหมาย ($\bar{E} = \sum |\theta_1 - \theta_1| / N$)

SEE หมายถึง ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ

roθ หมายถึง ค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ

$\bar{D}$ หมายถึง ค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ

โดยคิดเครื่องหมายด้วย [$\bar{D} = \sum (\theta_1 - \theta_1) / N$]

SD $\bar{D}$ หมายถึง ค่าความคลาดเคลื่อนมาตรฐานของค่า $\bar{D}$ (Standard Error of $\bar{D}$)

to หมายถึง ค่าอัตราส่วนที่ ของค่า $\bar{D}$

การนำเสนอผลการวิจัยในครั้งนี้ จะแบ่งออกเป็น 3 หัวข้อตามวัตถุประสงค์ของการวิจัย ดังนี้

1. ผลการศึกษาและเปรียบเทียบค่าความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ (E) จากการทดสอบเทเลอร์รูปปริมาตรที่มี รูปแบบ จำนวนชั้น และวิธีการให้คะแนนที่แตกต่างกัน

1.1 การศึกษาค่าความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ จากการทดสอบเทเลอร์รูปปริมาตรที่มี รูปแบบ จำนวนชั้น และวิธีการให้คะแนน ที่แตกต่างกัน จะเสนอผล ลำดับ ดังนี้

(1) การทดสอบความแตกต่างระหว่างความสามารถจริงกับความสามารถที่ได้จากการประมาณค่า ในแต่ละกลุ่มการทดลองโดยการทดลองซ้ำ 40 ครั้ง ปรากฏผล ดังแสดงใน ตาราง 12

(2) ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ ($\bar{E}$) ในแต่ละกลุ่มการทดลองโดยการทดลองซ้ำ 40 ครั้ง ปรากฏผล ดังแสดงใน ตาราง 13 และภาพประกอบ 21 ถึง 23

1.2 การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ ($\bar{E}$) จากการทดสอบเทเลอร์รูปปริมาตรที่มี รูปแบบ จำนวนชั้น และวิธีการให้คะแนน ที่แตกต่างกัน โดยการวิเคราะห์ความแปรปรวน ปรากฏผล ดังแสดงใน ตาราง 14 และทดสอบเปรียบเทียบค่าเฉลี่ยของค่าเฉลี่ยเป็นรายคู่ ดังแสดงใน ตาราง 15

ตาราง 12 แสดงผลการทดสอบความแตกต่างระหว่างความสามารถจริงกับความสามารถที่ได้
จากการประมาณค่า ในแต่ละกลุ่มการทดลองโดยการทดลองซ้ำ 40 ครั้ง

กลุ่ม ทดลอง	ช่วงของค่า $\bar{D}$	ช่วงของค่า SD_D	ช่วงของค่า t_D	จำนวนครั้งที่ นัยสำคัญ (α)
c7s1	.0653 ถึง .1248	.015 ถึง .016	4.23 ถึง 8.11	40(.01)
c10s1	.1602 ถึง .2201	.014 ถึง .015	11.37 ถึง 15.18	40(.01)
c13s1	.1929 ถึง .2517	.013 ถึง .014	14.21 ถึง 17.95	40(.01)
c16s1	.2422 ถึง .2752	.013 ถึง .014	16.97 ถึง 20.73	40(.01)
c7s2	-.3116 ถึง -.2553	.014 ถึง .016	-21.13 ถึง -15.57	40(.01)
c10s2	-.1995 ถึง -.1597	.013 ถึง .014	-15.36 ถึง -11.48	40(.01)
c13s2	-.1600 ถึง -.1284	.013 ถึง .014	-11.80 ถึง -9.11	40(.01)
c16s2	-.1477 ถึง -.1131	.012 ถึง .013	-11.92 ถึง -8.72	40(.01)
c7s3	-.0138 ถึง .0541	.013 ถึง .015	-0.96 ถึง 3.89	2(.01);2(.05)
c10s3	-.0280 ถึง .0257	.011 ถึง .013	-2.38 ถึง 3.11	1(.01);2(.05)
c13s3	-.0068 ถึง .0365	.010 ถึง .011	-0.63 ถึง 3.42	1(.01)
c16s3	-.0046 ถึง .0425	.009 ถึง .010	-0.46 ถึง 2.11	1(.05)
c7s4	-.0280 ถึง .0369	.015 ถึง .017	-1.81 ถึง 3.36	2(.01);3(.05)
c10s4	-.0373 ถึง .0276	.013 ถึง .014	-2.83 ถึง 2.08	2(.01);2(.05)
c13s4	-.0341 ถึง .0266	.011 ถึง .012	-2.93 ถึง 2.34	1(.01);1(.05)
c16s4	-.0150 ถึง .0198	.010 ถึง .011	-1.51 ถึง 1.99	1(.05)
v7s1	.0705 ถึง .1390	.019 ถึง .021	3.53 ถึง 6.86	40(.01)
v10s1	.2577 ถึง .3263	.017 ถึง .018	14.72 ถึง 18.49	40(.01)
v13s1	.3649 ถึง .4325	.015 ถึง .017	22.59 ถึง 27.13	40(.01)
v16s1	.4505 ถึง .5092	.014 ถึง .016	29.86 ถึง 34.79	40(.01)

ตาราง 12 (ต่อ)

กลุ่มทดลอง	ช่วงของค่า $\bar{D}$	ช่วงของค่า SD $\bar{D}$	ช่วงของค่า t_D	จำนวนครั้งที่มีความสำคัญ (α)
v7s2	-.5407 ถึง -.4220	.015 ถึง .017	-35.42 ถึง -25.50	40(.01)
v10s2	-.4802 ถึง -.4245	.013 ถึง .014	-36.05 ถึง -31.53	40(.01)
v13s2	-.4357 ถึง -.3977	.011 ถึง .012	-37.40 ถึง -33.22	40(.01)
v16s2	-.4112 ถึง -.3653	.010 ถึง .011	-38.74 ถึง -33.06	40(.01)
v7s3	-.0134 ถึง .0478	.014 ถึง .016	-0.89 ถึง 2.98	2(.01);3(.05)
v10s3	-.0110 ถึง .0339	.012 ถึง .014	-0.83 ถึง 2.61	1(.01);2(.05)
v13s3	-.0290 ถึง .0252	.011 ถึง .012	-2.67 ถึง 2.38	1(.01);2(.05)
v16s3	-.0150 ถึง .0230	.010 ถึง .011	-1.49 ถึง 2.24	2(.05)
v7s4	-.0273 ถึง .0345	.016 ถึง .018	-1.63 ถึง 2.92	3(.01);3(.05)
v10s4	-.0221 ถึง .0396	.013 ถึง .015	-1.61 ถึง 2.88	1(.01);3(.05)
v13s4	-.0183 ถึง .0279	.012 ถึง .013	-1.51 ถึง 2.86	1(.01);2(.05)
v16s4	-.0242 ถึง .0263	.010 ถึง .011	-2.16 ถึง 2.58	1(.01)

จากตาราง 12 จะเห็นว่า การให้คะแนนโดยให้ตามค่าความยากของข้อที่ยากที่สุดที่ทำถูก (s1) ไม่ว่าจะเป็นการทดสอบเทเลอรัปปริมาตรแบบขนาดขั้นคงที่ หรือแบบขนาดขั้นแปรผัน และไม่ว่าจะใช้จำนวนขั้นของการทดสอบเทเลอรัปปริมาตรที่เป็น 7, 10, 13 หรือ 16 ขั้น ก็ตาม ทุกกลุ่มและทุกครั้งของการทดลอง จะให้ค่าเฉลี่ยความคลาดเคลื่อน ($\bar{D}$) มีค่าเป็นบวกเสมอ นั่นคือ การให้คะแนนโดยวิธีนี้จะประมาณค่าความสามารถของผู้สอบสูงกว่าความสามารถจริงเสมอ และ เมื่อทดสอบความแตกต่างระหว่างความสามารถที่ได้จากการประมาณค่ากับความสามารถจริง โดยใช้ ที-เทสท์ (t-test) แล้ว ปรากฏว่า แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

ในทุกกลุ่มและทุกครั้งของการทดลอง

การให้คะแนนโดยให้ตามค่าเฉลี่ยความยากของข้อที่หาถูก (s2) จะเห็นว่า ไม่ว่าจะเป็นการทดสอบเทเลอร์รูปปริมาตรแบบขนาดชั้นคงที่หรือแบบขนาดชั้นแปรผัน และ ไม่ว่าจะใช้จำนวนชั้นของการทดสอบเทเลอร์รูปปริมาตรที่เป็น 7, 10, 13 หรือ 16 ชั้น ก็ตาม ทุกกลุ่มและทุกครั้งของการทดลองจะให้ค่าเฉลี่ยความคลาดเคลื่อน ($\bar{D}$) มีค่าเป็นลบเสมอ นั่นคือ การให้คะแนนโดยวิธีนี้ จะประมาณค่าความสามารถของผู้สอบต่ำกว่าความสามารถจริงเสมอ และเมื่อทดสอบความแตกต่างระหว่าง ความสามารถที่ได้จากการประมาณค่ากับความสามารถจริง โดยใช้ ที-เทสต์ (t-test) ปรากฏว่า แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ในทุกกลุ่มและทุกครั้งของการทดลอง

สำหรับการให้คะแนนโดยวิธีของเบส์ (s3) และให้คะแนนโดยวิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (s4) จะให้ค่าเฉลี่ยความคลาดเคลื่อน ($\bar{D}$) มีทั้งบวกและลบ ในแต่ละกลุ่มการทดลองจากการทดลองซ้ำ 40 ครั้ง จะมีจำนวนครั้งไม่ต่ำกว่า 34 ครั้ง ที่ความสามารถจริงกับความสามารถที่ได้จากการประมาณค่า ไม่แตกต่างกัน

ตาราง 13 แสดงค่าเฉลี่ยของค่าเฉลี่ย ($\bar{E}$) และส่วนเบี่ยงเบนมาตรฐานของค่าเฉลี่ย ($SD_{\bar{E}}$) ความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ ในแต่ละกลุ่มการทดลองโดยทำการทดลองซ้ำ 40 ครั้ง

82 n.f.u.

กลุ่มทดลอง	$\bar{E}$	$SD_{\bar{E}}$	กลุ่มทดลอง	$\bar{E}$	$SD_{\bar{E}}$
c7s1	.399	.008	v7s1	.508	.011
c10s1	.376	.010	v10s1	.493	.012
c13s1	.363	.016	v13s1	.513	.013
c16s1	.373	.009	v16s1	.548	.011
c7s2	.462	.069	v7s2	.592	.017
c10s2	.370	.007	v10s2	.532	.052

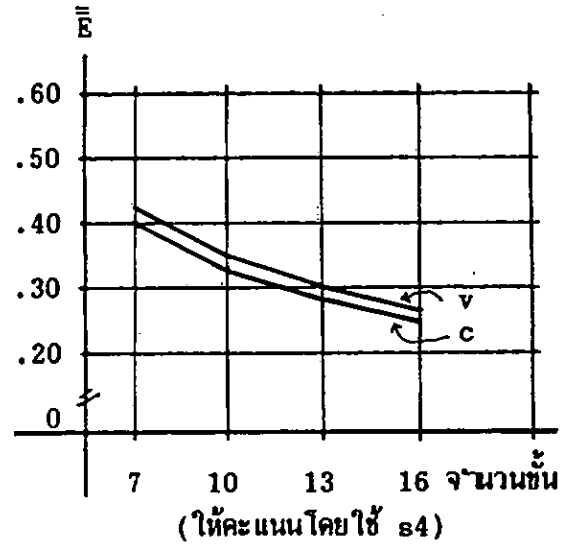
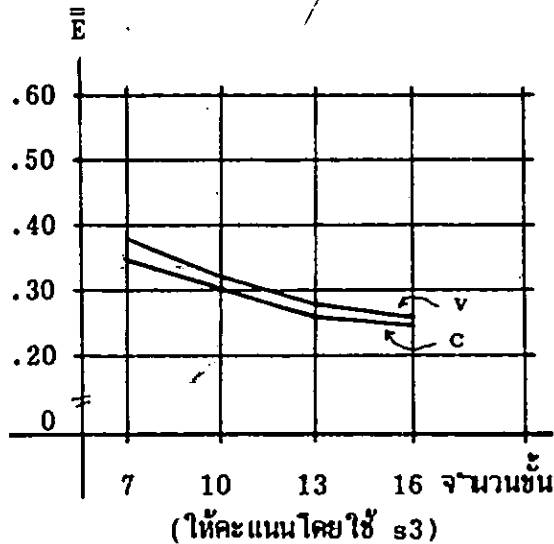
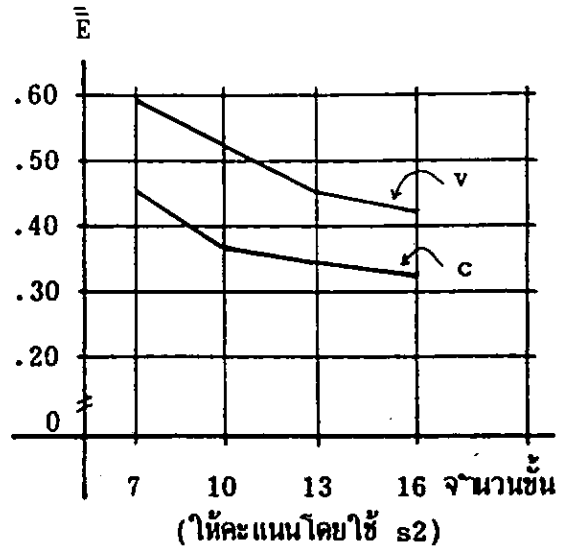
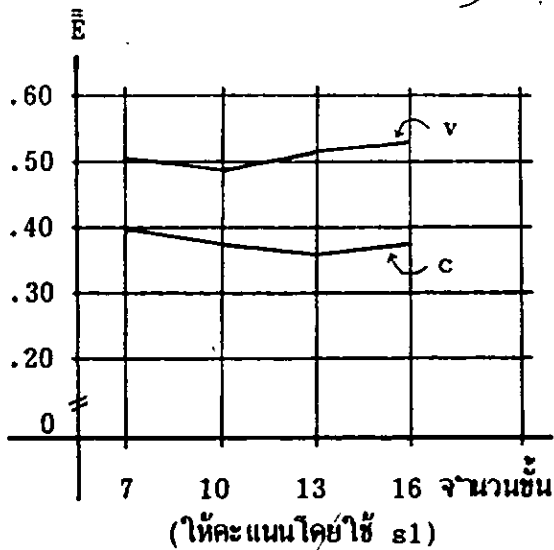
ตาราง 13 (ต่อ)

กลุ่มทดลอง	$\bar{E}$	$SD\bar{E}$	กลุ่มทดลอง	$\bar{E}$	$SD\bar{E}$
c13s2	.353	.016	v13s2	.469	.009
c16s2	.330	.006	v16s2	.431	.035
c7s3	.356	.009	v7s3	.381	.011
c10s3	.300	.008	v10s3	.321	.008
c13s3	.268	.004	v13s3	.280	.008
c16s3	.246	.005	v16s3	.258	.005
c7s4	.401	.011	v7s4	.425	.011
c10s4	.330	.008	v10s4	.350	.009
c13s4	.288	.008	v13s4	.303	.007
c16s4	.251	.007	v16s4	.271	.007

จากตาราง 13 จะเห็นว่า ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ ($\bar{E}$) มีค่าอยู่ระหว่าง .246 ถึง .592 และมีส่วนเบี่ยงเบนมาตรฐานของค่าเฉลี่ยความคลาดเคลื่อน ($SD\bar{E}$) มีค่าอยู่ระหว่าง .004 ถึง .069 โดยกลุ่มที่ทดสอบด้วยการทดสอบเทเลอร์รูปปริมาตรแบบขนาดชั้นคงที่ ที่มี 16 ชั้น ให้คะแนนโดยใช้วิธีของเบส์ (c16s3) จะให้ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ ($\bar{E}$) ต่ำสุด และกลุ่มที่ทดสอบด้วยการทดสอบเทเลอร์รูปปริมาตรแบบขนาดชั้นแปรผัน ที่มี 7 ชั้น โดยให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ถูกต้อง (v7s2) จะให้ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ ($\bar{E}$) สูงสุด

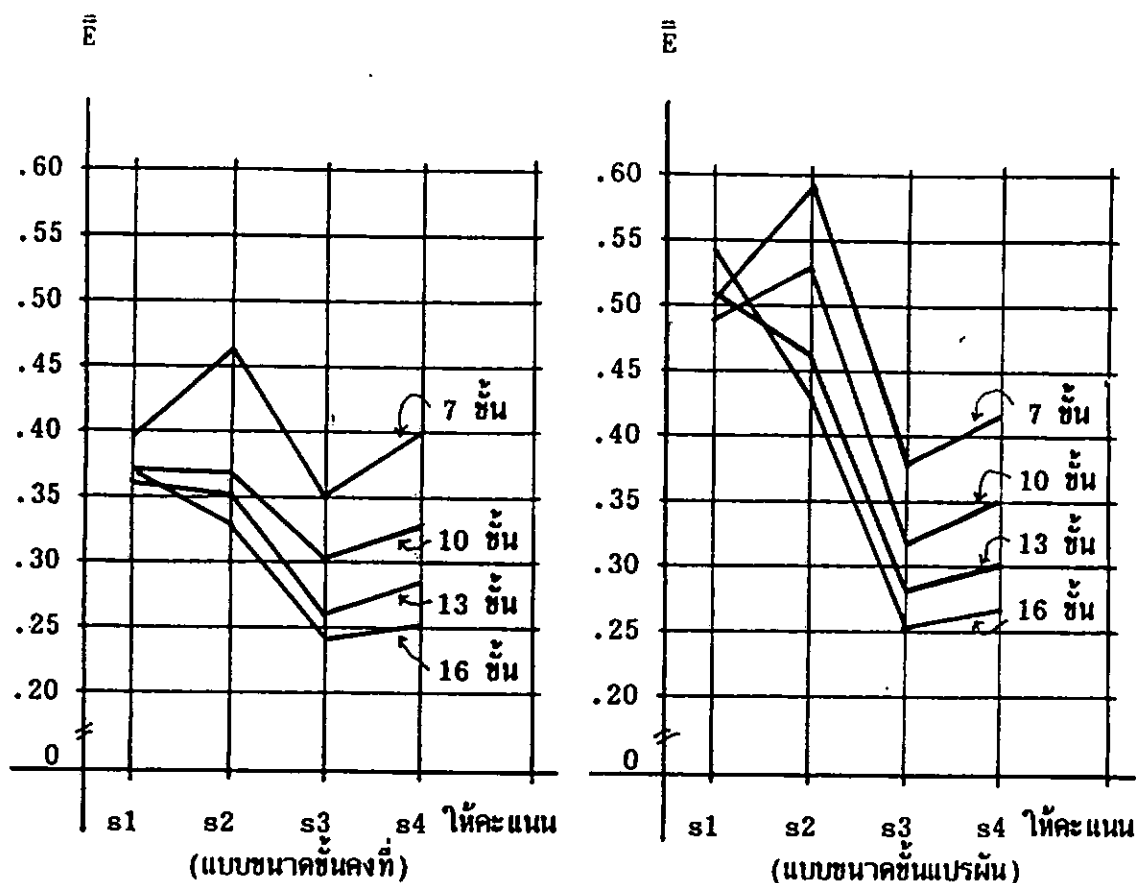
เมื่อนำค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ ($\bar{E}$) ในแต่ละกลุ่มการทดลอง มาเขียนกราฟ จะได้ดัง ภาพประกอบ 21 ถึง 23

- ① 1V
2 45° (ในแนวนอน)
3 45° (ในแนวตั้ง)



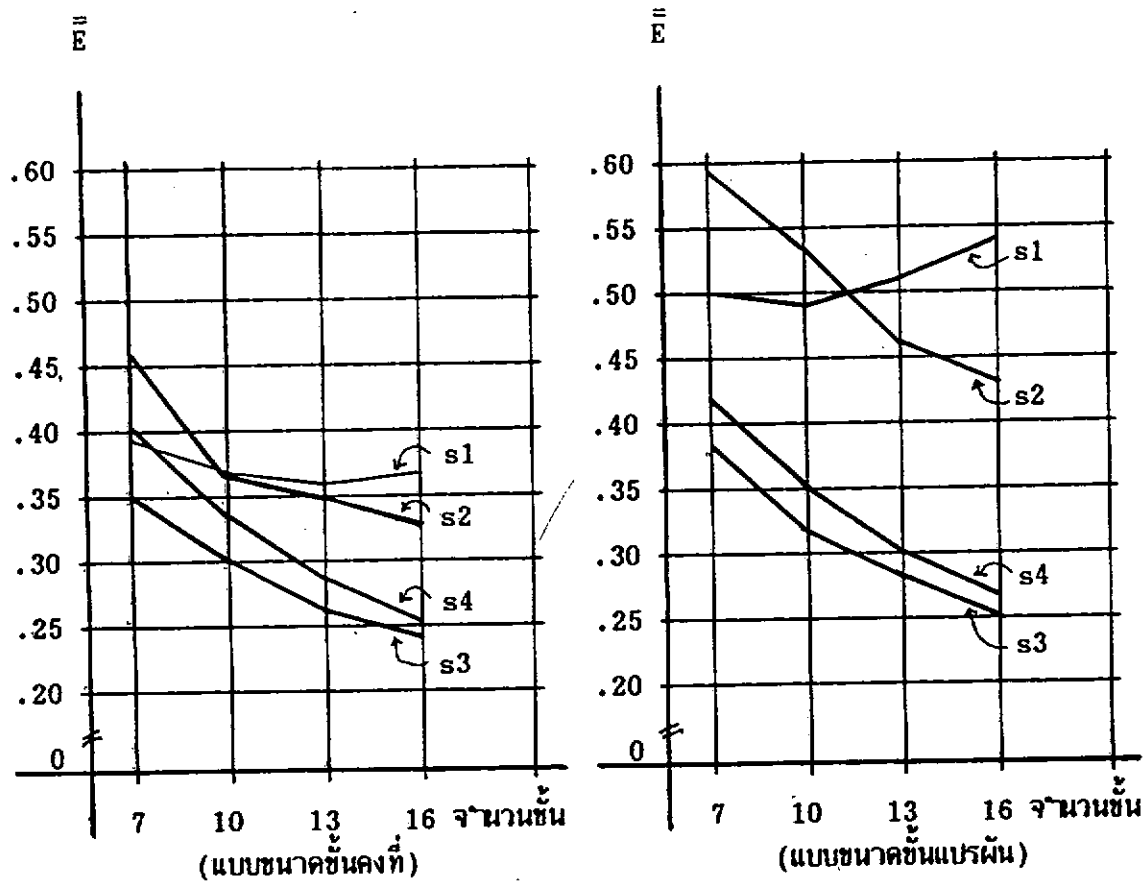
ภาพประกอบ 21 กราฟเปรียบเทียบค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ ($\bar{E}$) จากการทดสอบเทเลอรัปรีมิด ที่มีรูปแบบต่างกัน

จากภาพประกอบ 21 จะเห็นว่า ไม่ว่าจะใช้จำนวน 7, 10, 13 หรือ 16 ชั้น และ ไม่ว่าจะ ให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำได้ (s1) ให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทำได้ (s2) ให้คะแนนโดยใช้วิธีของเบส์ (s3) หรือให้คะแนนโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (s4) ก็ตาม การทดสอบเทเลอรัปรีมิดแบบขนาดชั้นคงที่ (c) จะให้ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ ($\bar{E}$) ต่ำกว่า การทดสอบเทเลอรัปรีมิดแบบขนาดชั้นแปรผัน (v) อยู่เสมอ



ภาพประกอบ 22 กราฟเปรียบเทียบค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ (E) จากการทดสอบเทเลอรัปรีมิด ที่มีจำนวนชั้นต่างกัน

จากภาพประกอบ 22 จะเห็นว่า การทดสอบเทเลอรัปรีมิด ไม่ว่าจะ ให้คะแนนค่าเฉลี่ยความยากของข้อที่ท้าทาย (s2) ให้คะแนนโดยใช้วิธีของเบส์ (s3) หรือให้คะแนนโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (s4) และ ไม่ว่าจะ เป็น การทดสอบเทเลอรัปรีมิดแบบขนาดชั้นคงที่หรือแบบขนาดชั้นแปรผันก็ตาม การทดสอบเทเลอรัปรีมิดที่มีจำนวน 16 ชั้น จะให้ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ (E) ต่ำสุด รองลงมาได้แก่ แบบที่มีจำนวน 13, 10 และ 7 ชั้น ตามลำดับ



ภาพประกอบ 23 กราฟเปรียบเทียบค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ ($\bar{E}$) จากการทดสอบเทเลอรัปรีมิด ที่ใช้วิธีการให้คะแนนที่ต่างกัน

จากภาพประกอบ 23 จะเห็นว่า การทดสอบเทเลอรัปรีมิด ไม่ว่าจะเป็น 7, 10, 13 หรือ 16 ชั้น และไม่ว่าจะเป็นแบบขนาดชั้นคงที่ หรือแบบขนาดชั้นแปรผันก็ตาม การให้คะแนนโดยใช้วิธีของเบส์ (s3) จะให้ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ ($\bar{E}$) ค่าที่ต่ำที่สุด รองลงมาได้แก่ การให้คะแนนโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (s4) และการให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ถูกต้อง (s2) ตามลำดับ

ตาราง 14 แสดงผลการวิเคราะห์ความแปรปรวนของค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ ($\bar{E}$) ในแต่ละกลุ่มการทดลองโดยทำการทดลองซ้ำ 40 ครั้ง

Source of Variation	Sum of Squares	df	Mean Square	F
Main Effects				
รูปแบบ	1.831	1	1.831	5118.755**
จำนวนชั้น	1.930	3	0.643	1798.186**
การให้คะแนน	5.543	3	1.848	5164.468**
2-Way Interaction				
รูปแบบxจำนวนชั้น	0.003	3	0.001	2.892*
รูปแบบxการให้คะแนน	1.041	3	0.347	970.186**
จำนวนชั้นxการให้คะแนน	0.724	9	0.080	224.870**
3-Way Interaction				
รูปแบบxจำนวนชั้นxการให้คะแนน	0.098	9	0.011	30.299**
Residual	0.446	1248	3.574×10^{-4}	-
Total	11.617	1279	0.009	-

* $p < .05$ ** $p < .01$

จากตาราง 14 เมื่อพิจารณาจากการทดสอบหลัก (Main Effects) จะเห็นว่า ไม่ว่าจะเป็นรูปแบบ จำนวนชั้น หรือวิธีการให้คะแนน มีนัยสำคัญทางสถิติที่ระดับ .01 นั่นคือ การใช้รูปแบบของการทดสอบเทเลอร์รูปปริมาตรที่ต่างกัน การใช้จำนวนชั้นในการทดสอบที่ไม่เท่ากัน หรือวิธีการให้คะแนนที่ต่างกัน จะส่งผลต่อค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่าความสามารถ

ของผู้สอบ ($\bar{E}$)

เมื่อพิจารณาจากการทดสอบปฏิสัมพันธ์สองทาง (2-Way Interaction) จะเห็นว่า ปฏิสัมพันธ์ระหว่างรูปแบบจำนวนชั้น มีนัยสำคัญที่ระดับ .05 ปฏิสัมพันธ์ระหว่างรูปแบบวิธีการ ให้คะแนน และจำนวนชั้นวิธีการให้คะแนน มีนัยสำคัญทางสถิติที่ระดับ .01 นั่นคือ การใช้รูปแบบกับจำนวนชั้นที่ต่างกัน รูปแบบกับวิธีการให้คะแนนที่ต่างกัน หรือ จำนวนชั้นกับวิธีการให้คะแนนที่ต่างกัน จะร่วมกันส่งผลต่อค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ

เมื่อพิจารณาจากการทดสอบปฏิสัมพันธ์สามทาง (3-Way Interaction) จะเห็นว่า ปฏิสัมพันธ์ระหว่าง รูปแบบจำนวนชั้นวิธีการให้คะแนน มีนัยสำคัญทางสถิติที่ระดับ .01 นั่นคือ การใช้รูปแบบ จำนวนชั้น และวิธีการให้คะแนนที่ต่างกัน จะร่วมกันส่งผลต่อค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ ($\bar{E}$)

จากภาพประกอบ 21 ปฏิสัมพันธ์ระหว่างรูปแบบจำนวนชั้น เมื่อแยกพิจารณาตามตัวแปรที่สามคือวิธีการให้คะแนน จะเห็นว่าเส้นกราฟทั้งหมดไม่ตัดกันและส่วนใหญ่ไม่ขนานกัน นั่นคือ ปฏิสัมพันธ์ดังกล่าว เป็นแบบ Ordinal ทั้งหมด

จากภาพประกอบ 22 ปฏิสัมพันธ์ระหว่างจำนวนชั้นวิธีการให้คะแนน เมื่อแยกพิจารณาตามตัวแปรที่สามคือรูปแบบ จะเห็นว่าเส้นกราฟส่วนใหญ่ไม่ตัดกันและไม่ขนานกัน นอกจากการให้คะแนนโดยใช้ s1 ที่ทำให้เส้นกราฟตัดกัน นั่นคือ ปฏิสัมพันธ์ดังกล่าว ส่วนใหญ่เป็นแบบ Ordinal นอกจากการให้คะแนนโดยใช้ s1 ที่ทำให้เกิดปฏิสัมพันธ์แบบ Disordinal

จากภาพประกอบ 23 ปฏิสัมพันธ์ระหว่างวิธีการให้คะแนนจำนวนชั้น เมื่อแยกพิจารณาตามตัวแปรที่สามคือรูปแบบ จะเห็นว่าเส้นกราฟส่วนใหญ่ไม่ตัดกันและไม่ขนานกัน นอกจากการให้คะแนนโดยใช้ s1 ที่ทำให้เส้นกราฟตัดกัน นั่นคือ ปฏิสัมพันธ์ดังกล่าว ส่วนใหญ่เป็นแบบ Ordinal นอกจากการให้คะแนนโดยใช้ s1 ที่ทำให้เกิดปฏิสัมพันธ์แบบ Disordinal

อนึ่ง สำหรับการวิจัยครั้งนี้มุ่งศึกษาประเด็นสำคัญที่ ปฏิสัมพันธ์สามทาง (3-Way Interaction) ทั้งนี้เพราะในการทดสอบเทเลอร์รูปปริมาตรแต่ละครั้งนั้น จะต้องใช้ตัวแปรทั้งสามนี้ คือ รูปแบบ จำนวนชั้น และวิธีการให้คะแนน พร้อมกันไปทีเดียวทั้ง 3 ตัวแปร จะแยกใช้ตัวแปร

ใดตัวแปรหนึ่งใด ๆ ไม่ได้ และเพื่อให้สามารถตอบวัตถุประสงค์ของการวิจัยในข้อที่ 1 และทดสอบสมมติฐานที่ตั้งไว้ จึงได้ทดสอบความแตกต่างของ ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ (E) ที่เกิดจากปฏิสัมพันธ์ระหว่างตัวแปรทั้งสามนี้ โดยทำการทดสอบเปรียบเทียบเป็นรายคู่ โดยใช้วิธีของ เชฟเฟ่ (Scheffe') ปรากฏผล ดังแสดงในตาราง 15

ตาราง 15 แสดงผลการทดสอบความแตกต่างของ ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่าความสามารถ
ของผู้สอบ (E) เป็นรายคู่ โดยใช้วิธีของ เชฟเฟ่ (Scheffe')

กลุ่ม	ค่าเฉลี่ย (E)	c16s3 .246	c16s4 .251	v16s3 .258	c13s3 .268	v16s4 .271	v13s3 .280	c13s4 .288	c10s3 .300	v13s4 .303	v10s3 .321	c16s2 .330	c10s4 .330	v10s4 .350	c13s2 .353	c7s3 .356	c13s1 .363
c16s3	.246	-	.005	.012	.022	.025	.034	.042	.054	.057	.075	.084	.084	.104	.107	.110	.117
c16s4	.251		-	.007	.017	.020	.029	.037	.049	.052	.070	.079	.079	.099	.102	.105	.112
v16s3	.258			-	.010	.013	.022	.030	.042	.045	.063	.072	.072	.092	.095	.098	.105
c13s3	.268				-	.003	.012	.020	.032	.035	.053	.062	.062	.082	.085	.088	.095
v16s4	.271					-	.009	.017	.029	.032	.050	.059	.059	.079	.082	.085	.092
v13s3	.280						-	.008	.020	.023	.041	.050	.050	.070	.073	.076	.083
c13s4	.288							-	.012	.015	.033	.042	.042	.062	.065	.068	.075
c10s3	.300								-	.003	.021	.030	.030	.050	.053	.056	.063
v13s4	.303									-	.018	.027	.027	.047	.050	.053	.060
v10s3	.321										-	.009	.009	.029	.032	.035	.042
c16s2	.330											-	.000	.020	.023	.026	.033
c10s4	.330												-	.020	.023	.026	.033
v10s4	.350													-	.003	.006	.013
c13s2	.353														-	.003	.010
c7s3	.356															-	.007
c13s1	.363																-

* p < .05 ** p < .01

ตาราง 15 (ต่อ)

กลุ่ม	ค่าเฉลี่ย ($\bar{E}$)	c10s2	c16s1	c10s1	v7s3	c7s1	c7s4	v7s4	v16s2	c7s2	v13s2	v10s1	v7s1	v13s1	v10s2	v16s1	v7s2
c16s3	.246	.124	.127	.130	.135	.153	.155	.179	.185	.216	.223	.247	.262	.267	.286	.302	.346
c16s4	.251	.119	.122	.125	.130	.148	.150	.174	.180	.211	.218	.242	.257	.262	.281	.297	.341
v16s3	.258	.112	.115	.118	.123	.141	.143	.167	.173	.204	.211	.235	.250	.255	.274	.290	.334
c13s3	.268	.102	.105	.108	.113	.131	.133	.157	.163	.194	.201	.225	.240	.245	.264	.280	.324
v16s4	.271	.099	.102	.105	.110	.128	.130	.154	.160	.191	.198	.222	.237	.242	.261	.277	.321
v13s3	.280	.090	.093	.096	.101	.119	.121	.145	.151	.182	.189	.213	.228	.233	.252	.268	.312
c13s4	.288	.082	.085	.088	.093	.111	.113	.137	.143	.174	.181	.205	.220	.225	.244	.260	.304
c10s3	.300	.070	.073	.076	.081	.099	.101	.125	.131	.162	.169	.193	.208	.213	.232	.248	.292
v13s4	.303	.067	.070	.073	.078	.096	.098	.122	.128	.159	.166	.190	.205	.210	.229	.245	.289
v10s3	.321	.049	.052	.055	.060	.078	.080	.104	.110	.141	.148	.172	.187	.192	.211	.227	.271
c16s2	.330	.040	.043	.046	.051	.069	.071	.095	.101	.132	.139	.163	.178	.183	.202	.218	.262
c10s4	.330	.040	.043	.046	.051	.069	.071	.095	.101	.132	.139	.163	.178	.183	.202	.218	.262
v10s4	.350	.020	.023	.026	.031	.049	.051	.075	.081	.112	.119	.143	.158	.163	.182	.198	.242
c13s2	.353	.017	.020	.023	.028	.046	.048	.072	.078	.109	.116	.140	.155	.160	.179	.195	.239
c7s3	.356	.014	.017	.020	.025	.043	.045	.069	.075	.106	.113	.137	.152	.157	.176	.192	.236
c13s1	.363	.007	.010	.013	.018	.036	.038	.062	.068	.099	.106	.130	.145	.150	.169	.185	.229

* p < .05 ** p < .01

กลุ่ม	ค่าเฉลี่ย ($\bar{E}$)	c10s2 .370	c16s1 .373	c10s1 .376	v7s3 .381	c7s1 .399	c7s4 .401	v7s4 .425	v16s2 .431	c7s2 .462	v13s2 .469	v10s1 .493	v7s1 .508	v13s1 .513	v10s2 .532	v16s1 .548	v7s2 .592
c10s2	.370	-	.003	.006	.011	.029	.031	.055	.061	.092	.099	.123	.138	.143	.162	.178	.222
c16s1	.373	-	-	.003	.008	.026	.028	.052	.058	.089	.096	.120	.135	.140	.159	.175	.219
c10s1	.376	-	-	.005	.023	.025	.025	.049	.055	.086	.093	.117	.132	.137	.156	.172	.216
v7s3	.381	-	-	.018	.020	.044	.050	.081	.088	.112	.127	.132	.151	.167	.211	.222	.222
c7s1	.399	-	-	.002	.026	.032	.063	.070	.094	.109	.114	.133	.149	.193	.222	.222	.222
c7s4	.401	-	-	.024	.030	.061	.068	.092	.107	.112	.131	.147	.191	.222	.222	.222	.222
v7s4	.425	-	-	.006	.037	.044	.068	.083	.088	.107	.123	.167	.222	.222	.222	.222	.222
v16s2	.431	-	-	.031	.038	.062	.077	.082	.101	.117	.161	.222	.222	.222	.222	.222	.222
c7s2	.462	-	-	.007	.031	.046	.051	.070	.086	.130	.222	.222	.222	.222	.222	.222	.222
v13s2	.469	-	-	.024	.039	.044	.063	.079	.123	.222	.222	.222	.222	.222	.222	.222	.222
v10s1	.493	-	-	.015	.020	.039	.055	.099	.222	.222	.222	.222	.222	.222	.222	.222	.222
v7s1	.508	-	-	.005	.024	.040	.084	.222	.222	.222	.222	.222	.222	.222	.222	.222	.222
v13s1	.513	-	-	.019	.035	.079	.222	.222	.222	.222	.222	.222	.222	.222	.222	.222	.222
v10s2	.532	-	-	.016	.060	.222	.222	.222	.222	.222	.222	.222	.222	.222	.222	.222	.222
v16s1	.548	-	-	.044	.222	.222	.222	.222	.222	.222	.222	.222	.222	.222	.222	.222	.222
v7s2	.592	-	-	.222	.222	.222	.222	.222	.222	.222	.222	.222	.222	.222	.222	.222	.222

* p < .05 ** p < .01

จากตาราง 15 เป็นการเปรียบเทียบค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ (E) จากการทดสอบเทเลอร์รูปปริมาตรที่มี รูปแบบ จำนวนชั้น และวิธีการให้คะแนน ที่แตกต่างกัน โดยการเปรียบเทียบเป็นรายคู่ ซึ่งมีกลุ่มตัวอย่างทั้งหมด 32 กลุ่ม จะเห็นว่า

หากที่ให้ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ (E) ค่าสุด หรือ มีประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบสูงสุด และรองลงมา ใน 3 อันดับแรก ได้แก่

อันดับที่ 1 ได้แก่ กลุ่ม c16s3 ซึ่งมีค่า E ค่าที่ต่ำสุด เมื่อทดสอบเปรียบเทียบกับกลุ่มอื่น ๆ แล้ว ปรากฏว่า

(1) มีค่า E ไม่แตกต่างจากกลุ่ม c16s4, v16s3, c13s3, และ v16s4

(2) มีค่า E ต่ำกว่ากลุ่มอื่น ๆ ที่เหลือ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

อันดับที่ 2 ได้แก่ กลุ่ม c16s4 เมื่อทดสอบเปรียบเทียบกับกลุ่มอื่น ๆ แล้ว ปรากฏว่า

(1) มีค่า E ไม่แตกต่างจากกลุ่ม v16s3, c13s3 และ v16s4

(2) มีค่า E ต่ำกว่ากลุ่ม v13s3 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และ

(3) มีค่า E ต่ำกว่ากลุ่มอื่น ๆ ที่เหลือ นอกจากกลุ่มที่อยู่ในอันดับที่ 1 อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

อันดับที่ 3 ได้แก่ กลุ่ม v16s3 เมื่อทดสอบเปรียบเทียบกับกลุ่มอื่น ๆ แล้ว ปรากฏว่า

(1) มีค่า E ไม่แตกต่างจากกลุ่ม c13s3, v16s4 และ v13s3

(2) มีค่า E ต่ำกว่ากลุ่ม c13s4 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และ

(3) มีค่า E ต่ำกว่ากลุ่มอื่น ๆ ที่เหลือ นอกจากกลุ่มที่อยู่ในอันดับที่ 1 และ 2 อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

หากที่ให้ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ (E) สูงสุด หรือ มีประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบค่าสุด และรองลงมา ใน 3 อันดับแรก ได้แก่

อันดับที่ 1 ได้แก่ กลุ่ม v7s2 ซึ่งมีค่า E สูงที่สุด เมื่อทดสอบเปรียบเทียบกับกลุ่มอื่น ๆ แล้ว ปรากฏว่า มีค่า E สูงกว่ากลุ่มอื่น ๆ ที่เหลือ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

อันดับที่ 2 ได้แก่ กลุ่ม v16s1 เมื่อทดสอบเปรียบเทียบกับกลุ่มอื่น ๆ แล้ว ปรากฏว่า

- (1) มีค่า E ไม่แตกต่างจากกลุ่ม v10s2 และ
- (2) มีค่า E สูงกว่ากลุ่มอื่น ๆ ที่เหลือ นอกจากกลุ่มที่อยู่ในอันดับที่ 1 อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

อันดับที่ 3 ได้แก่ กลุ่ม v10s2 เมื่อทดสอบเปรียบเทียบกับกลุ่มอื่น ๆ แล้ว ปรากฏว่า

- (1) มีค่า E ไม่แตกต่างจากกลุ่ม v13s1 และ v7s1
- (2) มีค่า E สูงกว่ากลุ่มอื่น ๆ ที่เหลือ นอกจากกลุ่มที่อยู่ในอันดับที่ 1 และ 2 อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

จากผลการวิจัยที่พบจะเห็นว่า กลุ่มที่ทดสอบด้วยการทดสอบเทเลอร์รูปปริมาตรแบบขนาดชั้นแปรผัน ที่มี 16 ชั้น ให้คะแนนโดยวิธีวิธีของเบส์ (v16s3) ถึงแม้จะมีค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ (E) ค่า อยู่ในอันดับ 3 รองจากกลุ่ม c16s3 และ c16s4 ก็ตาม แต่เมื่อทดสอบเปรียบเทียบค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบทั้ง 3 ค่านี้ แล้ว ปรากฏว่า ไม่แตกต่างกัน นั่นคือการทดสอบเทเลอร์รูปปริมาตรแบบขนาดชั้นแปรผันที่มี 16 ชั้น ให้คะแนนโดยวิธีวิธีของเบส์ (v16s3) จะให้ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบจัดอยู่ในพวกที่ให้ค่าต่ำสุด ซึ่งก็สอดคล้องกับสมมุติฐานที่ตั้งไว้

สำหรับกลุ่มที่ทดสอบด้วยการทดสอบเทเลอร์รูปปริมาตรแบบขนาดชั้นคงที่ ที่มี 7 ชั้น ให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ถูกต้อง (c7s1) ผลการศึกษาพบว่า ไม่ใช่กลุ่มที่ให้ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ (E) สูงสุด ซึ่งไม่สอดคล้องกับวัตถุประสงค์ที่ตั้งไว้

2. ผลการศึกษาและเปรียบเทียบค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ (SSE) จาการทดสอบเทเลอร์รูปปริมาตรที่มี รูปแบบ จำนวนชั้น และวิธีการให้คะแนน ที่แตกต่างกัน

2.1 การศึกษาค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ($\overline{SEE}$) ปรากฏผล ดังแสดงใน ตาราง 16 และภาพประกอบ 24 ถึง 26

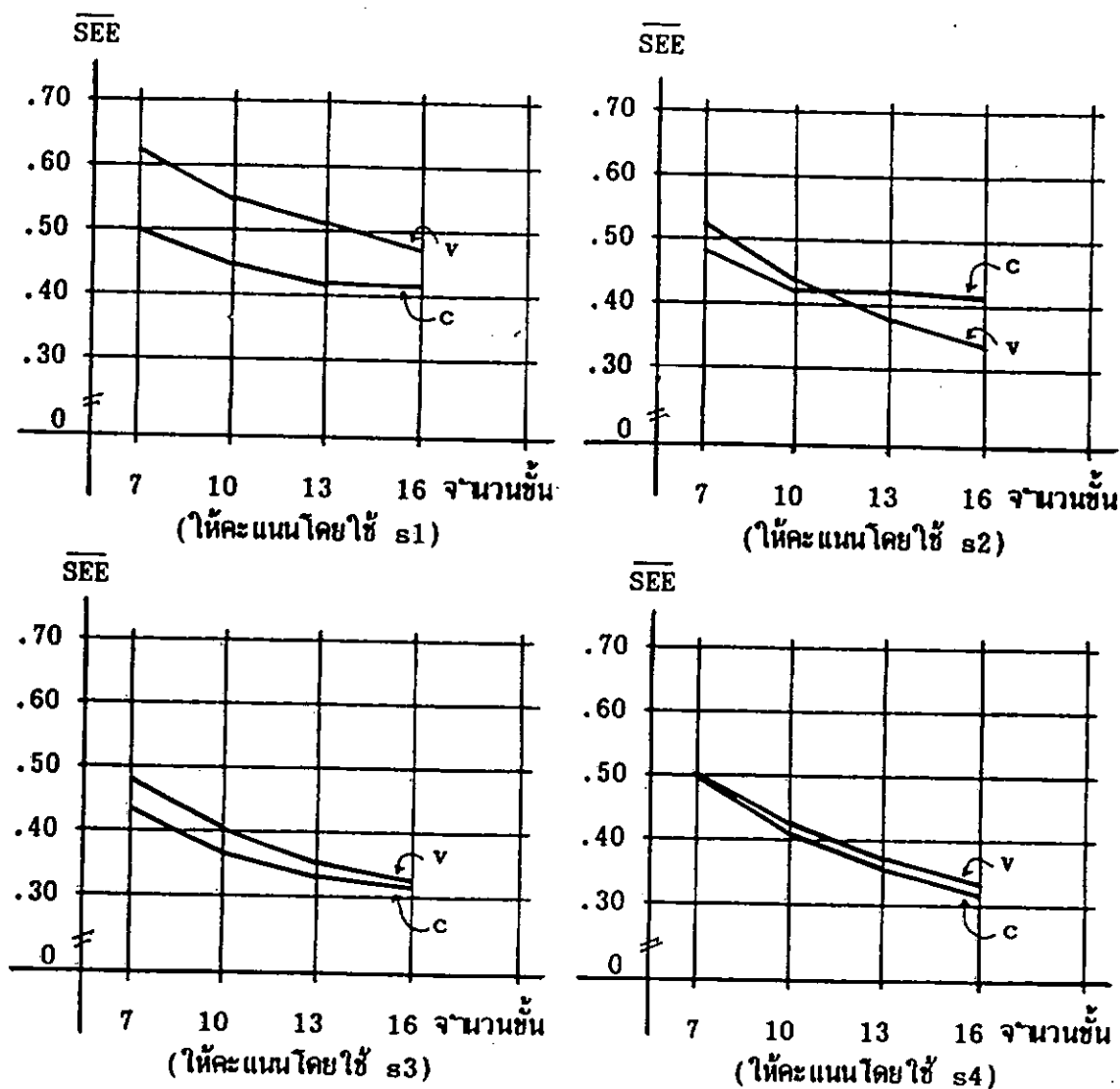
2.2 การเปรียบเทียบค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ (SEE) จากการทดสอบเทเลอร์รูปปริมาตรที่มี รูปแบบ จำนวนชั้น และวิธีการให้คะแนน ที่แตกต่างกัน โดยการวิเคราะห์ความแปรปรวน ปรากฏผล ดังแสดงใน ตาราง 17 และ ทดสอบเปรียบเทียบค่าเฉลี่ยเป็นรายคู่ ดังแสดงใน ตาราง 18

ตาราง 16 แสดงค่าเฉลี่ย ($\overline{SEE}$) และส่วนเบี่ยงเบนมาตรฐาน (SD_{SEE}) ของค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ในแต่ละกลุ่มการทดลองโดยทำการทดลองซ้ำ 40 ครั้ง

กลุ่มทดลอง	$\overline{SEE}$	SD_{SEE}	กลุ่มทดลอง	$\overline{SEE}$	SD_{SEE}
c7s1	.498	.010	v7s1	.631	.013
c10s1	.450	.013	v10s1	.551	.011
c13s1	.423	.009	v13s1	.512	.013
c16s1	.427	.012	v16s1	.470	.010
c7s2	.480	.014	v7s2	.535	.012
c10s2	.426	.011	v10s2	.441	.018
c13s2	.421	.030	v13s2	.381	.008
c16s2	.411	.094	v16s2	.340	.007
c7s3	.448	.011	v7s3	.480	.013
c10s3	.378	.009	v10s3	.406	.018
c13s3	.338	.005	v13s3	.352	.009
c16s3	.310	.006	v16s3	.324	.007
c7s4	.508	.013	v7s4	.501	.012
c10s4	.418	.010	v10s4	.428	.009
c13s4	.365	.009	v13s4	.376	.010
c16s4	.318	.008	v16s4	.336	.006

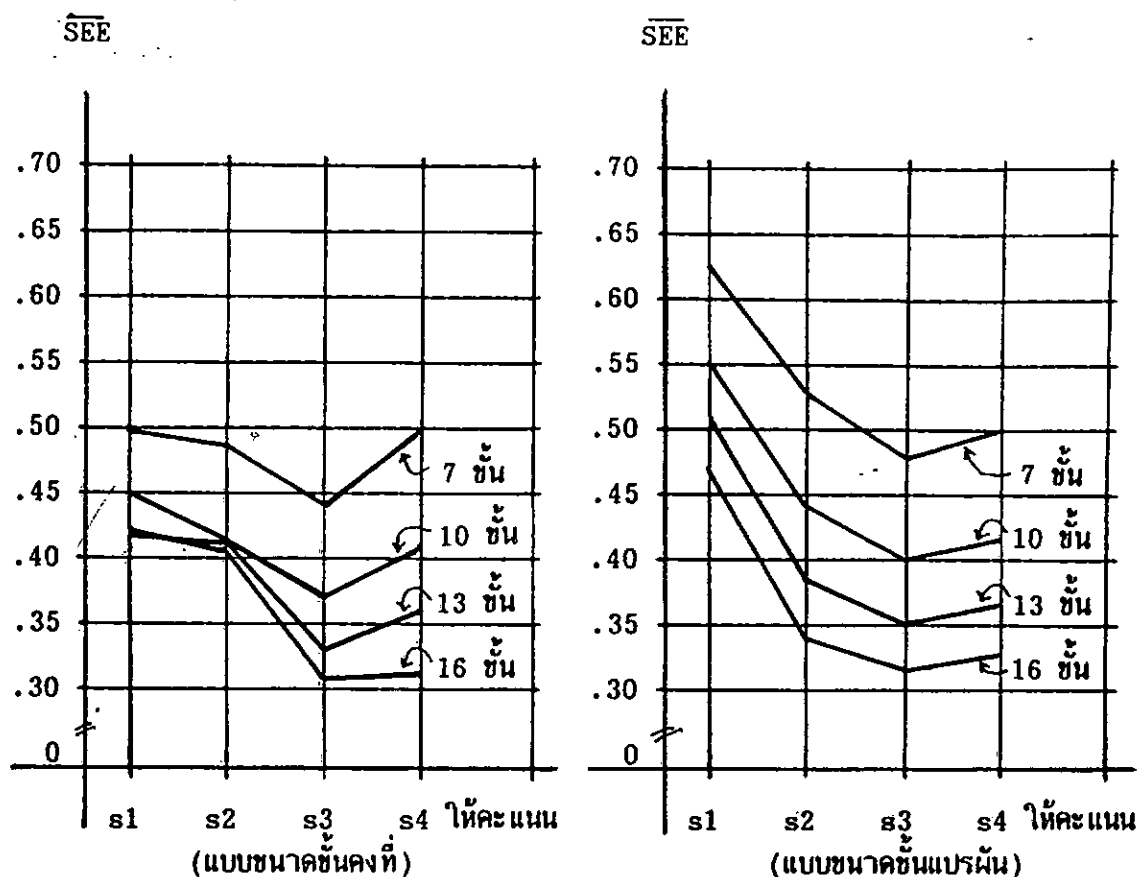
จากตาราง 16 จะเห็นว่าค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ($\overline{SEE}$) มีค่าอยู่ระหว่าง .310 ถึง .631 และมีส่วนเบี่ยงเบนมาตรฐานของค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ (SD_{SEE}) มีค่าอยู่ระหว่าง .005 ถึง .094 โดยกลุ่มที่ทดสอบด้วยการทดสอบเทเลอร์รูปปริมาตรแบบขนาดชั้นคงที่ ที่มี 16 ชั้น ให้คะแนนโดยใช้วิธีของเบส์ (c16s3) จะให้ค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ($\overline{SEE}$) ต่ำสุด และกลุ่มที่ทดสอบด้วยการทดสอบเทเลอร์รูปปริมาตรแบบขนาดชั้นแปรผันที่มี 7 ชั้น ให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำได้ (v7s1) จะให้ค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ($\overline{SEE}$) สูงสุด

เมื่อนำค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ($\overline{SEE}$) ในแต่ละกลุ่มการทดลอง มาเขียนกราฟ จะได้ดัง ภาพประกอบ 24 ถึง 26



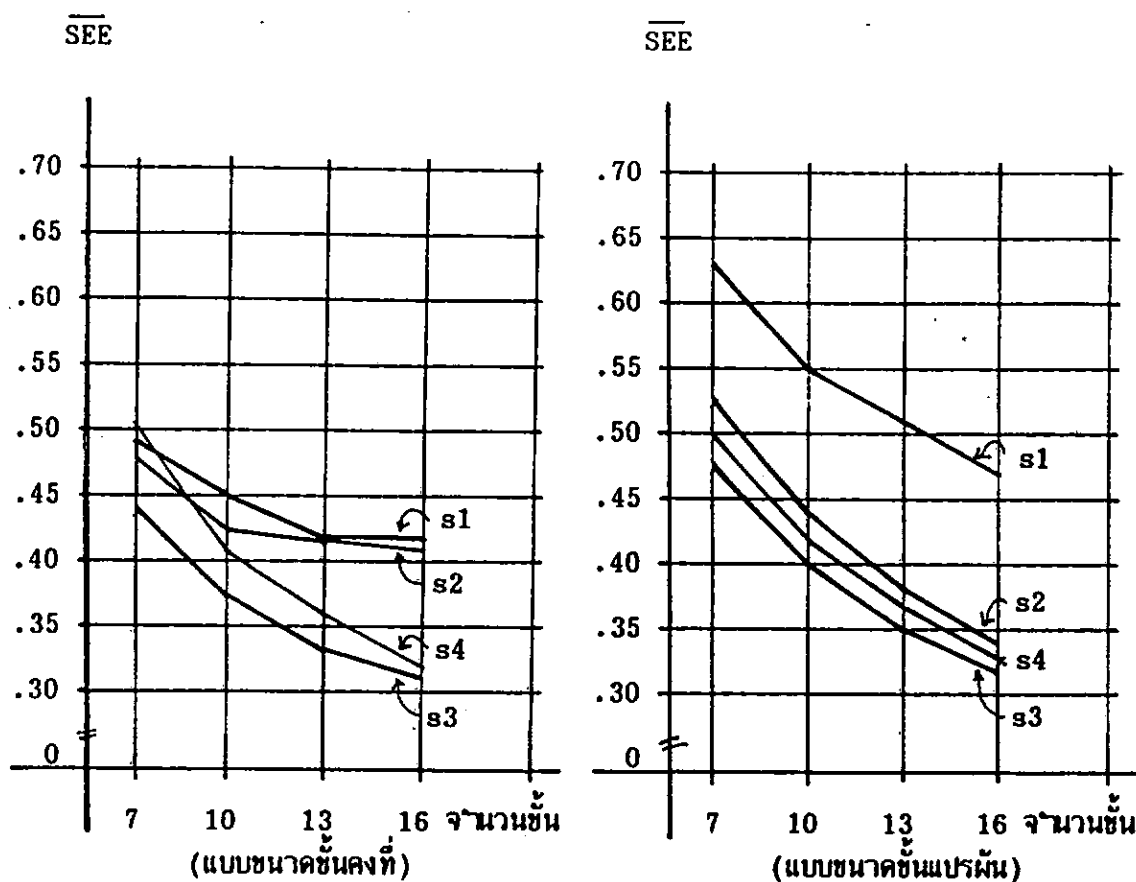
ภาพประกอบ 24 กราฟเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ($\overline{SEE}$) จากการทดสอบเทเลอรัปรีรามิด ที่มีรูปแบบต่างกัน

จากภาพประกอบ 24 จะเห็นว่า การทดสอบเทเลอรัปรีรามิด ที่มีตั้งแต่ 10 ชั้นขึ้นไป ไม่ว่าจะให้คะแนน ความค่าความยากของข้อที่ยากที่สุดที่ทำได้ (s1) ให้คะแนนโดยวิธีของเบส์ (s3) หรือ ให้คะแนนโดยวิธีความเป็นไปได้สูงสุดแบบเงื่อนไข (s4) ก็ตาม การทดสอบเทเลอรัปรีรามิดแบบขนาดชั้นคงที่ (c) จะให้ค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ($\overline{SEE}$) ต่ำกว่าการทดสอบเทเลอรัปรีรามิดแบบขนาดชั้นแปรผัน (v)



ภาพประกอบ 25 กราฟเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ($\overline{SEE}$) จากการทดสอบเทเลอ์รูปปิรามิด ที่มีจำนวนชั้นต่างกัน

จากภาพประกอบ 25 จะเห็นว่า การทดสอบเทเลอ์รูปปิรามิด ไม่ว่าจะ ให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทำถูก (s2) ให้คะแนนโดยใช้วิธีของเบส์ (s3) หรือให้คะแนนโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (s4) และ ไม่ว่าจะ เป็น การทดสอบเทเลอ์รูปปิรามิดแบบขนาดชั้นคงที่หรือแบบขนาดชั้นแปรผัน ก็ตาม การทดสอบเทเลอ์รูปปิรามิดที่มี จำนวน 16 ชั้น จะให้ค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ ($\overline{SEE}$) ต่ำสุด รองลงมา ได้แก่ แบบที่มีจำนวน 13, 10 และ 7 ชั้น ตามลำดับ



ภาพประกอบ 26 กราฟเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ($\overline{SEE}$) จากการทดสอบเทเลอร์รูปปิรามิด ที่ใช้วิธีการให้คะแนนที่ต่างกัน

จากภาพประกอบ 26 จะเห็นว่า การทดสอบเทเลอร์รูปปิรามิด ที่มีตั้งแต่ 10 ชั้นขึ้นไป ไม่ว่าจะเป็นแบบขนาดชั้นคงที่ หรือ แบบขนาดชั้นแปรผัน ก็ตาม การให้คะแนนโดยใช้วิธีของเบส์ (s3) จะให้ค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ($\overline{SEE}$) ต่ำที่สุด รองลงมาได้แก่ การให้คะแนนโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (s4) การให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ถูกต้อง (s2) และการให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ถูกต้อง (s1) ตามลำดับ

ตาราง 17 แสดงผลการวิเคราะห์ความแปรปรวนของค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ (SEE) ในแต่ละกลุ่มการทดลองโดยทำการทดลองซ้ำ 40 ครั้ง

Source of Variation	Sum of Squares	df	Mean Square	F
Main Effects				
รูปแบบ	0.249	1	0.249	592.353**
จำนวนชั้น	3.720	3	1.240	2955.236**
การให้คะแนน	2.348	3	0.783	1865.348**
2-Way Interaction				
รูปแบบxจำนวนชั้น	0.125	3	0.042	99.134**
รูปแบบxการให้คะแนน	0.475	3	0.158	377.623**
จำนวนชั้นxการให้คะแนน	0.096	9	0.011	25.531**
3-Way Interaction				
รูปแบบxจำนวนชั้นxการให้คะแนน	0.162	9	0.018	42.769**
Residual	0.524	1248	4.199×10^{-4}	-
Total	7.699	1279	0.006	-

** p < .01

จากตาราง 17 เมื่อพิจารณาจากการทดสอบหลัก (Main Effects) จะเห็นว่า ไม่ว่าจะเป็นรูปแบบ จำนวนชั้น หรือวิธีการให้คะแนน มีนัยสำคัญทางสถิติที่ระดับ .01 นั่นคือ การใช้รูปแบบการทดสอบเทเลอร์รูปปริมาตรที่ต่างกัน การใช้จำนวนชั้นในการทดสอบที่ไม่เท่ากัน หรือวิธี

การให้คะแนนที่ต่างกัน จะส่งผลต่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ (SEE)

เมื่อพิจารณาจากการทดสอบปฏิสัมพันธ์สองทาง (2-Way Interaction) จะเห็นว่า ปฏิสัมพันธ์ระหว่าง รูปแบบxจำนวนชั้น รูปแบบxวิธีการให้คะแนน และจำนวนชั้นxวิธีการให้คะแนน มีนัยสำคัญทางสถิติที่ระดับ .01 นั่นคือ การใช้รูปแบบกับจำนวนชั้นที่ต่างกัน รูปแบบกับวิธีการให้คะแนนที่ต่างกัน หรือ จำนวนชั้นกับวิธีการให้คะแนนที่ต่างกัน จะร่วมกันส่งผลต่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ (SEE)

เมื่อพิจารณาจากการทดสอบปฏิสัมพันธ์สามทาง (3-Way Interaction) จะเห็นว่า ปฏิสัมพันธ์ระหว่าง รูปแบบxจำนวนชั้นxวิธีการให้คะแนน มีนัยสำคัญทางสถิติที่ระดับ .01 นั่นคือ การใช้รูปแบบ จำนวนชั้น และวิธีการให้คะแนนที่ต่างกัน จะร่วมกันส่งผลต่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ (SEE)

จากภาพประกอบ 24 ปฏิสัมพันธ์ระหว่างรูปแบบxจำนวนชั้น เมื่อแยกพิจารณาตามตัวแปรที่สามคือวิธีการให้คะแนน จะเห็นว่าเส้นกราฟส่วนใหญ่ไม่ตัดกันและไม่ขนานกัน นอกจากการให้คะแนนโดยใช้ s2 กับการใช้จำนวน 7 ชั้น ที่ทำให้เส้นกราฟตัดกัน นั่นคือ ปฏิสัมพันธ์ดังกล่าวส่วนใหญ่เป็นแบบ Ordinal นอกจากการให้คะแนนโดยใช้ s2 และการใช้จำนวน 7 ชั้น ที่ทำให้เกิดปฏิสัมพันธ์แบบ Disordinal

จากภาพประกอบ 25 ปฏิสัมพันธ์ระหว่างจำนวนชั้นxวิธีการให้คะแนน เมื่อแยกพิจารณาตามตัวแปรที่สามคือรูปแบบ จะเห็นว่าเส้นกราฟส่วนใหญ่ไม่ตัดกันและไม่ขนานกัน นอกจากการให้คะแนนโดยใช้ s1 ที่ทำให้เส้นกราฟตัดกัน นั่นคือ ปฏิสัมพันธ์ดังกล่าว ส่วนใหญ่เป็นแบบ Ordinal นอกจากการให้คะแนนโดยใช้ s1 ที่ทำให้เกิดปฏิสัมพันธ์แบบ Disordinal

จากภาพประกอบ 26 ปฏิสัมพันธ์ระหว่างวิธีการให้คะแนนxจำนวนชั้น เมื่อแยกพิจารณาตามตัวแปรที่สามคือรูปแบบ จะเห็นว่าเส้นกราฟส่วนใหญ่ไม่ตัดกันและไม่ขนานกัน นอกจากจำนวน 7 ชั้น ที่ทำให้เส้นกราฟตัดกัน นั่นคือ ปฏิสัมพันธ์ดังกล่าว ส่วนใหญ่เป็นแบบ Ordinal นอกจากจำนวน 7 ชั้น ที่ทำให้เกิดปฏิสัมพันธ์แบบ Disordinal

อนึ่ง สำหรับการวิจัยครั้งนี้มุ่งศึกษาประเด็นสำคัญที่ ปฏิสัมพันธ์สามทาง (3-Way Interaction) ทั้งนี้เพราะในการทดสอบทฤษฎีรูบิรามิดแต่ละครั้งนั้น จะต้องใช้ตัวแปรทั้งสามนี้ คือ รูปแบบ จำนวนชั้น และวิธีการให้คะแนน พร้อมกันไปทีเดียวทั้ง 3 ตัวแปร จะแยกใช้ตัวแปรใดตัวแปรหนึ่งโดด ๆ ไม่ได้ และเพื่อให้สามารถตอบวัตถุประสงค์ของการวิจัยในข้อที่ 2 และทดสอบสมมติฐานที่ตั้งไว้ จึงได้ทดสอบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ (SEE) ที่เกิดจากปฏิสัมพันธ์ระหว่างตัวแปรทั้งสามนี้ โดยทำการทดสอบเปรียบเทียบเป็นรายคู่ โดยใช้วิธีของ เชฟเฟ่ (Scheffe') ปรากฏผล ดังแสดงใน ตาราง 18

ผู้สอบ (SEE) เป็นรายคู่ โดยใช้วิธีของ เชฟเฟ (Scheffe')

กลุ่ม	ค่าเฉลี่ย (SEE)	c16s3	c16s4	v16s3	v16s4	c13s3	c13s4	v13s3	v13s4	c10s3	v13s2	v10s3	c16s2	c10s4	c13s2	c13s1
c16s3	.310	-	.008	.014	.026	.028	.030	.042	.055	.066	.071	.096	.101	.108	.111	.113
c16s4	.318	-	.006	.018	.020	.022	.034	.047	.058	.060	.063	.088	.093	.100	.103	.105
v16s3	.324	-	.012	.014	.016	.028	.041	.052	.054	.057	.082	.087	.094	.097	.099	.099
v16s4	.336	-	.002	.004	.016	.029	.040	.042	.045	.070	.075	.082	.085	.087	.088	.087
c13s3	.338	-	.002	.014	.027	.038	.040	.043	.068	.073	.080	.083	.085	.086	.087	.085
v16s2	.340	-	.012	.025	.036	.038	.041	.066	.071	.078	.081	.083	.084	.085	.086	.083
v13s3	.352	-	.013	.024	.026	.029	.054	.059	.066	.069	.071	.072	.073	.074	.075	.071
c13s4	.365	-	.011	.013	.016	.041	.046	.053	.056	.058	.059	.060	.061	.062	.063	.058
v13s4	.376	-	.002	.005	.030	.035	.042	.045	.047	.048	.049	.050	.051	.052	.053	.047
c10s3	.378	-	.003	.028	.033	.040	.043	.045	.046	.047	.048	.049	.050	.051	.052	.045
v13s2	.381	-	.025	.030	.037	.040	.042	.043	.044	.045	.046	.047	.048	.049	.050	.045
v10s3	.406	-	.005	.012	.015	.017	.018	.019	.020	.021	.022	.023	.024	.025	.026	.021
c16s2	.411	-	.007	.010	.012	.013	.014	.015	.016	.017	.018	.019	.020	.021	.022	.012
c10s4	.418	-	.003	.005	.006	.007	.008	.009	.010	.011	.012	.013	.014	.015	.016	.005
c13s2	.421	-	.002	.003	.004	.005	.006	.007	.008	.009	.010	.011	.012	.013	.014	.002
c13s1	.423	-	.001	.002	.003	.004	.005	.006	.007	.008	.009	.010	.011	.012	.013	-

** p < .01

ตาราง 18 (ต่อ)

กลุ่ม	ค่าเฉลี่ย (SEE)	c10s2	c16s1	v10s4	v10s2	c7s3	c10s1	v16s1	c7s2	v7s3	c7s1	v7s4	c7s4	v13s1	v7s2	v10s1	v7s1
c16s3	.310	.116	.117	.118	.131	.138	.140	.160	.170	.170	.188	.191	.198	.202	.225	.241	.321
c16s4	.318	.108	.109	.110	.123	.130	.132	.152	.162	.162	.180	.183	.190	.194	.217	.233	.313
v16s3	.324	.102	.103	.104	.117	.124	.126	.146	.156	.156	.174	.177	.184	.188	.211	.227	.307
v16s4	.336	.090	.091	.092	.105	.112	.114	.134	.144	.144	.162	.165	.172	.176	.199	.215	.295
c13s3	.338	.088	.089	.090	.103	.110	.112	.132	.142	.142	.160	.163	.170	.174	.197	.213	.293
v16s2	.340	.086	.087	.088	.101	.108	.110	.130	.140	.140	.158	.161	.168	.172	.195	.211	.291
v13s3	.352	.074	.075	.076	.089	.096	.098	.118	.128	.128	.146	.149	.156	.160	.183	.199	.279
c13s4	.365	.061	.062	.063	.076	.083	.085	.105	.115	.115	.133	.136	.143	.147	.170	.186	.266
v13s4	.376	.050	.051	.052	.065	.072	.074	.094	.104	.104	.122	.125	.132	.136	.159	.175	.255
c10s3	.378	.048	.049	.050	.063	.070	.072	.092	.102	.102	.120	.123	.130	.134	.157	.173	.253
v13s2	.381	.045	.046	.047	.060	.067	.069	.089	.099	.099	.117	.120	.127	.131	.154	.170	.250
v10s3	.406	.020	.021	.022	.035	.042	.044	.064	.074	.074	.092	.095	.102	.106	.129	.145	.225
c16s2	.411	.015	.016	.017	.030	.037	.039	.059	.069	.069	.087	.090	.097	.101	.124	.140	.220
c10s4	.418	.008	.009	.010	.023	.030	.032	.052	.062	.062	.080	.083	.090	.094	.117	.133	.213
c13s2	.421	.005	.006	.007	.020	.027	.029	.049	.059	.059	.077	.080	.087	.091	.114	.130	.210
c13s1	.243	.003	.004	.005	.018	.025	.027	.047	.057	.057	.075	.078	.085	.089	.112	.128	.208

* p < .05 ** p < .01

กลุ่ม	ค่าเฉลี่ย (SEE)	c10s2 .426	c16s1 .427	v10s4 .428	v10s2 .441	c7s3 .448	c10s1 .450	v16s1 .470	c7s2 .480	v7s3 .480	c7s1 .498	v7s4 .501	c7s4 .508	v13s1 .512	v7s2 .535	v10s1 .551	v7s1 .631
c10s2	.426	-	.001	.002	.015	.022	.024	.044	.054	.054	.072	.075	.082	.086	.109	.125	.205
c16s1	.427		-	.001	.014	.021	.023	.043	.053	.053	.071	.074	.081	.085	.108	.124	.204
v10s4	.428			-	.013	.020	.022	.042	.052	.052	.070	.073	.080	.084	.107	.123	.203
v10s2	.441				-	.007	.009	.029	.039	.039	.057	.060	.067	.071	.094	.110	.190
c7s3	.448					-	.002	.022	.032	.032	.050	.053	.060	.064	.087	.103	.183
c10s1	.450						-	.020	.030	.030	.048	.051	.058	.062	.085	.101	.181
v16s1	.470							-	.010	.010	.028	.031	.038	.042	.065	.081	.161
c7s2	.480								-	.000	.018	.021	.028	.032	.055	.071	.151
v7s3	.480									-	.018	.021	.028	.032	.055	.071	.151
c7s1	.498										-	.003	.010	.014	.037	.053	.133
v7s4	.501											-	.007	.011	.034	.050	.130
c7s4	.508												-	.004	.027	.043	.123
v13s1	.512													-	.023	.039	.119
v7s2	.535															.016	.096
v10s1	.551															-	.080
v7s1	.631																-

* p < .05 ** p < .01

จากตาราง 18 เป็นการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ($\overline{SEE}$) จากการทดสอบเทเลอร์รูปปริมาตรที่มี รูปแบบ จำนวนชั้น และวิธีการให้คะแนน ที่แตกต่างกัน โดยการเปรียบเทียบเป็นรายคู่ ซึ่งมีกลุ่มตัวอย่างทั้งหมด 32 กลุ่ม จะเห็นว่า

หากทำให้ค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ($\overline{SEE}$) ต่ำสุด หรือมีประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบสูงสุด และรองลงมา ใน 3 อันดับแรก ได้แก่

อันดับที่ 1 ได้แก่ กลุ่ม c16s3 ซึ่งมีค่า $\overline{SEE}$ ต่ำที่สุด เมื่อทดสอบเปรียบเทียบกับกลุ่มอื่น ๆ แล้ว ปรากฏว่า

(1) มีค่า $\overline{SEE}$ ไม่แตกต่างจากกลุ่ม c16s4, v16s3, v16s4, c13s3 และ v16s2 และ

(2) มีค่า $\overline{SEE}$ ต่ำกว่ากลุ่มอื่น ๆ ที่เหลือ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

อันดับที่ 2 ได้แก่ กลุ่ม c16s4 เมื่อทดสอบเปรียบเทียบกับกลุ่มอื่น ๆ แล้ว ปรากฏว่า

(1) มีค่า $\overline{SEE}$ ไม่แตกต่างจากกลุ่ม v16s3, v16s4, c13s3 และ v16s2 และ

(2) มีค่า $\overline{SEE}$ ต่ำกว่ากลุ่มอื่น ๆ ที่เหลือ นอกจากกลุ่มที่อยู่ในอันดับที่ 1 อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

อันดับที่ 3 ได้แก่ กลุ่ม v16s3 เมื่อทดสอบเปรียบเทียบกับกลุ่มอื่น ๆ แล้ว ปรากฏว่า

(1) มีค่า $\overline{SEE}$ ไม่แตกต่างจากกลุ่ม v16s4, c13s3, v16s2 และ v13s3 และ

(2) มีค่า $\overline{SEE}$ ต่ำกว่ากลุ่มอื่น ๆ ที่เหลือ นอกจากกลุ่มที่อยู่ในอันดับที่ 1 และ 2 อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

หากทำให้ค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ($\overline{SEE}$) สูงสุด หรือมีประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบต่ำสุด และรองลงมา ใน 3 อันดับแรก ได้แก่

- อันดับที่ 1 ได้แก่ กลุ่ม v7s1 เป็นกลุ่มที่ให้ค่า $\overline{SEE}$ สูงที่สุด และเมื่อทดสอบเปรียบเทียบกับกลุ่มอื่น ๆ แล้ว ปรากฏว่า มีค่า $\overline{SEE}$ สูงกว่ากลุ่มอื่น ๆ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01
- อันดับที่ 2 ได้แก่ กลุ่ม v10s1 เมื่อทดสอบเปรียบเทียบกับกลุ่มอื่น ๆ แล้ว ปรากฏว่า
- (1) มีค่า $\overline{SEE}$ ไม่แตกต่างจากกลุ่ม v7s2 และ
 - (2) มีค่า $\overline{SEE}$ สูงกว่ากลุ่มอื่น ๆ นอกจากกลุ่มที่อยู่ในอันดับที่ 1 อย่างมีนัยสำคัญทางสถิติที่ระดับ .01
- อันดับที่ 3 ได้แก่ กลุ่ม v7s2 เมื่อทดสอบเปรียบเทียบกับกลุ่มอื่น ๆ แล้ว ปรากฏว่า
- (1) มีค่า $\overline{SEE}$ ไม่แตกต่างจากกลุ่ม v13s1 และ c7s4
 - (2) มีค่า $\overline{SEE}$ สูงกว่ากลุ่มอื่น ๆ ที่เหลือ นอกจากกลุ่มที่อยู่ในลำดับที่ 1 และ 2 อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

จากผลการวิจัยที่พบจะเห็นว่า กลุ่มที่ทดสอบด้วยการทดสอบเทเลอรัูปิรามิดแบบขนาดชั้นแปรผัน ที่มี 16 ชั้น ให้คะแนนโดยวิธีของเบส์ (v16s3) ถึงแม้จะมีค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ($\overline{SEE}$) ค่า อยู่ในอันดับ 3 รองจากกลุ่ม c16s3 และ c16s4 ก็ตาม แต่เมื่อทดสอบเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบทั้ง 3 นี้แล้ว ปรากฏว่า ไม่แตกต่างกัน นั่นคือ การทดสอบเทเลอรัูปิรามิดแบบขนาดชั้นแปรผันที่มี 16 ชั้น ให้คะแนนโดยใช้วิธีของเบส์ (v16s3) จะให้ค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ($\overline{SEE}$) จัดอยู่ในพวกที่ให้ค่าต่ำสุด ซึ่งก็สอดคล้องกับสมมุติฐานที่ตั้งไว้

สำหรับกลุ่มที่ทดสอบด้วยการทดสอบเทเลอรัูปิรามิดแบบขนาดชั้นคงที่ ที่มี 7 ชั้น ให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำได้ (c7s1) ผลการศึกษาพบว่า ไม่ใช่กลุ่มที่ให้ค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ($\overline{SEE}$) สูงสุด ซึ่งไม่สอดคล้องกับวัตถุประสงค์ที่ตั้งไว้

3. ผลการศึกษาและเปรียบเทียบค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ($re\hat{\theta}$) จากการทดสอบเทเลอร์รูปปริมาตรที่มี รูปแบบ จำนวนชั้น และวิธีการให้คะแนน ที่แตกต่างกัน

3.1 การศึกษาค่าเฉลี่ยความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ($re\hat{\theta}$) ปรากฏผล ดังแสดงใน ตาราง 19 และภาพประกอบ 27 ถึง 29

3.2 การเปรียบเทียบค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ($re\hat{\theta}$) จากการทดสอบเทเลอร์รูปปริมาตรที่มี รูปแบบ จำนวนชั้น และวิธีการให้คะแนน ที่แตกต่างกัน โดยการวิเคราะห์ความแปรปรวน ปรากฏผล ดังแสดงใน ตาราง 20 และเปรียบเทียบค่าเฉลี่ยเป็นรายคู่ ดังแสดงใน ตาราง 21

ตาราง 19 แสดงช่วงของค่า $re\hat{\theta}$, นัยสำคัญของค่า $re\hat{\theta}$, $\bar{Z}_r$, SD_{Z_r} และค่า $re\hat{\theta}$ ในแต่ละกลุ่มการทดลองโดยทำการทดลองซ้ำ 40 ครั้ง

กลุ่มทดลอง	ช่วงของค่า $re\hat{\theta}$	จำนวนครั้งที่มีการ นัยสำคัญ (μ)	$\bar{Z}_r$	SD_{Z_r}	$re\hat{\theta}$
c7s1	.7857 ถึง .8383	40 (.01)	1.136	.026	.813
c10s1	.8650 ถึง .8897	40 (.01)	1.378	.029	.881
c13s1	.8889 ถึง .9095	40 (.01)	1.458	.026	.897
c16s1	.8944 ถึง .9208	40 (.01)	1.522	.034	.909
c7s2	.8237 ถึง .9485	40 (.01)	1.276	.094	.855
c10s2	.8687 ถึง .9363	40 (.01)	1.421	.075	.890
c13s2	.9115 ถึง .9263	40 (.01)	1.584	.023	.919
c16s2	.9202 ถึง .9364	40 (.01)	1.645	.029	.928
c7s3	.8689 ถึง .8910	40 (.01)	1.367	.027	.878
c10s3	.9056 ถึง .9239	40 (.01)	1.558	.026	.915
c13s3	.9318 ถึง .9400	40 (.01)	1.699	.017	.935

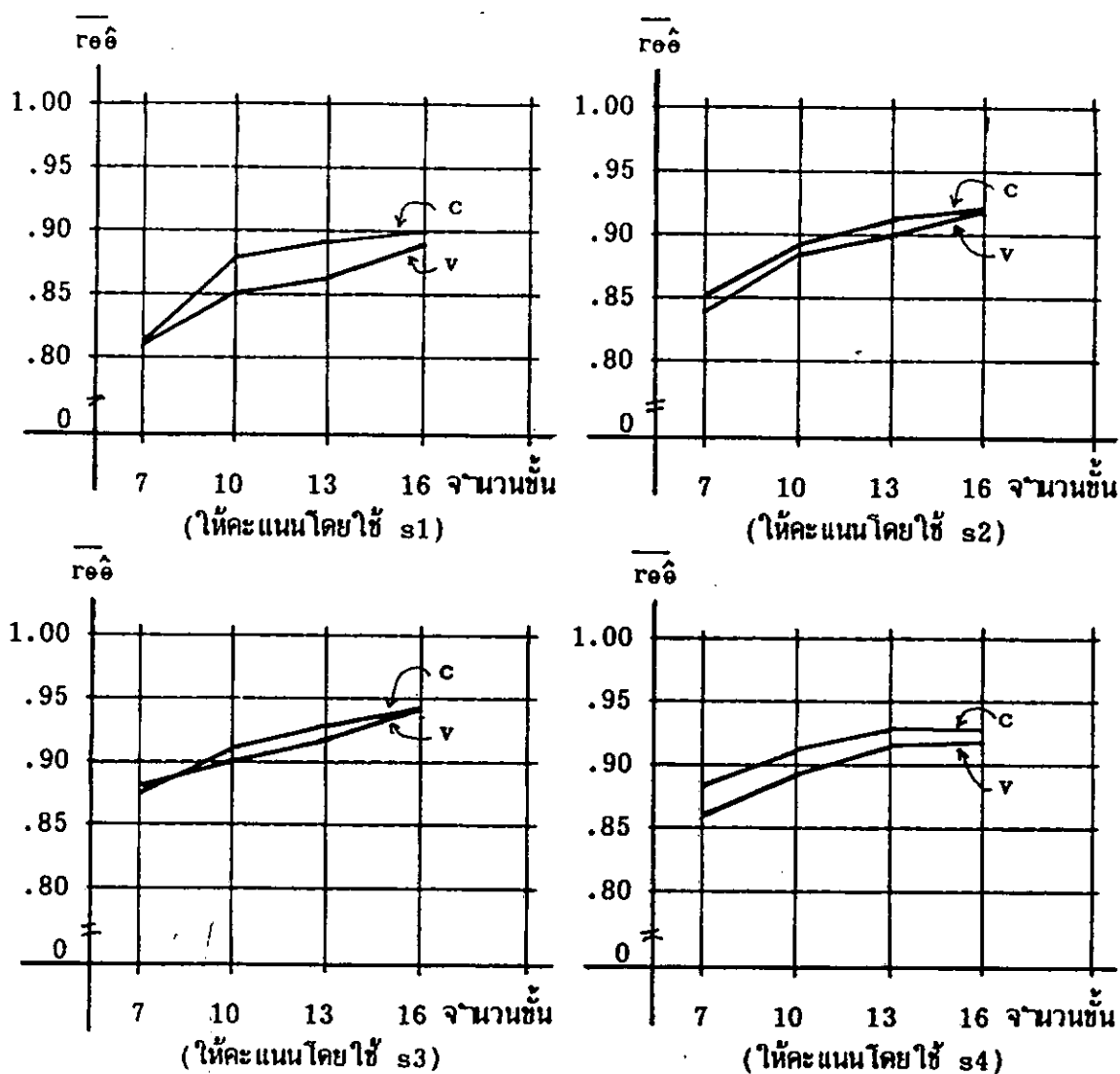
ตาราง 19 (ต่อ)

กลุ่มทดลอง	ช่วงของค่า $re\hat{\theta}$	จำนวนครั้งที่ นัยสำคัญ (%)	$\bar{Zr}$	SD_{Zr}	$\overline{re\hat{\theta}}$
c16s3	.9372 ถึง .9524	40 (.01)	1.813	.026	.948
c7s4	.8741 ถึง .8975	40 (.01)	1.397	.027	.885
c10s4	.9074 ถึง .9246	40 (.01)	1.558	.025	.915
c13s4	.9200 ถึง .9399	40 (.01)	1.677	.044	.932
c16s4	.9245 ถึง .9409	40 (.01)	1.690	.027	.934
v7s1	.7971 ถึง .8328	40 (.01)	1.133	.026	.812
v10s1	.8427 ถึง .8708	40 (.01)	1.279	.025	.856
v13s1	.7787 ถึง .8793	40 (.01)	1.307	.079	.864
v16s1	.8843 ถึง .9022	40 (.01)	1.441	.020	.894
v7s2	.8194 ถึง .8599	40 (.01)	1.232	.029	.843
v10s2	.8796 ถึง .9024	40 (.01)	1.418	.024	.889
v13s2	.9004 ถึง .9183	40 (.01)	1.521	.028	.909
v16s2	.9180 ถึง .9316	40 (.01)	1.619	.021	.924
v7s3	.8619 ถึง .9369	40 (.01)	1.387	.078	.883
v10s3	.8848 ถึง .9085	40 (.01)	1.473	.027	.900
v13s3	.9110 ถึง .9472	40 (.01)	1.601	.068	.922
v16s3	.9359 ถึง .9456	40 (.01)	1.744	.022	.941
v7s4	.8475 ถึง .8705	40 (.01)	1.295	.026	.860
v10s4	.8869 ถึง .9080	40 (.01)	1.455	.028	.897
v13s4	.9095 ถึง .9242	40 (.01)	1.575	.020	.918
v16s4	.9177 ถึง .9374	40 (.01)	1.638	.027	.927

จากตาราง 19 จะเห็นว่าค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ($\overline{re\hat{o}}$) จากการทดลองมาทั้งหมด มีค่าอยู่ระหว่าง .7787 ถึง .9524 มีนัยสำคัญทางสถิติที่ระดับ 0.01 ในทุกกลุ่ม และทุกครั้งในการทดลอง

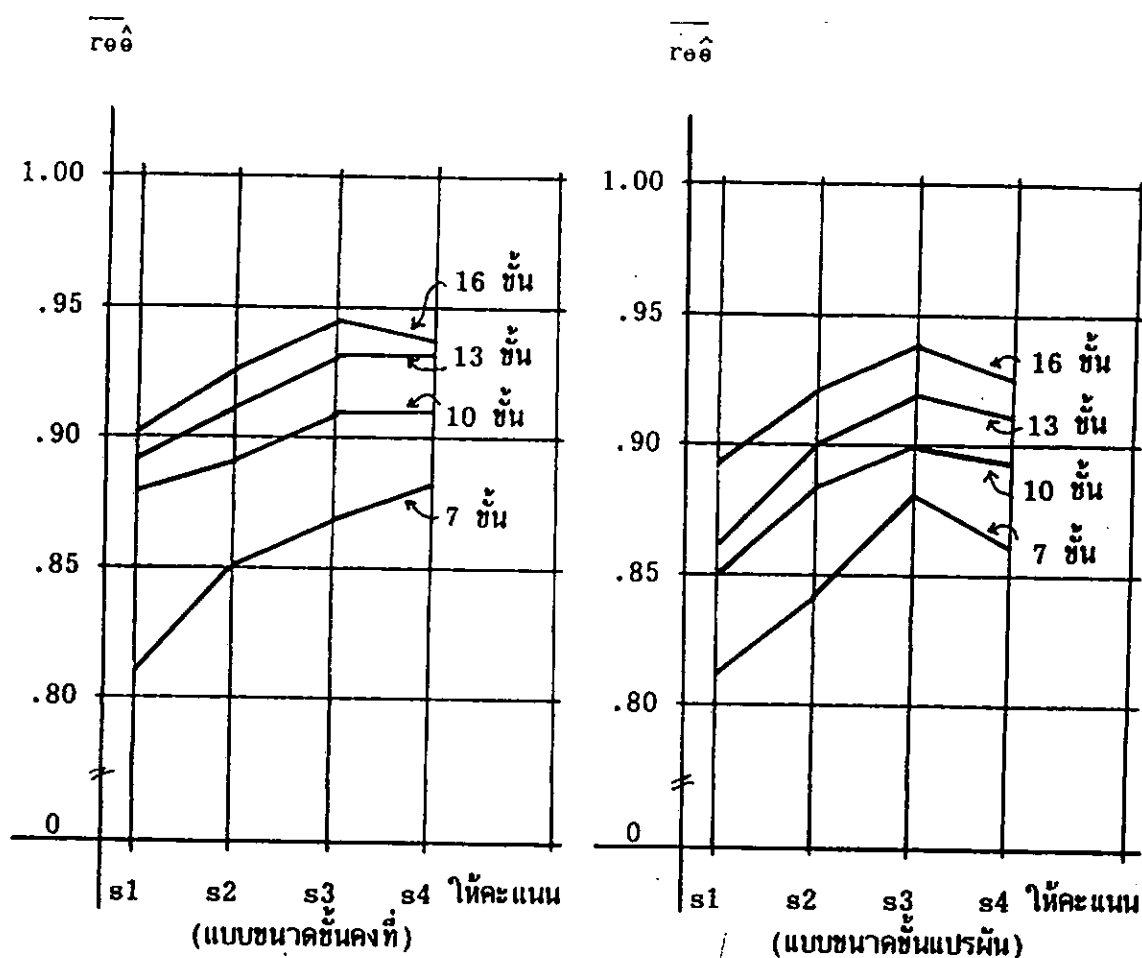
เมื่อพิจารณาความเที่ยงตรงโดยเฉลี่ยภายในกลุ่มทดลอง ($\overline{re\hat{o}}$) จะเห็นว่า กลุ่มที่ให้ค่าเฉลี่ยความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ($\overline{re\hat{o}}$) สูงสุด ได้แก่ กลุ่มที่ทดสอบด้วยการทดสอบเทเลอรรูปปริมาตรแบบขนาดชั้นคงที่ ที่มี 16 ชั้น ให้คะแนนโดยใช้วิธีของเบส์ (c16s3) โดยมีค่าเท่ากับ .948 และกลุ่มที่ให้ค่าเฉลี่ยความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ($\overline{re\hat{o}}$) ต่ำสุด ได้แก่ กลุ่มที่ทดสอบด้วยการทดสอบเทเลอรรูปปริมาตรแบบขนาดชั้นแปรผันที่มี 7 ชั้น ให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำได้ (v7s1) โดยมีค่าเท่ากับ .812

เมื่อนำค่าเฉลี่ยความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ($\overline{re\hat{o}}$) ในแต่ละกลุ่มการทดลอง มาเขียนกราฟ จะได้ดัง ภาพประกอบ 27 ถึง 29



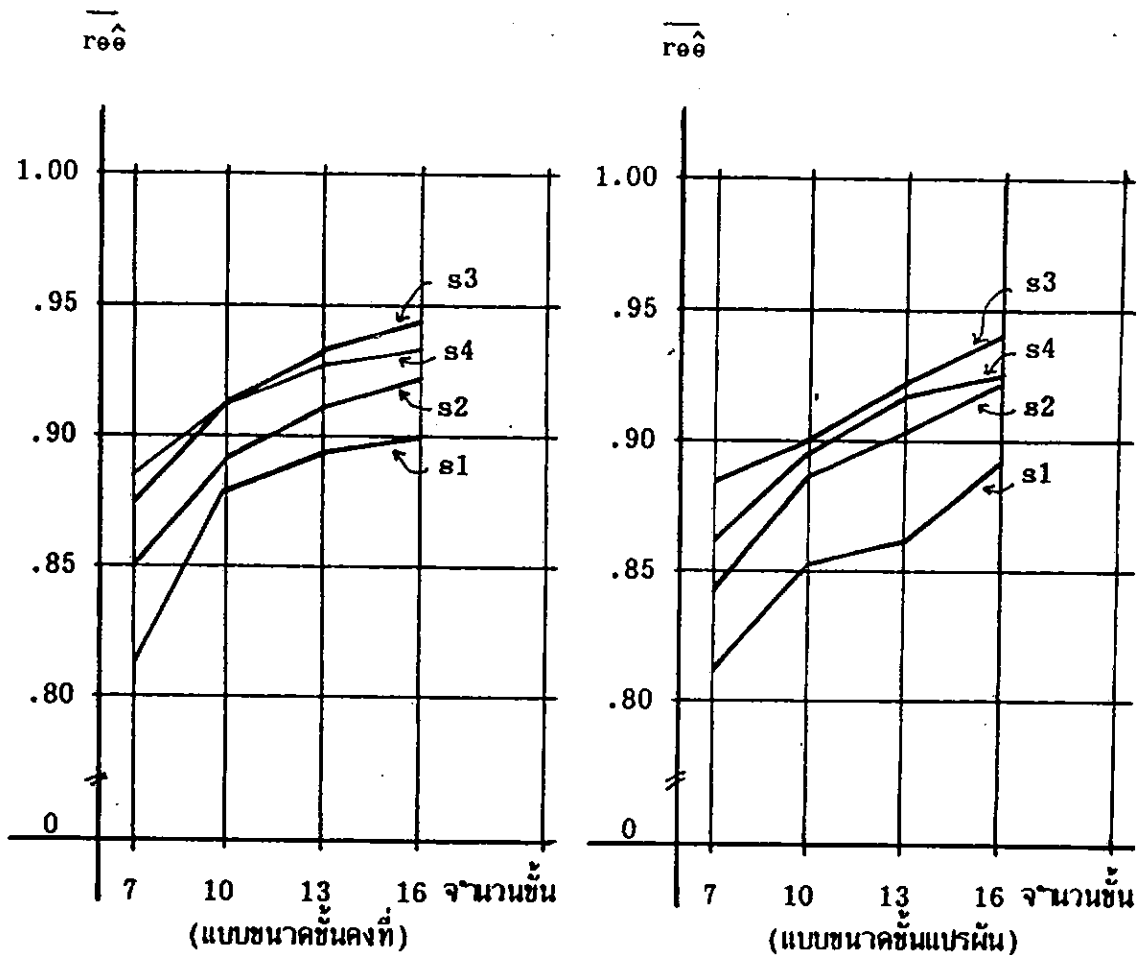
ภาพประกอบ 27 กราฟเปรียบเทียบค่าเฉลี่ยความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ($\overline{reô}$) จากการทดสอบเทเลอรัปรีดามิด ที่มีรูปแบบต่างกัน

จากภาพประกอบ 27 จะเห็นว่า การทดสอบเทเลอรัปรีดามิด ที่มีตั้งแต่ 10 ชั้น ขึ้นไปไม่ว่าจะให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำได้ (s1) ให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทำได้ (s2) ให้คะแนนโดยใช้วิธีของเบส์ (s3) หรือให้คะแนนโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (s4) ก็ความ การทดสอบเทเลอรัปรีดามิดแบบขนาดชั้นคงที่ (c) จะให้ค่าเฉลี่ยความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ($\overline{reô}$) สูงกว่า การทดสอบเทเลอรัปรีดามิดแบบขนาดชั้นแปรผัน (v)



ภาพประกอบ 28 กราฟเปรียบเทียบค่าเฉลี่ยความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ($\overline{re\theta}$) จากการทดสอบเทเลอรูปิรามิด ที่มีจำนวนชั้นต่างกัน

จากภาพประกอบ 28 จะเห็นว่า การทดสอบเทเลอรูปิรามิด ไม่ว่าจะให้ความแนบตามค่าความยากของข้อที่ยากที่สุดที่ทำได้ (s1) ให้ความแนบตามค่าเฉลี่ยความยากของข้อที่ทำได้ (s2) ให้ความแนบโดยวิธีของเบส์ (s3) หรือ ให้ความแนบโดยวิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (s4) และ ไม่ว่าจะเป็น การทดสอบเทเลอรูปิรามิดแบบขนาดชั้นคงที่หรือแบบขนาดชั้นแปรผันก็ตาม การทดสอบเทเลอรูปิรามิดที่มีจำนวน 16 ชั้น จะให้ค่าเฉลี่ยความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ($\overline{re\theta}$) สูงสุด รองลงมา ได้แก่ แบบที่มีจำนวน 13, 10 และ 7 ชั้น ตามลำดับ



ภาพประกอบ 29 กราฟเปรียบเทียบค่าเฉลี่ยความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ($\overline{re\theta}$) จากการทดสอบเทเลอ์รูปปริมาตร ที่ใช้วิธีการให้คะแนนที่ต่างกัน

จากภาพประกอบ 29 จะเห็นว่า การทดสอบเทเลอ์รูปปริมาตร ที่มีตั้งแต่ 10 ชั้น ขึ้นไปไม่ว่าจะเป็น แบบขนาดชั้นคงที่ หรือแบบขนาดชั้นแปรผัน ก็ตาม การให้คะแนนโดยใช้วิธีของเบส์ (s3) จะให้ค่าเฉลี่ยความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ($\overline{re\theta}$) สูงสุด รองลงมาได้แก่ การให้คะแนนโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (s4) การให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทำได้ (s2) และการให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำได้ (s1) ตามลำดับ

ตาราง 20 แสดงผลการวิเคราะห์ความแปรปรวนของค่าความเที่ยงตรงในการประมาณค่า
ความสามารถของผู้สอบ (ในรูปของ Fisher's $Z: Z_r$) ในแต่ละกลุ่มการทดลอง
โดยทำการทดลองซ้ำ 40 ครั้ง

Source of Variation	Sum of Squares	df	Mean Square	F
Main Effects				
รูปแบบ	1.405	1	1.405	420.294**
จำนวนชั้น	23.305	3	7.768	2323.760**
การให้คะแนน	11.301	3	3.767	1126.810**
2-Way Interaction				
รูปแบบxจำนวนชั้น	0.212	3	0.071	21.147**
รูปแบบxการให้คะแนน	0.157	3	0.052	15.626**
จำนวนชั้นxการให้คะแนน	0.472	9	0.052	15.674**
3-Way Interaction				
รูปแบบxจำนวนชั้นxการให้คะแนน	0.255	9	0.028	8.487**
Residual	4.172	1248	3.343×10^{-3}	-
Total	41.279	1279	0.032	-

** $p < .01$

จากตาราง 20 เมื่อพิจารณาจากการทดสอบหลัก (Main Effects) จะเห็นว่า ไม่
ว่าจะเป็นรูปแบบ จำนวนชั้น หรือวิธีการให้คะแนน มีนัยสำคัญทางสถิติที่ระดับ .01 นั่นคือ การใช้
รูปแบบการทดสอบทดสอบเลขรูปปริมาตรที่ต่างกัน การใช้จำนวนชั้นในการทดสอบที่ไม่เท่ากัน หรือวิธี

การให้คะแนนที่ต่างกัน จะส่งผลต่อค่าความความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ (reliability)

เมื่อพิจารณาจากการทดสอบปฏิสัมพันธ์สองทาง (2-Way Interaction) จะเห็นว่า ปฏิสัมพันธ์ ระหว่างรูปแบบxจำนวนชั้น รูปแบบxวิธีการให้คะแนน และจำนวนชั้นxวิธีการให้คะแนน มีนัยสำคัญทางสถิติที่ระดับ .01 นั่นคือ การใช้รูปแบบกับจำนวนชั้นที่ต่างกัน รูปแบบกับวิธีการให้คะแนนที่ต่างกัน หรือจำนวนชั้นกับวิธีการให้คะแนนที่ต่างกัน จะร่วมกันส่งผลต่อค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ (reliability)

เมื่อพิจารณาจากการทดสอบปฏิสัมพันธ์สามทาง (3-Way Interaction) จะเห็นว่า ปฏิสัมพันธ์ระหว่าง รูปแบบxจำนวนชั้นxวิธีการให้คะแนน มีนัยสำคัญทางสถิติที่ระดับ .01 นั่นคือ การใช้รูปแบบ จำนวนชั้น และวิธีการให้คะแนนที่ต่างกัน จะร่วมกันส่งผลต่อค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ

จากภาพประกอบ 27 ปฏิสัมพันธ์ระหว่างรูปแบบxจำนวนชั้น เมื่อแยกพิจารณาตามตัวแปรที่สามคือวิธีการให้คะแนน จะเห็นว่าเส้นกราฟส่วนใหญ่ไม่ตัดกันและไม่ขนานกัน นอกจากจำนวน 7 ชั้น ที่ทำให้เส้นกราฟตัดกัน นั่นคือ ปฏิสัมพันธ์ดังกล่าว ส่วนใหญ่เป็นแบบ Ordinal นอกจาก จำนวน 7 ชั้น ที่ทำให้เกิดปฏิสัมพันธ์แบบ Disordinal

จากภาพประกอบ 28 ปฏิสัมพันธ์ระหว่างจำนวนชั้นxวิธีการให้คะแนน เมื่อแยกพิจารณาตามตัวแปรที่สามคือรูปแบบ จะเห็นว่าเส้นกราฟทั้งหมดไม่ตัดกันและส่วนใหญ่ไม่ขนานกัน นั่นคือ ปฏิสัมพันธ์ดังกล่าว ทั้งหมดเป็นแบบ Ordinal

จากภาพประกอบ 29 ปฏิสัมพันธ์ระหว่างวิธีการให้คะแนนxจำนวนชั้น เมื่อแยกพิจารณาตามตัวแปรที่สามคือรูปแบบ จะเห็นว่าเส้นกราฟส่วนใหญ่ไม่ตัดกันและไม่ขนานกัน นอกจากจำนวน 7 ชั้น ที่ทำให้เส้นกราฟตัดกัน นั่นคือ ปฏิสัมพันธ์ดังกล่าว ส่วนใหญ่เป็นแบบ Ordinal นอกจากจำนวน 7 ชั้น ที่ทำให้เกิดปฏิสัมพันธ์แบบ Disordinal

อนึ่ง สำหรับการวิจัยครั้งนี้มุ่งศึกษาประเด็นสำคัญที่ ปฏิสัมพันธ์สามทาง (3-Way Interaction) ทั้งนี้เพราะในการทดสอบเทเลอร์รูปปริมาตรแต่ละครั้งนั้น จะต้องใช้ตัวแปรทั้งสามนี้

คือ รูปแบบ จำนวนชั้น และวิธีการให้คะแนน พร้อมกันไปทีเดียวทั้ง 3 ตัวแปร จะแยกใช้ตัวแปรใดตัวแปรหนึ่งใด ๆ ไม่ได้ และเพื่อให้สามารถตอบวัตถุประสงค์ของการวิจัยในข้อที่ 3 และทดสอบสมมติฐานที่ตั้งไว้ จึงได้ทดสอบค่าเฉลี่ยความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ($\overline{r_{00}}$) ที่เกิดจากปฏิสัมพันธ์ระหว่างตัวแปรทั้งสามนี้ โดยทำการทดสอบเปรียบเทียบเป็นรายคู่ โดยใช้วิธีของ เชฟเฟ่ (Scheffe') ปรากฏผล ดังแสดงใน ตาราง 21

ตาราง 21 แสดงผลการทดสอบความแตกต่างของ ค่าเฉลี่ยความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ในรูป

ของพิชเชอร์รี่ ($\bar{Z}_r$) เป็นรายคู่ โดยใช้วิธีของ เชฟเฟ (Scheffe')

กลุ่ม	ค่าเฉลี่ย ($\bar{Z}_r$)	v7s1	c7s1	v7s2	c7s2	v10s1	v7s4	v13s1	c7s3	c10s1	v7s3	c7s4	v10s2	c10s2	v16s1	v10s4	c13s1
v7s1	1.133	-	.003	.099	.143	.146	.162	.174	.234	.245	.254	.264	.285	.288	.308	.322	.325
c7s1	1.136	-	.096	.140	.143	.159	.171	.231	.242	.251	.261	.282	.285	.305	.319	.322	
v7s2	1.232	-	.044	.047	.063	.075	.135	.146	.155	.165	.186	.189	.209	.223	.226		
c7s2	1.276	-	.003	.019	.031	.091	.102	.111	.121	.142	.145	.165	.179	.182			
v10s1	1.279	-	.016	.028	.088	.099	.108	.118	.139	.142	.162	.176	.179				
v7s4	1.295	-	.012	.072	.083	.092	.102	.123	.126	.146	.160	.163					
v13s1	1.307	-	.060	.071	.080	.090	.111	.114	.134	.148	.151						
c7s3	1.367	-	.011	.020	.030	.051	.054	.074	.088	.091							
c10s1	1.378	-	.009	.019	.040	.043	.063	.077	.080								
v7s3	1.387	-	.010	.031	.034	.054	.068	.071									
c7s4	1.397	-	.021	.024	.044	.058	.061										
v10s2	1.418	-	.003	.023	.037	.040											
c10s2	1.421	-	.020	.034	.037												
v16s1	1.441	-	.014	.017													
v10s4	1.455	-	.003														
v13s1	1.458	-															

* p < .05 ** p < .01

กลุ่ม	ค่าเฉลี่ย ($\bar{Z}_r$)	v10s3	v13s2	c16s1	c10s4	c10s3	v13s4	c13s4	c13s2	v13s3	v16s2	v16s4	c16s2	c13s4	c16s4	c13s3	v16s3	c16s3
v7s1	1.133	.340	.388	.389	.425	.425	.442	.451	.451	.468	.486	.505	.512	.544	.557	.556	.611	.680
c7s1	1.136	.337	.385	.386	.422	.422	.439	.448	.448	.465	.483	.502	.509	.541	.554	.563	.608	.677
v7s2	1.232	.241	.289	.290	.326	.326	.343	.352	.352	.369	.387	.406	.413	.445	.458	.467	.512	.581
c7s2	1.276	.197	.245	.246	.282	.282	.299	.308	.308	.325	.343	.362	.369	.401	.414	.423	.468	.537
v10s1	1.279	.194	.242	.243	.279	.279	.296	.305	.305	.322	.340	.359	.366	.398	.411	.420	.465	.534
v7s4	1.295	.178	.226	.227	.263	.263	.280	.289	.289	.306	.324	.343	.350	.382	.395	.404	.449	.518
v13s1	1.307	.166	.214	.215	.251	.251	.268	.277	.277	.294	.312	.331	.338	.370	.383	.392	.437	.506
c7s3	1.367	.106	.154	.155	.191	.191	.208	.217	.217	.234	.252	.271	.278	.310	.323	.332	.377	.446
c10s1	1.378	.095	.143	.144	.180	.180	.197	.206	.206	.223	.241	.260	.267	.299	.312	.321	.366	.435
v7s3	1.387	.086	.134	.135	.171	.171	.188	.197	.197	.214	.232	.251	.258	.290	.303	.312	.357	.426
c7s4	1.397	.076	.124	.125	.161	.161	.178	.187	.187	.204	.222	.241	.248	.280	.293	.302	.347	.416
v10s2	1.418	.055	.103	.104	.140	.140	.157	.166	.166	.183	.201	.220	.227	.259	.272	.281	.326	.395
c10s2	1.421	.052	.100	.101	.137	.137	.154	.163	.163	.180	.198	.217	.224	.256	.269	.278	.323	.392
v16s1	1.441	.032	.080	.081	.117	.117	.134	.143	.143	.160	.178	.197	.204	.236	.249	.258	.303	.372
v10s4	1.455	.018	.066	.067	.103	.103	.120	.129	.129	.146	.164	.183	.190	.222	.235	.244	.289	.358
c13s1	1.458	.015	.063	.064	.100	.100	.117	.126	.126	.143	.161	.180	.187	.219	.232	.241	.286	.355

** p < .01

กลุ่ม	ค่าเฉลี่ย ($\bar{Z}_r$)	v10s3	v13s2	c16s1	c10s4	c10s3	v13s4	c13s4	c13s2	v13s3	v16s2	v16s4	c16s2	c13s4	c16s4	c13s3	v16s3	c16s3
v10s3	1.473	-	.048	.049	.085	.085	.102	**	**	**	**	**	**	**	**	**	**	**
v13s2	1.521	-	-	.001	.037	.037	.054	.063	.080	.098	.117	.124	.156	.178	.223	.292		
c16s1	1.522	-	-	.036	.036	.053	.062	.079	.097	.116	.123	.155	.168	.177	.222	.291		
c10s4	1.558	-	-	.000	.017	.026	.043	.061	.080	.087	.119	.132	.141	.186	.255			
c10s3	1.558	-	-	.017	.026	.043	.061	.080	.087	.119	.132	.141	.186	.255				
v13s4	1.575	-	-	.009	.026	.044	.063	.070	.102	.115	.124	.169	.238					
c13s2	1.584	-	-	-	.017	.035	.054	.061	.093	.106	.115	.160	.229					
v13s3	1.601	-	-	.018	.037	.044	.076	.089	.143	.212								
v16s2	1.619	-	-	.019	.026	.058	.071	.080	.125	.194								
v16s4	1.638	-	-	.007	.039	.052	.061	.106	.175									
c16s2	1.645	-	-	.032	.045	.054	.099	.168										
c13s4	1.677	-	-	.013	.022	.067	.136											
c16s4	1.690	-	-	.009	.054	.123												
c13s3	1.699	-	-	.045	.114													
v16s3	1.744	-	-	.069														
c16s3	1.813	-	-															

* p < .05 ** p < .01

จากตาราง 21 เป็นการเปรียบเทียบค่าเฉลี่ยความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ (F_{0.05}) จากการทดสอบเทเลอร์รูปปิรามิดที่มี รูปแบบ จำนวนชั้น และวิธีการให้คะแนนที่แตกต่างกัน แต่ความเที่ยงนี้ได้แปลงให้อยู่ในรูปของ ฟิชเชอร์ซี (Fisher's $Z:\bar{Zr}$) โดยการเปรียบเทียบเป็นรายคู่ ซึ่งมีกลุ่มตัวอย่างทั้งหมด 32 กลุ่ม จะเห็นว่า

พวกที่ให้ค่าเฉลี่ยความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ในรูปของ ฟิชเชอร์ซี ($\bar{Zr}$) ต่ำสุด หรือมีประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบต่ำสุด และรองลงมา ใน 3 อันดับแรก ได้แก่

อันดับที่ 1 ได้แก่ กลุ่ม v7s1 ซึ่งมีค่า $\bar{Zr}$ ต่ำที่สุด เมื่อทดสอบเปรียบเทียบกับกลุ่มอื่น ๆ แล้ว ปรากฏว่า

- (1) มีค่า $\bar{Zr}$ ไม่แตกต่างจากกลุ่ม c7s1 และ
- (2) มีค่า $\bar{Zr}$ ต่ำกว่ากลุ่มอื่น ๆ ที่เหลือ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

อันดับที่ 2 ได้แก่ กลุ่ม c7s1 เมื่อทดสอบเปรียบเทียบกับกลุ่มอื่น ๆ แล้ว ปรากฏว่า มีค่า $\bar{Zr}$ ต่ำกว่ากลุ่มอื่น ๆ นอกจากกลุ่มที่อยู่ในอันดับที่ 1 อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

อันดับที่ 3 ได้แก่ กลุ่ม v7s2 เมื่อทดสอบเปรียบเทียบกับกลุ่มอื่น ๆ แล้ว ปรากฏว่า

- (1) มีค่า $\bar{Zr}$ ไม่แตกต่างจากกลุ่ม c7s2, v10s1, v7s4 และ v13s1
- (2) มีค่า $\bar{Zr}$ ต่ำกว่ากลุ่มอื่น ๆ ที่เหลือ นอกจากกลุ่มที่อยู่ในอันดับที่ 1 และ 2 อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

พวกที่ให้ค่าเฉลี่ยความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ในรูปของ ฟิชเชอร์ซี (Zr) สูงสุด หรือมีประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบสูงสุด และรองลงมา ใน 3 อันดับแรก ได้แก่

อันดับที่ 1 ได้แก่ กลุ่ม c16s3 เป็นกลุ่มที่มีค่า $\bar{Zr}$ สูงที่สุด และเมื่อทดสอบเปรียบเทียบกับกลุ่มอื่น ๆ แล้ว ปรากฏว่า

- (1) มีค่า $\bar{Zr}$ ไม่แตกต่างจากกลุ่ม v16s3 และ
- (2) มีค่า $\bar{Zr}$ สูงกว่ากลุ่มอื่น ๆ ที่เหลือ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

อันดับที่ 2 ได้แก่ กลุ่ม v16s3 เมื่อทดสอบเปรียบเทียบกับกลุ่มอื่น ๆ แล้ว ปรากฏว่า

- (1) มีค่า $\bar{Z}_r$ ไม่แตกต่างจากกลุ่ม c13s3, c16s4, และ c13s4
- (2) มีค่า $\bar{Z}_r$ สูงกว่ากลุ่มอื่น ๆ ที่เหลือ นอกจากกลุ่มที่อยู่ในอันดับที่ 1 อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

อันดับที่ 3 ได้แก่ กลุ่ม c13s3 เมื่อทดสอบเปรียบเทียบกับกลุ่มอื่น ๆ แล้ว ปรากฏว่า

- (1) มีค่า $\bar{Z}_r$ ไม่แตกต่างจากกลุ่ม c16s4, c13s4, c16s2, v16s4 และ v16s2
- (2) มีค่า $\bar{Z}_r$ สูงกว่ากลุ่มอื่น ๆ ที่เหลือ นอกจากกลุ่มที่อยู่ในอันดับที่ 1 และ 2 อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

จากผลการวิจัยที่พบ จะเห็นว่า กลุ่มที่ทดสอบด้วยการทดสอบเทเลอรูปิรามิดแบบขนาดชั้นแปรผัน ที่มี 16 ชั้น ให้คะแนนโดยใช้วิธีของเบส์ (v16s3) ถึงแม้จะมีค่า $\bar{Z}_r$ สูงเป็นอันดับ 2 รองจากกลุ่ม c16s3 ก็ตาม แต่เมื่อทดสอบเปรียบเทียบค่า $\bar{Z}_r$ ทั้ง 2 นี้แล้ว ปรากฏว่า ไม่แตกต่างกัน นั่นคือ การทดสอบเทเลอรูปิรามิดแบบขนาดชั้นแปรผันที่มี 16 ชั้น ให้คะแนนโดยใช้วิธีของเบส์ (v16s3) จะให้ค่า $\bar{Z}_r$ หรือ ความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ (reô) จัดอยู่ในพวกที่ให้ค่าสูงสุด ซึ่งก็สอดคล้องกับสมมุติฐานที่ตั้งไว้

สำหรับกลุ่มที่ทดสอบด้วยการทดสอบเทเลอรูปิรามิดแบบขนาดชั้นคงที่ ที่มี 7 ชั้น ให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทัก (c7s1) ผลการศึกษาพบว่า มีค่า $\bar{Z}_r$ ค่าเป็นอันดับ 2 รองจากกลุ่ม v7s1 แต่เมื่อทดสอบเปรียบเทียบค่า $\bar{Z}_r$ ทั้ง 2 นี้แล้ว ปรากฏว่า ไม่แตกต่างกัน นั่นคือ การทดสอบเทเลอรูปิรามิดแบบขนาดชั้นคงที่ ที่มี 7 ชั้น ให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทัก (c7s1) จะให้ค่า $\bar{Z}_r$ หรือ ค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ (reô) จัดอยู่ในพวกที่ให้ค่าต่ำสุด ซึ่งก็สอดคล้องกับสมมุติฐานที่ตั้งไว้

จากผลการวิจัยที่กล่าวมา สามารถสรุปได้ ดังนี้

1. การให้คะแนนโดยให้ตามค่าความยากของข้อที่ยากที่สุดที่ทัก จะประมาณค่าความสามารถของผู้สอบสูงกว่าความสามารถจริงอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ในทุกกลุ่มและทุก

ครั้งของการทดลอง และการให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทาถูก จะประมาณค่าความสามารถของผู้สอบต่ำกว่าความสามารถจริงอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ในทุกกลุ่มและทุกครั้งของการทดลอง

ส่วนการให้คะแนนโดยใช้วิธีของเบส์ และวิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข ในแต่ละกลุ่มการทดลอง จากการทดลองซ้ำ 40 ครั้ง จะมีจำนวนครั้งไม่ต่ำกว่า 34 ครั้ง ที่ความสามารถจริงกับความสามารถที่ได้จากการประมาณค่า ไม่แตกต่างกัน

2. ในแต่ละกลุ่มการทดลองโดยการทดลองซ้ำ 40 ครั้ง พบว่า ความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ มีค่าเฉลี่ยของค่าเฉลี่ย ($\bar{E}$) อยู่ระหว่าง 0.246 ถึง 0.592 โดยมีส่วนเบี่ยงเบนมาตรฐานของค่าเฉลี่ย มีค่าอยู่ระหว่าง .004 ถึง .069 และความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบมีค่าเฉลี่ย ($\overline{SEE}$) มีค่าอยู่ระหว่าง .310 ถึง .631 โดยมีส่วนเบี่ยงเบนมาตรฐาน มีค่าอยู่ระหว่าง .005 ถึง .094

ส่วนความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ($ro6$) จากการทดลองมาทั้งหมด มีค่าอยู่ระหว่าง .7787 ถึง .9524 โดยมีนัยสำคัญทางสถิติที่ระดับ .01 ทุกค่า และเมื่อหาค่าเฉลี่ยภายในกลุ่มทดลอง ปรากฏว่า มีค่าอยู่ระหว่าง .812 ถึง .948

3. การทดสอบเทเลอร์รูปปริมาตร ทั้งแบบขนาดขั้นคงที่และแบบขนาดขั้นแปรผัน / เมื่อจำนวนขั้นเพิ่มขึ้น มีแนวโน้มว่าจะให้ประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบสูงขึ้น เมื่อให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทาถูก ให้คะแนนโดยใช้วิธีของเบส์ และให้คะแนนโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข

4. การทดสอบเทเลอร์รูปปริมาตรทั้งแบบขนาดขั้นคงที่และแบบขนาดขั้นแปรผัน ถ้าใช้จำนวนขั้นตั้งแต่ 10 ขั้นขึ้นไป การให้คะแนนโดยใช้วิธีของเบส์ จะให้ประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบสูงสุด รองลงมาได้แก่ การให้คะแนนโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข และให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทาถูก ตามลำดับ

5. การทดสอบเทเลอร์รูปปริมาตรที่ใช้ รูปแบบที่ต่างกัน จำนวนขั้นที่ต่างกัน หรือวิธีการให้คะแนนที่ต่างกัน จะมีผล (Main Effect) ต่อ ค่าความคลาดเคลื่อน ค่าความคลาดเคลื่อนมาตรฐาน และ ค่าความเที่ยงตรง อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และพบว่า มีปฏิสัมพันธ์

ระหว่าง รูปแบบจำนวนชั้น รูปแบบวิธีการให้คะแนน จำนวนชั้นวิธีการให้คะแนน และรูปแบบจำนวนชั้นวิธีการให้คะแนน

6. การทดสอบเทลอร์รูปปิรามิดแบบขนาดชั้นคงที่ ที่มี 16 ชั้น ให้คะแนนโดยใช้วิธีของเบส์ (c16s3) จะให้ประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบสูงสุด

บทที่ 5

สรุป อภิปรายผล และข้อเสนอแนะ

การวิจัยครั้งนี้ เป็นการศึกษาเปรียบเทียบประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบ จากการทดสอบเทเลอร์รูปปिरามิดที่มี รูปแบบ จำนวนชั้น และวิธีการให้คะแนน ที่แตกต่างกัน และ ได้ศึกษาโดยใช้เทคนิควิธี มอนติ คาร์โล (Monte Carlo Method) กล่าว คือ ใช้เครื่องคอมพิวเตอร์ในการสุ่มตัวอย่าง จำลองผลการตอบข้อสอบ และประมาณค่าความสามารถของผู้สอบจากการทดสอบเทเลอร์ ซึ่งมีรายละเอียดในการวิจัย ดังนี้

วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาเปรียบเทียบ ค่าความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ (E) จากการทดสอบเทเลอร์รูปปिरามิดที่มี รูปแบบ จำนวนชั้น และวิธีการให้คะแนน ที่แตกต่างกัน
2. เพื่อศึกษาเปรียบเทียบ ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ (SEE) จากการทดสอบเทเลอร์รูปปिरามิดที่มี รูปแบบ จำนวนชั้น และวิธีการให้คะแนน ที่แตกต่างกัน
3. เพื่อศึกษาเปรียบเทียบ ค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ (re $\hat{\theta}$) จากการทดสอบเทเลอร์รูปปिरามิดที่มี รูปแบบ จำนวนชั้น และวิธีการให้คะแนน ที่แตกต่างกัน

ตัวแปรที่ศึกษา

1. ตัวแปรอิสระ ได้แก่

1.1 รูปแบบของการทดสอบเทเลอร์รูปปिरามิด มี 2 รูปแบบ คือ

- (1) แบบขนาดขั้นคงที่ (Constant Step Size Pyramid)
- (2) แบบขนาดขั้นแปรผัน (Variable Step Size Pyramid)

1.2 จำนวนชั้นของการทดสอบเทเลอร์รูปปิรามิด มี 4 แบบ คือ

- (1) แบบที่มีจำนวน 7 ชั้น
- (2) แบบที่มีจำนวน 10 ชั้น
- (3) แบบที่มีจำนวน 13 ชั้น
- (4) แบบที่มีจำนวน 16 ชั้น

1.3 วิธีการให้คะแนนในการทดสอบเทเลอร์รูปปิรามิด มี 4 วิธี คือ

- (1) ให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำได้
- (2) ให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทำได้
- (3) ให้คะแนนโดยใช้วิธีของเบส์ (Bayesian Updating)
- (4) ให้คะแนนโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข

(Conditional Maximum Likelihood)

2. ตัวแปรตาม ได้แก่

- (1) ความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ (E)
- (2) ความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ
(SEE)
- (3) ความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ
(Validity: re $\hat{o}$)

สมมุติฐานในการวิจัย

การทดสอบเทเลอร์รูปปิรามิดแบบขนาดชั้นแปรผัน ที่มี 16 ชั้น ให้คะแนนโดยใช้วิธีของเบส์ จะมีประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบสูงสุด และการทดสอบเทเลอร์รูปปิรามิดแบบขนาดชั้นคงที่ ที่มี 7 ชั้น ให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำได้ จะมีประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบต่ำสุด

เมื่อประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบสูง หมายถึง (1) มีความคลาดเคลื่อนในการประมาณค่าต่ำ (2) มีความคลาดเคลื่อนมาตรฐานในการประมาณค่าต่ำ และ (3) มีความเที่ยงตรงในการประมาณค่าสูง

ประชากรและกลุ่มตัวอย่าง

ประชากรที่ใช้ในการทดลองครั้งนี้ ได้ใช้คะแนนจากการตรวจสอบคุณภาพทางการศึกษา วิชาภาษาไทย ปีการศึกษา 2531 ชั้นประถมศึกษาปีที่ 6 ทั่วประเทศ จำนวน 106,637 คน กลุ่มตัวอย่างได้มาจากการสุ่มจากประชากร มาจำนวน 32 กลุ่ม ๆ ละ 1,000 คน

รูปแบบการทดลอง

การวิจัยครั้งนี้ได้ออกแบบการวิจัย แบบคอมพลีทลี่ แรนดอมไมซ์ แฟกทอเรียล ดีไซน์ (Completely Randomized Factorial Design: CRF_{2x4x4}) (Kirk. 1982: 429-455) ซึ่งจะมีเซลล์ทั้งหมด 32 เซลล์ โดยทำการสุ่มกลุ่มตัวอย่างทั้ง 32 กลุ่ม เข้าไปในเซลล์ต่าง ๆ

ขั้นตอนในการทดลอง

การวิจัยครั้งนี้มีลำดับขั้นตอนในการทดลอง ดังนี้

1. จ้างกลุ่มตัวอย่างและความสามารถจริง (True-Ability: θ_i) ของตัวอย่าง โดยการนำคะแนนจากการตรวจสอบคุณภาพทางการศึกษา ทั่วประเทศ ปีการศึกษา 2531 วิชาภาษาไทย ชั้นประถมศึกษาปีที่ 6 จำนวน 106,637 คน มาแปลงเป็นคะแนนมาตรฐานซี (Z-Score) แล้วสุ่มคะแนนมาตรฐานซี มาเป็นกลุ่มตัวอย่างจำนวน 32 กลุ่ม ๆ ละ 1,000 คน ซึ่งจะใช้คะแนนมาตรฐานซี (Z-Score) ที่ได้นี้ แทนความสามารถจริงของผู้สอบ (True-Ability: θ_i) สำหรับการทดลองครั้งนี้
2. จ้างโครงสร้างของการทดสอบเทเลอรัปรีรามิดแบบขนาดขั้นคงที่ (Constant Step Size) และแบบขนาดขั้นแปรผัน (Variable Step Size) ที่มี 7, 10, 13 และ 16 ขั้น พร้อมทั้งหาค่าความยากของข้อสอบในแต่ละข้อ (β_j) ในการวิจัยครั้งนี้ ได้กำหนดค่าความยากของข้อสอบ ค่าสุด-สูงสุด ให้อยู่ในช่วง -3.000 ถึง +3.000 สำหรับการกำหนดค่าความยากของข้อสอบแต่ละข้อจะเป็นไปตามรูปแบบ และจำนวนขั้นของการทดสอบเทเลอรัปรีรามิดนั้น ๆ
3. หาค่าโอกาส (Probability: P_{ij}) ในการตอบข้อสอบถูกต้องของกลุ่มตัวอย่างที่

จำลองขึ้นในข้อ 1. กับข้อสอบที่จำลองขึ้นในข้อ 2. โดยใช้สูตร แบบ 1-พารามิเตอร์ (One-Parameter Logistic Model) (Hambleton and Swaminathan. 1985: 47)

4. หาผลการตอบข้อสอบของผู้สอบ (U_{ij}) จากโอกาสในการตอบข้อสอบถูก (P_{ij}) ที่ได้ใน ข้อ 3. มาหาผลการตอบ คือ ตอบถูก=1 ตอบผิด=0 ของผู้สอบแต่ละคนในข้อสอบแต่ละข้อ โดยการเรียกเลขสุ่มจากเครื่องคอมพิวเตอร์ที่มีการแจกแจงเป็นแบบยูนิฟอร์ม (Uniform) ที่มีค่าตั้งแต่ 0 ถึง 1 (X_{ij}) แล้วนำค่า X_{ij} ไปเปรียบเทียบกับค่า P_{ij} ถ้า $X_{ij} \leq P_{ij}$ แสดงว่าตกอยู่ในพื้นที่ที่ยอมรับว่าตอบข้อสอบถูก จึงให้ $U_{ij}=1$ และถ้า $X_{ij} > P_{ij}$ ก็จะปฏิเสธการตอบถูก นั่นคือ ตอบผิด ก็จะทำให้ $U_{ij}=0$ (Weiss and McBride. 1984: 276; Vijver. 1986: 46) ตัวอย่างนี้กับผู้สอบทุกคนในข้อสอบทุกข้อ ก็จะได้ผลการตอบ (0, 1) โดยจะทำการทดลองซ้ำทั้งหมด 40 ครั้ง ซึ่งเปรียบเหมือนผู้สอบทำข้อสอบชุดเดิมซ้ำ 40 ครั้ง

5. หาผลการตอบข้อสอบในข้อ 4. มาหาเส้นทางในการตอบข้อสอบการทดสอบเทเลอร์ รูปปริมาตร ตามที่จำลองขึ้นในข้อ 2. และให้คะแนนตามวิธีการที่กำหนดไว้ 4 วิธี ก็จะได้ค่าความสามารถของผู้สอบที่ประมาณค่าจากการทดสอบเทเลอร์

การวิเคราะห์ข้อมูล

หาค่าสถิติเบื้องต้น ทดสอบความแตกต่างโดยใช้ ที-เทสต์ (t-test) การวิเคราะห์ความแปรปรวนสามทาง (Tree-Way ANOVA) และเปรียบเทียบเป็นรายคู่โดยใช้วิธีของเชฟเฟ่ (Scheffe's Method)

สรุปผลการวิจัย

1. จากการศึกษาเปรียบเทียบ ค่าความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ (E) จากการทดสอบเทเลอร์รูปปริมาตรที่มี รูปแบบ จำนวนชั้น และวิธีการให้คะแนนที่ต่างกันไป พบว่า

1.1 การให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำถูก จะประมาณค่าความสามารถของผู้สอบสูงกว่าความสามารถจริงอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ในทุกกลุ่มและทุก

ครั้งของการทดลอง

1.2 การให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทำได้ จะประมาณค่าความสามารถของผู้สอบต่ำกว่าความสามารถจริงอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ในทุกกลุ่มและทุกครั้งของการทดลอง

1.3 การให้คะแนนโดยใช่วิธีของเบส์ และวิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไขในแต่ละกลุ่มการทดลอง จากการทดลองซ้ำ 40 ครั้ง จะมีจำนวนครั้งไม่ต่ำกว่า 34 ครั้ง ที่ความสามารถจริงกับความสามารถที่ได้จากการประมาณค่า ไม่แตกต่างกัน

1.4 ความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ มีค่าเฉลี่ย ($\bar{E}$) ภายในกลุ่มทดลอง มีค่าอยู่ระหว่าง .246 ถึง .592 โดยมีค่าส่วนเบี่ยงเบนมาตรฐานอยู่ระหว่าง 0.004 ถึง 0.069

1.5 การทดสอบเทเลอร์รูปปริมาตรทั้งแบบขนาดขั้นคงที่และแบบขนาดขั้นแปรผัน เมื่อจำนวนขั้นเพิ่มขึ้น ค่าความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ ($\bar{E}$) มีแนวโน้มว่าจะลดลง เมื่อให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทำได้ ให้คะแนนโดยใช่วิธีของเบส์ และให้คะแนนโดยใช่วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข ยกเว้นการให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำได้ ซึ่งยังสรุปไม่ได้

1.6 การทดสอบเทเลอร์รูปปริมาตรทั้งแบบขนาดขั้นคงที่และแบบขนาดขั้นแปรผัน การให้คะแนนโดยใช่วิธีของเบส์ จะให้ค่าความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ ($\bar{E}$) ต่ำสุด รองลงมาได้แก่ การให้คะแนนโดยใช่วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข และให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทำได้ ตามลำดับ ส่วนการให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำได้ ซึ่งยังสรุปไม่ได้

1.7 การทดสอบเทเลอร์รูปปริมาตรที่ใช้ รูปแบบที่ต่างกัน จำนวนขั้นที่ต่างกัน หรือวิธีการให้คะแนนที่ต่างกัน จะมีผล (Main Effect) ต่อค่าความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ ($\bar{E}$) อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

เมื่อทดสอบปฏิสัมพันธ์ (Interaction) พบว่า มีปฏิสัมพันธ์ระหว่าง รูปแบบ x จำนวนขั้น อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 รูปแบบ x วิธีการให้คะแนน จำนวนขั้น x วิธีการให้คะแนน และ

รูปแบบจำนวนชั้นวิธีการให้คะแนน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

1.8 เมื่อทดสอบเปรียบเทียบเป็นรายคู่ พบว่า การทดสอบเทเลอรูปิรามิดแบบขนาดชั้นคงที่ ที่มี 16 ชั้น ให้คะแนนโดยใช่วิธีของเบส์ (c16s3) จะให้ค่าความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบต่ำสุด รองลงมาได้แก่ แบบขนาดชั้นคงที่ ที่มี 16 ชั้น ให้คะแนนโดยใช่วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (c16s4) และแบบขนาดชั้นแปรผัน ที่มี 16 ชั้น ให้คะแนนโดยใช่วิธีของเบส์ (v16s3) ตามลำดับ เมื่อทดสอบค่าความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบทั้ง 3 ค่า นี้แล้ว ปรากฏว่า ไม่แตกต่างกัน

ส่วนการทดสอบเทเลอรูปิรามิด ที่ให้ค่าความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบสูงสุด ได้แก่ แบบขนาดชั้นแปรผัน ที่มี 7 ชั้น ให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ท้าทาย (v7s2)

2. จากการศึกษาเปรียบเทียบ ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ (SEE) จากการทดสอบเทเลอรูปิรามิดที่มี รูปแบบ จำนวนชั้น และวิธีการให้คะแนน ที่แตกต่างกัน พบว่า

2.1 ความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ มีค่าเฉลี่ย ($\overline{SEE}$) ภายในกลุ่มทดลอง มีค่าอยู่ระหว่าง .310 ถึง .631 โดยมีค่าส่วนเบี่ยงเบนมาตรฐานอยู่ระหว่าง .005 ถึง .094

2.2 การทดสอบเทเลอรูปิรามิดทั้งแบบขนาดชั้นคงที่และแบบขนาดชั้นแปรผัน เมื่อจำนวนชั้นเพิ่มขึ้น ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ (SEE) มีแนวโน้มว่าจะลดลง เมื่อให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ท้าทาย ให้คะแนนโดยใช่วิธีของเบส์ และให้คะแนนโดยใช่วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข ยกเว้นการให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ท้าทาย ซึ่งยังสรุปไม่ได้

2.3 การทดสอบเทเลอรูปิรามิดทั้งแบบขนาดชั้นคงที่และแบบขนาดชั้นแปรผัน ถ้าใช้จำนวนชั้นตั้งแต่ 10 ชั้นขึ้นไป การให้คะแนนโดยใช่วิธีของเบส์ จะให้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ (SEE) ต่ำสุด รองลงมาได้แก่ การให้คะแนนโดยใช่วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข ให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ท้าทาย

ถูก และให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำได้ ตามลำดับ ส่วนการทดสอบเทเลอรัป
ปิรามิดที่มี 7 ชั้น ซึ่งยังสรุปไม่ได้

2.4 การทดสอบเทเลอรัปปิรามิดที่ใช้ รูปแบบที่ต่างกัน จำนวนชั้นที่ต่างกัน หรือ
วิธีการให้คะแนนที่ต่างกัน จะมีผล (Main Effect) คือ ค่าความคลาดเคลื่อนมาตรฐานในการ
ประมาณค่าความสามารถของผู้สอบ (SEE) อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

เมื่อทดสอบปฏิสัมพันธ์ (Interaction) พบว่า มีปฏิสัมพันธ์ระหว่าง รูปแบบxจำนวนชั้น
รูปแบบxวิธีการให้คะแนน จำนวนชั้นxวิธีการให้คะแนน และรูปแบบxจำนวนชั้นxวิธีการให้คะแนน
อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

2.5 เมื่อทดสอบเปรียบเทียบเป็นรายคู่ พบว่า การทดสอบเทเลอรัปปิรามิดแบบ
ขนาดชั้นคงที่ ที่มี 16 ชั้น ให้คะแนนโดยใช้วิธีของเบส์ (c16s3) จะให้ค่าความคลาดเคลื่อน
มาตรฐานในการประมาณค่าความสามารถของผู้สอบต่ำสุด รองลงมาได้แก่ แบบขนาดชั้นคงที่ ที่มี
16 ชั้น ให้คะแนนโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (c16s4) และ แบบขนาดชั้น
แปรผัน ที่มี 16 ชั้น ให้คะแนนโดยใช้วิธีของเบส์ (v16s3) ตามลำดับ เมื่อทดสอบค่าความคลาด
เคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบทั้ง 3 ค่า นี้แล้ว ปรากฏว่า ไม่แตกต่าง

ส่วนการทดสอบเทเลอรัปปิรามิด ที่ให้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า
ความสามารถของผู้สอบ (SEE) สูงสุด ได้แก่ แบบขนาดชั้นแปรผัน ที่มี 7 ชั้น ให้คะแนนตามค่า
ความยากของข้อที่ยากที่สุดที่ทำได้ (v7s1)

3. จากการศึกษาเปรียบเทียบ ค่าความเที่ยงตรงในการประมาณค่าความสามารถของ
ผู้สอบ ($re\hat{\theta}$) จากการทดสอบเทเลอรัปปิรามิดที่มี รูปแบบ จำนวนชั้น และวิธีการให้คะแนน ที่
แตกต่างกัน พบว่า

3.1 ความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ($re\hat{\theta}$) จากการ
ทดลองทั้งหมด มีค่าอยู่ระหว่าง .7787 ถึง .9524 โดยมีนัยสำคัญทางสถิติที่ระดับ .01 ทุกค่า
และมีความเที่ยงตรงโดยเฉลี่ย ($\overline{re\hat{\theta}}$) ภายในกลุ่มทดลอง มีค่าอยู่ระหว่าง .812 ถึง .948

3.2 การทดสอบเทเลอรัปปิรามิดทั้งแบบขนาดชั้นคงที่และแบบขนาดชั้นแปรผัน

เมื่อจำนวนชั้นเพิ่มขึ้น ค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ (re $\hat{o}$) มีแนวโน้มว่าจะเพิ่มขึ้น ไม่ว่าจะเป็นให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำได้ ให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทำได้ ให้คะแนนโดยใช้วิธีของเบส์ หรือให้คะแนนโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข ก็ตาม

3.3 การทดสอบเทเลอรัปรีมิตทั้งแบบขนาดชั้นคงที่และแบบขนาดชั้นแปรผัน ถ้าใช้จำนวนชั้นตั้งแต่ 10 ชั้นขึ้นไป การให้คะแนนโดยใช้วิธีของเบส์ จะให้ค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ (re $\hat{o}$) สูงสุด รองลงมาได้แก่ การให้คะแนนโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข ให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทำได้ และให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำได้ ตามลำดับ ส่วนการทดสอบเทเลอรัปรีมิตที่มี 7 ชั้น ซึ่งยังสรุปไม่ได้

3.4 การทดสอบเทเลอรัปรีมิตที่ใช้ รูปแบบที่ต่างกัน จำนวนชั้นที่ต่างกัน หรือวิธีการให้คะแนนที่ต่างกัน จะมีผล (Main Effect) ต่อ ค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ (re $\hat{o}$) อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

เมื่อทดสอบปฏิสัมพันธ์ (Interaction) พบว่า มีปฏิสัมพันธ์ระหว่าง รูปแบบxจำนวนชั้น รูปแบบxวิธีการให้คะแนน จำนวนชั้นxวิธีการให้คะแนน และรูปแบบxจำนวนชั้นxวิธีการให้คะแนน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

3.5 เมื่อทดสอบเปรียบเทียบเป็นรายคู่ พบว่า การทดสอบเทเลอรัปรีมิตแบบแบบขนาดชั้นคงที่ ที่มี 16 ชั้น ให้คะแนนโดยใช้วิธีของเบส์ (c16s3) จะให้ค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบสูงสุด รองลงมาได้แก่ แบบขนาดชั้นแปรผัน ที่มี 16 ชั้น ให้คะแนนโดยใช้วิธีของเบส์ (v16s3) เมื่อทดสอบค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบทั้ง 2 ค่านี้แล้ว ปรากฏว่า ไม่แตกต่างกัน

ส่วนการทดสอบเทเลอรัปรีมิตที่ให้ค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ (re $\hat{o}$) ต่ำสุด ได้แก่ แบบขนาดชั้นแปรผัน ที่มี 7 ชั้น ให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำได้ (v7s1) รองลงมาได้แก่ แบบขนาดชั้นคงที่ ที่มี 7 ชั้น ให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำได้ (c7s1) เมื่อทดสอบค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบทั้ง 2 ค่านี้แล้ว ปรากฏว่า ไม่แตกต่างกัน

อภิปรายผลการวิจัย

การศึกษาเปรียบเทียบประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบ จากการทดสอบเทเลอ์รูปปริมาตรที่มี รูปแบบ จำนวนชั้น และวิธีการให้คะแนน ที่แตกต่างกัน ปรากฏผลที่สำคัญ และได้นำมาอภิปราย ดังต่อไปนี้

1. จากการศึกษาเปรียบเทียบ ค่าความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ (E) จากการทดสอบเทเลอ์รูปปริมาตรที่มี รูปแบบ จำนวนชั้น และวิธีการให้คะแนน ที่แตกต่างกัน พบว่า

1.1 การทดสอบเทเลอ์รูปปริมาตรแบบขนาดชั้นคงที่ ที่มี 16 ชั้น ให้คะแนนโดยใช้วิธีของเบส์ (c16s3) จะให้ค่าความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ (E) ต่ำสุด รองลงมา ได้แก่ แบบขนาดชั้นคงที่ ที่มี 16 ชั้น ให้คะแนนโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (c16s4) และแบบขนาดชั้นแปรผัน ที่มี 16 ชั้น ให้คะแนนโดยใช้วิธีของเบส์ (v16s3) ตามลำดับ เมื่อทดสอบความแตกต่างของค่าความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบทั้ง 3 ค่านี้แล้ว ปรากฏว่า ไม่แตกต่างกัน นั่นคือ ความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบที่ได้จากทั้งสามกลุ่มนี้ ถือว่า เป็นพวกที่อยู่ในกลุ่มค่าสุดกลุ่มเดียวกัน ซึ่งก็สอดคล้องกับสมมุติฐานที่ตั้งไว้ว่า การทดสอบเทเลอ์รูปปริมาตรแบบขนาดชั้นแปรผัน ที่มี 16 ชั้น ให้คะแนนโดยใช้วิธีของเบส์ (v16s3) จะให้ค่าความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบต่ำสุด

ส่วนการทดสอบเทเลอ์รูปปริมาตร ที่ให้ค่าความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ (E) สูงสุด ได้แก่ แบบขนาดชั้นแปรผัน ที่มี 7 ชั้น ให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ท้าทาย (v7s2) ซึ่ง ไม่สอดคล้องกับสมมุติฐานที่ตั้งไว้ว่า การทดสอบเทเลอ์รูปปริมาตรแบบขนาดชั้นคงที่ ที่มี 7 ชั้น ให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ท้าทาย (c7s1) จะให้ค่าความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบสูงสุด และไม่สอดคล้องกับคำกล่าวของ แพทเตอร์สันและลอร์ด (Weiss. 1974: 18; citing Paterson. 1972; Lord. 1971) ที่ว่า การทดสอบเทเลอ์รูปปริมาตรแบบขนาดชั้นแปรผัน มีประสิทธิภาพเกี่ยวกับความไวในการกำหนดเส้นทางในการตอบข้อสอบมากกว่าแบบขนาดชั้นคงที่ ทั้งนี้อาจจะเป็นเพราะ

กลุ่มตัวอย่างที่ใช้ในการทดลองครั้งนี้ ความสามารถจริง (θ_1) มีการแจกแจงเป็นโค้งปกติ นั้น คือ ผู้สอบส่วนใหญ่มีความสามารถจริงอยู่ในระดับปานกลาง ส่วนที่มีความสามารถค่อนข้างสูงหรือต่ำนั้นมีค่อนข้างน้อย จึงเหมาะกับการทดสอบเทเลอรัปรีรามิด แบบขนาดชั้นคงที่ ซึ่งมีข้อสอบส่วนใหญ่มีความยากปานกลาง ส่วนข้อสอบที่ค่อนข้างไปทางยากหรือง่ายมีจำนวนน้อย ซึ่งผิดกับการทดสอบเทเลอรัปรีรามิดแบบขนาดชั้นแปรผัน ที่มีข้อสอบส่วนใหญ่มีความยากค่อนข้างไปทางยาก หรือง่าย ส่วนข้อสอบที่มีความยากปานกลางมีจำนวนค่อนข้างน้อย แต่อย่างไรก็ตามควรจะ ได้มีการศึกษาในเรื่องนี้อีกครั้งหนึ่ง โดยศึกษาแยกกลุ่มตัวอย่างเป็น กลุ่มที่มีความสามารถสูง ปานกลาง และต่ำ ส่วนวิธีการให้คะแนนและจำนวนชั้นในการทดสอบเทเลอรัปรีรามิด ก็อาจจะ เป็นอีกสาเหตุหนึ่งที่ทำให้ผลการวิจัย ไม่สอดคล้องกับสมมุติฐานที่ตั้งไว้ กล่าวคือ การให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำได้ จะมีความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบค่อนข้างสูง ซึ่งจะทำให้การให้คะแนนโดยวิธีนี้ส่วนใหญ่ยังลงสรุปไม่ได้ และจำนวนชั้นของการทดสอบเทเลอรัปรีรามิดที่มี 7 ชั้น อาจจะมีจำนวนชั้นหรือจำนวนข้อสอบที่ผู้สอบต้องทำน้อยเกินไป จึงทำให้เกิดความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบค่อนข้างสูง ซึ่งจะทำให้ผลการวิจัยที่ใช้จำนวน 7 ชั้น ส่วนหนึ่งยังลงสรุปไม่ได้ เช่นเดียวกัน

1.2 การทดสอบเทเลอรัปรีรามิดทั้งแบบขนาดชั้นคงที่และแบบขนาดชั้นแปรผัน เมื่อจำนวนชั้นเพิ่มขึ้น ความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ มีแนวโน้มว่าจะมีค่าลดลง เมื่อให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทำได้ ให้คะแนนโดยใช้วิธีของเบส์ และให้คะแนนโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข ซึ่งก็สอดคล้องกับทฤษฎีการทดสอบ ที่ว่า ถ้าเพิ่มความยาวของข้อสอบ ค่าความเชื่อมั่นและค่าความเที่ยงตรงก็จะเพิ่มขึ้น หรือค่าความคลาดเคลื่อนมาตรฐานในการวัดก็จะลดลง (Gulliksen. 1967: 74-105; Lord and Novick. 1968: 112-118)

ส่วนการให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำได้ยังสรุปไม่ได้ ทั้งนี้ อาจจะเป็นเพราะการให้คะแนนโดยวิธีนี้ มีความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบค่อนข้างสูง จึงทำให้ผลการวิจัยยังลงสรุปไม่ได้

1.3 การให้คะแนนโดยให้ตามค่าความยากของข้อที่ยากที่สุดที่ทำได้ จะประมาณ

ค่าความสามารถของผู้สอบสูงกว่าความสามารถจริง อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ในทุกครั้งและทุกกลุ่มการทดลอง ซึ่งก็สอดคล้องกับผลการวิจัยของค่าย เชียงฉี (2533: 64) ที่ ศึกษาเปรียบเทียบความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ จากการทดสอบเทเลอรรูปปิรามิด ส่วนการให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทาถูก จะประมาณค่าความสามารถของผู้สอบต่ำกว่าความสามารถจริง อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ในทุกครั้งและทุกกลุ่มการทดลอง และการให้คะแนนโดยใช้วิธีของเบส์กับวิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข จะประมาณค่าความสามารถของผู้สอบได้ใกล้เคียงกับความสามารถจริงมาก ในแต่ละกลุ่มการทดลองจากการทดลองซ้ำ 40 ครั้ง มีจำนวนครั้งไม่ต่ำกว่า 34 ครั้ง ที่ความสามารถจริงกับความสามารถที่ได้จากการประมาณค่าไม่แตกต่างกัน แต่อย่างไรก็ตามการให้คะแนนโดยใช้วิธีของเบส์ และใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข จะต้องใช้สถิติที่ค่อนข้างซับซ้อน ซึ่งจำเป็นต้องใช้เครื่องคอมพิวเตอร์ช่วยในการประมาณค่าความสามารถของผู้สอบ นอกจากนั้นการให้คะแนนโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข ยังพบปัญหาอีกสองประการ คือ ประการแรก ในกรณีที่ผู้สอบตอบข้อสอบถูกหมดหรือผิดหมดไม่สามารถที่จะประมาณค่าได้ และ ประการที่สอง ในบางครั้งยังพบกรณีที่มีค่าความเป็นไปได้สูงสุดมีมากกว่าหนึ่งค่า โดยเฉพาะอย่างยิ่งการทดสอบเทเลอรรูปปิรามิดแบบขนาดขั้นแปดขั้น ซึ่งก็สอดคล้องกับผลการวิจัยของ แฮมมิลตัน (Hambleton and Swaminathan.1985: 87; citing Semijima.1973a) ที่พบว่าในกรณีที่ใช้ 3-พารามิเตอร์ ประมาณค่าความสามารถของผู้สอบโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข บางครั้งอาจจะมีค่าสูงสุดได้หลายค่า ในกรณีเช่นนี้จึง ไม่สามารถที่จะประมาณค่าความสามารถของผู้สอบได้

การทดสอบเทเลอรรูปปิรามิดทั้งแบบขนาดขั้นคงที่และแบบขนาดขั้นแปดขั้น การให้คะแนนโดยใช้วิธีของเบส์ จะให้ค่าความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบต่ำสุด ซึ่งก็เป็นไปตามที่คาดหมาย ทั้งนี้เพราะการให้คะแนนโดยใช้วิธีของเบส์ เป็นวิธีการให้คะแนนที่คิดขึ้นมาเพื่อใช้กับการทดสอบเทเลอรรูปปิรามิดโดยเฉพาะ (Owen. 1969; 1975: 351-356) นอกจากนี้ยังสอดคล้องกับผลการวิจัยของ กิฟฟอร์ดกับสวามินาธาน (Gifford and Swaminathan. 1990: 33) ที่พบว่า ถ้ากลุ่มตัวอย่างขนาดเล็ก และข้อสอบมีจำนวนน้อยข้อ การประมาณค่าพารามิเตอร์ต่าง ๆ โดยใช้วิธีของเบส์ (Joint Bayesian Estimate) จะให้ความถูกต้องมาก

กว่า เมื่อเทียบกับวิธีความเป็นไปได้สูงสุด (Joint Maximum Likelihood Estimate) นอก
จากนั้น ยังสอดคล้องกับผลการวิจัยของ สแคกส์กับสตีเวนสัน (Skaggs and Stevenson.
1989: 391-402) ที่พบว่าถ้ากลุ่มตัวอย่างขนาดเล็ก ($N < 500$) การประมาณค่าโดยวิธีของเบส์
(Pseudo-Bayesian) จะให้ค่าความคงที่ (Consistency) และความถูกต้องแม่นยำ
(Accuracy) มากกว่า การประมาณค่าโดยใช้วิธีความเป็นไปได้สูงสุด (Joint Maximum
Likelihood) รองลงมาได้แก่ การให้คะแนนโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข และ
การให้คะแนนตามค่าความยากเฉลี่ยของข้อที่ท้าทาย ตามลำดับ ส่วนการให้คะแนนตามค่าความยาก
ของข้อที่ยากที่สุดที่ท้าทาย ยังลงสรุปไม่ได้

2. จากการศึกษาเปรียบเทียบ ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความ
สามารถของผู้สอบ (SEE) จากการทดสอบเทเลอรัปรีรามิดที่มี รูปแบบ จำนวนชั้น และวิธีการ
ให้คะแนน ที่แตกต่างกัน พบว่า

2.1 การทดสอบเทเลอรัปรีรามิดแบบขนาดชั้นคงที่ ที่มี 16 ชั้น ให้คะแนนโดย
ใช้วิธีของเบส์ (c16s3) จะให้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ
ของผู้สอบต่ำสุด รองลงมาได้แก่ แบบขนาดชั้นคงที่ ที่มี 16 ชั้น ให้คะแนนโดยใช้วิธีความเป็นไป
ได้สูงสุดแบบมีเงื่อนไข (c16s4) และแบบขนาดชั้นแปรผัน ที่มี 16 ชั้น ให้คะแนนโดยใช้วิธีของ
เบส์ (v16s3) ตามลำดับ ซึ่งก็สอดคล้องกับค่าความคลาดเคลื่อนในการประมาณค่าความสามารถ
ของผู้สอบที่กล่าวไว้ในข้อ 1. เมื่อทดสอบค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความ
สามารถของผู้สอบทั้ง 3 ค่านี้แล้ว ปรากฏว่า ไม่แตกต่างกัน นั่นคือ ความคลาดเคลื่อนมาตรฐาน
ในการประมาณค่าความสามารถของผู้สอบที่ได้จากทั้งสามกลุ่มนี้ ถือว่า เป็นพวกที่อยู่ในกลุ่มต่ำสุด
กลุ่มเดียวกัน ซึ่งก็สอดคล้องกับสมมุติฐาน ที่ตั้งไว้ว่า การทดสอบเทเลอรัปรีรามิดแบบขนาดชั้น
แปรผัน ที่มี 16 ชั้น ให้คะแนนโดยใช้วิธีของเบส์ (v16s3) จะให้ค่าความคลาดเคลื่อนมาตรฐาน
ในการประมาณค่าความสามารถของผู้สอบต่ำสุด

ส่วนการทดสอบเทเลอรัปรีรามิด ที่ให้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า
ความสามารถของผู้สอบ (SEE) สูงสุด ได้แก่ แบบขนาดชั้นแปรผัน ที่มี 7 ชั้น ให้คะแนนตามค่า
ความยากของข้อที่ยากที่สุดที่ท้าทาย (v7s1) ซึ่งไม่สอดคล้องกับสมมุติฐานที่ตั้งไว้ว่า การทดสอบ

เทเลอร์รูปปริมาตรแบบขนาดชั้นคงที่ ที่มี 7 ชั้น ให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำได้ (c7s1) จะให้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบสูงสุด ทั้งนี้ อาจจะเป็นเพราะว่าความสามารถจริงของกลุ่มตัวอย่าง ไม่สอดคล้องกับค่าความยากของข้อสอบ การทดสอบเทเลอร์รูปปริมาตรแบบขนาดชั้นแปรผัน ดังกล่าวแล้วในข้อ 1.

2.2 การทดสอบเทเลอร์รูปปริมาตรทั้งแบบขนาดชั้นคงที่ และแบบขนาดชั้นแปรผัน เมื่อจำนวนชั้นเพิ่มขึ้น ความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ มีแนวโน้มว่าจะมีค่าลดลง เมื่อให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทำได้ ให้คะแนนโดยใช้วิธีของเบลล์ และให้คะแนนโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข ซึ่งก็สอดคล้องกับทฤษฎีการทดสอบ ที่ว่าถ้าเพิ่มความยาวของข้อสอบ ค่าความเชื่อมั่นและค่าความเที่ยงตรงก็จะเพิ่มขึ้น หรือค่าความคลาดเคลื่อนมาตรฐานในการวัดก็จะลดลง (Gulliksen. 1967: 74-105; Lord and Novick. 1968: 112-118) สำหรับการให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำได้ ซึ่งไม่เป็นไปตามทฤษฎี ทั้งนี้ อาจจะเป็นเพราะการให้คะแนนโดยวิธีนี้มีความคลาดเคลื่อนค่อนข้างสูงดังกล่าวแล้ว

2.3 การทดสอบเทเลอร์รูปปริมาตรทั้งแบบขนาดชั้นคงที่ และแบบขนาดชั้นแปรผัน ถ้าใช้จำนวนชั้นตั้งแต่ 10 ชั้นขึ้นไป การให้คะแนนโดยใช้วิธีของเบลล์ จะให้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบต่ำสุด รองลงมาได้แก่ การให้คะแนนโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข ให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทำได้ และ ให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำได้ ตามลำดับ ส่วนการทดสอบเทเลอร์รูปปริมาตรที่ใช้จำนวน 7 ชั้น ยังสรุปไม่ได้ ทั้งนี้ อาจจะเป็นเพราะจำนวนชั้นหรือจำนวนข้อที่ทำให้มีน้อยเกินไป จึงทำให้เกิดความคลาดเคลื่อนมาก

3. จากการศึกษาเปรียบเทียบ ค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ (reliability) จากการทดสอบเทเลอร์รูปปริมาตรที่มี รูปแบบ จำนวนชั้น และวิธีการให้คะแนน ที่แตกต่างกัน พบว่า

3.1 การทดสอบเทเลอร์รูปปริมาตรแบบขนาดชั้นคงที่ ที่มี 16 ชั้น ให้คะแนนโดยใช้วิธีของเบลล์ (c16s3) จะให้ค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบสูงสุด

เทเลอร์รูปปริมาตรแบบขนาดชั้นคงที่ ที่มี 7 ชั้น ให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำได้ (c7s1) จะให้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบสูงสุด ทั้งนี้ อาจจะเป็นเพราะว่าความสามารถจริงของกลุ่มตัวอย่าง ไม่สอดคล้องกับค่าความยากของข้อสอบ การทดสอบเทเลอร์รูปปริมาตรแบบขนาดชั้นแปรผัน ดังกล่าวแล้วในข้อ 1.

2.2 การทดสอบเทเลอร์รูปปริมาตรทั้งแบบขนาดชั้นคงที่ และแบบขนาดชั้นแปรผัน เมื่อจำนวนชั้นเพิ่มขึ้น ความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ มีแนวโน้มว่าจะมีค่าลดลง เมื่อให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทำได้ ให้คะแนนโดยใช้วิธีของเบส์ และให้คะแนนโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข ซึ่งก็สอดคล้องกับทฤษฎีการทดสอบ ที่ว่าถ้าเพิ่มความยาวของข้อสอบ ค่าความเชื่อมั่นและค่าความเที่ยงตรงก็จะเพิ่มขึ้น หรือ ค่าความคลาดเคลื่อนมาตรฐานในการวัดก็จะลดลง (Gulliksen. 1967: 74-105; Lord and Novick. 1968: 112-118) สำหรับการให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำได้ ซึ่งไม่เป็นไปตามทฤษฎี ทั้งนี้ อาจจะเป็นเพราะการให้คะแนนโดยวิธีนี้มีความคลาดเคลื่อนค่อนข้างสูงดังกล่าวแล้ว

2.3 การทดสอบเทเลอร์รูปปริมาตรทั้งแบบขนาดชั้นคงที่ และแบบขนาดชั้นแปรผัน ถ้าใช้จำนวนชั้นตั้งแต่ 10 ชั้นขึ้นไป การให้คะแนนโดยใช้วิธีของเบส์ จะให้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบต่ำสุด รองลงมาได้แก่ การให้คะแนนโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข ให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทำได้ และ ให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำได้ ตามลำดับ ส่วนการทดสอบเทเลอร์รูปปริมาตรที่ใช้จำนวน 7 ชั้น ยังสรุปไม่ได้ ทั้งนี้ อาจจะเป็นเพราะจำนวนชั้นหรือจำนวนข้อที่ให้ทำมีน้อยเกินไป จึงทำให้เกิดความคลาดเคลื่อนมาก

3. จากการศึกษาเปรียบเทียบ ค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ (reliability) จากการทดสอบเทเลอร์รูปปริมาตรที่มี รูปแบบ จำนวนชั้น และวิธีการให้คะแนน ที่แตกต่างกัน พบว่า

3.1 การทดสอบเทเลอร์รูปปริมาตรแบบขนาดชั้นคงที่ ที่มี 16 ชั้น ให้คะแนนโดยใช้วิธีของเบส์ (c16s3) จะให้ค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบสูงสุด

รองลงมาได้แก่ แบบขนาดชั้นแปรรัน ที่มี 16 ชั้น ให้คะแนนโดยวิธีของเบส์ (v16s3) เมื่อทดสอบค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบทั้ง 2 ค่านี้แล้ว ปรากฏว่า ไม่แตกต่างกัน นั่นคือ ความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ที่ได้จากทั้งสองกลุ่มนี้ ถือว่า เป็นพวกที่อยู่ในกลุ่มสูงสุดกลุ่มเดียวกัน ซึ่งก็สอดคล้องกับสมมุติฐาน ที่ตั้งไว้ว่าการทดสอบ เทเลอร์รูปปริมาตรแบบขนาดชั้นแปรรัน ที่มี 16 ชั้น ให้คะแนนโดยวิธีของเบส์ (v16s3) จะให้ค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบสูงสุด

ส่วนที่ให้ค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ (reô) ต่ำสุด ได้แก่แบบขนาดชั้นแปรรัน ที่มี 7 ชั้น ให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำได้ (v7s1) รองลงมาได้แก่ แบบขนาดชั้นคงที่ ที่มี 7 ชั้น ให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำได้ (c7s1) เมื่อทดสอบค่าความเที่ยงตรงในการประมาณค่าทั้ง 2 ค่า นี้แล้ว ปรากฏว่า ไม่แตกต่างกัน นั่นคือ ความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ที่ได้จากทั้งสองกลุ่มนี้ ถือว่า เป็นพวกที่อยู่ในกลุ่มต่ำสุดกลุ่มเดียวกัน ซึ่งก็สอดคล้องกับสมมุติฐานที่ตั้งไว้ว่า การทดสอบเทเลอร์รูปปริมาตรแบบขนาดชั้นคงที่ ที่มี 7 ชั้น ให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำได้ (c7s1) จะให้ค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบต่ำสุด

3.2 ความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ (reô) จากการทดลองมาทั้งหมด มีค่าอยู่ระหว่าง .7787 ถึง .9524 โดยมีนัยสำคัญทางสถิติที่ระดับ .01 ทุกค่า และมีความเที่ยงตรงโดยเฉลี่ย ($\overline{reô}$) ภายในกลุ่มทดลอง มีค่าอยู่ระหว่าง .812 ถึง .948 จะเห็นว่าไม่ว่าจะพิจารณาจากค่าทั้งหมด หรือจากค่าเฉลี่ยภายในกลุ่ม มีค่าไปในทางบวกสูงอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทุกค่า ซึ่งก็สอดคล้องกับผลการวิจัยของ ยูรี (Wiess and McBride. 1984: 274; citing Urry. 1971) ที่พบว่า ความสัมพันธ์ระหว่างความสามารถจริง กับความสามารถที่ได้จากการประมาณค่าโดยวิธีของเบส์ มีค่าอยู่ระหว่าง .919 ถึง .936 และยังสอดคล้องกับผลการวิจัยของ เจนเซมา (Weiss and McBride.1984: 274; citing Jansema. 1972) ที่พบว่า ความสัมพันธ์ระหว่างความสามารถจริงกับความสามารถที่ได้จากการประมาณค่าโดยวิธีของเบส์ มีค่าสูงถึง .92

3.3 การทดสอบเทเลอร์รูปปริมาตรทั้งแบบขนาดชั้นคงที่และแบบขนาดชั้นแปรรัน

เมื่อจำนวนชั้นเพิ่มขึ้น ค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ มีแนวโน้มว่าจะเพิ่มขึ้น ไม่ว่าจะให้คะแนนโดย ให้ตามค่าความยากของข้อที่ยากที่สุดที่ทำได้ ให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทำได้ ให้คะแนนโดยใช่วิธีของเบส์ หรือให้คะแนนโดยใช่วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข ก็ตาม ซึ่งก็สอดคล้องกับทฤษฎีการทดสอบ และ ผลการวิจัยในข้อ 1 และ 2 ดังกล่าวแล้ว

3.4 การทดสอบเทเลอร์ูปิรามิดทั้งแบบขนาดชั้นคงที่และแบบขนาดชั้นแปรผัน ถ้าใช้จำนวนชั้นตั้งแต่ 10 ชั้นขึ้นไป การให้คะแนนโดยใช่วิธีของเบส์ จะให้ความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ($re\hat{o}$) สูงสุด รองลงมาได้แก่ การให้คะแนนโดยใช่วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข ให้ตามค่าเฉลี่ยความยากของข้อที่ทำได้ และให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำได้ ตามลำดับ ซึ่งก็สอดคล้องกับผลการวิจัยในข้อ 1 และ 2 ดังกล่าวแล้ว

ข้อเสนอแนะ

จากผลการวิจัย และปัญหาอุปสรรคที่พบจากการวิจัยครั้งนี้ ผู้วิจัยใคร่จะ เสนอแนะสำหรับผู้ที่นำผลการวิจัยนี้ไปใช้ และผู้ที่คิดจะค้นคว้าวิจัยเรื่องในทำนองนี้ ต่อไป ดังนี้

1. ข้อเสนอแนะสำหรับผู้ที่จะนำผลการวิจัยนี้ไปใช้ จากผลการวิจัยครั้งนี้ได้ชี้ให้เห็นว่าการทดสอบเทเลอร์ูปิรามิดแบบขนาดชั้นคงที่ ที่มี 16 ชั้น ให้คะแนนโดยใช่วิธีของเบส์ เป็นรูปแบบที่มีประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบสูงสุด กล่าวคือ ให้ค่าความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบต่ำ ($E=.246$) ให้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบต่ำ ($SEE=.310$) และ ให้ค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบสูง ($re\hat{o}=.948$) นอกจากนั้นความสามารถที่ได้จากการประมาณค่ากับความสามารถจริง ส่วนใหญ่ไม่แตกต่างกัน จากการทดลองซ้ำ 40 ครั้ง ปรากฏว่าแตกต่างกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 มีเพียงครั้งเดียว แสดงว่าความสามารถของผู้สอบที่ได้จากการประมาณค่า จากการทดสอบเทเลอร์ูปิรามิดแบบขนาดชั้นคงที่ ที่มี 16 ชั้น ให้

คะแนนโดยใช้วิธีของเบส์ นั้น เป็นที่น่าเชื่อถือได้ จึงนำมาทำการทดสอบแบบนี้มาใช้สอบแทนการทดสอบด้วยการเขียนตอบที่ใช้กันอยู่ ทั้งนี้เพราะการทดสอบเทเลอร์มีข้อดีว่าการทดสอบแบบเขียนตอบที่ใช้กันอยู่ หลายประการ ที่สำคัญ คือ ใช้เวลาในการสอบน้อยกว่า เด็กเก่งไม่ต้องเสียเวลาไปทำข้อสอบที่ง่ายเกินไป และเด็กอ่อนก็ไม่เกิดความเครียดที่ต้องทำข้อสอบที่ยากเกินความสามารถ ด้วยเหตุนี้ผู้วิจัยใคร่จะเสนอแนะในการสร้างข้อสอบเทเลอร์รูปแบบนี้ ดังนี้

(1) สร้างตัวข้อสอบ (Item Pool) เนื่องจากการทดสอบเทเลอร์รูปปิรามิคแบบขนาดชั้นคงที่ ที่มี 16 ชั้น จะต้องใช้ข้อสอบทั้งหมด 136 ข้อ และข้อสอบแต่ละข้อจะต้องมีค่าความยากเท่ากับค่าความยากตามที่กำหนด (ดูภาพประกอบ 16) ในทางปฏิบัติอาจจะยากที่จะสร้างข้อสอบให้ได้ค่าความยากเท่ากับค่าที่กำหนดไว้ ดังแผนภาพประกอบ 16 ก็ขอให้มีค่าใกล้เคียงกับค่าดังกล่าว ก็ใช้ได้ ในทางปฏิบัติควรสร้างข้อสอบเกินไว้มาก ๆ แล้วนำไปทดลองใช้เพื่อหาค่าความยากของข้อสอบเป็นรายข้อ การหาค่าความยากของข้อสอบนี้ ไรส์กับคิงส์เบอรี (Weiss and Kingsbury. 1984: 362-364) ได้กล่าวว่า ควรใช้ทฤษฎีการตอบข้อสอบ (Item Response Theory) ที่เป็นแบบ 1-พารามิเตอร์, 2-พารามิเตอร์ หรือ 3-พารามิเตอร์ แบบโลจิสติก (Logistic Model) แบบใดแบบหนึ่งก็ได้ ถ้าใช้แบบ 1-พารามิเตอร์ กลุ่มตัวอย่างที่ใช้ทดลองอาจจะประมาณ 500 คน ก็พอแล้ว แต่ถ้าใช้แบบ 2-พารามิเตอร์ หรือ 3-พารามิเตอร์ กลุ่มตัวอย่างควรเป็นตั้งแต่ 1000 คนขึ้นไป เมื่อทดสอบแล้วก็ทำการวิเคราะห์เพื่อหาค่าความยากค่าอำนาจจำแนก และค่าการเดาของข้อสอบเป็นรายข้อ แล้วจัดข้อสอบให้ได้ค่าความยากตามโครงสร้างใน ภาพประกอบ 16 ก็จะได้ชุดของข้อสอบเทเลอร์รูปปิรามิคแบบขนาดชั้นคงที่ ที่มี 16 ชั้น ตามต้องการ

ตรงกับไรส์ (Weiss and Kingsbury. 1984: 362; citing Traub and Wolk. 1981: 401) ได้ยกปัญหาความเที่ยงตรงตามเนื้อหา (Content Validity) ในการทดสอบเทเลอร์ เพราะผู้สอบแต่ละคนได้รับข้อสอบไม่เหมือนกัน ไรส์และคิงส์เบอรี (Weiss and Kingsbury. 1984: 362) ได้กล่าวว่าปัญหานี้จะลดน้อยลง ถ้าได้กระจายเนื้อหาวิชาในโดเมน (Domain) ให้อยู่ในระดับความยากของข้อสอบแต่ละระดับให้เท่า ๆ กัน

(2) การดำเนินการสอบและการประมาณค่าความสามารถของผู้สอบ เนื่องจากการให้

คะแนนโดยใช้วิธีของเบลล์ มีการใช้สถิติที่ค่อนข้างซับซ้อน และจะต้องมีการประมาณค่าความสามารถของผู้สอบอยู่ตลอดเวลาในขณะที่ทำข้อสอบอยู่ เพื่อความสะดวกควรใช้คอมพิวเตอร์ในการดำเนินการสอบ และประมาณค่าความสามารถของผู้สอบ

2. ข้อเสนอแนะสำหรับผู้ที่จะทำวิจัยในทางนี้ต่อไป จากงานวิจัยนี้ชี้ให้เห็นว่า ถ้าจำนวนขั้นในการทดสอบเทเลอร์รูปปิรามิดเพิ่มขึ้น มีแนวโน้มว่าจะให้ประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบเพิ่มตามด้วย กล่าวคือ จะให้ค่าความคลาดเคลื่อน และค่าความคลาดเคลื่อนมาตรฐาน ในการประมาณค่าความสามารถของผู้สอบลดลง และให้ค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบเพิ่มขึ้น ประเด็นสำคัญอยู่ที่ว่า ควรจะใช้จำนวนขั้นในการทดสอบเทเลอร์รูปปิรามิดสักกี่ขั้นจึงจะเหมาะสม ในเมื่อมองในแง่ของการประหยัด ฉะนั้นจึงควรจะได้มีการวิจัยในเรื่องนี้ต่อ และตัวแปรที่น่าจะศึกษาอีกตัวหนึ่ง คือ ลักษณะของกลุ่มตัวอย่าง โดยแยกศึกษาเป็นกลุ่มที่มีความสามารถจริงสูง ปานกลาง และต่ำ

นอกจากนี้การทดสอบเทเลอร์แบบที่ไม่ได้กำหนดโครงสร้างและเส้นทางในการตอบข้อสอบไว้แน่นอน (Variable-Branching Model) แต่จะใช้คอมพิวเตอร์ในการคัดเลือกข้อสอบให้เหมาะสมกับระดับความสามารถของผู้สอบเป็นรายบุคคล (CAT) เป็นรูปแบบที่น่าสนใจ เพราะในทางปฏิบัติจริงการสร้างชุด หรือกลุ่มของข้อสอบ (Item Pool) ที่จะใช้ในการทดสอบเทเลอร์แบบนี้ สร้างได้ง่ายกว่าการทดสอบเทเลอร์แบบรูปปิรามิด ที่มีการกำหนดค่าความยากและโครงสร้างในการตอบข้อสอบในแต่ละข้อไว้แน่นอน ตัวแปรที่น่าจะศึกษา คือ ลักษณะของกลุ่มข้อสอบที่ใช้ (Item Pool) กฎในการคัดเลือกข้อสอบให้ผู้สอบทำ (Item Selection Rule) และจุดสิ้นสุดในการทดสอบเทเลอร์ (Termination Criterion) เป็นต้น

บรรณานุกรม

บรรณานุกรม

- เกรียงศักดิ์ สุวรรณาค. ความสัมพันธ์ระหว่างรูปแบบบางรูปแบบของแบบทดสอบแยกกลุ่มกับความ
ความสามารถทางการเรียน ในวิชาคณิตศาสตร์. ปรินทิฟอนท์ กศ.ม. กรุงเทพฯ :
มหาวิทยาลัย ศรีนครินทรวิโรฒ ประสานมิตร, 2532. อัดสำเนา.
- งามนิตย์ ธาตุทอง และ ปรีชา เครือธารณ. "การให้คะแนนตัวเองในการสอบแบบทดสอบชนิด
แฟล็กชีลเเวล," รายงานการวิจัย. ขอนแก่น: มหาวิทยาลัย ขอนแก่น, 2527.
- จิราพร ไกรสรคิ่วเวท. ความสัมพันธ์ระหว่างรูปแบบบางประการของแบบทดสอบรูปปริมาตรกับ
ความสามารถทางการเรียน ในวิชาคณิตศาสตร์. ปรินทิฟอนท์ กศ.ม. กรุงเทพฯ :
มหาวิทยาลัย ศรีนครินทรวิโรฒ ประสานมิตร, 2529. อัดสำเนา.
- ฉัตรณา พรหมมา. "Tailored Testing: ผลสำเร็จของการประยุกต์ใช้ Latent Trait
Theory," วารสารการวัดผลการศึกษา. 4(2):45-53; กันยายน-ธันวาคม 2525.
- ชวาล แพร่ดกุล. เทคนิคการวัดผล. กรุงเทพฯ : วัฒนาพานิช, 2509.
- คำย เชียงฉี. การศึกษาเปรียบเทียบความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้
สอบจากการทดสอบเทเลอรัปปริมาตรแบบต่าง ๆ. รายงานการวิจัย, 2533.
อัดสำเนา.
- นันทิยา พึ่งคำ. การเปรียบเทียบคุณภาพการทดสอบแบบซี เอ ที และแบบประเพณีนิยม ในการ
วัดความสามารถด้านคำศัพท์ภาษาอังกฤษ ของนักเรียนชั้นมัธยมศึกษาปีที่ 3. ปรินทิฟอนท์
กศ.ม. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย, 2531. อัดสำเนา.
- บัญชา พนเจริญสวัสดิ์. "ข้อมูลมอนติคาร์โลสำหรับการกระจายปกติ," วิทยาศาสตร์. 11(1):
700-704; พฤศจิกายน 2527.
- เรวดี อินทสระ. การเปรียบเทียบความคลาดเคลื่อนของการเทียบมาตรฐานระหว่างรูปแบบอิง
ทฤษฎีการตอบข้อสอบกับรูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบ. ปรินทิฟอนท์
กศ.ม. กรุงเทพฯ : มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2530. อัดสำเนา.

- เลิศลักษณ์ กลั่นหอม. การศึกษาความคลาดเคลื่อนชนิดที่ 1 และอำนาจของการทดสอบในแบบแผนการทดลองแบบสุ่มภายในบล็อกที่มีปฏิสัมพันธ์ระหว่างสิ่งทดลองกับบล็อก. ปริญญานิพนธ์ กศ.ค. กรุงเทพฯ ฯ: มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2532. อัดสำเนา. สำเร็จ บุญเรืองรัตน์. "การพัฒนาทฤษฎีเลเห็นเทรทเพื่อวิเคราะห์ข้อสอบ," วารสารการวัดผลการศึกษา. 4(2): 1-13; กันยายน-ธันวาคม, 2525.
- _____. ทฤษฎีการวัดและการประเมินผลการศึกษา. กรุงเทพฯ ฯ: สำนักทดสอบทางการศึกษา และจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2527.
- สมชาย ยืนนาน. การศึกษาโดยวิธีมอนติคาร์โลเปรียบเทียบอำนาจของการทดสอบ การเท่ากันของความแปรปรวนระหว่างประชากรสองกลุ่ม. ปริญญานิพนธ์ สด.ม. กรุงเทพฯ ฯ: จุฬาลงกรณ์มหาวิทยาลัย, 2528. อัดสำเนา.
- เสรี.ปรมชวลิตโรจน์. การศึกษาความเที่ยงตรงในการจำแนกความรอบรู้ และการประมาณค่าความสามารถ ในวิชาคณิตศาสตร์ของแบบทดสอบเพ็ล็กซ์เลเวลและแบบทดสอบรูปปิรามิดที่ใช้ขนาดชั้นแปรผัน. ปริญญานิพนธ์ กศ.ม. กรุงเทพฯ ฯ: มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2531. อัดสำเนา.
- ศิริจันทร์ ทองประเสริฐ. การจำลองแบบปัญหา. ภาควิชาวิศวกรรมอุตสาหกรรม จุฬาลงกรณ์มหาวิทยาลัย, 2532.
- Ayala, R.J. De. "A Comparison of the Nominal Response Model and the Three-parameter Logistic Model in Computerized Adaptive Testing," Educational and Psychological Measurement. 49(4): 789-805; 1989.
- Barton, Richard F. A Primer on Simulation and Gaming. Englewood Cliffs: Princetice-Hall, 1970.
- Betz, N.E. and B.J. Weiss. "An Empirical Study of Computer-Administered Two-Stage Ability Testing," Research Report. 73-4. Minneapolis: University of Minnesota, Psychometric Methods Program, Department of Psychology, 1973.
- Birnbaum, Allan. "Some Latent Trait Models and their Use in Inferring in Examinee's Ability," In F.M. Lord and M.R. Novick. Statistical Theories of Mental Test Scores. Reading, Massachusetts: Addison-Wesley, 1968.

- Blackmore, Lois Marie. "Computerized, Computerized Adaptive and Pencil-and-paper Test Administrations: A Comparative Study in a high School Setting," Dissertation Abstracts International. 47(07): 2554-A; January, 1987.
- Bloom, Benjamin. S. Human Characteristics and School Learning. New-York: McGraw-hill Book Company, 1976.
- _____. "The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring," Educational Researcher. 13(6): 4-15; June/July, 1984.
- Charter, Richard A. and Barbara S. Larsen. "Fisher's Z to R," Educational and Psychological Measurement. 43(1): 41-42; 1983.
- Divgi, D.R. "Estimating Reliabilities of Computerized Adaptive Tests," Applied Psychological Measurement. 13(2): 145-149; June, 1989.
- Dodd, Barbara and others. "Operational Characteristics of Adaptive Testing Procedures Using the Graded Response Model," Applied Psychological Measurement. 13(2): 129-143; June, 1989.
- Gifford, Janice A. and Hariharan Swaminathan. "Bias and the Effect of Priors in Bayesian Estimation of Parameters of Item Response Models," Applied Psychological Measurement. 14(1): 33-43; March, 1990.
- Green, Bert F. and others. "Technical Guidelines for Assessing Computerized Adaptive tests," Journal of Educational Measurement. 21(4): 347-360; Winter, 1984.
- Gulliksen, Harold. Theory of Mental Tests. New-York: John Wiley and Sons, 1967.
- Hambleton, R.K. and Hariharan Swaminathan. Item Response Theory. Boston: Kluwer-Nijhoff Publishing, 1985.
- Hambleton, R.K. and Linda L. Cook. "Latent Trait Models and their Use in the Analysis of Educational Test Data," Journal of Educational Measurement. 14(2): 75-96; summer, 1977.
- Hambleton, R.K. and R.E. Thumb. "Analysis of Empirical Data Using two Logistic Latent Trait Models," British Journal of Mathematical and Statistical Psychology. 26: 195-211; 1973.
- Hambleton, R.K. "Applications of Item Response Theory," International Journal of Educational Research. 13(2): 121-220; 1989.
- Hammersley, J.M. and D.C. Handscomb. Monte Carlo Methods. London: Nethuen and Co.Ltd., 1964.

- Hankins, Janette Anne. "The Effects of Variable Entry on Bias and Information of the Bayesian Adaptive Testing Procedure," Dissertation Abstracts International. 47(08): 3013-A; February, 1987.
- Henly, S. J. and others. "Adaptive and Conventional Versions of DAT: The First Complete Test Battery Comparison," Applied Psychological Measurement. 13(4): 363-371; December, 1989.
- Ho, Rong-Guey. "Using MicroCAT in Computerized Adaptive Testing: A Comparison of three Adaptive Testing Strategies," Dissertation Abstracts International. 50(2): 421-A; August, 1989.
- Janosky, Janine Eileen. "A Study of Microcomputer Based Pseudo-Random Number Generators for the Uniform and Normal Probability Distributions," Dissertation Abstracts International. 47(12): 4367-A; June, 1987.
- Kerlinger, Fred N. Foundations of Behavioral Research. Hong Kong: Holt, Rinehart and Winston. Inc. 1988.
- Kirk, Roger E. Experimental Design: Procedures for The Behavioral Sciences. California: Brooks/Cole Publishing Company, 1982.
- Krobe, Ronald W. and Lawren La Forge. The Manager's Guide to Statistics and Quantitative Methods. New York: McGraw-Hill, 1980.
- Larkin, Kevin C. and D.J. Weiss. "An Empirical Comparison of Two-Stage and Pyramidal Adaptive Ability Testing," Research Report. p 27; Psychometric Methods. Program, Department of Psychology, University of Minnesota Minneapolis, 1975.
- Lee, Sue-Sue Susan. "Development and Implementation of Adaptive Testing Strategies for Introductory Graduate Level Courses," Dissertation Abstracts International. 47(10): 3742-A; April, 1987.
- Linden, Wim J. Vander and Michel A. Zwarts. "Some Procedures for Computerized Ability Testing," International Journal of Educational Research. 13(2): 175-187; 1989.
- Lord, F.M. and Melvin R. Novick. Statistical Theories of Mental Test Scores. Reading, Massachusetts: Addison-Wesley, 1968.
- Lord, Frederic M. Applications of Item Response Theory to Practical Testing Problem. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1980.
- _____. "A Theoretical Study of the Measurement Effectiveness of Flexilevel Tests," Educational and Psychological Measurement. 31: 805-813; 1971.

- Lord, Frederic M. "Item Response Theory and Equating a Technical Summary," in Holland, Paul W. and Donald B. Rubin. Test Equating. New York: Academic Press, 1982.
- _____. "Maximum Likelihood and Bayesian Parameter Estimation in Item Response Theory," Journal of Educational Measurement. 23(2): 157-162; Summer, 1986.
- _____. "Robbins-Monro Procedures for Tailored Testing," Educational and Psychological Measurement. 31(1): 3-31; 1971.
- McCauley, Cynthia D. and Jorge Mendoza. "A Simulation Study of Item Bias Using a Two-Parameter Item Response Model," Applied Psychological Measurement. 9(4): 389-400; December, 1985.
- McDonald, R.P. "Linear Versus Non-linear Models in Item Response Theory," Applied Psychological Measurement. 6: 379-396; 1982.
- Meyer, Herbert A. Symposium on Monte Carlo Methods. New York: John Willy and Son, 1956.
- Mihram, G. Arthur. Simulation: Statistical Foundation and Methodology. New York: Academic Press, 1972.
- Moreno, K.E. and others. "Relationship Between Corresponding Armed Services Vocational Aptitude Battery and Computerized Adaptive Testing Subtests," Applied Psychological Measurement. 8: 155-163; 1984.
- Owen, Roger J. "A Bayesian Sequential Procedure for Quantal Response in the Context of Adaptive Mental Testing," Journal of the American Statistical Association. 70(350): 351-356; June, 1975.
- Payne, Jammes A. Introduction to Simulation Programming Techniques and Methods of Analysis. Singapore: McGRAW-HILL International Edition. 1988.
- Rasch, Georg. Probabilistic Model for Some Intelligence and Attainment Tests. Chicago: University of Chicago Press, 1980.
- Reckase, M.D. "Unifactor Latent Trait Models applied to Multifactor Tests: Results and Implications," Journal of Educational Statistics. 4: 207-230; 1979.
- Rudner, Lawrence. "A Closer Look at Latent Trait Parameter Invariance," Educational and Psychological Measurement. 43(4): 951-955; 1983.
- Rubinstein, R.Y. Simulation and the Monte Carlo Method. New York: John Willey and Son, 1981.

- Samejima, F. "A Use of the Information Function in Tailored Testing," Applied Psychological Measurement. 1: 233-247; 1977.
- Skaggs, Gary. and Jose Stevenson. "A Comparison of Pseudo-Bayesian and Joint Maximum Likelihood Procedures for Estimating Item Parameters in the Three-Parameter IRT Model," Applied Psychological Measurement. 13(4): 391-402; December, 1989.
- Spence, Ian. "Monte Carlo Simulation Studies," Applied Psychological Measurement. 7(4): 405-425; Fall, 1983.
- Swarup, Kanti., P.K. Gupta and Man Mohan. Operation Research. Sutan Chandel & Son Publisher, 1982.
- Taha, Hamdy A. Simulation Modeling and Simnet. Englewood Cliffs: Prentice-Hall, Inc., 1988.
- Thorndike, Robert L. Applied Psychometrics. Boston: Houghton Mifflin Company, 1982.
- Urry, Vern W. "Tailored Testing: A Successful Application of Latent Trait Theory," Journal of Educational Measurement. 14(2): 181-196; Summer, 1977.
- Vale, C.D. and David J. Weiss. "A Simulation Study of Stradative Ability Testing," Research Report. Psychometric Method Program. Department of Psychology, University of Minnesota, Minneapolis, 1975.
- Vijver, Fons J.R. Van De. "The Robustness of Rasch Estimates," Applied Psychological Measurement. 10(1): 45-57; 1986.
- Warm, Thomas A. "A Primer of Item Response Theory," Technical Report. 941278. Oklahoma: U.S. Coast Guard Institute, 1978.
- Waters, Carrie Wherry and A.G. Bayroff. "A Comparison of Computer Simulated Conventional and Branching Test," Educational and Psychological Measurement. 31(1): 125-136; 1971.
- Weiss, D.J. and G. Gage Kingsbury. "Application of Computerized Adaptive Testing to Educational Problems," Journal of Educational Measurement. 21(4): 361-375; Winter, 1984.
- Weiss, D.J. and James R. McBride. "Bias and Information of Bayesian Adaptive Testing," Applied Psychological Measurement. 8(3):273-285; Summer, 1984.
- Weiss, D.J. "Improving Measurement Quality and Efficiency with Adaptive Testing." Applied Psychological Measurement. 6: 473-492; 1982.

Weiss, D.J. "Strategies of Adaptive Ability Measurement," Research Report 74-5. Minneapolis: University of Minnesota, Psychometric Methods Program, Department of Psychology, 1974.

_____. "The Stratified Adaptive Computerized Aability Test," Research Report 73-3. Minneapolis: University of Minnesota, Psychometric Methods Program, Department of Psychology, 1973.

Wisniewski, Dennis Russell. "An Application of the Rasch Model to Computerized Adaptive Testing: The Binary Search Method," Dissertation Abstracts International. 47(01): 159-A; July, 1986.

Wright, B.D. and Geofferey N. Masters. Rating Scale Analysis. Chicago: MESA PRESS, 1982.

Wright, B.D. and Mark H. Stone. Best Test Design. Chicago: MESA PRESS, 1979.

การศึกษาเปรียบเทียบประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบ จากการทดสอบ
เทเลอร์รูปปรีมิตที่มี รูปแบบ จำนวนชั้น และวิธีการให้คะแนน ที่แตกต่างกัน
โดยใช้วิธีมอนติ คาร์โล

บทคัดย่อ
ของ
คำย เชียงฉี

เสนอต่อมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษาคุณวุฒิบัณฑิต สาขาการวิจัยและพัฒนาหลักสูตร
กุมภาพันธ์ 2534

การวิจัยนี้ได้นำใช้เทคนิคอิโมนติ-คาร์โล ศึกษาเปรียบเทียบ ค่าความคลาดเคลื่อน ค่าความคลาดเคลื่อนมาตรฐาน และค่าความเที่ยงตรง ในการประมาณค่าความสามารถของผู้สอบจากการทดสอบเทเลอ์รูปปริามิตที่มี รูปแบบ จำนวนชั้น และวิธีการให้คะแนน ที่แตกต่างกัน

วิธีดำเนินการวิจัย โดยการนำเอาคะแนนการตรวจสอบคุณภาพทางการศึกษาทั่วประเทศ ปีการศึกษา 2531 วิชาภาษาไทย ชั้นประถมศึกษาปีที่ 6 จำนวน 106,637 คน มาแปลงเป็นคะแนนมาตรฐานซี แล้วสุ่มคะแนนมาตรฐานซีมาเป็นกลุ่มตัวอย่าง จำนวน 32 กลุ่ม ๆ ละ 1,000 คน และใช้คะแนนมาตรฐานซีนี้ แทนความสามารถจริงของผู้สอบ เครื่องมือที่ใช้ในการวิจัยได้แก่การทดสอบเทเลอ์รูปปริามิต ซึ่งจะศึกษาในตัวแปรอิสระ 3 ตัว คือ หนึ่ง รูปแบบ มี 2 แบบ คือ แบบขนาดชั้นคงที่ และขนาดชั้นแปรผัน สอง จำนวนชั้น มี 4 แบบ คือ แบบ 7, 10, 13 และ 16 ชั้น และสาม วิธีการให้คะแนน มี 4 วิธี คือ ให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทําถูก ให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทําถูก ให้คะแนนโดยใช้วิธีของเบส์ และให้คะแนนโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข รูปแบบการทดลองใช้แบบ คอมพลีทีลี้ แรนดอมไมซ์ แพคทอเรียล ดีไซน์ (CRF_{2x4x4}) ซึ่งมีทั้งหมด 32 กลุ่มการทดลอง ในแต่ละกลุ่มการทดลองได้ทําการทดลองซ้ำ 40 ครั้ง สำหรับการหาค่าโอกาสในการตอบข้อสอบถูกโดยใช้ 1-พารามิเตอร์ แบบโลจิสติก โมเดล

การวิเคราะห์ข้อมูล โดยการหาค่าสถิติเบื้องต้น เปรียบเทียบโดยใช้ ที-เทสต์ การวิเคราะห์ความแปรปรวนสามทาง และเปรียบเทียบเป็นรายคู่โดยใช้วิธีของเซฟเฟ่ ปรากฏผลดังนี้

1. การให้คะแนนโดยให้ตามค่าความยากของข้อที่ยากที่สุดที่ทําถูก จะประมาณค่าความสามารถของผู้สอบสูงกว่าความสามารถจริงอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ในทุกกลุ่มและทุกครั้งของการทดลอง

2. การให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทําถูก จะประมาณค่าความสามารถของผู้สอบต่ำกว่าความสามารถจริงอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ในทุกกลุ่มและทุกครั้งของการทดลอง

3. การให้คะแนนโดยใช้วิธีของเบส์ และวิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข ในแต่ละกลุ่มการทดลอง จากการทดลองซ้ำ 40 ครั้ง จะมีจำนวนครั้งไม่ต่ำกว่า 34 ครั้ง ที่ความ

สามารถจริงกับความสามารถที่ได้จากการประมาณค่า ไม่แตกต่างกัน

4. ความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ มีค่าเฉลี่ย ($\bar{E}$) ภายในกลุ่มทดลอง มีค่าอยู่ระหว่าง .246 ถึง .592 โดยมีส่วนเบี่ยงเบนมาตรฐาน มีค่าอยู่ระหว่าง 0.004 ถึง 0.069

5. ความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ มีค่าเฉลี่ย ($\overline{SEE}$) ภายในกลุ่มทดลอง มีค่าอยู่ระหว่าง .310 ถึง .631 โดยมีส่วนเบี่ยงเบนมาตรฐาน มีค่าอยู่ระหว่าง .005 ถึง .094

6. ความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ($re\hat{o}$) จากการทดลองมาทั้งหมด มีค่าอยู่ระหว่าง .7787 ถึง .9524 โดยมีนัยสำคัญทางสถิติที่ระดับ .01 ทุกค่า และเมื่อหาค่าเฉลี่ยภายในกลุ่มทดลอง ปรากฏว่า มีค่าอยู่ระหว่าง .812 ถึง .948

7. การทดสอบเทเลอร์รูปปริมาตรทั้งแบบขนาดขั้นคงที่ และแบบขนาดขั้นแปรผัน เมื่อจำนวนขั้นเพิ่มขึ้น มีแนวโน้มจะให้ประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบสูงขึ้น เมื่อให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ท้าทาย ให้คะแนนโดยใช้วิธีของเบส์ และให้คะแนนโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข

8. การทดสอบเทเลอร์รูปปริมาตรทั้งแบบขนาดขั้นคงที่ และแบบขนาดขั้นแปรผัน ถ้าใช้จำนวนขั้นตั้งแต่ 10 ขั้นขึ้นไป การให้คะแนนโดยใช้วิธีของเบส์ จะให้ประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบสูงสุด รองลงมาได้แก่ การให้คะแนนโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข และให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ท้าทาย ตามลำดับ

9. การทดสอบเทเลอร์รูปปริมาตรที่ใช้ รูปแบบที่ต่างกัน จำนวนขั้นที่ต่างกัน หรือวิธีการให้คะแนนที่ต่างกัน จะมีผล (Main Effect) ต่อ ค่าความคลาดเคลื่อน ค่าความคลาดเคลื่อนมาตรฐาน และ ค่าความเที่ยงตรง ในการประมาณค่าความสามารถของผู้สอบ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และพบว่ามีปฏิสัมพันธ์ระหว่าง รูปแบบxจำนวนขั้น รูปแบบxวิธีการให้คะแนน จำนวนขั้นxวิธีการให้คะแนน และ รูปแบบxจำนวนขั้นxวิธีการให้คะแนน

10. เมื่อเปรียบเทียบค่าเฉลี่ยเป็นรายคู่ พบว่า การทดสอบเทเลอร์รูปปริมาตรแบบขนาดขั้นคงที่ ที่มี 16 ขั้น ให้คะแนนโดยใช้วิธีของเบส์ จะให้ประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบสูงสุด

A COMPARATIVE STUDY OF EXAMINEES' ABILITY-ESTIMATE EFFICIENCY FROM
PYRAMID TAILORED TESTING WITH DIFFERENT FORMS, STAGES
AND SCORING TECHNIQUES USING MONTE CARLO METHOD

ABSTRACT

BY

TAY CHIENGCHEE

Presented in partial fulfillment of the requirements for the Doctor
of Education Degree in Curriculum Research and Development
at Srinakharinwirot University

February 1991

In this research, Monte Carlo Method was employed to compare errors, standard errors and validity in estimating examinees' abilities of Pyramid Tailored Testing using different forms, stages and scoring techniques.

The nation-wide educational achievement scores in Prathom Suksa-6 Thai Language of 106,637 pupils in the academic year 1988 were transformed into standardized Z scores. These scores were then sampled to make 32 groups of a thousand each. The scores were used as examinees' true abilities. The research instrument was the Pyramid Tailored Testing. Three independent variables were 1) two test forms: constant step-size and variable step-size 2) four types of stages: 7, 10, 13 and 16 stages 3) four scoring techniques: scoring according to the most difficulty of items answered correctly, using average difficulty of items answered correctly, using Bayesian updating estimation and using conditional maximum likelihood estimation. Completely Randomized Factorial Design ($CRF_{2 \times 4 \times 4}$) was applied to the 32 experimental groups. The experiments were repeated 40 times for each group. The one-parameter logistic Model was used to calculate the probability for correct answers.

Data were analyzed through basic statistical methods, t-test and three-way ANOVA. Post hoc multiple comparisons were done by Scheffe's method. The findings were as follows:

1. Scoring according to the most difficulty of items answered correctly yielded higher estimate values than examinees' true abilities at the .01 significance level in all groups and all experiments.

2. Scoring according to average difficulty of items answered

correctly yielded lower estimate values than examinees' true abilities at the .01 significance level in all groups and all experiments.

3. Not less than 34 out of 40 experiments showed no difference between true abilities and estimated abilities, either through scoring by Bayesian updating estimation or by conditional maximum likelihood estimation.

4. the averages of errors ($\bar{E}$) in estimating examinees' abilities in experimental groups were between 0.246-0.592 and S.Ds were between 0.004-0.069.

5. The averages of standard errors ($\overline{SEE}$) in estimating examinees' abilities in experimental groups were between 0.310-0.631 and S.Ds 0.005-0.094.

6. The validity of estimate values ($r_{\theta\hat{\theta}}$) were between 0.7787-0.9524, all at the significance level of .01. The averages within experimental groups were at 0.812-0.948

7. Both types of Tailored Testing with increasing stages, tended to increase the efficiency of estimating when scoring by average difficulty of items answered correctly, by Bayesian updating estimation and by conditional maximum likelihood estimation.

8. When more than 10 stages were used in the test, scoring by Bayesian updating estimation was the most efficient in estimating examinees' abilities. Next were the conditional maximum likelihood estimation and average difficulty of items answered correctly, respectively.

9. Different forms, stages and scoring of Pyramid Tailored

Testing had "main effect" on errors, standard errors and validity in estimating examinees' abilities, significantly at the .01 level. Interaction between forms and stages; forms and scoring techniques; stages and scoring techniques; and between forms, stages and scoring techniques were also found.

10. In post hoc multiple comparisons, it was found that when the 16 stage and constant step-size Pyramid Tailored Testing was applied with Bayesian updating estimation, it was the most efficient in estimating examinees' abilities.