

371, 261

ร 314 ๗

ร.2

ผลของตัวแปรบางตัวต่อความเที่ยงตรงเชิงสภาพและจำนวนข้อสอบที่ใช้ใน
การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์

ปริญญานิพนธ์

ของ

รังสรรค์ มณีเล็ก

เสนอต่อมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษาดุฎิบัณฑิต สาขาการทดสอบและวัดผลการศึกษา

มีนาคม 2540

ลิขสิทธิ์เป็นของมหาวิทยาลัยศรีนครินทรวิโรฒ

คณะกรรมการควบคุมและคณะกรรมการสอบได้พิจารณาปริญญานิพนธ์ฉบับนี้แล้ว
เห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาการศึกษาดุฎบัณฑิต
สาขาการทดสอบและวัดผลการศึกษา ของมหาวิทยาลัยศรีนครินทรวิโรฒได้

คณะกรรมการควบคุม

..... รณ. ธีระพร ประธาน
(ดร.วัน สังข์สะอาด)
..... ธีระพร กรรมการ
(ดร.ชูศักดิ์ ชัมภลจิต)
..... ธีระพร กรรมการ
(รศ.ดร.ศิริชัย กาญจนวาสี)

คณะกรรมการสอบ

..... รณ. ธีระพร ประธาน
(ดร.วัน สังข์สะอาด)
..... ธีระพร กรรมการ
(ดร.ชูศักดิ์ ชัมภลจิต)
..... ธีระพร กรรมการ
(รศ.ดร.ศิริชัย กาญจนวาสี)
..... ธีระพร กรรมการที่แต่งตั้งเพิ่มเติม
(ดร.กมล ภูประเสริฐ)

บัณฑิตวิทยาลัยอนุมัติให้รับปริญญานิพนธ์ฉบับนี้ เป็นส่วนหนึ่งของการศึกษาตาม
หลักสูตรปริญญาการศึกษาดุฎบัณฑิต สาขาการทดสอบและวัดผลการศึกษา ของมหาวิทยาลัย
ศรีนครินทรวิโรฒ

..... ธีระพร คณบดีบัณฑิตวิทยาลัย
(ดร.ศรียา พูลสุวรรณ)
วันที่ 14 เดือน มีนาคม พ.ศ. 2540

ประกาศคุณูปการ

ปริญญานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยดีเป็นเพราะได้รับความกรุณาช่วยเหลือเป็นอย่างดีจาก ดร.วัน สังข์สะอาด ดร.ชูศักดิ์ ชัมภลขิต และรศ.ดร.ศิริชัย กาญจนวาสี ซึ่งได้ให้คำแนะนำในประเด็นต่างๆเสมอมาตั้งแต่เริ่มต้นทำวิจัย ดร.กมล ภูประเสริฐ เลขาธิการคณะกรรมการการประถมศึกษาแห่งชาติ กระทรวงศึกษาธิการ ซึ่งได้ให้ข้อเสนอแนะที่เป็นประโยชน์อย่างยิ่งต่อการวิจัยในครั้งนี้ รวมทั้ง รศ. ดร.ณรงค์ชัย ธนวิบูลย์ชัย ที่ได้กรุณาช่วยตรวจบทคัดย่อภาษาอังกฤษ ผู้วิจัยขอกราบขอบพระคุณคณาจารย์ทุกท่านด้วยความซาบซึ้ง ณ โอกาสนี้

ขอขอบคุณผู้บริหารโรงเรียน คณะครู และ นักเรียนโรงเรียนสังกัดสำนักงานการประถมศึกษากรุงเทพมหานคร สำนักการศึกษากรุงเทพมหานคร และกรมสามัญศึกษา ที่ให้ความร่วมมือเป็นอย่างดีในการรวบรวมข้อมูลสำหรับการวิจัยในครั้งนี้ ขอขอบคุณคุณพิษณุ ช่างสากรล จาก ศูนย์สารสนเทศ กระทรวงศึกษาธิการที่ให้คำแนะนำในการเขียนโปรแกรมคอมพิวเตอร์จนสำเร็จ

สุดท้ายนี้ผู้วิจัยต้องขอขอบคุณสำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ สำนักงานเลขาธิการคุรุสภา มหาวิทยาลัยศรีนครินทรวิโรฒ และสำนักพิมพ์ประสานมิตรที่กรุณา มอบทุนทรัพย์เพื่ออุดหนุนการทำวิจัยในครั้งนี้

รังสรรค์ มณีเล็ก

สารบัญ

บทที่		หน้า
1	บทนำ.....	1
	ความเป็นมาและความสำคัญของปัญหา.....	1
	วัตถุประสงค์ในการวิจัย.....	6
	สมมติฐานในการวิจัย.....	7
	ขอบเขตของการวิจัย.....	10
	ข้อดกลงเบื้องต้น.....	11
	ข้อจำกัดของการวิจัย.....	12
	คำนิยามศัพท์เฉพาะ.....	13
	ประโยชน์ที่คาดว่าจะได้รับ.....	14
2	เอกสารและงานวิจัยที่เกี่ยวข้อง.....	15
	การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ.....	15
	ลักษณะทั่วไป.....	15
	ประเภทของการทดสอบ.....	16
	องค์ประกอบของการทดสอบ.....	26
	การดำเนินการสอบ.....	29
	ตัวบ่งชี้คุณภาพของการทดสอบ.....	29
	การนำการทดสอบไปใช้ทางการศึกษา.....	33
	ข้อดีของการทดสอบ.....	37
	ทฤษฎีการตอบสนองข้อสอบ.....	39
	ประวัติความเป็นมา.....	39
	ลักษณะทั่วไป.....	41
	ข้อดกลงเบื้องต้นของทฤษฎี.....	42
	โมเดลตามทฤษฎี.....	44

สรุปผลการวิจัย..... 171

อภิปรายผลการวิจัย..... 175

ข้อเสนอแนะ..... 181

บรรณานุกรม..... 184

ภาคผนวก..... 189

ประวัติย่อของผู้วิจัย..... 215

บัญชีตาราง

ตาราง		หน้า
1	ลักษณะของข้อมูล โมเดลที่ใช้วิเคราะห์และผู้นำเสนอโมเดล.....	44
2	ค่าอนุพันธ์อันดับหนึ่งและอันดับสองของฟังก์ชันลอกลอไลส์.....	62
3	จำนวนประชากรและกลุ่มตัวอย่างที่ใช้สำหรับสร้างคลังข้อสอบ.....	95
4	โครงสร้างของข้อสอบวัดความเข้าใจ เรื่อง เศษส่วน ชั้นประถมศึกษาปีที่ 6 จำแนกตามเนื้อหาย่อย ฉบับที่ 1-7.....	96
5	จำนวนนักเรียนกลุ่มตัวอย่าง ในการวิจัยระยะที่ 2 จำแนกเป็นรายโรงเรียน และขนาดของโรงเรียน.....	101
6	เงื่อนไขของโปรแกรมคอมพิวเตอร์สำหรับทดสอบนักเรียนในการทดสอบ แบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์.....	103
7	เนื้อเรื่อง จำนวนคาบและข้อสอบที่ใช้วัดความสามารถทางคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6.....	104
8	หน่วยของการวิเคราะห์ความเที่ยงตรงเชิงสภาพและจำนวนข้อสอบที่ใช้ใน การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จำแนกตามความยากง่ายของข้อสอบที่ใช้เมื่อเริ่มต้น เกณฑ์การคัดเลือก ข้อสอบจากคลังข้อสอบ เกณฑ์การยุติการสอบ วิธีการประมาณค่าความ สามารถและความสามารถทางคณิตศาสตร์ของผู้สอบ.....	114
9	จำนวนข้อสอบ จำนวนผู้สอบ ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของค่า พหามิเตอร์ของข้อสอบที่ใช้วัดความเข้าใจ เรื่อง เศษส่วน ชั้นประถมศึกษา ปีที่ 6 ที่ยังไม่ได้ปรับสเกลการวัด.....	119
10	ค่าไอเกน (eigen value) ที่ได้จากการวิเคราะห์ตัวประกอบของแบบทดสอบ วัดความเข้าใจ เรื่อง เศษส่วน ฉบับที่ 1-7 และแบบทดสอบวัดความสามารถ ทางคณิตศาสตร์ ฉบับที่ 8 จำแนกเป็นรายตัวประกอบ (เฉพาะค่าไอเกนที่ มากกว่า1.00)	120

11	ค่าความเที่ยงตรงเชิงสภาพ จำนวนข้อสอบสูงสุด ต่ำสุด ค่าเฉลี่ย และ ค่าเบี่ยงเบนมาตรฐานของจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะ กับความสามารถของผู้สอบด้วยคอมพิวเตอร์ เรื่อง เศษส่วน จำแนกตามค่า ความยากง่ายของข้อสอบข้อแรก.....	131
12	ค่าความเที่ยงตรงเชิงสภาพ จำนวนข้อสอบสูงสุด ต่ำสุด ค่าเฉลี่ย และ ค่าเบี่ยงเบนมาตรฐานของจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะ กับความสามารถของผู้สอบด้วยคอมพิวเตอร์ เรื่อง เศษส่วน จำแนกตามเกณฑ์ การคัดเลือกข้อสอบในคลังข้อสอบ.....	131
13	ค่าความเที่ยงตรงเชิงสภาพ จำนวนข้อสอบสูงสุด ต่ำสุด ค่าเฉลี่ย และ ค่าเบี่ยงเบนมาตรฐานของจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะ กับความสามารถของผู้สอบด้วยคอมพิวเตอร์ เรื่อง เศษส่วน จำแนกตามวิธี การประมาณค่าความสามารถของผู้สอบ.....	132
14	ค่าความเที่ยงตรงเชิงสภาพ จำนวนข้อสอบสูงสุด ต่ำสุด ค่าเฉลี่ย และ ค่าเบี่ยงเบนมาตรฐานของจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะ กับความสามารถของผู้สอบด้วยคอมพิวเตอร์ เรื่อง เศษส่วน จำแนกตามเกณฑ์ ยุติการทดสอบ.....	133
15	ค่าความเที่ยงตรงเชิงสภาพ จำนวนข้อสอบสูงสุด ต่ำสุด ค่าเฉลี่ย และ ค่าเบี่ยงเบนมาตรฐานของจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะ กับความสามารถของผู้สอบด้วยคอมพิวเตอร์ เรื่อง เศษส่วน จำแนกตามความ สามารถทางคณิตศาสตร์ของผู้สอบ.....	133
16	ค่าความเที่ยงตรงเชิงสภาพและค่าไคสแควร์สำหรับทดสอบความแตกต่าง ของค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความ สามารถของผู้สอบด้วยคอมพิวเตอร์.....	134
17	ความแตกต่างของค่าความเที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะ กับความสามารถของผู้สอบด้วยคอมพิวเตอร์ เรื่อง เศษส่วน จำแนกตาม ความยากง่ายของข้อสอบข้อแรก.....	136

18	ความแตกต่างของค่าความเที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะ กับความสามารถของผู้สอบด้วยคอมพิวเตอร์ เรื่อง เศษส่วน จำแนกตาม เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ.....	136
19	ความเที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะกับความสามารถ ของผู้สอบด้วยคอมพิวเตอร์ เรื่อง เศษส่วน จำแนกตามวิธีประมาณค่า ความสามารถและสถิติทดสอบที	137
20	ความแตกต่างของค่าความเที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะ กับความสามารถของผู้สอบด้วยคอมพิวเตอร์ เรื่อง เศษส่วน จำแนกตาม เกณฑ์ยุติการทดสอบ.....	137
21	ความแตกต่างของค่าความเที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะ กับความสามารถของผู้สอบด้วยคอมพิวเตอร์ เรื่อง เศษส่วน จำแนกตาม ความสามารถทางคณิตศาสตร์ของผู้สอบ.....	138
22	ผลการวิเคราะห์ความแปรปรวนของจำนวนข้อสอบที่ใช้ในการทดสอบ แบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จำแนกตาม แหล่งความแปรปรวน	165
23	ความแตกต่างของค่าเฉลี่ยของจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะ กับความสามารถของผู้สอบด้วยคอมพิวเตอร์ เรื่อง เศษส่วน จำแนกตาม ความยากง่ายของข้อสอบข้อแรกที่ใช้เมื่อเริ่มต้นสอบ.....	168
24	ความแตกต่างของค่าเฉลี่ยของจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะ กับความสามารถของผู้สอบด้วยคอมพิวเตอร์ เรื่อง เศษส่วน จำแนกตาม เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ.....	169
25	ความแตกต่างของค่าเฉลี่ยของจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะ กับความสามารถของผู้สอบด้วยคอมพิวเตอร์ เรื่อง เศษส่วน จำแนกตาม เกณฑ์ยุติการทดสอบ.....	169
26	ความแตกต่างของค่าเฉลี่ยของจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะ กับความสามารถของผู้สอบด้วยคอมพิวเตอร์ เรื่อง เศษส่วน จำแนกตาม ความสามารถทางคณิตศาสตร์ของผู้สอบ.....	170

27	ความเที่ยงตรงเชิงสภาพและจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะกับความ ความสามารถของผู้สอบด้วยเครื่องคอมพิวเตอร์ จำแนกตามความยากง่ายของ ข้อสอบข้อแรก เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่า ความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบและความสามารถของผู้สอบ.....	172
----	---	-----

บัญชีภาพประกอบ

ภาพประกอบ

หน้า

1	ความถูกต้องแม่นยำในการวัดที่ระดับความสามารถต่างๆของแบบทดสอบ ที่มีความยากง่ายเฉพาะกลุ่ม แบบทดสอบที่มีความยากง่ายหลายระดับ และการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ.....	4
2	โครงสร้างของการทดสอบสองขั้นตอน.....	19
3	โครงสร้างของแบบทดสอบรูปพีระมิดแบบขนาดขั้นคงที่.....	20
4	โครงสร้างของแบบทดสอบรูปพีระมิดแบบขนาดขั้นแปรผัน.....	20
5	โครงสร้างของแบบทดสอบรูปพีระมิดแบบโรบินส์มอนโร.....	21
6	โครงสร้างของแบบทดสอบรูปพีระมิดแบบสกัดการสะท้อนกลับ.....	22
7	โครงสร้างของแบบทดสอบรูปพีระมิดแบบรักษากการสะท้อนกลับ.....	22
8	โครงสร้างของแบบทดสอบรูปพีระมิดแบบมีหลายข้อในแต่ละขั้น.....	23
9	โครงสร้างของแบบทดสอบรูปพีระมิดแบบให้น้ำหนักแก่ตัวเลือกของข้อสอบ เพื่อแยกทาง.....	24
10	โครงสร้างของการจัดแบบทดสอบแบบเฟลิกซ์เลเวล.....	24
11	โครงสร้างของแบบทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบแบบแบ่งชั้น..	25
12	ระดับผลสัมฤทธิ์ทางการเรียนและช่วงความเชื่อมั่นในการประมาณค่า 95% ของผู้สอบ A และ B.....	34
13	ระดับผลสัมฤทธิ์ทางการเรียนและช่วงความเชื่อมั่นในการประมาณค่า 95% ของผู้สอบ X และ O.....	35
14	ผลสัมฤทธิ์ทางการเรียนและช่วงความเชื่อมั่นในการประมาณค่าจากการทดสอบ 5 ครั้ง คือ ก่อนเรียนและในระหว่างเรียน.....	36
15	โค้งแสดงลักษณะของข้อสอบตามโมเดลโลจิสติกแบบ 3 พารามิเตอร์	46
16	โค้งแสดงลักษณะของข้อสอบตามโมเดลโลจิสติก 1 พารามิเตอร์	47
17	ช่วงความเชื่อมั่น (confidence interval)ของความสามารถที่ได้จากการประมาณค่า.....	50

18	ช่วงแห่งความเชื่อมั่นแอซิมโทติก 95% ของการประมาณค่าความสามารถ.....	51
19	ฟังก์ชันลอกลิเคิลฮูด(log likelihood)สำหรับโมเดลการตอบสนองข้อสอบ 3 แบบ.....	59
20	ที่มาของค่าต่างๆตามวิธีประมาณค่าของนิวตัน-ราฟสัน(Newton-Raphson).....	60
21	แบบแผนของการใช้ข้อสอบต่างชุดและผู้สอบต่างกลุ่ม.....	67
22	แบบแผนของการใช้ข้อสอบร่วมกัน.....	68
23	แบบแผนของการใช้ผู้สอบร่วมกัน.....	68
24	แบบแผนของการใช้ข้อสอบและผู้สอบร่วมกัน.....	69
25	แบบแผนของการใช้ข้อสอบร่วมแบบโยงสลับ.....	69
26	แบบแผนของการใช้ข้อสอบร่วมแบบโยงสลับเป็นกลุ่ม.....	70
27	ตัวอย่างผังงานแสดงขั้นตอนการทำงานของไมโครคอมพิวเตอร์ในการ ดำเนินการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วย คอมพิวเตอร์	107
28	ตัวอย่างเส้นทางของการตอบข้อสอบ ในการทดสอบแบบปรับเหมาะกับความ สามารถของผู้สอบด้วยคอมพิวเตอร์ วิธีหลายขั้นตอน แบบแยกทาง แปรผัน.....	111
29	ค่าไอเกน (eigen value) ที่ได้จากการวิเคราะห์ตัวประกอบของแบบทดสอบ วัดความเข้าใจ เรื่อง เศษส่วน ฉบับที่ 1-7 และแบบทดสอบวัดความสามารถ ทางคณิตศาสตร์ ฉบับที่ 8 จำแนกเป็นรายตัวประกอบ.....	122
30	จำนวนข้อสอบวัดความเข้าใจ เรื่อง เศษส่วน ชั้นประถมศึกษาปีที่ 6 ในคลังข้อสอบ 200 ข้อ จำแนกตามช่วงค่าอำนาจจำแนก.....	123
31	จำนวนข้อสอบวัดความเข้าใจ เรื่อง เศษส่วน ชั้นประถมศึกษาปีที่ 6 ในคลังข้อสอบ 200 ข้อ จำแนกตามช่วงค่าความยากง่าย.....	123
32	จำนวนข้อสอบวัดความเข้าใจ เรื่อง เศษส่วน ชั้นประถมศึกษาปีที่ 6 ในคลังข้อสอบ 200 ข้อ จำแนกตามช่วงค่าการเดา.....	124
33	ค่าเฉลี่ยและพิสัยของค่าอำนาจจำแนก ค่าความยากง่ายและค่าการเดา ของข้อสอบวัดความเข้าใจ เรื่อง เศษส่วน ชั้นประถมศึกษาปีที่ 6 จำนวน 200 ข้อในคลังข้อสอบ.....	125

34	ค่าความน่าจะเป็นโดยเฉลี่ยในการตอบข้อสอบ เรื่อง เศษส่วน ชั้นประถมศึกษา ปีที่ 6 ในคลังข้อสอบได้ถูกต้อง จำแนกแต่ละระดับความสามารถ 13 ระดับ.....	125
35	ค่าสารสนเทศของข้อสอบที่สูงสุด (Imax) ของข้อสอบ เรื่อง เศษส่วน ชั้นประถมศึกษาปีที่ 6 ในคลังข้อสอบ จำแนกแต่ละระดับความสามารถ 13 ระดับ	126
36	โค้งสารสนเทศของแบบทดสอบ (test information curve) วัดความเข้าใจ เรื่อง เศษส่วน ชั้นประถมศึกษาปีที่ 6 จำนวน 200 ข้อในคลังข้อสอบ.....	127
37	ความน่าจะเป็นโดยเฉลี่ยในการตอบข้อสอบวัดความสามารถทาง คณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ได้ถูกต้อง จำแนกแต่ละระดับความ สามารถ 13 ระดับ.....	127
38	ค่าสารสนเทศของข้อสอบที่สูงสุด (Imax) ของข้อสอบวัดความสามารถ ทางคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 จำแนกแต่ละระดับความ สามารถ 13 ระดับ.....	128
39	โค้งสารสนเทศของแบบทดสอบ (test information curve) วัดความสามารถ ทางคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 จำนวน 35 ข้อ.....	129
40	จำนวนนักเรียนจำแนกตามระดับความสามารถทางคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6.....	130
41	ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของ ผู้สอบด้วยคอมพิวเตอร์ จำแนกตามความยากง่ายของข้อสอบข้อแรกและเกณฑ์ การคัดเลือกข้อสอบจากคลังข้อสอบ.....	139
42	ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของ ผู้สอบด้วยคอมพิวเตอร์ จำแนกตามความยากง่ายของข้อสอบข้อแรกและวิธีการ ประมาณค่าความสามารถของผู้สอบ.....	140
43	ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของ ผู้สอบด้วยคอมพิวเตอร์ จำแนกตามความยากง่ายของข้อสอบข้อแรกและความ สามารถของผู้สอบ.....	141

[illegible]

61	ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จำแนกตามความยากง่ายของข้อสอบข้อแรก เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบและความสามารถของผู้สอบ.....	159
62	ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จำแนกตามความยากง่ายของข้อสอบข้อแรก เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบ.....	160
63	ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จำแนกตามความยากง่ายของข้อสอบข้อแรก เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ เกณฑ์ยุติการทดสอบและความสามารถของผู้สอบ.....	161
64	ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จำแนกตามความยากง่ายของข้อสอบข้อแรก วิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบและความสามารถของผู้สอบ.....	162
65	ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จำแนกตามเกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบและความสามารถของผู้สอบ.....	163
66	ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จำแนกตามความยากง่ายของข้อสอบข้อแรก เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบและความสามารถของผู้สอบ.....	164

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

การวัดและประเมินผลนับว่าเป็นองค์ประกอบที่สำคัญของกระบวนการจัดการศึกษา ทั้งนี้ เนื่องจากสิ่งที่ได้จากการวัดผลและประเมินผลนั้นทำให้เราทราบว่าคุณภาพของการจัดการศึกษาบรรลุเป้าหมายที่กำหนดไว้หรือไม่ มากน้อยเพียงไร เครื่องมือที่ถูกสร้างขึ้นมาสําหรับใช้วัดนั้นมีลักษณะแตกต่างกันออกไปตามวัตถุประสงค์ที่สร้าง เนื้อหาที่ใช้วัด ตลอดจนลักษณะของผู้สอบที่ถูกวัดความรู้ ความสามารถ ทักษะหรือเจตคติ ถึงแม้ว่าเครื่องมือวัดผลดังกล่าวอาจจะมีลักษณะที่แตกต่างกันออกไปก็ตาม แต่ก็มีลักษณะที่สอดคล้องกันอยู่ประการหนึ่ง นั่นคือ เครื่องมือวัดผลทุกชนิดต่างก็พยายามวัดคุณลักษณะของผู้สอบ เพื่อลงข้อสรุปเกี่ยวกับระดับความสามารถ หรือ ทักษะของผู้สอบให้ได้ถูกต้องแม่นยำมากที่สุด หรือ เกิดความคลาดเคลื่อนน้อยที่สุด โดยใช้เครื่องมือดังกล่าววัดตัวอย่างของพฤติกรรม แล้วแปลงพฤติกรรมที่วัดผลได้ให้ออกมาในรูปของปริมาณ (Lord, 1980 : 3,12) ทั้งนี้ ก็เพราะการวัดผลทางจิตวิทยา หรือ ทางพฤติกรรมศาสตร์นั้นไม่สามารถวัดคุณลักษณะ หรือ ความสามารถของผู้สอบได้โดยตรงเหมือนกับการวัดทางกายภาพ คุณลักษณะที่ผู้สร้างแบบทดสอบตั้งใจจะวัดเป็นสิ่งที่แฝงอยู่ภายในตัวผู้สอบ ดังนั้น ต้องประเมินโดยอาศัยผลของการตอบสนองต่อสิ่งเร้าในลักษณะใดลักษณะหนึ่งที่ใช้เป็นเครื่องมือวัดผลซึ่งเครื่องมือวัดผลที่ใช้กันอยู่นั้นมีหลายชนิดด้วยกัน แต่ที่ได้รับความนิยมมากก็คือแบบทดสอบ ทั้งนี้ อาจเนื่องมาจากกระบวนการสร้างและการดำเนินการสอบ ตลอดจนการแปลผลที่ได้จากการวัดนั้นค่อนข้างสะดวกและชัดเจน

การจัดชุดของแบบทดสอบแบบประเพณีนิยม (conventional test) ให้แก่กลุ่มผู้สอบนั้นมักจะจัดในลักษณะที่มีจำนวนข้อสอบแน่นอนและใช้สำหรับดำเนินการสอบกับผู้สอบทุกคน ซึ่งสามารถทำได้สองลักษณะ ลักษณะแรกเป็นแบบทดสอบที่มีค่าความยากง่ายเฉพาะระดับใดระดับหนึ่งของความสามารถ (peaked conventional test) ส่วนอีกลักษณะหนึ่งนั้นเป็นแบบทดสอบที่มีค่าความยากง่ายหลายระดับ (rectangular conventional test) แบบทดสอบที่มีความยากง่ายเฉพาะระดับใดระดับหนึ่งนั้นจะเหมาะสมกับผู้สอบที่มีความสามารถเท่ากับหรือใกล้เคียงกับค่าความยากง่ายของข้อสอบที่จัดไว้ แต่จะเป็นข้อสอบที่ยากเกินไปถ้านำไปใช้สอบกับผู้สอบที่มีความสามารถสูงหรือยากเกินไปถ้านำไปใช้สอบกับผู้สอบที่มีความสามารถต่ำ ซึ่งลักษณะเช่นนี้ทำให้

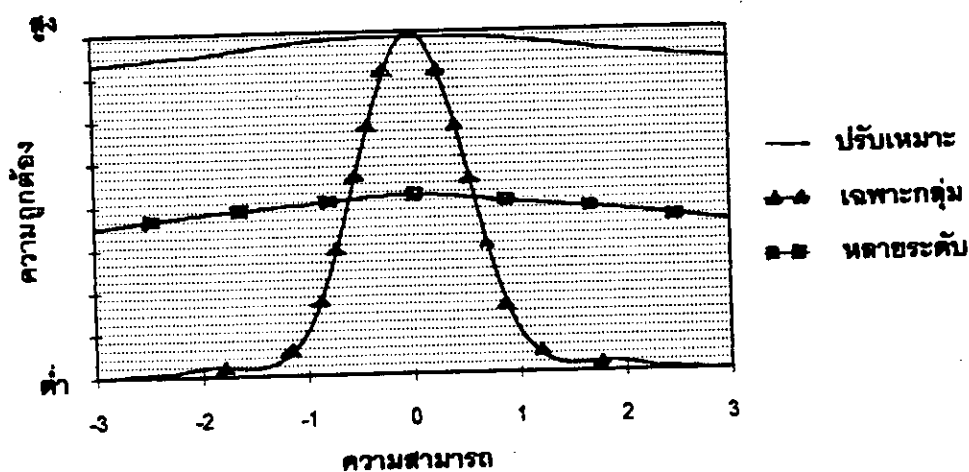
ความถูกต้องของการวัดลดลง ส่วนแบบทดสอบที่มีความยากง่ายหลายๆ ระดับนั้น ข้อสอบในแบบทดสอบส่วนหนึ่งจะเหมาะสำหรับผู้สอบที่มีความสามารถต่ำ ข้อสอบส่วนหนึ่งจะเหมาะสำหรับผู้สอบที่มีความสามารถปานกลางและอีกส่วนหนึ่งจะเหมาะสำหรับผู้สอบที่มีความสามารถสูง ซึ่งมีผลทำให้ความถูกต้องในการวัด (fidelity) มีค่าใกล้เคียงกันเกือบทุกลักษณะของผู้สอบที่สอบแต่ค่อนข้างต่ำ แบบทดสอบที่มีค่าความยากง่ายเฉพาะกลุ่มจะมีความถูกต้องในการวัดสูง โดยเฉพาะอย่างยิ่งกับกลุ่มผู้สอบที่มีความสามารถเท่ากับหรือใกล้เคียงกับค่าความยากง่ายของแบบทดสอบ แต่มีศักยภาพในการบอกความแตกต่างของผู้สอบ (bandwidth) ที่มีความสามารถระดับอื่นๆ ต่ำ ตรงกันข้ามกับกับแบบทดสอบที่มีค่าความยากง่ายหลายๆ ระดับ ซึ่งมีศักยภาพในการบอกความแตกต่างของผู้สอบที่ระดับความสามารถอื่นสูง แต่ความถูกต้องในการวัดจะต่ำ ในการสร้างแบบทดสอบตามลักษณะที่กล่าวมา โดยเฉพาะแบบทดสอบที่ความยากง่ายหลายๆ ระดับนั้นได้รับความนิยมน้อยมากเพราะผู้สร้างถือว่าเป็นการให้สิ่งไร้แก่นแก่นผู้สอบเหมือนกันทุกคน นับว่าเป็นวิธีการควบคุมตัวแปรแทรกซ้อนวิธีหนึ่งในการทดสอบ ทำให้สะดวกต่อการสรุปและแปลความหมายของผลการทดสอบอีกด้วย อย่างไรก็ตามแบบทดสอบที่สร้างและดำเนินการสอบตามทฤษฎีการทดสอบแบบเดิม (classical test theory) นั้นมีข้อจำกัดหลายประการ นั่นคือ เป็นแบบทดสอบที่มีความเหมาะสมกับผู้สอบที่มีความสามารถระดับปานกลาง แต่สำหรับผู้สอบที่มีความสามารถสูงและต่ำแล้ว แบบทดสอบลักษณะนี้จะให้สารสนเทศเกี่ยวกับผู้สอบต่ำ (Vale and Weiss, 1975 : 6) และภายในแบบทดสอบนั้นจะมีข้อสอบส่วนหนึ่งง่ายเกินไปสำหรับผู้สอบที่มีความสามารถสูง และยากเกินไปสำหรับผู้สอบที่มีความสามารถต่ำ ซึ่งนอกจากจะทำให้ผู้สอบเสียเวลาในการตอบข้อสอบที่ไม่เหมาะสมกับความสามารถของตนเองแล้ว ก็ยังอาจทำให้ผู้สอบเกิดความเบื่อหน่ายเมื่อยล้า เกิดการเดาในการตอบข้อที่ยากสำหรับผู้ที่มีความสามารถต่ำ และ เกิดความสับสนวุ่นวายสำหรับการตอบข้อที่ง่ายสำหรับผู้ที่มีความสามารถสูง ซึ่งก่อให้เกิดความผิดพลาดในการแปลผลการทดสอบตามมา นอกจากนั้นยังพบว่า ค่าสถิติที่เป็นดัชนีชี้คุณภาพของข้อสอบ และแบบทดสอบที่ได้มาโดยวิธีการวิเคราะห์ตามทฤษฎีการทดสอบแบบเดิม เช่น ค่าความยากง่าย ค่าอำนาจจำแนกของข้อสอบ ค่าความเชื่อมั่นของแบบทดสอบจะแปรผันไปตามลักษณะของกลุ่มผู้สอบ แบบทดสอบที่ถูกสร้างขึ้นมานั้นจะสามารถใช้ได้ผลดีกับกลุ่มผู้สอบที่มีลักษณะเหมือนกับกลุ่มตัวอย่างที่เคยนำแบบทดสอบไปทดลองใช้มา และในการแสดงความถูกต้องแม่นยำของการวัดนั้น นิยมใช้ค่าความคลาดเคลื่อนมาตรฐานในการวัด (standard error of measurement) เพียงค่าเดียวกับผู้สอบทุกคนที่สอบ ซึ่งเป็นการแสดงค่าแบบเฉลี่ย ดังนั้น จึงทำให้ไม่สามารถเห็นความแปรเปลี่ยนของความคลาดเคลื่อนในการวัดสำหรับผู้สอบที่มีความสามารถแตกต่างกัน

นอกจากนั้น ในการแปลความหมายของความคลาดเคลื่อนในการวัด ก็แปลความหมายเหมือนกับค่าเบี่ยงเบนมาตรฐาน ซึ่งการแปลความหมายในลักษณะนี้จะมีข้อจำกัดหากการแจกแจงไม่อยู่ในรูปโค้งปกติ (Feldt and others. 1985 : 351)

ลอร์ดและโนวิก (Lord and Novick. 1968 : 359) เสนอแนะว่า แบบทดสอบที่ใช้ในการทดสอบควรมีความเหมาะสมกับความสามารถของผู้สอบ นั่นคือ ผู้สอบที่มีความสามารถสูงก็ควรได้รับข้อสอบที่ยาก ส่วนผู้สอบที่มีความสามารถต่ำก็ควรได้รับข้อสอบที่ง่าย ๆ จากหลักการนี้จึงก่อให้เกิดการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ (adaptive testing) ขึ้นมาเพื่อแก้ปัญหาดังกล่าวข้างต้น การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบเป็นการทดสอบที่จัดข้อสอบให้มีความยากง่ายเหมาะสมกับความสามารถของผู้สอบ ซึ่งต่างจากการทดสอบแบบประเพณีนิยมที่ผู้สอบทุกคนต้องได้รับแบบทดสอบฉบับเดียวกันหรือแบบทดสอบที่เป็นคู่ขนานกัน ปัจจุบันการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบได้รับการพัฒนาไปอย่างกว้างขวาง ได้มีการนำคอมพิวเตอร์และทฤษฎีการตอบสนองข้อสอบ (item response theory) เข้ามาช่วยในการวิเคราะห์ข้อมูลและดำเนินการสอบ ในการดำเนินการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบนั้นเริ่มต้นให้ผู้สอบตอบข้อสอบก่อน 1 ข้อ (บางรูปแบบก็อาจให้ตอบมากกว่า 1 ข้อ) ข้อสอบข้อแรกนี้ตามปกติแล้วจะเป็นข้อสอบที่มีค่าความยากง่ายปานกลาง เมื่อผู้สอบตอบเสร็จแล้วก็จะตรวจคำตอบได้ทันทีว่าถูกหรือผิด หลังจากนั้นข้อสอบข้อต่อไปจากคลังข้อสอบก็จะถูกเลือกขึ้นมาให้ผู้สอบได้ตอบ โดยมีความยากง่ายที่เหมาะสมกับความสามารถของผู้สอบ กล่าวคือ ถ้าผู้สอบตอบข้อสอบข้อแรกได้ถูกต้อง ข้อสอบข้อต่อไปที่ถูกเลือกขึ้นมาจากคลังข้อสอบจะยากกว่าเดิม แต่ถ้าตอบข้อสอบข้อแรกผิด ข้อสอบข้อต่อไปจะง่ายกว่าเดิม การทดสอบจะดำเนินการเช่นนี้ไปเรื่อยๆจนกระทั่งถึงเกณฑ์ยุติการสอบ การทดสอบก็จะยุติลงและจะทราบว่าผู้สอบมีความสามารถในเรื่องที่ถูกวัดมาน้อยเพียงไร รูปแบบในการตอบนั้นมีทั้งใช้วิธีการเขียนตอบ และใช้คอมพิวเตอร์ดำเนินการสอบ ซึ่งการทดสอบลักษณะนี้ส่วนใหญ่จะถูกนำไปใช้ในการวัดความสามารถ (ability) ในด้านต่างๆของผู้สอบ นอกจากนี้ยังได้มีการนำเอาการทดสอบลักษณะนี้ไปประยุกต์ใช้ในการวัดบุคลิกภาพ (personality) และเจตคติ (attitude) ด้วยเหมือนกัน (Weiss. 1988 : 372-373) และในปัจจุบันมีแนวโน้มว่าจะมีการประยุกต์ใช้การทดสอบลักษณะนี้มากขึ้นกว่าเดิมโดยเฉพาะทางการศึกษาและอาชีพ

การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบนับว่ามีประโยชน์ต่อการวัดผลการศึกษาเป็นอย่างมาก เนื่องจากประหยัดค่าใช้จ่าย ทั้งนี้เพราะใช้ข้อสอบน้อยกว่าการทดสอบแบบเดิมและไม่จำเป็นต้องสร้างข้อสอบใหม่ทุกครั้งที่มีการสอบ เนื่องจากมีข้อสอบเก็บไว้ในคลัง

ข้อสอบอยู่แล้ว ผู้สอบไม่เกิดความเหน็ดเหนื่อยและความเบื่อหน่ายในการสอบ เพราะใช้เวลาในการทดสอบน้อยลงกว่าเดิม ข้อสอบที่ใช้มีความยากง่ายเหมาะสมกับความสามารถของผู้สอบและสิ่งที่สำคัญที่สุด ก็คือ ได้ผลการทดสอบที่มีความถูกต้องแม่นยำมากกว่าเดิม โดยเฉพาะกับกลุ่มผู้สอบที่มีความสามารถสูงและต่ำมากๆ (Hulin and others. 1983 : 232) ดังจะเห็นได้อย่างชัดเจนจากภาพประกอบต่อไปนี้ ซึ่งได้เปรียบเทียบให้เห็นถึงความถูกต้องในการวัดผลจาก แบบทดสอบ 3 ลักษณะ (Weiss. 1985 : 775)



ภาพประกอบ 1 ความถูกต้องในการวัดที่ระดับความสามารถต่างๆของแบบทดสอบที่มีค่าความยากง่ายเฉพาะกลุ่ม แบบทดสอบที่มีค่าความยากง่ายหลายระดับและการทดสอบแบบปรับเหมาะ กับความสามารถของผู้สอบ

จากที่กล่าวมาและจากผลการวิจัยในอดีตที่ผ่านมา พบว่า การทดสอบแบบปรับเหมาะ กับความสามารถของผู้สอบมีประสิทธิภาพสูงกว่าการทดสอบแบบประเพณีนิยม (Weiss. 1982 : 473-492 ; Morino and others. 1984 : 155-163 ; Garrison and Baumgarten. 1986) และเมื่อนำการทดสอบแบบปรับเหมาะ กับความสามารถของผู้สอบแบบเขียนตอบ (paper and pencil adaptive testing) มาทำการเปรียบเทียบกับ การทดสอบแบบปรับเหมาะ กับความสามารถของผู้สอบด้วยคอมพิวเตอร์ (computerized adaptive testing , CAT) ก็พบว่า การทดสอบแบบปรับเหมาะ กับความสามารถของผู้สอบด้วยคอมพิวเตอร์มีประสิทธิภาพสูงกว่าการทดสอบแบบปรับเหมาะ กับความสามารถของผู้สอบแบบเขียนตอบ (Urry. 1977 : 181) การทดสอบแบบปรับเหมาะ กับความสามารถของผู้สอบด้วยคอมพิวเตอร์มีองค์ประกอบที่สำคัญอยู่ 6 องค์ประกอบ คือ

1. โมเดลของการตอบสนองข้อสอบ (item response model)
2. คลังข้อสอบ (item pool)
3. ความยากง่ายของข้อสอบข้อแรก (entry level)
4. เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ (item selection rule)
5. วิธีการประมาณค่าความสามารถของผู้สอบ (ability estimation procedure)
6. เกณฑ์ยุติการทดสอบ (termination criteria)

นับตั้งแต่เริ่มมีการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์เป็นต้นมา ผู้ที่เกี่ยวข้องในวงการวัดผลการศึกษาที่มีความสนใจวิธีดำเนินการสอบลักษณะนี้ต่างก็ทำการศึกษา ค้นคว้าและพัฒนารูปแบบของการทดสอบ เพื่อให้เป็นการทดสอบที่มีประสิทธิภาพ ซึ่งการวิจัยดังกล่าวมีหลายประเด็นด้วยกัน เช่น เรื่อง โมเดลของการตอบสนองข้อสอบนั้น พบว่า โมเดลการตอบสนองข้อสอบตามทฤษฎีการตอบสนองข้อสอบแบบโลจิสติกแบบ 3 พารามิเตอร์นั้น มีความเหมาะสมกว่าโมเดลอื่นๆ (Urry. 1970,1977 ; McKinley and Reckase. 1980 ; Koch and Reckase. 1978) สำหรับในส่วนที่เกี่ยวกับคลังข้อสอบนั้น พบว่า ขนาดของคลังข้อสอบส่งผลต่อการประมาณค่าความสามารถของผู้สอบน้อยมาก (Ho. 1989 : 421-A) และขนาดของคลังข้อสอบนั้นควรมีประมาณ 100-200 ข้อ และคลังข้อสอบที่มีข้อสอบ 116-150 ข้อจะให้ผลดีที่สุด (Weiss. 1988 : 374-375) ซึ่งสอดคล้องกับยูริ (Urry. 1970,1977) ที่พบว่า คลังข้อสอบที่ใช้ควรมีอย่างน้อย 100 ข้อ ส่วนข้อสอบในคลังข้อสอบนั้น ควรมีค่าอำนาจจำแนกมากกว่า 0.8 ค่าความยากง่ายอยู่ระหว่าง -2.0 ถึง 2.0 และค่าสัมประสิทธิ์การเดาน้อยกว่า .3 นอกจากนี้ โฮ (Ho. 1989 : 421-A) ยังพบว่าชนิดของคลังข้อสอบ(มีข้อสอบแบบสุ่ม,ข้อสอบที่มีค่าอำนาจจำแนกสูงและข้อสอบที่ง่าย) นั้น มีปฏิสัมพันธ์กับวิธีการประมาณค่าความสามารถของผู้สอบ

ในประเด็นที่เกี่ยวกับระดับความยากง่ายของข้อสอบที่ใช้เมื่อเริ่มต้นในการสอบนั้น ลอร์ด (Lord. 1980 : 153 ; Hulin and others. 1983 : 219 ; citing Lord. 1977) พบว่า ค่าความยากง่ายของข้อสอบข้อแรกที่เลือกให้แก่ผู้สอบนั้นไม่มีผลต่อความถูกต้องในการประมาณค่าความสามารถของผู้สอบ แต่ โฮ (Ho. 1989 : 421-A) พบว่า ถ้าข้อสอบข้อแรกมีความยากง่ายเท่ากับหรือต่ำกว่าความสามารถของผู้สอบแล้ว จะทำให้การประมาณค่าความสามารถของผู้สอบมีความถูกต้องมากยิ่งขึ้น ในด้านวิธีการประมาณค่าความสามารถของผู้สอบนั้น ไวส์และแม็คไบรด์ (Weiss and McBride. 1984 : 274) พบว่า วิธีการประมาณค่าความสามารถตามวิธีของเบส์ที่ปรับปรุงใหม่ (bayesian updating) เป็นวิธีการประมาณค่าที่ไม่ลำเอียง ต่าย เชียงฉี (2534) พบว่า วิธีการประมาณค่าความสามารถผู้สอบตามวิธีของเบส์ที่ปรับปรุงใหม่ (bayesian updating) และวิธี

ความเป็นไปได้สูงสุดแบบมีเงื่อนไข (conditional maximum likelihood) ให้ผลไม่ต่างกัน ซึ่งให้ผลสอดคล้องกับข้อค้นพบของสเกจส์และสตีเวนสัน (Skaggs and Stevenson, 1989 : 391-402) ที่ได้ระบุว่า ถ้าผู้สอบมากกว่า 2000 คนแล้ว ผลที่ได้รับจากการประมาณค่าความสามารถทั้ง 2 วิธีไม่แตกต่างกัน แต่ถ้ามีผู้สอบน้อยกว่า 500 คนแล้ว วิธีการของเบส์จะมีความคงที่และถูกต้องมากกว่า และถ้าข้อสอบที่ใช้ในการสอบน้อยกว่า 15 ข้อแล้ว วิธีความเป็นไปได้สูงสุดจะมีความถูกต้องมากกว่า จากที่กล่าวมา มีบางส่วนสอดคล้องกับ ข้อค้นพบของ กิฟฟอร์ดและสวามินาธาน (Gifford and Swaminathan, 1990 : 33) ซึ่งพบว่า ถ้ากลุ่มผู้สอบมีขนาดเล็กและข้อสอบที่ใช้มีจำนวนน้อยข้อแล้ว การประมาณค่าความสามารถด้วยวิธีของเบส์แบบร่วมกัน (joint bayesian) จะมีความถูกต้องมากกว่า ส่วนอีกประเด็นหนึ่งนั้นไม่เกี่ยวกับองค์ประกอบของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์โดยตรง นั่นคือ เจตคติของผู้สอบต่อการสอบ จากการศึกษาพบว่า ผู้สอบมีเจตคติที่ดีต่อการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์และการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบแบบเขียนตอบ มากกว่า การทดสอบแบบเดิม (Wisniewski, 1985 ; Garrison and Baumgarten, 1986 ; Lee, 1987 ; Blackmore, 1987 ; Carroll, 1990 ; Vispoel, 1993 ; งามนิทย์ ธาตุทองและปรีชา เครือวัลย์, 2527)

จากผลการวิจัยที่กล่าวมาข้างต้น พบว่า มีตัวแปรบางตัวได้รับการยืนยันอย่างชัดเจนว่า ควรจะมีลักษณะเช่นไรจึงจะทำให้การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์มีประสิทธิภาพสูงสุด ประเด็นที่มีความชัดเจน ก็คือ โมเดลการตอบสนองข้อสอบ ขนาดของคลังข้อสอบและเจตคติต่อการสอบ ส่วนตัวแปรที่ยังไม่ได้มีการศึกษาหรือผลของการศึกษายังไม่ค่อยมีความชัดเจน ได้แก่ ค่าความยากง่ายของข้อสอบที่ใช้เมื่อเริ่มต้นการทดสอบ เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบและความสามารถของผู้สอบ ดังนั้น จึงทำให้ผู้วิจัยมีความสนใจที่จะศึกษาว่า ตัวแปรที่ยังไม่ได้รับการศึกษา หรือ ผลของการศึกษายังไม่ค่อยชัดเจนดังกล่าวจะส่งผลต่อความเที่ยงตรงเชิงสภาพและจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์หรือไม่ มากน้อยเพียงไร

วัตถุประสงค์ในการวิจัย

1. เพื่อเปรียบเทียบความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ เมื่อ

1.1 ความยากง่ายของข้อสอบข้อแรกต่างกัน

- 1.2 เกณฑ์การคัดเลือกข้อสอบต่างกัน
- 1.3 วิธีการประมาณค่าความสามารถของผู้สอบต่างกัน
- 1.4 เกณฑ์ยุติการทดสอบต่างกัน
- 1.5 ความสามารถของผู้สอบต่างกัน

2. เพื่อเปรียบเทียบจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะกับความ
สามารถของผู้สอบด้วยคอมพิวเตอร์ เมื่อ

- 2.1 ความยากง่ายของข้อสอบข้อแรกต่างกัน
- 2.2 เกณฑ์การคัดเลือกข้อสอบต่างกัน
- 2.3 วิธีการประมาณค่าความสามารถของผู้สอบต่างกัน ✓
- 2.4 เกณฑ์ยุติการทดสอบต่างกัน
- 2.5 ความสามารถของผู้สอบต่างกัน ✓

สมมติฐานในการวิจัย

การวิจัยครั้งนี้ ผู้วิจัยได้ตั้งสมมติฐานของการวิจัยและเหตุผลของการตั้งสมมติฐานแยก
เป็นประเด็นต่างๆไว้ โดยสอดคล้องกับวัตถุประสงค์ในการวิจัย และเพื่อให้ผู้อ่านได้มองเห็นภาพ
ของผลที่อาจเกิดขึ้นต่อตัวแปรตามได้ชัดเจนยิ่งขึ้น ผู้วิจัยจึงได้กล่าวถึงผลของตัวแปรตามทั้ง 2 ตัว
อันเนื่องมาจากตัวแปรอิสระ 1 ตัว ไว้ในสมมติฐานข้อเดียวกัน ดังนี้

1. เมื่อใช้ข้อสอบข้อแรกที่มีค่าความยากง่ายต่างกัน ค่าความเที่ยงตรงเชิงสภาพของ
แบบทดสอบน่าจะไม่ต่างกัน ส่วนจำนวนข้อสอบที่ใช้ในการทดสอบนั้นน่าจะต่างกัน ถ้าใช้ข้อสอบ
ข้อแรกที่มีความยากง่ายปานกลาง น่าจะใช้ข้อสอบน้อยกว่าเริ่มต้นทดสอบด้วยข้อสอบที่ยากและ
ง่าย

ทั้งนี้เนื่องจากลอร์ด (Lord. 1980 : 151 ; Hulin and others. 1983 : 219 ; citing Lord.
1977) พบว่า ข้อสอบที่ใช้เมื่อเริ่มต้นในการสอบนั้นไม่มีผลต่อความถูกต้องในการประมาณค่าความ
สามารถของผู้สอบ ซึ่งความถูกต้องในการประมาณค่าความสามารถของผู้สอบนั้นมีความสัมพันธ์
แบบผกผันกับค่าความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ และเนื่องจาก
ความเที่ยงตรงนั้น มีค่าไม่เกินรากที่สองของความเชื่อมั่น เมื่อข้อสอบที่ใช้เริ่มต้นในการสอบไม่มี
ผลต่อความถูกต้องในการประมาณค่าหรือความคลาดเคลื่อนในการประมาณค่า ความสามารถ
ของผู้สอบ ดังนั้นก็จะไม่มีผลต่อความเที่ยงตรงของแบบทดสอบด้วย ส่วนจำนวน ข้อสอบที่ใช้ใน
การทดสอบจะต่างกัน ทั้งนี้เนื่องจากข้อสอบเริ่มต้นที่มีค่าความยากง่ายใกล้เคียงกับความสามารถ

ของผู้สอบนั้นจะช่วยให้การประมาณค่าความสามารถของผู้สอบเป็นไปอย่างรวดเร็ว โดยใช้ข้อสอบน้อยกว่าการทดสอบที่ใช้ข้อสอบเริ่มต้นที่มีค่าความยากง่ายไม่ใกล้เคียงกับความสามารถของผู้สอบ

2. เมื่อใช้เกณฑ์การคัดเลือกข้อสอบต่างกัน ค่าความเที่ยงตรงเชิงสภาพของแบบทดสอบน่าจะไม่ต่างกัน แต่จำนวนข้อสอบที่ใช้น่าจะต่างกัน การคัดเลือกข้อสอบที่มีค่าสารสนเทศสูงสุดน่าจะใช้ข้อสอบน้อยกว่าการคัดเลือกข้อสอบด้วยวิธีการอื่น

ทั้งนี้เนื่องจากความเที่ยงตรงของแบบทดสอบนั้นจะขึ้นอยู่กับความถูกต้องในการวัด ซึ่งในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์นั้นสามารถจะกำหนดค่าความถูกต้องในการวัดให้เป็นค่าเดียวกันสำหรับผู้สอบทุกคนได้ ดังนั้น จึงทำให้ความเที่ยงตรงของแบบทดสอบไม่ต่างกัน แม้ว่าจะใช้วิธีการคัดเลือกข้อสอบวิธีการใดก็ตาม ส่วนจำนวนข้อสอบที่ใช้ในการทดสอบนั้นจะแตกต่างกัน เพราะข้อสอบที่ถูกเลือกขึ้นมาใช้ในแต่ละวิธีนั้นจะมีค่าสารสนเทศของข้อสอบแตกต่างกันไป ซึ่งค่าสารเทศดังกล่าวจะส่งผลต่อค่าความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ ดังนั้น จึงทำให้วิธีการคัดเลือกข้อสอบแต่ละวิธีสู่เข้าหาความคลาดเคลื่อนในการประมาณค่าที่กำหนดเป็นเกณฑ์ยุติการสอบเร็วช้าต่างกัน ซึ่งการเร็วช้าต่างกันนี้ก็คือการใช้จำนวนข้อสอบในการทดสอบต่างกันนั่นเอง

3. เมื่อใช้เกณฑ์ยุติการทดสอบต่างกัน ค่าความเที่ยงตรงเชิงสภาพของแบบทดสอบน่าจะต่างกันและจำนวนข้อสอบที่ใช้ในการทดสอบก็น่าจะต่างกันด้วย ถ้าใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถที่มีค่าต่ำๆเป็นเกณฑ์ยุติการทดสอบ ค่าความเที่ยงตรงเชิงสภาพของการทดสอบน่าจะสูงกว่าและใช้ข้อสอบมากกว่าใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถที่มีค่าสูงๆ

ทั้งนี้เนื่องจากความเที่ยงตรงของแบบทดสอบนั้นจะแปรผันตามค่าความคลาดเคลื่อนในการประมาณค่าดังเหตุผลที่กล่าวมาแล้วข้างต้น ดังนั้น เมื่อกำหนดค่าความคลาดเคลื่อนในการประมาณค่าความสามารถแตกต่างกันเป็นเกณฑ์ยุติการสอบ จึงทำให้ค่าความเที่ยงตรงของแบบทดสอบแตกต่างกันไปด้วย สำหรับจำนวนข้อสอบที่ใช้ในการทดสอบนั้น เมื่อจำนวนข้อสอบที่ใช้เพิ่มขึ้น ค่าความคลาดเคลื่อนในการประมาณค่าก็จะเกิดการเปลี่ยนแปลง ดังนั้น หากกำหนดค่าความคลาดเคลื่อนมาตรฐานให้มีค่าต่างกัน จำนวนข้อสอบที่ถูกนำมาใช้ในการทดสอบก็จะต่างกัน

4. เมื่อใช้วิธีการประมาณค่าความสามารถต่างกัน ค่าความเที่ยงตรงเชิงสภาพของแบบทดสอบน่าจะไม่ต่างกัน แต่จำนวนข้อสอบที่ใช้ในการทดสอบน่าจะต่างกัน วิธีของเบส์ที่ปรับใหม่น่าจะใช้ข้อสอบน้อยกว่าวิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข

ทั้งนี้เนื่องจากมีข้อค้นพบว่า ถ้ากลุ่มตัวอย่างมีขนาดใหญ่ (มากกว่า 2000 คน) และทดสอบโดยใช้ข้อสอบมากกว่า 35 ข้อแล้ว การประมาณค่าความสามารถตามวิธีการของเบส์ และวิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไขจะให้ผลการประมาณค่าไม่ต่างกันและมีความถูกต้องสูง และถึงแม้ว่ากลุ่มผู้สอบมีขนาดเล็ก (น้อยกว่า 500 คน) วิธีการของเบส์แบบประยุกต์ (pseudo-bayesian) จะให้ค่าการประมาณที่มีความคงเส้นคงวา (consistency) และมีความถูกต้องมากกว่าวิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไขก็ตาม แต่ก็ไม่ได้ทำให้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าแตกต่างกัน เพราะการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ครั้งนี้ได้กำหนดให้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถเป็นเกณฑ์ยุติการทดสอบ ซึ่งจะมีค่าเท่ากันสำหรับผู้สอบทุกคน ดังนั้นจึงส่งผลในความเที่ยงตรงของแบบทดสอบมีค่าไม่ต่างกันเมื่อใช้วิธีการประมาณค่าความสามารถของผู้สอบต่างกัน แต่จำนวนข้อสอบที่ใช้จะต่างกันเพราะที่ระดับความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถเดียวกันนั้น วิธีการประมาณค่าความสามารถตามวิธีการของเบส์จะใช้ข้อสอบน้อยกว่าวิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (ต่าย เชียงฉิน, 2530 : 122)

5. ผู้สอบที่มีความสามารถต่างกัน น่าจะไม่ส่งผลทำให้ค่าความเที่ยงตรงเชิงสภาพของแบบทดสอบต่างกัน ส่วนจำนวนข้อสอบที่ใช้ในการทดสอบนั้นน่าจะต่างกัน ผู้ที่มีความสามารถสูงน่าจะใช้ข้อสอบน้อยกว่าผู้ที่มีความสามารถต่ำ

ทั้งนี้เนื่องจากวัตถุประสงค์ของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์นั้นก็เพื่อวัดคุณลักษณะของผู้สอบโดยใช้ข้อสอบที่มีความเหมาะสมกับความสามารถของผู้สอบ ซึ่งผู้สอบที่มีความสามารถสูงก็จะได้รับข้อสอบที่ยาก สำหรับผู้สอบที่มีความสามารถต่ำก็จะได้รับข้อสอบที่ง่าย ดังนั้นจึงทำให้เกิดความเที่ยงตรงเชิงสภาพค่อนข้างสูงแต่ไม่แตกต่างกันถึงแม้ว่าผู้สอบจะมีความสามารถต่างกันก็ตาม ส่วนจำนวนข้อสอบที่ใช้ในการทดสอบจะต่างกัน เพราะผู้สอบที่มีความสามารถต่ำอาจตอบข้อสอบด้วยการเดาถึงแม้ว่าจะเป็นข้อสอบที่มีความเหมาะสมกับความสามารถของผู้สอบแล้วก็ตาม เมื่อการทดสอบเป็นลักษณะนี้จะทำให้ผู้สอบดังกล่าวจะต้องตอบข้อสอบมากขึ้นกว่าเดิมกว่าจะประมาณค่าความสามารถของผู้สอบให้มีค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าเท่ากับที่กำหนดไว้

ขอบเขตของการวิจัย

1. แบบทดสอบวัดความสามารถทางคณิตศาสตร์ของผู้สอบจะวัดเฉพาะสมรรถภาพ ความรู้ความเข้าใจเกี่ยวกับพื้นฐานทางจำนวน พีชคณิต การวัด เรขาคณิต ระดับชั้นประถมศึกษา ปีที่ 6 ตามที่กำหนดไว้ในหลักสูตรประถมศึกษา พุทธศักราช 2521

2. ตัวแปรที่ศึกษาในครั้งนี้ประกอบด้วย

2.1 ตัวแปรอิสระ

2.1.1 ความยากง่ายของข้อสอบข้อแรก

- (1) ข้อสอบที่ง่าย ($b=-1.3$)
- (2) ข้อสอบที่ยากปานกลาง ($b=0$)
- (3) ข้อสอบที่ยาก ($b=1.3$)

2.1.2 เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ

- (1) วิธีการจับคู่ระหว่างค่าความยากง่ายของข้อสอบ
(b) กับค่าความสามารถโดยประมาณของผู้สอบ ($\hat{\theta}$)
- (2) วิธีการจับคู่ระหว่างค่าความสามารถที่ทำให้ข้อสอบมี
สารสนเทศสูงสุด (θ_{max}) กับค่าความสามารถ
โดยประมาณของผู้สอบ ($\hat{\theta}$)
- (3) วิธีใช้ข้อสอบที่มีค่าสารสนเทศสูงสุด (maximum
item information , I_{max})

2.1.3 วิธีการประมาณค่าความสามารถของผู้สอบ

- (1) วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (conditional
maximum likelihood)
- (2) วิธีของเบย์ที่ปรับเปลี่ยน (Bayesian updating)

2.1.4 เกณฑ์ยุติการทดสอบ

- (1) ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า
ความสามารถ เท่ากับหรือน้อยกว่า .30
- (2) ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า
ความสามารถ เท่ากับหรือน้อยกว่า .40
- (3) ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า
ความสามารถ เท่ากับหรือน้อยกว่า .50

2.1.5 ความสามารถทางคณิตศาสตร์ของผู้เข้าสอบ

- (1) ความสามารถสูง $(\hat{\theta} > .60)$
- (2) ความสามารถปานกลาง $(-.50 < \hat{\theta} < .50)$
- (3) ความสามารถต่ำ $(\hat{\theta} < -.60)$

2.2 ตัวแปรตาม

2.2.1 ความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความ

ความสามารถของผู้สอบด้วยคอมพิวเตอร์ ซึ่งหาได้จากค่าสหสัมพันธ์ระหว่างค่าความสามารถทางคณิตศาสตร์กับค่าความสามารถเรื่องเศษส่วน

2.2.2 จำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะความสามารถของผู้สอบด้วยคอมพิวเตอร์ เมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถมีค่าเท่ากับหรือน้อยกว่าเกณฑ์ที่กำหนด

ข้อตกลงเบื้องต้น

1. ผู้สอบที่ตอบข้อสอบโดยใช้เวลาในการตอบนานกว่า 30 นาที จะถูกตัดชื่อออกจาก การประมาณค่าความสามารถของผู้สอบเนื่องจากข้อสอบชุดนี้ไม่เหมาะสมกับความสามารถของ ผู้สอบดังกล่าว
2. ข้อสอบที่นำผลการตอบมาวิเคราะห์หองค์ประกอบนั้น ถ้าตัวประกอบตัวที่แรกมีค่า ไอเกน (eigen value) มากกว่าตัวประกอบที่ 2 ตั้งแต่ 3 เท่าขึ้นไปและค่าไอเกนของตัวประกอบอื่นๆ มีขนาดใกล้เคียงกันแล้ว หรือ ตัวประกอบที่ 1 สามารถอธิบายความแปรปรวนได้ตั้งแต่ร้อยละ 70 ขึ้นไปถือว่าข้อสอบเหล่านั้นมีความเป็นมิติเดียว
3. การแบ่งระดับความยากง่ายของข้อสอบออกเป็น 3 ระดับ คือ ข้อสอบที่ง่าย ยาก ปานกลางและข้อสอบที่ยากนั้น ผู้วิจัยใช้เกณฑ์การแบ่งที่เป็นข้อค้นพบของคณิต ไชยมุทร์ (2533) และค่าความยากง่ายของข้อสอบ 3 ระดับที่ใช้เมื่อเริ่มต้นในการทดสอบนั้น ผู้วิจัยกำหนดค่าของ ความยากง่าย เท่ากับ -1.30 , 0.00 และ 1.30 ซึ่งค่าความยากง่ายเหล่านี้เป็นจุดกึ่งกลางของช่วง ความยากง่ายของข้อสอบทั้ง 3 ระดับดังกล่าว
4. ค่าความสามารถเบื้องต้นของผู้สอบในการประมาณค่าความสามารถวิธีความเป็นไปได้สูงสุดนั้น ได้ใช้ค่าความสามารถทางคณิตศาสตร์โดยเฉลี่ยของกลุ่มผู้สอบเป็นตัวกำหนด

ข้อจำกัดของการวิจัย

1. การวิจัยครั้งนี้มีข้อจำกัดในเรื่องของเวลา จึงทำให้ผู้วิจัยจำเป็นต้องใช้นักเรียนชั้นประถมศึกษาปีที่ 6 ปลายภาคเรียนที่ 2 ปีการศึกษา 2538 (กุมภาพันธ์ 2539) เป็นกลุ่มตัวอย่างสำหรับทดลองเครื่องมือเพื่อใช้ในการวิจัย ระยะที่ 2 และใช้นักเรียนชั้นมัธยมศึกษาปีที่ 1 ในตอนต้นภาคเรียนที่ 1 ปีการศึกษา 2539 (มิถุนายน 2539) ซึ่งยังไม่ได้เรียนเนื้อหาใหม่ของชั้นมัธยมศึกษาปีที่ 1 เพื่อศึกษาผลของตัวแปรอิสระต่อตัวแปรตาม

2. ค่าพารามิเตอร์ของข้อสอบในคลังข้อสอบไม่ได้กระจายเป็นช่วงเท่าๆกัน และในแต่ละช่วงก็ไม่มีข้อสอบหลายข้อ จึงทำให้การคัดเลือกข้อสอบในบางครั้งจากคลังข้อสอบไม่ได้ข้อสอบตามที่ต้องการและจำเป็นต้องใช้ข้อสอบที่มีคุณลักษณะใกล้เคียงขึ้นมาแทน

ค่านิยามศัพท์เฉพาะ

1. การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ หมายถึง การทดสอบความเข้าใจเรื่องเศษส่วน ด้วยข้อสอบที่มีความเหมาะสมกับความสามารถของผู้สอบ และดำเนินการสอบโดยใช้คอมพิวเตอร์ซึ่งข้อสอบที่ถูกนำเสนอแก่ผู้สอบแต่ละข้อนั้นขึ้นอยู่กับผลของการตอบข้อสอบข้อที่ผ่านมา

2. คลังข้อสอบ หมายถึง กลุ่มของข้อสอบที่มีคุณภาพ ซึ่งสร้างขึ้นตามหลักเกณฑ์ที่ถูกต้องและถูกเก็บไว้อย่างเป็นระบบ ในการเลือกข้อสอบขึ้นมาใช้นั้นต้องเลือกตามเกณฑ์ที่กำหนดขึ้น คลังข้อสอบที่ใช้ในการวิจัยครั้งนี้ เป็นกลุ่มข้อสอบที่วัดความเข้าใจ เรื่อง เศษส่วน มีข้อสอบทั้งหมด 200 ข้อ ลักษณะของข้อสอบเป็นแบบเลือกตอบ 4 ตัวเลือก แต่ละข้อได้รับการวิเคราะห์โดยใช้ทฤษฎีการตอบสนองข้อสอบ แต่ละข้อมีค่าอำนาจจำแนก ค่าความยากง่าย ค่าการเดาและค่าความสามารถที่ทำให้ข้อสอบมีสารสนเทศสูงสุดกำกับอยู่ โดยจะเก็บข้อสอบและค่าพารามิเตอร์ประจำตัวข้อสอบแยกกัน เมื่อต้องการจะเรียกข้อสอบขึ้นมาที่หน้าจอคอมพิวเตอร์ก็จะใช้ค่าพารามิเตอร์เป็นตัวชี้ คลังข้อสอบเรื่อง เศษส่วนที่สร้างขึ้นในครั้งนี้ มีค่าอำนาจจำแนกอยู่ในช่วง .41 ถึง 2.34 ค่าความยากง่ายอยู่ในช่วง -2.23 ถึง 2.72 และค่าการเดาอยู่ในช่วง .06 ถึง .30

3. ความยากง่ายของข้อสอบข้อแรก หมายถึง ค่าความยากง่ายของข้อสอบข้อแรกที่ถูกนำขึ้นมาจากคลังข้อสอบให้ผู้สอบได้ตอบบนหน้าจอคอมพิวเตอร์ ซึ่งกำหนดค่าความยากง่ายไว้ 3 ระดับ คือ ข้อที่ง่าย ($b = -1.30$) ข้อที่ยากปานกลาง ($b = 0.00$) และข้อที่ยาก ($b = 1.30$)

4. เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ หมายถึง วิธีการคัดเลือกข้อสอบแต่ละข้อจากคลังข้อสอบเพื่อให้ผู้สอบได้ตอบ 3 วิธี คือ วิธีการจับคู่ระหว่างค่าความยากง่ายของข้อสอบ

(b) กับค่าความสามารถโดยประมาณของผู้สอบ ($\hat{\theta}$) , วิธีการจับคู่ระหว่างค่าความสามารถที่ทำให้ข้อสอบมีสารสนเทศสูงสุด (θ_{max}) กับค่าความสามารถโดยประมาณของผู้สอบ ($\hat{\theta}$) และ วิธีใช้ข้อสอบที่มีค่าสารสนเทศสูงสุด (maximum item information , I_{max})

5. วิธีการประมาณค่าความสามารถของผู้สอบ หมายถึง วิธีการหาค่าความสามารถของผู้สอบโดยประมาณ โดยใช้วิธีการตามทฤษฎีการตอบสนองข้อสอบ 2 วิธี คือ วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไขและวิธีของเบย์แบบปรับใหม่

6. เกณฑ์ยุติการทดสอบ หมายถึง ค่าที่กำหนดไว้เพื่อเป็นตัวกำหนดการสิ้นสุดการทดสอบของผู้สอบแต่ละคน การวิจัยครั้งนี้ใช้ ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ 3 ระดับ คือ ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถเท่ากับหรือน้อยกว่า .30 , .40 และ .50 เป็นเกณฑ์ยุติการทดสอบ

7. ทฤษฎีการตอบสนองข้อสอบ หมายถึง ทฤษฎีที่แสดงความสัมพันธ์ระหว่างความสามารถของผู้สอบกับค่าประจำตัวของข้อสอบ คือค่าอำนาจจำแนก ค่าความยากง่าย และ ค่าการเดา

8. แบบทดสอบแบบประเพณีนิยมหรือแบบทดสอบแบบเดิม หมายถึง แบบทดสอบที่จัดพิมพ์เป็นชุดๆแต่ละชุดมีข้อสอบเหมือนกันหรือคู่ขนานกัน คุณภาพของข้อสอบและแบบทดสอบถูกวิเคราะห์โดยใช้ทฤษฎีการทดสอบแบบเดิม

9. ค่าพารามิเตอร์ของข้อสอบ หมายถึง ลักษณะประจำตัวของข้อสอบ ได้แก่ ค่าอำนาจจำแนก (a) ค่าความยากง่าย (b) และค่าการเดา (c) ที่วิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ

10. ความคลาดเคลื่อนในการประมาณค่าความสามารถ หมายถึง ผลต่างระหว่างค่าความสามารถจริง (θ) กับค่าความสามารถที่ได้จากการประมาณค่า ($\hat{\theta}$)

11. ความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ หมายถึง ค่าเบี่ยงเบนมาตรฐานของผลต่างระหว่างค่าความสามารถจริง(θ) กับค่าความสามารถของผู้สอบที่ได้จากการประมาณค่า ($\hat{\theta}$)

12. ความเที่ยงตรงเชิงสภาพ หมายถึง สหสัมพันธ์ระหว่างค่าความสามารถที่ได้จากการตอบแบบทดสอบวัดความสามารถทางคณิตศาสตร์กับค่าความสามารถที่ได้จากการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์เรื่องเศษส่วน

13. จำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะ หมายถึง จำนวนข้อสอบทั้งหมดที่ผู้สอบแต่ละคนได้รับจากการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์

ประโยชน์ที่คาดว่าจะได้รับ

ผลจากการวิจัยครั้งนี้คาดว่าจะมีประโยชน์ ดังนี้

1. ได้ข้อมูลสำหรับเป็นแนวทางในการเลือกใช้รูปแบบของการทดสอบแบบปรับเหมาะ กับความสามารถของผู้สอบด้วยคอมพิวเตอร์ที่มีประสิทธิภาพ นั่นคือ การทดสอบมีความเที่ยงตรง เชิงสภาพสูงและใช้ข้อสอบจำนวนน้อยๆ เนื่องจากผลของการวิจัยในครั้งนี้จะได้ข้อค้นพบว่า มี ตัวแปรใดบ้างที่เกี่ยวข้องกับความเที่ยงตรงเชิงสภาพและจำนวนข้อสอบที่ใช้ในการทดสอบแบบ ปรับเหมาะ กับความสามารถของผู้สอบด้วยคอมพิวเตอร์
2. ได้เผยแพร่ความรู้เกี่ยวกับการทดสอบแบบปรับเหมาะ กับความสามารถของผู้สอบ ด้วยคอมพิวเตอร์ให้แพร่หลายออกไปมากกว่าเดิม เนื่องจากขณะนี้การทดสอบลักษณะดังกล่าวยังไม่ค่อยแพร่หลายมากนักในประเทศไทย ทั้งนี้ อาจจะเนื่องมาจากมีข้อจำกัดในเรื่องของการดำเนินการสอบที่จะต้องใช้เครื่องคอมพิวเตอร์ ซึ่งมีราคาค่อนข้างสูง แต่ในปัจจุบันราคาของคอมพิวเตอร์ ได้ลดลงจากอดีตอย่างมากและคอมพิวเตอร์เริ่มเข้ามามีบทบาทต่อการจัดการเรียนการสอนใน สถานศึกษาระดับต่างๆมากขึ้น ดังนั้นผู้ที่มีหน้าที่เกี่ยวข้องกับการจัดการศึกษาโดยเฉพาะ ผู้ที่มี หน้าที่โดยตรงเกี่ยวกับการวัดผลการศึกษาควรที่จะมีความรู้ความเข้าใจเกี่ยวกับการทดสอบ ลักษณะนี้บ้างพอสมควร
3. ครูผู้สอนสามารถนำไปโปรแกรมคอมพิวเตอร์ที่ผู้วิจัยสร้างขึ้นไปใช้ในการสร้างคลัง ข้อสอบสำหรับวัดสมรรถภาพและเนื้อหาวิชาต่างๆตามที่ต้องการ ตลอดจนใช้โปรแกรมการทดสอบ แบบปรับเหมาะ กับความสามารถของผู้สอบด้วยคอมพิวเตอร์เพื่อปรับปรุงการเรียนการสอน โดย นำไปใช้เกี่ยวกับการเรียนเพื่อรอบรู้ การให้ระดับผลการเรียนและการตรวจสอบความก้าวหน้าของ นักเรียนแต่ละคน

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

ในบทนี้ได้นำเสนอเรื่องราวที่เกี่ยวข้องกับการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์หรือการทดสอบแบบซี เอ ที ซึ่งได้รวบรวมมาจากเอกสารและงานวิจัยที่เกี่ยวข้อง โดยนำเสนอเรียงตามลำดับ ดังนี้

1. การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ
 2. ทฤษฎีการตอบสนองข้อสอบ
 3. งานวิจัยที่เกี่ยวข้องกับการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ
- ซึ่งในแต่ละหัวข้อมีรายละเอียดดังต่อไปนี้

1. การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ

1.1 ลักษณะทั่วไป

นับตั้งแต่อดีตจนถึงปัจจุบัน วิธีการวัดผลที่ได้รับคามนิยมกันอย่างแพร่หลายก็คือ การทดสอบครั้งละหลายๆคนโดยใช้แบบทดสอบฉบับเดียวกันหรือแบบทดสอบที่มีความเป็นคู่ขนานกัน ซึ่งแบบทดสอบดังกล่าวประกอบด้วยข้อสอบที่มีค่าความยากง่ายครอบคลุมช่วงความสามารถของกลุ่มผู้สอบ จากการที่ใช้ข้อสอบที่มีค่าความยากง่ายครอบคลุมช่วงความสามารถของผู้สอบทั้งหมด จึงทำให้มีข้อสอบที่ง่ายเกินไปสำหรับผู้สอบที่มีความสามารถสูง และยากเกินไปสำหรับผู้ที่มีความสามารถต่ำ ซึ่งนอกจากจะทำให้ผู้สอบเสียเวลาในการตอบข้อสอบที่ไม่เหมาะสมกับความสามารถของตนแล้ว ก็ยังอาจทำให้ผู้สอบเกิดความเบื่อหน่าย เมื่อยล้าและเกิดการเดาในการตอบข้อสอบที่ยากสำหรับผู้ที่มีความสามารถต่ำและเกิดความสะเพร่าในการตอบข้อสอบที่ง่ายสำหรับผู้สอบที่มีความสามารถสูง ซึ่งก่อให้เกิดความคลาดเคลื่อนในการวัด ข้อสอบลักษณะนี้อาจมีถึงครึ่งหนึ่งของข้อสอบทั้งหมดในแบบทดสอบและเนื่องจากผู้สอบแต่ละคนมีความสามารถต่างกันในเรื่องที่ถูกวัด ดังนั้น จึงไม่ควรที่จะใช้แบบทดสอบฉบับเดียวกัน หรือ คู่ขนานกันกับผู้สอบทุกคน การทดสอบที่นับว่ามีประสิทธิภาพสูงสุด คือการทดสอบที่ไม่นำข้อสอบที่ง่ายเกินไปหรือยากเกินไปสำหรับผู้สอบมาให้ผู้สอบตอบ ผู้สอบที่มีความสามารถสูงควรได้รับข้อสอบที่ยาก ผู้สอบที่มีความสามารถต่ำควรได้รับข้อสอบที่ง่าย (Lord and Novick, 1968 : 359) นอกจากนั้นยังต้องมีการประมาณค่าความสามารถของผู้สอบลงบนคุณลักษณะที่ถูกวัด (trait) โดยใช้ข้อมูลจากการ

ตอบข้อสอบ การทดสอบลักษณะดังกล่าวมีชื่อและวิธีการแตกต่างกันไป เช่น adaptive testing , branched testing , individualized testing , response contingent testing , tailor testing , computerized testing , programmed testing , computerized automate testing , sequential item testing (Weiss. 1988 : 372 ; Lord. 1980 : 150) คำแรกที่ใช้เรียกการทดสอบลักษณะนี้ คือ การทดสอบเทเลอร์ (tailor testing) เป็นคำที่วิลเลียม ดับบลิว เทอร์นบูล (William W. Turnbull) เป็นผู้คิดขึ้นเมื่อปี 1951 (Lord. 1980 : 150-151)

ในระยะแรกๆ นั้นการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบได้ดำเนินการสอบโดยให้ผู้สอบเขียนตอบลงในกระดาษคำตอบ ซึ่งมีข้อจำกัดในเรื่องของการคัดเลือกข้อสอบข้อต่อไปหลังจากที่ผู้สอบตอบข้อสอบข้อแรกไปแล้วประกอบกับวิทยาการทางคอมพิวเตอร์ได้พัฒนาก้าวหน้าไปมาก ดังนั้นการดำเนินการทดสอบด้วยคอมพิวเตอร์จึงได้รับความนิยมมากขึ้นตามลำดับ เนื่องจากรักษาความลับของข้อสอบได้เป็นอย่างดี การบันทึกคำตอบและการแสดงผลการทดสอบสามารถทำได้ง่ายและรวดเร็วกว่าการใช้กระดาษคำตอบ และการตรวจข้อสอบแบบเดิม การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบต้องการคลังข้อสอบขนาดใหญ่ ข้อสอบในคลังข้อสอบดังกล่าวจะมีค่าพารามิเตอร์ประจำตัวของข้อสอบแต่ละข้อ ได้แก่ ค่าอำนาจจำแนก (a) ค่าความยากง่าย (b) ค่าการเดา (c) และค่าสารสนเทศของข้อสอบ (item information function) แสดงคุณภาพไว้เพื่อสะดวกต่อการเลือกใช้ข้อสอบ ดังกล่าว (Hambleton and Swaminathan. 1985 : 296)

ลักษณะสำคัญของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ ก็คือการนำเสนอข้อสอบที่มีความยากง่ายเหมาะสมกับความสามารถของผู้สอบแต่ละคน ทฤษฎีการตอบสนองข้อสอบถูกนำมาใช้ประโยชน์ในการทดสอบลักษณะนี้ ข้อสอบที่ถูกเลือกขึ้นมาจากคลังข้อสอบนั้นจะให้สารสนเทศเกี่ยวกับผู้สอบได้เป็นอย่างดี เพื่อให้ผู้อ่านได้เห็นขอบเขตและพัฒนาการของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบอย่างชัดเจน ผู้วิจัยจึงได้นำเสนอประเภทของการทดสอบดังกล่าว ดังนี้

1.2 ประเภทของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ

ไวส์ (Weiss. 1974 : 78-110) แฮมเบิลตันและสวามินาธาน (Hambleton and Swaminathan. 1985 : 297) ได้จำแนกการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบออกเป็น 2 ประเภทใหญ่ๆ คือ

1.2.1 การทดสอบสองขั้นตอน (two-stage strategies)

1.2.2 การทดสอบหลายขั้นตอน (multi-stage strategies)

(1) แบบแยกทางคงที่ (fixed branching model)

- (1.1) รูปพีระมิตขนาดขั้นคงที่ (constant step size pyramid)
- (1.2) รูปพีระมิตขนาดขั้นแปรผัน (variable step size pyramid)
- (1.3) รูปพีระมิตข้างตัด (truncated pyramid)
- (1.4) รูปพีระมิตที่มีหลายข้อในแต่ละขั้น (multiple item pyramid)
- (1.5) รูปพีระมิตที่ให้น้ำหนักแก่ตัวเลือก (differential response option branching)
- (1.6) แบบทดสอบเฟลิกซ์เลเวล (flexilevel test)

(2) แบบแยกทางแปรผัน (variable branching model)

ฮูลินและคนอื่นๆ (Hulin and others. 1983 : 211-219) ได้จำแนกประเภทของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบออกเป็น การทดสอบที่ไม่ใช้คอมพิวเตอร์และการทดสอบที่ใช้คอมพิวเตอร์

(1) การทดสอบที่ไม่ใช้คอมพิวเตอร์

- (1.1) แบบวัดความสามารถของบินเน็ต (Binet's ability test)
- (1.2) แบบทดสอบสองขั้นตอน (two-stage tests)
- (1.3) แบบทดสอบเฟลิกซ์เลเวล (flexilevel tests)
- (1.4) แบบทดสอบแบบแยกทาง (branching tests)

(2) การทดสอบที่ใช้คอมพิวเตอร์ หรือ การทดสอบแบบ ซี เอ ที

(computerized adaptive testing , CAT)

จะเห็นว่ามีควมพยายามที่จะจำแนกประเภทของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบออกเป็นหลายประเภทตามเกณฑ์ที่ได้กำหนดขึ้น อย่างไรก็ตาม เมื่อใช้เกณฑ์ในการจำแนกประเภทต่างกัน ก็ไม่ได้หมายความว่า จะทำให้การทดสอบดังกล่าวมีการเปลี่ยนแปลงไปจากเดิม เพียงแต่อาจมีการย้ายจากกลุ่มเดิมไปอยู่กลุ่มใหม่ เพื่อให้สอดคล้องกับเกณฑ์ที่กำหนดขึ้น ซึ่งรายละเอียดของการทดสอบแบบต่างๆพอสรุปได้ ดังนี้

การทดสอบความสามารถของบินเน็ต

การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบนั้นถูกประยุกต์ใช้เป็นครั้งแรกที่ประเทศฝรั่งเศส โดยอัลเฟรด บินเน็ต (Alfred Binet) ในตอนต้นของทศวรรษ 1900 โดยสร้างแบบวัดเชาว์ปัญญาของบินเน็ต (Binet Intelligence Test) (Weiss. 1985 : 776) ต่อมาได้ดัดแปลงเป็นแบบวัด

เชาว์ปัญญาสแตนฟอร์ด-บินเน็ต (Stanford Binet Intelligence Test) เพื่อนำไปใช้ในประเทศที่พูดภาษาอังกฤษ คลังข้อสอบสำหรับแบบวัดเชาว์ปัญญาของบินเน็ตจะใช้ระดับอายุเป็นโครงสร้างในการดำเนินการสอบนั้น เริ่มต้นโดยใช้ข้อสอบที่มีความเหมาะสมกับระดับอายุจริงของผู้สอบก่อน เมื่อตอบเสร็จก็จะตรวจให้คะแนนทันที ถ้าผู้สอบตอบได้ถูกต้องข้อสอบข้อต่อไปจะเป็นข้อสอบที่เหมาะสมกับระดับอายุที่สูงขึ้น แต่ถ้าตอบผิดข้อสอบข้อต่อไปจะเหมาะกับผู้ที่ระดับอายุต่ำกว่า การทดสอบจะมีลักษณะเช่นนี้ไปเรื่อยๆ จนกระทั่งถึงระดับอายุที่ผู้สอบถูกระบุว่ามีความสามารถอยู่ถึงระดับนั้น ซึ่งเรียกว่าระดับเพดาน (ceiling level) ซึ่งผู้สอบจะตอบคำถามในระดับนี้ผิดหมดทุกข้อ หรือจนกระทั่งถึงข้อสอบในระดับพื้นฐาน (basal level) ซึ่งผู้สอบจะตอบข้อสอบได้ถูกต้องทุกข้อในระดับนี้ ทั้งระดับเพดานและระดับพื้นฐานดังกล่าวจะถูกใช้เป็นเกณฑ์สำหรับยุติการทดสอบ ผู้สอบแต่ละคนจะยุติการสอบที่ระดับต่างๆ กัน มีผลทำให้จำนวนข้อสอบที่ใช้ในการทดสอบนั้นต่างกันไปด้วย

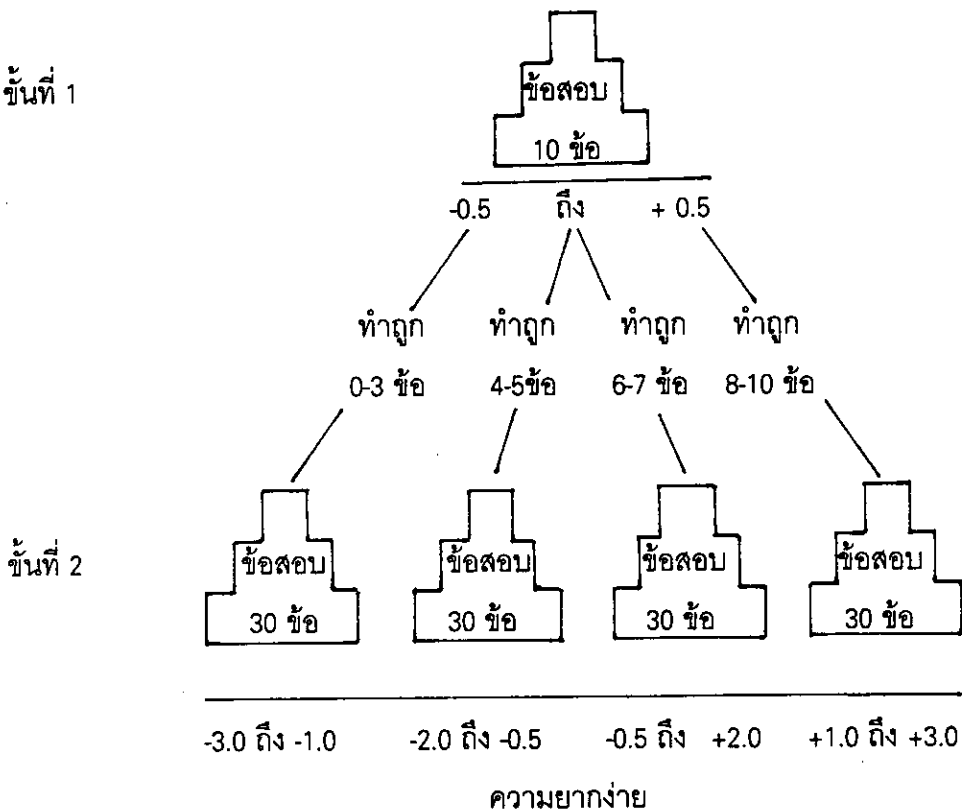
การทดสอบสองขั้นตอน

ในปี 1980 ลอร์ด (Lord) ได้นำเสนอการทดสอบสองขั้นตอน แบบทดสอบลักษณะนี้เป็นแบบทดสอบปรนัยชนิดเลือกตอบ การดำเนินการสอบเริ่มโดยให้ผู้สอบตอบข้อสอบในแบบทดสอบเพื่อแยกทาง (routing test) ผลของการตอบแบบทดสอบเพื่อแยกทางจะเป็นตัวชี้ว่า ผู้สอบจะต้องไปตอบแบบทดสอบชุดใดในการสอบขั้นที่ 2 แบบทดสอบสำหรับการสอบในขั้นที่ 2 เรียกว่า แบบทดสอบเพื่อวัดผล (measurement test) แนวคิดพื้นฐานของการทดสอบลักษณะนี้ ก็คือ ผู้ที่มีความสามารถต่ำก็ควรได้คะแนนน้อยๆ ในการทดสอบเพื่อแยกทางและควรได้รับข้อสอบที่ง่ายๆ ในการทดสอบเพื่อวัดผล ซึ่งตรงกันข้ามผู้ที่มีความสามารถสูงๆ ซึ่งก็ควรจะได้คะแนนสูงๆ จากการทดสอบเพื่อแยกทางและได้ตอบข้อสอบข้อที่ยากๆ ในการทดสอบเพื่อวัดผล ดังภาพประกอบ 2

การทดสอบหลายขั้นตอน

การทดสอบลักษณะนี้ ได้จัดเรียงข้อสอบในแบบทดสอบโดยมีโครงสร้างเป็นรูปพีระมิด และใช้ค่าความยากง่ายของข้อสอบเป็นเกณฑ์ในการจัดเรียง ข้อสอบที่อยู่ตรงจุดยอดของพีระมิดเป็นข้อสอบที่มีความยากง่ายปานกลาง ถ้าผู้สอบตอบข้อสอบข้อแรกถูกต้องข้อสอบข้อต่อไปก็จะยากขึ้น แต่ถ้าตอบข้อสอบข้อแรกผิดข้อสอบข้อต่อไปก็จะง่ายลง การทดสอบจะดำเนินเช่นนี้ไปเรื่อยๆ และจะยุติการสอบตามเกณฑ์ที่กำหนดไว้ การทดสอบลักษณะนี้จำแนกเป็นประเภทย่อยๆ ได้ดังนี้ (ต่าย เชียงฉี. 2530 : 38-48)

ขั้นที่ 1

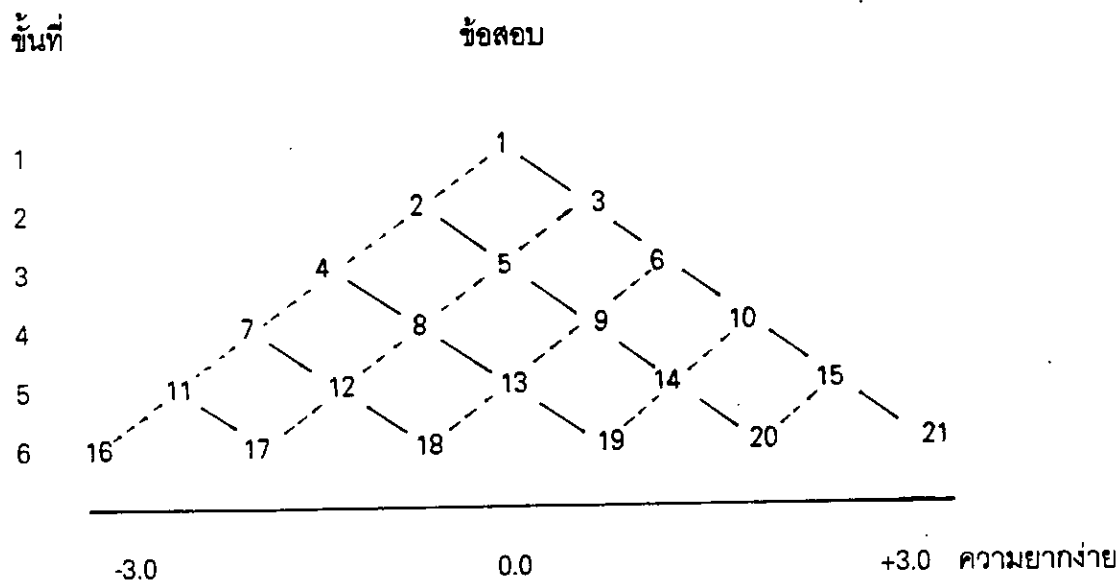


ขั้นที่ 2

ภาพประกอบ 2 โครงสร้างของการทดสอบสองขั้นตอน

(1) แบบแยกทางคงที่

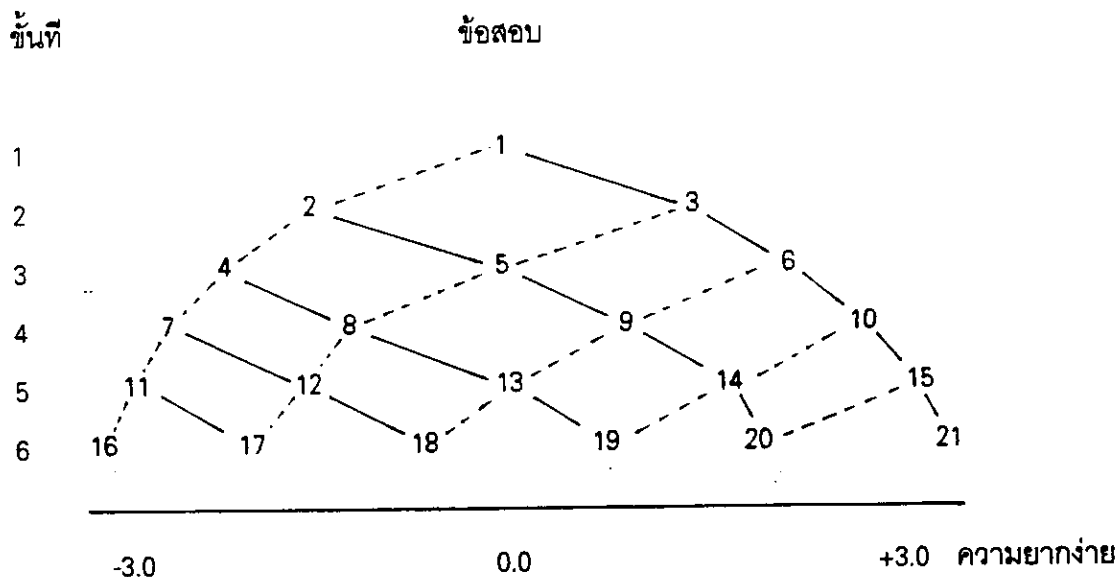
(1.1) รูปพีระมิดขนาดขั้นคงที่เป็นการจัดชุดแบบทดสอบเป็นรูปสามเหลี่ยม โดยมีจำนวนข้อสอบในแต่ละขั้นเท่ากับลำดับที่ของขั้น ผู้สอบจะต้องตอบข้อสอบขั้นละ 1 ข้อ โดยเริ่มตอบข้อ 1 ก่อน ถ้าตอบถูกก็จะไปตอบข้อที่ 3 ซึ่งยากกว่าข้อ 1 แต่ถ้าตอบผิดก็จะไปตอบข้อ 2 ซึ่งง่ายกว่าข้อ 1 ดังภาพประกอบ 3



ภาพประกอบ 3 โครงสร้างของแบบทดสอบรูปพีระมิตขนาดชั้นคงที่

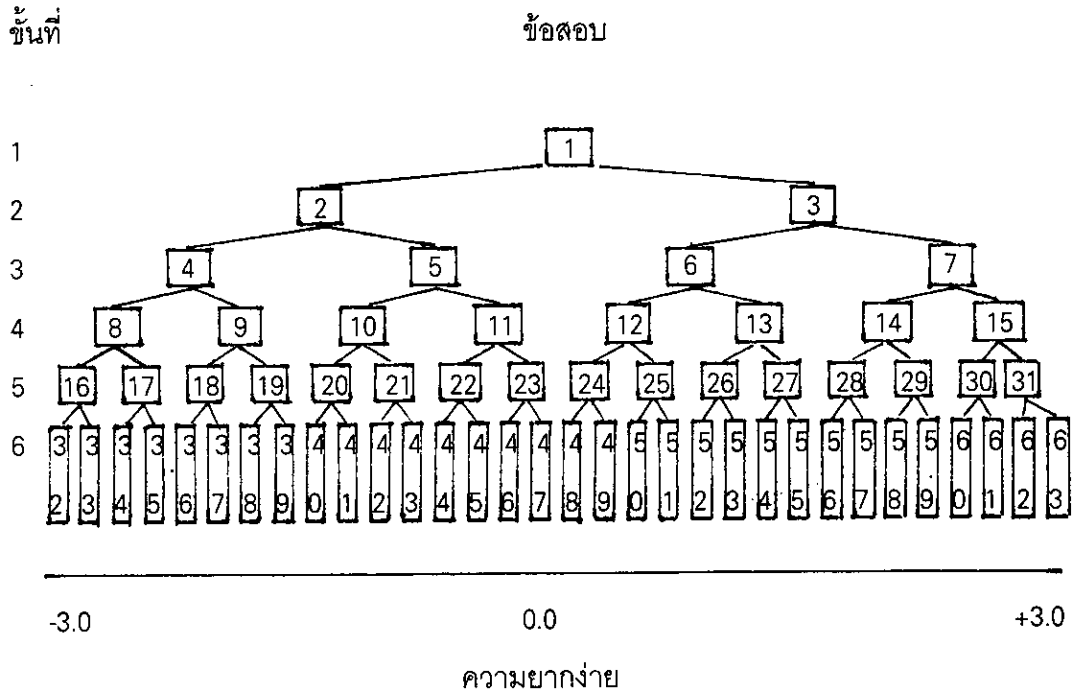
(1.2) รูปพีระมิตขนาดชั้นแปรผัน

การจัดชุดแบบทดสอบลักษณะนี้ ได้กำหนดช่วงห่างของค่าความยากง่ายของข้อสอบในชั้นเดียวกันไม่เท่ากัน โดยข้อสอบที่มีความยากง่ายปานกลางจะมีช่วงห่างของค่าความยากง่ายมากกว่าข้อสอบที่ยากหรือข้อสอบที่ง่าย ดังภาพประกอบ 4



ภาพประกอบ 4 โครงสร้างของแบบทดสอบรูปพีระมิตขนาดชั้นแปรผัน

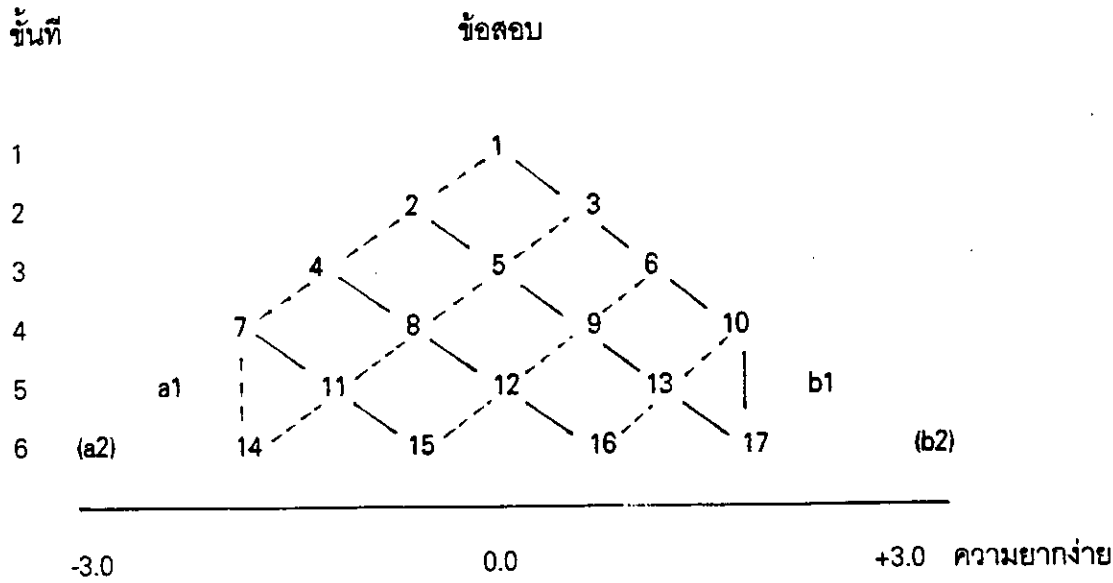
ต่อมาได้มีการพัฒนาโครงสร้างของแบบทดสอบรูปพีระมิดขนาดชั้นแปดใหม่โดยมีการเพิ่มจำนวนข้อสอบในแต่ละชั้นให้มีจำนวนเป็น 2 เท่าของจำนวนข้อสอบในชั้นที่ผ่านมา รูปแบบใหม่นี้มีชื่อว่า โรบินส์มอนโร (Robbins Monro) ซึ่งมีโครงสร้างดังภาพประกอบ 5



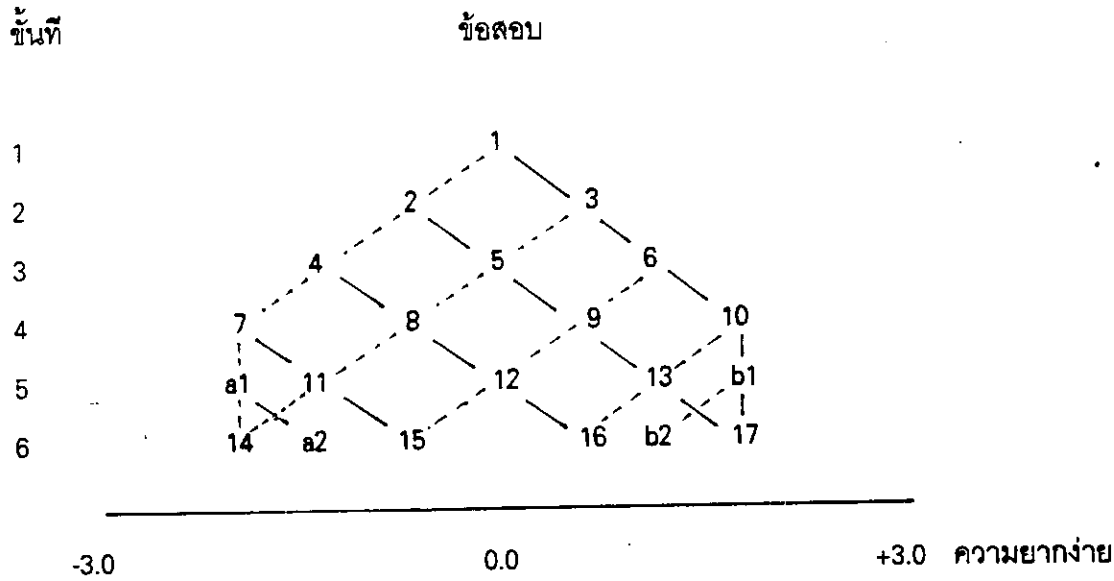
ภาพประกอบ 5 โครงสร้างของแบบทดสอบรูปพีระมิดแบบโรบินส์มอนโร

(1.3) รูปพีระมิดแบบข้างตัด

เป็นรูปแบบที่มุสสิโอ (Weiss, 1974, citing Mussio, 1973.) ที่สร้างขึ้นเพื่อลดจำนวนข้อสอบจากโครงสร้างแบบทดสอบรูปพีระมิดแบบขนาดชั้นคงที่ ซึ่งจำแนกเป็นแบบสกัดการสะท้อนกลับ (reflecting barrier) และ รักษาการสะท้อนกลับ (retaining barrier) ดังภาพประกอบ 6 และ 7 ตามลำดับ



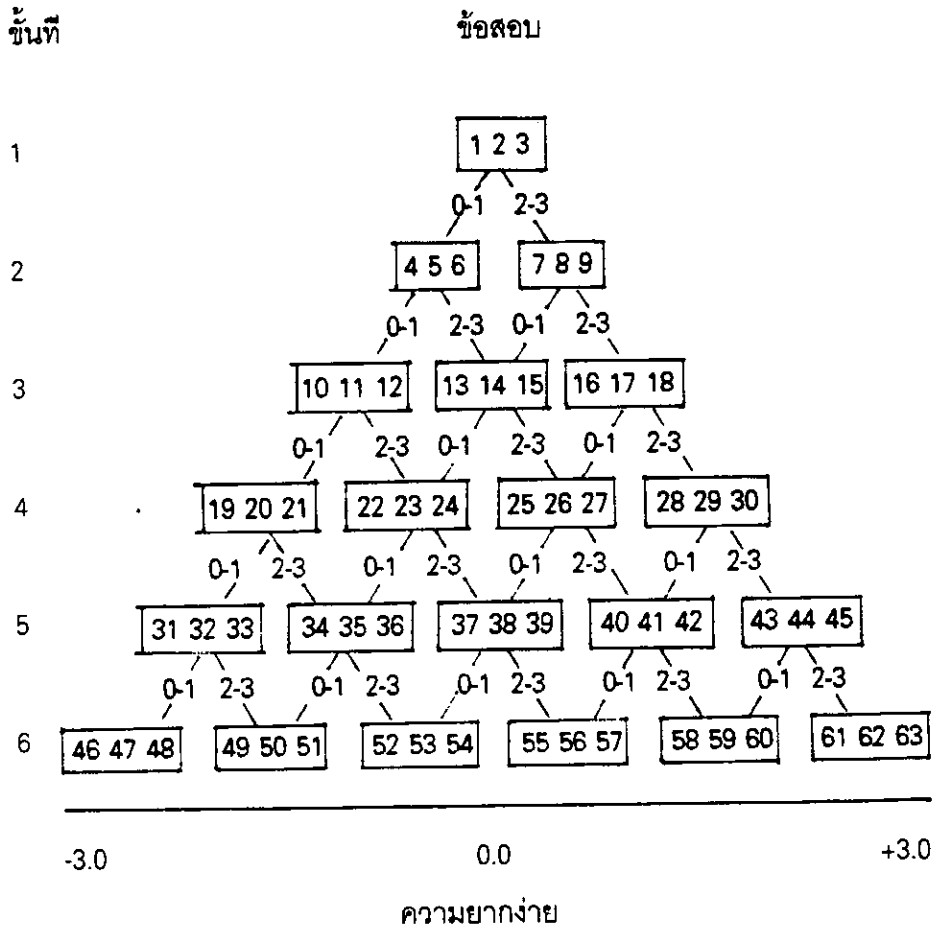
ภาพประกอบ 6 โครงสร้างของแบบทดสอบรูปพีระมิดแบบสกัดการสะท้อนกลับ



ภาพประกอบ 7 โครงสร้างของแบบทดสอบรูปพีระมิดแบบรักษาการสะท้อนกลับ

(1.4) รูปพีระมิดแบบมีหลายข้อในแต่ละชั้น

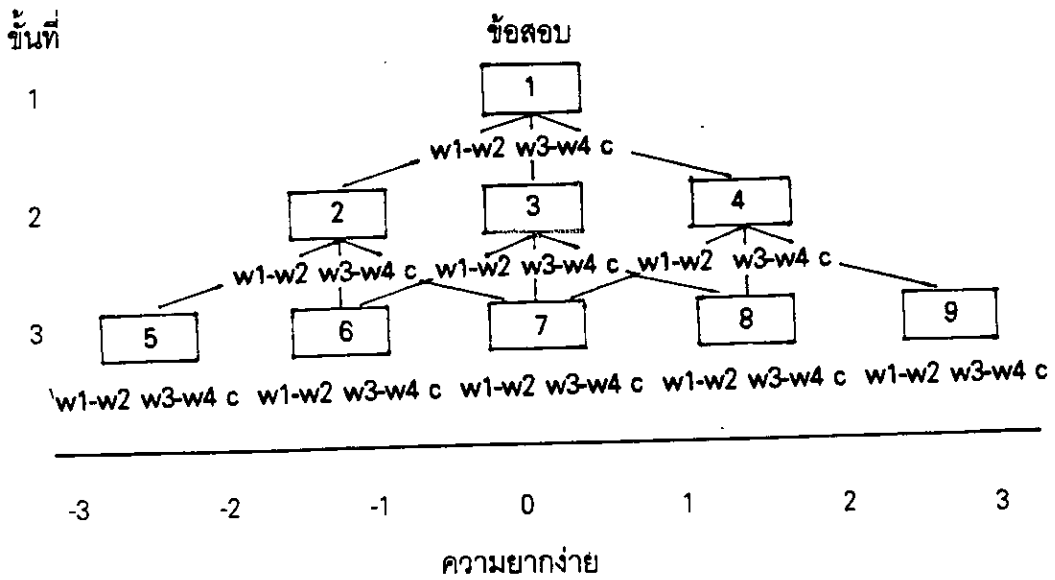
รูปแบบนี้กำหนดให้มีข้อสอบหลายข้อในแต่ละชั้น ทั้งนี้เพื่อให้การแยกทางมีความถูกต้องมากยิ่งขึ้น ซึ่งมีโครงสร้างของการจัดข้อสอบดังภาพประกอบ 8



ภาพประกอบ 8 โครงสร้างของแบบทดสอบรูปพีระมิดแบบมีหลายข้อในแต่ละชั้น

(1.5) รูปพีระมิดแบบให้น้ำหนักแก่ตัวเลือก

รูปแบบนี้คล้ายกับรูปแบบที่มีหลายข้อในแต่ละชั้น ต่างกันตรงที่จะมีการกำหนดน้ำหนักให้กับตัวเลือกแต่ละข้อ ดังภาพประกอบ 9



ภาพประกอบ 9 โครงสร้างของแบบทดสอบรูปพีระมิดแบบให้น้ำหนักแก่ตัวเลือกของข้อสอบ
เพื่อแยกทาง

(1.6) การทดสอบเฟลิกซีเลเวล

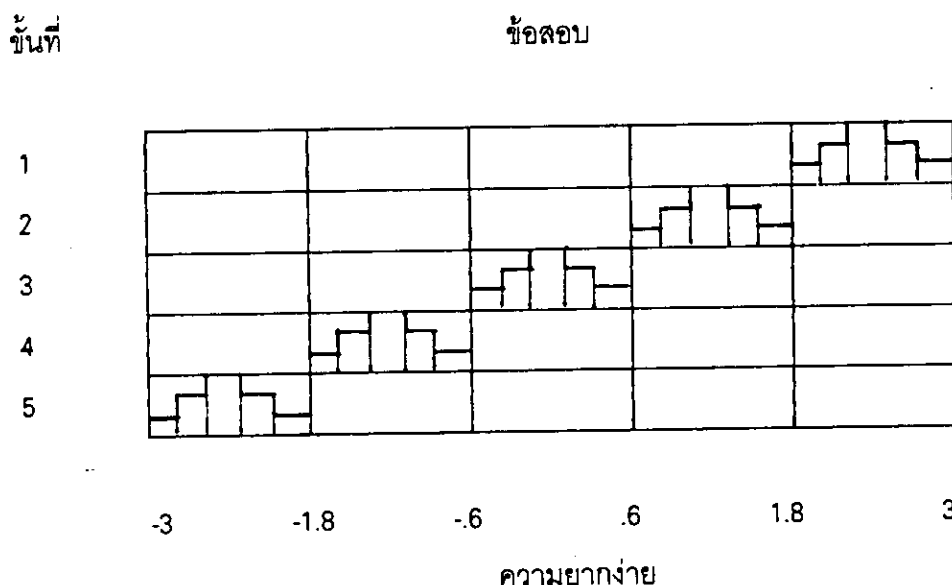
ลอร์ด (Lord 1980. : 114-117) เป็นผู้นำเสนอการทดสอบลักษณะนี้ขึ้นมา
การจัดเรียงข้อสอบจะเรียงจากง่ายไปหายาก ข้อที่มีความยากง่ายปานกลางจะถูกนำเสนอแก่
ผู้สอบก่อน เช่น ถ้าข้อสอบมีทั้งหมด 75 ข้อ โครงสร้างของการจัดข้อสอบลักษณะนี้สามารถ
แสดงได้ดังภาพประกอบ 10

ข้อที่ 38 ยากปานกลาง	
1. ข้อ 37 (ง่ายกว่าข้อ 38)	1. ข้อ 39 (ยากกว่าข้อ 38)
2.	2.
3.	3.
⋮	⋮
37. ข้อ 1 (ง่ายที่สุด) (พิมพ์ด้วยสีแดง)	37. ข้อ 75 (ยากที่สุด) (พิมพ์ด้วยสีน้ำเงิน)

ภาพประกอบ 10 โครงสร้างของการจัดแบบทดสอบเฟลิกซีเลเวล

กระดาคำตอบของการทดสอบลักษณะนี้จะทำให้ผู้สอบทราบได้ทันทีหลังจากการตอบแต่ละข้อว่า ถ้าตอบผิดจุดแดงจะขึ้นมา (ข้อสอบง่ายกว่าเดิม) ถ้าตอบถูกต้องจุดน้ำเงินจะขึ้นมา (ข้อสอบยากกว่าเดิม) ซึ่งอาจใช้วิธีการเจาะหรือแกะกระดาษที่ทับตัวเลือกแต่ละข้อ ผู้สอบจะต้องตอบข้อสอบครึ่งหนึ่งของข้อสอบทั้งหมด การทดสอบจึงจะยุติ วิธีการง่าย ๆ เพื่อให้ผู้สอบสามารถยุติการสอบได้ด้วยตนเอง ก็คือ จัดเรียงข้อสอบด้านยากใหม่โดยกลับหัวจากง่ายลงไปหายากที่สุด เป็นจากยากที่สุดลงไปหาง่าย (เหมือนการจัดข้อสอบด้านง่าย) การตอบก็จะมีลักษณะเหมือนเดิม คือตอบข้อสอบที่ยากง่ายปานกลางก่อน ถ้าตอบผิดก็จะไปตอบข้อที่ยากกว่าเดิม ถ้าตอบถูกก็จะไปตอบข้อที่ยากกว่าเดิม ตอบไปจนกระทั่งพบว่าในแต่ละแถวของข้อสอบ (มี 2 ข้อ) ถูกตอบครบทุกแถว แถวละ 1 ข้อ จึงยุติการสอบ

ต่อมาได้มีการปรับปรุงโครงสร้างของการทดสอบลักษณะนี้ โดยแบ่งความยากง่ายของข้อสอบออกเป็นช่วงๆ ในแต่ละช่วงความยากง่ายจะมีข้อสอบหลายข้อ การจัดแบบทดสอบลักษณะนี้ เรียกว่า แบบทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบแบบแบ่งชั้น (stradaptive test) ซึ่งมีโครงสร้างของการจัดข้อสอบ ดังภาพประกอบ 11



ภาพประกอบ 11 โครงสร้างของแบบทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบแบบแบ่งชั้น

(2) การทดสอบแบบหลายขั้นตอนแบบแยกทางแปรผัน

การทดสอบแบบแยกทางแปรผันนี้ต่างจากการทดสอบแบบแยกทางคงที่ตรงที่ไม่สามารถกำหนดเส้นทางของการตอบข้อสอบได้แน่นอนตายตัว ข้อสอบที่จะตอบข้อต่อไปนั้น ขึ้นอยู่กับผลการตอบข้อสอบข้อที่ผ่านมา ซึ่งต้องมีการประมาณค่าความสามารถของผู้สอบ ทุกครั้งหลังจากตอบข้อสอบแต่ละข้อ จึงเป็นเรื่องยุ่งยากที่จะจัดแบบทดสอบเป็นแบบเขียนตอบ (paper and pencil test) ดังนั้น จึงมีการนำไมโครคอมพิวเตอร์เข้ามาใช้เพื่อความสะดวกในการ ดำเนินการสอบและบันทึกผลการสอบ การทดสอบจะเริ่มต้นโดยให้ผู้สอบนั่งที่หน้าจอคอมพิวเตอร์ ให้ข้อมูลของตนเองตามที่ผู้ดำเนินการสอบต้องการ เช่น เลขที่ ชื่อ ชั้นเรียน ห้องเรียน เป็นต้น มีการแนะนำวิธีการตอบข้อสอบ หลังจากให้ผู้สอบเข้าใจวิธีการตอบดีแล้ว ข้อสอบข้อแรกก็จะถูก นำเสนอแก่ผู้สอบ ซึ่งตามปกติจะใช้ข้อสอบที่มีความยากง่ายปานกลางหรือใช้ข้อสอบที่ยากง่าย เหมาะกับผู้สอบ หากทราบสภาพภูมิหลังของผู้สอบมาก่อนข้อสอบถูกนำเสนอที่หน้าจอ คอมพิวเตอร์ครั้งละ 1 ข้อ หลังจากการตอบข้อสอบแต่ละข้อผ่านไปข้อสอบข้อใหม่ก็จะปรากฏ ขึ้นมาที่หน้าจอตามเกณฑ์การคัดเลือกข้อสอบที่กำหนดไว้ การทดสอบจะดำเนินไปเรื่อยๆ จน กระทั่งถึงเกณฑ์ยุติการสอบ การทดสอบก็จะสิ้นสุดลง

1.3 องค์ประกอบของการทดสอบ

การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์หรือ การทดสอบแบบ ซี เอ ที เป็นเทคโนโลยีของการทดสอบที่เข้ามามีส่วนช่วยแก้ไขจุดอ่อนของการ ทดสอบแบบเดิม องค์ประกอบของการทดสอบลักษณะนี้ ได้แก่ (Weiss and Kingsbury, 1984 : 362-364)

1.3.1 โมเดลการตอบสนองข้อสอบ (item response model) ในการประมาณค่า พารามิเตอร์ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์นั้นจะใช้ วิธีการประมาณค่าตามทฤษฎีการตอบสนองข้อสอบ ซึ่งอาจเป็นโมเดลโลจิสติก (logistic model) หรือ นอร์มอลโอจีฟ (normal ogive model) ชนิด 1 2 หรือ 3 พารามิเตอร์ก็ได้ ซึ่งการจะเลือกโมเดล ไตมาใช้ในการวิเคราะห์ข้อมูลนั้นก็จำเป็นต้องคำนึงถึงธรรมชาติของข้อมูลว่ามีลักษณะเป็นอย่างไร มีความสอดคล้องกับโมเดลใด โมเดลการตอบสนองข้อสอบนับว่ามีความสำคัญต่อการทดสอบใน ลักษณะนี้มาก เพราะการประมาณค่าความสามารถของผู้สอบนั้นจะเป็นอิสระจากข้อสอบที่สอบ ในทำนองเดียวกันการประมาณค่าพารามิเตอร์ของข้อสอบก็จะเป็นอิสระจากผู้สอบ ดังนั้น จึง สามารถเปรียบเทียบความสามารถของผู้สอบได้ถึงแม้ว่าผู้สอบแต่ละคนจะได้รับข้อสอบคนละ ชุดที่มีค่าความยากง่ายต่างกันก็ตาม (Hambleton and Swaminathan, 1985 : 296)

1.3.2 คลังข้อสอบ (item pool) การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์จะมีประสิทธิภาพสูงขึ้นถ้ามีคลังข้อสอบขนาดใหญ่ข้อสอบดังกล่าวได้รับการประมาณค่าพารามิเตอร์โดยใช้ ทฤษฎีการตอบสนองข้อสอบ ข้อสอบมีค่าอำนาจจำแนกสูง มีค่าความยากง่ายครอบคลุมช่วงความสามารถของผู้สอบและมีค่าการเดาต่ำ

1.3.3 ความยากง่ายของข้อสอบข้อแรก (entry level) หลักของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบทุกประเภท ก็คือ จัดข้อสอบให้เหมาะสมกับความสามารถของผู้สอบ ดังนั้นหากมีการเลือกข้อสอบข้อแรกให้กับผู้สอบอย่างเหมาะสมแล้วก็จะช่วยประหยัดเวลาในการทดสอบได้มาก เพราะจำนวนข้อสอบลดลง

1.3.4 เกณฑ์การคัดเลือกข้อสอบ (item selection rule) มีวิธีการเลือกข้อสอบจากคลังข้อสอบที่มีประสิทธิภาพ 2 วิธีการ คือ วิธีการเลือกข้อสอบที่มีค่าสารสนเทศสูงสุด (maximum information item selection) และวิธีการเลือกข้อสอบแบบเบย์ส์ (bayesian item selection) ซึ่งวิธีแรกจะให้ค่าสารสนเทศสูงสุด ส่วนวิธีหลังจะให้ค่าความแปรปรวนที่คาดหวังของการประมาณค่าความสามารถต่ำสุด

ฮูลินและคนอื่นๆ (Hulin and others. 1983 : 220-221) ได้เสนอแนะวิธีการเลือกข้อสอบจากคลังข้อสอบขึ้นมาให้ผู้สอบ ซึ่งมีความชัดเจนมากกว่าที่ไวส์ (Weiss) ได้เสนอแนะไว้ข้างต้น ซึ่งมีวิธีการดังนี้

(1) วิธีการจับคู่ระหว่างค่าความยากง่ายของข้อสอบกับค่าความสามารถของผู้สอบ วิธีการนี้เป็นวิธีการที่ง่ายที่สุดในการเลือกข้อสอบ หลักการคือ ข้อสอบในคลังข้อสอบที่มีค่าความยากง่ายเท่ากับ หรือ มีค่าใกล้เคียงกับค่าความสามารถมากที่สุดจะถูกเลือกขึ้นมาให้ผู้สอบได้ตอบ

(2) วิธีการจับคู่ระหว่างค่าความสามารถที่ทำให้ข้อสอบมีสารสนเทศสูงสุด(θ_{max}) กับค่าความสามารถของผู้สอบโดยประมาณ เมื่อมีการเดาเกิดขึ้นวิธีการเลือกข้อสอบโดยใช้ค่าของความยากง่ายเพียงอย่างเดียวไม่นับว่าไม่ค่อยเหมาะสม ทั้งนี้เพราะค่าฟังก์ชันสารสนเทศของข้อสอบ (item information function) ข้อที่ i จะสูงที่สุดที่ค่าของ m เมื่อ

$$m_i = b_i + \frac{1}{D_{ai}} \ln \left(\frac{1 + \sqrt{1 + 8c_i}}{2} \right)$$

ซึ่งค่าของ m ของข้อสอบข้อที่ i ก็คือจุดบนสเกลความสามารถที่ข้อสอบข้อที่ i มีค่าสารสนเทศสูงสุด โดยทั่วไปจะเรียกจุด m ว่า θ_{max}

(3) วิธีสารสนเทศสูงสุด (maximum information , I_{max}) วิธีการนี้จะมีการประมาณค่าสารสนเทศของข้อสอบ (item information function) เก็บไว้เป็นดัชนีในคลังข้อสอบทุกครั้งที่ค่าความสามารถของผู้สอบที่ได้จากการประมาณค่าเปลี่ยนแปลงไปข้อสอบที่ยังไม่ได้ถูกเลือกขึ้นมาใช้ที่มีค่าสารสนเทศของข้อสอบสูงสุดที่ระดับความสามารถที่ประมาณค่าได้ครั้งหลังสุดจะถูกเลือกขึ้นมาให้ผู้สอบได้ตอบ

1.3.5 วิธีการประมาณค่าความสามารถของผู้สอบ (scoring method) วิธีการประมาณค่าความสามารถของผู้สอบนั้น มีอยู่ด้วยกันหลายวิธีแต่วิธีที่ค่อนข้างได้รับความนิยม ก็คือวิธีของเบส์ (Bayesian) และวิธีความเป็นไปได้สูงสุด (maximum likelihood)

1.3.6 เกณฑ์ยุติการทดสอบ (termination criterion) การทดสอบแบบปรับเหมาะ กับความสามารถของผู้สอบด้วยคอมพิวเตอร์ นั้นการทดสอบจะดำเนินไปเรื่อยๆ ตั้งแต่เริ่มต้นจนกระทั่งถึงเกณฑ์ที่ระบุไว้ให้ยุติการทดสอบ การดำเนินการสอบก็จะสิ้นสุดลง เกณฑ์ที่ใช้กันอยู่ทั่วไปมีอยู่ 2 เกณฑ์ คือ การกำหนดจำนวนข้อสอบคงที่ เช่น กำหนดให้ผู้สอบทุกคนต้องทำข้อสอบคนละ 25 ข้อ เกณฑ์นี้เหมาะสำหรับการประมาณค่าโดยใช้วิธีการจำลองข้อมูลแบบมอนติคาร์โล เพราะสามารถเปรียบเทียบได้โดยตรง ระหว่างข้อสอบจากการทดสอบแบบปรับเหมาะ กับความสามารถของผู้สอบด้วยคอมพิวเตอร์กับการทดสอบแบบเดิม แต่ในทางปฏิบัติพบว่าไม่ค่อยเหมาะ เพราะผู้สอบบางคนอาจต้องการตอบข้อสอบน้อยกว่าที่กำหนดไว้ ซึ่งข้อสอบที่เกินความต้องการ อาจทำให้ผู้สอบเหนื่อย เสียเวลา ขาดความระมัดระวังหรือไม่เต็มใจที่จะสอบ สำหรับอีกเกณฑ์หนึ่งก็คือค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ ซึ่งหลังจากการตอบข้อสอบแต่ละข้อผ่านไปแล้วคอมพิวเตอร์ก็จะประมาณค่าความสามารถของผู้สอบ ค่าสารสนเทศของข้อสอบรวมทั้งค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถเมื่อค่าความคลาดเคลื่อนมาตรฐานดังกล่าวมีค่าเท่ากับหรือต่ำกว่าที่กำหนดไว้ การทดสอบก็จะยุติลงโดยอัตโนมัติ ซึ่งในการประมาณค่าความสามารถด้วยวิธีความเป็นไปได้สูงสุดนั้นได้ใช้ค่าส่วนกลับของค่ารากที่สองของค่าความสามารถเป็นค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ส่วนการประมาณค่าความสามารถตามวิธีของเบส์ (bayesian) นั้นใช้รากที่สองของความแปรปรวนภายหลังของการแจกแจงความสามารถ (posterior variance of the ability distribution) เป็นค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ

1.4 การดำเนินการสอบ

ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ มีขั้นตอนในการดำเนินการสอบ ดังนี้

1.4.1 เริ่มต้นการทดสอบโดยให้ผู้สอบนั่งที่หน้าจอคอมพิวเตอร์ซึ่งได้รับการติดตั้งโปรแกรมสำหรับดำเนินการสอบเรียบร้อยแล้ว

1.4.2 ก่อนที่จะตอบข้อสอบจากคอมพิวเตอร์ ควรมีการอธิบายวิธีการตอบให้เข้าใจตรงกัน ซึ่งอาจใช้วิธีการพิมพ์คำชี้แจงให้ผู้สอบอ่านเอง หรือผู้ดำเนินการสอบอาจชี้แจงให้ผู้สอบฟังพร้อมๆกัน ทั้งนี้เพราะการทดสอบมีลักษณะแตกต่างไปจากเดิม

1.4.3 ตรวจสอบเครื่องคอมพิวเตอร์ให้พร้อมที่จะดำเนินการสอบ

1.4.4 ให้ผู้สอบกรอกข้อมูลส่วนตัวตามที่ผู้ดำเนินการสอบต้องการ เช่น เลขที่ ชื่อ

1.4.5 คอมพิวเตอร์จะเลือกข้อสอบข้อแรกขึ้นมาที่หน้าจอให้ผู้สอบได้ตอบ เมื่อตอบเสร็จก็จะตรวจให้คะแนนความสามารถทันที นอกจากนั้นยังประมาณค่าอื่นๆ อีกด้วย เช่น ค่าสารสนเทศของข้อสอบ ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ

1.4.6 คอมพิวเตอร์จะคัดเลือกข้อสอบจากคลังข้อสอบขึ้นมาให้ผู้สอบครั้งละ 1 ข้อตามเกณฑ์การคัดเลือกข้อสอบที่ผู้ดำเนินการสอบกำหนดไว้

1.4.7 การทดสอบจะดำเนินไปเรื่อยๆจนกระทั่งพบเกณฑ์สำหรับยุติการสอบที่ผู้ดำเนินการสอบระบุไว้ การทดสอบก็จะสิ้นสุดลง ค่าต่างๆ ที่ได้จากการประมาณค่าก็จะถูกบันทึกเก็บไว้ในฐานแม่เหล็กโดยอัตโนมัติ

1.5 ตัวบ่งชี้คุณภาพของการทดสอบ

คุณภาพของเครื่องมือวัดผลนับว่าเป็นสิ่งที่มีความสำคัญเป็นอย่างยิ่ง ทั้งนี้เพราะเครื่องมือวัดผลที่มีคุณภาพนั้นจะให้ข้อมูลที่ถูกต้องและเชื่อถือได้ ซึ่งทำให้ผลที่ได้จากการวัดมีคุณค่า สิ่งที่ยอมรับเป็นตัวบ่งชี้คุณภาพของเครื่องมือวัดที่วิเคราะห์ข้อมูลตามทฤษฎีการทดสอบแบบเดิม ได้แก่ ค่าความเที่ยงตรง ความเชื่อมั่น ความยากง่ายและอำนาจจำแนก ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ นั้น ข้อมูลที่ได้จากการทดสอบจะถูกวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ ซึ่งตัวบ่งชี้คุณภาพของการทดสอบลักษณะนี้ ก็คือ (Green and others. 1984 :347-360)

1.5.1 มิติที่ใช้วัด (dimensionality) ข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ต้องมีความเป็นมิติเดียว (unidimensionality) คือ แต่ละข้อควรวัดเพียงโครงสร้างเดียวเท่านั้น ทั้งนี้เพื่อจะได้สอดคล้องกับข้อตกลงเบื้องต้นตาม

ทฤษฎีการตอบสนองข้อสอบ อย่างไรก็ตาม การที่จะสร้างเครื่องมือวัดผลให้มีความเป็นมิติเดียวอย่างแท้จริงนั้นนับว่าเป็นเรื่องที่ยากมาก โดยเฉพาะการวัดผลทางการศึกษา ดังนั้น หากเครื่องมือที่สร้างขึ้นมามีมิติใดมิติหนึ่งที่เด่น (dominant dimension) กว่ามิติอื่นก็นับว่าพอเพียงแล้ว ซึ่งเกณฑ์ที่ใช้พิจารณาก็คือ หากนำข้อมูลที่ได้จากการทดสอบมาทำการวิเคราะห์องค์ประกอบแล้ว หากพบว่ามีองค์ประกอบใดองค์ประกอบหนึ่งสามารถอธิบายความแปรปรวนร่วม (common variance) ได้ถึงร้อยละ 70 ขึ้นไป (Green and others. 1984 : 351) หรือถ้าค่าไอเกนของตัวประกอบตัวที่ 1 สูงกว่าค่าไอเกนของตัวประกอบที่ 2 ตั้งแต่ 3 เท่าขึ้นไป (วรณัฐ แหยมแสง. 2533 : 109) และค่าไอเกนของตัวประกอบที่เหลือมีค่าใกล้เคียงกันแล้ว ก็ถือได้ว่าแบบทดสอบนั้นมีความเป็นมิติเดียว สิ่งที่จะช่วยให้แบบทดสอบมีความเป็นมิติเดียว ก็คือ ข้อสอบในแบบทดสอบต้องมีค่าอำนาจจำแนกสูงๆ ซึ่งยูริ (Urry. 1981) แนะนำว่า ควรเลือกข้อสอบที่มีค่าอำนาจจำแนกอย่างน้อย 0.8 แต่อย่างไรก็ตามในบางครั้งข้อสอบที่มีค่าอำนาจจำแนกต่ำกว่านี้ก็อาจถูกเลือกขึ้นมาใช้ด้วย ทั้งนี้เพื่อให้มีข้อสอบที่ง่ายและยากๆบ้างในคลังข้อสอบ ซึ่งจะทำให้ค่าความยากง่ายของข้อสอบครอบคลุมช่วงความสามารถของผู้สอบตั้งแต่ต่ำถึงสูง

ในการสร้างคลังข้อสอบเพื่อใช้วัดผลสัมฤทธิ์ทางการเรียนนั้นอาจมีข้อจำกัดในเรื่องของความเป็นมิติเดียวกันของข้อสอบ ทั้งนี้เนื่องจากธรรมชาติของผลสัมฤทธิ์ทางการเรียนนั้นย่อมประกอบด้วยหลายโครงสร้างหลายเนื้อหาแม้ว่าจะเป็นวิชาเดียวกันก็ตาม แนวทางแก้ไขก็คือพยายามแบ่งโครงสร้างและเนื้อหาเหล่านั้นให้เป็นกลุ่มย่อยๆ ซึ่งแต่ละกลุ่มมีความเป็นมิติเดียวกัน

1.5.2 ความเที่ยงตรง (validity) จากงานวิจัยที่ผ่านมาพบว่า การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์นั้นมีประสิทธิภาพสูงกว่าการทดสอบแบบเดิม แต่คะแนนจากการทดสอบทั้งสองลักษณะดังกล่าวมีความสัมพันธ์กันสูง หมายความว่าค่าความเที่ยงตรงของแบบทดสอบแบบเดิมสามารถถ่ายโอนไปใช้กับการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ที่วัดคุณลักษณะเดียวกันได้ การศึกษาชนิดของความเที่ยงตรงสำหรับการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบนิยมศึกษา 2 ชนิด คือ ความเที่ยงตรงเชิงโครงสร้าง (construct validity) เป็นการศึกษาความสัมพันธ์ระหว่างความสามารถจริง (θ) กับความสามารถที่ได้จากการประมาณค่า ($\hat{\theta}$) ซึ่งสามารถหาได้จาก (Urry. 1977 : 183 , Cronbach Meehl. 1957)

$$\rho_{\theta\hat{\theta}} = \sqrt{1 - \sigma_e^2}$$

เมื่อ $\rho_{\theta\theta}$ แทน ความสัมพันธ์ระหว่างความสามารถจริงกับความสามารถ
ที่ได้จากการประมาณค่า ซึ่งใช้เป็นค่าความเที่ยงตรงเชิง
โครงสร้าง

σ_e แทน ความคลาดเคลื่อนมาตรฐานในการประมาณค่า
ความสามารถ

ความเที่ยงตรงอีกประเภทหนึ่งก็คือ ความเที่ยงตรงเชิงทำนาย (predictive validity)
เป็นการศึกษาผลของการทดสอบในปัจจุบัน กับคะแนนเกณฑ์ชนิดใดชนิดหนึ่งที่เกิดขึ้นในอนาคต
ค่าความเที่ยงตรงเชิงทำนายนี้สามารถหาค่าได้โดยอาศัยทฤษฎีการทดสอบแบบเดิม ดังนี้

$$r(T_p, X) = \frac{r(Y_p, X)}{\sqrt{\rho_p}}$$

$$r(T_c, X) = \frac{r(Y_c, X)}{\sqrt{\rho_c}}$$

เมื่อ X แทน คะแนนเกณฑ์

$r(T_p, X)$ แทน สหสัมพันธ์ระหว่างคะแนนจริงที่ได้จากการทดสอบ
แบบเดิมกับคะแนนเกณฑ์

$r(T_c, X)$ แทน สหสัมพันธ์ระหว่างคะแนนจริงที่ได้จากการทดสอบแบบ
ปรับเหมาะกับความสามารถของผู้สอบกับคะแนนเกณฑ์

Y_p, Y_c แทน คะแนนที่ได้จากการทดสอบแบบเดิมและการทดสอบ
แบบปรับเหมาะกับความสามารถของผู้สอบ

T_p, T_c แทน คะแนนจริงที่ได้จากการทดสอบแบบเดิมและการ
ทดสอบ แบบปรับเหมาะกับความสามารถของผู้สอบ

ρ_p, ρ_c แทน ค่าความเชื่อมั่นของแบบทดสอบแบบเดิมและการทดสอบ
แบบปรับเหมาะกับความสามารถของผู้สอบ

เนื่องจาก คะแนนจริงที่ได้จากการทดสอบแบบเดิมและแบบ ซี เอ ที่นั้นมีความ
สัมพันธ์กันเชิงเส้นตรง ดังนั้น จะได้

$$r(Y_c, X) = r(Y_p, X) \sqrt{\frac{\rho_c}{\rho_p}}$$

1.5.3 ความเชื่อมั่น (reliability)

ตามทฤษฎีการทดสอบแบบเดิมนั้นความเชื่อมั่นได้ถูกนิยามว่าเป็น อัตราส่วนระหว่างความแปรปรวนของคะแนนจริงกับความแปรปรวนของคะแนนที่สอบได้ของ ประชากรผู้สอบ ซึ่งความแปรปรวนของคะแนนจริงก็เป็นผลต่างของความแปรปรวนของคะแนนที่ สอบได้กับความแปรปรวนของความคลาดเคลื่อนซึ่งค่าความแปรปรวนของความคลาดเคลื่อน ดังกล่าวถูกนำไปใช้ในลักษณะที่เป็นค่าเฉลี่ย คือใช้เพียงค่าเดียวกับทุกค่าของคะแนนที่สอบได้ จากความสัมพันธ์ดังกล่าวจึงทำให้ค่าความเชื่อมั่นของแบบทดสอบขึ้นอยู่กับความเป็นเอกพันธ์ ของกลุ่มผู้สอบ กล่าวคือ ถ้ากลุ่มผู้สอบมีความเป็นเอกพันธ์กันสูง ค่าความแปรปรวนของคะแนน ก็จะต่ำ ซึ่งส่งผลให้ค่าความเชื่อมั่นสูงตามไปด้วย หากใช้ค่าความแปรปรวนของความคลาดเคลื่อน แบบมีเงื่อนไข (conditional error variance) ที่จุดต่างๆ กันของระดับความสามารถก็จะทำให้ได้รับ สารสนเทศมากกว่าเดิม ตามทฤษฎีการตอบสนองข้อสอบนั้น ค่าความแปรปรวนของความ คลาดเคลื่อนจะเป็นฟังก์ชันของความสามารถของผู้สอบ ซึ่งการประมาณค่าความสามารถด้วย วิธีความเป็นไปได้สูงสุด (maximum likelihood) นั้น ค่าความแปรปรวนของความคลาดเคลื่อน ในการวัด ก็คือ ส่วนกลับของค่าฟังก์ชันสารสนเทศของแบบทดสอบ ส่วนทฤษฎีของเบย์ (bayesian theory) นั้น ค่าความแปรปรวนของความคลาดเคลื่อน ก็คือค่าความแปรปรวนภายหลังของค่า ความสามารถ (posterior variance)

1.5.4 องค์ประกอบด้านตัวบุคคลและแบบแผนของระบบ (human factors and system design)

ในการดำเนินการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วย คอมพิวเตอร์จะต่างกับการทดสอบแบบเดิม ข้อสอบซึ่งตามปกติจะพิมพ์มาเป็นชุดๆ ในกระดาษ ก็จะเปลี่ยนเป็นการนำเสนอทางหน้าจอคอมพิวเตอร์ คำตอบจะถูกตอบผ่านทางแป้นพิมพ์ หรือ ปากกาแสงหรืออาจเป็นอุปกรณ์อื่นๆ ดังนั้นการดำเนินการสอบจึงควรเป็นไปอย่างระมัดระวัง สิ่งแวดล้อมในขณะที่สอบต้องเงียบสงบและสะดวกสบาย หากเป็นไปได้แล้วก็ควรแยกจอภาพ ของคอมพิวเตอร์ให้อยู่ห่างๆ กัน จอภาพไม่ควรมีแสงสะท้อนรบกวนผู้สอบผู้สอบสามารถมองเห็น ข้อสอบได้อย่างชัดเจน อุปกรณ์ที่ใช้สำหรับตอบควรเหมาะสมสำหรับผู้ที่นั่งตั่งมือซ้ายและขวา ปุ่มที่ใช้กดเพื่อตอบควรอยู่ตรงกลาง ไม่ควรอยู่ข้างใดข้างหนึ่ง หรืออย่างน้อยที่สุดก็ควรเคลื่อนย้าย อุปกรณ์ดังกล่าวได้สะดวก ควรอธิบายขั้นตอนของการทดสอบเหมือนการทดสอบแบบเดิม ตัวอย่างข้อสอบและวิธีการตอบควรถูกนำเสนอต่อผู้สอบอย่างชัดเจนการทดสอบแบบปรับเหมาะ กับความสามารถของผู้สอบด้วยคอมพิวเตอร์ไม่อนุญาตให้ผู้สอบเว้นหรือข้ามการตอบข้อสอบ เป็นบางข้อ หรือกลับไปแก้ไขคำตอบในข้อต้นๆ ที่ผ่านมาแล้ว ทั้งนี้เพราะการนำเสนอข้อสอบ

แต่ละข้อนั้นจะขึ้นอยู่กับผลการตอบข้อสอบข้อที่ผ่านมา ดังนั้นหากมีการเปลี่ยนแปลงคำตอบในข้อแรกๆก็จะทำให้ลำดับการนำเสนอข้อสอบเปลี่ยนแปลงไป

1.5.5 ความเท่าเทียมกัน (equity) ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์นั้น ผู้สอบแต่ละคนจะได้รับข้อสอบแตกต่างกันไปตามความสามารถของผู้สอบแต่ละคน ผู้สอบที่มีความสามารถสูงก็จะได้รับข้อสอบที่ยากผู้สอบที่มีความสามารถต่ำก็จะได้รับข้อสอบที่ง่ายกว่า ปัญหาตามมา ก็คือ จะอธิบายแก่นบุคคลทั่วไปให้เข้าใจได้อย่างไรว่าเราสามารถนำผลการสอบของผู้สอบซึ่งได้รับข้อสอบต่างกันมาเปรียบเทียบกันได้ ซึ่งในกรณีนี้สามารถอธิบายได้ว่าคลังข้อสอบที่มีความเป็นมิติเดียวนั้นจะผลิตแบบทดสอบคู่ขนานอย่างสุ่มและผลที่ได้จากการสอบแบบทดสอบต่างฉบับก็จะถูกเทียบคะแนนโดยอัตโนมัติ จากทฤษฎีการตอบสนองข้อสอบนอกจากนั้นในการยุติการสอบของผู้สอบแต่ละคนก็มักใช้เกณฑ์ความคลาดเคลื่อนมาตรฐานเท่ากันทุกคน

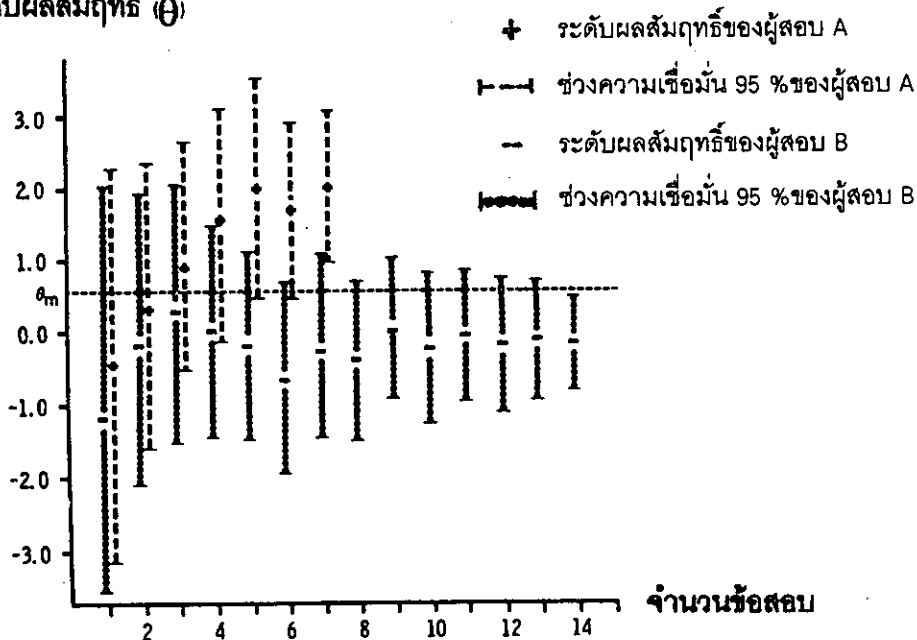
1.6 การนำการทดสอบแบบปรับเหมาะไปใช้ทางการศึกษา

การทดสอบแบบปรับเหมาะความสามารถของผู้สอบด้วยคอมพิวเตอร์นั้นเหมาะที่จะนำไปใช้วัดความสามารถ (ability) ของผู้สอบ เนื่องจากผู้ดำเนินการสอบสามารถที่จะกำหนดขอบเขตของสิ่งที่ต้องการวัดให้มีความเป็นมิติเดียวได้ง่ายและชัดเจนกว่าการสอบในลักษณะอื่น ซึ่งความเป็นมิติเดียวนั้นเป็นข้อตกลงเบื้องต้นที่มีความสำคัญในการวิเคราะห์ข้อมูลจากการทดสอบโดยใช้ทฤษฎีการตอบสนองข้อสอบ แม้ว่าการทดสอบดังกล่าวจะมีความเหมาะสมกับการวัดความสามารถก็ตามแต่ก็สามารถนำมาประยุกต์ใช้ในการวัดผลทางการศึกษาได้เช่นเดียวกัน ซึ่งสามารถทำได้หลายลักษณะ ดังนี้ (Weiss, 1988 : 375-376 ; Weiss and Kingbury, 1984 : 365-373)

1.6.1 เพื่อจำแนกและตัดสินความรอบรู้ (mastery) การทดสอบเพื่อจำแนกและตัดสินความรอบรู้ด้วยการทดสอบแบบปรับเหมาะ ซึ่งมีชื่อว่า การทดสอบเพื่อวัดความรอบรู้แบบปรับเหมาะความสามารถของผู้สอบ (adaptive mastery testing , AMT) ซึ่งออกแบบขึ้นมาเพื่อให้ผลของการทดสอบสามารถระบุได้ว่าระดับผลสัมฤทธิ์ทางการเรียนของนักเรียนอยู่สูงหรือต่ำกว่าระดับความรอบรู้ที่ได้กำหนดไว้เป็นเกณฑ์ก่อนที่จะสอบ ซึ่งเกณฑ์ดังกล่าวได้มาจากการแปลงค่าของสัดส่วนการตอบถูกไปเป็นคะแนนความสามารถตามทฤษฎีการตอบสนองข้อสอบโดยใช้ค่าเฉลี่ยของโค้งลักษณะข้อสอบ (item characteristic curve , ICC) ทุกข้อในคลังข้อสอบ (Weiss and Kingbury, 1984 : 364 ; citing Weiss and Kingbury, 1979: 2-4) ก่อนที่จะเริ่มสอบ จะมี การกำหนดว่า $\theta_0 = \theta_m$ ซึ่ง θ_0 ก็คือ ระดับเริ่มต้นโดยประมาณของผลสัมฤทธิ์ ในการเริ่มต้น

ทดสอบนั้น ข้อสอบในคลังข้อสอบที่มีค่าสารสนเทศสูงสุดที่ θ_m จะถูกเลือกขึ้นมาให้ผู้สอบได้ตอบ คำตอบของผู้สอบจะถูกให้คะแนนและระดับความสามารถหรือระดับผลสัมฤทธิ์ทางการเรียนจะถูกประมาณค่า ผลที่ได้ก็คือ ค่าโดยประมาณของผลสัมฤทธิ์ทางการเรียนที่ θ_1 และช่วงความเชื่อมั่น (confidence interval) ของการประมาณค่าผลสัมฤทธิ์ จะมีการพิจารณาว่า θ_m นั้น ตกอยู่ในช่วงความเชื่อมั่นหรือไม่ ถ้า θ_m ยังอยู่ในช่วงความเชื่อมั่น การทดสอบก็ยังคงดำเนินต่อไปในลักษณะเดิมไปเรื่อยๆ จนกระทั่งค่าของ θ_m ไม่ตกอยู่ในช่วงความเชื่อมั่น การดำเนินการสอบก็จะยุติลง ถ้าขีดจำกัดล่าง (lower limit) ของช่วงความเชื่อมั่นอยู่เหนือระดับความรอบรู้ (θ_m) ก็ถือว่าผู้สอบเป็นผู้ที่รอบรู้ในเรื่องที่ถูกวัด ในทางกลับกัน ถ้าขีดจำกัดบน (upper limit) ของช่วงความเชื่อมั่นอยู่ต่ำกว่าระดับความรอบรู้ (θ_m) ก็ถือว่าผู้สอบเป็นผู้ที่ไม่รอบรู้ในเรื่องที่ถูกวัด ซึ่งแสดงได้ดังภาพประกอบ 12

ระดับผลสัมฤทธิ์ (θ)



ภาพประกอบ 12 ระดับผลสัมฤทธิ์ทางการเรียนและช่วงความเชื่อมั่นในการ ประมาณค่า 95 % ของผู้สอบ A และ B

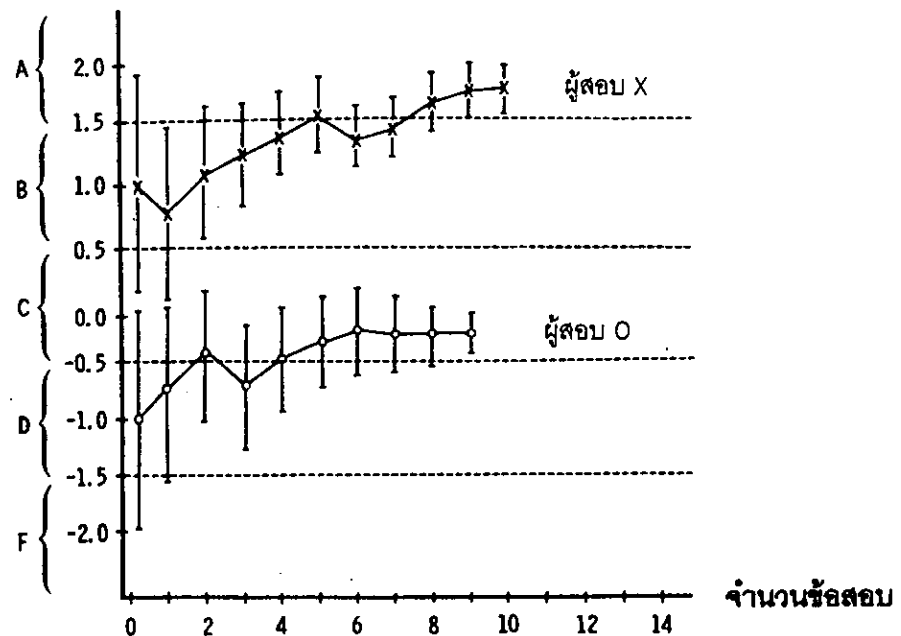
1.6.2. เพื่อกำหนดระดับผลการเรียน (assigning grades) การให้ระดับผลการเรียน จะเกี่ยวข้องกับการจัดตำแหน่งให้แก่ผู้สอบเป็นกลุ่มๆ โดยนำผลสัมฤทธิ์ทางการเรียนซึ่งได้ถูกวัด บนสเกลที่ต่อเนื่องมาใช้เป็นข้อมูลในการให้ระดับผลการเรียน มีการปรับเปลี่ยนการทดสอบเพื่อวัด

ความรู้แบบปรับเหมาะกับความสามารถของผู้สอบ (adaptive mastery testing, AMT) เพื่อให้กลายเป็นแบบทดสอบเพื่อกำหนดระดับผลการเรียนแบบปรับเหมาะกับความสามารถของผู้สอบ (adaptive grading test, AGT) ดังนี้

(1) กำหนดจุดตัด $N-1$ จุด ถ้าต้องการจำแนกผู้สอบออกเป็น N ระดับ โดยใช้วิธีการเหมือนกับการกำหนดจุดตัดเพื่อตัดสินความรู้ นั่นคือ แปลงเมตริกสัดส่วนของคะแนนการตอบถูกให้เป็นเมตริกของคะแนนความสามารถ (θ) ตามทฤษฎีการตอบสนองข้อสอบ

(2) การทดสอบจะยุติเมื่อช่วงความเชื่อมั่นของระดับผลสัมฤทธิ์โดยประมาณของผู้สอบตกอยู่ระหว่าง 2 จุดตัดอย่างสมบูรณ์ หรืออยู่เหนือระดับจุดตัดสูงสุดหรืออยู่ต่ำกว่าระดับจุดตัดต่ำสุด ซึ่งจากภาพประกอบ 13 ผู้สอบ O นั้นตอบข้อสอบเพียง 9 ข้อ ก็สามารถตัดสินได้ว่ามีผลการเรียนอยู่ในระดับ C ส่วนผู้สอบ X ต้องตอบข้อสอบ 10 ข้อจึงสามารถระบุได้ว่ามีผลการเรียนอยู่ในระดับ A

ระดับผลสัมฤทธิ์ (θ)

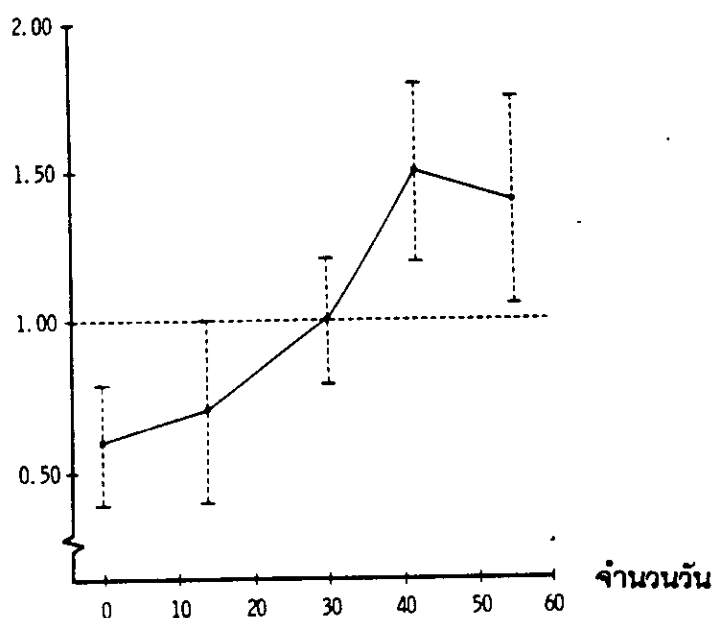


ภาพประกอบ 13 ระดับผลสัมฤทธิ์ทางการเรียนและช่วงความเชื่อมั่นในการประมาณค่า 95% ของผู้สอบ X และ O

1.6.3 เพื่อวัดความเปลี่ยนแปลงของผู้สอบ (individual change) นอกจากจะนำผลของการทดสอบไปใช้ประโยชน์ในการตัดสินความรู้และกำหนดระดับผลการเรียนของนักเรียน

แล้ว ก็ยังสามารถนำผลการสอบลักษณะนี้ไปใช้ในการวัดความเปลี่ยนแปลงของผู้สอบอันเนื่องมาจากการสอนได้อีกด้วย ในการวัดความเปลี่ยนแปลงของผู้สอบนี้อาจถูกพิจารณาว่าเป็นการวัดผลของการสอน โดยเฉพาะการสอนที่มีความยืดหยุ่น ซึ่งเรียนตามความแตกต่างระหว่างบุคคล และความแตกต่างระหว่างบุคคลนี้ก็เป็นจุดหมายสำคัญสำหรับการวินิจฉัยการวัดความเปลี่ยนแปลงของผู้สอบแต่ละคนโดยใช้การทดสอบแบบเดิมนั้นมีปัญหาสำคัญเกี่ยวกับความเชื่อมั่นของคะแนนการเปลี่ยนแปลง นั่นคือ คะแนนการเปลี่ยนแปลงของผู้สอบจากจุดหนึ่งถึงอีกจุดหนึ่งนั้น มักมีความเชื่อมั่นต่ำ (Lord, 1958 ; Traub, 1967) ในการวัดความเปลี่ยนแปลงของผู้สอบนั้นเป็นการหาความแตกต่างระหว่างผลสัมฤทธิ์ทางการเรียนในสองช่วงเวลา ซึ่งการเปลี่ยนแปลงดังกล่าวจะมีนัยสำคัญก็ต่อเมื่อช่วงของความเชื่อมั่นในการประมาณค่าความสามารถตามทฤษฎีการตอบสนองข้อสอบสองครั้งไม่มีส่วนที่ซ้อนกันเราจะสามารถเห็นแนวโน้มของการเปลี่ยนแปลงได้โดยนำผลการทดสอบแต่ละครั้งมาสร้างเป็นแผนภูมิดังภาพประกอบ 14

ระดับผลสัมฤทธิ์ (θ)



ภาพประกอบ 14 ผลสัมฤทธิ์ทางการเรียนและช่วงความเชื่อมั่นในการประมาณค่าจากการสอบวัด 5 ครั้ง คือ ก่อนเรียนและในระหว่างเรียน

จากภาพประกอบ 14 จะเห็นว่า ก่อนที่จะสอน(วันที่ 0) ผู้สอบมีผลสัมฤทธิ์เท่ากับ .60 ด้วยความเชื่อมั่น 95% หลังจากสอนไปได้ 14 วัน ผลสัมฤทธิ์เท่ากับ .65 ช่วงความเชื่อมั่นของวันที่

0 กับวันที่ 14 ทับกันสนิท แสดงว่าตั้งแต่วันแรกจนกระทั่งถึงวันที่ 14 นั้นผลสัมฤทธิ์ยังไม่มี การเปลี่ยนแปลง ในการทดสอบวันที่ 30 พบว่า ผลสัมฤทธิ์เท่ากับ 1 และช่วงของความเชื่อมั่นจะไม่ ข้อนกับการทดสอบเมื่อวันที่ 0 แสดงว่า เกิดการเปลี่ยนแปลงของผลสัมฤทธิ์ของผู้สอบแล้วระหว่าง วันที่ 0 กับวันที่ 30 เมื่อทดสอบวันที่ 42 ของการเรียน พบว่า ผลสัมฤทธิ์มีค่า 1.5 และช่วงความ เชื่อมั่นไม่ซ้อนทับกับวันที่ 0 , 14 , 30 เลย แสดงว่าผลสัมฤทธิ์เมื่อวันที่ 42 นั้น ได้เปลี่ยนแปลง ไปจากวันที่ 0 , 14 และ 30 ส่วนการทดสอบครั้งที่ 5 ในวันที่ 55 นั้น ปรากฏว่า การเปลี่ยนแปลง จะต่างจากวันที่ 0 และ 14 แต่ยังไม่ต่างจากวันที่ 30 และ 42

1.7 ข้อดีของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ

การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์นั้นมี ประสิทธิภาพสูงกว่าการทดสอบแบบเดิมหลายประการ คือ

1.7.1 ใช้เวลาในการทดสอบน้อยลงกว่าเดิมเนื่องจากผู้สอบได้รับข้อสอบน้อยกว่า เดิม ซึ่งข้อสอบที่ผู้สอบได้รับนี้เป็นข้อสอบที่มีค่าความยากง่ายเหมาะสมกับความสามารถของ ผู้สอบ (Urry. 1977 : 194 ; Weiss. 1982 : 473 - 492 ; Moreno and others. 1984 : 155-163 ; Weiss. 1984 : 362 ; Hansen 1969 ; Larkin and Weiss. 1974 ; Hulin and others. 1983.)

1.7.2 สามารถรายงานผลการทดสอบได้ทันทีที่การทดสอบลักษณะนี้ได้ใช้ คอมพิวเตอร์เข้ามาช่วยในการดำเนินการสอบ ข้อสอบจะถูกเลือกขึ้นมาที่หน้าจอคอมพิวเตอร์ ครั้งละ 1 ข้อตามเกณฑ์การคัดเลือกข้อสอบที่กำหนดไว้ คอมพิวเตอร์จะเป็นผู้ตรวจข้อสอบและ ประมวลผลค่าต่างๆ หลังจากที่คุณสอบตอบข้อสอบแต่ละข้อ เช่น ค่าความสามารถของผู้สอบ ค่า ความคลาดเคลื่อนมาตรฐานในการประมาณค่า เมื่อยุติการสอบลงค่าต่างๆ เหล่านี้สามารถ แสดงให้เห็นได้ทันที

1.7.3 สามารถบันทึกผลการทดสอบได้ทันที ในการทดสอบแต่ละครั้งนั้น หาก ผู้ดำเนินการสอบต้องการข้อมูลเกี่ยวกับการทดสอบส่วนใดส่วนหนึ่งเก็บไว้ เช่น เวลาในการตอบ ข้อสอบแต่ละข้อ เวลาที่ใช้ในการทดสอบทั้งหมด คะแนนความสามารถและความคลาดเคลื่อน มาตรฐานในการประมาณค่าความสามารถ ก็สามารถสั่งให้คอมพิวเตอร์เก็บข้อมูลต่างๆเหล่านี้ได้

1.7.4 ผู้สอบมีเจตคติที่ดีต่อการสอบเพราะผู้สอบแต่ละคนได้ข้อสอบที่เหมาะสม สมกับความสามารถของตนเอง ผู้สอบที่มีความสามารถสูงก็ไม่เกิดความเบื่อหน่ายเพราะไม่ต้อง ตอบข้อสอบที่ง่ายๆ ผู้ที่มีความสามารถต่ำก็ไม่เกิดความเครียด เพราะไม่ต้องตอบข้อสอบที่ยากๆ (Lee. 1987 : 3742-A ; Hulin and others 1983 : 210)

1.7.5 สามารถทดสอบกับนักเรียนเป็นรายบุคคลได้โดยไม่ต้องรอสอบพร้อมกัน นักเรียนคนใดพร้อมที่จะสอบก็เปิดเครื่องคอมพิวเตอร์แล้วทดสอบได้ทันที ซึ่งจะเป็นประโยชน์อย่างยิ่งต่อการจัดการเรียนการสอนในด้านการปรับปรุงการเรียนการสอน

1.7.6 ช่วยรักษาความลับและป้องกันการทุจริตในการสอบเป็นอย่างดี เพราะข้อสอบทั้งหมดถูกเก็บเอาไว้ในลักษณะของคลังข้อสอบ ซึ่งไม่สามารถเปิดดูได้อย่างง่ายดาย เหมือนการพิมพ์ข้อสอบไว้เป็นเอกสารเหมือนแต่เดิม สำหรับกรณีการทุจริตในการการสอบนั้น สามารถป้องกันได้ เพราะผู้สอบไม่สามารถลอกคำตอบจากเพื่อนได้ถึงแม้ว่าจะนั่งใกล้กันก็ตาม ทั้งนี้เพราะข้อสอบที่หน้าจอคอมพิวเตอร์นั้นไม่ใช่ข้อสอบข้อเดียวกัน

1.7.7 ประหยัดค่าใช้จ่าย การทดสอบลักษณะนี้จะมีค่าใช้จ่ายสูงในตอนเริ่มแรก กล่าวคือ จำเป็นต้องมีเครื่องคอมพิวเตอร์เพื่อให้เก็บข้อสอบและดำเนินการสอบ แต่เมื่อเทียบระยะเวลาที่นานๆ แล้ว การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์จะเสียค่าใช้จ่ายน้อยกว่า เพราะไม่จำเป็นต้องสร้างข้อสอบใหม่ทุกครั้ง แต่การทดสอบแบบเดิมนั้น ต้องสร้างแบบทดสอบขึ้นมาใหม่เกือบทุกครั้งที่มีการทดสอบ

1.7.8 สามารถกำหนดความถูกต้องในการประมาณค่าความสามารถของผู้สอบได้ โดยกำหนดค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบเป็น เกณฑ์ยุติการทดสอบ ถ้ากำหนดค่าความคลาดเคลื่อนสูงความถูกต้องในการประมาณค่าก็จะต่ำ แต่ถ้ากำหนดค่าความคลาดเคลื่อนในการประมาณค่าต่ำความถูกต้องในการประมาณค่าก็จะสูง ซึ่งต่างจากการทดสอบแบบเดิมที่ผู้ดำเนินการสอบไม่สามารถกำหนดค่าความถูกต้องในการวัดล่วงหน้าก่อนการทดสอบได้

1.7.9 เพิ่มค่าความเที่ยงตรงและความเชื่อมั่นของการทดสอบได้ เพราะการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบจะตัดข้อสอบที่ไม่เหมาะกับความสามารถของผู้สอบออกไป ซึ่งการทดสอบด้วยข้อสอบที่มีความเหมาะสมกับความสามารถของผู้สอบจะช่วยให้ความเที่ยงตรงและความเชื่อมั่นของการวัดมีค่าสูงขึ้น (Larkin and Weiss. 1974 : 1 ; Weiss. 1984 : 362)

2. ทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบเป็นทฤษฎีการวัดที่ถูกคิดขึ้นมาโดยมีวัตถุประสงค์เพื่อเสนอหลักการพื้นฐานที่สำคัญสำหรับการทำนาย ประเมินค่าและสรุปอ้างอิงเกี่ยวกับความสามารถ หรือคุณลักษณะที่สามารถวัดได้ด้วยแบบทดสอบ (Hambleton and Swaminathan. 1985 : 53) เพื่อให้เกิดความชัดเจนเกี่ยวกับทฤษฎีการตอบสนองข้อสอบ ผู้วิจัยจึงขอเสนอรายละเอียดเป็นด้านๆ ดังนี้

2.1 ประวัติความเป็นมา

ทฤษฎีการตอบสนองข้อสอบมีประวัติความเป็นมาเรียงตามลำดับดังนี้ (Hambleton and Swaminathan. 1985 : 4-5)

- ค.ศ.1916 บิเนตและไซมอน (Binet and Simon) ได้กำหนดระดับผลการปฏิบัติ (performance levels) ของผู้สอบตามตัวแปรอิสระต่างๆและใช้ผลการปฏิบัติดังกล่าวไปพัฒนาแบบทดสอบ
- ค.ศ.1936 ริชาร์ดสัน (Richardson) ได้หาความสัมพันธ์ระหว่างค่าพารามิเตอร์ของข้อสอบตามโมเดลทฤษฎีการตอบสนองข้อสอบและทฤษฎีการทดสอบแบบเดิม
- ค.ศ.1943-1944 ลอว์เลย์ (Lawley) ได้คิดวิธีการใหม่ๆ เพื่อใช้สำหรับการประมาณค่าพารามิเตอร์
- ค.ศ.1952 ลอร์ด (Lord) ได้นำเสนอโมเดลนอร์มอลโอไจฟ์ (normal ogive model) แบบ 2 พารามิเตอร์โดยนำเสนอมูลวิธีการประมาณค่าพารามิเตอร์และการนำโมเดลไปประยุกต์ใช้
- ค.ศ.1957-1958 เบอ์นบอม (Birnbaum) ได้นำเสนอโมเดลโลจิสติก (logistic model) เพื่อใช้แทนโมเดลนอร์มอลโอไจฟ์
- ค.ศ.1960 ราสช์ (Rasch) ได้พัฒนาโมเดลการตอบสนองข้อสอบ 3 โมเดล โดยพรรณนาไว้ในหนังสือ Probabilistic Model for some Intelligence and Attainment Tests ซึ่งผลงานชุดนี้มีอิทธิพลต่อไรท์ (Wright) ซึ่งอยู่ในอเมริกาและนักจิตวิทยาบางคน เช่น แอนเดอร์เซนและฟิชเชอร์ (Andersen and Fisher) ในยุโรป
- ค.ศ.1967 ไรท์ (Wright) ได้เป็นผู้นำและกระตุ้นให้มีการวิจัยในอเมริกาเกี่ยวกับโมเดลของราสช์ (Rasch) ตลอดช่วงทศวรรษ 1970

- ค.ศ.1968 ลอร์ดและโนวิก (Lord and Novick) ได้นำเสนอทฤษฎีคุณลักษณะแฝง (theory of latent trait)
- ค.ศ.1969 ไรท์และปันชาปากีสาน (Wright and Panchapakesan) ได้พรรณนาวิธีการประมาณค่าพารามิเตอร์ในโมเดลของราส์ช โดยใช้โปรแกรมไบคาล (BICAL) และในปีเดียวกันนี้ เซมีจิมา (Samejima) ได้พรรณนาโมเดลการตอบสนองข้อสอบโมเดลใหม่พร้อมทั้งการประยุกต์ใช้ โมเดลนี้ใช้ได้กับข้อมูลที่ได้จากการตอบหลายแบบและแบบต่อเนื่อง (polychotomous and continuous) และปรับปรุงโมเดลมิติเดียว (unidimension) ไปสู่มอเดลแบบหลายมิติ (multidimension)
- ค.ศ.1972 บอกซ์ (Bock) ได้สนับสนุนแนวคิดใหม่ๆหลายแนวคิดเกี่ยวกับการประมาณค่าพารามิเตอร์
- ค.ศ.1974 ลอร์ด (Lord) ได้พรรณนาวิธีการประมาณค่าพารามิเตอร์วิธีใหม่ โดยใช้โปรแกรมโลจิส (LOGIST) และฟิชเชอร์ (Fisher) ได้นำเสนอโปรแกรมการวิจัยเกี่ยวกับโมเดลโลจิสติกเชิงเส้นตรง (linear logistic models)
- ค.ศ.1976 ลอร์ด (Lord) ได้พัฒนาโปรแกรมโลจิสเพื่อให้สามารถประมาณค่าพารามิเตอร์ในโมเดลโลจิสติกได้ดียิ่งขึ้น
- ค.ศ.1977 เบเกอร์ (Baker) นำเสนอผลการทบทวนวิธีการประมาณค่าพารามิเตอร์ และได้ตีพิมพ์ประเด็นต่างๆเป็นกรณีพิเศษเกี่ยวกับการใช้ทฤษฎีการตอบสนองข้อสอบในการวัดผลในวารสาร Journal of Educational Measurement ซึ่งเขียนโดยนักวิจัยหลายคน เช่น บาซอร์ว (Bashaw), ลอร์ด (Lord) , มาร์โค (Marco) , เรนซ์ (Rentz) , ยูรี (Urry) และ ไรท์ (Wright) เป็นต้น
- ค.ศ.1979 ไรท์และสโตน (Wright and Stone) ได้เขียนหนังสือเกี่ยวกับแบบแผนของแบบทดสอบที่ดีที่สุด (best test design) โดยใช้โมเดลของราส์ช (Rasch)
- ค.ศ.1980 ลอร์ด (Lord) ได้ทบทวนพัฒนาการของทฤษฎีการตอบสนองข้อสอบ และการประยุกต์ใช้โมเดล 3 พารามิเตอร์ โดยเขียนในหนังสือ Application of Item Response Theory to Practical Testing Problems

- ค.ศ.1980 ไวส์ (Weiss) ได้บรรณาธิกรณรายงานการประชุมเกี่ยวกับการดำเนินการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบโดยคอมพิวเตอร์ หรือการทดสอบแบบ ซี เอ ที (computerized adaptive testing ,CAT) เมื่อ ปี 1979
- ค.ศ.1982 ลอร์ดและคณะ (Lord and Staff) ได้ปรับปรุงโปรแกรมโลจิส ครั้งที่ 2 เพื่อให้ใช้ได้สะดวกกว่าเดิม

2.2 ลักษณะทั่วไปของทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบเป็นทฤษฎีที่แสดงความสัมพันธ์ระหว่างการตอบข้อสอบกับคุณลักษณะที่ข้อสอบดังกล่าวต้องการจะวัด (ผจงจิต อินทสุวรรณ. 2534 : 5) ทฤษฎีนี้ได้เกิดขึ้นมาท่ามกลางข้อจำกัดของทฤษฎีการทดสอบแบบเดิมหลายประการ (Hambleton and Swaminathan. 1985 : 2) คือ

2.2.1 ค่าสถิติประจำตัวของผู้สอบ เช่น ค่าความยากง่าย ค่าอำนาจจำแนกนั้น จะขึ้นอยู่กับลักษณะของกลุ่มผู้สอบ กล่าวคือ ถ้าผู้สอบมีความสามารถสูงข้อสอบนั้นก็ง่าย แต่ถ้าผู้สอบมีความสามารถต่ำข้อสอบนั้นก็ยาก ทั้งๆที่เป็นแบบทดสอบฉบับเดิม ส่วนค่าอำนาจจำแนกก็จะขึ้นอยู่กับความเป็นเอกพันธ์ของกลุ่มผู้สอบ ถ้าผู้สอบมีความสามารถแตกต่างกันมาก ข้อสอบก็จะมีค่าอำนาจจำแนกสูง ซึ่งมีผลทำให้ค่าความเชื่อมั่นของแบบทดสอบมีค่าสูงตามไปด้วย ทั้งเนื่องจากค่าความเชื่อมั่นของแบบทดสอบมีความสัมพันธ์โดยตรงกับค่าความแปรปรวนของคะแนน

2.2.2 การเปรียบเทียบความสามารถของผู้สอบนั้นจำเป็นต้องใช้แบบทดสอบฉบับเดียวกันหรือแบบทดสอบที่คู่ขนานกันและค่าความเชื่อมั่นของแบบทดสอบถูกนิยามในรูปของผลที่ได้จากการใช้แบบทดสอบคู่ขนาน ซึ่งในการปฏิบัติจริงๆแล้วนับว่าเป็นเรื่องที่ยากมากที่จะให้การสอบทั้ง 2 ครั้งมีสภาพเหมือนกันทุกประการ ถึงแม้ว่าแบบทดสอบทั้งสองจะคู่ขนานกันจริง เพราะผู้สอบอาจมีลักษณะที่ต่างไปจากการสอบครั้งแรก เช่น แรงงูใจ ความกังวล การลืมหืมหรือการพัฒนาตนเองในบางทักษะ

2.2.3 ทฤษฎีการทดสอบแบบเดิมไม่สามารถบอกได้ว่าผู้สอบจะตอบข้อสอบได้ผลอย่างไร ยกเว้นว่าจะได้ใช้ข้อสอบนั้นกับผู้สอบที่มีลักษณะคล้ายคลึงกันมาก่อนแล้ว

2.2.4 แบบทดสอบวัดผลสัมฤทธิ์และแบบทดสอบวัดความถนัดนั้น ส่วนใหญ่แล้วจะเหมาะสมกับผู้สอบที่มีความสามารถปานกลาง ดังนั้น ความถูกต้องแม่นยำของการวัดผู้สอบที่มีความสามารถสูงและผู้สอบที่มีความสามารถต่ำจึงลดลง

2.2.5 ทฤษฎีการทดสอบแบบเดิมใช้ค่าความแปรปรวนของความคลาดเคลื่อนในการวัดเหมือนกันกับผู้สอบทุกคน ตามความจริงแล้วผู้สอบที่มีความสามารถสูงและต่ำจะมีค่าความแปรปรวนดังกล่าวต่างจากผู้สอบที่มีความสามารถปานกลาง

นอกจากที่กล่าวมาแล้ว ทฤษฎีการทดสอบแบบเดี่ยวยังหาคำตอบของปัญหาเกี่ยวกับการทดสอบไม่ชัดเจนอีกหลายประการ เช่น แบบแผนของการทดสอบ (design of test) ความลำเอียงของข้อสอบ (item bias) และการเทียบคะแนน (equating) เป็นต้น ทฤษฎีการตอบสนองข้อสอบมีชื่อเรียกหลายชื่อ เช่น ทฤษฎีคุณลักษณะแฝง (latent trait theory) ทฤษฎีโค้งคุณลักษณะข้อสอบ (item characteristic curve theory) แต่ชื่อที่ค่อนข้างได้รับความนิยม คือ ทฤษฎีการตอบสนองข้อสอบ (item response theory , IRT)

2.3 ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบ มีข้อตกลงเบื้องต้น ดังนี้

2.3.1 ข้อตกลงเกี่ยวกับมิติ (dimensionality) ทฤษฎีการตอบสนองข้อสอบมีความเชื่อว่า ผู้สอบนั้นมีความสามารถ k อย่าง ซึ่งความสามารถแต่ละอย่างก็จะตอบสนองต่อสิ่งเร้าต่างๆกันไป แบบทดสอบที่จะนำมาวิเคราะห์ด้วยวิธีการตามทฤษฎีการตอบสนองข้อสอบนั้นมีข้อตกลงว่าต้องมีความเป็นมิติเดียว (unidimension) นั่นคือ เป็นแบบทดสอบที่ประกอบด้วยข้อสอบหลายๆข้อสำหรับวัดความสามารถด้านใดด้านหนึ่งของผู้สอบ (สำเร็จ บุญเรืองรัตน์. 2529 : 43) สิ่งที่ยอมรับใช้เป็นดัชนีระบุความเป็นมิติเดียว ก็คือ องค์ประกอบที่เด่นชัดหรือมีความสำคัญต่อการอธิบายผลการตอบแบบทดสอบ สำหรับการกำหนดข้อตกลงเบื้องต้นเกี่ยวกับความเป็นมิติเดียวนี้ก็เพื่อประโยชน์ในการตีความหมายของคะแนนที่ได้จากการสอบนั่นเอง (Lord. 1980 : 19)

ลุมสเดน (Lumsden. 1961,1976) ได้เสนอแนะว่า การที่จะได้ข้อสอบที่วัดมิติเดียวนั้นต้องอาศัยวิธีการคัดเลือกข้อสอบและวิธีการตรวจสอบที่ดี ซึ่งวิธีการตรวจสอบความเป็นมิติเดียวนั้นมีหลายวิธีด้วยกัน แต่วิธีที่ได้รับความนิยมมาก ได้แก่วิธีวิเคราะห์องค์ประกอบ (factor analysis) โดยใช้อัตราส่วนของความแปรปรวนขององค์ประกอบแรกกับความแปรปรวนขององค์ประกอบที่สองเป็นดัชนีความเป็นมิติเดียว (unidimension) แต่ปัญหาที่พบในการวิเคราะห์องค์ประกอบ ก็คือ การเลือกใช้ค่าสหสัมพันธ์ ระหว่างค่าสหสัมพันธ์ฟี (phi correlation) และค่าสหสัมพันธ์เตตราคลอริก (tetrachoric correlation) ซึ่งการวิเคราะห์โดยใช้สหสัมพันธ์ฟีนั้นทำให้ได้องค์ประกอบหลายองค์ประกอบ และอาจทำให้ได้องค์ประกอบที่ไม่ใช่องค์ประกอบที่แท้จริง ซึ่งแมคโดนัลและออลาวัต (McDonald and Ahlawat 1974) ชี้ให้เห็นว่า ที่เกิดผลเช่นนี้ก็เพราะข้อมูลมีความสัมพันธ์กันไม่เป็นเส้นตรง (nonlinearity) ส่วนการใช้ค่าสหสัมพันธ์เตตราคลอริกในการวิเคราะห์นั้นค่อนข้างยุ่งยากใน

การคำนวณ ไม่เหมาะสมที่จะนำมาใช้เมื่อความสามารถของผู้สอบไม่แจ่มแจ้งเป็นปกติ ฟังก์ชันของการตอบสนองข้อสอบไม่เป็นแบบปกติสะสมและเมื่อมีการเดา (Lord, 1980 : 20) แต่สิ่งที่ดีของการใช้ค่าสหสัมพันธ์เตตราคลอริก ก็คือ ผลของการวิเคราะห์จะต้องประกอบที่เด่นชัดเพียงองค์ประกอบเดียว ซึ่งสิ่งนี้ก็เหมือนไขที่พอเพียงในการตรวจสอบความเป็นมิติเดียวของกลุ่มข้อสอบที่ถูกวิเคราะห์ (Lord and Novick, 1968) อย่างไรก็ตามบางครั้งการประเมินความเป็นมิติเดียวด้วยวิธีวิเคราะห์องค์ประกอบก็ได้ผลไม่เป็นที่น่าพอใจ เพราะการวิเคราะห์องค์ประกอบเป็นวิธีการที่ดำเนินการกับข้อมูลที่สัมพันธ์กันเชิงเส้นตรง ดังนั้นหากข้อมูลไม่สัมพันธ์กันเชิงเส้นตรง ผลของการวิเคราะห์ก็อาจจะไม่ได้มิติเดียว ถึงแม้ว่ากลุ่มข้อสอบจะมีความเป็นมิติเดียวก็ตามและส่วนใหญ่แล้วข้อมูลที่เหมาะ (fit) กับโมเดลของทฤษฎีการตอบสนองข้อสอบก็มักมีลักษณะไม่สัมพันธ์กันเป็นเส้นตรง แบบทดสอบที่น่าจะมีความเป็นมิติเดียว เช่น แบบทดสอบการสะกดคำ คำศัพท์ การอ่านเอาเรื่อง เหตุผลเชิงคณิตศาสตร์ อุปมาอุปไมย อนุกรมตัวเลขและมิติสัมพันธ์ สำหรับแบบทดสอบที่ไม่เป็นมิติเดียว เช่น แบบทดสอบวัดผลสัมฤทธิ์วิชาเคมี ซึ่งต้องอาศัยความสามารถหลายอย่าง ทั้งที่เกี่ยวข้องและไม่เกี่ยวข้องกับคณิตศาสตร์ในการตอบข้อสอบ (Lord, 1980 : 20)

2.3.2 ข้อตกลงเกี่ยวกับความเป็นอิสระของข้อสอบ (local independence)

ข้อตกลงข้อนี้ กล่าวว่า โอกาสในการตอบข้อสอบข้อใดข้อหนึ่งได้ถูกต้องนั้นจะไม่เกี่ยวข้องกับการตอบข้อสอบข้ออื่นๆ หากเขียนเป็นสมการทางคณิตศาสตร์จะได้ ดังนี้

$$\text{Prob. } \{U_1 = u_1, U_2 = u_2, \dots, U_n = u_n\} = \prod_{g=1}^n P_g^{u_g} Q_g^{1-u_g}$$

เมื่อ	U_g	แทน ผลการตอบข้อสอบข้อ g
	P_g	แทน ความน่าจะเป็นที่จะตอบข้อสอบข้อ g ถูกต้อง
	Q_g	แทน ความน่าจะเป็นที่จะตอบข้อสอบข้อ g ผิด

จากสมการนี้ กล่าวได้ว่า ความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องทุกข้อเท่ากับผลคูณของความน่าจะเป็นในการตอบข้อสอบในแต่ละข้อได้ถูกต้องและไม่ถูกต้อง การที่เราจะตรวจสอบข้อสอบแต่ละข้อว่าเป็นไปตามข้อตกลงเบื้องต้นข้อนี้หรือไม่นั้นก็สามารถตรวจสอบได้จากแบบแผนของการตอบข้อสอบของผู้สอบที่มีความสามารถเท่าๆกันว่า ข้อสอบแต่ละคู่ เช่น ข้อ 1 กับข้อ 2 ข้อ 2 กับข้อ 3 มีลักษณะของการตอบแตกต่างกันหรือไม่ ซึ่งสามารถทดสอบความเป็นอิสระของการตอบข้อสอบเหล่านี้ได้โดยใช้สถิติทดสอบไคสแควร์ (chi-square) (สำเร็จ บุญเรืองรัตน์, 2529 : 46)

สิ่งที่น่าสนใจเป็นอย่างยิ่ง ก็คือ ความเป็นอิสระของข้อสอบนี้จะได้มาโดยอัตโนมัติ ถ้าข้อสอบมีความเป็นมิติเดียว (Lord. 1980 : 19)

2.4 โมเดลตามทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบเป็นทฤษฎีที่อธิบายความสัมพันธ์ระหว่างปริมาณของการกระทำหรือคะแนนของผู้สอบกับปริมาณความสามารถของผู้สอบ ซึ่งแสดงได้ดังสมการ

$$P = f(\theta)$$

- เมื่อ P แทน ผลการสอบ (performance)
- θ แทน ความสามารถหรือคุณลักษณะ (ability หรือ trait)
- f แทน ฟังก์ชัน (function)

เนื่องจากข้อมูลที่ได้รับจากผู้สอบนั้นมีหลายลักษณะ ดังนั้นจึงได้มีการพัฒนาโมเดลที่แสดงความความสัมพันธ์ในลักษณะต่างๆกัน เพื่อให้แต่ละโมเดลมีความเหมาะสมกับข้อมูลที่ได้ โมเดลของทฤษฎีการตอบสนองข้อสอบสามารถจำแนกออกได้เป็น 3 กลุ่มใหญ่ๆตามลักษณะของข้อมูลที่ได้จากการสอบ 3 ลักษณะ คือ คำตอบที่เป็นข้อมูลแบบแบ่งสอง (dichotomous) ข้อมูลแบบพหุ (multichotomous) และข้อมูลแบบต่อเนื่อง (contineous) โมเดลที่มีความเหมาะสมกับลักษณะข้อมูลดังกล่าว มีดังนี้ (Hambleton and Swaminathan. 1985 : 33-52)

ตาราง 1 ลักษณะของข้อมูล โมเดลที่ใช้วิเคราะห์และผู้นำเสนอโมเดล

ลักษณะของข้อมูล	โมเดลที่ใช้วิเคราะห์	ผู้นำเสนอโมเดล
Dichotomous	Latent Linear	Lazarsfeld & Henry(1968)
	Latent Distance	
	Perfect scale	Guttman(1944)
	One,Two,Three Parameter	Lord(1952)
	Normal Ogive	
	One,Two,Three Parameter	Birnbbaum(1957,1958a,1958b 1968),Lord & Novick (1968)
	Logistic	
		Wright & Stone (1979)
	Four Parameter Logistic	McDonald (1967)
		Barton & Lord (1981)

ตาราง 1 (ต่อ)

ลักษณะของข้อมูล	โมเดลที่ใช้วิเคราะห์	ผู้นำเสนอโมเดล
Multicategory	Nominal Response	Bock (1972)
	Graded Response	Samejima (1969)
	Partial Credit Model	Master (1982)
Contineous	Contineous response	Samejima (1972)

ในแต่ละโมเดลก็มีฟังก์ชันทางคณิตศาสตร์แตกต่างกันออกไป ซึ่งสามารถแสดงได้ตามลำดับของการคิดค้นโมเดล ดังนี้

2.4.1 โมเดลนอร์มอลโอจีฟแบบ 2 พารามิเตอร์ (two parameter normal ogive model) ลอร์ด (Lord, 1952,1953a) เป็นผู้เสนอโมเดลนี้ ได้ังลักษณะข้อสอบจะได้จากสมการ

$$P(\theta) = \int_{-\alpha}^{a(\theta-b)} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$

- เมื่อ $P(\theta)$ แทน ความน่าจะเป็นที่ผู้สอบจะตอบข้อที่ i ได้ถูกต้อง ด้วยความสามารถ θ
- b แทน ค่าความยากของข้อสอบข้อที่ i
- a แทน ค่าอำนาจจำแนกของข้อสอบข้อที่ i
- Z แทน ความเบี่ยงเบนปกติ จากการแจกแจงที่ค่าเฉลี่ย b และความเบี่ยงเบนมาตรฐาน $1/a$

2.4.2 โมเดลโลจิสติกแบบ 2 พารามิเตอร์ (two parameter logistic model) เบอ์นบอม (Birnbbaum)เป็นผู้เสนอโมเดลนี้ ซึ่งได้ังลักษณะข้อสอบจะได้จากฟังก์ชันการแจกแจงดังนี้

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

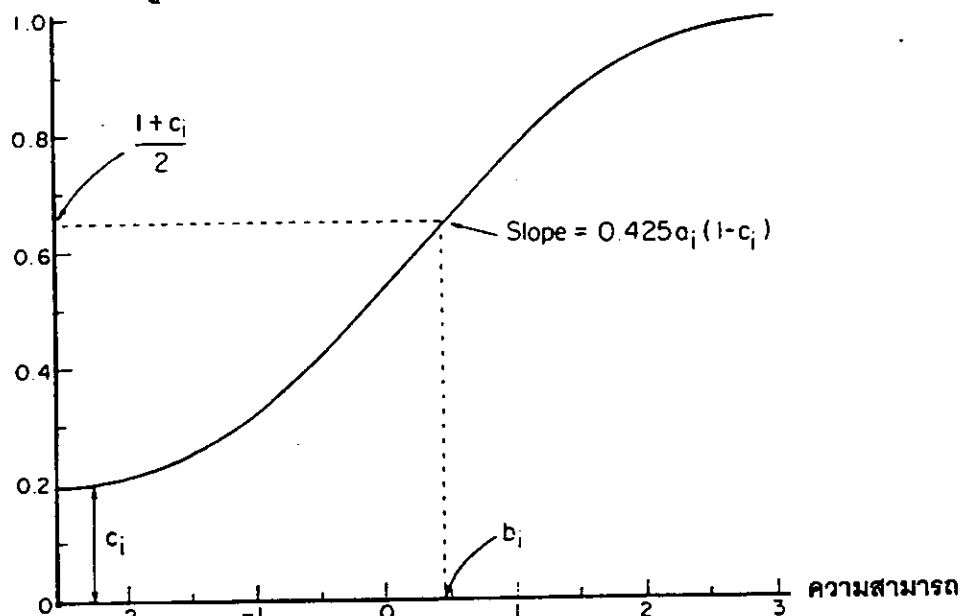
ค่าของ $P(\theta)$, a , b และ θ มีความหมายเหมือนเดิม ส่วน D เป็นค่าที่ใช้ในการปรับแก้ (scaling factor) ซึ่งถ้ากำหนดให้ $D=1.7$ แล้ว ค่าของ $P(\theta)$ จากโมเดลนอร์มอลโอโจईและโมเดลโลจิสติกแบบ 2 พารามิเตอร์จะต่างก็น้อยกว่า .01 ในทุกๆค่าของ θ ที่ประมาณได้ (Hambleton and Swaminathan, 1985 : 37 ; citing Haley, 1952) และโมเดลโลจิสติกนั้นนำไปใช้ได้สะดวกกว่า โมเดลนอร์มอลโอโจईในการประมาณค่า $P(\theta)$

2.4.3 โมเดลโลจิสติกแบบ 3 พารามิเตอร์ (three parameter logistic model)

โมเดลนี้พัฒนามาจากโมเดลโลจิสติกแบบ 2 พารามิเตอร์ โดยเพิ่มค่าพารามิเตอร์การเดา (c) เข้าไปในฟังก์ชันอีก 1 ค่า ทำให้ได้ฟังก์ชันของการแจกแจง ดังนี้

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

พารามิเตอร์ c เป็นค่าความน่าจะเป็นของผู้ สอบที่มีความสามารถต่ำมากแต่สามารถตอบข้อสอบได้ถูกต้อง ค่าความยากง่ายของข้อสอบ (b) จะอยู่บนสเกลความสามารถ (θ) ตรงจุดที่ค่าความชันของโค้งมีค่าสูงสุด ซึ่งค่าความชันของโค้งที่สูงสุดมีค่า เท่ากับ $0.425a(1-c)$ โค้งลักษณะข้อสอบที่มีค่าอำนาจจำแนกสูงจะมีความชันมากกว่าโค้งลักษณะข้อสอบที่มีค่าอำนาจจำแนกต่ำ ความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องของผู้สอบที่มีความสามารถเท่ากับค่าความยากง่ายของข้อสอบจะเท่ากับ $(1 + c)/2$ ซึ่งโค้งแสดงลักษณะของข้อสอบตามโมเดลนี้มีลักษณะดังภาพประกอบ 15 โอกาสตอบถูก

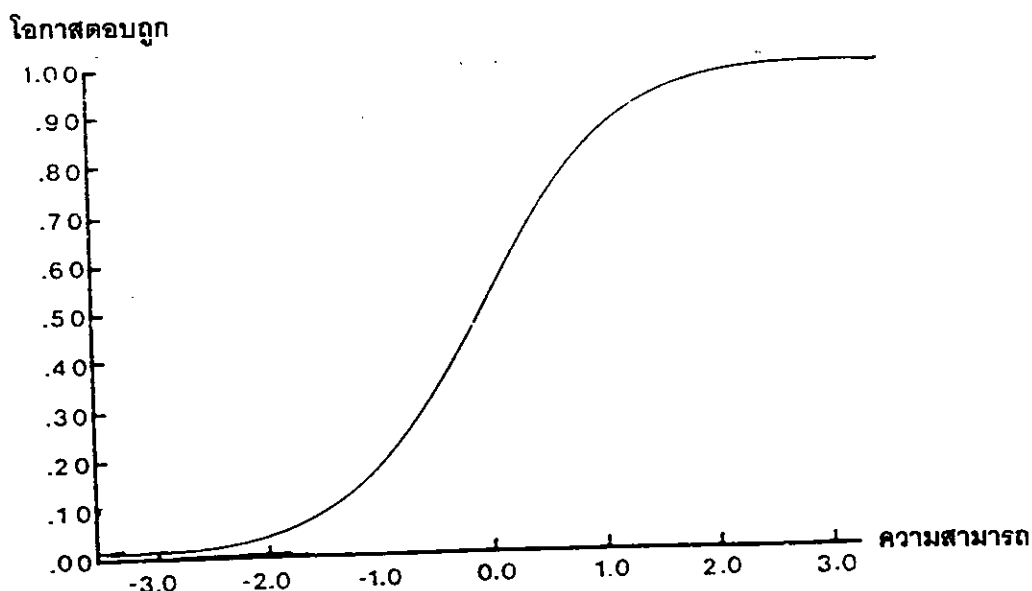


ภาพประกอบ 15 โค้งแสดงลักษณะของข้อสอบตามโมเดลโลจิสติกแบบ 3 พารามิเตอร์

2.4.4 โมเดลโลจิสติกแบบ 1 พารามิเตอร์หรือราสช์โมเดล (one parameter logistic model , rasch model) ราสช์ (Rasch) ซึ่งเป็นนักคณิตศาสตร์ชาวเดนมาร์กเป็นผู้เสนอโมเดลนี้ ซึ่งเป็นโมเดลที่เป็นกรณีเฉพาะของโมเดลโลจิสติก 3 พารามิเตอร์ของเบอร์นบอม (Birnbaum) ที่มีข้อตกลงว่า ข้อสอบทุกข้อมีค่าอำนาจจำแนกเท่ากันและมีค่าการเดาน้อยมากๆ ทำให้ได้ฟังก์ชันของการแจกแจง ดังนี้

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

เมื่อ a แทน ค่าอำนาจจำแนกของข้อสอบ ซึ่งกำหนดให้มีค่าเท่ากันทุกข้อ ซึ่งมีโค้งแสดงลักษณะของข้อสอบ ภาพประกอบ 16



ภาพประกอบ 16 โค้งแสดงลักษณะของข้อสอบตามโมเดลโลจิสติกแบบ 1 พารามิเตอร์

2.4.5 โมเดลโลจิสติกแบบ 4 พารามิเตอร์ (four parameter logistic parameter) แมคโดนัลด์ (McDonald, 1967) บาร์ตันและลอร์ด (Barton and Lord, 1981) เป็นผู้นำเสนอโมเดลนี้ โดยมีความเชื่อว่าผู้สอบที่มีความสามารถสูงๆนั้นบางครั้งอาจตอบข้อสอบผิดได้ด้วยความสะเพร่า หรือการเฉลยข้อสอบที่ผิดของผู้สร้างข้อสอบ ทำให้ได้ฟังก์ชันของการแจกแจง ดังนี้

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

เมื่อ γ เป็นค่าพารามิเตอร์ความสะเพร่า ซึ่งมีค่าต่ำกว่า 1 เล็กน้อย โมเดลนี้เป็นเพียงโมเดลที่ศึกษาตามทฤษฎีเท่านั้น ส่วนการปฏิบัติจริงนั้นยังไม่ค่อยได้นำมาใช้

2.4.6 โมเดลนอร์มอลโอไอจีแบบ 1 3 และ 4 พารามิเตอร์ (one three and four parameter normal ogive model) ลอร์ด (Lord, 1952) เป็นผู้นำเสนอโมเดลนี้ ซึ่งมีฟังก์ชันของการแจกแจง ดังนี้

$$P(\theta) = \int_{-\alpha}^{\theta - b} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz$$

$$P(\theta) = c + (1 - c) \int_{-\alpha}^{a(\theta - b)} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz$$

$$P(\theta) = c + (\gamma - c) \int_{-\alpha}^{a(\theta - b)} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz$$

เมื่อเปรียบเทียบกับโมเดลโลจิสติกแล้ว พบว่า สมการจะมีลักษณะคล้ายๆกัน แต่สมการของโมเดลโลจิสติกจะง่ายกว่า ซึ่งในทางปฏิบัติแล้วไม่ค่อยได้ใช้โมเดลนอร์มอลโอไอจีเท่าใดนัก เนื่องจากมีข้อจำกัดในเรื่องความยุ่งยากในการประมาณค่าและจะใช้โมเดลโลจิสติกแทน สำหรับโมเดลนอร์มอลโอไอจีนั้นเป็นการศึกษาตามทฤษฎีเท่านั้น

2.4.7 โมเดลการตอบสนองข้อสอบแบบนามบัญญัติ (nominal response model)

บ็อก (Bock, 1972) และ เซมีจิมา (Samejima, 1972) เป็นผู้เสนอโมเดลนี้ขึ้นมา เนื่องจากเห็นว่าโมเดลโลจิสติกและโมเดลนอร์มอลโอไอจีนั้นมุ่งตอบสนองข้อมูลที่มีลักษณะแบบแบ่งสอง (dichotomous) สำหรับโมเดลนี้เหมาะสำหรับข้อมูลที่ได้จากการทดสอบที่ให้คะแนนแบบพหุ (multidichotomous) มีฟังก์ชันทางคณิตศาสตร์หลายฟังก์ชันที่แสดงโค้งลักษณะตัวเลือก ความน่าจะเป็นที่ผู้สอบที่มีความสามารถ θ จะเลือกตัวเลือก k (จาก m ตัวเลือกในแต่ละข้อ) สำหรับข้อที่ i มีฟังก์ชันของการแจกแจง ดังนี้

$$P_k(\theta) = \frac{e^{b_k + a_k \theta}}{\sum_{h=1}^m e^{b_h + a_h \theta}}$$

เมื่อ b และ a เป็นค่าพารามิเตอร์ของข้อสอบที่สัมพันธ์กับตัวเลือก k ถ้า $m=2$ ข้อสอบก็จะถูกให้คะแนนแบบแบ่งสอง การประมาณค่าด้วยโมเดลโลจิสติกกับโมเดลนี้ก็จะมีความเท่ากัน

2.4.8 โมเดลการตอบสนองข้อสอบแบบให้ระดับคะแนน (grade response model)

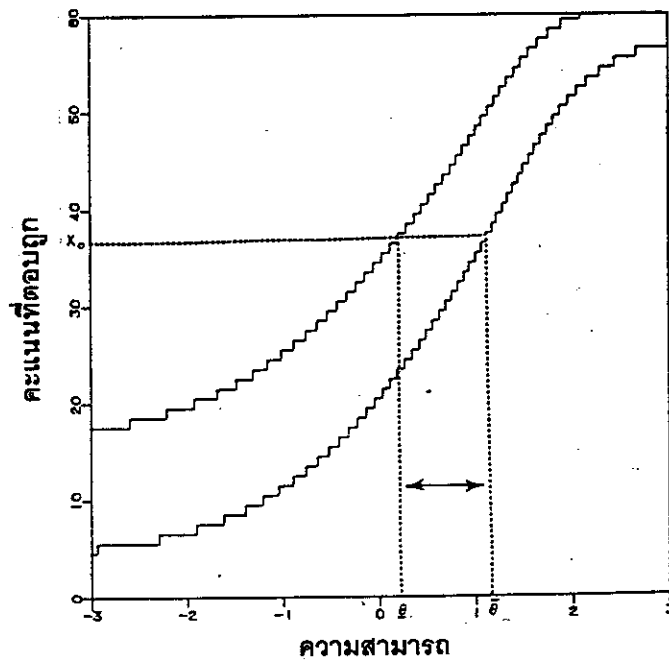
เซมีจิมา (Samejima) เป็นผู้นำเสนอโมเดลนี้ ซึ่งคะแนนของการทดสอบนั้นจะให้ตามระดับของความถูกต้องในการตอบข้อสอบแต่ละข้อ เช่น ตอบผิดทั้งหมด ตอบผิดบางส่วนหรือตอบได้ถูกต้องทั้งหมด ซึ่งในการทดสอบลักษณะนี้ ข้อสอบ 1 ข้อสามารถจำแนกคำตอบได้เท่ากับ $m+1$ ตัวเลือก ถ้าคะแนนเต็มของข้อสอบนี้ คือ m

2.4.9 โมเดลการตอบสนองสเกลการวัดแบบต่อเนื่อง (continuous scale) โมเดลนี้เป็นกรณีเฉพาะของโมเดลการตอบสนองข้อสอบแบบให้ระดับคะแนน ต่างกันตรงที่ผู้สอบจะตอบลงบนสเกลที่ต่อเนื่อง โมเดลนี้มีประโยชน์อย่างมากต่อนักจิตวิทยาสังคมและผู้ที่จะวัดเจตคติ

2.5 ฟังก์ชันสารสนเทศ (information function)

ทฤษฎีการตอบสนองข้อสอบได้กล่าวถึงฟังก์ชันสารสนเทศไว้ โดยแบ่งออกเป็น ฟังก์ชันสารสนเทศของคะแนน (score information function) ฟังก์ชันสารสนเทศของข้อสอบ (item information function) และฟังก์ชันสารสนเทศของแบบทดสอบ (test information function) ซึ่งแต่ละเรื่องมีรายละเอียด ดังนี้

2.5.1 ฟังก์ชันสารสนเทศของคะแนน (score information function) เมื่อเรานำคะแนนผลการสอบของผู้สอบแต่ละคนที่มีความสามารถต่าง ๆ กันมาแจกแจงความถี่และสร้างเป็นแผนภูมิ ก็จะพบว่า คะแนนในแต่ละระดับความสามารถ ส่วนใหญ่จะตกอยู่ในช่วง 95% ของการแจกแจงในแต่ละระดับความสามารถ ดังภาพประกอบ 17



ภาพประกอบ 17 ช่วงความเชื่อมั่น (confidence interval ; $\underline{\theta}, \bar{\theta}$) ของความสามารถที่ได้จากการประมาณค่า ($\hat{\theta}$)

เนื่องจากคะแนนนั้นเป็นจำนวนนับที่ไม่ต่อเนื่อง จึงทำให้ภาพที่เป็นเส้นราบขึ้นบันได จากภาพ กล่าวได้ว่า ผู้สอบที่ได้คะแนน x_0 จะมีความสามารถอยู่ในช่วง $(\underline{\theta}, \bar{\theta})$ ด้วยความเชื่อมั่น 95% ถ้าให้ $z = x/n$ ดังนั้น การถดถอยของ z ลงบน θ จะได้

$$\mu_{z|\theta} = \frac{1}{n} \sum_{i=1}^n p_i \theta = \xi$$

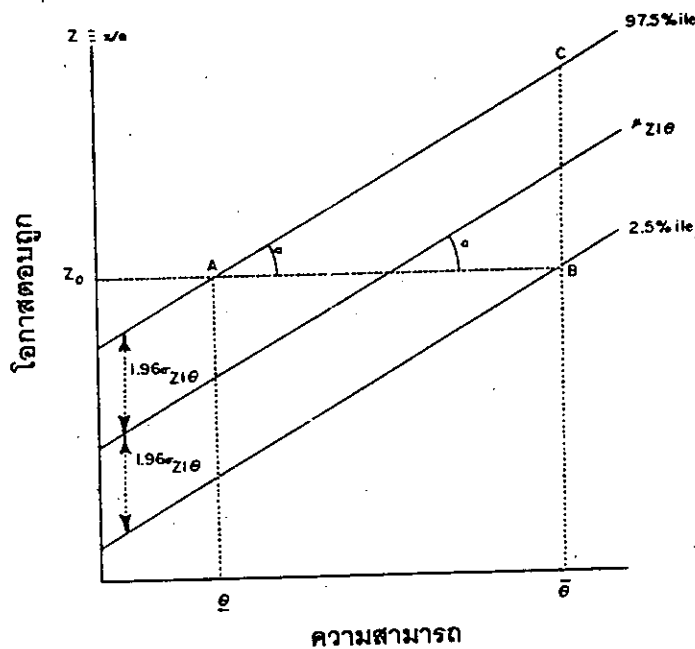
ถึงแม้ว่า n จะมากขึ้นแต่ก็ไม่ทำให้การถดถอยนี้เปลี่ยนไป ความแปรปรวนของ z สำหรับ θ ที่คงที่ ก็คือ

$$\begin{aligned} \sigma_{z|\theta}^2 &= \frac{1}{n} \sum_{i=1}^n p_i q_i \\ &= \left(\frac{1}{n} \right) \bar{p} \bar{q} - \sigma_{p|\theta}^2 \end{aligned}$$

ซึ่งเมื่อ n มีค่ามากขึ้น σ^2 ก็จะมีค่าเข้าใกล้ 0

เมื่อนำค่า z มาเขียนเป็นแผนภูมิของการแจกแจง ลักษณะของโค้งการแจกแจงก็จะมีลักษณะเหมือนกับ ภาพประกอบ 18 เมื่อลากเส้นตัดดอของ z ลงบน θ เส้นตัดดอนี้จะอยู่ระหว่างกลางฟังก์ชันแบบขั้นบันได 2 ฟังก์ชัน ซึ่งอยู่ที่ตำแหน่งเปอร์เซ็นต์ไทล์ที่ 2.5 และ 97.5 ของการแจกแจง Z เมื่อกำหนด θ ให้ และเมื่อ n มีค่าสูงขึ้นการตัดดอก็จะไม่มีการเปลี่ยนแปลง แต่ค่าความแปรปรวนจะลดลงเข้าหา 0 เมื่อฟังก์ชันแบบขั้นบันไดคู่เข้าหากัน จนกระทั่งการแจกแจงของ z เป็นโค้งปกติ ตำแหน่งเปอร์เซ็นต์ไทล์ที่ 2.5 และ 97.5 จะอยู่ห่างจากเส้นตัดดอประมาณ

$$1.96 \sigma_{z|\theta} \quad \text{หรือ} \quad \frac{1.96}{n} \sqrt{\sum_{i=1}^n P_i Q_i}$$



ภาพประกอบ 18 ช่วงแห่งความเชื่อมั่นแอสิมโทติก 95% ($\underline{\theta}, \bar{\theta}$) สำหรับความสามารถ θ

$$\tan \alpha = \frac{\overline{CB}}{\overline{AB}} = \frac{2 (1.96 \sigma_{z|\theta})}{\overline{AB}}$$

เมื่อ $\overline{AB}$ แทน ช่วงของความเชื่อมั่นสำหรับความสามารถ θ
(asymptotic confidence interval ; $\underline{\theta}, \bar{\theta}$)

$\overline{CB}$ แทน ค่าเบี่ยงเบนมาตรฐานของ Z
ดังนั้น

$$\frac{1}{(\overline{AB})^2} = \frac{\frac{(d \mu_{Z|\theta})^2}{d \theta}}{(3.92)^2 \text{Var}(Z|\theta)}$$

เมื่อ $\frac{1}{(\overline{AB})^2}$ แทน ฟังก์ชันสารสนเทศของคะแนน ใช้สัญลักษณ์ $I(\theta, y)$

ซึ่งเบอร์นบอม (ผจงจิต อินทสุวรรณ. 2528 : 100 ; อ้างอิงมาจาก Birnbaum. 1968) ได้นิยามฟังก์ชันสารสนเทศของคะแนน y โดยดังนี้

$$I(\theta, y) = \frac{\frac{(d \mu_{y|\theta})^2}{d \theta}}{\text{Var}(y|\theta)}$$

ค่าสารสนเทศของคะแนนนี้จะแปรเปลี่ยนไปตามระดับความสามารถ ซึ่งสรุปได้ว่า

- (1) ที่ความสามารถระดับเดียวกัน เส้นถดถอยที่มีความชันมากกว่าจะให้สารสนเทศของคะแนนสูงกว่าเส้นถดถอยที่มีความชันน้อยกว่า
- (2) ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ต่ำจะให้ค่าสารสนเทศของคะแนนสูง

2.5.2 ฟังก์ชันสารสนเทศของแบบทดสอบ (test information function) ค่าความสามารถโดยประมาณที่เป็นไปได้สูงสุด (maximum likelihood, θ) ในสภาวะปกติจะมีการแจกแจงในลักษณะเอซิมโทติก มีค่าเฉลี่ย เท่ากับ 0 และค่าความแปรปรวน เท่ากับ

$$\text{Var}(\hat{\theta} | \theta_0) = \frac{1}{\zeta \left(\left(\frac{d \ln L}{d \theta} \right)_{\theta_0}^2 \right)}$$

$$\begin{aligned}
\frac{1}{\text{Var}(\hat{\theta}|\theta_0)} &= \zeta \left\{ \left[\sum_{i=1}^n (u_i - P_i) P_i' / P_i Q_i \right]^2 \mid \theta_0 \right\} \\
&= \zeta \left\{ \left[\sum_{i=1}^n (u_i - P_i) P_i' / P_i Q_i \right] \left[\sum_{j=1}^n (u_j - P_j) P_j' / P_j Q_j \right] \mid \theta_0 \right\} \\
&= \sum_{i=1}^n \sum_{j=1}^n \frac{P_{i0}' P_{j0}'}{P_{i0} P_{j0} Q_{i0} Q_{j0}} \zeta \{ (u_i - P_i) (u_j - P_j) \mid \theta_0 \} \\
\frac{1}{\text{Var}(\hat{\theta}|\theta_0)} &= \sum_{i=1}^n \frac{p_i'^2}{P_i^2 Q_i^2} \text{Var} (u_{i0} \mid \theta_0) = \sum_{i=1}^n \frac{p_{i0}'^2}{P_{i0}^2 Q_{i0}^2} P_{i0} Q_{i0} \\
\text{Var}(\hat{\theta}|\theta_0) &= \frac{1}{\sum_{i=1}^n \frac{p_i'^2}{P_i Q_i}} \\
I(\theta) \equiv I(\theta, \hat{\theta}) &= \sum_{i=1}^n \frac{p_i'^2}{P_i Q_i}
\end{aligned}$$

เมื่อ L แทน ฟังก์ชันความเป็นไปได้ (likelihood function)

$I(\theta)$ แทน ฟังก์ชันสารสนเทศของแบบทดสอบ

P, Q แทน ความน่าจะเป็นในการตอบถูก และ ตอบผิดตามลำดับ

P_i' แทน ค่าความลาดชันของเส้นโค้ง

ลักษณะของฟังก์ชันสารสนเทศของแบบทดสอบ สรุปได้ดังนี้

- (1) ฟังก์ชันสารสนเทศของแบบทดสอบได้ถูกกำหนดขึ้นสำหรับชุดของข้อสอบที่แต่ละจุดของสเกลความสามารถ
- (2) ค่าของสารสนเทศของแบบทดสอบเป็นผลมาจากคุณภาพและจำนวนของข้อสอบ
- (3) ที่ความสามารถระดับเดียวกัน เส้นถดถอยที่มีความชันมากกว่าจะให้ค่าสารสนเทศของแบบทดสอบสูงกว่าเส้นถดถอยที่มีความชันน้อยกว่า
- (4) ข้อสอบที่มีค่าความแปรปรวนต่ำ จะส่งผลให้ค่าสารสนเทศของแบบทดสอบมีค่าสูง
- (5) ค่าสารสนเทศของแบบทดสอบจะไม่ขึ้นกับการจัดหมู่เฉพาะของข้อสอบ ข้อสอบแต่ละข้อเป็นอิสระจากกัน
- (6) ค่าสารสนเทศของแบบทดสอบมีความสัมพันธ์แบบผกผันกับค่าความ

คลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ

2.5.3 ฟังก์ชันสารสนเทศของข้อสอบ (item information function) ข้อสอบ

แต่ละข้อจะมีฟังก์ชันแสดงสารสนเทศ ได้แก่

$$I(\theta, u_i) = \frac{P_i'^2}{P_i Q_i}$$

หรือ เขียนในรูปอื่นได้อีก เช่น

$$I(\theta, u_i) = D^2 a_i^2 \frac{Q_i}{P_i} \left(\frac{P_i - c_i}{1 - c_i} \right)^2$$

$$I(\theta, u_i) = D^2 a_i^2 \frac{(1 - c_i)}{(c_i + e^{-D L_i})(1 - e^{-D L_i})^2}$$

$$\text{เมื่อ } L = a(\theta - b)$$

ฟังก์ชันสารสนเทศของข้อสอบแต่ละข้อจะรวมกันเป็นฟังก์ชันสารสนเทศของแบบทดสอบ ซึ่งค่าฟังก์ชันสารสนเทศของข้อสอบจะขึ้นอยู่กับค่าความชันของฟังก์ชันการตอบข้อสอบ และค่าความแปรปรวนที่มีเงื่อนไขที่แต่ละระดับของความสามารถ ค่าความชันสูงๆ และค่าความแปรปรวนต่ำๆ จะทำให้ค่าสารสนเทศของข้อสอบมีค่าสูง ซึ่งจะทำให้ค่าของความคลาดเคลื่อนมาตรฐานในการวัดมีค่าต่ำ การแจกแจงของฟังก์ชันสารสนเทศของข้อสอบมีลักษณะเป็นรูปประฆังคว่ำ ค่าสารสนเทศที่สูงสุดจะอยู่ที่จุด b บนสเกลความสามารถสำหรับโมเดลการตอบโลจิสติกแบบ 1 และ 2 พารามิเตอร์ ส่วนโมเดล 3 พารามิเตอร์นั้น ค่าสารสนเทศของข้อสอบข้อที่ i จะสูงที่สุดที่จุด $\theta_{\max}$

$$\theta_{\max} = b_i + \frac{1}{Da_i} \ln \left[\frac{1}{2} + \frac{\sqrt{1 + 8c_i}}{2} \right]$$

สำหรับโมเดล 1 พารามิเตอร์นั้นค่าสูงสุดของค่าสารสนเทศของข้อสอบจะคงที่ สำหรับโมเดล 2 พารามิเตอร์ ค่าสูงสุดของค่าสารสนเทศของข้อสอบจะเป็นสัดส่วนโดยตรงกับกำลังสองของค่าอำนาจจำแนก ถ้าค่าอำนาจจำแนกของข้อสอบสูงก็จะทำให้ค่าสารสนเทศของข้อสอบมีค่ามาก ส่วนโมเดล 3 พารามิเตอร์นั้น ค่าสารสนเทศของข้อสอบสูงสุดจะมีค่าเท่ากับ (Hambleton and Swaminathan, 1985 : 105 ; Lord, 1980a : 152)

$$I(\theta,u)_{\max} = \frac{D^2 a_i^2}{8(1-c_i)^2} [1-20c_i-8c_i^2+(1+8c_i)^{3/2}]$$

ลักษณะของสารสนเทศของข้อสอบ

- (1) ผลรวมค่าสารสนเทศของข้อสอบทุกข้อ คือ ค่าสารสนเทศของแบบทดสอบ
- (2) ค่าฟังก์ชันสารสนเทศของข้อสอบจะขึ้นอยู่กับค่าความชันของฟังก์ชันการตอบสนองข้อสอบและค่าความแปรปรวนที่มีเงื่อนไขในแต่ละระดับของความสามารถ ถ้าค่าความชันมากและค่าความแปรปรวนต่ำจะทำให้ค่าสารสนเทศของข้อสอบมีค่าสูงซึ่งจะทำให้ค่าความคลาดเคลื่อนมาตรฐานในการวัดมีค่าต่ำ

2.6 ประสิทธิภาพสัมพัทธ์ (relative efficiency)

หากต้องการเปรียบเทียบแบบทดสอบ 2 ฉบับที่วัดคุณลักษณะเดียวกันว่าฉบับใดจะมีประสิทธิภาพสูงกว่ากัน ก็สามารถทำได้โดยการใช้ประสิทธิภาพสัมพัทธ์ ซึ่งค่าประสิทธิภาพสัมพัทธ์ของคะแนนจากแบบทดสอบ y กับคะแนนจากแบบทดสอบ x เป็นอัตราส่วนของฟังก์ชันสารสนเทศของคะแนน ดังนี้

$$RE(y,x) = \frac{I(\theta,y)}{I(\theta,x)}$$

คะแนน x และ y อาจเป็นคะแนนจากแบบทดสอบสองฉบับที่ต่างกันแต่วัดความสามารถเดียวกันหรืออาจเป็นคะแนนจากแบบทดสอบฉบับเดียวกัน แต่ให้คะแนนด้วยวิธีที่ต่างกัน ค่าประสิทธิภาพสัมพัทธ์ของคะแนนจะแปรเปลี่ยนไปในแต่ละระดับความสามารถ ซึ่งสามารถเขียนให้อยู่ในรูปของการแจกแจงคะแนนจริงได้ดังนี้ .

$$RE(y,x) = \frac{(\sigma_{x|\theta}^2)^{p^2(\xi)}}{\sigma_{y|\eta(\xi)}^2 \sigma^2 \eta(\xi)}$$

ถ้าใช้คะแนนจากการสอบแทนคะแนนจริง จะได้

$$RE(y,x) = \frac{n_y x (n_x - x) t_x^2}{n_x y (n_y - y) t_y^2}$$

เมื่อ $P^2(\xi)$, $\sigma^2(\eta(\xi))$ แทน การแจกแจงความถี่ของคะแนนจริง ξ และ η
ตามลำดับ

$\sigma_{xi\xi}^2$, $\sigma_{y\eta(\xi)}^2$ แทน ค่าความแปรปรวนของคะแนนจากแบบทดสอบ x และ y
ตามลำดับ

x , y แทน คะแนนที่ตอบถูกจากแบบทดสอบ x และ y ที่ตำแหน่ง
เปอร์เซ็นต์ไทล์เดียวกัน

f_x , f_y แทน ความถี่สัมพัทธ์ระหว่างกลุ่ม x และ y ตามลำดับ

n_x , n_y แทน จำนวนข้อสอบในแบบทดสอบ x และ y ตามลำดับ

2.7 การประมาณค่าความสามารถของผู้สอบ

ทฤษฎีการตอบสนองข้อสอบมีความเชื่อว่าสิ่งที่อยู่เบื้องหลังและส่งผลต่อการทดสอบ ก็คือ ความสามารถของผู้สอบ (ability) ซึ่งอาจเป็นความสามารถเพียงอย่างเดียว หรือ หลายๆ อย่างต่อผลการสอบในครั้งหนึ่งๆ ก็ได้ ความสามารถดังกล่าวถูกกำหนดขึ้นเพื่อใช้ในการระบุคุณลักษณะภายใน (trait หรือ characteristic) ที่ต้องการวัดจากผู้สอบ ความสามารถนั้นเป็นสิ่งที่ถูกพรรณนาว่าเป็นสิ่งที่วัดได้ด้วยชุดของข้อสอบ อาจนิยามความสามารถอย่างกว้างๆ ว่าเป็น ตัวแปรด้านความถนัดหรือผลสัมฤทธิ์ทางการเรียน เช่น ความเข้าใจในการอ่าน ความสามารถทางจำนวน หรือ อาจนิยามในวงแคบ เช่น ความสามารถในการคูณจำนวน หรือ ตัวแปรที่เกี่ยวข้องกับบุคลิกภาพ เช่น สังกัปแห่งตน แรงจูงใจ ความสามารถเหล่านี้มิใช่เป็นสิ่งที่มืออยู่ตายตัว แต่สามารถเปลี่ยนแปลงไปได้ตลอดเวลาโดยเฉพาะเมื่อได้รับการเรียนการสอนเพิ่มขึ้น ในการแปลความหมายของความสามารถนั้น ก็จะใช้ผลที่ได้จากการศึกษาความเที่ยงตรงเชิงโครงสร้าง (Hambleton and Swaminathan, 1985 : 55)

ความสามารถ (ability, θ) และคะแนนจริง (true score, ξ) นั้นเป็นสิ่งเดียวกัน แต่ใช้มาตรวัดต่างกัน มาตรวัดคะแนนจริงขึ้นอยู่ข้อสอบทั้งหลายในแบบทดสอบ ส่วนมาตรวัดความสามารถนั้นจะเป็นอิสระจากข้อสอบทั้งหลายในแบบทดสอบ ดังนั้นค่าของความสามารถจึงมีประโยชน์มากกว่าคะแนนจริงเมื่อต้องการเปรียบเทียบแบบทดสอบต่างฉบับที่วัดความสามารถเดียวกัน ลอร์ดและโนวิก (Lord and Novick, 1980 : 45-46) กล่าวว่า คะแนนจริงเป็นค่าที่คาดหวังของคะแนนที่วัดได้ ผู้สอบทุกคนที่มีความสามารถเท่ากับ θ จะมีคะแนนจริงของจำนวนคำตอบถูกเท่ากัน คือ

$$\xi = \sum_{i=1}^n p_i(\theta)$$

ขั้นตอนในการประมาณค่าความสามารถของผู้สอบนั้นมี ดังนี้

- (1) รวบรวมข้อมูลการตอบของผู้สอบ
- (2) เปรียบเทียบความเหมาะสม (fit) ระหว่างโมเดลการวิเคราะห์ข้อมูลกับข้อมูลที่ได้มา และเลือกใช้โมเดลสำหรับวิเคราะห์ข้อมูลที่เหมาะสม
- (3) ประมาณค่าพารามิเตอร์ต่างๆ
- (4) กำหนดสเกลและแปลงค่าความสามารถให้อยู่ในรูปของสเกลง่าย ๆ

ในการประมาณค่าความสามารถของผู้สอบนั้นมีอยู่หลายวิธีด้วยกัน แต่พอจะจำแนกออกเป็น 2 กลุ่มใหญ่ๆ ได้แก่ วิธีความเป็นไปได้สูงสุด (maximum likelihood) และวิธีของเบส์ (bayesian) แต่ละวิธีมีรายละเอียดดังนี้ (Hambleton and Swaminathan. 1985 : 76-97)

2.7.1 วิธีความเป็นไปได้สูงสุด (maximum likelihood) ความน่าจะเป็นที่ผู้สอบซึ่งมีความสามารถ θ จะตอบข้อสอบได้ถูกต้องหรือผิดนั้น ก็คือ ฟังก์ชันการตอบสนองข้อสอบ (item response function) ซึ่งเขียนในรูปของสมการได้ ดังนี้

$$P(U_i | \theta) = P^U Q^{1-U}$$

เมื่อ $Q = 1 - P$ และ $U = 0$ หรือ 1

ถ้าผู้สอบที่มีความสามารถ θ ตอบข้อสอบ n ข้อ ความน่าจะเป็นร่วม (Joint probability) ของคำตอบ $u_1, u_2, \dots, u_n$ สามารถเขียนให้อยู่ในรูปสมการได้ดังนี้

$$\begin{aligned} P(U_1, U_2, \dots, U_n | \theta) &= P(U_1 | \theta) P(U_2 | \theta), \dots, P(U_n | \theta) \\ &= \prod_{i=1}^n P(U_i | \theta) \\ &= \prod_{i=1}^n P_i^U Q_i^{(1-U)} \end{aligned}$$

เมื่อตัวแปรสุ่ม $U_1, U_2, \dots, U_n$ รับเอาค่าเฉพาะ $u_1, u_2, \dots, u_n$ ซึ่ง u มีค่าได้ 2 ค่า คือ 0 กับ 1 ค่าความน่าจะเป็นร่วมที่เขียนให้อยู่ในรูปของฟังก์ชัน ซึ่งเรียกว่าฟังก์ชันความเป็นไปได้ (likelihood function) ดังนี้

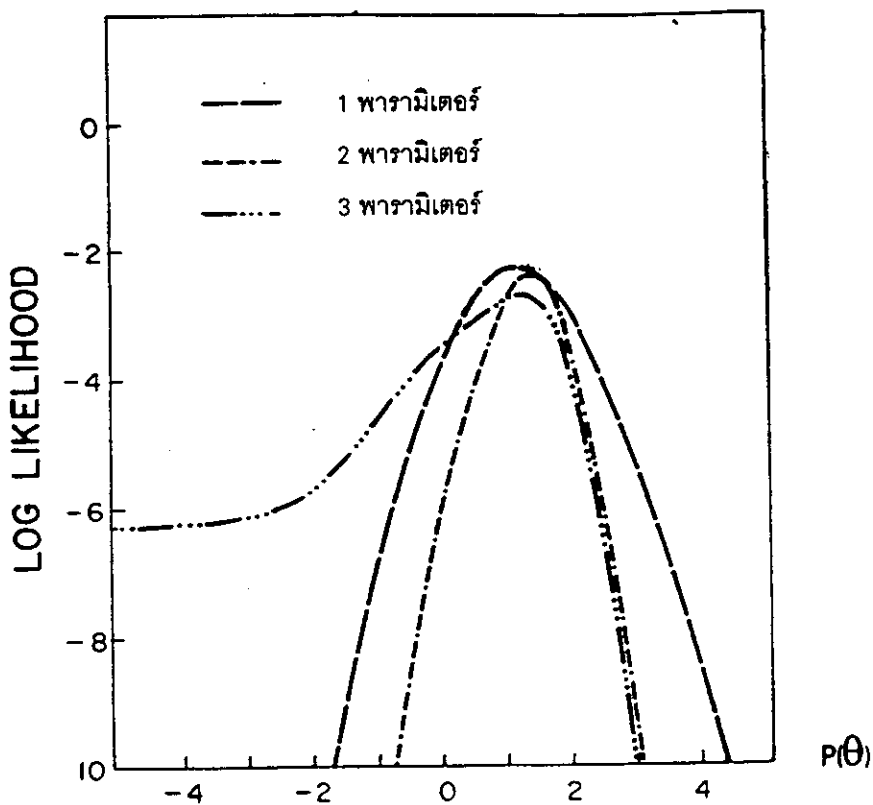
$$P(U_1, U_2, \dots, U_n | \theta) = \prod_{i=1}^n p_i^{u_i} (1-u_i)^{(1-u_i)}$$

ค่าของ θ ที่ทำให้ฟังก์ชันความเป็นไปได้มีค่าสูงสุดจะถูกใช้เป็นค่าความสามารถของผู้สอบโดยประมาณที่หาโดยวิธีความเป็นไปได้สูงสุด (maximum likelihood procedure) หากแปลงค่าฟังก์ชันความเป็นไปได้ให้อยู่ในรูปของลอการิทึมก็จะให้ผลเหมือนเดิม กล่าวคือ ค่าของ θ ที่ทำให้ฟังก์ชันความเป็นไปได้มีค่าสูงสุด ก็จะทำให้ฟังก์ชันลอการิทึมความเป็นไปได้มีค่าสูงสุดเช่นกัน แต่ผลดีก็คือ ผลคูณของค่าความน่าจะเป็นในการตอบถูกแต่ละข้อสามารถนำมารวมกันได้ตามสมบัติของลอการิทึม ดังสมการ

$$\begin{aligned} \ln L(u | \theta) &= \ln (P_1 P_2 P_3 P_4 P_5) \\ &= \ln P_1 + \ln P_2 + \ln P_3 + \ln P_4 + \ln P_5 \end{aligned}$$

แต่สิ่งที่ต่างกัน ก็คือ พหุคูณของ $L(u | \theta)$ คือ 0 ถึง 1 ส่วนพหุคูณของ $\ln L(u | \theta)$ คือ $-\infty$ ถึง 0 การประมาณค่าความสามารถของผู้สอบด้วยวิธีความเป็นไปได้สูงสุด แบ่งออกเป็น 2 วิธีการ คือ ประมาณค่าจากกราฟ และประมาณค่าโดยวิธีการคำนวณ ซึ่งมีรายละเอียด ดังนี้

(1) การประมาณค่าจากกราฟ ทำได้โดยการนำค่า $\ln L(u | \theta)$ มาสร้างเป็นกราฟ ซึ่งลักษณะของกราฟก็ขึ้นอยู่กับโมเดลของการตอบสนองข้อสอบที่เลือกใช้ ค่าของ θ ที่ทำให้ $\ln L(u | \theta)$ มีค่าสูงสุด คือค่าความสามารถของผู้สอบ แสดงได้ดังภาพประกอบ 19



ภาพประกอบ 19 ฟังก์ชันลอกลิเคิลฮูด (log likelihood) สำหรับโมเดลการตอบสนองข้อสอบ 3 แบบ

กราฟนี้ถูกสร้างขึ้นจากค่าของ $\ln L(u | \theta)$ เปรียบเทียบให้เห็นทั้ง 3 โมเดลของการตอบสนองข้อสอบ ค่า θ ที่ทำให้ค่า $\ln L(u | \theta)$ มีค่าสูงสุด คือ 1.18 ดังนั้น ค่าความสามารถของผู้สอบที่ได้จากการประมาณค่าวิธีความเป็นไปได้สูงสุดจากกราฟ คือ 1.18

(2) การประมาณค่าจากการคำนวณ ในการประมาณค่าความสามารถของผู้สอบจากกราฟนั้นไม่ค่อยได้รับความสะดวกและอาจเกิดความคลาดเคลื่อนได้ง่าย ค่าสูงสุดของฟังก์ชันใดๆ ก็คือ ค่าอนุพันธ์ของฟังก์ชันนั้นๆ ที่มีค่าเท่ากับ 0 ดังนั้น ค่าที่สูงสุดของฟังก์ชันความเป็นไปได้ ก็คือ

$$\frac{d}{d\theta} \ln L(u | \theta) = 0$$

เมื่อ

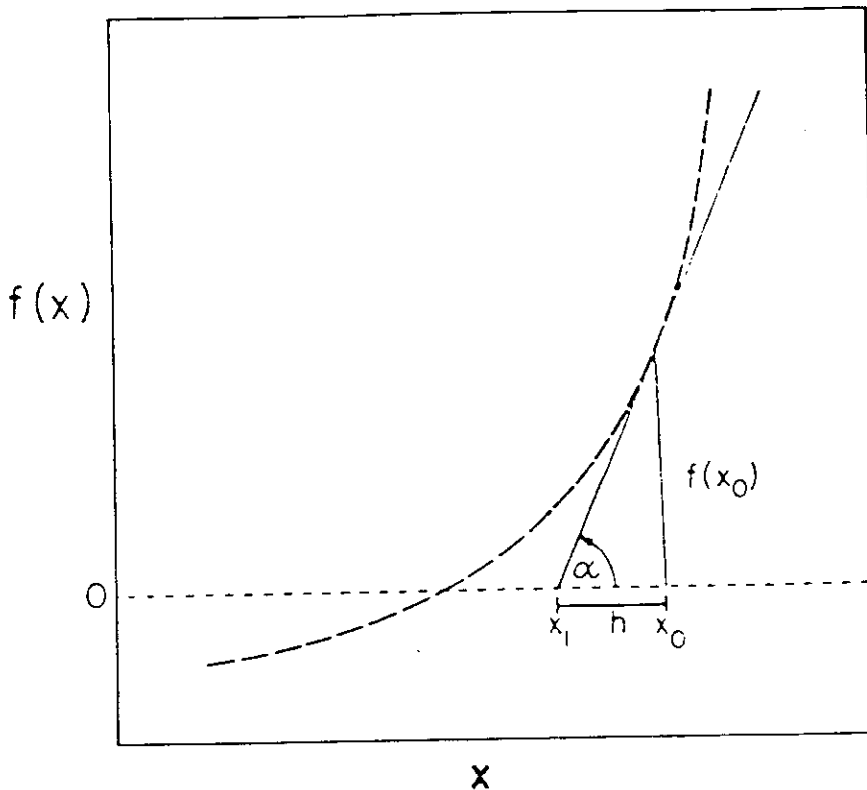
$$\ln L(u | \theta) = \sum_{i=1}^n [u_i \ln P_i + (1-u_i) \ln Q_i]$$

สมการอนุพันธ์นี้ เรียกว่า สมการความเป็นไปได้ (likelihood equation) ซึ่งไม่ใช่สมการเชิงเส้นและไม่สามารถแก้สมการได้ด้วยวิธีง่ายๆ วิธีที่รู้จักกันแพร่หลายในการแก้สมการลักษณะนี้ก็คือ วิธีนิวตัน-ราฟสัน (Newton-Raphson procedure) ซึ่งมีวิธีการ ดังนี้

(2.1) กำหนดให้สมการที่ต้องการหาคำตอบมีค่า เท่ากับ 0

(2.2) ผลลัพธ์โดยประมาณของสมการ คือ x_0 และค่าที่ถูกต้องมากยิ่งขึ้น

คือ x_1 ซึ่งได้จาก $x_1 = x_0 - h$



ภาพประกอบ 20 ที่มาของค่าต่างๆตามวิธีนิวตัน-ราฟสัน (Newton-Raphson)

เมื่อ
$$h = \frac{f(x_0)}{\tan \alpha}$$

$\tan \alpha$ แทน ค่าความชัน (slope) ของฟังก์ชัน $f(x)$ ที่ x_0 ซึ่ง $\tan \alpha = f'(x_0)$
 $f'(x_0)$ แทน อนุพันธ์ของฟังก์ชันที่ x_0

ดังนั้น
$$h = \frac{f(x_0)}{f'(x_0)}$$

แทนค่า h ลงในสมการ จะได้ $x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$

เมื่อได้ค่า x_1 แล้ว ค่าที่ดีกว่า x_1 ก็จะถูกประมาณค่าในทำนองเดียวกันกับที่กล่าวมา ซึ่งกระบวนการนี้จะทำซ้ำๆ กันไป ซึ่งโดยทั่วไปแล้ว ถ้า x_m เป็นผลลัพธ์โดยประมาณของการประมาณค่าครั้งที่ m แล้วผลลัพธ์ที่ดีกว่า ก็คือ x_{m+1} ซึ่งได้จาก

$$x_{(m+1)} = x_{(m+1)} - \frac{f(x_m)}{f'(x_m)}$$

วิธีการนี้จะถูกทำซ้ำ จนกระทั่งค่าของ $(x_{m+1} - x_m)$ มีค่าต่ำกว่าเกณฑ์ที่ได้กำหนดไว้ ซึ่งมักนิยมกำหนดไว้ เท่ากับ .001

เมื่อแทนค่า $f(x)$ ตามวิธีการ จะได้

$$f(x) = \frac{d}{d\theta} \ln L(u | \theta)$$

$$f'(x) = \frac{d^2}{d\theta^2} \ln L(u | \theta)$$

ดังนั้น ถ้า θ_m คือ ค่าความสามารถของผู้สอบในการประมาณค่าครั้งที่ m แล้ว ค่าความสามารถในการประมาณค่าครั้งที่ $m+1$ จะได้จากสมการ

$$\theta_{m+1} = \theta_m - \frac{\left[\frac{d}{d\theta} \ln L(u | \theta) \right]_m}{\left[\frac{d^2}{d\theta^2} \ln L(u | \theta) \right]_m}$$

กระบวนการนี้จะถูกทำซ้ำจนกระทั่งค่าของ θ_m และ θ_{m+1} แตกต่างกันอย่างไม่มีนัยสำคัญ ค่าที่ได้ก็คือค่าความสามารถของผู้สอบที่ได้จากการประมาณค่าด้วยวิธีความเป็นไปได้สูงสุด
สัญลักษณ์ที่ใช้แทน ก็คือ $\hat{\theta}$

ตาราง 2 ค่าอนุพันธ์อันดับหนึ่งและอันดับสองของฟังก์ชันลอกลิสต์

อนุพันธ์	1 พารามิเตอร์	2 พารามิเตอร์	3 พารามิเตอร์
$\frac{\partial \ln L}{\partial \theta}$	$D \sum_{i=1}^n (u_{i\theta} - P_{i\theta})$	$D \sum_{i=1}^n a_i (u_{i\theta} - P_{i\theta})$	$D \sum_{i=1}^n a_i (u_{i\theta} - P_{i\theta}) (P_{i\theta} - c_i) / P_{i\theta} (1 - c_i)$
$\frac{\partial^2 \ln L}{\partial \theta_a^2}$	$-D^2 \sum_{i=1}^n P_{i\theta} (1 - P_{i\theta})$	$-D^2 \sum_{i=1}^n a_i^2 P_{i\theta} (1 - P_{i\theta})$	$-D^2 \sum_{i=1}^n a_i^2 (u_{i\theta} c_i - P_{i\theta}^2) (P_{i\theta} - c_i) Q_{i\theta} / P_{i\theta}^2 (1 - c_i)^2$

ข้อจำกัดในการประมาณค่าความสามารถด้วยวิธีความเป็นไปได้สูงสุด

ในการประมาณค่าความสามารถของผู้สอบด้วยวิธีความเป็นไปได้สูงสุดนั้นได้ใช้กระบวนการทำซ้ำ (iterate) โดยใช้ค่าปรับแก้มาลบค่าความสามารถในแต่ละรอบ จนกระทั่งได้ค่าความสามารถของผู้สอบคงที่จึงจะทำข้อสอบข้อต่อไป ดังสมการ

$$\theta_{m+1} = \theta_m - \frac{\frac{d}{d\theta} \ln L(u | \theta)|_m}{\frac{d^2}{d\theta^2} \ln L(u | \theta)|_m}$$

เพื่อให้ได้ค่าอนุพันธ์ของ $\ln L$ มีค่าสูงสุดจึงกำหนดให้อนุพันธ์ของ $\ln L$ มีค่าเท่ากับ 0 ดังนั้น

$$\frac{\delta}{\delta \theta} \ln L = \sum_{i=1}^n k_i u_i - \sum_{i=1}^n k_i p_i = 0$$

เมื่อ	$k = D$	(1 พารามิเตอร์)
	$k = Da_i$	(2 พารามิเตอร์)
	$k = Da_i \frac{(P_i - c_i)}{P_i(1 - c_i)}$	(3 พารามิเตอร์)

ถ้าตอบผิด $u = 0 \quad \sum k_p = 0$ ในกรณีนี้จะเป็นไปได้ก็ต่อเมื่อ $\theta = -\alpha$

ถ้าตอบถูก $u = 1 \quad \sum k = \sum k_p$ ในกรณีนี้จะเป็นไปได้ก็ต่อเมื่อ $\theta = +\alpha$

ซึ่งในทางปฏิบัติจริงนั้นเป็นไปได้ไม่ได้ จนกว่าผู้สอบจะตอบผิดหรือถูกอย่างน้อย 1 ข้อ ค่า θ ก็จะเปลี่ยนไป ดังนั้นในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์หรือการทดสอบแบบซี เอ ที นั้น ถ้าผู้สอบตอบข้อแรกได้ถูกต้อง ข้อสอบข้อต่อไปจะยากที่สุดในคลังข้อสอบเพื่อให้ผู้สอบมีโอกาสตอบผิดมากที่สุด ในทำนองเดียวกันถ้าผู้สอบตอบข้อสอบข้อแรกผิด ข้อสอบข้อต่อไปจะง่ายที่สุดในคลังข้อสอบ เพื่อให้ผู้สอบมีโอกาสตอบถูกมากที่สุด (Weiss, 1975 : 62 ; citing Urry, 1970)

แนวทางแก้ไขข้อจำกัดดังกล่าว ก็คือ

วิธีที่ 1. ประมาณค่าความสามารถของผู้สอบตั้งแต่เริ่มต้นสอบด้วยวิธีการของเบส์ จนกระทั่งผู้สอบตอบข้อสอบถูกหรือผิดอย่างน้อย 1 ข้อ จึงเปลี่ยนวิธีการประมาณค่าความสามารถเป็นวิธีการความเป็นไปได้สูงสุด (Warm, 1978 : 120 - 121)

วิธีที่ 2 ถ้าผู้สอบตอบข้อสอบข้อแรกผิด จะได้คะแนนเท่ากับ 0 ($u=0$) ก็จะทำให้ $\sum k_p = .001$ ถ้าผู้สอบตอบข้อสอบข้อแรกถูก จะได้คะแนนเท่ากับ 1 ($u = 1$) ก็จะทำให้ $\sum k_p = \sum k - .001$ (Hambleton and swaminathan 1985 : 96)

วิธีที่ 3 ถ้าตอบถูก ก็จะทำให้ $\sum P = \sum c + .001$ ถ้าตอบผิดก็จะให้ $\sum P = Z - .001$ เมื่อ Z แทน จำนวนข้อสอบ

วิธีที่ 4 ให้คะแนนด้วยวิธีฟิชเชอร์ การประมาณค่าความสามารถด้วยวิธีความเป็นไปได้สูงสุดนั้น ต้องมีกระบวนการทำซ้ำเพื่อให้ได้ค่าความสามารถที่ดีกว่าเดิม ซึ่งวิธีการของนิวตัน-ราฟสันนั้นอาจไม่เร็วเท่าที่ควรในการหาค่าความสามารถ ทั้งนี้อาจเป็นเพราะที่บางจุดของการทำซ้ำนั้น เมตริกของอนุพันธ์อันดับสองไม่ได้ถูกนิยามไว้ (indefinite) วิธีการแก้ปัญหานี้ก็คือ นำเมตริกสารสนเทศ (information matrix) ไปแทน เมตริกอนุพันธ์อันดับสองในสูตรการหาค่าความสามารถของผู้สอบตามวิธีการของนิวตัน-ราฟสัน วิธีการใหม่นี้ เรียกว่าวิธีการให้คะแนนของฟิชเชอร์ (Fisher's method) (Hambleton and Swaminathan 1985 :135 ;citing Rao, 1965 : 302)

เมตริกสารสนเทศที่จะใช้แทนเมตริกของอนุพันธ์อันดับสอง ชนิด 3 พารามิเตอร์ ก็คือ

$$D^2 \sum_{i=1}^n \frac{(P_i - c_i) Q_i}{(1 - c_i)^2 P_i} (c_i - P_i)$$

2.7.2 วิธีของเบส์ (bayesian) ถ้าผู้สอบตอบข้อสอบถูกทุกข้อหรือผิดหมดทุกข้อ การประมาณค่าความสามารถด้วยวิธีความเป็นไปได้สูงสุดก็จะไม่เหมาะสม เว้นแต่จะนำเอาผลการตอบของผู้สอบเหล่านั้นออกไปจากกลุ่มผู้สอบก่อน หากทราบค่าสารสนเทศภายในแรก (prior information) การประมาณค่าความสามารถด้วยวิธีของเบส์ก็จะเกิดประสิทธิผลมากกว่า ค่าสารสนเทศภายในแรกที่มีกันิยมใช้กันก็คือ ความสามารถของผู้สอบที่มีการแจกแจงเป็นรูปโค้งปกติ มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 (Hambleton and Swaminathan. 1985 : 92. ; citing Owen. 1975)

การกำหนดรูปร่างของการแจกแจงเป็นรูปโค้งปกติถือว่าค่อนข้างสะดวก เบอร์นบอม (Birnbbaum) เชื่อว่า การแจกแจงภายในแรกของความสามารถเป็นฟังก์ชันโลจิสติก (logistic density function) ซึ่งมีสมการ ดังนี้

$$f(\theta) = \frac{\exp. (\theta)}{[1 + \exp.(\theta)]^2}$$

หัวใจของวิธีการประมาณค่าความสามารถตามวิธีการของเบส์ คือทฤษฎีของเบส์ (Bayes theorem)ซึ่งมีความสัมพันธ์กับ condition probability และ marginal probability

$$P(B| A) = \frac{P (A|B) P(B)}{P(A)}$$

ในการประมาณค่าความสามารถของผู้สอบ A อาจถูกกำหนดให้เป็น θ_a และ B เป็นชุดของคำตอบของข้อสอบ n ข้อ ได้คะแนน u แล้วสมการ $P(B | A)$ เขียนใหม่ได้เป็น

$$P(\theta_a| u) = \frac{P(u|\theta_a) P(\theta_a)}{P(u)}$$

เนื่องจาก θ_a เป็นตัวแปรต่อเนื่อง ดังนั้นจึงถูกแปลความว่าเป็นฟังก์ชันความหนาแน่น (density function) เพื่อเป็นการป้องกันการจำสับสนระหว่าง $P(\theta_a)$ กับ $P(\theta)$ ซึ่งเป็นฟังก์ชันการตอบสนองข้อสอบ (item response function) จึงใช้ $f(\theta_a)$ แทน $P(\theta)$ ดังนั้น จึงเขียนสมการใหม่ได้ดังนี้

$$f(\theta_a | u) = \frac{f(u|\theta_a) f(\theta_a)}{f(u)}$$

เมื่อ $f(u)$ แทน เป็นค่าคงที่

$f(\theta_a | u)$ แทน เป็นความหนาแน่นภายหลัง (posterior density) ของ θ_a

$f(\theta_a)$ แทน เป็นความหนาแน่นก่อน (prior density) ของ θ_a

แต่เนื่องจาก $f(u|\theta_a)$ เป็นฟังก์ชันความเป็นไปได้ (likelihood function) ของคะแนน ดังนั้น จึงสามารถเขียนสมการ $f(u|\theta_a)$ ให้อยู่ในรูปของการแปรผันได้ ดังนี้

$$f(\theta_a | u) \propto L(u|\theta_a) f(\theta_a) \text{ หรือ Posterior} \propto \text{Likelihood} \times \text{Prior}$$

วิธีการประมาณค่าความสามารถของผู้สอบตามวิธีการของเบส์นี้ ได้ถูกนำไปประยุกต์ เป็นวิธีต่างๆอีกหลายวิธี แต่วิธีที่ได้รับความนิยมมาก คือ วิธีของเบส์ที่ปรับใหม่ (bayesian updating) ซึ่งเสนอโดยโอเวน (Owen, 1969 ; 1975 : 353)) ซึ่งมีวิธีการประมาณค่า ดังนี้

ถ้าตอบถูกต้อง

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

$$\sigma_{m+1}^2 = \sigma_m^2 \left[1 - \left[\frac{1 - c}{1 + \frac{1}{a^2 \sigma_m^2}} \right] \left[\frac{\Theta(D)}{A} \right] \left[\frac{(1-c) \Theta(D)}{A} - D \right] \right]$$

ถ้าตอบผิด

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

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

- เมื่อ $\Theta(D)$ แทน ค่าออร์ดิเนต (ordinate) ของโค้งปกติที่จุด D
- $\Phi(D)$ แทน พื้นที่ใต้โค้งปกติ จากค่า D ต่ำสุด จนถึงจุด D ที่มีค่าเท่ากับ $\frac{b - \theta_m}{\sqrt{\frac{1}{a^2} + \sigma_m^2}}$ และ $A = c + (1 - c) \Phi(-D)$
- θ_m แทน ความสามารถของผู้สอบที่ประมาณค่าได้ก่อนตอบข้อสอบข้อที่ m+1 ซึ่งตามปกติแล้ว ถ้าไม่ทราบค่าความสามารถเบื้องต้นของผู้สอบ ก็มักกำหนดให้เท่ากับ 0 ซึ่งจากสูตรเดิมใช้สัญลักษณ์ M_0
- σ_m^2 แทน ความแปรปรวนในการประมาณค่าความสามารถของผู้สอบก่อนตอบข้อที่ m+1 ถ้าไม่ทราบค่าความแปรปรวนดังกล่าวมาก่อน ก็มักกำหนดให้เท่ากับ 1 ซึ่งจากสูตรเดิมใช้สัญลักษณ์ V_0
- θ_{m+1} แทน ค่าความสามารถของผู้สอบโดยประมาณหลังจากที่ตอบข้อที่ m+1 ซึ่งจากสูตรเดิมใช้สัญลักษณ์ $E(\theta | 1)$
- σ_{m+1}^2 แทน ค่าความแปรปรวนในการประมาณค่าความสามารถของผู้สอบ เมื่อตอบข้อสอบข้อที่ m+1 ซึ่งจากสูตรเดิมใช้สัญลักษณ์ Var
- a แทน ค่าอำนาจจำแนกของข้อสอบข้อที่ m+1 ซึ่งจากสูตรเดิมใช้สัญลักษณ์ p
- b แทน ค่าความยากง่ายของข้อสอบข้อที่ m+1 ซึ่งจากสูตรเดิมใช้สัญลักษณ์ d
- c แทน ค่าการเดาของข้อสอบข้อที่ m+1 ซึ่งจากสูตรเดิมใช้สัญลักษณ์ g

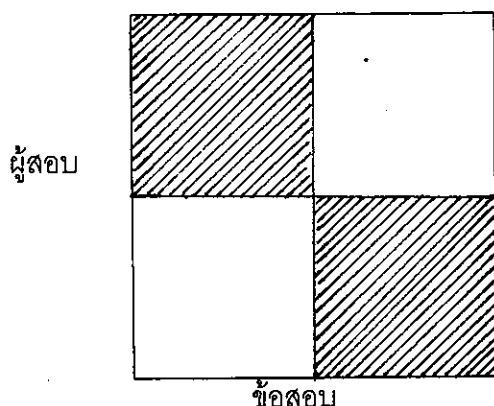
เหตุผลที่ผู้วิจัยไม่ได้ใช้สัญลักษณ์เดิมตามที่โอเวน (Owen) ได้เสนอไว้ในสูตรการหาความสามารถและความแปรปรวน ก็เนื่องจากไม่ต้องการให้เกิดความสับสนในการอ่านรายงานวิจัยฉบับนี้

2.8 การแปลงค่าพารามิเตอร์ให้อยู่บนสเกลเดียวกัน (rescale)

สิ่งที่นับว่าจำเป็นอย่างยิ่งสำหรับการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ก็คือคลังข้อสอบ หากต้องการให้การทดสอบดังกล่าวมีประสิทธิภาพแล้ว คลังข้อสอบต้องประกอบด้วยข้อสอบที่มีคุณภาพสูงที่ระดับต่างๆของความสามารถ ในการสร้างและพัฒนาคลังข้อสอบเพื่อให้ได้ข้อสอบที่คุณภาพกระจายไปทุกระดับความสามารถนั้น ในทางปฏิบัติอาจต้องใช้แบบทดสอบคู่ขนานหลายๆฉบับไปทดลองกับผู้สอบหลายๆกลุ่ม ค่าพารามิเตอร์ของข้อสอบที่ได้จากแบบทดสอบต่างชุดและผู้สอบต่างกลุ่มกันนั้น ไม่สามารถที่จะนำมาเปรียบเทียบกันได้ทันที เพราะสเกลของการวัดจะต่างกัน (Warm. 1978 : 113) แนวทางแก้ไขที่เป็นไปได้ก็คือ การแปลงค่าพารามิเตอร์ของข้อสอบให้อยู่บนสเกลเดียวกัน วิธีการแปลงค่าดังกล่าวประกอบด้วย 2 ขั้นตอน คือ การเลือกใช้แบบแผนของการใช้ข้อสอบหรือผู้สอบร่วมกัน (anchoring design) และวิธีการแปลงคะแนน (transformation method) ซึ่งในแต่ละขั้นตอนดังกล่าวมีรายละเอียดพอสังเขป ดังนี้

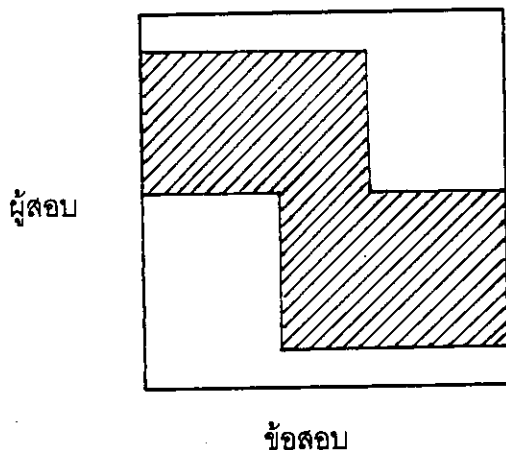
2.8.1 แบบแผนของการใช้ข้อสอบหรือผู้สอบร่วมกัน (anchoring design) ในขั้นนี้เป็นขั้นตอนการวางแผนเพื่อเก็บข้อมูล ซึ่งจะต้องดำเนินการก่อนที่จะมีการทดสอบ แบบแผนดังกล่าวมีดังนี้

(1) ใช้ข้อสอบต่างชุดและผู้สอบต่างกลุ่ม (unanchored) วิธีนี้ใช้แบบทดสอบคู่ขนานกันซึ่งได้จากการสุ่มข้อสอบลงฉบับ แบบทดสอบเหล่านี้สามารถใช้แทนกันได้ แต่ละฉบับจะถูกนำไปทดสอบกับผู้สอบต่างกลุ่มกัน ไม่มีข้อสอบข้อใดที่ถูกนำไปทดสอบกับผู้สอบทั้งสองกลุ่ม และไม่มีผู้สอบคนใดที่ต้องทำข้อสอบหลายๆชุด ซึ่งมีลักษณะดังภาพประกอบ 21



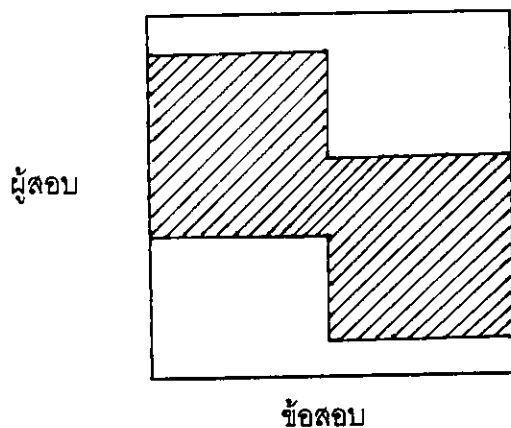
ภาพประกอบ 21 แบบแผนของการใช้ข้อสอบต่างชุดและผู้สอบต่างกลุ่ม

(2) ใช้ข้อสอบร่วมกัน (anchor-items) แบบแผนนี้ถือว่าเป็นแบบแผน ที่ได้รับความนิยมมาก ข้อสอบส่วนหนึ่งจะถูกใช้เป็นข้อสอบร่วม ซึ่งปรากฏอยู่ในแบบทดสอบทั้งสอง ฉบับ ค่าพารามิเตอร์ของข้อสอบร่วมจะเป็นตัวปรับค่าพารามิเตอร์ของข้อสอบจากฉบับต่างๆสู่ สเกลที่ใช้ร่วมกัน (common scale) ซึ่งมีลักษณะดังภาพประกอบ 22



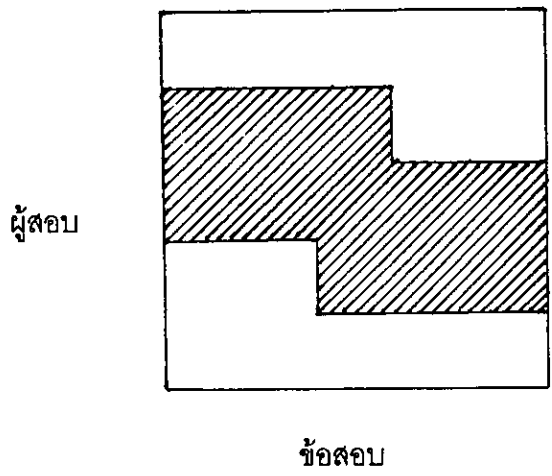
ภาพประกอบ 22 แบบแผนของการใช้ข้อสอบร่วมกัน

(3) ให้ผู้สอบร่วมกัน (anchor-group) แบบแผนนี้มีลักษณะคล้ายๆกับ แบบแผนการใช้ข้อสอบร่วมกัน แต่ต่างกันตรงที่ให้ผู้สอบร่วมกัน กล่าวคือ มีผู้สอบบางส่วนต้อง ทำ แบบทดสอบมากกว่า 1 ฉบับ ซึ่งมีลักษณะดังภาพประกอบ 23



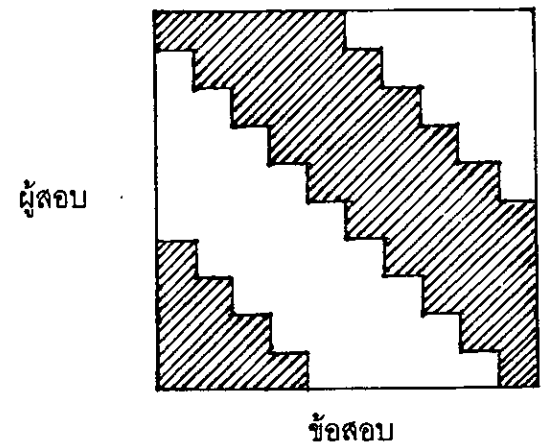
ภาพประกอบ 23 แบบแผนของการให้ผู้สอบร่วมกัน

(4) ใช้ข้อสอบและผู้สอบร่วมกัน (double anchor designs) สำหรับแผนนี้ ได้นำแบบแผนของการใช้ข้อสอบร่วมกันและผู้สอบร่วมกันมาผสมกัน เพื่อประหยัดค่าใช้จ่ายในการเก็บรวบรวมข้อมูล แต่ในทางปฏิบัติจริงนั้นไม่ค่อยได้มีการเลือกใช้แบบแผนนี้ เพราะยุ่งยากในเรื่องของการแปลงคะแนน ซึ่งมีแบบแผนดังภาพประกอบ 24



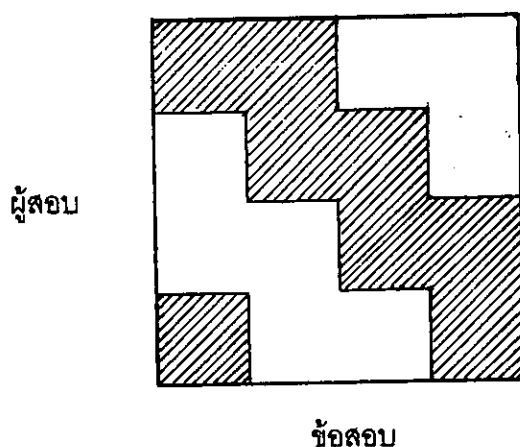
ภาพประกอบ 24 แบบแผนของการใช้ข้อสอบและผู้สอบร่วมกัน

(5) ใช้ข้อสอบร่วมกันแบบโยงสลับ (interlaced) แบบแผนนี้ใช้ข้อสอบร่วมกัน จำนวนของแบบทดสอบจะเท่ากับจำนวนของข้อสอบทั้งหมด และจำนวนข้อสอบในแต่ละฉบับจะเท่ากัน การจัดฉบับของแบบทดสอบจะมีลักษณะ ดังภาพประกอบ 25



ภาพประกอบ 25 แบบแผนของการใช้ข้อสอบร่วมกันแบบโยงสลับ

(6) ให้ข้อสอบร่วมกันแบบโยงสลับเป็นกลุ่ม (block interlaced anchor) แบบแผนนี้มีลักษณะคล้ายกับแบบแผนของการใช้ข้อสอบร่วมกันแบบโยงสลับ เป็นการแก้ปัญหาที่ต้องใช้แบบทดสอบหลายฉบับ ดังนั้น จึงจัดผู้สอบเป็นกลุ่มๆ ให้ทำแบบ ทดสอบฉบับเดียวกันดังภาพประกอบ 26 จึงทำให้จำนวนแบบทดสอบลดลงกว่าเดิม



ภาพประกอบ 26 แบบแผนของการใช้ข้อสอบร่วมกันแบบโยงสลับเป็นกลุ่ม

2.8.2 วิธีการแปลงคะแนน (transformation method) แบบทดสอบที่สามารถนำค่าพารามิเตอร์ของข้อสอบมาแปลงคะแนนได้นั้นต้องมีลักษณะ ดังนี้ (Wainer, 1990 : 146)

- (1) แบบทดสอบทุกฉบับต้องวัดโครงสร้าง (construct) เดียวกัน
- (2) การแปลงคะแนนต้องมีความเท่าเทียมกัน (equity) กล่าวคือ การแจกแจงอย่างมีเงื่อนไขของคะแนนของแบบทดสอบแต่ละฉบับต้องเท่าเทียมกัน

(3) ผลของการแปลงคะแนนต้องไม่แปรเปลี่ยน (invariant) เมื่อใช้กลุ่มประชากรต่างกัน

(4) คะแนนที่แปลงต้องมีลักษณะสมมาตร (symmetric) เช่น คะแนนที่แปลงจากแบบทดสอบ X ไปยังแบบทดสอบ Y จะเท่ากับคะแนนที่แปลงจากแบบทดสอบ Y ไปยังแบบทดสอบ X ดังนั้น จึงไม่สามารถแปลงคะแนนโดยใช้ลักษณะของการถดถอย (regression) ได้

ในการแปลงคะแนนไปสู่สเกลเดียวกันนั้นจะทำหลังจากที่ข้อสอบในแต่ละข้อแต่ละฉบับได้รับการประมาณค่าพารามิเตอร์แล้ว การแปลงคะแนนเป็นเพียงการจัดวางตำแหน่งของค่า พารามิเตอร์ของข้อสอบจากฉบับต่างๆ บนสเกลที่ใช้ร่วมกัน (common scale) วิธีการแปลง

คะแนนที่ง่ายที่สุด ก็คือ การแปลงคะแนนเชิงเส้น (linear transformation) ซึ่งจะได้

$$\theta^* = k \theta + m$$

$$b^* = kb + m$$

ค่าคงที่จะได้จาก

$$k = \frac{S^*}{S}$$

$$m = x^* - x \frac{S^*}{S}$$

เมื่อ θ^* , θ แทน ค่าความสามารถที่ได้รับการแปลงค่าและยังมิได้รับการแปลงค่า
 S^* , S แทน ค่าความคลาดเคลื่อนมาตรฐานที่แปลงค่าและยังมิได้รับการ
 แปลงค่า

x^* , x แทน คะแนนดิบที่ได้รับการแปลงค่าและยังมิได้รับการแปลงค่า

k , m แทน ค่าคงที่

จากแบบแผนของการใช้ข้อสอบร่วมกันและผู้สอบร่วมกัน ดังที่ได้กล่าวมาแล้ว สรุปเป็น
 กลุ่มใหญ่ๆ ได้ 3 กลุ่ม คือ กลุ่มที่ไม่ได้ใช้ข้อสอบและผู้สอบร่วมกันเลย กลุ่มที่ใช้ข้อสอบ ร่วมกัน
 และกลุ่มที่ใช้ผู้สอบร่วมกัน ซึ่งมีลักษณะและวิธีการแปลงคะแนน ดังนี้

(1) กลุ่มที่ไม่ได้ใช้ข้อสอบและผู้สอบร่วมกันเลย

ผลของการวิเคราะห์ข้อมูลจากผลการทดสอบจากกลุ่มนี้ ค่าพารามิเตอร์ที่ได้ ซึ่ง
 ได้แก่ คะแนนความสามารถ ค่าอำนาจจำแนก ค่าความยากง่ายและค่าการเดา ไม่สามารถนำมา
 เปรียบเทียบกันได้ เพราะ สเกลของ θ จะอยู่บนเมตริกที่ต่างกัน วิธีการแก้ไขก็ต้องรวมข้อมูล
 ของผู้สอบทั้งหมดเข้าด้วยกัน เหมือนเป็นผู้สอบกลุ่มเดียวกันแล้วจึงวิเคราะห์ ส่วนข้อสอบต่างชุดที่
 ผู้สอบไม่ได้ทำก็ถือเสมือนว่าเป็นข้อสอบที่ผู้สอบทำไม่ถึง (Lord. 1980 : 201)

(2) กลุ่มที่ใช้ข้อสอบร่วมกัน

ลักษณะของการทดสอบ ก็คือ จะมีข้อสอบอยู่จำนวนหนึ่งที่เหมือนกันอยู่ในแบบ
 ทดสอบทุกฉบับที่ใช้สอบ ซึ่งมีวิธีการทดสอบ ดังนี้

(2.1) นำแบบทดสอบไปให้ผู้สอบ 2 กลุ่ม กลุ่มละ 1 ฉบับ

(2.2) นำผลการทดสอบมาวิเคราะห์ค่าพารามิเตอร์แยกกัน โดยกำหนดค่าเฉลี่ย

และค่าเบี่ยงเบนมาตรฐานของความสามารถของผู้สอบ ซึ่งมักกำหนดให้เท่ากับ 0 และ 1 ตามลำดับ

(2.3) หาค่าเฉลี่ยเลขคณิตและค่าเบี่ยงเบนมาตรฐานของค่าความยากง่ายของข้อสอบรวม (anchor item) ในแบบทดสอบทั้ง 2 ฉบับ

(2.4) แปลงค่าความยากง่าย ค่าอำนาจจำแนกของข้อสอบ และค่าความสามารถของผู้สอบ จากฉบับหนึ่งไปสู่อีกฉบับหนึ่ง จากสูตร

$$b_1 = \left[\frac{SD_{b1}}{SD_{b2}} \right] b_2 + [\bar{b}_1 - \frac{SD_{b1}}{SD_{b2}} \bar{b}_2]$$

$$a_1 = a_2 \times \frac{SD_{b2}}{SD_{b1}}$$

$$\theta_1 = \left[\frac{SD_{b1}}{SD_{b2}} \right] \theta_2 + [\bar{b}_1 - \frac{SD_{b1}}{SD_{b2}} \bar{b}_2]$$

ค่าการเดาของข้อสอบไม่ต้องแปลงค่า เพราะอยู่บนสเกลเดียวกันกับค่าความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง ($P(\theta)$) อยู่แล้ว

หากต้องการวิเคราะห์ข้อมูลรวมกันก็สามารถทำได้โดยรวมข้อมูลผลการตอบของผู้สอบทั้ง 2 กลุ่ม ข้อสอบต่างฉบับที่ผู้สอบไม่ได้ตอบก็ถือเป็นข้อสอบที่ทำไม่ถึงหรือทำไม่ทัน จากนั้นก็วิเคราะห์ข้อสอบเหมือนกับว่าข้อสอบชุดใหญ่ถูกตอบโดยผู้สอบกลุ่มเดียวกัน

(3) กลุ่มที่ใช้ผู้สอบร่วมกัน

กรณีที่ใช้แบบทดสอบต่างฉบับกัน ผู้สอบต่างกลุ่มกัน แต่จะมีผู้สอบอยู่จำนวนหนึ่งที่ต้องตอบแบบทดสอบทั้ง 2 ฉบับ วิธีการแปลงคะแนนสามารถทำได้ ดังนี้

(3.1) รวมกระดาษคำตอบของผู้สอบรวม (anchor person) ทั้ง 2 ฉบับ ดังนั้น ก็จะเป็นเสมือนได้กระดาษคำตอบของแบบทดสอบฉบับยาว 1 ฉบับ

(3.2) วิเคราะห์ค่าพารามิเตอร์แยกเป็นรายฉบับก็จะได้ค่าความสามารถ 2 ค่า หาค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของค่าความสามารถแต่ละฉบับ แล้วแปลงค่าที่ได้ ดังนี้

$$b_1 = \left[\frac{SD_{\theta_1}}{SD_{\theta_2}} \right] b_2 + [\bar{\theta}_1 - \frac{SD_{\theta_1}}{SD_{\theta_2}} \bar{\theta}_2]$$

$$a_1 = a_2 \times \frac{SD_{\theta_2}}{SD_{\theta_1}}$$

$$\theta_1 = \left[\frac{SD_{\theta_1}}{SD_{\theta_2}} \right] \theta_2 + [\bar{\theta}_1 - \frac{SD_{\theta_1}}{SD_{\theta_2}} \bar{\theta}_2]$$

ค่าการเดานั้นไม่ต้องการแปลงค่า เพราะอยู่บนสเกลเดียวกันอยู่แล้ว

กรณีที่ไม่มีผู้สอบร่วมซึ่งต้องทำแบบทดสอบทั้ง 2 ฉบับ แต่ผู้ดำเนินการสอบต้องการที่จะใช้แบบแผนนี้ ก็อาจทำได้โดยนำข้อสอบจากทั้ง 2 ฉบับส่วนหนึ่งมาสร้างเป็นแบบทดสอบฉบับที่ 3 แล้วนำไปทดสอบกับผู้สอบกลุ่มที่ 3 แล้วจึงใช้แบบทดสอบฉบับที่ 3 นี้เป็นข้อสอบร่วมเพื่อปรับค่าพารามิเตอร์ของแบบทดสอบฉบับที่ 1 และฉบับที่ 2

การหาค่าคงที่สำหรับสมการเทียบคะแนน

ค่าความสามารถของผู้สอบและค่าพารามิเตอร์ของข้อสอบจากการทดสอบตั้งแต่ 2 ครั้งขึ้นไปสามารถนำมาจัดให้อยู่บนสเกลเดียวกันได้ โดยการใช้สมการเชิงเส้นเพื่อแปลงค่า ซึ่งสมการดังกล่าวมีค่าคงที่ 2 ค่า คือ α และ β ซึ่งมีวิธีการหาค่า 4 วิธี ดังนี้ (Hambleton and Swaminathan, 1985 : 206-210)

(1) Regression Method

ความสัมพันธ์เชิงเส้นระหว่าง 2 ตัวแปร ส่วนใหญ่แล้วมีลักษณะถดถอย
ดังสมการ

$$y = \alpha x + \beta + e$$

การแปลงค่าความสามารถทำได้โดย ให้ $y = \theta_y$ และ $x = \theta_x$ และในการแปลงค่าความยากง่ายของข้อสอบนั้น ทำได้โดยให้ $y = b_y$ และ $x = b_x$ ส่วนค่า e นั้น คือค่าความคลาดเคลื่อน ค่าสัมประสิทธิ์ถดถอย หาได้โดย

$$\alpha = r_{xy} \frac{s_y}{s_x}$$

$$\hat{\beta} = \bar{y} - \hat{\alpha}\bar{x}$$

เมื่อ r แทน สหสัมพันธ์ระหว่าง x กับ y

$\bar{y}, \bar{x}$ แทน ค่าเฉลี่ยของ y และ x ตามลำดับ

s_y, s_x แทน ค่าเบี่ยงเบนมาตรฐานของ y และ x ตามลำดับ

วิธีการนี้นับว่าไม่ค่อยเหมาะสมในการเทียบคะแนน เพราะมีลักษณะไม่สมมาตร

เนื่องจากแบบทดสอบที่ใช้เป็นฐาน (base test) จะมีผลต่อค่าสัมประสิทธิ์ถดถอย

(2) Mean and Sigma Method

วิธีนี้มีลักษณะความสัมพันธ์ของตัวแปร ดังสมการ

$$y = \alpha x + \beta$$

$$\bar{y} = \alpha \bar{x} + \beta$$

$$s_y = \alpha s_x$$

$$\alpha = \frac{s_y}{s_x}$$

$$\beta = \bar{y} - \alpha \bar{x}$$

ความสัมพันธ์นี้มีลักษณะสมมาตร หากใช้รูปแบบการวิเคราะห์ข้อมูลโดยราชโมเดลแล้ว ค่าคงที่ที่ได้จากวิธีนี้จะเท่ากับวิธีถดถอย

(3) Robust Mean and Sigma Method

ถึงแม้ว่าวิธีการหาค่าคงที่ วิธี Mean and Sigma Method จะมีลักษณะของการให้ค่าคงที่ที่แปลงคะแนนแล้วมีลักษณะสมมาตรก็ตาม แต่ก็ไม่สามารถตอบได้ว่า ค่าพารามิเตอร์ของข้อสอบและค่าความสามารถของผู้สอบจะถูกแปลงค่าได้อย่างถูกต้องแม่นยำกว่า ในการวิเคราะห์ข้อมูลแต่ละครั้ง ก็เกิดความคลาดเคลื่อนมาตรฐานในการประมาณค่าทุกครั้ง การถ่วงน้ำหนักให้กับข้อมูลแต่ละคู่ (x, y) ก็จะเป็นการปรับค่าความแปรปรวนที่มากกว่าจากการประมาณ ถ้าเป็นการแปลงค่าความสามารถ ค่าความสามารถที่มีความแปรปรวนมาก ๆ จะมีค่าถ่วงน้ำหนักน้อยๆ ถ้าใช้ค่าความยากง่ายเพื่อหาค่าคงที่ของสมการแล้ว information matrix ของข้อสอบแต่ละข้อจะถูกปรับค่า ส่วนประกอบของ diagonal ที่เหมาะสมจะถูกใช้เป็นค่าความแปรปรวน

วิธีการหาค่าคงที่ทำได้ ดังนี้

(3.1) กำหนดค่าถ่วงน้ำหนัก (w) ของ x และ y แต่ละคู่

$$w_j = \max \{ V(x_j), V(y_j) \} \quad \text{เมื่อ } j = 1, 2, \dots, k$$

เมื่อ $v(\theta_j)$ แทน ความแปรปรวนของการประมาณค่าความสามารถ ครั้งที่ j

(3.2) สเกลของค่าถ่วงน้ำหนัก คือ

$$w_j = \frac{w_j}{\sum_{j=1}^n w_j}$$

(3.3) หาค่าที่ได้รับการถ่วงน้ำหนักแต่ละคู่

$$x'_j = w'_j x_j$$

$$y'_j = w'_j y_j$$

(3.4) หาค่าเฉลี่ย และ ค่าเบี่ยงเบนมาตรฐานของค่าที่ถ่วงน้ำหนัก

(3.5) หาค่าคงที่ (α, β) จากสมการ

$$\alpha = \frac{s_y}{s_x}$$

$$\beta = y' - \alpha x'$$

ได้มีการปรับปรุงค่าถ่วงน้ำหนักให้มีประสิทธิภาพสูงขึ้น โดยนำระยะห่างระหว่างข้อมูลกับเส้นเทียบคะแนน ซึ่งเรียกว่า ตัวถ่วงน้ำหนักที่แกร่ง (robust weight) มาใช้ในการถ่วงน้ำหนัก

(3.6) เมื่อได้ค่าคงที่ของสมการแปลงคะแนนแล้ว เส้นเทียบคะแนนจะถูกกำหนดขึ้น ระยะห่างระหว่างข้อมูลกับเส้นเทียบคะแนน หาได้จากสมการ

$$d_j = \frac{(y_j - \alpha x_j - \beta)}{\sqrt{\alpha^2 + \beta^2}}$$

(3.7) หาค่าถ่วงน้ำหนัก (turkey weight) จากสมการ

$$T_j = [1 - (\frac{d_j}{6M})^2]^2 \quad ; \quad d_j < 6M$$

(3.8) ถ่วงน้ำหนักของข้อมูลแต่ละคู่ด้วยค่าถ่วงน้ำหนัก

$$u_j = w_j' T (\sum_{j=1}^n w_j' T_j)$$

(3.9) ทำซ้ำขั้นตอนที่ 3 โดยใช้ค่า u_j แทน w_j' แล้วหาค่าคงที่ α และ β ในขั้นตอนที่ 5

(3.10) ทำซ้ำขั้นตอนที่ 6 - 9 จนกระทั่งค่าคงที่ α และ β เปลี่ยนแปลงน้อยกว่าที่กำหนดไว้

(4) Characteristic Curve Method

วิธีการหาค่าคงที่ในสมการเทียบคะแนน วิธี Robust Mean and Sigma Method มีข้อบกพร่องตรงที่จะใช้ความสัมพันธ์ $b_y = \alpha b_x + \beta$ เท่านั้นในการหาค่าคงที่ แต่ไม่ได้ใช้ค่าอำนาจจำแนกในการหาค่า α แต่อย่างใด

คะแนนจริงของผู้สอบที่มีความสามารถ θ บนแบบทดสอบ x และความสามารถ θ บนแบบทดสอบ y ก็คือ

$$\xi_{xa} = \sum_{i=1}^n P(\theta_a, a_{xi}, b_{xi}, c_{xi})$$

$$\xi_{ya} = \sum_{i=1}^n P(\theta_a, a_{yi}, b_{yi}, c_{yi})$$

ค่าคงที่ α และ β ที่ดี ก็คือค่าที่ทำให้คะแนนจริง ξ_{xa} และ ξ_{ya} มีความแตกต่างกันน้อยที่สุดโดยใช้เกณฑ์ของสตอคคิงและลอร์ด (Stocking and Lord, 1983) ดังนี้

$$F = \frac{1}{N} \sum_{a=1}^N (\xi_{xa} - \xi_{ya})^2$$

เมื่อ N แทน จำนวนผู้สอบ

F แทน ฟังก์ชันของ α และ β ซึ่งมิต่ำน้อยที่สุดเมื่อ

เชื่อมั่นของแบบทดสอบ (reliability) หรือ ค่าความคลาดเคลื่อนมาตรฐานในการวัดตาม ทฤษฎี การทดสอบแบบเดิม (Hambleton and Cook. 1977 : 84)

วอร์ม (Warm. 1978 : 77) ได้แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่ากับค่าความเชื่อมั่นของแบบทดสอบไว้ ดังนี้

$$r_{tt} = 1 - \overline{SEE}^2$$

เมื่อ $\overline{SEE}$ แทน ค่าเฉลี่ยของความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ

เซมิจิม่า กรีน และคณะ (Samejima. 1977 : Green and others. 1984 : 352) มีความเห็นว่า ไม่ควรนำค่าความเชื่อมั่นมาใช้ในการทดสอบแบบปรับเหมาะกับความสามารถ ของผู้สอบ แต่ ดิฝกิ (Divgi. 1989 : 146) เห็นว่า ควรใช้ค่าความเชื่อมั่นของแบบทดสอบกับการทดสอบลักษณะนี้ ซึ่งค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าหาได้จากสูตร (ต่าย เชียงฉิน. 2534 : 9)

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

เมื่อ SEE แทน ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ

E แทน ค่าความคลาดเคลื่อนในการประมาณค่า

N แทน จำนวนผู้สอบ

2.9.3 ความเที่ยงตรง (validity) ความเที่ยงตรงนั้นนับว่ามีความสำคัญอย่างยิ่งต่อการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ (Green and others. 1984 : 354) ข้อสอบที่มีค่าอำนาจจำแนกสูงจะทำให้ข้อสอบนั้นมีความเที่ยงตรงสูง ความเที่ยงตรงที่นิยมใช้ก็คือ ความเที่ยงตรงเชิงโครงสร้าง (construct validity) ซึ่งเป็นความสัมพันธ์ระหว่างความสามารถจริงกับความสามารถที่ได้จากการประมาณค่า ซึ่งหาค่าได้จากสูตร (Urry. 1977 : 183 ; citing Cronbach and Meehl. 1957)

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

เมื่อ	σ^2	แทน ความเที่ยงตรงเชิงโครงสร้าง
	θ	แทน ความสามารถจริงของผู้สอบ
	θ	แทน ความสามารถของผู้สอบที่ได้จากการประมาณค่า
	N	แทน จำนวนผู้สอบ

2.10 การนำทฤษฎีการตอบสนองข้อสอบไปใช้

2.10.1 ใช้วิเคราะห์ข้อสอบ (Item analysis) นับแต่อดีตจนถึงปัจจุบันเรานิยมใช้ทฤษฎีการทดสอบแบบเดิมวิเคราะห์คุณภาพของข้อสอบเป็นรายข้อและแบบทดสอบทั้งฉบับ แต่เนื่องจากทฤษฎีดังกล่าวมีข้อจำกัดบางประการ เช่น ค่าสถิติของข้อสอบและแบบทดสอบที่ได้จากการวิเคราะห์จะแปรเปลี่ยนไปตามลักษณะของกลุ่มผู้สอบ ในการเปรียบเทียบความสามารถของผู้สอบนั้นจำเป็นต้องใช้แบบทดสอบฉบับเดียวกันหรือแบบทดสอบคู่ขนาน นอกจากนั้นผลการวิเคราะห์ก็ไม่ได้บ่งบอกว่าผู้สอบสามารถทำอะไรได้บ้างหรือมีข้อบกพร่องในเรื่องใด เพราะได้มีการนำเอาผลของการทดสอบของผู้สอบแต่ละคนไปเปรียบเทียบกับกลุ่มผู้สอบ ทำให้มองไม่เห็นในรายละเอียดย่อย

ค่าพารามิเตอร์ของข้อสอบตามทฤษฎีการตอบสนองข้อสอบซึ่ง ได้แก่ ค่าอำนาจจำแนก ค่าความยากง่ายและค่าการเดานั้นจะเป็นอิสระจากกลุ่มผู้สอบ ไม่แปรผันไปตามความสามารถของกลุ่มผู้สอบ ขณะเดียวกันความสามารถของผู้สอบก็จะเป็นอิสระจากข้อสอบเช่นเดียวกัน การที่ผู้สอบจะตอบข้อสอบได้ถูกต้องหรือไม่ นั้น ไม่ได้ขึ้นอยู่กับลักษณะของกลุ่มผู้สอบ แต่ขึ้นอยู่กับโค้งลักษณะข้อสอบ (item characteristic curve) และใช้ค่าสารสนเทศของแบบทดสอบ ซึ่งเป็นสิ่งที่แสดงความถูกต้องของการวัดแทนค่าความเชื่อมั่นของแบบทดสอบตามทฤษฎีการทดสอบแบบเดิม ทั้งนี้เนื่องจากส่วนกลับของรากที่สองของค่าสารสนเทศของแบบทดสอบ ก็คือ ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ

2.10.2 ใช้สร้างธนาคารข้อสอบ (Item bank) เนื่องจากค่าพารามิเตอร์ของข้อสอบที่ได้จากการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบนั้นจะไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบ ดังนั้น จึงสามารถที่จะเก็บข้อสอบพร้อมทั้งค่าพารามิเตอร์ประจำข้อไว้ใช้ในโอกาสต่อไปได้ การเก็บข้อสอบพร้อมทั้งค่าพารามิเตอร์ที่เป็นตัวบ่งชี้คุณภาพไว้อย่างเป็นระบบระเบียบสะดวกต่อการนำมาใช้ตามความต้องการ ตลอดจนสามารถเพิ่มหรือลดข้อสอบออกจากระบบได้โดยข้อสอบที่เพิ่มเข้าไปในระบบนั้น หากถูกวิเคราะห์โดยใช้ผลการทดสอบจากกลุ่มตัวอย่างคนละกลุ่มกับ

ข้อสอบชุดเดิมแล้ว ข้อสอบชุดใหม่ที่จะนำเข้าไปเก็บในระบบจำเป็นต้องได้รับการแปลงคะแนน (equating) ให้อยู่บนสเกลเดียวกันกับข้อสอบชุดเดิม นอกจากนั้นยังต้องมีการปรับปรุงข้อสอบให้มีความเป็นปัจจุบันอยู่เสมอ การดำเนินการตามลักษณะที่กล่าวมาก็คือ การสร้างธนาคารข้อสอบนั่นเอง ซึ่งในปัจจุบันได้มีการนำเอาคอมพิวเตอร์เข้ามาใช้ประโยชน์ในการสร้างธนาคารข้อสอบ ซึ่งช่วยให้ได้รับความสะดวกเป็นอย่างยิ่งในการดำเนินการ

2.10.3 ใช้ศึกษาลักษณะการตอบแบบทดสอบ (response pattern) ในการศึกษา ความถดถอยระหว่างค่าเศษเหลือของการตอบ (response residual) กับค่าความยากง่ายของข้อสอบ ถ้าเส้นถดถอยโค้งขึ้น แสดงว่าผู้สอบตอบข้อสอบด้วยการเดา ถ้าเส้นโค้งลงแสดงว่าผู้สอบตอบข้อสอบด้วยความสะเพร่า ถ้าค่าจากสมการถดถอยระหว่างค่าเศษเหลือของการตอบกับตำแหน่งของผู้สอบเป็นบวก แสดงว่า ผู้สอบตอบข้อสอบอย่างไม่เป็นระบบ ถ้าค่าเป็นลบ แสดงว่าผู้สอบตอบข้อสอบไม่ทันเวลา (สุพัฒน์ สุกมลสันต์. ม.ป.ป. : 8/5 ; อ้างอิงมาจาก Wright. 1977 : 112)

2.10.4 ใช้ตรวจสอบความลำเอียงของข้อสอบ (Item bias) ทฤษฎีการตอบสนอง ข้อสอบมีความเชื่อว่า ในการทดสอบครั้งหนึ่งครั้งใดก็ตาม ผู้สอบที่มีความสามารถเท่ากันย่อมมีความน่าจะเป็นในการตอบข้อสอบข้อเดียวกันได้ถูกต้องเท่ากัน แต่ก็มีบางกรณีที่ผู้สอบซึ่งมีความสามารถเท่าเทียมกัน แต่มีความน่าจะเป็นในการตอบข้อสอบข้อเดียวกันได้ถูกต้องไม่เท่ากัน ลักษณะเช่นนี้ถือว่าข้อสอบดังกล่าวมีความลำเอียง (bias) เป็นข้อสอบที่เอื้อต่อการสอบของผู้สอบคนหนึ่งมากกว่าอีกคนหนึ่ง ในการที่ผู้สอบซึ่งมีความสามารถเท่าเทียมกันในมิติที่ถูกวัด แต่ความน่าจะเป็นในการตอบถูกต่างกัน แสดงให้เห็นว่า ข้อสอบดังกล่าวไม่ได้วัดเพียงความสามารถเดียวที่ผู้สอบมีอยู่เท่านั้น แต่ยังวัดความสามารถด้านอื่นอีกที่ผู้สอบซึ่งมีโอกาสตอบถูกสูงกว่ามีอยู่ แต่ผู้สอบที่มีความน่าจะเป็นในการตอบถูกต่ำกว่าจะไม่มี ทฤษฎีการตอบสนองข้อสอบสามารถบอกได้เป็นอย่างดีว่าข้อสอบข้อใดมีความลำเอียง แต่อย่างไรก็ตามในเรื่องของสาเหตุของความลำเอียงของข้อสอบ ตลอดจนเกณฑ์ที่เหมาะสมที่จะชี้ความลำเอียงของข้อสอบก็เป็น เรื่องที่ยังไม่มีความชัดเจนเท่าที่ควร

2.10.5 ใช้ในการเทียบคะแนน (equating) ในกรณีที่มีผู้สอบกลุ่มเดียว แต่ตอบ แบบทดสอบหลายฉบับที่วัดสิ่งเดียวกัน ในเวลาที่ใกล้เคียงกัน หรือผู้สอบหลายกลุ่มที่มีความสามารถต่างกันแต่ตอบแบบทดสอบที่วัดสิ่งเดียวกัน หรือผู้สอบหลายๆกลุ่มที่มีความสามารถต่างกัน และแต่ละกลุ่มของผู้สอบก็ตอบแบบทดสอบแตกต่างกันออกไปแต่แบบทดสอบดังกล่าวแต่ละฉบับมีข้อสอบที่ร่วมกันอยู่บางส่วน ทั้ง 3 สถานการณ์ที่กล่าวมานี้ หากต้องการทราบ

ผู้สอบคนใดมีความสามารถมากกว่ากัน ก็สามารถทำได้โดยการเทียบคะแนน

ในการเทียบคะแนนนั้นจำแนกออกเป็นประเภทใหญ่ๆ ได้ 2 ประเภท คือ การเทียบคะแนนตามแนวนอน (horizontal equating) ซึ่งเป็นการใช้แบบทดสอบหลายฉบับที่มีค่าความยากง่ายใกล้เคียงกันกับกลุ่มผู้สอบกลุ่มเดียวกัน แล้วเทียบคะแนนระหว่างแบบทดสอบแต่ละชุดว่าที่คะแนนของแบบทดสอบชุดหนึ่งเท่ากับที่คะแนนของแบบทดสอบอีกชุดหนึ่ง ส่วนอีกประเภทหนึ่งนั้นเป็นการเทียบคะแนนตามแนวตั้ง (vertical equating) ซึ่งเป็นการใช้แบบทดสอบหลายฉบับที่มีค่าความยากง่ายต่างๆ กัน นำไปทดสอบกับผู้สอบหลายๆ กลุ่ม การเทียบคะแนนก็จะเทียบตามระดับความสามารถของผู้สอบ (สงบ ลักษณะ. 2525 : 23)

2.10.6 ใช้ออกแบบแบบทดสอบที่ดีที่สุด (best test design) แต่เดิมนั้นผู้สร้างแบบทดสอบไม่สามารถระบุได้ก่อนสอบว่า ผลของการทดสอบด้วยแบบทดสอบที่สร้างขึ้นจะมีความถูกต้องหรือความคลาดเคลื่อนในการวัดเพียงไร จะทราบได้ก็ต่อเมื่อการทดสอบได้ผ่านพ้นไปแล้วเท่านั้น จากประเด็นนี้ จึงนำไปสู่แนวคิดในการออกแบบแบบทดสอบที่ดีที่สุดโดยจะมีการออกแบบแบบทดสอบไว้ล่วงหน้าก่อนการทดสอบ ทั้งนี้เพื่อให้แบบทดสอบดังกล่าววัดได้ถูกต้องแม่นยำ มีความคลาดเคลื่อนในการวัดน้อยที่สุด

แนวคิดที่สำคัญในการออกแบบแบบทดสอบก็คือ การจัดชุดของแบบทดสอบให้เหมาะสมกับความสามารถของผู้สอบ ดังนั้นหากผู้สร้างแบบทดสอบมีข้อสงสัยเกี่ยวกับผู้สอบละเอียดมากเท่าไร การออกแบบแบบทดสอบก็จะมีประสิทธิภาพมากเท่านั้น แบบแผนของแบบทดสอบที่ดีที่สุดจะขึ้นอยู่กับความสัมพันธ์ระหว่างลักษณะของแบบทดสอบกับลักษณะ ของกลุ่มผู้สอบที่ทำให้เกิดความคลาดเคลื่อนในการวัดมีค่าต่ำสุด ซึ่งไรท์และสโตน (Wright and Stone. 1979 : 139-140) ได้ใช้โมเดลของราส์ช (Rasch model) ที่ใช้เฉพาะค่าความยากง่ายของข้อสอบเพียงอย่างเดียว ในการออกแบบแบบทดสอบ

2.10.7 ใช้ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ (adaptive testing) หลักการของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบนั้น ก็คือ การจัดข้อสอบให้มีความเหมาะสมกับความสามารถของผู้สอบ ผู้สอบที่มีความสามารถต่ำก็จะได้ข้อสอบที่ง่าย ส่วนผู้สอบที่มีความสามารถสูงก็จะได้ข้อสอบที่ยาก สมบัติในด้านความไม่แปรเปลี่ยน (invariant) ของค่าพารามิเตอร์ของข้อสอบไปตามกลุ่มข้อสอบตามทฤษฎีการตอบสนองข้อสอบได้ถูกนำมาใช้ประโยชน์อย่างมากในการทดสอบลักษณะนี้ นอกจากนั้นยังสามารถกำหนดค่าความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบในแต่ละครั้งและผู้สอบแต่ละคนได้อีก

ด้วย การทดสอบลักษณะนี้ใช้เวลาในการทดสอบน้อยลงกว่าเดิม ผู้สอบไม่เกิดความเบื่อหน่าย และยังมี เจตคติที่ดีต่อการสอบมากกว่าการทดสอบแบบเดิม

3. งานวิจัยที่เกี่ยวข้องกับการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ

งานวิจัยที่เกี่ยวข้องกับการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบนั้น พอสรุปเป็นประเด็นต่างๆได้ ดังนี้

3.1 โมเดลการตอบสนองข้อสอบ (item response model)

จากการศึกษากับข้อมูลจำลอง ยูรี (Urry. 1970,1977) พบว่า การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบนั้นจะไม่ค่อยมีประสิทธิภาพเลือกใช้โมเดลการวิเคราะห์ข้อมูลที่น้อยเหมาะสม เช่น ค่าความเที่ยงตรงเชิงโครงสร้างของแบบทดสอบจะลดลงเมื่อเทียบกับการใช้แบบทดสอบที่มีความยาวเท่ากันแต่เลือกใช้โมเดลการวิเคราะห์ข้อสอบที่ดีกว่า โมเดลของราล์ช (Rasch) นับว่าไม่ค่อยเหมาะสมในการนำมาใช้กับการวิเคราะห์ข้อสอบแบบเลือกตอบ ทั้งนี้เนื่องจากโมเดลของราล์ชมีข้อดกลงเบื้องต้นว่า ค่าอำนาจจำแนกของข้อสอบแต่ละข้อต้องคงที่ และค่าการเดาต้องมีค่าน้อยมากหรือเป็นศูนย์ ซึ่งตามสภาพความเป็นจริงแล้ว นับว่าเป็นเรื่องที่ยากมากที่จะทำให้ข้อสอบแต่ละข้อมีลักษณะตามข้อดกลงดังที่กล่าวมา โมเดลโลจิสติก 3 พารามิเตอร์นับว่ามีความเหมาะสมมากกว่าโมเดลอื่นๆ (Urry. 1977 : 183-184)

3.2 คลังข้อสอบ (Item pool)

โฮ (Ho. 1989 : 421-A) พบว่า ขนาดของคลังข้อสอบส่งผลน้อยมากต่อการประมาณค่าความสามารถของผู้สอบ ยกเว้นกลุ่มผู้สอบที่มีความสามารถต่ำ เมื่อใช้ข้อสอบเริ่มต้นที่มีค่าความยากง่ายสูงกว่าความสามารถที่แท้จริงของผู้สอบ และยังพบว่า มีปฏิสัมพันธ์ ระหว่างชนิดของคลังข้อสอบ (คลังข้อสอบที่มีข้อสอบแบบสุ่ม, ข้อสอบที่มีค่าอำนาจจำแนกสูงและมีเฉพาะข้อสอบที่ง่ายๆ) กับวิธีการประมาณค่าความสามารถ (วิธี bayesian , modal bayesian และ maximum likelihood) และระหว่างชนิดของคลังข้อสอบ กับระดับความสามารถของผู้สอบ (สูง ปานกลาง ต่ำ)

ไวส์ (Weiss. 1988 : 374-375) ได้เสนอแนะว่า คลังข้อสอบที่มีขนาดพอเหมาะนั้น ควรมีประมาณ 100-200 ข้อ ซึ่ง 116-150 ข้อ จะให้ผลดีที่สุด ซึ่งสอดคล้องกับข้อค้นพบของยูรี (Urry. 1970,1977) ที่พบว่า คลังข้อสอบควรมีข้อสอบอย่างน้อย 100 ข้อ และข้อสอบที่จะช่วยยก ระดับคุณภาพของการวัดนั้น ควรมีค่าอำนาจจำแนกมากกว่า 0.8 ค่าความยากอยู่ในช่วง-2.0 ถึง 2.0 และค่าการเดาน้อยกว่า 0.3 (Urry. 1977 : 184-187)

3.3 ความยากง่ายของข้อสอบข้อแรก (entry level)

ในการเลือกข้อสอบข้อแรกให้กับผู้สอบนั้น พบว่า ลักษณะของข้อสอบข้อแรก นั้น ไม่มีผลต่อความถูกต้องในการประมาณค่าความสามารถของผู้สอบเมื่อสิ้นสุดการสอบ (Hulin and others. 1983 : 219 ; citing Lord. 1977) แต่โฮ (Ho. 1989 : 421-A) พบว่า ถ้าใช้ ข้อสอบเริ่มต้นที่มีความยากง่ายเท่ากับหรือน้อยกว่าความสามารถจริงของผู้สอบแล้ว การประมาณค่าความสามารถของผู้สอบจะมีความถูกต้องมากยิ่งขึ้น ส่วนผลของค่าความยากง่าย ของข้อสอบที่ใช้เมื่อเริ่มต้นสอบต่อค่าความลำเอียงนั้น แฮนกินส์ (Hankins. 1987 : 3013-A) พบว่า แม้ว่าจะกำหนดค่าความยากง่ายของข้อสอบที่ใช้เมื่อเริ่มต้นแบบคงที่แน่นอน (fixed entry) หรือแบบแปรผัน (variable entry) สำหรับผู้สอบที่มีความสามารถปานกลาง ผลของการประมาณค่าจะไม่ลำเอียง แต่ถ้าผู้สอบที่มีความสามารถสูงและต่ำแล้ว การประมาณค่าจะลำเอียง

3.4 เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ (selection rule)

ฮูลินและคนอื่นๆ (Hulin and others. 1983 : 222-223) พบว่า การใช้เกณฑ์การคัดเลือกข้อสอบโดยการจับคู่ระหว่างค่าความสามารถของผู้สอบกับค่า θ_{max} เพื่อแก้ปัญหาการใช้ข้อสอบมากเกินไปนั้น พบว่า มีการสูญเสียสารสนเทศไปมากกว่าการคัดเลือกข้อสอบโดยวิธีสารสนเทศสูงสุด (I_{max}) อย่างไรก็ตามก็ยังให้สารสนเทศมากกว่าการทดสอบแบบเดิม กล่าวคือ ถ้าความสามารถของผู้สอบต่ำกว่า -1.25 แล้ว การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์จะให้สารสนเทศมากเป็น 2 เท่าของการทดสอบแบบเดิม

3.5 วิธีการประมาณค่าความสามารถของผู้สอบ (ability estimation procedure)

ไวส์และแม็คไบรด์ (Weiss and McBride) ได้ศึกษาเกี่ยวกับวิธีการให้คะแนน หรือวิธีการประมาณค่าความสามารถของผู้สอบตามวิธีการของเบส์ที่ปรับใหม่ (bayesian updating) พบว่า เป็นวิธีการประมาณค่าที่ไม่ลำเอียง นอกจากนี้ ยูรี (Urry. 1971) และ เจนเซมา (Jensem. 1972) ยังพบว่า ความสามารถจริงกับความสามารถที่ได้จากการประมาณค่าตามวิธีของเบส์มี ความสัมพันธ์กันสูงมาก (ต่าย เชียงฉี. 2534 : 13 ; อ้างอิงมาจาก Weiss and McBride. 1984 : 274)

จากการศึกษาของต่าย เชียงฉี (2534) พบว่า การประมาณค่าความสามารถตามวิธีการของเบส์ (bayesian) และวิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (conditional maximum likelihood) นั้นให้ผลการประมาณที่ไม่ต่างกัน การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบรูปพีระมิดที่มีรูปแบบ จำนวนชั้นและวิธีการให้คะแนนที่ต่างกันจะส่งผลต่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า นอกจากนี้ยังพบว่าค่าความเที่ยงตรงเชิงโครงสร้างของ แบบทดสอบพีระมิดแบบขนาดชั้นคงที่ที่มี 16 ชั้น ประมาณค่าความสามารถด้วยวิธีของเบส์มี ประสิทธิภาพ

ภาพในการประมาณค่าความสามารถของผู้สอบสูงสุด ซึ่งข้อค้นพบดังกล่าวมีบางส่วน สอดคล้องกับผลการศึกษาของ สเกจจ์และสตีเวนสัน (Skaggs and Stevenson. 1989 : 391-402) ซึ่งได้ศึกษาเปรียบเทียบผลของการประมาณค่าพารามิเตอร์ต่างๆ ด้วยวิธีของเบส์(pseudo-bayesian) กับวิธีความเป็นไปได้สูงสุด (joint maximum likelihood) โดยใช้โมเดล การวิเคราะห์แบบ 3 พารามิเตอร์ วิเคราะห์ข้อมูล จากผลการศึกษา พบว่า ถ้าใช้กลุ่มตัวอย่าง ขนาดใหญ่ ตั้งแต่ 2000 คนขึ้นไปและข้อสอบตั้งแต่ 35 ข้อขึ้นไปแล้ว การประมาณค่าทั้ง 2 วิธี นี้จะมีความถูกต้องสูงและให้ผลไม่ต่างกัน ถ้ากลุ่มผู้สอบน้อยกว่า 500 คน การประมาณค่าด้วย วิธีของเบส์ จะให้ความคงเส้นคงวา (consistency) และความถูกต้องแม่นยำ (accuracy) มากกว่า แต่ถ้าใช้ข้อสอบน้อยกว่า 15 ข้อแล้ว การประมาณค่าด้วยวิธีความเป็นไปได้สูงสุดจะให้ความถูกต้องแม่นยำมากกว่า นอกจากนี้ กิฟฟอร์ดและสวามินาธาน (Gifford and Swaminathan) ก็ได้ศึกษาในทำนองเดียวกันนี้ และยังพบว่าถ้ากลุ่มตัวอย่างมีขนาดเล็กและใช้ข้อสอบน้อยข้อแล้ว การประมาณค่าพารามิเตอร์ต่างๆ ด้วยวิธีของเบส์จะมีความถูกต้องมากกว่า เมื่อเทียบกับวิธีความเป็นไปได้สูงสุด (joint maximum likelihood)

3.6 เจตคติต่อการสอบ (examinee attitude)

จากการศึกษาเกี่ยวกับเจตคติต่อการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบนั้น พบว่า ผลการวิจัยของงานนิธย์ ธาตุทองกับปรีชา เครือวัลย์ (2527) สอดคล้องกับผลการวิจัยของลี (ต่าย เขียงจี. 2534 : 7 ; อ้างอิงมาจาก Lee. 1987 : 3742-A) ซึ่งพบว่า ผู้สอบส่วนใหญ่มีเจตคติที่ดีต่อการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ อีกทั้ง สอดคล้องกับผลการศึกษาของแบล็คมอร์ (Blackmore 1987 :2554-A) ซึ่งพบว่า ผู้สอบมีเจตคติต่อการทดสอบแบบเดิมแต่ใช้คอมพิวเตอร์ดำเนินการสอบ (computerized) การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ (computerized adaptive testing , CAT) และการทดสอบแบบเดิมแตกต่างกัน

3.7 ประสิทธิภาพของการวัด (efficiency)

นันทิยา พึ่งคำ (2531) พบว่า ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จะให้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบต่ำกว่าและใช้ข้อสอบน้อยกว่าการทดสอบแบบเดิม แต่ยังให้ความเที่ยงตรงในการประมาณค่าความสามารถเหมือนเดิม ซึ่งสอดคล้องกับผลการวิจัยของไวส์ (Weiss. 1982 : 473-492) และมอริโนและคนอื่นๆ (Moreno and others. 1984 : 155-163)

จากการศึกษาของต่าย เชียงฉี (2534) พบว่า ในการทดสอบแบบปรับเหมาะกับความ
ความสามารถของผู้สอบโดยใช้แบบทดสอบรูปพีระมิดนั้น เมื่อจำนวนชิ้นเพิ่มขึ้น (ทั้งแบบขนาดชั้น
คงที่และขนาดชั้นแปรผัน)จะทำให้ประสิทธิภาพของการประมาณค่าความสามารถของผู้สอบสูงขึ้น
ซึ่งสอดคล้องกับผลการศึกษาของธอร์นไดค์ (Thondike. 1982 : 321) ที่พบว่า การทดสอบด้วยแบบ
ทดสอบรูปพีระมิดนั้น หากแบบทดสอบมีจำนวนชิ้นมากขึ้น และข้อสอบมีค่าอำนาจจำแนกสูงขึ้น
ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าจะลดลง

เสรี ปรมชวลิตโรจน์ (2531) พบว่า การจัดข้อสอบแบบเฟล็กซิเลเวลและรูป
พีระมิดแบบขนาดชั้นแปรผัน จะมีความเที่ยงตรงในการจำแนกความรอบรู้อย่างมีนัยสำคัญ และ
ในการทดสอบแบบเฟล็กซิเลเวลนี้ ลอร์ด (Lord. 1971 : 805-813) พบว่า เมื่อขนาดของช่วงห่าง ของ
ค่าความยากง่ายของข้อสอบมากขึ้น จะทำให้ค่าสารสนเทศลดลงทุกระดับความสามารถ แบบ
ทดสอบจะมีประสิทธิภาพสูงสุดเมื่อมีช่วงห่างของความยากง่ายอยู่ระหว่าง 0.033 ถึง 0.067 และ
จะไม่มีประสิทธิภาพเมื่อใช้ช่วงห่างของความยากง่ายใกล้เคียงกับ 0.1 และเมื่อนำไป เปรียบเทียบ
กับการทดสอบแบบเดิม พบว่า การทดสอบแบบเฟล็กซิเลเวลจะให้สารสนเทศสูง กว่าทดสอบ
แบบเดิมเมื่อผู้สอบมีความสามารถสูงหรือต่ำ แต่สำหรับผู้สอบที่มีความสามารถปานกลางแล้ว
การทดสอบแบบเดิมจะให้ค่าสารสนเทศสูงว่าการทดสอบแบบเฟล็กซิเลเวล

สำหรับรายละเอียดของงานวิจัยแต่ละเรื่องดังกล่าว พอสรุปได้ ดังนี้

ดักลาส (Douglas. 1990) ได้เปรียบเทียบการดำเนินการสอบความถนัดโดยใช้แบบ
ทดสอบ DAT (differential aptitude tests) ซึ่งใช้วิธีการทดสอบแบบเดิมและการทดสอบแบบปรับ
เหมาะความสามารถของผู้สอบด้วยคอมพิวเตอร์ กลุ่มตัวอย่างเป็นนักเรียนระดับมัธยมศึกษา
จำนวน 40 คน วิเคราะห์ข้อมูลโดยใช้สหสัมพันธ์ การวิเคราะห์ความแปรปรวน การถดถอยเชิงเส้น
ตรวจสอบภาวะสารรูปสนิทสุด (goodness of fit) ด้วยสถิติอนพาราเมตริกคอลโมโกรอฟ-มินอฟ
(Kolmogorov smithnov) ผลจากการศึกษาพบว่า ประสิทธิภาพทางคอมพิวเตอร์และเจตคติต่อ
คอมพิวเตอร์มีความสัมพันธ์กับผลการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ แต่ในเรื่อง
ประสิทธิภาพเกี่ยวกับการพิมพ์นั้นไม่มีความสัมพันธ์กับผลของการทดสอบดังกล่าว

ควน (Kuan. 1991.) ได้ศึกษาเปรียบเทียบวิธีการทดสอบ 4 วิธี คือ การทดสอบ
แบบเดิม (PAP) การทดสอบด้วยคอมพิวเตอร์และแสดงข้อมูลย้อนกลับ (CFT) การทดสอบด้วย
คอมพิวเตอร์แต่ไม่แสดงข้อมูลย้อนกลับ (CAMT) การทดสอบแบบปรับเหมาะความสามารถ
ของผู้สอบด้วยคอมพิวเตอร์ (CADT) กลุ่มตัวอย่างได้แก่ นักศึกษาจำนวน 120 คนที่ลงทะเบียน

เรียนคอมพิวเตอร์ที่มหาวิทยาลัยมิสซิสซิปปี กลุ่มตัวอย่างดังกล่าวถูกแบ่งออกเป็น 4 กลุ่มย่อย สุ่มวิธีการทดสอบให้แต่ละกลุ่ม กลุ่มที่ 1 ได้แบบทดสอบแบบเดิม 2 ชุด กลุ่มที่ 2 ได้ PAP และ CAMT กลุ่มที่ 3 ได้ PAP และ CFT กลุ่มที่ 4 ได้ PAP และ CADT ผลการศึกษา พบว่า ผลการทดสอบจาก CAMT, CFT และ CADT มีความสัมพันธ์ทางบวกกับ PAP การเปรียบเทียบคะแนนเฉลี่ยโดยการวิเคราะห์ความแปรปรวนร่วม พบว่า คะแนนเฉลี่ยจากการทดสอบทั้ง 4 วิธีดังกล่าวไม่แตกต่างกัน เมื่อเปรียบเทียบเวลาที่ให้ทดสอบ พบว่า การทดสอบ CADT ใช้เวลาในการทดสอบน้อยกว่าวิธีอื่นๆ

คาร์โรล (Carroll, 1990) ได้เปรียบเทียบผลการทดสอบเพื่อจัดนักเรียนเข้ากลุ่ม โดยใช้การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์หรือการทดสอบแบบ CAT และการทดสอบแบบเดิมโดยใช้แบบสำรวจ HMPI (the publishers placement inventory) เพื่อวัดความเข้าใจในการอ่าน กลุ่มตัวอย่างได้แก่ นักเรียนระดับประถมศึกษา จำนวน 213 คน ผลการศึกษาพบว่า การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ เสียค่าใช้จ่ายน้อยกว่าและมีประสิทธิภาพมากกว่าการทดสอบด้วย HMPI คือใช้เวลาเพียง 1 ใน 3 และใช้ข้อสอบน้อยกว่าถึงครึ่งหนึ่ง นอกจากนั้นผู้สอบร้อยละ 92 ยังมีเจตคติที่ดีต่อการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบมากกว่า การทดสอบด้วยแบบสำรวจ HMPI

ไครสกี (Kirisici, 1990) ได้เปรียบเทียบผลของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ กับการทดสอบแบบแฟล็กซีเลเวล โดยการทดสอบแบบปรับเหมาะถูกตั้งชื่อเป็น การทดสอบแบบ PAT (predictive adaptive testing) เนื่องจากใช้วิธีการกำหนดค่าพารามิเตอร์ที่ไม่ทราบค่าแบบทำนายและประมาณค่าพารามิเตอร์ด้วยวิธีของเบส์ ข้อสอบที่ใช้มี 71 ข้อ ผู้สอบ 999 คน ซึ่งได้มาโดยการจำลองข้อมูล หาค่าพารามิเตอร์โดยใช้โมเดลโลจิสติก 3 พารามิเตอร์ ผู้สอบถูกแบ่งออกเป็น 3 กลุ่ม คือ กลุ่มที่มีความสามารถสูง ปานกลางและต่ำ ซึ่งสุ่มแบบแบ่งชั้น ดำเนินการสอบด้วยคอมพิวเตอร์โดยใช้โปรแกรมที่เขียนด้วยภาษาฟอร์แทรน 4 และโปรแกรมสำเร็จรูป MicroCAT วิเคราะห์ข้อมูลโดยใช้สถิติการวิเคราะห์ความแปรปรวน แบบแผนแฟคทอเรียล 2 ทาง มีตัวแปรตาม 3 ตัว คือ จำนวนข้อสอบที่ใช้ ผลต่างของความสามารถจริงกับความสามารถที่ได้จากการประมาณค่าและผลต่างของความแปรปรวนของความคลาดเคลื่อนจากการทดสอบแบบ PAT กับการทดสอบแบบเดิมรูปแบบ ผลการศึกษาพบว่า ที่ระดับความสามารถปานกลาง การทดสอบแบบ PAT ใช้ข้อสอบต่างจากการทดสอบแบบแฟล็กซีเลเวล ส่วนผลต่างของความแปรปรวนของความคลาดเคลื่อนและผลต่างของความสามารถจริงกับความสามารถที่ได้จาก

การประมาณที่ได้จากการทดสอบทุกแบบนั้นไม่ต่างกัน นอกจากนั้นความสามารถจริงและความสามารถที่ได้จากการประมาณยังมีความสัมพันธ์กันด้วย

ซาง (Chang, 1990) ได้ศึกษาวิธีการแยกทางในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ กรณีที่การทดสอบดังกล่าวมีหลายเนื้อหา ซึ่งอาจเป็นไปได้ที่แบบทดสอบดังกล่าวจะมีหลายมิติ ดังนั้นจึงได้จัดทำแบบทดสอบเป็นฉบับย่อยๆ และได้เปรียบเทียบวิธีการแยกทางไปสู่แบบทดสอบย่อย 4 วิธี คือ วิธีใช้ค่าสหสัมพันธ์พหุคูณของคะแนนดิบ ชุดที่ให้ค่าสหสัมพันธ์ที่สูงกว่าจะได้รับการคัดเลือก วิธีใช้ค่าสหสัมพันธ์พหุคูณของคะแนนดิบชุดที่ให้ค่าสหสัมพันธ์ที่ต่ำกว่าจะได้รับการคัดเลือก วิธีคัดเลือกแบบทดสอบย่อยด้วยการสุ่ม วิธีสุดท้ายคือไม่มีการแยกทางซึ่งใช้เป็นฐานในการเปรียบเทียบกับวิธีอื่นๆ กลุ่มตัวอย่างได้มาโดยการจำลองข้อมูล 1,000 คน สำหรับการวิเคราะห์หาค่าพารามิเตอร์ของข้อสอบ และ 200 คน สำหรับการทดสอบแบบปรับเหมาะความสามารถของผู้สอบด้วยคอมพิวเตอร์ ใช้แบบทดสอบ 4 ชุดย่อย แต่ละชุดย่อยมีข้อสอบ 100 ข้อ แบบทดสอบทั้ง 4 ชุดย่อยมีสหสัมพันธ์อยู่ในช่วง .14 ถึง .69 ผลการศึกษาพบว่า วิธีที่ใช้ค่าสหสัมพันธ์สูงจะให้ผลดีกว่าวิธีอื่นๆ โดยเฉพาะการทดสอบขั้นสุดท้ายที่แบบทดสอบชุดย่อยชุดที่ 3 และชุดที่ 4 ถูกดำเนินการสอบ อย่างไรก็ตามความแตกต่างระหว่างการแยกทางทั้ง 4 วิธีก็ไม่เด่นชัดพอที่จะอ้างอิงในการนำไปใช้ จำนวนชุดย่อยของข้อสอบและค่าสหสัมพันธ์ระหว่างชุดย่อยของข้อสอบควรเป็นตัวแปรที่จะต้องศึกษาต่อไป

โฮ (Ho, 1988) ได้เปรียบเทียบวิธีการทดสอบแบบปรับเหมาะความสามารถของผู้สอบ 3 วิธี คือ ประมาณค่าความสามารถของผู้สอบโดยวิธีของเบย์ (bayesian) วิธีโมดัลเบย์ (modal-bayesian) และวิธีความเป็นไปได้สูงสุด (maximum likelihood) โดยใช้โปรแกรมในการทดสอบนี้กับคลังข้อสอบ 3 ขนาด คือ 86 , 71 และ 56 ข้อ ลักษณะของคลังข้อสอบมี 3 แบบ คือ ข้อสอบได้มาโดยการสุ่ม ข้อสอบที่มีค่าอำนาจสูงๆและข้อสอบที่ง่ายๆ ทดลองกับผู้สอบที่มีความสามารถ 3 ระดับ คือ สูง กลาง และต่ำ จุดเริ่มต้นในการทดสอบนั้นใช้จุดเริ่มต้นเดียวกัน วิเคราะห์ข้อมูลโดยการวิเคราะห์ความแปรปรวน 3 ทาง แบบวัดซ้ำ ซึ่งตัวแปรตาม ได้แก่ จำนวนข้อสอบที่ใช้ ความแปรปรวนของความคลาดเคลื่อนเฉลี่ย และค่าสัมบูรณ์ของความแตกต่างระหว่างค่าความสามารถที่คาดหวังกับค่าความสามารถโดยประมาณ ซึ่งตัวแปรตาม 2 ตัวแรกนั้นใช้สำหรับประเมินประสิทธิภาพของการทดสอบ ตัวที่ 3 ใช้ประเมินความถูกต้องของการทดสอบ ผลการศึกษาพบว่า

(1) วิธีการประมาณค่าแบบโมดัลเบย์มีประสิทธิภาพสูงสุด วิธีการประมาณค่าของเบย์มีความเชื่อมั่นสูง แต่วิธีการประมาณค่าความเป็นไปได้สูงสุดไม่ค่อยคงเส้นคงวา

(2) ขนาดของคลังข้อสอบมีผลน้อยมาก ยกเว้นกลุ่มผู้สอบที่มีความสามารถต่ำและใช้ข้อสอบเริ่มต้นที่มีค่าความยากง่ายสูงกว่าความสามารถจริง

(3) การประมาณค่าความสามารถจะมีความถูกต้องมากยิ่งขึ้น ถ้าข้อสอบที่ใช้เริ่มต้นสอบนั้นมีค่าความยากง่ายเท่ากับหรือต่ำกว่าความสามารถจริงของผู้สอบ

(4) มีปฏิสัมพันธ์ระหว่างชนิดของคลังข้อสอบกับวิธีการประมาณค่าความสามารถและชนิดของคลังข้อสอบกับระดับความสามารถของผู้สอบ

(5) มีความแตกต่างกันน้อยมากระหว่างคุณลักษณะทางสถิติของข้อสอบที่ไม่เคยถูกเลือกใช้กับข้อสอบที่ถูกเลือกใช้บ่อย

แฮนกินส์ (Hankins. 1986) ได้ศึกษาผลของข้อสอบที่ใช้เมื่อเริ่มต้นสอบต่อค่าความลำเอียงและสารสนเทศของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ โดยประมาณค่าความสามารถของผู้สอบด้วยวิธีของเบส์ ข้อสอบที่ใช้เริ่มต้นในการทดสอบนั้นมี 2 ลักษณะ คือ กำหนดแน่นอน (fixed entry procedure) และกำหนดแบบแปรผัน (variable entry procedure) ผลการศึกษาพบว่า การใช้ข้อสอบเริ่มต้นทั้งสองลักษณะดังกล่าวไม่ก่อให้เกิดความลำเอียงเมื่อผู้สอบมีความสามารถปานกลาง แต่จะมีความลำเอียงถ้าผู้สอบมีความสามารถสูงและต่ำมากๆ โค้งสารสนเทศเป็นรูปสมมาตรแบบโด่ง (peaked and asymmetric) ทั้งสองวิธี โค้งประสิทธิภาพสัมพัทธ์แสดงให้เห็นว่าการทดสอบด้วยวิธีการกำหนดข้อสอบเริ่มต้นที่ต่างกันมีประสิทธิภาพเท่าเทียมกัน แต่ทั้ง 2 วิธีเลือกข้อสอบได้ต่างกัน วิธีการใช้ข้อสอบเริ่มต้นแบบแปรผันนั้นใช้ข้อสอบมากกว่า

แบล็คมอร์ (Blackmore. 1986) ได้ศึกษาเปรียบเทียบการทดสอบที่ดำเนินการสอบด้วยคอมพิวเตอร์ (computerized) การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ (computerized adaptive) และการทดสอบแบบเดิม (conventional) โดยทดลองกับนักเรียนเกรด 12 โรงเรียนมัธยมการ์เดนโกรฟ ในแคลิฟอร์เนีย จำนวน 24 คนในปี 1986 โดยสุ่มเป็น 3 กลุ่มเพื่อรับวิธีดำเนินการสอบแต่ละวิธี แบบทดสอบวัดความถนัดซึ่งมี 6 ฉบับย่อย ได้แก่ การให้เหตุผลทางภาษา (verbal reasoning) ความสามารถทางจำนวน (numerical) การให้เหตุผลเชิงนามธรรม (abstract reasoning) การให้เหตุผลเชิงกล (mechanical reasoning) มิติสัมพันธ์ (space relation) และการใช้ภาษา (language usage) ถูกใช้ในการสอบแต่ละกลุ่มวิเคราะห์ข้อมูลโดยใช้การทดสอบความแปรปรวนและการทดสอบแม็คเนมาร์ (McNemar test) ผลการศึกษาพบว่า คะแนนผลการสอบแบบทดสอบความถนัดจากการดำเนินการสอบ 3 วิธีไม่แตกต่างกัน แต่จะต่างกันหากพิจารณาตามคุณลักษณะของผู้เข้าสอบ เช่น ชาติพันธุ์และเพศ

ลี (Lee. 1986) ได้ศึกษาความเป็นไปได้ของการนำการทดสอบแบบปรับเหมาะกับความ
ความสามารถของผู้สอบด้วยคอมพิวเตอร์มาใช้ในการวัดผลสัมฤทธิ์ทางการเรียนโดยสร้างแบบ
ทดสอบ เอ็ม เอ เอ ที (microcomputerized adaptive achievement test , MAAT) จากคลังข้อสอบที่
สร้างโดยแผนกคลังข้อสอบของ พี อี ที เอ (pitt educational testing aids , PETA) กลุ่มตัวอย่างได้แก่
นิสิตระดับบัณฑิตศึกษาจำนวน 3 ห้องเรียนที่กำลังเรียนวิจัยการศึกษาเบื้องต้น วิธีการทดสอบ
แบบปรับเหมาะ 2 วิธี ได้แก่ วิธีเอส พี อาร์ ทีของวอลท์ (Wald's sequential probability ratio test)
และแบบเฟลิกซ์เลเวลของลอร์ด (Lord s flexilevel) นอกจากนั้นยังมีการวัดเจตคติของผู้สอบต่อ
การทดสอบและสัมภาษณ์ผู้ดำเนินการสอบด้วย ผลการศึกษาพบว่า มีความเป็นไปได้ที่จะนำการ
ทดสอบแบบ เอ็ม เอ เอ ที มาใช้วัดผลสัมฤทธิ์ทางการเรียน คะแนนจากการทดสอบด้วย
คอมพิวเตอร์มีความสัมพันธ์ทางบวกกับผลการทดสอบกลางเทอม ผู้สอบมีเจตคติที่ดีต่อการ
ทดสอบแบบ เอ็ม เอ เอ ที

วิสโนวสกี (Wisniewski. 1985) ได้ประยุกต์ใช้ราล์ซโมเดลในการทดสอบแบบ
ปรับเหมาะกับความความสามารถของผู้สอบด้วยคอมพิวเตอร์ โดยใช้วิธีการค้นหาแบบแบ่งสอง (binary
search method , BSM) โดยมีวัตถุประสงค์เพื่อตรวจสอบความถูกต้องแม่นยำของการประมาณค่า
ความสามารถของผู้สอบ เจตคติของผู้สอบต่อการสอบแบบปรับเหมาะกับการทดสอบแบบเดิม
และศึกษาความสัมพันธ์ระหว่างเวลาที่ใช้ในการทดสอบกับค่าความยากง่ายของข้อสอบ วิเคราะห์
ข้อมูลด้วยสถิติการถดถอยพหุคูณ ผลการวิจัย พบว่า เวลาที่ใช้ในการตอบข้อสอบในการทดสอบ
ด้วยคอมพิวเตอร์ไม่มีความสัมพันธ์กับวิธีการประมาณค่าความสามารถของผู้สอบ การประมาณ
ค่าความสามารถของผู้สอบที่ใช้วิธีการค้นหาแบบแบ่งสองเป็นตัวทำนายผลการทดสอบแบบเดิมที่
ดี การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ดึงดูดความสนใจของผู้สอบได้ดีและสามารถ
ควบคุมสภาพแวดล้อมของการทดสอบได้ดีกว่าการทดสอบแบบเดิม

อyalalaและคนอื่นๆ (Alaya and others. 1990) ได้เปรียบเทียบการทดสอบโดยใช้
แบบทดสอบเฟลิกซ์เลเวลกับการทดสอบแบบปรับเหมาะกับความความสามารถของผู้สอบด้วย
คอมพิวเตอร์ที่ประมาณค่าความสามารถด้วยวิธีของเบส์ ใช้ค่าความคลาดเคลื่อนมาตรฐานในการ
ประมาณค่าความสามารถเป็นเกณฑ์ยุติการสอบ โดยเปรียบเทียบที่จุดต่างๆกันของค่าของ
ความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ ผลจากการศึกษา พบว่า ค่าความ
สามารถของผู้สอบที่ได้จากการทดสอบสองวิธีมีความถูกต้องแม่นยำไม่ต่างกัน

แกรริสันและบวมการ์เตน (Garrison and Baumgarten. 1986) ได้ทดลองใช้การ
ทดสอบแบบปรับเหมาะกับความความสามารถของผู้สอบด้วยคอมพิวเตอร์ เพื่อวัดทักษะคณิตศาสตร์

ของนักศึกษา จำนวน 60 คน โดยมีการเปรียบเทียบความคงเส้นคงวาของค่าความสามารถที่ประมาณได้จากการทดสอบแบบเดิมและการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ ผลจากการศึกษา พบว่า การทดสอบแบบปรับเหมาะดังกล่าวมีประสิทธิภาพในการวัดสูงกว่าการทดสอบแบบเดิม ความยาวของแบบทดสอบลดลงจากเดิม ผู้สอบมีเจตคติที่ดีต่อการสอบสูงกว่าการทดสอบแบบเดิม

อลายาและโคช (Ayala and Koch, 1987) ได้เปรียบเทียบการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ ระหว่างโมเดลการตอบสนองแบบนามบัญญัติ (Nominal response model) กับโมเดลโลจิสติก 3 พารามิเตอร์ โดยศึกษาจากข้อมูลที่จำลองขึ้น ผลจากการศึกษา พบว่า ความสามารถที่ประมาณได้จากการทดสอบแบบปรับเหมาะดังกล่าวทั้ง 2 โมเดลมีความสัมพันธ์กันสูง

โมรีโนและคนอื่นๆ (Moreno and others, 1983) ได้ศึกษาความสัมพันธ์ระหว่างการเลือกแบบทดสอบย่อยจากชุดของข้อสอบ ASVAB (armed services vocational aptitude battery) ดำเนินการทดสอบแบบเดิมและการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ โดยใช้กลุ่มตัวอย่างจาก Marine recruits ผลการศึกษา พบว่า คะแนนจากการดำเนินการสอบแบบทดสอบย่อย ทั้งสองวิธีการมีความสัมพันธ์กัน แม้ว่าการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จะใช้ข้อสอบเพียงครั้งเดียวก็ตาม และจากผลการวิเคราะห์องค์ประกอบ พบว่า แบบทดสอบย่อยที่ดำเนินการสอบทั้งสองวิธีการนั้นวัดองค์ประกอบเดียวกัน

เวทเซล , ดักลาส และ แม็คไบรด์ (Wetzel, Douglas and McBride, 1983) ได้ศึกษาผลของความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ของข้อสอบ โดยใช้กลยุทธ์ในการเลือกข้อสอบต่างวิธีกัน ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์และการทดสอบแบบเดิม โดยศึกษาจากข้อมูลที่จำลองขึ้น โดยจำลองข้อมูลให้มี ความคลาดเคลื่อนมากและน้อยกว่าความเป็นจริง ผลการศึกษา พบว่า ความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ของข้อสอบมีผลน้อยมากต่อประสิทธิผล (efficiency) และความถูกต้องในการวัด (precision)

เบอร์กสตรอมและสตาห์ล (Bergstrom and Stahl, 1992) ได้ศึกษาขนาดที่พอเหมาะของคลังข้อสอบที่จะทำให้การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์มีประสิทธิภาพที่ระดับต่างๆของความสามารถของผู้สอบ ตัวแปรที่ศึกษา คือ ลักษณะ

เฉพาะของเนื้อหา ความยาวของแบบทดสอบและเกณฑ์ยุติการสอบ ผลจากการศึกษา พบว่า ขนาดของคลังข้อสอบที่พอเพียงนั้นขึ้นอยู่กับเกณฑ์ยุติการสอบ

คาร์ลสัน (Carlson. 1993) ได้ศึกษาประสิทธิภาพของการทดสอบแบบปรับเหมาะ กับความสามารถของผู้สอบด้วยคอมพิวเตอร์ เมื่อใช้วิธีการเลือกข้อสอบต่างกัน 4 วิธี เกณฑ์ยุติ การสอบ 2 เกณฑ์คือ ตอบข้อสอบครบ 20 ข้อและค่าความคลาดเคลื่อนมาตรฐานในการประมาณ ค่าความสามารถน้อยกว่า .40 ประมาณค่าความสามารถของผู้สอบโดยใช้กลยุทธ์ GSSS ของ เขียวโอ (Xiao's golden section search strategy) ดำเนินการสอบโดยใช้โปรแกรมที่เขียนโดยภาษา ฟอรัทเรน วิเคราะห์ข้อมูลโดยใช้สถิติสหสัมพันธ์ของเพียร์สันและเคนดอลล์เทาบี (pearson and kendall tau-b correlation) ผลจากการศึกษา พบว่า ผลจากการทดสอบแบบปรับเหมาะกับความ สามารถของผู้สอบด้วยคอมพิวเตอร์กับคะแนนรวมทั้งฉบับมีความสัมพันธ์กันสูงมาก วิธีการเลือก ข้อสอบจากคลังข้อสอบทั้ง 4 วิธีให้ผลไม่ต่างกัน อย่างไรก็ตามการเลือกข้อสอบด้วยวิธีใช้ค่าสาร สนเทศของข้อสอบที่สูงสุด (maximum information) ดูเหมือนจะดีกว่าวิธีอื่นๆ

แฮนเซน (Larkin and Weiss. 1975 ; citing Hansen. 1969) ได้ทดลองใช้แบบ ทดสอบพีระมิด 5 ชนิด ที่มีจำนวนชั้น 3 หรือ 4 ชั้นขึ้นอยู่กับคำตอบที่ตอบ โดยจะตอบทั้งหมด 17 ข้อ ขนาดชั้นคงที่คือค่าความยากง่ายต่างกัน .10 ให้คะแนนผลการสอบ 4 วิธี ผลจากการศึกษา พบว่า คะแนนจากแบบทดสอบพีระมิดมีความสัมพันธ์กับคะแนนที่ได้จากการทดสอบในชั้นเรียนที่ ใช้เวลาทดสอบ 1 ชั่วโมงโดยวัดเนื้อหาเดียวกันก่อนใช้แบบทดสอบพีระมิด 1 สัปดาห์ นอกจากนั้น ยังมีความสัมพันธ์กับระดับผลการเรียนอีกด้วย แบบทดสอบพีระมิดมีความคงเส้นคงวาในการวัด (internal consistency) สูงกว่าการทดสอบแบบเดิม การกระจายของคะแนนที่ได้จากแบบทดสอบปี ระมิดมีลักษณะเป็นสี่เหลี่ยมผืนผ้า (rectangular) ส่วนการทดสอบแบบเดิมนั้นคะแนนจะกระจาย แบบเบ้ทางลบ (negative skew)

วูด (Larkin and Weiss. 1975 ; citing Wood. 1969) ได้ศึกษาความสัมพันธ์ระหว่าง ผลจากการทดสอบจากแบบทดสอบพีระมิดกับการทดสอบแบบเดิม โดยใช้แบบทดสอบพีระมิด ชนิด 4 5 และ 6 ชั้น รูปแบบการแยกทางแบบถูกเพิ่ม 1 ผิดลด 1 ขนาดชั้นคงที่ ระยะห่างของความ ยากง่ายเท่ากับ .05 ให้คะแนนตามจำนวนข้อที่ทำถูก โดยทดสอบกับกลุ่มตัวอย่าง 91 คน ผลจาก การศึกษา พบว่า สหสัมพันธ์ระหว่างคะแนนรวมทั้ง 3 ฉบับจากแบบทดสอบพีระมิดกับระดับผล การเรียนเฉลี่ย เท่ากับ .51 หากใช้แบบทดสอบแบบเดิม 46 ข้อแล้วจะมีความสัมพันธ์กับระดับผล การเรียน เท่ากับ .68

วอเตอร์ (Larkin and Weiss. 1975 ; citing Water. 1964) ได้เปรียบเทียบแบบทดสอบพีระมิดชนิด 5 ชั้น ใช้รูปแบบการแยกทางแบบคงที่ ระยะห่างของค่าความยากง่าย เท่ากับ .01 แยกทางแบบถูกเพิ่ม 1 ผิดลด 1 กับแบบทดสอบแบบเดิม 4 ชุด ชุดละ 5 ข้อ ผลจากการศึกษา พบว่า คะแนนจากการทดสอบโดยใช้แบบทดสอบพีระมิดจะมีความสัมพันธ์กับความสามารถของผู้สอบสูงกว่าการทดสอบแบบเดิม และการกระจายของคะแนนมีลักษณะเป็นรูปสี่เหลี่ยมผืนผ้าและกระจายมากกว่าผลจากการทดสอบแบบเดิม

เบย์รอฟและซีเลย์ (Larkin and Weiss. 1975 ; citing Bayroff and Seeley. 1967) ได้หาความสัมพันธ์ของผลการทดสอบโดยใช้แบบทดสอบพีระมิดและการทดสอบแบบเดิม โดยใช้แบบทดสอบพีระมิดชนิด 8 ชั้น 2 ฉบับ ใช้ผู้สอบ 102 คน ดำเนินการสอบด้วยคอมพิวเตอร์ และผู้สอบต้องตอบแบบทดสอบแบบเดิมอีก 2 ฉบับ วัดด้านภาษา 50 ข้อ ด้านจำนวน 40 ข้อ ผลจากการศึกษา พบว่า สหสัมพันธ์ของผลการทดสอบจากแบบทดสอบทั้งสองลักษณะ มีค่าเท่ากับ 0.87 และ 0.79 และเมื่อใช้ข้อสอบจำนวนเท่ากัน การทดสอบด้วยแบบทดสอบพีระมิดจะมีค่าสหสัมพันธ์สูงกว่าการทดสอบแบบเดิม

ลันซ์และเบอร์กสตรอม (Lunz and Bergstrom. 1994) ได้ศึกษารูปแบบของการดำเนินการสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ต่อผลการสอบรูปแบบดังกล่าว ได้แก่ ค่าความยากง่ายของข้อสอบเริ่มต้นต่างกัน 3 ระดับ (ง่าย ปานกลาง ยาก) เป้าหมายของความยากง่ายของแบบทดสอบต่างกัน (โอกาสในการตอบถูกร้อยละ 50 60 70) ความยาวของแบบทดสอบ (50 , 100 , 240 ข้อ) และโอกาสในการควบคุมแบบทดสอบ (ข้ามข้อ , ทบทวนข้อเดิม , หน่วงเวลาในการตอบ และ ไม่มีเงื่อนไข) กลุ่มตัวอย่าง ได้แก่ นิสิตเทคนิคการแพทย์ วิเคราะห์ข้อมูลด้วยสถิติการวิเคราะห์ความแปรปรวนร่วม โดยใช้ผลการทดสอบแบบเดิมเป็นตัวแปรร่วม ผลจากการศึกษา พบว่า ผลการทดสอบจะต่างกันเมื่อโอกาสในการควบคุมแบบทดสอบต่างกัน ปฏิสัมพันธ์ของตัวแปรอิสระทั้งหลายไม่ส่งผลต่อตัวแปรตาม

นันทียา พิงคำ (2531) ได้เปรียบเทียบคุณภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ กับการทดสอบแบบเดิม กลุ่มตัวอย่างได้แก่นักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 120 คน แบบทดสอบแบบเดิมมีข้อสอบ 40 ข้อ ส่วนข้อสอบในคลังข้อสอบที่ใช้สำหรับการทดสอบแบบปรับเหมาะนั้น มี 361 ข้อ ข้อสอบทั้งสองลักษณะวัดความสามารถด้านคำศัพท์ภาษาอังกฤษ ผลจากการศึกษา พบว่า การทดสอบทั้งสองลักษณะมีความ

เพียงตรงเชิงเงื่อนไขสัมพันธ์ไม่ต่างกัน โดยที่การทดสอบแบบปรับเหมาะใช้ข้อสอบน้อยกว่าครึ่งหนึ่งของการทดสอบแบบเดิม นอกจากนั้นยังมีค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าต่ำกว่าการทดสอบแบบเดิมอีกด้วย

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการวิจัยเชิงทดลอง เพื่อศึกษาผลที่เกิดขึ้นจากตัวแปรอิสระ 5 ตัวแปร ได้แก่ ค่าความยากง่ายของข้อสอบข้อแรกที่ใช้เมื่อเริ่มต้นในการทดสอบ เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบ และความสามารถของผู้สอบที่มีต่อ ค่าความเที่ยงตรงเชิงสภาพ และ จำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ ซึ่งผู้วิจัยแบ่งการดำเนินการวิจัยออกเป็น 2 ระยะ กล่าวคือ ระยะแรกเป็นการสร้างคลังข้อสอบ เรื่อง เศษส่วน ชั้นประถมศึกษาปีที่ 6 ระยะที่สองเป็นการศึกษาผลของตัวแปรอิสระต่อตัวแปรตามดังกล่าวข้างต้น โดยมีวิธีดำเนินการวิจัยในแต่ละระยะ ดังนี้

การดำเนินการวิจัย ระยะที่ 1 การสร้างคลังข้อสอบ เรื่อง เศษส่วน ประชากรและกลุ่มตัวอย่าง

ประชากรสำหรับสร้างคลังข้อสอบครั้งนี้ ได้แก่ นักเรียนชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2537 ในโรงเรียนประถมศึกษาสังกัดสำนักงานการประถมศึกษากรุงเทพมหานคร ซึ่งมี 38 โรงเรียน จำนวน 7,531 คน โรงเรียนสังกัดกรุงเทพมหานคร 427 โรงเรียน จำนวน 37,994 คน รวมทั้งสิ้น 45,525 คน

กลุ่มตัวอย่างได้มาจากกลุ่มประชากร ด้วยการสุ่มแบบหลายขั้นตอน (multi stage random sampling) ซึ่งผู้วิจัยใช้กลุ่มตัวอย่างเพื่อสร้างคลังข้อสอบ จำนวน 8,720 คน โดยมีขั้นตอนในการสุ่ม ดังนี้

ขั้นตอนแรก สุ่มโรงเรียนด้วยวิธีการสุ่มแบบแบ่งชั้น (stratified random sampling) โดยใช้สังกัดและขนาดของโรงเรียนเป็นขั้นของการสุ่ม ซึ่งจำแนกเป็น 2 สังกัด สังกัดละ 3 ขนาด ได้แก่ ขนาดใหญ่ (นักเรียนมากกว่า 800 คน) ขนาดกลาง (นักเรียนระหว่าง 401-800 คน) และขนาดเล็ก (นักเรียนน้อยกว่า 401 คน) สุ่มได้โรงเรียนทั้งหมด 120 โรงเรียน เป็นขนาดใหญ่ 40 โรงเรียน ขนาดกลาง 28 โรงเรียนและขนาดเล็ก 52 โรงเรียน

ขั้นตอนที่สอง สุ่มห้องเรียนจากโรงเรียนที่สุ่มได้ในขั้นตอนแรก โดยใช้วิธีการสุ่มอย่างง่าย (simple random sampling) ซึ่งสุ่มห้องเรียนจากโรงเรียนขนาดใหญ่ ขนาดกลางและขนาดเล็ก

โรงเรียนละ 4 ห้อง 3 ห้อง และ 1 ห้องตามลำดับ ได้ห้องเรียนทั้งหมด 296 ห้อง และเนื่องจากมีข้อจำกัดในเรื่องของการสุ่มนักเรียนจากห้องเรียนเป็นรายบุคคล ผู้วิจัยจึงใช้นักเรียนทุกคนในแต่ละห้องเรียนที่สุ่มได้เป็นกลุ่มตัวอย่าง ซึ่งได้นักเรียนกลุ่มตัวอย่างทั้งหมด 8,720 คน จำแนกตามสังกัดและขนาดของโรงเรียนได้ดังตาราง 3

ตาราง 3 จำนวนประชากรและกลุ่มตัวอย่างที่ใช้สำหรับสร้างคลังข้อสอบ

ขนาด โรงเรียน	ส.ป.กทม.		กทม.		รวม	
	ประชากร	กลุ่ม ตัวอย่าง	ประชากร	กลุ่ม ตัวอย่าง	ประชากร	กลุ่ม ตัวอย่าง
ขนาดเล็ก	134	26	8059	1543	8193	1569
ขนาดกลาง	855	164	12023	2303	12878	2467
ขนาดใหญ่	6542	1253	17912	3431	24454	4684
รวม	7531	1443	37994	7277	45525	8720

เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้สำหรับสร้างคลังข้อสอบ คือ แบบทดสอบคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 เรื่อง เศษส่วน ซึ่งผู้วิจัยได้สร้างขึ้นรวมทั้งสิ้น 290 ข้อ จัดพิมพ์ข้อสอบเป็น 7 ฉบับ ฉบับละ 50 ข้อ (แต่ละฉบับมีข้อสอบต่างจากฉบับอื่นอยู่ 40 ข้อและมีข้อสอบที่เหมือนกันอยู่ 10 ข้อทุกฉบับซึ่งใช้เป็นข้อสอบรวม) ผู้วิจัยสร้างข้อสอบดังกล่าวให้เป็นข้อสอบคู่ขนาน โดยสร้างจากฟ้าขีดของข้อสอบชุดเดียวกัน สาเหตุที่ต้องมีการสร้างแบบทดสอบหลายฉบับ ก็เนื่องจากผู้วิจัยต้องการให้มีข้อสอบในคลังข้อสอบจำนวนมากๆ มีความยากง่ายของข้อสอบเหมาะกับผู้สอบที่มีความสามารถทุกระดับตั้งแต่ความสามารถต่ำไปจนถึงความสามารถสูง แต่เนื่องจากกลุ่มตัวอย่างที่ใช้ในการสร้างคลังข้อสอบเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 จึงมีศักยภาพไม่สูงมากนักในการตอบข้อสอบจำนวนมากข้อในแต่ละครั้ง ผู้วิจัยจึงได้แยกข้อสอบออกเป็นฉบับย่อยๆ 7 ฉบับ ดังกล่าวเมื่อนำไปทดลองใช้แล้ว จึงนำค่าพารามิเตอร์ของข้อสอบแต่ละฉบับมาปรับให้อยู่บนสเกลเดียวกัน ข้อสอบทั้งหมดเป็นชนิดปรนัย 4 ตัวเลือก ความเป็นมิติเดียว ใช้วัดสมรรถภาพความเข้าใจ เรื่อง เศษส่วน ข้อสอบที่มีคุณภาพจะถูกคัดเลือกเก็บไว้ในลักษณะของคลังข้อสอบ เพื่อให้สำหรับการ

ทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ ซึ่งสัดส่วนของจำนวนข้อสอบจำแนกตามรายเนื้อหาย่อยมีดังตาราง 4

ตาราง 4 โครงสร้างของข้อสอบวัดความเข้าใจ เรื่องเศษส่วน ชั้นประถมศึกษาปีที่ 6 จำแนกตามเนื้อหาย่อย ฉบับที่ 1 - 7

ที่	เนื้อหาย่อย	ข้อที่	หมายเหตุ
1	เศษส่วนของจำนวนนับ	1 , 2	* หมายถึงข้อสอบรวม
2	การเปรียบเทียบเศษส่วน	3 , 4 , 5 , 6* , 7*	
3	การบวกเศษส่วน	8 , 9 , 10	
4	การลบเศษส่วน	11* , 12 , 13 , 14	
5	การคูณเศษส่วน	15* , 16 , 17 , 18	
6	การหารเศษส่วน	19* , 20 , 21 , 22	
7	การแปลงเศษเกินเป็นจำนวนคละ	23* , 24 , 25	
8	ชนิดของเศษส่วน	26 , 27	
9	การนำเศษส่วนไปใช้ในชีวิตประจำวัน	28* , 29 , 30 , 31 , 32 , 33*	
10	เศษซ้อน	34 , 35 , 36	
11	เศษส่วนที่มีค่าเท่ากัน	37 , 38 , 39	
12	ความหมายของเศษส่วน	40 , 41 , 42	
13	เศษส่วนอย่างต่ำ	43* , 44 , 45	
14	การบวกลบคูณหารระคน	46 , 47 , 48 , 49 , 50*	

การสร้างเครื่องมือที่ใช้ในการวิจัย

การสร้างเครื่องมือวิจัยในระยะที่ 1 นี้ ผู้วิจัยได้สร้างแบบทดสอบคณิตศาสตร์ เรื่อง เศษส่วน ซึ่งมีขั้นตอนในการสร้าง ดังนี้

1. ศึกษาจุดประสงค์ เนื้อหาและเวลาของการเรียนการสอนคณิตศาสตร์ระดับประถมศึกษาจากหลักสูตรประถมศึกษา พุทธศักราช 2521 และคู่มือครูคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ของสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี กระทรวงศึกษาธิการ

2. ศึกษาสมรรถภาพที่ใช้วัดทางคณิตศาสตร์ ซึ่งมี 4 สมรรถภาพ ได้แก่ ความรู้ความเข้าใจ ทักษะการคิดคำนวณ การแก้โจทย์ปัญหาและการปฏิบัติงานทางคณิตศาสตร์ ซึ่งผู้วิจัยเลือกวัดสมรรถภาพความเข้าใจเรื่องเศษส่วน เนื่องจากเป็นสมรรถภาพและเนื้อหาที่ค่อนข้างเป็นปัญหาสำหรับนักเรียน

3. กำหนดลักษณะของเครื่องมือวัด ซึ่งผู้วิจัยเลือกใช้เลือกใช้รูปแบบของข้อสอบแบบปรนัยชนิดเลือกตอบ 4 ตัวเลือก ให้คะแนนระบบ 0 , 1

4. กำหนดวัตถุประสงค์ย่อยสำหรับใช้วัดสมรรถภาพความรู้ความเข้าใจ เรื่อง เศษส่วน

5. สร้างฟาเซต (facet) ของคำถามและตัวเลือก เพื่อใช้เป็นกรอบในการสร้างข้อสอบวัดความเข้าใจ เรื่อง เศษส่วน ซึ่งแสดงไว้ในภาคผนวก

4. สร้างข้อสอบเพื่อวัดสมรรถภาพความเข้าใจ เรื่อง เศษส่วน จำนวน 290 ข้อ โดยแยกออกเป็น 7 ฉบับ ฉบับละ 50 ข้อ แต่ละฉบับจะมีลักษณะใกล้เคียงกันโดยสร้างจากฟาเซต(facet) ของคำถามคำตอบชุดเดียวกัน ซึ่งแต่ละฉบับมีข้อสอบที่ต่างจากฉบับอื่นอยู่ 40 ข้อและ เหมือนกันทุกฉบับอยู่ 10 ข้อ ข้อสอบที่เหมือนกันทุกฉบับนี้จะถูกใช้เป็นข้อสอบร่วม (anchor item) สำหรับเทียบคะแนนของข้อสอบแต่ละฉบับให้อยู่บนสเกลเดียวกัน ซึ่งจำนวนข้อสอบร่วมที่ใช้ในทุกฉบับนั้น ผู้วิจัยได้ใช้จำนวนข้อสอบร่วมตามข้อเสนอนแนะของลอร์ด (Lord) ที่แนะนำว่าควรใช้ข้อสอบร่วมอย่างน้อยร้อยละ 20 ของข้อสอบทั้งหมดในฉบับ

5. นำข้อสอบไปให้ผู้เชี่ยวชาญด้านหลักสูตร การสอน การวัดและประเมินผลการศึกษา คณิตศาสตร์ตรวจสอบความเที่ยงตรงตามเนื้อหาและเที่ยงตรงเชิงโครงสร้าง โดยใช้เทคนิคของโรวินेलลี และ แฮมเบิลตัน (Rovinelli and Hambleton) กำหนดเกณฑ์ในการพิจารณา 3 เกณฑ์ คือ ตรง ไม่ตรงและไม่แน่ใจว่าตรงตามเนื้อหาและโครงสร้างที่ใช้วัดหรือไม่ ซึ่งวัดเนื้อหา เรื่องเศษส่วน วัดโครงสร้าง ความเข้าใจ

6. ปรับปรุงข้อสอบตามคำแนะนำของผู้เชี่ยวชาญ ซึ่งส่วนใหญ่แล้วจะปรับปรุงเกี่ยวกับภาษาที่ใช้ และรูปที่ใช้แทนเศษส่วน เพื่อให้สื่อความหมายได้ตรงตามสิ่งของเรื่องเศษส่วน

การเก็บรวบรวมข้อมูล

ในการเก็บรวบรวมข้อมูลเพื่อสร้างคลังข้อสอบ เรื่อง เศษส่วน ได้ดำเนินการ ดังนี้

1. นำหนังสือจากบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตรไปให้สำนักงานการประถมศึกษากรุงเทพมหานคร และ สำนักงานการศึกษากรุงเทพมหานคร เพื่อขอความร่วมมือในการวิจัย โดยขอทดสอบนักเรียนชั้นประถมศึกษาปีที่ 6 ในโรงเรียนที่ได้รับการสุ่มเป็น

กลุ่มตัวอย่างเพื่อทดลองเครื่องมือวิจัยที่ผู้วิจัยสร้างขึ้น

2. ประสานงานกับสำนักงานการประถมศึกษากรุงเทพมหานคร และ สำนักงานการศึกษา กรุงเทพมหานคร รวมทั้งโรงเรียนในสังกัดที่เป็นกลุ่มตัวอย่าง เพื่อชี้แจงวัตถุประสงค์ของการวิจัย และได้นัดหมาย วัน เวลาเพื่อทดสอบนักเรียน

3. ทดสอบนักเรียนชั้นประถมศึกษาปีที่ 6 ในโรงเรียนกลุ่มตัวอย่าง ซึ่งใช้วิธีการแจกแบบทดสอบให้แก่นักเรียนในห้องเรียนแบบเรียงสลับแบบเป็นระบบ กล่าวคือ คนที่ 1-7 จะได้รับแบบทดสอบฉบับที่ 1-7 ตามลำดับ คนที่ 8-14 จะได้รับแบบทดสอบฉบับที่ 1-7 ตามลำดับ การแจกแบบทดสอบจะกระทำเช่นนี้ไปเรื่อยๆจนครบนักเรียนทุกคนที่มีอยู่ในห้อง

การวิเคราะห์ข้อมูล

เมื่อได้ผลการตอบข้อสอบของนักเรียนครบตามจำนวนที่ต้องการแล้ว ผู้วิจัยได้นำ ข้อมูลดังกล่าวมาวิเคราะห์เพื่อหาคุณภาพของข้อสอบ โดยวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ ซึ่งมีขั้นตอน ดังนี้

1. ตรวจสอบความเป็นมิติเดียว ด้วยการวิเคราะห์องค์ประกอบ (factor analysis) ซึ่งหมุนแกนโดยวิธีแวนแมก (varimax) สกัดตัวประกอบโดยวิธีองค์ประกอบหลัก (principle component) พิจารณาค่าไอเกนของตัวประกอบที่ได้จากการวิเคราะห์ ถ้าหากตัวประกอบที่ 1 มีค่าไอเกนสูงกว่าค่าไอเกนของตัวประกอบที่ 2 ตั้งแต่ 3 เท่าขึ้นไปและค่าไอเกนของตัวประกอบ 2 มีค่าใกล้เคียงกับค่าไอเกนของตัวประกอบที่เหลือ หรือสามารถอธิบายความแปรปรวนร่วมได้ร้อยละ 70 ขึ้นไป แสดงว่า ข้อสอบดังกล่าวมีความเป็นมิติเดียว

2. วิเคราะห์คุณภาพของข้อสอบเป็นรายฉบับทั้ง 7 ฉบับ เพื่อหาค่าพารามิเตอร์ของข้อสอบ คือ ค่าอำนาจจำแนก ค่าความยากง่าย และ ค่าการเดาตามทฤษฎีการตอบสนองข้อสอบ โดยใช้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ การหาค่าพารามิเตอร์ของข้อสอบดังกล่าวนั้น วิเคราะห์ด้วยโปรแกรม BILOG 3

3. แปลงค่าพารามิเตอร์ของข้อสอบแต่ละฉบับที่ได้จากการวิเคราะห์ ให้อยู่บนสเกลเดียวกัน (common scale) โดยเทียบคะแนนตามแบบแผน การใช้ข้อสอบร่วมกัน (anchor item design) ซึ่งหาค่าคงที่สำหรับการแปลงคะแนนด้วยวิธีค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐาน (mean and sigma method) ซึ่งมีขั้นตอนในการแปลงค่าพารามิเตอร์ ดังนี้

3.1 ประมาณค่าพารามิเตอร์ของข้อสอบตามทฤษฎีการตอบสนองข้อสอบ โดยประมาณค่าพารามิเตอร์ดังกล่าวแยกกันเป็นรายฉบับ 7 ฉบับ โดยกำหนดค่าความสามารถของ

ผู้สอบให้เป็นค่าใดค่าหนึ่ง ซึ่งผู้วิจัยกำหนดให้ค่าความสามารถของผู้สอบมีค่าเฉลี่ยเท่ากับ 0 และค่าเบี่ยงเบนมาตรฐานเท่ากับ 1

3.2 หาค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของข้อสอบที่ใช้เป็นข้อสอบร่วมจากทุกฉบับ

3.3 ปรับค่าอำนาจจำแนกและค่าความยากง่ายของข้อสอบจากฉบับหนึ่งไปสู่อีกฉบับหนึ่ง โดยใช้สูตร

$$y = \alpha x + \beta$$

เมื่อ

$$\alpha = \frac{SD_{by}}{SD_{bx}}$$

$$\beta = \overline{b_y} - \frac{SD_{by}}{SD_{bx}} \overline{b_x}$$

$$a_y = \frac{a_x}{\alpha}$$

$$b_y = \alpha b_x + \beta$$

a_y, a_x แทน ค่าอำนาจจำแนกที่ได้รับการแปลงค่าและยังไม่ได้รับการแปลงค่า

b_y, b_x แทน ค่าความยากง่ายที่ได้รับการแปลงค่าและยังไม่ได้รับการแปลงค่า

$\overline{b_y}, \overline{b_x}$ แทน ค่าความยากง่ายเฉลี่ยที่ได้รับการแปลงค่าและยังไม่ได้รับการแปลงค่า

SD_{by}, SD_{bx} แทน ค่าเบี่ยงเบนมาตรฐานที่ได้รับการแปลงค่าและยังไม่ได้รับการแปลงค่า

3.4 หาค่าเฉลี่ยของความยากง่ายและค่าอำนาจจำแนกของข้อสอบรวมทั้ง 2 ฉบับ หลังจากปรับสเกลแล้ว ทั้งนี้ เนื่องจากเป็นข้อสอบข้อเดียวกัน ดังนั้นจึงควรมีค่าพารามิเตอร์ค่าเดียวกัน

4. คัดเลือกข้อสอบที่มีคุณภาพตามเกณฑ์ของยูริ (Urry, 1977) ซึ่งเป็นข้อสอบที่มี ค่าอำนาจจำแนก ตั้งแต่ 0.8 ขึ้นไป ค่าความยากง่าย อยู่ในช่วง -2.0 ถึง 2.0 ค่าการเดาน้อย กว่า 0.3 ซึ่งได้ข้อสอบทั้งหมด 150 ข้อ อย่างไรก็ตามเพื่อให้ได้ข้อสอบที่มีค่าความยากง่ายครอบคลุมช่วง

ความสามารถของผู้สอบที่เป็นกลุ่มตัวอย่างในการวิจัย ดังนั้นผู้วิจัยจึงเลือกข้อสอบเพิ่มอีก 50 ข้อ ซึ่งข้อสอบที่คัดเลือกเพิ่มนี้ มีค่าอำนาจจำแนก ตั้งแต่ 0.4 ขึ้นไป ค่าความยากง่ายอยู่ในช่วง-3.0 ถึง 3.0 ค่าการเดาต่ำกว่า 0.3 ดังนั้นจึงทำให้ได้ข้อสอบทั้งหมด 200 ข้อ

5. จัดเก็บข้อสอบที่ได้ไว้ในคลังข้อสอบ โดยใช้โปรแกรมคอมพิวเตอร์สำหรับสร้างคลังข้อสอบที่ผู้วิจัยสร้างขึ้น ซึ่งแยกเป็น 3 ไฟล์ คือ

5.1 ไฟล์ตัวอักษร มีชื่อว่า TEST.DAT สำหรับเก็บข้อมูลเกี่ยวกับเลขประจำข้อของข้อสอบ โจทย์ ตัวเลือกและเฉลยคำตอบ

5.2 ไฟล์รูปภาพ มีชื่อว่า T.PIC สำหรับเก็บรูปภาพของข้อสอบข้อต่างๆ

5.3 ไฟล์ค่าพารามิเตอร์ มีชื่อว่า COMPUTE.DAT สำหรับเก็บค่าพารามิเตอร์ของข้อสอบแต่ละข้อ คือ ค่าอำนาจจำแนก ค่าความยากง่ายและค่าการเดา นอกจากนั้นผู้วิจัยยังได้เก็บค่าความสามารถที่ทำให้ข้อสอบมาค่าสารสนเทศสูงสุด (θ_{max}) ของข้อสอบแต่ละข้อเอาไว้ด้วย ซึ่งค่าพารามิเตอร์ของข้อสอบดังกล่าวนี้จะถูกนำไปใช้ในการประมาณค่าความสามารถของผู้สอบ และใช้เป็นตัวชี้สำหรับคัดเลือกข้อสอบจากคลังข้อสอบออกมาให้ผู้สอบตอบที่หน้าจอตัวอย่างหน้าจอคอมพิวเตอร์สำหรับไฟล์ต่างๆ ผู้วิจัยแสดงไว้ในภาคผนวก

การดำเนินการวิจัย ระยะที่ 2 ศึกษาผลของตัวแปรอิสระต่อตัวแปรตาม

การวิจัยในระยะที่ 2 นี้เป็นการศึกษาผลของตัวแปรอิสระ 5 ตัว คือ ความยากง่ายของข้อสอบข้อแรก เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบและความสามารถของผู้สอบที่มีต่อค่าความเที่ยงตรงเชิงภาพและจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ ซึ่งมีวิธีการดำเนินการวิจัย ดังนี้

ประชากรและกลุ่มตัวอย่าง

ประชากรในการวิจัยระยะที่ 2 นี้ ได้แก่ นักเรียนชั้นมัธยมศึกษาปีที่ 1 ในโรงเรียน มัธยมศึกษาส่วนกลาง สังกัดกรมสามัญศึกษา กระทรวงศึกษาธิการ ช่วงเริ่มต้นของภาคเรียนที่ 1 ปีการศึกษา 2539 ซึ่งเป็นนักเรียนที่เพิ่งเรียนจบชั้นประถมศึกษาปีที่ 6 และยังไม่ได้เรียน เนื้อหาใหม่ของชั้นมัธยมศึกษาปีที่ 1 จำนวน 59,744 คน จาก 118 โรงเรียน

กลุ่มตัวอย่างของการวิจัยได้มาโดยการสุ่มแบบหลายขั้นตอน (multi-stage random sampling) โดยมีขั้นตอนในการสุ่ม ดังนี้

ขั้นตอนแรก สุ่มโรงเรียน 15 โรงเรียน จากโรงเรียนมัธยมศึกษาในส่วนกลางทั้งหมด 118 โรงเรียน โดยการสุ่มแบบแบ่งชั้น (stratified random sampling) ซึ่งใช้ขนาดของโรงเรียน เป็นชั้นของการสุ่ม ได้โรงเรียนขนาดใหญ่ 5 โรงเรียน ขนาดกลาง 8 โรงเรียนและขนาดเล็ก 2 โรงเรียน

ขั้นตอนที่สอง สุ่มห้องเรียนจากโรงเรียนที่สุ่มได้จากขั้นตอนแรก โดยการสุ่มอย่างง่าย (simple random sampling) โรงเรียนละ 5 ห้อง แต่เนื่องจากโรงเรียนขนาดเล็กมีห้องเรียนน้อย ดังนั้นโรงเรียนขนาดเล็ก 2 โรงเรียนจึงสุ่มได้เพียง 5 ห้อง ขนาดกลาง 8 โรงเรียนได้ 40 ห้องและขนาดใหญ่ 5 โรงเรียน ได้ 25 ห้องและสุ่มเพิ่มเติมเพื่อทดแทนโรงเรียนขนาดเล็กอีก 5 ห้อง จึงได้ห้องเรียนจากโรงเรียนขนาดใหญ่ 30 ห้องได้ห้องเรียนทั้งหมด 75 ห้อง นักเรียน 3,104 คน จากนั้นผู้วิจัยได้ใช้แบบทดสอบวัดความสามารถทางคณิตศาสตร์ทดสอบนักเรียนที่ได้จากการสุ่มในขั้นนี้ทุกคน

ขั้นตอนที่สาม สุ่มนักเรียนในแต่ละห้องที่ได้จากการสุ่มในขั้นที่สอง โดยการสุ่มแบบแบ่งชั้น ซึ่งใช้ความสามารถทางคณิตศาสตร์ของนักเรียนเป็นชั้นของการสุ่ม ได้แก่ ความสามารถสูง ปานกลาง และต่ำ ผู้วิจัยต้องการกลุ่มตัวอย่าง กลุ่มละ 540 คน รวมกลุ่มตัวอย่างทั้งหมดที่สุ่มได้ 1,620 คน จำแนกเป็นรายชื่อโรงเรียนได้ดังตาราง 5

ตาราง 5 จำนวนนักเรียนกลุ่มตัวอย่าง จำแนกเป็นรายชื่อโรงเรียนและขนาดของโรงเรียน

ที่	โรงเรียน	ขนาดโรงเรียน	จำนวนนักเรียน
1	สวนกุหลาบวิทยาลัย	ใหญ่	130
2	ราชดำริ	ใหญ่	130
3	ชิโนรสวิทยาลัย	ใหญ่	130
4	ศึกษานารี	ใหญ่	130
5	สันติราษฎร์วิทยาลัย	ใหญ่	130
6	วัดประดู่ในทรงธรรม	ปานกลาง	100
7	จันทร์ประดิษฐารามวิทยาคม	ปานกลาง	100
8	ปทุมคงคา	ปานกลาง	100
9	พระโขนงพิทยาลัย	ปานกลาง	100
10	สิริรัตนาร	ปานกลาง	100
11	สีกัน(วัฒนานันท์อุปถัมภ์)	ปานกลาง	100

ที่	โรงเรียน	ขนาดโรงเรียน	จำนวนนักเรียน
12	ดอนเมืองทหารอากาศบำรุง	ปานกลาง	100
13	ศิลาจารพิพัฒน์	ปานกลาง	100
14	วัดนายโรง	เล็ก	85
15	วัดบวรมงคล	เล็ก	85
รวม			1,620

เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยในระยะที่ 2 นี้ ได้แก่

1. แบบทดสอบวัดความสามารถทางคณิตศาสตร์ สำหรับนักเรียนชั้นประถมศึกษาปีที่ 6 ประกอบด้วยข้อสอบ 35 ข้อ เป็นข้อสอบปรนัยชนิดเลือกตอบ 4 ตัวเลือก เครื่องมือชุดนี้มีไว้สำหรับทดสอบนักเรียนกลุ่มตัวอย่างเพื่อแบ่งออกเป็น 3 กลุ่ม คือ กลุ่มที่มีความสามารถทางคณิตศาสตร์สูง ความสามารถปานกลางและความสามารถต่ำ นอกจากนั้นยังได้ใช้คะแนนความสามารถจากการทดสอบจากแบบทดสอบฉบับนี้เป็นคะแนนเกณฑ์ในการศึกษาความเที่ยงตรงเชิงสภาพของแบบทดสอบอีกด้วย

2. โปรแกรมคอมพิวเตอร์สำหรับการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ ด้วยคอมพิวเตอร์หรือการทดสอบแบบ ซี เอ ที ซึ่งมีจำนวนทั้งสิ้น 18 โปรแกรมย่อย ซึ่งได้จากการจัดหมู่ตัวแปรอิสระ 3 ตัว คือ ความยากง่ายของข้อสอบข้อแรก 3 ระดับ (ง่าย ปานกลาง ยาก) เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ 3 เกณฑ์ (ค่าความยากง่าย ค่าความสามารถที่ทำให้ข้อสอบมีสารสนเทศสูงสุด ค่าสารสนเทศของข้อสอบสูงสุด) วิธีการประมาณค่าความสามารถของผู้สอบ 2 วิธี (วิธีความเป็นไปได้สูงสุด วิธีของเบส์) ส่วนเกณฑ์ยุติการทดสอบ 3 เกณฑ์ (ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า เท่ากับหรือน้อยกว่า .30 .40 .50) และความสามารถของผู้สอบ 3 ระดับ (สูง กลาง ต่ำ) นั้นไม่ได้นำมารวมในการจัดหมู่ เพราะโปรแกรมสามารถใช้ร่วมกันได้ ซึ่งโปรแกรมย่อยทั้ง 18 โปรแกรมมีเงื่อนไขแสดงได้ดังตาราง 6

ตาราง 6 เงื่อนไขของโปรแกรมคอมพิวเตอร์สำหรับทดสอบนักเรียนในการทดสอบแบบปรับเหมาะ
กับความสามารถของผู้สอบด้วยคอมพิวเตอร์

โปรแกรมย่อย	ข้อสอบข้อแรก	เกณฑ์คัดเลือกข้อสอบ	วิธีประมาณค่า
1	ง่าย	$\hat{\theta} = b$	MLE
2	ง่าย	$\hat{\theta} = b$	Bayesian
3	ง่าย	$\hat{\theta} = \theta_{\max}$	MLE
4	ง่าย	$\hat{\theta} = \theta_{\max}$	Bayesian
5	ง่าย	$I_{\max}$	MLE
6	ง่าย	$I_{\max}$	Bayesian
7	ยากปานกลาง	$\hat{\theta} = b$	MLE
8	ยากปานกลาง	$\hat{\theta} = b$	Bayesian
9	ยากปานกลาง	$\hat{\theta} = \theta_{\max}$	MLE
10	ยากปานกลาง	$\hat{\theta} = \theta_{\max}$	Bayesian
11	ยากปานกลาง	$I_{\max}$	MLE
12	ยากปานกลาง	$I_{\max}$	Bayesian
13	ยาก	$\hat{\theta} = b$	MLE
14	ยาก	$\hat{\theta} = b$	Bayesian
15	ยาก	$\hat{\theta} = \theta_{\max}$	MLE
16	ยาก	$\hat{\theta} = \theta_{\max}$	Bayesian
17	ยาก	$I_{\max}$	MLE
18	ยาก	$I_{\max}$	Bayesian

การสร้างเครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยระยะที่ 2 นี้ ประกอบด้วย แบบทดสอบวัดความสามารถทาง
คณิตศาสตร์ และ โปรแกรมคอมพิวเตอร์สำหรับการทดสอบแบบปรับเหมาะกับความสามารถของ
ผู้สอบด้วยคอมพิวเตอร์ ซึ่งมีขั้นตอนในการสร้าง ดังนี้

แบบทดสอบวัดความสามารถทางคณิตศาสตร์

ผู้วิจัยได้สร้างแบบทดสอบวัดความสามารถทางคณิตศาสตร์สำหรับนักเรียนชั้นประถมศึกษาปีที่ 6 ตามขั้นตอนต่อไปนี้

1. ศึกษาโครงสร้างของหลักสูตรคณิตศาสตร์ระดับประถมศึกษาจากหลักสูตรประถมศึกษา พุทธศักราช 2521 พบว่ามี 5 ด้าน คือ ความรู้พื้นฐานเกี่ยวกับจำนวน พีชคณิต การวัด เรขาคณิตและสถิติ

2. ศึกษาสมรรถภาพที่สำคัญสำหรับวัดความสามารถทางคณิตศาสตร์ พบว่า มี 4 สมรรถภาพ ได้แก่ ความรู้ความเข้าใจ ทักษะการคิดคำนวณ ทักษะการแก้โจทย์ปัญหาและการปฏิบัติงานทางคณิตศาสตร์

3. กำหนดลักษณะของเครื่องมือวัด ซึ่งผู้วิจัยเลือกใช้รูปแบบของข้อสอบปรนัยชนิดเลือกตอบ 4 ตัวเลือก

4. สร้างฟาเซต (facet) ของคำถามและตัวเลือก เพื่อใช้เป็นกรอบในการสร้างข้อสอบวัดความรู้ความเข้าใจ เกี่ยวกับคณิตศาสตร์ ซึ่งแสดงไว้ในภาคผนวก

5. สร้างข้อสอบวัดความสามารถทางคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 โดยวัดสมรรถภาพความเข้าใจเกี่ยวกับคณิตศาสตร์ จากเนื้อหาตามโครงสร้าง 5 ด้าน ได้ข้อสอบรวม ทั้งหมด 50 ข้อ จำแนกตามปริมาณของเนื้อหาได้ดังตาราง 7

ตาราง 7 เนื้อเรื่อง จำนวนคาบและข้อสอบที่ใช้วัดความสามารถทางคณิตศาสตร์
ชั้นประถมศึกษาปีที่ 6

ที่	เนื้อหา	จำนวนคาบ	ข้อที่
1	จำนวนและตัวเลข	8	1 2 3
2	สมการและการแก้สมการ	34	4 5 6 7
3	ตัวประกอบของจำนวนนับ	43	8 9 10 11 12
4	มุมและส่วนของเส้นตรง	24	13 14 15
5	เส้นขนาน	17	18 19
6	ทศและแผนผัง	20	16 17
7	เศษส่วน	45	20 21 22 23 24
8	ทศนิยม	72	25 26 27 28 29 30 31

ตาราง 7 (ต่อ)

ที่	เนื้อหา	จำนวนคาบ	ข้อที่
9	รูปสามเหลี่ยม	23	32 33
10	รูปสี่เหลี่ยม	45	34 35 36 37 38 39
11	รูปวงกลม	13	40 41
12	บทประยุกต์	48	42 43 44 45 46
13	รูปทรงและปริมาตร	20	47 48
14	แผนภูมิและกราฟ	18	49 50

6. นำข้อสอบให้ผู้เชี่ยวชาญด้านหลักสูตร การสอน การวัดและประเมินผลการศึกษา คณิตศาสตร์ตรวจสอบ ความเที่ยงตรงตามเนื้อหาและเที่ยงตรงเชิงโครงสร้าง และปรับปรุงข้อสอบตามคำแนะนำของผู้เชี่ยวชาญ

7. นำแบบทดสอบไปทดลองใช้เพื่อหาคุณภาพ ครั้งที่ 1 โดยทดลองใช้กับนักเรียนชั้นประถมศึกษาปีที่ 6 ในโรงเรียนประถมศึกษาสังกัดสำนักงานการประถมศึกษากรุงเทพมหานคร และสำนักการศึกษากรุงเทพมหานคร ภาคเรียนที่ 2 ปีการศึกษา 2537 จำนวน 1,016 คน

8. วิเคราะห์คุณภาพของข้อสอบตามทฤษฎีการตอบสนองข้อสอบ โมเดลโลจิสติกแบบ 3 พารามิเตอร์ ซึ่งมีขั้นตอนการวิเคราะห์เหมือนกับการวิเคราะห์คุณภาพข้อสอบ เรื่อง เศษส่วน เพื่อตรวจสอบความเป็นมิตติเดียว หาค่าอำนาจจำแนก ค่าความยากง่ายและ ค่าการเดาของข้อสอบเป็นรายข้อ

9. คัดเลือกข้อสอบที่มีคุณภาพไว้โดยใช้เกณฑ์การคัดเลือกข้อสอบเหมือนกับเกณฑ์การคัดเลือกข้อสอบเรื่องเศษส่วน ซึ่งได้ข้อสอบวัดความสามารถทางคณิตศาสตร์ 35 ข้อ กระจายไปทุกเนื้อหา ยกเว้นเนื้อหาในโครงสร้างด้านสถิติ ซึ่งผู้วิจัยได้ตัดออกไป เพราะขณะทดลองเครื่องมือ นั้น นักเรียนยังไม่ได้เรียนเนื้อหาดังกล่าว

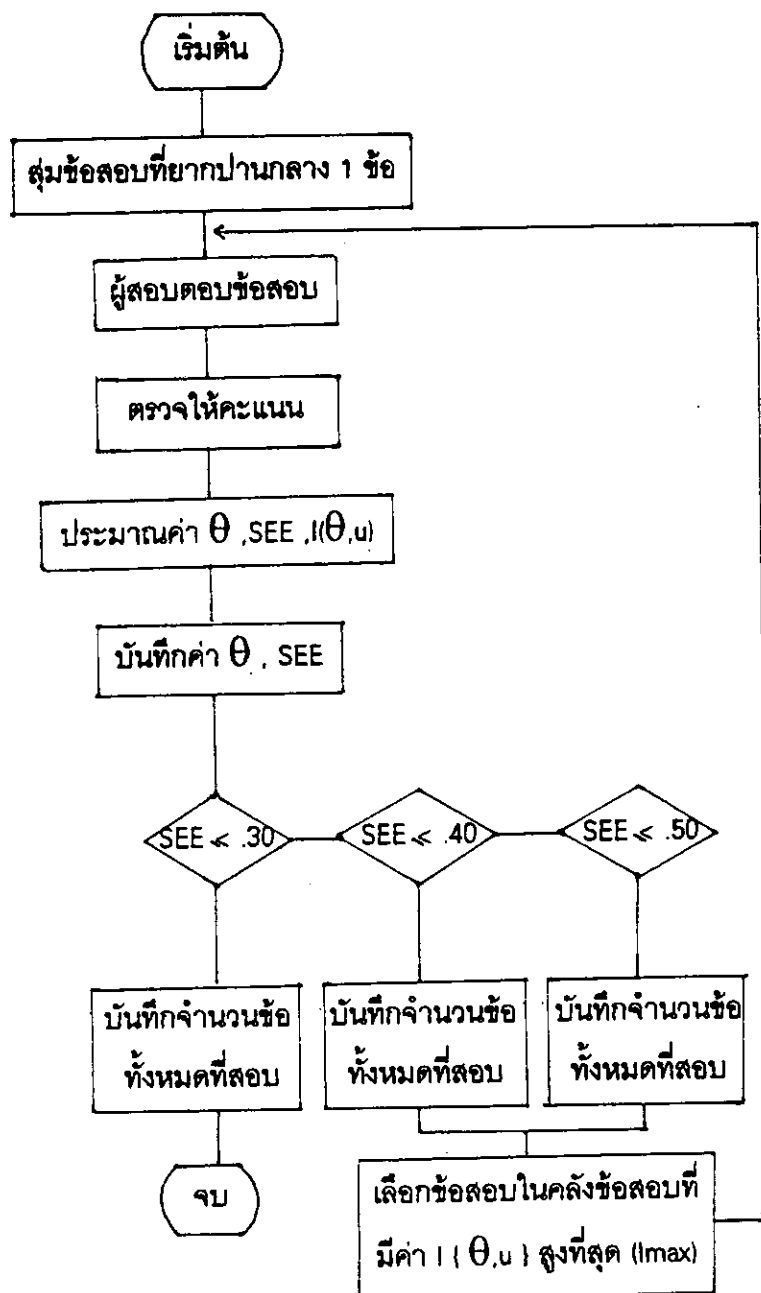
10. นำแบบทดสอบไปทดลองใช้ ครั้งที่ 2 เพื่อหาค่าคุณภาพประจำข้อ โดยทดสอบกับนักเรียนชั้นประถมศึกษาปีที่ 6 ในภาคเรียนที่ 2 ปีการศึกษา 2538 จำนวน 1900 คน จากโรงเรียน วัดพลับพลาชัย โรงเรียนดาราคาม โรงเรียนราชวินิต โรงเรียนพญาไท โรงเรียนวัดจันทรีโน โรงเรียนการเคหะท่าทรายและโรงเรียนประชานิเวศน์

โปรแกรมคอมพิวเตอร์สำหรับการทดสอบ

การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์หรือ การทดสอบแบบ ซี เอ ที นั้นดำเนินการสอบโดยให้ผู้สอบนั่งหน้าจอคอมพิวเตอร์และตอบข้อสอบ ครั้งละ 1 ข้อ โดยคอมพิวเตอร์เป็นผู้เลือกข้อสอบจากคลังข้อสอบขึ้นมาให้ผู้สอบได้ตอบ ข้อสอบที่ถูกเลือกขึ้นมานั้นจะมีความเหมาะสมกับความสามารถของผู้สอบแต่ละคน ในการดำเนินการสอบ ลักษณะนี้จำเป็นต้องมีโปรแกรมสั่งให้คอมพิวเตอร์จัดกระทำกับข้อมูลในลักษณะต่างๆ ตามตัวแปรอิสระที่ต้องการศึกษา ซึ่งผู้วิจัยได้สร้างโปรแกรมคอมพิวเตอร์ดังกล่าวตามขั้นตอน ดังนี้

1. เขียนผังงานแสดงขั้นตอนการทำงานของคอมพิวเตอร์ในการดำเนินการสอบ รวมทั้งวิเคราะห์และบันทึกผลการสอบ ซึ่งมีทั้งหมด 18 ผังงาน ตามจำนวนที่ได้จากการจัดหมู่ตัวแปรอิสระ 3 ตัวแปร คือ ความยากง่ายของข้อสอบข้อแรก 3 ระดับ (ข้อสอบที่ง่าย ยากปานกลางและยาก) เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ 3 เกณฑ์ (วิธีจับคู่ระหว่างความสามารถของผู้สอบกับค่าความยากง่ายของข้อสอบ วิธีจับคู่ระหว่างความสามารถของผู้สอบกับค่าความสามารถที่ทำให้ข้อสอบมีสารสนเทศสูงสุด วิธีใช้ค่าสารสนเทศของข้อสอบที่สูงที่สุด) วิธีการประมาณค่าความสามารถของผู้สอบ 2 วิธี (วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข และวิธีของเบส์) สำหรับตัวแปรอิสระอีก 2 ตัวแปร ได้แก่ เกณฑ์สำหรับยุติการทดสอบ 3 เกณฑ์ (ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าเท่ากับ หรือน้อยกว่า .30 .40 และ .50) และ ความสามารถของผู้สอบ 3 ระดับ (ความสามารถสูง ปานกลางและต่ำ) นั้นไม่ได้นำมารวมในการจัดหมู่ เพราะผู้สอบที่มีความสามารถทางคณิตศาสตร์ 3 ระดับดังกล่าวใช้โปรแกรมร่วมกันได้ นอกจากนี้ยังได้เขียนโปรแกรมสั่งให้คอมพิวเตอร์บันทึกค่าของตัวแปรตาม จำแนกตามเกณฑ์สำหรับยุติการทดสอบ 3 เกณฑ์ จึงทำให้ผู้สอบพบเกณฑ์การยุติการทดสอบทั้ง 3 เกณฑ์ ดังตัวอย่างของผังงานตามภาพประกอบ 27

2. ศึกษาเปรียบเทียบภาษาของคอมพิวเตอร์ที่จะนำมาเขียนโปรแกรมการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ ซึ่งผู้วิจัยคำนึงถึงความสะดวกในการ นำโปรแกรมไปใช้ และในการเขียนโปรแกรมในครั้งนี้ผู้วิจัยได้เลือกใช้ภาษาซี (C language) ในการเขียนโปรแกรม ทั้งนี้เนื่องจากภาษาซีมีความยืดหยุ่น ง่ายต่อการใช้คำสั่งและเกิดการผิดพลาดน้อย นอกจากนั้นยังได้นำโปรแกรม BTRIEVE ของ NETWARE เข้ามาช่วยในการจัดระบบข้อมูลด้วย



ภาพประกอบ 27 ตัวอย่างผังงานแสดงขั้นตอนการทำงานของคอมพิวเตอร์ในการดำเนินการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์

3. เขียนโปรแกรมคอมพิวเตอร์ตามผังงานที่ได้กำหนดไว้ 18 ผังงาน ซึ่งโปรแกรม ที่เขียนขึ้นตามโครงสร้างของตัวแปรที่ศึกษานั้น มีบางส่วนของโปรแกรมที่ต้องมีการประมาณค่า ต่างๆ ได้แก่

3.1 ความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง

ตามโมเดลโลจิสติกแบบ 3 พารามิเตอร์ ค่าความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องมีค่าดังนี้ (Hambleton and Swaminathan. 1985 : 37-38)

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

- เมื่อ $P_i(\theta)$ แทน ความน่าจะเป็นที่ผู้สอบซึ่งมีความสามารถเท่ากับ θ ตอบข้อสอบข้อที่ i ได้ถูกต้อง
- θ แทน ค่าความสามารถของผู้สอบ
- a_i แทน ค่าอำนาจจำแนกของข้อสอบข้อที่ i
- b_i แทน ค่าความยากง่ายของข้อสอบข้อที่ i
- c_i แทน ค่าความน่าจะเป็นของผู้สอบที่มีความสามารถต่ำมาก แต่สามารถตอบข้อสอบข้อที่ i ได้ถูกต้อง ซึ่งก็คือ ค่าการเดา
- D แทน ค่าคงที่มีค่าเท่ากับ 1.7
- e แทน ค่าคงที่มีค่าเท่ากับ 2.71828

3.2 ค่าความสามารถของผู้สอบ

3.2.1 วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (conditional maximum likelihood) ซึ่งมีวิธีการประมาณค่าความสามารถของผู้สอบ ดังนี้ (Hambleton and Swaminathan. 1985 : 81-91)

(1) ประมาณค่าความสามารถเริ่มต้นของผู้สอบโดยใช้สูตร

$$\theta_0 = \ln \frac{r}{n - r}$$

- เมื่อ θ_0 แทน ความสามารถเริ่มต้นของผู้สอบ
- $\ln$ แทน ลอการิทึมฐาน e
- r แทน คะแนนดิบของผู้สอบ
- n แทน จำนวนข้อสอบทั้งหมดที่ตอบ

(2) ประเมินค่า $\sum P(\theta)$ และ $\sum P(\theta)Q(\theta)$

(3) ประเมินค่าปรับแก้ : ซึ่งค่าปรับแก้นี้จะช่วยให้การประมาณค่าความ สามารถมีความถูกต้องมากยิ่งขึ้น ค่าปรับแก้ชนิด 3 พารามิเตอร์ (correction factor, h_m) หาได้ จากสูตร

$$h_m = \frac{D \sum_{i=1}^n \frac{a_i(u_i - P_i)(P_i - c_i)}{P_i(1 - c_i)}}{D^2 \sum_{i=1}^n \frac{a_i^2(P_i - c_i)(u_i c_i - P_i^2)Q_i}{P_i^2(1 - c_i)^2}}$$

(4) ประเมินค่า θ_1 จากสมการ

$$\theta_1 = \theta_0 - h_0$$

(5) ประเมินค่าในข้อ (2) ถึงข้อ (4) ซ้ำ จนกระทั่งตัวปรับแก้มีค่าน้อย กว่าที่ได้กำหนดไว้ (ξ) ซึ่งตามปกติจะกำหนดให้ $\xi = 0.001$ ก็จะสิ้นสุดการประมาณค่า นั่นคือ ถ้า $|h_0| = |\theta_1 - \theta_0| < \xi$ และค่า θ_1 ก็จะถูกใช้เป็นค่าความสามารถของผู้สอบ

3.3.2 วิธีของเบย์ส์ (bayesian) โดยเลือกใช้วิธีของโอเวน (Owen) ซึ่งมีวิธี การ ดังนี้ (Owen, 1975 : 351-356)

ถ้าตอบถูกต้อง

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

$$\sigma_{m+1}^2 = \sigma_m^2 \left[1 - \left[\frac{1 - c}{1 + \frac{1}{a^2 \sigma_m^2}} \right] \left[\frac{\Theta(D)}{A} \right] \left[\frac{(1-c) \Theta(D)}{A} - D \right] \right]$$

ถ้าตอบผิด

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

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

- เมื่อ
- $\Theta(D)$ แทน ค่าออร์ดิเนต (ordinate) ของโค้งปกติที่จุด D
 - $\Phi(D)$ แทน พื้นที่ใต้โค้งปกติ จากค่า D ต่ำสุด จนถึงจุด D ที่มีค่าเท่ากับ $\frac{b - \theta_m}{\sqrt{\frac{1}{a^2} + \sigma_m^2}}$ และ $A = c + (1 - c) \Phi(-D)$
 - θ_m แทน ความสามารถของผู้สอบที่ประมาณค่าได้ก่อนตอบข้อสอบข้อที่ $m+1$ ซึ่งตามปกติแล้ว ถ้าไม่ทราบค่าความสามารถเบื้องต้นของผู้สอบ ก็มักกำหนดให้เท่ากับ 0
 - σ_m^2 แทน ความแปรปรวนในการประมาณค่าความสามารถของผู้สอบก่อนตอบข้อที่ $m+1$ ถ้าไม่ทราบค่าความแปรปรวนดังกล่าวมาก่อน ก็มักกำหนดให้เท่ากับ 1
 - θ_{m+1} แทน ค่าความสามารถของผู้สอบโดยประมาณหลังจากที่ตอบข้อที่ $m+1$
 - σ_{m+1}^2 แทน ค่าความแปรปรวนในการประมาณค่าความสามารถของผู้สอบ เมื่อตอบข้อสอบข้อที่ $m+1$
 - a แทน ค่าอำนาจจำแนกของข้อสอบข้อที่ $m+1$
 - b แทน ค่าความยากง่ายของข้อสอบข้อที่ $m+1$
 - c แทน ค่าการเดาของข้อสอบข้อที่ $m+1$

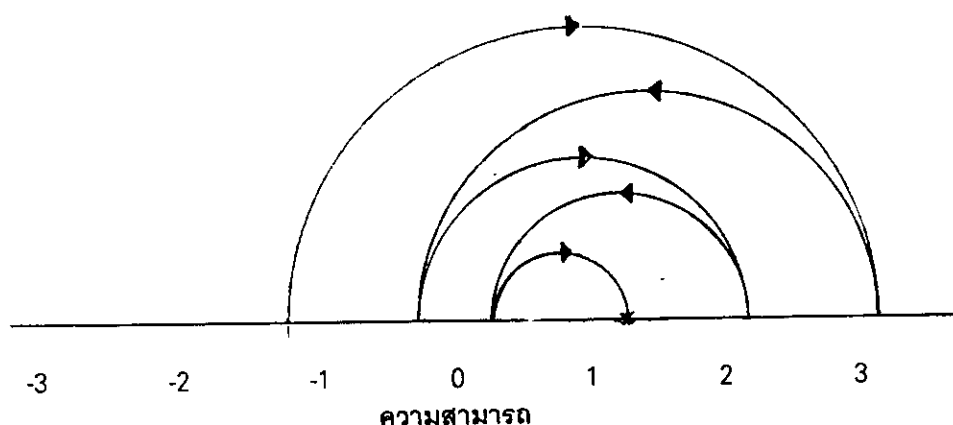
4. ทดลองใช้โปรแกรมคอมพิวเตอร์สำหรับสำหรับดำเนินการสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์กับนักเรียนชั้นประถมศึกษาปีที่ 6 ภาคเรียนที่ 2 ของปีการศึกษา 2538 ประมาณ 600 คน จากโรงเรียนวัดพลับพลาชัย โรงเรียนดาราคาม โรงเรียนราชวินิต โรงเรียนวัดจันทรีโน โรงเรียนการเคหะท่าทรายและโรงเรียนประชานิเวศน์

5. ปรับปรุงแก้ไขข้อบกพร่องของโปรแกรมคอมพิวเตอร์ให้สมบูรณ์มากยิ่งขึ้น

รูปแบบของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์

ในการวิจัยครั้งนี้ ผู้วิจัยเลือกใช้รูปแบบของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์แบบหลายขั้นตอน (multi-stage strategies) แบบแยกทางแปรผัน (variable branching model) ซึ่งในการทดสอบตามรูปแบบนี้ไม่สามารถ กำหนดจำนวนขั้นของการตอบได้อย่างแน่นอน ว่าผู้สอบจะต้องตอบกี่ขั้นก็ข้อ ทั้งนี้เนื่องจากผู้วิจัยได้ใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบเป็นเกณฑ์สำหรับยุติการทดสอบ

ผู้สอบแต่ละคนจะได้รับข้อสอบในจำนวนที่ไม่เท่ากัน นอกจากนั้นเส้นทาง ของการตอบข้อสอบข้อ ต่อๆไปก็ไม่สามารถกำหนดได้ล่วงหน้าเช่นเดียวกัน ทั้งนี้เพราะเส้นทาง ดังกล่าวจะขึ้นอยู่กับผล ของการตอบข้อที่ผ่านมาและเกณฑ์การคัดเลือกข้อสอบข้อต่อไปจากคลังข้อสอบ ซึ่งเส้นทางของ การตอบข้อสอบมีลักษณะดังภาพประกอบ 28



ภาพประกอบ 28 ตัวอย่างเส้นทางการตอบข้อสอบในการทดสอบแบบปรับเหมาะกับความ สามารถของผู้สอบด้วยคอมพิวเตอร์แบบหลายขั้นตอน แยกทางแปรผัน

จากภาพประกอบ 28 ข้อสอบข้อแรกมีความยากง่าย เท่ากับ -1.3 เมื่อตอบข้อแรกไป แล้ว ก็จะมีการประมาณค่าความสามารถ ข้อสอบข้อต่อไปจะถูกคัดเลือกขึ้นมาจากคลังข้อสอบ สมมติว่าได้ข้อสอบที่มีความยากง่าย เท่ากับ 3.1 การดำเนินการสอบจะเป็นเช่นนี้ไปเรื่อยๆจน กระทั่งถึงเกณฑ์ยุติการทดสอบ ตามภาพประกอบ 28 ผู้สอบได้ใช้ข้อสอบทั้งหมด 5 ข้อ ความ สามารถที่ประมาณค่าได้ คือ 1.4

การเก็บรวบรวมข้อมูลในการวิจัย ระยะที่ 2

ในการเก็บรวบรวมข้อมูลเพื่อศึกษาผลที่เกิดขึ้นต่อค่าความเที่ยงตรงเชิงสภาพและ จำนวนข้อสอบที่ใช้ในการทดสอบนั้นได้แบ่งออกเป็น 2 ขั้นตอน ขั้นตอนแรก วัดความรู้ความ สามารถทางคณิตศาสตร์ ขั้นตอนที่สอง วัดความเข้าใจ เรื่องเศษส่วนโดยการทดสอบแบบปรับ เหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ ซึ่ง มีขั้นตอน ดังนี้

1. นำหนังสือขอความร่วมมือในการวิจัยจากบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตรขอความร่วมมือไปยังกรมสามัญศึกษา กระทรวงศึกษาธิการ เพื่อขอทดสอบนักเรียนชั้นมัธยมศึกษาปีที่ 1 ในโรงเรียนที่ได้รับการสุ่มเป็นกลุ่มตัวอย่าง จำนวน 15 โรงเรียนในตอนต้นของภาคเรียนที่ 1 ปีการศึกษา 2539
2. ประสานงานกับกรมสามัญศึกษาและโรงเรียนที่ถูกสุ่มเป็นกลุ่มตัวอย่าง เพื่อชี้แจงวัตถุประสงค์ของการวิจัยและนัดหมายวัน เวลาเพื่อทดสอบนักเรียน
3. ทดสอบนักเรียนกลุ่มตัวอย่างด้วยแบบทดสอบวัดความสามารถทางคณิตศาสตร์ ซึ่งมีข้อสอบทั้งหมด 35 ข้อ
4. นำผลการตอบแบบทดสอบมาตรวจให้คะแนนความสามารถ แล้วแบ่งกลุ่มตัวอย่างดังกล่าวออกเป็น 3 กลุ่ม คือ กลุ่มที่มีความสามารถทางคณิตศาสตร์สูง ปานกลางและต่ำ โดยใช้ผลของการทดสอบเป็นข้อมูลในการแบ่งกลุ่มตามเกณฑ์ที่กำหนดไว้ในขอบเขตของการวิจัย ซึ่งผู้วิจัยต้องการกลุ่มตัวอย่างประมาณกลุ่มละ 540 คน รวมทั้งสิ้น 1,620 คน
5. เก็บคะแนนความสามารถจากการทดสอบในข้อ 4 ไว้ เพื่อนำไปใช้เป็นคะแนนเกณฑ์ในการศึกษาความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์หรือการทดสอบแบบ ซี เอ ที
6. แบ่งกลุ่มตัวอย่างออกเป็น 54 กลุ่มย่อย ด้วยการสุ่มแบบแบ่งชั้น โดยใช้ความสามารถของกลุ่มตัวอย่างเป็นชั้นของการสุ่ม ดังนั้น จึงทำให้กลุ่มตัวอย่างที่มีความสามารถทางคณิตศาสตร์สูง ปานกลางและต่ำแต่ละกลุ่มถูกแบ่งเป็นกลุ่มละ 18 กลุ่มย่อยเท่าๆกัน ประมาณกลุ่มละ 30 คน
7. สุ่มโปรแกรมคอมพิวเตอร์สำหรับดำเนินการสอบ ซึ่งมีอยู่ 18 โปรแกรมย่อยให้ กลุ่มตัวอย่างแต่ละกลุ่ม กลุ่มละ 1 โปรแกรม โดยการสุ่มอย่างง่าย
8. ทดสอบนักเรียนกลุ่มตัวอย่างโดยใช้คอมพิวเตอร์ดำเนินการสอบ ซึ่งก่อนสอบนั้นผู้ดำเนินการสอบจะชี้แจงเกี่ยวกับวิธีการตอบข้อสอบแก่ผู้สอบจนเข้าใจเป็นอย่างดีจึงเริ่มสอบ ผลการตอบของผู้สอบแต่ละคนจะถูกเก็บบันทึกไว้ในแผ่นจานแม่เหล็กโดยอัตโนมัติ
9. นำข้อมูลของผู้สอบแต่ละคนที่บันทึกไว้ในจานแม่เหล็ก คือ คะแนนความสามารถและจำนวนข้อสอบที่ตอบมาวิเคราะห์ทางสถิติ เพื่อศึกษาผลของตัวแปรอิสระต่อตัวแปรตาม

แบบแผนการวิจัย

การวิจัยครั้งนี้ ผู้วิจัยได้เลือกใช้แบบแผนการทดลอง $3 \times 3 \times 3 \times 2 \times 3$ Factorial Design ซึ่งมีโมเดลเชิงโครงสร้างของแบบแผน ดังนี้

$$\begin{aligned}
 Y = & \mu + \alpha + \beta + \gamma + \delta + \pi + \alpha\beta + \alpha\gamma + \alpha\delta + \alpha\pi \\
 & + \beta\gamma + \beta\delta + \beta\pi + \gamma\delta + \gamma\pi + \delta\pi + \alpha\beta\gamma + \alpha\beta\delta + \\
 & + \alpha\beta\pi + \alpha\gamma\delta + \alpha\gamma\pi + \alpha\delta\pi + \beta\gamma\delta + \beta\gamma\pi + \beta\delta\pi + \\
 & + \gamma\delta\pi + \alpha\beta\gamma\delta + \alpha\beta\gamma\pi + \alpha\beta\delta\pi + \alpha\gamma\delta\pi + \beta\gamma\delta\pi + \alpha\beta\gamma\delta\pi \\
 & + \epsilon
 \end{aligned}$$

เมื่อ	Y	แทน คะแนนความสามารถ/จำนวนข้อสอบ ที่ได้รับผลจากตัวแปรอิสระ
	μ	แทน ค่าเฉลี่ยหลัก ✓
	α	แทน ผลอันเนื่องมาจากความยากง่ายของข้อสอบข้อแรก ✓
	β	แทน ผลอันเนื่องมาจากเกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ ✓
	γ	แทน ผลอันเนื่องมาจากเกณฑ์สำหรับยุติการทดสอบ ✓
	δ	แทน ผลอันเนื่องมาจากวิธีการประมาณค่าความสามารถผู้สอบ ✓
	π	แทน ผลอันเนื่องมาจากความสามารถทางคณิตศาสตร์ของผู้สอบ
	ϵ	แทน ความคลาดเคลื่อน

จากแบบแผนดังกล่าวสามารถนำมาเขียนแยกเป็นหน่วยของการวิเคราะห์ข้อมูลได้

ดังตาราง 8

ตาราง 8 หน่วยของการวิเคราะห์ความเที่ยงตรงเชิงสภาพและจำนวนข้อสอบที่ใช้ในการทดสอบ
แบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์จำแนกตามความยากง่ายของ
ข้อสอบข้อแรก เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถ
ของผู้สอบ เกณฑ์ยุติการทดสอบและความสามารถของผู้สอบ

ยากง่าย	เกณฑ์ เลือก	เกณฑ์ ยุติ	วิธี MLE			วิธีของ Bay		
			สูง	กลาง	ต่ำ	สูง	กลาง	ต่ำ
ง่าย	b	.30	G1	G2	G3	G4	G5	G6
		.40	G1	G2	G3	G4	G5	G6
		.50	G1	G2	G3	G4	G5	G6
	Qmax	.30	G7	G8	G9	G10	G11	G12
		.40	G7	G8	G9	G10	G11	G12
		.50	G7	G8	G9	G10	G11	G12
	lmax	.30	G13	G14	G15	G16	G17	G18
		.40	G13	G14	G15	G16	G17	G18
		.50	G13	G14	G15	G16	G17	G18
	b	.30	G19	G20	G21	G22	G23	G24
		.40	G19	G20	G21	G22	G23	G24
		.50	G19	G20	G21	G22	G23	G24
กลาง	Qmax	.30	G25	G26	G27	G28	G29	G30
		.40	G25	G26	G27	G28	G29	G30
		.50	G25	G26	G27	G28	G29	G30
	lmax	.30	G31	G32	G33	G34	G35	G36
		.40	G31	G32	G33	G34	G35	G36
		.50	G31	G32	G33	G34	G35	G36
	b	.30	G37	G38	G39	G40	G41	G42
		.40	G37	G38	G39	G40	G41	G42
		.50	G37	G38	G39	G40	G41	G42
		.30	G43	G44	G45	G46	G47	G48

ยากง่าย	เกณฑ์ เลือก	เกณฑ์ ยุติ	วิธี MLE			วิธีของ Bay		
			สูง	กลาง	ต่ำ	สูง	กลาง	ต่ำ
ยาก	Qmax	.40	G43	G44	G45	G46	G47	G48
		.50	G43	G44	G45	G46	G47	G48
		.30	G49	G50	G51	G52	G53	G54
	lmax	.40	G49	G50	G51	G52	G53	G54
		.50	G49	G50	G51	G52	G53	G54

การวิเคราะห์ข้อมูล

ในการวิเคราะห์ข้อมูลนั้นได้วิเคราะห์ค่าของตัวแปรตามเพื่อศึกษาผลที่เกิดขึ้นอันเนื่องมาจากตัวแปรอิสระ ซึ่งตัวแปรตาม คือ ค่าความเที่ยงตรงเชิงสภาพของแบบทดสอบและจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ โดยดำเนินการ ดังนี้

- 1. หาค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จากสูตร (Glass and Stanley. 1970 : 114)

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

เมื่อ r_{xy} แทน ความสัมพันธ์ระหว่างคะแนนจากการตอบแบบทดสอบ
วัดความสามารถทางคณิตศาสตร์และความเข้าใจ เรื่อง เศษส่วน
 n แทน จำนวนผู้สอบ
 x, y แทน คะแนนความสามารถจากการตอบแบบทดสอบวัดความสามารถทางคณิตศาสตร์และความเข้าใจ เรื่อง เศษส่วน ตามลำดับ

- 2. แปลงค่าความเที่ยงตรงเชิงสภาพให้เป็นคะแนนฟิชเชอร์ซี (Fishe's Z) และแปลงคะแนนฟิชเชอร์ซี กลับเป็นค่าความเที่ยงตรงเชิงสภาพ จากสูตร (ต่าย เขียงฉี. 2534 : 99 ; อ้างอิงมาจาก Charterand Larsen. 1983 : 41-42)

$$Z_r = 0.5 [\ln (1+r) - \ln (1-r)]$$

$$r = \frac{(e^{2z} - 1)}{(e^{2z} + 1)}$$

เมื่อ Z_r แทน คะแนนฟิชเชอร์รี
 r แทน ค่าความเที่ยงตรงเชิงสภาพ

3. หาค่าเฉลี่ยของจำนวนข้อสอบที่ใช้ในการทดสอบ จากสูตร

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

เมื่อ $\bar{x}$ แทน จำนวนข้อสอบเฉลี่ยที่ใช้ในการทดสอบ
 $\sum x$ แทน ผลรวมของข้อสอบที่ใช้ในการทดสอบแต่ละคน
 n แทน จำนวนผู้สอบ

4. วิเคราะห์ความแปรปรวน เพื่อเปรียบเทียบค่าเฉลี่ยของจำนวนข้อสอบที่ใช้ในการทดสอบ จากสูตร (Glass and Stanley, 1970 : 353)

$$F = \frac{MS_B}{MS_W}$$

เมื่อ F แทน ค่าสถิติทดสอบเอฟ สำหรับทดสอบความแตกต่างของค่าเฉลี่ย
 MS_B แทน ค่ากำลังสองเฉลี่ยของข้อมูลระหว่างกลุ่ม (mean square between groups)
 MS_W แทน ค่ากำลังสองเฉลี่ยของข้อมูลภายในกลุ่ม (mean square within groups)

5. เปรียบเทียบความแตกต่างของค่าเฉลี่ยจำนวนข้อสอบเป็นรายคู่โดยใช้วิธีเอสของเซฟเฟ (Scheffe's S Method) จากสูตร (อุทุมพร ทองอุไทย, 2523 : 158)

$$S = \sqrt{(k-1) F_{\alpha, v_1, v_2} MS_e \left[\sum_{j=1}^k \frac{(c_j)^2}{n_j} \right]}$$

- เมื่อ F แทน ค่า F ที่ขึ้นความเป็นอิสระ V1 และ V2 ที่ระดับนัยสำคัญ α
 k แทน จำนวนระดับการทดลอง
 c_j แทน สัมประสิทธิ์การเปรียบเทียบ
 n_j แทน จำนวนคะแนนในระดับการทดลอง j

6. ทดสอบความแตกต่างของค่าความเที่ยงตรงเชิงสภาพ โดยใช้ค่าไคสแควร์

(Chi square) จากสูตร (อนันต์ ศรีโสภา. 2525 : 251)

$$\sum_{i=1}^j (Z_{rj} - Z_r)^2 (n_j - 3) \rightarrow \chi_{j-1}^2 ; \alpha$$

$$Z_r = \frac{\sum_{i=1}^j (n_j - 3) Z_{rj}}{\sum_{i=1}^j (n_j - 3)}$$

- เมื่อ Z_{rj} แทน ค่าฟิชเชอร์ซี (Fisher Z) ของกลุ่มผู้สอบ j
 n_j แทน จำนวนผู้สอบกลุ่มที่ j

7. เปรียบเทียบความแตกต่างของค่าความเที่ยงตรงเชิงสภาพเป็นรายคู่จากสูตร

(Glass and Stanley. 1970 : 311-313)

กรณีค่าสหสัมพันธ์เป็นอิสระจากกัน

$$Z = \frac{Z_{r1} - Z_{r2}}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}}$$

กรณีค่าสหสัมพันธ์ไม่เป็นอิสระจากกัน

$$Z = \frac{\sqrt{n} (r_{xy} - r_{xz})}{\sqrt{(1 - r_{xy}^2)^2 + (1 - r_{xz}^2)^2 - 2r_{yz}^3 - (2r_{yz} - r_{xy}r_{xz})(1 - r_{xy}^2 - r_{xz}^2 - r_{yz}^2)}}$$

เมื่อ	Z	แทน สถิติทดสอบที (Z test)
	Z_{r1}, Z_{r2}	แทน ค่าพีชเชอร์รี ซึ่งแปลงมาจากค่าสหสัมพันธ์ r_1 และ r_2 ตามลำดับ
	n	แทน ขนาดของกลุ่มตัวอย่าง
	r_{xy}	แทน สหสัมพันธ์ระหว่างตัวแปร x กับตัวแปร y
	r_{xz}	แทน สหสัมพันธ์ระหว่างตัวแปร x กับตัวแปร z
	r_{yz}	แทน สหสัมพันธ์ระหว่างตัวแปร y กับตัวแปร z

บทที่ 4

ผลการวิเคราะห์ข้อมูล

ในบทนี้ได้นำเสนอผลของการวิเคราะห์ข้อมูล ซึ่งได้แบ่งการนำเสนอออกเป็น 2 ตอน ตอนแรก เกี่ยวกับค่าพารามิเตอร์ของข้อสอบในคลังข้อสอบ เรื่อง เศษส่วน และข้อสอบวัดความสามารถทางคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ตอนที่สองเป็นผลการทดสอบความสามารถทางคณิตศาสตร์และความเข้าใจ เรื่อง เศษส่วน จำแนกตามตัวแปรอิสระที่ศึกษา 5 ตัวแปรซึ่งมีรายละเอียดของผลการวิเคราะห์ข้อมูลในแต่ละตอน ดังนี้

ตอนที่ 1 ค่าพารามิเตอร์ของข้อสอบในคลังข้อสอบ เรื่อง เศษส่วน

ข้อสอบที่นำมาสร้างเป็นคลังข้อสอบนั้นได้มาจาก ข้อสอบวัดความเข้าใจ เรื่อง เศษส่วน ชั้นประถมศึกษาปีที่ 6 ซึ่งจัดทำเป็นแบบทดสอบคู่ขนาน จำนวน 7 ฉบับ เพื่อนำไปทดลองใช้กับกลุ่มตัวอย่าง จากผลจากการทดลอง ได้ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของค่าอำนาจจำแนก ค่าความยากง่ายและค่าการเดาของแต่ละฉบับ ดังตาราง 9

ตาราง 9 จำนวนข้อสอบ จำนวนผู้สอบ ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของค่าพารามิเตอร์ของข้อสอบที่ใช้วัดความเข้าใจ เรื่อง เศษส่วน ชั้นประถมศึกษาปีที่ 6 ที่ยังไม่ได้ปรับเกลการวัด

ฉบับที่	จำนวน		ค่าอำนาจจำแนก		ค่าความยากง่าย		ค่าการเดา	
	ข้อสอบ	ผู้สอบ	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
1	50	1074	1.020	.308	1.099	.923	.220	.079
2	50	1101	1.030	.382	.877	1.074	.207	.082
3	50	1084	1.039	.504	.581	1.082	.233	.065
4	50	1042	.937	.374	.840	1.073	.217	.082
5	50	1055	.949	.381	.978	1.107	.210	.077
6	50	1036	.937	.418	.851	1.034	.210	.066
7	50	1025	1.050	.349	.829	1.166	.210	.071

จากตาราง 9 จะเห็นว่าแบบทดสอบที่ใช้วัดความเข้าใจ เรื่อง เศษส่วน ชั้นประถมศึกษาปีที่ 6 ที่ยังไม่ได้ปรับสเกลการวัดนั้นมีลักษณะใกล้เคียงกัน ซึ่งสังเกตได้จากค่าเฉลี่ยของค่าพารามิเตอร์ของข้อสอบที่วิเคราะห์ได้ ดังนี้ คือ ค่าอำนาจจำแนกประมาณ 1.00 ซึ่งอยู่ในเกณฑ์ดี ค่าความยากง่ายประมาณ .80 ซึ่งอยู่ในเกณฑ์ยากง่ายปานกลางค่อนข้างไปทางยาก ส่วนค่าการเดามีค่าประมาณ .20 ซึ่งอยู่ในเกณฑ์ดี

เมื่อนำผลของการตอบแบบทดสอบวัดความเข้าใจ เรื่อง เศษส่วน 7 ฉบับ และ ความสามารถทางคณิตศาสตร์ 1 ฉบับไปวิเคราะห์องค์ประกอบ เพื่อตรวจสอบความเป็นมิติเดียวของข้อสอบ ซึ่งผลของการวิเคราะห์ดังกล่าวแสดงได้ดังตาราง 10 และภาพประกอบ 29

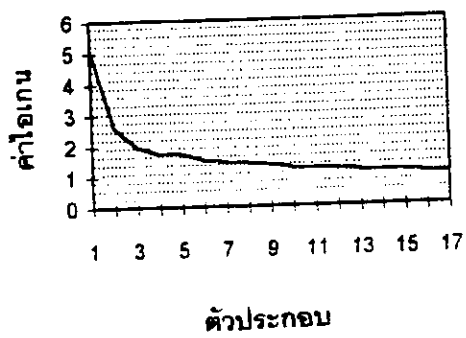
ตาราง 10 ค่าไอเกน (eigen value) ที่ได้จากการวิเคราะห์ตัวประกอบของแบบทดสอบวัดความเข้าใจ เรื่อง เศษส่วน ฉบับที่ 1-7 และ แบบทดสอบวัดความสามารถทางคณิตศาสตร์ ฉบับที่ 8 จำแนกเป็นรายตัวประกอบ(เฉพาะค่าไอเกนที่มากกว่า 1.00)

ตัวประกอบ	ฉบับที่							
	1	2	3	4	5	6	7	8
1	5.1472	9.3753	8.6276	9.6948	9.2754	10.454	10.930	10.607
2	2.6148	1.7921	2.6803	2.6384	2.6601	2.7913	3.8873	2.2267
3	1.9547	1.5650	1.8408	2.0334	1.7510	1.8761	1.7936	1.7005
4	1.7366	1.4086	1.5972	1.3904	1.4049	1.6499	1.5250	1.4726
5	1.7108	1.3493	1.4798	1.3278	1.2578	1.3580	1.3247	1.3388
6	1.5109	1.3103	1.3054	1.2302	1.2444	1.2585	1.2373	1.2583
7	1.4538	1.2602	1.1978	1.2123	1.2209	1.2450	1.1756	1.2143
8	1.4216	1.2233	1.1728	1.1772	1.1785	1.1350	1.1359	1.1455
9	1.3493	1.1431	1.1411	1.1444	1.1291	1.0936	1.0889	1.1283
10	1.2490	1.1284	1.1085	1.0919	1.0778	1.0824	1.0456	1.0922
11	1.2154	1.0857	1.0725	1.0405	1.0624	1.0523		1.0439
12	1.1981	1.0580	1.0636	1.0353	1.0302	1.0084		1.0267
13	1.1490	1.0330	1.0309					
14	1.1042	1.0165	1.0017					

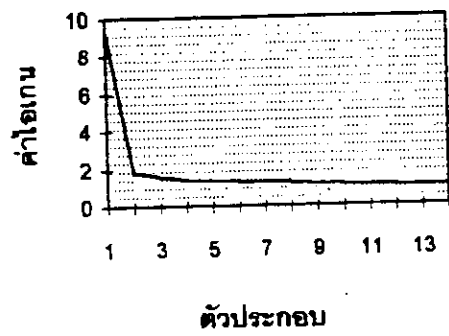
ตาราง 10 (ต่อ)

ตัวประกอบ	ฉบับที่							
	1	2	3	4	5	6	7	8
15	1.0931							
16	1.0348							
17	1.0188							
18	1.0005							

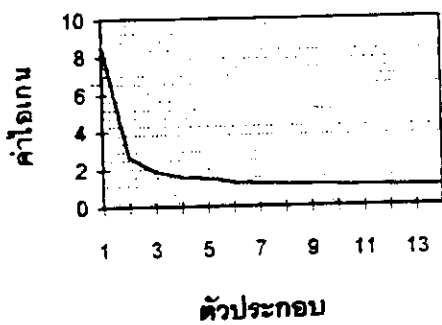
ฉบับที่ 1



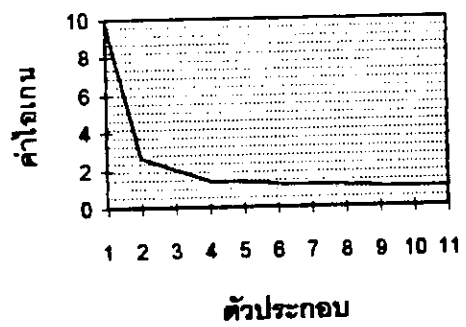
ฉบับที่ 2



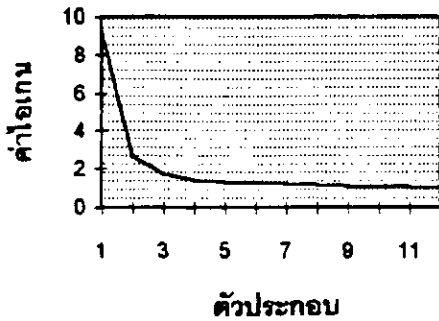
ฉบับที่ 3



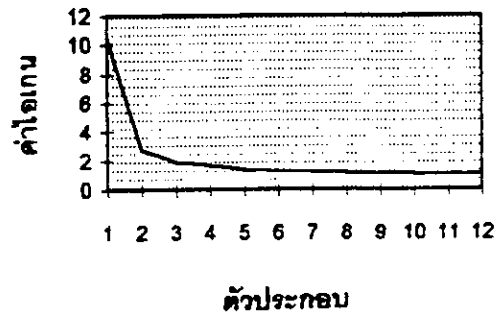
ฉบับที่ 4



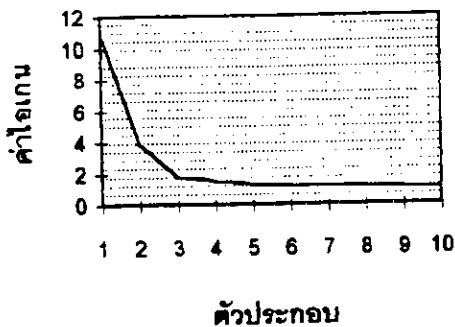
ฉบับที่ 5



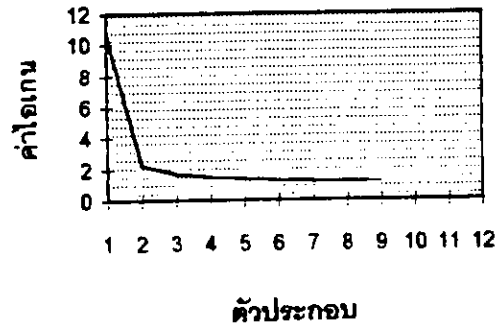
ฉบับที่ 6



ฉบับที่ 7



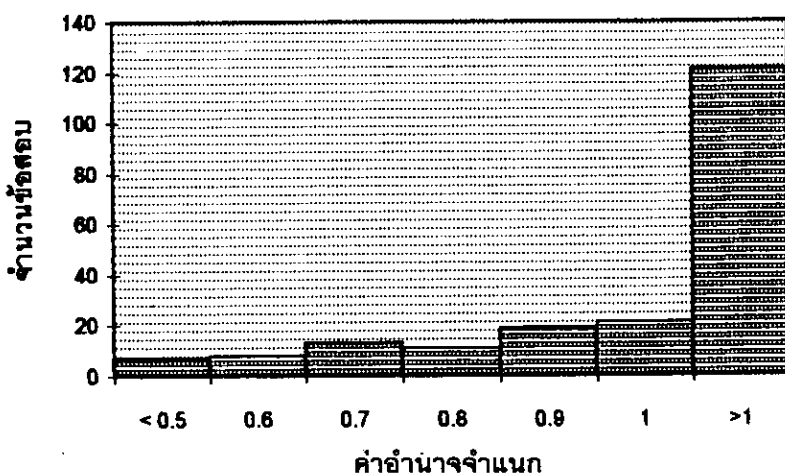
ฉบับที่ 8



ภาพประกอบ 29 ค่าไอเกน (eigen value) ที่ได้จากการวิเคราะห์ตัวประกอบของแบบทดสอบวัดความเข้าใจ เรื่องเศษส่วน ฉบับที่ 1-7 และ แบบทดสอบวัดความสามารถทางคณิตศาสตร์ ฉบับที่ 8 จำแนกเป็นรายตัวประกอบ

จากตาราง 10 และภาพประกอบ 29 จะเห็นว่า แบบทดสอบทั้ง 8 ฉบับ มีค่าไอเกนของตัวประกอบที่ 1 สูงกว่าค่าไอเกนของตัวประกอบอื่นๆที่เหลือ และตัวประกอบที่เหลือดังกล่าวก็มีค่าไอเกนที่ใกล้เคียงกัน ซึ่งลักษณะดังกล่าวนี้พออนุมานได้ว่า ข้อสอบที่ได้มีความเป็นมิติเดียว ตรงตามข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ

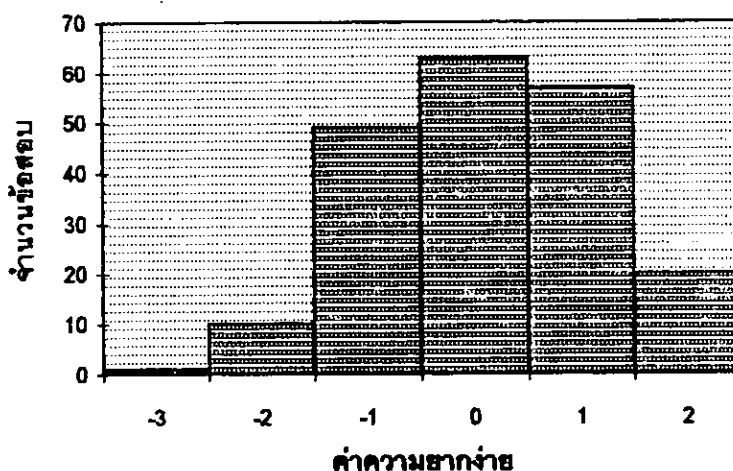
เพื่อให้เห็นจำนวนของข้อสอบในคลังข้อสอบในแต่ละช่วงของอำนาจจำแนก ค่าของความยากง่ายของข้อสอบและการเดา ผู้วิจัยจึงได้นำเสนอดังภาพประกอบ 30 - 32



ภาพประกอบ 30 จำนวนข้อสอบวัดความเข้าใจ เรื่อง เศษส่วน ชั้นประถมศึกษาปีที่ 6 ใน
คลังข้อสอบ 200 ข้อ จำแนกตามช่วงของค่าอำนาจจำแนก

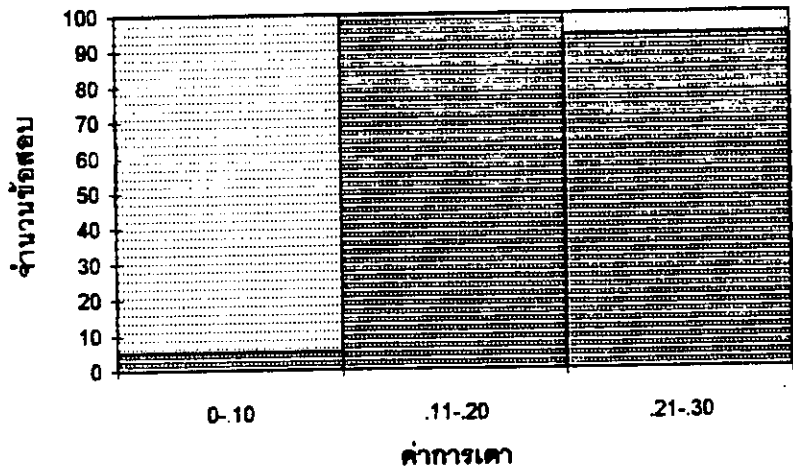
จากภาพประกอบจะเห็นว่าข้อสอบส่วนใหญ่ในคลังข้อสอบมีค่าอำนาจจำแนกมากกว่า
1.00 ซึ่งมีข้อสอบในกลุ่มนี้ประมาณ 120 ข้อ ถ้าพิจารณาโดยใช้เกณฑ์ค่าอำนาจจำแนกไม่ต่ำกว่า
.80 แล้ว ข้อสอบที่มีค่าอำนาจจำแนกมากกว่า .80 มีประมาณ 150 ข้อ

เมื่อจำแนกข้อสอบในคลังข้อสอบออกตามค่าความยากง่ายแล้ว สามารถจำแนกได้
ดังภาพประกอบ 31



ภาพประกอบ 31 จำนวนข้อสอบวัดความเข้าใจ เรื่อง เศษส่วน ชั้นประถมศึกษาปีที่ 6 ใน
คลังข้อสอบ 200 ข้อ จำแนกตามช่วงของค่าความยากง่าย

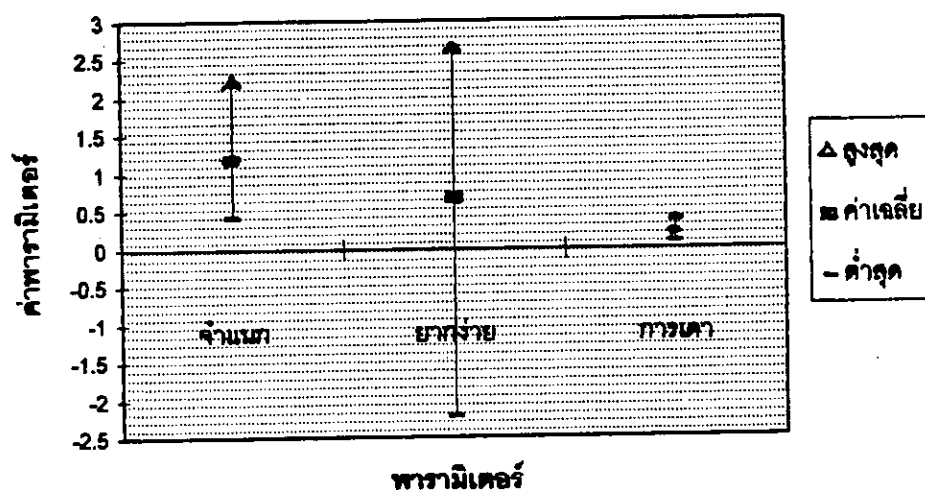
จากภาพประกอบจะเห็นว่าข้อสอบส่วนใหญ่ในคลังข้อสอบมีค่าความยากง่ายอยู่ในช่วง-1.5 ถึง 1.5 ซึ่งมีข้อสอบที่มีค่าความยากง่ายอยู่ในช่วงนี้ประมาณ 160 ข้อ ซึ่งการแจกแจงของค่าความยากง่ายดังกล่าวมีลักษณะเบ้ทางลบเล็กน้อย นั่นคือ ข้อสอบในคลังข้อสอบค่อนข้างยากเมื่อจำแนกข้อสอบในคลังข้อสอบออกตามค่าการเดาแล้ว สามารถจำแนกได้ดังภาพประกอบ 32



ภาพประกอบ 32 จำนวนข้อสอบวัดความเข้าใจ เรื่อง เศษส่วน ชั้นประถมศึกษาปีที่ 6 ในคลังข้อสอบ 200 ข้อ จำแนกตามช่วงของค่าการเดา

จากภาพประกอบจะเห็นว่าข้อสอบส่วนใหญ่ในคลังข้อสอบมีค่าการเดาอยู่ในช่วง .11 ถึง .30 ซึ่งมีข้อสอบในช่วงการเดา .11 ถึง .20 ประมาณ 100 ข้อ และช่วง .21 ถึง .30 ประมาณ 95 ข้อ

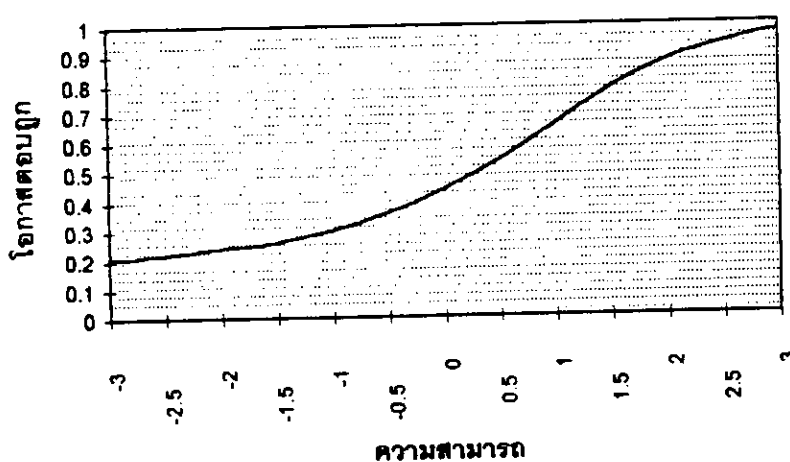
สำหรับค่าเฉลี่ยและพิสัยของค่าอำนาจจำแนก ค่าความยากง่ายและค่าการเดาของข้อสอบในคลังข้อสอบ 200 ข้อ แสดงได้ดังภาพประกอบ 33



ภาพประกอบ 33 ค่าเฉลี่ยและพิสัยของค่าอำนาจจำแนก ค่าความยากและค่าการเดาของข้อสอบ วัดความเข้าใจ เรื่อง เศษส่วน ชั้นประถมศึกษาปีที่ 6 จำนวน 200 ข้อในคลังข้อสอบ

จากภาพประกอบ 33 พบว่า ข้อสอบวัดความเข้าใจ เรื่อง เศษส่วนในคลังข้อสอบนั้นมีค่าอำนาจจำแนกเฉลี่ย เท่ากับ 1.11 มีพิสัยอยู่ในช่วง .41 ถึง 2.34 มีค่าความยากง่ายเฉลี่ย เท่ากับ .60 มีพิสัยของค่าความยากง่าย อยู่ในช่วง -2.23 ถึง 2.72 มีค่าการเดาเฉลี่ย เท่ากับ .19 มีพิสัยของค่าการเดา อยู่ในช่วง .06 ถึง .30

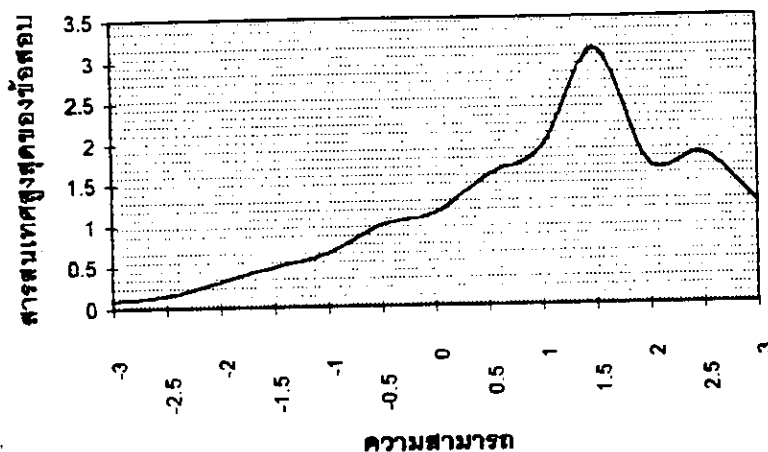
เมื่อพิจารณาถึงความน่าจะเป็นโดยเฉลี่ยในการตอบข้อสอบในคลังข้อสอบแต่ละข้อได้ถูกต้องที่ระดับความสามารถต่างกัน 13 ระดับ ก็แสดงให้เห็นได้ดังภาพประกอบ 34



ภาพประกอบ 34 ค่าความน่าจะเป็นโดยเฉลี่ยในการตอบข้อสอบ เรื่อง เศษส่วน ชั้นประถมศึกษาปีที่ 6 ในคลังข้อสอบได้ถูกต้อง จำแนกแต่ละระดับความสามารถ 13 ระดับ

จากภาพประกอบ 34 ผู้วิจัยได้นำค่าพารามิเตอร์ของข้อสอบแต่ละข้อในคลังข้อสอบมาหาค่าความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง แล้วนำค่าความน่าจะเป็น 200 ค่ามาหาค่าเฉลี่ย พบว่า ข้อสอบในคลังข้อสอบมีโค้งลักษณะข้อสอบสอดคล้องกับโมเดลโลจิสติก แบบ 3 พารามิเตอร์ตามทฤษฎีการตอบสนองข้อสอบ กล่าวคือ ผู้สอบที่มีความสามารถต่ำก็มีโอกาสที่จะตอบข้อสอบได้ถูกต้องต่ำกว่าผู้สอบที่มีความสามารถสูง

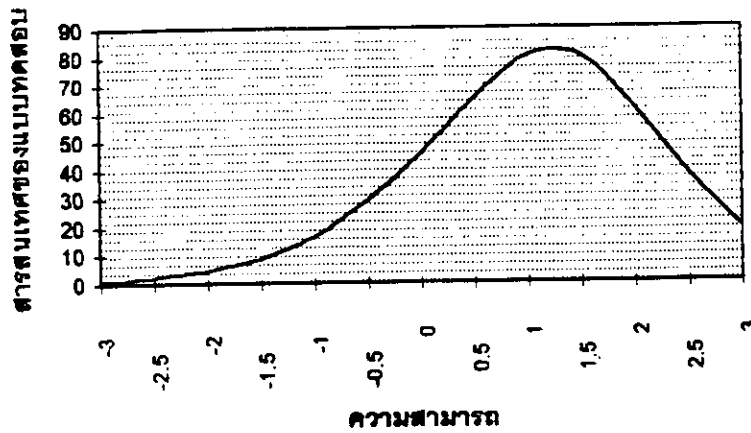
เมื่อนำค่าความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องที่ระดับต่างๆกันของความสามารถของผู้สอบไปหาค่าอนุพันธ์ และนำค่าอนุพันธ์ดังกล่าวนี้ไปแปลงเป็นค่าสารสนเทศของข้อสอบที่ระดับต่างๆกันของความสามารถ 13 ระดับ ซึ่งค่าสารสนเทศของข้อสอบที่สูงที่สุดในแต่ละระดับความสามารถ แสดงได้ดังภาพประกอบ 35



ภาพประกอบ 35 ค่าสารสนเทศของข้อสอบที่สูงสุด (I_{max}) ของข้อสอบ เรื่อง เศษส่วน ชั้นประถมศึกษาปีที่ 6 ในคลังข้อสอบ จำแนกแต่ละระดับความสามารถ 13 ระดับ

จากภาพประกอบ 35 จะเห็นว่า โค้งสารสนเทศที่สูงที่สุดของข้อสอบในคลังข้อสอบ เรื่อง เศษส่วนจะกระจายไปทุกระดับความสามารถ ตั้งแต่ระดับความสามารถ -3.00 ถึง 3.00 และค่าสารสนเทศของข้อสอบที่สูงที่สุดจะสูงสุดที่ระดับความสามารถ 1.5

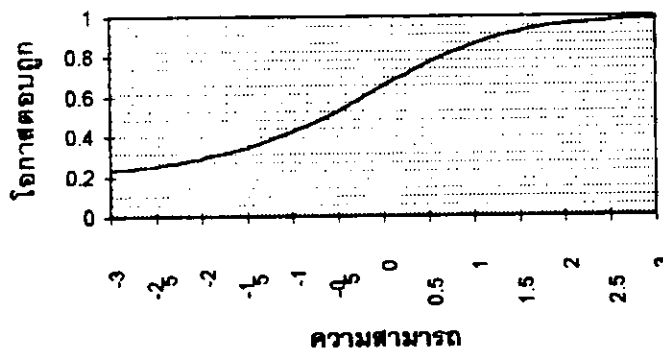
เมื่อนำค่าสารสนเทศของข้อสอบแต่ละข้อในคลังข้อสอบมารวมกัน เพื่อดูค่าสารสนเทศของแบบทดสอบในภาพรวม จะเห็นได้ดังภาพประกอบ 36



ภาพประกอบ 36 โค้งสารสนเทศของแบบทดสอบ (test information curve) วัดความเข้าใจ
เรื่อง เศษส่วน ชั้นประถมศึกษาปีที่ 6 จำนวน 200 ข้อในคลังข้อสอบ

จากภาพจะเห็นว่า โค้งสารสนเทศของแบบทดสอบ (ข้อสอบทุกข้อในคลังข้อสอบ) มี
ลักษณะเป็นไปทางลบ นั่นคือ ข้อสอบส่วนใหญ่ในคลังข้อสอบนั้นเหมาะสมกับผู้สอบที่มีความ
สามารถปานกลางถึงความสามารถสูง

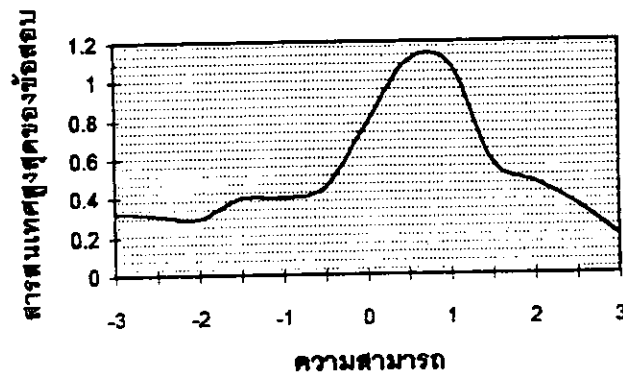
สำหรับแบบทดสอบวัดความสามารถทางคณิตศาสตร์นั้น ความน่าจะเป็นโดยเฉลี่ยใน
การตอบข้อสอบแต่ละข้อได้ถูกต้องที่ระดับความสามารถต่างๆกัน 13 ระดับแสดงได้ดังภาพ
ประกอบ 37



ภาพประกอบ 37 ค่าความน่าจะเป็นโดยเฉลี่ยในการตอบข้อสอบวัดความสามารถทาง
คณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ได้ถูกต้อง จำแนกแต่ละระดับความสามารถ 13 ระดับ

จากภาพประกอบ 37 พบว่า ข้อสอบวัดความสามารถทางคณิตศาสตร์มีโด่งลักษณะ ข้อสอบสอดคล้องกับ โมเดลโลจิสติก แบบ 3 พารามิเตอร์ตามทฤษฎีการตอบสนองข้อสอบ คือ ผู้สอบที่มีความสามารถต่ำมีโอกาสที่จะตอบข้อสอบได้ถูกต้องต่ำกว่าผู้สอบที่มีความสามารถสูง

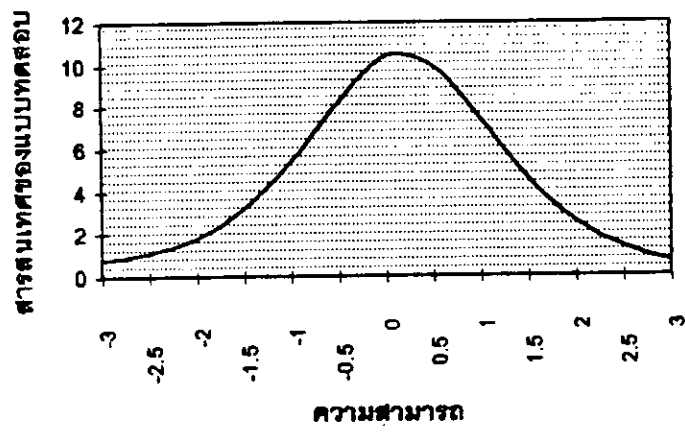
เมื่อนำค่าความน่าจะเป็นในการตอบข้อสอบวัดความสามารถทางคณิตศาสตร์ได้ถูกต้อง ที่ระดับต่างๆกันของความสามารถของผู้สอบไปหาค่าอนุพันธ์ และนำค่าอนุพันธ์ดังกล่าวนี้ไปแปลง เป็นค่าสารสนเทศของข้อสอบที่ระดับต่างๆกันของความสามารถ 13 ระดับ สามารถแสดงได้ดัง ภาพประกอบ 38



ภาพประกอบ 38 ค่าสารสนเทศของข้อสอบที่สูงที่สุด (Imax) ของข้อสอบวัดความสามารถทางคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 จำแนกแต่ละระดับความสามารถ 13 ระดับ

จากภาพประกอบ 38 จะเห็นว่าโด่งสารสนเทศที่สูงที่สุดของข้อสอบที่ระดับความสามารถต่างๆกระจายไปทุกระดับความสามารถ โดยเฉพาะในช่วงความสามารถ .5 ถึง 1.00 ค่าสารสนเทศของข้อสอบจะสูงกว่าช่วงความสามารถอื่นๆ

เมื่อนำค่าสารสนเทศของข้อสอบวัดความสามารถทางคณิตศาสตร์แต่ละข้อมารวมกัน ทั้ง 35 ข้อ จะได้โด่งสารสนเทศของแบบทดสอบวัดความสามารถทางคณิตศาสตร์ ซึ่งแสดงได้ดังภาพประกอบ 39

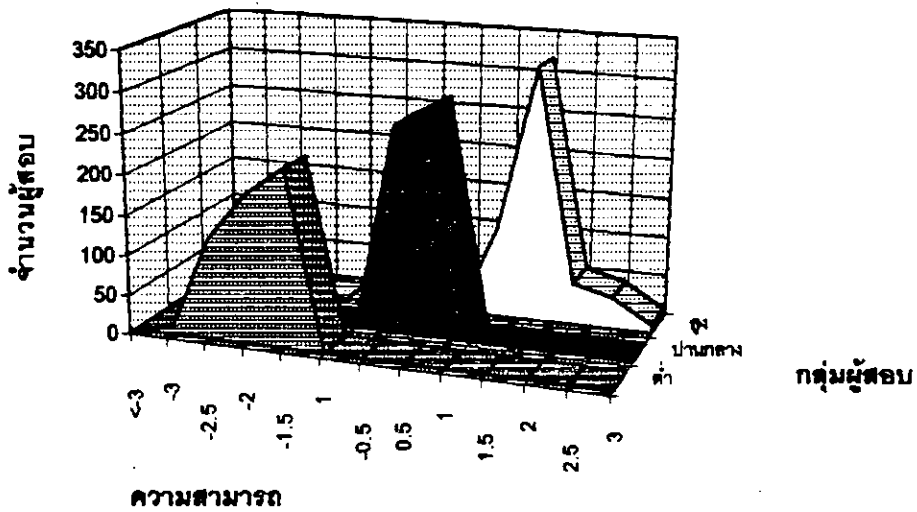


ภาพประกอบ 39 โค้งสารสนเทศของแบบทดสอบ(test information curve) วัดความสามารถทาง
คณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 จำนวน 35 ข้อ

จากภาพประกอบ 39 พบว่า แบบทดสอบวัดความสามารถทางคณิตศาสตร์ ชั้นประถม
ศึกษาปีที่ 6 จำนวน 35 ข้อนี้ เหมาะสมกับผู้สอบที่มีความสามารถปานกลาง ค่อนข้างไปทางความ
สามารถสูง

ตอนที่ 2 ผลการทดสอบความสามารถทางคณิตศาสตร์และความเข้าใจ เรื่อง เศษส่วน
จำแนกตามตัวแปรอิสระที่ศึกษา ซึ่งผลการวิเคราะห์ข้อมูล มีดังนี้

ผู้วิจัยได้แบ่งกลุ่มตัวอย่างออกเป็น 3 กลุ่ม คือ กลุ่มที่มีความสามารถทางคณิตศาสตร์สูง
(ค่าความสามารถสูงกว่า .60) ความสามารถปานกลาง (ค่าความสามารถระหว่าง -.50 ถึง +.50)
และความสามารถต่ำ (ค่าความสามารถต่ำกว่า -.60) ซึ่งลักษณะของการแจกแจงของกลุ่ม
ตัวอย่างดังกล่าวแสดงให้เห็นได้ดังภาพประกอบ 40



ภาพประกอบ 40 จำนวนนักเรียนจำแนกตามระดับความสามารถทางคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6

จากภาพประกอบ 38 พบว่า จำนวนนักเรียนในแต่ละกลุ่ม คือ กลุ่มที่มีความสามารถทางคณิตศาสตร์สูง ปานกลางและต่ำ มีจำนวนเท่ากัน คือ กลุ่มละ 540 คน รวมทั้งหมด 1,620 คน ซึ่งการกระจายของกลุ่มตัวอย่าง 3 กลุ่มมีลักษณะดังภาพประกอบ 40 จะเห็นว่ากลุ่มตัวอย่างที่มีความสามารถทางคณิตศาสตร์ต่ำมีลักษณะการแจกแจงคล้ายกับกลุ่มตัวอย่างที่มีความสามารถทางคณิตศาสตร์ปานกลาง ส่วนกลุ่มตัวอย่างที่มีความสามารถทางคณิตศาสตร์สูงนั้นมีการแจกแจงแตกต่างไปจากกลุ่มตัวอย่าง 2 กลุ่มเล็กน้อย กล่าวคือ การแจกแจงของกลุ่มตัวอย่างที่มีความสามารถทางคณิตศาสตร์สูงนั้นมีลักษณะโค้ง

ในการนำเสนอผลของตัวแปรตาม คือ ความเที่ยงตรงเชิงสภาพและจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์นั้น ในช่วงแรกได้เสนอเป็นภาพรวมจำแนกตามตัวแปรอิสระแต่ละตัว แต่ยังไม่ได้มีการเปรียบเทียบความแตกต่าง ในช่วงหลังจึงได้เปรียบเทียบค่าของตัวแปรตามจำแนกตามระดับของตัวแปรอิสระแต่ละตัว

เมื่อใช้ค่าความยากง่ายของข้อสอบข้อแรกเป็นเกณฑ์ในการศึกษาความเที่ยงตรงเชิงสภาพและจำนวนข้อสอบที่ใช้ในการทดสอบ ได้ข้อค้นพบดังตาราง 11

ตาราง 11 ค่าความเที่ยงตรงเชิงสภาพ จำนวนข้อสอบสูงสุด ต่ำสุด ค่าเฉลี่ย และค่าเบี่ยงเบน
มาตรฐานของจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ
ด้วยคอมพิวเตอร์ เรื่อง เศษส่วน จำแนกตามค่าความยากง่ายของข้อสอบข้อแรก

ความยากง่าย ของข้อแรก	ความเที่ยงตรงเชิงสภาพ		จำนวนข้อสอบ			
	r	Z	สูงสุด	ต่ำสุด	ค่าเฉลี่ย	ค่าเบี่ยงเบนฯ
ง่าย	.7419	.9547	28	8	14.58	5.25
ปานกลาง	.7103	.8877	28	7	13.81	5.00
ยาก	.7497	.9722	27	8	14.99	5.42

จากตาราง 11 พบว่า เมื่อเริ่มต้นทดสอบด้วยข้อสอบข้อแรกที่ง่าย ปานกลางและยาก
มีความเที่ยงตรงเชิงสภาพ เท่ากับ .7419 , .7130 และ .7497 ตามลำดับ โดยใช้ข้อสอบประมาณ
13 -14 ข้อ ค่าเบี่ยงเบนมาตรฐานของจำนวนข้อสอบที่ใช้ในการทดสอบมีค่าใกล้เคียงกัน ประมาณ
5.00

เมื่อใช้เกณฑ์การคัดเลือกข้อสอบข้อต่อไปในคลังข้อสอบเป็นเกณฑ์ในการศึกษาความ
เที่ยงตรงเชิงสภาพและจำนวนข้อสอบที่ใช้ ได้ผลการวิเคราะห์ ดังตาราง 12

ตาราง 12 ความเที่ยงตรงเชิงสภาพ จำนวนข้อสอบที่สูงที่สุด ต่ำสุด ค่าเฉลี่ยและค่าเบี่ยงเบน
มาตรฐานของจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ
ด้วยคอมพิวเตอร์ เรื่อง เศษส่วน จำแนกตามเกณฑ์การคัดเลือกข้อสอบในคลังข้อสอบ

เกณฑ์คัดเลือก ข้อสอบ	ความเที่ยงตรงเชิงสภาพ		จำนวนข้อสอบ			
	r	Z	สูงสุด	ต่ำสุด	ค่าเฉลี่ย	ค่าเบี่ยงเบนฯ
ค่าความยากง่าย	.7148	.8969	33	9	17.59	5.78
ค่า Qmax	.7404	.9513	29	8	14.45	5.23
ค่า Imax	.7444	.9603	22	6	11.34	3.76

จากตาราง 12 พบว่า เมื่อใช้เกณฑ์การคัดเลือกข้อสอบข้อต่อไปจากคลังข้อสอบด้วยวิธีใช้ค่าความยากง่าย ใช้ค่าความสามารถที่ทำให้ข้อสอบมีสารสนเทศสูงสุด และ ใช้ข้อสอบที่มีค่าสารสนเทศของข้อสอบสูงสุด จะมีความเที่ยงตรงเชิงสภาพ เท่ากับ .7148 , .7404 และ .7444 ตามลำดับ โดยใช้ข้อสอบประมาณ 17 , 14 และ 11 ข้อตามลำดับ เป็นที่น่าสังเกตว่า การคัดเลือกข้อสอบด้วยวิธีใช้ข้อสอบที่มีค่าสารสนเทศสูงสุดนั้นได้ใช้ข้อสอบน้อยกว่าการคัดเลือกข้อสอบด้วยวิธีอื่นๆ อีกทั้งมีค่าเบี่ยงเบนมาตรฐานของจำนวนข้อสอบที่ใช้ต่ำกว่าอีกด้วย

เมื่อใช้วิธีการประมาณค่าความสามารถของผู้สอบเป็นเกณฑ์ในการศึกษาความเที่ยงตรงเชิงสภาพและจำนวนข้อสอบที่ใช้ในการทดสอบ ได้ผลการวิเคราะห์ ดังตาราง 13

ตาราง 13 ความเที่ยงตรงเชิงสภาพ จำนวนข้อสอบที่สูงที่สุด ต่ำสุด ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ เรื่อง เศษส่วน จำแนกตามวิธีการประมาณค่าความสามารถของผู้สอบ

วิธีประมาณ ค่าความสามารถ	ความเที่ยงตรงเชิงสภาพ		จำนวนข้อสอบ			
	r	Z	สูงสุด	ต่ำสุด	ค่าเฉลี่ย	ค่าเบี่ยงเบนฯ
วิธี Bayesian	.7608	.9981	25	10	15.10	3.58
วิธี MLE	.7026	.8724	31	5	13.82	6.27

จากตาราง 13 พบว่า เมื่อใช้วิธีการประมาณค่าความสามารถด้วยวิธีของเบส์และวิธีความเป็นไปได้สูงสุด จะมีความเที่ยงตรงเชิงสภาพ เท่ากับ .7608 และ .7026 และใช้ข้อสอบประมาณ 15 และ 14 ข้อ ตามลำดับ วิธีของเบส์จะมีค่าเบี่ยงเบนมาตรฐานของจำนวนข้อสอบที่ใช้ในการทดสอบต่ำกว่าวิธีความเป็นไปได้สูงสุด

เมื่อใช้เกณฑ์ยุติการทดสอบเป็นเกณฑ์ในการศึกษาความเที่ยงตรงเชิงสภาพและจำนวนข้อสอบที่ใช้ในการทดสอบ ได้ข้อค้นพบบังตาราง 14

ตาราง 14 ความเที่ยงตรงเชิงสภาพ จำนวนข้อสอบที่สูงสุด ต่ำสุด ค่าเฉลี่ยและค่าเบี่ยงเบน
มาตรฐานของจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ
ด้วยคอมพิวเตอร์ เรื่อง เศษส่วน จำแนกตามเกณฑ์ยุติการทดสอบ

เกณฑ์ ยุติการทดสอบ	ความเที่ยงตรงเชิงสภาพ		จำนวนข้อสอบ			
	r	Z	สูงสุด	ต่ำสุด	ค่าเฉลี่ย	ค่าเบี่ยงเบนฯ
SEE ≤ .30	.7710	1.0228	46	13	23.14	8.11
SEE ≤ .40	.7323	.9337	24	7	12.14	4.18
SEE ≤ .50	.6869	.8421	14	4	7.80	2.49

จากตาราง 14 พบว่า เมื่อใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความ
สามารถ เท่ากับหรือน้อยกว่า .30 , .40 และ .50 เป็นเกณฑ์ยุติการทดสอบจะมีความเที่ยงตรง
เชิงสภาพ เท่ากับ .7710 , .7323 และ .6869 โดยใช้ข้อสอบประมาณ 23 12 และ 8 ข้อ ตามลำดับ
เป็นที่น่าสังเกตว่า เมื่อใช้เกณฑ์ความคลาดเคลื่อนมาตรฐานในการประมาณค่า เท่ากับหรือน้อย
กว่า .30 จำนวนข้อสอบที่ใช้ในการทดสอบมีค่าเบี่ยงเบนมาตรฐานสูงกว่าเกณฑ์อื่นๆ คือ 8.11

เมื่อใช้ความสามารถทางคณิตศาสตร์ของผู้สอบเป็นเกณฑ์ในการศึกษาความเที่ยงตรง
เชิงสภาพและจำนวนข้อสอบที่ใช้ ได้ผลการวิเคราะห์ ดังตาราง 15

ตาราง 15 ความเที่ยงตรงเชิงสภาพ จำนวนข้อสอบที่สูงสุด ต่ำสุด ค่าเฉลี่ยและค่าเบี่ยงเบน
มาตรฐานของจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ
ด้วยคอมพิวเตอร์ เรื่อง เศษส่วน จำแนกตามความสามารถทางคณิตศาสตร์ของผู้สอบ

ความสามารถ ทางคณิตศาสตร์	ความเที่ยงตรงเชิงสภาพ		จำนวนข้อสอบที่ใช้			
	r	Z	สูงสุด	ต่ำสุด	ค่าเฉลี่ย	ค่าเบี่ยงเบนฯ
สูง	.4449	.4783	24	6	11.33	4.28
ปานกลาง	.3173	.3286	26	8	13.67	4.16
ต่ำ	.2994	.3089	35	9	18.36	6.34

จากตาราง 15 พบว่า เมื่อผู้สอบที่มีความสามารถทางคณิตศาสตร์สูง ปานกลาง และต่ำ มาทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จะมีความเที่ยงตรงเชิงสภาพ เท่ากับ .4449 , .3173 และ .2994 โดยใช้ข้อสอบ ประมาณ 11 , 14 และ 18 ข้อ ตามลำดับ ข้อสอบที่ใช้ในกลุ่มที่มีความสามารถต่ำมีค่าเบี่ยงเบนมาตรฐานสูงกว่าอีก 2 กลุ่ม เมื่อศึกษาความแตกต่างของค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ โดยใช้ค่าไคสแควร์ ซึ่งได้ผลการวิเคราะห์ ดังตาราง 16

ตาราง 16 ค่าความเที่ยงตรงเชิงสภาพ และค่าไคสแควร์ สำหรับทดสอบความแตกต่างของค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์

ข้อแรก	ยากง่าย	เลือก	เกณฑ์ ยุดิ	วิธี MLE			วิธีของ Bay		
				สูง	กลาง	ต่ำ	สูง	กลาง	ต่ำ
ง่าย	b		.30	.3200	.4382	.5204	.6548	.3138	-.1007
			.40	.4536	.5602	.7125	.5765	.2417	.0323
			.50	.3659	.4860	.6864	.4084	.3380	.0108
	Qmax		.30	.4235	.3336	.4456	.5930	.2554	.0666
			.40	.1593	.4605	.4444	.6232	.2926	-.0139
			.50	.3878	.4095	.4472	.6060	.3651	-.2055
	Imax		.30	.3815	.5803	.4534	.2242	.3392	.1847
			.40	.3250	.5723	.3358	.3510	.2147	.1500
			.50	.4497	.4084	.4102	.3136	.1398	.2373
กลาง	b		.30	.7368	.0089	.2979	.1868	.3140	.2599
			.40	.7911	.2336	.6086	.2150	.2109	.3801
			.50	.8547	.1471	.5033	.1994	.2393	.3795
	Qmax		.30	.2912	.1292	.4620	.3643	.3232	.3705
			.40	.3785	.1437	.2334	.3240	.0664	.2942
			.50	.2687	.2439	.1969	.2040	.0520	.1417

ข้อแรก	เกณฑ์ เลือก	เกณฑ์ ยุติ	วิธี MLE			วิธีของ Bay		
			สูง	กลาง	ต่ำ	สูง	กลาง	ต่ำ
ยาก	lmax	.30	.4157	.3171	.4438	.3616	.2020	.2124
		.40	.1522	.2799	.2276	.4495	.0693	.0729
		.50	.3118	.3366	-.0084	.4608	.0169	-.0996
	b	.30	.4585	.2475	.1415	.5583	.5441	-.0538
		.40	.4641	.2682	.1321	.5776	.5240	.0026
		.50	.6086	.2411	.0491	.5479	.4941	-.0041
	Qmax	.30	.4811	.1921	.6742	.4240	.4924	.2569
		.40	.3691	.0867	.5556	.3826	.3484	.2194
		.50	.0808	.0997	.5938	.5113	.2712	.2006
	lmax	.30	.5192	.2185	.3421	.6147	.5773	.2957
		.40	.5282	.1089	.2038	.6552	.5225	.1634
		.50	.6166	.3950	.3169	.4521	.4182	.1033
$\chi^2 =$			248.23**					

** มีนัยสำคัญทางสถิติที่ระดับ .01

จากตาราง 16 พบว่า ผลการทดสอบภาพรวมของค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์นั้น มีอย่างน้อย 1 คู่ ที่แตกต่างกัน

เมื่อนำค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์มาทดสอบความแตกต่างเป็นรายคู่ในแต่ละระดับของตัวแปรอิสระ 5 ตัวแปร คือ ความยากง่ายของข้อสอบข้อแรกที่ใช้ในการทดสอบ เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบและความสามารถของผู้สอบ ซึ่งผลการทดสอบปรากฏ ดังตาราง 17 - 25

ตาราง 17 ความแตกต่างของค่าความเที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะกับความ
 สามารถของผู้สอบด้วยคอมพิวเตอร์ เรื่อง เศษส่วน จำแนกตามความยากง่ายของข้อสอบ
 ข้อแรก

ความเที่ยงตรง เชิงสภาพ	r ง่าย	r ปานกลาง	r ยาก
r ง่าย = .7419	-	.0316	.0078
r ปานกลาง = .7103		-	.0394
r ยาก = .7497			-

จากตาราง 17 พบว่า เมื่อใช้ข้อสอบข้อแรกที่มีค่าความยากง่ายต่างกันในการเริ่มต้น
 ทดสอบ ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแตกต่างกันอย่างไม่มีนัยสำคัญ

ตาราง 18 ความแตกต่างของค่าความเที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะกับความ
 สามารถของผู้สอบด้วยคอมพิวเตอร์ เรื่อง เศษส่วน จำแนกตามเกณฑ์การคัดเลือกข้อสอบจาก
 คลังข้อสอบ

ความเที่ยงตรง เชิงสภาพ	r ความยากง่าย	r Qmax	r lmax
r ความยากง่าย = .7148	-	.0256	.0296
r Qmax = .7404		-	.004
r lmax = .7444			-

จากตาราง 18 พบว่า เมื่อใช้เกณฑ์ในการคัดเลือกข้อสอบจากคลังข้อสอบต่างกัน ค่า
 ความเที่ยงตรงเชิงสภาพของการทดสอบแตกต่างกันอย่างไม่มีนัยสำคัญ

ตาราง 19 ความเที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ เรื่อง เศษส่วน จำแนกตามวิธีการประมาณค่าความสามารถและสถิติทดสอบซี

วิธีการประมาณค่า	ความเที่ยงตรง เชิงสภาพ		Z
ความสามารถ	r	Z_r	
วิธี Bayesian	.7608	.9981	4.3238**
วิธี MLE	.7026	.8724	

** มีนัยสำคัญทางสถิติที่ระดับ .01

จากตาราง 19 พบว่า เมื่อใช้วิธีการประมาณค่าความสามารถของผู้สอบด้วยวิธีของเบส์ และวิธีความเป็นไปได้สูงสุดในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ ทำให้ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแตกต่างกันอย่างมีนัยสำคัญที่ระดับ .01

ตาราง 20 ความแตกต่างของค่าความเที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ เรื่อง เศษส่วน จำแนกตามเกณฑ์ยุติการทดสอบ

ความเที่ยงตรง เชิงสภาพ	$r_{SEE \leq .30}$	$r_{SEE \leq .40}$	$r_{SEE \leq .50}$
$r_{SEE \leq .30} = .7710$	-	.0387 *	.0841*
$r_{SEE \leq .40} = .7323$		-	.0454*
$r_{SEE \leq .50} = .6869$			-

* มีนัยสำคัญทางสถิติที่ระดับ .05

จากตาราง 20 พบว่า เมื่อใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถเท่ากับหรือน้อยกว่า .30 ค่าความเที่ยงตรงเชิงสภาพจะต่างจากการใช้เกณฑ์ .40 และ .50 อย่างมีนัยสำคัญที่ระดับ .05 และเมื่อใช้เกณฑ์ .40 ในการยุติการทดสอบค่าความเที่ยงตรง

เชิงสภาพของการทดสอบแตกต่างกันกับการใช้เกณฑ์ .50 อย่างมีนัยสำคัญ ที่ระดับ .05 เช่นกัน

ตาราง 21 ความแตกต่างของค่าความเที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะกับความ
สามารถของผู้สอบด้วยคอมพิวเตอร์ เรื่อง เศษส่วน จำแนกตามความสามารถทางคณิตศาสตร์
ของผู้สอบ

ความเที่ยงตรง เชิงสภาพ	ร สูง	ร ปานกลาง	ร ต่ำ
ร สูง = .4449	-	.1276*	.1455*
ร ปานกลาง = .3173		-	.0179
ร ต่ำ = .2994			-

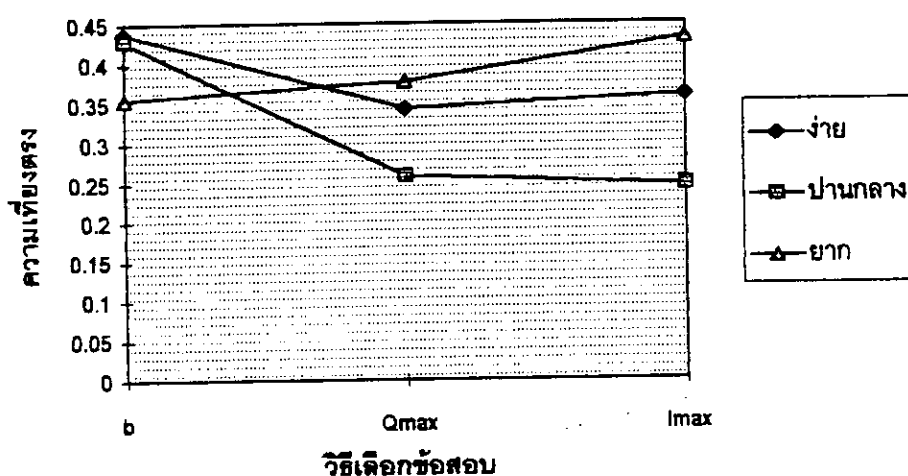
* มีนัยสำคัญทางสถิติที่ระดับ .05

จากตาราง 21 พบว่า กลุ่มผู้สอบที่มีความสามารถทางคณิตศาสตร์สูง จะมีค่าความ
เที่ยงตรงเชิงสภาพของการทดสอบแตกต่างจากกลุ่มผู้สอบที่มีความสามารถปานกลางและต่ำ
อย่างมีนัยสำคัญ ที่ระดับ .05 ส่วนผู้สอบที่มีความสามารถปานกลาง และต่ำ จะมีความเที่ยงตรง
เชิงสภาพไม่ต่างกัน

ในการพิจารณาปฏิสัมพันธ์ของตัวแปรอิสระที่ร่วมกันส่งผลต่อค่าความเที่ยงตรง
เชิงสภาพนั้น ผู้วิจัยได้นำค่าสหสัมพันธ์ในรูปของคะแนนพิชเชอร์ซีมาสร้างเป็นภาพ เพื่อแสดง
ปฏิสัมพันธ์ แบบ 2 ทาง 3 ทาง 4 ทางและ 5 ทาง ต่อความเที่ยงตรงเชิงสภาพ ซึ่งแสดงได้
ดังภาพประกอบ 41-66 เหตุผลที่ไม่สามารถประมาณค่าปฏิสัมพันธ์ออกมาเป็นจำนวนได้นั้น
เพราะค่าสหสัมพันธ์ระหว่างค่าความสามารถทางคณิตศาสตร์ กับค่าความสามารถ เรื่อง เศษส่วน
ที่ได้จากการทดสอบแบบปรับเหมาะความสามารถของผู้สอบด้วยคอมพิวเตอร์ที่ผู้วิจัยใช้เป็นค่า
ความเที่ยงตรงเชิงสภาพนั้น หาได้กลุ่มละ 1 ค่า ดังนั้นในแต่ละกลุ่มจึงไม่มีค่าความแปรปรวนของ
ค่าความเที่ยงตรงเชิงสภาพดังกล่าว จึงทำให้ไม่สามารถวิเคราะห์หาค่าปฏิสัมพันธ์ได้ แต่ผู้วิจัย
ต้องการจะทราบว่าตัวแปรอิสระทั้ง 5 ตัวมีตัวแปรใดบ้างที่ร่วมกันส่งผลต่อค่าความเที่ยงตรง
เชิงสภาพของการทดสอบ จึงได้เสนอในลักษณะของภาพประกอบ โดยเทียบเคียงกับการวิเคราะห์
ความแปรปรวน กล่าวคือ หากเส้นตรงที่ต่อเชื่อมระหว่างค่าของความเที่ยงตรงเชิงสภาพอันเนื่องมา

จากตัวแปรอิสระตัดกัน แสดงว่าตัวแปรอิสระในกลุ่มมีปฏิสัมพันธ์ร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพ

เมื่อพิจารณาปฏิสัมพันธ์ระหว่างความยากง่ายของข้อสอบข้อแรก เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบที่คาดว่าจะร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ แสดงได้ดังภาพประกอบ 41

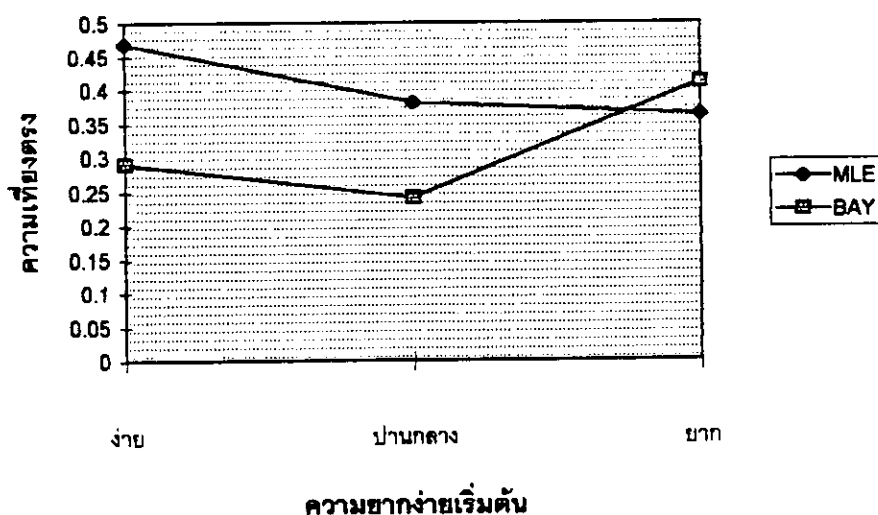


ภาพประกอบ 41 ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จำแนกตามความยากง่ายของข้อสอบข้อแรกและเกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ

จากภาพประกอบ 41 พบว่า ค่าความเที่ยงตรงเชิงสภาพมีพิสัยอยู่ในช่วง .2459 ถึง .4381 ความเที่ยงตรงเชิงสภาพมีค่าสูงสุด เมื่อเริ่มต้นทดสอบด้วยข้อสอบที่ง่าย คัดเลือกข้อสอบในคลังข้อสอบด้วยค่าความยากง่ายของข้อสอบ มีค่าต่ำสุด เมื่อเริ่มต้นทดสอบด้วยข้อสอบที่ยาก ปานกลาง คัดเลือกข้อสอบในคลังข้อสอบด้วยค่าสารสนเทศสูงสุดของข้อสอบ

จากลักษณะการตัดกันของเส้นตรงในภาพ พออนุมานได้ว่า ความยากง่ายของข้อสอบข้อแรกและเกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์

เมื่อพิจารณาปฏิสัมพันธ์ระหว่างความยากง่ายของข้อสอบข้อแรกและวิธีประมาณค่าความสามารถของผู้สอบที่คาดว่าจะร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ แสดงได้ดังภาพประกอบ 42

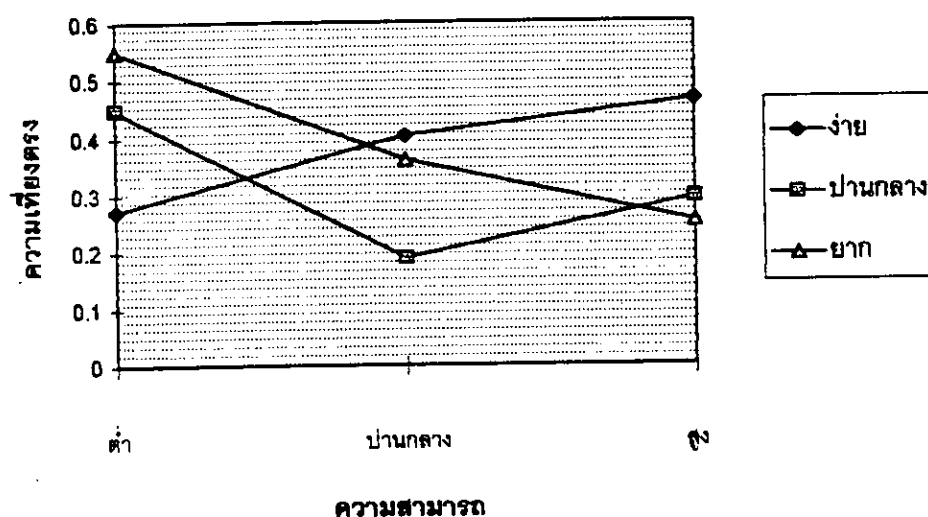


ภาพประกอบ 42 ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จำแนกตามความยากง่ายของข้อสอบข้อแรกและวิธีการประมาณค่าความสามารถของผู้สอบ

จากภาพประกอบ 42 พบว่า ความเที่ยงตรงเชิงสภาพมีพิสัยอยู่ในช่วง .2412 ถึง .4687 ความเที่ยงตรงเชิงสภาพมีค่าสูงสุด เมื่อเริ่มต้นทดสอบด้วยข้อสอบที่ง่าย ประมาณค่าความสามารถของผู้สอบด้วยวิธีความเป็นไปได้สูงสุด มีค่าต่ำสุด เมื่อเริ่มต้นทดสอบด้วยข้อสอบที่ยาก ปานกลาง ประมาณค่าความสามารถของผู้สอบด้วยวิธีของเบส์

จากลักษณะการตัดกันของเส้นตรงในภาพ พออนุมานได้ว่า ความยากง่ายของข้อสอบข้อแรก และวิธีการประมาณค่าความสามารถของผู้สอบร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์

เมื่อพิจารณาปฏิสัมพันธ์ระหว่างความยากง่ายของข้อสอบข้อแรกและความสามารถของผู้สอบที่คาดว่าจะร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ แสดงได้ดังภาพประกอบ 43

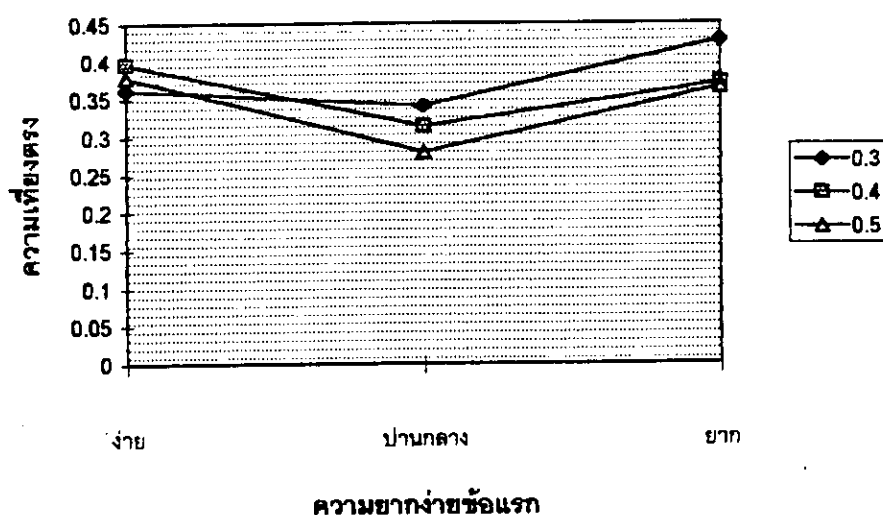


ภาพประกอบ 43 ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จำแนกตามความยากง่ายของข้อสอบข้อแรกและความสามารถของผู้สอบ

จากภาพประกอบ 43 พบว่า ค่าความเที่ยงตรงเชิงสภาพมีพิสัยอยู่ในช่วง .1896 ถึง .5496 ความเที่ยงตรงเชิงสภาพมีค่าสูงสุด เมื่อเริ่มต้นทดสอบด้วยข้อสอบที่ยาก ผู้สอบมีความสามารถต่ำ มีค่าต่ำสุด เมื่อเริ่มต้นทดสอบด้วยข้อสอบที่ยากปานกลาง ผู้สอบมีความสามารถปานกลาง

จากลักษณะการตัดกันของเส้นตรงในภาพ พออนุมานได้ว่า ความยากง่ายของข้อสอบข้อแรกและความสามารถของผู้สอบ ร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์

เมื่อพิจารณาปฏิสัมพันธ์ระหว่างความยากง่ายของข้อสอบข้อแรกและเกณฑ์ยุติการทดสอบที่คาดว่าจะร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ แสดงได้ดังภาพประกอบ 44

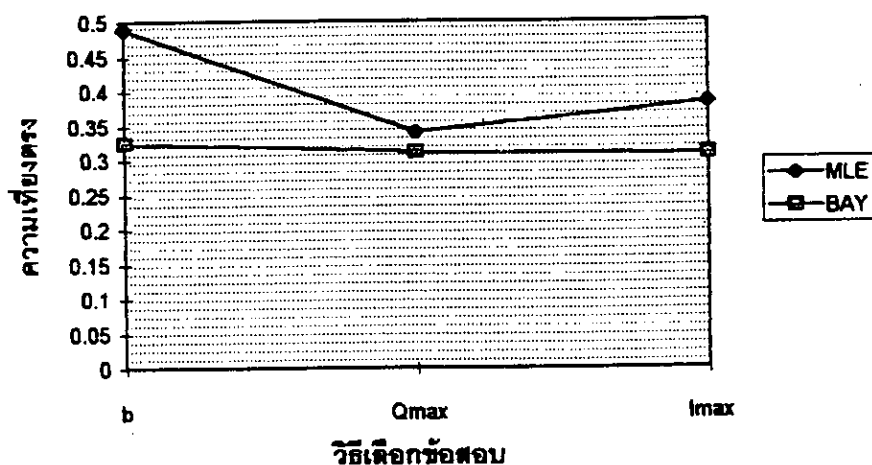


ภาพประกอบ 44 ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จำแนกตามความยากง่ายของข้อสอบข้อแรก เกณฑ์ยุติการทดสอบ

จากภาพประกอบ 44 พบว่า ค่าความเที่ยงตรงเชิงสภาพมีพิสัยอยู่ในช่วง .2792 ถึง .4279 ความเที่ยงตรงเชิงสภาพมีค่าสูงสุด เมื่อเริ่มต้นทดสอบด้วยข้อสอบที่ยาก ยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ เท่ากับหรือน้อยกว่า .30 มีค่าต่ำสุด เมื่อเริ่มต้นทดสอบด้วยข้อสอบที่ยากปานกลาง ยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ เท่ากับหรือน้อยกว่า .50

จากลักษณะการตัดกันของเส้นตรงในภาพ พออนุมานได้ว่า ความยากง่ายของข้อสอบข้อแรกและเกณฑ์ยุติการทดสอบร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์

เมื่อพิจารณาปฏิสัมพันธ์ระหว่าง เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบ ที่คาดว่าจะร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ แสดงได้ดังภาพประกอบ 45

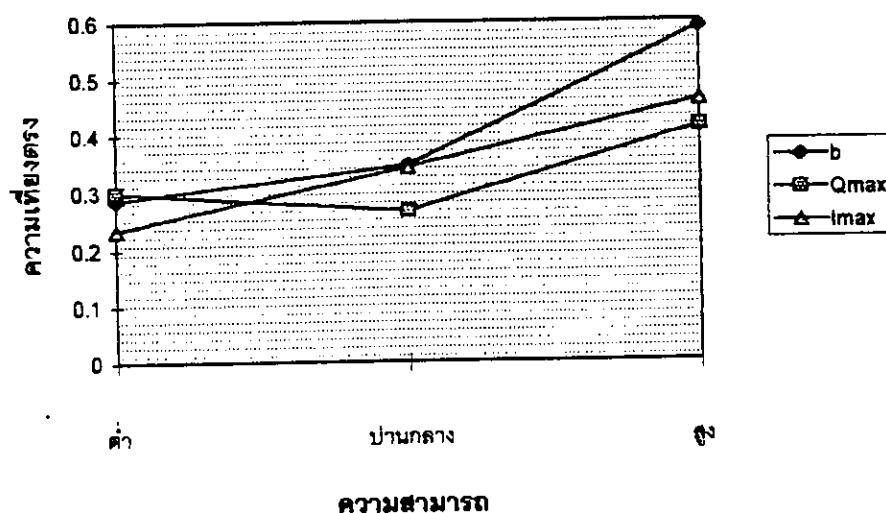


ภาพประกอบ 45 ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จำแนกตาม เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบ

จากภาพประกอบ 45 พบว่า ค่าความเที่ยงตรงเชิงสภาพมีพิสัยอยู่ในช่วง .3073 ถึง .4891 ความเที่ยงตรงเชิงสภาพมีค่าสูงสุด เมื่อคัดเลือกข้อสอบในคลังข้อสอบด้วยค่าความยากง่ายของข้อสอบ ประมาณค่าความสามารถของผู้สอบด้วยวิธีความเป็นไปได้สูงสุด มีค่าต่ำสุด เมื่อคัดเลือกข้อสอบในคลังข้อสอบด้วยค่าสารสนเทศสูงสุดของข้อสอบ ประมาณค่าความสามารถของผู้สอบด้วยวิธีของเบส์

จากลักษณะการตัดกันของเส้นตรงในภาพ พออนุมานได้ว่า เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบไม่ร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์

เมื่อพิจารณาปฏิสัมพันธ์ระหว่าง เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ และความสามารถของผู้สอบที่คาดว่าจะร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ แสดงได้ดังภาพประกอบ 46

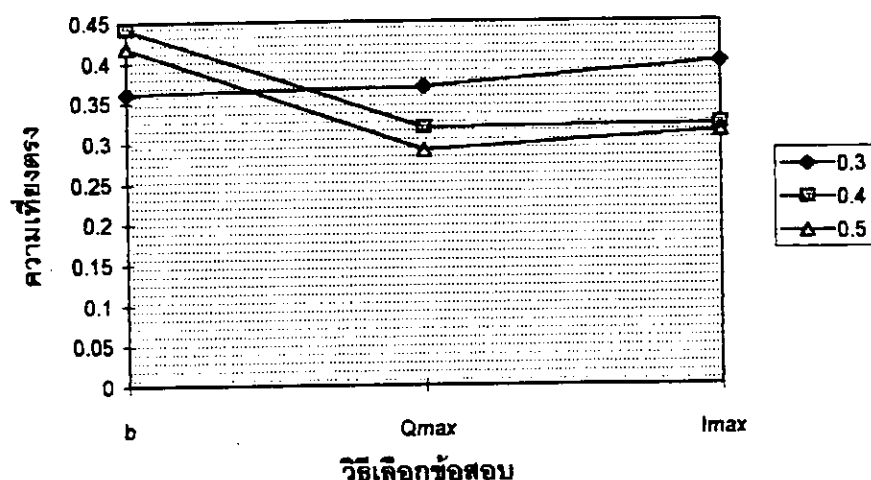


ภาพประกอบ 46 ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จำแนกตามเกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบและความสามารถของผู้สอบ

จากภาพประกอบ 46 พบว่า ค่าความเที่ยงตรงเชิงสภาพมีพิสัยอยู่ในช่วง .2338 ถึง .5883 ความเที่ยงตรงเชิงสภาพมีค่าสูงสุด เมื่อคัดเลือกข้อสอบในคลังข้อสอบด้วยค่าความยากง่ายของข้อสอบและผู้สอบมีความสามารถสูง มีค่าต่ำสุด เมื่อคัดเลือกข้อสอบในคลังข้อสอบด้วยค่าค่าสารสนเทศสูงสุดของข้อสอบและผู้สอบมีความสามารถต่ำ

จากลักษณะการตัดกันของเส้นตรงในภาพ พออนุมานได้ว่า เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบและความสามารถของผู้สอบ ร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์

เมื่อพิจารณาปฏิสัมพันธ์ระหว่างความยากง่ายของเกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ เกณฑ์สำหรับยุติการทดสอบซึ่งคาดว่าจะร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ แสดงได้ดังภาพประกอบ 47

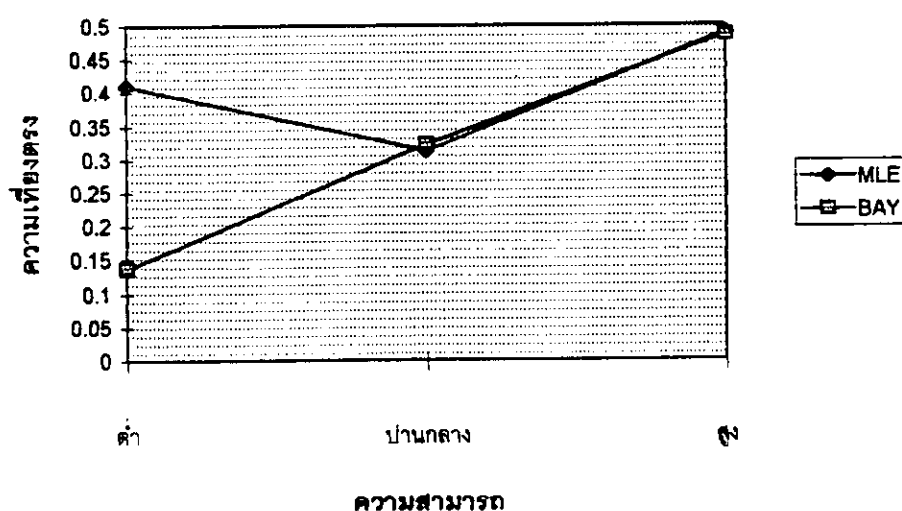


ภาพประกอบ 47 ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จำแนกตามเกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบและเกณฑ์ยุติการทดสอบ

จากภาพประกอบ 47 พบว่า ค่าความเที่ยงตรงเชิงสภาพมีพิสัยอยู่ในช่วง .2910 ถึง .4419 ความเที่ยงตรงเชิงสภาพมีค่าสูงสุด เมื่อคัดเลือกข้อสอบในคลังข้อสอบด้วยค่าความยากง่ายของข้อสอบ ยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถเท่ากับหรือน้อยกว่า .40 มีค่าต่ำสุด เมื่อคัดเลือกข้อสอบในคลังข้อสอบด้วยค่าความสามารถที่ทำให้ข้อสอบมีสารสนเทศสูงสุดและยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ เท่ากับหรือน้อยกว่า .50

จากลักษณะการตัดกันของเส้นตรงในภาพ พออนุมานได้ว่า เกณฑ์การคัดเลือกข้อสอบ จากคลังข้อสอบและเกณฑ์ยุติการทดสอบ ร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพของการ ทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์

เมื่อพิจารณาปฏิสัมพันธ์ระหว่างวิธีการประมาณค่าความสามารถของผู้สอบ และ ความสามารถของผู้สอบซึ่งคาดว่าจะร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพในการทดสอบ แบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ แสดงได้ดังภาพประกอบ 48



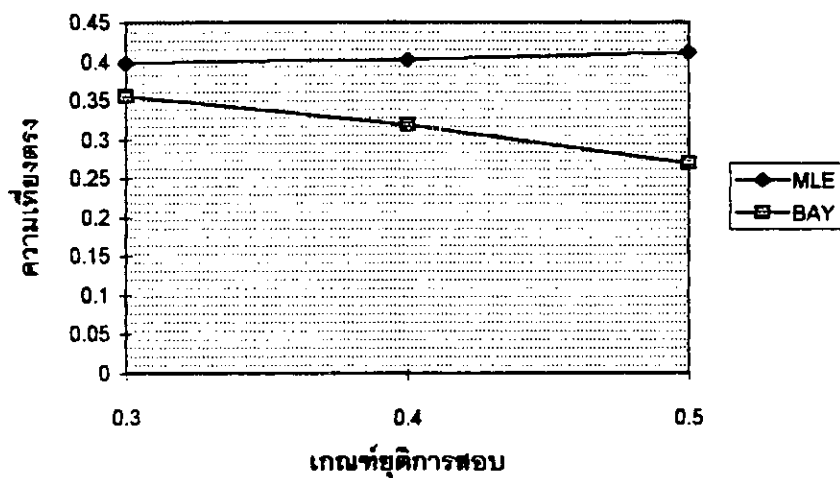
ภาพประกอบ 48 ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถ ของผู้สอบด้วยคอมพิวเตอร์ จำแนกตามวิธีการประมาณค่าความสามารถของผู้สอบและความ สามารถของผู้สอบ

จากภาพประกอบ 48 พบว่า ค่าความเที่ยงตรงเชิงสภาพมีพิสัยอยู่ในช่วง .1356 ถึง .4894 ความเที่ยงตรงเชิงสภาพมีค่าสูงสุดเมื่อ ประมาณค่าความสามารถของผู้สอบด้วยวิธีความเป็นไปได้สูงสุดและผู้สอบมีความสามารถสูง มีค่าต่ำสุด เมื่อประมาณค่าความสามารถของ ผู้สอบด้วยวิธีของเบย์และผู้สอบมีความสามารถต่ำ

จากลักษณะการตัดกันของเส้นตรงในภาพ พออนุมานได้ว่า วิธีการประมาณค่าความ สามารถของผู้สอบ และ ความสามารถของผู้สอบร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพของ

การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์

เมื่อพิจารณาปฏิสัมพันธ์ระหว่างข้อสอบข้อแรก วิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์สำหรับการยุติการทดสอบซึ่งคาดว่าจะร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ แสดงได้ดังภาพประกอบ 49

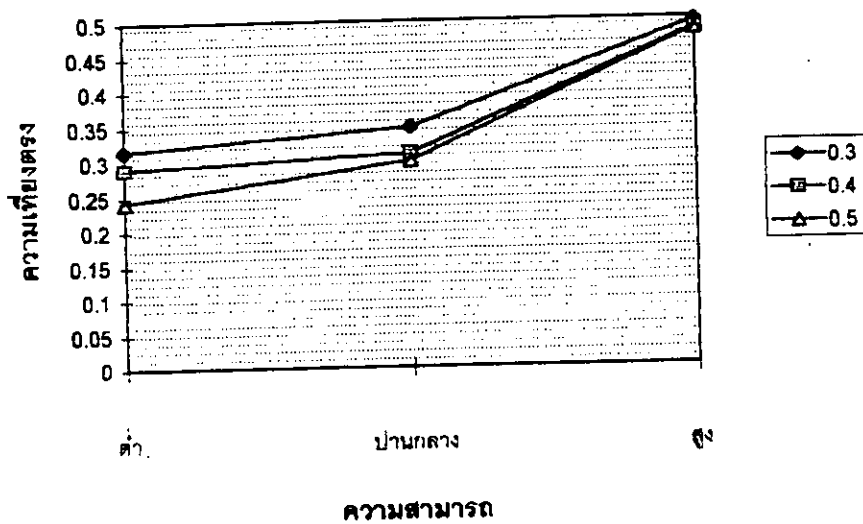


ภาพประกอบ 49 ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์จำแนกตามวิธีการประมาณค่าความสามารถของผู้สอบและเกณฑ์ยุติการทดสอบ

จากภาพประกอบ 49 พบว่า ค่าความเที่ยงตรงเชิงสภาพมีพิสัยอยู่ในช่วง .2696 ถึง .4122 ความเที่ยงตรงเชิงสภาพมีค่าสูงสุด เมื่อ ประมาณค่าความสามารถของผู้สอบด้วยวิธีความเป็นไปได้สูงสุดและยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ เท่ากับหรือน้อยกว่า.50 มีค่าต่ำสุด เมื่อ ประมาณค่าความสามารถของผู้สอบด้วยวิธีของเบส์และยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ เท่ากับหรือน้อยกว่า .50

จากลักษณะของเส้นตรงในภาพ พออนุมานได้ว่า วิธีการประมาณค่าความสามารถของผู้สอบ และ เกณฑ์ยุติการทดสอบ ไม่ร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์

เมื่อพิจารณาปฏิสัมพันธ์ระหว่าง เกณฑ์ยุติการทดสอบ และความสามารถของผู้สอบซึ่ง คาดว่าจะร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ แสดงได้ดังภาพประกอบ 50

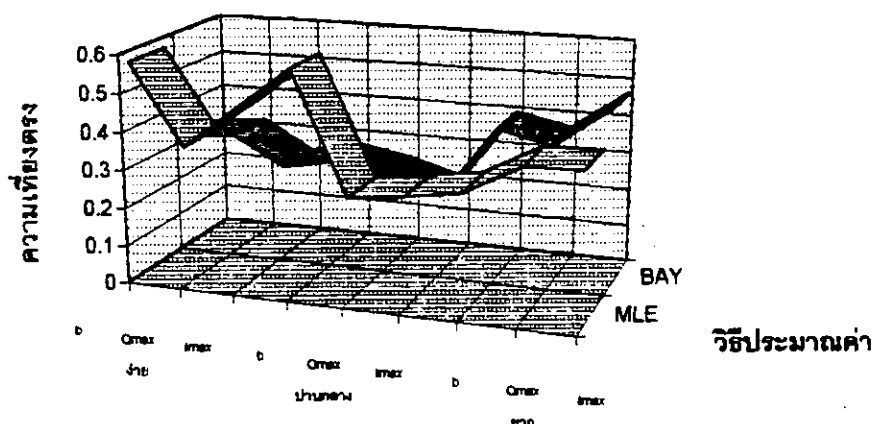


ภาพประกอบ 50 ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จำแนกตามเกณฑ์ยุติการทดสอบและความสามารถของผู้สอบ

จากภาพประกอบ 50 พบว่า ค่าความเที่ยงตรงเชิงสภาพมีพิสัยอยู่ในช่วง .2419 ถึง .4950 ความเที่ยงตรงเชิงสภาพมีค่าสูงสุด เมื่อยุติการทดสอบด้วยค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ เท่ากับหรือน้อยกว่า .30 และผู้สอบมีความสามารถสูง มีค่าต่ำสุด เมื่อยุติการทดสอบด้วยค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ เท่ากับหรือน้อยกว่า .50 และผู้สอบมีความสามารถต่ำ

จากลักษณะการตัดกันของเส้นตรงในภาพ พออนุมานได้ว่า เกณฑ์ยุติการทดสอบ และความสามารถของผู้สอบ ร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะ กับความสามารถของผู้สอบด้วยคอมพิวเตอร์

เมื่อพิจารณาปฏิสัมพันธ์ระหว่างความยากง่ายของข้อสอบข้อแรก เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบซึ่งคาดว่าจะร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะ กับความสามารถของผู้สอบด้วยคอมพิวเตอร์ แสดงได้ดังภาพประกอบ 51



ความยากง่ายข้อแรก/วิธีคัดเลือก

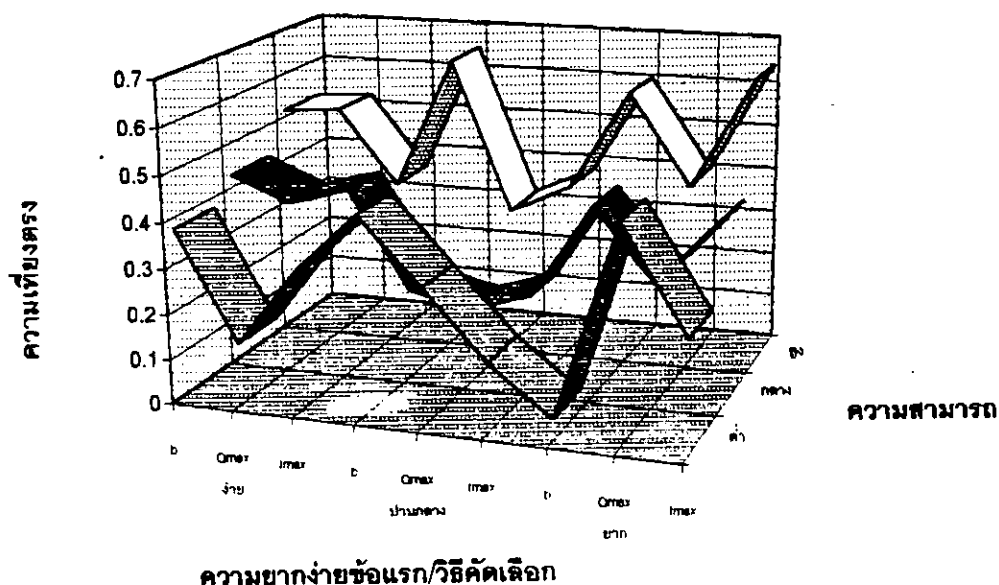
ภาพประกอบ 51 ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะ กับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จำแนกตามความยากง่ายของข้อสอบข้อแรก เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ และวิธีการประมาณค่าความสามารถของผู้สอบ

จากภาพประกอบ 51 พบว่า ค่าความเที่ยงตรงเชิงสภาพมีพิสัยอยู่ในช่วง .2046 ถึง .5838 ความเที่ยงตรงเชิงสภาพมีค่าสูงสุด เมื่อเริ่มต้นทดสอบด้วยข้อสอบที่ยากปานกลาง คัดเลือกข้อสอบในคลังข้อสอบด้วยค่าความยากง่ายของข้อสอบ ประมาณค่าความสามารถของผู้สอบด้วยวิธีความเป็นไปได้สูงสุด มีค่าต่ำสุด เมื่อเริ่มต้นทดสอบด้วยข้อสอบที่ยากปานกลาง

คัดเลือกข้อสอบในคลังข้อสอบด้วยค่าสารสนเทศสูงสุดของข้อสอบ ประมาณค่าความสามารถของผู้สอบด้วยวิธีของเบส์

จากลักษณะการตัดกันของเส้นตรงในภาพ พออนุมานได้ว่า ความยากง่ายของข้อสอบข้อแรก เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบ ร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์

เมื่อพิจารณาปฏิสัมพันธ์ระหว่างความยากง่ายของข้อสอบข้อแรก เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ และความสามารถของผู้สอบซึ่งคาดว่าจะร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์แสดงได้ดังภาพประกอบ 52



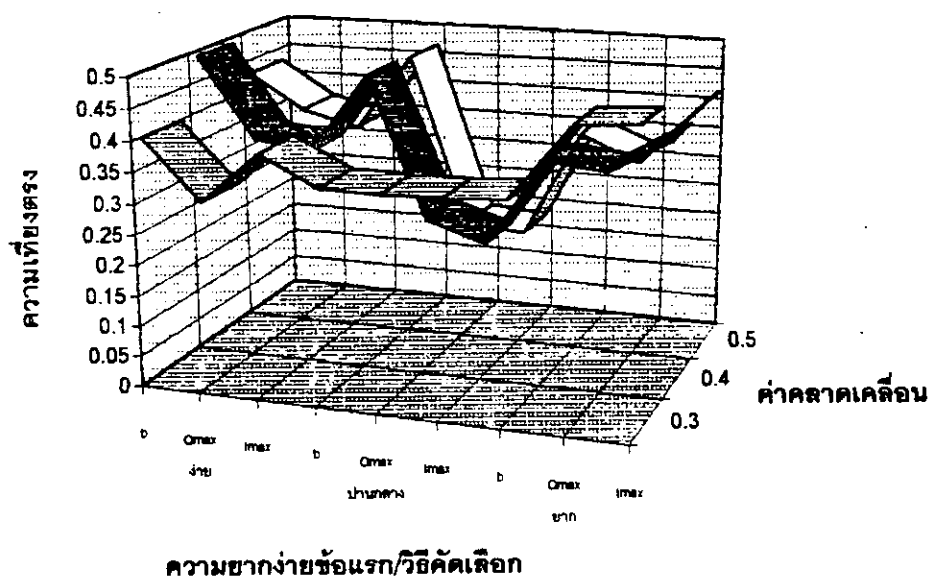
ภาพประกอบ 52 ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จำแนกตามความยากง่ายของข้อสอบข้อแรก เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบและความสามารถของผู้สอบ

จากภาพประกอบ 52 พบว่า ค่าความเที่ยงตรงเชิงสภาพมีพิสัยอยู่ในช่วง .0449 ถึง .6501 ความเที่ยงตรงเชิงสภาพมีค่าสูงสุด เมื่อเริ่มต้นทดสอบด้วยข้อสอบที่ยากปานกลาง คัดเลือกข้อสอบในคลังข้อสอบด้วยค่าความยากง่ายของข้อสอบและผู้สอบมีความสามารถสูง

มีค่าต่ำสุด เมื่อเริ่มต้นทดสอบด้วยข้อสอบที่ยาก คัดเลือกข้อสอบในคลังข้อสอบด้วยค่าความยากง่ายของข้อสอบและผู้สอบมีความสามารถต่ำ

จากลักษณะการตัดกันของเส้นตรงในภาพ พออนุมานได้ว่า ความยากง่ายของข้อสอบข้อแรก เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ และความสามารถของผู้สอบ ร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์

เมื่อพิจารณาปฏิสัมพันธ์ระหว่างความยากง่ายของข้อสอบข้อแรก เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบและเกณฑ์ยุติการทดสอบซึ่งคาดว่าจะร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ แสดงได้ดังภาพประกอบ 53



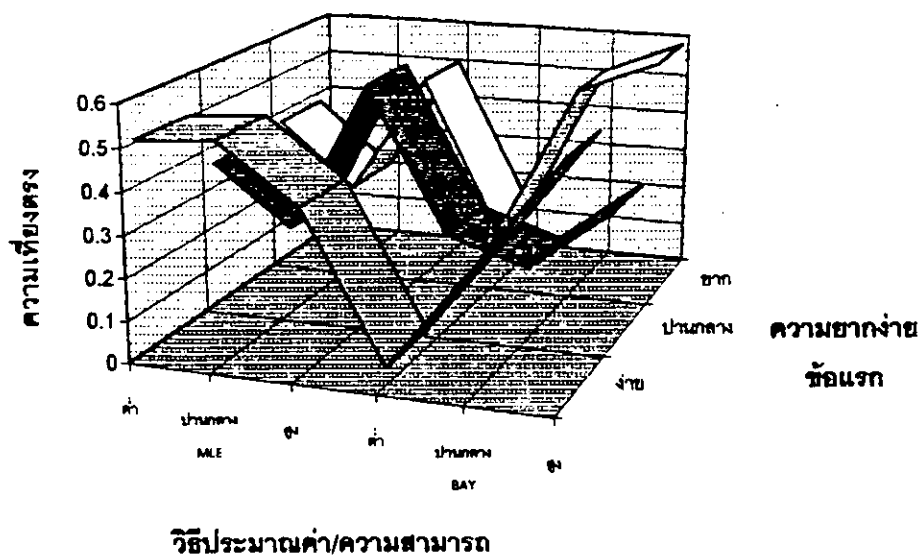
ภาพประกอบ 53 ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จำแนกตามความยากง่ายของข้อสอบข้อแรก เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบและเกณฑ์ยุติการทดสอบ

จากภาพประกอบ 53 พบว่า ค่าความเที่ยงตรงเชิงสภาพมีพิสัยอยู่ในช่วง .1800 ถึง .4914 ความเที่ยงตรงเชิงสภาพมีค่าสูงสุด เมื่อเริ่มต้นทดสอบด้วยข้อสอบที่ง่าย คัดเลือกข้อสอบในคลังข้อสอบด้วยค่าความยากง่ายของข้อสอบ และยุติการทดสอบเมื่อค่าความคลาดเคลื่อน

มาตรฐานในการประมาณค่าความสามารถ เท่ากับหรือน้อยกว่า .40 มีค่าต่ำสุด เมื่อเริ่มต้นทดสอบด้วยข้อสอบที่ยากปานกลาง คัดเลือกข้อสอบในคลังข้อสอบด้วยค่าสารสนเทศสูงสุดของข้อสอบและยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ เท่ากับหรือน้อยกว่า .50

จากลักษณะการตัดกันของเส้นตรงในภาพ พออนุมานได้ว่า ความยากง่ายของข้อสอบ ข้อแรก เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ เกณฑ์ยุติการทดสอบ ร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์

เมื่อพิจารณาปฏิสัมพันธ์ระหว่างความยากง่ายของข้อสอบข้อแรก วิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบ และความสามารถของผู้สอบซึ่งคาดว่าจะ ร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ แสดงได้ดังภาพประกอบ 54



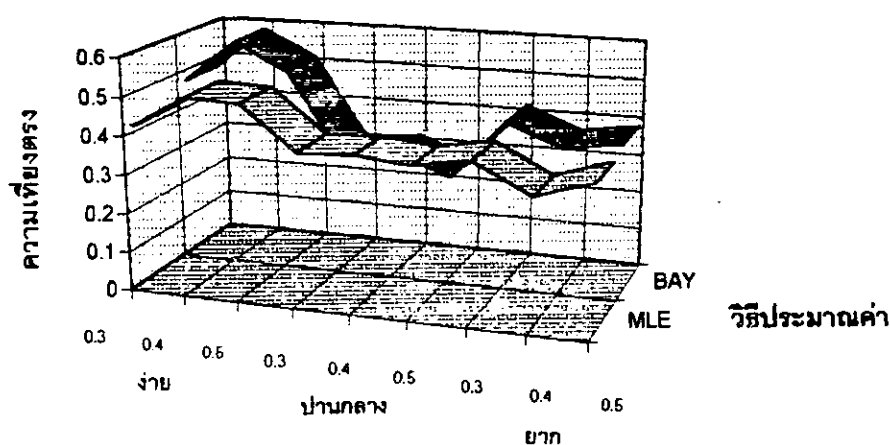
ภาพประกอบ 54 ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จำแนกตามความยากง่ายของข้อสอบข้อแรก วิธีการประมาณค่าความสามารถของผู้สอบและความสามารถของผู้สอบ

จากภาพประกอบ 54 พบว่า ค่าความเที่ยงตรงเชิงสภาพมีพิสัยอยู่ในช่วง .0405 ถึง

.5903 ความเที่ยงตรงเชิงสภาพมีค่าสูงสุด เมื่อเริ่มต้นทดสอบด้วยข้อสอบที่ยาก ประมาณค่าความสามารถของผู้สอบด้วยวิธีของเบส์และผู้สอบมีความสามารถสูง มีค่าต่ำสุด เมื่อเริ่มต้นทดสอบด้วยข้อสอบที่ง่าย ประมาณค่าความสามารถของผู้สอบด้วยวิธีของเบส์และผู้สอบมีความสามารถต่ำ

จากลักษณะการตัดกันของเส้นตรงในภาพ พออนุมานได้ว่า ความยากง่ายของข้อสอบข้อแรก วิธีการประมาณค่าความสามารถของผู้สอบและความสามารถของผู้สอบ ร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์

เมื่อพิจารณาปฏิสัมพันธ์ระหว่างความยากง่ายของข้อสอบข้อแรก วิธีการประมาณค่าความสามารถของผู้สอบและเกณฑ์สำหรับยุติการทดสอบซึ่งคาดว่าจะร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ แสดงได้ดังภาพประกอบ 55



ความยากง่ายข้อแรก/ความคลาดเคลื่อน

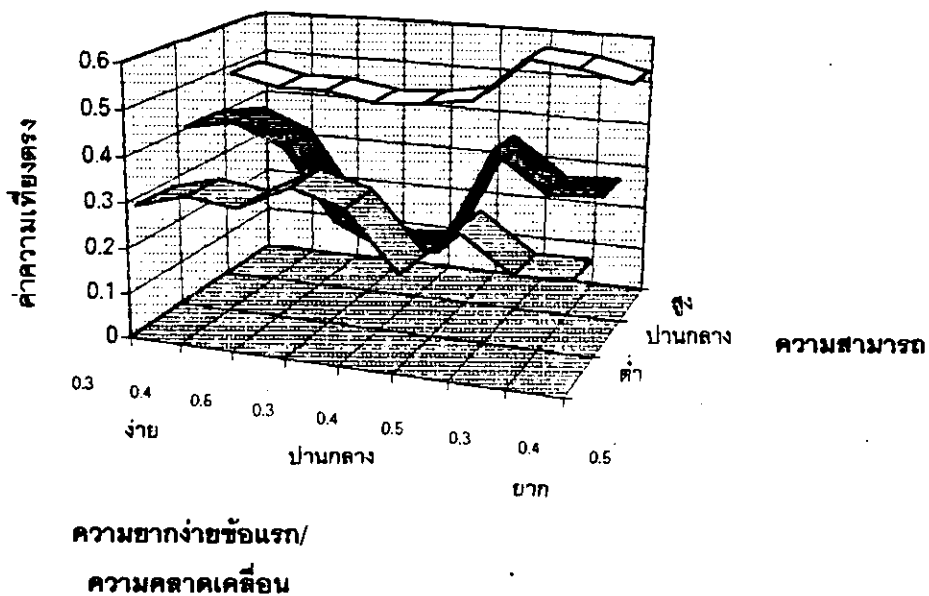
ภาพประกอบ 55 ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จำแนกตามความยากง่ายของข้อสอบข้อแรก วิธีการประมาณค่าความสามารถของผู้สอบและเกณฑ์ยุติการทดสอบ

จากภาพประกอบ 55 พบว่า ค่าความเที่ยงตรงเชิงสภาพมีพิสัยอยู่ในช่วง .2666 ถึง .5839 ความเที่ยงตรงเชิงสภาพมีค่าสูงสุด เมื่อเริ่มต้นทดสอบด้วยข้อสอบที่ง่าย ประมาณค่าความ

สามารถของผู้สอบด้วยวิธีของเบส์ ยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ เท่ากับหรือน้อยกว่า .40 มีค่าต่ำสุด เมื่อเริ่มต้นทดสอบด้วยข้อสอบที่ยากปานกลาง ประมาณค่าความสามารถของผู้สอบด้วยวิธีของเบส์ ยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ เท่ากับหรือน้อยกว่า .50

จากลักษณะการตัดกันของเส้นตรงในภาพ พออนุมานได้ว่า ความยากง่ายของข้อสอบข้อแรก วิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบ ร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์

เมื่อพิจารณาปฏิสัมพันธ์ระหว่างความยากง่ายของข้อสอบข้อแรก เกณฑ์ยุติการทดสอบ และความสามารถของผู้สอบซึ่งคาดว่าจะร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ แสดงได้ดังภาพประกอบ 56



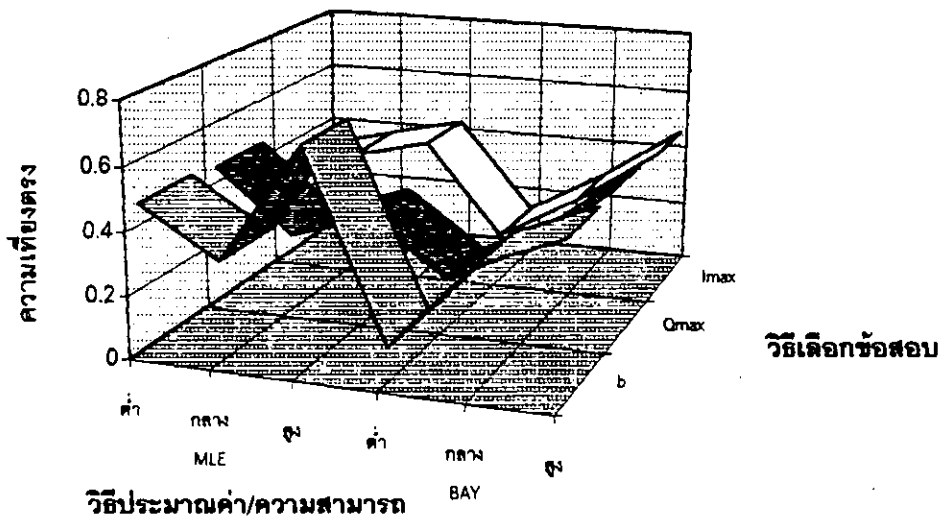
ภาพประกอบ 56 ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จำแนกตามความยากง่ายของข้อสอบข้อแรก เกณฑ์ยุติการทดสอบ และความสามารถของผู้สอบ

จากภาพประกอบ 56 พบว่า ค่าความเที่ยงตรงเชิงสภาพมีพิสัยอยู่ในช่วง .1701 ถึง .5657 ความเที่ยงตรงเชิงสภาพมีค่าสูงสุดเมื่อเริ่มต้นทดสอบด้วยข้อสอบที่ยาก ยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ เท่ากับหรือน้อยกว่า .30 และ

ผู้สอบมีความสามารถสูง มีค่าต่ำสุด เมื่อเริ่มต้นทดสอบด้วยข้อสอบที่ยากปานกลาง ยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ เท่ากับหรือน้อยกว่า .40 และผู้สอบมีความสามารถปานกลาง

จากลักษณะการตัดกันของเส้นตรงในภาพ พออนุมานได้ว่า ความยากง่ายของข้อสอบ ข้อแรก เกณฑ์ยุติการทดสอบและความสามารถของผู้สอบ ร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์

เมื่อพิจารณาปฏิสัมพันธ์ระหว่างเกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบ และความสามารถของผู้สอบซึ่งคาดว่าจะร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ แสดงได้ดังภาพประกอบ 57



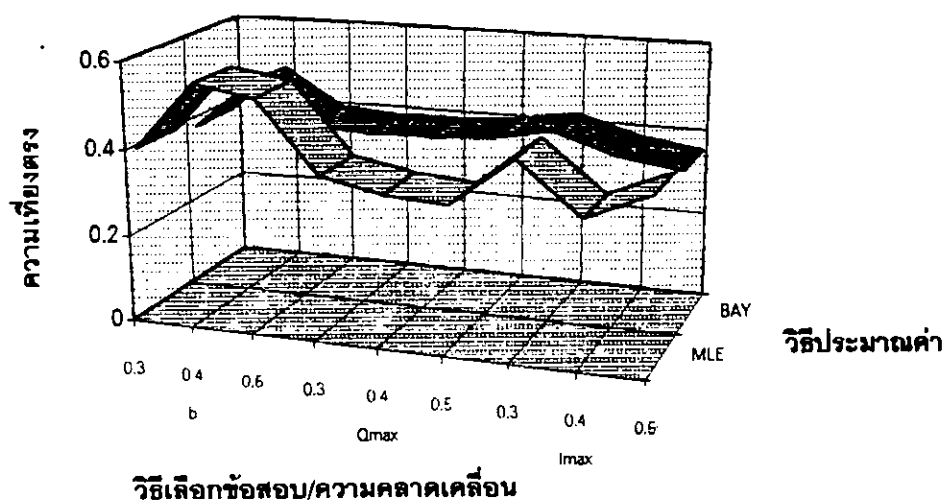
ภาพประกอบ 57 ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จำแนกตามเกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบและความสามารถของผู้สอบ

จากภาพประกอบ 57 พบว่า ค่าความเที่ยงตรงเชิงสภาพมีพิสัยอยู่ในช่วง .1056 ถึง .6889 ความเที่ยงตรงเชิงสภาพมีค่าสูงสุด เมื่อคัดเลือกข้อสอบในคลังข้อสอบด้วยค่าความยากง่ายของข้อสอบ ประมาณค่าความสามารถของผู้สอบด้วยวิธีความเป็นไปได้สูงสุดและผู้สอบมีความ

สามารถสูง มีค่าต่ำสุด เมื่อคัดเลือกข้อสอบในคลังข้อสอบด้วยค่าความยากง่ายของข้อสอบ
ประมาณค่าความสามารถของผู้สอบด้วย วิธีของเบย์และผู้สอบมีความสามารถต่ำ

จากลักษณะการตัดกันของเส้นตรงในภาพ พออนุมานได้ว่า เกณฑ์การคัดเลือกข้อสอบ
จากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบและความสามารถของผู้สอบ ร่วมกัน
ส่งผลต่อความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วย
คอมพิวเตอร์

เมื่อพิจารณาปฏิสัมพันธ์ระหว่างเกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการ
ประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบซึ่งคาดว่าจะร่วมกันส่งผลต่อความ
เที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์
แสดงได้ดังภาพประกอบ 58



ภาพประกอบ 58 ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถ
ของผู้สอบด้วยคอมพิวเตอร์ จำแนกตาม เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการ
ประมาณค่าความสามารถของผู้สอบและเกณฑ์ยุติการทดสอบ

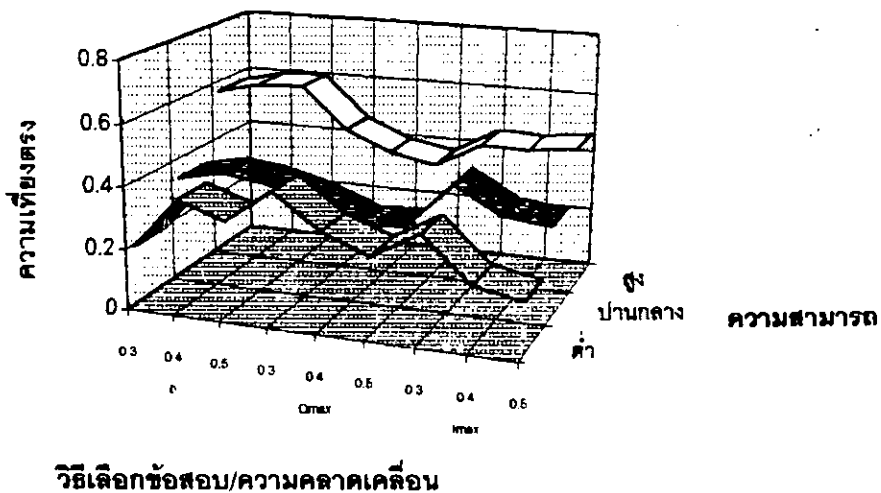
จากภาพประกอบ 58 พบว่า ค่าความเที่ยงตรงเชิงสภาพมีพิสัยอยู่ในช่วง .3238 ถึง
.5493 ความเที่ยงตรงเชิงสภาพมีค่าสูงสุด เมื่อคัดเลือกข้อสอบในคลังข้อสอบด้วยค่าความ

ยากง่ายของข้อสอบ ประเมินค่าความสามารถของผู้สอบด้วยวิธีความเป็นไปได้สูงสุด และยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ เท่ากับหรือน้อยกว่า .40 มีค่าต่ำสุด เมื่อคัดเลือกข้อสอบในคลังข้อสอบด้วยค่าความสามารถที่ทำให้ข้อสอบมีสารสนเทศสูงสุด ประเมินค่าความสามารถของผู้สอบด้วยวิธีความเป็นไปได้สูงสุด ยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ เท่ากับหรือน้อยกว่า .50

จากลักษณะการตัดกันของเส้นตรงในภาพ พออนุมานได้ว่า เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบและเกณฑ์ยุติการทดสอบ ร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์

เมื่อพิจารณาปฏิสัมพันธ์ระหว่างเกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ เกณฑ์ยุติการทดสอบและความสามารถของผู้สอบซึ่งคาดว่าจะร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ แสดงได้ดังภาพประกอบ

59



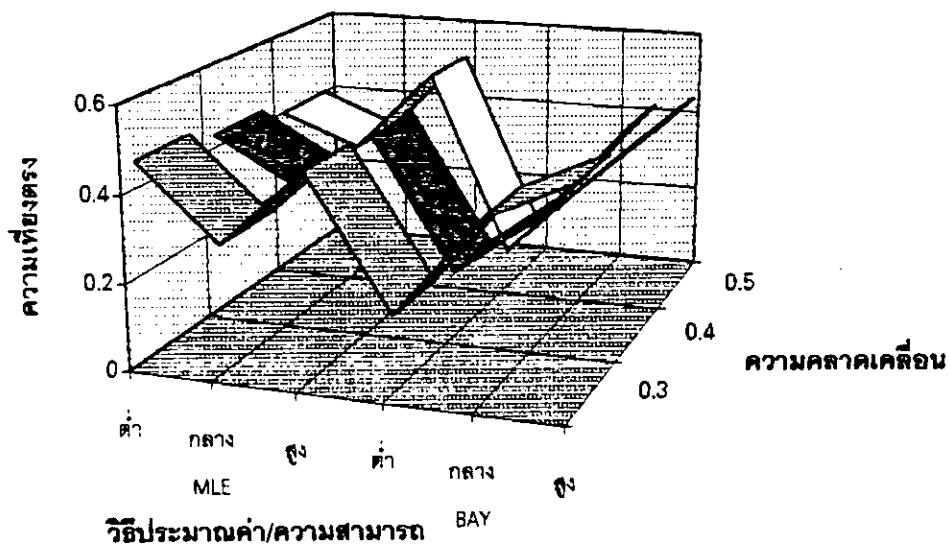
ภาพประกอบ 59 ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จำแนกตามเกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ เกณฑ์ยุติการทดสอบและความสามารถของผู้สอบ

จากภาพประกอบ 59 พบว่า ค่าความเที่ยงตรงเชิงสภาพมีพิสัยอยู่ในช่วง .1669 ถึง

.6025 ความเที่ยงตรงเชิงสภาพมีค่าสูงสุด เมื่อคัดเลือกข้อสอบในคลังข้อสอบด้วยค่าความยากง่ายของข้อสอบ ยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถเท่ากับหรือน้อยกว่า .50 และผู้สอบมีความสามารถสูง มีค่าต่ำสุด เมื่อคัดเลือกข้อสอบในคลังข้อสอบด้วยค่าสารสนเทศสูงสุดของข้อสอบ ยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ เท่ากับหรือน้อยกว่า .50 ผู้สอบมีความสามารถต่ำ

จากลักษณะการตัดกันของเส้นตรงในภาพ เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ เกณฑ์ยุติการทดสอบ และความสามารถของผู้สอบ ร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์

เมื่อพิจารณาปฏิสัมพันธ์ระหว่างวิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบ และความสามารถของผู้สอบที่ร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ แสดงได้ดังภาพประกอบ 60



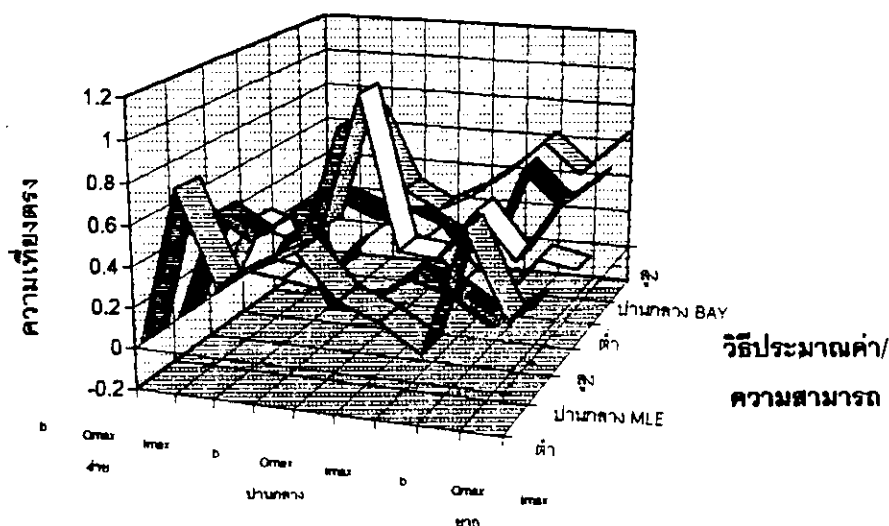
ภาพประกอบ 60 ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จำแนกตามวิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบ และความสามารถของผู้สอบ

จากภาพประกอบ 60 พบว่า ค่าความเที่ยงตรงเชิงสภาพมีพิสัยอยู่ในช่วง .0877 ถึง .5173 ความเที่ยงตรงเชิงสภาพมีค่าสูงสุดเมื่อประมาณค่าความสามารถของผู้สอบด้วยวิธีความเป็นไปได้สูงสุด ยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความ

สามารถ เท่ากับหรือน้อยกว่า .50 และผู้สอบมีความสามารถสูง มีค่าต่ำสุด เมื่อประมาณค่าความสามารถของผู้สอบด้วยวิธีของเบส์ ยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ เท่ากับหรือน้อยกว่า .50 และ ผู้สอบมีความสามารถต่ำ

จากลักษณะการตัดกันของเส้นตรงในภาพ พออนุมานได้ว่า วิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบและความสามารถของผู้สอบ ร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์

เมื่อพิจารณาปฏิสัมพันธ์ระหว่างความยากง่ายของข้อสอบข้อแรกที่ใช้เมื่อเริ่มต้นในการทดสอบ เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบและความสามารถของผู้สอบซึ่งคาดว่าจะร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ แสดงได้ดังภาพประกอบ 61



ความยากง่ายข้อแรก/วิธีคัดเลือก

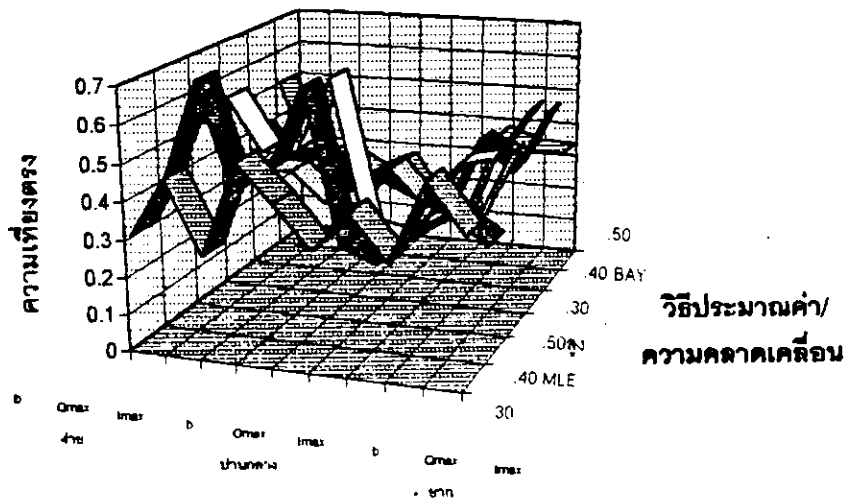
ภาพประกอบ 61 ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จำแนกตามความยากง่ายของข้อสอบข้อแรก เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบและความสามารถของผู้สอบ

จากภาพประกอบ 61 พบว่า ค่าความเที่ยงตรงเชิงสภาพมีพิสัยอยู่ในช่วง -.0185 ถึง 1.0970 ความเที่ยงตรงเชิงสภาพมีค่าสูงสุด เมื่อเริ่มต้นทดสอบด้วยข้อสอบที่ยากปานกลาง คัดเลือกข้อสอบในคลังข้อสอบด้วยค่าความยากง่ายของข้อสอบ ประมาณค่าความสามารถของ

ผู้สอบด้วยวิธีความเป็นไปได้สูงสุดและผู้สอบมีความสามารถสูง มีค่าต่ำสุด เมื่อเริ่มต้นทดสอบด้วยข้อสอบที่ยาก คัดเลือกข้อสอบในคลังข้อสอบด้วยค่าความยากง่ายของข้อสอบและประมาณค่าความสามารถของผู้สอบด้วยวิธีของเบส์ ผู้สอบมีความสามารถต่ำ

จากลักษณะการตัดกันของเส้นตรงในภาพ พออนุมานได้ว่า ความยากง่ายของข้อสอบข้อแรก เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบ และความสามารถของผู้สอบ ร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์

เมื่อพิจารณาปฏิสัมพันธ์ระหว่างความยากง่ายของข้อสอบข้อแรก เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบ และเกณฑ์ยุติการทดสอบ ซึ่งคาดว่าจะร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ แสดงได้ดังภาพประกอบ 62



ความยากง่ายข้อแรก/ วิธีเลือกข้อสอบ

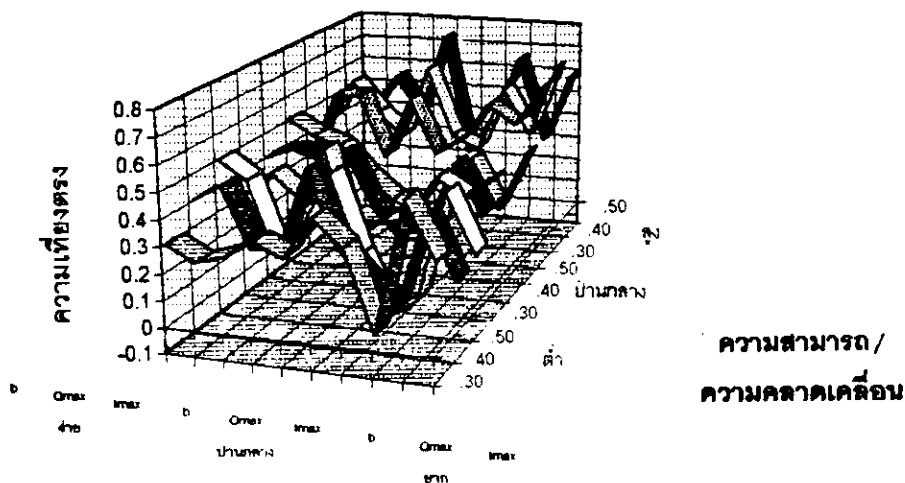
ภาพประกอบ 62 ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จำแนกตามความยากง่ายของข้อสอบข้อแรก เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบ

จากภาพประกอบ 62 พบว่า ค่าความเที่ยงตรงเชิงสภาพมีพิสัยอยู่ในช่วง .2088 ถึง .6730 ความเที่ยงตรงเชิงสภาพมีค่าสูงสุด เมื่อเริ่มต้นทดสอบด้วยข้อสอบที่ยากปานกลาง

คัดเลือกข้อสอบในคลังข้อสอบด้วยค่าความยากง่ายของข้อสอบ ประมาณค่าความสามารถของผู้สอบด้วยวิธีความเป็นไปได้สูงสุด ยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ เท่ากับหรือน้อยกว่า .40 มีค่าต่ำสุด เมื่อเริ่มต้นทดสอบด้วยข้อสอบที่ยากปานกลาง คัดเลือกข้อสอบในคลังข้อสอบด้วยค่าสารสนเทศสูงสุดของข้อสอบ ประมาณค่าความสามารถของผู้สอบด้วยวิธีของเบส์ ยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ เท่ากับหรือน้อยกว่า .40

จากลักษณะการตัดกันของเส้นตรงในภาพ พออนุมานได้ว่า ความยากง่ายของข้อสอบ ข้อแรก เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบ ร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความความสามารถของผู้สอบด้วยคอมพิวเตอร์

เมื่อพิจารณาปฏิสัมพันธ์ระหว่างความยากง่ายของข้อสอบข้อแรก เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ เกณฑ์ยุติการทดสอบและความสามารถของผู้สอบซึ่งคาดว่าจะร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะกับความความสามารถของผู้สอบด้วยคอมพิวเตอร์ แสดงได้ดังภาพประกอบ 63



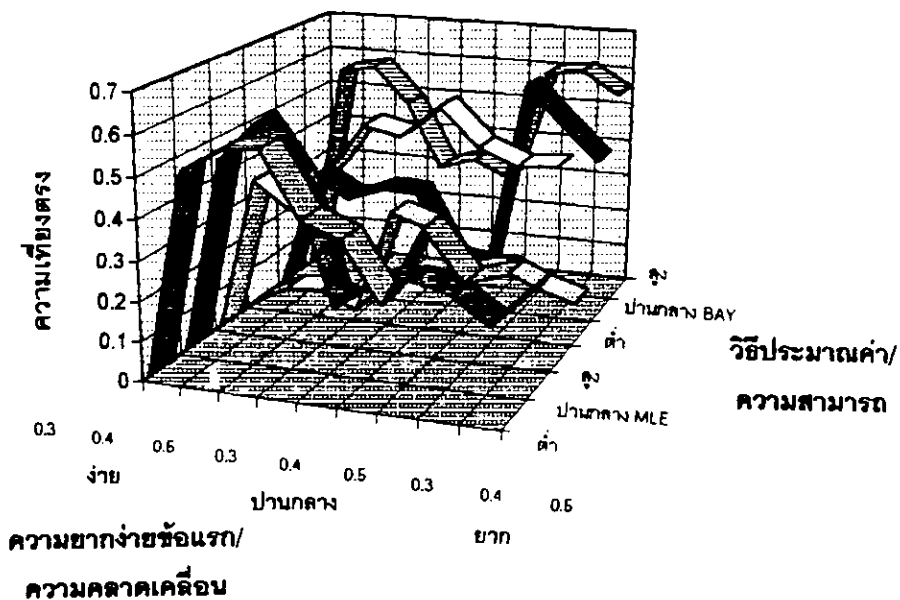
ความยากง่ายข้อแรก/ วิธีเลือกข้อสอบ

ภาพประกอบ 63 ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความความสามารถของผู้สอบด้วยคอมพิวเตอร์ จำแนกตามความยากง่ายของข้อสอบข้อแรก เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ เกณฑ์ยุติการทดสอบและความสามารถของผู้สอบ

จากภาพประกอบ 63 พบว่า ค่าความเที่ยงตรงเชิงสภาพมีพิสัยอยู่ในช่วง -0.0542 ถึง $.6860$ ความเที่ยงตรงเชิงสภาพมีค่าสูงสุด เมื่อเริ่มต้นทดสอบด้วยข้อสอบที่ยาก คัดเลือกข้อสอบในคลังข้อสอบโดยใช้ค่าสารสนเทศสูงสุดของข้อสอบ ยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ เท่ากับหรือน้อยกว่า $.40$ และผู้สอบมีความสามารถสูง มีค่าต่ำสุด เมื่อเริ่มต้นทดสอบด้วยข้อสอบที่ยากปานกลาง คัดเลือกข้อสอบในคลังข้อสอบด้วยค่าสารสนเทศสูงสุดของข้อสอบ ยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ เท่ากับหรือน้อยกว่า $.50$ และผู้สอบมีความสามารถต่ำ

จากลักษณะการตัดกันของเส้นตรงในภาพ พออนุมานได้ว่า ความยากง่ายของข้อสอบ ข้อแรก เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ เกณฑ์ยุติการทดสอบ ความสามารถของผู้สอบ ร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์

เมื่อพิจารณาปฏิสัมพันธ์ระหว่างความยากง่ายของข้อสอบข้อแรก เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบ และความสามารถของผู้สอบซึ่งคาดว่าจะร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ แสดงได้ดังภาพประกอบ 64

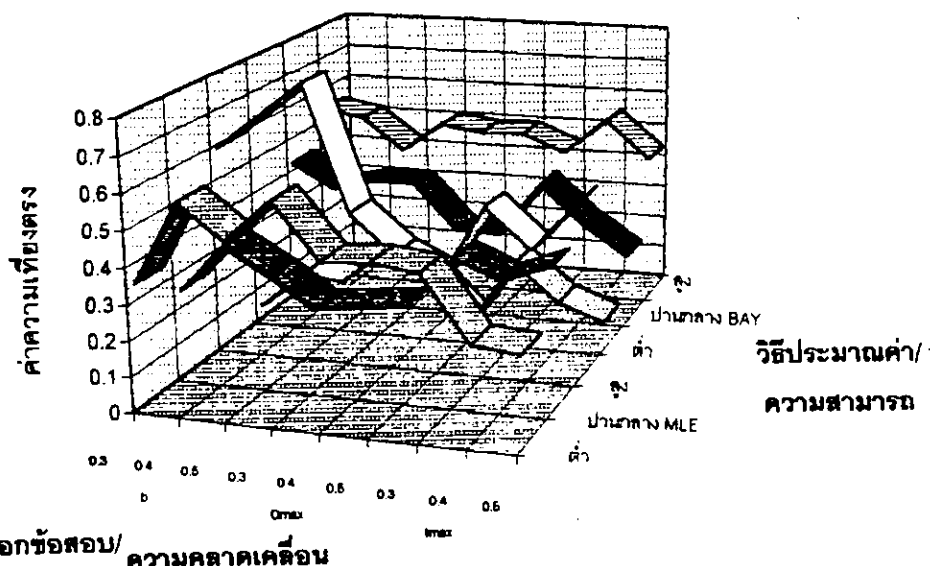


ภาพประกอบ 64 ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จำแนกตามความยากง่ายของข้อสอบข้อแรก วิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบและความสามารถของผู้สอบ

จากภาพประกอบ 64 พบว่า ค่าความเที่ยงตรงเชิงสภาพมีพิสัยอยู่ในช่วง .0147 ถึง .6155 ความเที่ยงตรงเชิงสภาพมีค่าสูงสุด เมื่อเริ่มต้นทดสอบด้วยข้อสอบที่ยาก ประมาณค่าความสามารถของผู้สอบด้วยวิธีของเบส์ ยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ เท่ากับหรือน้อยกว่า .40 และผู้สอบมีความสามารถสูง มีค่าต่ำสุด เมื่อเริ่มต้นทดสอบด้วยข้อสอบที่ง่าย ประมาณค่าความสามารถของผู้สอบด้วยวิธีของเบส์ ยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ เท่ากับหรือน้อยกว่า .50 และผู้สอบมีความสามารถต่ำ

จากลักษณะการตัดกันของเส้นตรงในภาพ พออนุมานได้ว่า ความยากง่ายของข้อสอบ ข้อแรก วิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบและความสามารถของผู้สอบ ร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์

เมื่อพิจารณาปฏิสัมพันธ์ระหว่างเกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบ และความสามารถของผู้สอบ ซึ่งคาดว่าจะร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ แสดงได้ดังภาพประกอบ 65

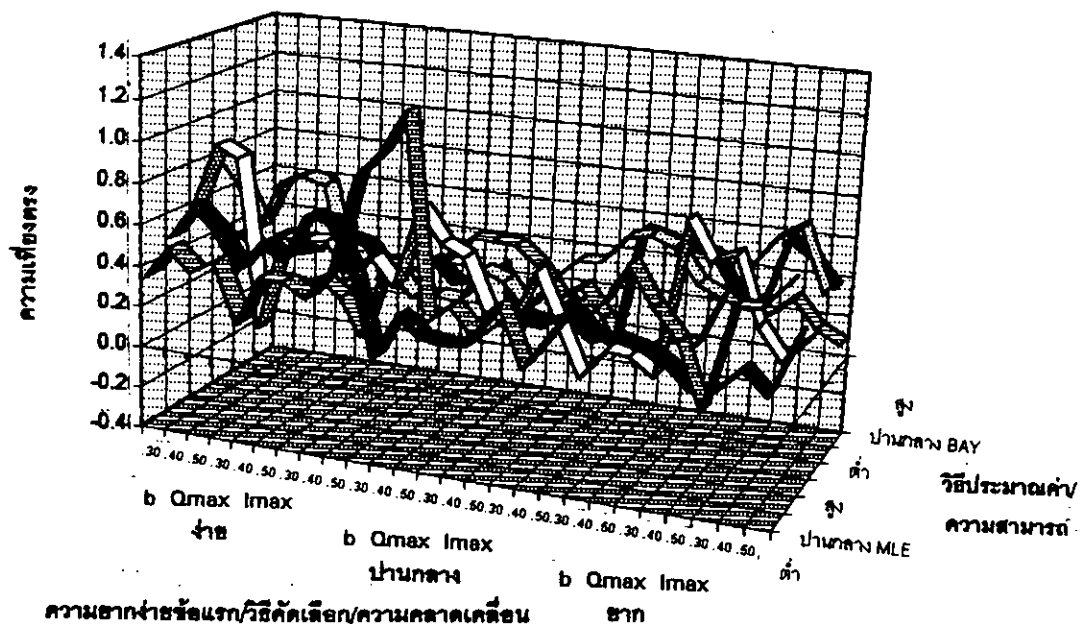


ภาพประกอบ 65 ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จำแนกตามเกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบและความสามารถของผู้สอบ

จากภาพประกอบ 65 พบว่า ค่าความเที่ยงตรงเชิงสภาพมีพิสัยอยู่ในช่วง .0370 ถึง .7879 ความเที่ยงตรงเชิงสภาพมีค่าสูงสุด เมื่อคัดเลือกข้อสอบในคลังข้อสอบด้วยค่าความยากง่ายของข้อสอบ ประมาณค่าความสามารถของผู้สอบด้วยวิธีความเป็นไปได้สูงสุด ยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ เท่ากับหรือน้อยกว่า .50 และผู้สอบมีความสามารถสูง มีค่าต่ำสุด เมื่อคัดเลือกข้อสอบในคลังข้อสอบด้วยค่าความยากง่ายของข้อสอบ ประมาณค่าความสามารถของผู้สอบด้วยวิธีของเบส์ ยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ เท่ากับหรือน้อยกว่า .30 ผู้สอบมีความสามารถต่ำ

จากลักษณะการตัดกันของเส้นตรงในภาพ พออนุมานได้ว่า เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบและความสามารถของผู้สอบ ร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์

เมื่อพิจารณาปฏิสัมพันธ์ระหว่างความยากง่ายของข้อสอบข้อแรก เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบ และความสามารถของผู้สอบซึ่งคาดว่าจะร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ แสดงได้ดังภาพประกอบ 66



ภาพประกอบ 66 ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จำแนกตามความยากง่ายของข้อสอบข้อแรก เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบและความสามารถของผู้สอบ

จากภาพประกอบ 66 พบว่า ค่าความเที่ยงตรงเชิงสภาพมีพิสัยอยู่ในช่วง -.0041 ถึง 1.2733 ความเที่ยงตรงเชิงสภาพมีค่าสูงสุด เมื่อเริ่มต้นทดสอบด้วยข้อสอบที่ยากปานกลาง คัดเลือกข้อสอบในคลังข้อสอบด้วยค่าความยากง่ายของข้อสอบ ประมาณค่าความสามารถของผู้สอบด้วยวิธีความเป็นไปได้สูงสุด ยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ เท่ากับหรือน้อยกว่า .50 และผู้สอบมีความสามารถต่ำ มีค่าต่ำสุด เมื่อเริ่มต้นทดสอบด้วยข้อสอบที่ยาก คัดเลือกข้อสอบในคลังข้อสอบด้วยค่าความยากง่ายของข้อสอบ ประมาณค่าความสามารถของผู้สอบด้วยวิธีความเป็นไปได้สูงสุด ยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถเท่ากับหรือน้อยกว่า .50 และผู้สอบมีความสามารถต่ำ

จากลักษณะการตัดกันของเส้นตรงในภาพ พออนุมานได้ว่า ความยากง่ายของข้อสอบ ข้อแรก เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบ ความสามารถของผู้สอบ ร่วมกันส่งผลต่อความเที่ยงตรงเชิงสภาพของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์

เมื่อศึกษาความแตกต่างของจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ โดยใช้เทคนิคการวิเคราะห์ความแปรปรวน ซึ่งได้ผลการวิเคราะห์ ดังตาราง 22

ตาราง 22 ผลการวิเคราะห์ความแปรปรวนของจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จำแนกตามแหล่งความแปรปรวน

แหล่งแปรปรวน	SS	df	MS	F
ผลหลัก				
ความยากง่ายเริ่มต้น (A)	1229.704	2	614.852	16.914**
เกณฑ์คัดเลือกข้อสอบ (B)	31372.397	2	15686.198	431.505**
วิธีประมาณค่า (C)	2066.461	1	2066.461	56.845**
ความสามารถเริ่มต้น (D)	40937.375	2	20468.679	563.064**
เกณฑ์ยุติการทดสอบ (E)	199798.461	2	99899.231	2748.083**
ปฏิสัมพันธ์ 2 ทาง				
AXB	723.146	4	180.787	4.973**

แหล่งแปรปรวน	SS	df	MS	F
AXC	917.132	2	458.566	12.614**
AXD	90.895	4	22.724	.625
AXE	72.988	4	18.247	.502
BXC	120.395	2	60.197	1.656
BXD	1520.427	4	380.107	10.456**
BXE	2988.863	4	747.216	20.555**
CXD	1806.856	2	903.428	24.852**
CXE	656.156	2	328.078	9.025**
DXE	17794.620	4	4448.655	122.376**
<u>ปฏิสัมพันธ์ 3 ทาง</u>				
AXBXC	96.451	4	24.113	.663
AXBXD	714.464	8	89.308	2.457*
AXBXE	451.828	8	56.478	1.554
AXCXD	183.280	4	45.820	1.260
AXCXE	118.890	4	29.723	.818
AXDXE	111.271	8	13.909	.383
BXCXD	417.747	4	104.437	2.873*
BXCXE	90.638	4	22.659	.623
BXDXE	652.330	8	81.541	2.243*
CXDXE	399.772	4	99.943	2.749*
<u>ปฏิสัมพันธ์ 4 ทาง</u>				
AXBXCXD	934.720	8	116.840	3.214**
AXBXCXE	182.905	8	22.863	.629
AXBXDXE	594.607	16	37.163	1.022
AXCXDXE	183.230	8	22.904	.630
BXCXDXE	142.905	8	17.863	.491
<u>ปฏิสัมพันธ์ 5 ทาง</u>				
A X B X C X D X E	552.950	16	34.559	.951

แหล่งแปรปรวน	SS	df	MS	F
ความคลาดเคลื่อน	170601.498	4693	36.352	

** มีนัยสำคัญทางสถิติที่ระดับ .01

* มีนัยสำคัญทางสถิติที่ระดับ .05

จากตาราง 22 เมื่อพิจารณาผลของตัวแปรอิสระทั้ง 5 ตัว ต่อจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ ถ้าไม่พิจารณาปฏิสัมพันธ์ระหว่างตัวแปรอิสระแล้ว พบว่า ตัวแปรอิสระทั้ง 5 ตัวต่างก็ส่งผลต่อจำนวนข้อสอบที่ใช้ในการทดสอบอย่างมีนัยสำคัญที่ระดับ .01

หากพิจารณาการร่วมกันของตัวแปรต้น 2 ตัวต่อจำนวนข้อสอบที่ใช้ พบว่า ความยากง่ายเริ่มต้นของข้อสอบกับเกณฑ์การคัดเลือกข้อสอบ ความยากง่ายเริ่มต้นกับวิธีการประมาณค่าความสามารถ เกณฑ์การคัดเลือกข้อสอบกับความสามารถเริ่มต้นของผู้สอบ เกณฑ์การคัดเลือกข้อสอบกับเกณฑ์ยุติการทดสอบ วิธีการประมาณค่าความสามารถกับความสามารถเริ่มต้น วิธีการประมาณค่าความสามารถกับเกณฑ์ยุติการทดสอบ ความสามารถเริ่มต้นกับเกณฑ์ยุติการทดสอบ ซึ่งรวมทั้งหมด 7 คู่ จาก 10 คู่ร่วมกันส่งผลต่อจำนวนข้อสอบที่ใช้ในการทดสอบอย่างมีนัยสำคัญ

เมื่อพิจารณาการร่วมกันของตัวแปรต้น 3 ตัวต่อจำนวนข้อสอบที่ใช้ พบว่า มี 4 กลุ่มที่ส่งผลต่อจำนวนข้อสอบที่ใช้ในการทดสอบอย่างมีนัยสำคัญ กลุ่มที่ 1 คือ ความยากง่ายของข้อสอบเริ่มต้น เกณฑ์การคัดเลือกข้อสอบ ความสามารถเริ่มต้นของผู้สอบ กลุ่มที่ 2 คือ เกณฑ์การคัดเลือกข้อสอบ วิธีการประมาณค่าความสามารถ ความสามารถเริ่มต้นของผู้สอบ กลุ่มที่ 3 คือ เกณฑ์การคัดเลือกข้อสอบ ความสามารถเริ่มต้นของผู้สอบ เกณฑ์ยุติการทดสอบ กลุ่มที่ 4 คือ วิธีการประมาณค่าความสามารถ ความสามารถเริ่มต้นและเกณฑ์ยุติการทดสอบ

หากพิจารณากลุ่มของตัวแปรอิสระ 4 ตัว พบว่า ตัวแปรอิสระทั้งหมด ยกเว้น เกณฑ์ยุติการทดสอบ ร่วมกันส่งผลต่อจำนวนข้อสอบที่ใช้ในการทดสอบ

เมื่อพิจารณาปฏิสัมพันธ์ร่วม 5 ทาง พบว่า ตัวแปรอิสระทั้ง 5 ตัวดังกล่าวข้างต้นร่วมกันส่งผลต่อจำนวนข้อสอบที่ใช้อย่างไม่มีนัยสำคัญ

เมื่อเปรียบเทียบค่าเฉลี่ยของจำนวนข้อสอบที่ใช้ในการทดสอบเป็นรายคู่ ตามระดับของตัวแปรอิสระ 5 ตัวแปร คือ ความยากง่ายของข้อสอบข้อแรก เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบและความสามารถของผู้สอบ ซึ่งได้ผลการวิเคราะห์ ดังตาราง 23 - 26

ตาราง 23 ความแตกต่างของค่าเฉลี่ยของจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะกับความ
ความสามารถของผู้สอบด้วยคอมพิวเตอร์ เรื่อง เศษส่วน จำแนกตามความยากง่ายของ
ข้อสอบข้อแรกที่ใช้เมื่อเริ่มต้นสอบ

ค่าเฉลี่ย	$\bar{X}_{ง่าย}$	$\bar{X}_{ปานกลาง}$	$\bar{X}_{ยาก}$
$\bar{X}_{ง่าย} = 14.58$	-	.77	-.41
$\bar{X}_{ปานกลาง} = 13.81$		-	1.18*
$\bar{X}_{ยาก} = 14.99$			-

* มีนัยสำคัญทางสถิติที่ระดับ .05

จากตาราง 23 พบว่า เมื่อเริ่มต้นทดสอบด้วยข้อสอบที่ยากง่ายปานกลางจะใช้ข้อสอบต่างจากการเริ่มต้นทดสอบด้วยข้อสอบที่ยากอย่างมีนัยสำคัญ ที่ระดับ .05 สำหรับการเริ่มต้นทดสอบด้วยข้อสอบที่ง่ายกับยาก และง่ายกับปานกลางนั้น จำนวนข้อสอบที่ใช้ไม่ต่างกัน

เมื่อเปรียบเทียบค่าเฉลี่ยของจำนวนข้อสอบที่ใช้ในการทดสอบตามเกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ ได้ผลดังตาราง 24

ตาราง 24 ความแตกต่างของค่าเฉลี่ยของจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะกับความ
ความสามารถของผู้สอบด้วยคอมพิวเตอร์ เรื่อง เศษส่วน จำแนกตามเกณฑ์การคัดเลือก
ข้อสอบจากคลังข้อสอบ

ค่าเฉลี่ย	$\bar{X}$ ความยากง่าย	$\bar{X}$ θ_{max}	$\bar{X}$ I_{max}
$\bar{X}$ ความยากง่าย = 17.59	-	3.14	6.25*
$\bar{X}$ ใช้ค่า θ_{max} = 14.45		-	3.11
$\bar{X}$ ใช้ค่า I_{max} = 11.34			-

* มีนัยสำคัญทางสถิติที่ระดับ .05

จากตาราง 24 พบว่า เกณฑ์การคัดเลือกข้อสอบโดยใช้ค่าความยากง่ายของข้อสอบกับ
การใช้ค่าสารสนเทศของข้อสอบที่สูงสุด จะทำให้ใช้ข้อสอบมากน้อยต่างกัน อย่างมีนัยสำคัญที่
ระดับ .05

เมื่อเปรียบเทียบค่าเฉลี่ยของจำนวนข้อสอบที่ใช้ในการทดสอบตามเกณฑ์ยุติการทดสอบ
ได้ผลดังตาราง 25

ตาราง 25 ความแตกต่างของค่าเฉลี่ยของจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะกับความ
ความสามารถของผู้สอบด้วยคอมพิวเตอร์ เรื่อง เศษส่วน จำแนกตามเกณฑ์ยุติการทดสอบ

ค่าเฉลี่ย	$\bar{X}$ $SEE_{\leq .30}$	$\bar{X}$ $SEE_{\leq .40}$	$\bar{X}$ $SEE_{\leq .50}$
$\bar{X}_{SEE_{\leq .30}}$ = 23.14	-	10.7*	15.34*
$\bar{X}_{SEE_{\leq .40}}$ = 12.44		-	4.64*
$\bar{X}_{SEE_{\leq .50}}$ = 7.80			-

* มีนัยสำคัญทางสถิติที่ระดับ .05

จากตาราง 25 พบว่า ทุกคู่ของเกณฑ์ยุติการทดสอบจะใช้ข้อสอบจำนวนแตกต่างกัน
อย่างมีนัยสำคัญที่ระดับ .05 กล่าวคือ เกณฑ์ยุติการทดสอบ .30 ใช้ข้อสอบมากน้อยต่างจาก

เกณฑ์ยุติการทดสอบ .40 และ .50 และ เกณฑ์ยุติการทดสอบ .40 ใช้ข้อสอบมากน้อยต่างจาก เกณฑ์ยุติการทดสอบ .50 อย่างมีนัยสำคัญที่ระดับ .05

เมื่อเปรียบเทียบค่าเฉลี่ยของจำนวนข้อสอบที่ใช้ในการทดสอบตามความสามารถทางคณิตศาสตร์ของผู้สอบ 3 ระดับ คือ ความสามารถสูง ปานกลาง และต่ำ ได้ผลดังตาราง 26

ตาราง 26 ความแตกต่างของค่าเฉลี่ยของจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะกับความ
ความสามารถของผู้สอบด้วยคอมพิวเตอร์ เรื่อง เศษส่วน จำแนกตามความสามารถทาง
คณิตศาสตร์ของผู้สอบ

ค่าเฉลี่ย	$\bar{X}$ สูง	$\bar{X}$ ปานกลาง	$\bar{X}$ ต่ำ
$\bar{X}_{\text{สูง}} = 11.33$	-	2.34	7.03*
$\bar{X}_{\text{ปานกลาง}} = 13.67$		-	4.69*
$\bar{X}_{\text{ต่ำ}} = 18.36$			-

* มีนัยสำคัญทางสถิติที่ระดับ .05

จากตาราง 26 พบว่า ผู้สอบที่มีความสามารถทางคณิตศาสตร์สูงใช้ข้อสอบมากน้อยต่างจากผู้สอบที่มีความสามารถทางคณิตศาสตร์ปานกลางและต่ำ อย่างมีนัยสำคัญที่ระดับ .05 ส่วนผู้สอบที่มีความสามารถปานกลาง และต่ำ จะใช้ข้อสอบจำนวนไม่ต่างกัน

บทที่ 5

สรุป อภิปรายผลและข้อเสนอแนะ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบความเที่ยงตรงเชิงสภาพ และ จำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ เมื่อมีตัวแปรอิสระ 5 ตัว ซึ่งประกอบด้วย ความยากง่ายของข้อสอบข้อแรก เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบและความสามารถของผู้สอบ โดยแบ่งการดำเนินการวิจัยออกเป็น 2 ระยะ

ระยะแรก เป็นการสร้างคลังข้อสอบ เรื่อง เศษส่วน เครื่องมือที่ใช้ คือ แบบทดสอบวัดความเข้าใจเรื่องเศษส่วน จำนวน 7 ฉบับ ฉบับละ 50 ข้อ รวม 290 ข้อ (แต่ละฉบับมีข้อสอบร่วมกันอยู่ฉบับละ 10 ข้อ) กลุ่มตัวอย่าง คือ นักเรียนชั้นประถมศึกษาปีที่ 6 ในโรงเรียนสังกัดสำนักงานการประถมศึกษากรุงเทพมหานครและสังกัดกรุงเทพมหานคร ภาคเรียนที่ 2 ปีการศึกษา 2537 จำนวน 8,720 คน

ระยะที่สอง เป็นการศึกษาผลของตัวแปรอิสระ 5 ตัวต่อตัวแปรตามที่กล่าวถึงข้างต้น เครื่องมือที่ใช้ คือ แบบทดสอบวัดความสามารถทางคณิตศาสตร์และโปรแกรมคอมพิวเตอร์ที่ใช้สำหรับการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ กลุ่มตัวอย่าง คือ นักเรียนชั้นมัธยมศึกษาปีที่ 1 สังกัดกรมสามัญศึกษาในส่วนกลาง ในตอนต้นของภาคเรียนที่ 1 ปีการศึกษา 2539 จำนวน 1,620 คน

สถิติที่ใช้ในการวิจัย คือ ค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐาน ค่าสหสัมพันธ์ ค่าไคสแควร์ และการวิเคราะห์ความแปรปรวน โดยทำการทดสอบภายหลังด้วยวิธีเอสของเซฟเฟ

สรุปผลการวิจัย

จากผลการวิเคราะห์ข้อมูลสามารถสรุปผลการวิจัยตามวัตถุประสงค์ของการวิจัยได้ดังตาราง 27

172

ตาราง 27 ความเที่ยงตรงเชิงสภาพและจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จำแนกตามความยากง่ายของข้อสอบข้อแรก เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบและความสามารถของผู้สอบ

ตัวแปรอิสระ	ความเที่ยงตรงเชิงสภาพ	จำนวนข้อสอบ
1. ความยากง่ายข้อแรก	ไม่ต่าง	ต่าง
- ง่าย	.7419	14.58
- ปานกลาง	.7103	13.81
- ยาก	.7497	14.99
2. เกณฑ์การคัดเลือกข้อสอบ	ไม่ต่าง	ต่าง
- $\theta = b$	.7148	17.59
- $\theta = \theta_{\max}$	.7404	14.45
- $l_{\max}$	.7444	11.34
3. ประมาณค่าความสามารถ	ต่าง	ต่าง
- วิธีความเป็นไปได้สูงสุด	.7026	13.82
- วิธีของเบส์	.7608	15.10
4. เกณฑ์ยุติการทดสอบ	ต่าง	ต่าง
- $SEE \leq .03$	.7710	23.14
- $SEE \leq .04$	.7323	12.14
- $SEE \leq .05$	.6869	7.80
5. ความสามารถของผู้สอบ	ต่าง	ต่าง
- สูง	.4449	11.33
- ปานกลาง	.3173	13.67
- ต่ำ	.2994	18.36

จากตาราง 27 สามารถสรุปผลการวิจัยได้ ดังนี้

- เมื่อใช้ข้อสอบที่มีความยากง่ายต่างกันเป็นข้อสอบข้อแรกในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ไม่ทำให้ค่าความเที่ยงตรงเชิงสภาพของการทดสอบต่างกัน

2. เมื่อใช้เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบต่างกันในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ไม่ทำให้ค่าความเที่ยงตรงเชิงสภาพของการทดสอบต่างกัน
3. เมื่อใช้วิธีการประมาณค่าความสามารถของผู้สอบต่างกันในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ทำให้ค่าความเที่ยงตรงเชิงสภาพของการทดสอบต่างกัน วิธีการประมาณค่าความสามารถวิธีของเบส์จะให้ค่าความเที่ยงตรงเชิงสภาพของการทดสอบสูงกว่าวิธีความเป็นไปได้สูงสุด
4. เมื่อใช้เกณฑ์ยุติการทดสอบต่างกันในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ทำให้ค่าความเที่ยงตรงเชิงสภาพของการทดสอบต่างกัน ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถเท่ากับหรือน้อยกว่า .30 ให้ค่าความเที่ยงตรงเชิงสภาพสูงกว่าเกณฑ์อื่น
5. เมื่อความสามารถทางคณิตศาสตร์ของผู้สอบต่างกันในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ทำให้ค่าความเที่ยงตรงเชิงสภาพของการทดสอบต่างกัน ความเที่ยงตรงเชิงสภาพจะสูงที่สุด เมื่อทดสอบกับผู้สอบที่มีความสามารถสูง
6. เมื่อใช้ข้อสอบที่มีความยากง่ายต่างกันเป็นข้อสอบข้อแรกในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ทำให้จำนวนข้อสอบที่ใช้ในการทดสอบต่างกัน ถ้าใช้ข้อสอบข้อแรกที่มีความยากง่ายปานกลางจะใช้ข้อสอบในการทดสอบน้อยกว่าการเริ่มต้นทดสอบด้วยข้อสอบที่ยากหรือง่าย
7. เมื่อใช้เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบต่างกันในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ทำให้จำนวนข้อสอบที่ใช้ในการทดสอบต่างกัน การคัดเลือกข้อสอบจากคลังข้อสอบด้วยการใช้ค่าสารสนเทศสูงสุดของข้อสอบจะใช้ข้อสอบน้อยกว่าการคัดเลือกข้อสอบวิธีอื่นๆ
8. เมื่อใช้วิธีการประมาณค่าความสามารถของผู้สอบต่างกันในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ทำให้จำนวนข้อสอบที่ใช้ในการทดสอบต่างกัน วิธีของเบส์ใช้ข้อสอบมากกว่าวิธีความเป็นไปได้สูงสุด
9. เมื่อใช้เกณฑ์ยุติการทดสอบต่างกันในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ทำให้จำนวนข้อสอบที่ใช้ในการทดสอบต่างกัน ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถเท่ากับหรือน้อยกว่า .30 ใช้ข้อสอบมากกว่าเกณฑ์อื่นๆ
10. เมื่อความสามารถทางคณิตศาสตร์ของผู้สอบต่างกันในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ทำให้ จำนวนข้อสอบที่ใช้ในการทดสอบต่างกัน ผู้สอบที่มีความสามารถสูงใช้ข้อสอบน้อยกว่าผู้สอบที่มีความสามารถต่ำ

ถ้าพิจารณาการร่วมกันส่งผลของตัวแปรอิสระหรือปฏิสัมพันธ์ของตัวแปรอิสระต่อค่าความเที่ยงตรงเชิงสภาพของการทดสอบ พบว่า เกณฑ์การเลือกข้อสอบจากคลังข้อสอบกับวิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบกับวิธีการประมาณค่าความสามารถ และเกณฑ์ยุติการทดสอบกับความสามารถของผู้สอบไม่ร่วมกันส่งผลต่อค่าความเที่ยงตรงเชิงสภาพ ส่วนกลุ่มของตัวแปรอิสระที่เหลือ ไม่ว่าจะจัดกลุ่มเป็นปฏิสัมพันธ์ 2 ทาง 3 ทาง 4 ทาง หรือ 5 ทางก็ตาม ต่างก็ร่วมกันส่งผลต่อค่าความเที่ยงตรงเชิงสภาพของการทดสอบทั้งสิ้น

หากพิจารณาการร่วมกันของตัวแปรอิสระ 2 ตัวต่อจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ พบว่า ความยากง่ายของข้อสอบข้อแรกกับเกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ ความยากง่ายของข้อสอบข้อแรกกับวิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบกับความสามารถของผู้สอบ เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบกับเกณฑ์ยุติการทดสอบ วิธีการประมาณค่าความสามารถของผู้สอบกับความสามารถของผู้สอบ วิธีการประมาณค่าความสามารถของผู้สอบกับเกณฑ์ยุติการทดสอบ ความสามารถของผู้สอบกับเกณฑ์ยุติการทดสอบ ซึ่งรวมทั้งหมด 7 คู่ จาก 10 คู่ร่วมกันส่งผลต่อจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์อย่างมีนัยสำคัญ

เมื่อพิจารณาการร่วมกันส่งผลของตัวแปรอิสระ 3 ตัวต่อจำนวนข้อสอบที่ใช้ พบว่า มี 4 กลุ่มที่ส่งผลต่อจำนวนข้อสอบที่ใช้ในการทดสอบอย่างมีนัยสำคัญ กลุ่มที่ 1 คือ ความยากง่ายของข้อสอบข้อแรก เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ ความสามารถของผู้สอบ กลุ่มที่ 2 คือ เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีประมาณค่าความสามารถของผู้สอบ ความสามารถของผู้สอบ กลุ่มที่ 3 คือ เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ ความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบ กลุ่มที่ 4 คือ วิธีการประมาณค่าความสามารถของผู้สอบ ความสามารถของผู้สอบและเกณฑ์ยุติการทดสอบ

หากพิจารณากลุ่มของตัวแปรอิสระ 4 ตัว พบว่า ตัวแปรอิสระทั้งหมด ยกเว้น เกณฑ์ยุติการทดสอบ ร่วมกันส่งผลต่อจำนวนข้อสอบที่ใช้ในการทดสอบ

เมื่อพิจารณาปฏิสัมพันธ์ร่วม 5 ทาง พบว่า ตัวแปรอิสระทั้ง 5 ตัวดังกล่าวข้างต้นร่วมกันส่งผลต่อจำนวนข้อสอบที่ใช้อย่างไม่มีนัยสำคัญ

ข้อค้นพบที่ได้จากการวิจัยในครั้งนี้ บางประเด็นก็ตรงตามสมมติฐานที่ผู้วิจัยกำหนดไว้ บางประเด็นก็ไม่ตรงตามสมมติฐานที่กำหนดไว้ ซึ่งผู้วิจัยได้ตั้งข้อสังเกตไว้ในการอภิปรายผลการวิจัยแยกเป็นประเด็นต่างๆ สอดคล้องกับวัตถุประสงค์ของการวิจัย ดังนี้

อภิปรายผลการวิจัย

ในการวิจัยครั้งนี้ ผู้วิจัยได้ตั้งสมมติฐานไว้ 10 ประเด็น ซึ่งเกิดจากการจัดกลุ่มตัวแปรอิสระ 5 ตัว คือ ค่าความยากง่ายของข้อสอบข้อแรก วิธีการคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบ และความสามารถของผู้สอบกับตัวแปรตาม 2 ตัว คือ ค่าความเที่ยงตรงเชิงสภาพของการทดสอบและจำนวนข้อสอบที่ใช้ในการทดสอบ จากการวิเคราะห์ข้อมูลในบทที่ 4 พบว่า ผลของการวิจัยตรงตามสมมติฐานที่ตั้งไว้ 8 ข้อจาก 10 ข้อ โดยเฉพาะผลของตัวแปรอิสระ 5 ตัวต่อจำนวนข้อสอบที่ใช้ในการทดสอบนั้น ผลการวิจัยตรงตามสมมติฐานทุกประเด็น ส่วนผลของตัวแปรอิสระต่อค่าความเที่ยงตรงเชิงสภาพนั้น ผลการวิจัยตรงตามสมมติฐาน 3 ประเด็น ประเด็นที่ไม่ตรงตามสมมติฐาน ก็คือ วิธีการประมาณค่าความสามารถของผู้สอบและความสามารถของผู้สอบต่อค่าความเที่ยงตรงเชิงสภาพ

สำหรับผลของการวิจัยที่ค้นพบในครั้งนี้ มีประเด็นที่น่าสนใจสำหรับการอภิปรายแยกตามสมมติฐานของการวิจัย ดังนี้

1. เมื่อใช้ข้อสอบที่มีความยากง่ายต่างกันเป็นข้อสอบข้อแรกในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์แล้ว ปรากฏว่า ค่าความเที่ยงตรงเชิงสภาพไม่ต่างกัน ซึ่งข้อค้นพบจากการวิจัยนี้ตรงตามสมมติฐานที่ตั้งไว้ และ สอดคล้องกับข้อค้นพบของฮอร์ดี (Hulin and others. 1983 : 219 ; citing Lord. 1977) ที่พบว่าข้อสอบที่ใช้เมื่อเริ่มต้นในการทดสอบนั้นไม่มีผลต่อความถูกต้องในการประมาณค่าความสามารถของผู้สอบ ซึ่งความถูกต้องในการประมาณค่าความสามารถของผู้สอบนั้นมีความสัมพันธ์แบบผกผันกับค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า และ เนื่องจากความเที่ยงตรงนั้นมีค่าสูงที่สุดเท่ากับรากที่สองของความเชื่อมั่น เมื่อข้อสอบที่ใช้เริ่มต้นในการทดสอบไม่มีผลต่อความถูกต้องในการประมาณค่า หรือความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ ดังนั้นก็จะไม่มีผลต่อความเที่ยงตรงของการทดสอบด้วย

2. เมื่อใช้เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบต่างกันในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ ปรากฏว่า ความเที่ยงตรงเชิงสภาพของการทดสอบไม่ต่างกัน ข้อค้นพบจากการวิจัยในครั้งนี้เป็นไปตามสมมติฐานที่ตั้งไว้ ซึ่งน่าจะเป็นเพราะความเที่ยงตรงของแบบทดสอบนั้นจะขึ้นอยู่กับความถูกต้องในการวัด ซึ่งในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์นั้นสามารถกำหนดค่าความถูกต้องในการวัดให้เป็นค่าเดียวกันสำหรับผู้สอบทุกคนได้ โดยใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าเป็นตัวกำหนด ดังนั้น จึงทำให้ค่าความเที่ยงตรงของการทดสอบไม่ต่างกัน

3. เมื่อใช้วิธีการประมาณค่าความสามารถของผู้สอบต่างกันในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ ปรากฏว่า ค่าความเที่ยงตรงเชิงสภาพของการทดสอบต่างกัน ผลการวิจัยที่ค้นพบในครั้งนี้ไม่สอดคล้องกับสมมติฐานที่ตั้งไว้ กล่าวคือ ผู้วิจัยได้ตั้งสมมติฐานไว้ว่าหากใช้วิธีการประมาณค่าความสามารถของผู้สอบต่างกันค่าความเที่ยงตรงเชิงสภาพของการทดสอบจะไม่ต่างกัน แต่ผลการวิจัยพบว่าวิธีการประมาณค่าความสามารถวิธีของเบส์นั้นจะให้ค่าความเที่ยงตรงเชิงสภาพสูงกว่าวิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข ซึ่งสอดคล้องกับผลการวิจัยของกีฟฟอร์ดและสวามินาธาน (Gifford and Swaminathan, 1990 : 33) ที่พบว่า ถ้ากลุ่มตัวอย่างขนาดเล็กและข้อสอบน้อยข้อ การประมาณค่าความสามารถวิธี joint bayesian estimate จะมีความถูกต้องมากกว่า joint maximum likelihood และข้อค้นพบของโฮ (Ho, 1988) ที่พบว่าวิธีการประมาณค่าความสามารถวิธีของเบส์มีความเชื่อมั่นและมีประสิทธิภาพสูงกว่าวิธีความเป็นไปได้สูงสุด ซึ่งมักไม่ค่อยคงเส้นคงวานัก

4. เมื่อใช้เกณฑ์ยุติการทดสอบต่างกันในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ทำให้ค่าความเที่ยงตรงเชิงสภาพของการทดสอบต่างกัน ผลของการวิจัยข้อนี้เป็นไปตามสมมติฐานที่กำหนดไว้ ทั้งนี้เนื่องจากความเที่ยงตรงเชิงสภาพของการทดสอบนั้นจะแปรผันไปตามค่าความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ หรือแปรผันตรงกับค่าของความถูกต้องในการวัด ดังนั้น เมื่อกำหนดให้เกณฑ์ยุติการทดสอบ ซึ่งผู้วิจัยใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบแตกต่างกันออกไป 3 ระดับ จึงส่งผลให้ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแตกต่างกันออกไป

5. เมื่อความสามารถทางคณิตศาสตร์ของผู้สอบต่างกันในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ ปรากฏว่า ค่าความเที่ยงตรงเชิงสภาพของการทดสอบต่างกัน ผลของการวิจัยนี้ไม่สอดคล้องกับสมมติฐานที่กำหนดไว้ ซึ่งผู้วิจัยกำหนดไว้ว่า เมื่อผู้สอบมีความสามารถทางคณิตศาสตร์ต่างกัน จะไม่ทำให้ค่าความเที่ยงตรงเชิงสภาพของการทดสอบต่างกัน ทั้งนี้อาจเป็น เพราะ ค่าความสามารถของผู้สอบนั้นจะส่งผลต่อค่าความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องและค่าความน่าจะเป็นดังกล่าวจะส่งผลต่อค่าสารสนเทศของข้อสอบ ถ้าค่าความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องมีค่าสูงก็จะส่งผลให้ค่าสารสนเทศของข้อสอบและค่าสารสนเทศของแบบทดสอบสูงตามไปด้วย กรณีเช่นนี้จะส่งผลทำให้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบมีค่าต่ำลง นั่นก็คือ มีความถูกต้องในการ

วัดสูง ในทางกลับกันถ้าความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องมีค่าต่ำก็จะส่งผลให้ค่าสารสนเทศของข้อสอบและสารสนเทศของแบบทดสอบมีค่าต่ำ และจะส่งผลให้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่ามีค่าสูง นั่นก็คือ มีความถูกต้องในการวัดต่ำนั่นเอง และค่าความถูกต้องในการวัดนั้น เกี่ยวข้องโดยตรงกับค่าความเที่ยงตรงของแบบทดสอบ ดังนั้น จึงเป็นไปได้ที่เมื่อทดสอบกับผู้สอบที่มีความสามารถสูงจะมีความเที่ยงตรงเชิงสภาพสูงกว่าเมื่อทดสอบกับผู้สอบที่มีความสามารถต่ำกว่า อีกประการหนึ่งที่น่าจะเป็นไปได้ ก็คือ ข้อสอบในคลังข้อสอบนั้น หากพิจารณาโครงสร้างสารสนเทศของแบบทดสอบแล้ว จะพบว่า เป็นข้อสอบที่เหมาะสมกับผู้สอบที่มีความสามารถปานกลางค่อนข้างสูง ซึ่งหากกำหนดข้อสอบให้มีความเหมาะสมหรือตรงกับความสามารถของผู้สอบแล้ว ก็จะทำให้ค่าของความคลาดเคลื่อนมาตรฐานมีค่าต่ำลง (Haladyna and Roid, 1983)

เป็นที่น่าสังเกตว่า ค่าความเที่ยงตรงเชิงสภาพของการทดสอบ จำแนกตามความสามารถของผู้สอบนั้นจะไม่สูงมากนัก ทั้งนี้อาจเป็นเพราะ เมื่อแบ่งกลุ่มผู้สอบออกเป็น 3 กลุ่ม คือ ความสามารถสูง ปานกลาง และ ต่ำ ทำให้พิสัยของค่าความสามารถของผู้สอบที่จะนำมาหาค่าสหสัมพันธ์แคบลง พิสัยของค่าความสามารถที่แคบลงนี้จะทำให้ความแปรปรวนของค่าความสามารถมีค่าต่ำ ซึ่งจะส่งผลให้ค่าสหสัมพันธ์ซึ่งผู้วิจัยใช้เป็นค่าความเที่ยงตรงเชิงสภาพมีค่าต่ำไปด้วย

6. เมื่อใช้ข้อสอบที่มีค่าความยากง่ายต่างกันเป็นข้อสอบข้อแรกในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ ปรากฏว่าจำนวนข้อสอบที่ใช้ในการทดสอบต่างกัน ผลการวิจัยนี้ตรงตามสมมติฐานที่ตั้งไว้ นั่นคือ เมื่อใช้ข้อสอบข้อแรกที่มีค่าความยากง่ายต่างกัน จะทำให้ใช้ข้อสอบในการทดสอบมากน้อยต่างกัน ซึ่งหากเริ่มต้นด้วยข้อสอบที่มีค่าความยากปานกลาง ก็จะใช้ข้อสอบน้อยกว่าเริ่มต้นด้วยข้อสอบยากหรือง่าย และการเริ่มต้นด้วยข้อสอบที่ยากหรือง่ายนั้นจะใช้ข้อสอบไม่ต่างกัน อาจเป็นไปได้ว่า ถ้าใช้ข้อสอบข้อแรกตรงกับระดับความสามารถของผู้สอบแล้ว ก็จะช่วยให้ลดจำนวนข้อสอบกว่าเดิม เหตุที่เมื่อใช้ข้อสอบข้อแรกง่ายหรือยากแล้วจะใช้ข้อสอบมากกว่าเริ่มต้นด้วยข้อสอบที่มีค่าความยากปานกลาง เพราะเมื่อเริ่มต้นด้วยข้อสอบที่ยาก ข้อสอบที่ยากนี้ก็จะไม่เหมาะสมกับผู้สอบที่มีความสามารถต่ำ ในทางกลับกันเมื่อใช้ข้อสอบที่ง่าย ข้อสอบดังกล่าวนี้ก็จะไม่เหมาะสมกับผู้สอบที่มีความสามารถสูง แต่ถ้าหากใช้ข้อสอบที่มีค่าความยากง่ายปานกลางแล้วข้อสอบที่ไม่เหมาะสมกับผู้สอบที่มีความสามารถสูงและต่ำก็จะน้อยลงกว่าเมื่อใช้ข้อสอบที่ง่ายหรือยาก

7. เมื่อใช้เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบต่างกันในการทดสอบแบบ

ปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ ปรากฏว่าจำนวนข้อสอบที่ใช้ในการทดสอบต่างกัน ผลการวิจัยนี้ตรงตามสมมติฐานที่กำหนดไว้ เพราะข้อสอบที่ถูกเลือกขึ้นมาใช้แต่ละวิธีนั้น จะมีสารสนเทศของข้อสอบแตกต่างกันออกไปซึ่งค่าสารสนเทศของข้อสอบดังกล่าวจะส่งผลต่อค่าความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ ตามสูตร $SEE = 1/\sqrt{I(\theta)}$ ซึ่ง $I(\theta)$ (สารสนเทศของแบบทดสอบ) ก็เป็นผลรวมของ $I(\theta, u)$ (สารสนเทศของข้อสอบ) ดังนั้นจึงทำให้วิธีการคัดเลือกข้อสอบแต่ละวิธีลู่เข้าหาความคลาดเคลื่อนในการประมาณค่าที่กำหนดเป็นเกณฑ์ยุติการทดสอบเร็วช้าต่างกัน ซึ่งการเร็วช้าต่างกันนี้ก็คือ การใช้ข้อสอบมากน้อยต่างกันในการทดสอบนั่นเอง

8. เมื่อใช้วิธีการประมาณค่าความสามารถของผู้สอบต่างกันในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ ปรากฏว่า จำนวนข้อสอบที่ใช้ในการสอบต่างกัน ผลการวิจัยนี้สอดคล้องกับสมมติฐานที่กำหนดไว้ นั่นคือ เมื่อใช้วิธีการประมาณค่าความสามารถของผู้สอบต่างกันแล้ว จำนวนข้อสอบที่ใช้ในการทดสอบก็จะต่างกันไปด้วย ซึ่งวิธีของเบส์จะใช้ข้อสอบมากกว่าวิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข ซึ่งต่างจากข้อค้นพบของต่าย เชียงจิ ที่พบว่า วิธีการประมาณค่าความสามารถของเบส์จะใช้ข้อสอบน้อยกว่าวิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (ต่าย เชียงจิ. 2530 :122)

9. เมื่อใช้เกณฑ์ยุติการสอบต่างกันในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ ปรากฏว่า จำนวนข้อสอบที่ใช้ในการสอบต่างกัน ผลการวิจัยนี้สอดคล้องกับสมมติฐานที่กำหนดไว้ ทั้งนี้ เนื่องมาจากเมื่อจำนวนข้อสอบที่ใช้ในการทดสอบเปลี่ยนไป ค่าความถูกต้องในการประมาณค่าความสามารถของผู้สอบ ซึ่งก็รวมไปถึงค่าความคลาดเคลื่อนในการประมาณค่าจะเกิดการเปลี่ยนแปลง กล่าวคือ เมื่อใช้ข้อสอบมากขึ้น ความเชื่อมั่นก็จะมีค่าสูงขึ้น นั่นก็คือ ค่าความคลาดเคลื่อนมาตรฐานก็จะมีค่าลดลงตามความสัมพันธ์ ดังสูตร $r_u = 1 - \overline{SEE}^2$ ดังนั้น หากกำหนดค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าต่างกัน จำนวนข้อสอบที่ถูกนำมาใช้ในการทดสอบก็จะต่างกันไป

10. เมื่อความรู้ความสามารถทางคณิตศาสตร์ของผู้สอบต่างกันในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ ปรากฏว่า จำนวนข้อสอบที่ใช้ในการสอบต่างกัน ผลการวิจัยนี้สอดคล้องกับสมมติฐานที่กำหนดไว้ นั่นคือ ผู้สอบที่มีความสามารถทางคณิตศาสตร์ต่างกันจะใช้ข้อสอบในการทดสอบไม่เท่ากัน ผู้สอบที่มีความสามารถสูง จะใช้ข้อสอบน้อยกว่าผู้สอบที่มีความสามารถปานกลาง และ ต่ำ ทั้งนี้อาจเป็นเพราะ เหตุผลเดียวกับที่กล่าวมาแล้วข้างต้น (ข้อ 5) กล่าวคือ ผู้สอบที่มีความสามารถสูงก็จะมีแนวโน้มจะเป็นในการตอบ

ข้อสอบได้ถูกต้องสูง ซึ่งจะส่งผลทำให้ค่าสารสนเทศของข้อสอบและค่าสารสนเทศของแบบทดสอบสูงตามไปด้วย แต่จะส่งผลทำให้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่ามีค่าต่ำลง และในการวิจัยครั้งนี้ ผู้วิจัยได้ใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าเป็นเกณฑ์ยุติการสอบ เมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าจากผู้สอบที่มีความสามารถสูงมีค่าต่ำลงเร็วกว่าผู้สอบที่มีความสามารถต่ำ นั้นแสดงว่า ผู้สอบที่มีความสามารถสูงใช้ข้อสอบน้อยกว่าผู้สอบที่มีความสามารถต่ำ

เมื่อพิจารณาเป็นภาพรวมในการเลือกตัวแปรอิสระไปใช้ในการทดสอบแบบปรับเหมาะ กับความสามารถของผู้สอบด้วยคอมพิวเตอร์แล้ว พบว่า ค่าความยากง่ายของข้อสอบข้อแรกที่ใช้เมื่อเริ่มต้นสอบนั้น ไม่ว่าจะเริ่มต้นสอบด้วยข้อสอบที่ง่าย ปานกลางหรือยากก็ตามก็ไม่ทำให้ความเที่ยงตรงเชิงสภาพต่างกันแต่อย่างใด แต่การเริ่มต้นทดสอบด้วยข้อสอบที่มีความยากง่ายปานกลางนั้นจะใช้ข้อสอบน้อยกว่า การเริ่มต้นทดสอบด้วยข้อสอบที่ยากหรือง่าย ทั้งนี้อาจเป็นไปได้ว่า ผู้สอบที่มีความสามารถสูงต้องเสียเวลาในการทำข้อสอบที่ง่าย และในทางกลับกัน ผู้สอบที่มีความสามารถต่ำก็เสียเวลาในการทำข้อสอบที่ยาก ซึ่งทั้ง 2 กรณีดังกล่าวมานี้ ผู้สอบจะได้ข้อสอบที่ไม่เหมาะสมกับความสามารถของตนเองส่วนหนึ่ง ดังนั้น จึงควรเริ่มต้นทดสอบด้วยข้อสอบที่มีค่าความยากง่ายปานกลาง ในกรณีที่เราไม่ทราบความสามารถพื้นฐานของผู้สอบก่อนสอบ แต่ถ้าหากทราบความสามารถพื้นฐานของผู้สอบก่อนที่จะทดสอบด้วยการทดสอบแบบปรับเหมาะแล้ว ก็ควรเริ่มต้นด้วยการใช้ข้อสอบที่มีความยากง่ายเหมาะกับระดับความสามารถของผู้สอบ

สำหรับการคัดเลือกข้อสอบข้อต่อไปในคลังข้อสอบ ซึ่งผู้วิจัยเลือกให้อยู่ 3 วิธี คือใช้ค่าความสามารถของผู้สอบจากการทำข้อสอบที่ผ่านมาแล้ว เทียบกับค่าความยากง่ายของข้อสอบในคลังข้อสอบ วิธีที่สองใช้ค่าความสามารถของผู้สอบจากการทำข้อสอบที่ผ่านมาแล้ว เทียบกับค่าความสามารถที่ทำให้ข้อสอบมีสารสนเทศสูงสุด และวิธีที่ 3 ใช้ค่าสารสนเทศของข้อสอบที่สูงที่สุด ซึ่งจากผลการวิจัย พบว่า ไม่ว่าจะใช้วิธีการใดในการเลือกข้อสอบจากคลังข้อสอบก็ตาม ค่าความเที่ยงตรงเชิงสภาพของการทดสอบก็ไม่ต่างกัน อย่างไรก็ตาม การเลือกข้อสอบจากคลังข้อสอบวิธีที่ 2 และวิธีที่ 3 นั้นจะใช้ข้อสอบน้อยกว่าวิธีแรก คือ ใช้ค่าความยากง่ายของข้อสอบเทียบกับความสามารถของผู้สอบ สำหรับการเลือกใช้วิธีการเลือกข้อสอบวิธีที่ 2 กับวิธีที่ 3 นั้น หากเรามีคลังข้อสอบขนาดใหญ่ ก็ควรเลือกข้อสอบด้วยวิธีที่ 3 คือ เลือกข้อสอบข้อที่มีสารสนเทศของข้อสอบสูงที่สุด ซึ่งพิจารณาจากข้อมูลแล้วมีแนวโน้มว่าจะใช้ข้อสอบน้อยกว่าวิธีที่ 2 แต่ถ้าเรามีคลังข้อสอบมีขนาดเล็ก ก็ควรเลือกข้อสอบด้วยวิธีที่ 2 เพราะถ้าใช้วิธีที่ 3 คอมพิวเตอร์จะเลือกข้อสอบที่มีค่าสาร

สนทนาล่าสุดในแต่ละครั้ง และใช้ข้อสอบเพียงไม่มากนัก ซึ่งจะทำให้ผู้สอบจำข้อสอบได้และจะไม่เกิดประโยชน์อันใดในการทดสอบเพื่อศึกษาความก้าวหน้าของนักเรียนในครั้งต่อไป

ในเรื่องของเกณฑ์ยุติการทดสอบนั้น พบว่า เมื่อใช้เกณฑ์ยุติการทดสอบต่างกันออกไป 3 เกณฑ์ คือ ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบเท่ากับหรือน้อยกว่า .30 .40 และ .50 ปรากฏว่า ทั้งจำนวนข้อสอบ และ ค่าความเที่ยงตรงเชิงสภาพของการทดสอบแตกต่างกันในแต่ละกรณี อย่างไรก็ตามหากพิจารณาแล้วจะพบว่า ถ้าใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าเท่ากับ .40 จะใช้ข้อสอบเพิ่มจากเกณฑ์ .50 ประมาณ 2 เท่า ถ้าใช้เกณฑ์ .30 จะใช้ข้อสอบเพิ่มจากเกณฑ์ .40 ประมาณ 2 เท่า และเพิ่มจากเกณฑ์ .50 ประมาณ 3 เท่า ส่วนค่าความเที่ยงตรงเชิงสภาพนั้นต่างกันประมาณ .04 ในแต่ละเกณฑ์ ดังนั้น ผู้วิจัยจึงเสนอแนะให้ใช้เกณฑ์ยุติการสอบ โดยใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถเท่ากับหรือน้อยกว่า .40 ซึ่งถ้าพิจารณาถึงความสัมพันธ์ระหว่างความเชื่อมั่นกับความคลาดเคลื่อนมาตรฐานในการประมาณค่าตามสูตร $r_{tt} = 1 - \text{SEE}^2$ จะพบว่า ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ .40 นี้ จะทำให้ข้อสอบมีความเชื่อมั่นในการวัดเท่ากับ .9165 ซึ่งนับว่าสูงมาก อย่างไรก็ตามหากต้องการให้การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์มีความถูกต้องสูงๆในการประมาณค่าความสามารถ เช่น การนำการทดสอบลักษณะนี้ไปใช้ในการคัดเลือกบุคคล ก็จำเป็นต้องใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ เท่ากับ .30 หรือ อาจจะต่ำกว่านี้ ถึงแม้ว่าจะใช้ข้อสอบมากขึ้น แต่ค่าความเที่ยงตรงเชิงสภาพก็สูงขึ้นด้วย มีเกณฑ์ยุติการทดสอบอีกเกณฑ์หนึ่งที่ใช้กันในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ นั่นก็คือ ใช้จำนวนข้อสอบที่คงที่ในการทดสอบแต่ละครั้งกับผู้สอบแต่ละคน เช่น กำหนดไว้ว่าเมื่อผู้สอบตอบข้อสอบครบ 20 ข้อก็จะยุติการทดสอบ ซึ่งการกำหนดเกณฑ์การยุติการสอบในลักษณะนี้นั้น ผู้วิจัยไม่ค่อยเห็นด้วยดังที่ได้กล่าวมาแล้วในตอนต้นๆ เพราะผู้สอบบางคนอาจไม่จำเป็นต้องทำข้อสอบเท่ากับที่กำหนดไว้ก็สามารถประมาณค่าความสามารถได้ ในขณะที่เดียวกันผู้สอบบางคนอาจต้องตอบข้อสอบมากกว่าที่กำหนดไว้จึงจะสามารถประมาณค่าความสามารถได้ การใช้จำนวนข้อสอบเป็นเกณฑ์ยุติการทดสอบนั้นก็จะมีอะไรที่ต่างไปจากการทดสอบแบบเดิม เพียงแต่เปลี่ยนวิธีการนำเสนอข้อสอบจากนำเสนอในกระดาษมาเป็นนำเสนอที่หน้าจอคอมพิวเตอร์ และจากผลการวิจัยที่ผ่านมา พบว่า ถ้าใช้จำนวนข้อสอบที่แน่นอนเป็นเกณฑ์ยุติการสอบแล้ว การประมาณค่าความสามารถของผู้สอบที่มีความสามารถต่ำจะมีความถูกต้องน้อยกว่าการประมาณค่าความสามารถของผู้สอบที่มีความสามารถสูงกว่า (Wainer, 1990 : 170)

การประมาณค่าความสามารถของผู้สอบด้วยวิธีความเป็นไปได้สูงสุดและวิธีของเบส์นั้น ต่างก็ทำให้ใช้ข้อสอบมากน้อยต่างกัน และความเที่ยงตรงเชิงสภาพก็ต่างกันไปด้วย ซึ่งวิธีของเบส์ ใช้ข้อสอบมากกว่าวิธีความเป็นไปได้สูงสุด 1 ข้อ แต่ก็ทำให้ค่าความเที่ยงตรงเชิงสภาพสูงกว่าถึง .06 ดังนั้น ผู้วิจัยจึงเสนอแนะให้ใช้วิธีการประมาณค่าความสามารถของผู้สอบด้วยวิธีของเบส์ เพราะไม่ค่อยยุ่งยากซับซ้อน ไม่ต้องมีการประมาณค่าซ้ำหลายรอบเพื่อให้ได้ค่าที่ดีที่สุด และ เพื่อเป็นการหลีกเลี่ยงปัญหาการตอบถูกทั้งหมดและผิดทั้งหมดของผู้สอบ ซึ่งเป็นข้อจำกัดของการ ประมาณค่าความสามารถด้วยวิธีความเป็นไปได้สูงสุดที่ไม่สามารถประมาณค่าความสามารถได้ แต่ในการวิจัยครั้งนี้ ผู้วิจัยได้แก้ปัญหาข้อจำกัดนี้ โดยเริ่มต้นด้วยการประมาณค่าความสามารถของผู้สอบด้วยวิธีการของเบส์ไปก่อน จนกว่าผู้สอบจะตอบถูกหรือผิดอย่างน้อย 1 ข้อ จึงจะกลับมาประมาณค่าความสามารถของผู้สอบด้วยวิธีความเป็นไปได้สูงสุด ซึ่งเป็นวิธีการที่วอร์ม (Warm) ได้แนะนำไว้

สำหรับความสามารถทางคณิตศาสตร์ของผู้สอบก่อนที่จะทดสอบด้วยการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ เรื่อง เศษส่วนนั้น พบว่า ผู้สอบที่มีความสามารถต่างกันจะใช้ข้อสอบมากน้อยต่างกัน และมีความเที่ยงตรงเชิงสภาพต่างกัน ผู้สอบที่มีความสามารถสูงจะใช้ข้อสอบน้อยข้อกว่า และมีความเที่ยงตรงเชิงสภาพสูงกว่าผู้สอบที่มีความสามารถต่ำกว่า สิ่งหนึ่งที่เป็นไปได้ที่ทำให้ค่าของความเที่ยงตรงเชิงสภาพไม่ตรงตามสมมติฐานที่ตั้งไว้ ก็คือ ข้อสอบในคลังข้อสอบ เรื่อง เศษส่วนในการวิจัยครั้งนี้ ถ้าพิจารณาค่าสารสนเทศของแบบทดสอบแล้ว พบว่า มีลักษณะเบ้ทางซ้าย หรือ เบ้ลบ นั่นก็หมายถึงข้อสอบส่วนใหญ่เหมาะสมกับผู้สอบที่มีความสามารถปานกลางค่อนข้างไปทางสูง วิธีแก้ปัญหาก็คือ คลังข้อสอบต้องมีข้อสอบมากๆ เพื่อให้ครอบคลุมระดับความสามารถของผู้สอบทุกระดับ อย่างไรก็ตามคงเลือกไม่ได้สำหรับตัวแปรอิสระเกี่ยวกับความสามารถของผู้สอบก่อนสอบที่จะเลือกว่า ผู้สอบที่มีความสามารถสูงให้สอบด้วยการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ ส่วนผู้สอบที่มีความสามารถปานกลางและต่ำให้ไปทดสอบแบบเดิม

ข้อเสนอแนะ

จากผลการวิจัยที่ได้ ผู้วิจัยมีข้อเสนอแนะแยกเป็น 2 ประเด็นใหญ่ๆคือ

1. การนำผลการวิจัยไปใช้

1.1 การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์นี้ เป็นการทดสอบรายบุคคลที่ค่อนข้างมีความถูกต้องเที่ยงตรงตามสภาพที่แท้จริงของนักเรียนตาม

โดเมนที่ถูกวัด การทดสอบแต่ละครั้งไม่จำเป็นต้องทำพร้อมกันทุกคน ดังนั้น ปริมาณของเครื่องคอมพิวเตอร์จึงไม่ใช่ข้อสรุปสำหรับการทดสอบลักษณะนี้ เมื่อผู้สอบต้องการที่จะทดสอบเมื่อไรก็เปิดเครื่องคอมพิวเตอร์และทดสอบด้วยตนเอง ผลของการทดสอบแต่ละครั้งอาจถูกเก็บไว้ในแฟ้มสะสมงาน (portfolio) ของนักเรียนแต่ละคน ซึ่งจะเป็นแหล่งข้อมูลแหล่งหนึ่งที่จะช่วยให้ครูผู้สอนสามารถติดตามความก้าวหน้าทางการเรียนการสอน นอกจากนั้นยังสามารถนำข้อมูลไปใช้ในการประเมินผลนักเรียนตามสภาพที่แท้จริง (authentic assessment) ได้อย่างมีประสิทธิภาพอีกด้วย

1.2 โรงเรียนและกลุ่มโรงเรียนที่มีความพร้อมอาจนำเอาการทดสอบลักษณะนี้ไปใช้ในลักษณะเครือข่าย (wide area network , WAN) โดยให้โรงเรียนใดโรงเรียนหนึ่งในกลุ่มโรงเรียนเป็นศูนย์กลางเก็บคลังข้อสอบไว้ ส่วนโรงเรียนที่เหลือในกลุ่มโรงเรียนจะเป็นโรงเรียนที่ใช้บริการ โดยมีเครื่องคอมพิวเตอร์อยู่ที่แต่ละโรงเรียน เมื่อมีการทดสอบเพื่อปรับปรุงการเรียนการสอนหรือตัดสินผลการเรียนในบางจุดประสงค์ บางเนื้อหา ก็สามารถใช้คลังข้อสอบร่วมกันได้ โดยไม่จำเป็นต้องสร้างข้อสอบใหม่ทุกครั้งที่มีการทดสอบ ส่วนการพัฒนาคลังข้อสอบเพื่อให้มีข้อสอบจำนวนมากและมีคุณภาพนั้น คณะครูในกลุ่มโรงเรียนต้องช่วยกันสร้าง ทดลองใช้ วิเคราะห์และคัดเลือกข้อสอบที่มีคุณภาพเก็บเข้าคลังข้อสอบอย่างต่อเนื่อง ซึ่งจะทำให้ได้คลังข้อสอบที่มีขนาดใหญ่ขึ้น มีข้อสอบหลากหลาย และมีคุณภาพสูง

2. การวิจัยในครั้งต่อไป

การวิจัยในครั้งนี้มีข้อจำกัดในเรื่องของเวลาและงบประมาณ ดังนั้น หากมีการวิจัยเกี่ยวกับการทดสอบลักษณะนี้อีก ผู้วิจัยควรคำนึงถึงสิ่งต่อไปนี้

2.1 เวลาที่ใช้ในการทดสอบเป็นรายข้อและเวลาที่ใช้สำหรับดำเนินการสอบของผู้สอบแต่ละคน ควรมีการศึกษาว่าถ้ากำหนดเวลาในการตอบข้อสอบแต่ละข้อแล้ว จะให้ผลต่างจากที่ไม่ได้กำหนดเวลาหรือไม่ และถ้าจะกำหนดเวลาในการทดสอบจะกำหนดเวลาในลักษณะใด

2.2 ควรมีการศึกษารูปแบบของการจัดข้อสอบให้กับผู้สอบในลักษณะของกลุ่มข้อสอบ (testlet) นั่นคือ ในแต่ละขั้นของการทดสอบนั้น แทนที่จะเลือกข้อสอบให้กับผู้สอบเพียง 1 ข้อ ก็เลือกข้อสอบให้ผู้สอบหลายๆข้อ เมื่อตอบข้อสอบในกลุ่มข้อสอบที่จัดให้หมดแล้ว จึงเลือกข้อสอบกลุ่มใหม่ที่เหมาะสมให้กับผู้สอบ ทำเช่นนี้ไปเรื่อยๆจนกระทั่งถึงเกณฑ์ยุติการทดสอบ การจัดข้อสอบในลักษณะของกลุ่มข้อสอบอาจช่วยสร้างความสมดุลของเนื้อหาและป้องกันความผิดพลาดในการประมาณค่าความสามารถแต่ละครั้งของผู้สอบ

2.3 ควรทดลองใช้การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์กับคลังข้อสอบเนื้อหาอื่นที่ไม่ใช่คณิตศาสตร์

2.4 ควรนำเอาการทดสอบลักษณะนี้ไปใช้กับการสอบวัดจิตพิสัย (affective domain) และทักษะพิสัย (psychomotor domain) ซึ่งมีระบบการให้คะแนนที่ไม่ใช่แบบแบ่งสองเป็น 0 กับ 1 แต่เป็นหลายชั้น (multiple categories) ซึ่งโมเดลตามทฤษฎีการตอบสนองข้อสอบที่นิยมใช้ คือ partial credit model (Masters. 1982) grade response model (Samejima. 1969) และ rating scale model (Andrich 1978a. 1978 b)

2.5 ควรสร้างรูปแบบของข้อสอบลักษณะของสื่อผสม (multimedia) อาจเป็น ภาพกราฟิกที่เคลื่อนไหวได้ มีเสียงประกอบ ซึ่งสามารถนำไปใช้ในการทดสอบความสามารถ ในการมองเห็นและการฟัง

2.6 ควรศึกษาในกรณีที่คลังข้อสอบมีข้อสอบวัดได้หลายมิติ (multidimension)

2.7 ควรมีการศึกษาเกี่ยวกับเกณฑ์ยุติการทดสอบที่ใช้เพียงเกณฑ์เดียวและหลาย เกณฑ์ผสมกัน ซึ่งเกณฑ์ยุติการทดสอบที่นิยมใช้กันอยู่ ก็คือ ค่าความคลาดเคลื่อนมาตรฐานใน การประมาณค่าความสามารถและจำนวนข้อสอบที่คงที่

2.8 ควรมีการทดลองสอบซ้ำ โดยใช้ข้อสอบชุดเดิมในการทดสอบแก่ผู้สอบแต่จัด ลำดับในการนำเสนอใหม่ แล้วพิจารณาว่า ได้ผลการทดสอบเหมือนเดิมหรือไม่

2.9 ควรมีการศึกษาเกี่ยวกับการข้ามข้อสอบเป็นบางข้อ (skip) การทบทวน ข้อสอบที่ทำไปแล้ว (review) การเลือกข้อสอบในคลังข้อสอบได้ด้วยตนเอง (self adapted) การสร้าง ความสมดุลของเนื้อหา (content balancing) ในรูปแบบต่างๆโดยเฉพาะในเรื่องของเนื้อหานั้น บุคคลโดยทั่วไปมักเกรงกันว่า การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วย คอมพิวเตอร์จะทดสอบได้ไม่ครอบคลุมเนื้อหา เพราะจำนวนข้อสอบน้อยลงกว่าการทดสอบ แบบเดิมประมาณครึ่งหนึ่ง

2.10 ควรศึกษาเกี่ยวกับรูปแบบของคลังข้อสอบเพิ่มเติม เช่น การแจกแจงของค่า พหาวามเตอร์ของข้อสอบในคลังข้อสอบควรมีลักษณะเป็นโค้งปกติ หรือ มีจำนวนข้อสอบเท่าๆกันใน แต่ละระดับความสามารถ (rectangular) หรือ ข้อสอบควรแจกแจงเป็นโค้งปกติในแต่ละระดับความ สามารถ

บรรณานุกรม

บรรณานุกรม

คณิต ไช้มุกด์. การพัฒนาเกณฑ์ในการคัดเลือกข้อสอบที่มีความเหมาะสมกับความสามารถของผู้สอบ. วิทยานิพนธ์ ค.ด. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย, 2533. อัดสำเนา.

งามนิตย ธาตุทองและ ปรีชา เครือวัลย์. “การให้คะแนนตนเองในการสอบแบบทดสอบชนิดเฟล็กซิเบิล,” รายงานการวิจัย. ขอนแก่น : มหาวิทยาลัยขอนแก่น, 2527.

ต่าย เขียงฉี. การศึกษาเปรียบเทียบประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบจากการทดสอบเทเลอร์รูปปิรามิดที่มีรูปแบบ จำนวนชั้นและวิธีการให้คะแนนที่แตกต่างกันโดยใช้วิธีมอนติคาร์โล. วิทยานิพนธ์ กศ.ด. กรุงเทพฯ : มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2534. อัดสำเนา.

✓ นันทิยา พิงคำ. การเปรียบเทียบภาพการทดสอบแบบ ซี เอ ที และแบบประเพณีนิยมในการวัดความสามารถด้านคำศัพท์ภาษาอังกฤษของนักเรียนชั้นมัธยมศึกษาปีที่ 3. วิทยานิพนธ์ ค.ม. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย, 2531. อัดสำเนา.

ผจงจิต อินทสุวรรณ. ทฤษฎีการตอบข้อคำถาม. สถาบันวิจัยพฤติกรรมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2528.

_____. “การวิเคราะห์แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนด้วยโมเดลโลจิสติก,” รายงานการวิจัย ฉบับที่ 45. สถาบันวิจัยพฤติกรรมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2534.

วรรณช แหยมแสง. การพัฒนาระบบการตรวจสอบความเป็นเอกมิตของแบบทดสอบ. วิทยานิพนธ์ ค.ด. กรุงเทพ : จุฬาลงกรณ์มหาวิทยาลัย, 2533. อัดสำเนา.

สำเร็จ บุญเรืองรัตน์. “การพัฒนาทฤษฎีลาเท้นเทรทเพื่อวิเคราะห์ข้อสอบ,” วารสารการวัดผลการศึกษา. 7(21) : 41-66 ; มกราคม-เมษายน, 2529.

เสรี ปรมชวลิตโรจน์. การศึกษาความเที่ยงตรงในการจำแนกความรู้และการประมาณค่าความสามารถในวิชาคณิตศาสตร์ของแบบทดสอบเฟล็กซิเบิลและแบบทดสอบรูปปิรามิดที่ใช้ขนาดชั้นแปรผัน. วิทยานิพนธ์ กศ.ม. กรุงเทพ : มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2531. อัดสำเนา.

สงบ ลักษณะ. “การเทียบคะแนนระหว่างแบบทดสอบ,” วารสารการวัดผลการศึกษา.

4(2) : 21-32 ; กันยายน-ธันวาคม 2525.

สุพัฒน์ สุกมลสันต์. Rasch model. ม.ป.ป. อัดสำเนา.

ศิริชัย กาญจนวาสี. การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ (Adaptive Testing).

กรุงเทพฯ : ศูนย์ทดสอบทางการศึกษา , คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย,
2538.

อนันต์ ศรีโสภณ. สถิติเบื้องต้น. กรุงเทพฯ : สำนักพิมพ์ไทยวัฒนาพานิชจำกัด , 2525.

Blackmore,L.M. “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.

Divgi,D.R. “Estimating Reliabilities of Computerized Adaptive Tests,” Applied Psychological

Measurement. 13(2) : 145-149 ; June, 1989.

Feldt,L.S. and others. “A Comparison of Five Methods for Estimating the Standard

Error of Measurement at Specific Score Levels,” Applied Psychological

Measurement. 9(4) : 351-361 ; December, 1985.

Gifford,J.A. and H.Swaminathan. “Bias and the Effect of Priors in Bayesian Estimation of

Parameters of Item Response Model,” Applied Psychological Measurement.

14(1) : 33-43 ; March, 1990.

Glass,G.V. and J.C.Stanley. Statistical Methods in Education and Psychology. Englewood

Cliffs,N.J. : Prentice-Hall, 1970.

Green,B.F. and others. “Technical Guidelines for Assessing Computerized Adaptive Tests,”

Journal of Educational Measurement. 21(4) : 347-360 ; Winter , 1984.

Hambleton,R.K. and H.Swaminathan. Item Response Theory : Principles and Application.

Boston : Kluwer-Nijhoff Publishing, 1985.

Hambleton,R.K. and 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.

Hankins,J.A. “The Effect of Variable Entry on Bias and Information of the Bayesian Adaptive

Testing Procedure,” Dissertation Abstracts International. 47(08) : 3013-A ; February, 1987.

Ho,R.G. “Using MicroCAT in Computerized Adaptive Testing : A Comparison of Three Adaptive

Testing Strategies,” Dissertation Abstracts International. 50(2) : 421-A ; August, 1989.

- Hulin, C.L., F. Drasgow, and C.K. Parsons. Item Response Theory : Application to Psychological Measurement. Homewood, IL : Dow-Jones Irwin, 1983.
- Lee, Sue Susan. Development and Implementation of Adaptive Testing Strategies for Introductory Graduate Level Courses, Dissertation Abstracts International. 47(10) : 3742-A ; April, 1987.
- Lord, F.M. and M.R. Novick. Statistical Theories of Mental Test Score. Massachusetts : Addison-Wesley Publishing Company, 1968.
- _____. Applications of Item Response Theory to Practical Testing Problem. New Jersey : Lawrence Erlbaum Associates, 1980.
- _____. "Maximum Likelihood and Bayesian Parameter Estimation in Item Response Theory," Journal of Educational Measurement. 31(1) : 3-31 ; 1971.
- Lumsden, J. The Construction of Unidimensional Tests, Psychological Bulletin. 58 : 122-131 ; 1961.
- _____. "Test Theory," in Annual Review of Psychology. Rosenzweig, M.R. and Porter, L.W. (Eds). Palo Alto, CA : Annual Reviews Inc., 1987.
- McDonald, R.P. and K.S. Ahlwardt. "Difficulty Factors in Binary Data," British Journal of Mathematical and Statistical Psychology. 27 : 82-89 ; 1974.
- McKinley, R. L. and M.D. Reckase. "A Comparison of the ANCILLES and LOGIST Parameter Estimation Procedure for the Three Parameter Logistic Model Using Goodness of Fit as a Criterion," Research report 80-2. Columbia, MD : University of Missouri, 1980.
- Moreno, K. C. and others. "Relationship Between Corresponding Armed Services Vocational Aptitude Battery and Computerized Adaptive Testing subtests," Applied Psychological Measurement. 8(1) : 155-163 ; 1984.
- Samegima, F. "A Use of the Information Function in Tailor Testing," Applied Psychological Measurement. 1 : 233-247 ; 1977.
- Skaggs, G. and J. Stevenson. "Comparison of Pseudo-Bayesian and Joint Maximum Likelihood Procedure for Estimating Item Parameters in the Three Parameters IRT Model," Applied Psychological Measurement. 13(4) : 391-402 ; December, 1989.

- Thorndike, R. K. Applied Psychometrics. Boston : Houghton Mifflin Company, 1982.
- Urry, V. W. "Tailored Testing : A Successful Application of Latent Trait Theory," Journal of Educational Measurement. 14(2) : 181-196 ; 1977.
- Vale, C. D. and D.J.Weiss "A Simulation Study of Stradaptive Ability Testing," Research report. Psychometric Method Program. Department of Psychology, University of Minesota, Minneapolis, 1975.
- Warm, T. A. "A Primer of Item Response Theory," Technical Report 941278. Oklahoma : U.S. Coast Guard Institute, 1978.
- Weiss, D. J. and G.G.Kingbury. "Aplication of Computerized Adaptive Testing to Education Problem," Journal of Educational Measurement. 21(4) : 361-375 ; winter, 1984.
- Weiss, D. J. and J.R.McBride. "Bias and Information of Bayesian Adaptive Testing," Applied Psychological Measurement. 8(3) : 273-285 ; Summer, 1984.
- Weiss, D. J. "Strategies of Adaptive Measurement," Research Report 74-5. Psychometric Method Program. Department of Psychology, University of Minesota, Minneapolis, 1974.
- _____. "Improving Measurement Quality and Efficiency with Adaptive Testing," Applied Psychological Measurement. 6 : 473-492 ; 1982.
- _____. "Adaptive Testing by Computer," Journal of consulting and Clinical Psychology. 53(6) : 774-789 ; 1985.
- _____. "Adaptive Testing," Educational Research Methodology and Measurement : International Handbook. New York : Pergamon Press, Inc., 1988 ; 372 - 376.
- Wisniewski, D. R. "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. "Solving Measurement Problems with the Rasch Model," Journal of Educational Measurement. 14(2) : 97-116 ; 1977.
- Wright, B. D. and M.H.Stone. Best Test Design. Chicago : Mesa Press, 1979.

ภาคผนวก

ฟาเซทสำหรับสร้างข้อสอบเรื่อง เศษส่วน
ชั้น ป.6 ฉบับที่ 1-7

จุดประสงค์/เนื้อหา/ข้อ	ฟาเซท	
	คำถาม	คำตอบ
1. มีความเข้าใจพื้นฐานเกี่ยวกับเรื่องเศษส่วน		
1.1 ความหมายของเศษส่วน (40, 41, 42)	ก. สัญลักษณ์ 1. รูปที่แบ่งเป็นส่วนเท่า ๆ กัน 2. เศษส่วน	ก. สัญลักษณ์ 1. รูปที่แบ่งเป็นส่วนเท่า ๆ กัน 2. เศษส่วน
1.2 ชนิดของเศษส่วน (26, 27)	ก. ชนิด 1. เศษส่วนแท้ 2. เศษเกิน 3. จำนวนคละ ข. จัดกลุ่ม 1. ประเภทเดียวกัน 2. คนละประเภท ค. สมบัติ 1. เศษส่วนแท้ 2. เศษเกิน 3. จำนวนคละ ง. การจัดกลุ่ม 1. เข้าพวก 2. ไม่เข้าพวก	ก. ชนิด 1. เศษส่วนแท้ 2. เศษเกิน 3. จำนวนคละ ข. จำนวนเศษส่วนที่เสนอ 1. เดี่ยว ($\frac{1}{2}$) 2. กลุ่ม ($\frac{1}{2}$, $\frac{1}{4}$, $\frac{4}{5}$)
1.3 เศษส่วนของจำนวนนับ (1, 2)	การนำเสนอ 1. เศษส่วนของจำนวนนับ 2. รูปที่แบ่งเป็นส่วนเท่า ๆ กัน ลักษณะ 1. สัญลักษณ์ 2. ความหมาย	การนำเสนอ 1. เศษส่วนของจำนวนนับ 2. รูปที่แบ่งเป็นส่วนเท่า ๆ กัน ลักษณะ 1. สัญลักษณ์ 2. ความหมาย

จุดประสงค์/เนื้อหา/ข้อ	พาเซท	
	คำถาม	คำตอบ
1.4 เศษส่วนอย่างต่ำ (43, 44, 45)	<u>การเปรียบเทียบ</u> 1. เท่ากัน 2. ไม่เท่ากัน <u>ชนิด</u> 1. เศษส่วนแท้ 2. เศษเกิน 3. จำนวนคละ <u>สัญลักษณ์เศษส่วน</u> 1. ตัวเลข 2. ตัวอักษร (จัดกลุ่มและไม่จัดกลุ่ม) <u>วิธีการ</u> 1. นำจำนวนเต็มหาร หรือคูณ 2. นำเศษส่วนที่มีเศษและส่วน เท่ากันหารหรือคูณ 3. นำเศษส่วนที่มีเศษและส่วน ไม่เท่ากันหารหรือคูณ	<u>ชนิด</u> 1. เศษส่วนแท้ 2. เศษเกิน 3. จำนวนคละ <u>วิธีการ</u> 1. นำจำนวนเต็มหาร หรือคูณ 2. นำเศษส่วนที่มีเศษและส่วน เท่ากันหารหรือคูณ 3. นำเศษส่วนที่มีเศษและส่วน ไม่เท่ากันหารหรือคูณ
	1.5 การแปลงเศษเกินเป็น จำนวนคละ <u>ชนิด</u> 1. เศษเกิน 2. จำนวนคละ <u>การเปรียบเทียบ</u> 1. เท่ากัน 2. ไม่เท่ากัน <u>สัญลักษณ์แทนจำนวน</u> 1. ตัวเลข 2. ตัวอักษร	<u>ชนิด</u> 1. เศษเกิน 2. จำนวนคละ <u>การระบุค่าที่เท่ากัน</u> 1. ระบุเฉพาะเศษหรือส่วน 2. ระบุทั้งเศษและส่วน <u>การแจกแจงองค์ประกอบ</u> 1. แจกแจง 2. ไม่แจกแจง

จุดประสงค์/เนื้อหา/ข้อ	ฟาเซท	
	คำถาม	คำตอบ
2. เมื่อกำหนดเศษส่วนให้สามารถเขียนเป็นเศษส่วนที่มีค่าเท่าเดิมโดยที่ตัวเศษหรือตัวส่วนมีค่าตามที่กำหนดให้ได้ (37, 38, 39)	ชนิด 1. เศษส่วนแท้ 2. เศษเกิน 3. จำนวนคละ การเปรียบเทียบ 1. เท่ากัน 2. ไม่เท่ากัน ลักษณะแทนจำนวน 1. รูป 2. สัญลักษณ์	
3. เมื่อกำหนดเศษส่วนสองจำนวนให้สามารถเปรียบเทียบและใช้สัญลักษณ์ $>$ $<$ หรือ $=$ ได้ (3 4 5 6 7)	ลักษณะของเศษส่วน 1. เศษเท่ากัน 2. ส่วนเท่ากัน 3. เศษน้อย/มากกว่าครึ่งหนึ่งของส่วน สัญลักษณ์ 1. ตัวเลข 2. ตัวอักษร 3. รูป การเรียงลำดับเศษส่วน 1. มากไปหาน้อย 2. น้อยไปหามาก	เครื่องหมายที่ใช้ 1. มากกว่า 2. น้อยกว่า 3. เท่ากับ ลักษณะของเศษส่วน 1. เศษเท่ากัน 2. ส่วนเท่ากัน 3. เศษน้อย/มากกว่าครึ่งหนึ่งของส่วน
4. เมื่อกำหนดเศษส่วนใด ๆ สามารถบวก ลบ คูณ หารระคนได้ - การบวก (8-10) - การลบ (11-14)	ก. รูปที่นำเสนอ 1. แรเงาตรงตามจำนวน  = 4/8 2. แรเงาเป็นสัดส่วน  = 1/2	ก. สมบัติ 1. สลับที่ 2. เปลี่ยนกลุ่ม ข. การนำเสนอ 1. จำนวนโดด

จุดประสงค์/เนื้อหา/ข้อ	พาเซท	
	คำถาม	คำตอบ
	ข. ตัวแหน่งที่ให้เติมส่วนที่หายไป 1. เศษ/ตัวคูณเศษ 2. เศษส่วนที่มีส่วนไม่เท่ากัน ง. สัญลักษณ์ 1. ตัวเลข 2. ตัวอักษร	2. จำนวนหลายจำนวนที่กระทำการกันอยู่
- การคูณ (15-18)	ก. ความสัมพันธ์การบวก, คูณ 1. จำนวนครั้งของการบวกเศษส่วนที่เท่ากัน 2. เศษส่วนที่เท่ากันที่นำมาบวก ประโยคสัญลักษณ์ ง. สัญลักษณ์ 1. ตัวเลข 2. ตัวอักษร ง. ส่วนที่ให้เติม 1. ตัวตั้ง 2. ตัวคูณ 3. เศษหรือส่วนของผลลัพธ์	ก. ชนิดของจำนวน 1. จำนวนเต็ม 2. เศษส่วน
- การหาร (19-22)	ก. ประโยคสัญลักษณ์ 1. จำนวนเต็ม 2. เศษส่วน 3. ตัวอักษร ข. ส่วนที่ให้เติม 1. ตัวตั้ง 2. ตัวหาร	ก. สัญลักษณ์ 1. ตัวเลข 2. ตัวอักษร 3. รูป

จุดประสงค์/เนื้อหา/ข้อ	ฟาเซท	
	คำถาม	คำตอบ
	3. ผลลัพธ์ ค. การเปรียบเทียบ 1. เท่ากัน 2. ไม่เท่ากัน	
- การบวกลบคูณหารระคน (46-50)	ก. เครื่องหมาย 1. บวก 2. ลบ 3. คูณ 4. หาร ข. ชนิด 1. จำนวนเต็ม 2. เศษส่วน ค. สัญลักษณ์ 1. ตัวเลข 2. ตัวอักษร 3. รูป ง. การเปรียบเทียบ 1. เท่ากัน 2. ไม่เท่ากัน (ต่างจาก)	ก. ชนิด 1. จำนวนเต็ม 2. เศษส่วน ข. สัญลักษณ์ 1. ตัวเลข 2. ตัวอักษร 3. รูป ค. วิธี 1. หา ค.ร.น. 2. แปลงจำนวนคละ 3. กลับเศษเป็นส่วน 4. เปลี่ยนเครื่องหมาย $\div \rightarrow \times$ ง. สมบัติ 1. สถานที่ 2. เปลี่ยนกลุ่ม 3. แจกแจง
5. เมื่อกำหนดเศษซ้อนอย่างง่ายที่เฉพาะตัวเศษหรือตัวส่วนมีการบวกลบคูณหารอย่างใดอย่างหนึ่งให้	ก. เครื่องหมาย 1. บวก 2. ลบ 3. คูณ	ก. ชนิดของจำนวน 1. เศษซ้อน 2. เศษส่วนธรรมดา 3. จำนวนเต็ม

จุดประสงค์/เนื้อหา/ข้อ	พาเซท	
	คำถาม	คำตอบ
สามารถหามลัทธิได้ (34-36)	4. <u>หาร</u> ข. <u>ชนิดของจำนวน</u> 1. <u>เศษส่วน</u> 2. <u>จำนวนเต็ม</u> ค. <u>สัญลักษณ์</u> 1. <u>ตัวเลข</u> 2. <u>ตัวอักษร</u> ง. <u>ส่วนที่ให้เต็ม</u> 1. <u>ตัวตั้ง</u> 2. <u>ตัวหาร</u> จ. <u>การแปลงรูป</u> 1. <u>เศษส่วน → เศษซ้อน</u> 2. <u>เศษซ้อน → เศษส่วน</u>	
6. เมื่อกำหนดโจทย์ปัญหา เกี่ยวกับการบวก ลบ คูณ หาร เศษส่วนให้สามารถ แสดงวิธีทำได้ (28-33)	ก. <u>เครื่องหมายการทำงาน</u> 1. <u>บวก</u> 2. <u>ลบ</u> 3. <u>คูณ</u> 4. <u>หาร</u> ข. <u>ชนิดของจำนวน</u> 1. <u>จำนวนเต็ม</u> 2. <u>เศษส่วน</u> ค. <u>สถานการณ์</u> 1. <u>เงิน</u> 2. <u>งาน</u> 3. <u>สิ่งของ</u> 4. <u>พื้นที่</u> 5. <u>ความยาว</u>	ก. <u>ลักษณะการตอบ</u> 1. <u>ประโยคสัญลักษณ์</u> 2. <u>คำตอบในรูปของการ</u> <u>การทำงาน</u>

จุดประสงค์/เนื้อหา/ข้อ	ฟächer	
	คำถาม	คำตอบ
	6. น้ำหนัก 7. เวลา 8. ปริมาตร	

ฟาเซทสำหรับสร้างข้อสอบ
ความรู้ความสามารถทางคณิตศาสตร์
ชั้น ป.6 ฉบับที่ 8

จุดประสงค์/เนื้อหา/ข้อ	ฟาเซท	
	คำถาม	คำตอบ
สามารถแปลความ, ตีความ, ขยายความเกี่ยวกับ 1. จำนวนและตัวเลข (1-3)	ก. <u>ผลต่างของค่าประจำหลัก</u> 1. หลักสิบ 2. หลักร้อย ข. <u>ค่าประมาณใกล้เคียง</u> 1. เต็มร้อย 2. เต็มพัน ค. <u>ค่าของจำนวน</u> 1. มากที่สุด 2. น้อยที่สุด	ก. <u>วิธีคิด</u> 1. จำนวนเท่า 2. ไม่คำนึงถึงค่าประจำหลัก 3. คิดเป็นหลักเดียวกัน ข. <u>ค่าประมาณใกล้เคียง</u> 1. เต็มพันมาก/น้อย 2. เต็มหมื่นมาก/น้อย ก. <u>จำนวน</u> 1. ค่าคงที่ 2. 1 ข. <u>เครื่องหมาย</u> + - x ÷
2. สมการและการแก้สมการ (4-7)	ก. <u>การนำเสนอ</u> 1. สมการมีตัวแปร 1 ตัว 2. โจทย์ปัญหา	ก. <u>สมบัติ</u> 1. การบวก 2. การลบ 3. การคูณ 4. การหาร ข. <u>การนำเสนอ</u> 1. สมการมีตัวแปร 1 ตัว 2. โจทย์ปัญหา

จุดประสงค์/เนื้อหา/ข้อ	พาเซท	
	คำถาม	คำตอบ
3. ตัวประกอบของจำนวนนับ (8-12)	ก. การนำเสนอสอ 1. แจกแจงตัวประกอบ 2. ไม่แจกแจง	ก. ลักษณะตัวประกอบ 1. ตัวประกอบเฉพาะ 2. ไม่ใช่ ตัวประกอบเฉพาะ ข. สิ่งที่ต้องการ 1. ค.ร.น. 2. ห.ร.ม.
4. มุมและส่วนของเส้นตรง (13-15)	ก. รูปมุม 1. รูปมุมมีชื่อกำกับ 2. รูปมุมตรง บอกขนาดมุม 1 มุม ให้หามุมประชิด 3. รูปมุมรอบจุดบอกขนาด 2 มุม ให้หามุมที่เหลือ	ก. สัญลักษณ์แทนมุม 1. ถูกต้อง 2. ไม่ถูกต้อง ข. ขนาดมุม 1. ถูกต้อง 2. ไม่ถูกต้อง
5. ทิศและแผนผัง (16-17)	ก. สิ่งที่ต้องการทราบ 1. ทิศ 2. มุม	ก. ตำแหน่ง 1. ทิศ 2. จุดดูทิศ
6. เส้นขนาน (18, 19)	ก. สิ่งที่น่าเสนอ 1. รูปเส้นขนานมีเส้นตัด 2. สมบัติเส้นขนาน	ก. สมบัติของเส้นขนาน 1. มุมแย้ง 2. มุมภายในบนข้างเดียวกันของเส้นตัด มีค่า 180° 3. มุมภายในและมุมภายนอกบนข้างเดียวกันของเส้นตัด
7. เศษส่วน (20-24)	ก. นิยามเศษส่วนอย่างต่ำ ข. การเปรียบเทียบเศษส่วน 1. เรียงจากน้อยไปมาก 2. เรียงจากมากไปน้อย	ก. ชนิดของเศษส่วน 1. เศษส่วนแท้ 2. เศษเกิน 3. จำนวนคละ

จุดประสงค์/เนื้อหา/ข้อ	พาเซท	
	คำถาม	คำตอบ
	ค. การกระทำการ 1. การบวก 2. การลบ 3. การคูณ 4. การหาร	ข. ลักษณะของเศษส่วน 1. มีส่วนเท่ากัน 2. มีเศษเท่ากัน 3. มีเศษเท่ากับส่วน
8. ทศนิยม (25-31)	เศษส่วน ทศนิยม ค่าประจำหลัก การกระจายทศนิยม การหารทศนิยม การคูณทศนิยม 1. ทศนิยมตัวตั้ง 2. ทศนิยมตัวคูณ	ทศนิยม เศษส่วน การเปรียบเทียบ 1. มากกว่า 2. น้อยกว่า ค่าประจำหลักทศนิยม จำนวนหลังจุดทศนิยม 1. ตัวตั้งบวกตัวคูณ 2. ตัวตั้งลบตัวคูณ 3. ตัวตั้งหารตัวคูณ 4. ตัวตั้งคูณตัวคูณ
9. รูปสามเหลี่ยม (32,33,37)	ก. พื้นที่ ข. ความยาวรอบรูป ค. สามเหลี่ยมคล้าย	ก. ชนิดของสามเหลี่ยม - หน้าจั่ว - ด้านเท่า - มุมฉาก - มุมแหลม ข. ส่วนประกอบ 1. ฐาน 2. สูง

จุดประสงค์/เนื้อหา/ข้อ	พาเซท	
	คำถาม	คำตอบ
		ค. วิธีการหาพื้นที่ 1. ฐานบวกสูง 2. ฐานลบสูง 3. ครึ่งหนึ่งของฐานบวกสูง 4. ครึ่งหนึ่งของฐานคูณสูง
10. รูปสี่เหลี่ยม (34-36, 38-39)	ก. สมบัติ - ด้าน - มุม - เส้นทแยงมุม - พื้นที่	ก. ชนิดของสี่เหลี่ยม 1. ผืนผ้า 2. คางหมู 3. ด้านขนาน 4. ขนมห้อยกปูน 5. จตุรัส 6. รูปว่าง ข. ส่วนประกอบ 1. ฐาน 2. สูง
11. วงกลม (40-41)	ก. สมบัติ 1. ถูก 2. ผิด	ก. ส่วนประกอบของวงกลม 1. รัศมี 2. เส้นผ่าศูนย์กลาง 3. เส้นรอบวง 4. พื้นที่ ข. ปริมาณ 1. เท่ากัน 2. มากที่สุด 3. น้อยที่สุด
12. หาประยุกต์ (ร้อยละ) (42-46)	ก. เหตุการณ์ 1. เป็นไปได้ 2. เป็นไปไม่ได้	ก. สภาพ 1. ทุน 2. กำไร

จุดประสงค์/เนื้อหา/ข้อ	พาเซท	
	คำถาม	คำตอบ
13. รูปทรงและปริมาตร (47-48)	ข. สลัภาพ 1. ทุน 2. กำไร 3. ขาดทุน ค. ปริมาณ 1. ร้อยละ 2. จำนวน ก. รูปทรง 1. สี่เหลี่ยม 2. พีระมิด 3. ทรงกลม ข. ขนาด 1. เท่ากัน 2. ไม่เท่ากัน	3. ขาดทุน ก. ส่วนประกอบ 1. ด้านกว้าง 2. ด้านยาว 3. พื้นฐาน 4. ระนาบ
14. แผนภูมิและกราฟ (49-50)	ก. การนำเสนอข้อมูล 1. รูปกราฟ 2. ข้อความ ข. 1. หน่วย 2. คู่/กลุ่ม	ก. ปริมาณ 1. มากที่สุด 2. น้อยที่สุด 3. เท่ากัน

ค่าพารามิเตอร์ของข้อสอบในคลังข้อสอบ

	Id	prior	key	a	b	c	amax	imax		Id	prior	key	a	b	c	amax	imax
1	1	101	4	1.089	.5350	.1550	.6548	.6362	34	34	221	1	1.500	1.686	.1540	1.773	1.209
2	2	102	3	1.064	.6670	.1520	.7878	.8108	35	35	222	4	1.042	1.626	.1640	1.757	.5726
3	3	104	2	.8560	2.214	.2680	2.439	.3162	36	36	229	1	1.170	1.281	.2580	1.441	.6025
4	4	105	3	.9790	2.211	.1320	2.329	.5371	37	37	230	4	1.299	1.839	.2780	1.991	.7140
5	5	106	2	1.198	1.350	.0920	1.422	.8676	38	38	234	2	1.207	2.540	.2480	4.052	.6539
6	6	107	1	.7070	1.137	.2400	1.389	.2279	39	39	235	2	1.631	1.240	.1980	1.336	1.315
7	7	109	4	.9610	.2410	.2490	.4314	.4137	40	40	240	1	1.384	.4120	.1290	.4940	1.079
8	8	111	3	.8100	.1910	.2390	.4107	.2997	41	41	241	4	1.684	.5900	.1570	.6682	1.516
9	9	113	4	1.298	1.461	.2240	1.592	.7922	42	42	242	3	1.180	1.253	.1330	1.351	.7787
10	10	114	2	1.513	2.055	.1920	2.156	1.145	43	43	244	3	1.086	.2950	.2040	.4415	.5784
11	11	115	2	.7530	-.075	.2900	.1940	.2343	44	44	245	3	.8040	.8990	.2310	1.115	.2999
12	12	119	3	.9440	2.002	.1870	2.161	.4498	45	45	247	1	1.005	.4050	.2220	.5733	.4768
13	13	123	1	.9320	.6420	.2180	.6212	.4132	46	46	249	2	1.133	2.234	.2140	2.379	.6154
14	14	126	4	1.613	2.720	.1800	2.810	1.331	47	47	302	3	1.169	1.679	.2130	1.820	.6563
15	15	127	1	.9460	1.223	.1030	1.323	.5298	48	48	303	2	1.651	1.481	.2670	1.597	1.179
16	16	128	3	.9660	.7270	.2950	.9390	.3817	49	49	312	4	1.291	2.295	.2350	2.431	.7672
17	17	134	2	1.025	1.667	.2260	1.834	.4921	50	50	313	3	1.499	.8750	.1510	.9604	1.215
18	18	135	4	1.439	1.562	.2070	1.674	1.006	51	51	318	4	1.108	.4010	.1470	.5142	.6686
19	19	136	2	1.110	1.044	.2570	1.212	.5434	52	52	322	2	.7980	.3520	.1890	.5411	.3202
20	20	137	3	1.473	.8130	.2840	.9486	.9072	53	53	324	2	1.090	-.205	.2040	-.059	.5806
21	21	139	1	.9380	.9570	.2010	1.125	.4324	54	54	325	1	1.610	-.302	.2210	-.197	1.226
22	22	140	1	1.364	.8910	.1490	.9839	1.009	55	55	326	3	1.301	-.354	.2340	-.219	.7807
23	23	142	2	.9290	.2540	.2190	.4343	.4098	56	56	329	1	1.197	-.194	.2480	-.042	.6431
24	24	144	4	1.313	.9870	.1530	1.085	.9283	57	57	330	2	1.558	.0400	.2470	.1568	1.092
25	25	145	1	1.197	1.355	.1870	1.480	.7233	58	58	332	3	.9840	1.780	.2870	1.984	.4024
26	26	150	2	1.173	2.633	.2150	2.774	.6583	59	59	335	2	.8820	-.266	.1720	-.106	.4041
27	27	202	4	.9180	2.118	.2300	2.307	.3917	60	60	336	1	1.623	.5580	.2350	.6644	1.213
28	28	203	4	.9380	1.248	.2020	1.417	.4316	61	61	341	2	1.046	.7460	.2790	.9348	.4620
29	29	205	3	1.203	1.232	.2050	1.365	.7058	62	62	342	4	1.569	.8150	.1610	.9005	1.306
30	30	208	2	1.710	1.706	.1760	1.790	1.507	63	63	344	3	.8250	-.474	.1740	-.302	.3522
31	31	213	4	1.297	.8020	.1360	.8931	.9355	64	64	345	1	1.040	.3150	.2240	.4786	.5086
32	32	216	2	1.744	1.572	.1530	1.646	1.638	65	65	347	2	1.375	1.179	.2960	1.328	.7718
33	33	218	4	.9040	.4550	.1100	.5658	.4774	66	66	349	2	.8320	2.292	.2820	2.531	.2906

	id	prior	key	a	b	c	qmax	imax		id	prior	key	a	b	c	qmax	imax
67	67	401	1	.8710	.6420	.1870	.8139	.3830	100	100	541	4	1.550	.6530	.1720	.7439	1.248
68	68	403	1	1.106	1.099	.2280	1.255	.5708	101	101	544	4	.9430	.0640	.1880	.2234	.4480
69	69	404	2	.8730	1.123	.1290	1.253	.4295	102	102	545	4	1.319	.3820	.1630	.4847	.9192
70	70	405	4	1.675	1.293	.1250	1.359	1.593	103	103	547	1	1.481	1.705	.1980	.1810	1.084
71	71	409	1	1.555	.7410	.2800	.8683	1.019	104	104	548	3	1.013	1.652	.2130	.1814	.4929
72	72	416	2	1.364	2.541	.1480	2.633	1.011	105	105	549	3	.9510	1.684	.2410	1.872	.4115
73	73	422	2	.8130	.6660	.2050	.8627	.3224	106	106	603	2	1.257	1.685	.1510	1.787	.8541
74	74	425	1	1.072	.1540	.2450	.3228	.5188	107	107	605	1	1.373	1.165	.1500	1.258	1.021
75	75	428	1	.8230	-.507	.1800	-.230	.3465	108	108	608	3	1.637	1.920	.0610	1.957	1.719
76	76	430	3	.8670	2.394	.2170	2.586	.3583	109	109	616	2	.8800	.5520	.1760	.7149	.3992
77	77	434	4	1.134	.5270	.2400	.6844	.5862	110	110	617	1	1.499	1.429	.1410	1.510	1.238
78	78	436	2	1.586	2.115	.2010	2.214	1.236	111	111	618	3	1.255	1.055	.1380	1.150	.8726
79	79	440	1	1.473	.9650	.1220	1.039	1.239	112	112	620	3	.9380	.0810	.2320	.2669	.4074
80	80	441	4	.9550	1.236	.1760	1.386	.4701	113	113	621	2	1.501	1.319	.1570	1.407	1.204
81	81	444	2	1.266	.6460	.2700	.7986	.6890	114	114	622	2	1.490	1.014	.2110	1.124	1.070
82	82	445	4	.8380	.5780	.1310	.7150	.3942	115	115	625	1	.9980	-.551	.2180	-.384	.4738
83	83	501	4	1.038	1.331	.1460	1.451	.5879	116	116	626	4	1.024	-.115	.2230	.0507	.4940
84	84	505	2	1.576	2.088	.2700	2.211	1.068	117	117	627	4	1.365	.1240	.2560	.2606	.8233
85	85	510	1	1.196	1.103	.2880	1.272	.5933	118	118	630	1	2.338	1.406	.1180	1.451	3.145
86	86	512	3	1.543	2.253	.0820	2.304	1.467	119	119	631	4	1.076	2.645	.1920	2.787	.5789
87	87	514	3	1.056	1.365	.1270	1.471	.6308	120	120	635	4	.8990	.4630	.1080	.5728	.4739
88	88	516	3	1.019	.0210	.2280	.1901	.4845	121	121	636	3	1.248	2.015	.1980	2.140	.7699
89	89	521	3	1.332	.2890	.2150	.432	.8489	122	122	637	3	1.218	1.052	.2110	1.186	.7153
90	90	522	3	1.283	.1190	.1520	.2192	.8881	123	123	639	2	1.514	.0400	.1850	.1382	1.162
91	91	525	3	1.350	-.383	.1740	-.278	.9431	124	124	640	2	1.929	.7520	.1190	.8072	2.137
92	92	529	2	1.001	.7350	.2220	.9039	.4730	125	125	642	1	1.371	.5070	.1430	.5966	1.032
93	93	532	4	1.701	2.417	.1560	2.494	1.549	126	126	645	3	1.054	1.323	.1800	1.461	.5683
94	94	534	2	1.000	1.222	.2220	1.391	.4720	127	127	701	3	1.078	.3920	.2050	.5403	.5668
95	95	535	1	.8340	.6480	.2240	.8520	.3271	128	128	703	3	1.145	-.161	.2940	.0175	.5373
96	96	536	1	1.835	1.040	.2020	1.126	1.652	129	129	705	4	1.403	1.392	.1070	1.462	1.156
97	97	537	3	1.600	1.156	.1420	1.232	1.408	130	130	708	1	1.704	1.607	.1680	1.688	1.520
98	98	538	4	1.449	.6210	.1710	.7178	1.093	131	131	709	3	1.884	.9110	.1580	.9813	1.893
99	99	540	1	1.285	.4710	.1820	.5898	.8256	132	132	713	3	1.198	.1400	.2380	.2872	.6584

	id	prior	key	a	b	c	qmax	imax		id	prior	key	a	b	c	qmax	imax
133	133	716	1	2.150	2.189	.2050	2.263	2.254	166	166	237	2	1.121	-.399	.1450	-.288	6870
134	134	718	2	1.420	1.463	.1110	1.534	1.176	167	167	238	1	1.062	-.747	.1130	-.651	6551
135	135	721	2	1.133	.9000	.2890	1.078	.5314	168	168	239	2	.6350	-.367	.1460	-.171	2200
136	136	724	3	1.552	.9760	.2450	1.093	1.087	169	169	246	4	.8610	-.064	.0880	.0331	4515
137	137	725	2	.9490	-.517	.1850	-.360	4.564	170	170	301	1	.6150	-.152	.2470	-.122	.1701
138	138	729	4	1.266	1.539	.1990	1.663	.7907	171	171	310	3	.7660	-.327	.1600	-.153	.3118
139	139	730	2	1.283	1.142	.2290	1.277	.7666	172	172	316	3	.4950	-.223	.2150	-.190	.1172
140	140	731	4	2.070	1.576	.1450	1.636	2.343	173	173	317	4	.5870	-.196	.2800	.1412	.1452
141	141	732	4	1.045	.7840	.3770	.9930	.3771	174	174	334	1	.8750	-.918	.1390	-.781	4.234
142	142	735	3	1.064	1.363	.2670	1.543	.4896	175	175	339	4	.4390	-.595	.2340	-.196	.0889
143	143	736	1	.9850	.1730	.2090	.3376	.4696	176	176	414	3	.6110	-.106	.2330	-.774	.1725
144	144	737	1	1.035	.3440	.2040	.4980	.5235	177	177	426	1	.4620	-.514	.1560	-.230	.1143
145	145	738	4	1.169	.1580	.2390	.3102	.6242	178	178	431	3	.6300	-.571	.2330	-.293	.1834
146	146	739	1	1.166	1.068	.2120	1.208	.6542	179	179	432	1	.6950	-.757	.1770	-.550	.2485
147	147	740	2	1.668	.7140	.1770	.8003	1.431	180	180	435	1	.6990	-.133	.1670	.0643	.2562
148	148	745	1	1.208	1.084	.1170	1.171	.8415	181	181	438	4	.7230	-.873	.1350	-.710	.2912
149	149	747	4	1.494	2.144	.0940	2.203	1.344	182	182	439	1	1.206	-.306	.1050	-.226	.8578
150	150	748	3	1.581	1.654	.2560	1.772	1.104	183	183	448	4	.5110	-.131	.1960	-.100	.1296
151	151	749	1	1.455	1.520	.1120	1.590	1.232	184	184	504	1	.5910	-.986	.1850	-.735	.1770
152	152	124	2	1.082	-.030	.1970	.1137	.5798	185	185	513	3	.3740	-.101	.1740	-.867	.4909
153	153	125	1	1.185	-.300	.2100	-.163	.6783	186	186	518	1	4.700	-.310	.1920	.0147	.1105
154	154	131	4	.7070	-.529	.1800	-.323	.2557	187	187	539	3	.6170	-.951	.1540	-.741	.2046
155	155	133	4	.5110	.0130	.1880	.3072	.1316	188	188	542	4	.6360	-.913	.2420	-.631	.1837
156	156	141	4	.4860	-.931	.2440	-.560	.1068	189	189	609	3	.7880	-.142	.1880	-.123	.3129
157	157	143	3	.4060	1.699	.1930	2.076	.0823	190	190	610	3	.7230	-.167	.1860	-.146	.2644
158	158	201	3	.6080	-.157	.2130	-.120	.1775	191	191	632	4	.6620	-.060	.2560	.2216	.1936
159	159	209	1	.5980	-.106	.1660	-.842	.1879	192	192	638	1	.7020	-.568	.1880	-.355	.2492
160	160	210	4	.4550	-.125	.2120	-.306	.0996	193	193	641	4	.7680	-.109	.2060	.0999	.2871
161	161	217	4	.6220	-.736	.2610	-.432	.1693	194	194	710	2	1.026	-.538	.1870	-.392	.5314
162	162	225	4	.8630	.031	.1920	.1459	.3724	195	195	714	2	1.202	-.707	.1430	-.605	.7929
163	163	231	3	.6980	-.521	.1410	-.347	.2684	196	196	722	1	.6460	-.465	.2110	-.212	.2012
164	164	232	4	.5820	-.049	.1780	.1993	.1739	197	197	726	3	.5750	-.378	.2020	-.103	.1622
165	165	236	1	.7060	-.010	.1140	.1359	.2890	198	198	727	4	.5230	-.220	.2220	.1033	.1291
									199	199	741	2	1.029	-.143	.2060	-.128	.5154
									200	200	742	3	.8190	-.151	.1870	.0319	.3386

โปรแกรมคอมพิวเตอร์สำหรับดำเนินการสอบ

/*examinee	first item	ability estimate	next item	terminate
H	simple	BAY	b,Qmax,Imax	.30 .40 .50
M	modurate	MLE		
L	difficulty			*/

```
#include <stdio.h>
#include <math.h>
#include <graphics.h>
#include <dos.h>
#include <string.h>
#include <stdlib.h>
#include <mem.h>
#include "struct.h"
#include "define.h"

#define P1 0.2316419
#define C1 0.31938153
#define C2 -0.356563782
#define C3 1.78147937
#define C4 -1.821255978
#define C5 1.330274429

extern struct st_test test;
extern struct st_compute com;
extern unsigned char group;

int MARK=0,ans,W,X,Y,Z,ZZ,MF;
float ABIT=0,P,Q,IIF,TIF=0;
float ab4,ab5,see4,see5; /* abit & see >0.5, =0.5, =0.4*/
float SEE,sumP=0,sumPQ=0; /* MLE */
float ABIT0=0,VAR0=1; /* BAY */
float ma[200],mb[200],mc[200],Qmax[200];
int U[200];
char yes[202],id[5];
int ch_see=0,ch_see_o=100;
/*----- INTERFACE -----*/
int cat(int c);
void title(int c);
void by();
void first_item(int c);
void answer(int c);
void para(int f),MLE(),BAY();
void select(int c);
int terminate();
/*----- MAIN FOUNTION CAT -----*/
int cat(int c)
{
    answer(c);
    if((group%2)==0 && (MF!=ZZ && MF!=0)) MLE(); else BAY();
    select(group);
    return(terminate());
}
```

```

/*----- Title -----*/
void title(int c)
{
    int cc[] = {0,0,0,0,0,0, 1,1,1,1,1,1, 2,2,2,2,2,2};
    int bq[] = {0,0,1,1,2,2, 0,0,1,1,2,2, 0,0,1,1,2,2};
    int mb[] = {0,1,0,1,0,1, 0,1,0,1,0,1, 0,1,0,1,0,1};
    int l;
    savescreen(199,99,461,301); message(200,100,460,300,2);
    color(0xF8); grattr(1); grprint(13,110,"ສ້າງກາລາສັກກອນ ກອນພາ"); grattr(0);
    l = cc[c];
    if(l==0) grprint(32,140," ລາຮຸກ ພື້ນ ກອນ ຈັດ ");
    else if(l==1) grprint(32,140," ລາຮຸກ ພື້ນ ກອນ ໂຄຟໂປດສໂຕ ");
    else grprint(32,140," ລາຮຸກ ພື້ນ ກອນ ໂຕ ");
    l = bq[c];
    if(l==0) grprint(32,170," ສັກກອນ ກອນພາ ສັກກອນ ຈັດ ສັກກອນ ");
    else if(l==1) grprint(32,170," ສັກກອນ ກອນພາ ສັກກອນ ຈັດ Qmax ");
    else grprint(32,170," ສັກກອນ ກອນພາ ສັກກອນ ຈັດ Imax ");
    l = mb[c];
    if(l==0) grprint(32,200," ສັກກອນ ພື້ນ ກອນ ສ້າງກາລາສັກກອນ MLE ");
    else grprint(32,200," ສັກກອນ ພື້ນ ກອນ ສ້າງກາລາສັກກອນ BAY ");
    wait();
    restorescreen(199,99,461,301);
    setfillstyle(1,11); bar(12,352,343,438);
}

/*----- Good By -----*/
void by( )
{
    savescreen(199,99,461,301); message(200,100,460,300,1);
    color(0xF5); grattr(1); grprint(13,180," ສັກກອນ ພື້ນ ກອນ ຈັດ "); grattr(0);
    wait();
    restorescreen(199,99,461,301);
}

/*----- FIRST ITEM-----*/
void first_item(int c)
{
    int cc[] = {0,0,0,0,0,0, 1,1,1,1,1,1, 2,2,2,2,2,2};
    setmem(yes,202,0);
    MARK = 0; ABIT=0; TIF=0;
    sumP=0; sumPQ=0;
    ABIT0=0; VAR0=1;
    W = X = Y = Z = MF = ZZ = 0;
    switch(cc[c])
    {
        case 0: selec_b(-1.30); te_seach(com.id); te_display(); break;
        case 1: selec_b(0); te_seach(com.id); te_display(); break;
        case 2: selec_b(1.30); te_seach(com.id); te_display(); break;
    }
}

```

```

/*-----ANSWER-----*/
void answer(int c)
/*correct=1,incorect=0*/
{
    color(0xEB); ZZ++;
    strcpy(id,com.id);
    if(test.key == '1'+c)
    {
        ans = 1; MARK++; MF++;
        grprint(4,360,"MARK = %4d",MARK);
    } else ans = 0;
    ma[Z] = com.a; mb[Z] = com.b; mc[Z] = com.c; U[Z] = ans;
}

/*-----PARAMETER-----*/
void para(int f)
/* P,Q,IIF,TIF,SEE*/
{
    P=com.c+((1.0-com.c)*((float)exp(1.7*com.a*(ABIT-com.b))/(1.0+(float)exp(1.7
*com.a*(ABIT-com.b)))));
    Q=1.0-P;
    IIF=(float)pow((1.7*com.a*Q*(P-com.c))/(1.0-com.c),2)/(P*Q);
    TIF+=IIF; /*---sum(1.7*1.7*a*a*Q(P-c)(P-c))/((1-c)(1-c)P)---*/
    if(!f) SEE=(float)sqrt(VARO);
    else
        if((float)sqrt(TIF)!=0) SEE = 1.0/(float)sqrt(TIF); else SEE = 0.0;
}

/*-----MAXIMUM LIKELIHOOD-----*/
void MLE()
{
    float hm=0,A,B,PC;
    int c,i; c = 0;

    ma[Z] = com.a; mb[Z] = com.b; mc[Z] = com.c; U[Z] = ans; Qmax[Z] = com.Qmax;
    ABIT = (float)(log((float)MARK/(float)(ZZ-MARK))/log((float)2.7182818));
    /*-----
    color(0x1F);
    grprint(10,50,"abit=%5.4f ZZ=%5d mark=%5d",ABIT,ZZ,MARK);
    getch();
    -----*/
    do {
        A=0; B=0;
        for(i=0;i<=Z;i++)
        {
            P=mc[i]+((1.0-mc[i])*((float)exp(1.7*ma[i]*(ABIT-mb[i]))/(1.0+(float)e
xp(1.7*ma[i]*(ABIT-mb[i])))));
            if(P==1.000) P=0.999;
            Q=1.0-P;
            if(P == mc[i]) PC = 0.01; else PC = P-mc[i];
            A += (ma[i] *((float)U[i] - P)*PC) / (P*(1.0-mc[i]));
            B += ((ma[i]*ma[i])*PC*((float)U[i]*mc[i])-(P*P)*Q) / ((P*P)*((1.0-mc[
i])*(1.0-mc[i])));
        }
    } while(1);
}

```

```

/*-----*/
grprint(4,2,"A=%-10.4f B=%-10.4f U=%2d P=%-10.4f ABIT=%-10.4f loop=%5
d",A,B,U[i],P,ABIT,c);
grprint(4,20,"a=%-10.4f b=%-10.4f c=%-10.4f Qmax=%-10.4f",ma[i],mb[i]
,mc[i],Qmax[i]);
delay(1000);
/*-----*/
}
hm = A/(1.7*B);
/*-----*/
grprint(65,20,"hm=%-10.4f",fabs(hm));
/*-----*/
ABIT = ABIT - hm;
c++;
} while(fabs(hm) >= 0.001 && c < 100);
para(1);
}

/*-----BAYESIAN-----*/
void BAY()
{
float D,A,t,tD,ord,area,areaD;
para(0);
D=(com.b-ABIT)/(float)sqrt((1.0/(float)pow(com.a,2)+VARO));
t=(float)(1/(1+(P1*D)));
tD=(float)(1/(1-(P1*D)));
ord=0.3989/(float)pow(2.7183,0.5*D*D);
area=1.0-(ord*((C1*t)+(C2*t*t)+(C3*t*t*t)+(C4*t*t*t*t)+(C5*t*t*t*t*t)));
areaD=1.0-(ord*((C1*tD)+(C2*tD*tD)+(C3*tD*tD*tD)+(C4*tD*tD*tD*tD)+(C5*tD*tD*tD
*tD*tD)));
A=com.c+(1.0-com.c)*areaD;
if(ans)
{
ABITO=ABITO+(((1.0-com.c)*((VARO*ord)/(float)(sqrt((1.0/(com.a*com.a))+VARO
)*
(com.c+(1.0-com.c)*areaD)))));
VARO=VARO*(1.0-(((1.0-com.c)/(1.0+(1.0/(com.a*com.a*VARO)))))*
(ord/A)*(((1.0-com.c)*ord)/A)-D));
ABIT=ABITO;
}
else
{
ABITO=ABITO-((VARO/(float)sqrt((1.0/(com.a*com.a))+VARO))*(ord/area));
VARO=VARO*(1.0-((ord/(1.0+(1.0/((com.a*com.a)*VARO)))))*((ord/area)+D)/area)
);
ABIT=ABITO;
}
para(0);
/*-----*/
color(0x1F);
grprint(10,50,"abit=%5.4f VARO=%5.4f SEE=%5.4f",ABIT,VARO,SEE);
getch();
/*-----*/
}

```

```

/*-----generate next item by abit=b,abit=Qmax,Imax-----*/
void select(int c)
{
    int cc[] = {0,0,1,1,2,2, 0,0,1,1,2,2, 0,0,1,1,2,2 };
    float Ps,Qs,Imax=-50,Iif;
    char str[10];
    switch(cc[c])
    {
        case 0: /*abit=b*/
            selec_b(ABIT); te_seach(com.id); te_display(); break;
        case 1: /*abit=Qmax*/
            selec_q(ABIT); te_seach(com.id); te_display(); break;
        case 2: /*Imax*/
            co_getbof();
            do {
                Ps=com.c+((1.0-com.c)*((float)exp(1.7*com.a*(ABIT-com.b))/(1.0+(float)exp(1.7*com.a*(ABIT-com.b)))));
                Qs=1.0-Ps;
                Iif=(float)pow((1.7*com.a*Qs*(Ps-com.c))/(1.0-com.c),2)/(Ps*Qs);
                if(Imax<Iif && !yes[atoi(com.id)]) { Imax=Iif; strcpy(str,com.id); }
            } while(!co_getnext());
            yes[atoi(str)] = 1; co_seach(str); te_seach(str); te_display(); break;
    }
}

/*-----terminate-----*/
int terminate()
{
    Z++;
    if(SEE<=0.30)
    {
        save_ID(id,1);
        W=Z; save_info(0);
        color(0xEB);
        grprint(4,380,"ABIT = %2.4f",ABIT);
        grprint(4,400,"SEE = %2.4f",SEE);
        return(1);
    }
    else if(SEE<=0.40) { ch_see = 2; if(ch_see != ch_see_o){ch_see_o = ch_see; save_ID(id,2); ab4=ABIT; see4=SEE; X=Z;} else save_ID(id,0); }
    else if(SEE<=0.50) { ch_see = 3; if(ch_see != ch_see_o){ch_see_o = ch_see; save_ID(id,3); ab5=ABIT; see5=SEE; Y=Z;} else save_ID(id,0); }
    else { ch_see = 4; if(ch_see != ch_see_o){ch_see_o = ch_see; save_ID(id,4);} else save_ID(id,0); }
    color(0xEB);
    grprint(4,380,"ABIT = %2.4f",ABIT);
    grprint(4,400,"SEE = %2.4f",SEE);
    return(0);
}

```

หน้าจอคอมพิวเตอร์เมื่อต้องการป้อนข้อสอบเข้าคลังข้อสอบ

MODE TEST

เพิ่มข้อมูล

ค้นหาข้อมูล

ลบข้อมูล

ดูข้อมูลจริง

กลับ MENU หลัก

เลขข้อ.....

โจทย์

.....

.....

คำตอบ ก.....

ข.....

ค.....

ง.....

คำตอบถูก.....

F3 แก้ไข

F2 SAVE

หน้าจอคอมพิวเตอร์เมื่อต้องการป้อนค่าพารามิเตอร์ของข้อสอบ

MODE COMPUTE

เพิ่มข้อมูล

ค้นหาข้อมูล

ลบข้อมูล

ดูข้อมูลจริง

กลับ MENU หลัก

เลขข้อ

ค่าอำนาจจำแนก

ค่าความยากง่าย

ค่าการเดา

ค่า θ_{max}

F3 แก้ไข

F2 SAVE

หน้าจอคอมพิวเตอร์เมื่อเริ่มต้นทดสอบ

COMPUTERIZED ADAPTIVE TESTING

รหัส

ชื่อ

โรงเรียน

F2 SAVE

Esc CANCEL

ALT X EXIT

F10 SETUP

ตัวอย่างการแสดงผลข้อสอบหน้าจอคอมพิวเตอร์

COMPUTERIZED ADAPTIVE TESTING

ข้อ 00

ถ้า $\frac{A}{B}$ เป็นเศษส่วนแท้ ข้อใดถูกต้อง

ก. $A < B$

ข. $B < A$

ค. A อาจเท่ากับ B

ง. ข้อ ก และ ค ถูกต้อง

เลือกคำตอบที่ถูกต้อง

ก ข ค ง

Esc CANCEL

ALT X EXIT

F10 SETUP

รายชื่อผู้เชี่ยวชาญด้านหลักสูตร การสอน การวัดและประเมินผลคณิตศาสตร์

- | | |
|---------------------------------|--|
| 1. ว่าที่ร้อยตรีสุวรร กาญจนมยุร | สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี |
| 2. ดร.ปานทอง กุลนาถศิริ | สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี |
| 3. นายประสาท สอ้านวงศ์ | ศูนย์พัฒนาหลักสูตร กรมวิชาการ กระทรวงศึกษาธิการ |
| 4. รศ.ดร.ดวงเดือน อ่อนน้อม | คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย |
| 5. รศ.ดร.ปิยวดี วงษ์ใหญ่ | คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ |
| 6. ผศ.สุภาพร ศรีบุญรินทร์ | คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ |
| 7. รศ.ปรีชา เนาว์เย็นผล | คณะศึกษาศาสตร์ มหาวิทยาลัยสุโขทัยธรรมาธิราช |
| 8. นางบุญทวี แสงวงศ์ | โรงเรียนสายน้ำทิพย์ สำนักงานการประถมศึกษา
กรุงเทพมหานคร |
| 9. นายนิรันดร์ ตันตัยย์ | โรงเรียนวัดหงส์รัตนาราม สำนักงานการประถมศึกษา
กรุงเทพมหานคร |

ประวัติย่อของผู้วิจัย

ชื่อ	นายรังสรรค์ มณีเล็ก
วันเดือนปีเกิด	20 เมษายน 2500
สถานที่เกิด	ต.บ้านสร้าง อ.บ้านสร้าง จ.ปราจีนบุรี
สถานที่อยู่ปัจจุบัน	71/189 หมู่ 8 ต.บางเมืองใหม่ อ.เมือง จ.สมุทรปราการ 12700
ตำแหน่งหน้าที่ปัจจุบัน	นักวิชาการศึกษา สำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ กระทรวงศึกษาธิการ

ประวัติการศึกษา

พ.ศ.2522	กศ.บ.(ฟิสิกส์-คณิตศาสตร์) มหาวิทยาลัยศรีนครินทรวิโรฒ บางแสน
พ.ศ.2528	ค.ม.(การวัดและประเมินผลการศึกษา) จุฬาลงกรณ์มหาวิทยาลัย
พ.ศ.2540	กศ.ด.(การทดสอบและวัดผลการศึกษา) มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร

ประวัติการทำงาน

พ.ศ.2518	ครูโรงเรียนวัดวังซอน อ.บ้านสร้าง จ.ปราจีนบุรี
พ.ศ.2525	ศึกษานิเทศก์สำนักงานการประถมศึกษาอำเภอเมืองปราจีนบุรี
พ.ศ.2527	ศึกษานิเทศก์สำนักงานการประถมศึกษาจังหวัดปราจีนบุรี
พ.ศ.2530	นักวิชาการศึกษา สำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ กระทรวงศึกษาธิการ

ผลของตัวแปรบางตัวต่อความเที่ยงตรงเชิงสภาพและจำนวนข้อสอบที่ใช้ใน
การทดสอบแบบปรับเหมาะจะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์

บทคัดย่อ
ของ
รังสรรค์ มณีเล็ก

เสนอต่อมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษาดุชฎิบัณฑิต สาขาการทดสอบและวัดผลการศึกษา
มีนาคม 2540

การวิจัยครั้งนี้มีวัตถุประสงค์ที่จะศึกษาผลของความยากง่ายของข้อสอบข้อแรกที่ใช้ในการทดสอบ เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบเกณฑ์ยุติการทดสอบและความสามารถของผู้สอบต่อความเที่ยงตรงเชิงสภาพและจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์

การวิจัยแบ่งออกเป็นสองระยะ ระยะแรก ดำเนินการสร้างคลังข้อสอบเรื่อง เศษส่วน ชั้นประถมศึกษาปีที่ 6 กลุ่มตัวอย่างที่ใช้คือ นักเรียนชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2537 จำนวน 8,720 คน วิเคราะห์คุณภาพของข้อสอบตามทฤษฎีการตอบสนองข้อสอบ ได้ข้อสอบทั้งหมด 200 ข้อ มีค่าอำนาจจำแนกอยู่ในช่วง 0.406 ถึง 2.338 ค่าความยากง่ายอยู่ในช่วง -2.233 ถึง 2.720 ค่าการเดาอยู่ในช่วง 0.061 ถึง 0.30 ระยะที่สองเป็นการศึกษาผลของตัวแปรอิสระ 5 ตัว ดังกล่าวต่อความเที่ยงตรงเชิงสภาพและจำนวนข้อสอบที่ใช้ในการทดสอบ กลุ่มตัวอย่างคือ นักเรียนชั้นมัธยมศึกษาปีที่ 1 ที่เรียนเป็นเทอมต้น ในกรุงเทพมหานคร ปีการศึกษา 2539 จำนวน 1,620 คน เครื่องมือที่ใช้ในการวิจัยคือ คลังข้อสอบ เรื่อง เศษส่วนแบบทดสอบวัดความสามารถทางคณิตศาสตร์และโปรแกรมคอมพิวเตอร์สำหรับการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์วิเคราะห์ข้อมูลโดยใช้ค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐาน ค่าสหสัมพันธ์ ค่าไคสแควร์และการวิเคราะห์ความแปรปรวน

ผลการวิจัยพบว่า วิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบและความสามารถของผู้สอบมีผลต่อค่าความเที่ยงตรงเชิงสภาพของการทดสอบ ส่วนผลต่อตัวแปรตามอีกตัวแปรหนึ่งนั้น, ความยากง่ายของข้อสอบข้อแรก, เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบและความสามารถของผู้สอบต่างก็มีผลต่อจำนวนข้อสอบที่ใช้ในการทดสอบ

THE EFFECT OF SOME VARIABLES ON CONCURRENT VALIDITY AND
ITEM NUMBER OF COMPUTERIZED ADAPTIVE TESTING

ABSTRACT

BY

RANGSAN MANEELEK

Presented in partial fulfillment of the requirements for the Doctor
of Education Degree in Testing and Measurement
at Srinakharinwirot University

March 1997

The purpose of this study was to investigate the effect of the difficulty of the first entry item in testing, item selection criteria, ability estimation procedure, termination criteria, ability of examinee on concurrent validity and number of items of computerized adaptive testing.

The process of this study has two phases. The first phase was the creation of item bank on fraction. The samples of this phase were consisted of 8,720 students from prathomsuksa 6 of 1994 academic year, analyzing the qualities of items by using item response theory. A total of 200 items, having the discrimination between 0.406 and 2.338, the difficulty of items were between -2.233 and 2.720, guessing value were between 0.061 and 0.30. The second phase was the study of the effect of five independent variables on concurrent validity and number of test items. The samples were 1,620 students of M.1 at the beginning of the first semester of 1996 academic year in Bangkok Metropolitan. The instruments used in this analysis were item pools on fraction, mathematics comprehension test and computer program for computerized adaptive testing. Data were analysed by using means, standard deviation, correlation coefficient, chi-square and analysis of variance

The finding of this study was that concurrent validity was affected by ability estimation procedure, termination criteria and ability of examinee whereas effects on another dependent variable were the difficulty of the first entry item, item selection criteria, abilities of examinees estimate, termination criteria and ability of examinee all affected the number of items in the testing