

371.271

๑๒๘๒๗

๘.๓

การศึกษาอิทธิพลของการให้คะแนนและการคัดเลือกข้อสอบที่มีต่อคุณภาพ
แบบทดสอบเลือกตอบ

ปริญญานิพนธ์

ของ

ชาลิต โพธิ์นคร

25 88.8. 2540

เสนอต่อมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษาดุขุฎิบัณฑิต สาขาการทดสอบและวัดผลการศึกษา

มีนาคม 2540

ลิขสิทธิ์เป็นของมหาวิทยาลัยศรีนครินทรวิโรฒ

h 35948

63970

คณะกรรมการควบคุมและคณะกรรมการสอบได้พิจารณาปริญญานิพนธ์ฉบับนี้แล้ว
เห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาการศึกษาดุฎิบัณฑิต
สาขาการทดสอบและวัดผลการศึกษา ของมหาวิทยาลัยศรีนครินทรวิโรฒได้

คณะกรรมการควบคุม

..... ศ.ดร. วาสนา ประธาน
(รศ.ดร.ส วาสนา ประवालพฤษ) กรรมการ
(รศ.ดร.สมบูรณ์ ชิตพงศ์) กรรมการ
(ดร.ชูศักดิ์ ชัมภลลิขิต)

คณะกรรมการสอบ

..... ศ.ดร. วาสนา ประธาน
(รศ.ดร.ส วาสนา ประवालพฤษ) กรรมการ
(รศ.ดร.สมบูรณ์ ชิตพงศ์) กรรมการ
(ดร.ชูศักดิ์ ชัมภลลิขิต) กรรมการที่แต่งตั้งเพิ่มเติม
(รศ.ดร.ศิริชัย กาญจนวาสี)

บัณฑิตวิทยาลัยอนุมัติให้ได้รับปริญญานิพนธ์ฉบับนี้ เป็นส่วนหนึ่งของการศึกษาตาม
หลักสูตรปริญญาการศึกษาดุฎิบัณฑิต สาขาการทดสอบและวัดผลการศึกษา ของมหาวิทยาลัย
ศรีนครินทรวิโรฒ

..... ศ.ดร. วาสนา คณบดีบัณฑิตวิทยาลัย
(ดร.ศิริภา พูลสุวรรณ)

วันที่ 14 เดือน มิถุนายน พ.ศ. 2540

ปริญญานิพนธ์นี้ได้รับพระราชทานทุน “ ภูมิพล “
ผู้วิจัยรู้สึกสำนึกในพระมหากรุณาธิคุณเป็นล้นเกล้าล้นกระหม่อมหาที่สุดมิได้

ชวลิต โพธิ์นคร

ประกาศคุณูปการ

ปริญญานิพนธ์ฉบับนี้สำเร็จสมบูรณ์ได้เพราะได้รับความกรุณาจาก รองศาสตราจารย์ ดร.ส วาสนา ประवालพฤษ์ รองศาสตราจารย์ ดร.สมบูรณ์ ชิตพงศ์ ดร.ชูศักดิ์ ชัมภลขิต รองศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี ที่ให้คำปรึกษาแนะนำ ตลอดจนแก้ไขข้อบกพร่องต่างๆ จนปริญญานิพนธ์สำเร็จสมบูรณ์ ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

ขอขอบพระคุณ ศาสตราจารย์ ดร.สำเริง บุญเรืองรัตน์ ดร.วัน สังข์สะอาด รองศาสตราจารย์ ดร.บุญเชิด ภิญโญนันต์พงษ์ และคณะกรรมการบริหารหลักสูตรที่ได้กรุณาให้ คำปรึกษาแนะนำ ตลอดจนแก้ไขข้อบกพร่องต่างๆ ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

ขอโน้มรำลึกถึงพระคุณของ ศาสตราจารย์ ดร.อนันต์ ศรีโสภา ที่ได้ให้แนวคิด และคำปรึกษาซึ่งนับว่าเป็นจุดเริ่มต้นของปริญญานิพนธ์ไว้ ณ โอกาสนี้

ขอขอบพระคุณ สำนักทดสอบทางการศึกษาและจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร ที่ให้ความอนุเคราะห์โปรแกรม BIGSTEPS สำหรับใช้ในการวิเคราะห์ข้อมูล

ขอขอบพระคุณ ดร.บัญชา แสงทวิ คุณรังสรรค์ มณีเล็ก ดร.สุรัช มีชาญ ที่กรุณาให้คำแนะนำ ปรึกษาเกี่ยวกับการวิเคราะห์ข้อมูลโดยใช้โปรแกรมคอมพิวเตอร์ SPSSPC และ BIGSTEPS คุณปริษา อรุณสวัสดิ์ และคุณธรรมรัตน์ รัชมณต์ ที่ให้ความเอื้อเฟื้อเกี่ยวกับแบบทดสอบสำหรับใช้เป็นเครื่องมือในการวิจัย ตลอดจนคณะกรรมการผู้เชี่ยวชาญทุกท่านที่มีส่วนร่วมในการสร้างแบบทดสอบ

ขอขอบพระคุณ คุณจิตสุตา นราธร ที่ให้ความกรุณาช่วยเหลือในการตรวจให้คะแนน และบันทึกคะแนน ศึกษานิเทศก์จังหวัด เพื่อนๆ รุ่นพี่ รุ่นน้องวัดผล ผู้บริหารโรงเรียน คณะครู นักเรียน และผู้ที่มีส่วนเกี่ยวข้องในการเก็บรวบรวมข้อมูลทุกคน ที่ได้ให้ความช่วยเหลืออย่างดียิ่ง จนงานบรรลุผลสำเร็จตามเป้าหมาย

ขอขอบพระคุณ ดร.ฉันทยา วงศ์ศิลปกรมย์ ผู้บังคับบัญชาตามสายงาน และผู้ร่วมงาน ในหน่วยงานทุกท่านที่ให้ความสนใจและมีส่วนช่วยเหลือในการทำปริญญานิพนธ์ฉบับนี้

สุดท้ายนี้ ขอโน้มรำลึกถึงพระคุณของบิดามารดา ครูบาอาจารย์ ผู้มีพระคุณตลอดจน ภรรยาและญาติพี่น้อง ที่ให้กำลังใจและสนับสนุนการศึกษาด้วยดีตลอดมา

ชวลิต โพธิ์นคร

สารบัญ

บทที่		หน้า
1	บทนำ	1
	ภูมิหลัง	1
	วัตถุประสงค์ของการวิจัย	16
	สมมติฐานการวิจัย	17
	ขอบเขตและความจำกัดของการวิจัย	18
	ความสำคัญของการวิจัย	19
	คำนิยามศัพท์เฉพาะ	20
2	เอกสารและงานวิจัยที่เกี่ยวข้อง	22
	มโนทัศน์เกี่ยวกับทฤษฎีพื้นฐานในการพัฒนาแบบทดสอบเลือกตอบ	22
	ทฤษฎีมาตรฐานเดิม (CTT)	22
	แบบจำลองคะแนนจริง (True score model)	24
	คุณภาพแบบทดสอบ (Test quality)	26
	ความเที่ยงตรงของแบบทดสอบ (Test validity)	27
	ความเชื่อมั่นของแบบทดสอบ (Test validity)	43
	คุณภาพแบบทดสอบเป็นรายข้อ	50
	การประมาณค่าความยากของแบบทดสอบ	52
	การประมาณค่าอำนาจจำแนกของข้อสอบ	54
	การวิเคราะห์ประสิทธิภาพตัวลวง (Distractor power analysis)	59
	ปฏิสัมพันธ์ระหว่างคุณสมบัตินี้ของข้อสอบต่างๆ	60
	การคัดเลือกข้อสอบ (Item selecting)	61
	ทฤษฎีการสรุปอ้างอิง (G-theory)	62
	กรอบความคิดของ GT	64
	โมเดลทางสถิติ ของ G-theories	66

การวิเคราะห์สัมประสิทธิ์ GT	68
ทฤษฎีการตอบสนองข้อสอบ	72
ลักษณะทั่วไปของทฤษฎีการตอบสนองข้อสอบ	73
ข้อตกลงเบื้องต้นเกี่ยวกับทฤษฎีการตอบสนองข้อสอบ	74
โมเดลการตอบสนองข้อสอบ	75
ฟังก์ชันสารสนเทศ (Information function)	96
ประสิทธิภาพสัมพัทธ์ (Relative efficiency)	100
ความสัมพันธ์ระหว่าง IRT กับ CTT	101
มโนทัศน์เกี่ยวกับกระบวนการพัฒนาแบบทดสอบเลือกตอบ	104
ข้อได้เปรียบและข้อจำกัดของแบบทดสอบเลือกตอบ	105
กระบวนการพัฒนาแบบทดสอบเลือกตอบ	105
การวิเคราะห์ข้อสอบเพื่อการพัฒนาแบบทดสอบ	107
กฎเกณฑ์ในการเขียนข้อสอบเลือกตอบ	115
หลักเกณฑ์ในการประเมินคุณภาพของแบบทดสอบเลือกตอบ	117
มโนทัศน์เกี่ยวกับการพัฒนาโครงสร้างของแบบทดสอบเลือกตอบ	
วิธีการตอบและการตรวจให้คะแนน	118
มโนทัศน์เกี่ยวกับการศึกษาความลำเอียงของข้อสอบ	123
งานวิจัยที่เกี่ยวข้องกับการพัฒนาแบบทดสอบเลือกตอบ	126
 3. วิธีดำเนินการวิจัย	152
ขอบเขตของการวิจัย	152
ประชากรและกลุ่มตัวอย่าง	152
ตัวแปรในการวิจัย	156
เครื่องมือการวิจัย	157
การวิเคราะห์และจัดกระทำข้อมูล	159

บทที่	หน้า
4 ผลการวิเคราะห์ข้อมูล	173
สัญลักษณ์และอักษรย่อที่ใช้ในการนำเสนอผลการวิเคราะห์ข้อมูล	173
การนำเสนอผลการวิเคราะห์ข้อมูล	176
การศึกษาอิทธิพลของการให้คะแนน	177
การศึกษาอิทธิพลของวิธีการคัดเลือกข้อสอบ	184
การศึกษารูปแบบการให้คะแนนและวิธีการคัดเลือกข้อสอบที่ส่งผล ต่อคุณภาพของแบบทดสอบ	196
การศึกษารูปแบบความสัมพันธ์เชิงสาเหตุของรูปแบบการให้คะแนน และวิธีการคัดเลือกข้อสอบ ที่มีต่อคุณภาพแบบทดสอบ	215
5 สรุป อภิปรายผล และข้อเสนอแนะ	291
วัตถุประสงค์ของการวิจัย	291
ขอบเขตของการวิจัย	291
วิธีดำเนินการวิจัย	292
สรุปผลการวิจัย	294
อภิปรายผล	301
ข้อเสนอแนะ	310
บรรณานุกรม	311
ภาคผนวก	325
ภาคผนวก ก. คณะกรรมการสร้าง และตรวจสอบความเที่ยงตรงเชิงสภาพและ เชิงโครงสร้างของแบบทดสอบ	326
ภาคผนวก ข. เครื่องมือที่ใช้ในการวิจัย	329
ภาคผนวก ค. ค่าสถิติ และคุณสมบัติของเครื่องมือที่ใช้ในการวิจัย	343
ภาคผนวก ง. การพัฒนาสูตรการให้คะแนนแบบ 2-SR	356
ประวัติย่อของผู้วิจัย	
บทคัดย่อ	

บัญชีตาราง

ตาราง	หน้า
1 สรุปการประมาณค่าความเชื่อมั่นวิธีต่างๆ	47
2 รูปแบบ(design) และแหล่งแปรปรวนที่ได้ตามทฤษฎี G-theories	69
3 แหล่งความแปรปรวนตามทฤษฎี G-theories	71
4 ลักษณะของข้อมูล โมเดลที่ใช้วิเคราะห์ และผู้นำเสนอโมเดล	76
5 การเปรียบเทียบความแตกต่างระหว่างโมเดลเรขาคณิตกับโมเดล เกรดเรสพอน	87
6 ความสัมพันธ์ระหว่าง a_g และ ρ_g	102
7 ความสัมพันธ์ระหว่างคุณลักษณะแฝง และคะแนนจริง	103
8 จำนวนนักเรียนกลุ่มตัวอย่าง จำแนกตามเพศ โรงเรียน อำเภอ จังหวัด และ เขตการศึกษา	155
9 คะแนนเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน คะแนนสูงสุด และต่ำสุดของ คะแนนสอบวัดความสามารถในการอ่านจับใจความ และการแก้โจทย์ ปัญหาทางคณิตศาสตร์ซึ่งให้คะแนนตามรูปแบบปกติ (x1) และ รูปแบบ 2 - SR (x2 และ x3)	177
10 ความสามารถเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ระดับความสามารถสูงสุด และต่ำสุดในการตอบแบบทดสอบวัดความสามารถในการอ่านจับใจความ และแก้โจทย์ปัญหาทางคณิตศาสตร์ ที่ให้คะแนนตามรูปแบบปกติ (A1) และ รูปแบบ 2- SR (A2)	178
11 ค่าความเชื่อมั่น ความคลาดเคลื่อนมาตรฐานในการวัด ของแบบทดสอบ วัดความสามารถในการอ่านจับใจความและการแก้โจทย์ปัญหา คณิตศาสตร์ซึ่งให้คะแนนตามรูปแบบปกติ และรูปแบบ 2 -SR	179
12 ค่าสารสนเทศของแบบทดสอบ ความคลาดเคลื่อนมาตรฐานในการประมาณ ค่าและประสิทธิภาพสัมพัทธ์ของแบบทดสอบวัดความสามารถในการอ่าน จับใจความและการแก้โจทย์ปัญหาคณิตศาสตร์ซึ่งให้คะแนนตามรูปแบบปกติ และรูปแบบ 2-SR	180

13	ค่าเฉลี่ยความยากแบบ CTT และ IRT สัมประสิทธิ์การจำแนก ดัชนีการ จำแนก ค่าความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ ค่าความเป็นมิติเดียว ค่าความเชื่อมั่นรายข้อ ค่าความเหมาะสมของข้อสอบกับโมเดล และ ค่าสารสนเทศของข้อสอบของแบบทดสอบวัดความสามารถในการอ่านจับ ใจความและการแก้โจทย์ปัญหาทางคณิตศาสตร์ซึ่งให้คะแนนตามรูปแบบปกติ และรูปแบบ 2 - SR	183
14	ค่าความเชื่อมั่นของแบบทดสอบวัดความสามารถในการอ่านจับใจความ ที่ให้คะแนนตามรูปแบบปกติ และรูปแบบ 2-SR เมื่อคัดเลือกข้อสอบเข้าฉบับ โดยตัวแปรต่างๆ และค่าความเชื่อมั่นของแบบทดสอบเมื่อปรับขยาย จำนวนข้อสอบเต็มฉบับ (40ข้อ)	185
15	ค่าความเชื่อมั่นของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทาง คณิตศาสตร์ ที่ให้คะแนนตามรูปแบบปกติ และรูปแบบ 2-SR เมื่อคัดเลือก ข้อสอบเข้าฉบับโดยตัวแปรต่างๆ และค่าความเชื่อมั่นของแบบทดสอบ เมื่อปรับขยายจำนวนข้อสอบเต็มฉบับ(40ข้อ)	187
16	ค่าสารสนเทศของข้อสอบที่ ระดับความสามารถสูงของแบบทดสอบวัดความสามารถ ในการอ่านจับใจความ ที่ให้คะแนนตามรูปแบบปกติและรูปแบบ 2-SR จำแนก ตามตัวแปรต่างๆที่ใช้ในการคัดเลือกข้อสอบเข้าฉบับตามกรอบทฤษฎี CTT และ IRT	189
17	ค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลางของแบบทดสอบวัด ความสามารภในการอ่านจับใจความ ที่ให้คะแนนตามรูปแบบปกติ และ รูปแบบ 2-SR จำแนกตามตัวแปรต่างๆ ที่ใช้ในการคัดเลือกข้อสอบเข้าฉบับตาม กรอบทฤษฎี CTT และ IRT	190
18	ค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำของแบบทดสอบวัดความ สามารถในการอ่านจับใจความ ที่ให้คะแนนตามรูปแบบปกติ และรูปแบบ 2-SR จำแนกตามตัวแปรต่างๆที่ใช้ในการคัดเลือกข้อสอบเข้าฉบับ ตามกรอบทฤษฎี CTT และ IRT	191

19	ค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูงของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนตามรูปแบบปกติ และรูปแบบ 2- SR จำแนกตามตัวแปรต่างๆที่ใช้ในการคัดเลือกข้อสอบเข้าฉบับตามกรอบทฤษฎี CTT และ IRT	192
20	ค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลางของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนตามรูปแบบปกติ และรูปแบบ 2-SR จำแนกตามตัวแปรต่างๆที่ใช้ในการคัดเลือกข้อสอบเข้าฉบับตามกรอบทฤษฎี CTT และIRT.....	194
21	ค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนตามรูปแบบปกติ และรูปแบบ 2-SR จำแนกตามตัวแปรต่างๆที่ใช้ในการคัดเลือกข้อสอบเข้าฉบับตามกรอบทฤษฎี CTT และ IRT	195
22	ค่าสหสัมพันธ์ระหว่างตัวแปรที่ใช้ในการคัดเลือกข้อสอบเข้าฉบับ ความเชื่อมั่นรายข้อ สารสนเทศของข้อสอบของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนตามรูปแบบปกติ	197
23	ค่าสหสัมพันธ์ระหว่างตัวแปรที่ใช้ในการคัดเลือกข้อสอบเข้าฉบับ ความเชื่อมั่นรายข้อ สารสนเทศของข้อสอบของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนตามรูปแบบ 2 - SR	199
24	ค่าสหสัมพันธ์ระหว่างตัวแปรที่ใช้ในการคัดเลือกข้อสอบเข้าฉบับ ความเชื่อมั่นรายข้อ สารสนเทศของข้อสอบของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนตามรูปแบบปกติ	201
25	ค่าสหสัมพันธ์ระหว่างตัวแปรที่ใช้ในการคัดเลือกข้อสอบเข้าฉบับ ความเชื่อมั่นรายข้อ สารสนเทศของข้อสอบของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนตามรูปแบบ 2 - SR	203
26	สมการพยากรณ์ สหสัมพันธ์พหุคูณ สัมประสิทธิ์การพยากรณ์ ประสิทธิภาพในการพยากรณ์ ความคลาดเคลื่อนมาตรฐานในการประมาณค่า และรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนของสมการพยากรณ์ค่าความเชื่อมั่นของแบบทดสอบวัดความสามารถในการอ่านจับใจความและการแก้โจทย์ปัญหา	

	ทางคณิตศาสตร์ที่ได้จากการให้คะแนนแบบปกติ และแบบ 2-SR โดยใช้ตัวแปร คัดเลือกข้อสอบตามกรอบทฤษฎี CTT	205
27	สมการพยากรณ์ สหสัมพันธ์พหุคูณ สัมประสิทธิ์การพยากรณ์ ประสิทธิภาพในการพยากรณ์ ความคลาดเคลื่อนมาตรฐานในการประมาณค่า และรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนของสมการพยากรณ์ค่าสารสนเทศ ของข้อสอบที่ระดับความสามารถสูง (IIFH) ของแบบทดสอบวัดความสามารถ ในการอ่านจับใจความและการแก้โจทย์ปัญหาทางคณิตศาสตร์ ที่ได้จากการ ให้คะแนนแบบปกติ และแบบ 2-SR โดยใช้ตัวแปรในการคัดเลือกข้อสอบ ตามกรอบทฤษฎี IRT	206
28	สมการพยากรณ์ สหสัมพันธ์พหุคูณ สัมประสิทธิ์การพยากรณ์ ประสิทธิภาพ ในการพยากรณ์ ความคลาดเคลื่อนมาตรฐานในการประมาณค่า และรากที่สอง ของค่าเฉลี่ยความคลาดเคลื่อนของสมการพยากรณ์ค่าสารสนเทศของข้อสอบ ที่ระดับความสามารถปานกลาง (IIFM) ของแบบทดสอบวัดความสามารถ ในการอ่านจับใจความและการแก้โจทย์ปัญหาทางคณิตศาสตร์ ที่ได้จากการให้ คะแนนแบบปกติ และแบบ 2-SR โดยใช้ตัวแปรในการคัดเลือกข้อสอบตามกรอบ ทฤษฎี IRT	207
29	สมการพยากรณ์ สหสัมพันธ์พหุคูณ สัมประสิทธิ์การพยากรณ์ ประสิทธิภาพใน การพยากรณ์ ความคลาดเคลื่อนมาตรฐานในการประมาณค่า และรากที่สองของ ค่าเฉลี่ยความคลาดเคลื่อนของสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับ ความสามารถต่ำ (IIFL) ของแบบทดสอบวัดความสามารถในการอ่านจับใจความ และการแก้โจทย์ปัญหาทางคณิตศาสตร์ ที่ได้จากการให้คะแนนแบบปกติ และ แบบ 2-SR โดยใช้ตัวแปรในการคัดเลือกข้อสอบตามกรอบทฤษฎี IRT	209
30	สมการพยากรณ์ สหสัมพันธ์พหุคูณ สัมประสิทธิ์การพยากรณ์ ประสิทธิภาพ ในการพยากรณ์ ความคลาดเคลื่อนมาตรฐานในการประมาณค่า และรากที่สอง ของค่าเฉลี่ยความคลาดเคลื่อนของสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ ระดับความสามารถสูง (IIFH) ของแบบทดสอบวัดความสามารถในการอ่านจับใจ ความและการแก้โจทย์ปัญหาทางคณิตศาสตร์ ที่ได้จากการให้คะแนนแบบปกติและ แบบ 2-SR โดยใช้ตัวแปรในการคัดเลือกข้อสอบตามกรอบทฤษฎี CTT และ IRT	210

31	สมการพยากรณ์ สหสัมพันธ์พหุคูณ สัมประสิทธิ์การพยากรณ์ ประสิทธิภาพในการพยากรณ์ ความคลาดเคลื่อน มาตรฐานในการประมาณค่า และรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนของสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลาง(IFM)ของแบบทดสอบวัดความสามารถในการอ่านจับใจความและการแก้โจทย์ปัญหาทางคณิตศาสตร์ ที่ได้จากการให้คะแนนแบบปกติ และแบบ 2-SR โดยใช้ตัวแปรในการคัดเลือกข้อสอบตามกรอบทฤษฎี CTT และ IRT	212
32	สมการพยากรณ์ สหสัมพันธ์พหุคูณ สัมประสิทธิ์การพยากรณ์ ประสิทธิภาพในการพยากรณ์ ความคลาดเคลื่อนมาตรฐานในการประมาณค่า และรากที่สองของค่าเฉลี่ย ความคลาดเคลื่อนของสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำ (IFL) ของแบบทดสอบวัดความสามารถในการอ่านจับใจความและการแก้โจทย์ปัญหาทางคณิตศาสตร์ ที่ได้จากการให้คะแนนแบบปกติ และแบบ 2 - SR โดยใช้ตัวแปรในการคัดเลือกข้อสอบตามกรอบทฤษฎี CTT และ IRT	214
33	เปรียบเทียบความสอดคล้อง แผนภาพเส้นทางเชิงประจักษ์กับแผนภาพเส้นทางเต็มรูปของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนตามรูปแบบปกติ ตามกรอบทฤษฎี CTT	224
34	ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมที่มีต่อความเชื่อมั่นของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนตามรูปแบบปกติ ตามกรอบทฤษฎี CTT	224
35	เปรียบเทียบความสอดคล้อง แผนภาพเส้นทางเชิงประจักษ์กับแผนภาพเส้นทางเต็มรูปของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนตามรูปแบบ 2 - SR ตามกรอบทฤษฎี CTT	228
36	ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมที่มีต่อความเชื่อมั่นของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนโดยวิธีแบบ 2 - SR ตามกรอบทฤษฎี CTT	228

37	เปรียบเทียบความสอดคล้อง แผนภาพเส้นทางเชิงประจักษ์กับแผนภาพเส้นทาง เติมรูปของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ ที่ให้คะแนน ตามรูปแบบปกติ ตามกรอบทฤษฎี CTT	231
38	ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมที่มีต่อ ความเชื่อมั่นของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหา ทางคณิตศาสตร์ที่ให้คะแนนโดยวิธีแบบปกติ ตามกรอบทฤษฎี CTT	232
39	เปรียบเทียบความสอดคล้อง แผนภาพเส้นทางเชิงประจักษ์กับแผนภาพเส้นทาง เติมรูปของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ ที่ให้คะแนน ตามรูปแบบ 2 - SR ตามกรอบทฤษฎี CTT	235
40	ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมที่มีต่อ ความเชื่อมั่นของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทาง คณิตศาสตร์ที่ให้คะแนนโดยวิธีแบบ 2 - SR ตามกรอบทฤษฎี CTT	235
41	เปรียบเทียบความสอดคล้อง แผนภาพเส้นทางเชิงประจักษ์กับแผนภาพเส้นทาง เติมรูปของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนน ตามรูปแบบปกติ ตามกรอบทฤษฎี IRT	239
42	ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมที่มีต่อ ค่าสารสนเทศที่ระดับความสามารถสูง ปานกลาง และต่ำของแบบทดสอบ วัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนตามรูปแบบปกติ ตามกรอบทฤษฎี	240
43	เปรียบเทียบความสอดคล้อง แผนภาพเส้นทางเชิงประจักษ์กับแผนภาพเส้นทาง เติมรูปของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนน ตามรูปแบบ 2 - SR ตามกรอบทฤษฎี IRT	244
44	ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมที่มีต่อ ค่าสารสนเทศที่ระดับความสามารถสูง ปานกลาง และต่ำของแบบทดสอบวัด ความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนโดยวิธีแบบ 2 - SR ตามกรอบทฤษฎี IRT	245
45	เปรียบเทียบความสอดคล้อง แผนภาพเส้นทางเชิงประจักษ์กับแผนภาพเส้นทาง เติมรูปของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์	

	ที่ให้คะแนน ตามรูปแบบปกติ ตามกรอบทฤษฎี IRT	249
46	ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลพลอมที่มีต่อ ค่าสารสนเทศที่ระดับความสามารถสูง ปานกลาง และต่ำของแบบทดสอบ วัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ได้จากการให้คะแนน ตามรูปแบบปกติ ตามกรอบทฤษฎี IRT	250
47	เปรียบเทียบความสอดคล้อง แผนภาพเส้นทางเชิงประจักษ์กับแผนภาพเส้นทาง เต็มรูปแบบของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ ที่ให้คะแนน ตามรูปแบบ 2 - SR ตามกรอบทฤษฎี IRT	253
48	ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลพลอมที่มีต่อ ค่าสารสนเทศที่ระดับความสามารถสูง ปานกลาง และต่ำของแบบทดสอบวัด ความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ได้จากการให้คะแนน ตามรูปแบบ 2 - SR ตามกรอบทฤษฎี IRT	254
49	เปรียบเทียบความสอดคล้อง แผนภาพเส้นทางเชิงประจักษ์กับแผนภาพเส้นทาง เต็มรูปแบบของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนน ตาม รูปแบบปกติ ตามกรอบทฤษฎี CTT และ IRT	261
50	ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลพลอมที่มีต่อ ค่าสารสนเทศที่ระดับความสามารถสูง ของแบบทดสอบวัดความสามารถใน การอ่านจับใจความที่ได้จากการให้คะแนนตามรูปแบบปกติ ตามกรอบทฤษฎี CTT และ IRT	261
51	ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลพลอมที่มีต่อ ค่าสารสนเทศที่ระดับความสามารถปานกลาง ของแบบทดสอบวัดความสามารถ ในการอ่านจับใจความที่ได้จากการให้คะแนนตามรูปแบบปกติ ตามกรอบทฤษฎี CTT และ IRT	262
52	ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลพลอมที่มีต่อ ค่าสารสนเทศที่ระดับความสามารถต่ำของแบบทดสอบวัดความสามารถในการ อ่านจับใจความที่ได้จากการให้คะแนนตามรูปแบบปกติ ตามกรอบทฤษฎี CTT และ IRT	263

53	เปรียบเทียบความสอดคล้อง แผนภาพเส้นทางเชิงประจักษ์กับแผนภาพเส้นทาง เต็มรูปของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนน ตามรูปแบบ 2 - SR ตามกรอบทฤษฎี CTT และ IRT	270
54	ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมที่มีต่อ ค่าสารสนเทศที่ระดับความสามารถสูงของแบบทดสอบวัดความสามารถในการ อ่านจับใจความที่ได้จากการให้คะแนนตามรูปแบบ 2 - SR ตามกรอบทฤษฎี CTT และ IRT	270
55	ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมที่มีต่อ ค่าสารสนเทศที่ระดับความสามารถปานกลาง ของแบบทดสอบวัดความสามารถ ในการอ่านจับใจความที่ได้จากการให้คะแนนตามรูปแบบ 2 - SR ตามกรอบทฤษฎี CTT และ IRT	271
56	ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมที่มีต่อ ค่าสารสนเทศที่ระดับความสามารถต่ำของแบบทดสอบวัดความสามารถในการ อ่านจับใจความที่ได้จากการให้คะแนนตามรูปแบบ 2 - SR ตามกรอบทฤษฎี CTT และ IRT	272
57	เปรียบเทียบความสอดคล้อง แผนภาพเส้นทางเชิงประจักษ์กับแผนภาพเส้นทาง เต็มรูปของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ ที่ให้คะแนน ตามรูปแบบปกติ ตามกรอบทฤษฎี CTT และ IRT.....	279
58	ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมที่มีต่อ ค่าสารสนเทศที่ระดับความสามารถสูง ของแบบทดสอบวัดความสามารถใน การแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ได้จากการให้คะแนนตามรูปแบบปกติ ตามกรอบทฤษฎี CTT และ IRT	279
59	ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมที่มีต่อ ค่าสารสนเทศที่ระดับความสามารถปานกลาง ของแบบทดสอบวัดความสามารถ ในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ได้จากการให้คะแนนตามรูปแบบปกติ ตามกรอบทฤษฎี CTT และ IRT	280
60	ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมที่มีต่อ ค่าสารสนเทศที่ระดับความสามารถต่ำของแบบทดสอบวัดความสามารถใน	

	การแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ได้จากการให้คะแนนตามรูปแบบปกติ	
	ตามกรอบทฤษฎี CTT และ IRT	282
61	เปรียบเทียบความสอดคล้อง แผนภาพเส้นทางเชิงประจักษ์กับแผนภาพเส้นทาง เต็มรูปของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ ที่ให้คะแนน ตามรูปแบบ 2 - SR ตามกรอบทฤษฎี CTT และ IRT.....	288
62	ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมที่มีต่อ ค่าสารสนเทศที่ระดับความสามารถสูงของแบบทดสอบวัดความสามารถใน การแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ได้จากการให้คะแนนตามรูปแบบ 2-SR ตามกรอบทฤษฎี CTT และ IRT	288
63	ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมที่มีต่อ ค่าสารสนเทศที่ระดับความสามารถปานกลางของแบบทดสอบวัดความสามารถ ในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ได้จากการให้คะแนนตามรูปแบบ 2 - SR ตามกรอบทฤษฎี CTT และ IRT	289
64	ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมที่มีต่อ ค่าสารสนเทศที่ระดับความสามารถต่ำของแบบทดสอบวัดความสามารถในการ แก้โจทย์ปัญหาทางคณิตศาสตร์ที่ได้จากการให้คะแนนตามรูปแบบ 2-SR ตามกรอบทฤษฎี CTT และ IRT	290
65	ค่าสถิติของข้อสอบวัดความสามารถในการอ่านจับใจความ ตามกรอบทฤษฎี CTT และ IRT ที่ให้คะแนนตามรูปแบบปกติ และรูปแบบ 2 -SR	344
66	สารสนเทศของข้อสอบที่ระดับความสามารถต่าง ๆ ของแบบทดสอบวัดความสามารถ ในการอ่านจับใจความที่ให้คะแนนตามรูปแบบปกติ และรูปแบบ 2 -SR	347
67	ค่าสถิติของข้อสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ ตามกรอบทฤษฎี CTT และ IRTที่ให้คะแนนตามรูปแบบปกติ และรูปแบบ 2 -SR	350
68	สารสนเทศของข้อสอบที่ระดับความสามารถต่าง ๆ ของแบบทดสอบวัดความสามารถ ในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนตามรูปแบบปกติ และรูปแบบ 2 -SR	353

บัญชีภาพประกอบ

ภาพประกอบ	หน้า
1 ความสัมพันธ์ระหว่างความสามารถที่แท้จริงกับพฤติกรรมการตอบ ข้อสอบแบบ MC	12
2 การกระจายของจุดคู่ลำดับตามลักษณะของค่าสัมประสิทธิ์สหสัมพันธ์	42
3 การนำเสนอผลการทดสอบ	43
4 ค่าพารามิเตอร์ของโมเดลลอจิสติกแบบ 2 พารามิเตอร์	78
5 ความหมายค่าพารามิเตอร์ของโมเดลโลจิสติกแบบ 1 พารามิเตอร์	79
6 ความหมายค่าพารามิเตอร์ของโมเดลโลจิสติกแบบ 2 พารามิเตอร์	80
7 ความหมายค่าพารามิเตอร์ของโมเดลโลจิสติกแบบ 3 พารามิเตอร์	81
8 โค้งแสดงลักษณะการตอบสำหรับข้อความในโมเดลเกรดเรสพอน	84
9 โค้งแสดงลักษณะลำดับขั้นของข้อความโมเดลเรทติงสเกล	86
10 โค้งความน่าจะเป็นในการเลือกตอบระดับที่ 1 และ 2 ของข้อ i	89
11 เส้น ICC แสดงความน่าจะเป็นในการเลือกตอบข้อความระดับต่างๆ (3 ระดับ) ตามโมเดลพาร์เชียลเครดิต	91
12 โค้งลักษณะข้อคำถามที่มีรูปแบบของการตอบเพียงขั้นตอนเดียว (One-Step Item)	93
13 โค้งลักษณะข้อคำถามสำหรับการตอบ 7 ลำดับขั้นแรกของปัวซองเคาท์โมเดล	95
14 ช่วงความเชื่อมั่น (Confidence interval ; $\underline{\theta}$, $\bar{\theta}$) ของความสามารถที่ได้จากการ ประมาณค่า	96
15 ช่วงแห่งความเชื่อมั่นเอซิมโทติก 95% ($\underline{\theta}$, $\bar{\theta}$) สำหรับความสามารถ θ	98
16 แผนภาพการดำเนินงานเก็บรวบรวมและวิเคราะห์ผลการวิจัย	158
17 วิธีการให้คะแนนการตอบแบบ 2-SR	160
18 แผนภาพการให้คะแนนแบบ 2-SR (5, 4, และ 3 ตัวเลือก)	161
19 เส้นต่อเนื่องแสดงระดับคะแนนความสามารถของผู้สอบในการตอบข้อสอบ เลือกตอบ ชนิด 4 ตัวเลือก	166
20 สารสนเทศของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้ คะแนนตามรูปแบบปกติ และรูปแบบ 2-SR ที่ช่วงระดับความสามารถ -3.00 ถึง 3.00	181

21	สารสนเทศของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ ที่ให้คะแนนตามรูปแบบปกติ และรูปแบบ 2-SR ที่ช่วงระดับความสามารถ -3.00 ถึง 3.00	182
22	แผนภาพเส้นทางสมมติฐานแสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่าง ๆ กับความเชื่อมั่นรายข้อ ตามกรอบทฤษฎี CTT	216
23	แผนภาพเส้นทางสมมติฐานแสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่าง ๆ กับสารสนเทศของข้อสอบ ตามกรอบทฤษฎี IRT	218
24	แผนภาพเส้นทางสมมติฐานแสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่าง ๆ กับสารสนเทศของข้อสอบ ตามกรอบทฤษฎี CTT และ IRT	220
25	แผนภาพเส้นทางเต็มรูปแสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่าง ๆ กับความเชื่อมั่นรายข้อของแบบทดสอบวัดความสามารถในการอ่านจับใจความ ที่ให้คะแนนแบบปกติ ตามกรอบทฤษฎี CTT	222
26	แผนภาพเส้นทางเชิงประจักษ์ แสดงอิทธิพลทางตรงและทางอ้อมที่มีต่อความเชื่อมั่นของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนตามรูปแบบปกติ ตามกรอบทฤษฎี CTT	223
27	แผนภาพเส้นทางเต็มรูปแสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่าง ๆ กับความเชื่อมั่นรายข้อของแบบทดสอบวัดความสามารถในการอ่านจับใจความ ที่ให้คะแนนแบบ 2- SR ตามกรอบทฤษฎี CTT	226
28	แผนภาพเส้นทางเชิงประจักษ์แสดงอิทธิพลทางตรงและทางอ้อมที่มีต่อความเชื่อมั่นของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนตามรูปแบบ 2 - SR ตามกรอบทฤษฎี CTT	227
29	24 แผนภาพเส้นทางแสดงอิทธิพลทางตรงและทางอ้อมที่มีต่อความเชื่อมั่นของ แผนภาพเส้นทางเต็มรูปแสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่าง ๆ กับความเชื่อมั่นรายข้อของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาคณิตศาสตร์ ที่ให้คะแนนตามรูปแบบปกติ ตามกรอบทฤษฎี CTT	230

- 30 แผนภาพเส้นทางเชิงประจักษ์แสดงอิทธิพลทางตรงและทางอ้อมของตัวแปรต่างๆที่มีต่อความเชื่อมั่นรายข้อของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาคณิตศาสตร์ที่ให้คะแนนตามรูปแบบปกติ ตามกรอบทฤษฎี CTT 231
- 31 แผนภาพเส้นทางเต็มรูปแสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่าง ๆ กับความเชื่อมั่นรายข้อของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาคณิตศาสตร์ที่ให้คะแนนตามรูปแบบ 2 - SR ตามกรอบทฤษฎี CTT 233
- 32 แผนภาพเส้นทางเชิงประจักษ์แสดงอิทธิพลทางตรงและทางอ้อมที่มีต่อความเชื่อมั่นรายข้อของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ได้จากการให้คะแนนตามรูปแบบ 2 - SR ..ตามกรอบทฤษฎี 234
- 33 แผนภาพเส้นทางเต็มรูปแสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่าง ๆ กับค่าสารสนเทศของข้อสอบที่ระดับความสามารถ สูง ปานกลาง ต่ำ ของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนตามรูปแบบปกติตามกรอบทฤษฎี IRT 237
- 34 แผนภาพเส้นทางเชิงประจักษ์แสดงอิทธิพลทางตรงและทางอ้อมที่มีต่อสารสนเทศที่ระดับความสามารถ สูง ปานกลาง และต่ำของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนตามรูปแบบปกติ ตามกรอบทฤษฎีตามรูปแบบ IRT 238
- 35 แผนภาพเส้นทางเต็มรูปแสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่าง ๆ กับค่าสารสนเทศของข้อสอบที่ระดับความสามารถ สูง ปานกลาง ต่ำ ของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนตามรูปแบบ 2 - SR ตามกรอบทฤษฎี IRT 242
- 36 แผนภาพเส้นทางเชิงประจักษ์แสดงอิทธิพลทางตรงและทางอ้อมที่มีต่อสารสนเทศที่ระดับความสามารถ สูง ปานกลาง และต่ำของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนตามรูปแบบ 2 - SR ตามกรอบทฤษฎี IRT 243

- 37 แผนภาพเส้นทางเต็มรูปแสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่าง ๆ กับค่า
 สารสนเทศของข้อสอบที่ระดับความสามารถ สูง ปานกลาง ต่ำ ของแบบทดสอบ
 วัดความสามารถในการแก้โจทย์ปัญหาคณิตศาสตร์ที่ให้คะแนนตามรูปแบบปกติ
 ตามกรอบทฤษฎี IRT 247
- 38 แผนภาพเส้นทางเชิงประจักษ์แสดงอิทธิพลทางตรงและทางอ้อมที่มีต่อสารสนเทศที่
 ระดับความสามารถ สูง ปานกลาง และต่ำของแบบทดสอบวัดความสามารถ
 ในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ได้จากการให้คะแนนตามรูปแบบปกติ
 ตามกรอบทฤษฎี ตามกรอบทฤษฎี IRT 248
- 39 แผนภาพเส้นทางเต็มรูปแสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่าง ๆ กับ
 ค่าสารสนเทศของข้อสอบที่ระดับความสามารถ สูง ปานกลาง ต่ำ ของแบบทดสอบ
 วัดความสามารถในการแก้โจทย์ปัญหาคณิตศาสตร์ที่ให้คะแนนตามรูปแบบ 2-SR
 ตามกรอบทฤษฎี IRT 251
- 40 แผนภาพเส้นทางเชิงประจักษ์แสดงอิทธิพลทางตรงและทางอ้อมที่มีต่อสารสนเทศที่
 ระดับความสามารถ สูง ปานกลาง และต่ำของแบบทดสอบวัดความสามารถ
 ในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ได้จากการให้คะแนนตามรูปแบบ 2 - SR
 ตามกรอบทฤษฎี IRT 252
- 41 แผนภาพเส้นทางเต็มรูปแสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่าง ๆ กับ
 ค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูงของแบบทดสอบ
 วัดความสามารถในการอ่านจับใจความที่ให้คะแนนตามรูปแบบปกติ
 ตามกรอบทฤษฎี CTT และ IRT 255
- 42 แผนภาพเส้นทางเต็มรูปแสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่าง ๆ กับ
 ค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลางของแบบทดสอบ
 วัดความสามารถในการอ่านจับใจความที่ให้คะแนนตามรูปแบบปกติ
 ตามกรอบทฤษฎี CTT และ IRT 256
- 43 แผนภาพเส้นทางเต็มรูปแสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่าง ๆ กับ
 ค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำของแบบทดสอบ
 วัดความสามารถในการอ่านจับใจความที่ให้คะแนนตามรูปแบบปกติ
 ตามกรอบทฤษฎี CTT และ IRT 257

- 44 แผนภาพเส้นทางเชิงประจักษ์แสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่างๆ กับ
 สารสนเทศของข้อสอบที่ระดับความสามารถ สูง ปานกลาง และต่ำของ
 แบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนน
 ตามรูปแบบปกติ ตามกรอบทฤษฎี CTT และ IRT 259
- 45 แผนภาพเส้นทางเต็มรูปแสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่าง ๆ กับ
 ค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูงของแบบทดสอบ
 วัดความสามารถในการอ่านจับใจความที่ให้คะแนนตามรูปแบบ 2 - SR
 ตามกรอบทฤษฎี CTT และ IRT 264
- 46 แผนภาพเส้นทางเต็มรูปแสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่าง ๆ กับ
 ค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลางของแบบทดสอบ
 วัดความสามารถในการอ่านจับใจความที่ให้คะแนนตามรูปแบบ 2 - SR
 ตามกรอบทฤษฎี CTT และ IRT 265
- 47 แผนภาพเส้นทางเต็มรูปแสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่าง ๆ กับ
 ค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำของแบบทดสอบ
 วัดความสามารถในการอ่านจับใจความที่ให้คะแนนตามรูปแบบ 2 - SR
 ตามกรอบทฤษฎี CTT และ IRT 266
- 48 แผนภาพเส้นทางเชิงประจักษ์แสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่างๆ กับ
 สารสนเทศของข้อสอบที่ระดับความสามารถ สูง ปานกลาง และต่ำของ
 แบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนน
 ตามรูปแบบ 2 - SR ตามกรอบทฤษฎี CTT และ IRT 268
- 49 แผนภาพเส้นทางเต็มรูปแสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่างๆ กับค่า
 สารสนเทศที่ระดับความสามารถสูง ของแบบทดสอบวัดความสามารถ
 ในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ได้จากการให้คะแนนตามรูปแบบปกติ
 ตามกรอบทฤษฎี CTT และ IRT 273
- 50 แผนภาพเส้นทางเต็มรูปแสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่างๆ กับค่า
 สารสนเทศที่ระดับความสามารถปานกลาง ของแบบทดสอบวัดความสามารถ
 ในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ได้จากการให้คะแนนตามรูปแบบปกติ
 ตามกรอบทฤษฎี CTT และ IRT 274

- 51 แผนภาพเส้นทางเติมรูปแสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่างๆ กับค่า
 สารสนเทศที่ระดับความสามารถต่ำ ของแบบทดสอบวัดความสามารถ
 ในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ได้จากการให้คะแนนตามรูปแบบปกติ
 ตามกรอบทฤษฎี CTT และ IRT 275
- 52 แผนภาพเส้นทางเชิงประจักษ์แสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่างๆ กับ
 ค่าสารสนเทศที่ระดับความสามารถสูง ปานกลาง และ ต่ำของแบบทดสอบ
 วัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ได้จากการให้คะแนน
 ตามรูปแบบปกติตามกรอบทฤษฎี CTT และ IRT 277
- 53 แผนภาพเส้นทางเติมรูปแสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่างๆ กับค่า
 สารสนเทศที่ระดับความสามารถสูง ของแบบทดสอบวัดความสามารถ
 ในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ได้จากการให้คะแนนตามรูปแบบ 2 - SR
 ตามกรอบทฤษฎี CTT และ IRT 282
- 54 แผนภาพเส้นทางเติมรูปแสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่างๆ กับค่า
 สารสนเทศที่ระดับความสามารถปานกลาง ของแบบทดสอบวัดความสามารถ
 ในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ได้จากการให้คะแนนตามรูปแบบ 2 - SR
 ตามกรอบทฤษฎี CTT และ IRT 283
- 55 แผนภาพเส้นทางเติมรูปแสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่างๆ กับค่า
 สารสนเทศที่ระดับความสามารถต่ำ ของแบบทดสอบวัดความสามารถ
 ในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ได้จากการให้คะแนนตามรูปแบบ 2 - SR
 ตามกรอบทฤษฎี CTT และ IRT 284
- 56 แผนภาพเส้นทางเชิงประจักษ์แสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่างๆ กับ
 ค่าสารสนเทศที่ระดับความสามารถสูง ปานกลาง และ ต่ำของแบบทดสอบ
 วัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ได้จากการให้คะแนน
 ตามรูปแบบ 2 - SR ตามกรอบทฤษฎี CTT และ IRT 286
- 57 รูปแบบการพัฒนาแบบทดสอบ MC ตามกรอบทฤษฎี CTT และ IRT 309

บทที่ 1

บทนำ

ภูมิหลัง

แบบทดสอบเลือกตอบเป็นเครื่องมือวัดทางจิตวิทยาประเภทหนึ่งซึ่งนิยมใช้กันอย่างแพร่หลายนับตั้งแต่ราว ๆ ปี ค.ศ. 1920 เป็นต้นมา (Choppin in Keeves. 1988 : 384 ; Rogers in Husen and Postlethwaite. 1994 : 3996) โดยเฉพาะการวัดทางด้านสติปัญญา (Intelligence) และด้านการรู้คิด (Cognition) ดังเห็นได้จากแบบทดสอบมาตรฐานในทุกระดับตั้งแต่วัยเด็กปฐมวัยจนถึงระดับอุดมศึกษา จุดประสงค์การใช้แบบทดสอบเลือกตอบมีหลากหลาย เช่น การวัดความพร้อม การวินิจฉัยข้อบกพร่อง การสอบคัดเลือก การจัดตำแหน่ง การเลื่อนขั้นเลื่อนตำแหน่ง นิยมใช้แบบทดสอบเลือกตอบแทบทั้งสิ้น ทั้งนี้อาจเป็นเพราะว่าแบบทดสอบชนิดนี้ สามารถออกได้ครอบคลุมเนื้อหาที่ต้องการวัดได้ นอกจากนั้นโครงสร้างของข้อสอบสามารถโน้มนำให้ได้คำตอบที่ตรงประเด็น และโดยเฉพาะอย่างยิ่ง การตรวจให้คะแนนมีหลักเกณฑ์ชัดเจน สะดวกรวดเร็วเหมาะสมกับการสอบวัดกับบุคคลจำนวนมาก ๆ และสอดคล้องกับการใช้เทคโนโลยีทางคอมพิวเตอร์

การพัฒนาแบบทดสอบในระยะเริ่มแรกนั้นดำเนินการภายใต้กรอบความคิดพื้นฐานจากทฤษฎีการวัดแบบมาตรฐานเดิม (Classical test theory ; CTT) ที่นำเสนอโดย ชาร์ล สเปียร์แมน (Charles spearman) เมื่อประมาณปี ค.ศ. 1904-1913 ซึ่งเน้นมโนทัศน์เกี่ยวกับคะแนนจริง (True Score) ความคลาดเคลื่อนในการวัด (Measurement error) และดัชนีแสดงความเชื่อมั่นของแบบทดสอบ (Test reliability) (Crocker and Algina. 1986 : 106 ; citing Spearman. 1904, 1907, 1913 ; Keeves. 1988 : 242) ซึ่งต่อมา กิลฟอร์ด (Guilford) และ กัลลิกเซน (Gulliksen) ได้ขยายแนวความคิดของสเปียร์แมน โดยแสดงให้เห็นว่าคะแนนจากการทดสอบ (Observed test score) ประกอบด้วยส่วนประกอบสำคัญสองส่วนคือ คะแนนจริง และความคลาดเคลื่อนแบบสุ่ม (Random error component) ซึ่งแสดงในรูปของสมการเส้นตรงเชิงคณิตศาสตร์ได้ว่า $X = T + E$ เมื่อ X คือคะแนนที่วัดได้จากการทดสอบ T คือคะแนนจริงของผู้สอบแต่ละบุคคล และ E คือความคลาดเคลื่อนแบบสุ่มจากการทดสอบ โดยมีข้อตกลงสำคัญ 3 ประการคือ ค่าเฉลี่ยของคะแนนคลาดเคลื่อนของประชากรผู้สอบเท่ากับ 0 ($\mu_E = 0$) ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนจริงและคะแนนคลาดเคลื่อนของประชากรผู้สอบต้องไม่สัมพันธ์กัน ($\rho_{TE} = 0$) และคะแนนคลาดเคลื่อนของการ

ทดสอบทั้งสองครั้งไม่ว่าจะเป็นการสอบซ้ำ (Test-retests) ด้วยแบบทดสอบชุดเดียวกัน หรือด้วยแบบทดสอบที่คู่ขนานกัน (Parallel forms) ต้องไม่มีความสัมพันธ์กัน ($\rho_{E1E2} = 0$) (Gulliksen. 1950 : 4-11 ; Lord. 1980 : 4-5 ; Crocker and Algina. 1986 : 106-113 ; citing Guilford. 1936 , Magnussen. 1967 ; Wilcox in Keeves. 1988 : 267 ; Keat in Husen and Postlethwaite. 1994 : 785-786)

กระบวนการพัฒนาแบบทดสอบเลือกตอบภายใต้กรอบความคิดตามทฤษฎีมาตรฐานเดิมได้ดำเนินการอย่างต่อเนื่อง โดยมีแนวดำเนินการ 2 กลุ่มใหญ่ ๆ คือ กลุ่มที่มุ่งพัฒนาตัวบ่งชี้คุณภาพ และประสิทธิภาพของแบบทดสอบในเชิงสถิติ ได้แก่ ความเที่ยงตรง (Validity) และความเชื่อมั่น กลุ่มนี้พยายามศึกษาแนวทาง และพัฒนาสูตรทางคณิตศาสตร์เพื่อประมาณค่าดัชนีความเที่ยงตรง และความเชื่อมั่น รวมทั้งการใช้กระบวนการทางการวิจัย ศึกษาแหล่งของความคลาดเคลื่อนในการวัด ซึ่งเป็นที่มาของทฤษฎีการสรุปอ้างอิง (Generalizability theory ; G-theory) ส่วนอีกกลุ่มหนึ่งมุ่งประเด็นไปที่คะแนนสอบวัด โดยพยายามศึกษาวิธีการเพื่อแก้ไขจุดอ่อนของแบบทดสอบเลือกตอบเกี่ยวกับเรื่องของการเดา (Guessing) ซึ่งมีแนวทางการศึกษาสองแนวทางคือ ใช้สูตรปรับคะแนนหลังจากการสอบวัด และวิธีกำหนดการตอบและการให้คะแนนขึ้นใหม่ที่มิใช่แบบสองทาง (Dichotomous scoring) แต่เป็นแบบหลายทาง (Polychotomous scoring) หรือให้คะแนนความรู้บางส่วน (Partial knowledge scoring)

ครอนบัคและคณะ (Cronbach et al.) ได้กล่าวถึงการประเมินโดยใช้แบบทดสอบว่าจะต้องเผชิญกับปัญหาสำคัญ 2 ประการที่มีความแตกต่างกัน แต่ก็มีส่วนซ้อนทับกัน ประการแรกก็คือ จะต้องศึกษาว่าสาระของแบบทดสอบครอบคลุมคำตอบที่เป็นไปได้ทั้งหมดในสิ่งที่สนใจมากน้อยเพียงใด และประเด็นที่สองก็คือ ความสัมพันธ์ระหว่างกลุ่มของคำตอบที่เป็นไปได้ทั้งหมดนั้นกับคุณลักษณะที่แฝงในผู้สอบเชื่อถือได้เพียงใด ซึ่งคำถามในประเด็นแรกก็คือความเชื่อมั่น ส่วนในประเด็นหลัง ก็คือความเที่ยงตรงของแบบทดสอบนั่นเอง (Thorndike in Keeves. 1988 :330 ; citing Cronbach et al. 1972) นอกจากนั้น เมสสิค (Messick) ได้กล่าวถึงคะแนนสอบว่าปกติจะต้องถูกนำไปอ้างอิงเกี่ยวกับพฤติกรรมของผู้สอบในสถานการณ์ที่นอกเหนือไปจากสถานการณ์สอบ หรือเป็นการอ้างอิงไปสู่พฤติกรรมทั้งหมด (Domain of behaviors) ของผู้สอบที่แสดงในปัจจุบันหรือในอนาคต การนำคะแนนสอบไปใช้อ้างอิงก็ขึ้นอยู่กับจุดประสงค์ของผู้วัด ดังนั้น จึงมีความจำเป็นที่ผู้วัดจะต้องทำให้คะแนนสอบนั้นมีความคงที่และมีหลักประกันได้ว่าสามารถอ้างอิงไปถึงสิ่งที่ต้องการศึกษาในตัวของผู้สอบ (Crocker and Algina. 1986 : 217 ; citing Messick. 1981) จากแนวคิดดังกล่าวน่าจะเป็นเหตุผลสำคัญที่ทำให้การพิจารณาคุณภาพหรือประสิทธิภาพของแบบทดสอบเน้น

ไปที่ความเที่ยงตรงและความเชื่อมั่น โดยให้ความสำคัญกับความเที่ยงตรงเป็นประการแรก และถือว่าเป็นหัวใจสำคัญของการพัฒนาคุณภาพการนำเสนอ และการใช้ผลการวัดทางการศึกษา (Oosterhof. 1994 : 53)

สำหรับความเกี่ยวข้องระหว่างความเที่ยงตรงและความเชื่อมั่นนี้ไคลน์ (Kline) กล่าวสรุปไว้ว่าแบบทดสอบใดมีความเที่ยงตรงก็ย่อมจะมีความเชื่อมั่นด้วย แต่แบบทดสอบใดที่มีค่าความเชื่อมั่นในระดับสูงก็ไม่สามารถบอกได้ว่าแบบทดสอบนั้นจะมีความเที่ยงตรงสูงตามไปด้วย (Kline. 1993 : 26-27) การพัฒนาแบบทดสอบให้มีความเที่ยงตรงและมีความเชื่อมั่นนั้น ลินด์ควิสต์ (Linguist) ได้เสนอแนวคิดในการพัฒนาแบบทดสอบเพื่อการตัดสินใจสองคำถามหลัก คือ จะวัดอะไรและจะวัดอย่างไร และได้เสนอขั้นตอนในการพัฒนาไว้ 5 ขั้นตอน คือ 1) กำหนดรูปแบบ 2) วิเคราะห์รูปแบบกับการนำไปใช้ 3) คัดเลือกและฝึกผู้เขียนข้อสอบ 4) เขียนข้อสอบ และ 5) ติดตามความก้าวหน้าของผู้ออกข้อสอบและคุณภาพข้อสอบ (Crocker and Algina. 1986 : 75 ; citing Linquist.1936) ซึ่งต่อมาได้มีการขยายแนวความคิดโดยนักวัดผลต่าง ๆ เกี่ยวกับกระบวนการพัฒนาแบบทดสอบเลือกตอบขึ้นอย่างหลากหลาย ซึ่งสรุปได้ว่าประกอบด้วยกระบวนการย่อยสำคัญ 4 กระบวนการคือ การวางแผนการสร้าง (Test plan) การสร้าง (Test construction) การวิเคราะห์ (Item analysis) และการดำเนินการจัดการ (Test administration) โดยเริ่มต้นจากการกำหนดเป้าหมายของการทดสอบ และขั้นสุดท้ายคือ การนำเสนอผลการสอบ

การวิเคราะห์ข้อสอบ นับว่าเป็นกระบวนการสำคัญทางสถิติที่อธิบายผลการตอบของผู้สอบในแต่ละข้อคำถาม ซึ่งจะส่งผลต่อความเที่ยงตรงและความเชื่อมั่นของแบบทดสอบ คุณสมบัติสำคัญของข้อสอบที่นิยมวิเคราะห์กันมากมีสองค่าคือ ค่าความยาก (Difficulty) และ ค่าอำนาจจำแนก (Discrimination) ค่าความยากของข้อสอบเป็นดัชนีที่บ่งบอกว่าข้อสอบข้อนั้นมีคนตอบถูกคิดเป็นสัดส่วนเท่าไร เมื่อเปรียบเทียบกับผู้เข้าสอบทั้งหมด หรือสัดส่วนข้อสอบข้อนั้นมีคนตอบถูกคิดเป็นสัดส่วนเท่าไร เมื่อเปรียบเทียบกับผู้เข้าสอบทั้งหมดหรือสัดส่วนของคนตอบถูกเมื่อข้อสอบนั้นมีการให้คะแนนแบบ 0-1 โดยทั่วไปสัญลักษณ์แทนความยากข้อสอบคือ p ดังนั้นค่า p_i คือ สัดส่วนคนตอบถูก สำหรับข้อ i ตามปกติ ค่า p_i ก็คือค่าเฉลี่ยข้อสอบ (Item mean) ที่ให้คะแนนแบบ 0-1 ซึ่งค่า p จะกระจายอยู่ในช่วง 0 ถึง 1 และถ้านำไปลบออกจาก 1 จะได้สัดส่วนคนตอบผิดหรือ q เมื่อนำ p และ q มาคูณกันจะแสดงความแปรปรวนของข้อสอบ (Item variance) (Gulliksen. 1950 : 376 ; Crocker and Algina. 1986 : 90-92) และค่าความแปรปรวนของคะแนนสอบจะมีค่าสูงสุดเมื่อ p มีค่าเท่ากับ .50 ดังนั้น ถ้าสามารถจัดแบบทดสอบให้ค่า p ของข้อสอบเท่ากับ

.50 ทุกข้อก็น่าจะทำให้ค่าความเชื่อมั่นมีค่าเท่ากับ 1.00 สำหรับในกรณีที่การให้คะแนนไม่ใช่แบบ 0-1 (Multicategory scoring) ให้พิจารณาจากค่าเฉลี่ยของคะแนน ($\bar{X}$) และ ส่วนเบี่ยงเบนมาตรฐาน (S.D) หรือความแปรปรวน (S^2) ของแต่ละข้อแทนดัชนีความยากของข้อสอบ (Millman and Greene in Linn. 1989 : 358 ; citing University of Iowa. 1984 ; Nunnally and Bernstein.1994 : 302) สำหรับค่าอำนาจจำแนก เป็นดัชนีที่ต้องการบอกคุณลักษณะของข้อสอบเกี่ยวกับความแตกต่างระหว่างบุคคลในโครงสร้างที่ต้องการวัดโดยแบบทดสอบ หรือเกณฑ์ภายนอกบางประการ ปกติจะใช้คะแนนรวมเพื่อจัดกลุ่มผู้สอบที่มีคะแนนสูงซึ่งมั่นใจว่ามีความน่าจะเป็นในการตอบถูก สูงกว่าผู้สอบที่มีคะแนนต่ำซึ่งคาดว่าจะมีโอกาสในการตอบข้อสอบถูกน้อยมาก ค่าอำนาจจำแนกมีค่ากระจายอยู่ในช่วงระหว่าง -1.00 ถึง 1.00 ซึ่งประมาณค่าได้ 2 แนวทางใหญ่ ๆ คือ ใช้คะแนนจุดตัดแบ่งกลุ่มผู้สอบออกเป็นกลุ่มได้คะแนนสูง และกลุ่มได้คะแนนต่ำกลุ่มละ 50 % หรือ 30% หรือ 27 % ตามขนาดของกลุ่มตัวอย่าง แล้วพิจารณาความแตกต่างของสัดส่วนการตอบถูกของผู้สอบทั้ง 2 กลุ่ม ซึ่งความแตกต่างที่ได้ก็คือ ค่าดัชนีการจำแนก ซึ่งแทนด้วยสัญลักษณ์ D (Crocker and Algina. 1986 : 313 ; citing Kelly. 1939, Elglehert. 1965, Beuchert and Mendoza. 1979) ส่วนอีกแนวทางหนึ่งนั้น เป็นการหาความสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนรวม หรือเกณฑ์ภายนอกอื่น ๆ โดยใช้สูตรประมาณค่าสัมประสิทธิ์สหสัมพันธ์ (r) 4 แบบ ได้แก่ 1) ρ_{bis} (Biserial correlation coefficient) 2) ρ_{pbis} (Point biserial correlation coefficient) 3) ϕ (Phi coefficient) และ 4) r_{tet} (Tetrachoric correlation coefficient) (Crocker and Algina. 1986 : 319) ในการคัดเลือกข้อสอบเข้าฉบับควรพิจารณาค่าอำนาจจำแนกเป็นประเด็นแรก ทั้งนี้เพราะว่าข้อสอบที่มีค่าอำนาจจำแนกสูงจะทำให้ความแปรปรวนของคะแนนสอบสูง ซึ่งจะมีผลในการเพิ่มค่าความเชื่อมั่นสูงตามไปด้วย ในทางตรงข้ามถ้าค่าอำนาจจำแนกของข้อสอบต่ำจะมีผลทำให้ค่าความเชื่อมั่นต่ำลงไปด้วย ในทางปฏิบัติน่าจะยอมรับค่าอำนาจจำแนกตั้งแต่ .30 ขึ้นไป แต่ถ้าข้อสอบมีจำนวนมาก ๆ ค่าอำนาจจำแนกของข้อสอบมากกว่า .20 ก็น่าจะยอมรับได้ (Nunnally and Bernstein. 1994 : 305-306) อย่างไรก็ตาม อีเบล (Ebel) เสนอว่า ค่า $D > .40$ เหมาะสมสำหรับแบบทดสอบมากที่สุด (Crocker and Algina.1986 : 315 ; citing Ebel. 1965) สำหรับความสัมพันธ์ระหว่างค่าความยากกับค่าอำนาจจำแนกพบว่า ค่าอำนาจจำแนกจะมีค่าสูงสุดเมื่อค่าความยากเท่ากับ .50 และเมื่อใดที่ค่าความยากเท่ากับ 0 หรือ 1 ข้อสอบนั้นก็ไม่สามารถจำแนกผู้สอบได้เลย (Murphy and Davidshofer. 1988 : 137) นอกจากนั้นในการคำนวณ ค่าอำนาจจำแนกจาก ρ_{pbis} และ ρ_{bis} ต้องอาศัยค่าความยาก

ในการคำนวณ หรือคำนวณค่าความสัมพันธ์ระหว่างข้อจากสูตร ϕ จะมีค่าเท่ากับ 1.00 ได้นั้น ค่าความยากของข้อสอบทั้ง 2 ข้อ จะต้องเท่ากัน ทั้งนี้เพราะค่า ϕ เหมือนกับค่า ρ_{pbis} ตรงที่แปลงโดยตรงจากสูตรสหสัมพันธ์ของเพียร์สัน (Crocker and Algina. 1986 : 317-319)

ในกระบวนการวิเคราะห์ข้อสอบนอกจากนิยามพิจารณาค่าความยากและค่าอำนาจจำแนกของข้อสอบแล้ว การวิเคราะห์ตัวลวง (Distractor analysis) ก็นับว่ามีการใช้กันอย่างกว้างขวาง ซึ่งจะทำให้ทราบแบบแผนการตอบว่าในแต่ละตัวเลือกของแต่ละข้อมีผู้เลือกตอบจำนวนเท่าไร หรือคิดเป็นร้อยละเท่าไร อันจะนำไปสู่การอธิบายใน 2 ประเด็นคือ ตัวเลือกสะท้อนให้เห็นความรู้บางส่วน (Partial knowledge) หรือโครงสร้างของคำถามมีจุดอ่อน นอกจากนั้นยังแสดงความเที่ยงตรงในการวัดตามโดเมนเนื้อหา ความยากของข้อสอบ และโดยเฉพาะ ถ้ากลุ่มเก่งเลือกตัวลวงมาก ๆ จะทำให้ค่าความเชื่อมั่นและความเที่ยงตรงของแบบทดสอบต่ำลง นอกจากนั้นยังพบว่าข้อสอบที่มีค่าอำนาจจำแนกน้อย ๆ ส่วนใหญ่แสดงให้เห็นความด้อยประสิทธิภาพของตัวลวง และข้อสอบที่มีตัวลวงที่ด้อยเหตุผลจะให้ความยากลดลง และถ้าข้อสอบมีความยากหรือง่ายจนเกินไป ค่าความเที่ยงตรงเชิงจำแนกจะมีค่าน้อยมาก (Murphy and Davidshofer. 1988 : 127-138 ; Gay. 1991 : 255-257 ; Oosterhof. 1994 : 199-202)

นอกจากการวิเคราะห์คุณสมบัติของข้อสอบและแบบทดสอบดังกล่าวแล้ว มิลล์แมนและกรีน ยังได้แสดงความสนใจเกี่ยวกับมิติของเนื้อหาที่กำลังวัด โดยมีความเห็นว่ามีมิติของเนื้อหาเป็นสิ่งที่สำคัญที่สุดในการอ้างอิงมโนทัศน์หรือทฤษฎีเกี่ยวกับความเป็นเอกพันธ์ของเนื้อหา (Homogeneity) หรือเป็นเนื้อหาแบบวิวิธพันธ์ (Heterogeneity) หรือโดเมนที่คาดว่าจะส่งผลกระทบต่อคะแนนสอบและความเชื่อมั่นของแบบทดสอบ ความเป็นมิติเดียวของเนื้อหา (Unidimensionality of test content) จะทำให้เกิดคะแนนที่มีความหมายเดียวหรือคะแนนแบบเอกนัย (Single test score) ที่สืบเนื่องมาจากความเป็นเอกพันธ์ของชุดข้อสอบ ในทำนองเดียวกันถ้าแบบทดสอบวัดหลายมิติ (Multidimensionality) คะแนนที่ได้ก็จะมีหลายความหมาย หรือเป็นคะแนนแบบพหุนัย (Multiple test scores) ซึ่งเป็นผลสืบเนื่องมาจากแบบทดสอบที่หลากหลายชุด หรือข้อสอบมีลักษณะเป็นวิวิธพันธ์ (Millman and Greene. in Linn. 1989 : 341) ไคลน์ (Kline. 1993 : 134-148) ได้แสดงความเห็นว่ามันนอกเหนือจากการใช้กระบวนการวิเคราะห์ข้อสอบในการสร้างข้อสอบแล้วอีกวิธีหนึ่งที่ใช้กันอย่างแพร่หลายโดยเฉพาะแบบทดสอบที่ไม่วัดสติปัญญา (Noncognitive test) ก็คือการวิเคราะห์องค์ประกอบ โดยเน้นข้อสอบที่จะใช้ในการวิเคราะห์องค์ประกอบจะต้องมีค่าความยากระหว่าง .20 ถึง .80 ซึ่งเป็นเกณฑ์ปกติเช่นเดียวกับการวิเคราะห์ข้อสอบ เกณฑ์การคัดเลือกข้อสอบจะพิจารณาจากค่าน้ำหนักองค์ประกอบ (Factor loading) ที่มีนัยสำคัญตั้งแต่ .30 สำหรับองค์

ประกอบแรกเท่านั้น นอกจากนั้น ไคลน์ ได้ให้ความสนใจในความแตกต่างระหว่างเพศทั้งในการวิเคราะห์ข้อสอบหรือการวิเคราะห์องค์ประกอบโดยให้เหตุผลว่าจะทำให้ขนาดความคลาดเคลื่อนมาตรฐานของน้ำหนักของข้อสอบเพิ่มขึ้นถ้ามีการจัดลำดับของข้อสอบในองค์ประกอบที่จำเป็นจะต้องเหมือนกัน และในการนำเสนอองค์ประกอบต่าง ๆ ไม่ควรให้เกิดความแตกต่างระหว่างเพศในเรื่องของน้ำหนักองค์ประกอบ หรือค่าความยากในการวิเคราะห์ข้อสอบ ซึ่งแนวคิดของไคลน์นี้ นับว่าสอดคล้องกับนักพัฒนาแบบทดสอบอีกหลาย ๆ คนที่สนใจเกี่ยวกับความลำเอียง และความยุติธรรมของข้อสอบ (Bias and Fairness) ซึ่งใช้ในการตัดสินข้อสอบแต่ละข้อว่าเป็นฟังก์ชันของการสอบสำหรับผู้สอบ 2 กลุ่มในทิศทางเดียวกันหรือไม่ โดยปกติจะเกี่ยวข้องกับเพศ พื้นฐานทางเชื้อชาติ อายุ ประสบการณ์ หรือเงื่อนไขความพิการ รวมทั้งภาษาถิ่น การศึกษาความลำเอียงของข้อสอบนี้เหมาะที่จะนำไปประยุกต์ใช้เมื่อปรากฏว่ากลุ่มผู้สอบที่สนใจสองกลุ่มมีความแตกต่างของระดับเฉลี่ยของความสามารถ ความรู้หรือทักษะที่กำลังวัดในลักษณะการเปรียบเทียบทางตรงระหว่างพฤติกรรมของผู้สอบ (Scheuneman and Bleistein in Husen and Postlethwaite. 1994 : 3043) หรือกล่าวได้ว่า ถ้ากลุ่มที่มีความสามารถเท่ากันแต่คะแนนสอบที่ได้ มีคะแนนแตกต่างกัน แสดงว่าแบบทดสอบนั้นมีความลำเอียง (Bias) ซึ่งมีความสำคัญมากต่อการจัดตำแหน่งของผู้สอบ (Oosterhof. 1994 : 413-427) การศึกษาความลำเอียงของข้อสอบเริ่มต้นเมื่อปลาย ๆ ปี ค.ศ.1960 และต้น ๆ ปี 1970 (Crocker and Algina. 1986 : 267 ;Scheuneman and Bleistein in Husen and Postlethwaite. 1994 : 3043) ซึ่งศึกษาใน 2 รูปแบบ คือ ความลำเอียงในการวัดและความลำเอียงในการพยากรณ์

การพัฒนาแบบทดสอบเลือกตอบนอกเหนือจากพัฒนาในรูปแบบทางคณิตศาสตร์และสถิติดังที่กล่าวมาแล้วนั้น นักวัดผลกลุ่มหนึ่งได้พยายามพัฒนาแบบทดสอบเลือกตอบโดยเน้นที่วิธีตอบและการให้คะแนน ซึ่งมีเป้าหมายเพื่อที่จะแก้จุดอ่อนของแบบทดสอบเลือกตอบเกี่ยวกับการเดาเพื่อให้คะแนนสอบที่ได้สามารถนำไปประมาณค่าความเที่ยงตรงและความเชื่อมั่นของแบบทดสอบให้มีความคลาดเคลื่อนน้อยที่สุด แนวคิดของกลุ่มนี้มีแนวปฏิบัติ 2 แนวทางสำคัญ คือ ปรับแก้การเดาและการให้คะแนนที่เน้นการแสดงความรู้บางส่วน แต่การใช้สูตรปรับแก้การเดาดังกล่าวเห็นได้ว่า เน้นที่คะแนนรวมทั้งฉบับ (Total score) ซึ่งไม่ได้ให้รายละเอียดในส่วนที่เป็นรายข้อเท่าไรนัก จึงได้มีแนวทางที่สองที่เน้นการปรับวิธีการตอบ และให้คะแนนเป็นรายข้อ ซึ่งคะแนนแต่ละข้อจะไม่ใช้แบบ 0-1 แต่จะเป็นลักษณะหลาย ๆ คะแนน (Polychotomous scores) และเน้นให้คะแนนความรู้บางส่วน ได้แก่ 1) วิธีการเฉพาะข้อผิดของคัมป์ส (Coombs. 1953 : 308-310 : 2) วิธีให้ตอบได้มากกว่า 1 ตัวเลือกของเดรสเซลและชมิท (Dressel and Schmid. 1953 : 574-595) และวิธี

ของสำราญ มีแจ้ง (2525 : 10-11) 3) วิธีกำหนดน้ำหนักคะแนนทุกตัวเลือกโดยใช้คะแนนมาตรฐานเฉลี่ย จากการเลือกตอบของกลุ่มตัวอย่างในแต่ละตัวเลือก (Davis and Fifer. 1959 : 159-170) และวิธีกำหนดความถูกต้องของแต่ละตัวเลือกโดยผู้เชี่ยวชาญ (Patnaik and Traub. 1973 : 281-286) นอกจากนั้น ยังมีการกำหนดน้ำหนักจากค่าดัชนี ρ_{pbis} และดัชนีของบรอกเดน (Brogden's Coefficient) (Waters. 1982 : 256-260 : 4) วิธีกำหนดคะแนนความเชื่อมั่นของแต่ละข้อโดยผู้สอบ วิธีนี้ให้ผู้สอบบอกคะแนนที่ควรจะได้เมื่อตอบถูก ตั้งแต่ 1 ถึง 5 หรือบอกความมั่นใจหรือไม่มั่นใจในการตอบ (Rippey. 1970 : 165-170 ; Echternacht. 1972 : 217-234 ; Abu-Sayf. 1975 : 160-162 ; Pugh and Brunza. 1975 : 73-78 ; citing Ahlgren. 1969) และ 5) วิธีตอบทุกตัวเลือกของ อนันต์ ศรีโสภา (2516 : 41-46) 6) วิธีใช้เครื่องบันทึกคำตอบของ เกริกชัย ฮวบเจริญ (2525 : 5) จากวิธีการที่กล่าวมาทั้ง 6 วิธีนั้นรูปแบบข้อสอบยังมีโครงสร้างเช่นเดียวกับข้อสอบเลือกตอบปกติคือ มีตัวเลือกถูกเพียงตัวเลือกเดียว แต่ยังมีข้อสอบเลือกตอบอีก 2 รูปแบบที่แตกต่างจากแบบปกติ คือ 1) แบบทดสอบเลือกตอบให้ตอบว่าถูกหรือผิดทุกตัวเลือก (Multiple true-false test items) วิธีนี้โครงสร้างของข้อสอบจะมีสองส่วนเช่นเดียวกับแบบปกติคือ ส่วนคำถาม และส่วนตัวเลือก แต่ในส่วนตัวเลือกอาจมีคำตอบถูกเพียงข้อเดียวหรือมากกว่า หรือไม่มีข้อถูกเลย ผู้สอบจะต้องตอบทุกตัวเลือก (Cronbach. 1941 : 533-543 ; Ebel. 1978 : 37-44 ; Frisbie and Sweeney. 1982 : 29-34 ; Albanese and Sabers. 1988 : 111-123 ; Oosterhof. 1994 : 169-171) 2) แบบทดสอบเลือกตอบชนิดตัวเลือกต่อเนื่อง (Sequential true-false items) แบบทดสอบลักษณะนี้ประกอบด้วยข้อสอบแต่ละข้อจะมีตัวเลือกเกี่ยวเนื่องกันเป็นลำดับขั้น ดังนั้น ทุกตัวเลือกจึงมีส่วนถูกทุกข้อและคะแนนในแต่ละตัวเลือกจะมีน้ำหนักไม่เท่ากันตามลำดับความถูกต้องสมบูรณ์ของตัวเลือก (Haladyna. 1994 : 44-46 ; Oosterhof. 1994 : 171) .

ขณะที่การพัฒนาแบบทดสอบเลือกตอบภายใต้กรอบความคิดพื้นฐานทฤษฎีมาตรฐานเดิม (CTT) ดำเนินไปอย่างต่อเนื่อง และข้อปัญหาหลาย ๆ ประการไม่ได้รับการอธิบายอย่างชัดเจน โดยเฉพาะใน 5 ประเด็นสำคัญคือ 1) ค่าสถิติของข้อคำถาม เช่น ความยาก อำนาจจำแนกขึ้นอยู่กับความสามารถของผู้สอบ กล่าวคือ ถ้าผู้สอบมีความสามารถสูงข้อคำถามนั้นจะง่าย แต่ถ้าผู้สอบมีความสามารถต่ำข้อคำถามนั้นจะยาก ส่วนค่าอำนาจจำแนกขึ้นอยู่กับความเป็นเอกพันธ์ของผู้สอบ (Group homogeneity) ถ้าผู้สอบมีความสามารถแตกต่างกันมาก (Group heterogeneity) ข้อคำถามก็จะมีอำนาจจำแนกสูง ซึ่งมีผลทำให้ความเชื่อมั่นสูงตามไปด้วย 2) การเปรียบเทียบความสามารถของผู้สอบนั้นจะต้องใช้แบบทดสอบฉบับเดียวกัน หรือแบบทดสอบคู่ขนาน ปัญหาที่เกิดขึ้นก็คือแบบทดสอบวัดผลสัมฤทธิ์ และแบบทดสอบวัดความถนัดนั้น ส่วนใหญ่แล้วจะเหมาะสม

กับผู้สอบที่มีความสามารถปานกลาง ดังนั้นความถูกต้องแม่นยำของการวัดผู้สอบที่มีความสามารถสูงและผู้สอบที่มีความสามารถต่ำจึงลดลง 3) ค่าความเชื่อมั่นของแบบทดสอบถูกนิยามในรูปของผลที่ได้จากการวัดซ้ำหรือใช้แบบทดสอบคู่ขนาน ซึ่งในการปฏิบัติจริงนั้นนับว่าเป็นเรื่องยากที่จะให้การสอบทั้งสองครั้งมีสภาพที่เหมือนกัน หรืออาจใช้แบบทดสอบคู่ขนานกันจริง แต่ผู้สอบจะมีลักษณะที่แตกต่างไปจากการสอบครั้งแรกเกี่ยวกับแรงจูงใจ ความกังวล การลืมหืมหรือการพัฒนาตนเองในบางทักษะ 4) ทฤษฎีการทดสอบแบบเดิมไม่สามารถบอกได้ว่าผู้สอบจะตอบข้อคำถามอย่างไร (ถูกหรือผิด) เว้นแต่ว่าจะใช้คำถามที่มีลักษณะคล้ายคลึงกับคำถามที่เคยใช้สอบมาก่อนแล้ว 5) ทฤษฎีการทดสอบแบบเดิมใช้ความแปรปรวนของความคลาดเคลื่อนในการวัดค่าเดียวกันกับผู้สอบทุกคน ซึ่งตามความเป็นจริงแล้วผู้สอบที่มีความสามารถสูงหรือต่ำจะมีค่าความแปรปรวนของความคลาดเคลื่อนในการวัดต่างจากผู้สอบที่มีความสามารถปานกลาง นอกเหนือจากข้อจำกัด 5 ประการดังกล่าวแล้ว ทฤษฎีมาตรฐานเดิมไม่สามารถตอบปัญหาเกี่ยวกับการทดสอบอีกหลายประการ เช่น การออกแบบการทดสอบที่เหมาะสมกับผู้สอบ การศึกษาขนาดความลำเอียงของข้อสอบ และการเทียบคะแนนสอบ (Equating of test scores) (Hambleton and Swaminathan, 1985 :1-4 ; Hambleton et al. 1991 : 2-5) ด้วยเหตุผลดังกล่าวข้างต้น นักพัฒนาแบบทดสอบรุ่นต่อมาจึงมีความพยายามที่จะอธิบายกระบวนการวัดในรูปของคุณลักษณะแฝง (Latent trait) และพัฒนาต่อมาจนกลายเป็นทฤษฎีการตอบสนองข้อสอบ (Item response theory ; IRT) ซึ่งมีข้อตกลงว่า การตอบข้อสอบเป็นฟังก์ชันของระดับคุณลักษณะในตัวผู้สอบและข้อสอบ และมุ่งเน้นถึงความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องโดยมีความคิดพื้นฐานว่า ถ้ามีระดับความสามารถสูงโอกาสในการตอบข้อสอบได้ถูกต้องจะสูงตามไปด้วย ซึ่งโอกาสความน่าจะเป็นนี้จะมีผลมาจากคุณลักษณะหรือพารามิเตอร์ของข้อสอบในรูปของฟังก์ชันการตอบสนอง (Item response functions ; IRF) ทางคณิตศาสตร์และเขียนแทนได้ด้วยเส้นแสดงคุณลักษณะ (Trace line) ที่เรียกว่าโค้งแสดงคุณลักษณะข้อสอบ (Item characteristic curve ; ICC) ซึ่งเป็นโมโนโทนสำคัญของ IRT เส้น ICC ตามทฤษฎี IRT มีลักษณะเป็นเส้นที่แสดงถึงการเพิ่มหรือลดในลักษณะทิศทางเดียว (Monotonically) ซึ่งจะบ่งบอกถึงความสัมพันธ์ระหว่างระดับความสามารถ (θ) ของผู้สอบและคุณลักษณะของข้อสอบกับโอกาสการตอบข้อสอบได้ถูกต้อง $P(\theta)$ หรืออาจกล่าวได้ว่าโอกาสที่ผู้สอบจะตอบข้อสอบได้ถูกนั้นขึ้นอยู่กับความสามารถของผู้สอบและคุณลักษณะของข้อสอบซึ่งอาจจะอธิบายด้วยพารามิเตอร์ของข้อสอบ 3 ค่า ได้แก่ ความยาก (b) ค่าอำนาจจำแนก (a) และโอกาสการได้คะแนนโดยบังเอิญ (Pseudo-chance score level) หรือพารามิเตอร์การเดาตอบ (Guessing parameter) ซึ่งแทนด้วยสัญลักษณ์ c การสร้างเส้น ICC สร้างได้จากฟังก์ชันลอจิสติก (Logistic function) หรือ

ฟังก์ชันลอจิสติก (Logistic function) และภายหลังต่อมาได้มีผู้นำเสนอค่าพารามิเตอร์ γ ซึ่งหมายถึงความไม่ถี่ถ้วนหรือความสะเพร่า (Careless) ในการตอบของผู้สอบ อย่างไรก็ตามค่าพารามิเตอร์นี้ ลอร์ดกล่าวว่าคงเป็นเพียงแนวคิดเชิงทฤษฎีเท่านั้นเพราะไม่สามารถพบได้ว่าเมื่อนำโมเดลนี้ไปใช้แล้วจะได้รับประโยชน์อะไรตามมา

นอกเหนือจากโมเดลที่ใช้กับการตอบแบบสองทาง (0-1) ดังที่กล่าวมาแล้ว ทฤษฎี IRT ยังสามารถนำไปประยุกต์ใช้กับการตอบแบบหลายทาง เช่น การให้คะแนนเป็นระดับจากมาตรฐานวัดแบบลิเคิร์ท การวัดเจตคติ การให้คะแนนแบบทดสอบอัตนัย รวมทั้งแบบทดสอบเลือกตอบที่มีการให้คะแนนแบบเรียงลำดับ หรือแบบไม่เรียงลำดับ ซึ่งจะพบในแบบทดสอบที่มีการพิจารณาเกี่ยวกับตัวลวงด้วย การให้คะแนนรูปแบบนี้ ได้แก่ โมเดลการตอบแบบนามบัญญัติ (Nominal response model) โมเดลการตอบแบบเรียงลำดับ (Ordinal response model) และโมเดลการตอบสนองแบบต่อเนื่อง (Continuous response model) นอกจากนี้ยังมีโมเดลการตอบในตระกูลราชคโมเดลอีก ได้แก่ โมเดลการให้คะแนนความรู้บางส่วน (Partial credit model) (Hambleton and Swaminathan. 1985 : 4-52 ; Andrich and Masters in Keeves. 1988 : 298-301 ; Masters in Keeves. 1988 : 292-293 ; Hambleton in Linn. 1989 : 147-160 ; Hambleton et al. 1991 : 12-31 ; Husen and Postlethwaite. 1994 : 409-414 ; Oosterhof. 1994 : 208-211)

การพิจารณาคุณภาพเครื่องมือตามกรอบทฤษฎี IRT นี้สามารถกระทำได้ทั้งรายข้อและรายฉบับเช่นเดียวกับ CTT โดยพิจารณาจากฟังก์ชันสารสนเทศของข้อสอบ (Item information function ; IIF) และฟังก์ชันสารสนเทศแบบทดสอบ (Test information function; TIF) หรือพิจารณาจากความคลาดเคลื่อนในการประมาณค่า (Standard error of estimation) ซึ่งเขียนแทนด้วย $SE(\theta)$ อย่างไรก็ตามแม้ว่าในปัจจุบัน การใช้ IRT ยังยอมรับข้อตกลงของความเป็นมิติเดียว แต่ถ้าโมเดลเหมาะสม (Fit) กับข้อมูลจะทำให้เกิดข้อได้เปรียบดังนี้คือ 1) ค่าพารามิเตอร์ไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบที่เป็นกลุ่มประชากรเดียวกัน 2) ระดับความสามารถของผู้สอบ ไม่ขึ้นอยู่กับชุดข้อสอบที่นำมาใช้กับผู้สอบ 3) ความแม่นยำของการประมาณคุณลักษณะหรือความสามารถหรือความคลาดเคลื่อนในการวัด สามารถแสดงให้เห็นในแต่ละระดับความสามารถซึ่งสิ่งนี้จะสามารถอธิบายแทนความเชื่อมั่นตามทฤษฎี CTT ได้ 4) มโนทัศน์เกี่ยวกับสารสนเทศข้อสอบหรือแบบทดสอบใช้เป็นเงื่อนไขในการชี้แนะสำหรับการสร้างแบบทดสอบ โดยแสดงให้เห็นถึงความเหมาะสมเป็นรายข้อ ซึ่งจะทำให้เกิดความแม่นยำในการวัดและจะเป็นพื้นฐานในการคัดเลือกข้อสอบ 5) คุณสมบัติของความไม่แปรเปลี่ยนของพารามิเตอร์ที่เป็นคุณลักษณะ หรือความสามารถจะเป็นประโยชน์ต่อการดำเนินงานสอบเมื่อข้อสอบและผู้สอบมีความแตกต่างกัน และสามารถนำไป

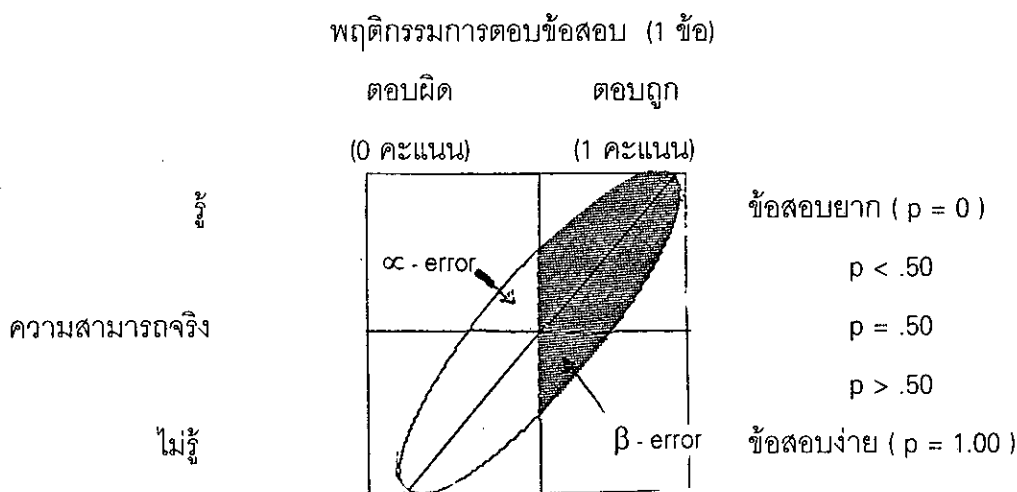
ปรับปรุงกระบวนการทดสอบให้เหมาะสมกับระดับของผู้สอบซึ่งการใช้วิธีการทดสอบแบบจัดข้อสอบให้เหมาะสมกับผู้สอบ (Adaptive testing procedure) นี้แสดงให้เห็นว่าสามารถลดเวลาการทดสอบลงไปได้อย่างน้อยร้อยละ 50 (Husen and Postlethwaite. 1994 : 3250) นอกเหนือจากนี้แล้ว ทฤษฎี IRT ยังสามารถนำไปประยุกต์ใช้เพื่ออธิบายปรากฏการณ์ต่าง ๆ เช่น การจัดข้อสอบเฉพาะบุคคล (Tailored testing) การสร้างโปรแกรมคอมพิวเตอร์เพื่อการทดสอบ (Computerized adaptive testing ; CAT) การตรวจสอบความรอบรู้ (Mastery testing) การเทียบคะแนน (Equating) การศึกษาความลำเอียงของข้อสอบ (Bias) ในรูปของ DIF (Differential item function)

แม้ว่าแนวทางและวิธีการของ CTT และ IRT ดูเหมือนว่าจะแตกต่างกันอย่างสิ้นเชิง แต่เมื่อพิจารณาในรายละเอียดจะพบว่า ทฤษฎี IRT พยายามจะอธิบายในสิ่งที่ข้อจำกัดของพารามิเตอร์ตามแนวของ CTT เดิม ได้แก่ คะแนนจริง ความยาก อำนาจจำแนก การเดา ความเชื่อมั่น ความเที่ยงตรงและความคลาดเคลื่อนในการวัด โดยที่ลอร์ดเองได้แสดงความเห็นในเชิงวิพากษ์วิจารณ์ถึงข้อจำกัดต่าง ๆ เหล่านั้น แต่มีโทษหมายความว่าสิ่งที่เขาอธิบายจะไปปลงล้างแนวคิดของ CTT ทั้งหมด ซึ่งเขาเองได้แสดงให้เห็นถึงความสัมพันธ์บางประการของคะแนนสอบ และค่าสถิติข้อสอบของ CTT และ IRT (Lord. 1980 : 3, 27-30) นอกจากนั้น ลอร์ลีย์ (1943) ได้แสดงความสัมพันธ์ระหว่างโมเดลนอร์มัลโอไบเควท์กับสัมประสิทธิ์ความเชื่อมั่นแบบวัดซ้ำ (r_{x1x2}) ซึ่งเขาแสดงให้เห็นว่า ค่า r_{x1x2} นี้ ประมาณค่าได้จากความแปรปรวนของคะแนนคลาดเคลื่อนและคะแนนสอบเมื่อการกระจายของคุณลักษณะเป็น $N(0, 1)$ กล่าวคือพารามิเตอร์ความยากของข้อสอบที่มีการให้คะแนนแบบ 0-1 จำนวน n ข้อ และมีการกระจายแบบปกติมีค่าเฉลี่ยเป็น 0 และความคลาดเคลื่อนเป็น 1 ในปี ค.ศ.1952 ลอร์ดได้แสดงให้เห็นว่าดัชนีการจำแนกของข้อสอบตาม CTT ณ ระดับคุณลักษณะที่กำหนดให้ (ซึ่งมีความสัมพันธ์มากกับสารสนเทศของแบบทดสอบ) มีความสัมพันธ์กับ r_{x1x2} ณ ระดับคุณลักษณะที่กำหนด เมื่อเส้นถดถอยของคะแนนสอบ ณ ระดับ 0 ได้รับการประมาณแบบเส้นตรง ต่อมาในปี 1977 ซาเมจิมาได้แสดงให้เห็นว่า r_{x1x2} สามารถประมาณค่าได้จากสารสนเทศของแบบทดสอบ (TIF) และการกระจายของคุณลักษณะของประชากรเป้าหมาย และในปี ค.ศ.1983 ลอร์ดได้อธิบายถึงตัวประมาณค่าที่ไม่ลำเอียงของ r_{x1x2} ที่ได้จากแบบทดสอบคู่ขนานด้วยโมเดลลอจิสติก 3 พารามิเตอร์ ซึ่งแนวคิดนี้นำมาสู่การใช้ TIF ใน IRT มาใช้อธิบายแทน r_{x1x2} และความคลาดเคลื่อนในการวัด (SEM) ตามมโนทัศน์ของ CTT ซึ่งซาเมจิมาได้ดัดแปลงสูตรของ TIF ที่ใช้ฟังก์ชันที่ลำเอียงของ MLE (Maximum likelihood estimation) ที่เรียกว่า MLEBF. ในการพยากรณ์ค่า r_{x1x2} และ SEM ตาม CTT ของกลุ่มผู้สอบ (Samejima. 1994 : 229-244 ; citing Lawley. 1943, Lord. 1952, Samejima. 1977b, Lord. 1983, Samejima. 1987, 1996a,

1993b) จากสิ่งที่ศึกษาดังกล่าวข้างต้นผู้วิจัยจึงคิดว่าถ้า TIF ครอบคลุมระดับความสามารถของผู้สอบได้ตั้งแต่ $-\infty$ ถึง $+\infty$ และมีรูปร่างเช่นเดียวกับโค้งปกติแล้ว ค่าความคลาดเคลื่อนในการประมาณค่าเฉลี่ย $SE(\theta)$ ของ IRT ก็คือสัมประสิทธิ์ความเชื่อมั่นที่คำนวณโดยสูตร KR-20 นั่นเอง ดังนั้น การประยุกต์ทฤษฎีทั้งสองมาใช้ในการพัฒนาแบบทดสอบ โดยใช้ทฤษฎี CTT ในส่วนของการสร้าง วิเคราะห์และคัดเลือกข้อสอบที่จะทำให้แบบทดสอบมีความเที่ยงตรงและความเชื่อมั่น และใช้ทฤษฎี IRT ประมาณค่าพารามิเตอร์ของคนและข้อสอบแยกจากกัน เพื่อนำไปประยุกต์ใช้กับการเทียบคะแนน การจัดข้อสอบให้เหมาะสมกับคนการจัดทำคลังข้อสอบรวมทั้งการใช้คอมพิวเตอร์เพื่อการทดสอบ (CAT) นับว่าเป็นสิ่งที่น่าสนใจทำการศึกษาต่อไป

กระบวนการวัดการทดสอบในประเทศไทยปัจจุบันไม่ว่าจะด้วยจุดประสงค์ใดก็ตามสิ่งที่ประจักษ์ให้เห็นส่วนใหญ่ก็คือ การใช้คะแนนสอบหรือคะแนนดิบ (Observed or raw score) ที่ได้จากการใช้แบบทดสอบเป็นดัชนีชี้บ่งถึงระดับความรู้ความสามารถของผู้สอบ ทั้งที่คะแนนดิบนี้ยังไม่เป็นที่ยอมรับของนักวัดผลว่ามีความบริสุทธิ์เพียงพอที่จะแทนความรู้ความสามารถที่แท้จริงของผู้สอบได้หรือไม่ โดยเฉพาะคะแนนสอบที่เกิดจากการใช้แบบทดสอบเลือกตอบ ทั้งตามกรอบทฤษฎีการวัดแบบ CTT และ IRT เนื่องจากนักวัดผลเชื่อว่าคะแนนดิบที่ได้นี้น่าจะมีส่วนของความคลาดเคลื่อนเจือปนอยู่ด้วย ตรงกับที่ที่ไม่สามารถแสดงหรือพิสูจน์ให้เห็นว่าในการวัดการทดสอบครั้งนั้นไม่เกิดความคลาดเคลื่อนซึ่งคาดว่ามันจะเกิดจากแหล่งใหญ่ ๆ 3 แหล่งคือ แบบทดสอบ ตัวผู้สอบ และวิธีดำเนินการสอบ อย่างไรก็ตามนักวัดผลส่วนใหญ่ก็ยังมองว่า ถ้าสร้างแบบทดสอบที่มีคุณภาพและมาตรฐานก็น่าจะช่วยลดความคลาดเคลื่อนในการวัดไปได้มาก ดังนั้น ผู้วิจัยจึงมีความสนใจที่จะพัฒนาแบบทดสอบโดยให้ความสำคัญในสองประเด็นสำคัญคือการให้คะแนน (Item scoring) และการคัดเลือกข้อสอบเข้าฉบับ (Item selecting)

รูปแบบการให้คะแนนแบบทดสอบเลือกตอบในปัจจุบันพบว่า ส่วนใหญ่ให้แบบถูก 1 ผิด 0 ซึ่งผู้วิจัยคิดว่าไม่น่าจะเหมาะสมนัก โดยเฉพาะอย่างยิ่งการนำคะแนนสอบไปใช้ในการตัดสินใจต่าง ๆ ทั้งนี้ เพราะการให้คะแนนแบบ 0-1 น่าจะทำให้เกิดความคลาดเคลื่อนในการวัดได้ง่าย และถ้าลองพิจารณาความสัมพันธ์ระหว่างความสามารถที่แท้จริงกับพฤติกรรมการตอบข้อสอบแบบเลือกตอบ (MC) พบว่ามีโอกาสเกิดความคลาดเคลื่อนได้ดังแสดงภาพประกอบ 1



ภาพประกอบ 1 ความสัมพันธ์ระหว่างความสามารถที่แท้จริงกับพฤติกรรมการตอบข้อสอบ
แบบ MC

จากภาพประกอบที่ 1 ในส่วนที่แรเงาแสดงถึงส่วนที่ได้คะแนนจากการตอบ ซึ่งเมื่อพิจารณาแล้วพบว่า มีคนอยู่ 2 กลุ่ม คือ กลุ่มรู้และกลุ่มไม่รู้ ในการทำข้อสอบถูกนั้นน่าจะมีเฉพาะกลุ่มรู้ เท่านั้น ดังนั้น การที่กลุ่มไม่รู้ตอบได้คะแนนเช่นเดียวกันกับกลุ่มรู้ถือว่าเป็นความคลาดเคลื่อนของการวัด ซึ่งเกี่ยวข้องโดยตรงกับเครื่องมือวัดหรือแบบทดสอบที่นำมาใช้วัด ในขณะที่เดียวกันการที่กลุ่มรู้ได้คะแนนศูนย์เช่นเดียวกับกลุ่มไม่รู้ก็ถือว่าเป็นความคลาดเคลื่อนของการวัดเช่นเดียวกัน และน่าจะมีความสำคัญกว่าประเด็นแรก ซึ่งในการวิจัยจัดว่าเป็นความคลาดเคลื่อนประเภทที่ 1 (α - error) ส่วนในประเด็นแรกคือความคลาดเคลื่อนประเภทที่ 2 (β - error) ดังนั้น ในความคิดของผู้วิจัยเห็นว่า รูปแบบการให้คะแนนน่าจะเป็นการให้คะแนนความรู้บางส่วน ซึ่งจากผลงานวิจัยพบว่า ทำให้เกิดการจำแนกบุคคลได้ถูกต้องแม่นยำ และแสดงดัชนีความเชื่อมั่นของแบบทดสอบสูงกว่าการให้คะแนนแบบ 0-1 และถึงแม้ว่าจะมีวิธีการให้คะแนนความรู้บางส่วนหลาย ๆ วิธีด้วยกัน แต่ผู้วิจัยเห็นว่าวิธีการต่าง ๆ เหล่านั้น ล้วนแต่มีข้อจำกัดโดยเฉพาะการให้คะแนนติดลบ ซึ่งผู้วิจัยไม่เห็นด้วยนัก ทั้งนี้ เพราะเราไม่อาจวินิจฉัยได้ชัดเจนว่าผู้ที่ตอบผิดเป็นเพราะเดาตอบหรือสับสนหรือเกิดความเข้าใจผิดหรือเป็นเพราะความบกพร่องของข้อสอบ และอีกประการหนึ่งน่าจะไม่สอดคล้องกับหลักการทางจิตวิทยาที่ว่าการศึกษาคือการงอกงาม นั่นคือผู้ที่ไม่น่าจะได้แค่ 0 แต่การให้คะแนนติดลบไม่น่าจะมีความหมายอะไรในการแปลความหมายคะแนนที่เป็นผลการทดสอบ นอกจากนั้นการให้ผู้สอบกำหนดคะแนนตามระดับความมั่นใจของ

ผู้ที่ไม่วางใจจะได้แค่ 0 แต่การให้คะแนนติดลบไม่น่าจะมีความหมายอะไรในการแปลความหมายคะแนนที่เป็นผลการทดสอบ นอกจากนั้นการให้ผู้สอบกำหนดคะแนนตามระดับความมั่นใจของตนเองน่าจะไม่ใช่วิธีที่เหมาะสมนัก ทั้งนี้ ขึ้นอยู่กับความกล้าของผู้ตอบเสียมากกว่า และเกณฑ์ความมั่นใจเกิดขึ้นในแต่ละบุคคล ทำให้การเปรียบเทียบคะแนนดูเหมือนว่าจะอยู่ต่างสเกลกัน ส่วนวิธีของสำรวจนั้น สูตรในการประมาณค่าคะแนนจำกัดอยู่เฉพาะข้อสอบที่มี 4 ตัวเลือกเท่านั้น สำหรับวิธีของเกริกชัยจำเป็นต้องสร้างเครื่องมือบันทึกคำตอบโดยเฉพาะ ซึ่งนำไปสู่ความไม่สะดวกในการนำไปใช้ทดสอบจริงกับคนกลุ่มใหญ่ ๆ และรูปแบบสุดท้ายคือ ให้ตอบว่าถูกหรือผิด แม้ว่าจะช่วยป้องกันการเดาตอบได้ดีกว่าวิธีอื่น แต่ค่อนข้างยากในการสร้างตัวเลือกให้มีคำตอบถูกมากกว่า 1 คำตอบขึ้นไป โดยแต่ละตัวเลือกจะต้องเป็นอิสระต่อกัน นอกจากนั้นยังใช้เวลาทดสอบค่อนข้างยาวนานกว่าแบบอื่น ๆ กล่าวคือผู้สอบจะต้องใช้เวลาพิจารณาทุกตัวเลือก นอกจากนั้น การใช้แบบทดสอบเลือกตอบลำดับขั้น (SMC) ปัญหาที่เกิดขึ้นก็คือ การกำหนดน้ำหนักแต่ละตัวเลือก เช่นเดียวกับแบบทดสอบเลือกตอบปกติแบบกำหนดน้ำหนักคะแนนในแต่ละตัวเลือก จากข้อจำกัดของการให้คะแนนแบบปกติ 0-1 วิธีปรับแก้การเดา และวิธีให้คะแนนบางส่วนดังกล่าว ผู้วิจัยจึงคิดวิธีการตอบและให้คะแนนโดยเน้นความรู้บางส่วน และความมั่นใจในการตอบของผู้สอบที่เรียกว่า การตอบ 2 ขั้นตอน (Two step response or 2-SR) และเรียกแบบทดสอบที่ใช้ลักษณะการตอบแบบนี้ว่าแบบทดสอบเลือกตอบชนิดตอบ 2 ขั้นตอน (Two step response multiple choice test or 2-SRMC) โดยที่แบบทดสอบชนิดนี้ ผู้วิจัยตั้งใจที่จะทำให้เกิดการพัฒนาแบบทดสอบเลือกตอบเพื่อนำมาใช้ในการวัดความสามารถที่แท้จริงของผู้ถูกวัด โดยใช้รูปแบบการตอบและการสร้างแบบทดสอบเลือกตอบที่ไม่ต่างไปจากที่ใช้กันอยู่มากนัก แต่มีการปรับวิธีการตอบโดยให้ผู้สอบตอบ 2 ขั้นตอน คือ ขั้นตอนที่หนึ่งให้เลือกคำตอบที่คิดว่าเป็นข้อถูกโดยไม่จำกัดจำนวน และขั้นตอนที่สองให้เลือกคำตอบที่คิดว่าถูกที่สุดจากตัวเลือกที่คัดเลือกไว้ในขั้นตอนที่หนึ่งเพียงข้อเดียว ซึ่งจากแนวความคิดนี้น่าจะทำให้คะแนนที่เป็นผลการวัดนั้นบ่งบอกถึงความสามารถที่แท้จริงได้มากที่สุด ทั้งนี้เพราะว่า

1. การให้ตอบ 2 ครั้ง เพื่อต้องการตรวจสอบความมั่นใจ และความคงที่ในการตอบ
2. การเปิดโอกาสให้ผู้สอบเลือกคำตอบได้โดยไม่จำกัดจำนวน ในขั้นตอนแรกเป็นเสมือนการแสดงการคิดและตัดสินใจของผู้สอบขณะตัดสินใจเลือกตัวเลือก ซึ่งจะเป็นการฝึกกระบวนการคิดและตัดสินใจ และจะทำให้สามารถจำแนกความรอบรู้ของผู้ตอบได้ นอกจากนั้นยังทำให้ผู้สอบไม่ลำบากใจที่จะต้องตัดสินใจเลือกคำตอบอันจะนำไปสู่การเดาตอบแบบสุ่ม (Blind guessing) ให้

น้อยที่สุดหรืออาจกล่าวได้ว่าถ้าผู้ตอบจะเดาตอบก็ควรเดาตอบแบบรู้เป็นบางส่วน (Informed guessing) มากกว่า

นอกจากนั้นผลการวัดที่ได้จะเป็นประโยชน์อย่างมากในการนำมาใช้เป็นข้อมูลสำหรับวิเคราะห์คุณภาพของแบบทดสอบ ทั้งตามแนวของทฤษฎี CTT และ IRT จากแนวคิดดังกล่าว ผู้วิจัยจึงได้กำหนดกรอบการพัฒนาแบบทดสอบเลือกตอบไว้ดังนี้

1. โครงสร้างแบบทดสอบเช่นเดียวกับแบบทดสอบเลือกตอบปกติ คือมีส่วนที่เป็นคำถาม และส่วนที่เป็นตัวเลือกที่มีคำตอบถูกหรือถูกที่สุดเพียงข้อเดียว
2. เปิดโอกาสให้ผู้ตอบเลือกตัวเลือกอย่างเสรีตามความรู้ ความสามารถของผู้ตอบ โดยเชื่อว่าผู้ที่มีความรู้ความสามารถจะเลือกตอบน้อยกว่าผู้ที่มีความสามารถต่ำกว่า และผู้ที่ยังมีความคลุมเครือสูงน่าจะเลือกตอบมากกว่าผู้ที่มีความคลุมเครือน้อยกว่า
3. ผู้ที่มีความรู้ความสามารถสูงต้องได้คะแนนมากกว่าผู้ที่มีความสามารถต่ำกว่าและผู้ที่ไม่ตอบ และจากกรอบความคิดดังกล่าว ผู้วิจัยจึงกำหนดโครงสร้าง วิธีการตอบและให้คะแนนแบบทดสอบเลือกตอบชนิด 2 ขั้นตอน ดังนี้

3.1 แบบทดสอบมีวิธีการสร้างเช่นเดียวกับแบบทดสอบเลือกตอบหลายตัวเลือกปกติ ที่มีคำตอบถูกเพียงคำตอบเดียว

3.2 กระดาษคำตอบจัดแบ่งเป็น 2 ส่วน ส่วนที่หนึ่งสำหรับการตอบโดยเสรี และส่วนที่สองให้ผู้ตอบเลือกคำตอบจากส่วนที่หนึ่งที่เห็นว่าถูกต้องที่สุดเพียงตัวเลือกเดียว

3.3 ตรวจให้คะแนนการตอบทั้ง 2 ส่วน และหาคะแนนรวมรายชื่อจากการแทนค่าลงในสูตรที่ผู้วิจัยสร้างขึ้น เพื่อกระจายสเกลการวัดแบบ 0-1 ให้วัดได้ละเอียดมากขึ้นในลักษณะของการให้คะแนนความรู้เป็นส่วน ซึ่งจะทำให้เกิดข้อได้เปรียบดังนี้

3.3.1 ข้อสอบแต่ละข้อจะมีการกระจายของคะแนนตั้งแต่ 0-8 ในกรณี 5 ตัวเลือก หรือ 0-6 ในกรณี 4 ตัวเลือก หรือ 0-4 ในกรณี 3 ตัวเลือก ทั้งนี้ จะเห็นได้ว่าค่าคะแนนสูงสุดรายข้อขึ้นอยู่กับจำนวนตัวเลือก ซึ่งน่าจะสอดคล้องกับหลักการที่ว่าข้อสอบที่มีตัวเลือกมากน่าจะมีแนวโน้มยากกว่าข้อสอบที่มีจำนวนตัวเลือกน้อยกว่า ดังนั้น ข้อสอบที่มีลักษณะที่ยากต่อการตอบมากกว่าน้ำหนักคะแนนก็น่าจะมากกว่า

3.3.2 ผู้มีความสามารถสูงจะได้คะแนนมากกว่าผู้ที่มีความสามารถต่ำกว่า เพราะผู้ที่มีความสามารถสูง มีแนวโน้มจะเลือกตัวเลือกในส่วนที่หนึ่งน้อยกว่าผู้ที่มีความสามารถต่ำและจำนวนตัวเลือกในส่วนที่ 1 แสดงถึงความมั่นใจของผู้ตอบ ถ้ามั่นใจมากจำนวนตัวเลือกจะน้อยกว่าผู้ที่มั่นใจน้อย หรือยังมีความคลุมเครือ แม้ว่าในส่วนที่ 2 จะตอบถูกเหมือนกัน

3.3.3 ผู้มีความรู้บางส่วนมีโอกาสได้คะแนนมากกว่าแบบถูก 1 ผิด 0 ในกรณีที่เกิดความคลุมเครือ หรือลังเลในการตัดสินใจ ถ้าเป็นแบบถูก 1 ผิด 0 การตัดสินใจนั้นน่ามีความสำคัญมาก ถ้าตัดสินใจถูกจะได้ 1 คะแนน ถ้าตัดสินใจผิดจะไม่มีคะแนนเลย แต่การตอบแบบ 2 ขั้นตอนนี้ ในส่วนที่ 1 มีโอกาสได้คะแนน ถ้าเลือกคำตอบที่มีตัวถูกอยู่ด้วย แม้ว่าในส่วนที่ 2 จะตัดสินใจผิดก็ตาม

3.3.4 จำแนกความรู้ของผู้ตอบได้ดีกว่าแบบ 0-1 ซึ่งจำแนกคนได้ 2 กลุ่ม คือ ตอบถูกได้คะแนน ซึ่งในความเป็นจริงแล้วอาจรวมทั้งคนที่รู้จริง รู้เป็นบางส่วนแต่ตัดสินใจถูกหรือพวกที่ไม่รู้แต่เดาถูก และกรณีได้ 0 อาจรวมทั้งคนที่ไม่รู้ หรือรู้บางส่วนที่ตัดสินใจผิดพลาดหรือเข้าใจข้อความของข้อสอบผิด แต่ลักษณะการตอบแบบ 2 ขั้นตอน ผู้ที่รู้จริงเท่านั้นมี โอกาสได้คะแนนเต็ม ผู้ที่มีความรู้บางส่วนก็จะได้คะแนนลดหลั่นกันลงมา จนถึงผู้ที่ไม่รู้หรือผู้เดาตอบมีโอกาสได้คะแนน 0 มาก หรืออาจจะมีพวกที่ไม่รู้เลยแต่ชอบเสี่ยงโดยการเดาทั้งในส่วนที่ 1 และส่วนที่ 2 โดยกาเครื่องหมายให้ตรงกันทั้ง 2 ส่วน ถ้าถูกก็จะได้คะแนนสูงสุด ถ้าผิดก็จะได้ 0 เหตุการณ์ในลักษณะเช่นนี้อาจมีน้อยมากหรือถ้าเกิดขึ้นจริงก็คงจะไม่เกิดขึ้นกับทุกข้อและมีโอกาสที่จะได้คะแนนสูงสุดเพียง 100/K % เท่านั้น ซึ่งจะทำให้มีโอกาสได้คะแนนน้อยกว่าผู้ที่ตอบด้วยความรู้จริง หรือมีความรู้เพียงบางส่วนมาก ในกรณีเช่นนี้วิธีป้องกันก็คือ ไม่บอกวิธีการตรวจให้คะแนนกับผู้สอบอย่างละเอียด อาจบอกเพียงหลักใหญ่ ๆ เช่น

ให้ตอบในส่วนที่ 1 อย่างเสรี แต่ไม่ควรตอบทุกตัวเลย เพราะจะทำให้ได้ 0

ตัวเลือกในส่วนที่ 2 จะต้องเป็นตัวเลือกที่เลือกไว้ในส่วนที่ 1 มิฉะนั้นจะได้ 0

และจากช่วงของคะแนนที่มากขึ้น วิธีการตรวจแบบนี้จะทำให้ค่าความยากของข้อสอบเปลี่ยนไป ค่าอำนาจจำแนก ค่าความเชื่อมั่น ค่าความเที่ยงตรง สูงขึ้น และค่าความคลาดเคลื่อนในการวัดต่ำกว่าแบบทดสอบเลือกตอบปกติ

นอกจากประเด็นของการให้คะแนนดังกล่าวแล้ว กระบวนการคัดเลือกข้อสอบเข้าฉบับ นับว่าเป็นประเด็นหนึ่งที่น่าสนใจ การใช้แบบทดสอบในปัจจุบันส่วนใหญ่ พบว่าไม่ค่อยมีการวิเคราะห์เชิงสถิติ จึงเกิดความคลงแคลงใจต่อคุณภาพของเครื่องมือที่จะนำไปใช้ เพื่อให้ได้ผลการวัดสำหรับการตัดสินใจตามจุดมุ่งหมายต่างๆ โดยเฉพาะการคัดเลือกหรือเลื่อนขั้น แบบทดสอบที่ไม่มีคุณภาพย่อมทำให้ผลการสอบไม่มีคุณภาพ ขาดความเชื่อถือ และนำมาสู่ความคลาดเคลื่อนในการวัดและตัดสินใจ อย่างไรก็ตาม การสร้างแบบทดสอบในปัจจุบันยังมีการยึดถือตารางวิเคราะห์หลักสูตรหรือกรอบสมรรถภาพเป็นเบื้องต้น ดังนั้น จึงทำให้แบบทดสอบอย่างน้อยก็มีคุณภาพในเรื่องของความเที่ยงตรง ซึ่งอาจจะเป็นความเที่ยงตรงเชิงเนื้อหา ความเที่ยงตรงเชิง

โครงสร้าง หรือความเที่ยงตรงเชิงประจักษ์ และประเด็นสำคัญสิ่งที่ผู้วิจัยสนใจศึกษาในครั้งนี้ก็คือ ถ้ามีการคัดเลือกข้อสอบเข้าฉบับโดยผ่านกระบวนการวิเคราะห์ข้อสอบ ที่พิจารณาจากค่าความยาก และค่าอำนาจจำแนก ตามแนวทฤษฎี CTT หรือพิจารณาจากความเหมาะสมกับโมเดล (Model fit) หรือค่าความยาก (b) ตามแนวของ IRT (ราชคโโมเดล) จะทำให้ข้อสอบมีคุณภาพสูงขึ้นเพียงไร ถ้าเพิ่มตัวแปรความเป็นมิติเดียว ซึ่งเป็นข้อตกลงสำคัญของ IRT ซึ่งมีคุณลักษณะคล้ายกับความเป็นเอกพันธ์ของข้อสอบตามแนว CTT และความลำเอียงของข้อสอบซึ่งสัมพันธ์กับความเป็นมิติเดียวที่เกี่ยวข้องกับความเป็นวิวิธพันธ์ของกลุ่มผู้สอบตามแนวของ CTT เป็นเกณฑ์ในการคัดเลือกข้อสอบจะทำให้คุณภาพของแบบทดสอบตามแนวของ CTT และ IRT สูงขึ้นหรือไม่ และถ้าจะผสมผสานแนวทฤษฎีทั้งสองเข้าด้วยกัน โดยใช้หลักการของ CTT ในการวิเคราะห์ข้อสอบ และคัดเลือกข้อสอบเป็นเบื้องต้น สำหรับการประมาณค่าพารามิเตอร์ของผู้สอบและข้อสอบแยกจากกันตามแนวของ IRT จะทำให้ความคลาดเคลื่อนในการประมาณค่าต่ำลงหรือไม่อย่างไร ซึ่งจากประเด็นการศึกษาทั้งหมดข้างต้นน่าจะทำได้รูปแบบสมการโครงสร้างที่เหมาะสมที่สุดสำหรับการพัฒนาแบบทดสอบเลือกตอบตามแนวของ CTT และ IRT เพื่อให้ได้ผลการวัดที่คลาดเคลื่อนน้อยที่สุดสำหรับการใช้ประโยชน์ต่อไป

วัตถุประสงค์ของการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์สำคัญเพื่อศึกษาอิทธิพลของการให้คะแนนและการคัดเลือกข้อสอบเพื่อกำหนดรูปแบบสมการโครงสร้างที่เหมาะสมที่สุดในการพัฒนาแบบทดสอบเลือกตอบ โดยมีวัตถุประสงค์เฉพาะดังนี้

1. เพื่อศึกษาอิทธิพลของรูปแบบการให้คะแนน และวิธีการคัดเลือกข้อสอบ ที่มีต่อคุณภาพของแบบทดสอบเลือกตอบภายใต้กรอบทฤษฎี CTT
2. เพื่อศึกษาอิทธิพลของรูปแบบการให้คะแนน และวิธีการคัดเลือกข้อสอบ ที่มีต่อคุณภาพของแบบทดสอบเลือกตอบภายใต้กรอบทฤษฎี IRT (ราชคโโมเดล)
3. เพื่อศึกษาอิทธิพลของรูปแบบการให้คะแนน วิธีการคัดเลือกข้อสอบและคุณภาพข้อสอบตามกรอบทฤษฎีของ CTT ที่มีต่อคุณภาพของแบบทดสอบเลือกตอบตามแนวของ IRT (ราชคโโมเดล)
4. เพื่อศึกษารูปแบบความสัมพันธ์เชิงสาเหตุของการให้คะแนน และวิธีการคัดเลือกข้อสอบ ที่มีต่อคุณภาพของแบบทดสอบตามกรอบทฤษฎีของ CTT และ IRT (ราชคโโมเดล)

สมมติฐานการวิจัย

1. รูปแบบการให้คะแนนและวิธีการคัดเลือกข้อสอบมีอิทธิพลต่อคุณภาพของแบบทดสอบเลือกตอบตามกรอบทฤษฎี CTT ซึ่งมีสมมติฐานเฉพาะคือ

1.1 รูปแบบการให้คะแนนแบบ 2-SR ทำให้คะแนนสอบ ค่าความยาก ค่าอำนาจจำแนก ความเป็นมิติเดียว ความลำเอียงเนื่องจากตัวแปรเพศ ความเชื่อมั่นของข้อสอบสูงขึ้นกว่าการให้คะแนนแบบปกติ

1.2 รูปแบบการให้คะแนนแบบ 2-SR ทำให้ค่าความลำเอียงเนื่องจากตัวแปรเพศของข้อสอบ ความคลาดเคลื่อนในการวัดลดลงจากการให้คะแนนตามแบบปกติ

1.3 รูปแบบการให้คะแนนแบบ 2-SR ทำให้ค่าความเชื่อมั่นของแบบทดสอบสูงกว่าการให้คะแนนแบบปกติ

1.4 การคัดเลือกข้อสอบจากตัวแปรที่ได้จากการให้คะแนนตามรูปแบบ 2-SR ทำให้ค่าความเชื่อมั่นของแบบทดสอบสูงขึ้นกว่าการให้คะแนนแบบปกติ และทำให้ค่าความคลาดเคลื่อนในการวัดต่ำกว่าการให้คะแนนแบบปกติ

2. รูปแบบการให้คะแนน และวิธีการคัดเลือกข้อสอบมีอิทธิพลต่อคุณภาพของแบบทดสอบเลือกตอบตามกรอบทฤษฎี IRT ซึ่งมีสมมติฐานเฉพาะคือ

2.1 รูปแบบการให้คะแนนแบบ 2-SR ทำให้ระดับความสามารถ ความเหมาะสมของข้อสอบกับโมเดลและสารสนเทศของข้อสอบสูงกว่าการให้คะแนนตามรูปแบบปกติ

2.2 รูปแบบการให้คะแนนแบบ 2-SR ให้ค่าสารสนเทศของแบบทดสอบที่มีประสิทธิภาพสูงกว่าการให้คะแนนตามรูปแบบปกติ

2.3 รูปแบบการให้คะแนนแบบ 2-SR ให้ค่าความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ข้อสอบต่ำกว่าการให้คะแนนแบบรูปแบบปกติ

2.4 ข้อสอบที่ได้จากการคัดเลือกตามเกณฑ์ความเป็นมิติเดียว ความลำเอียงเนื่องจากตัวแปรเพศ ความยาก (β) และความเหมาะสมของข้อสอบกับโมเดลที่ให้คะแนนตามรูปแบบ 2-SR และแบบปกติ ให้ค่าสารสนเทศของข้อสอบสูงกว่าข้อสอบที่ถูกคัดออกจากเกณฑ์การคัดเลือกคือ ความเป็นมิติเดียว ความลำเอียงเนื่องจากตัวแปรเพศ ความยาก (β) และความเหมาะสมกับโมเดล

3. รูปแบบการให้คะแนน และวิธีการคัดเลือกข้อสอบ และคุณภาพข้อสอบตามกรอบทฤษฎี CTT มีอิทธิพลต่อคุณภาพของแบบทดสอบตามกรอบทฤษฎี IRT ซึ่งมีจุดประสงค์เฉพาะคือ

- ข้อสอบที่ได้จากการคัดเลือกตามเกณฑ์ความยากแบบ CTT ค่าอำนาจจำแนกที่ให้คะแนนตามรูปแบบ 2-SR และรูปแบบปกติให้ค่าสารสนเทศของข้อสอบสูงกว่าข้อสอบที่ถูกคัดออกจากฉบับโดยเกณฑ์ค่าความยากแบบ CTT และค่าอำนาจจำแนก

4. รูปแบบการให้คะแนน และวิธีการคัดเลือกข้อสอบมีความสัมพันธ์เชิงสาเหตุกับคุณภาพแบบทดสอบเลือกตอบ ตามกรอบทฤษฎี CTT และ IRT ซึ่งมีสมมติฐานเฉพาะคือ

4.1 แผนภาพเส้นทางแสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่าง ๆ ตามกรอบทฤษฎี CTT ที่ได้จากการให้คะแนนตามรูปแบบ 2-SR และแบบปกติให้ค่าสัมประสิทธิ์พยากรณ์ในการอธิบายความแปรปรวนของความเชื่อมั่นของข้อสอบของแบบทดสอบสูงกว่าการใช้สมการถดถอยพหุคูณ

4.2 แผนภาพเส้นทางแสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่าง ๆ ตามกรอบทฤษฎี IRT ที่ได้จากการให้คะแนนตามรูปแบบ 2-SR และแบบปกติให้ค่าสัมประสิทธิ์พยากรณ์ในการอธิบายความแปรปรวนของสารสนเทศของข้อสอบของแบบทดสอบสูงกว่าการใช้สมการถดถอยพหุคูณ

4.3 แผนภาพเส้นทางแสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่าง ๆ ตามกรอบทฤษฎี CTT ที่ได้จากการให้คะแนนตามรูปแบบ 2-SR และแบบปกติให้ค่าสัมประสิทธิ์พยากรณ์ในการอธิบายความแปรปรวนของสารสนเทศของข้อสอบของแบบทดสอบสูงกว่าการใช้สมการถดถอยพหุคูณ

ขอบเขตและความจำกัดของการวิจัย

การวิจัยครั้งนี้ ต้องการศึกษารูปแบบสมการโครงสร้าง เพื่อเสนอให้ผู้ที่เกี่ยวข้องกับการใช้แบบทดสอบเลือกตอบนำไปเป็นแนวทางในการพัฒนาแบบทดสอบของตน ทั้งตามกรอบทฤษฎีแบบ CTT และ IRT ในทุกชั้น ทุกระดับ และทุกสาขาวิชาที่สามารถสร้างเป็นแบบทดสอบเลือกตอบได้ แต่เนื่องจากมีข้อจำกัดบางประการเกี่ยวกับระยะเวลา ผู้วิจัยจึงเน้นศึกษากับแบบทดสอบในระดับชั้นประถมศึกษา โดยจะเลือกออกข้อสอบตามสมรรถภาพของกลุ่มทักษะที่สำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ ได้กำหนดไว้ซึ่งถือว่าเป็นกลุ่มประสบการณ์สำคัญที่คาดว่าจะมีกระบวนการคิดที่แตกต่างกัน ได้แก่ ความสามารถในการอ่านจับใจความ และความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ ดังนั้น เครื่องมือที่ใช้ในการศึกษาวิจัยครั้งนี้ ประกอบด้วยแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนกลุ่มทักษะระดับชั้นประถมศึกษา จำนวน 2 ฉบับ แต่ละฉบับมีข้อสอบ จำนวน 40 ข้อ วัดความรู้ความสามารถในการอ่านจับใจความ และการแก้โจทย์

ปัญหาทางคณิตศาสตร์

สำหรับกลุ่มตัวอย่างเป้าหมายที่ทำการศึกษา ได้แก่ นักเรียนระดับชั้นประถมศึกษาปีที่ 6 ในสังกัดสำนักงานคณะกรรมการการประถมศึกษาแห่งชาติซึ่งเป็นระดับสุดท้ายของระดับชั้นประถมศึกษาโดยสุ่มกลุ่มตัวอย่างในทุกเขตการศึกษา จำนวนประมาณ 1,200 คน

และสำหรับตัวแปรสำคัญในการศึกษาวิจัยครั้งนี้ ได้แก่ 1) วิธีการให้คะแนน ซึ่งมี 2 วิธี ได้แก่ การให้คะแนนแบบ ปกติ (0-1) และวิธีการให้คะแนนแบบ 2-SR 2) วิธีการคัดเลือกข้อสอบเข้าฉบับ มีหลักเกณฑ์พิจารณาดังนี้ คือ

1. การคัดเลือกข้อสอบตามกรอบทฤษฎี CTT ได้แก่ การพิจารณาค่าความยาก ค่าอำนาจจำแนกรายข้อ ความเป็นมิติเดียว และความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ

2. การคัดเลือกข้อสอบตามกรอบทฤษฎี IRT แบบราชคโหมเดล ได้แก่ การพิจารณาค่าความยาก (b) ความเหมาะสมกับโมเดล (Item fit) ความเป็นมิติเดียว และความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ

3. คุณภาพของข้อสอบ

3.1 ตามกรอบ CTT ได้แก่ ค่าความเชื่อมั่นรายข้อ และรายฉบับ

3.2 ตามกรอบ IRT ได้แก่ ค่าสารสนเทศของข้อสอบและแบบทดสอบ (Item and test information function)

ความสำคัญของการวิจัย

การศึกษาวิจัยในครั้งนี้ คาดว่าจะเป็นประโยชน์ต่อนักพัฒนาแบบทดสอบที่จะได้เลือกใช้รูปแบบของกระบวนการพัฒนาที่เหมาะสมกับจุดประสงค์ที่กำหนด และความรู้ความสามารถของผู้ที่เกี่ยวข้องกับการวัด นอกจากนั้นผลของการศึกษาวิจัยจะทำให้ทราบถึงประเด็นสำคัญที่เกิดประโยชน์ต่อการกำหนดรูปแบบที่เหมาะสมในการพัฒนาคุณภาพแบบทดสอบ ได้แก่

1. ความสัมพันธ์ระหว่างการให้คะแนนกับวิธีการคัดเลือกข้อสอบ และคุณภาพของแบบทดสอบตามกรอบทฤษฎี CTT และ IRT

2. ความสัมพันธ์ระหว่างค่าพารามิเตอร์ของข้อสอบกับความเป็นมิติเดียว และความลำเอียงของข้อสอบ

3. ความสัมพันธ์ระหว่างค่าพารามิเตอร์ของข้อสอบ ตามแนวทฤษฎี CTT และ IRT (ราชคโหมเดล)

4. แนวทางในการประยุกต์ทฤษฎี CTT ใช้อย่างเหมาะสม

คำนิยามศัพท์เฉพาะ

1. การศึกษาอิทธิพลของการให้คะแนนและวิธีการคัดเลือกข้อสอบ หมายถึง การศึกษาใน 2 ลักษณะคือ

1.1 การศึกษาความเปลี่ยนแปลงของคะแนนสอบ ค่าพารามิเตอร์ของข้อสอบและแบบทดสอบ ที่ถูกคัดเลือกโดยเกณฑ์ต่าง ๆ ที่ให้คะแนนตามรูปแบบ 2-SR เปรียบเทียบกับการให้คะแนนตามรูปแบบปกติ และ

1.2 การศึกษาการส่งผลต่อตัวแปรเกณฑ์ของตัวแปรอิสระที่ใช้เป็นเกณฑ์คัดเลือกข้อสอบที่ได้จากการให้คะแนนแบบ 2-SR และแบบปกติจากเทคนิคสมการถดถอยพหุคูณ และแผนภาพเส้นทางแสดงความสัมพันธ์เชิงสาเหตุ

2. รูปแบบการให้คะแนน หมายถึง การกำหนดคะแนนการตอบถูกหรือผิดของผู้สอบ มี 2 รูปแบบ คือ

2.1 แบบปกติ ได้แก่ การให้คะแนนแบบถูก 1 ผิด 0 ซึ่งได้จากการตรวจการตอบในส่วนที่สองของกระดาษคำตอบ

2.2 แบบ 2-SR ได้แก่ การให้คะแนนแบบความรู้บางส่วนซึ่งคำนวณจากสูตรที่ผู้วิจัยกำหนดขึ้น

3. การพัฒนาแบบทดสอบเลือกตอบ หมายถึง การใช้รูปแบบการให้คะแนนและกระบวนการคัดเลือกข้อสอบที่แตกต่างไปจากวิธีการปฏิบัติตามปกติ

4. วิธีการคัดเลือกข้อสอบ หมายถึง กระบวนการในการคัดเลือกข้อสอบเข้าฉบับโดยใช้พารามิเตอร์ของข้อสอบตามกรอบทฤษฎี CTT ได้แก่ ค่าความยาก และอำนาจจำแนก ซึ่งมี 2 ค่าคือสัมประสิทธิ์การจำแนก (Discriminant coefficient) และดัชนีการจำแนก (Index of discriminant) ตามกรอบทฤษฎี IRT ได้แก่ ค่าความยาก (β) และความเหมาะสมของข้อสอบกับโมเดล (Model-item fit) และคุณสมบัติบางประการ ได้แก่ ความลำเอียงเนื่องจากตัวแปรเพศ และความเป็นมิติเดียวของข้อสอบ

5. ความเป็นมิติเดียวของแบบทดสอบ หมายถึง กลุ่มของข้อสอบที่อยู่ในองค์ประกอบเดียวกัน โดยพิจารณาจากค่าน้ำหนักองค์ประกอบ (Factor loading) ตั้งแต่ .30 ขึ้นไปขององค์ประกอบที่ 1

6. ความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ หมายถึง คุณสมบัติของข้อสอบที่เอื้อประโยชน์ต่อเพศใดเพศหนึ่ง ซึ่งพิจารณาจากแบบแผนการตอบในแต่ละเพศโดยพิจารณาจากค่าไคสแควร์ที่แสดงความแตกต่างของ Likelihood ratio อย่างมีนัยสำคัญทางสถิติตั้งแต่ .05 ขึ้นไป

7. ความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ หมายถึง ความสามารถในการคำนวณหาคำตอบของข้อคำถามทางคณิตศาสตร์เกี่ยวกับเนื้อหาที่เป็นสถานการณ์ต่าง ๆ โดยใช้ทักษะพื้นฐานของการบวก ลบ คูณ หาร และกระบวนการทางคณิตศาสตร์ (จำแนก จัดกลุ่ม หาความสัมพันธ์ สรุปข้อมูล)

8. ความสามารถในการอ่านจับใจความ หมายถึง ความสามารถในการจับใจความสำคัญ แปลความ ตีความ ขยายความ การวิเคราะห์ (จำแนก เปรียบเทียบ ให้เหตุผล) และ ประเมินค่า (ประเมิน วินิจฉัยเกณฑ์ วิพากษ์วิจารณ์)



เอกสารและงานวิจัยที่เกี่ยวข้อง

ในการวิจัยครั้งนี้ ผู้วิจัยศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการพัฒนาแบบทดสอบเลือกตอบซึ่งนำเสนอโดยสรุปเรียงลำดับตามหัวข้อต่อไปนี้

1. มโนทัศน์เกี่ยวกับทฤษฎีพื้นฐานในการพัฒนาแบบทดสอบเลือกตอบ
 - 1.1 ทฤษฎีมาตรฐานเดิม (Classical test theory ; CTT)
 - 1.2 ทฤษฎีการสรุปอ้างอิง (Generalizability theory ; G-Theory ; GT)
 - 1.3 ทฤษฎีการตอบสนองข้อสอบ (Item response theory ; IRT)
2. มโนทัศน์เกี่ยวกับกระบวนการพัฒนาแบบทดสอบเลือกตอบ
3. งานวิจัยที่เกี่ยวข้องกับการพัฒนาแบบทดสอบเลือกตอบ

มโนทัศน์เกี่ยวกับทฤษฎีพื้นฐานในการพัฒนาแบบทดสอบเลือกตอบ

ทฤษฎีมาตรฐานเดิม (CTT)

ในช่วงระหว่างปี 1890-1901 แคทเทล (Cattell) และลูกศิษย์ของเขา คือ คลาร์ก ไวส์เลอร์ (Clark Wissler) แห่งมหาวิทยาลัยโคลัมเบีย (Columbia University) ได้พยายามนำคะแนนจากการวัดด้วยแบบทดสอบทางจิตวิทยาไปหาความสัมพันธ์กับระดับคะแนน (Grades) ของวิชาต่าง ๆ และพบว่า ค่าความสัมพันธ์มีค่าอยู่รอบ ๆ ศูนย์ โดยมีค่าสูงสุดเท่ากับ .19 ต่อมาในปี 1894 กิลเบิร์ต (Gilbert) แห่งมหาวิทยาลัยเยล (Yale University) ได้ศึกษาในลักษณะเดียวกัน และได้ผลเช่นเดียวกัน

สำหรับการวัดความแตกต่างรายบุคคลที่จัดว่าเชื่อถือได้ในลักษณะทางวิทยาศาสตร์ น่าจะเริ่มต้นจากมาตรวัดของบิเน็ต (Binet) และไซมอน (Simon) และเทคนิคเชิงปริมาณที่ได้รับการพัฒนาโดยคาร์ล เพียร์สัน (Karl Pearson) และ ชาร์ล สเปียร์แมน (Charles Spearman) ในราว ๆ ต้นศตวรรษที่ 20 ซึ่งจากงานเขียนของสเปียร์แมนในช่วงระหว่างปี 1904-1913 แสดงให้เห็นทั้งการพัฒนาทฤษฎีและการปฏิบัติเกี่ยวกับการวัดความถนัดและผลสัมฤทธิ์ทางการเรียน ซึ่งใช้กันมากในแวดวงทางการศึกษาและอุตสาหกรรมในขณะนั้น และตั้งแต่ปี 1900 เป็นต้นมา พฤติกรรมการทดสอบด้วยข้อสอบและคะแนนสอบก็ได้รับการอธิบายในเชิงคณิตศาสตร์และสถิตินับแต่นั้นมา

(Gulliksen. 1950 : 1 ; citing Cattell. 1890, Cattell and Farrand. 1896, Crocker and Algina. 1986 : 106 ; Keeves. 1988 : 241-243) กรอบความคิดทางทฤษฎีของสเปียร์แมนได้กล่าวถึงสาระสำคัญ 3 ประการคือ คะแนนจริง (True score) ความคลาดเคลื่อนในการวัด (Measurement error) และความเชื่อมั่นของแบบทดสอบ (Test reliability) ซึ่งนับว่าเป็นการเริ่มต้นของทฤษฎีการวัดแบบมาตรฐานเดิม (Classical test theory ; CTT) (Keeves. 1988 : 242) กัลลิคเซน (Gulliksen. 1950 : 4-11) ได้แสดงความเห็นว่าการวัดทางจิตวิทยามีความคลาดเคลื่อนทางการวัดมากกว่าการวัดทางกายภาพ และปัญหาที่สำคัญของการวัดทางสถิติปัญหา ก็คือการประมาณค่าความคลาดเคลื่อน (Error) ของคะแนนการสอบ และความแตกต่างระหว่างคะแนนคลาดเคลื่อนกับคะแนนจริง และได้แสดงความสัมพันธ์ของคะแนนจริง (T) คะแนนสอบ (X) และคะแนนคลาดเคลื่อน (E) ในรูปสมการเส้นตรงเชิงคณิตศาสตร์ดังนี้

$$X_i = T_i + E_i \text{ หรือ } E_i = X_i - T_i$$

เมื่อ X_i คือ คะแนนสอบของผู้สอบคนที่ i

T_i คือ คะแนนจริงของผู้สอบคนที่ i

E_i คือ องค์ประกอบของความคลาดเคลื่อนของบุคคลที่ i

จากสมการดังกล่าวทำให้เห็นข้อตกลงสำคัญสองประการคือ ความสัมพันธ์ระหว่างคะแนนจริงและคะแนนสอบ หรือเป็นการแสดงความหมายของความคลาดเคลื่อน ซึ่งอาจกล่าวได้ว่าเป็นความแตกต่างระหว่างคะแนนจริงกับคะแนนสอบของผู้สอบคนเดียวกัน และเพื่อให้สมการดังกล่าวสามารถนำไปใช้ได้เชิงปฏิบัติจึงได้สร้างข้อตกลงที่สำคัญดังนี้

1. ความคลาดเคลื่อนที่เกิดขึ้นในการวัดเป็นความคลาดเคลื่อนอย่างสุ่ม (Random error) และถ้าวัดกับผู้สอบจำนวนมากพอค่าเฉลี่ยของความคลาดเคลื่อนจะมีค่าเท่ากับ 0 ($\mu_E=0$)
2. เมื่อเป็นความคลาดเคลื่อนแบบสุ่ม เราจึงไม่สามารถบอกได้ว่าคนที่มีคะแนนจริงสูงจะมีความคลาดเคลื่อนสูงหรือต่ำ หรือมีคะแนนจริงต่ำจะมีความคลาดเคลื่อนต่ำไปด้วย ดังนั้นความสัมพันธ์ระหว่างคะแนนจริงกับความคลาดเคลื่อนแบบสุ่มจะมีค่าเท่ากับ 0 ($\rho_{TE}=0$)
3. ในการทดสอบด้วยแบบทดสอบต่างฉบับที่ขนานกัน ความคลาดเคลื่อนแบบสุ่มที่เกิดขึ้นจะเป็นอิสระต่อกัน ดังนั้น ความสัมพันธ์ของความคลาดเคลื่อนแบบสุ่มที่เกิดจากการทดสอบด้วยแบบทดสอบคู่ขนานจะมีค่าเท่ากับ 0 ($\rho_{E1E2}=0$)

แบบจำลองคะแนนจริง (True score model)

การนิยามคะแนนจริงในระยะแรกพิจารณาจากสมการ $X = T + E$ จะได้ว่าคะแนนจริงก็คือ ค่าความแตกต่างระหว่างคะแนนสอบ (X) กับคะแนนความคลาดเคลื่อน (E) หรือเขียนแทนด้วยสมการ $T_i = X_i - E_i$ ซึ่งกัลลิกเซน (Gulliksen : 8-11) ได้อธิบายโดยใช้วิธีการทางคณิตศาสตร์ จะได้ว่า $S_x^2 = S_t^2 + S_e^2$

คร็อกเกอร์และแอลจินา (Crocker and Algina. 1986 : 109) ได้กล่าวเสริมว่า คะแนนจริงของผู้สอบสามารถแสดงได้โดยค่าเฉลี่ยของคะแนนสอบเมื่อทำการทดสอบซ้ำแบบไม่มีการจำกัดด้วยแบบทดสอบฉบับเดิม หรือเท่ากับค่าคาดประมาณของ $X(EX)$ หรือ μ ซึ่งแทนค่าได้จากสูตร

$$\mu = \sum_{k=1}^K X_k P_k$$

เมื่อ X_k คือค่าตัวแปรที่เกิดแบบสุ่มค่าที่ k (คะแนนที่ได้จากการทดสอบครั้งที่ K) และ p_k คือความน่าจะเป็นที่จะทำให้เกิดค่า X_k (คะแนนนั้น ๆ) ตัวอย่างในกรณีโยนลูกเต๋า

ค่า X_k คือ {1,2,3,4,5,6} ค่า p_k คือ 1/6

จากสูตรจะได้ว่า $\mu = 1/6 (1+2+3+4+5+6) = 3.5$

นอกเหนือจากนั้นยังได้นิยามคะแนนจริงในรูปของ

$$T_i = \frac{\lim_{k \rightarrow \infty} \sum_{g=1}^k X_{ig}}{K}$$

เมื่อ T_i คือคะแนนจริงของผู้สอบคนที่ i , X_{ig} คือ คะแนนสอบของคนที่ i ในการทำแบบทดสอบชุดที่ g ตั้งแต่ 1 ถึง k ฉบับ ซึ่งแบบทดสอบทั้งหมดจะต้องเป็นแบบทดสอบคู่ขนาน (Parallel tests) ซึ่งมีคุณสมบัติดังนี้

1. คะแนนจริงของผู้สอบต้องเท่ากันทุกฉบับ

$$T_{i1} = T_{i2}, \dots = T_{ik} \text{ (แบบทดสอบมี } k \text{ ฉบับ)}$$

2. ความคลาดเคลื่อนมาตรฐานของการวัดต้องเท่ากันทุกฉบับ

$$S_{e1} = S_{e2}, \dots = S_{ek}$$

3. ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบต้องเท่ากันทุกฉบับ

$$M_{x1} = M_{x2}, \dots = M_{xk}$$

$$S_{x1} = S_{x2}, \dots = S_{xk}$$

4. ค่าสหสัมพันธ์ภายในระหว่างแบบทดสอบจะต้องเท่ากัน

$$r_{x1x2} = r_{x1x3}, \dots = r_{xk-1xk}$$

ค่าความแปรปรวนของคะแนนจริง S_t^2 มีค่าเท่ากับ $S_x^2 r_{xggh}$ (เมื่อ g และ h เป็นแบบทดสอบคู่ขนาน) หรือ

$$S_t^2 = S_x^2 r_{xggh} = S_g^2 r_{gh}$$

ดังนั้น ส่วนเบี่ยงเบนมาตรฐานของคะแนนจริง คือ

$$S_t = S_x \sqrt{r_{xggh}} = S_g \sqrt{r_{gh}}$$

(Gulliksen, 1950 : 11-14, 28-31)

อย่างไรก็ตาม ในเชิงปฏิบัติการสอบคงทำได้เพียงครั้งเดียว ดังนั้นการประมาณค่าคะแนนจริงของผู้สอบอาจทำได้ 2 วิธีคือ วิธีการใช้สมการถดถอย และประมาณค่าแบบช่วง (Crocker and Algina, 1986 : 146-152)

1. การใช้สมการถดถอย (Regression equation)

$$\text{จาก } Y' = \rho_{xy} \frac{\sigma_y}{\sigma_x} (X - \mu_x) + \mu_y$$

แทนค่า ρ_{xy} ด้วย ρ_{xt} และ Y' ด้วย T'

$$\text{จะได้ว่า } T' = \rho_{xt} \frac{\sigma_t}{\sigma_x} (X - \mu_x) + \mu_t$$

เพราะว่า $\sigma_t / \sigma_x = \rho_{xt} / \rho_{xx}$ และ $\mu_t = \mu_x$ จะได้ว่า

$$T' = \rho_{xx} (X - \mu_x) + \mu_t$$

ถ้าให้ $t' = (T' - \mu_t)$ และ $x' = (X - \mu_x)$

$$\text{ดังนั้น } t' = \rho_{xx} (x')$$

เพราะว่าการประมาณค่าของ ρ_{xx} ขึ้นอยู่กับข้อมูลจากการทดสอบจริง ซึ่งส่วนมากแล้วมีค่าน้อยกว่า 1.00 เสมอ จึงทำให้ค่าของ t' มีค่าน้อยกว่า x ดังนั้น การพยากรณ์คะแนนจริงของผู้สอบแต่ละคนจะถูกนำไปอ้างอิงจากค่าเฉลี่ยของกลุ่มมากกว่าคะแนนสอบเสมอ อย่างไรก็ตาม

ถ้าสัมประสิทธิ์ของความเชื่อมั่นมีค่าน้อย ค่าคะแนนจริงพยากรณ์ของผู้สอบก็จะมีค่าใกล้เคียงกับค่าเฉลี่ยของกลุ่ม และเมื่อ $p_{xx} = 0$ ค่า T ของผู้สอบทุกคนจะมีค่าเท่ากับ X หรืออาจกล่าวได้อีกอย่างหนึ่งว่า ถ้าค่าความเชื่อมั่นมีค่าเท่ากับ 1.00 ค่าของ T' ก็จะมีค่าใกล้เคียงกับ X

2. การประมาณค่าคะแนนจริงแบบช่วง

การประมาณค่าแบบนี้จะใช้ความสัมพันธ์ระหว่างคะแนนสอบกับค่าคาดประมาณความคลาดเคลื่อน ซึ่งเคลลี (Kelley) เสนอวิธีคาดประมาณความคลาดเคลื่อนไว้ 3 แนวทางคือ

1. พิจารณาความแตกต่างระหว่างคะแนนจริงกับคะแนนสอบของผู้สอบ
2. พิจารณาความแตกต่างระหว่างคะแนนสอบของแบบทดสอบฉบับหนึ่งกับคะแนนพยากรณ์ของแบบทดสอบคู่ขนานกับฉบับแรกของผู้สอบ
3. พิจารณาความแตกต่างระหว่างคะแนนจริงกับค่าประมาณคะแนนจริงของผู้สอบในทางปฏิบัติแล้ว วิธีที่ 1 และวิธีที่ 2 มีความเป็นไปได้ โดยที่

วิธีที่ 1 ค่าทำนายคะแนนจริง $E(T) = X \pm Z_\alpha SEM$

เมื่อ X คือ คะแนนสอบ

Z_α คือ พื้นที่ภายใต้โค้งปกติ ณ ระดับนัยสำคัญ α

SEM คือ ความคลาดเคลื่อนมาตรฐานของการวัด (Standard error of measurement) ซึ่งแทนค่าได้จากสูตร

$$SEM = S_x \sqrt{1 - r_{xx}}$$

เมื่อ r_{xx} คือ ค่าความเชื่อมั่นของแบบทดสอบ

วิธีที่ 2 ค่าทำนายคะแนนจริง $E(T)$ จากการใช้แบบทดสอบคู่ขนาน

$$E(T) = X \pm Z_\alpha SEE$$

เมื่อ SEE คือ ความคลาดเคลื่อนมาตรฐานในการประมาณค่า (Standard error of estimate) ซึ่งแทนค่าได้จากสูตร

$$SEE = S_x \sqrt{1 - r_{xx}^2}$$

ซึ่งค่า $E(T)$ ที่ประมาณจาก SEE จะมีช่วงกว้างกว่าประมาณจากค่า SEM

(Gulliksen. 1950 : 15, 41 ; Crocker and Algina. 1986 : 150-152 ; citing Kelly. 1927)

คุณภาพของแบบทดสอบ (Test quality)

จอร์แดน (Jordan. 1953 : 14-26) กล่าวถึงเครื่องมือวัดที่ดีควรมีคุณลักษณะสำคัญดังนี้ คือ 1) มีความเที่ยงตรง (Validity) 2) มีความเชื่อมั่น (Reliability) 3) สะดวกในการใช้และการดำเนิน

การ (Administrability) 4) มีการนำเสนอคุณสมบัติในเชิงการเปรียบเทียบความสามารถ (Interpretation and Comparability) ของแบบทดสอบ เช่น เกณฑ์มาตรฐาน (Norm) ค่าความเชื่อมั่นและความเที่ยงตรง และ 5) ประหยัด (Economy) เช่น งบประมาณ เวลาของนักเรียนและผู้สอน ส่วนครอนบัคและคณะ (Cronbach et al.) ได้กล่าวถึงการประเมินโดยใช้แบบทดสอบว่าจะต้องเผชิญกับปัญหาสำคัญที่มีความแตกต่างกันแต่ก็มีส่วนซ้อนทับกัน 2 ประการคือ 1) สาระของแบบทดสอบที่แสดงถึงเอกภพ (Universe) ของคำตอบในสิ่งที่สนใจมาน้อยเพียงใด และ 2) ศึกษาถึงความสัมพันธ์ของเอกภพนั้นกับคุณลักษณะที่แฝงในตัวผู้สอบว่ามีความเชื่อถือได้เพียงใด ซึ่งปัญหาในประเด็นแรกก็คือความเชื่อมั่น ส่วนในประเด็นหลังก็คือความเที่ยงตรงนั่นเอง (Thorndike in Keeves. 1988 : 330 ; citing Cronbach et al. 1972) ดังนั้น จึงอาจกล่าวได้ว่า ความเที่ยงตรงและความเชื่อมั่นเป็นคุณสมบัติสำคัญที่บ่งชี้ถึงคุณภาพของแบบทดสอบ

ความเที่ยงตรงของแบบทดสอบ (Test validity)

ความเที่ยงตรงนับว่าเป็นคุณสมบัติสำคัญที่สุดของแบบทดสอบที่ต้องพิจารณาเป็นลำดับแรก ถ้าแบบทดสอบใดขาดความเที่ยงตรงแบบทดสอบนั้นก็ใช้ประโยชน์ไม่ได้ นอกจากนั้น ข้อมูลที่ได้จากการทดสอบไม่สามารถนำไปสู่การตัดสินใจที่ถูกต้องได้ (Jordan. 1953 : 14 ; Airasian. 1994 : 2 ; citing Messick. 1989 ; Hoy and Gregg. 1994 : 72, Oosterhof. 1994 : 53 ; Venn. 1994 : 49)

การนิยามความหมายของความเที่ยงตรงของแบบทดสอบมีลักษณะที่หลากหลายในแต่ละยุคสมัยซึ่งสรุปได้ 4 กลุ่มคือ

1) เป็นคุณสมบัติของแบบทดสอบที่สามารถวัดได้ตรงกับสิ่งที่จะวัด (Jordan. 1953 : 14 ; Lord and Novick. 1968 : 261 ; Good. 1973 : 635 ; Zeller in Keeves. 1988 : 322 ; Linn in Alkin 1992 : 9 ; Kline. 1993 : 15 ; Hoy and Gregg. 1994 : 72 ; Zeller in Husen and Postlethwaite. 1994 : 6569 ; Nunnally and Bernstein. 1994 : 83 ; Oosterhof. 1994 : 53 ; Venn. 1994 : 49 ; สำเร็จ บุญเรืองรัตน์. 2535 : 101)

2) เป็นความสัมพันธ์ของแบบทดสอบกับเกณฑ์บางประการหรือตัวแปรอื่น ๆ (Gulliksen. 1950 : 88 ; Messick in Linn. 1989 : 18 ; citing Guilford. 1946 : 429 ; Friedenber. 1995 : 221)

3) เป็นกระบวนการใช้คะแนนสอบอ้างอิงสิ่งต่าง ๆ ตามจุดประสงค์ที่ต้องการ (Crocker and Algina. 1986 : 217 ; citing Cronbach. 1971)

4) เป็นการตัดสินใจเชิงประเมินเกี่ยวกับคะแนนสอบหรือข้อมูลอื่นที่เป็นผลมาจากการประเมินโดยพิจารณาจากหลักฐานหรือข้อมูลที่สังเกตพบ และเหตุผลเชิงทฤษฎีเพื่ออธิบายข้อสรุปและพฤติกรรมแสดงออกได้อย่างเหมาะสมและเพียงพอ หรืออาจกล่าวได้อย่างกว้าง ๆ ว่าเป็นข้อสรุปเชิงอนุมานจากหลักฐานข้อมูลที่ปรากฏชัดเจน และแฝงความสำคัญสำหรับใช้และนำเสนอผลของคะแนนสอบ (Messick in Linn. 1989 : 13)

ประเภทความเที่ยงตรง (Type of validity)

ประเภทความเที่ยงตรงที่ยอมรับและใช้กันมากที่สุดในปัจจุบันจัดแบ่งเป็น 3 ประเภทตามแนวทางของสมาคมจิตวิทยาอเมริกัน (American Psychological Association ; APA) ดังนี้

1. ความเที่ยงตรงเชิงเนื้อหา (Content validity) เป็นการศึกษาความสอดคล้องความสัมพันธ์ หรือความครอบคลุมของเนื้อหาสาระของข้อสอบในแบบทดสอบกับโดเมนเนื้อหา ทักษะ หรือพฤติกรรมแสดงออกที่สนใจศึกษา หรืออาจกล่าวได้อย่างง่าย ๆ ว่าเป็นคุณสมบัติของแบบทดสอบที่วัดได้ตรงกับเนื้อหาหรือพฤติกรรมแสดงออกที่ต้องการวัด

2. ความเที่ยงตรงเชิงโครงสร้าง (Construct validity) เป็นการศึกษาคุณสมบัติของแบบทดสอบที่เน้นการประเมินคุณลักษณะของผู้สอบตามกรอบทฤษฎีโครงสร้างทางจิตวิทยาหรืออาจกล่าวได้อย่างง่าย ๆ ว่าเป็นคุณสมบัติของแบบทดสอบที่วัดได้ตรงหรือครอบคลุมโครงสร้างหรือคุณลักษณะเชิงทฤษฎีที่ต้องการศึกษา

3. ความเที่ยงตรงเกณฑ์สัมพันธ์ (Criterion related validity) เป็นการศึกษาความสัมพันธ์ของแบบทดสอบกับเกณฑ์อิสระต่าง ๆ แบ่งเป็น 2 ประเภท คือ

3.1 ความเที่ยงตรงเชิงพยากรณ์ (Predictive validity) เป็นการศึกษาความสัมพันธ์ของแบบทดสอบกับเกณฑ์พยากรณ์ความสามารถในอนาคต

3.2 ความเที่ยงตรงเชิงสภาพ (Concurrent validity) เป็นการศึกษาความสัมพันธ์ของแบบทดสอบกับเกณฑ์ที่บ่งบอกความสามารถในปัจจุบันและในอดีต (Gronlund. 1982 : 125-138 ; Messick in Linn. 1989 : 16-20 ; citing APA. 1954, 1966, 1985 ; Anastasi. 1990 : 140-162 ; Zeller in Keeves. 1988 : 324-329 ; Kline. 1993 : 17-26 ; Hoy and Gregg. 1994 : 72-73 ; Nunnally and Birnstien. 1994 : 83-104 ; Oosterhof. 1994 : 53-68 ; Venn. 1994 : 50-53 ; Frledenberg. 1995 : 225-250)

ความเที่ยงตรงประเภทอื่น ๆ

1. ความเที่ยงตรงโครงร่าง (Empirical validity) ความเที่ยงตรงภายนอก (External validity) และความเที่ยงตรงเชิงสถิติ (Statistical validity) คือ ความเที่ยงตรงเกณฑ์สัมพัทธ์ (Jordan. 1953 : 26 ; Messick in Linn. 1989 : 19)

2. ความเที่ยงตรงภายใน (Intrinsic validity or Internal validity) และความเที่ยงตรงแบบวงแหวน (Circular validity) คือ ความเที่ยงตรงเชิงเนื้อหา (Jordan. 1959 : 26 ; Messick in Linn. 1989 : 18-19 ; citing Anastasi. 1954)

3. ความเที่ยงตรงคุณลักษณะ (Trait validity) และความเที่ยงตรงแบบพหุ (Factorial validity) คือความเที่ยงตรงเชิงโครงสร้าง

4. ความเที่ยงตรงเชิงประจักษ์ (Face validity) เป็นความเที่ยงตรงที่เกิดจากความรู้สึกของผู้เชี่ยวชาญหรือผู้เกี่ยวข้องในการตัดสินใจว่าเครื่องมือวัดในสิ่งที่ต้องการวัดและวัดได้เหมาะสมกับระดับผู้สอบ (Kline. 1993 : 16 ; citing Cattell and Warbuton. 1967 ; Hoy and Gregg. 1994 : 72-73 ; Nunnally and Birnstien. 1994 : 109-110 ; Oosterhof. 1994 : 59 ; Venn. 1994 : 51)

5. ความเที่ยงตรงเชิงการค้า (Cash validity) เป็นการพิจารณาแบบทดสอบในเชิงพาณิชย์ (Venn. 1994 : 51)

6. ความเที่ยงตรงตามหลักสูตรและการเรียนการสอน (Curricular and instructional validity) เป็นการพิจารณาความครอบคลุมเนื้อหาตามหลักสูตรและการเรียนการสอนของแบบทดสอบ (Oosterhof. 1994 : 59) หรืออาจกล่าวได้ว่าเป็นความสอดคล้องระหว่างเนื้อหาของแบบทดสอบและเนื้อหาในหลักสูตรและจุดประสงค์ของแบบทดสอบและจุดประสงค์ของหลักสูตร ซึ่งก็คือ ความเที่ยงตรงเชิงเนื้อหา (Good. 1973 : 635)

7. ความเที่ยงตรงข้ามกลุ่ม (Cross validation) หมายถึงความเที่ยงตรงของแบบทดสอบชุดหนึ่งที่มีต่อกลุ่มตัวอย่างอื่นที่ไม่ใช่กลุ่มตัวอย่างแรกที่ใช้ศึกษาความเที่ยงตรงของแบบทดสอบชุดนั้น (สำเร็จ บุญเรืองรัตน์. 2535 : 123 ; Good. 1973 : 635 ; (Anastasi. 1990)

8. ความเที่ยงตรงเอกนัย (Convergent validity) เป็นค่าสัมประสิทธิ์ที่แสดงความสัมพันธ์ของการวัดแบบต่าง ๆ ที่วัดโครงสร้างของพฤติกรรมหรือคุณลักษณะเดียวกัน (monotrait-hetero method) (Good. 1973 : 635 ; Crocker and Algina. 1986 : 232-233 ; citing Campbell and Fiske. 1959)

9. ความเที่ยงตรงอนกนัย (Divergent validity) หรือความเที่ยงตรงเชิงจำแนก (Discriminant validity) เป็นการวัดความแตกต่างของบางสิ่งบางอย่าง โดยวิธีการที่มีอยู่ซึ่งมีอยู่ 2 ลักษณะ คือ

9.1 วิธีการเดียวแต่วัดหลายคุณลักษณะ (Heterotrait-monomethod)

9.2 หลาย ๆ วิธีวัดหลาย ๆ คุณลักษณะ (Heterotrait-heteromethod)

(สำเร็จ บุญเรืองรัตน์. 2535 : 118-121 ; Good. 1973 : 635 ; Crocker and Algina. 1986 : 232-233 ; citing Campbell and Fiske. 1959 ; Kline. 1993 : 15-27 ; Nunnally and Bernstein. 1994 : 9293)

กระบวนการประเมินความเที่ยงตรงของแบบทดสอบ (Test Validation)

จอร์แดน (Jordan. 1953 : 14-15) กล่าวว่า การวัดความเที่ยงตรงของแบบทดสอบมีเป้าหมายสำคัญ 2 ประการคือ

1. การวัดความเที่ยงตรงในการสร้างข้อสอบ (Curricular or internal validity) ได้แก่ การตรวจสอบความสัมพันธ์กับจุดประสงค์การเรียนการสอน และเพื่อค้นหารูปแบบของการตอบสนองซึ่งกำหนดไว้ในจุดประสงค์

2. การวัดความเที่ยงตรงโดยการหาความสัมพันธ์กับเกณฑ์ภายนอกบางประการ (External validity) เช่น คะแนนการเรียน แบบทดสอบวัดสติปัญญาอื่น ความสามารถความสำเร็จในการทำงาน ระดับการศึกษา ช่วงเวลา และเกณฑ์อื่นที่เกี่ยวข้อง

ความเที่ยงตรงเชิงเนื้อหา (Content-related validation)

กรอนลินด์ (Gronlund. 1982 : 125-138) เสนอวิธีการไว้ดังนี้

1. กำหนดนิยามหัวข้อในสาขาวิชาหรือผลความรู้ที่จะวัด
2. จัดเตรียมลักษณะเฉพาะข้อสอบ (Item spec.) และตัวอย่างข้อสอบที่จะนำไปใช้
3. สร้างแบบทดสอบจากลักษณะเฉพาะของข้อสอบ

อนาสตาซี (Anastasi. 1990 : 141-144)

1. กำหนดตารางลักษณะเฉพาะของแบบทดสอบ (Test-specification) เป็นตารางแบบสองทางระหว่างขอบเขตหรือหัวข้อของเนื้อหาและจุดประสงค์การเรียนการสอนหรือกระบวนการ
2. กำหนดจำนวนข้อตามความสำคัญของหัวข้อและกระบวนการโดยผู้เชี่ยวชาญ
3. จัดทำคู่มือของแบบทดสอบอธิบายวิธีการเพื่อเป็นหลักประกันว่าเนื้อหาของแบบทดสอบมีความเหมาะสมและเป็นตัวแทนหรือแสดงสารสนเทศของแบบทดสอบที่

3. จัดทำคู่มือของแบบทดสอบอธิบายวิธีการเพื่อเป็นหลักประกันว่าเนื้อหาของแบบทดสอบมีความเหมาะสมและเป็นตัวแทนหรือแสดงสารสนเทศของแบบทดสอบที่ครอบคลุมขอบเขตเนื้อหาและทักษะหรือจุดประสงค์การเรียนรู้การสอน

4. ตรวจสอบแบบทดสอบว่า

4.1 แบบทดสอบครอบคลุมทักษะและความรู้ที่กำหนดไว้หรือไม่

4.2 พฤติกรรมแสดงออกต่อข้อสอบเป็นอิสระจากอิทธิพลของตัวแปรที่ไม่เกี่ยวข้องอย่างมีเหตุผลหรือไม่ โดยเฉพาะแบบทดสอบอิงเกณฑ์ที่แสดงพฤติกรรมแสดงออกตามความหมายของเนื้อหา

ครอกเกอร์และแอลจินา (Crocker and Algina. 1986 : 218 - 222) เสนอขั้นตอนดำเนินการตามลำดับขั้นดังนี้

1. นิยามโดเมนพฤติกรรมแสดงออกที่สนใจ (Performance domain of interest) ส่วนใหญ่พบว่านิยามในรูปของจุดประสงค์การเรียนรู้ (Instructional objectives)

2. คัดเลือกคณะผู้ตัดสินที่มีความเชี่ยวชาญโดเมนเนื้อหา นั้น ๆ

3. ก่อนที่จะทำการวิเคราะห์ความสอดคล้องตามข้อ 2 คณะผู้ตัดสินต้องทำความเข้าใจกับหลักวิชา กระบวนการคิด ระดับความซับซ้อน รูปแบบคำถาม วิธีการนำเสนอ และคำตอบที่คาดว่าจะเป็นไปได้ของข้อสอบและโดเมนให้ชัดเจน

4. การตัดสินข้อสอบที่เหมาะสมเน้นที่คุณภาพมากกว่าปริมาณซึ่งอาจพิจารณาจากตัวบ่งชี้เชิงปริมาณบางประการ ได้แก่

4.1 ร้อยละข้อสอบที่สอดคล้องกับจุดประสงค์ (ข้อสอบอย่างน้อย 100 ข้อ)

4.2 ร้อยละข้อสอบที่สอดคล้องกับจุดประสงค์ในระดับสำคัญอย่างยิ่ง (จำนวนข้อสอบอย่างน้อย 100 ข้อ)

4.3 สหสัมพันธ์ระหว่างน้ำหนักความสำคัญของจุดประสงค์กับจำนวนข้อสอบที่ใช้วัดในแต่ละจุดประสงค์ (เฉพาะกรณีกำหนดน้ำหนักความสำคัญของจุดประสงค์ไม่เท่ากัน)

4.4 ร้อยละของจุดประสงค์ที่ไม่ปรากฏว่าสอดคล้องกับข้อสอบ

4.5 พิจารณาจากดัชนีความสอดคล้อง (Index of congruence ; IOC) ของโรวินELLI และแฮมเบิลตัน (Crocker and Algina . 1986. : 218-222 ; citing Rovinelli and Hambleton. 1977)

$$IOC_{ik} = \frac{N}{2N-2} (\mu_k - \mu)$$

เมื่อ IOC_k คือ ค่าดัชนีความสอดคล้องของข้อสอบข้อที่ i กับจุดประสงค์ข้อที่ k
 N คือ จำนวนของจุดประสงค์ทั้งหมด

μ_k คือ ค่าระดับเฉลี่ยของการตัดสินข้อสอบข้อที่ i ของจุดประสงค์ข้อที่ k

μ คือ ค่าระดับเฉลี่ยของการตัดสินข้อสอบข้อที่ i จากจุดประสงค์ทั้งหมด

ค่าสูงสุดของ $IOC = 1.00$ เมื่อข้อสอบ 1 ข้อ ได้รับการตัดสินว่าสอดคล้องกับ
 จุดประสงค์เพียง 1 จุดประสงค์ โดยผู้ตัดสินทุกคน

$IOC < 1.00$ ในกรณีที่ข้อสอบ 1 ข้อ ถูกตัดสินว่าวัดมากกว่า 1 จุดประสงค์

ในทางอุดมคติ ค่า IOC ควรมีค่าสูงมากกับจุดประสงค์ที่ต้องการให้สอดคล้องและมีค่า
 ต่ำมากกับจุดประสงค์อื่น และถ้าแบบทดสอบวัดหลายจุดประสงค์แต่ข้อสอบทุกข้อสอดคล้องกับ
 จุดประสงค์เดียวดัชนีความสอดคล้องจะมีค่าต่ำ

ออสเตอร์ฮอฟ (Oosterhof, 1994 : 54-60) กล่าวถึงการกำหนดความเที่ยงตรงที่สัมพันธ์
 กับเนื้อหาว่าบ่อยครั้งที่พบว่าการประเมินพร้อม ๆ กับการประเมินความเที่ยงตรงเชิงโครงสร้าง
 การกำหนดจะกระทำโดยผู้เชี่ยวชาญโดยการเปรียบเทียบคำถามในแบบทดสอบกับทักษะที่ออก
 แบบไว้สำหรับการวัด ซึ่งกระบวนการนี้ต้องผ่านการวิเคราะห์เชิงระบบ เพื่อพิจารณาว่าแบบ
 ทดสอบวัดอะไรบ้าง ซึ่งมี 2 วิธีการคือ

1. จัดทำตารางกำหนดลักษณะเฉพาะข้อสอบ (Table of specifications) ซึ่งเป็นตาราง
 สองทางโดยแนวดิ่งแสดงขอบเขตของเนื้อหา และแนวนอนแสดงกลุ่มของพฤติกรรมแสดงออก
 ที่ต้องการวัด เช่น ลำดับชั้นการรู้คิดของบลูม (Bloom's taxonomy) ที่เสนอขึ้นในปี 1956 หรือผล
 ผลการเรียนรู้ของกานเย่ (Gagne's categories of learning outcomes) ที่เสนอในปี 1985 เป็นต้น
 จากนั้นดำเนินการตามขั้นตอนดังนี้

1.1 กำหนดจำนวนข้อคำถามทั้งหมดในแบบทดสอบ

1.2 กระจายจำนวนข้อคำถามในแต่ละขอบเขตเนื้อหาตามความสำคัญ

1.3 กระจายจำนวนข้อตามความสำคัญของพฤติกรรมแสดงออกในแต่ละ
 ขอบเขตเนื้อหา

1.4 คัดเลือกหรือปรับจำนวนข้อคำถามสำหรับพฤติกรรมแสดงออกในแต่ละ
 ขอบเขตเนื้อหา

2. กำหนดจุดประสงค์ของพฤติกรรมแสดงออก (Performance objectives) จากเนื้อหา
 และกลุ่มของพฤติกรรมแสดงออกหรือสมรรถภาพ (Capability) ในรูปของทักษะ (Skill) การเขียนจุด
 ประสงค์จะต้องประกอบด้วยส่วนสำคัญคือ เนื้อหาหรือหลักวิชา (Content) พฤติกรรม (Behavior)

สถานการณ์ (Situation) และเงื่อนไขเฉพาะ (Special conditions) ซึ่งจุดประสงค์นี้จะนำมาใช้แทนตารางกำหนดลักษณะเฉพาะ จากนั้นกำหนดจำนวนข้อคำถามตามความสำคัญของจุดประสงค์ในแต่ละทักษะตามจำนวนที่กำหนดไว้ในแบบทดสอบ

เวนน์ (Venn, 1994 : 50-51) เสนอกระบวนการกำหนดความเที่ยงตรงเชิงเนื้อหาไว้ดังนี้

1. พัฒนาลักษณะเฉพาะของแบบทดสอบ (Test specifications) ที่ครอบคลุมขอบเขตเนื้อหา

2. เขียนข้อสอบให้สอดคล้องกับคุณลักษณะเฉพาะของแบบทดสอบ
3. ควบคุมการสอบภาคสนามโดยกำหนดแบบทดสอบให้เหมาะสมกับกลุ่มเป้าหมาย
4. พิจารณาปรับปรุงแบบทดสอบหลังจากทดสอบภาคสนามครั้งแรก
5. จัดเรียงเรียงแบบทดสอบเข้าฉบับ

ซึ่งขั้นตอนดังกล่าวนี้จำเป็นต้องใช้ความคิดเห็นจากผู้เชี่ยวชาญที่มีความชำนาญในสาขาวิชานั้น ๆ ได้แก่ ผู้สอน ผู้เชี่ยวชาญหลักสูตร อาจารย์ในระดับวิทยาลัย และผู้เชี่ยวชาญในการพัฒนาแบบทดสอบในลักษณะเป็นคณะ ถ้าเป็นเชิงพาณิชย์ควรเป็นคณะที่ใหญ่

ผู้สอนและนักพัฒนาแบบทดสอบจะสร้างความเที่ยงตรงของแบบทดสอบหลังจากประเมินคุณภาพข้อสอบเป็นรายข้อ ซึ่งส่วนใหญ่จะทำก่อนพิมพ์เป็นฉบับ

ผู้ใช้แบบทดสอบประเมินความเที่ยงตรงเชิงเนื้อหาโดยการทบทวนจากคู่มือแบบทดสอบรวมทั้งการทบทวนเกี่ยวกับคุณภาพข้อสอบเป็นรายข้อ

ไฟรเดนเบิร์ก (Friedenberg, 1995 : 248-252) แสดงทรรศนะเกี่ยวกับความเที่ยงตรงเชิงเนื้อหาว่าเป็นกระบวนการเชิงคุณภาพซึ่งเป็นการเปรียบเทียบข้อทดสอบกับโดเมนของแบบทดสอบโดยการตัดสินจากผู้เชี่ยวชาญ และถ้าแบบทดสอบนั้นได้รับการตัดสินว่ามีความเหมาะสม อาจสรุปได้ว่าความแตกต่างในคะแนนสอบ (σ^2_x) สะท้อนถึงความแตกต่างของโดเมนคุณลักษณะที่เกี่ยวข้อง (σ^2_{μ}) ซึ่งกระบวนการกำหนดความเที่ยงตรงเชิงเนื้อหาดำเนินการตามลำดับดังนี้

1. วิเคราะห์โดเมน ทำความชัดเจนเกี่ยวกับความรู้และทักษะเฉพาะ ซึ่งต้องการให้แบบทดสอบวัดได้ครอบคลุม ซึ่งอาจทำได้ทั้งก่อนหรือหลังการคัดเลือกข้อสอบ

2. เปรียบเทียบแบบทดสอบกับเนื้อหาและโครงสร้างของโดเมนตามข้อคำถาม 3 ประการ คือ

- 2.1 เนื้อหาของข้อสอบในแบบทดสอบเหมาะสมกับโดเมนหรือไม่
- 2.2 รูปแบบของคำถามเหมาะสมหรือไม่

2.3 ข้อสอบเป็นตัวแทนของโดเมนหรือไม่ กล่าวคือ วัดคุณลักษณะที่สำคัญของโดเมนมากกว่าคุณลักษณะย่อยของโดเมนหรือไม่

กระบวนการประเมินความเที่ยงตรงเชิงเนื้อหาวิธีการอื่น

1. บราวน์ (Brown. 1983) พัฒนามาตรวัดประมาณค่ามาตรฐาน เพื่อตัดสินในเชิงความเห็น (Degree of agreement) และประมาณค่าความเชื่อมั่นของผู้ให้คะแนน (Interrater reliability) (Friedenberg. 1995 : 252 ; citing Brown. 1983) ซึ่งอาจใช้สูตรคำนวณ ซึ่งเสนอโดย ไวลรี ไบเลย์และซูไก สำหรับผู้ตัดสิน 2 คน ดังนี้

$$\text{ร้อยละของความเห็นที่สอดคล้อง} = \frac{\text{จำนวนความเห็นที่สอดคล้อง} \times 100}{\text{จำนวนความเห็นที่สอดคล้อง} + \text{ไม่สอดคล้อง}}$$

2. บาร์โลว์และเฮร์เซน (Barlow and Hersen) แสดงขั้นตอนของการประมาณความเชื่อมั่นของผู้ตัดสินไว้ดังนี้

1. คัดเลือกพฤติกรรมที่ต้องการให้พิจารณา
2. กำหนดเงื่อนไขภายใต้ข้อมูลที่ได้รับ
3. เลือกหน่วยวิเคราะห์

4. เปลี่ยนแปลงข้อมูลที่รวบรวมไว้ถ้าจำเป็น (Venn. 1994 : 49 ; citing Barlow and Hersen. 1984, Wolery, Bailey and Sugai. 1988)

3. ครอนบาค (Cronbach) ประเมินความเที่ยงตรงเชิงเนื้อหา โดยหาความสัมพันธ์ของคะแนนจากแบบทดสอบ 2 ฉบับ ที่วัดโครงสร้างเดียวกัน โดยเริ่มจากวิเคราะห์โดเมนแล้วตั้งคณะทำงาน 2 ชุด ให้แต่ละชุดสร้างแบบทดสอบตามโดเมนที่วิเคราะห์แล้วนำไปทดสอบกับกลุ่มผู้สอบกลุ่มเดียวกัน นำคะแนนสอบทั้งสองฉบับมาหาค่าสัมประสิทธิ์สหสัมพันธ์ ถ้ามีค่าสูงแสดงว่าแบบทดสอบแต่ละชุดมีความเที่ยงตรงเชิงเนื้อหา (Friedenberg. 1995 : 252 ; citing Cronbach. 1971)

ความเที่ยงตรงเชิงโครงสร้าง (Construct-Related Validation)

กรอนลันด์ (Gronlund. 1982 : 125-138) กล่าวถึงเป้าหมายในการประเมินความเที่ยงตรงเชิงโครงสร้างมี 2 ประการคือ

1. เพื่อกำหนดองค์ประกอบ (Factors) ที่มีอิทธิพลต่อพฤติกรรมการตอบสนองต่อแบบทดสอบ

2. เพื่อศึกษาระดับอิทธิพลของแต่ละองค์ประกอบ โดยมีกระบวนการตามลำดับขั้นดังนี้

2.1 จำแนกโครงสร้างที่จะอธิบายพฤติกรรมการตอบสนองต่อแบบทดสอบ

2.2 ตั้งสมมติฐานจากทฤษฎีต่าง ๆ ในแต่ละโครงสร้าง

2.3 เก็บรวบรวมข้อมูลเพื่อทดสอบสมมติฐาน

ครอกเกอร์และแอลจินา (Crocker and Algina. 1986 : 230-235) เสนอกระบวนการพิจารณาความเที่ยงตรงเชิงโครงสร้างดังนี้

1. ตั้งสมมติฐานเกี่ยวกับโครงสร้างที่คาดว่าจะมีความแตกต่างภายในบุคคล ได้แก่คุณลักษณะที่เป็นข้อเท็จจริงของประชากร เกณฑ์พฤติกรรมแสดงออกหรือการวัดโครงสร้างอื่น ๆ ที่สัมพันธ์กับเกณฑ์พฤติกรรมแสดงออกที่มีความเที่ยงตรง สมมติฐานต้องตั้งอยู่บนพื้นฐานทางทฤษฎี ที่แสดงถึงโครงสร้างและได้รับการยอมรับภายใต้เงื่อนไขของการนิยามที่ชัดเจน

2. คัดเลือกเครื่องมือวัด ซึ่งประกอบด้วยข้อสอบที่วัดพฤติกรรมเฉพาะที่แสดงถึงโครงสร้างที่เป็นรูปธรรมชัดเจน

3. รวบรวมข้อมูลเพื่อตรวจสอบความสัมพันธ์กับสมมติฐาน

4. พิจารณาตัดสินข้อมูลที่สอดคล้องกับสมมติฐาน และควรอธิบายสิ่งที่ปรากฏจากการสังเกตโดยทฤษฎีหรือข้ออธิบายอื่นที่เหมาะสม ถ้าความสัมพันธ์เชิงสมมติฐานได้รับการยืนยัน ในลักษณะของการพยากรณ์โดยโครงสร้างเชิงทฤษฎี แบบทดสอบนั้นสามารถนำมาใช้ประโยชน์ได้ แต่ถ้าสมมติฐานไม่ได้รับการยืนยันเกี่ยวกับความเที่ยงตรง นักพัฒนาแบบทดสอบจะไม่สามารถทราบเลยว่าโครงสร้างเชิงทฤษฎีหรือแบบทดสอบ หรือทั้งสองสิ่งยังมีส่วนบกพร่อง

วิธีการตรวจสอบความเที่ยงตรงเชิงโครงสร้าง

1. ความแตกต่างระหว่างกลุ่ม (Known groups technique or experimental manipulation) เป็นวิธีการทดสอบความแตกต่างของกลุ่มที่ตั้งสมมติฐานว่ามีโครงสร้างแตกต่างกัน เช่น โครงสร้างเกี่ยวกับการรับรู้ บทบาทตามเพศของกลุ่มเพศชายและหญิง หรือออกแบบให้กลุ่มทดลองได้รับเงื่อนไขตามโครงสร้างที่ต้องการศึกษา เปรียบเทียบผลกับกลุ่มควบคุม (กลุ่มที่ไม่ได้รับเงื่อนไข) หรือเปรียบเทียบระหว่างคะแนนก่อนสอบ (Pretest) กับคะแนนหลังสอบ (Posttest) (Gronlund. 1982 : 125-138 ; Citing Cronbach. 1971 ; Crocker and Algina. 1986 : 231 ; Anastasi. 1990 : 158-159 ; Venn. 1994 : 52 ; citing Angoff. 1988 ; Friedenberg. 1995 : 255)

2. การเปลี่ยนแปลงของพฤติกรรมแสดงออกจากการฝึกฝนหรือการวิเคราะห์พัฒนาการ โดยการพิจารณาแบบแผนของการเปลี่ยนแปลงหรือพัฒนาที่เป็นไปอย่างต่อเนื่อง (Gronlund. 1982 : 125-138 ; citing Cronbach. 1971 ; Anastasi. 1990 : 153)

3. การหาความสัมพันธ์กับแบบทดสอบฉบับอื่น เป็นวิธีการอย่างง่ายโดยการหาความสัมพันธ์ของคะแนนสอบกับคะแนนหรือผลการวัดที่เกี่ยวข้องหรือวัดโครงสร้างเดียวกัน (Gronlund. 1982 : 125-138 ; citing Cronbach. 1971 ; Crocker and Algina. 1986 : 231 ; citing Dorlington. 1970 ; Anastasi. 1990 : 154 ; Venn. 1994 : 52 ; citing Angoff. 1988 ; Friedenberg. 1995 : 255)

4. การวัดความสอดคล้องภายใน (Internal consistency) เป็นการพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างข้อสอบภายในฉบับ ถ้าสหสัมพันธ์มีค่าสูงแสดงว่าข้อสอบเหล่านั้นวัดคุณลักษณะเดียว (One-trait) หรือข้อสอบเป็นเอกพันธ์ (Homogeneity) หรืออาจพิจารณาจากคะแนนของแบบทดสอบย่อยกับคะแนนรวมทั้งฉบับ หรืออาจใช้วิธีแบ่งกลุ่มคะแนนของผู้สอบ โดยกำหนดคะแนนเกณฑ์และจัดแบ่งผู้สอบเป็นกลุ่มสูงและกลุ่มต่ำกว่าเกณฑ์และทดสอบสัดส่วนของการสอบผ่าน หรืออาจใช้วิธีการของค่าสหสัมพันธ์ เช่น สหสัมพันธ์แบบไบซีเรียลระหว่างการสอบผ่าน-ไม่ผ่านของข้อสอบแต่ละข้อกับคะแนนรวม (Gronlund. 1982 : 125-138 ; citing Cronbach. 1971 ; Anastasi. 1990 : 155-156)

5. วิเคราะห์กระบวนการตอบข้อสอบ โดยสอบถามผู้ตอบถึงแนวความคิดขณะที่ตอบข้อสอบและวิเคราะห์กระบวนการทางสมองที่ผู้สอบใช้ขณะตอบข้อคำถาม (Gronlund. 1982 : 125-138 ; citing Cronbach. 1971) ซึ่งใช้แนวคิดเกี่ยวกับกระบวนการสารสนเทศ (Information processing) ในการศึกษา (Anastasi. 1990 : 159-161)

6. การวิเคราะห์องค์ประกอบ (Factor analysis) การศึกษาวิธีนี้มีเงื่อนไขว่าจะต้องวัด n ครั้ง หรือ n รูปแบบกับกลุ่มผู้สอบกลุ่มเดียว และคำนวณเมตริกซ์สหสัมพันธ์ระหว่างการวัดทั้งหมด และใช้เทคนิคของการวิเคราะห์องค์ประกอบเพื่อลดจำนวนตัวแปรให้น้อยลง ซึ่งมีวิธีการโดยทั่วไป 2 กรณีคือ

6.1 พิจารณาจากเมตริกซ์สหสัมพันธ์ภายในระหว่างข้อสอบภายในฉบับ ซึ่งถูกจัดไว้เป็นกลุ่มภายใต้โครงสร้างทางทฤษฎีหรือโครงสร้างที่สนใจ และผลที่เกิดจากการวิเคราะห์องค์ประกอบควรเป็นไปตามที่คาดหมาย

6.2 พิจารณาจากเมตริกซ์สหสัมพันธ์ของแบบทดสอบหรือการวัดอื่น ๆ ที่แตกต่างกัน n ชุด จากคะแนนที่ได้จากการวัดที่แสดงลักษณะของความแปรปรวนของหนึ่งหรือหลาย

องค์ประกอบร่วมหรืออาจกล่าวได้ว่า แบบทดสอบฉบับย่อย (Subtest) หรือแบบทดสอบที่ถูกกำหนดให้วัดโครงสร้างเดียวกัน จะต้องมียุทธวิธีประกอบร่วมกันหลังจากวิเคราะห์องค์ประกอบและสามารถที่จะอธิบายองค์ประกอบของแบบทดสอบแต่ละชุด โดยหาค่าสหสัมพันธ์ของคะแนนสอบกับน้ำหนักองค์ประกอบ (Factor loading) ขององค์ประกอบหลัก (Major factors) ซึ่งเรียกว่าความเที่ยงตรงแฟคตอเรียล (Factorial validity) (Crocker and Algina. 1986 : 232 ; Anastasi. 1990 : 154-155 ; Friedenberg. 1995 : 255)

7. เมตริกซ์คุณลักษณะหลายวิธีหลาย (Multitrait-multimethod) แคมป์เบลและฟิสเคีย (Campbell and Fiske. 1959) ได้แสดงแนวคิดเกี่ยวกับการวัดโครงสร้างของแบบทดสอบว่ามีความเหมาะสมมากกว่าการตัดสินโครงสร้าง โดยการยืนยันจากการทำนายความสัมพันธ์เชิงทฤษฎีด้วยการวัดโครงสร้างอื่น ๆ สำหรับเทคนิควิธีการนี้จำเป็นต้องวัดโครงสร้างที่สนใจด้วยวิธีการอย่างน้อยสองวิธี นอกจากนั้นจำเป็นต้องแยกโครงสร้างอื่นที่มีความแตกต่างกัน ซึ่งสามารถวัดได้อย่างเหมาะสมโดยวิธีการเดียวกันในกลุ่มตัวอย่างกลุ่มเดียว และการวัดแต่ละโครงสร้างโดยแต่ละวิธี ซึ่งจะทำให้ได้เมตริกซ์ของค่าสหสัมพันธ์ ซึ่งประกอบด้วยสัมประสิทธิ์สหสัมพันธ์ 3 รูปแบบ คือ

1. สัมประสิทธิ์ความเชื่อมั่น เป็นความสัมพันธ์ระหว่างการวัดโครงสร้างอย่างเดียวกัน โดยวิธีวัดอย่างเดียวกัน ซึ่งควรจะมีค่าสูง

2. สัมประสิทธิ์ความเที่ยงตรงเอกนัย (Convergent validity coefficients) เป็นความสัมพันธ์ระหว่างการวัดโครงสร้างอย่างเดียวกัน โดยใช้วิธีการวัดที่แตกต่างกัน ซึ่งควรจะมีค่าสูงแต่เป็นไปได้ที่จะมีค่าลดลงสาเหตุมาจากวิธีการวัดไม่มีความเชื่อมั่น

3. สัมประสิทธิ์ความเที่ยงตรงจำแนก (Discriminant validity coefficients) หรือความเที่ยงตรงอนกนัย (Divergent validity) เป็นความสัมพันธ์ระหว่างการวัดโครงสร้างที่แตกต่างกัน โดยใช้วิธีการวัดอย่างเดียวกัน (Heterotrait-monomethod coefficients) หรือความสัมพันธ์ระหว่างการวัดโครงสร้างที่ต่างกันโดยใช้วิธีการที่ต่างกัน (Heterotrait-heteromethod coefficients) ซึ่งควรจะมีค่าต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นหรือความเที่ยงตรงเอกนัย (Crocker and Algina. 1986 : 232-234 ; Anastasi. 1990 : 156-157 ; Friedenberg. 1995 : 255)

4. การสังเกตความไม่แปรเปลี่ยนโครงสร้างเป็นรายบุคคลเมื่อวัดด้วยวิธีการที่แตกต่างกัน

เคน (Kane) ได้ศึกษาเกี่ยวกับเรื่องนี้โดยใช้หลักการวิเคราะห์ความแปรปรวนองค์ประกอบ (Analysis of variance components) โดยการประยุกต์ทฤษฎี การสรุปอ้างอิง

(G-theories) โดยตั้งข้อสมมติว่า การวัดแต่ละวิธีเป็นตัวแทนตัวอย่างแบบสุ่มจากเอกภพของการวัดที่เป็นไปได้ทั้งหมด ซึ่งเอกภพนี้อาจจะแปรผันไปตามมิติที่หลากหลายที่เรียกว่าองค์ประกอบ (Facets) ซึ่งอาจจะประกอบด้วยวิธีการวัดที่แตกต่างกัน การจำแนกองค์ประกอบที่เหมาะสมของโครงสร้างขึ้นอยู่กับทฤษฎีที่สนใจและเอกภพที่ต้องการสรุปอ้างอิง ตัวอย่างเช่น การประเมินโครงสร้างของทักษะการอ่านจับใจความ วิธีการวัดโครงสร้างนี้สามารถทำได้โดยใช้ข้อสอบหลากหลายรูปแบบ เช่น แบบเลือกตอบ แบบถูกผิด แบบปลายเปิด แบบจับคู่ เป็นต้น การสรุปอ้างอิงโครงสร้างจะพิจารณาจากคะแนนของผู้สอบควรจะได้ผลการวัดอย่างเดียวกันไม่ว่าจะวัดโดยแบบทดสอบรูปแบบใดก็ตาม ซึ่งศึกษาจากค่า G (G-study) ที่ออกแบบศึกษาโดยให้ผู้สอบแต่ละคนถูกทดสอบด้วยข้อสอบที่แตกต่างกัน n รูปแบบ และถ้าต้องการพิจารณารูปแบบข้อสอบใน G-study ในลักษณะของตัวอย่างสุ่มจากรูปแบบที่เป็นไปได้ทั้งหมดอาจจำเป็นต้องศึกษาจากค่า D (D-studies) ซึ่งรูปแบบข้อสอบก็คือองค์ประกอบแบบสุ่ม และค่าสัมประสิทธิ์การสรุปอ้างอิง (Generalizability) ที่เหมาะสมคือ

$$\rho^2 = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_{pt}^2 + \sigma_e^2}$$

σ_p^2 คือ ความแปรปรวนของผู้สอบ

σ_{pt}^2 คือ ความแปรปรวนของปฏิสัมพันธ์ระหว่างผู้สอบกับรูปแบบข้อสอบ

σ_e^2 คือ ความแปรปรวนที่เหลือทั้งหมดในคะแนนที่ไม่ได้อธิบายไว้

ซึ่ง เคนได้อ้างถึงว่า ค่า ρ^2 คือ สัมประสิทธิ์ความเที่ยงตรง ซึ่งนำเสนอได้เช่นเดียวกับค่าเฉลี่ยสัมประสิทธิ์ความเที่ยงตรงเอกนัยโดยการเลือกวิธีการวัดที่แตกต่างกัน แต่วัดคุณลักษณะอย่างเดียวกันจากเอกภพของวิธีการที่เป็นไปได้ทั้งหมดอย่างสุ่ม

ครอกเกอร์และแอลจินา ได้ให้ข้อสังเกตเกี่ยวกับการตรวจสอบความเที่ยงตรงเชิงโครงสร้างของเคน โดยใช้สัมประสิทธิ์การสรุปอ้างอิงว่าไม่แตกต่างจากการประมาณค่าความเชื่อมั่นของแบบทดสอบโดยทฤษฎีสรุปอ้างอิง จึงทำให้มีคำถามเกิดขึ้นตามมาถึงความแตกต่างระหว่างความเที่ยงตรงและความเชื่อมั่น ซึ่งสามารถอธิบายได้ 3 กรณี คือ

1. เคน แสดงสูตร $(\sigma_p^2 + \sigma_{pt}^2) / (\sigma_p^2 + \sigma_{pt}^2 + \sigma_e^2)$ ว่าเป็นค่าสัมประสิทธิ์สำหรับประมาณค่าความเชื่อมั่น ดังนั้น เขาจึงจำกัดวงการประมาณค่าความเชื่อมั่นสำหรับสัมประสิทธิ์การสรุปอ้างอิงในลักษณะวิธีการวัดที่แตกต่างกันเพื่อกำหนดองค์ประกอบคงที่มากกว่าสุ่มอย่างใดอย่างหนึ่ง (ซึ่ง ลอร์ดและโนวิก (Lord and Novick, 1968) เรียกว่าสัมประสิทธิ์ความเชื่อมั่นเฉพาะ)

2. มโนทัศน์ความแตกต่างระหว่างสัมประสิทธิ์สหสัมพันธ์เชิงอันดับในการประมาณค่าความเที่ยงตรงหรือความเชื่อมั่นขึ้นอยู่กับโครงสร้างและเอกภพของการอ้างอิง ตัวอย่างเช่น ถ้านักพัฒนาแบบทดสอบพิจารณารูปแบบของแบบทดสอบที่ต่างกันเป็นวิธีการวัดอย่างเดียวกัน การศึกษาลักษณะนี้คือความเชื่อมั่น แต่ถ้าพิจารณาว่าเป็นวิธีการวัดที่ต่างกัน การศึกษาลักษณะนี้ก็คือ ความเที่ยงตรง (Crocker and Algina. 1986 : 234-235 ; citing Kane. 1982, Lord and Novick. 1968)

ความเที่ยงตรงเชิงเกณฑ์สัมพันธ์ (Criterion-Related Validation)

ครอกเกอร์และแอลจินา (Crocker and Algina. 1986 : 224) เสนอแนวทางการศึกษาเกี่ยวกับความเที่ยงตรงเชิงเกณฑ์สัมพันธ์ตามลำดับขั้นดังนี้

1. กำหนดเกณฑ์พฤติกรรมและวิธีการวัดที่เหมาะสม
2. กำหนดกลุ่มผู้สอบให้เป็นตัวแทนของกลุ่มที่จะนำแบบทดสอบไปใช้
3. ดำเนินการสอบและเก็บบันทึกข้อมูลของคะแนนแต่ละคน
4. เมื่อได้ข้อมูลที่จะนำมาใช้เป็นเกณฑ์แล้ว ทำการวัดพฤติกรรมการแสดงออกตามเกณฑ์ของผู้สอบแต่ละคน
5. ประมาณค่าความสัมพันธ์ระหว่างคะแนนสอบและเกณฑ์พฤติกรรมแสดงออกที่เหมาะสมกับลักษณะของข้อมูล

วิธีการตรวจสอบความเที่ยงตรงเชิงเกณฑ์สัมพันธ์

กัลลิกเซน (Gulliksen. 1950 : 386-395) ได้แสดงดัชนีความเที่ยงตรงในลักษณะของสัมประสิทธิ์สหสัมพันธ์ของคะแนนสอบกับเกณฑ์บางประการ โดยใช้สูตรประมาณค่าสหสัมพันธ์ของเพียร์สัน (Pearson product-moment correlation)

$$r_{xg} = \frac{N \sum_{i=1}^N A_{ig} X_i - \sum_{i=1}^N A_{ig} \sum_{i=1}^N X_i}{\sqrt{N \sum_{i=1}^N A_{ig}^2 - (\sum_{i=1}^N A_{ig})^2} \sqrt{N \sum_{j=1}^N X_j^2 - (\sum_{j=1}^N X_j)^2}} \quad (1)$$

เมื่อ $A = 1$ เมื่อตอบถูก และ $A = 0$ เมื่อตอบผิด

$$N \text{ จำนวนผู้สอบ} \quad ; N_g = \sum_{i=1}^N A_{ig} = N_p \quad (N_p \text{ ผู้ตอบข้อสอบข้อ } p) \quad (2)$$

$$\text{ในกรณีตอบถูก} \quad \sum_{i=1}^N A_{ig} X_i = \sum_{i=1}^N X_{ig} = N_g X_g = N_{pg} X_g \quad (3)$$

$$N_{pg} = \sqrt{N \sum_{i=1}^N A_{ig}^2 - \left(\sum_{i=1}^N A_{ig} \right)^2} \quad (4)$$

ถ้าให้ Y แทน X และแทนค่าสมการ 2, 3 และ 4 ในสมการ 1 และคูณด้วย S_g ทั้งสองข้าง
จะได้ว่า

$$r_{yg} S_g = \frac{N_g \sum_{i=1}^N Y_{ig} - N_g \sum_{i=1}^N Y_i}{\sqrt{N \sum_{i=1}^N Y_{ig}^2 - \left(\sum_{i=1}^N Y_{ig} \right)^2}} \quad (5)$$

$$\sum_{i=1}^N Y_i = NY \text{ และ } N_r = N s_y = \sqrt{N \sum_{i=1}^N Y_{ig}^2 - \left(\sum_{i=1}^N Y_{ig} \right)^2} \quad (6)$$

แทนสมการ (3) และ (6) ในสมการ (5) หาคะและส่วน โดย N^2 และถอด p_g ออกนอกวงเล็บ จะได้ว่า

$$r_{xg} S_g = p_g \frac{(\bar{Y}_g - \bar{Y})}{\bar{Y}}$$

เมื่อ N คือ จำนวนผู้เข้าสอบทั้งหมด

N_g คือ จำนวนผู้สอบที่ตอบข้อสอบข้อที่ g ถูกต้อง ($g = 1 \dots K$)

P_g คือ N_g/N หรือ สัดส่วนของคนตอบข้อสอบข้อที่ g ได้ถูกต้อง

Y_i คือ คะแนนเกณฑ์ของบุคคลที่ i ($i=1 \dots N$)

$\bar{Y}$ คือ ค่าประมาณของค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของการกระจายของคะแนนเกณฑ์

Y_g คือ คะแนนเกณฑ์ของแต่ละบุคคลที่ตอบข้อ g ถูกต้อง

$\bar{Y}_g$ คือ ค่าเฉลี่ยของคะแนนเกณฑ์ของการตอบข้อ g ถูกต้อง

$r_{vg}S_g$ คือ ดัชนีความเที่ยงตรงของข้อ g

$$r_{vg} = \frac{\sum_{i=1}^N r_{vg} S_g}{\sum_{i=1}^N S_g} = \frac{\sum_{i=1}^N r_{vg} S_g}{\sum_{i=1}^N S_g}$$

เมื่อ $r_{vg}S_g$ คือ ความเชื่อมั่นรายข้อ = $\frac{p_g(\bar{X}_g - \bar{X})}{S_x}$

เมื่อ $\bar{X}_g$ คือ ค่าเฉลี่ยของคะแนนสอบของผู้ตอบข้อ g ถูก

$\bar{X}$ คือ ค่าเฉลี่ยของคะแนนสอบ

S_x คือ ส่วนเบี่ยงเบนมาตรฐานของการกระจายของคะแนนสอบ

r_{vg} คือ ค่าสหสัมพันธ์ระหว่างคะแนนสอบรายข้อกับคะแนนเกณฑ์

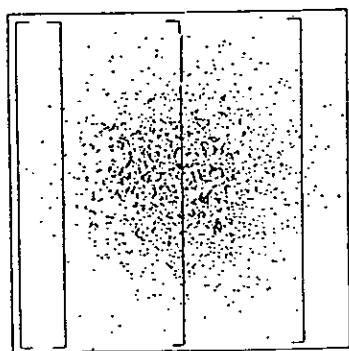
r_{xg} คือ ค่าสหสัมพันธ์ระหว่างคะแนนสอบรายข้อกับคะแนนรวม

S_g คือ ส่วนเบี่ยงเบนมาตรฐานของข้อสอบข้อที่ g หรือ $= \sqrt{P_g(1 - p_g)}$

(Gulliksen. 1950 : 386-395)

นอกจากวิธีการพิจารณาความเที่ยงตรงเชิงเกณฑ์สัมพันธ์โดยเทคนิคทางสถิติแล้ว ออสเตอร์ฮอฟ (Oosterhof. 1994 : 63-66) เสนอวิธีหาความเที่ยงตรงเชิงเกณฑ์สัมพันธ์โดยใช้เทคนิคของการสร้างกราฟโดยพิจารณาจากรูปร่างของจุดคู่ลำดับ (Scatter plot) ของคะแนนพยากรณ์บนแกน X และคะแนนเกณฑ์บนแกน Y โดยมีแนวการพิจารณาภาพประกอบ 2 ดังนี้

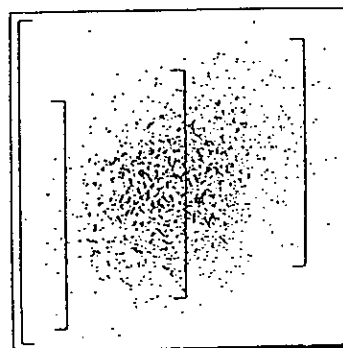
คะแนนเกณฑ์



คะแนนพยากรณ์

$$r_{xy} = .00$$

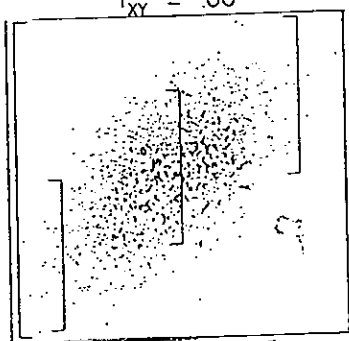
คะแนนเกณฑ์



คะแนนพยากรณ์

$$r_{xy} = .30$$

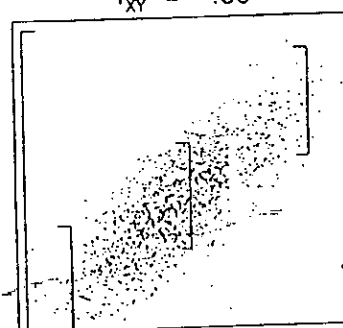
คะแนนเกณฑ์



คะแนนพยากรณ์

$$r_{xy} = .60$$

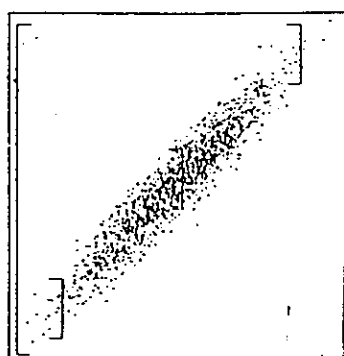
คะแนนเกณฑ์



คะแนนพยากรณ์

$$r_{xy} = .80$$

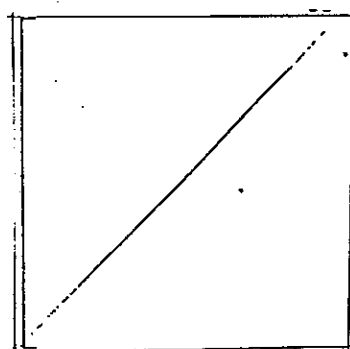
คะแนนเกณฑ์



คะแนนพยากรณ์

$$r_{xy} = .95$$

คะแนนเกณฑ์



คะแนนพยากรณ์

$$r_{xy} = 1.00$$

ภาพประกอบ 2 การกระจายของจุดคู่ลำดับตามลักษณะของค่าสัมประสิทธิ์สหสัมพันธ์

อย่างไรก็ตาม แม้ว่าส่วนใหญ่ยังยอมรับการแบ่งประเภทความเที่ยงตรงออกเป็นความเที่ยงตรงเชิงเนื้อหา ความเที่ยงตรงเชิงโครงสร้าง และความเที่ยงตรงเชิงเกณฑ์สัมพันธ์ แต่ข้อสรุป

ของครอนบัค ในปี 1984 แสดงให้เห็นว่าความเที่ยงตรงทั้ง 3 ชนิดไม่อาจแบ่งแยกกันได้อย่างชัดเจน โดยพยายามอธิบายเชื่อมโยงความเที่ยงตรงทั้ง 3 ชนิดเข้าด้วยกัน ดังนั้นจึงอาจกล่าวได้ว่าความเที่ยงตรงทุกประเภทคือ ความเที่ยงตรงเชิงโครงสร้าง และ APA อธิบายความเที่ยงตรงเพื่อเป็นมาตรฐานในการใช้แบบทดสอบในปี 1985 โดยสรุปว่าความเที่ยงตรงไม่สามารถแบ่งแยกได้ซึ่งเป็นการใช้คะแนนสอบในการสรุปอ้างอิง แสดงความหมายและใช้ประโยชน์ได้อย่างเหมาะสม นอกจากนั้น เมสสิก แสดงแนวคิดเกี่ยวกับการเสนอค่าความเที่ยงตรงที่เน้นเกี่ยวกับการใช้และการนำเสนอแบบทดสอบ ซึ่งออสเตอร์ฮอฟ สรุปไว้ดังภาพประกอบ 3 ดังนี้

	การนำเสนอผลการทดสอบ (Test interpretation)	การใช้ผลการทดสอบ (Test used)
ความชัดเจน (Evidence)	แบบทดสอบวัดโครงสร้างอะไร (Construct validity)	ความเกี่ยวข้องและการใช้ประโยชน์ที่แสดงถึงแสดงถึงความเที่ยงตรงเชิงโครงสร้าง (Construct validity+relevance/utility)
ความสำคัญ (Consequence)	ความรู้จากการวัดโดยแบบทดสอบเหมาะสมและมีคุณค่าต่อสังคม (Value implication)	การนำผลการสอบไปใช้ต้องคำนึงถึงความสำคัญต่อสังคม (Social consequence)

ภาพประกอบ 3 การนำเสนอผลการทดสอบ Messick in Linn. 1989 : 18-22 ; citing Cronbach. 1984, APA. 1985 ; Oosterhof. 1994 : 69)

ความเชื่อมั่นของแบบทดสอบ (Test reliability)

ความเชื่อมั่นนับว่าเป็นคุณสมบัติที่สำคัญของแบบทดสอบเป็นลำดับสองรองจากความเที่ยงตรง (Airasian. 1994 : 2) ซึ่งแสดงคุณภาพของแบบทดสอบเกี่ยวกับความแม่นยำ (Accuracy) และความคงที่ (Consistency) ของคะแนนสอบหรือผลการวัดอื่น ๆ เช่น ทักษะ ความสามารถและพฤติกรรมของผู้เรียน แบบทดสอบที่มีความเชื่อมั่นย่อมให้ผลการสอบที่เหมือนกันไม่ว่าจะอยู่ในเงื่อนไขและสถานการณ์ใด ๆ รวมทั้งผู้ประเมินและสิ่งแวดล้อมในการสอบที่ต่างกัน (Venn. 1994 : 46) นอกจากนั้นยังเป็นมูลฐานของความเที่ยงตรงจนอาจกล่าวได้ว่าแบบทดสอบใดขาดความเชื่อมั่นย่อมขาดความเที่ยงตรงด้วย (Oosterhof. 1994 : 74)

ประเภทของการวิเคราะห์ความเชื่อมั่น (Type of reliability analysis)

1. หาค่าความสัมพันธ์จากคะแนน 2 ชุด โดยใช้สูตรของเพียร์สัน (Pearson product moment correlation) โดยมีพื้นฐานการนิยามคะแนนจริงจากสมการ $X = T + E$

1.1 การวัดซ้ำ (Test-Retest) เทคนิควิธีการนี้จะใช้แบบทดสอบเพียงฉบับเดียวทดสอบกับกลุ่มผู้สอบหนึ่งกลุ่ม จำนวน 2 ครั้ง โดยทิ้งช่วงเวลา ซึ่งความแตกต่างของการตอบทั้งสองครั้งสะท้อนให้เห็นถึงความคลาดเคลื่อนในการวัด (Friedenberg, 1995 : 188-189) แต่ไม่ได้สืบค้นความสอดคล้องภายในระหว่างพฤติกรรมการตอบครั้งแรกและครั้งที่สอง (Oosterhof, 1994 : 81-82) หรืออาจกล่าวได้ว่า การวิเคราะห์ความเชื่อมั่นแบบวัดซ้ำเป็นการตรวจสอบความคงที่ (Stability) ของคะแนนในการวัดหลาย ๆ ครั้ง ซึ่งสอดคล้องกับแนวทฤษฎีแบบมาตรฐานเดิมการบอกถึงความแตกต่างของคะแนนสอบย่อมสะท้อนให้เห็นถึงความแตกต่างของคะแนนจริง อย่างไรก็ตาม ค่าสัมประสิทธิ์ก็ไม่ได้บอกถึงธรรมชาติของความแตกต่างนั้น ข้อจำกัดที่สำคัญของเทคนิควิธีการก็คือระยะเวลาความห่างระหว่างการสอบทั้งสองครั้ง ซึ่งจะมีอิทธิพลต่อการจดจำคำถามของผู้สอบ และการเรียนรู้ข้อสอบ (Test wiseness) (Kline, 1993 : 6 ; Friedenberg, 1995 : 190) นอกจากนี้ความรู้สึก ความเจ็บป่วย ปัญหาทางอารมณ์ ที่นั่งสอบ ดินสอหรือปากกาที่ไม่ดี สิ่งดังกล่าวนี้ล้วนแล้วแต่ทำให้ค่าความเชื่อมั่น โดยเทคนิควัดซ้ำมีค่าลดลงได้นอกจากนี้แล้วคุณสมบัติหลาย ๆ ประเด็นของแบบทดสอบ ได้แก่ กระบวนการสร้างที่ไม่ดี การให้คะแนนที่เป็นอัตนัย การเดา ระดับความยากของข้อสอบ และรวมถึงลักษณะของผู้สอบและขนาดของกลุ่มผู้สอบก็ย่อมส่งอิทธิพลต่อการวิเคราะห์ความเชื่อมั่น โดยวิธีวัดซ้ำได้เช่นกัน ไคลน์ และเวนน์ สรุปว่า ความเชื่อมั่นโดยเทคนิคการวัดซ้ำนี้ ตั้งแต่ .80 ขึ้นไปน่าจะเป็นที่ยอมรับได้การทิ้งช่วงเวลาลงสอบควรใช้เวลาประมาณ 3 เดือน หรืออย่างน้อยประมาณ 2 สัปดาห์ กลุ่มผู้สอบควรเป็นตัวแทนประชากรและมีจำนวนมากพอ อย่างน้อยที่สุดประมาณ 100 คน (Kline, 1993 : 6-9 ; Venn, 1994 : 47)

1.2 การใช้แบบทดสอบต่างฉบับ (Alternate Form) เทคนิควิธีการนี้จะใช้แบบทดสอบ 2 ฉบับ ทดสอบกับกลุ่มผู้สอบกลุ่มเดียว ในเวลาติดต่อกันหรือห่างกันเล็กน้อย ทั้งนี้ขึ้นอยู่กับความยาวหรือเนื้อหาที่วัด ซึ่งความแตกต่างของการตอบแบบทดสอบทั้งสองฉบับสะท้อนให้เห็นถึงความคลาดเคลื่อนในการวัด (Friedenberg, 1995 : 191-192) การใช้แบบทดสอบต่างฉบับในระยะแรกเน้นที่ความเป็นคู่ขนานของแบบทดสอบตามทฤษฎีมาตรฐานเดิม (Parallel test forms) กล่าวคือแบบทดสอบทั้งสองฉบับจะต้องมีคุณสมบัติเท่ากัน 4 ประเด็น สำคัญคือ

1. คะแนนจริงเท่ากัน
2. ความคลาดเคลื่อนในการวัดเท่ากัน ($S_{eg} = S_{eh}$)

3. ค่าสหสัมพันธ์ภายในของแบบทดสอบต้องเท่ากัน

4. ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของคะแนนทั้งสองฉบับต้องเท่ากัน

(Gulliksen, 1950 : 11-14) ข้อจำกัดของแบบทดสอบคู่ขนานแบบมาตรฐานเดิมก็คือ การยึดถือตามเงื่อนไขข้อตกลงทุกข้อในทางปฏิบัติเป็นไปได้ยาก หรือเป็นไปได้เลย ดังนั้น ในระยะต่อมาจึงมีการนำเสนอในลักษณะของแบบทดสอบสมมูล (Tau-equivalent form) ซึ่งมีการผ่อนปรนเงื่อนไขของความคลาดเคลื่อน และคะแนนสอบ คะแนนจริงที่ได้อยู่ในรูปของคะแนนจริงสมมูล (Tau-equivalent) ซึ่งเขียนเป็นสมการเชิงคณิตศาสตร์ได้คือ $T_{pf} = T_{pq} + T_{fg}$ เมื่อ T_{pf} และ T_{pq} คือคะแนนจริงของบุคคลที่ p ที่ทำข้อสอบฉบับ p และ q และ C_p คือค่าคงที่ซึ่งขึ้นอยู่กับจำนวนข้อของแบบทดสอบ แต่ไม่ได้ขึ้นอยู่กับผู้สอบลักษณะของแบบทดสอบนี้จึงเน้นเฉพาะการวัดโดเมนความรู้และคุณลักษณะอย่างเดียวกัน โดยพิจารณารูปแบบคำถาม เนื้อหา รวมทั้งความเท่าเทียมกันของความยาก การพิจารณาความเชื่อมั่นที่ยอมรับได้ควรมีค่าสัมประสิทธิ์สหสัมพันธ์ของแบบทดสอบทั้งสองฉบับไม่ควรต่ำกว่า .90 (Lord & Novick, 1968 : 49-50 ; Crocker and Algina, 1986 : 132 ; Feldt and Brennan in Linn, 1989 : 109-110 ; Anastasi, 1990 : 118-120 ; Gronlund and Linn, 1990 : 82-83 ; Kline, 1993 : 12 ; Oosterhof, 1994 : 82-83 ; Venn, 1994 : 47-48 ; Friedenberg, 1995 : 191-192)

2. ใช้คะแนนชุดเดียว เทคนิควิธีนี้ใช้คะแนน 1 ชุด จากแบบทดสอบชุดใดชุดหนึ่งสอบกับกลุ่มผู้สอบหนึ่งกลุ่ม เพื่อใช้คะแนนสอบสะท้อนถึงความแตกต่างของความรู้หรือคุณลักษณะที่แท้จริงของผู้สอบ หรืออาจกล่าวได้ว่าเป็นการอ้างอิงพฤติกรรมการตอบจากแบบทดสอบฉบับหนึ่งไปยังพฤติกรรมการตอบแบบทดสอบฉบับอื่นที่วัดโดเมนเดียวกัน การประมาณความเชื่อมั่นในลักษณะนี้เรียกว่าเป็นการศึกษาความสอดคล้องภายในฉบับ (Internal consistency) หรือความเป็นเอกพันธ์ของแบบทดสอบ (Homogeneity) กล่าวคือ ถ้าแบบทดสอบมีความสอดคล้องภายในซึ่งแสดงถึงข้อสอบทุกข้อในฉบับวัดในสิ่งที่เกี่ยวข้องกับขอบเขตของความรู้หรือคุณลักษณะของผู้สอบอย่างแท้จริง หรือกล่าวได้อีกอย่างหนึ่งว่า ข้อสอบมีความเป็นเอกพันธ์ซึ่งแสดงว่าอยู่ในโดเมนเดียวกัน (Friedenberg, 1995 : 193-194) ความสอดคล้องภายในของข้อสอบนับว่าเป็นเป้าหมายสำคัญที่นักสร้างแบบทดสอบให้ความสำคัญเป็นเป้าหมายหลักถ้าค่าความสอดคล้องภายในมีค่าต่ำย่อมแสดงให้เห็นว่ามีข้อสอบเพียงส่วนหนึ่งเท่านั้นที่วัดตัวแปรที่ต้องการศึกษา แต่ยังมีอีกส่วนหนึ่งไม่ได้วัดตัวแปรที่ศึกษา และถ้าข้อสอบมีความเที่ยงตรง กล่าวคือ วัดในสิ่งที่จะวัด ค่าความสอดคล้องภายในจะสูงขึ้นไปด้วย ซึ่งกิลฟอร์ด และนันทอลลี ได้แสดงความเห็นว่าค่าความสอดคล้องภายในยิ่งสูงเท่าไรย่อมจะทำให้ความเที่ยงตรงสูงขึ้นไปด้วย แต่ก็ได้รับการโต้แย้งจากแคทเทล (Cattell) ว่าค่าความสอดคล้องภายในยิ่งสูงมากเท่าไรย่อมเป็นการขัดขวางต่อทฤษฎีของ

ความเที่ยงตรง เขาได้ยกตัวอย่างเกี่ยวกับกรณีของแบบทดสอบทางบุคลิกภาพที่วัดหลายคุณลักษณะ ถ้าข้อสอบในฉบับมีความสอดคล้องภายในสูง แสดงถึงข้อสอบชุดนั้นวัดเพียงคุณลักษณะเดียวนับว่าไม่มีความเที่ยงตรง ดังนั้น จึงไม่สามารถสรุปได้ว่าความเชื่อมั่นที่สูงจะทำให้ความเที่ยงตรงสูงตามไปด้วยเสมอไป อย่างไรก็ตาม บาเรตต์และไคลน์ (Barrett and Kline) ได้ให้ข้อสรุปว่า การปรับปรุงแบบทดสอบโดยยึดถือความเชื่อมั่นจากค่าความสอดคล้องภายในนั้นยึดถือได้บ้างเล็กน้อยหรืออาจสรุปได้ว่าความเชื่อมั่นโดยวิธีหาความสอดคล้องภายในยังเป็นสิ่งที่จำเป็น แต่ยังไม่เพียงพอที่จะบอกถึงคุณสมบัติของแบบทดสอบที่ดี (Kline. 1993 : 9-10 ; citing Guilford. 1956, Nunnally. 1978, Cattell. 1977, Barrett and Kline. 1982) การวิเคราะห์ความเชื่อมั่นแบบหาความสอดคล้องภายในแบ่งออกตามแนวคิดพื้นฐานของคะแนนจริง 3 รูปแบบ คือ คะแนนจริงแบบมาตรฐานเดิม (Classical true score) คะแนนจริงสมมูล (Essential tau-equivalent true score) และคะแนนจริงสัมพันธ์ (Congeneric true score) ซึ่งเขียนเป็นรูปแบบเชิงคณิตศาสตร์ดังนี้

1. คะแนนจริงมาตรฐานเดิม

$$T_{pf} = T_{fg}$$

2. คะแนนจริงสมมูล

$$T_{pf} = T_{pg} + C_{fg}$$

3. คะแนนจริงสัมพันธ์

$$T_{pf} = (b)_{fg} T_{pg} + C_{fg}$$

เมื่อ T_{pf} คือ คะแนนจริงของผู้สอบ p ตอบแบบทดสอบฉบับ f

T_{pg} คือ คะแนนจริงของผู้สอบ p ตอบแบบทดสอบฉบับ g

b_{fg}, C_{fg} คือ ค่าคงที่ที่ขึ้นอยู่กับการออกแบบทดสอบแต่ไม่ขึ้นอยู่กับผู้สอบ

(Feldt and Brennan in Linn. 1989 : 110)

การวัดความเชื่อมั่นแบบความสอดคล้องภายในนี้ ไคลน์ (Kline. 1993 : 10-12, 30-39) ให้ข้อคิดเห็นในเชิงสรุปไว้ดังนี้

1. กลุ่มผู้สอบต้องเป็นตัวแทนประชากรที่เป็นกลุ่มเป้าหมายสำหรับการสอบวัดและมีขนาดเพียงพอเพื่อให้เกิดความคลาดเคลื่อนเชิงสถิติที่น้อยที่สุด และขนาดกลุ่มตัวอย่างอย่างน้อยที่สุดควรมีจำนวน 100 คน

2. ค่าสัมประสิทธิ์ α เป็นดัชนีแสดงความสอดคล้องภายในที่ดีที่สุด และความเชื่อมั่นแบบแบ่งครึ่งข้อสอบ แสดงถึงธรรมชาติของความสอดคล้องภายในได้ดี

3. การวิเคราะห์ความเชื่อมั่นโดยแบ่งครึ่งข้อสอบซึ่งส่วนใหญ่ใช้ข้อคู่ข้อคี่เป็นเกณฑ์ในการแบ่งแยก สิ่งสำคัญก็คือควรมีการจัดเรียงลำดับข้อสอบตามความยาก

4. ค่าความเชื่อมั่นแบบแบ่งครึ่งข้อสอบมีความสัมพันธ์กับความยาวของแบบทดสอบถ้าแบบทดสอบมีความยาวมากขึ้น ค่าความเชื่อมั่นจะสูงขึ้น ดังนั้น แบบทดสอบที่มีจำนวนข้อน้อย

กว่า 10 ข้อ ยากที่จะมีค่าความเชื่อมั่นที่สูงขึ้น ซึ่งการใช้ข้อสอบเพียง 10 ข้อ น่าจะเพียงพอสำหรับการเป็นแบบทดสอบที่ดีแต่ข้อสอบทั้ง 10 ข้อนั้น ต้องมีความเป็นเอกพันธ์กัน ซึ่งพิจารณาจากค่าสหสัมพันธ์ภายในของข้อสอบตั้งแต่ .20 ขึ้นไป หรืออาจใช้สูตรของสเปียร์แมน-บราวน์

$$r_{kk} = Kr_{ij}/1-(K-1)r_{ij}$$

เมื่อ Kr_{ij} คือ ค่าความเชื่อมั่นของแบบทดสอบ

K คือ จำนวนข้อสอบ

r_{ij} คือ ค่าสหสัมพันธ์ภายในของข้อสอบเฉลี่ย

5. ค่าความเชื่อมั่นแบบสอดคล้องภายในเป็นการศึกษาเกี่ยวกับความสัมพันธ์ของข้อสอบภายในฉบับ ซึ่งข้อสอบเหล่านี้ได้รับการพิจารณาว่าเป็นตัวอย่างแบบสุ่มของเอกภาพหรือโดเมนของข้อสอบ เช่นเดียวกับการหาความเชื่อมั่นแบบใช้แบบทดสอบต่างฉบับ แต่ข้อสอบถูกจัดแบ่งบรรจุในแบบทดสอบมากกว่าหนึ่งฉบับสำหรับความเชื่อมั่นแบบวัดซ้ำมีลักษณะเหมือนกับ ความสอดคล้องภายในคือ เป็นการหาความสัมพันธ์ของข้อสอบภายในฉบับแต่มีความแตกต่างที่ การดำเนินการสอบที่จัดแบ่งเป็นสองครั้งจึงทำให้เพิ่มแหล่งความคลาดเคลื่อนระหว่างการทดสอบ อย่างไรก็ตามการประมาณค่าความเชื่อมั่นทุกวิธีควรมีลักษณะที่เหมือนกันและการหาความเชื่อมั่นแบบวัดซ้ำจะได้ค่าความเชื่อมั่นที่ต่ำสุดสำหรับการประมาณค่าความเชื่อมั่นสรุปได้ดังตาราง 1 ดังนี้

ตาราง 1 สรุปการประมาณค่าความเชื่อมั่นวิธีต่างๆ

ลักษณะการทดสอบ	สูตร	รูปแบบคะแนนจริง
แบบทดสอบ 2 ฉบับ	สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน	มาตรฐานเดิม
แบบทดสอบ 1 ฉบับ		
ก วัดซ้ำ	สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน	มาตรฐานเดิม
ข แบ่งครึ่งฉบับ ($X = X_1 + X_2$)		
1. สเปียร์แมน-บราวน์ (Spearman-Brown)	$\frac{2r_{x1x2}}{1 + r_{x1x2}}$	มาตรฐานเดิม
2. แองกอฟ-เฟลด์ต์ (Angoff-Feldt)	$\frac{4r_{x1x2}}{\frac{2}{s_x^2} - (\frac{2}{s_{x1}^2} + \frac{2}{s_{x2}^2})/s_x^2}$	มาตรฐานเดิม(ไม่ทราบความยาว)

ลักษณะการทดสอบ	สูตร	รูปแบบคะแนนจริง
3. ฟลานาแกน (Flanagan) ค แบ่งสามส่วน ($X = X_1 + X_2 + X_3$) 1. คริสตอฟ (Kristof)	$\frac{4s_{x_1x_2}^2}{s_x^2}$ $\frac{(s_{x_1x_2}s_{x_1x_3} + s_{x_1x_2}s_{x_2x_3} + s_{x_1x_3}s_{x_2x_3})^2}{s_{x_1x_2}s_{x_1x_3}s_{x_2x_3}s_x^2}$	สมมูล สัมพันธ์
จ. แบ่งหลายส่วน ($X = X_1 + X_2 + \dots + X_n$) 1. ครอนบัค แอลฟา (Cronbach alpha)	$\frac{n}{n-1} \left \frac{s_x^2 - \sum s_{xj}^2}{s_x^2} \right $	สมมูล
2. คูเดอร์- ริชาร์ดสัน 20 (Kuder-Richardson)	$\frac{n}{n-1} \left \frac{s_x^2 - \sum pq_i}{s_x^2} \right $	สมมูล
3. คูเดอร์- ริชาร์ดสัน 21	$\frac{n}{n-1} \left \frac{1 - \bar{X}(n - \bar{X})}{ns_x^2} \right $	สมมูล
4. ราชู (Raju)	$\frac{1}{1 - \sum \lambda_x^2} \left \frac{s_x^2 - \sum s_x^2}{s_x^2} \right $	สัมพันธ์(รู้ความยาว)
5. เฟลด์-กิลเมอร์ (Feldt-Gilmer)	$\frac{(\sum D_i)^2}{(\sum D_i)^2 - \sum D_i^2} \left \frac{s_x^2 - \sum s_\lambda^2}{s_\lambda^2} \right $	สัมพันธ์ (ไม่รู้ความยาว)

(Crocker and Algina. 1986 : 141,153 ; Feldt and Brennan in Linn. 1989 : 115)

ปัจจัยที่ส่งอิทธิพลต่อสัมประสิทธิ์ความเชื่อมั่น

1. ความเป็นเอกพันธ์ของกลุ่มผู้สอบ

ความเป็นเอกพันธ์ของกลุ่มผู้สอบนับว่าเป็นสิ่งสำคัญต่อการพัฒนาและคัดเลือกแบบทดสอบ ซึ่งจะมีค่าความเชื่อมั่นลดลงเมื่อกลุ่มผู้สอบแสดงถึงลักษณะของความเป็นเอกพันธ์มาก

ขึ้นหรือมีความเป็นวิวิธพันธ์ (Heterogeneous) ลดลง และในกรณีข้อสอบยากหรือง่ายเกินไปจะทำให้ช่วงพิสัยของคะแนนและความแปรปรวนคะแนนจริงแคบลง แม้ว่าขนาดของกลุ่มและความแปรปรวนของคะแนนคลาดเคลื่อนจะเท่ากันก็ตาม

2. เวลาในการสอบ

ในการทดสอบที่มีการกำหนดเวลาพบว่า ผู้สอบบางคนทำเสร็จบางคนทำไม่เสร็จ ซึ่งอัตราการทำงานของผู้สอบนั้นจะมีอิทธิพลต่อพฤติกรรมการตอบข้อสอบ การทวนซ้ำของผู้สอบ ความแปรปรวนของอัตราการทำงานจะกลายเป็นส่วนหนึ่งของความแปรปรวนของคะแนนจริงแบบทดสอบบางอย่างมีเป้าหมายเพื่อวัดความเร็วในการปฏิบัติงาน บางแบบทดสอบไม่ได้วัดอัตราการปฏิบัติงานเกี่ยวข้องกับคุณลักษณะที่ต้องการวัด ดังนั้น แบบทดสอบประเภทหลังนี้จึงมีการปล่อยให้ผู้สอบทำเสร็จทุกคนหรือทำเสร็จเป็นส่วนใหญ่ จึงทำให้ค่าประมาณของความเชื่อมั่นสูงขึ้นผิดปกติ โดยเฉพาะกรณีแบบทดสอบเน้นความเร็ว (Speeded Test) และใช้เทคนิคแบ่งครึ่งแบบทดสอบเป็นข้อคู่ข้อคี่ ทั้งนี้ เพราะว่าเมื่อผู้สอบใช้เวลาจนหมดก็จะเกิดความไม่สมบูรณ์เมื่อแบ่งข้อสอบเป็นข้อคู่ข้อคี่ซึ่งก็เป็นเหตุผลเดียวกันกับแนวคิดเกี่ยวกับความสอดคล้องภายในอื่น ๆ ดังนั้นวิธีการวัดซ้ำหรือใช้แบบทดสอบสมมูลน่าจะเป็นวิธีที่เหมาะสมที่สุดสำหรับแบบทดสอบเน้นความเร็วทั้งนี้ เพราะว่าวิธีการดังกล่าวจะแยกการดำเนินการสอบออกเป็น 2 ครั้ง แต่ละครั้ง (ชุด) ก็จะกำหนดเวลาของแต่ละฉบับ ความเชื่อมั่นของแบบทดสอบเต็มฉบับได้รับการประมาณค่า โดยใช้สูตรของ สเปียร์แมน-บราวน์ หรือวิธีการของกัทแมน อย่างไรก็ตามการประมาณค่าความเชื่อมั่นของแบบทดสอบเน้นความเร็วต้องนำเสนออย่างระมัดระวังเกี่ยวกับงานที่ต้องการมากกว่าความสามารถในการทำงานอย่างง่าย ๆ โดยใช้เวลาที่รวดเร็ว

3. ความยาวข้อสอบ

ความยาวของข้อสอบนับว่าเป็นตัวแปรสำคัญต่อความแปรปรวนของคะแนนจริงและคะแนนสอบ ถ้าพิจารณาสูตรของสเปียร์แมนและบราวน์ คือ

$$r_{xx} = \frac{K r_{tt}}{1 + (K-1) r_{tt}}$$

เมื่อ r_{xx} คือ ค่าความเชื่อมั่นใหม่

r_{tt} คือ ค่าความเชื่อมั่นของแบบทดสอบฉบับเดิม

K คือ จำนวนเท่าของความยาวแบบทดสอบฉบับใหม่เมื่อเปรียบเทียบกับความยาวของแบบทดสอบฉบับเดิม

จากสูตรดังกล่าวจะเห็นได้ว่า ถ้าจำนวนข้อมากขึ้น ค่า Kr_u จะมากขึ้น ขณะที่ผลต่างของ $K-1$ จะน้อยลง ซึ่งส่งผลให้ $1+(K-1)r_u$ น้อยลงไปด้วย ในทางตรงข้าม ถ้าความยาวข้อสอบลดค่าของ Kr_u จะลดลง ขณะที่ $1+(K-1)r_u$ เพิ่มขึ้น อย่างไรก็ตามการอธิบายอิทธิพลของความยาวต่อค่าความเชื่อมั่นตามกรณีนี้จะสะท้อนประสิทธิภาพของค่าความเชื่อมั่นเฉพาะกรณีเพิ่มหรือลดข้อสอบที่เป็นคู่ขนานในเชิงเนื้อหาและความยากจากแบบทดสอบฉบับเดิมเท่านั้น (Gulliksen. 1950 : 74-86, 108-126 ; Crocker and Algina. 1986 : 143-146) นอกจากนั้นชูเอเกอร์และคณะได้ทำการศึกษาเกี่ยวกับความคลาดเคลื่อนในการวัดแบบสุ่มซึ่งเป็นตัวบอกประสิทธิภาพของการวัดในเชิงของความเชื่อมั่น โดยใช้เทคนิควิเคราะห์สมการถดถอย ซึ่งพบว่า อายุและสถานภาพของผู้สอบจำนวนข้อ ค่าสหสัมพันธ์ภายในของข้อสอบ และค่าความเชื่อมั่นแบบสอดคล้องภายในเป็นตัวแปรพยากรณ์ที่มีนัยสำคัญต่อการทำนายค่าความเชื่อมั่นแบบวัดซ้ำและความคลาดเคลื่อนในการวัด (Kline. 1993 : 29-30 ; citing Schuerger et al. 1989)

การแสดงผลความเชื่อมั่น

ค่าสัมประสิทธิ์ความเชื่อมั่นมีค่าตั้งแต่ -1.00 ถึง +1.00 โดยที่ความเชื่อมั่น (r) เท่ากับ 0 แสดงว่า แบบทดสอบขาดความเชื่อมั่น ขณะที่ $r = 1.00$ แสดงถึงค่าความเชื่อมั่นที่สมบูรณ์ ขณะที่ $r = .90$ หมายความว่า ผลการวัดหรือคะแนนสอบมีส่วนความแปรปรวนของคะแนนจริง 90% และอีก 10% คือความคลาดเคลื่อนในการวัด โดยทั่วไปแบบทดสอบที่ยอมรับได้ควรเริ่มตั้งแต่ $r = .90$ ถ้าต่ำกว่า .90 ปกติจะแสดงให้เห็นว่า แบบทดสอบมีความไม่เหมาะสม ยกเว้นหลักสูตรพิเศษและการประเมินพฤติกรรมยาก ๆ เช่น ปฏิสัมพันธ์ในสังคม มโนทัศน์แห่งตน และปัญหาทางพฤติกรรม (Venn. 1994 : 46) เช่นเดียวกับไคลน์ (Kline. 1993 : 11-13) ที่เห็นว่าค่าความเชื่อมั่นควรอยู่ประมาณ .90 แต่อย่างน้อยที่สุดไม่ควรต่ำกว่า .70

คุณภาพแบบทดสอบเป็นรายข้อ

คุณภาพรายข้อนับว่ามีความสำคัญต่อคุณภาพของแบบทดสอบเป็นอย่างยิ่ง การศึกษาคุณภาพรายข้อปกติจะผ่านกระบวนการที่เรียกว่า การวิเคราะห์ข้อสอบ (Item analysis) ซึ่งเป็นกระบวนการคำนวณและตรวจสอบคุณสมบัติการตอบข้อสอบของผู้สอบเป็นรายข้อเชิงสถิติ โดยปกติแล้วเป็นการตรวจสอบค่าพารามิเตอร์ของข้อสอบเป็น 3 กลุ่มใหญ่ ๆ คือ

1. แสดงให้เห็นถึงการกระจายการตอบข้อสอบแต่ละข้อ (ค่าเฉลี่ยและความแปรปรวนของการตอบข้อสอบ)

2. แสดงให้เห็นระดับของความสัมพันธ์ระหว่างการตอบข้อสอบกับเกณฑ์ที่สนใจบางประการ

3. แสดงให้เห็นถึงฟังก์ชันของทั้งความแปรปรวนและความสัมพันธ์ของข้อสอบกับเกณฑ์ ซึ่งผลการวิเคราะห์นี้จะเป็ประโยชน์ต่อการคัดเลือกข้อสอบ (Item selecting) เข้าฉบับ หรือการทบทวนข้อสอบ (Test revision) เพื่อแก้ไขปรับปรุงข้อสอบให้มีคุณภาพที่ดีการวิเคราะห์ข้อสอบโดยทั่วไปแล้วเน้นคุณสมบัติ 3 ประการคือ ความยาก (Difficulty) อำนาจจำแนก (Discrimination) และคุณภาพของตัวลวง (Distractor power) ซึ่งคุณสมบัติของข้อสอบเหล่านี้จะทำให้เกิดความเข้าใจเกี่ยวกับระดับความเชื่อมั่นและความเที่ยงตรงของแบบทดสอบได้ดียิ่งขึ้น (Crocker and Algina. 1986 : 311 ; Murphy and Davidshofer. 1988 : 127 ; Nunnally and Bernstein. 1994 : 301 ; Oosterhof. 1994 : 195-196 ; Friedenbergr. 1995 : 260-261)

นอกจากนั้น เมอร์เรนและลีแมน (Mehrens and Lehmann. 1984 : 195) สรุปการใช้ประโยชน์จากการวิเคราะห์ข้อสอบ 7 ประการคือ

1. ช่วยตัดสินคุณค่าหรือคุณภาพของแบบทดสอบ
2. ช่วยให้เกิดข้อเท็จจริงที่ตามมาในการพิจารณาทบทวนแบบทดสอบ
3. ช่วยในการสร้างแฟ้มแบบทดสอบในอนาคต
4. ช่วยเพิ่มทักษะในการสร้างแบบทดสอบ
5. ให้คุณค่าในเชิงวินิจัยและช่วยวางแผนจัดกิจกรรมทางการเรียนในอนาคต
6. เป็นพื้นฐานในการอภิปรายเกี่ยวกับผลของแบบทดสอบ
7. เพิ่มประสบการณ์การเรียนรู้ให้กับผู้เรียน

ฟุสท์ (1958) เฮล์มสตัดเทอร์ (1974) และกรอนลันด์ (1981) (Mehrens and Lehmann. 1984 : 196 ; citing Furst. 1958, Cox. 1964, Helmstadter. 1974 and Gronlund. 1981) ได้ให้ข้อสังเกตเกี่ยวกับการวิเคราะห์ข้อสอบดังนี้ คือ

1. ผลการวิเคราะห์ข้อสอบไม่ได้แสดงคุณสมบัติเชิงเหตุผลเกี่ยวกับความเที่ยงตรงของข้อสอบ

2. ค่าอำนาจจำแนกไม่ได้วัดเกี่ยวกับคุณภาพข้อสอบรายข้อเสมอไป

3. ผลการวิเคราะห์ข้อสอบเป็นสิ่งที่ไม่น่าเชื่อถือ และได้รับอิทธิพลจากกลุ่มผู้เข้าสอบ

4. หลีกเลี่ยงการคัดเลือกข้อสอบรายข้อจากค่าสถิติที่เป็นผลของการวิเคราะห์เท่านั้น ทั้งนี้ เนื่องจากค่าสถิตินั้นอาจได้รับผลกระทบจากการเดาตอบ นอกจากนั้นในบางโอกาสข้อสอบอาจไม่ใช่ตัวแทนของโดเมนของข้อสอบ และบางครั้งเกิดความลำเอียง

การประมาณค่าความยากของแบบทดสอบ

ค่าความยาก (p) หมายถึง สัดส่วนของคนตอบถูกในกรณีการให้คะแนนแบบสองทาง (Dichotomous) ถูก 1 ผิด 0

$$p_j = \frac{\text{จำนวนคนตอบถูกและได้คะแนน 1 ข้อที่ } j}{\text{จำนวนคนตอบข้อที่ } j \text{ ทั้งหมด}}$$

$$\text{เมื่อ } p_j = \frac{H+K}{N}$$

p_j คือ ค่าความยากของข้อที่ j

H คือ จำนวนคนในกลุ่มสูงที่ทำข้อนั้นถูก (จำนวนร้อยละ 20, 27, 30 หรือ 50 นับจากคะแนนสูงสุดลงมา)

L คือ จำนวนคนในกลุ่มต่ำที่ทำข้อนั้นถูก (จำนวนร้อยละ 20, 27, 30 หรือ 50 นับจากคะแนนต่ำสุดขึ้นไป)

N คือ จำนวนคนทั้งหมดที่นำมาใช้วิเคราะห์

ในกรณีให้คะแนนแบบ 1-0 ค่า $p_j = \dots X_{ij} = X_j$ (ค่าเฉลี่ยรายข้อ) โดยมีความแปรปรวน (S^2_j) เท่ากับ $p_j q_j$ หรือ $S^2_j = p_j q_j$ ซึ่งมีค่าสูงสุดของ S^2_j เท่ากับ .25

ในกรณีแบบทดสอบเลือกตอบซึ่งมีโอกาสดังที่ผู้ตอบจะใช้การเดาตอบสูงมากดังนั้นจึงมีการประมาณค่า p ปรับแก้การเดาจากสูตร

$$p_o = p_j + .50/m = p_o = .50/m$$

(เมื่อ m คือจำนวนตัวเลือกและ p_o ปรับแก้การเดา)

$$\text{ค่าความยากของแบบทดสอบ}(\bar{p}) = \frac{\sum p_j}{K} \text{ (เมื่อ } K \text{ คือจำนวนข้อสอบ)}$$

ค่าความยากมาตรฐาน (Δ) หรือดัชนี บรอลเลอร์ (Brolyer's index) เป็นการแปลงร้อยละการตอบถูก (ค่าความยาก) โดยใช้พื้นฐานของโค้งปกติ

$$\Delta = m_w + S_w Z_p$$

เมื่อ Δ คือ ค่าความยากมาตรฐาน

Z_p คือ การเบี่ยงเบนปกติที่สัมพันธ์กับ P (ถ้า $p > .50$, $Z_p < 0$ และถ้า

$$p < .50, Z_p > 0)$$

m_w คือ ค่าเฉลี่ยของคะแนนสอบของกลุ่มผู้สอบ

S_w คือ ส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบของกลุ่ม

สูตรปรับแก้ความยากของกิลฟอร์ด (Guilford. 1936a)

$$P = R_i - \frac{w_i}{f-1}$$

T

เมื่อ R_i คือ จำนวนคนตอบข้อสอบ i ถูก

W_i คือ จำนวนคนตอบข้อสอบ i ผิด

f คือ จำนวนตัวเลือกในแต่ละข้อ

T คือ จำนวนข้อในแบบทดสอบทั้งหมด

สูตรปรับแก้ความยากที่เหมาะสมของไฟร์เดนแบร์ก

$$\text{ความยากที่เหมาะสมเมื่อปรับแก้การเดา} = \frac{1.0 - (\text{โอกาสในการเดาถูก}) + \text{โอกาสในการเดาผิด}}{2}$$

(Gulliksen. 1950 :246-252 ; citing Guilford. 1936a ; Crocker and Algina. 1986 : 311-313 ; Murphy and Davidshofer. 1988 : 130-131 ; Wood in Keeves. 1988 : 376-377 ; citing University of Iowa. 1984 ; Anastasi. 1990 : Kubiszyn and Borich. 1993 : 129-132 ; Oosterhof. 1994 : 196-197 ; Friedenber. 1995 : 260-271)

นอกจากประมาณค่าความยากโดยใช้สูตรต่าง ๆ แล้ว จุง เตห์ฟาน (Fan. 1952 : 3-5) ได้จัดทำตารางสำเร็จ โดยประมาณค่า pH และ pL แล้วเปิดหาค่า p และ Δ จากตาราง

เมื่อ pH คือ สัดส่วนของผู้ตอบถูกในกลุ่มสูง (27% ของคนทั้งหมด)

pL คือ สัดส่วนของผู้ตอบถูกในกลุ่มต่ำ (27% ของคนทั้งหมด)

Δ คือ ความยากมาตรฐาน = $13 + 4X$

X คือ ค่า p ที่เป็นบวกน้อยกว่า .50 และเป็นลบเมื่อมากกว่า .50

เลข 13.4 คือ จำนวนที่ได้มาจากการตัดสิน เมื่อกำหนดค่าที่เป็นลบ และยอมรับว่าเป็นค่าที่ใช้ทั่วไป ซึ่งได้มาจากค่าอินทิเกรต

ในกรณีที่ข้อสอบมีการให้คะแนนไม่ใช่แบบ 0-1 หรือแบบสองทาง (Dichotomous) ให้พิจารณาจากค่าเฉลี่ยและความแปรปรวนในแต่ละข้อแทนความยาก (Millman and Greene in Linn. 1989 : 358 ; citing University of Iowa. 1984 ; Nunnally and Bernstein. 1994 : 302) นอกจากนั้น

ในกรณีที่ข้อสอบมีการให้คะแนนไม่ใช่แบบ 0-1 หรือแบบสองทาง (Dichotomous) ให้พิจารณาจากค่าเฉลี่ยและความแปรปรวนในแต่ละข้อแทนความยาก (Millman and Greene in Linn. 1989 : 358 ; citing University of Iowa. 1984 ; Nunnally and Bernstein. 1994 : 302) นอกจากนั้น อาจใช้สูตรการวิเคราะห์ค่าความยากของแบบทดสอบอัตนัย ซึ่ง วิทเนย์และเซเบอร์ (Whitney and Saber) เสนอไว้ในปี ค.ศ.1970 ดังนี้

$$P = \frac{\sum H + \sum L - (2N\text{Score}_{\min})}{2N(\text{Score}_{\max} - \text{Score}_{\min})}$$

เมื่อ P คือ ค่าความยาก

$\sum H$ คือ คะแนนรวมของกลุ่มสูง (25%)

$\sum L$ คือ คะแนนรวมของกลุ่มต่ำ (25%)

N คือ 25% ของจำนวนคนเข้าสอบ

$\text{Score}_{\max}$ คือ คะแนนสูงสุดที่ผู้สอบได้ในแต่ละข้อ

$\text{Score}_{\min}$ คือ คะแนนต่ำสุดที่ผู้สอบได้ในแต่ละข้อ

(Mehrens and Lehmann. 1984 : 198-199 ; citing Whitney and Sabers. 1970)

การประมาณค่าอำนาจจำแนกของข้อสอบ

ค่าอำนาจจำแนก (r) หมายถึง ดัชนีที่บ่งบอกถึงคุณลักษณะของข้อสอบเกี่ยวกับความแตกต่างระหว่างบุคคลในโครงสร้างที่ต้องการวัดโดยแบบทดสอบหรือเกณฑ์ภายนอกบางประการ ค่าอำนาจจำแนกมีค่าการกระจายอยู่ในช่วงระหว่าง -1.00 ถึง 1.00 ซึ่งมีวิธีการประมาณค่าได้ 2 วิธี คือ ประมาณค่าดัชนีการจำแนก (D) และประมาณค่าจากสัมประสิทธิ์สหสัมพันธ์ (r)

1. ค่าดัชนีการจำแนก (D) ได้จากการใช้คะแนนรวมเพื่อจัดกลุ่มผู้สอบที่มีคะแนนสูงซึ่งมั่นใจว่ามีความน่าจะเป็นในการตอบข้อสอบถูกต้องมากกว่ากลุ่มที่ถูกจัดให้อยู่ในกลุ่มต่ำซึ่งคาดว่าจะมีโอกาสในการตอบข้อสอบถูกน้อยมาก ดังนั้นค่าดัชนีการจำแนกจึงเป็นค่าที่บอกความแตกต่างของจำนวนผู้ตอบถูกในกลุ่มสูงและกลุ่มต่ำเปรียบเทียบกับจำนวนคนในกลุ่มใดกลุ่มหนึ่งซึ่งเขียนแทนได้โดยสูตรดังนี้ (Crocker and Algina. 1986 : 313 ; citing Kelly. 1939 , Elglehert. 1965 , Beuchert and Mendoza. 1979)

$$D = \frac{H - L}{NH}$$

เมื่อ D คือ ดัชนีการจำแนก

H คือ จำนวนคนที่ตอบถูกในกลุ่มสูง

L คือ จำนวนคนที่ตอบถูกในกลุ่มสูง

NH คือ จำนวนคนในกลุ่มสูง

อนึ่ง ในการจำแนกกลุ่มผู้สอบเป็นกลุ่มสูงและกลุ่มต่ำนั้นอาจใช้จำนวนข้อมูลของผู้สอบตามเทคนิค 25%, 27%, 30%, 33% หรือ 50% เช่นเดียวกับค่าความยาก สำหรับค่า D นี้ อีเบล (Ebel) เสนอเมื่อปี 1965 เกี่ยวกับเกณฑ์ในการนำเสนอค่า D ดังนี้

1. ถ้า $D > .40$ ข้อสอบมีประสิทธิภาพ
2. ถ้า $.30 < D < .39$ ควรปรับปรุงเล็กน้อยหรือไม่ปรับปรุงก็ได้
3. ถ้า $.20 < D < .29$ เกือบใช้ไม่ได้และจำเป็นต้องมีการปรับปรุง
4. ถ้า $D < .19$ ข้อสอบควรถูกตัดทิ้งหรือปรับปรุงทั้งหมด

2. ค่าสัมประสิทธิ์สหสัมพันธ์ (r) เป็นค่าความสัมพันธ์ระหว่างผลการตอบในแต่ละข้อกับคะแนนรวม หรือเกณฑ์ภายนอกอื่น ๆ โดยใช้สูตรประมาณค่า 4 แบบคือ

2.1 สัมประสิทธิ์สหสัมพันธ์แบบพอยท์ไบซีเรียล (Point biserial correlation ; r_{pbis})

เป็นการศึกษาที่มุ่งสนใจถึงความสัมพันธ์ของผลการตอบข้อสอบรายข้อที่มีการให้คะแนน 1-0 กับคะแนนรวม หรือเกณฑ์ที่มีการกระจายแบบต่อเนื่องอื่น ๆ สูตรการคำนวณทำให้ง่ายขึ้น จากสูตรของเพียร์สันดังนี้

$$\text{or } r_{pbis} = \frac{(\mu_+ - \mu_x) (p/q)^{1/2}}{\sigma_x}$$

$$r_{pbis} = \frac{(\mu_+ - \mu_o) (p/q)^{1/2}}{\sigma_x}$$

เมื่อ μ_o คือ คะแนนเกณฑ์เฉลี่ยสำหรับผู้สอบที่ตอบข้อนั้นผิด

μ_+ คือ คะแนนเกณฑ์เฉลี่ยสำหรับผู้สอบที่ตอบข้อนั้นถูก

μ_x คือ คะแนนเฉลี่ยของคะแนนรวมทั้งหมด

σ_x คือ ความเบี่ยงเบนมาตรฐานของคะแนนทั้งหมด

p คือ ค่าความยาก

q คือ $1 - p$

ค่าของ r_{pbis} ระหว่างข้อสอบกับคะแนนรวมบางครั้งดูเหมือนว่าเป็นค่าไม่แท้จริง ทั้งนี้ เพราะว่าคะแนนสอบเป็นส่วนที่ทำให้เกิดคะแนนรวมของผู้สอบแต่ละคน ถ้าจำนวนของข้อสอบมี มาก ๆ (25 ข้อขึ้นไป) อย่างไรก็ตาม ถ้าจำนวนข้อสอบน้อยข้อปัญหานี้อาจแก้ไขโดยใช้สูตร

$$\rho_{i(x-i)} = \frac{\rho_{xi} \sigma_x - \sigma_i}{\sqrt{\sigma_i^2 - \sigma_x^2 - 2\rho_{xi}\sigma_x\sigma_i}}$$

เมื่อ $\rho_{i(x-i)}$ คือ ค่าความสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนรวม เมื่อข้อสอบข้อนั้นถูกตัดออกไป และ σ_x และ σ_i คือ ส่วนเบี่ยงเบนมาตรฐานของคะแนนรวมและคะแนนรายข้อและ ρ_{xi} คือ ความสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนรวม

2.2 สัมประสิทธิ์สหสัมพันธ์ไบซีเรียล (Biserial correlation coefficient) r_{bis} ในกรณีที่น่าจะมั่นใจว่าตัวแปรแฝง ซึ่งอยู่ในรูปของการตอบข้อสอบที่มีลักษณะการกระจายแบบโค้งปกติ อาจใช้สูตร

$$r_{bis} = \frac{\mu_+ - \mu_x}{\sigma_x} \sqrt{p/y}$$

เมื่อ ρ_{bis} คือ สัมประสิทธิ์สหสัมพันธ์แบบไบซีเรียล

$\bar{y}$ คือ ค่าออร์ดิเนตหรือความสูงของโค้งปกติ ณ ที่คะแนน Z

แสดงค่าของ p ของข้อสอบ

$$\text{ค่า } r_{bis} = \frac{\sqrt{pq}}{\bar{y}} r_{pbis}$$

(Lord and Novick. 1968 : 314-318 ; Crocker and Algina. 1986 : 314-318 ; Wood in Keeves. 1988 : 377-379 ; Anastasi. 1990 : 210-219 ; Nunnally and Bernstein. 1994 : 126-127, 301-304 ; Oosterhof. 1994 : 126-127,301-304 ; Friedenberg. 1995 : 271-279)

ครอกเกอร์ และแอลจินา (Crocker and Algina. 1986 : 319-320) ได้กล่าวสรุปถึงการเลือกใช้เทคนิควิธีการประมาณค่าอำนาจจำแนกของข้อสอบที่มีการให้คะแนนแบบถูก 1 ผิด 0 ดังนี้

1. เมื่อข้อสอบมีความยากปานกลาง การหาค่าสถิติของอำนาจจำแนกด้วยวิธีต่าง ๆ มีความแตกต่างกันน้อยมาก

2. ถ้าเป้าหมายต้องการคัดเลือกข้อสอบที่มีช่วงความยากใดความยากหนึ่งควรใช้ r_{bis} (ทั้งนี้ต้องสร้างข้อตกลงว่าคุณลักษณะภายใต้การตอบข้อสอบนั้นมีการกระจายแบบปกติ)

3. ถ้านักพัฒนาแบบทดสอบคาดหวังว่า กลุ่มผู้ที่สอบจะแสดงความแตกต่างทางความสามารถจากกลุ่มที่ใช้วิเคราะห์ข้อสอบ ควรใช้ r_{bis} เพราะว่าค่า r_{bis} จะให้ค่าคงที่แม้ว่ากลุ่มตัวอย่างจะแปรเปลี่ยนความสามารถ หรืออาจกล่าวได้ว่า ค่า r_{bis} มีค่าต่ำสำหรับกลุ่มความสามารถสูงทุกระดับ แต่ค่า r_{bis} จะมีค่าต่ำเมื่อกลุ่มผู้สอบมีความสามารถต่ำ (หรือกลุ่มที่มีความสามารถสูง) ซึ่งที่จริงแล้วอาจจะมีความยากของข้อสอบ ดังนั้น จึงไม่อาจกล่าวได้ว่าข้อสอบมีค่าอำนาจจำแนกต่ำ

4. ถ้านักพัฒนาแบบทดสอบมั่นใจว่ากลุ่มผู้สอบที่เป็นเป้าหมายมีลักษณะคล้ายคลึงกับกลุ่มผู้สอบที่ใช้วิเคราะห์และมีเป้าหมายที่จะคัดเลือกข้อสอบที่มีความสอดคล้องภายในสูง ควรจะใช้ r_{bis} ซึ่งค่า r_{bis} นี้จะมีค่าสูงมากเมื่อข้อสอบมีความยากปานกลาง และข้อสอบจะให้ค่าความแปรปรวนร่วมสูงสุด ดังนั้น จะทำให้ค่าสัมประสิทธิ์ α มีค่าสูงสุด

5. ในบางกรณีที่ข้อสอบและเกณฑ์มีการให้คะแนนแบบสองทางควรใช้การประมาณค่าโดยสัมประสิทธิ์ Phi (Φ) หรือสัมประสิทธิ์เตทระคอรริก (Tetrachoric ; r_{tet}) สำหรับค่า ϕ (Φ) จะคำนวณง่ายแต่มีข้อจำกัดเกี่ยวกับสัดส่วนในการแบ่งแบบสองทางไม่เท่ากัน ซึ่งตัวแปรทั้งสองจะต้องเป็นคะแนนแบบสองทางแท้จริง (True dichotomous)

$$\phi = \frac{p_{ik} - p_j p_k}{p_j q_j p_k q_k}$$

เมื่อ p_{ik} คือ สัดส่วนร่วมที่แสดงถึงผู้สอบตอบข้อ j และ k ถูกต้อง

$p_j p_k$ คือ สัดส่วนของผู้ตอบถูกข้อ j และ k

$q_j q_k$ คือ $1 - p_j$, $1 - p_k$

ส่วน r_{tet} จะต้องอยู่บนพื้นฐานข้อตกลงว่าคะแนนสอบและคะแนนเกณฑ์เป็นตัวแปรที่มีการกระจายแบบปกติแต่ถูกจัดแบ่งให้เป็นคะแนนแบบสองทาง

ในกรณีที่ข้อสอบมีการให้คะแนนไม่ใช่ 0-1 อาจหาค่าอำนาจจำแนกได้ 2 วิธีคือ

1. ประมาณค่าดัชนีการจำแนก (D) จากสูตรของไวท์เนย์และเซเบอร์ (Whitney and Saber)

$$D = \frac{\sum H - \sum L}{N(\text{Score}_{\max} - \text{Score}_{\min})}$$

เมื่อ $\sum H$ คือ คะแนนรวมของคนในกลุ่มสูง (25% ของคนที่ได้คะแนนตั้งแต่สูงสุดลงมา)

$\sum L$ คือ คะแนนรวมของคนในกลุ่มสูง (25% ของคนที่ได้คะแนนตั้งแต่ต่ำสุดขึ้นไป)

N คือ จำนวนคนในกลุ่มสูงหรือกลุ่มต่ำ

$Score_{max}$ คือ คะแนนสูงสุดที่ผู้สอบได้ในแต่ละข้อ

$Score_{min}$ คือ คะแนนต่ำสุดที่ผู้สอบได้ในแต่ละข้อ

(Mehrens and Lehnann. 1984 : 198 ; citing Whitney and Sabers. 1970)

2. ประมาณค่าความสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนรวม จากสูตรของเพียร์สัน

$$r_{xy} = \frac{(\sum xy/N) - (\bar{X})(\bar{Y})}{\sigma_x \sigma_y}$$

x คือ คะแนนรายข้อ

y คือ คะแนนรวมทั้งฉบับ

$\bar{x}$ คือ คะแนนเฉลี่ยรายข้อ

$\bar{y}$ คือ คะแนนเฉลี่ยรวมทั้งฉบับ

σ_x คือ ความแปรปรวนของคะแนนรายข้อ

σ_y คือ ความแปรปรวนของคะแนนรวมทั้งฉบับ

N คือ จำนวนผู้เข้าสอบ

(Friedenberg. 1995 : 277)

ในกรณีศึกษาค่าอำนาจจำแนกเฉลี่ยของแบบทดสอบให้เปลี่ยนค่า r เป็นค่า Fisher Z แล้วใช้สูตร $\bar{Z} = \frac{\sum \text{Fisher } Z}{K}$ เมื่อ K คือ จำนวนข้อ จากนั้นเปลี่ยนค่า Z เป็น r

ค่าความสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนรวมทั้งฉบับเมื่อนำมาคูณกับส่วนเบี่ยงเบนมาตรฐานรายข้อจะได้ค่าดัชนีความเชื่อมั่นรายข้อ (Item reliability index)

$$r_{ii} = r_{ir} s_i$$

เมื่อ r_{ii} คือ ความเชื่อมั่นรายข้อ

r_{ir} คือ ความสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนรวม (อำนาจจำแนก)

s_i คือ ส่วนเบี่ยงเบนมาตรฐานรายข้อ ซึ่งเท่ากับ p_q หรือ

$$\sqrt{p(1-p)} \text{ หรือ } \sqrt{p-p^2} \text{ (เมื่อ } p \text{ คือค่าความยาก)}$$

และในกรณีเช่นเดียวกัน ค่าความสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนเกณฑ์ภายนอกอื่นเมื่อนำมาคูณกับส่วนเบี่ยงเบนมาตรฐานรายข้อจะได้ค่าดัชนีความเที่ยงตรงรายข้อ (Item validity index)

$$r_{cc} = r_{icS_i}$$

เมื่อ r_{cc} คือ ค่าความเที่ยงตรงรายข้อ

r_{ic} คือ สหสัมพันธ์ระหว่างคะแนนรายข้อกับเกณฑ์ภายนอก

s_i คือ ส่วนเบี่ยงเบนมาตรฐานรายข้อ

(Gulliksen. 1950 : 365-395 ; Crocker and Algina. 1986 : 320-321)

การวิเคราะห์ประสิทธิภาพตัวลวง (Distractor power analysis)

การวิเคราะห์ตัวลวงเป็นกระบวนการตรวจสอบคุณภาพของตัวลวงในแบบทดสอบเลือกตอบ โดยพิจารณาจากร้อยละของผู้ที่เลือกตอบในแต่ละตัวเลือก ตัวลวงนับว่ามีความสำคัญต่อการเดาตอบมาก ถ้าตัวลวงไม่ดีโอกาสที่ผู้สอบจะเลือกตอบคำตอบโดยวิธีการเดาย่อมมีมาก การเดานับว่าช่วยลดความเชื่อมั่นและสร้างข้อจำกัดต่อภาพของความเที่ยงตรงของแบบทดสอบเป็นอย่างมาก ดังนั้น การวิเคราะห์ประสิทธิภาพของตัวลวงจึงนับว่ามีส่วนในการเพิ่มความเชื่อมั่นและความเที่ยงตรงของแบบทดสอบอีกทางหนึ่ง โดยปกติแล้วพฤติกรรมตอบข้อสอบเลือกตอบของผู้สอบมี 2 ลักษณะคือ

1. ผู้ที่มีความรู้หรือทักษะเกี่ยวกับแบบทดสอบส่วนใหญ่จะค้นหาตัวเลือกถูก
2. ผู้ที่ไม่มีความรู้จะสุ่มเลือกจากตัวเลือกที่เป็นไปได้

ในกรณีของผู้สอบที่เลือกตอบแบบสุ่ม โอกาสที่ตัวเลือกทุกตัวจะได้รับการเลือกตอบจะมีโอกาสเท่ากัน บางคนอาจตอบถูก ส่วนที่เหลือคือคนตอบผิดซึ่งน่าจะมีการกระจายในแต่ละตัวลวงเท่า ๆ กัน จำนวนผู้สอบที่คาดหวังว่าจะเลือกตัวลวงแต่ละตัวลวงโดยวิธีแบบสุ่มเรียกว่า ประสิทธิภาพที่คาดหวังของตัวลวง (Expected distractor power) ซึ่งคำนวณได้จากสูตร

จำนวนผู้ที่ตอบผิดในแต่ละข้อ

จำนวนของตัวลวง

จากนั้นนำค่าที่ได้ไปเปรียบเทียบกับประสิทธิภาพเชิงประจักษ์ของตัวลวง (Actual distractor power) หรือจำนวนผู้เลือกตอบในแต่ละตัวลวง ตัวลวงใดที่ไม่มีผู้เลือกตอบหรือเลือกตอบน้อยกว่าประสิทธิภาพที่คาดหวังของตัวลวงควรเน้นพิจารณาเป็นพิเศษหรืออาจจะต้องเขียนใหม่อย่างไรก็ตามอาจเป็นไปได้ว่าตัวลวงเหล่านั้นเป็นคำตอบผิดจริง ๆ สำหรับตัวลวงที่มี

ผู้เลือกตอบมากกว่าค่าประสิทธิภาพที่คาดหวังอย่างมากจำเป็นต้องเขียนใหม่เช่นเดียวกัน ทั้งนี้ เพราะว่าตัวเลือกเหล่านี้อาจจะมีลักษณะที่คล้ายกับตัวเลือกถูก ซึ่งแสดงความเป็นไปได้ 2 กรณี คือ ตัวลวงสะท้อนความรู้บางส่วน (Partial knowledge) หรือโครงสร้างของคำถามมีจุดอ่อนจนทำให้ผู้สอบที่มีความรู้เกิดการไขว้เขวในการเลือกคำตอบถูก (Murphy and David Shoffer. 1989 : 128-130 ; Gay. 1991 : 255-257, Kubiszyn and Borich. 1993 : 130-140 ; Oosterhof. 1994 : 199-202 ; Friedenber. 1995 : 285-288)

ปฏิสัมพันธ์ระหว่างคุณสมบัติของข้อสอบต่าง ๆ

แม้ว่าค่าความยาก ค่าอำนาจจำแนก และการวิเคราะห์ประสิทธิภาพตัวลวงดูเหมือนจะมีความแตกต่างกันในเชิงมโนทัศน์ แต่สิ่งที่ปรากฏพบว่ามีความสัมพันธ์กัน ในการตรวจสอบตัวลวงพบว่ามี mốiเกี่ยวข้องในบางประการกับค่าความยาก ค่าความยากมีผลโดยตรงกับประสิทธิภาพการจำแนกโดยที่ค่าความยากจะเป็นตัวจำกัดอำนาจจำแนกของข้อสอบ กล่าวคือ ถ้าทุกคนทำถูกหมดหรือผิดหมด ค่าความยากจะเท่ากับ 1.00 และ 0 ($P = 1.00$ และ $P = 0$) ตามลำดับ ค่าอำนาจจำแนกจะมีค่าเท่ากับ 0 หรืออาจกล่าวได้ว่าข้อสอบไม่มีอำนาจจำแนก ซึ่งความสัมพันธ์นี้ของ P และ D แสดงได้ดังนี้

ค่า P (ความยาก)	ค่าสูงสุดของ D (อำนาจจำแนก)
1.00	.00
.90	.20
.80	.40
.70	.60
.60	.80
.50	1.00
.40	.80
.30	.60
.20	.40
.10	.20
.00	.00

สำหรับประสิทธิภาพของตัวลวงกับค่าอำนาจจำแนกมีส่วนสัมพันธ์กันเช่นกัน ข้อสอบที่ขาดประสิทธิภาพของการจำแนกแสดงถึงคุณลักษณะของตัวลวงที่ไม่มีประสิทธิภาพ ตัวลวงที่ไม่มี

เหตุผลจะทำให้ค่าความยากของข้อสอบลดลง ในกรณีที่ค่า D หรือค่าความสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนรวมมีค่าเป็นลบ ควรพิจารณาตัวลงอย่างระมัดระวัง และในกรณีที่ตัวลงแสดงเหตุผลที่ทำให้ผู้สอบระลึกถึงคำตอบที่ถูกต้อง อาจเป็นไปได้ว่า ผู้สอบนั้นมีความรู้บางส่วนในการทำข้อสอบ (Murphy and Davidsofer. 1988 : 136-138 ; Friedenber : 1995 ;288-289)

การคัดเลือกข้อสอบ (Item selecting)

กระบวนการคัดเลือกข้อสอบส่วนใหญ่จะกระทำต่อเนื่องจากกระบวนการวิเคราะห์ข้อสอบ โดยมีความเชื่อว่าถ้าสามารถคัดเลือกข้อสอบที่มีคุณภาพเข้าฉบับแล้วจะทำให้แบบทดสอบฉบับนั้น ๆ มีคุณภาพไปด้วย โดยทั่วไปในการคัดเลือกข้อสอบพิจารณาจากค่าความยากและค่าอำนาจจำแนกของข้อสอบ อีเบล (Ebel) เสนอแนะเกณฑ์การพิจารณาดังนี้การจำแนกของข้อสอบสำหรับการคัดเลือกข้อสอบควรมีค่ามากกว่า .00 อย่างมีนัยสำคัญ สำหรับสัมประสิทธิ์สหสัมพันธ์ พี (ϕ) และพอยท์ไบซีเรียล (r_{pbis}) จะประมาณค่าความคลาดเคลื่อนมาตรฐานเมื่อเปรียบเทียบกับสหสัมพันธ์แบบเพียร์สัน จากสูตร

$$\hat{\sigma}_p = \frac{1}{\sqrt{N-1}}$$

เมื่อ N คือขนาดกลุ่มตัวอย่าง (ในการใช้สูตรนี้ต้องมีข้อตกลงว่าขนาดกลุ่มตัวอย่าง อย่างน้อยเท่ากับ 50 คน) โดยปกติค่าวิกฤตต่ำสุดกำหนดความคลาดเคลื่อนมาตรฐานมากกว่า .00 เท่ากับ 2 ดังนั้น ถ้าใช้กลุ่มตัวอย่าง 101 คน จะได้ว่า

$$\hat{\sigma}_p = \frac{1}{\sqrt{101-1}} = \frac{1}{10} = .10$$

ดังนั้น ถ้าต้องการคัดเลือกข้อสอบด้วยค่า r_{pbis} หรือ ϕ จะต้องพิจารณาค่าอำนาจจำแนกตั้งแต่ .20 ขึ้นไป ซึ่ง $.20 = .00 + 2(.10)$

สำหรับค่าสหสัมพันธ์แบบไบซีเรียล r_{bis} ค่าความคลาดเคลื่อนมาตรฐานประมาณค่าได้จากสูตร

$$r_{bis} = \sqrt{pq(N-1)/Y}$$

เมื่อ p คือ ค่าความยาก ; q = (1-p) ; N คือ ขนาดกลุ่มตัวอย่าง และ Y คือ ออร์ดิเนตของโค้งปกติ ค่าของ P ค่า r_{bis} จะมีค่าน้อยที่สุดเมื่อ p = .50 และจะเพิ่มขึ้นเมื่อค่า p เท่ากับ 0 หรือ 1.00 ดัง

นั้น ความคงที่ของค่าสถิติจากกลุ่มตัวอย่างหนึ่งไปยังอีกกลุ่มตัวอย่างหนึ่งซึ่งได้รับอิทธิพลอย่างมากจากค่าความยากข้อสอบ ด้วยเหตุนี้จึงดูเหมือนว่าค่าความยากข้อสอบเป็นเกณฑ์เบื้องต้นในกระบวนการคัดเลือกข้อสอบแบบอิงกลุ่ม แต่โดยทั่วไปแล้วมีความสำคัญน้อยกว่าค่าอำนาจจำแนก นอกจากนั้นค่าความยากจะแปรเปลี่ยนไปตามกลุ่มตัวอย่าง ดังนั้น ถ้าจะใช้แบบทดสอบเพื่อจำแนกความสามารถควรเลือกข้อสอบที่มีความยากปานกลางแต่ในเชิงพาณิชย์นิยมที่จะให้มีข้อสอบง่าย ปานกลางและยากจำนวนเท่า ๆ กัน อย่างไรก็ตามในเชิงปฏิบัตินิยมใช้ข้อสอบที่มีความยากปานกลางเพื่อช่วยแก้ปัญหาการเดาตอบ และจากการวิจัยของนักวัดผลหลายคนพบว่า ถ้าความยากข้อสอบปานกลาง ค่าความสัมพันธ์ของคะแนนรายข้อกับคะแนนรวมจะมีค่าปานกลางเช่นเดียวกัน ตัวอย่างเช่น ค่าเฉลี่ย r_{bis} .30 ถึง .40 ค่าความยากที่ควรเป็นน่าจะอยู่ในช่วงระหว่าง .40 ถึง .60 แต่ถ้าค่าเฉลี่ย r_{bis} สูงกว่า .60 ช่วงของความยากอาจจะมากขึ้น (Crocker and Algina. 1986 : 322-327 ; citing Lord, 1952, Warrington. 1952, Ebel. 1965, and Henryssen. 1971) อย่างไรก็ตาม ไคลน์ (Kline. 1993 : 149) เสนอเกณฑ์ในการคัดเลือกข้อสอบควรพิจารณาค่าความยากในช่วง 0.2-0.8 และค่าอำนาจจำแนกแบบความสัมพันธ์คะแนนรายข้อกับคะแนนรวมตั้งแต่ 0.3 ขึ้นไปเช่นเดียวกับไฟรเดนแบร์ก (Friedenberg. 1995 : 289-293) ซึ่งเสนอค่าอำนาจจำแนกที่เหมาะสมตั้งแต่ 0.3 ขึ้นไป แต่ค่าความยากที่เหมาะสมควรอยู่ระหว่าง .40 ถึง .60 และค่าทั้งสองสามารถนำมาสร้างเป็นเส้นภาพแสดงความสัมพันธ์ในลักษณะคู่ลำดับ นอกจากนั้นไฟรเดนแบร์กยังเสนอแนวทางการพิจารณาข้อสอบโดยพิจารณาจากเส้นคุณลักษณะที่แสดงถึงการจำแนกความสามารถของผู้สอบในแต่ละช่วงคะแนน โดยสร้างกราฟความสัมพันธ์ระหว่างสัดส่วนการตอบถูกกับช่วงคะแนนรวม

ทฤษฎีการสรุปอ้างอิง (G-theory ; GT)

ทฤษฎีการสรุปอ้างอิงเริ่มต้นแนวคิดโดยครอนบัคและคณะ (Cronbach et al) ในปี 1963 เป็นทฤษฎีทางสถิติที่กล่าวถึงความเชื่อถือได้ (Dependability) ของการวัดพฤติกรรมที่อ้างถึงประสิทธิภาพในการสรุปอ้างอิงจากคะแนนที่สังเกตได้ของผู้สอบจากแบบทดสอบหรือการวัดตัวอย่างใดอย่างหนึ่ง (เช่น การสังเกตพฤติกรรม การสำรวจความคิดเห็น) ไปยังคะแนนเฉลี่ยซึ่งผู้สอบผู้นั้นควรจะได้รับภายใต้เงื่อนไขที่เป็นไปได้ทั้งหมดที่ผู้ใช้แบบทดสอบให้การยอมรับ จากแนวคิดนี้จำเป็นต้องสร้างข้อตกลงว่า ความรู้ เจตคติ ทักษะ หรือคุณลักษณะต่าง ๆ ที่ทำการวัดจะต้องอยู่ในสถานะไม่เปลี่ยนแปลง นอกจากนั้นความแตกต่างของคะแนนของผู้สอบแต่ละคนที่เกิดจากการวัด

หลาย ๆ สถานการณ์ย่อมเกิดความคลาดเคลื่อนจากแหล่งใดแหล่งหนึ่งหรือหลายแหล่งและความคลาดเคลื่อนนี้ไม่มีการเปลี่ยนแปลงอย่างมีระบบเนื่องจากวุฒิภาวะหรือการเรียนรู้ของผู้สอบ

แนวคิดของ GT เกิดจากความไม่มั่นใจในความเชื่อมั่นของเครื่องมือที่เกิดจากการทดสอบเพียงครั้งเดียว แบบทดสอบฉบับเดียว และผู้ดำเนินการสอบเพียงคนเดียว และนำผลการทดสอบไปใช้ในการสรุปอ้างอิงคะแนนจริงของผู้สอบภายใต้สถานการณ์ แบบทดสอบและผู้ดำเนินการทดสอบที่เป็นไปได้ทั้งหมด ซึ่งปกติแล้วคะแนนที่ได้จากการสอบในสถานการณ์ต่างกัน แบบทดสอบต่างฉบับกันหรือผู้ดำเนินการสอบต่างกัน น่าจะทำให้ผลการสอบมีความแตกต่างกัน นอกจากนั้นการประมาณค่าความคลาดเคลื่อนตามแนว CTT นั้น ในการทดสอบหนึ่งครั้ง ประมาณค่าความคลาดเคลื่อนจากเพียงแหล่งเดียวเท่านั้น หรือมองว่าความคลาดเคลื่อนเป็นสิ่งที่แบ่งแยกไม่ได้ตัวอย่างเช่นจากสถานการณ์การทดสอบโดยเทคนิคการวัดซ้ำจากแบบทดสอบโดยการให้แบบทดสอบคู่ขนานหรือจากข้อสอบซึ่งใช้เทคนิคความสอดคล้องภายในประมาณค่าความคลาดเคลื่อน แต่ GT สามารถที่จะประมาณค่าความคลาดเคลื่อนพร้อมกันครั้งละหลายแหล่ง และสามารถแยกวิเคราะห์ได้ นอกจากนั้น GT ยังสามารถให้ข้อมูลเกี่ยวกับจำนวนครั้งของสถานการณ์การสอบ จำนวนแบบทดสอบ จำนวนผู้ดำเนินการสอบที่จำเป็นต่อการตัดสินใจ เพื่อที่จะกำหนดการสอบให้ได้ผลการสอบที่มีความเชื่อถือได้ และโดยกระบวนการของ GT จะทำให้ข้อสรุปที่สะท้อนให้เห็นระดับของความเชื่อถือได้ ซึ่งสัมพันธ์ของการสรุปอ้างอิงนี้มีความเป็นเหตุเป็นผลอย่างมากกับสัมพันธ์ความเชื่อมั่นตามแนวของ CTT และยิ่งไปกว่านี้ GT สามารถจะนำไปใช้ศึกษาความเชื่อถือได้ของคะแนนที่มีการนำเสนอในรูปแบบต่าง ๆ กัน (Allal in Keeves. 1988 : 272 ; Feldt and Brennan in Linn. 1989 : 138, Shavelson and Webb. 1991 : 1-2)

นูนอลลีและเบิร์นสไตน์ (Nunnally and Bernstein. 1994 : 279-280) ได้แสดงความเห็นเกี่ยวกับ GT ว่าเป็นทฤษฎีที่ขยายออกไปจากทฤษฎีการวัด CTT โดยเน้นการวัดที่แตกต่างกันกับบุคคลเดียวกัน ซึ่งอาจจะเกิดการแปรเปลี่ยนเนื่องจากความแตกต่างของวิธีวัดด้วยสาเหตุของความคลาดเคลื่อนในการวัดแบบสุ่ม นอกจากนั้นยังขยายทฤษฎีการสุ่มโดเมน (Domain sampling theory) ไปยังสถานการณ์ที่ดำเนินการ เนื่องจากมีการสุ่มอย่างหลากหลายจากหลาย ๆ โดเมน ทำให้ GT มีความเกี่ยวข้องอย่างมากกับการออกแบบการทดลอง ซึ่งจำเป็นต้องใช้เทคนิควิธีการของการวิเคราะห์ความแปรปรวนที่เน้นประมาณค่าองค์ประกอบของความแปรปรวนของคะแนนมากกว่าค่าสถิติของแบบทดสอบซึ่งเป็นการใช้เทคนิคการวิเคราะห์ความแปรปรวนเพื่อประมาณค่าสัมพันธ์ความเชื่อมั่นตามแนว CTT อีกประเด็นหนึ่งที่ GT แตกต่างไปจาก CTT คือ เงื่อนไขการยอมรับความเป็นคู่ขนานของแบบทดสอบ ใน GT ถ้าแบบทดสอบคู่ใด

บรรจด้วยข้อสอบที่ได้รับการสุ่มมาจากเอกภพ (Universe) เดียวกัน ยอมรับได้ว่าเป็นคู่ขนานกัน (Feldt and Brennan in Linn. 1989 : 127-128)

กรอบความคิดของ GT

เนื่องจาก GT ไม่ใช่ส่วนหนึ่งของ CTT ทั้งหมด ดังนั้น จึงมีคำสำคัญเพิ่มขึ้นจากมโนทัศน์ของ CTT ได้แก่

1. องค์ประกอบ (Facet) หมายถึง กลุ่มของเงื่อนไขในการวัดที่มีลักษณะเช่นเดียวกันหรือกลุ่มของตัวแปรที่คาดว่าจะส่งผลต่อการวัดตัวอย่างเช่น การประเมินการปฏิบัติของพนักงานในโรงงาน โดยให้ผู้คุมงานเป็นผู้กำหนดระดับการทำงาน 'หนัก ปานกลาง และเบา' จากสถานการณ์การประเมินครั้งนี้มีองค์ประกอบเป็นตัวแปรที่เกี่ยวข้องกับการวัด 2 องค์ประกอบ คือ ผู้คุมงานและระดับการทำงาน หรือในการศึกษาเกี่ยวกับการเขียนเรียงความของนักเรียน จำนวน 4 ครั้ง ให้แต่ละคนเขียนเรียงความที่แตกต่างกัน 2 หัวเรื่อง และให้มีการตรวจคะแนนโดยผู้ตรวจ 3 คน องค์ประกอบของการวัดครั้งนี้คือ จำนวนครั้ง หัวข้อเรื่องและผู้ตรวจให้คะแนน

2. เงื่อนไขของการวัด (Condition of measurement) ได้แก่ ระดับขององค์ประกอบหรือตัวแปร เช่น เพศแบ่งเป็น 2 ระดับ คือ ชาย หญิง หรือข้อสอบแต่ละข้อเป็นเงื่อนไขการวัดขององค์ประกอบแบบทดสอบ

3. รูปแบบความสัมพันธ์ระหว่างองค์ประกอบ เป็นสิ่งที่ชี้ถึงแหล่งความแปรปรวนของค่าที่วัดได้ ซึ่งมี 3 ลักษณะคือ

3.1 ความสัมพันธ์แบบไขว้ (Crossed) เป็นความสัมพันธ์ในลักษณะที่แต่ละระดับขององค์ประกอบถูกวัดภายใต้เงื่อนไขเดียวกันทั้งหมด ตัวอย่างเช่น นักเรียนทุกคน (p) ทำแบบทดสอบทุกข้อ (i) และผู้ตรวจ (r) ตรวจคะแนนทุกข้อรูปแบบความสัมพันธ์เขียนได้ว่า $p \times i \times r$

3.2 ความสัมพันธ์แบบแฝง (Nested) เป็นความสัมพันธ์ในลักษณะที่แต่ละระดับขององค์ประกอบถูกวัดภายใต้เงื่อนไขการวัดที่แตกต่างกัน ตัวอย่างเช่น ข้อสอบแต่ละข้อ (i) วัดในแต่ละจุดประสงค์ (O) รูปแบบความสัมพันธ์เขียนได้ว่า $i : O$

3.3 ความสัมพันธ์แบบผสม (Confounded) เป็นความสัมพันธ์ที่มีทั้งความสัมพันธ์แบบไขว้และความสัมพันธ์แบบแฝง ตัวอย่างเช่น ผู้สอบทุกคน (p) ทำข้อสอบ (i) ที่วัดในแต่ละจุดประสงค์ (O) รูปแบบความสัมพันธ์เขียนได้ว่า $p \times (i : O)$

4. องค์ประกอบความแปรปรวน (Variance component) ใน GT ประกอบด้วย 3 พารามิเตอร์หลักคือ

4.1 ความแปรปรวนของคะแนนเอกภพ (Universe score variance) เขียนแทนด้วยสัญลักษณ์ $\sigma^2(\mu)$ ซึ่งหมายถึง ความแปรปรวนค่าคาดหวังของคะแนนที่แสดงเอกภพของค่าสังเกตที่ยอมรับได้ (Universe of admissible observation) ที่สะท้อนความแปรปรวนอย่างมีระบบ อันเนื่องมาจากความแตกต่างระหว่างวัตถุประสงค์ของการวัด ตัวอย่างเช่นการอธิบายรูปแบบของ $p \times i$ องค์ประกอบความแปรปรวนประมาณค่าได้จากความแปรปรวนของผู้สอบ σ_p^2

4.2 ความแปรปรวนของความคลาดเคลื่อนสัมพัทธ์ (Relative error variance) เขียนแทนด้วยสัญลักษณ์ $\sigma^2(S)$ ซึ่งประมาณค่าได้จากผลรวมของน้ำหนักองค์ประกอบความแปรปรวนที่ประมาณค่าจากแหล่งความคลาดเคลื่อนที่เกี่ยวข้อง ซึ่งแต่ละองค์ประกอบถูกกำหนดน้ำหนักในทางกลับกันกับจำนวนครั้งที่ส่งผลกระทบต่อผู้สอบเมื่อทำการคำนวณคะแนนเฉลี่ยของผู้สอบแต่ละคนในแต่ละจุดประสงค์ของการวัด ตัวอย่างเช่น การอธิบายรูปแบบของ $p \times i$ การประมาณค่าความแปรปรวนของความคลาดเคลื่อนสัมพัทธ์ขึ้นอยู่กับประมาณองค์ประกอบความแปรปรวนเดียว $\sigma^2(S) = 1/\eta_i \sigma_{pic}^2$ เมื่อ σ_{pic}^2 คือ ความแปรปรวนที่เป็นผลกระทบจากปฏิสัมพันธ์ระหว่างผู้สอบและข้อคำถามกับผลที่เกิดจากเหตุการณ์ต่าง ๆ ที่ไม่สามารถระบุได้อย่างชัดเจน และ i คือจำนวนของระดับขององค์ประกอบข้อสอบ

4.3 ความแปรปรวนของความคลาดเคลื่อนสัมบูรณ์ (Absolute error variance) เขียนแทนด้วยสัญลักษณ์ $\sigma^2(\Delta)$ ซึ่งประมาณค่าได้จากความแปรปรวนทุกตัวยกเว้นความแปรปรวนเนื่องจากบุคคลจะอยู่รวมในความคลาดเคลื่อนในการวัด ดังนั้น รูปแบบของความแปรปรวนที่ก่อให้เกิดความคลาดเคลื่อนสัมบูรณ์คือ σ_i^2 และ σ_{pie}^2 หรือ $\sigma^2(\Delta) = 1/N, \sigma_{pie}^2 + 1/N, \sigma_i^2$ หรือ $\sigma^2(\Delta) = \sigma_i^2/\eta_i + \sigma_{pie}^2/\eta_i$

5. การศึกษาเชิงสรุปอ้างอิง (Generalizability studies ; G-studies) มีเป้าหมายเพื่อมุ่งศึกษาการใช้วิธีการวัดที่หลากหลายและสารสนเทศเกี่ยวกับแหล่งความแปรปรวนในการวัดเท่าที่จะเป็นไปได้ทั้งหมด การศึกษาแบบ G ควรเน้นศึกษาแหล่งความแปรปรวนที่มีศักยภาพและอยู่ในเอกภพของค่าสังเกตที่ยอมรับได้ (Universe of admissible observations) ซึ่งสามารถวัดหรือสังเกตได้ในแต่ละองค์ประกอบ (Facet) เท่าที่จะเป็นไปได้

6. การศึกษาเชิงตัดสินใจ (Decision studies ; D-studies) เป็นการศึกษา โดยใช้สารสนเทศจากการศึกษาแบบ G เพื่อนำมาใช้ออกแบบการวัดเป้าประสงค์ย่อยทางสังคมศาสตร์ในเชิงปฏิบัติที่เป็นไปได้มากที่สุด ในการวางแผนศึกษาแบบ D จะเกี่ยวข้องกับเอกภพของการอ้างอิง (Universe of generalization) ซึ่งเป็นเงื่อนไขการวัดทั้งหมดขององค์ประกอบที่ผู้ศึกษาต้องการอ้างอิง และเป็นส่วนหนึ่งของเอกภพของค่าสังเกตที่ยอมรับได้

7. สัมประสิทธิ์การสรุปอ้างอิง (Generalizability coefficient) เขียนแทนด้วยสัญลักษณ์ EP_2 แสดงประสิทธิภาพของการสรุปอ้างอิงจากคะแนนผลการวัดพฤติกรรมที่เป็นตัวแทนของผู้สอบไปยังคะแนนเอกภพ (Universse score) ของผู้สอบ ซึ่งมีความคล้ายกับค่าความเชื่อมั่นตามแนวทางฎีของ CTT ค่า EP_2 สะท้อนสัดส่วนของความแปรปรวนของคะแนนรายบุคคลที่เกิดขึ้นอย่างเป็นระบบกับความแปรปรวนของคะแนนเอกภพ หรืออาจกล่าวได้ว่าค่า EP_2 เป็นสัดส่วนระหว่างความแปรปรวนของคะแนนเอกภพกับความแปรปรวนของคะแนนสอบที่คาดหวัง ซึ่งมีวิธีการประมาณค่า 2 แนวทางคือ

7.1 ในกรณีเปรียบเทียบคะแนนสอบเชิงสัมพัทธ์ ความแปรปรวนของคะแนนสอบที่คาดหวังจะประกอบด้วยความแปรปรวนของคะแนนเอกภพรวมกับความแปรปรวนของความคลาดเคลื่อนสัมพัทธ์ ตัวอย่างเช่น รูปแบบของ $p \times i$ ค่า $EP = \sigma_p^2 / (\sigma_p^2 + \sigma_{(s)}^2)$

7.2 ในกรณีเปรียบเทียบคะแนนสอบเชิงสัมบูรณ์การประมาณค่าคล้ายกับการเปรียบเทียบเชิงสัมพัทธ์แต่ใช้ $\sigma^2(\Delta)$ แทน $(\sigma_{(s)}^2)$ ดังนั้นกรณีของ $p \times i$

$$\text{ค่า } EP_{(\Delta)}^2 = \sigma^2 / (\sigma^2 + \sigma_{(\Delta)}^2)$$

โมเดลทางสถิติของ G-Theories

1. โมเดลหนึ่งองค์ประกอบ (Single-facet model) แบบไขว้ แสดงแหล่งของความแปรปรวนดังนี้

แหล่งของความแปรปรวน	ชนิดของความแปรปรวน	สัญลักษณ์ของความแปรปรวน
ผู้สอบ (p)	คะแนนเอกภพ	σ_p^2
ข้อสอบ (i) - ---	ความคลาดเคลื่อน	σ_i^2
ส่วนที่เหลือ (pi.e)	ความคลาดเคลื่อน	$\sigma_{pi.e}^2$

$$X_{pi} = \mu + (\mu_p - \mu) + (\mu_i - \mu) + (X_{pi} - \mu_p - \mu_i + \mu)$$

เมื่อ X_{pi} คือ คะแนนสอบของผู้สอบคนที่ p จากการทำข้อสอบข้อที่ i

μ คือ ค่าเฉลี่ยรวมของทุกคน

$\mu_p - \mu$ คือ ผลกระทบที่เกิดจากคน

$\mu_i - \mu$ คือ ผลกระทบที่เกิดจากข้อสอบ

$X_{pi} - \mu_p - \mu_r + \mu$ คือ ผลที่สะท้อนถึงส่วนที่เหลือที่เกิดจากอิทธิพลของปฏิสัมพันธ์ระหว่างผู้สอบกับข้อสอบ รวมทั้งแหล่งความคลาดเคลื่อนที่เกิดขึ้นอย่างมีระบบอื่น ๆ ที่ไม่ได้กำหนดรวมไว้

องค์ประกอบความแปรปรวนเขียนได้ว่า

$$\sigma^2(X_{pi}) = \sigma_p^2 + \sigma_r^2 + \sigma_{pi.e}^2$$

2. โมเดลหลายองค์ประกอบ (Multifaceted model)

ตัวอย่างเช่น 2 องค์ประกอบ $p \times r \times o$ (p = ผู้ถูกวัด, r = ผู้ให้คะแนน, o = จำนวนครั้ง)

$$\begin{aligned} X_{pro} = & \mu + (\mu_p - \mu) + (\mu_r - \mu) + (\mu_o - \mu) + (\mu_{pr} - \mu_p - \mu_r + \mu) \\ & + (\mu_{po} - \mu_p - \mu_o + \mu) + (X_{pro} - \mu_{pr} - \mu_{po} - \mu_{ro}) \\ & + (\mu_{pro} - \mu_{po} - \mu_{ro} + \mu_p + \mu_r + \mu_o - \mu) \end{aligned}$$

เมื่อ X_{pro} คือ คะแนนผลการประเมินของผู้ถูกวัดคนที่ P จากกรรมการคนที่ r ครั้งที่ o)

μ

คือ ค่าเฉลี่ยคะแนนรวม

$\mu_p - \mu$

คือ ผลกระทบจาก

$\mu_r - \mu$

คือ ผลกระทบจาก

$\mu_o - \mu$

คือ ผลกระทบจาก

$\mu_{pr} - \mu_p - \mu_r + \mu$

คือ ผลกระทบของผู้ถูกประเมิน

$\mu_{po} - \mu_p - \mu_o + \mu$

คือ ผลกระทบของผู้สอบจากครั้ง

$\mu_{ro} + \mu_r - \mu_o - \mu$

คือ ผลกระทบของกรรมการจากครั้งที่

$X_{pro} - \mu_{pr} - \mu_{po} - \mu_{ro} + \mu_p + \mu_r + \mu_o - \mu$ คือ ส่วนที่เหลือ

องค์ประกอบความแปรปรวนเขียนได้ว่า

$$\sigma^2(X_{pro}) = \sigma_p^2 + \sigma_r^2 + \sigma_o^2 + \sigma_{pr}^2 + \sigma_{po}^2 + 2\sigma_{ro}^2 + \sigma_{pro.e}^2$$

3. โมเดลหนึ่งองค์ประกอบแบบแฝง (One-facet nested design)

ตัวอย่าง เช่น ในกรณีที่มีผู้สอบ 2 กลุ่ม ตอบข้อสอบไม่เหมือนกัน คือ ใช้ข้อสอบแฝงในผู้สอบเขียนแทนด้วยสัญลักษณ์ $i : p$ หรือ $i(p)$

$$X_{pi} = \mu + (\mu_p - \mu) + (\mu_i - \mu) + (X_{pi} - \mu_p)$$

เมื่อ μ คือ คะแนนเฉลี่ยรวมของผู้สอบทุกคน

$\mu_p - \mu$ คือ ผลกระทบจากผู้สอบ

$\mu_{pi} - \mu_p$ คือ ผลกระทบส่วนตัวเหลือ

องค์ประกอบความแปรปรวน เขียนได้ว่า

$$\sigma^2(X_{pi}) = \sigma_p^2 + \sigma_{i:pi}^2$$

การวิเคราะห์สัณยศาสตร์ GT

1. การศึกษาเชิงสรุปอ้างอิง (G-study) เป็นขั้นแรกของการวิเคราะห์ข้อมูล เพื่อที่จะประมาณค่าความแปรปรวนจากแหล่งต่าง ๆ ภายใต้เอกภพของค่าสังเกตที่ยอมรับได้ โดยมีกระบวนการดังนี้

1.1 กำหนดสิ่งที่วัด (Object of measurement) เช่น นักเรียนชั้นประถมศึกษาเป็นต้น

1.2 กำหนดองค์ประกอบ (Facet) ของการวัด เช่น องค์ประกอบของข้อสอบ องค์ประกอบของแบบทดสอบ

1.3 กำหนดเอกภพของค่าสังเกตที่ยอมรับได้ เช่น ในการวิจัยครั้งนี้ คือ องค์ประกอบข้อสอบ และองค์ประกอบของแบบทดสอบ

1.4 กำหนดความสัมพันธ์ระหว่างองค์ประกอบ เช่น ถ้าต้องการศึกษาแบบให้ผู้สอบทุกคนทำข้อสอบทุกข้อ ก็จะเป็นแบบ Crossed คือ $p \times i$ (p คือผู้สอบ และ i คือข้อสอบ) แต่ถ้าให้ผู้สอบทุกคนทำข้อสอบบางข้อก็เป็นแบบ Nested คือ $p : i$ ซึ่งจะแล้วแต่ความสนใจของผู้ศึกษาว่าจะกำหนดความสัมพันธ์ขององค์ประกอบแบบใด

1.5 กำหนดชนิดขององค์ประกอบหรือตัวประกอบว่าเป็นชนิดใดต่อไปนี้

(1) องค์ประกอบสุ่มแท้จริง (Purely random facet) ได้แก่ องค์ประกอบทุกเงื่อนไขการวัดได้มาจากการสุ่มอย่างง่ายจากเอกภพที่มีขนาดไม่จำกัด (∞)

(2) องค์ประกอบสุ่มจำกัด (Finite random facet) ได้แก่ องค์ประกอบที่ทุกเงื่อนไขการวัดได้มาจากการสุ่มอย่างง่ายจากเอกภพที่มีขนาดจำกัด

(3) องค์ประกอบกำหนด (Fixed facet) ได้แก่ องค์ประกอบที่ประกอบด้วยเงื่อนไขการวัดจำนวนจำกัดอยู่ในเอกภพ และเงื่อนไขการวัดเหล่านั้นถูกเลือกมาศึกษาทั้งหมด

1.6 กำหนดรูปแบบ (Model) การวัดว่าเป็นรูปแบบใดในต่อไปนี้

(1) รูปแบบกำหนด (Fixed model) เป็นรูปแบบที่ประกอบด้วยองค์ประกอบกำหนดทั้งหมด

(2) รูปแบบสุ่ม (Random model) เป็นรูปแบบที่ประกอบด้วยองค์ประกอบสุ่มแท้จริงและหรือองค์ประกอบสุ่มจำกัด

(3) รูปแบบผสม (Mixed model) เป็นรูปแบบที่ประกอบด้วยองค์ประกอบสุ่มแท้จริงและหรือสุ่มจำกัด และองค์ประกอบกำหนดผสมกัน

ตาราง 2 รูปแบบ (Design) และแหล่งแปรปรวนที่ได้ตามทฤษฎี G-theories

แบบ (Design)	แหล่งแปรปรวนที่ได้	
	ผลหลัก (Main effect)	ผลของปฏิสัมพันธ์ (Interaction effect)
$p \times i$	p, i	pi
$p : i$	$p, p : i$	-
$p \times i \times t$	p, i, t	pi, pt, it, pit
$p \times (i:t)$	$p, t, i : t$	$pt, pi:t$
$(i:p) \times t$	$p, t, i : p$	$pt, it:p$
$i : (p \times t)$	$p, t, i : pt$	pt
$(i \times t) : p$	$p, i:p, t:p$	$it:p$
$i : t : p$	$p, t:p, i:t:p$	-

เมื่อ p คือผู้สอบ, i คือข้อสอบ, และ t คือแบบทดสอบ

เมื่อได้แหล่งแปรปรวนแล้วนำมาวิเคราะห์โดยวิธี ANOVA (Analysis of variance) จะได้ค่าความแปรปรวนจากแหล่งต่าง ๆ ($\sigma^2_{(p)}$) เช่นในแบบ $p \times i$ จะได้ความแปรปรวนของผลหลักเป็น $\sigma^2_{(p)}$ ซึ่งในการตีความจะพิจารณาผู้สอบแต่ละคนในประชากรผู้สอบ (p) โดยผู้สอบแต่ละคนจะมีคะแนนที่ได้จากการทำข้อสอบทุกข้อ (n) ในเอกภพของค่าสังเกตที่ยอมรับได้ ดังนั้น นักเรียนแต่ละคนจะได้คะแนนเฉลี่ย μ_p และความแปรปรวนของคะแนนเฉลี่ยดังกล่าวของประชากรผู้สอบคือ

$\sigma^2_{(\mu)}$ ซึ่งก็คือ ส่วนความแปรปรวนของข้อสอบ (i) ก็จะตีความได้ในทำนองเดียวกัน ผลที่ได้จากการประมาณค่าความแปรปรวนในครั้งนี้เป็นการศึกษา เพื่อการสรุปอ้างอิง (G-study) ซึ่งจะเป็นพื้นฐานสำหรับใช้เพื่อการตัดสินใจเกี่ยวกับสิ่งที่จะวัดในการศึกษาเพื่อตัดสินใจ (D-study) ต่อไป

2. การศึกษาเพื่อการตัดสินใจ (D-study) เป็นขั้นการนำผลที่ได้จากการประมาณความแปรปรวนในการศึกษาเพื่อการสรุปอ้างอิง (G-study) มาเลือกวิธีวัดที่เหมาะสมเพื่อให้ได้ประโยชน์สูงสุดที่ต้องการ มีขั้นตอนดังนี้

2.1 กำหนดเอกภพเชิงสรุปอ้างอิง (Universe of generalization) หมายถึง การกำหนดลักษณะเฉพาะของเอกภพของการอ้างอิงที่ผู้วิจัยต้องการอ้างอิงถึง ซึ่งอาจจะประกอบด้วยเงื่อนไขทั้งหมดในเอกภพค่าสังเกตที่ยอมรับได้ หรืออาจเป็นเซตย่อย (Set) ของเอกภพค่าสังเกตที่ยอมรับได้

2.2 กำหนดรูปแบบ (Model)ว่าจะใช้รูปแบบใด เช่น รูปแบบสุ่ม (Random model) หรือรูปแบบกำหนด (Fixed model) หรือรูปแบบผสม (Mixed model) ซึ่งโดยหลักการแล้วควรใช้โมเดลแบบสุ่มดีที่สุด เพราะจะทำให้การอ้างอิงสรุปได้กว้างขวางกว่าโมเดลแบบผสมและโมเดลแบบกำหนดตามลำดับ

2.3 ขนาดของกลุ่มตัวอย่างของการศึกษาเพื่อการตัดสินใจ (D-Study sample sizes) จำนวนเงื่อนไขขององค์ประกอบใน D-study ไม่จำเป็นต้องเหมือนใน G-study การกำหนดจำนวนเงื่อนไขหรือขนาดตัวอย่างจะขึ้นอยู่กับความสนใจของผู้วิจัย

2.4 กำหนดโครงสร้างของแผนแบบที่ศึกษา (D-Study design structure) โครงสร้างของแผนแบบที่ศึกษา หรือความสัมพันธ์ขององค์ประกอบที่ศึกษา อาจจะเหมือน หรือต่างจาก G-study ก็ได้ เช่นใน G-study ตัดสินใจให้นักเรียนทุกคนทำข้อสอบทุกข้อ และผู้ตรวจทุกคนตรวจข้อสอบทุกข้อเหมือนกัน โครงสร้างของแบบจะเป็น $p \times i \times r$ แต่เมื่อเป็น D-study ตัดสินใจให้นักเรียนทุกคนทำข้อสอบทุกข้อ โดยผู้ตรวจแต่ละคนทำการตรวจชุดข้อสอบต่างกันในกรณีนี้จะได้โครงสร้างของแบบเป็น $p \times (i:r)$ เป็นต้น

2.5 การประมาณองค์ประกอบความแปรปรวนในขั้น D-study (Estimate D-study variance components) ในขั้นนี้จะมีการประมาณค่าความแปรปรวนขึ้นมาใหม่ โดยอาศัยผลการประมาณค่าในขั้น G-study โดยให้สอดคล้องกับรูปแบบ (Model) ของการวัดขนาดกลุ่มตัวอย่าง และโครงสร้างของแบบที่ศึกษา ค่าความแปรปรวนใน D-study หาได้จากการหารค่าความ

แปรปรวนที่ได้จาก G-study ของผลหลักและผลปฏิสัมพันธ์ได้ ด้วยขนาดของกลุ่มตัวอย่างที่กำหนดใน D-study เช่นในแบบ $p \times i \times r$ ขนาดตัวอย่างข้อสอบ = n_i และขนาดของผู้ตรวจ = n_r จะหาความแปรปรวนใหม่ได้ ดังนี้

ตาราง 3 แหล่งความแปรปรวนตามทฤษฎี G-theories

ความแปรปรวนใหม่	ความแปรปรวนเดิม/ขนาดของกลุ่มตัวอย่าง
$\sigma^2(p)$	$\sigma^2(p)$
$\sigma^2(i)$	$\sigma^2(i)/n_i$
$\sigma^2(r)$	$\sigma^2(r)/n_r$
$\sigma^2(pi)$	$\sigma^2(pi)/n_i$
$\sigma^2(pr)$	$\sigma^2(pr)/n_r$
$\sigma^2(ir)$	$\sigma^2(ir) / n_i n_r$
$\sigma^2(pir)$	$\sigma^2(pir) / n_i n_r$

ในแบบ $p \times (i:r)$ จะได้ความแปรปรวนใหม่ดังนี้

ตาราง 3 (ต่อ)

ความแปรปรวนใหม่	ความแปรปรวนเดิม/ขนาดของกลุ่มตัวอย่าง
$\sigma^2(p)$	$\sigma^2(p)$
$\sigma^2(r)$	$\sigma^2(r)/n_r$
$\sigma^2(pr)$	$\sigma^2(pr)/n_r$
$\sigma^2(ir)$	$\sigma^2(ir) / n_i n_r$
$\sigma^2(pi:r)$	$\sigma^2(pi:r) / n_i n_r$

2.6 การประมาณค่าความแปรปรวนของความคลาดเคลื่อนแบ่งเป็น 2 ชนิดคือ

(1) ความแปรปรวนคลาดเคลื่อนแบบสัมบูรณ์ (Absolute error variance) ใช้สัญลักษณ์ว่า $\sigma^2(\Delta)$ เป็นความแปรปรวนของความแตกต่างระหว่างคะแนนที่ได้จากการสังเกตกับคะแนนเอกภพ คำนวนจากผลบวกของค่าองค์ประกอบความแปรปรวนทั้งหมดยกเว้นความแปรปรวนของคะแนนเอกภพ [$\sigma^2_{(p)}$] จากโครงสร้างแบบ $p \times (I:T)$ จะได้

$$\sigma^2(\Delta) = \sigma^2_{(t)} + \sigma^2_{(I:t)} + \sigma^2_{(P:t)} + \sigma^2_{(PI:t)}$$

(2) ความแปรปรวนคลาดเคลื่อนแบบสัมพัทธ์ (Relative error variance) ใช้สัญลักษณ์ว่า $\sigma^2_{(s)}$ เป็นค่าความแปรปรวนของผลต่างระหว่างส่วนเบี่ยงเบนคะแนนที่สังเกตได้ และส่วนเบี่ยงเบนคะแนนเอกภพ คำนวณจากผลบวกค่าขององค์ประกอบความแปรปรวนร่วมระหว่างสิ่งที่ถูกวัดกับองค์ประกอบ (Facet) อื่น ๆ

$$\sigma^2_{(s)} = \sigma^2_{(PT)} + \sigma^2_{(PI:T)}$$

2.7 การประมาณค่าสัมประสิทธิ์การสรุปอ้างอิง เป็นขั้นสุดท้ายของการศึกษา เพื่อการตัดสินใจ (D-study) ซึ่งเป็นดัชนีชี้บ่งถึงความเชื่อถือได้ของการวัด ซึ่งคำนวณได้จากอัตราส่วนระหว่างความแปรปรวนของคะแนนเอกภพ และความแปรปรวนของคะแนนที่สังเกตที่คาดหวัง

(1) วิธีของเบรนนอน (Brennan, 1983 : 108-109) สำหรับการวัดที่มีการตัดสินใจเชิงเปรียบเทียบหรือการวัดแบบอิงกลุ่ม ความคลาดเคลื่อนในการวัดจะเป็นความคลาดเคลื่อนสัมพัทธ์ ค่าสัมประสิทธิ์การสรุปอ้างอิงประมาณค่าได้จากสูตรดังนี้

$$E\rho^2 = \sigma^2_{(P)} / [\sigma^2_{(P)} + \sigma^2_{(S)}]$$

(2) วิธีของเบรนนอน (Brennan, 1983 : 109) สำหรับการวัดที่มีการวัดการตัดสินใจแบบสัมบูรณ์ หรือการวัดแบบอิงเกณฑ์ ความคลาดเคลื่อนในการวัดจะเป็นความคลาดเคลื่อนแบบสัมบูรณ์ ค่าสัมประสิทธิ์การอ้างอิงสรุปประมาณค่าได้จากสูตรดังนี้

$$E\rho^2 = \sigma^2_{(P)} / [\sigma^2_{(P)} + \sigma^2_{(\Delta)}]$$

(3) วิธีของเบรนนอนและเคน (Brennan, 1983 : 108 ; citing Brennan and Kane, 1977) สำหรับการวัดที่มีการตัดสินใจแบบสัมบูรณ์หรือการวัดแบบอิงเกณฑ์ที่กำหนดคะแนนจุดตัด ค่าสัมประสิทธิ์การอ้างอิงสรุปประมาณค่าได้จากสูตรดังนี้

$$E\rho^2 = \phi(\lambda) = [\sigma^2_{(t)} + (\mu - \lambda)^2] / [\sigma^2_{(P)} + (\mu - \lambda)^2 + \sigma^2_{(\Delta)}]$$

แทนสัดส่วนของคะแนนจุดตัด

ทฤษฎีการตอบสนองข้อสอบ (IRT)

ทฤษฎีการตอบสนองข้อสอบเป็นทฤษฎีที่เน้นวัตถุประสงค์เพื่อเสนอหลักการพื้นฐานสำคัญ สำหรับการประมาณค่า ความสามารถหรือคุณลักษณะ (Ability or Trait) ของผู้สอบ ซึ่งวัดด้วยแบบทดสอบ ทฤษฎี IRT เชื่อกันว่าเริ่มต้นตั้งแต่ปี ค.ศ.1916 โดยบินท์ และ ซีมอน (Binet and Simon) เป็นผู้เริ่มต้นโดยได้กำหนดระดับผลการปฏิบัติ (Performance levels) ของผู้สอบในตัวแปร

อิสระต่าง ๆ และใช้ผลการปฏิบัติดังกล่าวไปพัฒนาแบบทดสอบและ ในปี 1943 ลอร์เลย์ (Lawley) ได้เริ่มคิดวิธีการประมาณค่าพารามิเตอร์ ต่อมาในปี 1952 ลอร์ด (Lord) ได้นำเสนอโมเดลนอร์มัล โอโจไฟแบบ 2 พารามิเตอร์ เพื่อประมาณค่าพารามิเตอร์และนำไปประยุกต์ใช้ ต่อจากนั้นในปี 1957 เบอร์นบอม (Birnbau) ได้นำเสนอโมเดลโลจิสติก (Logistic model) เพื่อใช้แทน โมเดลนอร์มัลโอโจไฟ ขณะเดียวกันราสค์ (Rasch) นำเสนอโมเดลขึ้นใหม่จากโมเดลความน่าจะเป็น (Probabilistic model) โดยใช้ชื่อว่า ราสค์โมเดล (Rasch model) ในปี 1960 และจนกระทั่งในปี 1968 ลอร์ด และโนวิก (Lord and Novick) นำเสนอทฤษฎีคุณลักษณะแฝง (Theory of latent trait) ขึ้นมาและในปี 1969 ไรท์ และปันชาปเกसान (Wright and Panchapakesan) ได้เสนอโปรแกรมคอมพิวเตอร์ไบคาล (BICAL) เพื่อประมาณค่าพารามิเตอร์ตามแนวของราสค์โมเดล และในปี 1970 ลอร์ด ได้เสนอทฤษฎี IRT เพื่ออธิบายปรากฏการณ์ต่างๆแทนทฤษฎีคุณลักษณะแฝงเดิม และได้คิดโปรแกรมคอมพิวเตอร์ ลอจิส (LOGIST) เพื่อประมาณค่าพารามิเตอร์ ตามโลจิสติกโมเดล (Logistic models) ขึ้นในปี 1974 และปรับปรุงโปรแกรมเพิ่มเติมในปี ค.ศ. 1982 (Hambleton and Swaminathan, 1985 : 4-52 ; Hambleton in Linn. 1989 : 147-160 ; Hambleton et al. 1991 :12-31)

แนวคิดพื้นฐานของทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบ (IRT) มีแนวคิดมาจากพื้นฐาน 2 ประการคือ

1. พฤติกรรมการแสดงออกของผู้สอบในการตอบข้อทดสอบสามารถพยากรณ์ได้จากกลุ่มขององค์ประกอบที่เรียกว่าคุณลักษณะ (Traits) หรือคุณลักษณะแฝง (Latent traits) หรือ ความสามารถ (Abilities) และ
2. ความสัมพันธ์ระหว่าง การแสดงพฤติกรรมการตอบข้อสอบของผู้สอบกับชุดของคุณลักษณะที่อยู่ภายใต้คุณลักษณะที่แสดงออกของข้อสอบสามารถอธิบายได้โดยฟังก์ชันที่เพิ่มขึ้นในลักษณะทิศทางเดียว (Monotonically) ที่เรียกว่าฟังก์ชันคุณลักษณะข้อสอบ (Item Characteristic Function) หรือโค้งลักษณะข้อสอบ (Item Characteristic Curve ICC) (Hambleton and Others, 1991:7)

ลักษณะทั่วไปของทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีนี้ได้เกิดขึ้นท่ามกลางข้อจำกัดของทฤษฎีการทดสอบแบบเดิมหลายประการ คือ (Hambleton and Swaminathan. 1985 : 1-4 ; Hambleton et al.1991 : 2-5)

วิธีการเลือกข้อสอบที่มีค่าอำนาจจำแนกสูง ๆ (ไม่ต่ำกว่า .80) แต่อาจจะทำให้ข้อสอบขาดคุณสมบัติอื่น ๆ อีกได้

2. ความเป็นอิสระในการตอบข้อสอบ (Local Independence) หมายถึง โอกาสในการตอบข้อสอบแต่ละข้อได้ถูก ต้องเป็นอิสระจากกัน นั่นคือ การตอบข้อสอบข้อใดข้อหนึ่งได้ถูกต้องหรือผิด จะไม่มีผลกระทบต่อคำตอบข้อสอบข้ออื่น ๆ ด้วย

3. โค้งแสดงลักษณะข้อสอบ (Item characteristic curve : ICC) เป็นฟังก์ชันทางคณิตศาสตร์ที่แสดงถึงความสัมพันธ์ระหว่างโอกาสในการตอบข้อสอบข้อนั้นได้ถูกต้องกับระดับความสามารถที่วัดได้โดยใช้ชุดของข้อสอบหรือแบบทดสอบฉบับนั้น

โมเดลของทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบเป็นทฤษฎีที่อธิบายความสัมพันธ์ระหว่างปริมาณของการกระทำหรือคะแนนของผู้สอบกับปริมาณความสามารถของผู้สอบ ซึ่งแสดงได้ดังสมการ (สำเร็จ บุญเรืองรัตน์ . 2523 : 3)

$$P = f(\theta)$$

เมื่อ P แทน ผลการสอบ (Performance)

θ แทน ความสามารถ (Ability หรือ Trait)

f แทน ความสัมพันธ์ (Function)

โมเดลของทฤษฎีการตอบสนองข้อสอบสามารถจำแนกออกได้เป็น 3 กลุ่มใหญ่ ๆ ตามลักษณะของคำตอบที่ได้จากการตอบข้อคำถาม ซึ่งมี 3 ลักษณะ คือ คำตอบที่เป็นข้อมูลแบบสองทาง (Dichotomous) ข้อมูลแบบพหุ (Multicategory) และข้อมูลแบบต่อเนื่อง (Continuous) โมเดลที่มีความเหมาะสมกับลักษณะของข้อมูลดังกล่าวมีดังนี้ (Hambleton and Swaminathan. 1985. 34-52)

ตาราง 4 ลักษณะของข้อมูล โมเดลที่ใช้วิเคราะห์และผู้นำเสนอโมเดล

ลักษณะของข้อมูล	โมเดลที่ใช้วิเคราะห์	ผู้นำเสนอโมเดล
ข้อมูลแบบแบ่งสอง (Dichotomous)	Latent Linear	Lazarsfeld & Henry
	Latent Distance	(1968)
	Perfect scale	Guttman (1944)
	One, Two, Three Parameter	
	Normal Ogive	Lord (1952)
	One, Two, Three Parameter	
	Logistic	Birnbaum (1957, 1958a, 1958b, 1968) Lord & Novick (1968)
ข้อมูลแบบหลายชั้น (Multicategory)		Lord (1980a) Rasch (1960)
		Wright & Stone (1979)
	Four Parameter Logistic	McDonald (1967)
		Barton & Lord (1981)
	Nominal Response	Bock (1972)
ข้อมูลแบบต่อเนื่อง (Continuous)	Graded Response	Samejima (1969)
	Partial Credit Model	Master (1982)
	Continuous Response	Samejima (1972)

จากตาราง 4 ในแต่ละโมเดลจะมีฟังก์ชันทางคณิตศาสตร์ต่าง ๆ กัน ซึ่งจะเสนอตามลักษณะของข้อมูลจากการสอบดังนี้

1. โมเดลที่ใช้กับข้อมูลแบบสองทาง (Dichotomous หรือ Binary item) เป็นโมเดลที่ใช้กับข้อสอบหรือข้อคำถามที่ให้คะแนนในแต่ละข้อเป็น 2 ค่า คือตอบถูกให้ 1 และตอบผิดให้ 0 ซึ่งประกอบด้วยโมเดลต่าง ๆ ดังต่อไปนี้

1) โมเดลนอร์มัลอจีฟแบบ 2 พารามิเตอร์ (Two parameter normal ogive model) โมเดลนี้ลอว์เลย์ (Lawley) เป็นผู้เริ่มต้นพัฒนาขึ้นในราวปี ค.ศ.1943-1944 แต่เนื่องจากมีข้อจำกัดหลายอย่างต่อมัลลอร์ด(Lord.1952:1953a)จึงได้ศึกษาและนำมาปรับปรุง และนำไปประยุกต์ใช้กับ

ข้อคำถามที่ให้คะแนนเป็นสอง (Binary item) ซึ่งเน้นที่พารามิเตอร์ 2 ตัว คือค่าความยาก (b) และค่าอำนาจจำแนก (a) ซึ่งเขียนเป็นสมการทางคณิตศาสตร์ได้ดังนี้

$$P_i(\theta) = \int_{-\infty}^{a_i(\theta-b_i)} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \quad (2 \text{ พารามิเตอร์})$$

เมื่อ $P_i(\theta)$ แทน ความน่าจะเป็นที่ผู้สอบจะตอบข้อที่ 1 ได้ถูกต้องด้วย

ความสามารถ θ

b_i แทน ค่าพารามิเตอร์ความยากของข้อที่ 1

a_i แทน ค่าพารามิเตอร์อำนาจจำแนกของข้อที่ 1

Z_i แทน ความเบี่ยงเบนปกติจากการแจกแจงที่ค่าเฉลี่ย b_i

และส่วนเบี่ยงเบนมาตรฐาน $1/a_i$

2) โมเดลนอร์มัลลอจิสต์แบบ 1, 3 และ 4 พารามิเตอร์ (One, three and four parameter normal ogive models) หลังจากทีลอร์ด (Lord) ได้เสนอโมเดลนอร์มัลลอจิสต์แบบ 2 พารามิเตอร์ ต่อมาก็ได้เสนอโมเดลนอร์มัลลอจิสต์แบบ 1, 3 พารามิเตอร์ตามโมเดลทางคณิตศาสตร์ดังนี้

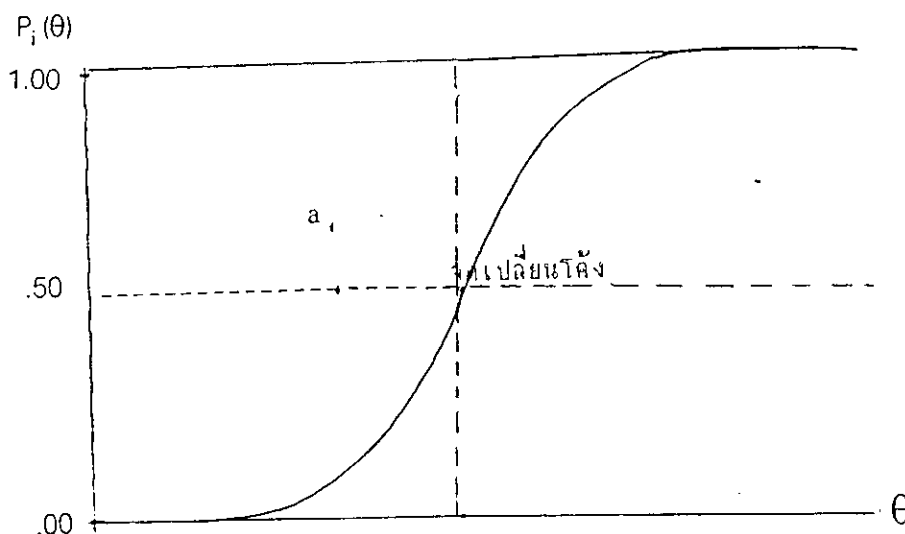
$$P_i(\theta) = \int_{-\infty}^{\theta-b_i} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \quad (1 \text{ พารามิเตอร์})$$

$$P_i(\theta) = c_i + (1 - c_i) \int_{-\infty}^{a_i(\theta-b_i)} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \quad (3 \text{ พารามิเตอร์})$$

$$P_i(\theta) = c_i + (\gamma_i - c_i) \int_{-\infty}^{a_i(\theta-b_i)} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \quad (4 \text{ พารามิเตอร์})$$

จากโมเดลนอร์มัลลอจีฟ เมื่อนำมาเขียนกราฟโค้งลักษณะของข้อสอบจะได้

ดังภาพประกอบ 4



ภาพประกอบ 4 ค่าพารามิเตอร์ของโมเดลนอร์มัลลอจีฟแบบ 2 พารามิเตอร์

2) โมเดลโลจิสติก (Logistic model) เป็นโมเดลที่เบิร์นบอม (Birnbbaum) พัฒนาขึ้นเพื่อใช้แทนโมเดลนอร์มัลลอจีฟ ทั้งนี้เพราะโมเดลทั้งสองมีลักษณะที่ใกล้เคียงกัน แต่โมเดลโลจิสติกสามารถคำนวณได้ง่ายกว่าโมเดลนอร์มัลลอจีฟจึงทำให้โมเดลโลจิสติกได้รับความนิยมนำไปใช้ในทางปฏิบัติมากกว่า ซึ่งโมเดลโลจิสติกนี้เบิร์นบอมได้พัฒนาโลจิสติกโมเดลแบบ 2 พารามิเตอร์ ที่เน้นพารามิเตอร์ค่าความยาก (b) และพารามิเตอร์ค่าอำนาจจำแนก (a) ซึ่งเป็นโมเดลที่มีโค้งแสดงลักษณะข้อสอบ (ICC) คล้ายคลึงกับโมเดลนอร์มัลลอจีฟแบบ 2 พารามิเตอร์ที่ลอร์ดได้เสนอไว้ โดยโค้งแสดงลักษณะข้อสอบของโมเดลทั้งสองมีความแตกต่างกันไม่เกิน .01 สำหรับทุก ๆ ค่าของความสามารถ (θ) เนื่องจากโมเดลโลจิสติกเป็นโมเดลทางคณิตศาสตร์ที่สามารถนำไปประยุกต์ใช้ได้ได้ง่ายกว่าโมเดลนอร์มัลลอจีฟ ต่อมาเบิร์นบอมได้เพิ่มพารามิเตอร์ค่าการเดา (c) เข้าไปอีก 1 ตัว เป็นโมเดลโลจิสติก 3 พารามิเตอร์ นอกจากนั้น ยังตัดค่าพารามิเตอร์อำนาจจำแนก (a) ออกจากโมเดลโลจิสติก 2 พารามิเตอร์เหลือเพียงพารามิเตอร์ (b) เพียงอย่างเดียว จึงเกิดโมเดลโลจิสติกแบบ 1 พารามิเตอร์ สำหรับโมเดลทางคณิตศาสตร์ของแต่ละโมเดลเขียนได้ดังนี้

2.1 โมเดลโลจิสติกแบบ 1 พารามิเตอร์ (One - parameter logistic model) เป็นโมเดลพัฒนาขึ้น โดยถือว่าข้อสอบทุกข้อมีค่าพารามิเตอร์อำนาจจำแนกเท่ากัน และไม่มีการเดา ($a=1$, $c=0$) ซึ่งมีโมเดลทางคณิตศาสตร์ดังนี้

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

เมื่อ $P_i(\theta)$ แทน ความน่าจะเป็นที่ผู้สอบจะตอบข้อที่ i ได้ถูกต้องด้วย

ความสามารถ θ

b_i แทน ค่าพารามิเตอร์ความยากของข้อที่ i คือค่า θ เมื่อ $P_i(\theta) = .50$

a_i แทน ค่าพารามิเตอร์อำนาจจำแนกเฉลี่ยของแบบทดสอบหรือเท่ากับ 1

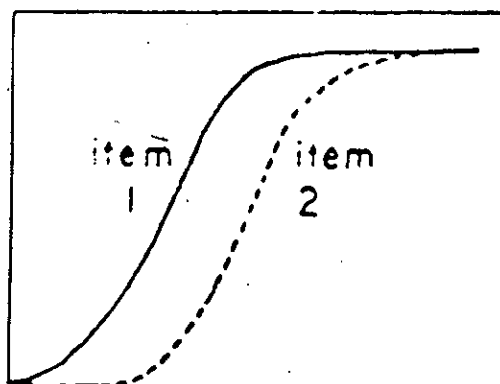
θ แทนระดับความสามารถของผู้สอบ

D แทน ค่าคงที่ที่ใช้ปรับแก้ค่าของ $P_i(\theta)$ มีค่าเท่ากับ 1.702

e แทน ค่าเอ็กซ์โปเนนเชียล (Exponential) เป็นค่าคงที่ซึ่งมีค่าเท่ากับ 2.718

จากโมเดลโลจิสติกแบบ 1 พารามิเตอร์ เมื่อนำมาเขียนเป็นกราฟโค้งลักษณะของข้อสอบ (ICC) จะได้ดังภาพประกอบ 5

โอกาสการ
ตอบถูก $P_i(\theta)$



ความสามารถ (θ)

ภาพประกอบ 5 แสดงความหมายค่าพารามิเตอร์ของโมเดลโลจิสติกแบบ 1 พารามิเตอร์ (2 ข้อ)

2.2 โมเดลโลจิสติกแบบ 2 พารามิเตอร์ (Two - parameter logistic model) เป็นโมเดลที่พัฒนามาจากนอร์มัลลอจีฟ เพื่อให้ง่ายต่อการคำนวณ ซึ่งมีโมเดลทางคณิตศาสตร์ดังนี้

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

เมื่อ $P_i(\theta)$ แทน ความน่าจะเป็นที่ผู้สอบจะตอบข้อที่ i ได้ถูกต้องด้วย

ความสามารถ θ

b_i แทน ค่าพารามิเตอร์ความยากของข้อที่ i จุด คือค่า θ เมื่อ $P_i(\theta) = .50$

a_i แทน ค่าพารามิเตอร์อำนาจจำแนกข้อที่ i คือ สัดส่วนของความชัน
ของ $P_i(\theta)$ ที่ $\theta = b_i$

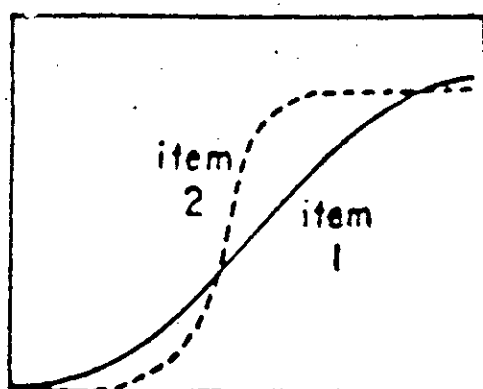
D แทน ค่าคงที่มีค่าเท่ากับ 1.702

e แทน ค่าเอ็กซ์โปเนนเชียล เป็นค่าคงที่ซึ่งมีค่าเท่ากับ 2.718

โมเดลโลจิสติกแบบ 2 พารามิเตอร์ เมื่อนำมาเขียนกราฟโค้งลักษณะของข้อสอบ
(ICC) จะได้ดังภาพประกอบ 6

โอกาสการ

ตอบถูก $P_i(\theta)$



ความสามารถ (θ)

ภาพประกอบ 6 แสดงความหมายค่าพารามิเตอร์ของโมเดลโลจิสติกแบบ 2 พารามิเตอร์

2.3 โมเดลโลจิสติกแบบ 3 พารามิเตอร์ (Three - parameter logistic model) เป็นโมเดลที่พัฒนามาจากโมเดลโลจิสติกแบบ 2 พารามิเตอร์เพื่อความเหมาะสมกับการทดสอบที่มีอิทธิพลของการเดาเข้ามาเกี่ยวข้อง โดยเพิ่มค่าพารามิเตอร์ขึ้นอีกตัวหนึ่งคือค่าการเดา ซึ่งเขียนเป็นโมเดลทางคณิตศาสตร์ได้ดังนี้

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

เมื่อ $P_i(\theta)$ แทน ความน่าจะเป็นที่ผู้สอบจะตอบข้อที่ i ได้ถูกต้องด้วยความสามารถ θ

b_i แทน ค่าพารามิเตอร์ความยากของข้อที่ i คือค่า θ เมื่อ $P_i(\theta) = (1 - c_i)/2$

a_i แทน ค่าอำนาจจำแนกของข้อที่ i คือ ค่าสัดส่วนของความชันของ $P_i(\theta) =$
ที่ $\theta = b_i$

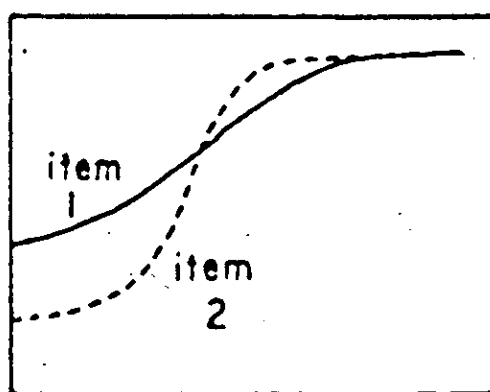
c_i แทน ค่าพารามิเตอร์ของการเดาของข้อสอบข้อที่ i คือ ความน่าจะเป็นที่
ผู้สอบไม่มีความรู้เลยทำข้อสอบข้อนั้นถูก

D แทน ค่าคงที่มีค่าเท่ากับ 1.702

e แทน ค่าเอ็กซ์โปเนนเชียลเป็นค่าคงที่ซึ่งมีค่าเท่ากับ 2.718

โมเดลโลจิสติกแบบ 3 พารามิเตอร์ เมื่อนำมาเขียนกราฟโค้งลักษณะของข้อสอบ (ICC) จะได้ดังภาพประกอบ 7

โอกาสการ
ตอบถูก $P_i(\theta)$



ความสามารถ (θ)

ภาพประกอบ 7 แสดงความหมายค่าพารามิเตอร์ของโมเดลโลจิสติกแบบ 3 พารามิเตอร์

2.4 โมเดลโลจิสติกแบบ 4 พารามิเตอร์ (Four - parameter logistic model) โมเดลนี้เสนอโดยแม็คโดแนลด์ (McDonald) ในปี ค.ศ.1967 ต่อมาลอร์ด (Lord) และบาร์ตัน (Barton) ได้พัฒนาขึ้นในปี ค.ศ.1981 โดยเพิ่มค่าพารามิเตอร์ ค่าความรอบคอบ (Careless) ซึ่งเกิดจากความคิดว่าในโมเดลโลจิสติกแบบ 3 พารามิเตอร์ มีบางครั้งที่กรณีผู้สอบที่มีความสามารถสูงจะตอบข้อสอบผิด ซึ่งอาจจะเป็นเพราะความไม่รอบคอบ ทำให้ผู้นั้นไม่ได้เลือกคำตอบที่ถูกต้อง จะเห็นได้จากปลายสูงสุดของโค้งแสดงลักษณะของข้อสอบ (ICC) ต่ำกว่า 1 ซึ่งเขียนในรูปโมเดลทางคณิตศาสตร์ได้ดังนี้

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

เมื่อ $P_i(\theta)$ แทน ความน่าจะเป็นที่ผู้สอบจะตอบข้อที่ i ได้ถูกต้องด้วยความสามารถ θ

b_i แทน ค่าพารามิเตอร์ความยากของข้อที่ i คือค่า θ ที่ $P_i(\theta)$ เท่ากับ $(1+c_i)/2$

a_i แทน ค่าพารามิเตอร์อำนาจจำแนกของข้อที่ i คือ ค่าสัดส่วนของความชันของ $P_i(\theta)$ ที่ $\theta = b_i$

c_i แทน ค่าพารามิเตอร์ของการเดาของข้อสอบข้อที่ i คือ ความน่าจะเป็นที่ผู้สอบไม่มีความรู้เลยตอบข้อสอบข้อนั้นถูก

D แทน ค่าคงที่มีค่าเท่ากับ 1.702

e แทน ค่าเอ็กซ์โปเนนเชียลเป็นค่าคงที่ซึ่งมีค่าเท่ากับ 2.718

y_i แทน ค่าความรอบคอบในการตอบข้อสอบข้อที่ i ซึ่งมีค่าน้อยกว่า 1

อย่างไรก็ตามค่าพารามิเตอร์การรอบคอบนี้ยังไม่สามารถที่จะระบุได้ชัดเจน ทำให้ผู้ใช้โมเดลโลจิสติกไม่นิยมใช้โมเดลโลจิสติกแบบ 4 พารามิเตอร์ ยังคงนิยมใช้โมเดลโลจิสติกแบบ 1-2 และ 3 พารามิเตอร์ ซึ่งจะใช้โมเดลใดก็จะขึ้นอยู่กับจุดมุ่งหมายของการสอบและการใช้คะแนนจากการสอบข้อสอบนั้น ๆ

2. โมเดลที่ใช้กับข้อมูลแบบหลายชั้น (Multicategory)

เป็นโมเดลที่ใช้กับแบบทดสอบที่มีผลการตอบที่ให้คะแนนมากกว่า 2 ค่า ให้คะแนนเป็นไปตามลำดับ เช่น แบบทดสอบวัดเจตคติ (Attitude-Test) แบบทดสอบวัดบุคลิกภาพ

(Personality test) ฯลฯ โมเดลที่ใช้กับข้อมูลแบบหลายขั้นนี้จะพัฒนามาจากโมเดลที่ใช้กับข้อมูลแบบแบ่งสอง (Dichotomous) ซึ่งประกอบโมเดลต่าง ๆ ดังนี้

2.1 โมเดล เกรด เรสพอน (Grade response model) โมเดลนี้มีแนวคิดหลักพื้นฐานมาจากการขยายโมเดลแบบ 2 พารามิเตอร์ ทั้งแบบโลจิสติกและนอร์มัล ออไจน์ โดยมีซาเมจิมา (Samejima) ได้พัฒนาขึ้นในปี ค.ศ.1969 โดยมีข้อตกลงเบื้องต้นที่เกี่ยวกับการตอบข้อสอบต่อคำถามหรือข้อความ i ของผู้ตอบ ซึ่งสามารถแบ่งออกได้เป็น m_i+1 ลำดับชั้น (Categories) และให้คะแนนแต่ละระดับ (x_i) เป็น 0, 1, ..., m_i ตามลำดับ ดังนั้น การตอบข้อสอบแต่ละลำดับของข้อความ (x_i) ของผู้ตอบแต่ละคน จากการตอบข้อความทั้งหมด n รายการ สามารถแทนด้วยเวกเตอร์ของจำนวนเต็ม คือ $V = (x_1, x_2, \dots, x_n)$ ความน่าจะเป็นหรือโอกาส (Probability) ในการตอบข้อคำถามหรือข้อคำถามแต่ละรายการสำหรับผู้ตอบแต่ละคน จากการตอบในลำดับชั้นใด ๆ ของโมเดล เกรด เรสพอน สามารถเขียนเป็นโมเดลของคณิตศาสตร์ได้ดังนี้

$$P_{x_i}(\theta) = P_{x_i}^*(\theta) - P_{(x_i+1)}^*(\theta).$$

เมื่อ $P_{x_i}(\theta)$ แทน โอกาสของผู้ตอบแต่ละคน ซึ่งมีคุณสมบัติภายใน ที่จะได้คะแนนในลำดับชั้นใด ๆ หรือสูงกว่าของข้อความ i ซึ่งแทนด้วยสมการดังนี้

$$P_{x_i}^*(\theta) = \frac{\exp(Da_i(\theta - b_{x_i}))}{1 + \exp(Da_i(\theta - b_{x_i}))}$$

เมื่อ D แทน ค่าคงที่ (Sealing constant) เท่ากับ 1.702

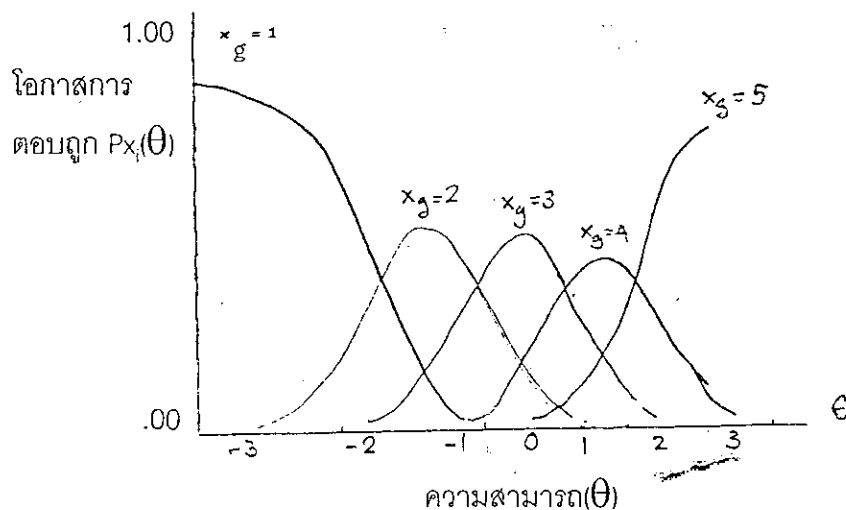
a แทน ค่าพารามิเตอร์อำนาจจำแนกของข้อความรายการ i

θ แทน ความสามารถของผู้ตอบหรือคุณลักษณะของผู้ตอบ

b_{x_i} แทน ค่าพารามิเตอร์ความยากของข้อความรายการที่ i ในชั้นที่ x_i

เมื่อ $x_i = 0, 1, \dots, m_i$

อนึ่ง การนำโมเดลเกรดเรสพอนไปประยุกต์ใช้ ส่วนมากกำหนดให้ค่าอำนาจจำแนกภายในแต่ละระดับชั้น ($x_i = 0, 1, \dots, m_i$) ของข้อความรายการใด ๆ หรือความชันของ Trace lines ทั้งหมดใน $P_{x_i}(\theta) = 1$ และ $(m_i + 1)(\theta) = 0$ ดังนั้น ความน่าจะเป็นหรือโอกาสในการตอบข้อความในระดับชั้น $x_i P_{x_i}(\theta)$ มีค่าดังอสมการ $P_{x_i}(\theta) - P_{(m_i+1)}(\theta) > 0$ ซึ่งสามารถนำมาเขียนกราฟแสดงลักษณะการตอบข้อความในแต่ละระดับชั้นได้ดังภาพประกอบ 8



ภาพประกอบ 8 โค้งแสดงลักษณะการตอบสำหรับข้อความในโมเดล เกรด เรสพอน

จากภาพ 8 จะเห็นได้ว่าผู้ตอบที่มีคุณลักษณะภายใน (θ) หรือความสามารถสูงย่อมมีความน่าจะเป็นที่จะตอบข้อความในระดับชั้นที่สูงกว่า เช่นเดียวกันกับผู้มีความสามารถ (θ) หรือคุณลักษณะภายในต่ำกว่าย่อมมีโอกาสที่จะตอบข้อความในระดับชั้นต่ำ ๆ มากกว่าผู้ที่มีคุณลักษณะภายในสูงกว่า จากแนวคิดหลักพื้นฐานของโมเดล เกรด เรสพอน ดังกล่าว นักวัดผลการศึกษาได้นำโมเดลไปประยุกต์ใช้ในการวัดเจตคติการทดสอบแบบเทลเลอร์ ฯลฯ และได้พัฒนาโปรแกรมสำเร็จรูป Multilog 5 เพื่อนำมาใช้ในการประมาณค่าพารามิเตอร์ต่าง ๆ ของโมเดลนี้

2.2 โมเดลเรตติ้งสเกล (Rating scale model ; SCALE)

เป็นโมเดลที่แอนดริช (Andrich) พัฒนาขึ้นในปี ค.ศ.1978 โดยการนำราชคโมเดลที่ใช้กับข้อคำถามที่ตรวจให้คะแนนเป็นแบบสองค่าจากการวัดผลสัมฤทธิ์ทางการเรียนมาประยุกต์ใช้กับการวัดเจตคติของบุคคล คือ จากผลการตอบถูกให้ 1 ผิดให้ 0 ก็มาเป็นเห็นด้วย และไม่เห็นด้วย แทน จากนั้นก็ขยายผลการตอบเป็น 3 ช่วง คือ ไม่เห็นด้วยให้ 0 คะแนน ไม่แน่ใจให้ 1 คะแนน และเห็นด้วยให้ 2 คะแนน ซึ่งตัวเลขที่กำหนดเป็นการแทนลำดับชั้นของการตอบคำถาม เพื่อแบ่ง

ความคิดเห็นให้ละเอียดขึ้น ความคิดเห็นของผู้ตอบจะมีลักษณะเป็นลำดับขั้น ซึ่งจะมีลักษณะเป็นช่วงที่ต่อเนื่องกัน (Psychological continuum) ดังนั้น การเลือกตอบแต่ละขั้นจะต้องผ่านขั้นต้นมาก่อน ซึ่งการผ่านลำดับขั้นของความคิดเห็นในแต่ละระดับนั้น จะขึ้นอยู่กับค่าประจำลำดับขั้น หรือที่เรียกว่า ค่าเทรชโฮลด์ (Threshold value ; T_k) ค่าเทรชโฮลด์ประจำลำดับขั้นเดียวกันของข้อความในแต่ละรายการในแบบทดสอบฉบับเดียวกันของโมเดลจะถูกกำหนดให้เท่ากัน และการกำหนดค่าเทรชโฮลด์เริ่มต้นเป็นศูนย์ ($T_0 = 0$) โดยเมื่อรวมค่าเทรชโฮลด์ในระดับขั้นต่าง ๆ ของข้อความแต่ละรายการกับค่าความยากประจำข้อ (Scale value ; δ_i) ซึ่งมีค่าแปรเปลี่ยนไปตามคุณสมบัติของข้อความแต่ละรายการแล้ว ก็จะได้ค่าความยากของแต่ละขั้น (Difficulty of the steps) ดังสมการ

$$\delta_{ik} = \delta_i + T_k$$

เมื่อ δ_{ik} แทน ค่าความยากของข้อความ i ในระดับขั้นที่ k

δ_i แทน ค่าความยากประจำข้อ (Scale value) ของข้อความรายการที่ i

T_k แทน ค่าเทรชโฮลด์ (Threshold) ในระดับขั้นที่ k

ดังนั้น ความน่าจะเป็นที่ผู้ตอบคนที่ n จะเลือกตอบในระดับขั้นที่ k ของข้อความรายการที่ $P_i(\theta_{nik})$ หรือ π_{nik} คือ

$$(\phi_{nik}) = \frac{\pi_{nik}}{\pi_{nik-1} + \pi_{nik}} = \frac{\exp [\beta_n - (\delta_i + T_k)]}{1 + \exp [\beta_n - (\delta_i + T_k)]} ; k=1, 2, \dots, m$$

เมื่อ π_{nik} แทน ความน่าจะเป็นของผู้ตอบคนที่ n ที่จะเลือกตอบในระดับขั้นที่ k ของข้อความรายการที่ i

π_{nik-1} แทน ความน่าจะเป็นของผู้ตอบคนที่ n ที่จะเลือกตอบในระดับขั้นที่ $k-1$ ของข้อความรายการที่ i

β_n แทน คุณลักษณะ (Traits) ของผู้ตอบคนที่ n

T_k แทน ค่าเทรชโฮลด์ในระดับขั้นที่ k

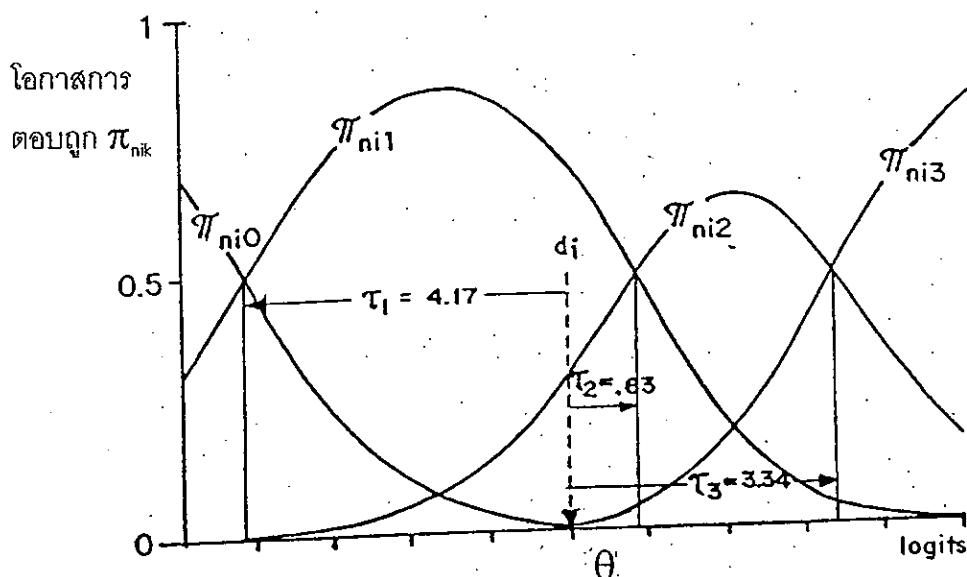
จากสมการดังกล่าวเมื่อพิจารณาความเป็นจริงเกี่ยวกับความเป็นไปได้ที่ผู้ตอบจะต้อง

เลือกตอบตามความคิดเห็นในระดับใดระดับหนึ่งของข้อความ ซึ่งมีทั้งหมด m_{i+} ระดับแล้ว ดังนั้นสามารถแสดงให้อยู่ในรูปของความน่าจะเป็นหรือโอกาส (Probability) ของผู้ตอบคนที่ n ที่จะตอบข้อความรายการที่ i โดยผ่านในขั้นที่ j หรือได้คะแนน x (π_{nik}) ได้ดังนี้

$$\pi_{nix} = \frac{\exp \left\{ \sum_{j=0}^x [\beta_n - (\delta_i + \tau_j)] \right\}}{\sum_{k=0}^m \exp \left\{ \sum_{j=0}^k [\beta_n - (\delta_i + \tau_j)] \right\}} \quad ; x = 0, 1, 2, \dots, m$$

เมื่อกำหนดให้ $\tau_0 = 0$ เพื่อที่จะให้ $\exp \sum_{j=0}^0 [\beta_n - (\delta_i + \tau_j)] = 1$

สำหรับการนำไปประยุกต์ใช้จะประมาณค่าคุณลักษณะของผู้ตอบคนที่ $n(n)$ ประจำข้อความรายการที่ i (δ_i) และค่าเทรตโฮลด์ $T_1, T_2, \dots, T_m$ ในข้อความที่มีระดับการตอบ m_{i+} ระดับ จากฟังก์ชันทางคณิตศาสตร์ที่แทนความสัมพันธ์ระหว่างค่าพารามิเตอร์ต่าง ๆ ที่ประมาณค่าจากโมเดลนี้ เมื่อนำมาเขียนโค้งแสดงความน่าจะเป็นของการเลือกตอบในแต่ละลำดับขั้นของข้อความ (Response category probability curves) ได้ดังภาพประกอบ 9



ภาพประกอบ 9 แสดงโค้งลักษณะลำดับขั้นของข้อความโมเดลเรตติงสเกล

จากภาพ 9 จะเห็นว่าค่าความยากประจำแต่ละระดับขึ้นอยู่กับค่าประจำข้อค่า T_1 , T_2 , T_3 มีค่าแตกต่างกันในแต่ละระดับขั้นของข้อความรายการ ผลรวมของค่าเทรตโฮลด์ทั้งหมดรอบค่าประจำข้อมีค่าเท่ากับ 0 ($T = 0$) ซึ่งรูปแบบ (Pattern) ของเส้นโค้งต่าง ๆ รวมทั้งค่าเทรตโฮลด์ของแต่ละระดับขั้นกำหนดให้มีค่าคงที่ (Fixed) ทุกข้อความ ดังนั้น เมื่อโค้งแสดงลักษณะการตอบแต่ละระดับขั้นของข้อความรายการอื่น ๆ ที่อยู่ในแบบทดสอบฉบับเดียวกับข้อความในภาพประกอบ 9 ก็จะมีลักษณะรูปแบบที่เหมือนกัน เพียงแต่ค่าประจำข้อของข้อความรายการต่าง ๆ เหล่านั้นจะมีค่าแตกต่างกันไป ซึ่งก็จะมีผลต่อค่าความยากในแต่ละระดับขั้นด้วย โดยอาจจะขยับไปทางด้านขวาของภาพ ถ้าข้อความนั้นมีค่าประจำข้อสูงกว่า หรืออาจจะขยับไปทางด้านซ้ายของภาพ ถ้าข้อความมีค่าประจำข้อต่ำกว่าบนมาตราวัดความสามารถหรือคุณลักษณะเดียวกัน

โมเดลเรตติงสเกล แตกต่างไปจากโมเดลเกรตเรสพอน พอดีสรุปโดยสังเขปได้ดังตาราง 5

ตาราง 5 การเปรียบเทียบความแตกต่างระหว่างโมเดลเรตติงสเกลกับโมเดลเกรตเรสพอน

โมเดล เรตติง สเกล (SCALE)	โมเดลเกรต เรสพอน
1) ค่าเทรตโฮลด์ในลำดับขั้นเดียวกันของข้อความทุกรายการกำหนดให้เท่ากัน	1) ค่าเทรตโฮลด์ในแต่ละระดับขั้นมีค่าแตกต่างกัน
2) ค่าอำนาจจำแนกที่ตำแหน่งของค่าเทรตโฮลด์ทั้งหมดสำหรับข้อความทุกรายการสมมติให้มีค่าเท่ากัน	2) ค่าอำนาจจำแนกของข้อความแต่ละรายการมีค่าแตกต่างกัน
3) ผลรวมของคะแนนดิบ (คะแนนรวม) เป็นสถิติที่พอเพียง (Sufficient statistics) สำหรับการประมาณค่าพารามิเตอร์ของผู้สอบ	3) ผลรวมของคะแนนดิบเป็นสถิติที่ไม่พอเพียง
4) ค่าพารามิเตอร์ของข้อความและของผู้ตอบสามารถแยกกันประมาณค่าได้อย่างเป็นอิสระหรือมีคุณสมบัติที่เรียกว่า 'Specific objectivity'	4) ค่าพารามิเตอร์ของข้อความและของผู้ตอบไม่สามารถแยกกันประมาณค่าได้

2.3 โมเดล พาเชียล เครดิต (Partial credit model)

เป็นโมเดลที่มาสเตอร์ (Masters) ได้พัฒนาขึ้นในปี ค.ศ. 1982 โดยมีแนวคิดหลักพื้นฐานของการพัฒนามาจากการขยายโมเดลราชค์ เพื่อให้สามารถนำมาใช้กับแบบทดสอบที่มีรูปแบบการตรวจให้คะแนนแบบลำดับขั้น หรือแบบให้คะแนนเป็นหลายค่า (Polychotomous) เช่นเดียวกับโมเดลเรตติงสเกลของแอนดริช แต่ให้ผลการวิเคราะห์ที่มีรายละเอียดมากกว่า

การนำโมเดลนี้มาประยุกต์ใช้กับข้อคำถามหรือข้อความต่าง ๆ จะทำให้สามารถประมาณค่าความสามารถหรือคุณลักษณะของผู้ตอบได้แม่นยำกว่าการให้คะแนนแบบ 0-1 เนื่องจากโมเดล พาเชียล เครดิต พัฒนามาจากราซค์โมเดล ดังนั้นจึงสามารถช่วยอธิบายความน่าจะเป็น (Probability) ของผู้ตอบคนที่ n ที่จะได้คะแนน 1 มากกว่า 0 (ผ่านขั้นที่ 1) ในการตอบข้อความ i (ϕ_{ni1})

$$\phi_{ni1} = \frac{\pi_{ni1}}{\pi_{ni0} + \pi_{ni1}} = \frac{\exp(\theta_n - \delta_{i1})}{1 + \exp(\theta_n - \delta_{i1})}$$

เมื่อ ϕ_{ni1} แทน ความน่าจะเป็นของผู้ตอบคนที่ n ที่จะได้คะแนน 1 มากกว่าคะแนน 0 (หรือผ่านขั้นที่ 1) จากการตอบข้อ i

π_{ni0} แทน ความน่าจะเป็นของผู้ตอบคนที่ n ที่จะได้คะแนน 0 (ไม่ผ่านขั้นที่ 1) จากการตอบข้อ i

π_{ni1} แทน ความน่าจะเป็นของผู้ตอบคนที่ n ที่จะได้คะแนน 1 จากการตอบข้อ i

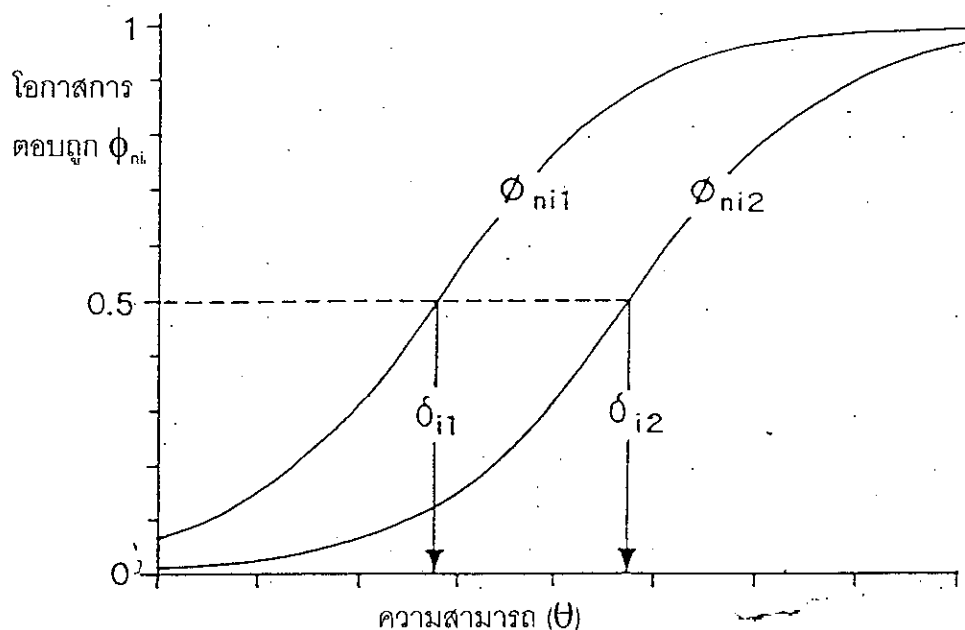
θ_n แทน ระดับความสามารถหรือคุณลักษณะของผู้ตอบคนที่ n

δ_{i1} แทน ค่าความยากประจำขั้นที่ 1 (First step) ของข้อที่ i

เนื่องจาก $\pi_{ni0} + \pi_{ni1} < 1$ ทั้งนี้ เพราะผลรวมของความน่าจะเป็นของการตอบในระดับทั้งสองเป็นเพียงความน่าจะเป็นที่จะตอบได้คะแนน 1 ($x_i = 1$) หรือผ่านในขั้นที่ 1 แต่ยังมีความน่าจะเป็นในการตอบที่จะได้คะแนน 2, 3, ... 4 สำหรับผู้ตอบที่ตอบข้อความผ่านในขั้นที่ 1 เข้าสู่ระดับขั้นที่ 2 แล้วสามารถผ่านขั้นที่ 2 ได้ ดังนั้น ความน่าจะเป็นที่จะได้คะแนนเป็น 2 ($x_i = 2$) สามารถคำนวณได้จากโมเดลทางคณิตศาสตร์ดังนี้

$$\phi_{ni2} = \frac{\pi_{ni2}}{\pi_{ni1} + \pi_{ni2}} = \frac{\exp(\theta_n - \delta_{i2})}{1 + \exp(\theta_n - \delta_{i2})}$$

จากค่า ϕ_{ni1} และ ϕ_{ni2} สามารถนำมาเขียนโค้งลักษณะของการตอบข้อคำถามที่มีลำดับการตอบ 2 ขั้น (Two-step Item) ดังภาพประกอบ 10



ภาพประกอบ 10 โค้งความน่าจะเป็นในการเลือกตอบระดับที่ 1 และ 2 ของข้อ i

จากภาพประกอบ 10 จะเห็นว่าความชัน (Slope) ของโค้งลักษณะความน่าจะเป็นในการเลือกตอบระดับที่ 1 และที่ 2 ของข้อ i มีค่าเท่ากัน และพบว่าระดับ 2 มีค่าความยากมากกว่าระดับ 1 ($\delta_{i1} < \delta_{i2}$) ซึ่งบางครั้งถ้าเป็นโจทย์ปัญหาอาจจะมีความยากขั้นที่ 2 น้อยกว่าขั้นที่ 1 ก็ได้ หลักการสำคัญของโมเดลพาสเซียล เครดิต ต้องการเพียงให้มีข้อความที่เรียงลำดับการตอบ (Ordered in sequence) โดยไม่จำเป็นที่ข้อความจะต้องมีค่าความยากประจำมากกว่ากัน แต่ขั้นต้องเรียงลำดับกัน ถ้าข้อสอบมีระดับการตอบมากกว่า 2 ระดับก็สามารถคำนวณหาความน่าจะเป็นในการตอบลำดับขั้นอื่น ๆ ของผู้ตอบที่ n ได้ดังสมการนี้

$$\phi_{nik} = \frac{\pi_{nik}}{\pi_{nik-1} + \pi_{nik}} = \frac{\exp(\theta_n - \delta_{ik})}{1 + \exp(\theta_n - \delta_{ik})} \quad ; k = 1, 2, \dots, m_i$$

จากสูตรทั่วไปนี้แสดงให้เห็นเกี่ยวกับการตอบของผู้ตอบคนที่ n ในการเลือกระดับใด ๆ ของข้อที่ i จากจำนวนระดับการตอบที่เป็นไปได้ทั้งหมด $m_i + 1$ ระดับ ดังนั้น ความน่าจะเป็นของผู้ตอบคนที่ n ที่จะได้คะแนน x ในการตอบข้อที่ i อาจเขียนแทนในรูปสูตรทั่ว ๆ ไป คือ

$$\pi_{nix} = \frac{\exp \left[\sum_{j=0}^x (\theta_n - \delta_{ik}) \right]}{\sum_{k=0}^m \exp \left[\sum_{j=0}^k (\theta_n - \delta_{ik}) \right]} ; x = 0, 1, 2, \dots, m_i$$

โดยกำหนดให้ค่าพารามิเตอร์ความยากของข้อใด ๆ ในระดับชั้น 0 เท่ากับ 0

$$(\delta_{i0} = 0) \text{ เพื่อให้ } \sum_{j=0}^0 (\theta_n - \delta_{ik}) = 0 \text{ ดังนั้น } \exp \sum_{j=0}^0 (\theta_n - \delta_{ik}) = 1$$

เมื่อ π_{nix} แทน ความน่าจะเป็นของผู้ตอบคนที่ n ที่จะได้คะแนน x ในการตอบข้อความที่ i

x แทน คะแนนผลของการตอบในแต่ละระดับที่เลือกตอบ ซึ่งมีค่าเท่ากับ $0, 1, 2, \dots, m_i$

δ_{ij} แทน ค่าความยากประจำชั้นที่ j ของข้อความที่ i โดยมีค่าแต่ละชั้นเป็น $\delta_{i1}, \delta_{i2}, \dots, \delta_{ix}$

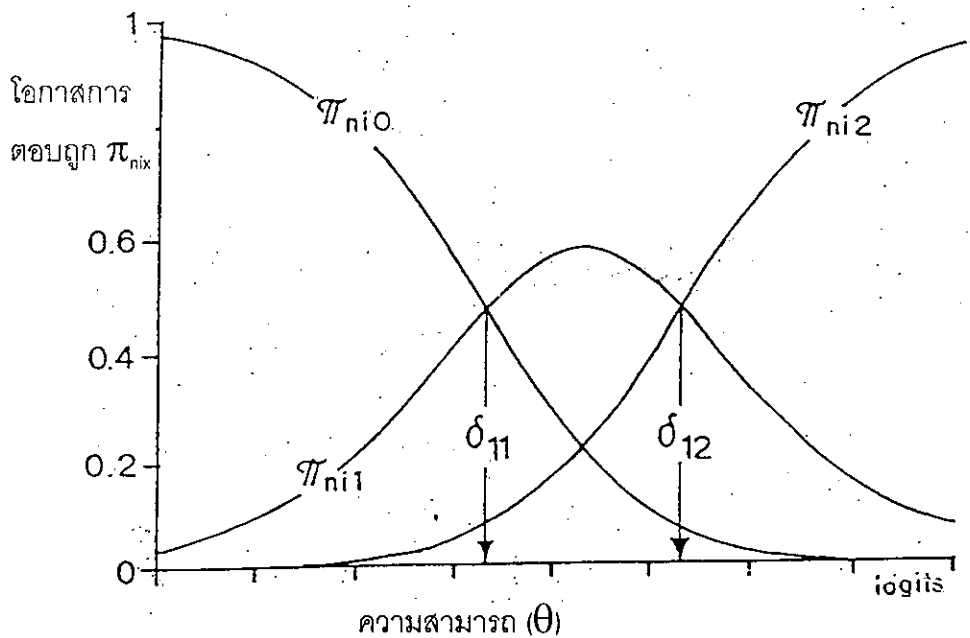
n แทน ระดับความสามารถหรือคุณลักษณะของผู้ตอบคนที่ n หรือเขียนให้อยู่ในรูปใหม่ดังนี้

$$P_{nix} = \frac{1}{1 + \sum_{k=1}^{m_i} \exp \left[\sum_{j=1}^k (\theta_n - \delta_{ij}) \right]} \quad \text{สำหรับ } x = 0$$

$$P_{nix} = \frac{\exp \left[\sum_{j=1}^x (\theta_n - \delta_{nj}) \right]}{1 + \sum_{k=1}^{m_i} \exp \left[\sum_{j=1}^k (\theta_n - \delta_{nj}) \right]} \quad \text{สำหรับ } x = 1, 2, \dots, m_i$$

เมื่อ P_{nix} แทน ความน่าจะเป็นของผู้ตอบคนที่ n ที่จะได้คะแนน x จากการตอบของ ข้อความที่ i

จากโมเดลทางคณิตศาสตร์ที่แทนความสัมพันธ์ระหว่างค่าพารามิเตอร์ของข้อความ (δ_{nj}) และความน่าจะเป็นของการตอบข้อความในแต่ละระดับชั้น (π_{nik}) เมื่อนำมาเขียนโค้งแสดงความน่าจะเป็นในการตอบแต่ละลำดับชั้น สำหรับข้อความที่มีรูปแบบการตอบมี 2 ขั้นตอน ดังภาพประกอบ 11



ภาพประกอบ 11 แสดงความน่าจะเป็นในการเลือกตอบข้อความระดับต่างๆ (3 ระดับ) ตามโมเดล พาร์เซียล เครดิต

เมื่อ π_{ni0} แทน ความน่าจะเป็นที่ผู้ตอบจะได้คะแนน 0 ($x_i=0$) (หรือเลือกตอบ
ในระดับไม่เห็นด้วย) จากการตอบข้อที่ i

π_{ni1} แทน ความน่าจะเป็นที่ผู้ตอบจะได้คะแนน 1 ($x_i=1$) (หรือเลือก
ตอบในระดับไม่แน่ใจ) จากการตอบข้อที่ i

π_{ni2} แทน ความน่าจะเป็นที่ผู้ตอบจะได้คะแนน 2 ($x_i=2$) (หรือเลือก
ตอบในระดับเห็นด้วย) จากการตอบข้อที่ i

จากภาพประกอบ 11 จะเห็นว่า δ_{i1} และ δ_{i2} เป็นจุดตัด (Intersection) ของโค้งความน่าจะเป็นในการตอบแต่ละลำดับขั้นบนแกน x ณ จุดตัดโค้งนี้ คือ ค่าความยากประจำแต่ละลำดับขั้นหรือค่าประจำขั้น (Step value) คือ δ_{i1} เป็นค่าประจำระดับไม่แน่ใจ δ_{i2} ประจำระดับเห็นด้วย ส่วนระดับไม่เห็นด้วย ซึ่งเป็นความรู้สึกเริ่มต้นกำหนดให้เป็นศูนย์ ($\delta_{i0}=0$) การกำหนดตำแหน่ง ณ จุดตัดของโค้งลักษณะการตอบแต่ละลำดับขั้นของข้อเป็นความยากประจำขั้นที่ 1 และ 2 แสดงว่าความรู้สึกของผู้ตอบผู้หนึ่ง (n) มีค่ามากกว่าค่าความยากประจำขั้นที่ 1 (δ_{i1}) แล้วผู้นั้นย่อมมีความน่าจะเป็นหรือโอกาสที่จะเลือกในระดับไม่แน่ใจมากกว่าการเลือกตอบในระดับไม่เห็นด้วย

โมเดล พาเชียล เครดิต อาจเรียกสั้น ๆ ว่า เครดิต โมเดล เป็นโมเดลหนึ่งที่ถูกจัดให้อยู่ในกลุ่มทฤษฎีการตอบข้อคำถาม (Item response theory ; IRT)ตระกูลราชค์โมเดล (The family of Rasch Models) ซึ่งประกอบไปด้วยโมเดลต่าง ๆ ดังนี้ (Wright and Masters, 1982; ; Wright, 1988(a) : 286-287)

1. ไดโคโตมัส โมเดล (Dichotomous model) หรือมักนิยมเรียกว่า 'ราชค์ โมเดล' (Rasch model) โมเดลนี้ถือได้ว่าเป็นโมเดลหลักพื้นฐานที่ จอร์จ ราชค์ (Gorg Rasch) นักคณิตศาสตร์ชาวเดนมาร์กพัฒนาขึ้นในระหว่างปี ค.ศ. 1951 ถึง 1959 โดยใช้กับข้อมูลการตอบที่มีรูปแบบง่ายที่สุด กล่าวคือใช้กับข้อมูลผลการตอบที่แบ่งแยกเป็น 2 ระดับ หรือ 2 ทาง คือ ผ่าน/ไม่ผ่าน สำเร็จ/ไม่สำเร็จ หรือถูก/ผิด เป็นต้น และส่วนใหญ่จะตรวจให้คะแนน สำหรับข้อคำถามที่ตอบถูก (ผ่าน) เป็น 1 คะแนน ข้อคำถามที่ตอบผิด (ไม่ผ่าน) เป็น 0 คะแนน ดังนั้น ไดโคโตมัส โมเดล จึงใช้สำหรับวิเคราะห์ข้อคำถามในแบบทดสอบเฉพาะที่มีรูปแบบของการตอบเพียงขั้นตอนเดียว (One-step item) เช่น ข้อคำถามในแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนที่ตรวจให้คะแนนเป็น 1 เมื่อผู้สอบสามารถตอบข้อคำถามได้อย่างถูกต้องและจะได้คะแนนเป็น 0 เมื่อตอบข้อคำถามผิด โดยจะไม่มี การให้คะแนนในขั้นตอนอื่น ๆ ที่มากกว่า 2 ระดับนี้ ซึ่งมีฟังก์ชันทางคณิตศาสตร์ ดังนี้

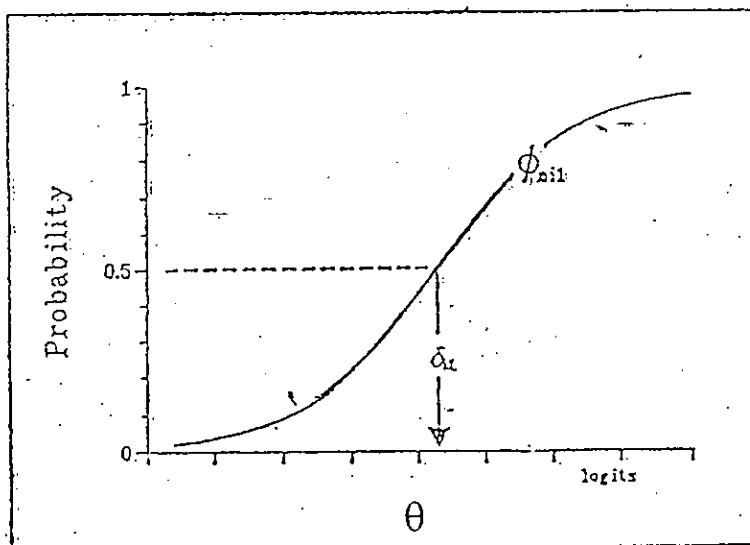
$$\phi_{ni1} = \frac{\exp(\theta_n - \delta_{ni1})}{1 + \exp(\theta_n - \delta_{ni1})}$$

เมื่อ ϕ_{ni1} แทน ความน่าจะเป็น หรือโอกาส (Probability) ของผู้สอบคนที่ n จะได้รับ คะแนนเป็น 1 (ตอบถูก) มากกว่า 0 (ตอบผิด) จากการตอบข้อคำถามข้อที่ 1

θ_n แทน ความสามารถ หรือคุณลักษณะของผู้สอบ (Ability or Traits)

δ_{ni1} แทน ความยากของข้อคำถามในขั้นที่ i (ซึ่งในโมเดลนี้มีเพียงขั้นเดียว) ของข้อคำถามข้อที่ i

จากฟังก์ชันทางคณิตศาสตร์สามารถนำมาเขียนแทนด้วยโค้งลักษณะการตอบข้อคำถาม หรือโค้งลักษณะข้อคำถาม (Operating characteristic curve ; OCC or Item characteristic curve ; ICC) ดังภาพประกอบ 12



ภาพประกอบ 12 โค้งลักษณะข้อคำถามที่มีรูปแบบของการตอบเพียงขั้นตอนเดียว (One-Step item)

2. ไบนอมิยัล ไทรอัลส์ โมเดล (Binomial trial model) เป็นโมเดลที่ราซค์และแอนดริซได้พัฒนาขึ้นในระหว่างปี ค.ศ. 1972 ถึง 1978 โดยพัฒนาขึ้นมาจากไดโคโตมัส เพื่อให้ใช้ได้ข้อมูลประมาณการนับจำนวนครั้งของความสำเร็จ และจำนวนครั้งที่ไม่สำเร็จในการตอบข้อคำถามแต่ละข้อ ทั้งหมด m ครั้งซึ่งเป็นอิสระจากกัน โมเดล ดังกล่าวนี้นำไปประยุกต์ใช้กับการทดสอบ

ภาคปฏิบัติทางด้านทักษะพิสัย (Psychomotor skills) โดยอาศัยการนับหรือสังเกตจำนวนครั้งของการปฏิบัติงานที่กระทำได้เป็นผลสำเร็จ (x) จากการกำหนดทางคณิตศาสตร์ของโมเดลนี้คือ

$$P(X = x | \theta) = \binom{n}{x} P(\theta)^x Q(\theta)^{n-x}$$

เมื่อ $P(x=x | \theta)$ แทน ความน่าจะเป็นของผู้สอบในการปฏิบัติเป็นผลสำเร็จ x ครั้ง
จากการปฏิบัติ n ครั้ง

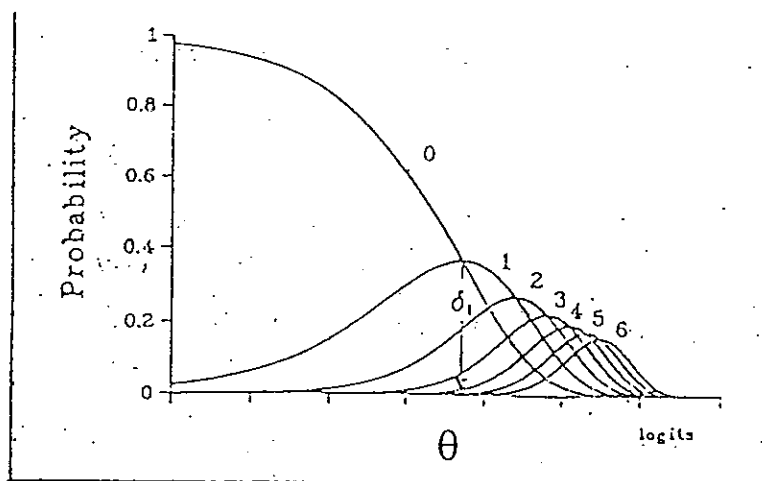
3. ปัวซอง เคาท์ โมเดล (Poisson counts model) เป็นโมเดลที่ราซด์ ได้พัฒนาขึ้นในปี ค.ศ.1960 เพื่อนำมาใช้กับข้อมูลที่มีลักษณะเช่นเดียวกับไบโนเมียลไทรอัลส์ โมเดล กล่าวคือ เป็นข้อมูลประเภทการนับจำนวนครั้งความสำเร็จ (x) ของผู้ตอบในการตอบข้อคำถามจำนวน m ครั้ง เพียงแต่โมเดลนี้แตกต่างจากไบโนเมียล ไทรอัลส์ โมเดล ตรงที่ไม่สามารถกำหนดขอบเขตจำนวนครั้งความสำเร็จ หรือไม่สำเร็จแน่นอนลงไปได้ ($x = 0, 1, \dots$) เช่น จำนวนครั้งของการเดาะลูกฟุตบอลไม่ตกกลิ้งสู่พื้นสนามภายในระยะเวลา 5 นาที หรือการทดสอบการอ่านปากเปล่า (Oral reading test) จากเนื้อหาและระยะเวลาที่กำหนดให้ จากตัวอย่างข้างต้นจะพบว่า ไม่สามารถนับจำนวนครั้งของการเดาะลูกฟุตบอลไม่ตกกลิ้งสู่พื้นสนาม (สำเร็จ) หรือการอ่านคำผิด (ไม่สำเร็จ) แน่แน่นอนลงไปได้ ทั้งนี้ เพราะอย่างน้อยที่สุดโอกาสของความสำเร็จดังกล่าว ก็ยังมีโอกาสเกิดขึ้นตามความสามารถของผู้สอบแต่ละคนภายในระยะเวลาที่กำหนดให้ เพียงแต่โอกาสเหล่านั้นจะมีโอกาสเกิดขึ้นน้อยลงไปตามระยะเวลาของการปฏิบัติ ดังมีฟังก์ชันทางคณิตศาสตร์ดังนี้

$$P(X = x | \theta, b) = \frac{e^{x(\theta-b)}}{x! e^{b(\theta-b)}}$$

เมื่อ x คือจำนวนครั้งที่ปฏิบัติได้เป็นผลสำเร็จในเวลาที่กำหนด

b คือความยากของงาน

จากฟังก์ชันทางคณิตศาสตร์สามารถนำมาเขียนแทนด้วยโค้งลักษณะการตอบแต่ละระดับ
ขั้นของข้อคำถาม (Operating characteristic curves ; OCCs) หรือโค้งลักษณะของข้อคำถาม
ดังภาพประกอบ 13



ภาพประกอบ 13 โค้งลักษณะข้อคำถามสำหรับการตอบ 7 ลำดับขั้นแรกของบัวของ เคาท์ โมเดล

2.4 โมเดลการตอบนามบัญญัติ (Nominal response model)

โมเดลนี้ บ็อค (Bock) และซามะจิมา (Samejima) เป็นผู้เสนอขึ้นมา เนื่องจากมีความเห็นว่าโมเดลโลจิสติกและโมเดลนอมัลลอจีฟนั้น มุ่งตอบสนองข้อมูลที่มีลักษณะแบบแบ่งสอง (Dichotomous) ซึ่งโมเดลนี้จะให้คะแนนแบบพหุ (Multidichotomous) มีสมการทางคณิตศาสตร์ที่แสดงโค้งลักษณะของข้อคำถามที่เชื่อว่าความน่าจะเป็นที่ผู้สอบมีความสามารถจะเลือกตัวเลือก k (จาก m ตัวเลือกต่อข้อ) สำหรับข้อ i หาได้จากสมการ

$$P_{ik}(\theta) = \frac{e^{b_{ik}^* + a_{ik}^* \theta}}{\sum_{h=1}^m e^{b_{ih}^* + a_{ih}^* \theta}} \quad (i = 1, 2, \dots, n; k = 1, 2, \dots, m)$$

$\sum_{k=1}^m P_{ik}$ คือ ผลรวมของค่าโอกาสที่น่าจะเป็นของรายการ $m = 1$

b_{ik}^* , a_{ik}^* คือ พารามิเตอร์ของรายการที่ i ที่เลือกข้อรายการที่ k จากรายการทั้งหมด m

3. โมเดลที่ใช้กับข้อมูลแบบต่อเนื่อง (Continuous)

โมเดลที่ใช้กับข้อมูลแบบต่อเนื่องนี้ มีโมเดลที่ใช้วิเคราะห์โมเดลหนึ่ง คือ โมเดลการตอบแบบต่อเนื่อง (Continuous scale model)

โมเดลนี้จะถูกพิจารณาว่าเป็นกรณีเฉพาะของโมเดลการตอบแบบให้ระดับคะแนนที่ต่างกับโมเดลที่กล่าวมาคือผู้สอบจะตอบคำถามลงบนสเกลที่ต่อเนื่อง (Continuous Scale) ซึ่งโมเดลนี้จะมีประโยชน์ต่อนักจิตวิทยาสังคมและการวัดเจตคติ

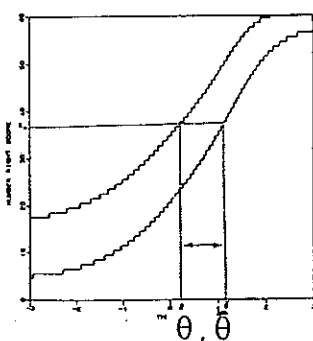
โมเดลของทฤษฎีการตอบข้อสอบจะเห็นได้ว่ามีหลากหลายและมีโอกาสที่จะพัฒนามากขึ้น เพื่อให้เหมาะสมกับการสอบ จะได้ช่วยให้การตีความหมายคะแนนจากการสอบเหมาะสมหรือทราบความสามารถของผู้สอบที่แท้จริงได้ นั่นคือ ให้เกิดความคลาดเคลื่อนจากการวัดให้น้อยที่สุด ส่วนการนำไปประยุกต์ใช้จะเลือกใช้แบบใดนั้น คงจะต้องพิจารณาถึงข้อตกลงเบื้องต้น ข้อดี-ข้อจำกัดของโมเดลให้ดีกว่าก่อน เพื่อจะได้โมเดลที่เหมาะสมกับข้อมูล และสรุปผลได้อย่างมั่นใจ

ฟังก์ชันสารสนเทศ (Information function)

ทฤษฎีการตอบข้อคำถามได้กล่าวถึงฟังก์ชันสารสนเทศไว้ โดยจำแนกออกเป็นฟังก์ชันสารสนเทศของคะแนน (Score information function) ฟังก์ชันสารสนเทศของข้อคำถาม (Item information function) และฟังก์ชันสารสนเทศของแบบทดสอบ (Test information function) ซึ่งแต่ละเรื่องมีรายละเอียดดังนี้

1. ฟังก์ชันสารสนเทศของคะแนน

เมื่อเรานำคะแนนผลการสอบของผู้สอบแต่ละคนที่มีความสามารถต่าง ๆ กันมาแจกแจงความถี่และสร้างเป็นแผนภูมิ ก็จะพบว่า คะแนนในแต่ละระดับความสามารถ ส่วนใหญ่จะตกอยู่ในช่วง 95 % ของการแจกแจงในแต่ละระดับความสามารถ ดังภาพประกอบ 14



ภาพประกอบ 14 ช่วงความเชื่อมั่น (Confidence interval; $\theta, \bar{\theta}$) ของความสามารถที่ได้ จากการประมาณค่า (θ)

เนื่องจาก x เป็นคะแนนที่เป็นจำนวนนับซึ่งไม่ต่อเนื่อง จึงทำให้ภาพที่ได้เป็นรูปขั้นบันได จากภาพประกอบ 14 กล่าวได้ว่า ผู้สอบที่ได้คะแนน x ต้องมีความสามารถอยู่ในช่วง $(\theta, \bar{\theta})$ ด้วย ความเชื่อมั่น 95 %

ถ้าให้ $z=x/n$ ดังนั้น การถดถอยของ z ลงบน θ จะได้

$$\mu_{z|\theta} = (1/n) \sum_{i=1}^n P_i(\theta) = \xi$$

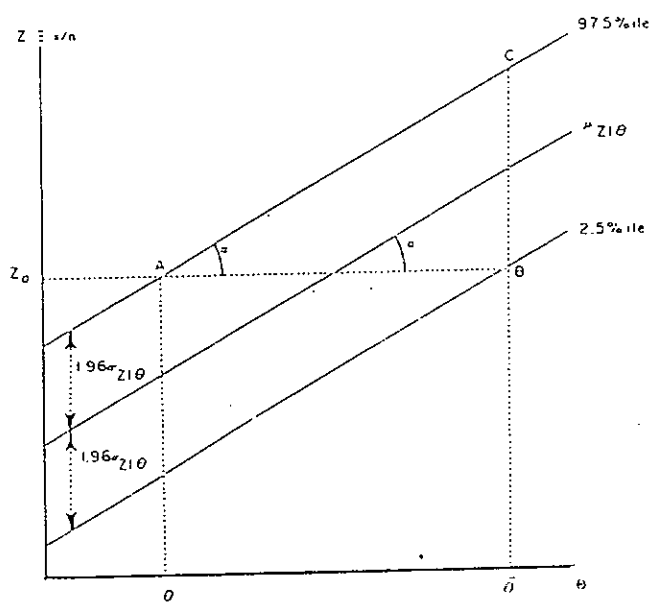
ซึ่งถึงแม้ว่า n จะมากขึ้นก็ไม่ทำให้การถดถอยนี้เปลี่ยนไป ความแปรปรวนของ z สำหรับ ที่คงที่ θ ก็คือ

$$\sigma_{z|\theta}^2 = (1/n^2) \sum_{i=1}^n P_i Q_i = (1/n) \overline{(P_i Q_i)} - \sigma_{p|\theta}^2$$

ซึ่งเมื่อ n เพิ่มขึ้น $\sigma_{z|\theta}^2$ ก็จะเข้าใกล้ 0

เมื่อนำค่า z มาเขียนเป็นแผนภูมิของการแจกแจง ลักษณะของโค้งการแจกแจงก็จะ เหมือนกับภาพประกอบ 14 เมื่อลากเส้นถดถอยของ z ลงบน ซึ่งเส้นถดถอยนี้จะอยู่ระหว่างกลาง ฟังก์ชันแบบขั้นบันได 2 ฟังก์ชัน ซึ่งอยู่ที่ตำแหน่งเปอร์เซนไทล์ที่ 2.5 และ 97.5 ของการแจกแจง z เมื่อกำหนด ให้ เมื่อ n มีค่าสูงขึ้น การถดถอยก็จะไม่เปลี่ยนแปลง แต่ความแปรปรวนจะลดลงเข้า หา 0 เมื่อฟังก์ชันแบบขั้นบันไดลู่เข้าหากัน จนกระทั่งการแจกแจงของ z เป็นโค้งปกติ ตำแหน่งเปอร์เซนไทล์ที่ 2.5 และ 97.5 จะอยู่ห่างเส้นถดถอยประมาณ $1.96 z | \theta$

หรือ $(1.96/n) \sqrt{\sum_{i=1}^n P_i Q_i}$ ดังภาพประกอบ 15



ภาพประกอบ 15 ช่วงแห่งความเชื่อมั่นอะซิมโทติก 95% ($\underline{\theta}$, $\bar{\theta}$) สำหรับความสามารถ θ
(Lord, 1980 : 67)

$$\text{จากภาพ } \tan \alpha = \overline{CB}/\overline{AB} = 2(1.96) \sigma_z | \theta / \overline{AB}$$

เมื่อ $\overline{AB}$ แทน ช่วงแห่งความเชื่อมั่นแอสิมโทติก (Asymptotic confidence interval, $\underline{\theta}$, $\bar{\theta}$)
สำหรับความสามารถ θ

$\overline{CB}$ แทน ค่าเบี่ยงเบนมาตรฐานของ Z

$$\overline{AB} = \overline{CB} / \tan \alpha = 3.92 \sigma_z | \theta \tan \alpha$$

$$\text{ดังนั้น } \frac{1}{\overline{AB}^2} \equiv \frac{\left(\frac{d}{d\theta} \mu_{z|\theta} \right)^2}{(3.92)^2 \text{Var}(z|\theta)}$$

ซึ่ง $1/(\overline{AB})^2$ แทน ฟังก์ชันสารสนเทศของคะแนน (Score information function, $I(\cdot, y)$)

ซึ่งเบอร์นบอม (Birnbaum) ได้นิยามฟังก์ชันแสดงสารสนเทศของคะแนน y ใด ๆ ดังนี้

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

ค่าสารสนเทศของคะแนนนี้จะแปรเปลี่ยนไปตามค่าของ θ ซึ่งสรุปได้ว่า

1. ที่ความสามารถเดียวกัน เส้นถดถอยที่มีความชันมากกว่าจะให้ค่าสารสนเทศของคะแนนสูงกว่าเส้นถดถอยที่มีความชันน้อยกว่า

2. ค่าความคลาดเคลื่อนมาตรฐานของการวัดที่ต่ำจะให้ค่าสารสนเทศของคะแนนสูง

2. ฟังก์ชันสารสนเทศของแบบทดสอบ

ตัวประมาณค่าที่เป็นไปได้สูงสุด (Maximum likelihood, θ) ของพารามิเตอร์ในสถานะปกติ จะมีการแจกแจงในลักษณะเอซิมโทติก มีค่าเฉลี่ย 0 และความแปรปรวนเท่ากับ

$$\text{Var}(\hat{\theta}|\theta_{\infty}) = \frac{1}{\mathcal{E}\left[\left(\frac{d \ln L}{d\theta}\right)_{\theta_{\infty}}^2\right]}$$

เมื่อ L เป็นฟังก์ชันความเป็นไปได้ (Likelihood function)

ลักษณะของฟังก์ชันสารสนเทศของแบบทดสอบ สามารถสรุปได้ดังนี้

1) ฟังก์ชันสารสนเทศของแบบทดสอบเป็นสิ่งที่กำหนดขึ้นสำหรับชุดของข้อคำถามที่แต่ละจุดของสเกลความสามารถ

2) ผลรวมของสารสนเทศเป็นผลมาจากคุณภาพและจำนวนของข้อคำถาม

3) ที่ความสามารถเดียวกัน เส้นถดถอยที่มีความชันมากกว่าจะให้ค่าสารสนเทศของแบบทดสอบสูงกว่าเส้นถดถอยที่มีความชันน้อยกว่า

4) ค่าความแปรปรวนของข้อคำถามที่ต่ำจะให้ค่าสารสนเทศของแบบทดสอบสูง

5) ค่าสารสนเทศของแบบทดสอบไม่ขึ้นอยู่กับการจัดหมู่เฉพาะของข้อคำถามการสนับสนุนของข้อคำถามแต่ละข้อก็เป็นอิสระจากกัน

6) ค่าสารสนเทศของแบบทดสอบมีความสัมพันธ์แบบผกผันกับค่าความคลาดเคลื่อนของการประมาณค่าความสามารถที่จุดเดียวกัน นั่นคือ

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

เมื่อ SEE แทน ค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถ

3. ฟังก์ชันสารสนเทศของข้อคำถาม (Item Information Function) ข้อคำถามแต่ละข้อจะมีฟังก์ชันแสดงสารสนเทศของข้อคำถาม ซึ่งได้แก่

$$I(\theta, u) = (P_i)^2 / P_i Q_i$$

ฟังก์ชันสารสนเทศของข้อคำถามแต่ละข้อจะรวมกันเป็นฟังก์ชันสารสนเทศของแบบทดสอบ ซึ่งค่าฟังก์ชันสารสนเทศของข้อคำถามจะขึ้นอยู่กับค่าความชันของฟังก์ชันการตอบข้อคำถาม (Item response function) และความแปรปรวนที่มีเงื่อนไขในแต่ละระดับของ ค่าความชันมาก ๆ และค่าความแปรปรวนต่ำ ๆ จะทำให้ค่าสารสนเทศของข้อคำถามมีค่าสูง ซึ่งก็จะทำให้ค่าความคลาดเคลื่อนมาตรฐานในการวัดมีค่าต่ำ

การแจกแจงของฟังก์ชันสารสนเทศของข้อคำถามมีลักษณะเป็นรูประฆังคว่ำ (Bell shape) ค่าสารสนเทศที่สูงที่สุดจะอยู่ที่จุด b_i บนสเกลความสามารถที่จุด $\theta_{\max}$ เมื่อ

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

ค่าสูงสุดของสารสนเทศของข้อคำถามจะคงที่ในโมเดล 1 พารามิเตอร์ ขณะที่โมเดล 2 พารามิเตอร์นั้น ค่าสูงสุดจะเป็นสัดส่วนโดยตรงกับกำลังสองของค่าอำนาจจำแนกถ้าอำนาจจำแนกสูงก็จะทำให้ค่าสารสนเทศของข้อคำถามมีค่ามาก ส่วนโมเดล 3 พารามิเตอร์นั้น ค่าสารสนเทศสูงสุดจะมีค่าดังนี้ (Hambleton and Swaminathan, 1985 : 105 ; citing Lord, 1980a : 152)

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

ถ้าค่า c_i ลดลง ค่าสารสนเทศของข้อคำถามก็จะเพิ่มขึ้น และเนื่องจากค่าพารามิเตอร์ของข้อคำถามจะมีผลต่อการเปลี่ยนแปลงค่าสารสนเทศของข้อคำถาม ดังนั้นเราจึงสามารถกำหนดแบบทดสอบที่มีวัตถุประสงค์เฉพาะได้

ประสิทธิภาพสัมพัทธ์ (Relative efficiency)

หากต้องการเปรียบเทียบแบบทดสอบ 2 ฉบับที่วัดคุณลักษณะเดียวกันว่าฉบับใดจะมีประสิทธิภาพสูงกว่ากัน ก็สามารถทำได้โดยใช้ค่าประสิทธิภาพสัมพัทธ์ ซึ่งค่าประสิทธิภาพสัมพัทธ์

ของคะแนนจากแบบทดสอบ y กับคะแนนจากแบบทดสอบ x ก็เป็นอัตราส่วนของฟังก์ชันสารสนเทศของคะแนน ซึ่งสามารถเขียนให้อยู่ในรูปของการแจกแจงคะแนนจริงได้ดังนี้

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

คะแนน x และ y อาจเป็นคะแนนจากแบบทดสอบ 2 ฉบับที่ต่างกัน แต่วัดความสามารถเดียวกัน หรืออาจเป็นคะแนนจากแบบทดสอบฉบับเดียวกัน แต่ให้คะแนนด้วยวิธีที่ต่างกัน ค่าประสิทธิภาพสัมพัทธ์ของคะแนน 2 ชุดนี้ จะแปรเปลี่ยนไปในแต่ละระดับความสามารถ

ประสิทธิภาพสัมพัทธ์สามารถเขียนให้อยู่ในรูปของการแจกแจงคะแนนจริงได้ดังนี้

$$RE\{y,x\} = \frac{V(Y)}{V(X)}$$

องอาจ นัยพัฒน์.2535:44-45 ; Lord. 1980 ; 44-103 ; Hambleton and Swaminathan. 1985 : 33-147 ; Crocker and Algina. 1986 : 340-371 ; Andrich and Masters in Keeves. 1988 : 297-302 ; Douglas in Keeves. 1988 : 282-285 ; Lord and Stocking in Keeves. 1988 : 269-271 ; Masters in Keeves. 1988 : 292-296 ; Wright in Keeves 1988 ; 286-291 ; Hambleton in Linn. 1989 : 147-200 ; Hambleton and Others. 1991 : 32-91 ; Kline. 1993 : 62-71 ; Nunnally and Birnstien. 1994 : 393-443 ; Oosterhof. 1994 : 208-211.

ความสัมพันธ์ระหว่าง IRT กับ CTT

1. พารามิเตอร์จาก IRT และค่าสถิติของข้อสอบตามแนว CTT ถ้าให้ ρ_g แสดงค่าความสัมพันธ์ไบนารีระหว่างคะแนนของข้อ g กับคะแนนที่แสดงถึงคุณลักษณะแฝงหนึ่ง ซึ่งก็คือดัชนีค่าอำนาจจำแนกในการวิเคราะห์ข้อสอบแบบ CTT และถ้ามีการกระจายเป็นโค้งปกติที่มีค่าเฉลี่ยเป็น 0 และความเบี่ยงเบนมาตรฐานเป็น 1 ดังนั้น

$$a_g = \frac{\rho_g}{\sqrt{1-\rho_g^2}}$$

เมื่อ a_g คือ ค่าอำนาจจำแนกตามทฤษฎี IRT หรือค่าความชันของ ICC และแสดงความสัมพันธ์กับ g ได้ดังตาราง 6

ตาราง 6 ความสัมพันธ์ระหว่าง a_g และ p_g

a_g	p_g
.1	.10
.2	.20
.3	.31
.4	.44
.5	.58
.6	.75
.7	.98
.8	1.33
.9	2.06

จากตาราง 6 พบว่า ถ้า p_g เข้าใกล้ 1 ค่า a_g จะเพิ่มมากขึ้น และภายใต้เงื่อนไขของโค้งปกติ เช่นเดียวกัน จะได้ว่า

$$b_g = \frac{\Phi^{-1}(p_g)}{p_g}$$

เมื่อ p_g คือ ค่าความยากของข้อสอบ g ตามแนวของ CTT และ $\Phi^{-1}(p_g)$ คือ คะแนน z ณ จุดตัดของพื้นที่ p_g ทางด้านซ้ายของ z

2. โค้งคุณลักษณะของแบบทดสอบ (Test characteristic curves) และความเชื่อมั่นของแบบทดสอบ

คะแนนจริงของแบบทดสอบมีความสัมพันธ์กับ θ ที่นิยามจาก ICC เขียนแทนด้วยสูตรคณิตศาสตร์ดังนี้

$$T = \frac{\sum P_x(\theta)}{x}$$

จากสมการไม่ได้คำนึงถึงเกี่ยวกับรูปแบบของ ICC อย่างจริงจัง และแสดงให้เห็นว่าความสัมพันธ์ระหว่าง T และ θ ไม่ใช่ลักษณะทางสถิติ ซึ่งที่จริงแล้วการแปลงคะแนนคุณลักษณะแฝงเป็นคะแนนจริงจะเป็นไปในลักษณะที่ไม่ใช้เส้นตรงดังข้อมูลในตารางที่ ซึ่งแสดงความสัมพันธ์ระหว่าง θ และ T ในแต่ละโมเดลของ IRT

ตาราง 7 ความสัมพันธ์ระหว่างคุณลักษณะแฝงและคะแนนจริง

θ	$P_1(\theta)$	$P_2(\theta)$	$P_3(\theta)$	คะแนนจริง
-2	.022	.105	.382	.51
-1	.158	.226	.460	.84
0	.500	.401	.539	1.44
1	.841	.636	.617	2.09
2	.977	.773	.691	2.44

$$a_1 = 1, \quad b_1 = 0; \quad a_2 = .5, \quad b_2 = .5; \quad a_3 = .2, \quad b_3 = -.5$$

จากตาราง แสดงความสัมพันธ์เชิงสถิติระหว่างคะแนนสอบ x และคุณลักษณะแฝง (θ) โค้งถดถอย (Regression curve) สำหรับพยากรณ์ x จาก θ เรียกว่า โค้งแสดงคุณลักษณะของแบบทดสอบ (Test characteristic curve) ซึ่งโค้งถดถอยนี้มีลักษณะเช่นเดียวกับโค้งความสัมพันธ์ของ T ไปยัง θ ดังนั้น สมการประมาณค่า x คือ

$$x = \frac{\sum P_k(\theta) + E}{k}$$

เมื่อ E คือ ค่าความคลาดเคลื่อนในการวัด และเนื่องจากความสัมพันธ์ระหว่าง x และ ไม่ใช่ลักษณะของเส้นตรง ค่าสหสัมพันธ์ที่ไม่ใช่เส้นตรงที่เหมาะสม คือ

$$\eta_x - \theta = \frac{1 - \Sigma \sigma^2_{(x|\theta)}}{\sqrt{\sigma^2_x}}$$

ในการวัดระดับความสัมพันธ์ระหว่าง x และ θ นั้น ค่า $\sigma^2_{(x|\theta)}$ คือ ความแปรปรวนของ x สำหรับประชากรย่อยที่มีความเป็นเอกพันธ์ ซึ่งสามารถแสดงให้เห็นว่าความแปรปรวนนี้มีค่าเท่ากับ $b^2_{(E|\theta)}$ หรือความแปรปรวนของความคลาดเคลื่อนของประชากรย่อย ดังนั้นค่า $b^2_{(x|\theta)}$ จึงมีค่ากับ $b^2_{(E|\theta)}$ ซึ่งชี้ให้เห็นค่าเฉลี่ยความแปรปรวนความคลาดเคลื่อนทุก ๆ ค่าของ θ นอกจากนั้น ยังแสดงได้อีกว่า $b^2_{(x|\theta)}$ เท่ากับความแปรปรวนของความคลาดเคลื่อนที่นิยามตามทฤษฎี CTT ดังนั้น

$$\eta_x - \theta = \sqrt{1 - \frac{\Sigma \sigma^2_E}{\sigma^2_x}}$$

เมื่อพิจารณาพจน์ทางด้านขวาของสมการก็คือค่าดัชนีความเชื่อมั่น (Index of reliability) ซึ่งอาจกล่าวได้ว่าความสัมพันธ์ระหว่าง η_x เท่ากับดัชนีความเชื่อมั่นสำหรับคะแนนที่ได้จาก x (Crocker and Algina. 1986 : 350-352)

มโนทัศน์เกี่ยวกับกระบวนการพัฒนาแบบทดสอบเลือกตอบ

แบบทดสอบเลือกตอบเป็นแบบทดสอบที่ให้ผู้สอบเลือกคำตอบจากตัวเลือกที่กำหนดให้ โครงสร้างของแบบทดสอบเลือกตอบประกอบด้วยข้อความ 2 ส่วนคือ

1. ส่วนที่เป็นคำถามหรือคำสั่ง หรือข้อความที่ไม่สมบูรณ์
2. ส่วนที่เป็นตัวเลือกเพื่อให้นักเรียนเลือกตอบ ที่พบส่วนใหญ่อยู่ในระหว่าง 3-6 ตัวเลือก ซึ่งตัวเลือกเหล่านี้แสดงถึงความเป็นไปได้ที่จะตอบคำถามหรือคำสั่งหรือทำให้ข้อความนั้นสมบูรณ์ถูกต้อง โดยปกติแล้วจะมีเพียงหนึ่งตัวเลือกเท่านั้นที่เป็นคำตอบถูก ส่วนตัวเลือกที่เหลือไว้สำหรับนักเรียนที่ไม่มีความรู้หรือมีความรู้เพียงบางส่วนเลือกตอบ ซึ่งเรียกว่า ตัวลวง (Distractor)

ข้อได้เปรียบและข้อจำกัดของแบบทดสอบเลือกตอบ

แบบทดสอบเลือกตอบมีข้อได้เปรียบที่สำคัญ 4 ประการคือ

1. สามารถวัดได้ครอบคลุมเนื้อหา เช่นเดียวกับแบบทดสอบแบบตอบสั้น (Short answer) โดยที่แบบทดสอบหนึ่งฉบับสามารถบรรจุข้อคำถาม จำนวนมากได้ซึ่งตรงข้ามกับแบบทดสอบแบบความเรียง (Essay items) และจุดเด่นนี้ทำให้แบบทดสอบทั้งสองประเภทสามารถที่จะวัดเนื้อหาได้มาก อย่างไรก็ตาม แบบทดสอบเลือกตอบสามารถที่จะวัดทักษะทางการรู้คิด (Cognitive skills) ได้สูงกว่าแบบทดสอบแบบตอบสั้นเช่นเดียวกับแบบทดสอบแบบความเรียง และข้อได้เปรียบที่สำคัญของแบบทดสอบเลือกตอบก็คือเป็นการประเมินความสามารถของผู้สอบ โดยมีความเกี่ยวข้องในด้านการใช้ภาษาโดยเฉพาะทักษะการเขียนน้อยที่สุด
2. โน้มนำให้ผู้สอบตอบปัญหาได้ตรงประเด็นกับคำถาม ในการใช้แบบทดสอบเลือกตอบบ่อยครั้งที่คำตอบช่วยขยายคำถามที่จะวัดให้ชัดเจนมากยิ่งขึ้น ซึ่งต่างจากแบบทดสอบแบบความเรียงหรือแบบตอบสั้นที่ผู้ตอบอาจตอบได้ไม่ตรงประเด็นกับคำถาม
3. มีระบบการให้คะแนนที่มีประสิทธิภาพ เนื่องจากการตอบข้อสอบแบบเลือกตอบนี้ผู้สอบจะต้องกาเครื่องหมายได้เพียงเครื่องหมายเดียว ดังนั้น จึงทำให้การให้คะแนนเป็นไปอย่างรวดเร็ว โดยเฉพาะสามารถนำไปใช้กับเครื่องตรวจข้อสอบหรือคอมพิวเตอร์
4. คะแนนที่ได้มีความเป็นปรนัย กล่าวคือ เมื่อมีการเฉลยคำตอบไว้แล้วใครนำไปตรวจก็จะได้คะแนนเท่ากัน

สำหรับข้อจำกัดของแบบทดสอบเลือกตอบที่สำคัญมีดังนี้

1. เปิดโอกาสให้ผู้สอบสามารถตอบถูกโดยการเดา
2. การวัดทักษะการรู้คิดในขั้นสูงเป็นไปโดยอ้อม
3. ใช้เวลาในการสร้างค่อนข้างมาก

กระบวนการพัฒนาแบบทดสอบเลือกตอบ

กรีน (Greene, 1952:42-58) เสนอคุณลักษณะของเครื่องมือวัดที่ดีควรใช้เวลาน้อย ประหยัดค่าใช้จ่าย และสะดวกในการจัดการ รวมทั้งการให้คะแนนและการนำเสนอที่เชื่อถือได้ สำหรับในเชิงปฏิบัติแบบทดสอบที่ดีควรมีคุณลักษณะดังนี้ คือ

1. ข้อสอบมีความยากเหมาะสม
2. การให้คะแนนต้องมีความเป็นปรนัย (Objectively scores) (ซึ่งควรใช้แบบทดสอบที่มี

รูปแบบเป็นปรนัย ได้แก่ แบบทดสอบเลือกตอบหรือแบบทดสอบเติมคำ) และมีการตรวจคะแนนที่รวดเร็วและมีประสิทธิภาพ (เหมาะสำหรับข้อสอบถูก-ผิด, แบบทดสอบเลือกตอบหรือจับคู่)

3. มีเกณฑ์มาตรฐาน (Norms) สำหรับเปรียบเทียบอายุ เพศ และสถานภาพที่เข้าใจง่าย และมีความเที่ยงตรงเชิงพยากรณ์ความสำเร็จ โดยมีสัมประสิทธิ์ความสัมพันธ์ตั้งแต่ .75 ขึ้นไป (กลุ่มตัวอย่างในการหาความสัมพันธ์ควรมีจำนวนตั้งแต่ 300 คนขึ้นไป) สำหรับค่าสหสัมพันธ์แต่ละระดับ กริเนได้แสดงความคิดเห็นดังนี้

- .95 ถึง 1.00 แสดงถึงความเชื่อถือได้ของแบบทดสอบในการพยากรณ์ได้ดีมาก
- .75 ถึง .95 พยากรณ์เป็นรายบุคคลได้ดีและพยากรณ์เป็นกลุ่มได้ดีที่สุด แต่อาจเกิดความผันแปรในบางครั้ง
- .50 ถึง .75 ค่าสัมประสิทธิ์ไม่สูงพอที่จะทำนายเป็นรายบุคคลได้ดีนัก เพราะว่ามีคนจำนวนมากได้คะแนนต่ำกว่าค่าเฉลี่ย แต่อาจนำไปใช้ เพื่อศึกษาแนวโน้มของกลุ่มได้
- .25 ถึง .50 ค่าสัมประสิทธิ์ต่ำเกินไปที่จะใช้พยากรณ์เป็นรายบุคคล แต่อาจใช้พิจารณากลุ่มได้อย่างหยاب ๆ
- 0 ถึง .25 เป็นค่าที่ต่างจาก 0 อย่างไม่มีนัยสำคัญ

นอกจากนี้ ควรมีความเชื่อมั่น ความเป็นหน่วยเดียวกัน (Uniqueness) ที่แสดงรูปแบบของการแสดงพฤติกรรมอย่างเดียวกัน ซึ่งอาจตรวจสอบได้โดยการวิเคราะห์องค์ประกอบหรือการสุ่มตัวอย่างข้อสอบ 2 ชุดมาศึกษาความสัมพันธ์ หรือศึกษาความสอดคล้องภายในของข้อสอบ

4. สะดวกในเชิงบริหารดำเนินงาน

- 4.1 ใช้เวลาน้อย
- 4.2 ใช้คำแนะนำน้อย
- 4.3 คำสั่งชัดเจนเข้าใจง่าย
- 4.4 ใช้วัสดุน้อยที่สุด
- 4.5 มีข้อสอบสมมูล 2 ชุด หรือมากกว่า
- 4.6 สถานการณ์ในแบบทดสอบสร้างความสนใจผู้สอบ

ครอกเกอร์และแอลจินา (Crocker and Algina. 1986:66-83, 311-318)เสนอกระบวนการสร้างแบบทดสอบและวิเคราะห์ข้อสอบเพื่อพัฒนาแบบทดสอบดังนี้

กระบวนการสร้างแบบทดสอบดำเนินตามขั้นตอนดังนี้

1. กำหนดเป้าประสงค์เบื้องต้นในการนำคะแนนสอบไปใช้

2. กำหนดพฤติกรรมที่แสดงถึงโครงสร้างหรือนิยามโดเมนโดยวิธีการต่าง ๆ เช่น วิเคราะห์เนื้อหา ศึกษางานวิจัย สังเกต ให้ผู้เชี่ยวชาญกำหนดจุดประสงค์การเรียนรู้การสอนหรือการใช้วิจารณ์จากเหตุการณ์ต่าง ๆ

3. จัดทำตารางลักษณะเฉพาะของแบบทดสอบ กำหนดสัดส่วนของข้อสอบที่ควรเน้นในแต่ละพฤติกรรมที่นิยามไว้ในขั้นตอนที่ 2

4. กำหนดโครงสร้างคลังข้อสอบ

5. ทบทวนข้อสอบ (และปรับปรุงเมื่อจำเป็น)

6. ทดลองใช้ข้อสอบ (และปรับปรุงเมื่อจำเป็น)

7. ทดสอบจริงกับกลุ่มผู้สอบขนาดใหญ่ที่เป็นตัวแทนของประชากรที่ต้องถูกทดสอบด้วยแบบทดสอบ

8. แสดงค่าสถิติที่เหมาะสมของคะแนนสอบ เช่น ความยาก อำนาจจำแนก และขจัดข้อสอบที่ไม่เป็นไปตามเกณฑ์ที่กำหนดไว้

9. ประมาณค่าความเชื่อมั่นและความเที่ยงตรงของแบบทดสอบ หลังจากคัดเลือกข้อสอบที่ดีเข้าฉบับ

10. พัฒนาคู่มือการดำเนินการสอบ การให้คะแนน และการนำเสนอผลการสอบ (เช่น การจัดทำตารางมาตรฐาน (Norm) การแสดงจุดตัดหรือมาตรฐานในการแสดงพฤติกรรม)

การวิเคราะห์ข้อสอบเพื่อการพัฒนาแบบทดสอบ

ในการสร้างข้อสอบตามแนวของ CTT เป้าหมายโดยรวมก็เพื่อต้องการให้แบบทดสอบมีความยาวนานที่สุด แต่ให้ผลคะแนนที่นำไปสู่ระดับของความเชื่อมั่นและความเที่ยงตรงตามความประสงค์ของการนำไปใช้ ในการสร้างแบบทดสอบโดยปกติแล้วก่อนที่จะได้แบบทดสอบฉบับหนึ่งจะผ่านกระบวนการวิเคราะห์ข้อสอบ ซึ่งแสดงถึงค่าเชิงสถิติของผลการตอบข้อสอบแต่ละข้อที่เรียกว่า ค่าพารามิเตอร์ข้อสอบ ได้แก่

1. ค่าความยาก ค่าเฉลี่ย และค่าความแปรปรวน

2. ค่าอำนาจจำแนก ซึ่งมี 2 ชนิด คือ ดัชนีการจำแนก (Index of discrimination) และดัชนีความสัมพันธ์ของค่าอำนาจจำแนกข้อสอบ (Correlational indices of item discrimination) ที่เน้นความสัมพันธ์ระหว่างผลการตอบข้อสอบรายข้อกับคะแนนรวมหรือผลการตอบข้อสอบรายข้อของข้อสอบภายในฉบับ

3. ดัชนีความเชื่อมั่นและความเที่ยงตรงรายข้อ

สำหรับการพัฒนาแบบทดสอบตามแนวของ IRT นั้น เนื่องจากเป้าประสงค์ของ IRT นั้น ต้องการทราบว่าผู้สอบที่มีระดับความสามารถต่างกันในคุณลักษณะใดลักษณะหนึ่งจะมีลักษณะการตอบข้อสอบต่างกันอย่างไร ดังนั้น จึงเน้นที่ผลของการวิเคราะห์ข้อสอบจากกลุ่มมากกว่านำข้อมูลของกลุ่มไปใช้วิเคราะห์ข้อสอบ ด้วยเหตุผลดังกล่าว กระบวนการพัฒนาข้อสอบตามแนว IRT จึงขึ้นอยู่กับโมเดลสำคัญ 3 ประการ คือ

1. คุณลักษณะแฝงและโค้งแสดงคุณลักษณะของข้อสอบ (ICC) ที่แสดงพารามิเตอร์ของข้อสอบ ได้แก่ ความยากข้อสอบ (b) ค่าอำนาจจำแนกข้อสอบ (a) และโอกาสการเดาข้อสอบ (c)

2. ความเป็นมิติเดียว และความเป็นอิสระของข้อสอบ เพื่อแสดงการวัดคุณลักษณะใดลักษณะหนึ่งเพียงลักษณะเดียว และการตอบข้อสอบแต่ละข้อในแบบทดสอบจะไม่ขึ้นแก่กัน ค่าสถิติของข้อสอบจะแสดงความเป็นอิสระต่อกันในแต่ละกลุ่มประชากรย่อยของผู้สอบ ซึ่งเป็นสมมติฐานที่คาดว่ามีความเป็นเอกพันธ์กันในคุณลักษณะนั้น ๆ

3. การวัดที่เป็นอิสระจากแบบทดสอบ (Test-free measurement) เป็นจุดสำคัญของ IRT ที่ยอมรับเกี่ยวกับการเปรียบเทียบความสามารถของผู้สอบในกรณีที่ไม่ได้ทำข้อสอบข้อเดียวกัน

มิลล์แมนและกรีน (Millman and Greene in Linn. 1989:335-364) เสนอแนวคิดเกี่ยวกับลักษณะเฉพาะและการพัฒนาแบบทดสอบวัดผลสัมฤทธิ์และความสามารถ ดังนี้

1. กำหนดเป้าหมายในการทดสอบ

1.1 จำแนกเชิงอธิบายเกี่ยวกับเป้าหมายของการทดสอบ

1.1.1 ข้อสรุปเกี่ยวกับโดเมนของหลักสูตร

1.1.2 ข้อสรุปเกี่ยวกับโดเมนด้านการรู้คิด

1.1.3 ข้อสรุปเกี่ยวกับการกำหนดเกณฑ์ในอนาคต

1.2 กำหนดหัวข้อสำคัญในการจำแนกเป้าหมายของการทดสอบ

1.3 แสดงตัวอย่างของเป้าหมายของการทดสอบในการพัฒนาแบบทดสอบ

2. กำหนดลักษณะเฉพาะของแบบทดสอบ

2.1 กำหนดองค์ประกอบที่เป็นบริบทภายนอก ได้แก่

2.1.1 คุณลักษณะของประชากรผู้สอบ

2.1.2 เวลาที่ใช้ในการทดสอบ

2.1.3 พิจารณาการบริหารการทดสอบเป็นกลุ่มหรือรายบุคคล

2.1.4 มาตรฐานในการทดสอบ

2.2 คุณลักษณะภายในของแบบทดสอบ

2.2.1 เนื้อหาของแบบทดสอบ ควรพิจารณาถึงสิ่งต่อไปนี้

- แหล่งที่มาของเนื้อหาแบบทดสอบ
- มิติเนื้อหาของแบบทดสอบ
- การนำเสนอเพื่ออ้างอิงโดเมนหรือกลุ่ม
- ปริมาณของเนื้อหากับความแม่นยำหรือความเชื่อมั่นในการนำเสนอ

คะแนนสอบ

- การกระจายของเนื้อหา

2.2.2 กำหนดรูปแบบของข้อสอบ

2.2.3 พิจารณาคุณลักษณะทางสถิติของข้อสอบ ได้แก่

- ช่วงของความยากข้อสอบ
- ช่วงของอำนาจจำแนกของข้อสอบ
- พิจารณาแบบแผนการตอบข้อสอบ

2.2.4 กำหนดเกณฑ์และวิธีการให้คะแนน

- ให้คะแนน 0-1
- ให้คะแนนแบบความรู้เป็นบางส่วน
- กำหนดน้ำหนัก โดยอาจพิจารณาความแตกต่างของจำนวนจัดสรรในแต่ละ

ละองค์ประกอบของเนื้อหา หรือกำหนดน้ำหนักในการให้คะแนนที่เป็นการปรับความแตกต่างของพารามิเตอร์ข้อสอบ หรือแบบทดสอบย่อยหรือทั้งสอง ซึ่งจะเน้นถึงความเที่ยงตรงเชิงเนื้อหา

2.2.5 กำหนดจำนวนของข้อสอบ

2.3 การพิจารณาลักษณะเฉพาะของแบบทดสอบ

2.3.1 วางแผนในการสร้างแบบทดสอบที่เฉพาะเจาะจง

2.3.2 วางแผนเกี่ยวกับความเที่ยงตรงโดยใช้ผู้เชี่ยวชาญและผู้ใช้แบบทดสอบ

2.3.3 พิจารณาบททวนเกี่ยวกับแผนการสร้างแบบทดสอบ

3. พัฒนาข้อสอบโดยพิจารณาในสิ่งต่อไปนี้ คือ

3.1 ชนิดของข้อสอบ

3.1.1 คุณลักษณะของรูปแบบ (เขียนตอบในกระดาษหรือใช้คอมพิวเตอร์)

3.1.2 ความสำคัญของรูปแบบ

3.2 เขียนข้อสอบให้พิจารณาสิ่งต่อไปนี้

3.2.1 กำหนดเรื่องในการเขียนข้อสอบ

3.2.2 กฎในการเขียนข้อสอบ

4. ประเมินข้อสอบ

4.1 ประเมินเนื้อหาข้อสอบ

4.1.1 ประสิทธิภาพและการสื่อความของข้อสอบ

4.1.2 ความเหมาะสมของข้อสอบ

4.1.3 ความสอดคล้องกันของข้อสอบกับสิ่งแวดล้อม ซึ่งเน้นความเที่ยงตรงเชิง

หลักสูตรและความเที่ยงตรงในการเรียนการสอน

4.2 วิธีการตอบข้อสอบและการทดลองวิธีดำเนินงาน

4.2.1 เตรียมทดลองใช้ข้อสอบ

4.2.2 กำหนดข้อสอบที่จะใช้ทดลองในรูปแบบของแบบทดสอบ ซึ่งมีวิธีวิธีที่

สำคัญ 3 ประการ คือ

- ใช้ข้อสอบทดลองเป็นแบบทดสอบเชิงปฏิบัติการ
- ใช้ข้อสอบทดลองเป็นส่วนหนึ่งของแบบทดสอบเชิงปฏิบัติการ
- แยกประเมินข้อสอบที่ใช้ทดลอง

4.2.3 จำนวนผู้สอบที่เหมาะสมประมาณ 100-200 คน ซึ่งพอจะประมาณค่าระดับความยากที่เชื่อถือได้ แต่กลุ่มผู้สอบที่มากกว่า 2,000 คน จะทำให้ประมาณค่าผู้สอบในกลุ่มต่ำที่อยู่ปลายล่างสุด (Asymptote) ของโมเดล 3 พารามิเตอร์ได้ดี

4.2.4 กำหนดจำนวนข้อสอบ

4.3 วิธีการตอบข้อสอบและดัชนีทางสถิติ

4.3.1 แนวความคิดทั่วไปในการวิเคราะห์ข้อสอบ

- ตามแนว CTT จะคำนวณค่าความยาก อำนาจจำแนกและดัชนีทางสถิติอื่น ๆ ของกลุ่มตัวอย่างที่ตอบข้อสอบ

- ตามแนว IRT เริ่มต้นจากการคำนวณค่าพารามิเตอร์ของ ICC ซึ่งเป็นการประมาณค่าความน่าจะเป็นในการตอบข้อสอบถูกต้อง ซึ่งเป็นฟังก์ชันที่แสดงสถานะภาพของผู้สอบบนโครงสร้างพื้นฐานที่กำลังสอบวัด

4.3.2 ดัชนีความยาก

- แบบ 0-1 ค่า p หมายถึง สัดส่วนของคนตอบถูกจากคนทั้งหมด

- แบบคะแนนหลายค่า ดัชนีความยากพิจารณาจากคะแนนเฉลี่ยรายข้อ หรือค่าแปรปรวนบางอย่าง เช่น อัตราส่วนระหว่างคะแนนเฉลี่ยรายข้อกับคะแนนสูงสุดในข้อนั้น ฯ (University of Iowa. 1984)

4.3.3 ดัชนีการจำแนก

4.3.4 ดัชนีพื้นฐานเกี่ยวกับแบบแผนการตอบข้อสอบ

-วิธีการที่ขึ้นอยู่กับการแบบแผนของดัชนีความยาก

-วิธีการที่ขึ้นอยู่กับการแบบแผนความสัมพันธ์ของข้อสอบ

เมื่อชุดข้อสอบได้รับการตั้งสมมติฐานอ้างอิงมิติเดียว ผู้ที่ตอบข้อสอบข้อใดข้อหนึ่งผิดก็ควรแสดงแนวโน้มในการตอบข้ออื่นผิด ในทำนองเดียวกัน เมื่อตอบข้อสอบข้อใดข้อหนึ่งถูกก็ควรจะมีแนวโน้มในการตอบข้ออื่นถูก

4.3.5 การวิเคราะห์ตัวลง

5. การคัดเลือกข้อสอบดำเนินการดังนี้

5.1 กำหนดสถานการณ์การคัดเลือก

5.2 กำหนดเกณฑ์การประเมินข้อสอบและข้อสรุปแบบทดสอบ

5.2.1 กำหนดวิธีการคัดเลือก

5.2.2 โดเมนของหลักสูตร

5.2.3 โดเมนด้านการรู้คิด

5.2.4 กำหนดเกณฑ์ในอนาคต

6. การจัดข้อสอบเข้าฉบับพิจารณาในสิ่งต่อไปนี้

6.1 ลำดับข้อสอบ

6.2 โครงร่างของข้อสอบ

6.3 จัดทำสำเนา

6.4 การรักษาความลับของแบบทดสอบ

แองเจอริโลไซ (Cangelosi. 1990:41-74) แสดงแนวคิดเชิงระบบในการสร้างแบบทดสอบ

7 กระบวนการ ได้แก่

กระบวนการที่ 1 กำหนดเป้าหมายการเรียนรู้ให้ชัดเจน

1. กำหนดชุดของจุดประสงค์

2. กำหนดความสำคัญเชิงสัมพันธ์ของจุดประสงค์

3. กำหนดน้ำหนักจุดประสงค์

4. จัดทำตารางแสดงคุณลักษณะเฉพาะ

กระบวนการที่ 2 พัฒนารายกำหนดลักษณะแบบทดสอบ (Test blue-print)

1. ความซับซ้อนของการออกแบบแบบทดสอบ

- จะให้นักเรียนสอบอย่างไร
- ธรรมชาติของเป้าหมายการเรียนรู้

2. บริหารเวลาทดสอบ

3. กำหนดเวลาตรวจให้คะแนน

4. คัดเลือกรูปแบบของแบบทดสอบ

5. จำนวนข้อสอบ

6. ความยากข้อสอบ

7. ประเมินค่าจำนวนต่ำสุดของคะแนนแบบทดสอบ

8. จำนวนของคะแนนในแต่ละจุดประสงค์

9. วิธีการกำหนดคะแนนจุดตัด (สำหรับแบบทดสอบอิงเกณฑ์)

10. โครงร่างแบบทดสอบ

กระบวนการที่ 3 จัดทำคลังข้อสอบ

1. กำหนดลักษณะของคลังข้อสอบ

- องค์ประกอบข้อสอบ
- รูปแบบข้อสอบที่หลากหลาย
- ความยากข้อสอบ (ข้อสอบง่าย $p > .75$, ปานกลาง $.25 < p < .75$, ยาก $p < .25$)

2. รูปแบบการจัดทำคลังข้อสอบ เช่น แฟ้ม หรือระบบคอมพิวเตอร์

กระบวนการที่ 4 สังเคราะห์แบบทดสอบ

1. ปฏิสัมพันธ์ระหว่างข้อสอบ

2. ลำดับขั้นของข้อสอบ

3. คำชี้แจงแก่ผู้สอบ

กระบวนการที่ 5 การบริหารแบบทดสอบ

1. ตัวลวงให้น้อยสุด

2. ดำเนินตามคำชี้แจง

3. ติดตามผู้เรียน

กระบวนการที่ 6 การให้คะแนนแบบทดสอบ

1. เกี่ยวกับการใช้ประโยชน์
 2. เกี่ยวกับความเชื่อมั่นของผู้ให้คะแนนภายนอกและผู้ให้คะแนนภายใน
- กระบวนการที่ 7 กำหนดคะแนนจุดตัด
1. สำหรับแบบทดสอบอิงเกณฑ์
 2. สำหรับแบบทดสอบอิงกลุ่ม

ไคลน์ (Kline, 1993:134-157) เสนอแนวทางในการสร้างแบบทดสอบตามแนว CTT 2 แนวทาง คือ การวิเคราะห์องค์ประกอบ และการวิเคราะห์ข้อสอบ ซึ่งแต่ละแนวทางดำเนินการดังนี้1. วิเคราะห์องค์ประกอบ วิธีนี้มีเป้าหมายเพื่อที่จะสร้างแบบทดสอบที่วัดเพียงหนึ่งองค์ประกอบ ซึ่งมีวิธีดำเนินงานตามขั้นตอนดังนี้

1.1 กำหนดจำนวนข้อสอบ

1.2 กำหนดกลุ่มตัวอย่างผู้สอบอย่างต่ำไม่ควรต่ำกว่า 100 หรือใกล้เคียงในอัตราส่วน 3:1 ระหว่างผู้สอบกับจำนวนองค์ประกอบ นอกจากนั้น กลุ่มตัวอย่างต้องแสดงความเป็นตัวแทนของประชากรผู้สอบ

1.3 หาความสัมพันธ์ระหว่างข้อสอบ ในกรณีแบบทดสอบให้คะแนน 0-1 อาจประมาณค่าความสัมพันธ์ โดยใช้ค่า Φ หรือเตตระคอร์ริค นอกจากนั้น อาจพิจารณาจากค่าดัชนี G หรือความแปรปรวนต่ำในเมตริกซ์ของ Φ การพิจารณากลุ่มข้อสอบและรวมทั้งการพิจารณามิติข้อสอบที่แสดงความเป็นเอกพันธ์ขององค์ประกอบ (Factored homogeneous item dimension; FHIDS) สำหรับในกรณีข้อสอบที่ให้คะแนนหลายค่าจะพิจารณาจากเมตริกซ์ ความสัมพันธ์แบบโปรดักโมเมนต์ระหว่างข้อสอบหรือค่า Φ ที่ประมาณได้จากกลุ่มตัวอย่างอย่างน้อย 200 คน

1.4 วิเคราะห์องค์ประกอบจากเมตริกซ์สหสัมพันธ์ โดยพิจารณาองค์ประกอบที่สำคัญเลือกหมุนแกนองค์ประกอบที่สำคัญ เลือกวิธีหมุนแกน พิจารณาค่าความยากตั้งแต่ .20-.80

1.5 คัดเลือกข้อสอบหลังจากวิเคราะห์องค์ประกอบ โดยพิจารณาจากค่าน้ำหนักองค์ประกอบที่มากกว่า .30 ขึ้นไป ในองค์ประกอบเดียวเท่านั้น

1.6 ปัญหาเชิงปฏิบัติในการสร้างแบบทดสอบ โดยการวิเคราะห์องค์ประกอบได้แก่ ความแตกต่างระหว่างเพศ องค์ประกอบมีความแปรปรวนน้อย และมีค่าน้ำหนักองค์ประกอบน้อย ค่าน้ำหนักองค์ประกอบของข้อสอบจะแสดงค่ามากกว่า 1 องค์ประกอบ การเลือกข้อสอบจึงต้องพิจารณาจากชุดข้อสอบที่มีค่าน้ำหนักองค์ประกอบสูง นอกจากนี้ ปัญหาอีกประการหนึ่งก็คือ ความยากในการหมุนแกน

- 1.7 การจำลององค์ประกอบโดยการวิเคราะห์ซ้ำก่อนจัดชุดข้อสอบครั้งสุดท้าย
- 1.8 หาค่าความเที่ยงตรงของแบบทดสอบ
- 1.9 หาค่าความเชื่อมั่นของแบบทดสอบ
- 1.10 จัดทำแบบทดสอบมาตรฐาน
2. การสร้างแบบทดสอบโดยการวิเคราะห์แบบทดสอบ
 - 2.1 สร้างข้อสอบ
 - 2.2 กำหนดจำนวนกลุ่มตัวอย่าง เพื่อทดลองข้อสอบ ประมาณ 100 คน
 - 2.3 พิจารณาความแตกต่างระหว่างเพศ โดยแยกวิเคราะห์ข้อสอบในแต่ละเพศข้อสอบที่ดีจะต้องไม่แสดงความแตกต่างของค่าสถิติในทั้งสองเพศ
 - 2.4 วิเคราะห์ค่าสถิติข้อสอบ ได้แก่ ค่าความยาก ควรอยู่ระหว่าง .20-.80 ค่าความสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนรวม (ค่าอำนาจจำแนก) หรือค่าความสัมพันธ์ภายในระหว่างข้อสอบ
 - 2.5 การคัดเลือกข้อสอบ พิจารณาจากค่าความสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนรวมตั้งแต่ .30 ขึ้นไป และค่าความยาก .20-.80 ในการแยกวิเคราะห์ทั้ง 2 เพศ
 - 2.6 ตรวจสอบค่าที่ค้นพบในกรณีที่ข้อสอบไม่มีความเชื่อมั่น โดยทำการวิเคราะห์กับกลุ่มตัวอย่างใหม่อีก 2 กลุ่ม
 - 2.7 ประมาณค่าความเที่ยงตรงและจัดทำแบบทดสอบมาตรฐาน
- การสร้างแบบทดสอบตามโมเดลของราชค
 1. เขียนข้อสอบ
 2. กำหนดกลุ่มตัวอย่างอย่างน้อยประมาณ 1,000 คน
 3. พิจารณาวิเคราะห์แยกตามเพศชายและหญิง
 4. แบ่งครั้งกลุ่มตัวอย่างเป็นคะแนนสูงและคะแนนต่ำ
 5. คำนวณค่าพารามิเตอร์ของราชค
 6. คัดเลือกข้อสอบโดยพิจารณาจากข้อสอบที่แสดงคุณลักษณะทั้งสองกลุ่มของข้อสอบ
 7. ปรับปรุงข้อสอบที่บกพร่องและทดลองซ้ำเป็นครั้งที่สอง
 8. การวัดความเป็นอิสระของข้อสอบ
 - 8.1 แบ่งข้อสอบออกเป็น 2 ฉบับ ฉบับที่หนึ่งเป็นข้อสอบที่ง่าย และอีกฉบับเป็นข้อสอบยาก พิจารณาผู้สอบควรได้รับคะแนนที่ต่างกัน

8.2 คำนวณค่า t พิจารณาจากสภาพภาพของผู้สอบในคุณลักษณะใดลักษณะหนึ่ง และความคลาดเคลื่อนมาตรฐานของผู้สอบแต่ละคนในการทดสอบด้วยแบบทดสอบ 2 ฉบับ ถ้าข้อสอบมีความเหมาะสม (fit) กับโมเดล ผู้สอบแต่ละคนต้องได้รับคะแนนเท่ากันทั้ง 2 ฉบับ ถ้าไม่เป็นเช่นนั้นข้อสอบต้องถูกตัดออกหรือเขียนใหม่

8.3 การศึกษาความเที่ยงตรงข้ามกลุ่ม

8.4 การศึกษาความแตกต่างระหว่างเพศ

8.5 การศึกษาเรื่องการเดา

8.6 มีค่าพารามิเตอร์อยู่ระหว่าง -4 กับ +4

ออสเตอร์ฮอฟ (Oosterhof, 1994:186-188) เสนอขั้นตอนในการพัฒนาแบบทดสอบในชั้นเรียนไว้ 6 ขั้นตอนดังนี้

1. กำหนดว่าจะนำผลการทดสอบไปใช้อย่างไร
2. กำหนดชนิดและจำนวนของทักษะที่ต้องการวัดโดยใช้แบบทดสอบ
3. กำหนดชนิดและจำนวนของข้อสอบที่จะนำไปใช้ในการทำแบบทดสอบ
4. กำหนดจำนวนของข้อสอบที่จะวัดในแต่ละจุดประสงค์หรือวางแผนโดยใช้ตาราง

กำหนดเฉพาะ (Table of specifications) ที่แสดงความสัมพันธ์ระหว่างเนื้อหากับจุดประสงค์ หรือเนื้อหากับพฤติกรรมเชิงโครงสร้างที่ต้องการวัด

5. จัดเตรียมข้อทดสอบที่พึงประสงค์จากตารางกำหนดเฉพาะที่วัดในแต่ละทักษะรวมทั้งการประเมินแบบทดสอบ เพื่อให้ได้ข้อสอบตามจำนวนที่ต้องการ

6. จัดข้อสอบเข้าฉบับ

กฎเกณฑ์ในการเขียนข้อสอบเลือกตอบ

1. มีประสิทธิภาพและความเที่ยงตรง

1.1 ข้อสอบหนึ่งข้อต้องมีคำตอบถูกหรือคำตอบที่ดีที่สุดเพียงข้อเดียว เว้นไว้แต่มีคำสั่งเป็นอย่างอื่น

1.2 การสร้างข้อสอบควรทำตามเกณฑ์ประเมินที่ได้จากตารางกำหนดลักษณะเฉพาะของข้อสอบ (Table specification) หรือนำหนักความสำคัญในแต่ละจุดประสงค์/เนื้อหา

1.3 แบบทดสอบที่วัดความรู้ความสามารถควรเขียนคำถามดังนี้

- หลีกเลี่ยงการถามในสิ่งที่ไม่สำคัญหรือไม่มีสาระ และรวมทั้งคำถาม ที่เป็นข้อเท็จจริงหรือมโนทัศน์ที่รู้ ๆ กันอยู่แล้ว

- ควรเขียนคำถามที่วัดระดับความสามารถทางความคิดมากกว่าคำถามที่เน้น

ระลึกจำ

- 1.4 ระดับความยากเหมาะสมกับผู้สอบ
- 1.5 หลีกเลี่ยงข้อสอบที่ถามความคิดเห็น
- 1.6 หลีกเลี่ยงการใช้ตัวเลือก "ไม่มีข้อใดถูก" และ "ถูกหมดทุกข้อ"
- 1.7 หลีกเลี่ยงการใช้แบบทดสอบเลือกตอบถ้ามีแบบทดสอบรูปแบบอื่นวัดได้ เหมาะ

สมกว่า

2. มีการสื่อความที่ดี

2.1 ใช้คำถามที่ตรงไปตรงมา หรือใช้คำถามที่เป็นข้อความที่ไม่สมบูรณ์เพื่อให้ผู้สอบ
หาคำตอบที่เหมาะสมที่ทำให้ข้อความนั้นสมบูรณ์

2.2 เขียนคำถามโดยใช้ภาษาและคำศัพท์ง่าย ๆ ชัดเจนเท่าที่เป็นไปได้

2.3 คำถามต้องเน้นในสิ่งที่จะถามวัดอย่างชัดเจนและสมบูรณ์

2.4 แต่ละข้อต้องถามเพียงเรื่องเดียว และตัวเลือกทุกตัวต้องมีความเป็นเอกพันธ์ใน

เชิงของเนื้อหา

2.5 ข้อคำถามและตัวเลือกต้องมีความสอดคล้องกันอย่างชัดเจน รวมทั้งหลีกเลี่ยงการ
ใช้ข้อความที่ซ้ำกันในตัวเลือก ควรนำมาเขียนในตัวคำถามดีกว่า

2.6 ในกรณีที่ข้อคำถามเป็นปฏิเสธควรเน้นให้เห็นอย่างชัดเจน และหลีกเลี่ยงที่จะนำ
ไปใช้กับผู้สอบที่เป็นเด็กเล็ก

2.7 ตัวเลือกต้องอยู่ท้ายคำถามเสมอ ไม่ควรวางตัวเลือตรงกลางของข้อคำถาม

2.8 จัดเรียงตัวเลือกตามลำดับเหตุผล เช่น จำนวนตัวเลขที่มีค่าน้อยไปหามากหรือ
เรียงจากข้อความสั้นไปหาข้อความยาว

2.9 การเขียนคำถามแต่ละข้อควรเป็นอิสระจากกัน

2.10 ในกรณีที่ข้อคำถามวัดเกี่ยวกับความเข้าใจควรใส่สาระที่คาดว่าเป็นประโยชน์ต่อผู้
สอบเท่าที่เป็นไปได้ แต่ไม่ควรใช้ข้อความที่เคยใช้ในการเรียนการสอนมาแล้ว

3. ประสิทธิภาพของตัวเลือก

3.1 ตัวเลือกทุกตัวเลือกมีความเป็นไปได้ สามารถดึงดูดใจผู้สอบที่ขาดความรู้ความ
สามารถ และเกี่ยวข้องกับประสบการณ์การเรียนรู้ของผู้สอบ

3.2 หลีกเลี่ยงความเกี่ยวข้องระหว่างข้อคำถามกับตัวเลือกโดยไม่ได้ตั้งใจ ซึ่งอาจ
เกิดจาก

- ความสอดคล้องเกี่ยวกับการใช้ภาษาระหว่างข้อคำถามกับตัวเลือก
- การใช้คำซ้ำซ้อนในข้อคำถามและตัวเลือกถูก
- การระลึกจำหรือความสัมพันธ์ทางภาษาด้านอื่นระหว่างคำหลักในข้อคำถาม

และตัวเลือกถูก

3.3 หลีกเลี่ยงการชี้นำซึ่งเกิดจากตัวเลือก

- ตัวเลือกถูกยาวผิดปกติ
- ตัวเลือกถูกบ่งบอกถึงระดับของคุณสมบัติบางประการ หรือใช้คำว่า 'ไม่เคย' และ

'สม่ำเสมอ' ในข้อความที่เป็นตัวลวง

- ตัวเลือกว่าความเป็นอิสระจากกัน มีความหมายครอบคลุมซึ่งกันและกัน
- คำตอบถูกไม่กระจายให้อยู่ในทุกตัวเลือกในจำนวนที่เท่า ๆ กัน

3.4 ระวังตัวเลือกว่าชี้นำการตอบข้ออื่น ๆ

3.5 หลีกเลี่ยงการเขียนตัวเลือกที่มีความหมายตรงข้ามกับตัวเลือกอื่นในข้อเดียวกัน

3.6 ตัวเลือกถูกและตัวลวงไม่ควรแสดงเหตุผลที่ผิด

หลักเกณฑ์ในการประเมินคุณภาพของแบบทดสอบเลือกตอบ

1. ข้อสอบวัดทักษะเฉพาะเจาะจงตามตารางกำหนดลักษณะข้อสอบหรือไม่
2. ผู้เชี่ยวชาญทางด้านเนื้อหาส่วนใหญ่หรือทั้งหมดยอมรับคำตอบถูกหรือไม่
3. การใช้คำศัพท์หรือประโยคเหมาะสมกับระดับของผู้สอบหรือไม่
4. ข้อความที่เป็นคำถามมีความชัดเจนหรือไม่
5. ถ้าข้อสอบต้องการวัดทักษะอื่นมากกว่าการระลึกจำ ประโยคคำถาม หรือสถานการณ์ปัญหานั้น เป็นสิ่งที่ผู้สอบเคยประสบมาแล้วหรือไม่ ในทางที่เหมาะสมควรเป็นบริบทใหม่สำหรับผู้สอบ

6. ข้อสอบวัดเนื้อหาของผู้สอบหรือไม่

7. เมื่อข้อคำถามหรือตัวเลือกมีความหมายในทางตรงกันข้ามหรือแตกต่างจากที่เข้าใจ

ทั่ว ๆ ไป มีการเน้นความสำคัญคำวิเศษณ์ให้เด่นชัดหรือไม่

8. มีการใช้คำซ้ำในทุกตัวเลือกหรือไม่

9. การใช้ภาษาในตัวเลือกสอดคล้องกับคำถามหรือไม่

10. ตัวเลือกทุกตัวเกี่ยวข้องกับเนื้อหาเดียวกันหรือไม่

11. มีตัวเลือก 'ถูกหมดทุกข้อ' หรือ 'ไม่มีข้อใดถูก' หรือไม่

12. มีการเรียงลำดับตัวเลือกอย่างมีเหตุผลหรือไม่ ตัวถูกควรวางตำแหน่งให้

การจัดกระจาย ไม่ควรวางไว้ที่อักษรใดอักษรหนึ่ง เช่น ก ทั้งหมดหรือ ข ทั้งหมด หรือวางไว้ตรงกลางของกลุ่มตัวเลือก เพราะจะทำให้ผู้สอบจับแบบแผนของการตอบได้ นอกจากนั้นการจัดเรียงลำดับตัวเลือกควรจัดอย่างมีระบบ เช่น ตัวเลือกที่แสดงค่าจำนวนควรจัดเรียงจากต่ำไปสูงหรือถ้าเป็นข้อความควรจัดเรียงจากข้อความสั้นไปจนถึงข้อความที่ยาวที่สุด

13. ตัวลองแต่ละตัวมีความเป็นไปได้หรือไม่

(Congelosi. 1990 : 94-102 ; Kline. 1993 : 209-210 ; Millman and Greene in Linn. 1989 : 353 ; Herman in Husen and Postlethwaite. 1994 : 3062 ; Nunnally and Bernstein. 1994 : 299 ; Oosterhof. 1994 : 127-147 ; Friedenber. 1995 : 69-71)

มโนทัศน์เกี่ยวกับการพัฒนาโครงสร้างของแบบทดสอบเลือกตอบ วิธีการตอบ และการตรวจให้คะแนน

โดยปกติแบบทดสอบชนิดเลือกตอบหลายตัวเลือก (MC) ประกอบไปด้วยส่วนสำคัญ 2 ส่วน คือ ส่วนที่เป็นสิ่งเร้าหรือคำถาม และส่วนที่เป็นตัวเลือก หรือคำตอบ ซึ่งจะมีเพียงตัวเลือกที่เป็นคำตอบถูก และตรวจให้คะแนนแบบ ถูก 1 ผิด 0 เท่านั้น ซึ่งนักวัดผลเห็นว่ามีข้อจำกัดหลาย ๆ ประเด็น โดยเฉพาะการเปิดโอกาสให้ผู้ตอบเดาตอบ และได้คะแนนนับว่าเป็นความคลาดเคลื่อนอย่างมาก จึงได้มีแนวคิดในการปรับวิธีการตอบ การตรวจให้คะแนน และโครงสร้างข้อสอบจากแบบปกติ ซึ่งแบ่งออกเป็นกลุ่มใหญ่ ๆ ได้ 4 กลุ่ม ดังนี้

1. ปรับการตรวจให้คะแนน เป็นวิธีการอย่างง่ายที่สุด เป็นวิธีการปรับคะแนนหลังจากที่ผู้เข้าสอบได้สอบเสร็จสิ้นแล้ว โดยพยายามที่จะขจัดคะแนนที่ได้ เนื่องจากการเดาออกไป นอกเหนือจากวิธีให้คะแนนติดลบสำหรับข้อตอบผิดแล้ว ยังมีการใช้สูตรปรับคะแนน ได้แก่

1.1 สูตรของแอบบอร์ต (Abbot's formular)(Nunnally and Bernstein. 1994 : 340-341 ; citing Finney. 1947)

$$X_c = R - \frac{W}{c - 1}$$

เมื่อ X_c แทน คะแนนที่แก้การเดา
 R แทน จำนวนข้อที่ตอบถูก
 W แทน จำนวนข้อที่ตอบผิด
 C แทน จำนวนตัวเลือกของข้อสอบแต่ละข้อ

ข้อจำกัดของสูตรนี้ ได้รับการวิพากษ์วิจารณ์ว่า การตอบผิดไม่ใช่การเดาเสมอไป แต่อาจเกิดจากการตอบผิดโดยการไม่รู้ หรือได้รับความรู้ที่ผิดก็เป็นได้

1.2 สูตรของกัลลิคเซน (Gulliksen, 1950 ; 250-251)

$$X_u = R + \frac{U}{A}$$

เมื่อ X_u แทน คะแนนที่แก้การเดา

R แทน จำนวนข้อที่ตอบถูก

U แทน จำนวนข้อที่เว้นไว้

A แทน จำนวนตัวเลือกของข้อสอบแต่ละข้อ

วิธีการนี้มีจุดประสงค์เพื่อให้คะแนนดึงดูดใจผู้ตอบไม่ให้เดาตอบ โดยมุ่งหวังว่า ถ้าไม่รู้ให้เว้นว่างไว้ จึงเกิดข้อจำกัดว่า ข้อที่เว้นว่างไว้เป็นเพราะผู้ตอบไม่ทำ เพราะไม่รู้ หรือว่าเพราะทำไม่ทัน

นอกจากวิธีการใช้สูตรแล้ว วิธีที่นิยมวิจัยกันมากที่สุดก็คือ วิธีบอกระดับความมั่นใจในการตอบ (Confidence weight scoring) โดยให้ผู้ตอบบอกถึงความมั่นใจ ซึ่งส่วนมากเป็นมาตรจัดลำดับที่มี 5 สเกล ถ้าตอบถูกก็จะได้คะแนนตามระดับความมั่นใจ แต่ถ้าตอบผิดจะถูกหักคะแนนตามระดับความมั่นใจ (Echternacht, 1972 : 217-234) ข้อจำกัดของวิธีนี้คือ ผู้ตอบที่รับรู้ผิดพลาดซึ่งกำหนดระดับความมั่นใจไว้สูงมาก เมื่อตอบผิดจะถูกหักคะแนนมาก ซึ่งบางครั้งอาจได้คะแนนติดลบ ซึ่งไม่สอดคล้องกับหลักความเป็นจริงของการวัดทางจิตวิทยา

นอกจากนั้นยังมีวิธีการของ เดวิส และฟิฟเฟอร์ (Davis and Fiffer, 1959 : 159-169) ที่กำหนดน้ำหนักคะแนนแก่ตัวเลือกทุกตัวเลือกไม่เท่ากัน โดยใช้ผู้เชี่ยวชาญหลาย ๆ ท่าน เป็นผู้กำหนดน้ำหนัก ซึ่งจะช่วยให้แบบทดสอบมีความเชื่อมั่นสูงขึ้น แต่มีข้อยุ่งยากในการปฏิบัติมาก จึงไม่เหมาะที่จะนำมาใช้ในแบบทดสอบที่ครูสร้างขึ้น

2. ปรับวิธีการตอบ

2.1 วิธีการของคัมบัส (Coombs, 1953 : 308-310) ได้เสนอให้มีการเลือกคำตอบที่ผิดแทนการเลือกคำตอบที่ถูก โดยเชื่อว่าผู้ตอบต้องใช้ส่วนของความรู้ย่อมมาพิจารณามากกว่าการเลือกคำตอบถูกเพียงคำตอบเดียว ความรู้ที่วัดได้จะสมบูรณ์มากกว่าเพราะผู้ตอบรู้ว่าสิ่งใดผิดสิ่งใดถูก วิธีนี้สามารถวัดส่วนของความรู้ได้ละเอียดกว่าวิธีธรรมดา สำหรับการตรวจให้คะแนนให้ดังนี้ ถ้านักเรียนทำเครื่องหมายผิดแก่ตัวเลือกที่ถูกเลือกนั้น จะได้คะแนน $1 - K$ เมื่อ K คือ จำนวนตัวเลือกทั้งหมดในแต่ละข้อ ดังนั้น ถ้ามี 4 ตัวเลือก คะแนนที่ได้จะอยู่ระหว่าง - 3 ถึง + 3 ข้อจำกัดคือ การให้คะแนนติดลบ

2.2 วิธีการของอนันต์ (อนันต์ ศรีโสภณ. 2516 : 13-19) ได้เสนอวิธีการตอบโดยให้ตอบทุกตัวเลือก ทั้งตัวเลือกถูกและผิด ซึ่งจะทำให้เป็นการวัดความรู้ในข้อคำถามนั้น ๆ ได้สมบูรณ์ยิ่งกว่าการเลือกคำตอบถูกเพียงคำตอบเดียว และยังสามารถทราบความรู้ที่ผิดของนักเรียนได้ดีกว่าวิธีตอบแบบธรรมชาติอีกด้วย การให้คะแนนให้เป็นรายตัวเลือก ถ้าทำเครื่องหมายถูกหรือผิดได้ตรงกับเฉลยว่าตัวเลือกนั้นถูกหรือผิด จะได้ตัวเลือกละ 1 คะแนน ถ้าทำเครื่องหมายตรงข้ามกับสภาพความเป็นจริง จะมีคะแนนติดลบ 1 ข้อ จำกัดของวิธีนี้คือ การกาเครื่องหมายผิดไม่ได้เกิดจากการเดาเสมอไปแต่อาจเกิดจากความไม่รู้ และการให้คะแนนติดลบนับว่าเป็นสิ่งที่ไม่เหมาะสม

2.3 วิธีการตอบอย่างเสรี มี 2 แนวคิด คือ

2.3.1 วิธีของเดรสเซล และชมิท (Dressel and Schmid. 1953 : 574-595) เป็นวิธีที่ให้ผู้ตอบเลือกตอบอย่างเสรี ผู้ตอบสามารถเลือกคำตอบเท่าไรก็ได้ ซึ่งมีวิธีการให้คะแนนดังนี้ (กรณีมี 5 คำตอบ)

จำนวนคำตอบที่ ทำเครื่องหมาย (n)	คะแนน		
	ผู้สอบทำเครื่องหมาย ข้อถูก	ผู้สอบไม่ได้ทำเครื่องหมาย ข้อถูก	
1	4	-1	* k คือจำนวน ตัวเลือก ในที่นี้เท่ากับ 5 เท่านั้น
2	3	-2	
3	2	-3	
4	1	-4	
5	0	-5	

ข้อจำกัด คือ กรณีที่ผู้ตอบเลือกทุกตัวเลือกจะต้องมีข้อถูกอยู่ด้วย 1 ข้อ ในกรณีนี้คะแนนลบห้า จึงไม่น่าจะเกิดขึ้น และการให้คะแนนติดลบไม่สอดคล้องกับการวัดทางจิตวิทยา

2.3.2 วิธีของลำราญ (ลำราญ มีแจ้ง, 2525) เป็นวิธีการตอบเหมือนกับของเดรสเซล คือ ให้นักเรียนเลือกจำนวนคำตอบได้อย่างเสรี แต่มีวิธีการให้ คะแนนต่างกัน ถ้าตอบถูกตัวเดียว แล้วถูกจะได้ 4 คะแนน แต่ถ้าตอบผิดจะได้ 0 คะแนนถ้าตอบ 2 ตัวเลือกขึ้นไป ใช้สูตร

$$X = 4 - 4C/(K - 1)$$

เมื่อ C แทน จำนวนตัวเลือกที่ตอบ

K แทน จำนวนตัวเลือกทั้งหมด

X แทน คะแนนสอบ

คะแนนจะกระจายตั้งแต่ 0 ถึง 4 คะแนน ข้อจำกัดการกำหนดสูตรใช้ได้เฉพาะ บางกรณีเท่านั้น ความเป็นจริงแล้วน่าจะใช้ทุกกรณี เช่น กรณีมี 5 ตัวเลือก ถ้าเลือก 1 ตัวเลือก แล้วตอบถูก แทนค่าในสูตรแล้วน่าจะได้เพียง 3 คะแนน เลือก 2 ตัวเลือก ถ้าตอบถูกได้ 2 คะแนน เลือก 3 ตัวเลือก ถ้าตอบถูกได้ 1 คะแนน เลือก 4 ตัวเลือก ถ้าตอบถูกได้ 0 คะแนน แต่ถ้าไม่เลือกเลยค่า $C = 0$ กลับเป็นได้ 4 คะแนน หรือกรณีถ้าเป็น 4 ตัวเลือก คะแนนที่ออกมาจะไม่เป็นคะแนนเต็ม เช่น ถ้าเลือก 2 ตัวเลือก แล้วตอบถูกจะได้คะแนน $4 - 4 \times 2/3$ หรือประมาณ 1.4 คะแนน ซึ่งเป็นข้อยุ่งยากในเชิงปฏิบัติ

3. ปรับโครงสร้างข้อสอบ

3.1 ปรับตัวเลือกให้มีตัวถูกมากกว่า 1 ตัวเลือก หรือไม่มีตัวถูกเลยลักษณะข้อสอบแบบนี้ จัดว่าเป็นแนวคิดเกี่ยวกับการให้คะแนนสำหรับความรู้บางส่วน (Partial knowledge) ซึ่งได้พัฒนามาจากหลักการสร้างแบบทดสอบเลือกตอบที่มีตัวถูกถูกตัวเดียว และแบบทดสอบถูก-ผิด โดยมีส่วนตอนนำของแต่ละข้อเป็นคำถามและมีหลายตัวเลือก ซึ่งแต่ละตัวเลือกจะต้องเป็นข้อความ หรือประโยคที่สามารถตอบได้ว่าถูกหรือผิด แต่ไม่ได้กำหนดว่าจะต้องมีตัวถูกถูกหรือผิดในข้อนั้น ๆ ซึ่งจะเป็นที่ตัวเลือกนั้น ขึ้นอยู่กับความเหมาะสมของเนื้อหา ในคำถามข้อหนึ่ง ๆ อาจประกอบด้วยตัวเลือกที่เป็นตัวถูกทั้งหมด ผิดทั้งหมด ถูกและผิดเท่า ๆ กัน หรือถูกผิดไม่เท่ากันก็ได้ ซึ่งหลักการเขียนตัวเลือกก็ใช้หลักการเดียวกับแบบทดสอบเลือกตอบแบบปกติ แบบทดสอบลักษณะนี้ น่าจะช่วยลดปัญหาความคลาดเคลื่อนในการประมาณค่าความสามารถที่แท้จริงของผู้สอบ นอกจากนั้นน่าจะบอกได้ถึงความรู้ความสามารถของผู้ตอบ ทั้งระดับที่รู้จักจริง รู้บางส่วน และรู้ไม่สมบูรณ์กว่า แบบถูก 1 ผิด 0 ซึ่งแบบทดสอบลักษณะนี้ มี 2 ลักษณะคือ

1. แบบตัวเลือกไม่ต่อเนื่อง (Multiple True-False Items) แบบทดสอบลักษณะนี้ ตัวเลือกทุกตัวเลือกจะต้องมีอิสระจากกัน

2. แบบตัวเลือกต่อเนื่อง (Sequential True-False Items) แบบทดสอบลักษณะนี้ ตัวเลือกทุกตัวเลือกจะแสดงความเกี่ยวข้องต่อเนื่องกันเป็นลำดับขั้น ซึ่งเหมาะสำหรับการวินิจฉัยอย่างมาก

วิธีการตรวจให้คะแนน มี 3 วิธีใหญ่ ๆ คือ

1. ตอบถูกต้องตามเฉลยทุกตัวเลือกให้ 1 คะแนน กรณีอื่น ๆ ให้ 0 เรียกว่าวิธี Multiple-response method (MR) หรือวิธี MTF-A (Multiple true-false items with absolute knowledge scoring)

2. ให้ 1 คะแนน เมื่อตอบได้ถูกต้องในแต่ละตัวเลือก (Multiple true-false method (MTF) วิธีนี้ได้ 0 เมื่อตอบไม่ถูกต้องเลยแม้แต่ตัวเลือกเดียว ซึ่งคะแนนจะกระจายตั้งแต่ 0 จนถึงคะแนนเต็ม ซึ่งเท่ากับจำนวนตัวเลือก บางครั้งมีผู้ให้คะแนนแบบเศษส่วน เช่น $1/4$ เมื่อตอบถูก 1 ตัวเลือก แต่อาจไม่สะดวกในแง่ของการปฏิบัติ

3. กำหนดสัดส่วนการตอบถูก (Count for correct options method) วิธีนี้กำหนดจำนวนตัวเลือกที่ถูกเพื่อกำหนด 1 คะแนน เช่น ตอบถูกต้องตั้งแต่ 3 ตัวเลือกขึ้นไปให้ 1 คะแนน กรณีอื่น ๆ ให้ 0 (Oosterhof, 1994 : 169-171)

ข้อจำกัดของวิธีนี้ก็คือ การสร้างข้อสอบในบางเนื้อหาอาจทำได้ยากมากและจำนวนข้อไม่ควรมีมากเกินไป ซึ่งผู้สอบต้องใช้เวลาคิดทุกตัวเลือก

3.2 ปรับรูปแบบข้อสอบจากวิธีเขียนตอบมาใช้เครื่องมือบันทึกคำตอบที่ประดิษฐ์ไว้เฉพาะ (เกริกชัย สวบทวีญ. 2525 : 5) วิธีการนี้ใช้ปากกาถูกลิ้นจิ้มลงไปในช่องวงกลมที่มีอักษร ก, ข, ค, ง หรือ จ ของแต่ละตัวเลือก ถ้าเป็นตัวเลือกถูกจะสามารถจิ้มลงไปถึงแผ่นล่างได้ แต่ถ้าไม่ถูกต้องไม่สามารถจิ้มลงถึงแผ่นล่างได้ เมื่อทำถูกต้องให้ทำข้อต่อไปได้เลย ถ้าไม่ถูกต้องนักเรียนสามารถตัดสินใจได้อย่างเสรีว่า จะทำข้อเดิมต่อไปหรือไม่ทำ การตรวจให้คะแนนใช้วิธีดังนี้

จำนวนครั้งที่เลือก คะแนนที่ได้เมื่อตอบถูก คะแนนที่ได้เมื่อตอบผิด

1	4	- 1
2	2	- 2
3	0	- 3
4	- 2	- 4
5	- 4	- 5

สำหรับข้อที่เว้น ได้คะแนน 0

ข้อจำกัดของวิธีนี้ คือ ต้องสร้างเครื่องมือเฉพาะซึ่งไม่ใช่การเขียนตอบในกระดาษแบบปกติ และการให้คะแนนติดลบ นับว่าไม่สอดคล้องกับการวัดทางจิตวิทยา

4. ปรับวิธีการดำเนินการสอบ วิธีนี้มีความคิดพื้นฐานว่า ผู้สอบทุกคนไม่จำเป็นต้องทำข้อสอบจำนวนเท่ากัน การใช้ข้อสอบที่ยาวเกินไปสำหรับผู้มีความสามารถสูง หรือข้อสอบที่ยาวเกินไปสำหรับผู้ที่มีความสามารถต่ำ นับว่าเป็นการเสียเวลา ซึ่งอาจทำให้เกิดความเบื่อหน่ายของผู้

สอบ หรือเกิดปัญหาการเดาสำหรับผู้ที่มีความสามารถต่ำในการที่ทำข้อสอบที่มีความยากเกินความสามารถ ซึ่งจะทำให้เกิดความคลาดเคลื่อนในการประมาณค่าความสามารถที่แท้จริงของผู้สอบ ซึ่งลอร์ดและนอร์ริคได้เสนอความเห็นใน ปี 1968 ว่าผู้สอบควรจะได้รับข้อสอบที่มีความยากเหมาะสมกับความสามารถของเขา ต่อมาจึงได้เกิดแนวคิดของข้อสอบที่เรียกว่า Tailor testing และอีกหลาย ๆ ชื่อ ได้แก่ Adaptive testing, Branched testing, Individualized testing, Response contingent testing, Programmed testing, Sequential item testing ซึ่งต่อมาได้นำมาใช้กับเทคโนโลยีทางคอมพิวเตอร์ ซึ่งเรียกต่อมาว่า Computerized adaptive testing ; CAT หรือ Computerized automate testing ซึ่งแบบทดสอบลักษณะนี้ ผู้ตอบจะต้องทำข้อสอบทีละข้อตามระดับความสามารถที่ประมาณค่าจากโปรแกรมคอมพิวเตอร์ และจะหยุดเมื่อเป็นไปตามเกณฑ์ที่กำหนด ส่วนใหญ่นิยมประมาณค่าความคลาดเคลื่อนมาตรฐาน ในการประมาณค่าความสามารถเข้าใกล้ศูนย์ให้มากที่สุด ข้อจำกัดของแนวคิดนี้จำเป็นต้องสร้างคลังข้อสอบ และต้องประมาณค่าต่าง ๆ ตามแนวของ IRT ซึ่งอาจจะก่อให้เกิดความยุ่งยากในการนำไปปฏิบัติจริงกับแบบทดสอบที่ครูผลิตเอง

มโนทัศน์เกี่ยวกับการศึกษาความลำเอียงของข้อสอบ

การศึกษาความลำเอียงของข้อสอบหรือแบบสอบนั้น เริ่มมาจากการสังเกตปรากฏการณ์ของผลการสอบเพื่อการคัดเลือก (Selection bias) ซึ่งพบว่าไม่เป็นไปตามสัดส่วน ระดับสติปัญญาหรือโครงสร้างของประชากร ไม่ว่าจะเป็นการคัดเลือกเพื่อการศึกษาต่อ เพื่อการบรรจุเข้าทำงาน เพื่อเลื่อนตำแหน่งหรืออื่น ๆ ส่งผลให้เกิดการศึกษาความลำเอียงของแบบสอบและข้อสอบ รายข้อในแบบสอบที่ใช้ (Test item bias) ในระยะต่อ ๆ มา

อย่างไรก็ตาม แม้แนวคิดในการตรวจสอบความยุติธรรมของการใช้ข้อสอบ เพื่อการคัดเลือกบุคคลจะมีมาตั้งแต่ปี 1951 ดังได้กล่าวมาแล้ว แต่เหตุการณ์ที่ทำให้มีการตื่นตัวเป็นอันมากในเรื่องของความลำเอียงของข้อสอบ ได้แก่ กรณีที่เรียกว่า เดอฟูนิส และมอดการ์ด (Defunis and Odegaard) ซึ่งเกิดขึ้นในปี 1971 เป็นกรณีที่มาร์โค เดอฟูนิส (Marco Defunis) และคณะซึ่งถูกปฏิเสธจากโรงเรียนกฎหมายของมหาวิทยาลัยแห่งวอชิงตันได้ฟ้องร้องว่า เขาได้คะแนนการสอบสูงกว่าผู้ได้รับการคัดเลือกบางคนที่ยังมหาวิทยาลัยรับเข้าศึกษา และยื่นฎีกาฟ้องร้อง ชาร์ลส์ มอดการ์ด (Charles Odegaard) อธิการบดีมหาวิทยาลัยดังกล่าวและคณะเป็นจำเลย เพื่อให้พิจารณาทบทวนการคัดเลือกนักศึกษาใหม่และจากนั้นเป็นต้นมา การพิจารณาตรวจสอบความลำเอียงของข้อสอบระหว่างผู้สอบกลุ่มย่อยเป็นสิ่งที่มีการปฏิบัติกันจนปัจจุบันเป็นเสมือนส่วนหนึ่งของขั้นตอนการพัฒนาแบบทดสอบโดยทั่ว ๆ ไป โดยเฉพาะอย่างยิ่งสำหรับการพัฒนาแบบทดสอบมาตรฐาน

ตัวอย่างเช่น ศูนย์บริการการทดสอบทางการศึกษา (Educational testing service) ของประเทศสหรัฐอเมริกา ซึ่งให้บริการด้านการทดสอบด้วยแบบสอบมาตรฐานทั่วประเทศสหรัฐอเมริกา ได้มีการตรวจสอบความลำเอียงทั้งก่อนและหลังจากการนำแบบทดสอบไปทดลองใช้การตรวจสอบความลำเอียงของข้อสอบก่อนนำไปทดลองใช้จะใช้การตรวจสอบโดยผู้เชี่ยวชาญ ซึ่งประกอบด้วยคณะบุคคลหลายฝ่ายตรวจสอบความไว (Sensitivity) ของแบบสอบ โดยพิจารณาถึงรูปแบบของข้อสอบ เนื้อหา ภาพประกอบ คำที่ใช้และอื่น ๆ เพื่อไม่ให้มีความลำเอียงเป็นประโยชน์หรือเสียประโยชน์สำหรับผู้สอบกลุ่มย่อยใด ๆ เช่น ในกลุ่มผู้สอบที่ต่างเพศ ต่างเชื้อชาติ และวัฒนธรรม เป็นต้น

เท่าที่ผ่านมาจนถึงปัจจุบัน มีผู้เสนอวิธีการตรวจสอบความลำเอียงของข้อสอบโดยพิจารณาจากค่าสถิติอยู่หลายวิธีดังนี้

1) การพิจารณาความแตกต่างกันในค่าเฉลี่ยระหว่างกลุ่ม กรณีเช่นนี้อาจไม่ได้แสดงถึงความลำเอียงของข้อสอบหรือแบบสอบ เพราะกลุ่มย่อยทั้งสองกลุ่มอาจมีความแตกต่างกันอยู่แล้วในสิ่งที่วัด

2) การวิเคราะห์ตัวประกอบ เป็นการดูความสัมพันธ์ระหว่างคะแนนย่อย ๆ ในแบบสอบ สำหรับกลุ่มย่อย ๆ โดยมากใช้เทคนิควิเคราะห์แบบ Maximum likelihood ปัญหาอยู่ที่การเปรียบเทียบโครงสร้างระหว่างกลุ่มว่าจะใช้เกณฑ์ได้ดีที่สุด รวมทั้งความซับซ้อนของวิธีการและการวิเคราะห์

3) การเปรียบเทียบค่าความยาก เป็นการพิจารณาลำดับที่สัมพันธ์ของค่าความยากที่ได้จากการวิเคราะห์แยกกลุ่ม เป็นการพิจารณาปฏิสัมพันธ์ระหว่างข้อกระทงและกลุ่มผู้สอบข้อกระทงใดที่มีลักษณะของความสัมพันธ์สำหรับกลุ่มบางกลุ่มแตกต่างไปจากข้ออื่น ๆ ในแบบสอบเดียวกัน จะเป็นข้อสอบที่ถือว่ามีความลำเอียง วิธีการทำได้ทั้งการวิเคราะห์ความแปรปรวนและการใช้แผนภูมิเส้น

ผู้ที่พัฒนาวิธีการนี้ได้สมบูรณ์ที่สุด คือ แองกอฟฟ์ (Angoff) ซึ่งแปลงค่า p ให้เป็นค่า z มาตรฐานเพื่อจัดความสัมพันธ์เชิงเส้นโค้งระหว่างค่า p ของข้อสอบ 2 ชุด แล้วแปลงค่า z เป็นเดลต้า (Δ) มีค่าเฉลี่ย = 13 และค่าส่วนเบี่ยงเบนมาตรฐาน = 4 เพื่อกำจัดค่าที่เป็นลบ แล้วพิจารณาค่าเดลต้าของข้อสอบข้อเดียวกันจากกลุ่มที่ต่างกันโดยใช้แนวเส้นทะแยงที่เป็นเส้นแกนหลัก ข้อสอบจะลำเอียงหรือไม่พิจารณาได้จากระยะทางจากจุดค่าเดลต้าที่ลากไปตั้งฉากกับเส้นแกนหลัก

4) การใช้ทฤษฎีการตอบข้อสอบ เป็นการเปรียบเทียบความน่าจะเป็นในการตอบข้อสอบ ถูกของกลุ่มผู้สอบย่อย โดยพิจารณาถึงค่าพารามิเตอร์ของข้อสอบ และผู้สอบตามแบบจำลองที่เลือกใช้ซึ่งมีทั้งชนิด 1, 2 และ 3 พารามิเตอร์

5) การวิเคราะห์ด้วยค่าไคสแควร์ ซึ่งมีผู้เสนอหลายคน เช่น วิธีของซูเนแมน (Scheueman. 1979) เป็นการวิธีการที่แปลงค่าความสามารถของสมาชิกในกลุ่มเป็นช่วง ๆ อาจเป็น 3 หรือ 5 ช่วง แล้วใส่ข้อมูลในตาราง 2 ทาง (จำนวนกลุ่ม X การตอบถูก-ผิด) จำนวนตารางเท่ากับ จำนวนช่วงคะแนนจากนั้นคำนวณค่าไคสแควร์แต่ละตาราง และเมื่อรวมค่าทุกตาราง จะได้ค่าไคสแควร์ที่มีชั้นแห่งความเป็นอิสระเท่ากับจำนวนกลุ่ม

การจัดกลุ่มตามระดับคะแนนเช่นนี้ ถ้ากลุ่มผู้สอบ 2 กลุ่ม มีคะแนนเฉลี่ยต่างกันแล้วจะเกิดผลที่ดูเป็นความลำเอียงได้

6) เมลเลนเบอร์ก (Mellenbergh. 1982) ได้เสนอเทคนิค Loglinear หรือแบบจำลอง โลกิที่เบี่ยงเบนจากไคสแควร์และได้มีการนำมาขยายโดยการทำซ้ำ หลังจากการจัดข้อสอบที่ลำเอียงออกไปแล้วในแต่ละรอบของการวิเคราะห์

7) ที่แตกแยกออกมาจาก Loglinear ได้แก่ วิธีการแมนเทล-แฮนส์เซล (Mantel-Haenszel) ที่ศึกษาตัวแปรประเภทแบ่งสองในกลุ่มที่แตกต่างกันด้วยตัวแปรบลิค และนำมาใช้เป็นตัวบ่งชี้ความลำเอียงของข้อสอบ

8) ลินน์ และฮาร์นิสซ์ (Linn and Harnisch. 1981) ได้เสนอวิธีวิเคราะห์ในลักษณะทฤษฎีการตอบข้อสอบแบบ 3 พารามิเตอร์ โดยรวมกลุ่มย่อยเข้าด้วยกัน แล้วแบ่งความสามารถเป็นช่วงดูความแตกต่างระหว่างความน่าจะเป็นของการตอบถูกที่คาดหวัง ซึ่งประมาณได้ด้วยทฤษฎีการตอบข้อสอบกับความน่าจะเป็นของการตอบถูกจากข้อมูลเชิงประจักษ์พิจารณาเป็นช่วงความสามารถแล้วนำมารวมกันเป็นดัชนีรวมของความลำเอียงของแต่ละข้อ วิธีนี้เรียกว่าวิธีทฤษฎีการตอบข้อสอบแบบเทียม และใช้ได้ในกรณีที่มีกลุ่มตัวอย่างน้อยทำให้ใช้ทฤษฎีการตอบข้อสอบแบบเต็มรูปไม่ได้

9) วิธีล่าสุดที่เสนอโดยชีลลีและสเตาท์ (Shealey and Stout. 1989) ได้แก่การทดสอบ SIB (หรือ SIBTEST) เป็นวิธีที่ดัดแปลงจากวิธีทฤษฎีการตอบข้อสอบ สามารถใช้ได้ทั้งเพื่อการตรวจสอบความลำเอียงของข้อสอบรายข้อ การตรวจสอบความลำเอียงของกลุ่มข้อสอบหรือแบบสอบทั้งฉบับ (Lord. 1980 : 212-223 ; Hambleton and Swaminathan. 1985 : 281-296 ; Crocker and Algina. 1986 : 267-283, 376-393 ; Citing Scheuneman. 1979, Angoff and Ford. 1973 ; Adams and Rowe in Keeves. 1988 : 398-403 ; Murphy and Davidshofer. 1988 : 214-221 ; Title in Keeves. 1988 :

1994 : 416-417 ; Oosterhof.1994 : 426-427, 434-435 ; Friedenbergh. 1995 : 327-334 : กาญจนา วัฒนสุนทร. 2537 : 18-20 ; อ้างอิงมาจาก Mellenbergh. 1982, Linn and Hornisch. 1981, and Harnisch. 1981, and Shealey and toud. 1989)

งานศึกษาวิจัยที่เกี่ยวข้องกับการพัฒนาแบบทดสอบเลือกตอบ

ในการศึกษางานวิจัยครั้งนี้ ได้ทำการศึกษางานวิจัยเกี่ยวกับการใช้ Partial credit model และที่เกี่ยวกับการปรับแบบทดสอบเลือกตอบปกติในมิติต่าง ๆ ซึ่งจะขอแยกเสนอและเรียงลำดับ ดังนี้คือ

ก. งานวิจัยเกี่ยวกับ Partial credit model

งานวิจัยในต่างประเทศ

ดอดด์และค็อคซ์ (Dodd and Koch. 1987 : 371-384) ใช้วิธีการสร้างสถานการณ์จำลอง (Simulation study) เพื่อศึกษาค่าอินฟอร์เมชันของข้อความ (Item information function) และค่าประสิทธิภาพสัมพัทธ์ (Relative efficiency) ของแบบทดสอบ เพื่อศึกษาความแม่นยำของการวัด (The precision of measurement) ที่เป็นผลมาจากการเปลี่ยนแปลงค่าความยากประจำขั้น (Step difficulties) ของพาเรียล เครดิต โมเดล พบว่าค่าฟังก์ชันอินฟอร์เมชันของข้อความมีความเกี่ยวข้องกับค่าความยากประจำขั้น ตามปัจจัยดังต่อไปนี้ 1) ระยะห่างระหว่างค่าความยากประจำขั้นแรกกับขั้นสุดท้าย 2) จำนวนของค่าความยากประจำแต่ละขั้นที่ไม่เรียงลำดับกัน (เรียงแบบผันกลับ ; Reverse) 3 ขนาด (Magnitude) ของระยะระหว่างค่าความยากประจำแต่ละขั้นที่เรียงผันกลับ ดังมีสาระสำคัญที่พอกกล่าวได้โดยย่อคือ เมื่อข้อความมีจำนวนขั้นและระยะระหว่างค่าความยากประจำขั้นแรกกับขั้นสุดท้ายเท่ากันแล้ว ข้อความที่มีค่าความยากเรียงลำดับกันจากมากไปหาน้อย จะมีค่าฟังก์ชันอินฟอร์เมชันสูงสุด ส่วนค่าความยากที่เรียงแบบอื่น ๆ จะมีค่าต่ำกว่า โดยเฉพาะอย่างยิ่งถ้าค่าความยากประจำขั้นเรียงจากน้อยไปหามาก ค่าฟังก์ชันอินฟอร์เมชันจะมีค่าต่ำสุด และถ้ามีข้อความบางรายการที่ค่าความยากบางขั้นเรียงผันกลับกันบ้าง ข้อความที่มีความยากขั้นแรก (Earlier step) สูงกว่าขั้นท้าย ๆ (Later step) จะมีค่าฟังก์ชันอินฟอร์เมชันสูงกว่าข้อความที่มีค่าความยากประจำขั้นแรกต่ำกว่าขั้นท้าย ซึ่งเป็นลักษณะเช่นเดียวกับข้อความที่มีค่าความยากไม่

เรียงลำดับกันที่ค่าฟังก์ชันอินฟอร์เมชันสูงนอกจากนี้ ข้อความที่มีขนาดของระยะห่างระหว่างค่าประจำชั้นที่เรียงผันกลับกันมากค่าฟังก์ชันอินฟอร์เมชันของข้อความจะมีค่าสูง เช่น ข้อความ A มีค่าความยากประจำชั้นแต่ละชั้นเป็น - 1.0, 0.5, -0.5 และ 1.0 ข้อความ B มีค่าเป็น - 1.0, -0.5, 1.0 และ 0.5 ข้อความทั้งสองรายการนี้ มีค่าความยากประจำชั้นไม่เรียงลำดับกันโดยข้อความ A มีการเรียงผันกลับกันในชั้นที่ 2 และ 3 ส่วนข้อความ B เรียงผันกลับกันในชั้นที่ 3 และ 4 ซึ่งจะเห็นได้ว่าขนาดของระยะระหว่างชั้นที่ผันกลับกันในข้อความ A ค่าเท่ากับ 1 มากกว่าข้อความ B ที่มีค่าเท่ากับ 0.5 ดังนั้นข้อความ A ย่อมมีค่าฟังก์ชันอินฟอร์เมชันสูงกว่าข้อความ B เป็นต้น สำหรับค่าประสิทธิภาพสัมพัทธ์ของแบบทดสอบพบว่าข้อความที่มีพิสัยของค่าความยากแคบ (น้อยกว่า 1.5) จะส่งผลต่อค่าประสิทธิภาพสัมพัทธ์ของแบบทดสอบมากกว่าข้อความที่มีพิสัยของค่าความยากกว้าง (มากกว่า 2.0) เมื่อเทียบจากแบบทดสอบที่ใช้เป็นเกณฑ์ฉบับเดียวกันแล้ว (Base test) นอกจากนี้ยังพบอีกว่าค่าประสิทธิภาพสัมพัทธ์ของแบบทดสอบ ขึ้นอยู่กับการเรียงผันกลับกันของข้อความในแบบทดสอบกล่าวคือแบบทดสอบที่มีข้อความเรียงผันกลับกันจะมีค่าประสิทธิภาพสัมพัทธ์สูงกว่าแบบทดสอบที่ไม่มีข้อความเรียงผันกลับ เมื่อเทียบกับแบบทดสอบฉบับเดียวกัน กล่าวโดยสรุปข้อความที่มีค่าความยากเรียงลำดับกัน และมีพิสัยของค่าความยากเหล่านั้นกว้างจะมีค่าประสิทธิภาพสัมพัทธ์ของแบบทดสอบต่ำที่สุด ข้อความที่มีค่าความยากเรียงแบบผันกลับกัน และมีพิสัยของค่าความยากแคบจะมีประสิทธิภาพสัมพัทธ์ของแบบทดสอบสูงสุด และสอดคล้องกับข้อคิดยังได้ เสนอแนะให้เลือกข้อความที่มีค่าความยากประจำชั้นแรก ๆ ทั้งนี้เพื่อให้การวัดคุณลักษณะของผู้ตอบ (θ) ด้วยการให้ข้อความดังกล่าวเป็นไปด้วยความแม่นยำสูงมากขึ้น และสอดคล้องกับข้อคิด ยังได้เสนอแนะให้เลือกข้อความที่มีค่าความยากประจำชั้นแรก ๆ ทั้งนี้ เพื่อให้การวัดคุณลักษณะของผู้ตอบ (θ) ด้วยการให้ข้อความดังกล่าวเป็นไปด้วยความแม่นยำสูงมากขึ้น

งานวิจัยในประเทศ

องอาจ นัยพัฒน์ (องอาจ นัยพัฒน์, 2535 : บทคัดย่อ) ได้ทำการศึกษาประสิทธิภาพของผลการประมาณค่าพารามิเตอร์ของพาเซียล เครดิต โมเดล ระหว่างวิธี PROX และวิธี UCON ในแบบทดสอบวัดเจตคติแบบลิเคิร์ทสเกล จำนวน 50 รายการ โดยใช้กลุ่ม ตัวอย่าง 500 คน คำนวณค่าพารามิเตอร์เจตคติของผู้ตอบ (θ_n) และค่าความยากประจำชั้นของข้อความ (I_j) โดยใช้โปรแกรม BIGSTEPS และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย และ Test information function (TIF) จากโปรแกรม PCM (Partial credit model) จากการวิจัยพบว่า ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยที่เป็นผลมาจากการประมาณค่าเจตคติของผู้ตอบ ด้วยวิธียูคอนมีค่าต่ำกว่าวิธี

กำลังสองเฉลี่ยที่เป็นผลมาจากการประมาณค่าเจตคติของผู้ตอบ ด้วยวิธียูคอนมีค่าต่ำกว่าวิธีพริกซ์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยที่เป็นผลมาจากการประมาณค่าความยากประจำชั้นของข้อความระหว่างวิธีทั้งสอง มีค่าไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สำหรับความเที่ยงตรงเชิงสภาพของการประมาณค่าพารามิเตอร์เจตคติของผู้ตอบที่ประมาณค่าด้วยวิธีพริกซ์ มีค่าสูงกว่าวิธียูคอนอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนค่า TIF ของแบบทดสอบที่ประมาณค่าพารามิเตอร์ด้วยวิธีพริกซ์สูงกว่าวิธียูคอน ในช่วงของเจตคติระดับปานกลางส่วนในช่วงเจตคติระดับต่ำและสูงพบว่าวิธียูคอนมีค่าสูงกว่า นอกจากนั้นยังพบว่าความสัมพันธ์ระหว่างค่าเจตคติที่เป็นผลมาจากการประมาณค่าพารามิเตอร์ด้วยวิธีพริกซ์และวิธียูคอน มีความสัมพันธ์ในทางบวกอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 และค่าความยากประจำชั้นของข้อความที่ประมาณค่าด้วยวิธีทั้งสองก็พบว่ามีความสัมพันธ์ในทางบวกระดับสูงมากเช่นเดียวกัน

สุรัชย์ มีชาญ (2539 : 79 - 112) ใช้พาร์เทียล เครดิต โมเดล ของมาสเตอร์ ศึกษาเปรียบเทียบความยากประจำชั้น และประสิทธิภาพในการประมาณค่าเจตคติ ของมาตรวัดเจตคติของลิเคิร์ท ที่มีรูปแบบการตอบ จำนวนของลำดับขั้น และทิศทางของการเรียงลำดับขั้นที่แตกต่างกัน โดยจัดชุดแบบทดสอบออกเป็น 8 ชุด ทำการศึกษากับกลุ่มตัวอย่างซึ่งเป็นนักเรียนระดับชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2539 ในจังหวัดลพบุรี และสิงห์บุรีจำนวน 3,226 คน จากการวิจัยพบว่า

1. มาตรวัดเจตคติที่มี 4 ระดับ ที่มีรูปแบบการตอบ และทิศทางของการเรียงลำดับขั้นต่างกัน ค่าเฉลี่ยความยากประจำชั้นที่ 1 และ 3 ของแบบทดสอบทั้ง 4 ฉบับมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วน มาตรวัดเจตคติที่มี 5 ระดับ ที่มีรูปแบบการตอบ และทิศทางของการเรียงลำดับขั้นต่างกัน ค่าเฉลี่ยความยากประจำชั้นของทุกชั้นของแบบทดสอบทั้ง 4 ฉบับมีค่าไม่แตกต่างกัน

2. มาตรวัดเจตคติที่มี 4 ระดับ ที่มีรูปแบบการตอบ และทิศทางของการเรียงลำดับขั้นต่างกัน ค่าความยากประจำชั้นที่ 2 และ 3 ของแบบทดสอบทั้ง 4 ฉบับมีความสัมพันธ์กันในเชิงบวก ส่วนมาตรวัดเจตคติที่มี 5 ระดับ ที่มีรูปแบบการตอบ และทิศทางของการเรียงลำดับขั้นต่างกัน ค่าความยากประจำชั้นที่ 2 , 3 และ 4 ของแบบทดสอบทั้ง 4 ฉบับมีความสัมพันธ์กันในเชิงบวก

3. มาตรวัดเจตคติที่มี 4 ระดับให้ค่าเฉลี่ยของค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยในการประมาณค่าเจตคติของผู้ตอบน้อยกว่ามาตรวัดเจตคติที่มี 5 ระดับ

4. สัดส่วนของผู้ตอบที่เหมาะสมกับโมเดลที่ได้จาก มาตรวัดเจตคติที่มีรูปแบบการตอบ

จำนวนของลำดับชั้น และทิศทางของการเรียงลำดับชั้นต่างกัน มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

งานวิจัยเกี่ยวกับการปรับแบบทดสอบเลือกตอบในมิติต่างๆ

เอกสารและงานวิจัยในต่างประเทศ

เดรสเซล และชมิท (Dressel and Schmid. 1953 : 574-595) ได้ศึกษาวิธีการตอบ และวิธีการตรวจให้คะแนนที่ต่างกัน เพื่อปรับปรุงอำนาจจำแนกของแบบทดสอบเลือกตอบโดยวิธีการ 4 วิธี คือ

1. แบบเลือกตอบเสรี (Free choice test) เป็นแบบทดสอบเลือกตอบที่มี 5 ตัวเลือก โดยผู้เข้าสอบจะเลือกตอบกี่ตัวเลือกก็ได้
2. แบบบอกระดับความมั่นใจในการตอบ (A degree of certainly test) เป็นแบบที่ผู้เข้าสอบจะต้องแสดงระดับความมั่นใจในการตอบ ซึ่งมี 4 ระดับ แต่คำถามแต่ละข้อให้ตอบได้เพียงคำตอบเดียว
3. แบบมีคำตอบถูกมากกว่าหนึ่ง (A multiple-answer test) แบบทดสอบชนิดนี้จะมีคำตอบถูกอยู่หลายคำตอบ ผู้เข้าสอบจะต้องทำเครื่องหมายทุกข้อที่ถูก
4. แบบมีคำตอบถูกสองคำตอบ (A two-answer test) ผู้เข้าสอบจะเลือกตอบได้เพียงสองคำตอบเท่านั้น

ผลการศึกษาพบว่า การตอบข้อสอบของนักเรียนเก่ง (ซึ่งพิจารณาจากคะแนนการสอนจะแตกต่างกับนักเรียนประเภทปานกลาง และอ่อน อย่างมีนัยสำคัญทางสถิติ เมื่อใช้แบบทดสอบเลือกตอบชนิดตอบเสรี คือ กลุ่มเก่งจะทำเครื่องหมายคำตอบน้อยกว่ากลุ่มอ่อน ทั้งในข้อสอบที่มีระดับความยากสูง ปานกลาง และต่ำ ในทางตรงกันข้าม แบบทดสอบชนิดทดสอบความมั่นใจ ไม่ได้จำแนกนักเรียนเก่ง ปานกลาง และอ่อน ได้ดีกว่าวิธีแรก เพราะเมื่อพิจารณาการตอบข้อสอบแล้ว พบว่าผู้เข้าสอบตอบข้อสอบที่มีความยากปานกลาง และยากมากด้วยระดับความมั่นใจอย่างเดียวกัน เดรสเซลและชมิท ได้สรุปว่าองค์ประกอบของความมั่นใจ ที่วัดโดยการให้ตอบแบบเสรีนั้น แตกต่างกับที่วัดโดยให้ผู้เข้าสอบบอกระดับความมั่นใจ

คัมบัส (Coombs. 1953 : 308-310) ได้เสนอวิธีการตอบและตรวจให้คะแนน โดยให้ผู้เลือกตอบเลือกคำตอบที่ผิดแทนคำตอบที่ถูก ซึ่งคัมบัสให้ความเห็นว่า การเลือกคำตอบที่ผิดซึ่งมีอยู่หลายตัวเลือกในข้อสอบข้อหนึ่งๆ นั้น ผู้ตอบต้องใช้ส่วนของความรู้ย่อย (Partial knowledge)

มาพิจารณามากกว่าการเลือกคำตอบถูกเพียงคำตอบเดียว เพราะต้องแยกแยะออกว่าผิดเพราะสาเหตุใด ความรู้ที่วัดได้จึงสมบูรณ์กว่า วิธีนี้จึงสามารถวัดส่วนของความรู้ได้ละเอียดกว่าวิธีตอบแบบปกติ และการตรวจให้คะแนนนั้น ถ้านักเรียนทำเครื่องหมายผิดแก่ตัวเลือกที่ผิดจริง ให้คะแนนตัวเลือกละ 1 คะแนน แต่ถ้านักเรียนทำเครื่องหมายผิดแก่ตัวเลือกที่ถูก ตัวเลือกนั้นจะได้คะแนน 1-k (เมื่อ k คือจำนวนตัวเลือกทั้งหมดในแต่ละข้อ)

คูลมส์ มิลฮอลแลนด์ และโวมเมอร์ (Coombs, Milholland and Womer. 1956 : 13-17)

ได้นำวิธีการตอบและให้คะแนนของคูลมส์มาศึกษาเปรียบเทียบกับวิธีธรรมดา โดยใช้แบบทดสอบ 3 ฉบับๆ ละ 40 ข้อ คือ แบบทดสอบที่เกี่ยวกับคำศัพท์ แบบทดสอบเกี่ยวกับความรู้ของนักขับรถ การใช้สายตา มิติสัมพันธ์ วิธีการตรวจให้คะแนนมี 3 วิธี คือ วิธี 0-1 วิธีของคูลมส์ และวิธี 0-1 วิธีของคูลมส์รวมกัน กลุ่มตัวอย่างเป็นนักศึกษาปีที่ 3 และ 4 ของแจคสันไฮสคูล จำนวน 855 คน แบ่งเป็น 3 กลุ่ม ผลการศึกษาปรากฏว่า

1. ค่าความเชื่อมั่นจากวิธีการตรวจของคูลมส์สูงกว่าวิธี 0-1
2. ค่าความเที่ยงตรงของแบบทดสอบแตกต่างกันอย่างไม่มีนัยสำคัญ
3. วิธีการตรวจของคูลมส์ป้องกันการเดาได้ดีกว่า
4. ทักษะของผู้สอบ ชอบวิธีการตรวจของคูลมส์มากกว่า

เดวิส และฟีฟเฟอร์ (Davis and Fifer. 1959 : 159-169) ได้ศึกษาเกี่ยวกับความเชื่อมั่นและความเที่ยงตรงของแบบทดสอบวัดความถนัดและแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ใช้วิธีกำหนดน้ำหนักของคะแนนแต่ละตัวเลือกไม่เท่ากัน เปรียบเทียบกับวิธีการตรวจให้คะแนนแบบ 0-1 โดยเดวิสและฟีฟเฟอร์คิดว่าการคิดโจทย์คณิตศาสตร์ปัญหานั้น ตัวเลือกแต่ละตัวเลือกมีความยากง่ายไม่เท่ากัน และการใช้สมรรถภาพสมองในการคิดไม่เท่ากัน จึงกำหนดคะแนนของแต่ละตัวเลือกไม่เท่ากัน กลุ่มตัวอย่างที่ใช้เป็นนักเรียนการบินที่ Lackland air force base จำนวน 65 คน วิธีตรวจให้คะแนนมี 2 วิธี คือ วิธี 0-1 กับวิธีการให้น้ำหนักคะแนนทุกตัวเลือก โดยผู้เชี่ยวชาญเป็นผู้กำหนดน้ำหนักคะแนนของตัวเลือกแต่ละตัวเลือก พบว่า

1. การตรวจให้คะแนนโดยกำหนดน้ำหนักคะแนนของแต่ละตัวเลือกไม่เท่ากัน ให้ค่าความเชื่อมั่นสูงกว่าวิธีการตรวจแบบ 0-1
2. การตรวจให้คะแนนโดยกำหนดน้ำหนักคะแนนของแต่ละตัวเลือกไม่เท่ากัน ให้ค่าความเที่ยงตรงสูงกว่าวิธีการตรวจแบบ 0-1

จะเห็นได้ว่า การกำหนดคะแนนของตัวเลือกแต่ละตัวเลือกไม่เท่ากันนี้ เป็นอีกวิธีหนึ่งที่จะช่วยให้คะแนนจากการสอบมีค่าความเชื่อมั่นสูงขึ้น แต่ข้อยุ่งยากของวิธีนี้คือการกำหนดน้ำหนักคะแนนของแต่ละตัวเลือก ข้อสอบต้องสร้างโดยผู้ชำนาญ และต้องใช้ผู้เชี่ยวชาญหลายๆ ท่าน ในการกำหนดน้ำหนักคะแนนของแต่ละตัวเลือก แล้วนำมาหาค่าเฉลี่ยว่าจะได้คะแนนของแต่ละตัวเลือกเป็นเท่าใด จึงมีความยุ่งยากในการปฏิบัติ ไม่เหมาะสำหรับครูที่สร้างข้อสอบเอง

อัลเบิร์ต (Albert. 1970 : 1619A) ได้ศึกษาวิธีการตรวจให้คะแนนความรู้บางส่วน ในวิชาคณิตศาสตร์ โดยใช้โจทย์ปัญหาคณิตศาสตร์ จำนวน 25 ปัญหา ศึกษาแก่นักเรียนเกรด 11 จำนวน 2 กลุ่ม กลุ่มละ 146 คน มีวิธีการตรวจให้คะแนน 2 ลักษณะ คือ

1. แบบทดสอบเลือกตอบที่มีตัวเลือกถูกตัวเดียว มีการให้คะแนน 3 วิธี คือ

1.1 วิธี 0-1

1.2 วิธีกำหนดน้ำหนักตัวเลือก โดยดูจากโอกาสผิดพลาดมากหรือน้อย

1.3 วิธีกำหนดน้ำหนักตัวเลือก โดยครูเป็นผู้กำหนด

2. แบบขยายคำตอบ เป็นแบบที่มีการแก้ปัญหาลายขั้นตอน มีวิธีการตรวจให้คะแนน

2 วิธี คือ

2.1 ให้คะแนนขั้นละ 1 คะแนน ในการตอบที่ถูกต้อง

2.2 ให้คะแนนขั้นละ 1 คะแนน ถ้านักเรียนเริ่มต้นจากข้อมูลที่ถูกต้องถึงคำตอบที่

ถูกต้อง

ผลการศึกษาพบว่า

1. แบบขยายคำตอบสามารถพิสูจน์ความรู้บางส่วนได้

2. แบบขยายคำตอบให้ค่าอำนาจจำแนกสูงกว่าแบบธรรมดา

3. แบบขยายคำตอบให้ค่าความเชื่อมั่นสูงกว่าแบบธรรมดา

4. แบบขยายคำตอบให้ค่าความเที่ยงตรงตามสภาพสูงกว่าแบบธรรมดา และสามารถหาความคลาดเคลื่อนที่เกิดจากการเดาได้

จากการศึกษาจะเห็นได้ว่า แบบขยายคำตอบ ซึ่งเป็นแบบที่ตรวจสอบความรู้บางส่วน ของนักเรียนในการตอบด้วย จะทำให้ค่าอำนาจจำแนกและค่าความเชื่อมั่นสูงขึ้น เปอร์เซนต์การเดาลดลง และยังให้ค่าความเที่ยงตรงตามสภาพสูงด้วย

แพทเนอิก และเทราบ (Patnaic and Traub. 1973 : 281-285) ได้ศึกษาวิธีการตรวจให้คะแนนที่คล้ายกับวิธีของเดวิสและฟิฟเฟอร์ โดยกำหนดคะแนนของตัวเลือกถูกในแต่ละข้อให้

ต่างกัน เปรียบเทียบกับการตอบวิธี 0-1 โดยใช้แบบทดสอบ Domain group test of learning capacity (DLC) จำนวน 60 ข้อ ให้ความเวลาในการทำแบบทดสอบ 20 นาที กลุ่มตัวอย่างเป็นนักเรียนเกรด 10-13 จำนวน 1,054 คน ในการกำหนดน้ำหนักคะแนนของตัวเลือกที่ถูกในแต่ละข้อ ใช้การตัดสินจากครูและผู้เชี่ยวชาญ 61 คน จัดอันดับน้ำหนักคะแนนของตัวเลือกที่ถูกในแต่ละข้อ แล้วหาค่าเฉลี่ยออกมาเป็นคะแนนในแต่ละข้อ ผลการศึกษาพบว่า สอดคล้องกับการศึกษาของเดวิสและฟิฟเฟอร์ คือ ค่าความเชื่อมั่นของแบบทดสอบที่มีวิธีการตรวจให้คะแนนแบบ 0-1 และแบบให้น้ำหนักคะแนนไม่เท่ากัน เมื่อคำนวณค่าความเชื่อมั่นของแบบทดสอบ โดยวิธีแบ่งครึ่งฉบับ ได้สัมประสิทธิ์ความเชื่อมั่นเป็น .881 และ .915 ส่วนค่าความเที่ยงตรงของแบบทดสอบ ซึ่งหาความเที่ยงตรงเชิงพยากรณ์ โดยคำนวณจากค่าสัมประสิทธิ์ระหว่างคะแนนแบบทดสอบ และคะแนนผลสัมฤทธิ์ทางการเรียนเฉลี่ย พบว่า วิธีการตรวจให้คะแนนแบบให้น้ำหนักคะแนนแต่ละข้อไม่เท่ากัน มีความเที่ยงตรงเชิงพยากรณ์น้อยกว่าวิธี 0-1 จะเห็นได้ว่าการกำหนดให้แต่ละตัวเลือกมีคะแนนไม่เท่ากัน โดยกำหนดคะแนนของตัวเลือกถูกไม่เท่ากัน จะทำให้ค่าความเชื่อมั่นของแบบทดสอบสูงขึ้นกว่าวิธีการตรวจให้คะแนนแบบ 0-1 แต่เป็นการยากในการปฏิบัติจริงๆ เพราะต้องใช้ผู้เชี่ยวชาญจำนวนมากในการกำหนดน้ำหนักคะแนนเหล่านี้

ฟิฟ และบรัสซา (Pugh and Brunza. 1975 : 73-78) ได้เสนอวิธีการตอบโดยให้กำหนดความมั่นใจในการตอบ และน้ำหนักความมั่นใจในการตอบและใช้เป็นส่วนหนึ่งในการคิดคะแนนด้วย นอกจากนี้ยังได้ศึกษาค่าความเชื่อมั่นของแบบทดสอบ ที่มีวิธีการตอบและตรวจให้คะแนน 2 วิธี คือ วิธีการตอบแบบธรรมดา และวิธีการตอบโดยให้ผู้ตอบบอกระดับความมั่นใจในการตอบข้อสอบนั้นด้วย โดยใช้แบบทดสอบเลือกตอบวัดด้านสติปัญญา (Intelligence scale : I.E.R.) จำนวน 48 ข้อ แบ่งเป็น 2 ส่วนๆ ละ 24 ข้อ กลุ่มตัวอย่างเป็นนักศึกษามหาวิทยาลัยอินเดียนา (Indiana University) จำนวน 84 คน แบ่งออกเป็น 3 กลุ่มๆ ละ 28 คน แต่ละกลุ่มได้รับแบบทดสอบที่ต่างกัน ดังนี้ กลุ่มแรกได้รับแบบทดสอบชุด A ซึ่งมีวิธีการตอบแบบธรรมดาทั้งสองส่วน กลุ่มที่สองได้รับแบบทดสอบชุด B ซึ่งมีวิธีการตอบแบบธรรมดาในส่วนที่หนึ่ง และมีวิธีการตอบโดยบอกความมั่นใจในการตอบในส่วนที่สอง กลุ่มที่สามได้รับแบบทดสอบชุด C ซึ่งมีวิธีการตอบโดยบอกความมั่นใจในการตอบทั้งสองส่วน ผลการศึกษาพบว่า ค่าความเชื่อมั่นของแบบทดสอบที่มีวิธีการตอบแบบบอกความมั่นใจในการตอบ สูงกว่าแบบทดสอบที่มีวิธีการตอบแบบธรรมดา คือ .85 และ .57 ตามลำดับ

อาบู ซาอีฟ (Abu-Sayf. 1975 : 160-161) ได้ศึกษาวิธีการตอบที่บอกระดับความมั่นใจ กลุ่มตัวอย่างเป็นนักเรียนระดับมัธยมศึกษา จำนวน 152 คน ให้ตอบแบบทดสอบเลือกตอบชนิด 4

ตัวเลือก วิชาภาษาอังกฤษ จำนวน 60 ข้อ ให้นักเรียนเลือกคำตอบที่ถูกที่สุดเพียงคำตอบเดียว และบอกระดับความมั่นใจในการตอบด้วยว่า ตอบด้วยความมั่นใจ ไม่ค่อยมั่นใจ หรือเดาตอบแล้ว เปรียบเทียบเปอร์เซ็นต์ในการตอบถูกหรือผิด ผลการศึกษาปรากฏว่าจากคำตอบทั้งหมดที่นักเรียน บอกว่า ตอบด้วยความมั่นใจ 3,491 ข้อ มีนักเรียนตอบผิด 846 ข้อ หรือร้อยละ 24.23 คำตอบ ที่นักเรียนบอกว่าไม่ค่อยมั่นใจ 2,468 ข้อ มีผิด 1,683 ข้อ หรือ 61.92 เปอร์เซ็นต์ และคำตอบ ที่นักเรียนบอกว่าตอบโดยการเดา 2,782 ข้อ มีนักเรียนตอบผิด 2,045 ข้อ หรือร้อยละ 73.31 จะเห็น ได้ว่าการตอบโดยอาศัยการเดามีโอกาสตอบถูกต้องน้อยมาก การเปิดโอกาสให้นักเรียนตอบโดย การเดาไม่ได้ช่วยให้นักเรียนได้คะแนนเพิ่มขึ้นมากนัก ถ้านักเรียนไม่มีความรู้ในเรื่องนั้นจริง ดังนั้น แบบทดสอบเลือกตอบก็มีใช้จะช่วยให้การตอบถูกโดยการเดามีมากนัก

ชี ซุง เวน (Shih-Soun Wen. 1975 : 197-199) ได้ศึกษาเกี่ยวกับวิธีทดสอบความมั่นใจ โดยศึกษาความสัมพันธ์ระหว่างคะแนนที่นับจากผลรวมของจำนวนข้อที่ตอบถูก กับคะแนนการ ทดสอบความมั่นใจ พบว่ามีความสัมพันธ์กันสูงถึง .97 ซึ่งในทางปฏิบัติแล้วคะแนนที่ได้จากทั้งสอง วิธีนี้ สามารถนำมาจัดอันดับ (Rank order) ได้เหมือนกัน แต่เขาให้ข้อคิดว่าถึงแม้สหสัมพันธ์ ระหว่างคะแนนที่ได้จากสองวิธีนี้จะสูง แต่ไม่ได้เป็นข้ออ้างที่จะปฏิเสธวิธีการทดสอบความมั่นใจ ถึงอย่างไรการทดสอบความมั่นใจ ก็สามารถให้รายละเอียดหรือข้อมูลจากการตอบข้อสอบได้ มากกว่า

อาบู ซาอีฟ และไดมอนด์ (Abu-Sayf and Daimond. 1976 : 62-65) ได้ศึกษาเกี่ยวกับ ค่าความเชื่อมั่น และความเที่ยงตรงของแบบทดสอบ เมื่อตรวจให้คะแนน 3 วิธี คือ วิธีที่ 1 ให้ คะแนนเฉพาะข้อที่ถูกและตอบอย่างมั่นใจ วิธีที่ 2 ให้คะแนนข้อที่ถูกโดยไม่ได้เดา วิธีที่ 3 ให้ คะแนนทุกข้อที่ตอบถูก โดยไม่คำนึงถึงว่ามีความมั่นใจในการตอบถูกหรือไม่เพียงใด แล้วหาค่า ความเชื่อมั่นโดยใช้สูตรครอนบัค แอลฟา (Cronbach's Alpha) ผลการศึกษาพบว่าการให้คะแนน เฉพาะข้อที่ถูกและมีความมั่นใจในการตอบ มีค่าความเชื่อมั่นสูงสุด คือ .898 และการให้คะแนน เฉพาะข้อที่ตอบถูกและไม่ได้เดา มีค่าความเชื่อมั่นสูงกว่าการให้คะแนนทุกข้อที่ตอบถูก โดยมีค่า ความเชื่อมั่น .871 และ .774 ตามลำดับ สำหรับค่าความเที่ยงตรง พบว่า วิธีการให้คะแนนเฉพาะข้อ ที่ตอบถูกและไม่ได้เดามีค่าสูงสุด คือ .260 รองลงมาคือ วิธีการให้คะแนนเฉพาะข้อที่ตอบถูกมีค่า ความเที่ยงตรงเป็น .236 และ .219 ตามลำดับ

เอสเซก (Essex. 1976 : 565-572) ได้ศึกษาวิธีการตรวจให้คะแนน 2 วิธี คือ วิธี 0-1 และ วิธีพิจารณาส่วนความรู้ย่อย โดยมีสูตรการให้คะแนนข้อสอบหนึ่งข้อ คือ (จำนวนตัวเลือกถูกทั้ง หมดที่นักเรียนเลือกหารด้วยจำนวนตัวเลือกถูกทั้งหมดในข้อนั้นๆ) คูณกับ (จำนวนตัวเลือกผิด

ทั้งหมดที่นักเรียนเลือกหารด้วยจำนวนตัวเลือกทั้งหมดในข้อนั้นๆ) โดยใช้แบบทดสอบเกี่ยวกับอาการของโรคหัวใจ จำนวน 150 ข้อคำถาม และมีแบบทดสอบวัดเจตคติแบบลิเคิร์ต 5 สเกล วัดความเห็นต่อวิธีการตอบและตรวจให้คะแนนด้วย กลุ่มตัวอย่างเป็นนักศึกษาคณะแพทย ชั้นปีที่ 1 ของมหาวิทยาลัยอิลลินอยส์ จำนวน 73 คน ผลการศึกษาพบว่า ค่าเฉลี่ยของคะแนนของวิธีการตรวจแบบพิจารณาส่วนความรู้ย่อย สูงกว่าวิธีการตรวจแบบ 0-1 อย่างมีนัยสำคัญ ส่วนค่าความเชื่อมั่นของแบบทดสอบที่ตรวจแบบพิจารณา ส่วนความรู้ย่อยสูงกว่าวิธี 0-1 อย่างไม่มีนัยสำคัญ นักศึกษาชอบวิธีการตรวจแบบพิจารณาส่วนความรู้ย่อยมากกว่า จะเห็นได้ว่าวิธีการตรวจให้คะแนนแบบพิจารณาส่วนความรู้ย่อยนั้น สามารถวัดความรู้ย่อยของผู้ตอบได้ดีกว่าวิธี 0-1

อัลบานีสและคณะ (Albanese et al. 1979 : 948-949) ได้ศึกษาค่าความเชื่อมั่นค่าความยาก ของแบบทดสอบเลือกตอบชนิดถูกผิดหลายตัวเลือก ที่มีวิธีการตรวจให้คะแนนโดย วิธี 0-1 และวิธีตรวจให้คะแนนแบบให้คะแนนตัวเลือกละ 1 คะแนน ถ้าผู้ตอบเลือกตอบได้ตรงกับสภาพเป็นจริงของตัวเลือกนั้น เปรียบเทียบกับแบบทดสอบเลือกตอบชนิดคำตอบซ้อน (Type K) โดยแบ่งออกเป็น 4 ชุดย่อย ชุดละ 40 ข้อ กลุ่มตัวอย่างเป็นนักศึกษาเภสัชวิทยา ภายภาพบับริงส์กายภาพ จำนวน 157 คน แบ่งออกเป็น 2 กลุ่มเท่าๆ กัน กลุ่มหนึ่งให้ตอบแบบทดสอบเลือกตอบชนิดถูกผิดหลายตัวเลือก อีกกลุ่มหนึ่งให้ตอบแบบทดสอบเลือกตอบชนิดคำตอบซ้อน และได้ศึกษากับกลุ่มตัวอย่างที่เป็นนักศึกษาแพทย์ชั้นปีที่ 2 จำนวน 175 คน แต่จำนวนข้อในแต่ละแบบทดสอบชุดย่อยเพิ่มเป็นอย่างละ 30 ข้อ ผลการศึกษาพบว่าแบบทดสอบเลือกตอบชนิดถูกผิดหลายตัวเลือกที่ใช้วิธีการตรวจแบบ 0-1 มีความยากกว่าตรวจแบบให้คะแนนตัวเลือกละ 1 คะแนน แต่มีความเชื่อมั่นต่ำกว่าการตรวจแบบให้คะแนนตัวเลือกละ 1 คะแนน อีกทั้งวิธีการตรวจให้คะแนนแบบที่สองนี้สามารถวิเคราะห์ความรู้ที่แท้จริงของผู้ตอบได้ และการสร้างแบบทดสอบเลือกตอบถูกผิดหลายตัวเลือกนี้สร้างง่ายกว่าแบบทดสอบเลือกตอบชนิดคำตอบซ้อน เพราะแบบทดสอบเลือกตอบชนิดคำตอบซ้อนนี้ ต้องสร้างตัวลองที่ดีก่อน แล้วจึงนำมาจัดกลุ่มของตัวลองอีกครั้งหนึ่งให้เป็นแบบคำตอบซ้อน และแบบทดสอบเลือกตอบถูกผิดหลายตัวเลือกนั้น นักศึกษาชอบมากกว่า อัลบานีสแนะนำว่าควรใช้แบบทดสอบเลือกตอบชนิดถูกผิดหลายตัวเลือก ในการวัดความรู้ของผู้ตอบ เพราะสามารถวัดส่วนของความรู้ย่อย ความผิดพลาด หรือจุดด้อยของผู้ตอบได้ดี และสร้างง่ายกว่าด้วย

ลอร์ด (Lord. 1975 : 7-11) ได้วิเคราะห์แบบทดสอบทางภาษา SCAT II Form 2A) ซึ่งเป็นแบบทดสอบเลือกตอบ 4 ตัวเลือก จำนวน 50 ข้อ กลุ่มตัวอย่างเป็นนักศึกษาระดับไฮสคูล โดยศึกษาเปรียบเทียบผลของการให้คะแนนข้อสอบแต่ละข้อแตกต่างกัน ลอร์ด พบว่า การให้คะแนน

ข้อสอบแต่ละข้อ โดยใช้วิธีค่าน้ำหนักคะแนนที่เหมาะสม โดยสูตรของลอว์ด จะให้รายละเอียด (Information) หรือผลของการสอบมากที่สุด และการให้ค่าน้ำหนักคะแนนรายข้อเท่ากัน ตามทฤษฎีดั้งเดิม จะให้รายละเอียดของการสอบน้อยที่สุด

เต คี ชู และชูศักดิ์ ชัมภลิจิต (Tse-Chi Hsu and Choosak Khampalikit. 1982 : 152-158) ได้ศึกษาเปรียบเทียบผลของการให้คะแนน โดยวิธีให้น้ำหนักคะแนนรายข้อที่เหมาะสม ตามทฤษฎีการตอบสนองข้อสอบ (IRT) โมเดลพารามิเตอร์ 3 ตัว กับวิธีการให้น้ำหนักคะแนนรายข้อเท่ากัน ตามทฤษฎีดั้งเดิม (Classical theory) โดยศึกษาข้อมูลกับสภาพการสอบคัดเลือกจริง ในข้อสอบวิชาภาษาจีน และข้อสอบวิชาประวัติศาสตร์ซึ่งสร้างขึ้นใหม่ และยังไม่มีการทดลองใช้มาก่อน โดยมีสมมติฐานของการวิจัยว่า การให้น้ำหนักคะแนนรายข้อที่เหมาะสมตามค่าพารามิเตอร์ของข้อสอบ น่าจะมีส่วนได้เปรียบกว่าวิธีการให้คะแนนแบบธรรมดา ในเรื่องที่ว่าผลของการให้คะแนนรายข้อที่เหมาะสมนั้น ค่าพารามิเตอร์ของข้อสอบ จะขึ้นอยู่กับคุณลักษณะของแบบทดสอบ น้อยกว่าการให้คะแนนแบบอื่น ผลการวิจัยพบว่า ข้อมูลที่ได้จากการทดสอบครั้งนี้ สามารถนำมาใช้กับทฤษฎีการตอบสนองข้อสอบ (IRT) ได้ และแก้ปัญหาแบบทดสอบที่ขาดคุณภาพได้ด้วย นั่นคือการให้น้ำหนักคะแนนรายข้อที่เหมาะสม จะให้ค่าความเชื่อมั่น ค่าอำนาจจำแนก สูงกว่าการให้คะแนนรายข้อเท่ากันตามวิธีดั้งเดิมในด้านความสัมพันธ์ของคะแนน และพบว่าคะแนนที่ได้จากการให้น้ำหนักคะแนนรายข้อตามวิธีเดิม กับคะแนนที่ได้จากการให้น้ำหนักคะแนนที่เหมาะสม (Optimal scoring method) โมเดลที่มีพารามิเตอร์ 3 ตัว มีความสอดคล้องกันสูงถึงร้อยละ 87 ของจำนวนผู้สมัครสอบและร้อยละ 10 ที่ถูกคัดเลือกโดยคะแนนทั้งสองแบบ

ฟรีสบาย และสวีนีย์ (Frisbie and Sweeney. 1982 : 29-35) ได้ศึกษาข้อได้เปรียบของ *MTF* แบบทดสอบวัดผลสัมฤทธิ์ที่มีตัวเลือกถูกหลายตัว โดยเทียบกับ แบบทดสอบที่มีตัวเลือกตัวเดียว กลุ่มตัวอย่างเป็นนักศึกษาระดับอุดมศึกษา ที่เรียนวิชาชีววิทยา จำนวน 574 คน ภาคเรียนแรกของปีการศึกษา 1980 โดยการกำหนดค่าน้ำหนักคะแนน แบบทดสอบเลือกตอบที่มีตัวเลือกถูกหลายตัวในแต่ละตัวเลือกเป็น 1 คะแนนเท่ากันหมด ถ้าผู้ตอบเลือกตอบได้ตรงกับสภาพเป็นจริงของตัวเลือกนั้นๆ และให้ 0 คะแนน แก่การตอบที่ตรงกันข้ามกับสภาพความเป็นจริงของตัวเลือกนั้นๆ ผลการศึกษาพบว่า

1. ใน 10 นาทีแรก อัตราส่วนแบบทดสอบเลือกตอบที่มีตัวเลือกถูกหลายตัว (ชนิดตัวเลือก) ต่อแบบทดสอบที่มีตัวเลือกถูกตัวเดียวเป็น .7 : 1 หรือ 3.44 : 1
2. ค่าความเชื่อมั่นของแบบทดสอบเลือกตอบที่มีตัวเลือกถูกหลายตัว สูงกว่าแบบทดสอบเลือกตอบที่มีตัวเลือกถูกตัวเดียว

3. ค่าความเที่ยงตรงของแบบทดสอบเลือกตอบที่มีตัวเลือกถูกหลายตัว สูงกว่าแบบทดสอบเลือกตอบที่มีตัวเลือกถูกตัวเดียว

4. ค่าความเที่ยงตรงตามสภาพ ของแบบทดสอบเลือกตอบที่มีตัวเลือกถูกหลายตัวกับแบบทดสอบเลือกตอบที่มีตัวเลือกถูกตัวเดียวของแบบทดสอบแต่ละชุด มีค่าความเที่ยงตรงตามสภาพสูง

5. นักศึกษาจำนวนร้อยละ 47.8 ชอบแบบทดสอบที่มีตัวเลือกถูกหลายตัว ร้อยละ 32 แบบทดสอบที่มีตัวเลือกถูกตัวเดียว ส่วนที่เหลือไม่ได้ชอบแบบทดสอบแบบใดมากกว่าแบบทดสอบอีกแบบหนึ่ง

บริตบาร์ด และกรีน (Bradbard and Green. 1985 : 68-72) ได้ศึกษาวิธีการตอบของคูมส์ สำหรับแบบทดสอบที่ใช้ในชั้นเรียน กลุ่มตัวอย่างเป็นนักศึกษาที่เรียนวิชาสถิติเบื้องต้นสำหรับธุรกิจ จำนวน 29 คน โดย 19 คน เป็นนักศึกษาชั้นปีที่ 2 และกลุ่มที่เป็นนักศึกษาชั้นปีที่ 3 เวลาที่ใช้ในการศึกษา 10 สัปดาห์ โดยใช้แบบทดสอบเลือกตอบชนิด 5 ตัวเลือก จำนวน 5 ฉบับ คือ ฉบับที่ 1 เรื่องความน่าจะเป็นและสถิติ เบื้องต้น ฉบับที่ 2 เรื่องตัวแปรต่อเนื่องและไม่ต่อเนื่อง ฉบับที่ 3 เรื่องการแจกแจงของกลุ่มตัวอย่าง และการทดสอบสมมติฐานที่มีกลุ่มตัวอย่างเดียว ฉบับที่ 4 เรื่องการกะประมาณ การทดสอบสมมติฐานที่มีกลุ่มตัวอย่างสองกลุ่ม และสมการรีเกรสชันอย่างง่าย ส่วนฉบับที่ 5 เป็นข้อสอบวัดความเข้าใจตั้งแต่ฉบับที่ 1 ถึงฉบับที่ 4 แบบทดสอบ 4 ฉบับแรกมีข้อสอบฉบับละ 12 ข้อ ส่วนฉบับที่ 5 มีข้อสอบจำนวน 25 ข้อ แบบทดสอบทุกข้อเป็นแบบทดสอบเลือกตอบที่มีตัวเลือกถูกตัวเดียว โดยใช้วิธีการตอบของคูมส์ คือ ให้เลือกตอบตัวเลือกที่ผิด หลังการตรวจให้คะแนนแบบทดสอบแต่ละฉบับ ต้องคืนแบบทดสอบให้กับนักศึกษาทุกครั้ง พร้อมทั้งวิจารณ์และให้ข้อเสนอแนะเกี่ยวกับการตอบครั้งต่อไป การให้คะแนนจะให้ 1 คะแนน สำหรับการกาข้อที่ผิดจริง และให้ -4 คะแนน ถ้ากาข้อที่ถูก โดยมีการห้ามการเดา ให้นักเรียนทำ โดยไม่จำกัดเวลา เปรียบเทียบการตรวจให้คะแนนกับวิธี 0-1 พบว่า

1. ความแตกต่างของค่าเฉลี่ยคะแนนรวมทั้งหมดของแบบทดสอบ 5 ฉบับ ที่ใช้วิธีของคูมส์ แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ส่วนวิธี 0-1 คะแนนเฉลี่ยแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ซึ่งจากจุดนี้ ถ้าใช้วิธีของคูมส์มาช่วยในการตัดเกรด จะทำให้เกรดที่ได้มีความเหมาะสมยิ่งขึ้น

2. การคิดคะแนนจากวิธีของคูมส์จะแยกเป็น 4 กรณี คือ ได้ 4 คะแนน ถือว่าเป็นคะแนนที่ถูกต้อง ได้คะแนนระหว่าง 0 และ 4 คะแนน ถือว่าเป็นคะแนนที่มีความรู้บางส่วน ได้คะแนน 0 คะแนน ถือว่าเป็นคะแนนที่ผิด และถ้าได้คะแนนน้อยกว่า 0 คะแนน คือติดลบ ถือว่า

เป็นคะแนนที่เกิดจากการเดา ผลการศึกษาค่าความแตกต่างของคะแนนเฉลี่ยใน 4 กรณีของแบบทดสอบทั้ง 5 ฉบับ พบว่า คะแนนที่ถูกต้องและคะแนนที่เกิดจากการเดามีแนวโน้มลดลงเรื่อยๆ จากการสอบด้วยแบบทดสอบฉบับแรกๆ ไปจนถึงการสอบด้วยแบบทดสอบฉบับที่ 5 ส่วนคะแนนที่มีความรู้บางส่วน และคะแนนที่ผิด มีแนวโน้มที่จะเพิ่มขึ้นการตีความหมายของคะแนนใน 4 กรณีนี้จะทำให้ผู้ตอบมีความระมัดระวังในการตอบมากขึ้น

3. ความเชื่อมั่นของแบบทดสอบทั้ง 5 ฉบับ ตามวิธีการของคูมบ์สจะให้ค่าความเชื่อมั่นสูงกว่าวิธีธรรมดา ผลจากการศึกษาครั้งนี้สอดคล้องกับผลวิจัยที่มีมาก่อน จากการวิเคราะห์ความแปรปรวนตามข้อ 2 เป็นการวิเคราะห์เพื่อดูคะแนนที่ถูกต้อง คะแนนบางส่วน คะแนนที่ผิด และคะแนนการเดา ซึ่งการวิเคราะห์แบบนี้จะทำให้การเดาลดลงเรื่อยๆ ผลที่ตามมาก็จะทำให้ค่าความเชื่อมั่นสูงขึ้น แต่ปรากฏว่าไม่เป็นไปตามนั้น ความเชื่อมั่นเฉลี่ยตามวิธีของคูมบ์สมีค่าเท่ากับ .66 ซึ่งไม่เป็นไปตามที่อีเบล (Ebel, 1969 : 565-570) ได้ศึกษาว่าแบบทดสอบวัดผลสัมฤทธิ์ที่ใช้ในชั้นเรียนควรมีค่าความเชื่อมั่น .70 ซึ่งค่าความเชื่อมั่นของแบบทดสอบทั้ง 5 ฉบับนี้ ไม่เหมาะที่จะเป็นแบบทดสอบในชั้นเรียน การที่ค่าความเชื่อมั่นต่ำนี้ อาจมาจากการหักคะแนนการเดานั้นเอง

ผลจากการทดลองของบริตบาร์ด และกรีนนี้ ไม่สามารถบอกได้ว่าวิธีการตอบและตรวจให้คะแนนตามวิธีของคูมบ์สนี้ จะทำให้แบบทดสอบในชั้นเรียนมีคุณภาพดีขึ้น แต่ก็สามารถเอาชนะการเดาได้ ทำให้คะแนนที่ได้มีความยุติธรรมขึ้น และนอกจากนี้ก็ทำให้มีค่าอำนาจจำแนกสูงขึ้นกว่าวิธีการตรวจแบบ 0-1 ด้วย

จาราดีท และซาเวจ (Jaradat and Sawaged, 1986 : 369-375) ได้เปรียบเทียบวิธีการให้คะแนน 3 วิธี คือ วิธี 0-1 วิธีใช้สูตรแก้การเดา และวิธีของกิบบอนส์และคณะที่เรียกว่า เทคนิคการเลือกกลุ่มย่อย โดยศึกษาค่าความเชื่อมั่น และความเที่ยงตรงของแบบทดสอบ กลุ่มตัวอย่างเป็นนักศึกษาระดับ 9 ของโรงเรียนในประเทศจอร์แดน จำนวน 160 คน เครื่องมือที่ใช้เป็นแบบทดสอบวิชาเคมี เรื่องทฤษฎีอะตอม และปฏิกิริยาเคมี แบบทดสอบที่สร้างขึ้นได้มีการปรึกษาและทำงานร่วมกันระหว่างผู้เชี่ยวชาญทางเนื้อหาและนักวัดผล จำนวน 6 ท่าน มีคุณภาพรายข้อตามเกณฑ์ที่กำหนด ผลการศึกษาค้นพบว่าความเชื่อมั่นของแบบทดสอบที่ตรวจให้คะแนนแบบ 0-1 สูงกว่าวิธีการให้คะแนนของกิบบอนส์และคณะอย่างมีนัยสำคัญ แต่สูงกว่าวิธีการให้คะแนนแบบแก้การเดาอย่างไม่มีนัยสำคัญ ในด้านความเที่ยงตรงของแบบทดสอบกว่าวิธีการตรวจให้คะแนนของกิบบอนส์และคณะสูงกว่าวิธีการตรวจให้คะแนนแบบ 0-1 และแก้การเดาอย่างมีนัยสำคัญ

จากการศึกษาของอัลบานีสและซาเบอร์ (Albanese and Sabers. 1988 : 111-123) โดยศึกษาค่าความเชื่อมั่นของแบบทดสอบเลือกตอบชนิดถูกผิดหลายตัวเลือก ที่มีจำนวนเลือกทั้งหมด 4 ตัวเลือก โดยมีวิธีการตรวจต่างกัน 4 วิธี คือ (1) วิธี 0-1 (2) วิธีการให้คะแนน 0.5 ในแต่ละข้อที่เลือกตอบได้ 3 ตัวเลือกที่ตรงกับเฉลย และให้ 1 คะแนนถ้าตอบได้ตรงกับเฉลย 4 ตัวเลือก (3) วิธีให้คะแนนลดลงครึ่งละ $\frac{1}{3}$ จากคะแนน 1 เมื่อตอบถูกลดลงทีละตัวเลือก และ (4) วิธีให้คะแนนตามตัวเลือกละ 1 คะแนน ถ้าตอบได้ตรงกับสภาพจริงของตัวเลือกนั้นๆ *MTF*
ผลการศึกษาพบว่า ค่าความเชื่อมั่นที่ตรวจให้คะแนนโดยวิธี (4) มีค่าสูงที่สุด และสหสัมพันธ์
 ของแบบทดสอบเลือกตอบถูกผิดหลายตัวเลือก ที่มีข้อคำถามเดียวกันในแต่ละรูปแบบสูงกว่าแบบทดสอบเลือกตอบถูกผิดหลายตัวเลือกที่มีข้อคำถามต่างกันในแต่ละรูปแบบ

ฟูจู ไต และ ฮอย เค ซุน (Fu-Ju Tsai and Hoi K. Suen. 1993 : 404) ได้ศึกษาค่าความเชื่อมั่น ค่าความยาก อำนาจจำแนก ของแบบทดสอบเลือกตอบถูกผิดหลายตัวเลือก ที่มีวิธีการตรวจให้คะแนนต่างกัน 6 วิธี คือ วิธีตรวจแบบ 0-1 วิธีให้น้ำหนักคะแนนตัวเลือกละ 1 คะแนนทุกตัวเลือก วิธี Count 2 วิธี Count 3 วิธีแก้การเดา และวิธีเพิ่มคะแนนตัวเลือกที่เว้นไว้ โดยใช้แบบทดสอบถูกผิดหลายตัวเลือก วิชาสถิติเบื้องต้นที่มีจำนวนตัวเลือก 4 ตัวเลือก จำนวน 20 ข้อ กลุ่มตัวอย่างเป็นนักศึกษาระดับปริญญาตรี จำนวน 199 คน โดยกระจายคำตอบของนักศึกษาแต่ละคนจะได้รับการตรวจทั้ง 6 วิธี ผลการศึกษาปรากฏว่าเมื่อตรวจวิธี 0-1 มีค่าความยากสูงสุด คือ 0.72 และแบบทดสอบที่ได้รับการตรวจด้วยวิธีต่างๆ ทั้ง 6 วิธี มีค่าความยากแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ยกเว้นวิธีตรวจแบบให้น้ำหนักคะแนนตัวเลือกละ 1 คะแนน วิธีแก้การเดา และวิธีเพิ่มคะแนนตัวเลือกที่เว้นไว้ ส่วนค่าความเชื่อมั่นและค่าอำนาจจำแนก เมื่อตรวจด้วยวิธีทั้ง 6 วิธี แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ผู้วิจัยได้แนะนำว่าจะเลือกใช้วิธีการตรวจให้คะแนนแบบใดขึ้นอยู่กับจุดประสงค์ หรือเป้าหมายของการสอบ เช่นจะเป็นแบบอิงเกณฑ์หรืออิงกลุ่ม เพราะถ้าเป็นการวัดแบบอิงกลุ่มแล้ว การตรวจให้คะแนนแบบใดคงไม่ให้ผลที่ต่างกันอย่างมากนัก แต่ถ้าแบบอิงเกณฑ์แล้ว ค่าความยากก็ต้องนำมาพิจารณาด้วย เพราะจะต้องอาศัยเกณฑ์ขั้นต่ำในการตัดสินการผ่านของผู้เรียน จึงต้องพิจารณาวิธีการตรวจให้คะแนนด้วย

เอกสารและงานวิจัยภายในประเทศ

อรรณณ ดันท์เจริญรัตน์ (2517 : บทคัดย่อ) ได้ทำการศึกษาเปรียบเทียบวิธีการตอบและการตรวจให้คะแนนวิธี 0 - 1 วิธีของคูมบัส และวิธีของ อนันต์ ศรีโสภา โดยใช้แบบทดสอบวิชาภาษาไทย กลุ่มตัวอย่างที่ใช้เป็นนักศึกษาวิทยาลัยครูหมู่บ้านจอมบึง จำนวน 214 คน โดยแบ่ง

ออกเป็นสามกลุ่มแต่ละกลุ่มใช้วิธีการตอบและการตรวจให้คะแนนตามวิธี 0 - 1 วิธีของคูมบัส และวิธีของอนันต์ ตามลำดับ

ในการวิจัยครั้งนี้ ได้ศึกษาเฉพาะค่าความเชื่อมั่น และเปอร์เซ็นต์การเดาตามเงื่อนไข การตอบและการตรวจให้คะแนนทั้งสามวิธีดังกล่าว

ผลจากการศึกษาปรากฏว่า

1. เปอร์เซ็นต์การเดาตามวิธีการตอบของคูมบัสสูงที่สุด และตามวิธี 0 - 1 ต่ำที่สุด
2. ค่าความเชื่อมั่นตามวิธีของอนันต์มีค่าสูงสุด และวิธี 0 - 1 ต่ำสุด

กาญจนา ศิริวัฒนพงษ์ (2520 : บทคัดย่อ) ได้ศึกษาเปรียบเทียบวิธีการตอบและตรวจให้คะแนนแบบทดสอบปรนัยเลือกตอบ ที่มีลักษณะต่างกัน คือ แบบทดสอบที่มีตัวเลือกถูกตัวเดียวแบบตัวเลือกถูกที่สุดเพียงตัวเดียว และแบบตัวเลือกถูกหลายตัว โดยใช้วิธีการตรวจให้คะแนนต่างกัน 3 วิธี คือ วิธี 0-1 วิธีของคูมบัส วิธีของอนันต์ โดยแบบทดสอบที่ใช้เป็นแบบทดสอบวัดผลสัมฤทธิ์ วิชาวิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ของโรงเรียนในจังหวัดปทุมธานี จำนวน 810 คน ผลการศึกษาพบว่า

1. เมื่อตอบและตรวจด้วยวิธี 0-1 และวิธีของคูมบัส ค่าความเชื่อมั่น ค่าความเที่ยงตรง และค่าอำนาจจำแนก ของแบบทดสอบทั้งสามลักษณะไม่แตกต่างกัน แต่เมื่อตอบและตรวจด้วยวิธีของอนันต์ ค่าความเชื่อมั่นของแบบทดสอบที่มีตัวเลือกถูกตัวเดียว สูงกว่าแบบที่มีตัวเลือกถูกที่สุดเพียงตัวเดียว และตัวเลือกถูกหลายตัว ส่วนค่าความเที่ยงตรงและค่าอำนาจจำแนก ไม่แตกต่างกัน

2. แบบทดสอบเลือกตอบที่มีตัวเลือกถูกตัวเดียวและตัวเลือกถูกที่สุดเพียงตัวเดียว เมื่อตอบและตรวจให้คะแนนโดยวิธีของอนันต์ ให้ค่าความเชื่อมั่นและค่าความเที่ยงตรงสูงกว่าวิธีตอบและตรวจของคูมบัส แต่ค่าความเที่ยงตรง และค่าอำนาจจำแนกไม่แตกต่างกัน ส่วนแบบทดสอบแบบปรนัยชนิดเลือกตอบที่มีตัวเลือกถูกหลายตัว เมื่อตอบและตรวจให้คะแนนด้วยวิธีทั้งสาม คุณภาพไม่แตกต่างกัน

3. เมื่อตอบและตรวจให้คะแนนทั้ง 3 วิธี แบบทดสอบปรนัยเลือกตอบที่มีตัวเลือกถูกตัวเดียว และตัวเลือกถูกหลายตัว มีการเดาน้อยกว่าตัวเลือกถูกที่สุดเพียงตัวเดียว ส่วนแบบทดสอบที่มีตัวเลือกถูกตัวเดียว การเดาจะน้อยกว่าแบบที่มีตัวเลือกถูกหลายตัว เมื่อตอบและตรวจให้คะแนนโดยวิธีของอนันต์ แต่ไม่พบว่าแตกต่างกันเมื่อตอบและตรวจด้วยวิธี 0-1 และวิธีของคูมบัส

4. แบบทดสอบปรนัยเลือกตอบทั้ง 3 ลักษณะเมื่อตอบและตรวจให้คะแนนด้วยวิธีของอนันต์ มีการเดาน้อยกว่าเมื่อตอบและตรวจให้คะแนนด้วยวิธี 0-1 และวิธีของคูมส์ ส่วนแบบทดสอบปรนัยเลือกตอบที่มีตัวเลือกถูกตัวเดียว และตัวเลือกถูกหลายตัวนั้น วิธี 0-1 มีการเดาน้อยกว่าวิธีของคูมส์ แต่ไม่พบความแตกต่างในแบบทดสอบปรนัยเลือกตอบ ที่มีตัวเลือกถูกที่สุดเพียงตัวเดียว

ธีรศักดิ์ อินทรมาศย์ (2520 : บทคัดย่อ) ได้ศึกษาถึงอิทธิพลของวิธีตอบแบบทดสอบปรนัยชนิดเลือกตอบวิธีต่าง ๆ ที่มีต่อค่าความเชื่อมั่น ค่าความเที่ยงตรงและปริมาณการเดา การวิจัยครั้งนี้มีจุดมุ่งหมาย เพื่อจะศึกษาผลของวิธีตอบแบบทดสอบปรนัยชนิดเลือกตอบที่มีวิธีตอบธรรมดา วิธีตอบโดยบอกความมั่นใจในการตอบ และวิธีตอบทุกตัวเลือกที่มีต่อค่าความเชื่อมั่น ค่าความเที่ยงตรง และปริมาณการเดาของแบบทดสอบ โดยใช้แบบทดสอบวิชาวิทยาศาสตร์ทดสอบกับนักเรียนชั้นประถมศึกษาปีที่ 7 ของโรงเรียนเมืองสมุทรสงคราม จำนวน 216 คน ในการดำเนินการสอบให้กลุ่มตัวอย่างที่หนึ่งได้รับแบบทดสอบที่มีวิธีตอบธรรมดา กลุ่มตัวอย่างที่สองได้รับแบบทดสอบที่มีวิธีตอบโดยบอกความมั่นใจในการตอบ กลุ่มที่สามได้รับแบบทดสอบที่มีวิธีตอบทุกตัวเลือก

ผลการศึกษาพบว่า วิธีการตอบโดยบอกความมั่นใจในการตอบมีค่าความเชื่อมั่นสูงสุดสูงกว่าวิธีตอบธรรมดาและวิธีตอบทุกตัวเลือก อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนวิธีการตอบทุกตัวเลือกมีค่าความเชื่อมั่นสูงกว่าวิธีตอบธรรมดา อย่างไม่มีนัยสำคัญทางสถิติ

ในด้านความเที่ยงตรงของแบบทดสอบพบว่า แบบทดสอบที่มีวิธีตอบทุกตัวเลือก มีค่าความเที่ยงตรงสูงสุด แบบทดสอบที่มีวิธีตอบโดยบอกความมั่นใจในการตอบมีค่าความเที่ยงตรงต่ำสุด แต่ทั้ง 3 วิธีแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

สำหรับปริมาณการเดานั้น แบบทดสอบที่มีวิธีตอบแตกต่างกันทั้ง 3 วิธี มีปริมาณการเดาแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยแบบทดสอบที่มีวิธีตอบธรรมดา มีปริมาณการเดาสูงสุด แบบทดสอบที่มีวิธีตอบทุกตัวเลือกมีปริมาณการเดาต่ำสุด และนักเรียนที่มีความสามารถในระดับต่ำจะตอบแบบทดสอบโดยการเดามากกว่านักเรียนที่มีความสามารถสูงทุกวิธีตอบ

ในปีเดียวกันนี้ เพ็ญศรี สว่างเนตร (2520 : บทคัดย่อ) ได้ศึกษาค่าความเชื่อมั่นของแบบทดสอบชนิดเลือกตอบด้วยเทคนิคการให้คะแนนที่ต่างกันเครื่องมือที่ใช้เป็นแบบทดสอบชนิดเลือกตอบ 5 ตัวเลือก ซึ่งเป็นแบบทดสอบวัดผลสัมฤทธิ์วิชาภาษาอังกฤษ นำไปสอบกับนักศึกษา

ระดับประกาศนียบัตรวิชาชีพการศึกษาระดับสูง ที่เรียนวิชาภาษาอังกฤษเป็นวิชาเอกหรือวิชาโท จำนวน 218 คน พร้อมกระบวนการทดสอบความมั่นใจจากเทคนิค การให้คะแนน 4 วิธี คือวิธีกำหนดคะแนนให้ข้อที่ตอบถูกข้อละหนึ่งคะแนน วิธีลดคะแนนข้อที่ตอบผิด วิธีเพิ่มคะแนนให้ข้อที่เว้น และวิธีการทดสอบความมั่นใจ ได้คำนวณค่าความเชื่อมั่นโดยวิธีวิเคราะห์ความแปรปรวนของฮอยด์ (Hoyd's analysis of variance) ได้ข้อค้นพบว่า ค่าความเชื่อมั่นภายใต้เทคนิคการให้คะแนนความมั่นใจสูงกว่า ความเชื่อมั่นที่ได้จากวิธีกำหนดคะแนนให้ ข้อที่ตอบถูกข้อละหนึ่งคะแนน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 คือ .61 และ .55 ตามลำดับ ค่าความเชื่อมั่น โดยใช้เทคนิคการลดคะแนนของข้อที่ตอบผิด .56 เพิ่มขึ้นเล็กน้อยจาก .55 ซึ่งเป็นค่าความเชื่อมั่นจากวิธีให้คะแนนที่ใช้กันทั่วไป แต่ความแตกต่างนี้ไม่มีนัยสำคัญทางสถิติส่วนเทคนิคการเพิ่มคะแนนให้ข้อที่เว้นไม่ทำให้ค่าความเชื่อมั่นเพิ่มขึ้นจะเห็นได้ว่าเทคนิคการให้คะแนนความเชื่อมั่นมีประสิทธิภาพที่สุดในการเพิ่มค่าความเชื่อมั่นของการวัดผลสัมฤทธิ์ทางการเรียนของแบบทดสอบชนิดเลือกตอบ

สำราญ มีแจ้ง (2525 : บทคัดย่อ) ได้ทำการศึกษาผลของคำสั่ง และการให้คะแนนที่ต่างกันต่อค่าความเชื่อมั่น ความเที่ยงตรงและค่าอำนาจจำแนกของแบบทดสอบเลือกตอบ เพื่อศึกษาคำสั่งและการให้คะแนนที่ต่างกัน 4 วิธี คือ วิธี 0 - 1 วิธีของคูมบัส วิธีของอนันต์ และวิธีของสำราญ ว่ามีผลต่อค่าความเชื่อมั่น ความเที่ยงตรง และค่าอำนาจจำแนกของแบบทดสอบเลือกตอบที่มีตัวเลือกถูกตัวเดียวหรือไม่ สำราญ มีแจ้ง ได้กำหนดวิธีการตอบและตรวจให้คะแนน 4 กรณี คือ

1. ถ้าผู้สอบทำเครื่องหมายตรงกับตัวเลือกที่ถูกที่สุดเพียงตัวเดียวแล้วถูกต้อง จะได้ 4 คะแนน
2. ถ้าผู้สอบไม่มั่นใจในการตอบและทำเครื่องหมาย 2 ตัวเลือก และมีตัวเลือกถูกอยู่ด้วย จะได้ 2 คะแนน ถ้าไม่มีตัวเลือกถูกปนอยู่ด้วยจะได้ 0 คะแนน
3. ถ้าผู้สอบไม่มั่นใจในการตอบ และทำเครื่องหมาย 3 ตัวเลือกถ้ามีตัวเลือกถูกอยู่ด้วย จะได้ 1 คะแนน ถ้าไม่มีตัวเลือกถูกปนอยู่ด้วยจะได้ 0 คะแนน
4. ถ้าผู้ตอบเว้นว่างหรือทำเครื่องหมายทั้ง 4 ตัวเลือก จะได้ 0 คะแนน

กลุ่มตัวอย่างที่ใช้ในการวิจัยเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 4 ของโรงเรียนสุรศักดิ์มนตรี และโรงเรียนวัดน้อยใน ซึ่งเรียนวิชาคณิตศาสตร์ (ค 011) ปีการศึกษา 2524 จำนวน 240 คน ผลการวิจัยพบว่า

วิธีของคัมภีร์กับวิธีของอนันต์ และวิธีของอนันต์กับวิธีของสำราญ พบว่าค่าความเชื่อมั่นไม่มี ความแตกต่างกัน

2. ค่าความเที่ยงตรงของแบบทดสอบที่ใช้คำสั่ง และการให้คะแนนตามวิธีของสำราญ และวิธีของอนันต์ มีความแตกต่างกับวิธี 0 - 1 ส่วนวิธี 0 - 1 กับวิธีของคัมภีร์ วิธีของคัมภีร์กับวิธีของอนันต์ และวิธีของอนันต์กับวิธีของสำราญ พบว่าความเที่ยงตรงไม่แตกต่างกัน

3. คำสั่งและการให้คะแนนทั้ง 4 วิธี ไม่ทำให้ค่าอำนาจจำแนกของแบบทดสอบ แตกต่างกัน

เกริกชัย ฮวบเจริญ (2525 : บทคัดย่อ) ได้ศึกษาเปรียบเทียบค่าความเชื่อมั่นและค่า คะแนนการเดา ของการตอบและตรวจให้คะแนนแบบทดสอบปรนัย ชนิดเลือกตอบแบบธรรมดา แบบทดสอบความมั่นใจ และวิธีตอบและตรวจให้คะแนนแบบใหม่ ซึ่งวิธีการตอบและตรวจให้ คะแนนแบบใหม่นี้ คือ วิธีตอบที่มีรูปแบบของกระดาษคำตอบชนิดพิเศษ เมื่อตอบแล้วนักเรียน สามารถทราบได้ทันทีว่า ตนเองทำผิดหรือถูกในข้อนั้น ๆ และถ้านักเรียนทำผิดสามารถตัดสินใจ ได้อย่างเสรีว่า จะทำข้อสอบเดิมต่อไปหรือไม่ทำ หรือตอบจนกระทั่งถูก ผลการศึกษาปรากฏว่า แบบทดสอบเลือกตอบเมื่อใช้วิธีการตอบ และตรวจให้คะแนนด้วยวิธีใหม่ มีค่าความเชื่อมั่นสูงสุด และเมื่อตอบและตรวจให้คะแนนด้วยวิธีธรรมดามีค่าความเชื่อมั่นต่ำสุด เมื่อทดสอบความแตก- ต่างของค่าความเชื่อมั่นพบว่า แบบทดสอบที่มีวิธีการตอบและตรวจให้คะแนนแบบใหม่และแบบ ทดสอบที่มีวิธีการตอบและตรวจให้คะแนนแบบความมั่นใจ มีค่าความเชื่อมั่นแตกต่างจากแบบ ทดสอบที่มีวิธีการตอบและตรวจให้คะแนนแบบธรรมดา อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนแบบทดสอบที่มีวิธีการตอบและตรวจให้คะแนนแบบใหม่ กับแบบทดสอบที่มีวิธีการตอบและ ตรวจให้คะแนนแบบวิธีทดสอบความมั่นใจ มีค่าความเชื่อมั่นแตกต่างกันอย่างไม่มีนัยสำคัญ ทางสถิติ

ทวี ทองคำ (2526 : บทคัดย่อ) ศึกษาเปรียบเทียบคำสั่งและวิธีการตรวจให้คะแนน ต่างกัน 3 วิธี คือวิธี 0-1 วิธีของสำราญ วิธีของกิบบอนส์และคณะ ที่มีต่อค่าความเชื่อมั่น ความเที่ยงตรง ค่าอำนาจจำแนก ของแบบทดสอบเลือกตอบชนิด 5 ตัวเลือก จำนวน 60 ข้อ วิชา คณิตศาสตร์พานิชยกรรม 2 กลุ่มตัวอย่างเป็นนักศึกษาระดับประกาศนียบัตรวิชาชีพชั้นปีที่ 1 ของ วิทยาลัยเทคนิคยะลา 180 คน โดยแบ่งกลุ่มตัวอย่างออกเป็น 3 กลุ่ม แต่ละกลุ่มได้รับการสอบโดย แบบทดสอบที่มีวิธีการตอบและตรวจให้คะแนนแต่ละวิธี ผลการศึกษาพบว่า

1. ค่าความเชื่อมั่นของแบบทดสอบเมื่อตรวจให้คะแนนด้วยวิธีต่างกัน แตกต่างกัน อย่างไม่มีนัยสำคัญที่ระดับ .05 โดยวิธีการตรวจแบบสำราญแสดงแนวโน้มว่าค่าความเชื่อมั่น

ของวิทยาลัยเทคนิคยะลา 180 คน โดยแบ่งกลุ่มตัวอย่างออกเป็น 3 กลุ่ม แต่ละกลุ่มได้รับการสอบโดยแบบทดสอบที่มีวิธีการตอบและตรวจให้คะแนนแต่ละวิธี ผลการศึกษาพบว่า

1. ค่าความเชื่อมั่นของแบบทดสอบเมื่อตรวจให้คะแนนด้วยวิธีต่างกัน แตกต่างกัน อย่างไม่มีนัยสำคัญที่ระดับ .05 โดยวิธีการตรวจแบบสำราญแสดงแนวโน้มว่าค่าความเชื่อมั่นสูงสุด วิธีของกิบบอนส์มีค่าความเชื่อมั่นต่ำสุด

2. ค่าความเที่ยงตรงตามสภาพของแบบทดสอบ โดยใช้ผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์พานิชยกรรม 1 เป็นเกณฑ์ พบว่า แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 เมื่อตรวจให้คะแนนแตกต่างกันทั้ง 3 วิธี โดยแบบทดสอบที่ใช้คำสั่งและวิธีการตรวจให้คะแนนของกิบบอนส์ ให้ค่าความเที่ยงตรงตามสภาพสูงสุด

3. ค่าความตรงตามสภาพ เมื่อให้ครูผู้สอนเป็นผู้จัดกลุ่มความสามารถเป็น 3 กลุ่ม พบว่า วิธีการให้คะแนนทั้ง 3 วิธี มีค่าความตรงตามสภาพทั้ง 3 วิธี ที่ระดับนัยสำคัญ .05 แบบทดสอบที่ตรวจด้วยวิธีของสำราญ วิธีของกิบบอนส์และคณะสามารถจำแนกนักเรียนระหว่างกลุ่มได้ทุกกลุ่ม

4. ค่าอำนาจจำแนกของแบบทดสอบ เมื่อใช้คำสั่งและวิธีการตรวจให้คะแนนต่างกัน 3 วิธี แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

ดร.น. หาญตระกูล และ นิโบล นิมกิงรัตน์ (2526 : 210-214) ได้ศึกษาเปรียบเทียบคุณภาพของแบบทดสอบ ชนิดเลือกตอบที่มีตัวถูกเดียว ที่มีการตรวจให้คะแนนที่แตกต่างกัน 4 วิธี คือ วิธี 0-1 วิธีของคูลส์ วิธีของอนันต์ และวิธีให้ผู้เรียนกำหนดน้ำหนักคะแนน ว่าต้องการกี่คะแนน จาก 1 ถึง 5 คะแนน ถ้าตอบผิดจะถูกหักคะแนนตามที่ระบุ ถ้าตอบถูกก็จะได้คะแนนตามนั้น โดยการศึกษาครั้งนี้มีจุดมุ่งหมายที่จะเปรียบเทียบค่าความเที่ยงตรง ค่าความเชื่อมั่น และค่าอำนาจจำแนก ของวิธีการตรวจให้คะแนนทั้ง 4 วิธี กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้านี้เป็นนักศึกษาคณะศึกษาศาสตร์ มหาวิทยาลัยเชียงใหม่ที่ลงทะเบียนเรียนวิชาการวัดและประเมินผลการศึกษา จำนวน 256 คน แบ่งนักศึกษาออกเป็น 4 กลุ่ม กลุ่มละ 64 คน ใช้ในการทดลองแต่ละวิธี

ผลการศึกษาพบว่า

1. ค่าความเที่ยงตรงของแบบทดสอบที่มีการตรวจให้คะแนนทุกวิธีแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 นอกจากวิธี 0 - 1 ให้ค่าความเที่ยงตรงของแบบทดสอบแตกต่างกับวิธีของอนันต์ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 การตรวจให้คะแนนตามวิธีของคูลส์

มีแนวโน้มว่าค่าความเที่ยงตรงของแบบทดสอบสูงสุด รองลงมาคือวิธีกำหนดน้ำหนักคะแนน และวิธีของอนันต์ให้ค่าความเที่ยงตรงของแบบทดสอบต่ำสุด

2. ค่าความเชื่อมั่นของแบบทดสอบที่มีการตรวจให้คะแนน ตามวิธีของอนันต์แตกต่างกับอีก 3 วิธี อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 วิธีของคุมส์ให้ค่าความเชื่อมั่นของแบบทดสอบแตกต่างกับวิธีกำหนดน้ำหนักอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 วิธี 0 - 1 และวิธีกำหนดน้ำหนักให้ค่าความเชื่อมั่นแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 การตรวจตามวิธีของอนันต์มีแนวโน้มว่า ค่าความเชื่อมั่นของแบบทดสอบสูงสุด วิธีกำหนดน้ำหนักให้ค่าความเชื่อมั่นต่ำสุด

3. ค่าอำนาจจำแนกของแบบทดสอบทุกวิธี แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ วิธีของอนันต์มีค่าอำนาจจำแนกสูงสุด และวิธีกำหนดน้ำหนักให้ค่าอำนาจจำแนกต่ำสุด

จินดา โตอนันต์ (2527 : บทคัดย่อ) ได้ศึกษาเปรียบเทียบคุณลักษณะของแบบทดสอบวัดผลสัมฤทธิ์เลือกตอบ ที่มีตัวถูกหลายตัว หรือแบบถูกผิดทุกตัวเลือกกับแบบที่มีตัวถูกถูกตัวเดียว ด้านค่าความเชื่อมั่น ความเที่ยงตรง ความยาก และอำนาจจำแนก ในแบบทดสอบวิชาวิทยาศาสตร์ กลุ่มตัวอย่างที่ใช้เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ในเขตกรุงเทพมหานคร จำนวน 432 คน

ผลการศึกษาปรากฏว่า

1. ค่าความยากของแบบทดสอบ ที่มีตัวถูกถูกตัวเดียว และตัวถูกถูกหลายตัวไม่เท่ากัน แบบทดสอบเลือกตอบที่มีตัวถูกถูกตัวเดียว มีความยากมากกว่าแบบทดสอบที่มีตัวถูกถูกหลายตัวอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

2. ค่าอำนาจจำแนกของแบบทดสอบ ที่มีตัวถูกถูกตัวเดียว สูงกว่าแบบทดสอบที่มีตัวถูกถูกหลายตัว อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

3. ค่าความเชื่อมั่นของแบบทดสอบทั้งสองลักษณะไม่แตกต่างกัน

4. นักเรียนกลุ่มต่ำชอบแบบทดสอบที่มีตัวถูกถูกหลายตัว มากกว่านักเรียนกลุ่มสูง ในขณะที่นักเรียนกลุ่มสูงต้องการให้ครูออกข้อสอบเลือกตอบ ที่มีตัวถูกถูกตัวเดียว

นาภา กาญจนกิจโสภณ (2528 : บทคัดย่อ) ศึกษาเปรียบเทียบคะแนนสอบ ระหว่างวิธีให้น้ำหนักคะแนนรายข้อตามทฤษฎีดั้งเดิม กับวิธีการให้คะแนนรายข้อต่างกันตามทฤษฎีการตอบข้อคำถาม โมเดลที่มีพารามิเตอร์ 3 ตัว ใช้แบบทดสอบวิชาคณิตศาสตร์ วิทยาศาสตร์ และพลานามัย ซึ่งเป็นแบบทดสอบของโครงการตรวจสอบคุณภาพทางการศึกษา กรมวิชาการ

กระทรวงศึกษาธิการ กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526 กลุ่มวิชาละ 3,907 คน โดยนำข้อมูลเดิมซึ่งตรวจให้คะแนนแบบน้ำหนักรายข้อเท่ากัน มาตรวจให้คะแนนแต่ละระดับความสามารถตามสูตร

$$W_i(\theta) = \frac{Da_i}{1 + c_i e^{-DL_i}}, i = 1, 2, 3, \dots, n$$

นำค่าคะแนนที่ได้ ไปเปรียบเทียบกับคะแนนที่ได้จากการตรวจให้คะแนนน้ำหนักรายข้อเท่ากัน ผลการศึกษาพบว่า

1. ค่ามัธยฐานเลขคณิต และส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบโดยวิธีให้น้ำหนักคะแนนรายข้อเท่ากัน แตกต่างกับคะแนนสอบ ที่ได้จากวิธีให้น้ำหนักคะแนนรายข้อต่างกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ทั้ง 3 วิชา

2. ความสัมพันธ์ของคะแนนสอบ ระหว่างการตรวจให้คะแนนทั้ง 2 วิธี ในวิชาคณิตศาสตร์เป็น .937 วิชาพลานามัยมีค่า .897 และพบว่ามีนัยสำคัญทางสถิติที่ระดับ .01 ทั้ง 3 วิชา

3. ค่าความเที่ยงตรงของแบบทดสอบวิชาคณิตศาสตร์ วิทยาศาสตร์ และพลานามัยของวิธีการให้คะแนนรายข้อต่างกันตามทฤษฎีตอบข้อคำถาม สูงกว่าค่าความเที่ยงตรงของวิธีที่ให้น้ำหนักคะแนนรายข้อเท่ากัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ทั้ง 3 วิชา

สุพัฒน์ สุกมลสันต์ (2528 : บทคัดย่อ) เปรียบเทียบวิธีการให้น้ำหนักคะแนนที่ต่างกัน 5 วิธี คือ (1) วิธีใช้ค่าเดลตา (Δ) (2) วิธีใช้ค่าเบต้าแปลง (β_i) วิธีใช้ค่าน้ำหนักคะแนนที่เหมาะสมของลอร์ด (4) วิธีใช้ค่าผลบวกของความแปรปรวนร่วม (h^2) (5) วิธีประเพณีนิยม เครื่องมือที่ใช้เป็นแบบทดสอบการอ่านจับใจความภาษาอังกฤษ ที่มีโครงสร้างต่างกัน 2 ฉบับ ฉบับที่ 1 เป็นฉบับที่ไม่มีชื่อเรื่องและภาพประกอบ ส่วนฉบับที่ 2 เป็นฉบับที่มีชื่อเรื่องและภาพประกอบ กลุ่มตัวอย่างเป็นนิสิตชั้นปีที่ 1 ที่เรียนวิชาภาษาอังกฤษพื้นฐาน เฉพาะการอ่าน 2 ของจุฬาลงกรณ์มหาวิทยาลัย จำนวน 1,872 คน ผลการศึกษาพบว่า

1. ค่าความเที่ยงตรงเชิงพยากรณ์ของแบบทดสอบอิงกลุ่ม

1.1 ค่าความเที่ยงตรงเชิงพยากรณ์ของแบบทดสอบที่ไม่มีชื่อเรื่องและภาพประกอบ พบว่า วิธีการให้คะแนนโดยใช้ผลบวกความแปรปรวนร่วม (h^2) จะมีค่าสูงสุดและสูงกว่าความเที่ยงตรงเชิงพยากรณ์ ของแบบทดสอบที่ให้คะแนนโดยวิธีใช้ค่าน้ำหนักคะแนนที่เหมาะสม (w) ซึ่งมีค่าต่ำสุดอย่างมีนัยสำคัญทางสถิติ

1.2 ในแบบทดสอบที่มีชื่อเรื่องและภาพประกอบพบว่า ค่าความเที่ยงตรงเชิงพยากรณ์ ที่ให้ค่าน้ำหนักคะแนนที่เหมาะสม (w) มีค่าสูงสุด และสูงกว่าวิธีเบต้าแปลง (β_j) ซึ่งมีค่าต่ำสุดอย่างมีนัยสำคัญทางสถิติ

1.3 ค่าความเที่ยงตรงเชิงพยากรณ์ทั้ง 2 แบบ ที่ให้คะแนนโดยวิธีต่างๆ มีลำดับที่ต่างกัน

2. ค่าความเชื่อมั่นของแบบทดสอบอิงกลุ่ม

2.1 ในแบบทดสอบที่ไม่มีชื่อเรื่องและภาพประกอบ ค่าความเชื่อมั่นที่ให้คะแนนโดยวิธีน้ำหนักคะแนนที่เหมาะสม (w) มีค่าสูงสุด และสูงกว่าค่าความเชื่อมั่นที่ให้คะแนนโดยวิธีอื่นๆ อย่างมีนัยสำคัญ โดยวิธีใช้ค่าเบต้าแปลง (β_j) จะมีค่าต่ำสุด

2.2 ในแบบทดสอบที่มีชื่อเรื่องและภาพประกอบ ค่าความเชื่อมั่นที่ให้คะแนนโดยวิธีน้ำหนักคะแนนที่เหมาะสม (w) มีค่าสูงสุด โดยสูงกว่าค่าความเชื่อมั่นที่ให้คะแนนโดยวิธีอื่นๆ อย่างมีนัยสำคัญ โดยค่าความเชื่อมั่นที่ใช้วิธีเบต้าแปลง (β_j) จะมีค่าต่ำสุด

2.3 ลำดับที่ของค่าความเชื่อมั่นของแบบทดสอบทั้ง 2 ฉบับ ที่ให้คะแนนโดยวิธีต่างๆ เหมือนกันทั้ง 2 แบบทดสอบ

3. ค่าความเชื่อมั่นของแบบทดสอบแบบอิงเกณฑ์ เมื่อจุดตัดอยู่ระหว่างร้อยละ 0 ถึง 100 พบว่า วิธีการให้คะแนนตามประเพณีนิยมมีค่าสูงสุด และวิธีให้คะแนนโดยใช้ค่าน้ำหนักคะแนนที่เหมาะสมมีค่าต่ำสุด ทั้งใน 2 แบบทดสอบ แต่ถ้าจุดตัดอยู่ระหว่างร้อยละ 50 ถึง 100 วิธีการให้คะแนนโดยใช้ค่าเบต้าแปลง (β_j) จะมีค่าความเชื่อมั่นสูงสุด และวิธีให้น้ำหนักคะแนนเหมาะสมให้ค่าต่ำสุดทั้งใน 2 แบบทดสอบเช่นกัน และค่าลำดับที่ของแบบทดสอบทั้งสองเหมือนกันทั้ง 2 กรณี

4. ความคงที่ของอันดับที่ของคะแนน พบว่า อันดับที่ของคะแนนของแบบทดสอบที่ให้คะแนนโดยวิธีใช้ค่าผลบวกของความแปรปรวนร่วม (h^2) มีความคงที่มากที่สุด ทั้งในแบบทดสอบ 2 ฉบับ และสำหรับแบบทดสอบที่ไม่มีชื่อเรื่องและภาพประกอบ วิธีการให้น้ำหนักคะแนนที่

เหมาะสมมีค่าความคงที่ต่ำสุด ส่วนแบบสองที่มีชื่อเรื่องและภาพประกอบ วิธีใช้ค่าเบต้าแปลง (β) มีค่าต่ำสุด

5. ค่าความสัมพันธ์ของคะแนน พบว่า วิธีการให้คะแนนด้วยวิธีต่างๆ มีความสัมพันธ์กันอย่างมีนัยสำคัญทั้งใน 2 แบบทดสอบ

พนิชา สวัสดิ์มงคล (2530 : บทคัดย่อ) ศึกษาเปรียบเทียบค่าความเที่ยงตรง ค่าความเชื่อมั่นของแบบทดสอบเลือกตอบ ที่วัดระดับความรู้ขั้นต่ำ และขั้นสูง โดยมีวิธีการให้คะแนนต่างกัน 2 วิธี คือ วิธีธรรมดา และวิธีการให้คะแนนความมั่นใจ ในกลุ่มผู้สอบที่มีผลสัมฤทธิ์ทางการเรียนต่างกัน คือ สูง ปานกลาง ต่ำ เครื่องมือที่ใช้คือ แบบทดสอบวิชาการพยาบาลผู้ป่วยอายุรกรรมและศัลยกรรม จำนวน 60 ข้อ กลุ่มตัวอย่างคือ นักเรียนพยาบาลที่เรียนวิชาดังกล่าว

ผลการวิจัยพบว่า ด้านความเที่ยงตรง ความเชื่อมั่นของแบบทดสอบในกลุ่มผู้สอบที่มีผลสัมฤทธิ์ทางการเรียนสูง ปานกลาง และต่ำ 1) แบบทดสอบที่วัดระดับความรู้ขั้นต่ำและมีวิธีการให้คะแนนวิธีธรรมดา ค่าความเชื่อมั่นของแบบทดสอบ ระหว่างกลุ่มต่ำกับกลุ่มปานกลางและสูง มีค่าแตกต่างกันอย่างมีนัยสำคัญที่ระดับ .01 ระหว่างกลุ่มสูง กับปานกลาง ไม่แตกต่างกัน ค่าความเที่ยงตรงของแบบทดสอบที่ได้จากทั้ง 3 กลุ่ม มีค่าไม่แตกต่างกัน 2) แบบทดสอบที่วัดระดับความรู้ขั้นสูง และมีวิธีให้คะแนนวิธีธรรมดา ค่าความเชื่อมั่น ค่าความเที่ยงตรง ของแบบทดสอบที่ได้จากทั้ง 3 กลุ่ม มีค่าไม่แตกต่างกันอย่างมีนัยสำคัญ 3) แบบทดสอบที่วัดระดับความรู้ขั้นต่ำ และมีวิธีการให้คะแนนความมั่นใจ ค่าความเชื่อมั่นของแบบทดสอบที่ได้จากทั้ง 3 กลุ่ม ไม่แตกต่างกันอย่างมีนัยสำคัญ ค่าความเที่ยงตรงของแบบทดสอบระหว่างกลุ่มสูงและกลุ่มต่ำ มีค่าแตกต่างกันอย่างมีนัยสำคัญที่ระดับ .05 ระหว่างกลุ่มปานกลางกับกลุ่มสูงและต่ำ มีค่าไม่แตกต่างกัน 4) แบบทดสอบที่วัดระดับความรู้ขั้นสูงและมีวิธีให้คะแนนความมั่นใจ ค่าความเชื่อมั่นของแบบทดสอบ ระหว่างกลุ่มสูงและกลุ่มต่ำแตกต่างกันอย่างมีนัยสำคัญที่ระดับ .05 ระหว่างกลุ่มปานกลางกับสูงและต่ำ มีค่าไม่แตกต่างกัน ค่าความเที่ยงตรงของแบบทดสอบที่ได้จากทั้ง 3 กลุ่ม มีค่าไม่แตกต่างกัน ผลการเปรียบเทียบค่าความเชื่อมั่น พบว่า 1) แบบทดสอบที่วัดระดับความรู้ขั้นต่ำ ค่าความเชื่อมั่นของแบบทดสอบที่ได้จากวิธีให้คะแนนทั้ง 2 วิธี มีค่าไม่แตกต่างกันอย่างไม่มีนัยสำคัญที่ระดับ .01 ในกลุ่มสูง ส่วนกลุ่มปานกลางและต่ำมีค่าไม่แตกต่างกัน 2) แบบทดสอบที่วัดระดับความรู้ขั้นสูง ค่าความเชื่อมั่นของแบบทดสอบที่ได้จากการวิธีการให้คะแนนทั้ง 2 วิธี มีค่าไม่แตกต่างกันอย่างมีนัยสำคัญ ทั้งในกลุ่มสูง ปานกลาง และต่ำ ค่าความเที่ยงตรงของแบบทดสอบ มีค่าแตกต่างกันอย่างมีนัยสำคัญที่ระดับ .01 ในกลุ่มต่ำส่วนกลุ่มปานกลางและสูง ค่าไม่แตกต่างกัน

เพลินพิศ คุณคำ (2532 : บทคัดย่อ) ได้ศึกษาเปรียบเทียบค่าความเชื่อมั่น ค่าความเที่ยงตรง เปอร์เซนต์การเดา และเปอร์เซนต์ของคะแนนที่ถูก คะแนนบางส่วนและคะแนนที่ผิดที่ใช้การตอบและการตรวจให้คะแนน 3 วิธี คือ วิธี 0 - 1 วิธีของ คูมบัสและวิธีของอนันต์ กลุ่มตัวอย่างที่ใช้เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 โรงเรียนปรางค์กู่ จังหวัดศรีสะเกษ จำนวน 30 คน กลุ่มตัวอย่างถูกสุ่มแยกออกเป็นกลุ่มทดลอง 23 กลุ่ม โดยทำแบบทดสอบ 2 ฉบับ ในเนื้อหาวิชาวิทยาศาสตร์ ระดับ ม.2 โดยทั้ง 2 ฉบับ มีเนื้อหาแตกต่างกัน กลุ่มที่ 1 ถามให้ตอบโดยวิธี 0 - 1 กลุ่มที่ 2 ถามให้ตอบโดยวิธีของคูมบัส และกลุ่มที่ 3 ถามให้ตอบโดยวิธีของอนันต์

ผลการศึกษาปรากฏว่า

1. วิธีของอนันต์ให้ค่าความเชื่อมั่นสูงกว่าวิธี - 1 และวิธีของคูมบัส
2. ค่าความเที่ยงตรงของแบบทดสอบที่ถามให้ตอบโดยวิธี 0 - 1 วิธีของคูมบัสและวิธีของอนันต์ พบว่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ
3. วิธีของอนันต์มีการเดาต่ำกว่าวิธี 0 - 1 และวิธีของคูมบัส
4. วิธีของอนันต์มีเปอร์เซนต์คะแนนที่ถูก ต่ำกว่าวิธีของคูมบัสแต่เปอร์เซนต์ของคะแนนบางส่วนมากกว่าวิธีของคูมบัส

จะเห็นได้ว่าวิธีของอนันต์สามารถพัฒนาประสิทธิภาพของข้อสอบวัดผลสัมฤทธิ์แบบเลือกตอบทั้งค่าความเชื่อมั่นและการเดา

และในปี! พ.ศ.2533 อนันต์ ศรีโสภา (2533) ได้เสนอผลงานวิจัยเปรียบเทียบประสิทธิภาพของแบบทดสอบที่ใช้วิธีการตอบ และการตรวจให้คะแนนแบบใหม่ (ที่ได้เสนอวิธีการตอบและการตรวจไว้แล้ว เมื่อ ปี พ.ศ.2516) แบบของคูมบัส และวิธีแบบ 0 - 1 โดยใช้แบบทดสอบวิชาวิจัยเบื้องต้น กลุ่มตัวอย่างที่ใช้เป็นนักศึกษาวิทยาลัยพยาบาลกรุงเทพปีที่ 4 จำนวน 144 คน แบ่งออกเป็น 3 กลุ่ม กลุ่มละ 48 คน แต่ละกลุ่มให้ตอบแบบวิธีอนันต์ วิ!!!!วิธีของคูมบัสและวิธี 0 - 1 ตามลำดับ

ผลการศึกษาพบว่า วิธีการตอบและการตรวจให้คะแนนแบบใหม่ มีการเดาน้อยกว่าวิธี 0 - 1 อย่างมีนัยสำคัญ แต่ไม่แตกต่างอย่างมีนัยสำคัญ เมื่อเปรียบเทียบกับวิธีของคูมบัส จำนวนข้อที่ถูกและข้อที่ผิดของวิธีแบบใหม่แตกต่างจากวิธี 0 - 1 อย่างมีนัยสำคัญแต่ไม่แตกต่างกับวิธีของคูมบัส ส่วนจำนวนข้อที่ถูกเพียงบางส่วนของวิธีแบบใหม่ และวิธีของคูมบัสไม่แตกต่างกันอย่างมีนัยสำคัญ จากข้อค้นพบเหล่านี้สรุปได้ว่า คะแนนการเดาคะแนนข้อถูก คะแนนข้อผิด และคะแนนข้อที่ถูกเพียงบางส่วนของวิธีการใหม่ไม่แตกต่างจากวิธี 0 - 1 อย่างมีนัยสำคัญ และพบว่าค่าความเชื่อมั่นของวิธีแบบใหม่สูงกว่าวิธีของคูมบัส และวิธี 0 - 1 อย่างมีนัยสำคัญ

เยาวลักษณ์ น้อยนาแสง (2535 : บทคัดย่อ) ได้ศึกษาเปรียบเทียบค่าความเชื่อมั่น ความเที่ยงตรง อำนาจจำแนก เปอร์เซนต์การเดา เปอร์เซนต์ความถนัดทั้งหมด เปอร์เซนต์ความถนัดบางส่วน และเปอร์เซนต์ไม่มีความถนัด ที่ใช้การตอบและตรวจให้คะแนน 3 วิธี คือ วิ-0-1 วิธีของคัมป์ส วิธีของอนันต์ กลุ่มตัวอย่างเป็นนักเรียนชายและหญิงชั้นมัธยมศึกษาปีที่ 3 ของโรงเรียนในสังกัดกรมสามัญศึกษา จังหวัดร้อยเอ็ด ปีการศึกษา 2534 ผลการศึกษาพบว่า

1. แบบทดสอบที่มีตัวเลือกถูกตัวเดียว มีค่าความเชื่อมั่นสูงกว่าแบบทดสอบที่มีตัวเลือกหลายตัว เมื่อตอบและตรวจด้วยวิธี 0-1 และเมื่อตอบและตรวจด้วยวิธีของคัมป์ส มีค่าความเที่ยงตรงสูงกว่าวิธีตรวจแบบ 0-1 ส่วนแบบทดสอบที่มีตัวเลือกถูกหลายตัว เมื่อตรวจด้วยวิธีของคัมป์ส และวิธีของอนันต์ มีค่าความเชื่อมั่นสูงกว่าวิธี 0-1

2. ค่าอำนาจจำแนกของแบบทดสอบวัดความถนัดทางการเรียน ที่มีตัวเลือกถูกตัวเดียว และมีตัวเลือกถูกหลายตัว เมื่อตอบและตรวจด้วยวิธี 0-1 วิธีของคัมป์ส และวิธีของอนันต์ แตกต่างกันอย่างมีนัยสำคัญทางสถิติ

3. วิธี 0-1 มีเปอร์เซนต์การเดาต่ำที่สุด วิธีของคัมป์สมีการเดาสูงสุด และเพศชายมีเปอร์เซนต์การเดาน้อยกว่าเพศหญิง

4. แบบทดสอบความถนัดทางการเรียน แบบที่มีตัวเลือกถูกตัวเดียว เมื่อตอบและตรวจด้วยวิธีของอนันต์ มีเปอร์เซนต์ความถนัดบางส่วนสูงกว่าวิธีของคัมป์ส ส่วนแบบทดสอบที่มีตัวเลือกถูกหลายตัว เมื่อตอบและตรวจด้วยวิธีของอนันต์ มีเปอร์เซนต์ความถนัดทั้งหมดและความถนัดบางส่วน สูงกว่าวิธีของคัมป์ส แต่เปอร์เซนต์ไม่มีความถนัดต่ำกว่าวิธีของคัมป์ส

พินิจ อุไรรักษ์ (2533 : บทคัดย่อ) ศึกษาผลของวิธีการให้คะแนนที่มีต่อคะแนนสอบ ค่าความเชื่อมั่น และค่าความเที่ยงตรง จากวิธีการให้คะแนน 4 วิธี คือ วิธีการให้น้ำหนักคะแนนรายข้อเท่ากัน วิธีการให้น้ำหนักคะแนนรายข้อต่างกันตามระดับความมั่นใจในการตอบ วิธีการให้น้ำหนักคะแนนรายข้อต่างกัน ตามค่าพารามิเตอร์ของข้อกระทงโดยสูตรของลอร์ด และวิธีการให้น้ำหนักคะแนนรายข้อต่างกัน ตามค่าความสามารถของผู้สอบ กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2534 จังหวัดบุรีรัมย์ จำนวน 1,011 คน เครื่องมือที่ใช้เป็นแบบทดสอบวัดผลสัมฤทธิ์ วิชาวิทยาศาสตร์ ผลการวิจัยพบว่า

1. ความสัมพันธ์ของคะแนนสอบ ระหว่างวิธีการให้คะแนนทั้ง 4 วิธี มีค่าตั้งแต่ .941 ถึง .983

2. ค่าความเชื่อมั่นของแบบทดสอบพบว่า ค่าความเชื่อมั่นของแบบทดสอบจากวิธีการให้น้ำหนักคะแนนรายข้อเท่ากัน วิธีการให้น้ำหนักคะแนนรายข้อต่างกัน ตามระดับความมั่นใจ

วิธีการให้น้ำหนักคะแนนรายข้อต่างกัน ตามค่าพารามิเตอร์ของข้อกระทงโดยสูตรของลอร์ด และวิธีการให้น้ำหนักคะแนนรายข้อต่างกัน ตามค่าความสามารถของผู้ตอบ มีค่าเท่ากับ .705

3. วิธีการให้คะแนนน้ำหนักรายข้อต่างกัน ตามระดับความมั่นใจในการตอบ วิธีการให้น้ำหนักคะแนนรายข้อต่างกัน ตามค่าพารามิเตอร์ของข้อกระทงโดยสูตรของลอร์ด และวิธีการให้น้ำหนักคะแนนรายข้อต่างกัน ตามค่าความสามารถของผู้ตอบ โดยเฉลี่ยแล้วให้ค่าความเชื่อมั่นของแบบทดสอบสูงกว่าวิธีการให้น้ำหนักคะแนนรายข้อเท่ากัน

วัชรภรณ์ จิตธมาศ (2538 : 72 - 87) ศึกษาเปรียบเทียบความยาก ค่าอำนาจจำแนก และค่าความเชื่อมั่นของแบบทดสอบเลือกตอบชนิดถูกผิดหลายตัวเลือก ที่มีจำนวน 4 ตัวเลือกและ 5 ตัวเลือก โดยใช้วิธีการตรวจให้คะแนนที่ต่างกัน 5 วิธี คือ แบบ 0-1 แบบ MTF แบบ COUNT 2 แบบ COUNT 3 และแบบ CFG เครื่องมือที่ใช้เป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนกลุ่มสร้างเสริมประสบการณ์ชีวิตระดับชั้นประถมศึกษาปีที่ 5 จำนวน 4 ฉบับ ความยาวฉบับละ 30 ข้อ ทำการทดสอบกับกลุ่มตัวอย่างซึ่งเป็นนักเรียนชั้นประถมศึกษาปีที่ 5 สังกัดสำนักงานการประถมศึกษาจังหวัดอุบลราชธานีในปีการศึกษา 2538 จำนวน 800 คน ได้ผลวิจัยดังนี้

1. แบบทดสอบถูกผิดหลายตัวเลือกเมื่อให้คะแนนโดยวิธีต่างกัน มีความยากแตกต่างกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยพบว่าเมื่อตรวจให้คะแนนแบบ 0-1 มีความยากสูงสุด และเมื่อตรวจด้วยวิธี MTF มีค่าความยากต่ำสุด

2. แบบทดสอบถูกผิดหลายตัวเลือกที่มีจำนวนตัวเลือก 4 ตัวเลือกเมื่อให้คะแนนแบบ CFG และแบบ 0-1 พบว่ามีค่าอำนาจจำแนกสูงกว่า แบบ MTF และพบว่าเมื่อให้คะแนนแบบ CFG พบว่ามีค่าอำนาจจำแนกสูงกว่าแบบ COUNT2 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนแบบทดสอบถูกผิดหลายตัวเลือกที่มีจำนวน 5 ตัวเลือก เมื่อให้คะแนนทั้ง 5 วิธี พบว่าค่าอำนาจจำแนกแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

3. แบบทดสอบถูกผิดหลายตัวเลือกที่มีจำนวนตัวเลือก 4 ตัวเลือกเมื่อให้คะแนนแบบ 0-1 พบว่ามีความเชื่อมั่นสูงกว่าแบบทดสอบถูกผิดหลายตัวเลือกที่มีจำนวนตัวเลือก 5 ตัวเลือก แต่เมื่อให้คะแนนตามวิธี MTF, COUNT2, COUNT3 และ CFG พบว่ามีความเชื่อมั่นต่ำกว่าอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 เมื่อทำการเปรียบเทียบเฉพาะความเชื่อมั่นของแบบทดสอบถูกผิดหลายตัวเลือกที่มีจำนวนตัวเลือก 4 ตัวเลือกเมื่อให้คะแนนทั้ง 5 วิธีพบว่า แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ และเมื่อศึกษากับแบบทดสอบถูกผิดหลายตัวเลือกที่มีจำนวนตัวเลือก 5 ตัวเลือก พบว่าค่าความเชื่อมั่นของแบบทดสอบเมื่อตรวจให้คะแนนแบบ 0-1 มีค่าน้อยกว่าเมื่อตรวจให้คะแนนแบบอื่นๆ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 สำหรับค่าความเชื่อมั่นของแบบทดสอบที่

ตรวจให้คะแนนแบบ MTF พบว่ามีค่าน้อยกว่าเมื่อตรวจให้คะแนนแบบ COUNT2 และ COUNT3 และพบว่าความเชื่อมั่นของแบบทดสอบที่ตรวจให้คะแนนแบบ CFG มีค่าน้อยกว่าเมื่อตรวจให้คะแนนแบบ COUNT2 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้ผู้วิจัยมีวัตถุประสงค์สำคัญคือศึกษาอิทธิพลของการให้คะแนนและการคัดเลือกข้อสอบเข้าฉบับ เพื่อประโยชน์ในการศึกษารูปแบบสมการโครงสร้างที่เหมาะสมในการพัฒนาแบบทดสอบเลือกตอบ รายละเอียดของวิธีดำเนินการวิจัยมีดังนี้

ขอบเขตของการวิจัย

ประชากรและกลุ่มตัวอย่าง

ประชากรในการวิจัยครั้งนี้เป็นนักเรียนระดับชั้นประถมศึกษาปีที่ 6 สังกัดสำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ ปีการศึกษา 2539 จำนวน 794,740 คน (กองนโยบายและแผน สำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ. 2538: 9-13)

กลุ่มตัวอย่างที่ใช้เป็นแหล่งข้อมูลเพื่อการวิจัยครั้งนี้ได้จากการเลือกจังหวัดเป้าหมายโดยให้ความน่าจะเป็นของการเลือกเป็นปฏิภาคกับขนาดของข้อมูลที่ต้องการสำรวจ (Sampling with probability proportional to size ; PPS) และกำหนดกลุ่มตัวอย่างโดยการสุ่มตัวอย่างหลายขั้นตอน (Multi-stage sampling) ซึ่งมีขั้นตอนดังนี้

1. กำหนดขนาดโดยประมาณ จากสูตรของ ทาโร ยามาเน (Yamane. 1973 :727) เพื่อให้ทราบขนาดของกลุ่มตัวอย่างที่ยอมรับได้ในเชิงของการวิจัยจากสูตร

$$n = \frac{N}{1 + Ne^2}$$

เมื่อ n คือ ขนาดกลุ่มตัวอย่าง (Sample size)

N คือ จำนวนประชากร

e^2 คือ ความคลาดเคลื่อนที่ยอมให้เกิดขึ้น ซึ่งในการวิจัยครั้งนี้ผู้วิจัยกำหนดไว้ที่ 5% หรือระดับนัยสำคัญที่ .05

จากการคำนวณตามสูตร ได้ค่าประมาณของขนาดกลุ่มตัวอย่างประมาณ

400 คน

2. กำหนดขนาดกลุ่มตัวอย่างที่เหมาะสมกับการวิเคราะห์ตามกรอบทฤษฎี CTT และ IRT

ในการศึกษาวิจัยครั้งนี้เพื่อให้ครอบคลุมเงื่อนไขต่าง ๆ ที่สนใจ ได้แก่การวิเคราะห์ข้อสอบตามแนวของ CTT และ IRT (ราซซีโมเดล และพาร์เซียลเครดิทโมเดล) การวิเคราะห์องค์ประกอบความลำเอียงเนื่องจากตัวแปรเพศ ผู้วิจัยจึงกำหนดกลุ่มตัวอย่างในการศึกษาจำนวนประมาณ 1200 คนโดยให้ครอบคลุมทั้ง 12 เขตการศึกษา และดำเนินการเลือกจังหวัดเป้าหมาย และสุ่มตัวอย่างตามขั้นตอนต่อไปนี้

1. กำหนดจังหวัดสำหรับเก็บข้อมูล โดยใช้วิธีการ PPS แบบสุ่มสลาก (The Lottery method of PPS selection) (นิยม ปุราคำ.2520:22-23 ; Ross in Keeves,1988:525-536) ซึ่งมีลำดับวิธีการดังนี้

1. จัดเรียงจังหวัดตามเขตการศึกษาทั้ง 12 เขตตามการกำหนดของกระทรวงศึกษาธิการ ยกเว้นกรุงเทพมหานคร โดยให้จังหวัดนครปฐมเป็นลำดับที่หนึ่งและเรียงลำดับของจังหวัดต่อไปจนถึงลำดับที่ 76 ได้แก่จังหวัดตราดเป็นจังหวัดสุดท้าย

2. หาจำนวนนักเรียนสะสมในแต่ละจังหวัดเพื่อเปรียบเทียบกับค่าคำนวณ (ยกเว้นกรุงเทพมหานคร)

3. กำหนดจำนวนจังหวัดที่ต้องการ ซึ่งในการศึกษาวิจัยครั้งนี้กำหนดเขตการศึกษาละ 1 จังหวัด ดังนั้นรวมจังหวัดทั้งสิ้น 12 จังหวัดและกรุงเทพมหานคร

4. กำหนดระยะช่วง (Interval) ของการสุ่มจากสูตร

$$\frac{I}{n} = M$$

เมื่อ I คือ ระยะช่วงของการสุ่ม

M คือ ยอดของจำนวนนักเรียนสะสมของจังหวัดสุดท้าย (ไม่รวม กรุงเทพมหานคร)

n คือ จำนวนจังหวัดที่ต้องการ

5. กำหนดจำนวนนักเรียนเริ่มต้นของการสุ่ม (Random start ; R) ซึ่ง $R < I$ และในการวิจัยครั้งนี้ใช้จำนวนนักเรียนของจังหวัดในเขตการศึกษา 1 ที่ได้จากการสุ่มอย่างง่ายแทนค่า R และหาจังหวัดต่อไปโดยวิธีการต่อไปนี้

จังหวัดที่ 1 มีจำนวนนักเรียนสะสมเท่ากับ R

จังหวัดที่ 2 มีจำนวนนักเรียนสะสมเท่ากับ $R + 1$

จังหวัดที่ 3 มีจำนวนนักเรียนสะสมเท่ากับ $R + 21$

.....

.....

.....

จังหวัดที่ 12 มีจำนวนนักเรียนสะสมเท่ากับ $R + 111$

6. กำหนดช่วงของจำนวนนักเรียน

7. เปรียบเทียบจำนวนนักเรียนที่เหมาะสมจากการคำนวณว่าตกอยู่ในช่วงของจังหวัดใด เลือกจังหวัดนั้น

8. ในกรณีเขตการศึกษาใดมีจำนวนจังหวัดมากกว่า 1 จังหวัดใช้วิธีเปรียบเทียบความแตกต่างระหว่างจำนวนที่คำนวณได้กับจำนวนนักเรียนสะสมของแต่ละจังหวัด โดยเลือกจังหวัดที่มีค่าความแตกต่างน้อยกว่าเพียง 1 จังหวัด

9. ในกรณีที่ไม่มีจังหวัดใดในเขตการศึกษาได้รับการสุ่ม ใช้วิธีการสุ่มอย่างง่ายเฉพาะจังหวัดในเขตการศึกษานั้นๆ

10. คำนวณขนาดกลุ่มตัวอย่างเป้าหมายที่เหมาะสมในแต่ละจังหวัดจากสูตร

$$n_h = \frac{n_h}{N} n_o$$

เมื่อ n_h คือ ขนาดกลุ่มตัวอย่างเป้าหมายที่เหมาะสมของจังหวัด h

N_h คือ จำนวนนักเรียนชั้นประถมศึกษาปีที่ 6 ในจังหวัดกลุ่มตัวอย่าง

N คือ จำนวนนักเรียนชั้นประถมศึกษาปีที่ 6 สังกัดสำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ ปีการศึกษา

2539

n_o คือ ขนาดกลุ่มตัวอย่างเป้าหมายทั้งหมดที่ต้องการ

11. เลือกโรงเรียน / ห้องเรียนกลุ่มตัวอย่างในแต่ละจังหวัดโดยวิธีสุ่มอย่างง่าย

จากวิธีการสุ่มดังกล่าวได้กลุ่มตัวอย่างจำนวน 1,304 คน เมื่อนำไปแทนค่าในสูตรของ ทาโร ยามาเน พบว่าให้ค่าความคลาดเคลื่อนที่ยอมรับได้ (α) ณ ระดับ .03 ดังตาราง 8

ตาราง 8 จำนวนนักเรียนกลุ่มตัวอย่าง จำแนกตามเพศ โรงเรียน อำเภอ จังหวัด และเขตการศึกษา

เขตการศึกษา	จังหวัด	อำเภอ	โรงเรียน	จำนวนนักเรียน		
				ชาย	หญิง	รวม
1	นครปฐม	เมือง	วัดไผ่ล้อม	41	34	75
2	ยะลา	เมือง	อนุบาลยะลา	35	35	70
3	สงขลา	เมือง	วิเชียรชม	64	65	129
4	ภูเก็ต	เมือง	อนุบาลภูเก็ต	20	20	40
5	กาญจนบุรี	พนมทวน	วัดดอนเจดีย์	17	12	29
		ท่าม่วง	เขาดินวิทยาการ	9	14	23
		ท่าม่วง	วัดกร่างทองราษฎร์	11	14	25
6	สิงห์บุรี	เมือง	วัดพรหมสาคร	18	22	40
7	กำแพงเพชร	ชาณุวรลักษณบุรี	บ้านเขาพริกไทย	11	8	19
			บ้านปางมะค่า	6	9	15
			บ้านวังน้ำ	20	27	47
8	เชียงใหม่	พร้าว	ชุมชนบ้านโป่ง	10	9	19
			ชุมชนสหกรณ์นิคมฯ	14	14	28
			บ้านแจ่งกูเรือ	53	53	106
9	ขอนแก่น	กระนวน	ชุมชนบ้านหนองโก	100	88	188
10	กาฬสินธุ์	ห้วยเม็ก	คำใหญ่วิทยา	38	31	69
		กมลาไสย	ชุมชนบ้านโพนงาม	24	28	52
11	สุรินทร์	เมือง	เมืองสุรินทร์	90	101	191

ตาราง 8 (ต่อ)

เขต การ ศึกษา	จังหวัด	อำเภอ	โรงเรียน	จำนวนนักเรียน		
				ชาย	หญิง	รวม
12	ระยอง	เมือง	บ้านวังหิน	34	17	51
			บ้านมาบตาพุด	21	14	35
กทม.			วัดหงส์รัตนาราม	28	25	53
รวม				664	640	1304

ตัวแปรในการวิจัย

ตัวแปรในการศึกษาวิจัยครั้งนี้จัดแบ่งประเภทดังนี้คือ

1. ตัวแปรอิสระ / ตัวแปรพยากรณ์

1.1 รูปแบบการให้คะแนน (แบบปกติ, แบบ 2-SR)

1.2 วิธีการคัดเลือกข้อสอบ โดยใช้ค่าพารามิเตอร์ และคุณสมบัติบางประการ
ของข้อสอบ ได้แก่

1.2.1 ค่าความยาก

1.2.2 ค่าอำนาจจำแนก

1.2.2.1 ค่าสัมประสิทธิ์การจำแนก (สหสัมพันธ์ระหว่างคะแนน

รายข้อกับคะแนนรวม)

1.2.2.2 ค่าดัชนีการจำแนก

1.2.3 ความเป็นมิติเดียว

1.2.4 ความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ

1.2.5 ความยากแบบราชคโหมเดล

1.2.6 ความเหมาะสมของข้อสอบกับราชคโหมเดล

2. ตัวแปรตาม/ตัวแปรเกณฑ์

2.1 ค่าความยากรายข้อ

2.2 ค่าอำนาจจำแนกรายข้อ

2.2.1 ค่าสัมประสิทธิ์การจำแนก (สหสัมพันธ์ระหว่างคะแนนรายข้อกับ
คะแนนรวม)

2.2.2 ค่าดัชนีการจำแนก

2.3 คำนำน้หนักองค์ประกอบ

2.4 ค่าสัมประสิทธิ์สหสัมพันธ์แบบ Contingency (ความลำเอียงของข้อสอบ)

2.5 ค่าความยากแบบราซซีโมเดล

2.6 ค่า t - test ที่แสดงความเหมาะสมของข้อสอบกับราซซีโมเดล

2.7 ความเชื่อมั่นของแบบทดสอบรายข้อ และรายฉบับ

2.8 สารสนเทศของข้อสอบ และแบบทดสอบ

2.9 คะแนนผลการสอบ

2.10 คะแนนความสามารถของผู้สอบ

เครื่องมือการวิจัย

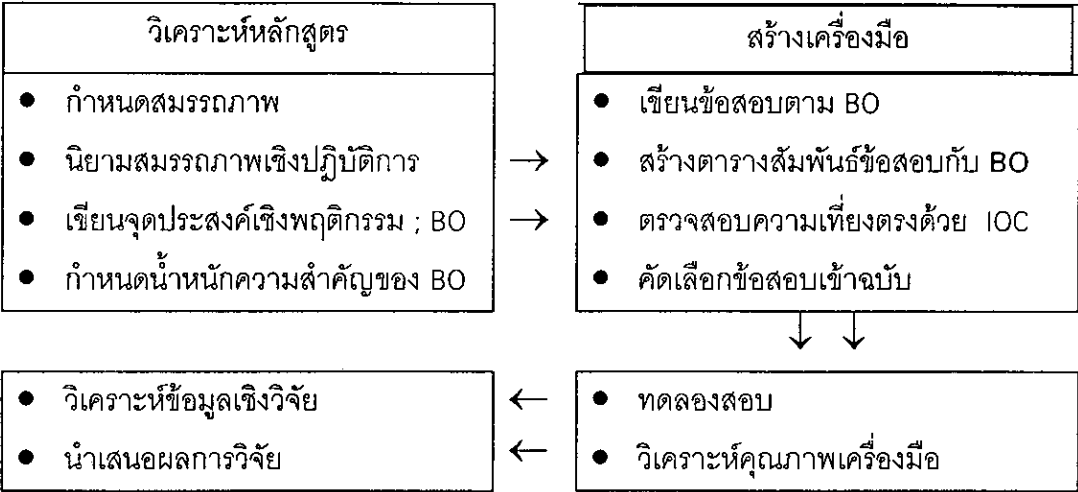
เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบเลือกตอบ 4 ตัวเลือก มีคำตอบถูก
เพียงคำตอบเดียว จำนวน 2 ฉบับ แต่ละฉบับมีความยาว 40 ข้อ คือ

1. แบบทดสอบวัดความสามารถในการอ่านจับใจความ (กลุ่มทักษะภาษาไทย)
ค่าความเชื่อมั่น .66

2. แบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ (กลุ่ม
ทักษะคณิตศาสตร์) มีความเชื่อมั่น .77

วิธีการดำเนินงาน

ในการดำเนินงานศึกษาวิจัยครั้งนี้ สรุปการดำเนินงานดังภาพประกอบ 16 ดังนี้



ภาพประกอบ 16 แผนภาพการดำเนินงานเก็บรวบรวม และวิเคราะห์ผลการวิจัย

จากแผนภาพตามภาพประกอบ 16 แสดงขั้นตอนการดำเนินงานวิจัยตามลำดับ ดังนี้

1. วิเคราะห์หลักสูตร กำหนดนิยามเชิงปฏิบัติการของสมรรถภาพที่วัดและเขียนจุดประสงค์เชิงพฤติกรรม
2. กำหนดน้ำหนักความสำคัญของจุดประสงค์เชิงพฤติกรรมโดยพิจารณาจากความเห็นของผู้เชี่ยวชาญในแต่ละสมรรถภาพ
3. จัดสร้างข้อสอบตามจุดประสงค์เชิงพฤติกรรมประมาณฉบับละ 60 ข้อ
4. จัดสร้างตารางสัมพันธ์ระหว่างจุดประสงค์เชิงพฤติกรรมกับข้อสอบเพื่อตรวจสอบความเที่ยงตรงเชิงเนื้อหาและความเที่ยงตรงเชิงโครงสร้างโดยผู้เชี่ยวชาญ และพิจารณาค่าความสอดคล้องของความเห็นจากดัชนี IOC (Index of congruence) ที่เสนอโดยโรวินลลี และแฮมเบิลตัน (Oosterhof. 1994:54-60 ; citing Rovinelli and Hambleton. 1977) จากสูตร

$$IOC_{ik} = \frac{N}{2N - 2} (\mu_k - \mu)$$

- เมื่อ IOC_k คือ ค่าดัชนีความสอดคล้องของข้อสอบข้อที่ i กับจุดประสงค์ข้อที่ k
- N คือ จำนวนของจุดประสงค์ทั้งหมด
- μ_k คือ ค่าระดับเฉลี่ยของการตัดสินข้อสอบข้อที่ i ของจุดประสงค์ข้อ k
- μ คือ ค่าระดับเฉลี่ยของการตัดสินข้อสอบข้อที่ i จากจุดประสงค์ทั้งหมด

5. คัดเลือกข้อสอบที่มีค่า IOC เรียงลำดับจากสูงมาหาต่ำเข้าฉบับๆ ละ 40 ข้อ
6. เก็บรวบรวมข้อมูลเพื่อศึกษารูปแบบที่เหมาะสมในการพัฒนาแบบทดสอบ
7. ปรับปรุงข้อสอบ
8. เก็บรวบรวมข้อมูลเพื่อทดลองรูปแบบ
9. วิเคราะห์สรุปผลเขียนรายงาน

การวิเคราะห์และจัดกระทำข้อมูล

1. เมื่อนำแบบทดสอบแต่ละฉบับไปทดสอบกับกลุ่มตัวอย่างให้คะแนนตามขั้นตอนดังนี้
 - 1.1 พิจารณาการเลือกตอบในส่วนที่ 1 โดยที่
 - 1.1.1 นับจำนวนตัวเลือกที่ผู้สอบกาเครื่องหมาย /
 - 1.1.2 พิจารณาตัวเลือกที่ผู้สอบเลือกตอบว่ามีตัวถูกรวมอยู่หรือไม่
- เหล่านั้นหรือไม่ ถ้ามีให้ 1 คะแนน ถ้าไม่มีหรือไม่เลือกตอบให้ 0
- 1.2 พิจารณาการตอบในส่วนที่ 2 และให้คะแนนตามปกติคือ ตอบถูกได้ 1 คะแนน ตอบผิดหรือไม่ตอบหรือเลือกคำตอบมากกว่า 1 คำตอบให้ 0
 - 1.3 นำคะแนนทั้งสองส่วน แทนค่าในสูตรดังนี้

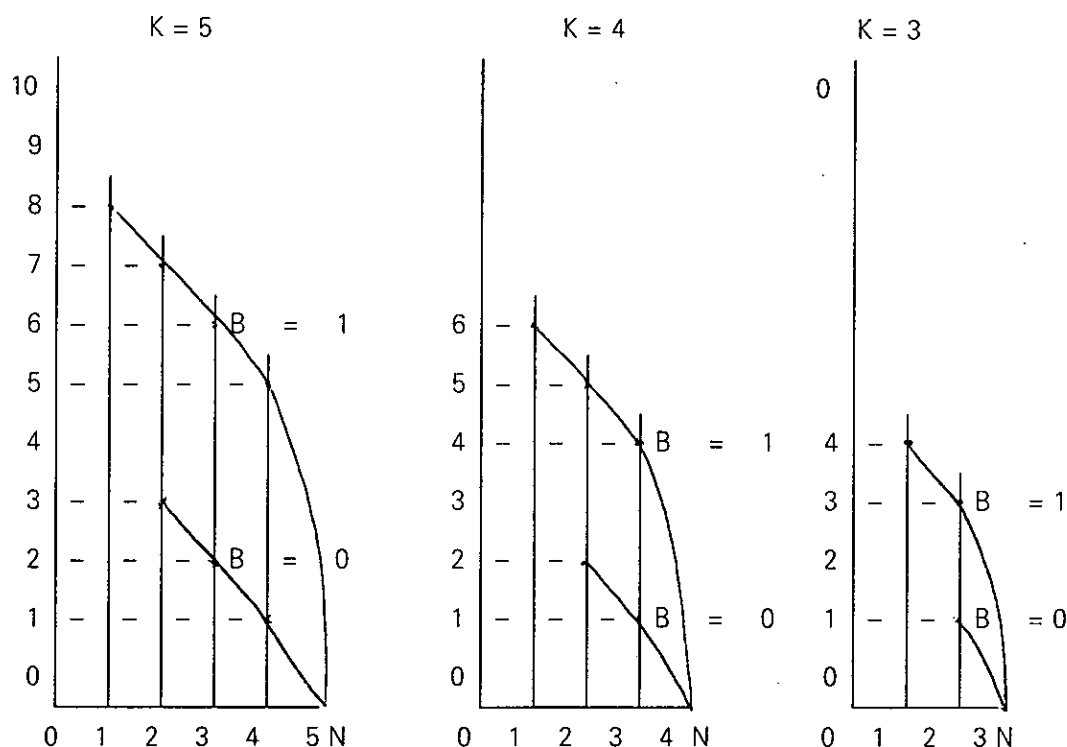
$$X = \left| \frac{(K-N)A + \frac{(K-1)(N-K)B}{N-K}}{C} \right|$$

- เมื่อ
- X คือ คะแนนรายข้อ
- K คือ จำนวนตัวเลือกทั้งหมด
- N คือ จำนวนตัวเลือกที่ผู้สอบเลือกตอบในส่วนของ 1
- A คือ คะแนนในส่วนที่ 1
- B คือ คะแนนในส่วนที่ 2
- C คือ ความสอดคล้องของการตอบในส่วนที่ 1 และส่วนที่ 2 โดยที่
- C = 1 เมื่อตัวเลือกในส่วนที่ 2 ปรากฏในส่วนที่ 1
- C = 0 เมื่อตัวเลือกในส่วนที่ 2 ไม่ปรากฏในส่วนที่ 1 หรือ เลือกคำตอบเพียงส่วนใดส่วนหนึ่ง

ในกรณีนี้อาจให้คะแนนโดยใช้ตารางคะแนนตามภาพประกอบ 18 หรือรูปตามภาพประกอบ 19 โดยใช้สัญลักษณ์เช่นเดียวกับการใช้สูตร

		ส่วนที่ (A)						
		K	5		4		3	
			A = 1	A = 0	A = 1	A = 0	A = 1	A = 0
		N	5 4 3 2 1 0	5 4 3 2 1 0	4 3 2 1 0	4 3 2 1 0	3 2 1 0	3 2 1 0
C = 1	ส่วนที่ 1	1	0 5 6 7 8 0		0 4 5 6 0		0 3 4 0	
	2 (B)	0	0 1 2 3 - 0	0 0 0 0 0 0	0 1 2 - 0	0 0 0 0 0	0 1 - 0	0 0 0 0
C = 0	ส่วนที่ 1	1	0 0 0 0 0 0	0 0 0 0 0 0	0 0 0 0 0	0 0 0 0 0	0 0 0 0	0 0 0 0
	2(B)	0	0 0 0 0 0 0	0 0 0 0 0 0	0 0 0 0 0	0 0 0 0 0	0 0 0 0	0 0 0 0

ภาพประกอบ 17 วิธีการให้คะแนนการตอบแบบ 2-SR



ภาพประกอบ 18 แผนภาพการให้คะแนนแบบ 2 - SR (5 , 4 และ 3 ตัวเลือก)

สำหรับการศึกษาเกี่ยวกับการให้คะแนนแบบปกติจะใช้ผลการตอบและตรวจให้คะแนนในส่วนที่ 2 เท่านั้น

2. ประมาณค่าสถิติของข้อสอบตามแนวของ CTT

2.1 ประมาณค่าความยากและค่าอำนาจจำแนกรายข้อ ในการวิจัยครั้งนี้จะใช้เทคนิค 25% ตามรูปแบบการให้คะแนน 2 วิธีคือ

2.1.1 สำหรับแบบทดสอบที่ให้คะแนนแบบ 0-1

$$P = \frac{H + L}{N}$$

$$D = \frac{H - L}{N/2}$$

เมื่อ P คือ ค่าความยากของข้อสอบ

D คือ ค่าดัชนีการจำแนกของข้อสอบ

H คือ จำนวนคนในกลุ่มสูง (25%) ที่ตอบถูก

L คือ จำนวนคนในกลุ่มต่ำ (25%) ที่ตอบถูก

N คือ จำนวนผู้สอบที่นำมาวิเคราะห์

(Gulliksen. 1950 : 373 ; citing Cook. 1932, Hirshstien and Finch. 1932, Guilford. 1936b, Kelly. 1939, Davis. 1946)

2.1.2 สำหรับแบบทดสอบที่ให้คะแนนแบบ 2-SR

$$P = \frac{\sum H + \sum L - (2N \text{ Score}_{\min})}{2N(\text{Score}_{\max} - \text{Score}_{\min})}$$

$$D = \frac{\sum H - \sum L}{N(\text{Score}_{\max} - \text{Score}_{\min})}$$

เมื่อ P คือ ค่าความยากของข้อสอบ

D คือ ค่าดัชนีการจำแนกของข้อสอบ

$\sum H$ คือ คะแนนรวมของกลุ่มสูง (25%) ในแต่ละข้อ

$\sum L$ คือ คะแนนรวมของคนกลุ่มต่ำ (25%) ในแต่ละข้อ

N คือ 50% ของจำนวนคนที่นำมาวิเคราะห์

$\text{Score}_{\max}$ คือ คะแนนสูงสุดของผู้สอบแต่ละข้อ

$\text{Score}_{\min}$ คือ คะแนนต่ำสุดของผู้สอบแต่ละข้อ (Mehrens and Lehmann.

1984 : 198-199 ; citing Whitney and Saber. 1970)

เนื่องจากค่าความยากที่น่าจะมีผลต่อค่าความเชื่อมั่นมากที่สุดคือ .5 ดังนั้นการ
ศึกษาอิทธิพลของค่าความยากในการวิจัยครั้งนี้จึงศึกษาความยากในรูปของตัวแปรจำลอง
(Dummy variable) โดยจัดเป็น 2 ระดับดังนี้ (เมื่อ p คือ ระดับความยากรายข้อ)

$$P = 0 \text{ เมื่อ } .20 < p ; p > .80$$

$$P = 1 \text{ เมื่อ } .20 \leq p \leq .80$$

และได้ศึกษาอิทธิพลของดัชนีการจำแนกในรูปของตัวแปรจำลองเช่นเดียวกันโดย
ประยุกต์จากข้อเสนอของอีเบล (Ebel) เมื่อปี 1965 (Crocker and ALgina.1986 : 315 ; citing
Ebel. 965) ที่เสนอไว้ดังนี้

ถ้า $D < .19$ ข้อสอบควรตัดออกหรือทบทวนอย่างมาก

$.20 < D < .29$ ข้อสอบพอใช้ได้และควรมีการทบทวน

$.30 < D < .39$ ควรมีการทบทวนบ้างหรือไม่ต้องทบทวนก็ได้

$D > .40$ เป็นข้อสอบที่มีค่าอำนาจจำแนกเหมาะสม

จากแนวคิดดังกล่าว ในการศึกษาวิจัยครั้งนี้ได้แบ่งกลุ่มตัวแปรจำลองของดัชนีการจำแนกดังนี้

$$D = 0 \text{ เมื่อ } D < .20$$

$$D = 1 \text{ เมื่อ } D \geq .20$$

นอกจากนั้นได้ศึกษาค่าอำนาจจำแนกรายข้อจากค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนรวมของผู้สอบจากสูตรดังนี้

1. สำหรับแบบทดสอบที่ให้คะแนนแบบ 0-1

$$r = r_{pbis} = \frac{M_T - M_W}{S_T} \sqrt{pq} = \frac{M_T - M_I}{S_T} \sqrt{pq}$$

เมื่อ r คือ ค่าอำนาจจำแนกของข้อสอบซึ่งได้จากค่าสัมประสิทธิ์สหสัมพันธ์แบบพ้อยท์ไบซีเรียลระหว่างคะแนนรายข้อกับคะแนนรวม

M_T คือ คะแนนรวมเฉลี่ยเฉพาะผู้สอบที่ตอบข้อสอบข้อนั้นได้ถูกต้อง

M_W คือ คะแนนรวมเฉลี่ยเฉพาะผู้สอบที่ตอบข้อสอบข้อนั้นผิด

M_T คือ คะแนนรวมเฉลี่ยของผู้สอบทั้งหมด

S_T คือ ส่วนเบี่ยงเบนมาตรฐานของคะแนนรวมของผู้สอบทั้งหมด

p คือ สัดส่วนการตอบถูก

q คือ $1 - p$

(Crocker and Algina.1986 : 313-318 ; Wood in Keeves. 1988 : 378 ; Kline. 1993 : 71)

2. สำหรับแบบทดสอบให้คะแนนแบบ 2-SR

$$r = r_{xy} = \frac{N \sum XY - \sum X \sum Y}{\sqrt{N \sum X^2 - (\sum X)^2} \sqrt{N \sum Y^2 - (\sum Y)^2}}$$

เมื่อ r คือ ค่าอำนาจจำแนกของข้อสอบ ซึ่งได้จากค่าสัมประสิทธิ์

สหสัมพันธ์แบบเพียร์สัน (Pearson product moment correlation)

N คือ จำนวนผู้สอบทั้งหมด

X คือ คะแนนรายข้อของผู้สอบแต่ละคน

Y คือ คะแนนรวมของผู้สอบแต่ละคน

(Kline. 1993 : 572)

เนื่องจากค่าสัมประสิทธิ์สหสัมพันธ์มีการกระจายตั้งแต่ -1.00 ถึง +1.00 จึงทำให้ค่าอำนาจจำแนกของข้อสอบมีทั้งค่าบวกและค่าลบ ดังนั้นในการศึกษาวิจัยครั้งนี้จึงศึกษาสัมประสิทธิ์การจำแนกในรูปของตัวแปรจำลอง โดยจัดช่วงของค่าอำนาจจำแนกดังนี้

$$r = 0 \text{ เมื่อ } r < .20$$

$$r = 1 \text{ เมื่อ } r \geq .20$$

2.2 ค่าความเชื่อมั่นรายข้อ (Item reliability index)

ค่าความเชื่อมั่นรายข้อเป็นการศึกษาฟังก์ชันของความแปรปรวนของคะแนนรายข้อและความสัมพันธ์ระหว่างคะแนนรายข้อกับเกณฑ์ภายใน (คะแนนรวม) ไปพร้อม ๆ กัน โดยประมาณค่าจากสูตร $s_{r_{ix}}$ เมื่อ s_i คือส่วนเบี่ยงเบนมาตรฐานรายข้อ และ r_{ix} คือ ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนรวม (Gulliksen. 1950 : 377 ; Crocker and Algina. 1986 : 320)

จากสูตรดังกล่าว ในการศึกษาวิจัยครั้งนี้ได้ประมาณค่าความเชื่อมั่นรายข้อโดยประยุกต์สูตรดังนี้

$$r_{ii} = s_{r_{ix}}$$

เมื่อ r_{ii} คือ ดัชนีความเชื่อมั่นรายข้อ

r_{ix} คือ สัมประสิทธิ์การจำแนกซึ่งเป็นดัชนีแสดงความสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนจริง

s_i คือ ส่วนเบี่ยงเบนมาตรฐานรายข้อซึ่งประมาณค่าได้ 2 กรณีคือ

1. กรณีให้คะแนนแบบ 0-1

$$s_i = p_i q_i$$

เมื่อ p_i คือ สัดส่วนของคนตอบถูก หรือคะแนนเฉลี่ย หรือค่าความยากของข้อที่ i

q_i คือ $1 - p_i$

(Gulliksen. 1950 : 376 ; Crocker and Algina. 1986 : 311-312)

2. กรณีให้คะแนนแบบ 2-SR

$$s_i = \frac{\sqrt{N \sum X_i^2 - (\sum X_i)^2}}{\sqrt{N}}$$

เมื่อ X_i คือ คะแนนรายข้อของผู้สอบแต่ละคน

N คือ จำนวนผู้สอบทั้งหมด

(Gulliksen. 1950 : 386)

2.3 ค่าความเชื่อมั่น และค่าความคลาดเคลื่อนมาตรฐานในการวัดของแบบทดสอบ ประเมินค่าได้จากสูตร

$$r_{tt} = \frac{k}{k-1} \left| 1 - \frac{\sum s_i^2}{s_T^2} \right|$$

$$SEM = s_T \sqrt{(1 - r_{tt})}$$

เมื่อ r_{tt} คือ ค่าความเชื่อมั่นของแบบทดสอบ

SEM คือ ค่าความคลาดเคลื่อนมาตรฐานในการวัด

k คือ จำนวนข้อ

3. ประเมินค่าสถิติของข้อสอบ ตามแบบราชคโมเดล

หลังจากตรวจให้คะแนนข้อสอบทั้ง 2 วิธีแล้ว ได้ดำเนินการต่อไปดังนี้

1. คัดเลือกผู้สอบที่สอบได้คะแนนเต็มและคะแนน 0 ออก
2. คัดเลือกข้อสอบที่มีคนตอบถูกทั้งหมด หรือตอบผิดทั้งหมด หรือผู้สอบทุกคน

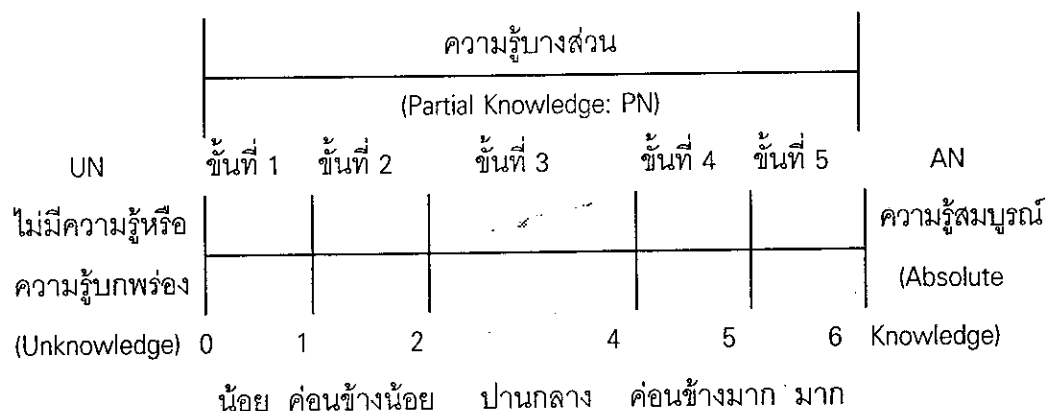
ได้คะแนนเท่ากัน (เฉพาะกรณีแบบ 2-SR) ออกจากการวิเคราะห์

3.1 การประมาณค่าความยากของข้อสอบ (β_i) และความสามารถของผู้สอบ (θ_i) ในกรณีให้คะแนนแบบ 0-1 ประเมินค่าโดยใช้โปรแกรม BIGSTEPS (Model R)

3.2 การประมาณค่าความยากของข้อสอบ (β_i) และความสามารถของผู้สอบ (θ_i) ในกรณีให้คะแนนแบบ 2 - SR

เนื่องจากการให้คะแนนแบบ 2-SR แตกต่างไปจากแบบปกติถูก 1 ผิด 0 กล่าวคือ คะแนนมีหลายระดับตั้งแต่ 0 , 1... ผู้ที่ได้คะแนน 0 คือผู้ตอบผิดซึ่งแสดงถึงความไม่รู้หรือมีความรู้บกพร่อง ส่วนผู้ที่ได้คะแนนตั้งแต่ 1 ขึ้นไปแสดงถึงความเป็นผู้มีความรู้ตั้งแต่รู้เป็นบางส่วน (Partial knowledge) จนถึงเป็นผู้มีความรู้อย่างสมบูรณ์ (Absolute knowledge) ซึ่งเป็นการอธิบายความสามารถของผู้สอบมากกว่าเพียงแค่อธิบายว่าตอบถูกเท่านั้น ดังนั้นในการประมาณค่าความยากของข้อสอบ จึงหมายความว่าค่าโอกาสการได้คะแนนของผู้สอบที่

ระดับความสามารถต่างๆ ซึ่งแต่ละระดับแสดงถึงปริมาณความรู้ ในการศึกษาวิจัยครั้งนี้ เครื่องมือ วิจัยคือแบบทดสอบเลือกตอบ 4 ตัวเลือกที่มีการกระจายของคะแนนตั้งแต่ 0 ถึง 6 ซึ่งแสดงโดยเส้นต่อเนื่อง (Continuous line) ดังภาพประกอบ 19



ภาพประกอบ 19 เส้นต่อเนื่องแสดงระดับคะแนนความสามารถของผู้สอบในการตอบข้อสอบเลือกตอบชนิด 4 ตัวเลือก

จากภาพประกอบ 17 แสดงเส้นต่อเนื่องระดับความรู้ 6 ระดับ (Level) คือ 0,1,2,4,5, และ 6 ตามลำดับ และมีจำนวนขั้น 5 ขั้น (Step) การที่ผู้สอบได้ 2 คะแนน แสดงว่าผู้สอบมีความรู้ในระดับ “ค่อนข้างน้อย” ในข้อนั้นๆ ซึ่งก่อนที่จะได้คะแนนในระดับนี้ ผู้สอบต้องมีความรู้ในระดับ “น้อย” กับข้อสอบข้อนี้มาก่อนแล้ว และในทำนองเดียวกัน การที่ผู้สอบได้ 6 คะแนน แสดงว่าผู้สอบมี “ความรู้สมบูรณ์” ซึ่งแสดงถึงผู้สอบมีความรู้ผ่านขั้นที่ 1,2,4, และ 5 มาแล้วตามลำดับ

ดังที่กล่าวมาแล้วข้างต้น เห็นได้ว่าการให้คะแนนแบบ 2 - SR สามารถที่จะประยุกต์ให้เข้ากับโมเดลพาเชิลเครดิต (PCM) ตามข้อเสนอของดอดด์ และค็อกซ์ (Dodd and Koch, 1987 : 373) ที่เสนอว่า หลักการสำคัญของ PCM ต้องการให้ข้อความมีลักษณะของการตอบเรียงลำดับกันเท่านั้น โดยไม่จำเป็นที่ข้อความจะต้องมีค่าความยากประจำแต่ละขั้นเรียงลำดับกัน ดังนั้นการประมาณค่าพารามิเตอร์ของข้อสอบ และความเหมาะสมกับโมเดลของข้อสอบที่ให้คะแนนแบบ 2 -SR จึงประยุกต์ใช้แนวการประมาณค่าตามวิธีการของ PCM โดยใช้โปรแกรม BIGSTEPS (Model R)

นอกจากนั้นในการศึกษาอิทธิพลของความยาก และความเหมาะสมของข้อสอบกับโมเดลได้กำหนดตัวแปรจำลองดังนี้

$F = 0$ เมื่อข้อสอบไม่เหมาะสมกับโมเดล

$F = 1$ เมื่อข้อสอบเหมาะสมกับโมเดล

$BTA = 0$ เมื่อข้อสอบมีค่าความยากน้อยกว่า -3.00 หรือมากกว่า 3.00

$BTA = 1$ เมื่อ $-3.00 \leq BTA \leq 3.00$

และในการศึกษาความเหมาะสมของข้อสอบในกรณีเป็นตัวแปรตาม ใช้ค่า t จากการคำนวณเป็นข้อมูลในการวิเคราะห์

4 การประมาณค่าสารสนเทศของข้อสอบ และแบบทดสอบ (Item and test information function)

$$IIF = I(\theta_i) = \sum I\{d_{ij}\}$$

เมื่อ IIF คือ $I(\theta_i)$ (สารสนเทศของข้อสอบข้อที่ i)

$I\{d_{ij}\}$ คือ สารสนเทศของขั้นที่ j ของข้อที่ i ซึ่งหาได้จากสูตร

$$I\{d_{ij}\} = \frac{D^2 a_i^2 Q_i (P_i - c_i)^2}{P_i (1 - c_i)^2} \quad (\text{Lord, 1980 : 73})$$

$$\text{เมื่อ } P_i = \frac{\exp(\theta_i - \beta_i)}{1 + \exp(\theta_i - \beta_i)} \quad (\text{ราซคีมเดล (Wright, 1980 : 374)})$$

เนื่องจากราซคีมเดลไม่เน้นค่าอำนาจจำแนก และค่าการเดา ดังนั้น $a = 1$, $c = 0$

จึงได้ว่า

$$IIF = I\{d_{ij}\} = I(\theta_i) = Q_i P_i$$

เมื่อ θ_i แทน ระดับความสามารถในการตอบข้อสอบข้อที่ i

β_i แทน ความยากของข้อสอบข้อที่ i

P_i แทน โอกาสในการตอบข้อสอบข้อที่ i ได้ถูกต้อง

$$Q_i = 1 - P_i$$

$$TIF = \sum IIF = \sum I(\theta_i) \text{ คือ สารสนเทศของแบบทดสอบ}$$

$$SE(\theta) = \frac{1}{TIF}$$

เมื่อ $SE(\theta)$ คือ ความคลาดเคลื่อนมาตรฐานในการประมาณค่าพารามิเตอร์ของแบบทดสอบ (Wright. 1980 : 374 ; Hambleton and Swaminathan. 1985 : 105) ของข้อสอบข้อที่ i

5. เปรียบเทียบค่า TIF ของแบบทดสอบที่ให้คะแนนโดยวิธีต่างกันจากค่าประสิทธิภาพสัมพัทธ์ (Relative precision ; RE(TIF) จากสูตร

$$RE(TIF) = \frac{TIF(A)}{TIF(B)}$$

การพิจารณาค่า RE(TIF) พิจารณาจากค่าจำนวนที่ได้ ถ้าค่ามากกว่า 1 แสดงว่าแบบทดสอบที่มีค่า TIF(A) มีประสิทธิภาพสูงกว่าแบบทดสอบที่มีค่า TIF(B) แต่ถ้าค่า RE(TIF) น้อยกว่า 1 แสดงว่าแบบทดสอบที่มีค่า TIF(A) มีประสิทธิภาพต่ำกว่า แบบทดสอบที่มีค่า TIF(B) (Hambleton . 1989 : 164)

6. การศึกษาความเป็นมิติเดียวของแบบทดสอบ (Unidimensionality) หลังจากนำแบบทดสอบแต่ละฉบับไปทดสอบกับกลุ่มผู้สอบแล้ว นำข้อมูลที่ได้มาจำแนกโดยเทคนิควิเคราะห์องค์ประกอบหลัก (Principal component ; PC) แล้วหมุนแกนด้วยเทคนิควิธีแปรimax (Varimax) การพิจารณาความเป็นมิติเดียวดูจากค่าไอเกน (Eigen value) ขององค์ประกอบที่ 1 ซึ่งควรจะมีค่าสูงกว่าค่าไอเกนขององค์ประกอบที่ 2 อย่างมาก และค่าไอเกนขององค์ประกอบที่ 2 ควรสูงกว่าค่าไอเกนขององค์ประกอบที่เรียงลำดับถัดไป (Lord. 1980 : 19- 21 ; Hambleton and Swaminathan. 1985 : 12-174 ; Hambleton et al. 1991 : 9-10) สำหรับการคัดเลือกข้อสอบโดยเกณฑ์ความเป็นมิติเดียวในการศึกษาวิจัยครั้งนี้ได้ทำการวิเคราะห์องค์ประกอบ จากการกำหนดจำนวนองค์ประกอบเพียง 2 องค์ประกอบเท่านั้น และใช้ค่าน้ำหนักองค์ประกอบ (Factor loading) ของแต่ละข้อในองค์ประกอบที่ 1 เป็นข้อมูลในการวิเคราะห์ผล นอกจากนั้นในการศึกษาอิทธิพลของความเป็นมิติเดียวได้กำหนดตัวแปรจำลองเพื่อจัดกลุ่มน้ำหนักองค์ประกอบตามข้อเสนอของไคลน์ดังนี้ (Kline. 1993 : 141)

$U = 0$ เมื่อ ค่าน้ำหนักองค์ประกอบน้อยกว่า .30 (ไม่เป็นมิติเดียวกับข้ออื่น)

$U = 1$ เมื่อ ค่าน้ำหนักองค์ประกอบตั้งแต่ .30 ขึ้นไป (เป็นมิติเดียวกับข้อ

อื่นๆ)

7. การศึกษาความลำเอียงของข้อสอบ (Item bias)

ในการศึกษาครั้งนี้ศึกษาความลำเอียงเฉพาะตัวแปรเพศโดยวิธีวิเคราะห์แบบไคแอสควร์ เนื่องจากคะแนนแบบ 2-SR เป็นคะแนนแบบหลายทาง (Polychotomous scores) และวิธีการไม่ซับซ้อนมากสามารถวิเคราะห์ได้โดยไม่ต้องใช้คอมพิวเตอร์ ซึ่งมีสูตรดังนี้

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad df = (r-1)(c-1)$$

เมื่อ i คือ ระดับคะแนน

j คือ ตัวแปรเพศ

O_{ij} คือ จำนวนคนเพศ j ที่ตอบข้อสอบข้อที่ i ถูก

E_{ij} คือ จำนวนคนที่คาดหวังในกลุ่มเพศ j ที่ได้คะแนนในระดับที่ i

ในการตัดสินความลำเอียงของข้อสอบ พิจารณาจากค่าความมีนัยสำคัญของอัตราส่วนความเป็นไปได้ (Likelihood ratio) ที่แสดงความแตกต่างของผลการตอบระหว่างเพศทั้งสองอย่างมีนัยสำคัญทางสถิติ เมื่อให้คะแนนโดยวิธีแบบปกติ และแบบ 2 - SR และได้สร้างตัวแปรจำลองเพื่อศึกษาอิทธิพลของความลำเอียงดังนี้

$B = 0$ เมื่อ อัตราส่วนความเป็นไปได้ (LR) แสดงความแตกต่างของระดับคะแนนของทั้งสองเพศอย่างไม่มีนัยสำคัญ (ข้อสอบไม่ลำเอียง)

$B = 1$ เมื่อ อัตราส่วนความเป็นไปได้ (LR) แสดงความแตกต่างของระดับคะแนนของทั้งสองเพศอย่างมีนัยสำคัญ (ข้อสอบลำเอียง)

นอกจากนั้นในการศึกษาความลำเอียงในกรณีที่เป็นตัวแปรตาม และหาความสัมพันธ์กับตัวแปรอื่นได้ใช้ค่าสัมประสิทธิ์สหสัมพันธ์แบบ Contingency เป็นข้อมูลในการวิเคราะห์ผล

8. สถิติและโปรแกรมคอมพิวเตอร์ ที่ใช้ในการวิเคราะห์ เครื่องมือและข้อมูลเชิงวิจัย

8.1 โปรแกรม SPSSPC & SPSS FOR WINDOWS

8.1.1 โปรแกรม FREQUENCIES ใช้แจกแจงความถี่ของคะแนนสอบ

8.1.2 โปรแกรม CROSSTABS สำหรับประมาณค่าความยาก และดัชนีการจำแนกของข้อสอบตามแนวของ CTT และศึกษาความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ

8.1.3 โปรแกรม RELIABILITY ใช้ประมาณค่าสัมประสิทธิ์การจำแนก

ของข้อสอบ และความเชื่อมั่นของแบบทดสอบ

8.1.4 โปรแกรม FACTOR ใช้ศึกษาความเป็นมิติเดียวของข้อสอบ

8.1.5 โปรแกรม DESCRIPTIVES ประมาณค่า สถิติพื้นฐาน ได้แก่ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน คะแนนรวม/คะแนนสูงสุด - ต่ำสุด

8.1.6 โปรแกรม MEANS / BREAKDOWN ใช้ทดสอบความแตกต่างของตัวแปรตามที่เกิดจากตัวแปรต้นที่เป็นอิสระต่อกัน

8.1.7 โปรแกรม T - TEST (PAIRS) ใช้ทดสอบความแตกต่างของคะแนนสอบที่เกิดจากการให้คะแนนที่ต่างกัน

8.1.8 โปรแกรม NONPARAMETRIC สถิติ Wilcoxon match pairs ใช้ทดสอบความแตกต่างของค่าสถิติของข้อสอบที่เกิดจากการให้คะแนนที่ต่างกัน

8.1.9 โปรแกรม CORRELATION เพื่อศึกษาระดับความสัมพันธ์ของตัวแปรต่าง ๆ

8.1.10 โปรแกรม REGRESSION เพื่อศึกษาสัมประสิทธิ์สหสัมพันธ์พหุคูณและความแปรปรวนร่วมกันของตัวแปรต่างๆ ที่เป็นปัจจัยสำคัญในการอธิบายหรือแสดงระดับการส่งผลต่อคุณภาพของแบบทดสอบ และตรวจสอบแผนภาพเส้นทางแสดงความสัมพันธ์เชิงสาเหตุซึ่งมีขั้นตอนดังนี้

8.1.10.1 กำหนดโมเดลเต็มรูป จากการศึกษาเอกสารงานวิจัยที่เกี่ยวข้อง และข้อมูลเชิงประจักษ์จากการวิเคราะห์สมการถดถอยพหุคูณเชิงเส้นตรง

8.1.10.2 วิเคราะห์โมเดล โดยใช้โปรแกรม REGRESSION ENTER) เพื่อประมาณค่าสัมประสิทธิ์เส้นทางของตัวแปรในโมเดล และตรวจสอบความมีนัยสำคัญของสัมประสิทธิ์เส้นทาง

8.1.10.3 ปรับโมเดล โดยพิจารณาจากความมีนัยสำคัญของสัมประสิทธิ์เส้นทาง

8.1.10.4 ทดสอบความสอดคล้องของโมเดลที่ปรับแล้วกับโมเดลเต็มรูป จากสูตร

$$R_{FM}^2 = 1 - (1 - R_1^2)(1 - R_2^2) \dots (1 - R_p^2)$$

เมื่อ R_1^2 สัมประสิทธิ์การพยากรณ์ของเส้นทางที่ 1 โมเดลเต็มรูป

$$R_{OM}^2 = 1 - (1 - R_1^2)(1 - R_2^2) \dots (1 - R_p^2)$$

เมื่อ R^2_j สัมประสิทธิ์การพยากรณ์ของเส้นทางที่ j ในโมเดลที่ปรับ

8.1.10.5 คำนวณค่าดัชนีความสอดคล้อง (Measure of goodness of fit ; Q) จากสูตร

$$Q = (1 - R^2_{FM}) / (1 - R^2_{OM})$$

8.1.1.6 คำนวณค่า W ซึ่งมีการแจกแจงใกล้เคียงการแจกแจง χ^2 สำหรับกลุ่มขนาดใหญ่

$$\chi^2 \approx W = - (N - d) \ln Q \text{ ที่ } df = d$$

เมื่อ N = ขนาดของกลุ่มตัวอย่าง

d = จำนวนเส้นทางที่ลบออกจากโมเดลเต็มรูป

(ศิริชัย กาญจนวาสี. 2536 : 63-108 ; อ้างอิงมาจาก Padhazur. 1982)

8.2 โปรแกรม BIGSTEPS ใช้ประมาณค่าความยาก และความเหมาะสมของข้อสอบตามแบบราชาคิโมเดล และ PCM (Partial credit model)

9. สถิติที่ใช้ทดสอบความแตกต่างของค่าสหสัมพันธ์ กรณีกลุ่มตัวอย่างสัมพันธ์กัน

$$t = \frac{(r_{11} - r_{22}) \sqrt{(n-2)}}{\sqrt{2(1 - r_{11}^2 - r_{22}^2 + 2r_{11}r_{22}r_{12})}}$$

เมื่อ r_{11} แทน ความเชื่อมั่นของแบบทดสอบฉบับที่ 1

r_{22} แทน ความเชื่อมั่นของแบบทดสอบฉบับที่ 2

r_{12} แทน ค่าสหสัมพันธ์ของคะแนนที่ได้จากการให้คะแนนทั้งสองวิธี

n แทนจำนวนผู้สอบ

10. ประเมินค่าประสิทธิภาพในการพยากรณ์ (E)

$$E = 100(1 - \sqrt{1 - r^2_{xy}})$$

r_{xy} แทน ค่าสัมประสิทธิ์สหสัมพันธ์
(วิเชียร เกตุสิงห์. 2526 : 40 , 70)

11. สูตรขยายความยาวของข้อสอบ

$$r_{nn} = \frac{nr_{tt}}{1 + (n-1) r_{tt}}$$

เมื่อ r_{nn} คือ ความเชื่อมั่นที่ปรับขยายความยาวข้อสอบแล้ว

r_{tt} คือ ความเชื่อมั่นของแบบทดสอบ เมื่อยังไม่ปรับขยายความยาวข้อสอบ

n คือ จำนวนเท่าของข้อสอบที่ต้องการเพิ่ม

(Gulliksen . 1950 : 74 - 86)

บทที่ 4

ผลการวิเคราะห์ข้อมูล

เพื่อให้เกิดความเข้าใจที่ตรงกันในการศึกษาผลการวิเคราะห์ข้อมูลจากการวิจัยจึงขอ
แสดงความหมายของสัญลักษณ์และอักษรย่อที่ใช้ดังต่อไปนี้

สัญลักษณ์และอักษรย่อที่ใช้ในการนำเสนอผลการวิเคราะห์ข้อมูล

- A1 แทน ระดับความสามารถของผู้สอบ ที่ประมาณค่าจากคะแนนแบบปกติ
(0-1)
- A2 แทน ระดับความสามารถของผู้สอบที่ประมาณค่าจากคะแนนที่ให้โดยวิธี
แบบ 2 - SR
- B แทน ความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ
- BIN แทน ข้อสอบที่ถูกคัดเลือกเข้าฉบับโดยพิจารณาจากความไม่ลำเอียงเนื่อง
จากตัวแปรเพศ (ค่าอัตราส่วนไลทิลฮูด (Likelihood ratio) ของค่าไค-
แสควร์ที่แสดงความแตกต่างของผลการตอบระหว่างเพศชายกับเพศ
หญิงอย่างไม่มีนัยสำคัญทางสถิติ)
- BOUT แทน ข้อสอบที่ถูกคัดออกจากฉบับโดยพิจารณาจากความลำเอียงเนื่องจาก
ตัวแปรเพศ (ค่าอัตราส่วนไลทิลฮูด (Likelihood ratio) ของค่าไค- แสควร์ที่
แสดงความแตกต่างของผลการตอบระหว่างเพศชายกับเพศหญิงอย่างมี
นัยสำคัญทางสถิติ ณ ระดับ .05)
- BTA แทน ค่าความยากของข้อสอบแบบราชคัมภีร์โมเดล
- BTIN แทน ข้อสอบที่ถูกคัดเลือกเข้าฉบับโดยพิจารณาจากค่าความยาก (BTA)
($-3.00 \leq BTA \leq 3.00$)
- BTOUT แทน ข้อสอบที่ถูกคัดออกจากฉบับโดยพิจารณาจากค่าความยาก (BTA)
($BTA < -3.00$; $BTA > 3.00$)
- CORR แทน สัมประสิทธิ์สหสัมพันธ์
- D แทน ดัชนีการจำแนกของข้อสอบ
- DIN แทน ข้อสอบที่ถูกคัดเลือกเข้าฉบับโดยพิจารณาจากค่าดัชนีการจำแนก
($D \geq .20$)

DOUT	แทน	ข้อสอบที่ถูกคัดออกจากฉบับโดยพิจารณาจากค่าดัชนีการจำแนก ($D < .20$)
E	แทน	ประสิทธิภาพในการพยากรณ์
F	แทน	ความเหมาะสมของข้อสอบกับราชคัมภีร์โมเดล
FIN	แทน	ข้อสอบที่เหมาะสมกับราชคัมภีร์โมเดล
FOUT	แทน	ข้อสอบที่ไม่เหมาะสมกับราชคัมภีร์โมเดล
IIF	แทน	ค่าสารสนเทศของข้อสอบ (Item information function)
IIFH	แทน	ค่าสารสนเทศของข้อสอบที่ระดับความสามารถ 3.00
IIFM	แทน	ค่าสารสนเทศของข้อสอบที่ระดับความสามารถ 0.00
IIFL	แทน	ค่าสารสนเทศของข้อสอบที่ระดับความสามารถ - 3.00
K	แทน	จำนวนข้อสอบ
MAX	แทน	ค่าสูงสุด
MEAN	แทน	ค่าเฉลี่ย
MIN	แทน	ค่าต่ำสุด
P	แทน	ค่าความยากของข้อสอบ
PIN	แทน	ข้อสอบที่ถูกคัดเลือกเข้าฉบับโดยพิจารณาจากค่าความยาก ($.20 \leq P \leq .80$)
POUT	แทน	ข้อสอบที่ถูกคัดออกจากฉบับโดยพิจารณาจากค่าความยาก ($P < .20$ หรือ $P > .80$)
r	แทน	สัมประสิทธิ์ของการจำแนกของข้อสอบ
R	แทน	ค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณ
R ²	แทน	ค่าสัมประสิทธิ์การพยากรณ์ของตัวแปรอิสระ ในสมการถดถอยแบบ เส้นตรง
R ² _{FM}	แทน	ค่าสัมประสิทธิ์การพยากรณ์ของตัวแปรอิสระ ในแผนภาพเส้นทาง เต็มรูป หรือ โมเดลสมมติฐาน
R ² _{OM}	แทน	ค่าสัมประสิทธิ์การพยากรณ์ของตัวแปรอิสระ ในแผนภาพเส้นทาง ไม่เต็มรูป ที่ปรับแล้ว หรือ โมเดลเชิงประจักษ์
RE	แทน	ค่าสัมประสิทธิ์สัมพัทธ์ซึ่งแสดงการเปรียบเทียบสารสนเทศของ แบบทดสอบที่มีวิธีให้คะแนนแตกต่างกัน

REM	แทน ค่าเฉลี่ยกำลังสองของส่วนที่เหลือจากการพยากรณ์ของสมการ
RII	แทน ความเชื่อมั่นรายข้อ
rIN	แทน ข้อสอบที่ถูกคัดเลือกเข้าฉบับโดยพิจารณาจากค่าสัมประสิทธิ์การ จำแนก ($r \geq .20$)
rOUT	แทน ข้อสอบที่ถูกคัดออกจากฉบับโดยพิจารณาจากค่าสัมประสิทธิ์การ จำแนก ($r < .20$)
RTT	แทน ความเชื่อมั่นของแบบทดสอบทั้งฉบับ
SEB	แทน ความคลาดเคลื่อนในการประมาณค่าความยากของข้อสอบแบบ ราชคโหมเดล
SE(θ)	แทน ความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ
SEM	แทน ความคลาดเคลื่อนมาตรฐานในการวัดของแบบทดสอบ
SD	แทน ส่วนเบี่ยงเบนมาตรฐานของคะแนน
TIF	แทน ค่าสารสนเทศของแบบทดสอบ (Test information function)
U	แทน ความเป็นมิติเดียวของข้อสอบ
UIN	แทน ข้อสอบที่ถูกคัดเลือกเข้าฉบับโดยพิจารณาจากค่าน้ำหนักองค์ ประกอบ (Factor loading) ที่แสดงความเป็นมิติเดียวกับข้ออื่นๆในองค์ ประกอบที่ 1 (น้ำหนักองค์ประกอบตั้งแต่ .30 ขึ้นไป)
UOUT	แทน ข้อสอบที่ถูกคัดออกจากฉบับโดยพิจารณาจากค่าน้ำหนักองค์ ประกอบ (Factor loading) ไม่แสดงความเป็นมิติเดียวกับข้ออื่นๆใน องค์ประกอบที่ 1 (น้ำหนักองค์ประกอบน้อยกว่า .30)
W	แทน ค่าการทดสอบความแตกต่างระหว่างแผนภาพเส้นทางสมมติฐาน กับแผนภาพเส้นทางเชิงประจักษ์
X1	แทน คะแนนรวมที่ได้จากการให้คะแนนตามรูปแบบปกติ (ถูก 1 ผิด 0)
X2	แทน คะแนนรวมที่ได้จากการให้คะแนนตามรูปแบบ 2- SR (0 - 6)
X3	แทน คะแนนรวมที่ได้จากการปรับคะแนนตามรูปแบบ 2- SR ให้มีฐาน คะแนนเต็มเท่ากับวิธีให้คะแนนแบบปกติ

การนำเสนอผลการวิเคราะห์ข้อมูล

ในการนำเสนอผลการวิเคราะห์ข้อมูลจากการวิจัยครั้งนี้ เพื่อให้การศึกษาเป็นไปอย่างมีลำดับขั้นตอน และสอดคล้องกับวัตถุประสงค์ของการวิจัย ผู้วิจัยจึงขอนำเสนอผลโดยเรียงลำดับตามหัวข้อต่อไปนี้

1. การศึกษาอิทธิพลของการให้คะแนนที่มีต่อ
 - 1.1 คะแนนสอบตามกรอบทฤษฎี CTT และความสามารถตามกรอบทฤษฎี IRT
 - 1.2 ค่าความเชื่อมั่น และ TIF ของแบบทดสอบ
 - 1.3 ค่าพารามิเตอร์ตามกรอบทฤษฎี CTT ความลำเอียง ความเป็นมิติเดียว และ ความเชื่อมั่นของข้อสอบ
 - 1.4 ค่าสารสนเทศ และความคลาดเคลื่อนมาตรฐานในการประมาณค่าของแบบทดสอบตามกรอบทฤษฎี IRT (ราชคโหมเดล)
 - 1.5 ความเหมาะสมกับโมเดล พารามิเตอร์ความยาก และสารสนเทศของข้อสอบตามกรอบทฤษฎี IRT (ราชคโหมเดล)
2. การศึกษาอิทธิพลของวิธีการคัดเลือกข้อสอบที่มีต่อ
 - 2.1 ค่าความเชื่อมั่นรายข้อ และความเชื่อมั่นของแบบทดสอบ
 - 2.2 ค่าสารสนเทศของข้อสอบ และแบบทดสอบ
3. การศึกษารูปแบบการให้คะแนน และวิธีการคัดเลือกข้อสอบ ที่ส่งผลต่อคุณภาพของแบบทดสอบ
 - 3.1 ความสัมพันธ์ระหว่างตัวแปรที่ใช้ในการคัดเลือกข้อสอบ ตามกรอบทฤษฎี CTT และ IRT (ราชคโหมเดล) ความเชื่อมั่นรายข้อ และสารสนเทศของข้อสอบ
 - 3.2 สหสัมพันธ์พหุคูณตามกรอบทฤษฎี CTT
 - 3.3 สหสัมพันธ์พหุคูณตามกรอบทฤษฎี IRT (ราชคโหมเดล)
 - 3.4 สหสัมพันธ์พหุคูณตามกรอบทฤษฎี CTT ร่วมกับ IRT (ราชคโหมเดล)
4. การศึกษารูปแบบความสัมพันธ์เชิงสาเหตุของการให้คะแนน และการคัดเลือกข้อสอบที่มีต่อคุณภาพแบบทดสอบ
 - 4.1 การศึกษาแผนภาพเส้นทางสมมติฐาน
 - 4.2 การตรวจสอบแผนภาพเส้นทางเชิงประจักษ์

การศึกษาอิทธิพลของการให้คะแนน

ในการศึกษาวิจัยครั้งนี้ผู้วิจัยได้เสนอวิธีการตอบและวิธีการให้คะแนนแบบทดสอบเลือกตอบขึ้นใหม่ ดังนั้นเพื่อให้เห็นความแตกต่างของผลที่เกิดขึ้นเมื่อเปรียบเทียบกับวิธีการให้คะแนนแบบปกติ (ถูก 1 ผิด 0) จึงขอแสดงผลการวิเคราะห์ตามตารางเรียงลำดับดังนี้

ตาราง 9 คะแนนเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน คะแนนสูงสุด และต่ำสุดของคะแนนสอบวัดความสามารถในการอ่านจับใจความ และการแก้โจทย์ปัญหาทางคณิตศาสตร์ซึ่งให้คะแนนตามรูปแบบปกติ (X1) และรูปแบบ 2 - SR (X3 และ X2)

	อ่านจับใจความ			แก้โจทย์ปัญหา		
	X1	X3	X2	X1	X3	X2
MEAN	19.20	18.86	113.14	17.01	16.88	101.29
SD	4.69	4.92	29.53	5.76	5.80	34.81
MAX	32.00	30.67	184.00	36	36	216
MIN	10.00	7.00	42.00	4	0	0
X1 vs X3	t = 2.09* , p = .040, CORR = .964			t = - 2.80**, p = .005, CORR = .966		

* มีระดับนัยสำคัญที่ .05

** มีระดับนัยสำคัญที่ .01

จากตาราง 9 พบว่าคะแนนเฉลี่ยของคะแนนสอบที่ได้จากการให้คะแนนตามรูปแบบปกติ (X1) สูงกว่าคะแนนเฉลี่ยของคะแนนสอบที่ได้จากการให้คะแนนแบบ 2-SR (X3) ในแบบทดสอบทั้งสองฉบับอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .05 และเมื่อพิจารณาคะแนนสูงสุด ต่ำสุดก็ให้ผลที่สอดคล้องกันยกเว้นแบบทดสอบการแก้โจทย์ปัญหาที่คะแนนสูงสุดของทั้งสองวิธีเท่ากัน สำหรับค่าความสัมพันธ์ของการให้คะแนนทั้งสองวิธีพบว่าอยู่ในระดับสูงคือ .964 และ .966 ตามลำดับ

ตาราง 10 ความสามารถเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ระดับความสามารถสูงสุดและต่ำสุดในการตอบแบบทดสอบวัดความสามารถในการอ่านจับใจความ และแก้โจทย์ปัญหาทางคณิตศาสตร์ ที่ให้คะแนนตามรูปแบบปกติ (A1) และรูปแบบ 2- SR (A2)

	อ่านจับใจความ			แก้โจทย์ปัญหา		
	A1	A2	t-test	A1	A2	t-test
MEAN	0.1762	- 0.0448	12.56**	-.2412	- 0.0912	- 2.10*
SD	0.664	0.187	p = .000	2.879	0.443	p = .036
SEE	0.018...	0.005		0.080	0.012	
MAX	1.76	0.41		2.52	0.57	
MIN	- 2.68	- 2.35		- 2.61	- 1.94	
CORR with X	0.282**	0.763**		.5882**	0.733**	
CORR A1- A2		0.292**			0.724**	

* มีระดับนัยสำคัญที่ .05

** มีระดับนัยสำคัญที่ .01

จากตาราง 10 พบว่า ค่าความสามารถเฉลี่ยของผู้สอบแบบทดสอบการอ่านจับใจความที่ให้คะแนนแบบปกติสูงกว่าค่าความสามารถเฉลี่ยที่ให้คะแนนแบบ 2-SR อย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01 เท่ากับ .1762 และ -0.0448 ตามลำดับ โดยมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.664 และ 0.187 ความคลาดเคลื่อนมาตรฐานในการประมาณค่าเท่ากับ .018 และ 0.005 ระดับความสัมพันธ์ของระดับความสามารถกับคะแนนสอบเท่ากับ 0.282 และ 0.763 ตามลำดับ ส่วนระดับความสัมพันธ์ของระดับความสามารถจากการให้คะแนนของทั้งสองวิธีอยู่ในระดับต่ำคือเท่ากับ 0.292 เท่านั้น

สำหรับแบบทดสอบการแก้โจทย์ปัญหา พบว่าค่าความสามารถเฉลี่ยที่ได้จากการให้คะแนนแบบ 2-SR สูงกว่าค่าความสามารถเฉลี่ยที่ได้จากการให้คะแนนแบบปกติอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .05 เท่ากับ -0.0912 และ -.2412 ตามลำดับ โดยมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.443 และ 2.879 ความคลาดเคลื่อนมาตรฐานในการประมาณค่าเท่ากับ 0.012 และ 0.080 ระดับความสัมพันธ์ของความสามารถกับคะแนนสอบเท่ากับ 0.733 และ 0.588 ตามลำดับ ส่วนระดับความสัมพันธ์ระหว่างระดับความสามารถจากการให้คะแนนทั้งสองวิธีอยู่ในระดับค่อนข้างสูงคือเท่ากับ 0.724

เมื่อพิจารณาค่าความสามารถสูงสุด-ต่ำสุดพบว่า ค่าสูงสุดของความสามารถที่ได้จากการให้คะแนนแบบปกติสูงกว่าความสามารถที่ได้จากการให้คะแนนแบบ 2- SR ในแบบทดสอบทั้งสองฉบับ เมื่อพิจารณาค่าความสามารถต่ำสุดพบว่าค่าความสามารถที่ได้จากการให้คะแนนโดยวิธีแบบ 2 - SR มีค่าสูงกว่าค่าความสามารถที่ได้จากการให้คะแนนแบบปกติในแบบทดสอบทั้งสองฉบับ

ตาราง 11 ค่าความเชื่อมั่น ความคลาดเคลื่อนมาตรฐานในการวัด ของแบบทดสอบวัดความสามารถในการอ่านจับใจความ และการแก้โจทย์ปัญหาคณิตศาสตร์ซึ่งให้คะแนนตามรูปแบบปกติ และรูปแบบ 2 - SR

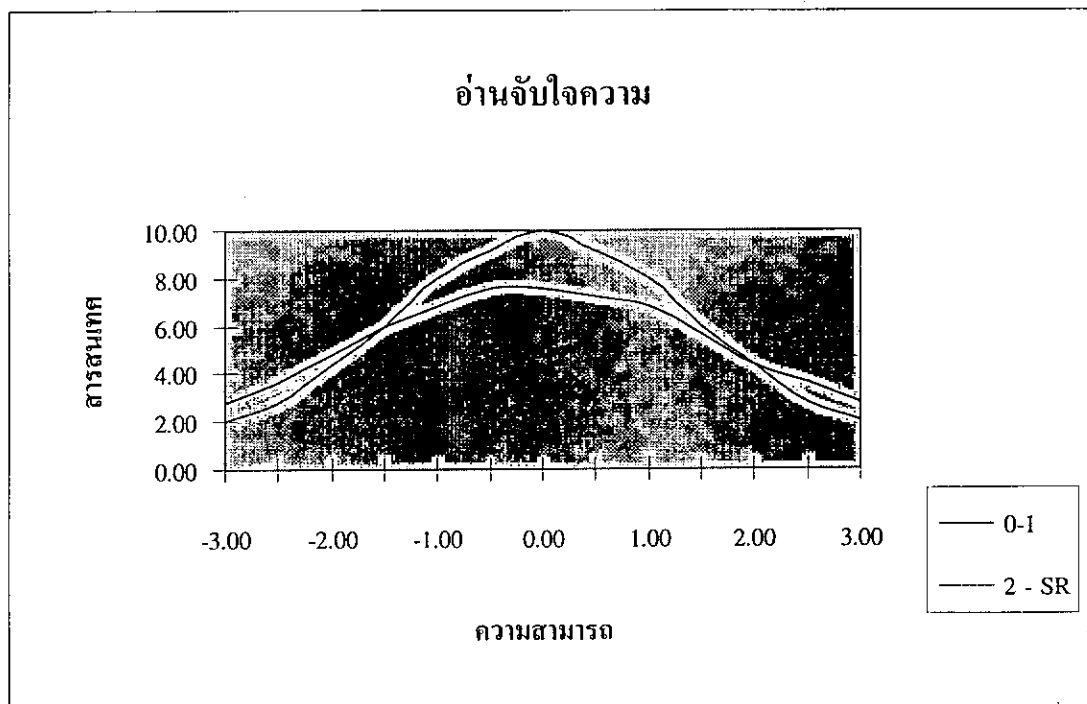
วิธีการ ให้คะแนน	การอ่านจับใจความ			การแก้โจทย์ปัญหา		
	RTT	SEM	t - test	RTT	SEM	t - test
ปกติ	.6578	2.74	1.747	.7678	4.82	.0645
2 - SR	.7261	2.58		.7915	3.20	

จากตาราง 11 พบว่าค่าความเชื่อมั่นของแบบทดสอบการอ่านจับใจความ และการแก้โจทย์ปัญหาที่ให้คะแนนโดยวิธีแบบปกติ และวิธีให้คะแนนแบบ 2-SR ไม่แตกต่างกัน แต่เมื่อพิจารณาค่าความคลาดเคลื่อนมาตรฐานของการวัดพบว่าวิธีการให้คะแนนแบบ 2-SR ให้ค่าที่ต่ำกว่าค่าที่ได้จากการให้คะแนนแบบปกติในแบบทดสอบทั้งสองฉบับ

ตาราง 12 ค่าสารสนเทศของแบบทดสอบ ความคลาดเคลื่อนมาตรฐานในการประมาณค่า และประสิทธิภาพสัมพัทธ์ของแบบทดสอบวัดความสามารถในการอ่านจับใจความ และการแก้โจทย์ปัญหาทางคณิตศาสตร์ซึ่งให้คะแนนตามรูปแบบปกติ และรูปแบบ 2 - SR

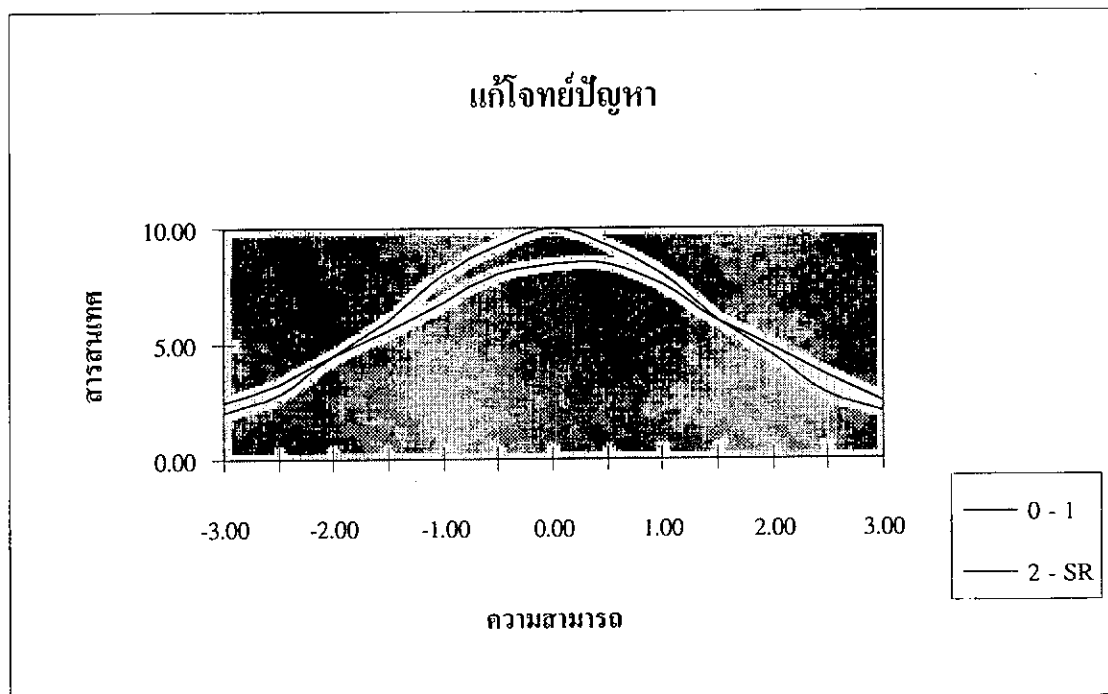
ระดับความสามารถ/ วิธีการให้คะแนน		การอ่านจับใจความ			การแก้โจทย์ปัญหา		
		TIF	SE (θ)	RE	TIF	SE (θ)	RE
ระดับความสามารถ สูง ($\theta = 3.00$)	ปกติ	2.628	0.617	0.706	2.348	0.623	0.082
	2 - SR	1.856	0.734		1.836	0.738	
ระดับความสามารถ ปานกลาง ($\theta = 0.00$)	ปกติ	7.720	0.360	1.271	8.388	0.345	1.178
	2 - SR	9.816	0.319		9.884	0.318	
ระดับความสามารถ ต่ำ ($\theta = - 3.00$)	ปกติ	2.616	0.618	0.709	2.560	0.625	0.719
	2 - SR	1.856	0.734		1.840	0.737	

จากตาราง 12 พบว่าค่าสารสนเทศของแบบทดสอบที่ให้คะแนนทั้ง 2 รูปแบบที่ระดับความสามารถปานกลางมีค่าสูงกว่าสารสนเทศของแบบทดสอบที่ระดับความสามารถสูงและต่ำในแบบทั้งสองฉบับ และเมื่อพิจารณาค่าประสิทธิภาพสัมพัทธ์ของแบบทดสอบทั้งสองฉบับพบว่าค่าสารสนเทศที่ได้จากการให้คะแนนแบบ 2 - SR ต่ำกว่าค่าที่ได้จากการให้คะแนนแบบปกติที่ระดับความสามารถสูงและต่ำ แต่เมื่อพิจารณาที่ ระดับความสามารถปานกลาง พบว่าค่าสารสนเทศของแบบทดสอบที่ให้คะแนนแบบ 2- SR สูงกว่าค่าสารสนเทศของแบบทดสอบที่ได้จากการให้คะแนนแบบปกติ



ภาพประกอบ 20 สารสนเทศของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนตามรูปแบบปกติ และรูปแบบ 2 - SR ที่ช่วงระดับความสามารถตั้งแต่ -3.00 ถึง 3.00

จากภาพประกอบ 20 พบว่าสารสนเทศของแบบทดสอบที่ให้คะแนนตามรูปแบบ 2 - SR แสดงแนวโน้มให้ค่าสูงกว่าเมื่อให้คะแนนตามรูปแบบปกติที่ระดับความสามารถปานกลาง ประมาณช่วงตั้งแต่ -2.50 ถึง 2.00 แต่แสดงแนวโน้มให้ค่าต่ำกว่าที่ระดับความสามารถสูง และต่ำ คือน้อยกว่า -2.50 และ มากกว่า 3.00



ภาพประกอบ 21 สารสนเทศของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนตามรูปแบบปกติ และรูปแบบ 2 - SR ที่ช่วงระดับความสามารถตั้งแต่ -3.00 ถึง 3.00

จากภาพประกอบ 21 พบว่าสารสนเทศของแบบทดสอบที่ให้คะแนนตามรูปแบบ 2 - SR แสดงแนวโน้มให้ค่าสูงกว่าเมื่อให้คะแนนตามรูปแบบปกติที่ระดับความสามารถปานกลาง ประมาณช่วงตั้งแต่ -2.00 ถึง 1.50 แต่แสดงแนวโน้มให้ค่าต่ำกว่าที่ระดับความสามารถสูง และต่ำ คือน้อยกว่า -2.00 และมากกว่า 1.50

ตาราง 13 ค่าเฉลี่ยความยากแบบ CTT และ IRT สัมประสิทธิ์การจำแนก ดัชนีการจำแนก

ค่าความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ ค่าความเป็นมิติเดียว ค่าความเชื่อมั่นรายข้อ
ค่าความเหมาะสมของข้อสอบกับโมเดล และค่าสารสนเทศของข้อสอบของแบบทดสอบวัด
ความสามารถในการอ่านจับใจความและการแก้โจทย์ปัญหาทางคณิตศาสตร์ซึ่งให้คะแนนตามรูปแบบ
ปกติ และรูปแบบ 2 - SR

	อ่านจับใจความ			แก้โจทย์ปัญหา		
	ปกติ	2 - SR	t - test	ปกติ	2 - SR	t - test
P	.4905	.4151	.858**	.4325	.4312	0.12
r	.1905	.2233	-6.86**	.2410	.2603	- 9.92**
D	.2174	.2669	-5.42**	.2921	.3749	8.58**
B	.0523	.0903	- 9.48**	.0320	.0708	- 6.86**
U	.1987	.2577	- 8.20**	.1977	.2463	- 5.42**
RII	.0892	.0957	- 3.54**	.1111	.1151	- 5.00**
BTA	.0003	.0002	0.00	-.0022	.0000	- 0.02
F	-0.4175	-0.3100	-0.64	-.6025	-.6300	.04
IIFH	.0652	.0464	2.28	.0587	.0459	2.36*
IIFM	.1828	.2455	-0.91**	.2097	.2471	- 4.87**
IIFL	.0659	.0464	2.68*	.0604	.0460	1.80

* นัยสำคัญทางสถิติ ณ ระดับ .05

** นัยสำคัญทางสถิติ ณ ระดับ .01

จากตาราง 13 พบว่าค่าสัมประสิทธิ์การจำแนก ดัชนีการจำแนก ความลำเอียงเนื่องจาก
ตัวแปรเพศ ความเป็นมิติเดียว ความเชื่อมั่นรายข้อ และสารสนเทศของข้อสอบที่ระดับความ
สามารถปานกลางที่ได้จากการให้คะแนนตามรูปแบบ 2-SR สูงกว่าค่าที่ได้จากการให้คะแนนแบบ
ปกติอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01 ทั้งในแบบทดสอบวัดความสามารถในการอ่าน
จับใจความ และการแก้โจทย์ปัญหาทางคณิตศาสตร์ ส่วนค่าความยากแบบ CTT พบว่าค่าที่ได้
จากการให้คะแนนแบบ 2 - SR ต่ำกว่าค่าที่ได้จากการให้คะแนนแบบปกติอย่างมีนัยสำคัญทาง
สถิติ ณ ระดับ .01 เฉพาะแบบทดสอบวัดความสามารถในการอ่านจับใจความเท่านั้น ส่วนในแบบ
ทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์พบว่าไม่แตกต่างกัน สำหรับค่า

สารสนเทศของข้อสอบที่ระดับความสามารถสูงพบว่าค่าที่ได้จากการให้คะแนนแบบปกติสูงกว่าค่าที่ได้จากการให้คะแนนแบบ 2 - SR อย่างมีนัยสำคัญทางสถิติ ณ ระดับ .05 ในแบบทดสอบทั้งสองฉบับ ส่วนสารสนเทศของข้อสอบที่ระดับความสามารถต่ำ พบว่าค่าที่ได้จากการให้คะแนนแบบปกติสูงกว่าค่าที่ได้จากการให้คะแนนแบบ 2-SR อย่างมีนัยสำคัญทางสถิติ ณ ระดับ .05 เฉพาะแบบทดสอบวัดความสามารถในการอ่านจับใจความเท่านั้น ส่วนในแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์พบว่าให้ผลไม่แตกต่างกัน และเมื่อพิจารณาความยากแบบ IRT และความเหมาะสมของข้อสอบกับโมเดล พบว่าค่าที่ได้จากการให้คะแนนทั้งสองวิธีไม่แตกต่างกันในแบบทดสอบทั้งสองฉบับ

การศึกษาอิทธิพลของวิธีการคัดเลือกข้อสอบ

ในการวิเคราะห์ข้อสอบเพื่อคัดเลือกข้อสอบเข้าฉบับตามกรอบทฤษฎี CTT ในปัจจุบัน นิยมใช้ค่าความยาก และค่าดัชนีการจำแนก (D) หรือสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนรายข้อ กับคะแนนรวม (r) ซึ่งในส่วนของผู้วิจัยเห็นว่าน่าจะมีตัวแปรอื่นที่เข้าแทนหรือช่วยเสริมค่าทั้งสองดังกล่าว โดยเฉพาะความเป็นมิติเดียว และความลำเอียงของข้อสอบดังนั้นในการศึกษาวิจัยครั้งนี้ จึงสนใจที่จะศึกษาอิทธิพลของตัวแปรต่างๆ ที่มีต่อค่าความเชื่อมั่น และสารสนเทศของข้อสอบดังผลการวิเคราะห์ตามตารางเป็นลำดับต่อไป

ตาราง 14 ค่าความเชื่อมั่นของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่หาคะแนนตามรูปแบบปกติ และรูปแบบ 2 - SR เมื่อคัดเลือกข้อสอบเข้าฉบับโดยตัวแปรต่างๆ และค่าความเชื่อมั่นของแบบทดสอบเมื่อปรับขยายจำนวนข้อสอบเต็มฉบับ (40 ข้อ)

วิธีการคัดเลือกข้อสอบ/ รูปแบบการให้คะแนน		แบบทดสอบไม่ปรับขยายความ ยาว			แบบทดสอบปรับขยายความ ยาวเต็มฉบับ (40 ข้อ)	
		K	RTT	SEM	K	RTT
PIN	ปกติ	34	.7066	2.64	40	.7382
	2-SR	34	.7621	2.37	40	.7902
rIN	ปกติ	18	.7478	1.81	40	.8680
	2-SR	22	.8048	1.83	40	.8822
DIN	ปกติ	31	.7284	2.28	40	.7757
	2-SR	30	.7855	1.03	40	.8299
BIN	ปกติ	23	.4239	2.15	40	.5613
	2-SR	18	.4694	1.70	40	.6625
UIN	ปกติ	12	.7239	1.37	40	.8972
	2-SR	19	.8036	1.67	40	.8961
ปกติ VS 2 - SR		z = -2.02* 2-tailed P = .0431			z = -1.753 2-tailed P = .0796	

* มีนัยสำคัญทางสถิติ ณ ระดับ .05

** มีนัยสำคัญทางสถิติ ณ ระดับ .01

จากตาราง 14 พบว่าจำนวนข้อสอบของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้รับการคัดเลือกโดยตัวแปรต่าง ๆ คือความยาก สัมประสิทธิ์การจำแนก ดัชนีการจำแนกความลำเอียงเนื่องจากตัวแปรเพศ และความเป็นมิติเดียวของข้อสอบ ที่ได้จากการให้คะแนนตามรูปแบบ 2 - SR อยู่ในช่วงระหว่าง 18 - 34 ข้อ ค่าความเชื่อมั่นของแบบทดสอบอยู่ในช่วงระหว่าง .4694 - .8048 เมื่อปรับขยายความยาวข้อสอบเป็น 40 ข้อ พบว่าอยู่ในช่วง .6625 - .8961 ส่วนจำนวนข้อสอบของแบบทดสอบที่ได้จากการให้คะแนนแบบปกติ พบว่าอยู่ในช่วงระหว่าง 12 - 34 ข้อ ค่าความเชื่อมั่นของแบบทดสอบอยู่ในช่วงระหว่าง .4239 - .7478 เมื่อปรับขยายความยาวข้อสอบเป็น 40 ข้อ พบว่าอยู่ในช่วง .5613 - .8972

จากการทดสอบความแตกต่างความเชื่อมั่นของแบบทดสอบที่ถูกคัดเลือกข้อสอบเข้าฉบับ โดยใช้ตัวแปรต่างๆ เมื่อให้คะแนนตามรูปแบบทั้งสองพบว่า ค่าความเชื่อมั่นของแบบทดสอบที่ได้จากการให้คะแนนตามรูปแบบ 2 - SR มีค่าสูงกว่าค่าที่ได้จากการให้คะแนนตามรูปแบบปกติ อย่างมีนัยสำคัญทางสถิติ ณ ระดับ .05 ทุกวิธีการคัดเลือกข้อสอบ แต่เมื่อปรับขยายข้อสอบให้มีจำนวนเท่ากัน พบว่าให้ผลที่ไม่แตกต่างกัน นอกจากนั้นยังพบว่าเมื่อใช้ค่าความยากที่ได้จากการให้คะแนนทั้ง 2 รูปแบบเป็นเกณฑ์ในการคัดเลือกข้อสอบเข้าฉบับ จำนวนข้อสอบที่ได้มีจำนวนมากกว่าตัวแปรอื่นๆ รองลงมาคือ ค่าดัชนีการจำแนก ส่วนความเป็นมิติเดียวพบว่าคัดเลือกข้อสอบเข้าฉบับน้อยกว่าตัวแปรอื่นๆ เมื่อให้คะแนนตามรูปแบบ 2 - SR แต่เมื่อให้คะแนนตามรูปแบบปกติ พบว่าตัวแปรความลำเอียงเนื่องจากตัวแปรเพศคัดเลือกข้อสอบเข้าฉบับได้น้อยกว่าตัวแปรอื่น และเมื่อพิจารณาค่าความเชื่อมั่นของแบบทดสอบที่ปรับขยายความยาวจากการคัดเลือกโดยวิธีต่างๆ พบว่าความเชื่อมั่นของแบบทดสอบที่ได้จากการคัดเลือกเข้าฉบับโดยความเป็นมิติเดียวมีค่าสูงกว่าตัวแปรอื่นๆ และความลำเอียงเนื่องจากตัวแปรเพศ คัดเลือกข้อสอบเข้าฉบับให้ค่าความเชื่อมั่นของแบบทดสอบต่ำกว่าตัวแปรอื่นๆ ในการให้คะแนนทั้งสองรูปแบบ

ตาราง 15 ค่าความเชื่อมั่นของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ ที่ให้คะแนนตามรูปแบบปกติ และรูปแบบ 2 - SR เมื่อคัดเลือกข้อสอบเข้าฉบับโดยตัวแปรต่างๆ และค่าความเชื่อมั่นของแบบทดสอบเมื่อปรับขยายจำนวนข้อสอบเต็มฉบับ (40 ข้อ)

วิธีการคัดเลือกข้อสอบ/ รูปแบบการให้คะแนน		แบบทดสอบไม่ปรับขยายความยาว			แบบทดสอบปรับขยายความยาวเต็มฉบับ (40 ข้อ)	
		K	RTT	SEM	K	RTT
PIN	ปกติ	35	.7572	1.37	40	.7807
	2-SR	37	.7917	1.23	40	.8042
rIN	ปกติ	31	.8112	1.12	40	.8471
	2-SR	31	.8300	1.02	40	.8630
DIN	ปกติ	30	.8078	0.99	40	.8485
	2-SR	33	.8151	1.09	40	.8423
BIN	ปกติ	33	.7321	1.37	40	.7681
	2-SR	31	.7602	1.19	40	.8035
UIN	ปกติ	12	.7379	0.80	40	.9036
	2-SR	18	.7739	0.91	40	.8838
ปกติ VS 2 - SR		z = -2.02* 2-tailed P = .0431			z = -.3439 2-tailed P = .3452	

* มีนัยสำคัญทางสถิติ ณ ระดับ .05

** มีนัยสำคัญทางสถิติ ณ ระดับ .01

จากตาราง 15 พบว่าจำนวนข้อสอบของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ได้รับการคัดเลือกโดยตัวแปรต่าง ๆ คือความยาก สัมประสิทธิ์การจำแนกดัชนีการจำแนก ความลำเอียงเนื่องจากตัวแปรเพศ และความเป็นมิติเดียวของข้อสอบ ที่ได้จากการให้คะแนนตามรูปแบบ 2 - SR อยู่ในช่วงระหว่าง 18 - 37 ข้อ ค่าความเชื่อมั่นของแบบทดสอบอยู่ในช่วงระหว่าง .7602 - .8300 เมื่อปรับขยายความยาวข้อสอบเป็น 40 ข้อ พบว่าอยู่ในช่วง .8035 - .8838 ส่วนจำนวนข้อสอบของแบบทดสอบที่ได้จากการให้คะแนนแบบปกติ พบว่าอยู่ในช่วงระหว่าง 12 - 35 ข้อ ค่าความเชื่อมั่นของแบบทดสอบอยู่ในช่วงระหว่าง .7321 - .8112 เมื่อปรับขยายความยาวข้อสอบเป็น 40 ข้อ พบว่าอยู่ในช่วง .7681 - .9036

จากการทดสอบความแตกต่างความเชื่อมั่นของแบบทดสอบที่ถูกคัดเลือกข้อสอบเข้าฉบับ โดยใช้ตัวแปรต่างๆ เมื่อให้คะแนนตามรูปแบบทั้งสองพบว่า ค่าความเชื่อมั่นของแบบทดสอบที่ได้จากการให้คะแนนตามรูปแบบ 2 - SR มีค่าสูงกว่าค่าที่ได้จากการให้คะแนนตามรูปแบบปกติ อย่างมีนัยสำคัญทางสถิติ ณ ระดับ .05 ทุกวิธีการคัดเลือกข้อสอบ แต่เมื่อปรับขยายข้อสอบให้มีจำนวนเท่ากัน พบว่าให้ผลที่ไม่แตกต่างกัน และพบว่าเมื่อใช้ค่าความยากที่ได้จากการให้คะแนน ทั้ง 2 รูปแบบเป็นเกณฑ์ในการคัดเลือกข้อสอบเข้าฉบับ จำนวนข้อสอบที่ได้มีจำนวนมากกว่าตัวแปรอื่นๆ และพบว่าตัวแปรความลำเอียงเนื่องจากตัวแปรเพศคัดเลือกข้อสอบเข้าฉบับได้น้อยกว่าตัวแปรอื่น และเมื่อพิจารณาค่าความเชื่อมั่นของแบบทดสอบที่ปรับขยายความยาวจากการคัดเลือกโดยวิธีต่างๆ พบว่าความเชื่อมั่นของแบบทดสอบที่ได้จากการคัดเลือกเข้าฉบับโดยความเป็นมิติเดียวมีค่าสูงกว่าตัวแปรอื่นๆ และความลำเอียงเนื่องจากตัวแปรเพศ คัดเลือกข้อสอบเข้าฉบับให้ค่าความเชื่อมั่นของแบบทดสอบต่ำกว่าตัวแปรอื่นๆ ในการให้คะแนนทั้งสองรูปแบบ

ตาราง 16 ค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูงของแบบทดสอบวัดความสามารถในการอ่านจับใจความ ที่ให้คะแนนตามรูปแบบปกติ และรูปแบบ 2-SR จำแนกตามตัวแปรต่างๆที่ใช้ในการคัดเลือกข้อสอบเข้าฉบับ ตามกรอบทฤษฎี CTT และ IRT

		0 - 1				2 - SR			
		K	IIFH	F- test	p	K	IIFH	F - test	p
P	POUT	6	.1814	51.5**	0.010	6	.0706	33.7**	0.73
	PIN	34	.0453			34	.0494		
r	rOUT	22	.1127	29.9**	0.000	18	.0559	.415**	.0000
	rIN	18	.0272			22	.0387		
D	DOUT	9	.1313	49.3**	0.000	10	.0599	47.6**	.0000
	DIN	31	.0303			30	.0406		
B	BOUT	17	.0442	3.45	.070	22	.0451	.5683	0.455
	BIN	23	.0816			18	.0480		
U	UOUT	28	.0863	12.1**	.001	21	.0544	37.1**	0.000
	UIN	12	.0175			19	.0377		
F	FOUT	12	.0401	2.78	102	14	.0439	0.9435	0.337
	FIN	28	.0766			26	.0478		
BTA	BTOUT	-	-	-	-	-	-	-	-
	BTIN	40	.0657			40	.0464		

* มีนัยสำคัญทางสถิติ ณ ระดับ .05

** มีนัยสำคัญทางสถิติ ณ ระดับ .01

จากตาราง 16 พบว่าค่าเฉลี่ยของค่าสารสนเทศของข้อสอบที่ระดับที่ระดับความสามารถสูงของแบบทดสอบวัดความสามารถในการอ่านจับใจความ ที่ได้จากการให้คะแนนตามรูปแบบปกติ และรูปแบบ 2 - SR เมื่อจำแนกตามกลุ่มความยาก สัมประสิทธิ์การจำแนก ดัชนีการจำแนก และความเป็นมิติเดียวมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01 ส่วนการจำแนกโดยพิจารณาจากความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ และความเหมาะสมของข้อสอบกับโมเดลพบว่าให้ผลไม่แตกต่างกัน และเมื่อพิจารณาจากตัวแปรค่าความยากแบบ IRT พบว่าไม่มีข้อสอบข้อใดถูกคัดออกจากฉบับโดยเกณฑ์ความยากแบบราชาโมเดล

ตาราง 17 ค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลางของแบบทดสอบ
วัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนตามรูปแบบปกติ และรูปแบบ 2-SR
จำแนกตามตัวแปรต่างๆที่ใช้ในการคัดเลือกข้อสอบเข้าฉบับ ตามกรอบทฤษฎี CTT และ IRT

		0 - 1				2 - SR			
		K	IIFM	F- test	p	K	IIFM	F - test	p
P	POUT	6	.1101	23.6**	.000	6	.2343	66.3**	0.00
	PIN	34	.2077			34	.2473		
r	rOUT	22	.1842	.789	.380	18	.2441	1.569	.218
	rIN	18	.2003			22	.2463		
D	DOUT	9	.1673	4.83*	.034	10	.2416	9.27**	.004
	DIN	31	.2069			30	.2469		
B	BOUT	17	.1867	.3541	.555	22	.2449	.3678	.547
	BIN	23	.1977			18	.2459		
U	UOUT	28	.2006	1.674	.203	21	.2450	.2013	.656
	UIN	12	.1754			19	.2457		
F	FOUT	12	.2175	3.339	.075	14	.2462	.4462	.508
	FIN	28	.1826			26	.2449		
BTA	BTOUT	-	-	-	-	-	-	-	-
	BTIN	40	.1930			40	.2454		

* มีนัยสำคัญทางสถิติ ณ ระดับ .05

** มีนัยสำคัญทางสถิติ ณ ระดับ .01

จากตาราง 17 พบว่าค่าเฉลี่ยของค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลางของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนตามรูปแบบปกติ และรูปแบบ 2 - SR เมื่อจำแนกตามกลุ่มความยาก และดัชนีการจำแนก มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .05 ส่วนการจำแนกโดยพิจารณาจากสัมประสิทธิ์การจำแนก ความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ ความเป็นมิติเดียว และความเหมาะสมของข้อสอบกับโมเดลพบว่าให้ผลไม่แตกต่างกัน และเมื่อพิจารณาจากตัวแปรค่าความยากแบบ IRT พบว่าไม่มีข้อสอบข้อใดถูกคัดออกจากฉบับโดยเกณฑ์ความยากแบบราชคโสมิเดล

ตาราง 18 ค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนตามรูปแบบปกติ และรูปแบบ 2-SR จำแนกตามตัวแปรต่างๆ ที่ใช้ในการคัดเลือกข้อสอบเข้าฉบับ ตามกรอบทฤษฎี CTT และ IRT

		0 - 1				2 - SR			
		K	IIFL	F- test	p	K	IIFL	F - test	p
P	POUT	6	.0198	5.31*	0.026	6	.0300	11.1**	0.005
	PIN	34	.0734			34	.0420		
r	rOUT	22	.0262	27.5**	0.000	18	.0378	41.0**	0.000
	rIN	18	.0974			22	.0541		
D	DOUT	9	.0204	21.8**	0.000	10	.0358	25.6**	0.000
	DIN	31	.0896			30	.0514		
B	BOUT	17	.0920	8.05**	0.007	22	.0486	1.348	0.252
	BIN	23	.0457			18	.0444		
U	UOUT	28	.0396	41.3**	0.000	21	.0390	40.6**	0.000
	UIN	12	.1257			19	.0553		
F	FOUT	12	.0701	.1221	0.728	14	.0493	1.065	0.308
	FIN	28	.0634			26	.0454		
BTA	BTOUT	-	-	-	-	-	-	-	-
	BTIN	40	.0654			40	.0464		

* มีนัยสำคัญทางสถิติ ณ ระดับ .05

** มีนัยสำคัญทางสถิติ ณ ระดับ .01

จากตาราง 18 พบว่าค่าเฉลี่ยของค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนตามรูปแบบปกติ และรูปแบบ 2 - SR เมื่อจำแนกตามกลุ่มความยาก สัมประสิทธิ์การจำแนก ดัชนีการจำแนก และความเป็นมิติ เดียวมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .05 และพบว่าการจำแนกโดยพิจารณาจากความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01 เฉพาะค่าที่ได้จากการให้คะแนนแบบปกติเท่านั้น ส่วนค่าที่ได้จากการให้คะแนนแบบ 2 - SR พบว่าให้ผลที่ไม่แตกต่างกัน สำหรับการจำแนกโดยความเหมาะสมของข้อ

สอบกับโมเดลพบว่าให้ผลไม่แตกต่างกัน และเมื่อพิจารณาจากตัวแปรค่าความยากแบบ IRT พบว่าไม่มีข้อสอบข้อใดถูกคัดออกจากฉบับโดยเกณฑ์ความยากแบบราชคโสมเดล

ตาราง 19 ค่าสารสนเทศของข้อสอบที่ ระดับความสามารถสูงของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนตามรูปแบบปกติ และรูปแบบ 2 - SR จำแนกตามตัวแปรต่างๆที่ใช้ในการคัดเลือกข้อสอบเข้าฉบับตามกรอบทฤษฎี CTT และ IRT

		0 - 1				2 - SR			
			IIFH	F- test	p	K	IIFH	F - test	p
P	POUT	5	.0772	1.096	0.302	3	.0567	3.27	.083
	PIN	35	.0561			37	.0500		
r	rOUT	9	.0939	9.94**	.0031	9	.0533	8.02**	0.007
	rIN	31	.0485			31	.0445		
D	DOUT	10	.1082	26.3**	0.000	7	.0544	11.7**	0.001
	DIN	30	.0443			33	.0442		
B	BOUT	7	.0585	.0002	.9896	9	.0467	0.004	0.950
	BIN	33	.0587			31	.0465		
U	UOUT	22	.0771	26.0**	0.000	22	.0450	1.39	0.245
	UIN	18	.0204			18	.0483		
F	FOUT	12	.0548	.2510	.6192	15	.0453	.4039	0.529
	FIN	28	.0616			25	.0472		
BTA	BTOUT	-	-	-	-	-	-	-	-
	BTIN	40	.0587			40	.0459		

* มีนัยสำคัญทางสถิติ ณ ระดับ .05 ** มีนัยสำคัญทางสถิติ ณ ระดับ .01

จากตาราง 19 พบว่าค่าเฉลี่ยของค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูงของแบบทดสอบวัดความสามารถแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ได้จากการให้คะแนนตามรูปแบบปกติ และรูปแบบ 2 - SR เมื่อจำแนกตามสัมประสิทธิ์การจำแนก และดัชนีการจำแนก มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01 และพบว่าเมื่อจำแนกโดยพิจารณาจากความเป็น

มิติเดียว ให้ผลแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01 เฉพาะการให้คะแนนแบบปกติ เท่านั้น ส่วนค่าที่ได้จากการให้คะแนนแบบ 2 - SR พบว่าให้ผลที่ไม่แตกต่างกัน สำหรับการจำแนก โดยความยากแบบ CTT ความลำเอียงเนื่องจากตัวแปรเพศ และความเหมาะสมของข้อสอบกับ โมเดลพบว่าให้ผลไม่แตกต่างกัน และเมื่อพิจารณาจากตัวแปรค่าความยากแบบ IRT พบว่าไม่มีข้อ สอบข้อใดถูกคัดออกจากฉบับโดยเกณฑ์ความยากแบบราซค์โมเดล

ตาราง 20 ค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลางของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนตามรูปแบบปกติ และรูปแบบ 2-SR จำแนกตามตัวแปรต่างๆที่ใช้ในการคัดเลือกข้อสอบเข้าฉบับตามกรอบทฤษฎี CTT และ IRT

		0 - 1				2 - SR			
		K	IIFM	F- test	p	K	IIFM	F - test	p
P	POUT	5	.0937	79.2**	0.000	3	.2433	42.4**	0.000
	PIN	35	.2263			37	.2500		
r	rOUT	9	.2090	0.002	0.964	9	.2478	.0003	0.986
	rIN	31	.2100			31	.2477		
D	DOUT	10	.1939	0.997	0.324	7	.2478	.0003	0.986
	DIN	30	.2143			33	.2477		
B	BOUT	7	.1711	4.76*	.0355	9	.2478	0.000	0.986
	BIN	33	.2179			31	.2477		
U	UOUT	22	.2245	7.22**	0.011	22	.2459	6.75*	0.013
	UIN	18	.1790			18	.2500		
F	FOUT	12	.2168	0.504	0.482	15	.2480	0.052	0.820
	FIN	28	.2045			25	.2476		
BTA	BTOUT		-	-	-	-	-	-	-
	BTIN	40	.2097			40	.2471		

* มีนัยสำคัญทางสถิติ ณ ระดับ .05

** มีนัยสำคัญทางสถิติ ณ ระดับ .01

จากตาราง 20 พบว่าค่าเฉลี่ยของค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลางของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ได้จากการให้คะแนนตามรูปแบบปกติ และรูปแบบ 2 - SR เมื่อจำแนกตามกลุ่มความยาก และความเป็นมิติเดียวของข้อสอบแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .05 และพบว่าเมื่อจำแนกโดยพิจารณาจากความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศให้ค่าที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .05 เฉพาะค่าที่ได้จากการให้คะแนนแบบปกติเท่านั้น ส่วนค่าที่ได้จากการให้คะแนนแบบ 2 - SR พบว่าให้ผลที่ไม่แตกต่างกัน สำหรับการจำแนกโดยสัมประสิทธิ์การจำแนก

ดัชนีการจำแนก และความเหมาะสมของข้อสอบกับโมเดลพบว่าค่าที่ได้จากการให้คะแนนทั้งสองวิธี ให้ผลไม่แตกต่างกัน และเมื่อพิจารณาจากตัวแปรค่าความยากแบบ IRT พบว่าไม่มีข้อสอบข้อใดถูกคัดออกจากฉบับโดยเกณฑ์ความยากแบบราซซีโมเดล

ตาราง 21 ค่าสารสนเทศของข้อสอบที่ ระดับความสามารถต่ำของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนโดยวิธีแบบปกติ และแบบ 2-SR จำแนกตามตัวแปรต่างๆที่ใช้ในการคัดเลือกข้อสอบเข้าฉบับตามกรอบทฤษฎี CTT และ IRT

		0 - 1				2 - SR			
		K	IIFL	F- test	p	K	IIFL	F - test	p
P	POUT	5	.1384	12.6**	0.001	3	.0367	2.504	0.126
	PIN	35	.0492			37	.0409		
r	rOUT	9	.0255	4.26*	.0458	9	.0378	7.82**	0.008
	riN	31	.0705			31	.0474		
D	DOUT	10	.0211	5.59*	.023	7	.0389	5.37*	0.026
	DIN	30	.0718			33	.0471		
B	BOUT	7	.0876	1.793	0.188	9	.0444	.0756	0.784
	BIN	33	.0546			31	.0455		
U	UOUT	22	.0300	46.2**	0.000	22	.0473	2.112	0.154
	UIN	18	.1234			18	.0428		
F	FOUT	12	.0577	.0586	.8100	15	.0447	0.082	0.776
	FIN	28	.0623			25	.0456		
	BTOUT		-	-	-	-	-	-	-
	BTIN	40	.0604			40	.0460		

* มีนัยสำคัญทางสถิติ ณ ระดับ .05 ** มีนัยสำคัญทางสถิติ ณ ระดับ .01

จากตาราง 21 พบว่าค่าเฉลี่ยของค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ได้จากการให้คะแนนตามรูปแบบปกติ และรูปแบบ 2 - SR เมื่อจำแนกตามสัมประสิทธิ์การจำแนก และดัชนีการจำแนก

แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .05 และพบว่าเมื่อจำแนกโดยพิจารณาจากความยาก และความเป็นมิติเดียว แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01 เฉพาะค่าที่ได้จากการให้คะแนนแบบปกติเท่านั้น ส่วนค่าที่ได้จากการให้คะแนนแบบ 2 - SR พบว่าให้ผลที่ไม่แตกต่างกัน สำหรับการจำแนกโดยความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ และความเหมาะสมของข้อสอบกับโมเดลพบว่าให้ผลไม่แตกต่างกัน และเมื่อพิจารณาจากตัวแปรค่าความยากแบบ IRT พบว่าไม่มีข้อสอบข้อใดถูกคัดออกจากฉบับโดยเกณฑ์ความยากแบบราชคโสมิเดล

จากผลการวิเคราะห์ข้างต้นสรุปได้ว่าการให้คะแนนตามรูปแบบ 2-SR ทำให้ค่าสัมประสิทธิ์การจำแนก ดัชนีการจำแนก ความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ ความเป็นมิติเดียว ความเชื่อมั่นรายข้อ ค่าสารสนเทศของข้อสอบ และประสิทธิภาพของสารสนเทศของแบบทดสอบที่ระดับความสามารถปานกลาง ความเชื่อมั่นของแบบทดสอบที่คัดเลือกด้วยตัวแปรเกณฑ์ต่างๆ ตามกรอบทฤษฎี CTT สูงกว่าค่าที่ได้จากการให้คะแนนตามรูปแบบปกติ อย่างมีนัยสำคัญทางสถิติ ณ ระดับ .05 แต่พบว่าค่าความยากแบบ CTT สารสนเทศของข้อสอบ และประสิทธิภาพของสารสนเทศของแบบทดสอบที่ระดับความสามารถสูง และต่ำ มีค่าน้อยกว่าค่าที่ได้จากการให้คะแนนตามรูปแบบปกติ อย่างมีนัยสำคัญทางสถิติ ณ ระดับ .05 ส่วนค่าความเชื่อมั่นของแบบทดสอบพบว่าการให้คะแนนตามรูปแบบทั้งสองให้ผลที่ไม่แตกต่างกัน

การศึกษารูปแบบการให้คะแนน และวิธีการคัดเลือกข้อสอบที่ส่งผลต่อคุณภาพของแบบทดสอบ

ในการศึกษานี้สิ่งที่ผู้วิจัยสนใจก็คือ ความสามารถในการอธิบายความแปรปรวนของคุณภาพของแบบทดสอบร่วมกันระหว่างรูปแบบการให้คะแนน และวิธีการคัดเลือกข้อสอบเพื่อให้ได้ข้อสรุปที่จะนำไปสู่รูปแบบในการพัฒนาคุณภาพของแบบทดสอบเลือกตอบตามกรอบทฤษฎี CTT และ IRT ที่เหมาะสม ดังนั้นจึงขอนำเสนอผลการวิเคราะห์เรียงตามหัวข้อดังนี้

1. ความสัมพันธ์ระหว่างตัวแปรที่ใช้ในการคัดเลือกข้อสอบ ความเชื่อมั่นรายข้อ และสารสนเทศของข้อสอบ ทั้งตามกรอบทฤษฎี CTT และ IRT
2. สหสัมพันธ์พหุคูณ และตัวแปรพยากรณ์ตามกรอบทฤษฎี CTT
3. สหสัมพันธ์พหุคูณ และตัวแปรพยากรณ์ตามกรอบทฤษฎี IRT
4. สหสัมพันธ์พหุคูณ และตัวแปรพยากรณ์ตามกรอบทฤษฎี CTT และ IRT

ตาราง 22 ค่าสหสัมพันธ์ระหว่างตัวแปรที่ใช้ในการคัดเลือกข้อสอบเข้าฉบับ ความเชื่อมั่นรายข้อ
 สารสนเทศของข้อสอบ ของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนน
 ตามรูปแบบปกติ

	P	r	D	B	U	RII	BT A	F	IIFH	IIFM	IIFL
P	1.00										
r	.88**	1.00									
D	.78**	.92**	1.00								
B	.44**	-.55**	.44**	1.00							
U	.91**	.91**	.82**	.53**	1.00						
RII	.84**	.98**	.94**	.52**	.86**	1.00					
BT A	-.96**	-.84**	-.77**	-.42**	-.91**	-.77**	1.00				
F	.54**	-.72**	-.61**	-.55**	-.58**	-.74**	.43**	1.00			
IIFH	-.91**	-.82**	-.84**	-.32	-.88**	-.80**	.94**	.34	1.00		
IIFM	.22	.27	.49**	-.14	.25	.38	-.19	.09	-.51**	1.00	
IIFL	.86**	.71**	.56**	.48**	.80**	.61**	-.91**	-.46*	-.72**	-.22	1.00

* มีนัยสำคัญทางสถิติ ณ ระดับ .05 ** มีนัยสำคัญทางสถิติ ณ ระดับ .01

จากตาราง 22 พบว่าค่าความเชื่อมั่นรายข้อของแบบทดสอบวัดความสามารถในการอ่าน
 จับใจความที่ได้จากการให้คะแนนแบบปกติมีความสัมพันธ์เชิงบวกอย่างมีนัยสำคัญทางสถิติ
 ณ ระดับ .01 กับสัมประสิทธิ์การจำแนก ดัชนีการจำแนก ความเป็นมิติเดียว ค่าความยากแบบ
 CTT สารสนเทศของข้อสอบที่ระดับความสามารถต่ำ และความลำเอียงเนื่องจากตัวแปรเพศ ซึ่งมี
 ค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .98 , .94 , .86 , .84 , .61 และ .52 ตามลำดับ และมีความ
 สัมพันธ์เชิงลบอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01 กับสารสนเทศของข้อสอบที่ระดับความ
 สามารถสูง ค่าความยากแบบ IRT และค่าความเหมาะสมของข้อสอบกับโมเดล ซึ่งมีค่า
 สัมประสิทธิ์สหสัมพันธ์เท่ากับ -.80 , -.77 , และ - .74 ตามลำดับ

สำหรับค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูง พบว่ามีความสัมพันธ์เชิงบวก
 อย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01 กับ ค่าความยากแบบ IRT ซึ่งมีค่าสัมประสิทธิ์
 สหสัมพันธ์เท่ากับ .94 และมีความสัมพันธ์เชิงลบอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01 กับ

ค่าความยากแบบ CTT ความเป็นมิติเดียว ดัชนีการจำแนก สัมประสิทธิ์การจำแนก และความเชื่อมั่นรายข้อซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ $-.91$, $-.88$, $-.84$, $-.82$, $-.80$ ตามลำดับ

ส่วนค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลาง พบว่ามีความสัมพันธ์เชิงบวกอย่างมีนัยสำคัญทางสถิติ ณ ระดับ $.01$ กับดัชนีการจำแนกซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ $.49$ และมีความสัมพันธ์เชิงลบอย่างมีนัยสำคัญทางสถิติ ณ ระดับ $.01$ กับค่าสารสนเทศของข้อสอบ ที่ระดับความสามารถสูง ซึ่งมีค่าเท่ากับ $-.51$ •

นอกจากนั้นพบว่า ค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำมีความสัมพันธ์เชิงบวกอย่างมีนัยสำคัญทางสถิติ ณ ระดับ $.01$ กับค่าความยากแบบ CTT ความเป็นมิติเดียว สัมประสิทธิ์การจำแนก ความเชื่อมั่นของข้อสอบรายข้อ ดัชนีการจำแนก และความลำเอียงเนื่องจากตัวแปรเพศ ซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ $.86$, $.80$, $.71$, $.61$, $.56$ และ $.48$ ตามลำดับ และมีความสัมพันธ์เชิงลบอย่างมีนัยสำคัญทางสถิติ ณ ระดับ $.01$ กับค่าความยากแบบ IRT ค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูง และค่าความเหมาะสมของข้อสอบกับโมเดล ซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์ เท่ากับ $-.91$, $-.72$ และ $-.46$ ตามลำดับ

ตาราง 23 ค่าสหสัมพันธ์ระหว่างตัวแปรที่ใช้ในการคัดเลือกข้อสอบเข้าฉบับ ความเชื่อมั่นรายข้อ
 สารสนเทศของข้อสอบของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนน
 ตามรูปแบบ 2 - SR

	P	r	D	B	U	RII	BTA	F	IIFH	IIFM	IIFL
P	1.00										
r	.88**	1.00									
D	.78**	.95**	1.00								
B	.23	.34*	.29**	1.00							
U	.94**	.97**	.91**	.33**	1.00						
RII	.83**	.98**	.99**	.34**	.94**	1.00					
BTA	-.99**	-.87**	-.76**	-.23	-.93**	-.81**	1.00				
F	-.65**	-.85**	-.82**	-.41**	-.77**	-.87**	.61**	1.00			
IIFH	-.98**	-.88**	-.80**	-.20	-.94**	-.83**	.99**	.59**	1.00		
IIFM	.25	.35**	.54**	-.14	.35**	.43**	-.26	-.03	-.38**	1.00	
IIFL	.99**	.85**	.71**	.25	.91**	.77**	-.99**	-.62**	-.97**	.13	1.00

* มีนัยสำคัญทางสถิติ ณ ระดับ .05 ** มีนัยสำคัญทางสถิติ ณ ระดับ .01

จากตาราง 23 พบว่าค่าความเชื่อมั่นรายข้อของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนแบบ 2 - SR มีความสัมพันธ์เชิงบวกอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01 กับ ดัชนีการจำแนก สัมประสิทธิ์การจำแนก ความเป็นมิติเดียว ค่าความยากแบบ CTT ค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำ สารสนเทศของข้อสอบที่ระดับความสามารถปานกลาง และความลำเอียงเนื่องจากตัวแปรเพศ ซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .99 , .98 , .94 , .83 , .77, 43 และ .34 ตามลำดับ และมีความสัมพันธ์เชิงลบอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01 กับค่าความเหมาะสมของข้อสอบกับโมเดล ค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูง และความยากแบบ IRT ซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ -.87 , -.83 , และ -.81 ตามลำดับ

สำหรับค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูง พบว่ามีความสัมพันธ์เชิงบวกอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01 กับ ค่าความยากแบบ IRT ซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์

เท่ากับ .99 และมีความสัมพันธ์เชิงลบอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01 กับค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำ ค่าความยากแบบ CTT ความเป็นมิติเดียว สัมประสิทธิ์การจำแนก ความเชื่อมั่นรายข้อ และดัชนีการจำแนก ซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ -.99 , -.98 , -.94 , -.88 , -.83 และ -.80 ตามลำดับ

ส่วนค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลาง พบว่ามีความสัมพันธ์เชิงบวกอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01 กับดัชนีการจำแนก ค่าความเชื่อมั่นรายข้อ สัมประสิทธิ์การจำแนก และความเป็นมิติเดียว ซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .54 , .43 , .35 , และ .35 ตามลำดับ และมีความสัมพันธ์เชิงลบกับค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูง ซึ่งมีค่าเท่ากับ -.38

นอกจากนั้นพบว่า ค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำมีความสัมพันธ์เชิงบวกอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01 กับค่าความยากแบบ CTT ความเป็นมิติเดียว สัมประสิทธิ์การจำแนก ความเชื่อมั่นของข้อสอบรายข้อ และดัชนีการจำแนก ซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .99 , .91 , .85 , .77 , และ .71 ตามลำดับ และมีความสัมพันธ์เชิงลบอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01 กับค่าความยากแบบ IRT ค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูง และค่าความเหมาะสมของข้อสอบกับโมเดล ซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ -.99 , -.97 และ -.62 ตามลำดับ

ตาราง 24 ค่าสหสัมพันธ์ระหว่างตัวแปรที่ใช้ในการคัดเลือกข้อสอบเข้าฉบับ ความเชื่อมั่นรายข้อ
 สารสนเทศของข้อสอบของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์
 ที่ให้คะแนนตามรูปแบบปกติ

	P	r	D	B	U	RII	BTA	F	IIFH	IIFM	IIFL
P	1.00										
r	.50**	1.00									
D	.46**	.95**	1.00								
B	.15	-.01	-.13	1.0							
U	.91**	.68**	.67**	.08	1.00						
RII	.40**	.98**	.98**	-.07	.61**	1.00					
BTA	-.99**	-.93**	-.42**	-.15	-.89**	-.36	1.00				
F	-.31	-.24	-.28	.03	-.30	-.22	.31	1.00			
IIFH	-.88**	-.53**	-.55**	.03	-.80**	-.48**	.89**	.32	1.00		
IIFM	-.41**	.03	.20	-.40	-.35	.17	.41**	.01	-.05	1.00	
IIFL	.94**	.36	.25	.25	.84**	.22	-.94**	-.25	-.68**	-.69**	1.00

* มีนัยสำคัญทางสถิติ ณ ระดับ .05 ** มีนัยสำคัญทางสถิติ ณ ระดับ .01

จากตาราง 24 พบว่าค่าความเชื่อมั่นรายข้อของแบบทดสอบวัดความสามารถในการ
 แก้โจทย์ปัญหาทางคณิตศาสตร์ที่ได้จากการให้คะแนนแบบปกติมีความสัมพันธ์เชิงบวกอย่างมีนัย
 สำคัญทางสถิติ ณ ระดับ .01 กับ ดัชนีการจำแนก สัมประสิทธิ์การจำแนก ความเป็นมิติเดียว และ
 ค่าความยากแบบ CTT ซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .98 , .98 , .61 , และ .40 ตามลำดับ
 และมีความสัมพันธ์เชิงลบอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01 กับค่าสารสนเทศของข้อสอบที่
 ระดับความสามารถสูง ซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ -.48

สำหรับค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูง พบว่ามีความสัมพันธ์เชิงบวก
 อย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01 กับ ค่าความยากแบบ IRT ซึ่งมีค่าสัมประสิทธิ์
 สหสัมพันธ์เท่ากับ .89 และมีความสัมพันธ์เชิงลบอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01 กับค่า
 ความยากแบบ CTT ความเป็นมิติเดียว ค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำ
 ค่าดัชนีการจำแนก สัมประสิทธิ์การจำแนก และความเชื่อมั่นรายข้อ ซึ่งมีค่าสัมประสิทธิ์

สหสัมพันธ์เท่ากับ $-.88$, $-.80$, $-.68$, $-.55$, $-.53$ และ $-.48$ ตามลำดับ

ส่วนค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลาง พบว่ามีความสัมพันธ์เชิงบวกอย่างมีนัยสำคัญทางสถิติ ณ ระดับ $.01$ กับค่าความยากแบบ IRT :ซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ $.41$ และมีความสัมพันธ์เชิงลบกับค่าความยากแบบ CTT และความเหมาะสมของข้อสอบกับโมเดลซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ $-.41$ และ $-.40$ ตามลำดับ

นอกจากนั้นพบว่า ค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำมีความสัมพันธ์เชิงบวกอย่างมีนัยสำคัญทางสถิติ ณ ระดับ $.01$ กับค่าความยากแบบ CTT และความเป็นมิติเดียวซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ $.94$ และ $.84$ ตามลำดับ และมีความสัมพันธ์เชิงลบอย่างมีนัยสำคัญทางสถิติ ณ ระดับ $.01$ กับค่าความยากแบบ IRT ค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลาง และ ค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูง ซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์ เท่ากับ $-.94$, $-.69$ และ $-.68$ ตามลำดับ

ตาราง 25 ค่าสหสัมพันธ์ระหว่างตัวแปรที่ใช้ในการคัดเลือกข้อสอบเข้าฉบับ ความเชื่อมั่นรายข้อ
 สารสนเทศของข้อสอบของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิต
 ศาสตร์ที่ให้คะแนนตามรูปแบบ 2 - SR

	P	r	D	B	U	RII	BTA	F	IIFH	IIFM	IIFL
P	1.00										
r	.54**	1.00									
D	.66**	.84**	1.00								
B	.18	.04	.09	1.00							
U	-.17	.63**	.47**	.12	1.00						
RII	.43**	.98**	.83**	.02	.67**	1.00					
BTA	-.92**	-.52**	-.47**	-.10	.25	-.41**	1.00				
F	-.43**	-.92**	-.74**	.03	-.58**	-.93**	.44**	1.00			
IIFH	-.92**	-.55**	-.50**	-.12	.21	-.44**	.99**	.46**	1.00		
IIFM	-.39**	.02	.07	.04	.40**	.18	.43**	.06	.31**	1.00	
IIFL	.92**	.49**	.43**	.09	-.29	.35**	-.99**	-.42**	-.97**	-.54**	1.00

* มีนัยสำคัญทางสถิติ ณ ระดับ .05 ** มีนัยสำคัญทางสถิติ ณ ระดับ .01

จากตาราง 25 พบว่าค่าความเชื่อมั่นรายข้อของแบบทดสอบวัดความสามารถในการแก้
 โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนแบบ 2 - SR มีความสัมพันธ์เชิงบวกอย่างมีนัยสำคัญทาง
 สถิติ ณ ระดับ .01 กับสัมประสิทธิ์การจำแนก ดัชนีการจำแนก ความเป็นมิติเดียว ค่าความยาก
 แบบ CTT และค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำ ซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์
 เท่ากับ .98 , .83 , .67 , .43 , และ .35 ตามลำดับ และมีความสัมพันธ์เชิงลบอย่างมีนัยสำคัญทาง
 สถิติ ณ ระดับ .01 กับค่าความเหมาะสมของข้อสอบกับโมเดล ค่าสารสนเทศของข้อสอบที่ระดับ
 ความสามารถสูง และความยากแบบ IRT ซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ -.93 , -.44 , และ -
 .41 ตามลำดับ

สำหรับค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูง พบว่ามีความสัมพันธ์เชิงบวก
 อย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01 กับ ค่าความยากแบบ IRT ค่าความเหมาะสมของ

โมเดลกับข้อสอบ และค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลาง ซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .99 , .46 และ .31 ตามลำดับ และมีความสัมพันธ์เชิงลบอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01 กับค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำ ค่าความยากแบบ CTT สัมประสิทธิ์การจำแนก ดัชนีการจำแนก และความเชื่อมั่นรายข้อ ซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ -.97 , -.92 , -.55 , -.50 , และ -.44 ตามลำดับ

ส่วนค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลาง พบว่ามีความสัมพันธ์เชิงบวกอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01 กับ ความยากแบบ IRT ความเป็นมิติเดียว และค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลางซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .43 , .40 และ .31 ตามลำดับ และมีความสัมพันธ์เชิงลบอย่างมีนัยสำคัญทางสถิติ ณ ระดับ.01 กับค่าความยากแบบ CTT ซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ -.39

นอกจากนั้นพบว่า ค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำมีความสัมพันธ์เชิงบวกอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01 กับค่าความยากแบบ CTT ความเป็นมิติเดียว สัมประสิทธิ์การจำแนก ดัชนีการจำแนก และความเชื่อมั่นของข้อสอบรายข้อ ซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .92 , .49 , .43 , และ .35 ตามลำดับ และมีความสัมพันธ์เชิงลบอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01 กับค่าความยากแบบ IRT ค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูง และระดับความสามารถปานกลาง และความสามารถของข้อสอบกับโมเดล ซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์ เท่ากับ -.99 , -.97, -.54 และ .62 ตามลำดับ

ตาราง 26 สมการพยากรณ์ สหสัมพันธ์พหุคูณ สัมประสิทธิ์การพยากรณ์ ประสิทธิภาพในการพยากรณ์ ความคลาดเคลื่อนมาตรฐานในการประมาณค่า และรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนของสมการพยากรณ์ค่าความเชื่อมั่นของแบบทดสอบวัดความสามารถในการอ่านจับใจความและการแก้โจทย์ปัญหาทางคณิตศาสตร์ ที่ได้จากการให้คะแนนแบบปกติ และแบบ 2-SR โดยใช้ตัวแปรคัดเลือกข้อสอบตามกรอบทฤษฎี CCT

แบบทดสอบ	รูปแบบการให้คะแนน	สมการพยากรณ์	R	R ²	E	SE	REM	F-test
อ่านจับใจความ	ปกติ	.41r** + .12D** - .04U* - .0048	.988	.975	84	.001	.000	490**
	2 - SR	1.49D** + 1.04r** - .0549**	.997	.994	92	.033	.001	3136**
การแก้โจทย์ปัญหา	ปกติ	.33r** + .15D** - .03P** + .0030**	.994	.988	89	.001	.000	1053**
	2 - SR	2.45r** + .07**	.936	.876	65	.145	.021	276**

** มีนัยสำคัญทางสถิติ ณ ระดับ .01

จากตาราง 26 พบว่าสมการพยากรณ์ค่าความเชื่อมั่นของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนแบบปกติประกอบด้วยตัวแปรคือ สัมประสิทธิ์การจำแนก ดัชนีการจำแนก และความเป็นมิติเดียวซึ่งมีระดับความสัมพันธ์พหุคูณเท่ากับ .988 สัมประสิทธิ์การพยากรณ์เท่ากับ .975 และประสิทธิภาพในการพยากรณ์เท่ากับ 84%

ส่วนสมการพยากรณ์ค่าความเชื่อมั่นของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนแบบ 2 - SR พบว่าประกอบด้วยตัวแปรคือ สัมประสิทธิ์ในการจำแนก และดัชนีการจำแนก ซึ่งมีระดับความสัมพันธ์พหุคูณเท่ากับ .997 สัมประสิทธิ์การพยากรณ์เท่ากับ .994 และประสิทธิภาพในการพยากรณ์เท่ากับ 92 %

สำหรับสมการพยากรณ์ค่าความเชื่อมั่นของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนแบบปกติ พบว่าตัวแปรในสมการคือ สัมประสิทธิ์การ

จำแนก ดัชนีการจำแนก และค่าความยาก ซึ่งมีระดับความสัมพันธ์พหุคูณเท่ากับ .994สัมประสิทธิ์พยากรณ์เท่ากับ .988 และประสิทธิภาพในการพยากรณ์เท่ากับ 89 %

และเมื่อพิจารณาสมการพยากรณ์ค่าความเชื่อมั่นของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนแบบ 2 - SR พบว่าตัวแปรในสมการคือสัมประสิทธิ์การจำแนก ซึ่งมีระดับความสัมพันธ์พหุคูณเท่ากับ .936 สัมประสิทธิ์พยากรณ์เท่ากับ .876 และประสิทธิภาพในการพยากรณ์เท่ากับ 65 %

ตาราง 27 สมการพยากรณ์ สหสัมพันธ์พหุคูณ สัมประสิทธิ์การพยากรณ์ ประสิทธิภาพในการพยากรณ์ ความคลาดเคลื่อนมาตรฐานในการประมาณค่า และรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนของสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูง (IIFH) ของแบบทดสอบวัดความสามารถในการอ่านจับใจความและการแก้โจทย์ปัญหาทางคณิตศาสตร์ ที่ได้จากการให้คะแนนแบบปกติ และแบบ 2-SR โดยใช้ตัวแปรในการคัดเลือกข้อสอบตามกรอบทฤษฎี IRT

แบบทดสอบ	รูปแบบการให้คะแนน	สมการพยากรณ์	R	R ²	E	SE	REM	F-test
อ่านจับใจความ	ปกติ	.05BTA** + .0651**	.936	.876	65	.022	.0005	277**
	2 - SR	.04BTA** + .0464**	.992	.984	87	.000	.0000	2376**
การแก้โจทย์ปัญหา	ปกติ	.29B* - .04BTA** + .051**	.901	.812	57	.019	.0004	79**
	2 - SR	.04BTA** - .002U* + .05**	.992	.985	88	.001	.0000	1176**

** มีนัยสำคัญทางสถิติ ณ ระดับ .01

จากตาราง 27 พบว่าสมการพยากรณ์ค่าสารสนเทศที่ระดับความสามารถสูงของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนแบบปกติ และแบบ 2 - SR ประกอบด้วยตัวแปรได้แก่ ค่าความยาก ซึ่งมีระดับความสัมพันธ์พหุคูณเท่ากับ .936 และ .992 สัมประสิทธิ์การพยากรณ์เท่ากับ .876 และ .984 และประสิทธิภาพในการพยากรณ์เท่ากับ 65 และ 87 % ตามลำดับ

สำหรับสมการพยากรณ์ค่าสารสนเทศที่ระดับความสามารถสูงของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนแบบปกติ พบว่าตัวแปรในสมการได้แก่ ความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ และค่าความยาก ซึ่งมีระดับความสัมพันธ์พหุคูณเท่ากับ .901 สัมประสิทธิ์พยากรณ์เท่ากับ .812 และประสิทธิภาพในการพยากรณ์เท่ากับ 57 %

และเมื่อพิจารณาสมการพยากรณ์ค่าสารสนเทศที่ระดับความสามารถสูงของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนแบบ 2 - SR พบว่าตัวแปรในสมการได้แก่ ค่าความยาก และความเป็นมิติเดียวซึ่งมีระดับความสัมพันธ์พหุคูณเท่ากับ .992 สัมประสิทธิ์พยากรณ์เท่ากับ .985 และประสิทธิภาพในการพยากรณ์เท่ากับ 88 %

ตาราง 28 สมการพยากรณ์ สหสัมพันธ์พหุคูณ สัมประสิทธิ์การพยากรณ์ ประสิทธิภาพในการพยากรณ์ ความคลาดเคลื่อนมาตรฐานในการประมาณค่า และรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนของสมการพยากรณ์ ค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลาง (IIFM) ของแบบทดสอบวัดความสามารถในการอ่านจับใจความและการแก้โจทย์ปัญหาคณิตศาสตร์ ที่ได้จากการให้คะแนนแบบปกติ และแบบ 2-SR โดยใช้ตัวแปรในการคัดเลือกข้อสอบตามกรอบทฤษฎี IRT

แบบทดสอบ	รูปแบบการให้คะแนน	สมการพยากรณ์	R	R ²	SE	REM	F-TEST	
อ่านจับใจความ	ปกติ	.1828 - .037BTA**	.948	.899	68	.015	.002	348**
	2 - SR	.06U** + .03BTA** + .002F** + .2319**	.663	.439	25	.000	.000	9**
การแก้โจทย์ปัญหา	ปกติ	.02BTA** - .81B** + .2358	.537	.288	16	.047	.0022	7.5**
	2 - SR	.008BTA* - .008U* + .245**	.524	.275	15	.004	.0002	7.0**

** มีนัยสำคัญทางสถิติ ณ ระดับ .01

จากตาราง 28 พบว่าสมการพยากรณ์ค่าสารสนเทศที่ระดับความสามารถปานกลาง

ของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนแบบปกติ ประกอบด้วยตัวแปรคือ ค่าความยาก ซึ่งมีระดับสหสัมพันธ์พหุคูณเท่ากับ .948 สัมประสิทธิ์พยากรณ์เท่ากับ .899 และประสิทธิภาพในการพยากรณ์เท่ากับ 68 %

ส่วนสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลางของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนแบบ 2 - SR ประกอบด้วยตัวแปร ความเป็นมิติเดียวของข้อสอบ ความยาก และความเหมาะสมของข้อสอบกับโมเดล ซึ่งมีระดับสหสัมพันธ์พหุคูณเท่ากับ .663 สัมประสิทธิ์พยากรณ์เท่ากับ .439 และประสิทธิภาพในการพยากรณ์เท่ากับ 25 %

สำหรับสมการพยากรณ์ค่าสารสนเทศที่ระดับความสามารถปานกลางของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนแบบปกติ พบว่าตัวแปรในสมการได้แก่ ความยาก และความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ ซึ่งมีระดับความสัมพันธ์พหุคูณเท่ากับ .537 สัมประสิทธิ์พยากรณ์เท่ากับ .288 และประสิทธิภาพในการพยากรณ์เท่ากับ 16 %

และเมื่อพิจารณาสมการพยากรณ์ค่าสารสนเทศที่ระดับความสามารถปานกลางของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนแบบ 2 - SR พบว่าตัวแปรในสมการได้แก่ ค่าความยาก และความเป็นมิติเดียวซึ่งมีระดับความสัมพันธ์พหุคูณเท่ากับ .527 สัมประสิทธิ์พยากรณ์เท่ากับ .275 และประสิทธิภาพในการพยากรณ์เท่ากับ 15 %

ตาราง 29 สมการพยากรณ์ สหสัมพันธ์พหุคูณ สัมประสิทธิ์การพยากรณ์ ประสิทธิภาพในการพยากรณ์ ความคลาดเคลื่อนมาตรฐานในการประมาณค่า และรากที่สองของค่าเฉลี่ยคลาดเคลื่อนของสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำ (IIFL) ของแบบทดสอบวัดความสามารถในการอ่านจับใจความและการแก้โจทย์ปัญหาทางคณิตศาสตร์ ที่ได้จากการให้คะแนนแบบปกติ และแบบ 2-SR โดยใช้ตัวแปรในการคัดเลือกข้อสอบตามกรอบทฤษฎี IRT

แบบทดสอบ	รูปแบบการให้คะแนน	สมการพยากรณ์	R	R ²	E	SE	REM	F-TEST
อ่านจับใจความ	ปกติ	.065 - .041BTA**	.912	.832	59	.022	.0005	193**
	2 - SR	.046 - .04BTA**	.992	.983	87	.000	.0000	2275**
การแก้โจทย์ปัญหา	ปกติ	.29B** - .05BTA** + .051**	.950	.902	69	.019	.0004	169**
	2 - SR	.046** - .044BTA** - .002U*	.994	.987	88	.001	.0000	1400**

** มีนัยสำคัญทางสถิติ ณ ระดับ .01

จากตาราง 29 พบว่าสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนแบบปกติ และแบบ 2 - SR ประกอบด้วยตัวแปรได้แก่ ความยากซึ่งมีระดับความสัมพันธ์พหุคูณเท่ากับ .912 และ .992 สัมประสิทธิ์การพยากรณ์เท่ากับ .832 และ .983 และประสิทธิภาพในการพยากรณ์เท่ากับ 59 % และ 87 % ตามลำดับ

สำหรับสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนแบบปกติ พบว่าตัวแปรในสมการได้แก่ ความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ และค่าความยาก ซึ่งมีระดับความสัมพันธ์พหุคูณเท่ากับ .950 สัมประสิทธิ์พยากรณ์เท่ากับ .902 และประสิทธิภาพในการพยากรณ์เท่ากับ 69 %

และเมื่อพิจารณาสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนแบบ 2 - SR

พบว่าตัวแปรในสมการได้แก่ ค่าความยาก และความเป็นมิติเดียวซึ่งมีระดับความสัมพันธ์พหุคูณเท่ากับ .994 สัมประสิทธิ์พยากรณ์เท่ากับ .987 และประสิทธิภาพในการพยากรณ์เท่ากับ 88 %

ตาราง 30 สมการพยากรณ์ สหสัมพันธ์พหุคูณ สัมประสิทธิ์การพยากรณ์ ประสิทธิภาพในการพยากรณ์ ความคลาดเคลื่อนมาตรฐานในการประมาณค่า และรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนของสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูง (IIFH) ของแบบทดสอบวัดความสามารถในการอ่านจับใจความและการแก้โจทย์ปัญหาทางคณิตศาสตร์ ที่ได้จากการให้คะแนนแบบปกติ และแบบ 2-SR โดยใช้ตัวแปรในการคัดเลือกข้อสอบตามกรอบทฤษฎี CTT และ IRT

แบบทดสอบ	รูปแบบการให้คะแนน	สมการพยากรณ์	R	R ²	E	SE	REM	F-TEST
อ่านจับใจความ	ปกติ	.02BTA*-.24D**-.12P* -.01F**+.1759**	.977	.955	79	.014	.0002	187**
	2 - SR	.04BTA**-.02D**-.001F .051**	.999	.998	95		.000	5799**
การแก้โจทย์ปัญหา	ปกติ	.54P**+.17r*+.14BTA** -.22D**+.15*	.940	.880	65	.015	.0002	66**
	2 - SR	.04BTA** .002U* +.047**	.992	.985	88	.001	.0000	1177**

** มีนัยสำคัญทางสถิติ ณ ระดับ .01

จากตาราง 30 พบว่าสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูงของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนแบบปกติประกอบด้วยตัวแปรคือ ความยากแบบ CTT และ IRT ดัชนีการจำแนก และความเหมาะสมของข้อสอบกับโมเดลซึ่งมีระดับความสัมพันธ์พหุคูณเท่ากับ .977 สัมประสิทธิ์การพยากรณ์เท่ากับ .955 และประสิทธิภาพในการพยากรณ์เท่ากับ 79 %

ส่วนสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูงของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนแบบ 2 - SR พบว่าประกอบด้วยตัวแปรได้แก่ ความยากแบบ IRT ดัชนีการจำแนก และความเหมาะสมของข้อสอบกับโมเดลซึ่งมีระดับความสัมพันธ์พหุคูณเท่ากับ .999 สัมประสิทธิ์การพยากรณ์เท่ากับ .998 และประสิทธิภาพในการพยากรณ์เท่ากับ 95 %

สำหรับสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูงของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนแบบปกติ พบว่าตัวแปรในสมการได้แก่ความยากแบบ CTT และ IRT สัมประสิทธิ์การจำแนก และดัชนีการจำแนก ซึ่งมีระดับความสัมพันธ์พหุคูณเท่ากับ .940 สัมประสิทธิ์พยากรณ์เท่ากับ .880 และประสิทธิภาพในการพยากรณ์เท่ากับ 65 %

และเมื่อพิจารณาสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูงของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนแบบ 2 - SR พบว่าตัวแปรในสมการได้แก่ ความยากแบบ IRT และ ความเป็นมิติเดียว ซึ่งมีระดับความสัมพันธ์พหุคูณเท่ากับ .992 สัมประสิทธิ์พยากรณ์เท่ากับ .985 และประสิทธิภาพในการพยากรณ์เท่ากับ 88 %

ตาราง 31 สมการพยากรณ์ สัมพันธ์พันธุ์พหุคูณ สัมประสิทธิ์การพยากรณ์ ประสิทธิภาพในการพยากรณ์ ความคลาดเคลื่อน มาตรฐานในการประมาณค่า และรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนของสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลาง (IIFM) ของแบบทดสอบวัดความสามารถในการอ่านจับใจความและการแก้โจทย์ปัญหาทางคณิตศาสตร์ ที่ได้จากการให้คะแนนแบบปกติ และแบบ 2-SR โดยใช้ตัวแปรในการคัดเลือกข้อสอบตามกรอบทฤษฎี CTT และ IRT

แบบทดสอบ	รูปแบบการให้คะแนน	สมการพยากรณ์	R	R ²	E	SE	REM	F-TEST
อ่านจับใจความ	ปกติ	.83D*+ .44P**+ .08BTA** +.01F** - .32r** - .1369*	.843	.711	46	.033	.001	16**
	2 - SR	.06D** + .002F** - .01P** +0.235**	.930	.864	63	.000	.0000	78**
การแก้โจทย์ปัญหา	ปกติ	.59D** - .18BTA** - 1.19P** - .49r** + .67**	.750	.560	.34	.037	.0013	11**
	2 - SR	.01BTA** + .01D* + .244**	.530	.281	.15	.004	.0002	7**

** มีนัยสำคัญทางสถิติ ณ ระดับ .01

จากตาราง 31 พบว่าสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลางของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนแบบปกติประกอบด้วยตัวแปรได้แก่ ดัชนีการจำแนก ความยากแบบ CTT และ IRT ความเหมาะสมของข้อสอบกับโมเดล และสัมประสิทธิ์การจำแนก ซึ่งมีระดับความสัมพันธ์พหุคูณเท่ากับ .843 สัมประสิทธิ์การพยากรณ์เท่ากับ .711 และประสิทธิภาพในการพยากรณ์เท่ากับ 46 %

ส่วนสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลางของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนแบบ 2 - SR พบว่าประกอบด้วยตัวแปรได้แก่ ดัชนีการจำแนก ความเหมาะสมของข้อสอบกับโมเดลและความยาก ซึ่งมีระดับความสัมพันธ์พหุคูณเท่ากับ .930 สัมประสิทธิ์การพยากรณ์เท่ากับ .864 และประสิทธิภาพในการพยากรณ์เท่ากับ 63 %

สำหรับสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลางของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนแบบปกติ พบว่าตัวแปรในสมการได้แก่ ดัชนีการจำแนก ความยากแบบ CTT และ IRT และสัมประสิทธิ์การจำแนก ซึ่งมีระดับความสัมพันธ์พหุคูณเท่ากับ .750 สัมประสิทธิ์พยากรณ์เท่ากับ .560 และประสิทธิภาพในการพยากรณ์เท่ากับ 34 %

และเมื่อพิจารณาสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลางของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนแบบ 2 - SR พบว่าตัวแปรในสมการได้แก่ ความยากแบบ IRT และดัชนีการจำแนก ซึ่งมีระดับความสัมพันธ์พหุคูณเท่ากับ .530 สัมประสิทธิ์พยากรณ์เท่ากับ .281 และประสิทธิภาพในการพยากรณ์เท่ากับ 15 %

ตาราง 32 สมการพยากรณ์ สหสัมพันธ์พหุคูณ สัมประสิทธิ์การพยากรณ์ ประสิทธิภาพในการพยากรณ์ ความคลาดเคลื่อนมาตรฐานในการประมาณค่า และรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนของสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำ (IIFL) ของแบบทดสอบวัดความสามารถในการอ่านจับใจความและการแก้โจทย์ปัญหาทางคณิตศาสตร์ ที่ได้จากการให้คะแนนแบบปกติ และแบบ 2-SR โดยใช้ตัวแปรในการคัดเลือกข้อสอบตามกรอบทฤษฎี CTT และ IRT

แบบทดสอบ	รูปแบบการให้คะแนน	สมการพยากรณ์	R	R ²	E	SE	REM	F-TEST
อ่านจับใจความ	ปกติ	.20** + .15r* - .34D** - .19P** - .08BTA** - .01F**	.970	.950	78	.02	.0002	123**
	2 - SR	.05** - .04BTA** - .02D** - .001F**	.999	.997	94	.001	.0000	4732**
การแก้โจทย์ปัญหา	ปกติ	.33P** + .20r** - .24D** - .019	.970	.940	81	.015	.0001	204**
	2 - SR	0.47** - .045BTA** - .003D*	.993	.987	88	.001	.0000	1420**

** มีนัยสำคัญทางสถิติ ณ ระดับ .01

จากตาราง 32 พบว่าสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนแบบปกติประกอบด้วยตัวแปรได้แก่ สัมประสิทธิ์การจำแนก ดัชนีการจำแนก ความยากแบบ CTT และ IRT และความเหมาะสมของข้อสอบกับโมเดลซึ่งมีระดับความสัมพันธ์พหุคูณเท่ากับ .970 สัมประสิทธิ์การพยากรณ์เท่ากับ .950 และประสิทธิภาพในการพยากรณ์เท่ากับ 78 %

ส่วนสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนแบบ 2 - SR พบว่าประกอบด้วยตัวแปรได้แก่ ความยากแบบ IRT ดัชนีการจำแนก และความเหมาะสมของข้อสอบกับโมเดลซึ่ง

มีระดับความสัมพันธ์พหุคูณเท่ากับ .999 สัมประสิทธิ์การพยากรณ์เท่ากับ .997 และประสิทธิภาพในการพยากรณ์เท่ากับ 94 %

สำหรับสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนแบบปกติ พบว่าตัวแปรในสมการได้แก่ ความยากแบบ CTT สัมประสิทธิ์การจำแนก และดัชนีการจำแนก ซึ่งมีระดับความสัมพันธ์พหุคูณเท่ากับ .970 สัมประสิทธิ์พยากรณ์เท่ากับ .940 และประสิทธิภาพในการพยากรณ์เท่ากับ 81%

และเมื่อพิจารณาสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนแบบ 2 - SR พบว่าตัวแปรในสมการได้แก่ ความยากแบบ IRT และดัชนีการจำแนก ซึ่งมีระดับความสัมพันธ์พหุคูณเท่ากับ .993 สัมประสิทธิ์พยากรณ์เท่ากับ .987 และประสิทธิภาพในการพยากรณ์เท่ากับ 88 %

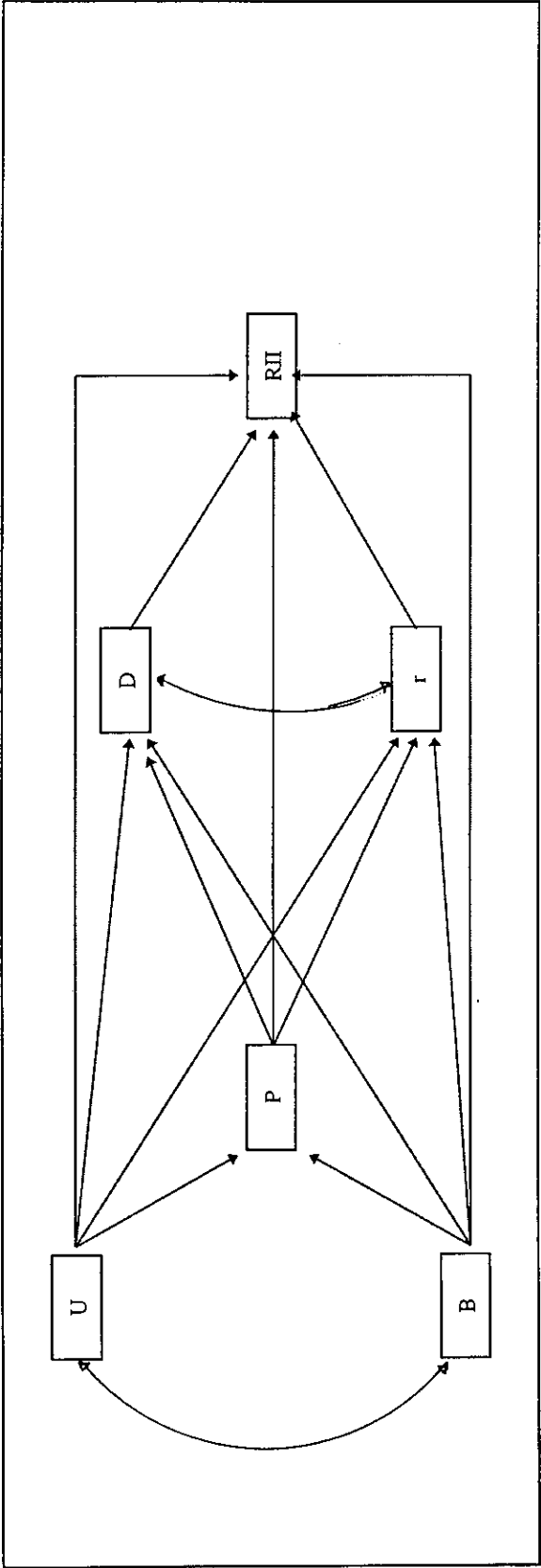
การศึกษารูปแบบความสัมพันธ์เชิงสาเหตุระหว่างรูปแบบการให้คะแนน วิธีการคัดเลือกข้อสอบและคุณภาพของแบบทดสอบ

การศึกษารูปแบบความสัมพันธ์เชิงสาเหตุในครั้งนี้ เพื่อต้องการทราบว่าตัวแปรต่างๆที่นำมาใช้เป็นเกณฑ์ในการคัดเลือกข้อสอบเข้าฉบับนั้น นอกจากตัวแปรที่ส่งผลโดยตรงแล้ว ตัวแปรอื่นๆที่เหลือมีการส่งผลในทางอ้อมต่อคุณภาพของแบบทดสอบอีกหรือไม่ และเมื่อให้คะแนนโดยรูปแบบที่ต่างกันแล้ว รูปแบบความสัมพันธ์เป็นเช่นเดียวกันหรือแตกต่างกันอย่างไร และเพื่อให้การศึกษามีผลวิเคราะห์เป็นไปอย่างมีลำดับ จึงขอเสนอผลตามหัวข้อดังนี้

1. แผนภาพเส้นทางสมมติฐานแสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่าง ๆ
2. การศึกษาความสัมพันธ์เชิงสาเหตุ และสัมประสิทธิ์เส้นทาง

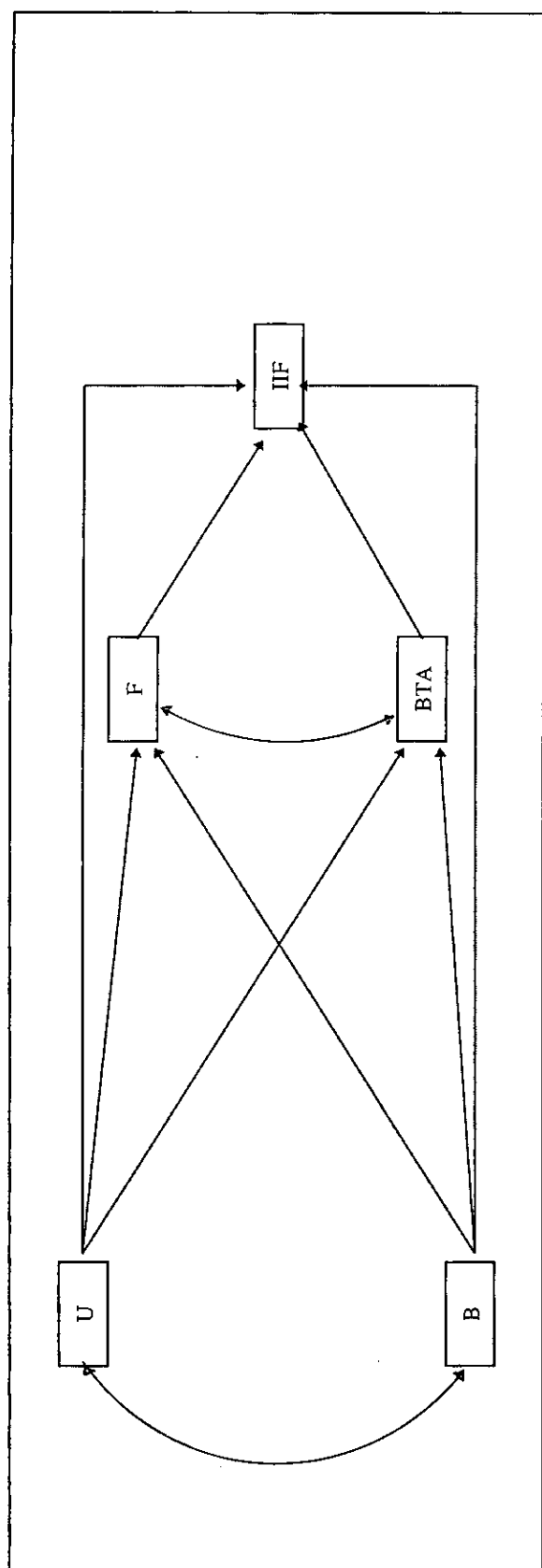
แผนภาพเส้นทางสมมติฐาน

จากการศึกษาทฤษฎี การตรวจสอบอิทธิพลและความสัมพันธ์ของตัวแปรอิสระเบื้องต้น และสมการถดถอยพหุคูณของตัวแปรต่างๆ ตามกรอบทฤษฎี CTT และ IRT ผู้วิจัยได้เขียนแผนภาพเส้นทางสมมติฐาน ดังนี้



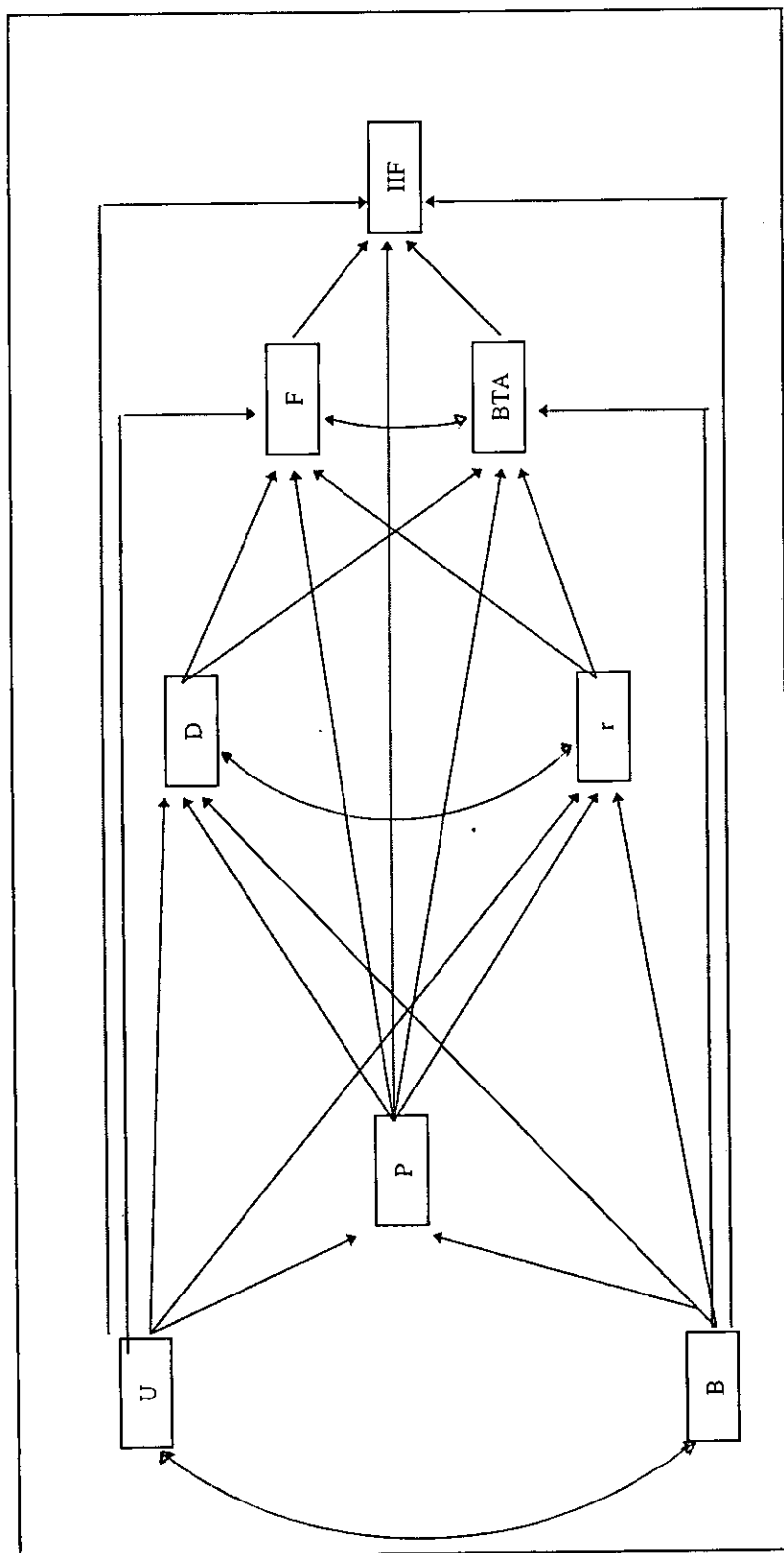
ภาพประกอบ 22 แผนภาพเส้นทางสมมติฐาน แสดงความสัมพันธ์เชิงสาเหตุ ของตัวแปรต่างๆ กับความเชื่อมั่นรายชั้ดตามกรอบทฤษฎี CTT

จากภาพประกอบ 22 ตัวแปรอิสระภายนอกคือ ความเป็นมิติเดียว และความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ ซึ่งมีความสัมพันธ์กันแต่ไม่สามารถบอกทิศทางเชิงเหตุผลได้ และคาดว่าส่งอิทธิพลทางตรง และทางอ้อมผ่านทางตัวแปรความยาก สัมประสิทธิ์การจำแนก และดัชนีการจำแนกต่อความเชื่อมั่นรายข้อ สำหรับตัวแปรอิสระภายในคือ ความยาก คาดว่าส่งอิทธิพลทางตรง และทางอ้อมผ่านสัมประสิทธิ์การจำแนก และดัชนีการจำแนกต่อความเชื่อมั่นรายข้อ และค่าสัมประสิทธิ์การจำแนก และดัชนีการจำแนก ซึ่งคาดว่าส่งอิทธิพลเฉพาะทางตรงต่อค่าความเชื่อมั่นรายข้อเท่านั้น และทั้งสองค่ามีความสัมพันธ์กันแต่ไม่สามารถบอกทิศทางเชิงเหตุผลได้



ภาพประกอบ 23 แผนภาพเส้นทางสมมติฐาน แสดงความสัมพันธ์เชิงสาเหตุ ของตัวแปรต่างๆ กับค่าสารสนเทศของข้อสอบตามกรอบทฤษฎี IRT

จากภาพประกอบ 23 ตัวแปรอิสระภายนอกคือ ความเป็นมิติเดียว และความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ ซึ่งมีความสัมพันธ์กันแต่ไม่สามารถบอกทิศทางเชิงเหตุผลได้ และคาดว่าส่งอิทธิพลทางตรง และทางอ้อมผ่านทางตัวแปรความยาก และความเหมาะสมของข้อสอบกับโมเดล ต่อค่าสารสนเทศของข้อสอบ สำหรับตัวแปรอิสระภายในคือ ความยาก และความเหมาะสมของข้อสอบกับโมเดล ซึ่งคาดว่าส่งอิทธิพลเฉพาะทางตรงเท่านั้น และทั้งสองค่ามีความสัมพันธ์กันแต่ไม่สามารถบอกทิศทางเชิงเหตุผลได้



ภาพประกอบ 24 แผนภาพเส้นทางสมมติฐาน แสดงความสัมพันธ์เชิงสาเหตุ ของตัวแปรต่างๆ กับค่าสารสนเทศของข้อสอบตามกรอบ

ทฤษฎี CTT และ IRT

จากภาพประกอบ 24 ตัวแปรอิสระภายนอกคือ ความเป็นมิติเดียว และความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ ซึ่งมีความสัมพันธ์กันแต่ไม่สามารถบอกทิศทางเชิงเหตุผลได้ และคาดว่าส่งอิทธิพลทางตรง และทางอ้อมผ่านทางตัวแปรความยากแบบ CTT และ IRT สัมประสิทธิ์การจำแนก ดัชนีการจำแนก และความเหมาะสมของข้อสอบกับโมเดลต่อค่าสารสนเทศของข้อสอบ สำหรับตัวแปรอิสระภายในคือ ความยากแบบ CTT คาดว่าส่งอิทธิพลทางตรง และทางอ้อมผ่านสัมประสิทธิ์การจำแนก ดัชนีการจำแนก ความยากแบบ IRT และความเหมาะสมของข้อสอบกับโมเดลต่อค่าสารสนเทศของข้อสอบ

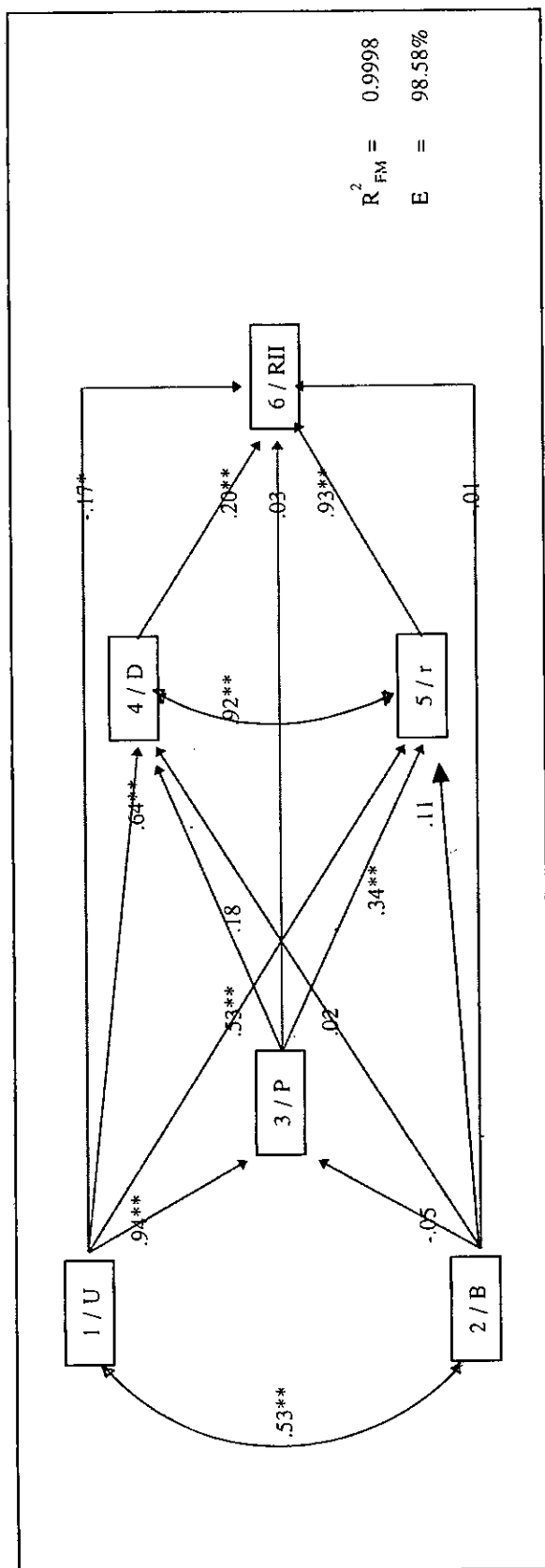
ค่าสัมประสิทธิ์การจำแนก และดัชนีการจำแนก คาดว่าส่งอิทธิพลทางตรง และทางอ้อมผ่านความยากแบบ IRT และความเหมาะสมของข้อสอบกับโมเดลต่อค่าสารสนเทศของข้อสอบ และทั้งสองค่ามีความสัมพันธ์กันแต่ไม่สามารถบอกทิศทางเชิงเหตุผลได้

และสำหรับความยากแบบ IRT และความเหมาะสมของข้อสอบกับโมเดล คาดว่าส่งอิทธิพลเฉพาะทางตรงต่อค่าสารสนเทศของข้อสอบเท่านั้น และทั้งสองค่ามีความสัมพันธ์กันแต่ไม่สามารถบอกทิศทางเชิงเหตุผลได้

การศึกษาความสัมพันธ์เชิงสาเหตุ และสัมประสิทธิ์เส้นทาง

จากการศึกษารูปแบบความสัมพันธ์เชิงสาเหตุของตัวแปรอิสระต่างๆที่ใช้เป็นเกณฑ์ในการคัดเลือกข้อสอบเข้าฉบับ ซึ่งให้คะแนนตามรูปแบบปกติและรูปแบบ 2-SR ตามกรอบทฤษฎี CTT และ IRT เพื่อให้เกิดความเข้าใจอย่างต่อเนื่องจึงขอลำดับขั้นตอนผลการศึกษาดังนี้

1. แผนภาพเส้นทางเต็มรูป
2. แผนภาพเส้นทางเชิงประจักษ์
3. การเปรียบเทียบความสอดคล้องของเส้นทางเชิงประจักษ์กับเส้นทางเต็มรูป
4. การศึกษาอิทธิพลของตัวแปรอิสระต่างๆที่ใช้เป็นเกณฑ์คัดเลือกข้อสอบที่มีต่อคุณภาพของแบบทดสอบเลือกตอบตามกรอบทฤษฎี CTT และ IRT

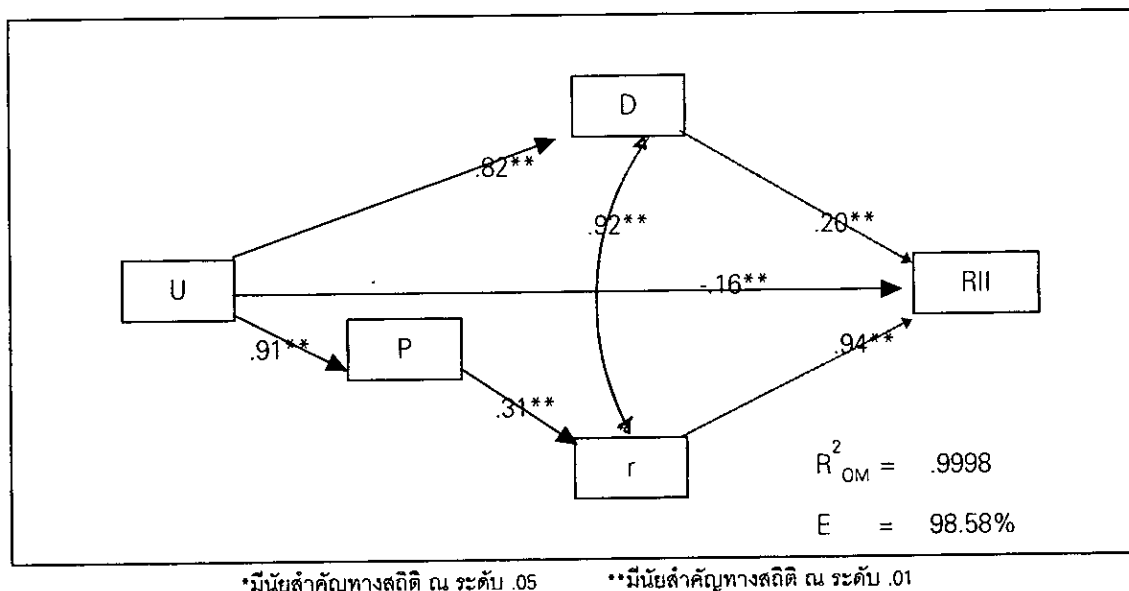


* มีนัยสำคัญทางสถิติ ณ ระดับ .05

** มีนัยสำคัญทางสถิติ ณ ระดับ .01

ภาพประกอบ 25 แผนภาพเส้นทางเต็มรูป แสดงความสัมพันธ์เชิงสาเหตุ ของตัวแปรต่างๆ กับความเชื่อมั่นรายข้อของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนแบบปกติ ตามกรอบทฤษฎี CTT

จากภาพประกอบที่ 25 พบว่าค่าสัมประสิทธิ์พหุคูณ และประสิทธิภาพในการพยากรณ์ของตัวแปรต่างๆในแผนภาพเส้นทางเท่ากับ 0.9998 และ 98.58 % ตามลำดับ เมื่อพิจารณาสัมประสิทธิ์เส้นทางที่มีนัยสำคัญทางสถิติ ณ ระดับ .01 พบว่ามี 9 ค่าคือ p_{21} , p_{31} , p_{41} , p_{51} , p_{61} , p_{53} , p_{54} , p_{64} และ p_{65} ซึ่งมีค่าเท่ากับ .53 , .94 , .64 , .53 , -.17 , .34 , .92 , .20 และ .93 ตามลำดับ



ภาพประกอบ 26 แผนภาพเส้นทางเชิงประจักษ์แสดงอิทธิพลทางตรงและทางอ้อมของตัวแปรต่างๆที่มีต่อความเชื่อมั่นของข้อสอบของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนตามรูปแบบปกติ ตามกรอบทฤษฎี CTT

จากภาพประกอบ 26 พบว่า สัมประสิทธิ์พหุคูณ และประสิทธิภาพในการพยากรณ์ของตัวแปรต่างๆในแผนภาพเส้นทางมีค่าเท่ากับ 0.9998 และ 98.58 % ตามลำดับ และพบว่าสัมประสิทธิ์การจำแนก และดัชนีการจำแนกส่งอิทธิพลทางตรงเชิงบวก ส่วนความเป็นมิติเดียวส่งอิทธิพลทางตรงเชิงลบต่อความเชื่อมั่นรายข้อของแบบทดสอบวัดความสามารถในการอ่านจับใจความซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ .94 , .20 และ -.16 ตามลำดับ และพบว่าความเป็นมิติเดียวส่งอิทธิพลทางอ้อมเชิงบวกผ่านความยาก ดัชนีการจำแนก และสัมประสิทธิ์การจำแนก ซึ่งมี

ค่าสัมประสิทธิ์เส้นทางเท่ากับ .91 , .82 และ .62 ตามลำดับ และพบว่าความยากส่งอิทธิพลทางอ้อมผ่านสัมประสิทธิ์การจำแนก มีค่าสัมประสิทธิ์เส้นทางเท่ากับ .31 นอกจากนั้นยังพบความสัมพันธ์ระหว่างดัชนีการจำแนก และสัมประสิทธิ์การจำแนก เท่ากับ .92 ซึ่งไม่สามารถบอกทิศทางเชิงสาเหตุได้

ตาราง 33 เปรียบเทียบความสอดคล้องแผนภาพเส้นทางเชิงประจักษ์กับแผนภาพเส้นทางเต็มรูปของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนตามรูปแบบปกติ ตามกรอบทฤษฎี CTT

	แผนภาพเต็มรูป	แผนภาพเชิงประจักษ์	W
R^2	0.9998	0.9998	0**
E	98.58 %	98.58 %	

*มีนัยสำคัญทางสถิติ ณ ระดับ .05

**มีนัยสำคัญทางสถิติ ณ ระดับ .01

จากตาราง 33 พบว่าแผนภาพเส้นทางเชิงประจักษ์สอดคล้องกับแผนภาพเส้นทางเต็มรูป

ตาราง 34 ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมที่มีต่อความเชื่อมั่นของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนตามรูปแบบปกติตามกรอบทฤษฎี CTT

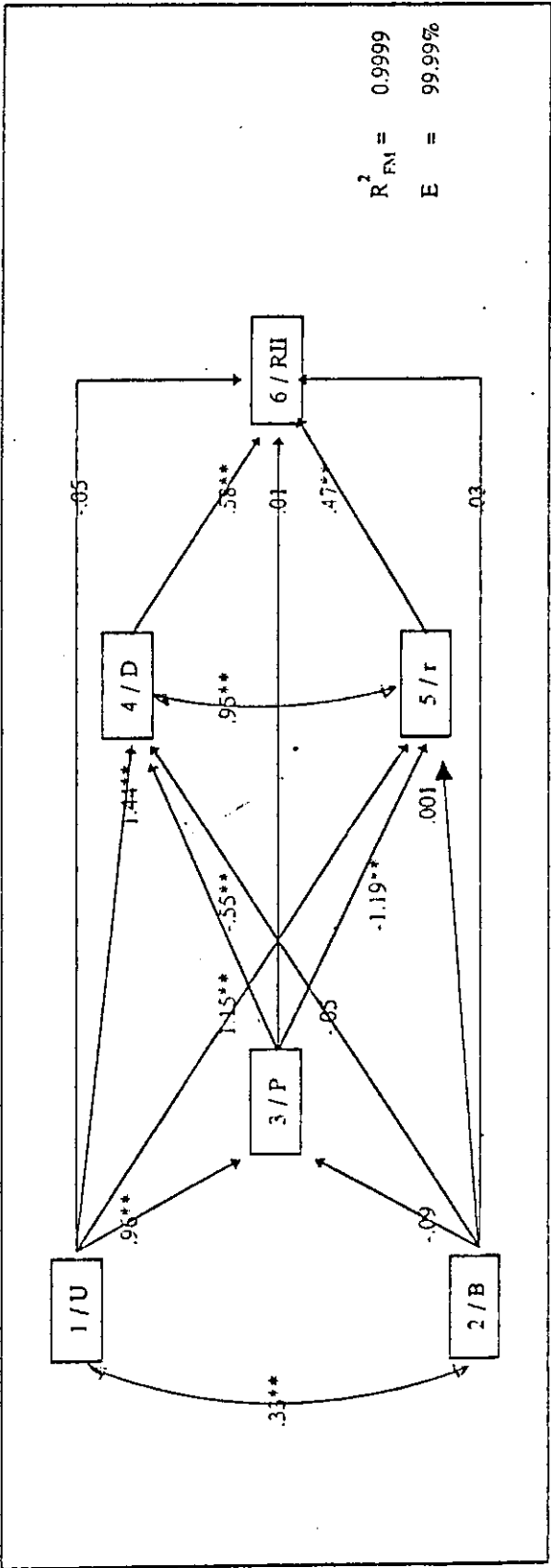
ตัวแปรสาเหตุ	อิทธิพลทางตรง	อิทธิพลทางอ้อม	อิทธิพลปลอม	รวม
r	0.94	.23	-	1.17
U	- 0.16	1.71	-	1.55
D	0.20	0.71	-	0.91
P	-	.34	-	.34

*มีนัยสำคัญทางสถิติ ณ ระดับ .05

**มีนัยสำคัญทางสถิติ ณ ระดับ .01

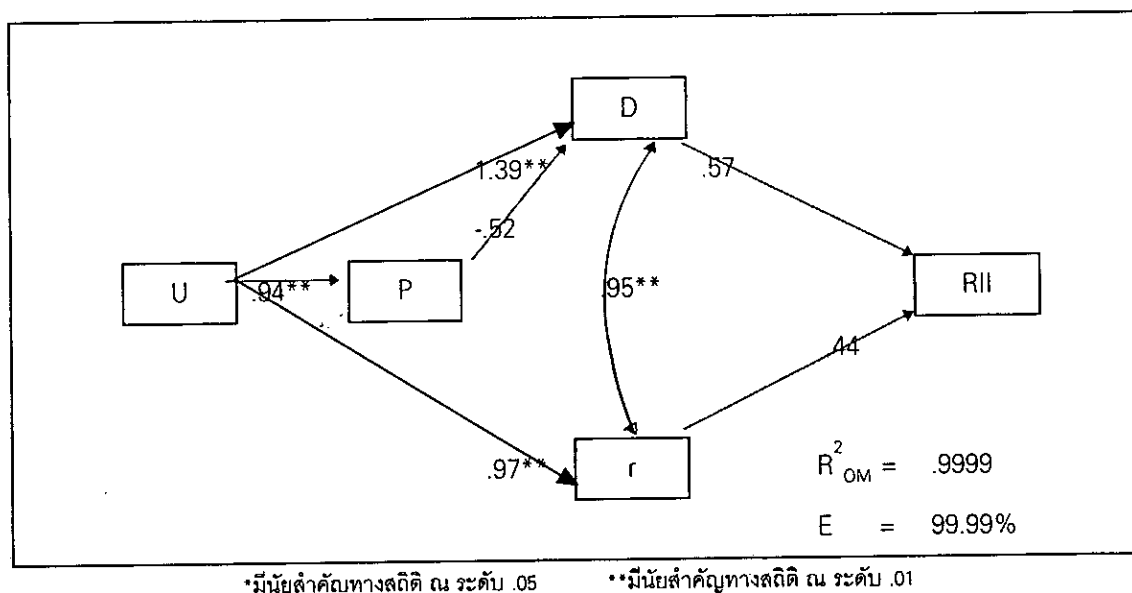
จากตาราง 34 พบว่าค่าความเป็นมิติเดียว มีอิทธิพลต่อความเชื่อมั่นของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนโดยวิธีแบบปกติ มากที่สุดคือเท่ากับ 1.55 รองลง

มาได้แก่ สัมประสิทธิ์การจำแนก ดัชนีการจำแนก และความยาก ซึ่งมีค่าอิทธิพลเท่ากับ 1.17 , .91 และ .34 ตามลำดับ



ภาพประกอบ 27 แผนภาพเส้นทางเริ่มรูป แสดงความสัมพันธ์เชิงสาเหตุ ของตัวแปรต่างๆ กับความเชื่อมั่นรายข้อของแบบทดสอบวัดความ สามารถในการอ่านจับใจความที่ให้คะแนนแบบ 2-SR ตามกรอบทฤษฎี CTT

จากภาพประกอบที่ 27 พบว่าค่าสัมประสิทธิ์พยากรณ์ และประสิทธิภาพในการพยากรณ์ของตัวแปรต่างๆในแผนภาพเส้นทางเท่ากับ 0.9999 และ 99.99 % ตามลำดับ เมื่อพิจารณาสัมประสิทธิ์เส้นทางที่มีนัยสำคัญทางสถิติ ณ ระดับ .01 พบว่ามี 9 ค่าคือ p_{21} , p_{31} , p_{41} , p_{51} , p_{43} , p_{53} , p_{54} , p_{64} และ p_{65} ซึ่งมีค่าเท่ากับ .33, .96, 1.44, 1.15, -.55, -1.19, .95, .58 และ .47 ตามลำดับ



ภาพประกอบ 28 แผนภาพเส้นทางเชิงประจักษ์ แสดงอิทธิพลทางตรงและทางอ้อมของตัวแปรต่างๆที่มีต่อความเชื่อมั่นของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนตามรูปแบบ 2-SR ตามกรอบทฤษฎี CTT

จากภาพประกอบ 28 พบว่า สัมประสิทธิ์พยากรณ์ และประสิทธิภาพในการพยากรณ์ของตัวแปรต่างๆในแผนภาพเส้นทางมีค่าเท่ากับ 0.9999 และ 99.99 % ตามลำดับ และพบว่า สัมประสิทธิ์การจำแนก และดัชนีการจำแนกส่งอิทธิพลทางตรงเชิงบวกต่อความเชื่อมั่นรายข้อของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนแบบ 2 - SR ซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ .57, และ .44 ตามลำดับ ส่วนความยากส่งอิทธิพลทางอ้อมเชิงลบผ่านทางดัชนีจำแนกซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ -.52 และสำหรับความเป็นมิติเดียวส่งอิทธิพลทางอ้อมเชิงบวกผ่านทางดัชนีการจำแนก สัมประสิทธิ์การจำแนก และความยากซึ่งมีค่าสัมประสิทธิ์เส้น

ทางเท่ากับ 1.39 , .97 และ .94 ตามลำดับ นอกจากนั้นยังพบว่า นอกจากนั้นยังพบความสัมพันธ์ระหว่างดัชนีการจำแนก และสัมประสิทธิ์การจำแนก เท่ากับ .95 ซึ่งไม่สามารถบอกทิศทางเชิงสาเหตุได้

ตาราง 35 เปรียบเทียบความสอดคล้องแผนภาพเส้นทางเชิงประจักษ์กับแผนภาพเส้นทางเต็มรูปของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนตามรูปแบบ 2 - SR ตามกรอบทฤษฎี CTT

	แผนภาพเต็มรูป	แผนภาพเชิงประจักษ์	W
R^2	0.9999	0.9999	0**
E	99.99 %	99.99 %	

*มีนัยสำคัญทางสถิติ ณ ระดับ .05 **มีนัยสำคัญทางสถิติ ณ ระดับ .01

จากตาราง 35 พบว่าแผนภาพเส้นทางเชิงประจักษ์สอดคล้องกับแผนภาพเส้นทางเต็มรูป

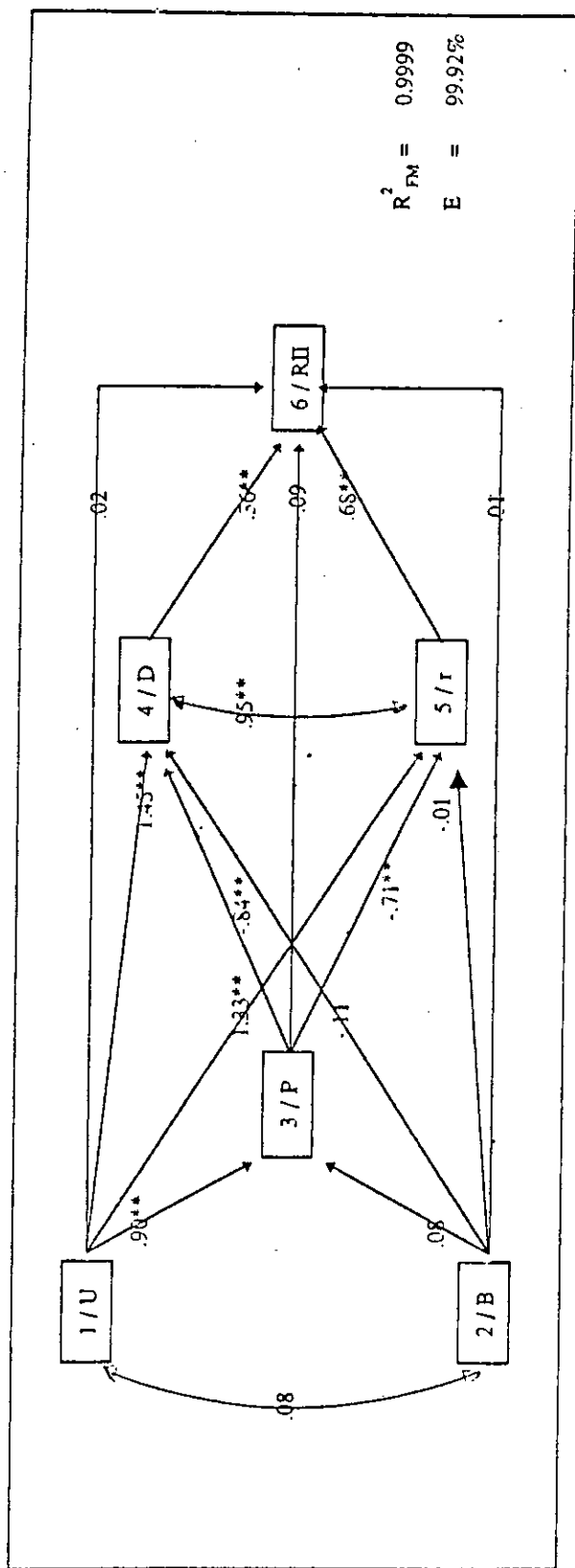
ตาราง 36 ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมที่มีต่อความเชื่อมั่นของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนโดยวิธีแบบ 2 - SR ตามกรอบทฤษฎี CTT

ตัวแปรสาเหตุ	อิทธิพลทางตรง	อิทธิพลทางอ้อม	อิทธิพลปลอม	รวม
r	.44	.94	-	1.38
U	-	2.22	-	2.22
D	.57	.73	-	1.30
P	-	-.68	-	-.68

*มีนัยสำคัญทางสถิติ ณ ระดับ .05 **มีนัยสำคัญทางสถิติ ณ ระดับ .01

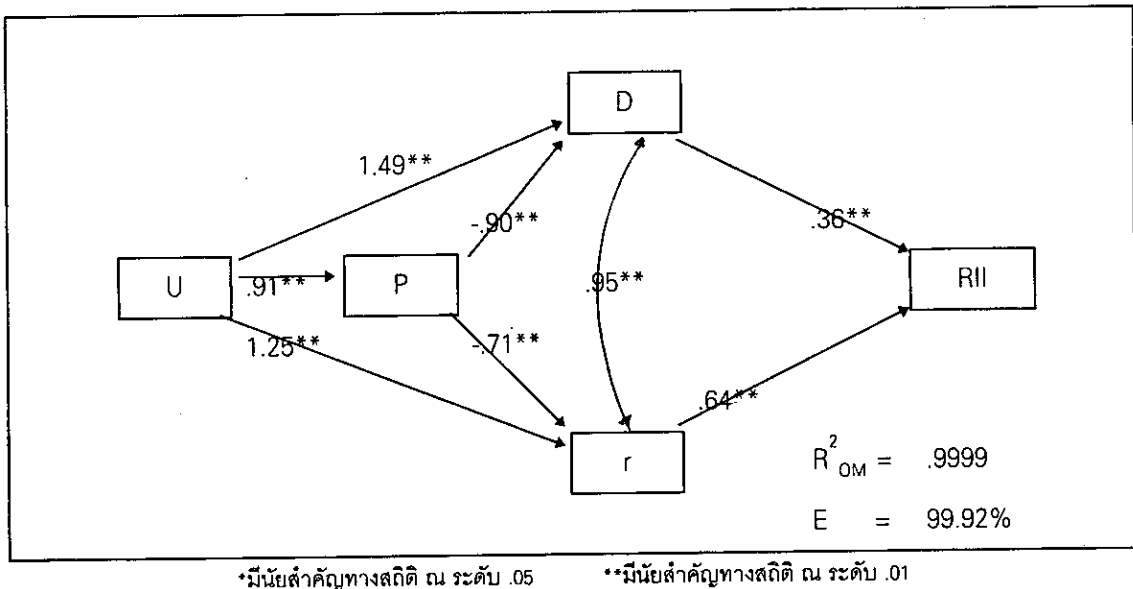
จากตาราง 36 พบว่าความเป็นมิติเดียวมีอิทธิพลต่อความเชื่อมั่นของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนโดยวิธีแบบ 2 - SR สูงสุดเท่ากับ 2.22 รองลงมา

ได้แก่ สัมประสิทธิ์การจำแนก และดัชนีการจำแนก ซึ่งมีค่าอิทธิพลเท่ากับ 1.38 และ 1.30 ตามลำดับ และพบว่าความยากส่งอิทธิพลเชิงลบมีค่าเท่ากับ -.68



ภาพประกอบ 29 แผนภาพเส้นทางเต็มรูป แสดงความสัมพันธ์เชิงสาเหตุ ของตัวแปรต่างๆ กับความเชื่อมั่นรายข้อของแบบทดสอบวัดความสามารถในการแก้ปัญหาทางคณิตศาสตร์ที่ให้คะแนนแบบปกติ ตามกรอบทฤษฎี CTT

จากภาพประกอบที่ 29 พบว่าค่าสัมประสิทธิ์พยากรณ์ และประสิทธิภาพในการพยากรณ์ของตัวแปรต่างๆในแผนภาพเส้นทางเท่ากับ 0.9999 และ 99.92 % ตามลำดับ เมื่อพิจารณาสัมประสิทธิ์เส้นทางที่มีนัยสำคัญทางสถิติ ณ ระดับ .01 พบว่ามี 8 ค่าคือ p_{31} , p_{41} , p_{51} , p_{43} , p_{53} , p_{54} , p_{64} และ p_{65} ซึ่งมีค่าเท่ากับ .90 , 1.45 , 1.33 , -.84 , -.71 , -.09 , .95 , .36 และ .68 ตามลำดับ



ภาพประกอบ 30 แผนภาพเส้นทางเชิงประจักษ์แสดงอิทธิพลทางตรงและทางอ้อมของตัวแปรต่างๆที่มีต่อความเชื่อมั่นของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ได้จากการให้คะแนนตามรูปแบบปกติ ตามกรอบทฤษฎี CTT

จากภาพประกอบ 30 พบว่า สัมประสิทธิ์พยากรณ์ และประสิทธิภาพในการพยากรณ์ของตัวแปรต่างๆในแผนภาพเส้นทางมีค่าเท่ากับ 0.9999 และ 99.99 % ตามลำดับ และพบว่าสัมประสิทธิ์การจำแนก และดัชนีการจำแนกส่งอิทธิพลทางตรงเชิงบวกต่อความเชื่อมั่นรายข้อของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนแบบปกติ ซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ .64 , และ .36 ตามลำดับ สำหรับความยากพบว่าส่งอิทธิพลทางอ้อมเชิงลบผ่านทางดัชนีจำแนก สัมประสิทธิ์การจำแนกซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ -.90 และ -.71ตามลำดับ และพบว่าความเป็นมิติเดียวส่งอิทธิพลทางอ้อมเชิงบวกผ่านทางดัชนีการจำแนก สัมประสิทธิ์การจำแนก และความยากซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ 1.49 , 1.33 และ .91

ตามลำดับ นอกจากนั้นยังพบความสัมพันธ์ระหว่างดัชนีการจำแนก และสัมประสิทธิ์การจำแนก เท่ากับ .95 ซึ่งไม่สามารถบอกทิศทางเชิงสาเหตุได้

ตาราง 37 เปรียบเทียบความสอดคล้องแผนภาพเส้นทางเชิงประจักษ์กับแผนภาพเส้นทางเต็มรูป ของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนตามรูปแบบปกติ ตามกรอบทฤษฎี CTT

	แผนภาพเต็มรูป	แผนภาพเชิงประจักษ์	W
R^2	0.9999	0.9999	0**
E	99.92 %	99.92 %	

*มีนัยสำคัญทางสถิติ ณ ระดับ .05

**มีนัยสำคัญทางสถิติ ณ ระดับ .01

จากตาราง 37 พบว่าแผนภาพเส้นทางเชิงประจักษ์สอดคล้องกับแผนภาพเส้นทางเต็มรูป

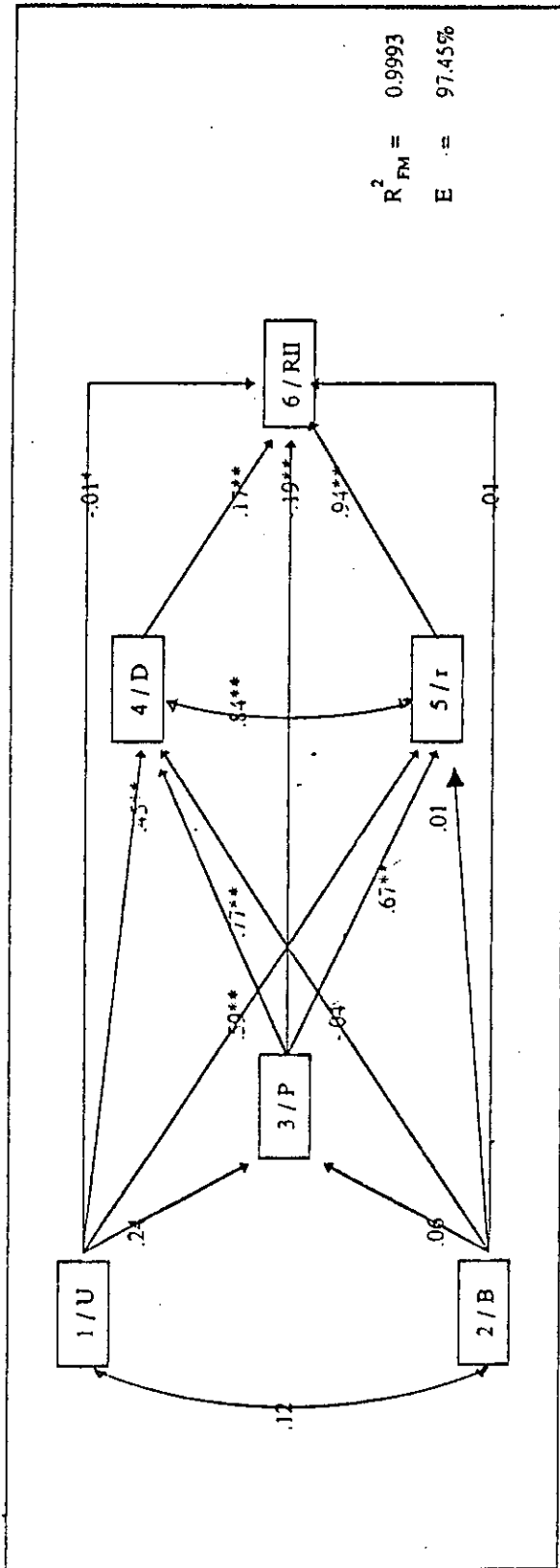
ตาราง 38 ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมที่มีต่อความเชื่อมั่นของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนโดยวิธีแบบแบบปกติ ตามกรอบทฤษฎี CTT

ตัวแปรสาเหตุ	อิทธิพลทางตรง	อิทธิพลทางอ้อม	อิทธิพลปลอม	รวม
r	.64	-.11	-	.53
U	-	.48	-	.48
D	.36	-.20	-	.16
P	-	-.52	-	-.52

*มีนัยสำคัญทางสถิติ ณ ระดับ .05

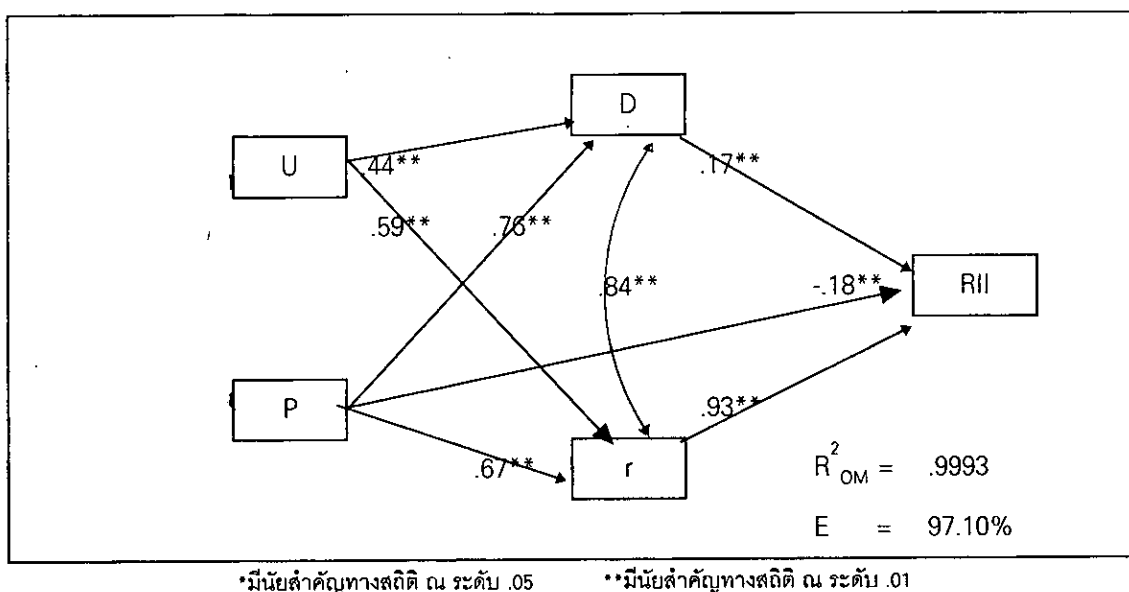
**มีนัยสำคัญทางสถิติ ณ ระดับ .01

จากตาราง 38 พบว่าสัมประสิทธิ์การจำแนก ส่งอิทธิพลเชิงบวกต่อความเชื่อมั่นรายข้อของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์มากที่สุด รองลงมาได้แก่ ความเป็นมิติเดียว และดัชนีการจำแนก ซึ่งมีค่าเท่ากับ .53 , .48 และ .16 ตามลำดับ และพบว่าค่าความยากมีอิทธิพลเชิงลบมีค่าเท่ากับ -.52



ภาพประกอบ 31 แผนภาพเส้นทางเต็มรูป แสดงความสัมพันธ์เชิงสาเหตุ ของตัวแปรต่างๆ กับความเชื่อมั่นรายข้อของแบบทดสอบวัดความสามารถในการแก้ปัญหาทางคณิตศาสตร์ที่ให้คะแนนแบบ 2-SR ตามกรอบทฤษฎี CTT

จากภาพประกอบที่ 31 พบว่าค่าสัมประสิทธิ์พยากรณ์ และประสิทธิภาพในการพยากรณ์ของตัวแปรต่างๆในแผนภาพเส้นทางเท่ากับ 0.9993 และ 97.45 % ตามลำดับ เมื่อพิจารณาสัมประสิทธิ์เส้นทางที่มีนัยสำคัญทางสถิติ ณ ระดับ .01 พบว่ามี 8 ค่าคือ p_{41} , p_{51} , p_{43} , p_{53} , p_{63} , p_{54} , p_{64} และ p_{65} ซึ่งมีค่าเท่ากับ .43 , .59 , .77 , .67 , -.19 , .84, .17 และ .94 ตามลำดับ



ภาพประกอบ 32 แผนภาพเส้นทางแสดงอิทธิพลทางตรงและทางอ้อมของตัวแปรต่างๆที่มีต่อความเชื่อมั่นของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ได้จากการให้คะแนนโดยวิธีแบบ 2 - SR ตามกรอบทฤษฎี CTT

จากภาพประกอบ 32 พบว่า สัมประสิทธิ์พยากรณ์ และประสิทธิภาพในการพยากรณ์ของตัวแปรต่างๆในแผนภาพเส้นทางมีค่าเท่ากับ 0.9993 และ 97.10 % ตามลำดับ และพบว่า สัมประสิทธิ์การจำแนก และดัชนีการจำแนกส่งอิทธิพลทางตรงเชิงบวกต่อความเชื่อมั่นรายข้อของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนแบบ 2 - SR ซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ .93 และ .17 ตามลำดับ ส่วนความยากพบว่าส่งอิทธิพลทางตรงเชิงลบซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ -.18 และพบว่าความเป็นมิติเดียวส่งอิทธิพลทางอ้อมเชิงบวกผ่านทางสัมประสิทธิ์การจำแนก และดัชนีการจำแนก ซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ .59 และ .44 ตามลำดับ ส่วนความยากพบว่าส่งอิทธิพลทางอ้อมเชิงบวกผ่านทาง ดัชนีการจำแนก และ

สัมประสิทธิ์การจำแนก ซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ .76 และ .67 ตามลำดับ นอกจากนั้นยังพบความสัมพันธ์ระหว่างดัชนีการจำแนก และสัมประสิทธิ์การจำแนก เท่ากับ .84 ซึ่งไม่สามารถบอกทิศทางเชิงสาเหตุได้

ตาราง 39 เปรียบเทียบความสอดคล้องแผนภาพเส้นทางเชิงประจักษ์กับแผนภาพเส้นทางเต็มรูปของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนตามรูปแบบ 2 - SR ตามกรอบทฤษฎี CTT

	แผนภาพเต็มรูป	แผนภาพเชิงประจักษ์	W
R^2	0.9993	0.9993	0**
E	97.45 %	97.45 %	

*มีนัยสำคัญทางสถิติ ณ ระดับ .05 **มีนัยสำคัญทางสถิติ ณ ระดับ .01

จากตาราง 39 พบว่าแผนภาพเส้นทางเชิงประจักษ์สอดคล้องกับแผนภาพเส้นทางเต็มรูป

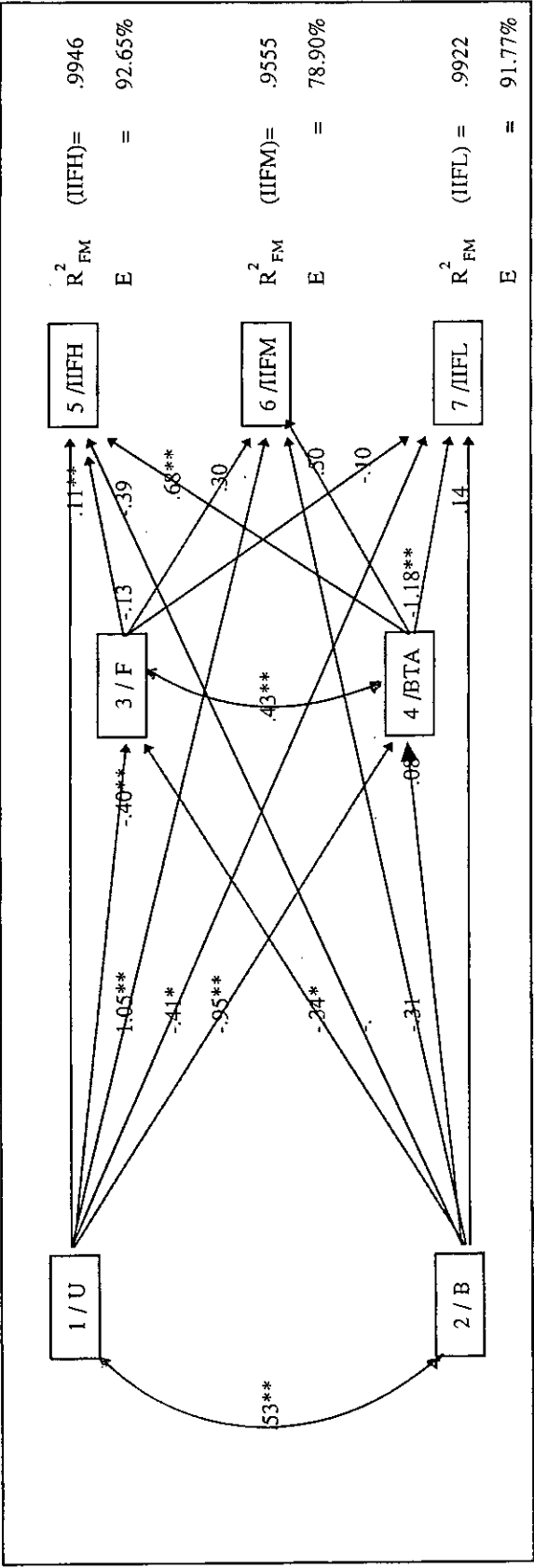
ตาราง 40 ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมที่มีต่อความเชื่อมั่นของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนโดยวิธีแบบ 2 - SR ตามกรอบทฤษฎี CTT

ตัวแปรสาเหตุ	อิทธิพลทางตรง	อิทธิพลทางอ้อม	อิทธิพลปลอม	รวม
r	.93	.12	-	1.05
U	-	.99	-	.99
D	.17	.65	-	.82
P	-.18	1.32	-	1.14

*มีนัยสำคัญทางสถิติ ณ ระดับ .05 **มีนัยสำคัญทางสถิติ ณ ระดับ .01

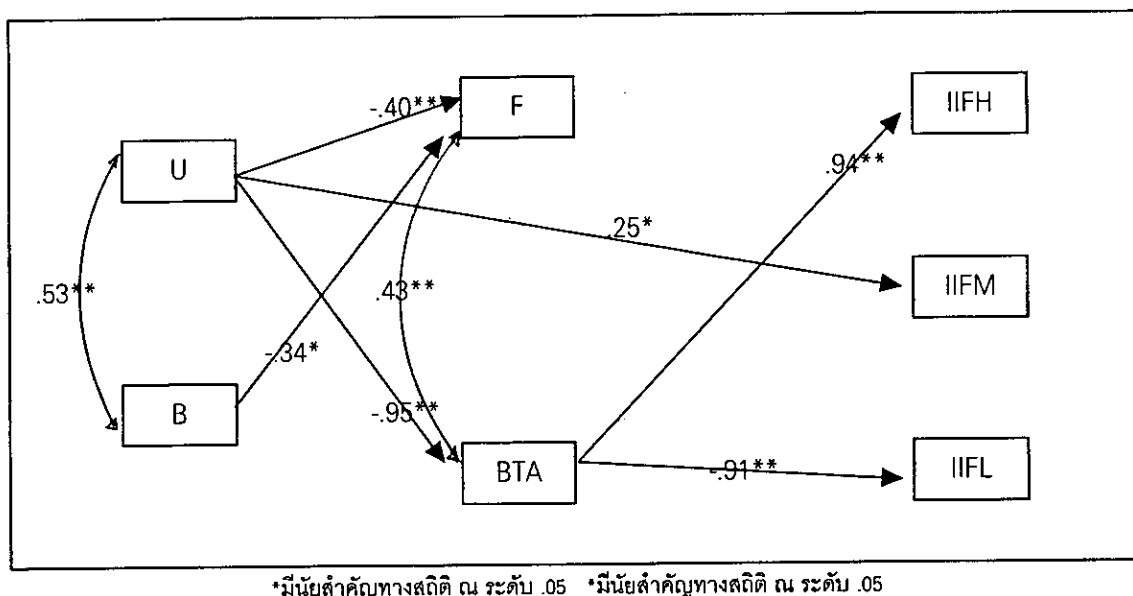
จากตาราง 40 พบว่าความยากส่งอิทธิพลเชิงบวกต่อความเชื่อมั่นรายข้อของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนโดยวิธีแบบ 2 - SR สูง

ชุดซึ่งมีค่าเท่ากับ 1.14 รองลงมาได้แก่สัมประสิทธิ์การจำแนก ความเป็นมิติเดียว และดัชนีการจำแนก ซึ่งมีค่าเท่ากับ 1.05 , .99 และ .82 ตามลำดับ



ภาพประกอบ 33 แผนภาพเส้นทางเต็มรูปแบบแสดงความสัมพันธ์เชิงสาเหตุ ของตัวแปรต่างๆ กับค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูง ปานกลาง และต่ำ ของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้ความแม่นยำในการตอบทฤษฎี IRT

จากภาพประกอบที่ 33 พบว่าค่าสัมประสิทธิ์พหุคูณ ของตัวแปรต่างๆในแผนภาพเส้นทาง แสดงความสัมพันธ์เชิงสาเหตุกับค่าสารสนเทศของข้อสอบที่ระดับความสามารถ สูง ปานกลางและต่ำ เท่ากับ 0.9946 , .9555 และ .9922 ตามลำดับโดยมีประสิทธิภาพในการพยากรณ์เท่ากับ 92.65 % , 78.90 % และ 91.77 % ตามลำดับ เมื่อพิจารณาสัมประสิทธิ์เส้นทางที่มีนัยสำคัญทางสถิติ ณ ระดับ .01 พบว่ามี 8 ค่าคือ p_{21} , p_{31} , p_{41} , p_{51} , p_{61} , p_{43} , p_{54} และ p_{74} ซึ่งมีค่าเท่ากับ .53 , -.40 , -.95 , .11 , 1.05 , .43 , .68 และ -1.18 ตามลำดับ และสัมประสิทธิ์เส้นทางที่มีนัยสำคัญทางสถิติ ณ ระดับ .05 พบว่ามี 2 ค่าคือ p_{32} , และ p_{71} ซึ่งมีค่าเท่ากับ -.34 และ -.41 ตามลำดับ



ภาพประกอบ 34 แผนภาพเส้นทางเชิงประจักษ์แสดงอิทธิพลทางตรงและทางอ้อมของตัวแปรต่างๆ ที่มีต่อสารสนเทศที่ระดับความสามารถสูง ปานกลาง และต่ำของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนตามรูปแบบปกติ ตามกรอบทฤษฎี IRT

จากภาพประกอบ 34 พบว่าความยากส่งอิทธิพลทางตรงเชิงบวก และเชิงลบต่อสารสนเทศของข้อสอบที่ระดับความสามารถสูง และต่ำ ของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนแบบปกติ ซึ่งมีค่าเท่ากับ .94 และ -.91 ตามลำดับ และพบว่าความเป็นมิติเดียวส่งอิทธิพลทางตรงต่อสารสนเทศของข้อสอบที่ระดับความสามารถปานกลาง ซึ่งมีสัมประสิทธิ์เส้นทางเท่ากับ .25 และส่งอิทธิพลทางอ้อมเชิงลบต่อสารสนเทศของข้อสอบที่ระดับความสามารถสูง และต่ำผ่านทางค่าความยากซึ่งมีค่าเท่ากับ -.95 นอกจากนี้ยังพบว่าความลำเอียงของข้อสอบ

เนื่องจากตัวแปรเพศส่งอิทธิพลทางอ้อมเชิงลบต่อสารสนเทศของข้อสอบที่ระดับความสามารถสูง และต่ำผ่านทางความเหมาะสมของข้อสอบกับโมเดลที่แสดงความสัมพันธ์กับความยากโดยไม่สามารถบอกทิศทางเชิงสาเหตุได้ และแสดงความสัมพันธ์กับความเป็นมิติเดียวโดยไม่สามารถบอกทิศทางเชิงสาเหตุได้เช่นเดียวกัน

ตาราง 41 เปรียบเทียบความสอดคล้องของแผนภาพเส้นทางเชิงประจักษ์กับแผนภาพเส้นทางเต็มรูปของแบบทดสอบวัดความสามารถในการอ่านจับใจความ ที่ให้คะแนนตามรูปแบบปกติ ตามทฤษฎี IRT

	แผนภาพเต็มรูป		แผนภาพเชิงประจักษ์		W
	R^2_{FM}	E	R^2_{OM}	E	
IIFH	.9946	92.65	.9946	92.65	0**
IIFM	.9555	78.90	.9555	78.90	0**
IIFL	.9922	91.77	.9922	91.77	0**

*มีนัยสำคัญทางสถิติ ณ ระดับ .05 *มีนัยสำคัญทางสถิติ ณ ระดับ .05

จากตาราง 41 พบว่า แผนภาพเส้นทางเชิงประจักษ์มีความสอดคล้องกับแผนภาพเต็มรูป ทุกระดับความสามารถ อย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01

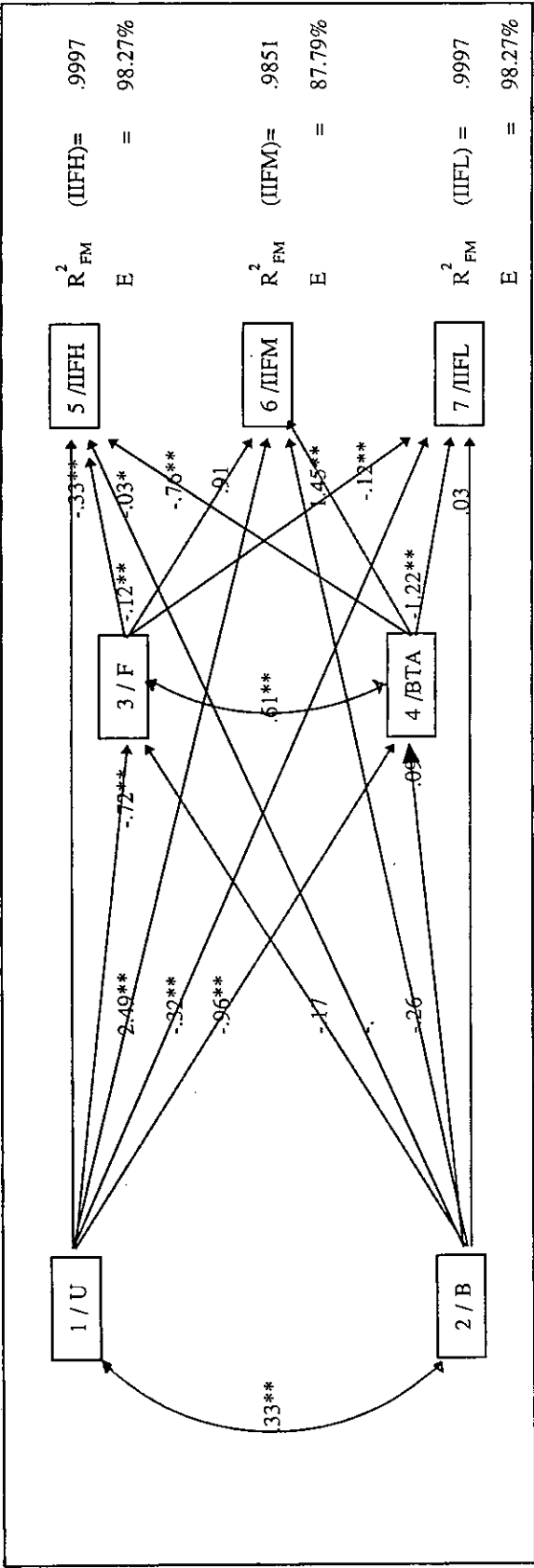
ตาราง 42 ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมต่อค่าสารสนเทศ ที่ระดับความสามารถสูง ปานกลาง และต่ำของแบบทดสอบวัดความสามารถในการอ่านจับใจ ใจความที่ได้จากการให้คะแนนตามรูปแบบปกติ ตามกรอบทฤษฎี IRT

ระดับความสามารถ/ อิทธิพล		ตัวแปรสาเหตุ			
		BTa	F	U	B
IIFH	ทางตรง	.94	-	-	
	ทางอ้อม	-	.53	-1.19	-.29
	ปลอม	-	-	-	-
	รวม	.94	.53	-1.19	-.29
IIFM	ทางตรง	-	-	.25	-
	ทางอ้อม	-	-	-	.13
	ปลอม	-	-	-	-
	รวม	-	-	.25	.13
IIFL	ทางตรง	-.91	-		-
	ทางอ้อม	-	-.51	1.15	.78
	ปลอม	-	-	-	-
	รวม	-.91	-.51	1.15	.78

จากตาราง 42 พบว่าค่าความยาก และความเหมาะสมของข้อสอบกับโมเดล ส่งอิทธิพลเชิงบวกต่อค่าสารสนเทศที่ระดับความสามารถสูงของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนแบบปกติ ซึ่งมีค่าเท่ากับ .94 และ .53 ตามลำดับ ส่วนความเป็นมิติเดียว และความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ ส่งอิทธิพลเชิงลบมีค่าเท่ากับ -1.19 และ -.29 ตามลำดับ

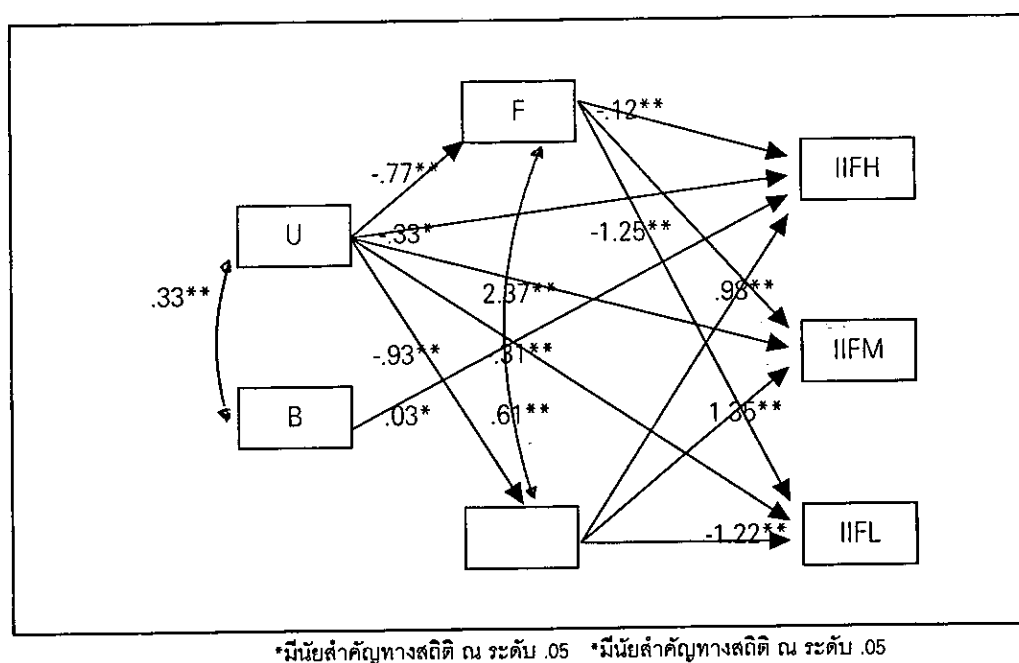
ส่วนตัวแปรที่ส่งอิทธิพลเชิงบวกต่อค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลางได้แก่ความเป็นมิติเดียว และความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ ซึ่งมีค่าเท่ากับ .25 และ .13 ตามลำดับ

และสำหรับตัวแปรที่ส่งอิทธิพลเชิงบวกต่อค่าสารสนเทศของข้อสอบที่ระดับความ
สามารถต่ำ ได้แก่ความเป็นมิติเดียว และความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ ซึ่งมีค่า
เท่ากับ 1.15 และ .78 ตามลำดับ และพบว่าความยาก และความเหมาะสมของข้อสอบกับโมเดล
ส่งอิทธิพลเชิงลบมีค่าเท่ากับ -.91 และ -.51 ตามลำดับ



ภาพประกอบ 35 แผนภาพเส้นทางเต็มรูปแบบแสดงความสัมพันธ์เชิงสาเหตุ ของตัวแปรต่างๆ กับค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูง ปานกลาง และต่ำ ของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนตามรูปแบบ 2-SR ตามกรอบทฤษฎี IRT

จากภาพประกอบที่ 35 พบว่าค่าสัมประสิทธิ์พหุคูณ ของตัวแปรต่างๆในแผนภาพเส้นทาง แสดงความสัมพันธ์เชิงสาเหตุกับค่าสารสนเทศของข้อสอบ ณ ระดับความสามารถ สูง ปานกลางและต่ำ เท่ากับ 0.9997 , .9851 และ .9997 ตามลำดับโดยมีประสิทธิภาพในการพยากรณ์เท่ากับ 98.27 % , 87.79 % และ 98.27 % ตามลำดับ เมื่อพิจารณาสัมประสิทธิ์เส้นทางที่มีนัยสำคัญทางสถิติ ณ ระดับ .01 พบว่ามี 13 ค่าคือ p_{21} , p_{31} , p_{41} , p_{51} , p_{61} , p_{71} , p_{43} , p_{53} , p_{63} , p_{73} , p_{54} , p_{64} และ p_{74} ซึ่งมีค่าเท่ากับ .33 , -.72 , -.96 , -.33 , 2.49 , -.42 , .61 , -.12 , .91 , -.12 , .76 , 1.45 และ -1.22 ตามลำดับ และสัมประสิทธิ์เส้นทางที่มีนัยสำคัญทางสถิติ ณ ระดับ .05 คือ p_{52} เท่ากับ -.03



ภาพประกอบ 36 แผนภาพเส้นทางเชิงประจักษ์แสดงอิทธิพลทางตรงและทางอ้อมของตัวแปรต่างๆที่มีต่อสารสนเทศของข้อสอบที่ระดับความสามารถสูง ปานกลาง และต่ำของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนตามรูปแบบ 2 - SR ตามกรอบทฤษฎี IRT

จากภาพประกอบ 36 พบว่า ความยากและความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศส่งอิทธิพลทางตรงเชิงบวกต่อค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูงของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนแบบ 2-SR ซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ .76 และ .03 ตามลำดับ ส่วนความเป็นมิติเดียวและความเหมาะสมของข้อสอบกับโมเดล

พบว่าอิทธิพลทางตรงเชิงลบซึ่งมีค่าเท่ากับ -.33 และ -.12 ตามลำดับ

สำหรับตัวแปรที่ส่งอิทธิพลทางตรงเชิงบวกต่อสารสนเทศของข้อสอบที่ระดับความสามารถปานกลางคือ ความเป็นมิติเดียว ความยาก และความเหมาะสมของข้อสอบกับโมเดล ซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ 2.37 ,1.35 และ .98 ตามลำดับ

และเมื่อพิจารณาตัวแปรที่ส่งอิทธิพลต่อสารสนเทศของข้อสอบที่ระดับความสามารถต่ำ พบว่าความเหมาะสมของข้อสอบกับโมเดล ความยาก และความเป็นมิติเดียว ส่งอิทธิพลทางตรงเชิงลบ ซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ -1.25 , -1.20 , และ -.31 ตามลำดับ

และพบว่าความเป็นมิติเดียวส่งอิทธิพลทางอ้อมต่อค่าสารสนเทศต่างๆ ผ่านทางความยาก และความเหมาะสมของข้อสอบกับโมเดล ซึ่งมีสัมประสิทธิ์เส้นทางเท่ากับ -.93 และ -.77 ตามลำดับ และมีความสัมพันธ์กับความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศที่ไม่สามารถบอกทิศทางเชิงสาเหตุได้ ซึ่งมีความสัมพันธ์เท่ากับ .33 และพบว่าความเหมาะสมของข้อสอบกับโมเดลมีความสัมพันธ์กับความยากแต่ไม่สามารถบอกทิศทางเชิงสาเหตุได้ ซึ่งมีค่าเท่ากับ .81

ตาราง 43 เปรียบเทียบความสอดคล้องของแผนภาพเส้นทางเชิงประจักษ์กับแผนภาพเส้นทางเต็มรูปของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนตามรูปแบบ 2 - SR ตามกรอบทฤษฎี IRT

	แผนภาพเต็มรูป		แผนภาพเชิงประจักษ์		W
	R^2_{FM}	E	R^2_{OM}	E	
IIFH	.9997	98.27	.9997	98.27	0**
IIFM	.9851	87.79	.9851	87.79	0**
IIFL	.9997	98.27	.9997	98.27	0**

*มีนัยสำคัญทางสถิติ ณ ระดับ .05 *มีนัยสำคัญทางสถิติ ณ ระดับ .05

จากตาราง 43 พบว่า แผนภาพเส้นทางเชิงประจักษ์มีความสอดคล้องกับแผนภาพเต็มรูป ทุกระดับความสามารถ อย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01

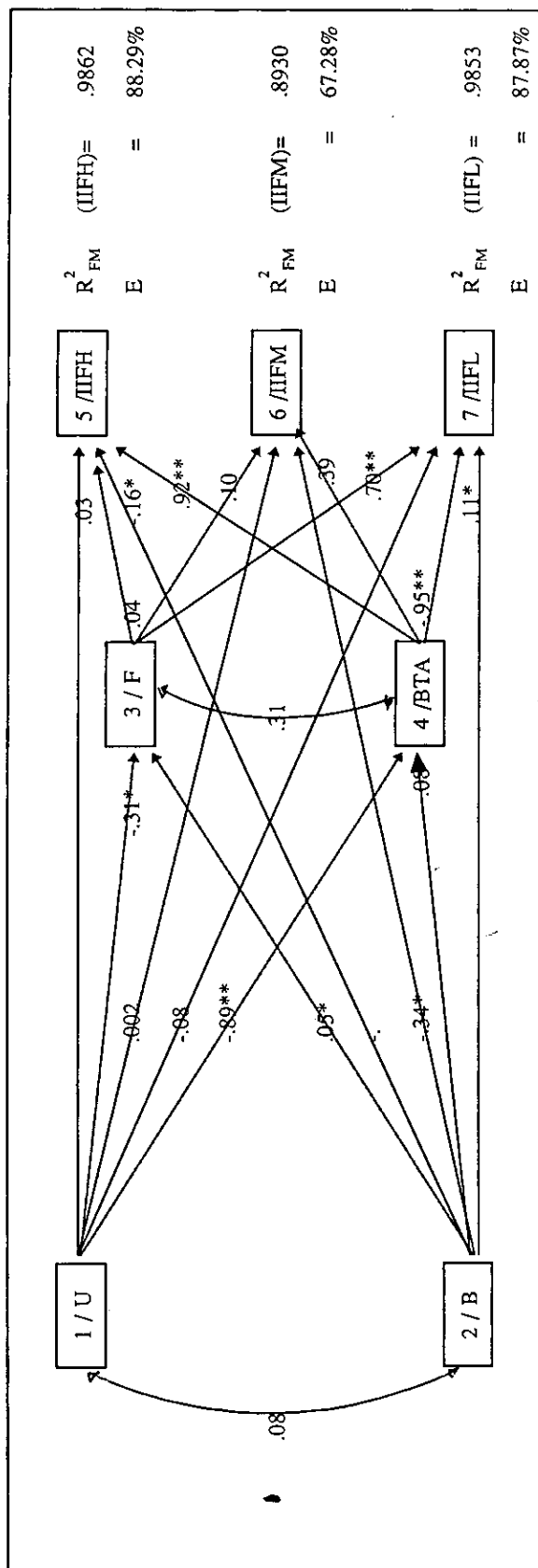
ตาราง 44 ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมต่อค่าสารสนเทศ ที่ระดับความสามารถสูง ปานกลาง และต่ำของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนโดยวิธีแบบ 2-SR ตามกรอบทฤษฎี IRT

ระดับความสามารถ/ อิทธิพล		ตัวแปรสาเหตุ			
		BTA	F	U	B
IIFH	ทางตรง	.76	-.12	-.33	.03
	ทางอ้อม	.01	-.08	-.55	-.09
	ปลอม	-	-	-	-
	รวม	.77	-.20	-.88	-.06
IIFM	ทางตรง	1.35	.98	2.37	-
	ทางอ้อม	-.11	-.15	-1.79	.08
	ปลอม	-	-	-	-
	รวม	1.24	.83	.58	.08
IIFL	ทางตรง	-1.20	-1.25	-.31	-
	ทางอ้อม	.14	.13	1.84	.51
	ปลอม	-	-	-	-
	รวม	-1.06	-1.12	1.53	.51

จากตาราง 44 พบว่า ตัวแปรที่มีอิทธิพลเชิงบวกต่อสารสนเทศของข้อสอบที่ระดับความสามารถสูงของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนแบบ 2-SR มากที่สุดคือ ความยากซึ่งมีค่าเท่ากับ .77 ส่วนอิทธิพลเชิงลบสูงสุดได้แก่ ความเป็นมิติเดียว มีค่าเท่ากับ -.88 รองลงมาได้แก่ความเหมาะสมของข้อสอบกับโมเดล และความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ ซึ่งมีค่าเท่ากับ -.20 และ -.06 ตามลำดับ

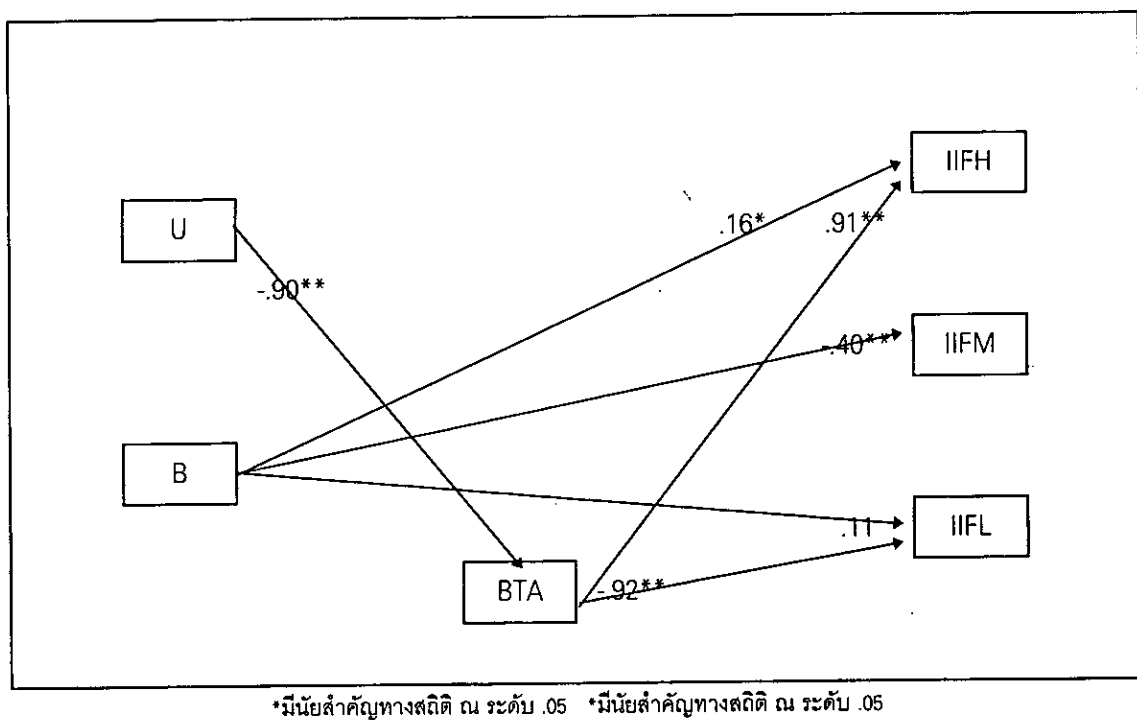
สำหรับตัวแปรที่มีอิทธิพลเชิงบวกต่อสารสนเทศของข้อสอบที่ระดับความสามารถปานกลางมากที่สุด ได้แก่ ความยากมีค่าเท่ากับ 1.24 รองลงมาได้แก่ความเหมาะสมของโมเดล ความเป็นมิติเดียว และความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ ซึ่งมีค่าเท่ากับ .83 , .58 และ .08 ตามลำดับ

และพบว่า ความเป็นมิติเดียวมีอิทธิพลเชิงบวกต่อสารสนเทศของข้อสอบที่ระดับความสามารถต่ำมีค่าเท่ากับ 1.53 รองลงมาได้แก่ความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ ซึ่งมีค่าเท่ากับ .51 ส่วนความเหมาะสมกับโมเดลพบว่ามีอิทธิพลเชิงลบสูงสุดเท่ากับ -1.12 รองลงมาได้แก่ความยากซึ่งมีค่าเท่ากับ -1.06



ภาพประกอบ 37 แผนภาพเส้นทางเต็มรูปแบบแสดงความสัมพันธ์เชิงสาเหตุ ของตัวแปรต่างๆ กับค่าสารสนเทศของข้อสอบที่ระดับความสามารถ สูง ปานกลาง และต่ำ ของแบบทดสอบวัดความสามารถในการแก้ปัญหาทางคณิตศาสตร์ที่ให้คะแนนตามรูปแบบปกติ ตามกรอบทฤษฎี IRT

จากภาพประกอบที่ 37 พบว่าค่าสัมประสิทธิ์พยากรณ์ ของตัวแปรต่างๆในแผนภาพเส้นทาง แสดงความสัมพันธ์เชิงสาเหตุกับค่าสารสนเทศของข้อสอบที่ระดับความสามารถ สูง ปานกลางและต่ำ เท่ากับ 0.9862 , .8930 และ .9853 ตามลำดับโดยมีประสิทธิภาพในการพยากรณ์เท่ากับ 88.29 % , 67.28 % และ 87.87 % ตามลำดับ เมื่อพิจารณาสัมประสิทธิ์เส้นทางที่มีนัยสำคัญทางสถิติ ณ ระดับ .01 พบว่ามี 4 ค่าคือ p_{41} , p_{54} , p_{73} และ p_{74} ซึ่งมีค่าเท่ากับ -.89 , .92 , .70 และ -.95 ตามลำดับ และสัมประสิทธิ์เส้นทางที่มีนัยสำคัญทางสถิติ ณ ระดับ .05 พบว่ามี 4 ค่าคือ p_{31} , p_{52} , p_{62} และ p_{72} ซึ่งมีค่าเท่ากับ -.31 , .16 , -.34 และ .11 ตามลำดับ



ภาพประกอบ 38 แผนภาพเส้นทางเชิงประจักษ์แสดงอิทธิพลทางตรงและทางอ้อมของตัวแปรต่างๆที่มีต่อสารสนเทศที่ระดับความสามารถสูง ปานกลาง และต่ำของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ได้จากการให้คะแนนตามรูปแบบปกติ ตามกรอบทฤษฎี IRT

จากภาพประกอบ 39 พบว่าตัวแปรที่มีอิทธิพลทางตรงเชิงบวกต่อค่าสารสนเทศที่ระดับความสามารถสูงของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้

คะแนนตามรูปแบบปกติ ได้แก่ ความยาก และความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศมีค่าสัมประสิทธิ์เส้นทางเท่ากับ .91 และ .16 ตามลำดับ

สำหรับตัวแปรที่ส่งอิทธิพลทางตรงเชิงลบต่อค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลาง ได้แก่ ค่าความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ -.40

และเมื่อพิจารณาสารสนเทศของข้อสอบที่ระดับความสามารถต่ำ พบว่า ตัวแปรที่ส่งอิทธิพลทางตรงเชิงบวกได้แก่ความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ .11 ส่วนความยาก พบว่า ส่งอิทธิพลเชิงลบซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ -.92

นอกจากนี้ยังพบว่า ความเป็นมิติเดียวส่งอิทธิพลทางอ้อมผ่านทางความยากเฉพาะสารสนเทศของข้อสอบที่ระดับความสามารถสูงและต่ำเท่านั้น โดยมีค่าสัมประสิทธิ์เส้นทางเท่ากับ -.90 และพบว่าความเหมาะสมของข้อสอบกับโมเดลไม่มีอิทธิพลต่อค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูง ปานกลาง และต่ำ

ตาราง 45 เปรียบเทียบความสอดคล้องของแผนภาพเส้นทางเชิงประจักษ์กับแผนภาพเส้นทางเต็มรูปของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนตามรูปแบบปกติ ตามกรอบทฤษฎี IRT

	แผนภาพเต็มรูป		แผนภาพเชิงประจักษ์		W
	R^2_{FM}	E	R^2_{OM}	E	
IIFH	.9862	88.29	.9654	81.40	1194
IIFM	.8930	67.28	.16	08.00	2678
IIFL	.9853	88.87	.9820	87.87	.026**

*มีนัยสำคัญทางสถิติ ณ ระดับ .05 *มีนัยสำคัญทางสถิติ ณ ระดับ .05

จากตาราง 45 พบว่าแผนภาพเส้นทางเชิงประจักษ์สอดคล้องกับแผนภาพเต็มรูป อย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01 เฉพาะ ค่าสารสนเทศข้อสอบที่ระดับความสามารถต่ำเท่านั้น ส่วนค่าสารสนเทศข้อสอบที่ระดับความสามารถสูง และปานกลางพบว่าไม่สอดคล้องกัน

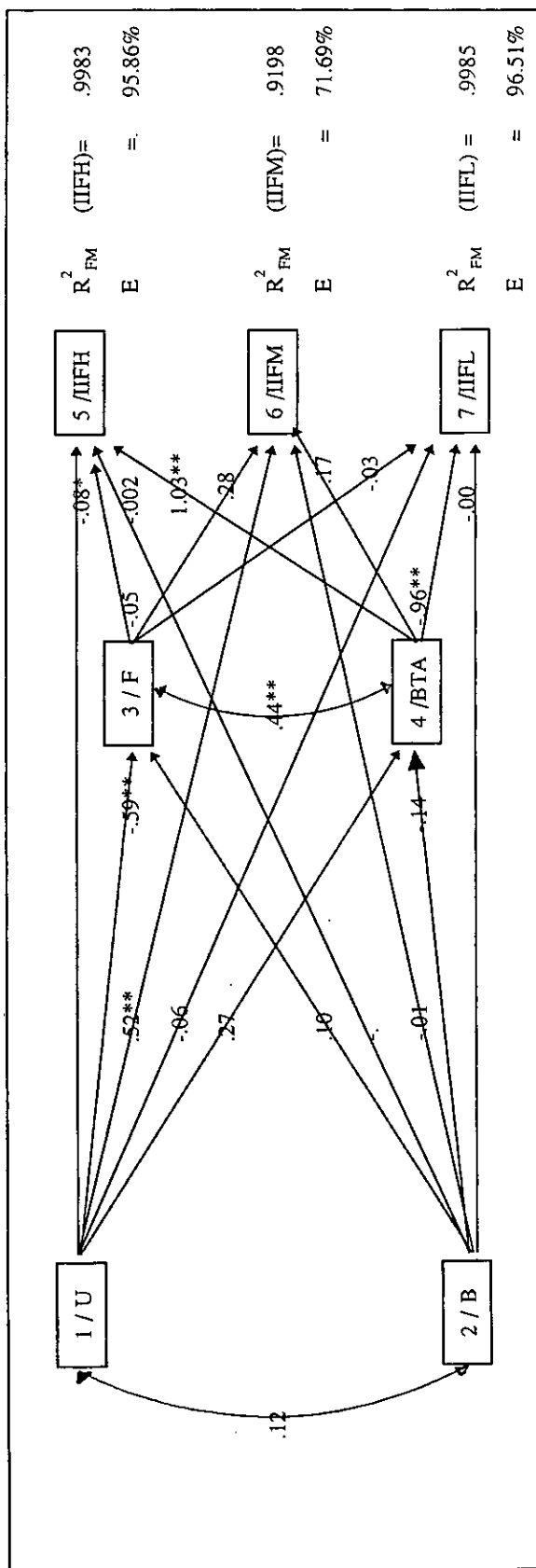
ตาราง 46 ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมที่มีต่อค่าสารสนเทศที่ระดับความสามารถสูง ปานกลาง และต่ำของแบบทดสอบวัดความสามารถในการแก้ไขโจทย์ปัญหาทางคณิตศาสตร์ที่ได้จากการให้คะแนนตามรูปแบบปกติ ตามกรอบทฤษฎี IRT

สารสนเทศ/อิทธิพล.		ตัวแปรสาเหตุ			
		BTA	F	U	B
IIFH	ทางตรง	.91	–	–	.16
	ทางอ้อม	–	–	-.82	–
	ปลอม	–	–	–	–
	รวม	.91	–	-.82	.16
IIFM	ทางตรง	–	–	–	-.40
	ทางอ้อม	–	–	–	–
	ปลอม	–	–	–	–
	รวม	–	–	–	-.40
IIFL	ทางตรง	-.92	–	–	.11
	ทางอ้อม	–	–	.83	–
	ปลอม	–	–	–	–
	รวม	-.92	–	.83	.11

จากตาราง 46 พบว่า ความยากมีอิทธิพลเชิงบวกต่อสารสนเทศข้อสอบที่ระดับความสามารถสูงของแบบทดสอบวัดความสามารถในการแก้ไขโจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนตามรูปแบบปกติมากที่สุดเท่ากับ .91 รองลงมา ได้แก่ ความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศมีค่าเท่ากับ .16 ส่วนอิทธิพลเชิงลบ ได้แก่ ความเป็นมิติเดียว มีค่าเท่ากับ -.82

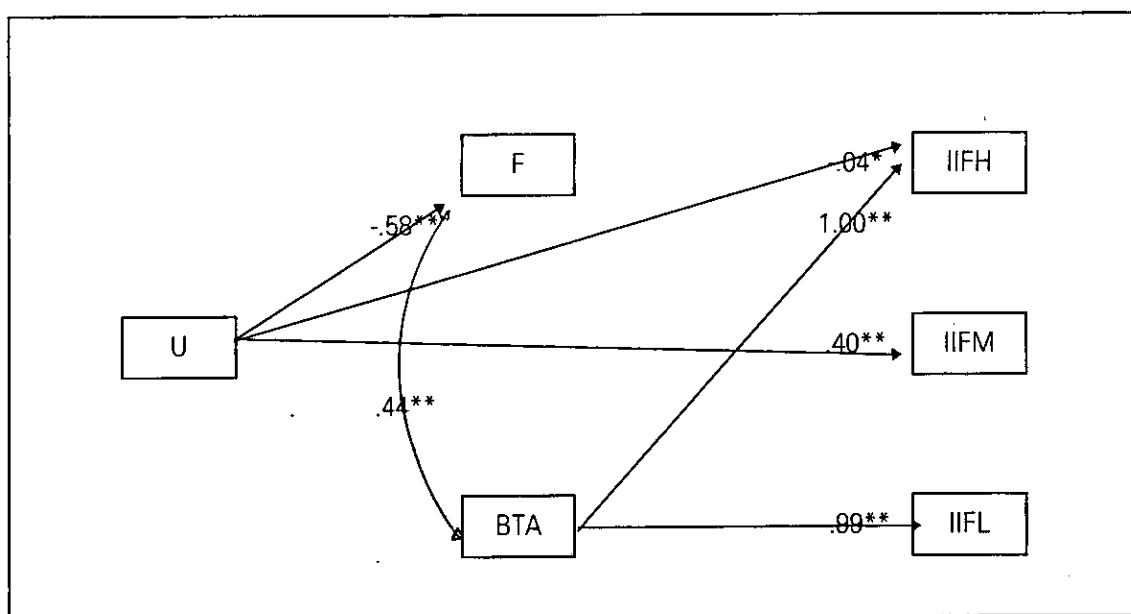
สำหรับค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลางพบว่า ตัวแปรที่มีอิทธิพลเชิงลบ ได้แก่ ความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ ซึ่งมีค่าเท่ากับ -.40

และพบว่า ตัวแปรที่มีอิทธิพลเชิงบวกต่อค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำ ได้แก่ ความเป็นมิติเดียว และความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ มีค่าเท่ากับ .83 และ .11 ตามลำดับ ส่วนความยากพบว่า มีอิทธิพลเชิงลบ มีค่าเท่ากับ -.92



ภาพประกอบ 39 แผนภาพเส้นทางเต็มรูปแบบแสดงความสัมพันธ์เชิงสาเหตุ ของตัวแปรต่างๆ กับค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูง ปานกลาง และต่ำ ของแบบทดสอบวัดความสามารถในการแก้ไขปัญหาทางคณิตศาสตร์ที่ให้คะแนนตามรูปแบบ 2-SR ตามกรอบทฤษฎี IRT

จากภาพประกอบที่ 39 พบว่าค่าสัมประสิทธิ์พยากรณ์ ของตัวแปรต่างๆในแผนภาพเส้นทาง แสดงความสัมพันธ์เชิงสาเหตุกับค่าสารสนเทศของข้อสอบที่ระดับความสามารถ สูง ปานกลาง และต่ำ เท่ากับ 0.9983 , .9198 และ .9985 โดยมีประสิทธิภาพในการพยากรณ์เท่ากับ 95.86 % , 71.69 % และ 96.51 % ตามลำดับ เมื่อพิจารณาสัมประสิทธิ์เส้นทางที่มีนัยสำคัญทางสถิติ ณ ระดับ .01 พบว่ามี 4 ค่าคือ p_{31} , p_{43} , p_{54} และ p_{74} ซึ่งมีค่าเท่ากับ -.59 , .44 , 1.03 และ -.96 ตามลำดับ และสัมประสิทธิ์เส้นทางที่มีนัยสำคัญทางสถิติ ณ ระดับ .05 พบว่ามี 2 ค่าคือ p_{51} และ p_{61} ซึ่งมีค่าเท่ากับ -.08 และ .52 ตามลำดับ



ภาพประกอบ 40 แผนภาพเส้นทางเชิงประจักษ์แสดงอิทธิพลทางตรงและทางอ้อมของตัวแปรต่างๆที่มีต่อสารสนเทศของข้อสอบ .ที่ระดับความสามารถสูง ปานกลาง และต่ำของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ได้จากการให้คะแนนตามรูปแบบ 2 - SR ตามกรอบทฤษฎี IRT

จากภาพประกอบ 40 พบว่า ตัวแปรที่ส่งอิทธิพลทางตรงเชิงบวกต่อค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูงของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ได้จากการให้คะแนนตามรูปแบบ 2-SR ได้แก่ค่าความยากซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ 1.00 และตัวแปรที่ส่งอิทธิพลทางตรงเชิงลบ ได้แก่ ความเป็นมิติเดียวซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ -.04

สำหรับตัวแปรที่มีอิทธิพลทางตรงเชิงบวกต่อค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลาง ได้แก่ ความเป็นมิติเดียวซึ่งมีค่าสัมประสิทธิ์เส้นทาง เท่ากับ .40

และพบว่าตัวแปรที่มีอิทธิพลทางตรงเชิงลบต่อค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำ ได้แก่ ความยากซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ -.99

นอกจากนี้ยังพบว่า ความเป็นมิติเดียวมีอิทธิพลทางอ้อมเชิงลบต่อสารสนเทศของข้อสอบที่ระดับความสามารถสูงและต่ำ ผ่านทางความเหมาะสมของข้อสอบกับโมเดลซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ -.58 และพบว่า ความเหมาะสมของข้อสอบกับโมเดลแสดงความสัมพันธ์กับความยากเท่ากับ .44 ซึ่งไม่สามารถระบุทิศทางเชิงสาเหตุได้

ตาราง 47 เปรียบเทียบความสอดคล้องของแผนภาพเส้นทางเชิงประจักษ์กับแผนภาพเส้นทางเต็มรูปของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนตามรูปแบบ 2 - SR ตามกรอบทฤษฎี IRT

	แผนภาพเต็มรูป		แผนภาพเชิงประจักษ์		W
	R^2_{FM}	E	R^2_{OM}	E	
IIFH	.9983	88.29	.9919	91.02	1856
IIFM	.9198	71.69	.2970	16.15	2819
IIFL	.9985	96.15	.9915	91.02	2190

*มีนัยสำคัญทางสถิติ ณ ระดับ .05 *มีนัยสำคัญทางสถิติ ณ ระดับ .05

จากตาราง 47 พบว่าแผนภาพเส้นทางเชิงประจักษ์ไม่สอดคล้องกับแผนภาพเต็มรูป ในทุกระดับความสามารถ

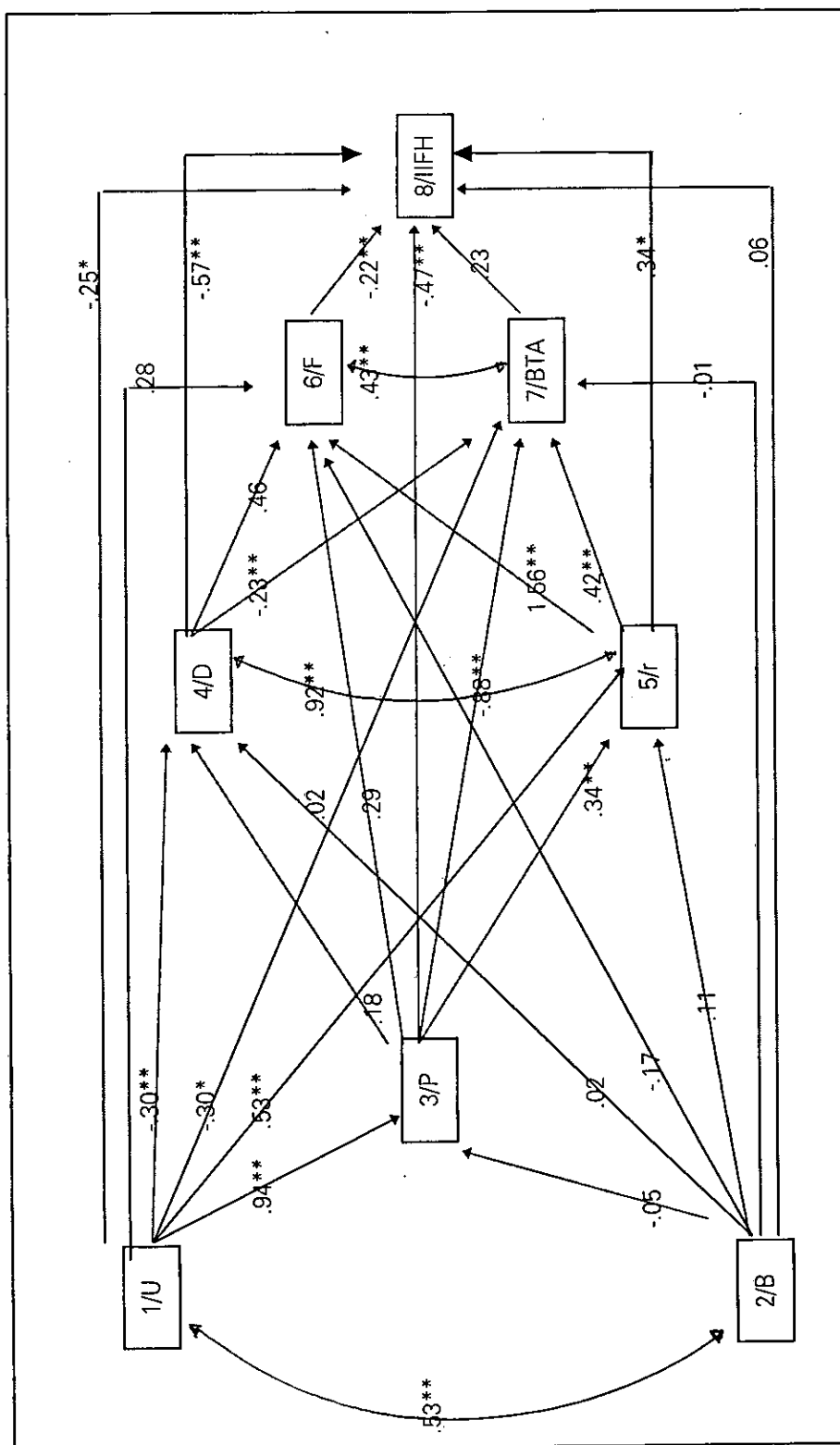
ตาราง 48 ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมที่มีต่อค่าสารสนเทศที่ระดับความสามารถสูง ปานกลาง และต่ำของแบบทดสอบวัดความสามารถในการแก้ไขโจทย์ปัญหาทางคณิตศาสตร์ที่ได้จากการให้คะแนนตามรูปแบบปกติ ตามกรอบทฤษฎี IRT

สารสนเทศ/อิทธิพล		ตัวแปรสาเหตุ			
		BTA	F	U	B
IIFH	ทางตรง	1.00	-	-.04	-
	ทางอ้อม	-	.44	-.26	-
	ปลอม	-	-	-	-
	รวม	1.00	.44	-.30	-
IIFM	ทางตรง	-	-	.40	-
	ทางอ้อม	-	-	-	-
	ปลอม	-	-	-	-
	รวม	-	-	.40	-
IIFL	ทางตรง	-.99		-	-
	ทางอ้อม	-	-.44	.25	
	ปลอม	-	-	-	-
	รวม	-.99	-.44	.25	

จากตาราง 47 พบว่า ความยากมีอิทธิพลเชิงบวกสูงสุดต่อค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูงของแบบทดสอบวัดความสามารถในการแก้ไขโจทย์ปัญหาที่ให้คะแนนแบบ 2-SR มีค่าเท่ากับ 1.00 รองลงมาคือความเหมาะสมของข้อสอบกับโมเดลมีค่าเท่ากับ .44 ส่วนความเป็นมิติเดียว พบว่ามีอิทธิพลเชิงลบ มีค่าเท่ากับ -.30

สำหรับตัวแปรที่ส่งอิทธิพลเชิงบวกต่อค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลาง ได้แก่ ความเป็นมิติเดียวซึ่งมีค่าเท่ากับ .40

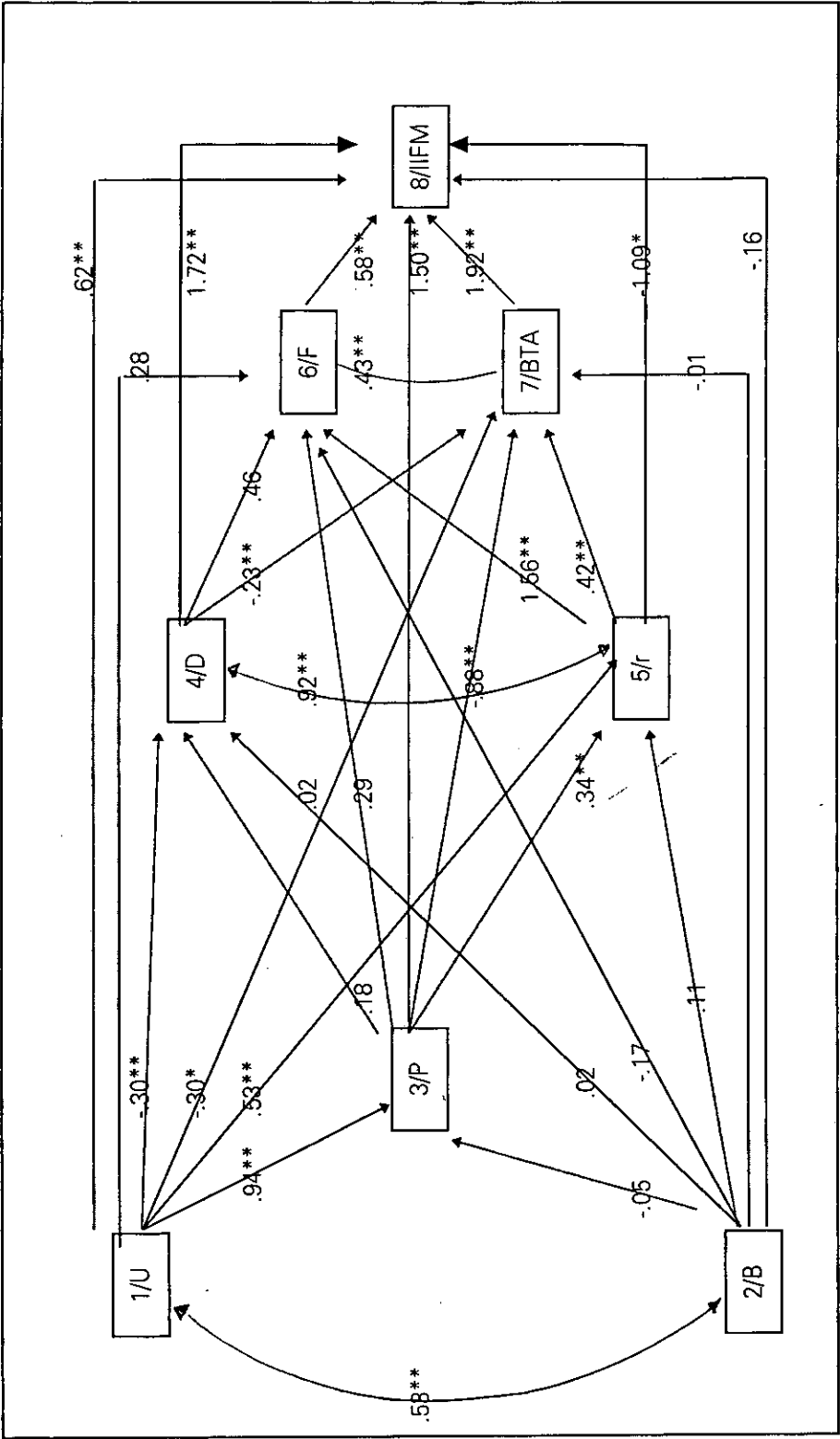
และพบว่าความเป็นมิติเดียว ส่งอิทธิพลเชิงบวกต่อค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำ ซึ่งมีค่าเท่ากับ .25 ส่วนตัวแปรที่ส่งอิทธิพลเชิงลบ ได้แก่ ความยาก และความเหมาะสมของข้อสอบกับโมเดล ซึ่งมีค่าเท่ากับ -.99 และ -.44 ตามลำดับ



** มีนัยสำคัญทางสถิติ ณ ระดับ .01

* มีนัยสำคัญทางสถิติ ณ ระดับ .05

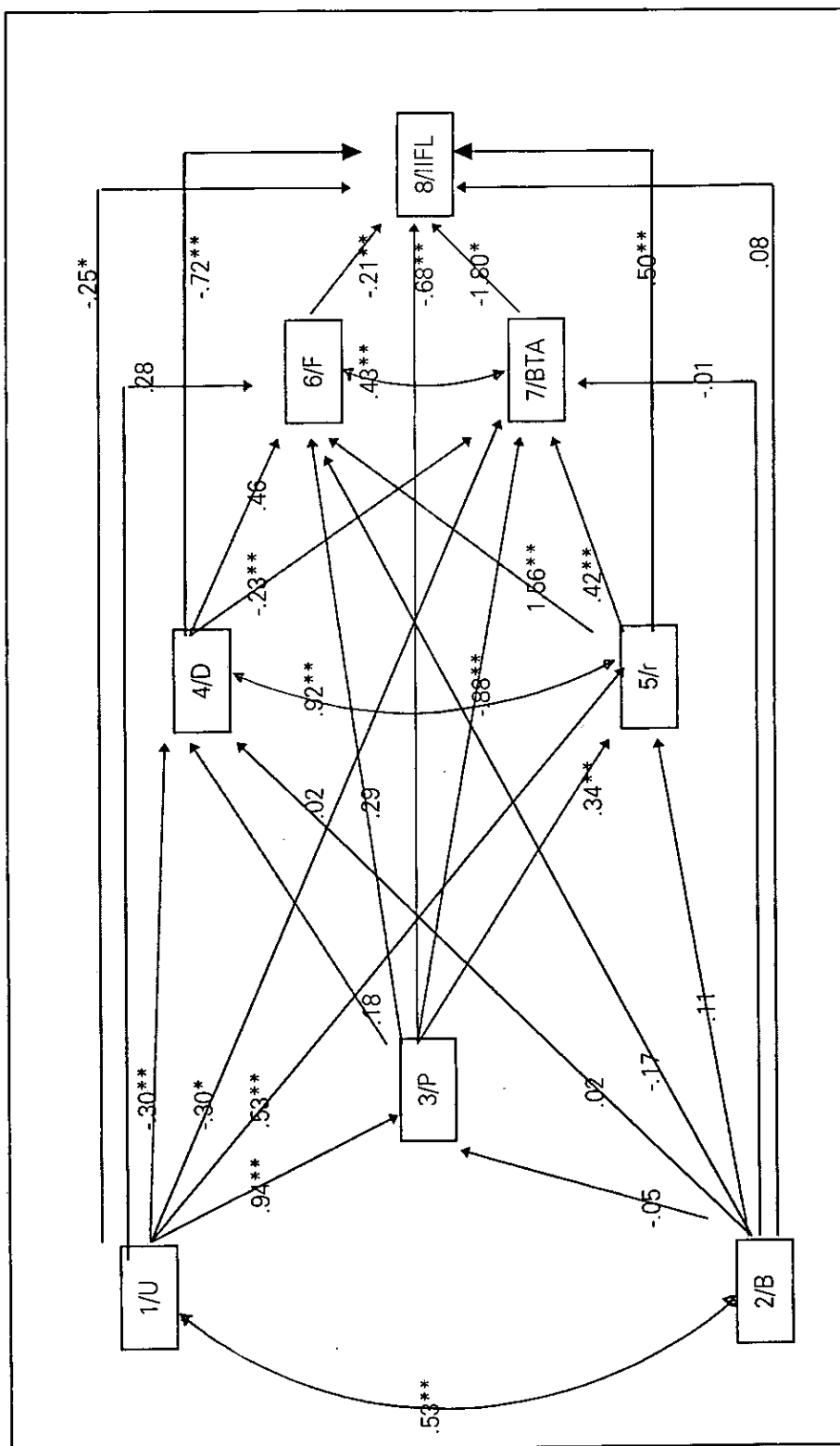
ภาพประกอบ 41 แผนภาพเส้นทางเต็มรูป แสดงความสัมพันธ์เชิงสาเหตุ ของตัวแปรต่างๆ กับค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูง ของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนตามรูปแบบปกติ ตามกรอบทฤษฎี CTT และ IRT



* มีนัยสำคัญทางสถิติ ณ ระดับ .05

** มีนัยสำคัญทางสถิติ ณ ระดับ .01

ภาพประกอบ 42 แผนภาพเส้นทางเต็มรูป แสดงความสัมพันธ์เชิงสาเหตุ ของตัวแปรต่างๆ กับค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลาง ของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนตามรูปแบบปกติ ตามกรอบทฤษฎี CTT และ IRT



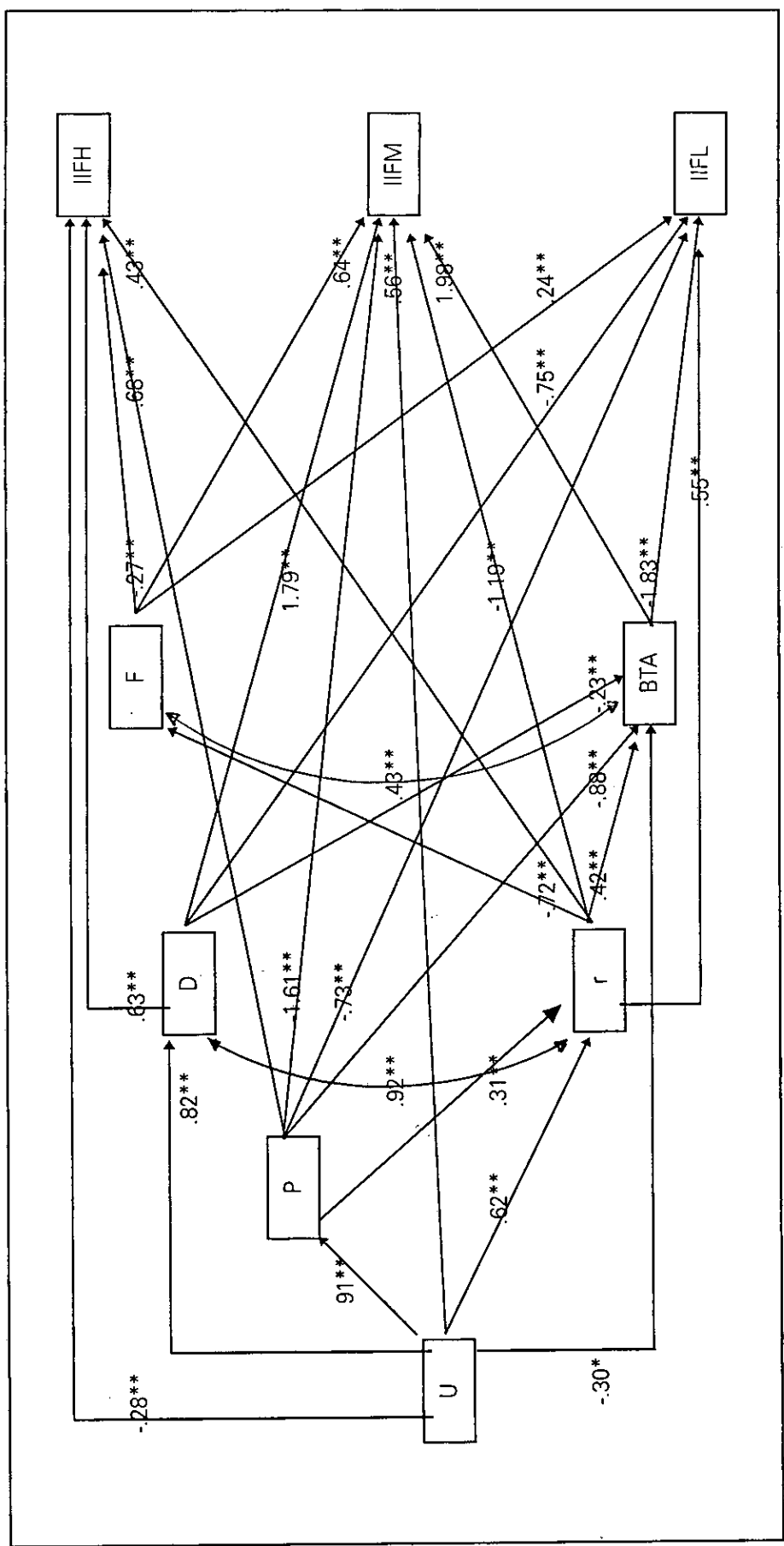
* มีนัยสำคัญทางสถิติ ณ ระดับ .05 ** มีนัยสำคัญทางสถิติ ณ ระดับ .01

ภาพประกอบ 43 แผนภาพเส้นทางเต็มรูป แสดงความสัมพันธ์เชิงสาเหตุ ของตัวแปรต่างๆ กับค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำ ของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้แก่แบบตามรูปแบบปกติ ตามกรอบทฤษฎี CTT และ IRT

จากภาพประกอบ 41 พบว่า สัมประสิทธิ์พยากรณ์ของตัวแปรต่างๆในแผนภาพเส้นทาง แสดงความสัมพันธ์เชิงสาเหตุกับค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูง มีค่าเท่ากับ 0.9999 และค่าประสิทธิภาพในการพยากรณ์เท่ากับ 99.93 % และพบว่าค่าสัมประสิทธิ์เส้นทางที่มี นัยสำคัญทางสถิติ ณ ระดับ .01 จำนวน 14 ค่า คือ p_{21} , p_{31} , p_{41} , p_{51} , p_{53} , p_{73} , p_{83} , p_{54} , p_{65} , p_{74} , p_{75} , p_{76} , p_{84} และ p_{86} ซึ่งมีค่าเท่ากับ .53 , .94 , -.30 , .53 , .34 , -.88 , -.47 , .92 , -1.56 , -.23 , .42 , .93 , -.57 และ -.22 ตามลำดับ และค่าสัมประสิทธิ์เส้นทางที่มีนัยสำคัญทางสถิติ ณ ระดับ .05 จำนวน 3 ค่า คือ p_{81} , p_{71} และ p_{85} ซึ่งมีค่าเท่ากับ -.25 , -.30 และ .34 ตามลำดับ

จากภาพประกอบ 42 พบว่า สัมประสิทธิ์พยากรณ์ของตัวแปรต่างๆในแผนภาพเส้นทาง แสดงความสัมพันธ์เชิงสาเหตุกับค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลาง มีค่า เท่ากับ 0.9999 และค่าประสิทธิภาพในการพยากรณ์เท่ากับ 99.93 % และพบว่าค่าสัมประสิทธิ์เส้น ทางที่มีนัยสำคัญทางสถิติ ณ ระดับ .01 จำนวน 17 ค่า คือ p_{21} , p_{31} , p_{41} , p_{51} , p_{61} , p_{53} , p_{73} , p_{83} , p_{54} , p_{65} , p_{74} , p_{75} , p_{76} , p_{84} , p_{85} , p_{86} และ p_{87} ซึ่งมีค่าเท่ากับ .53 , .94 , .84 , .53 , .62 , -.34 , -.88 , 1.50 , .92 , -1.56 , -.23 , .42 , .43 , 1.72 , -1.09 , .58 และ 1.92 ตามลำดับ และค่า สัมประสิทธิ์เส้นทางที่มีนัยสำคัญทางสถิติ ณ ระดับ .05 คือ p_{71} ซึ่งมีค่าเท่ากับ -.30

จากภาพประกอบ 43 พบว่า สัมประสิทธิ์พยากรณ์ของตัวแปรต่างๆในแผนภาพเส้นทาง แสดงความสัมพันธ์เชิงสาเหตุกับค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำ มีค่าเท่ากับ 0.9999 และค่าประสิทธิภาพในการพยากรณ์เท่ากับ 99.94 % และพบว่าค่าสัมประสิทธิ์เส้นทางที่มี นัยสำคัญทางสถิติ ณ ระดับ .01 จำนวน 15 ค่า คือ p_{21} , p_{31} , p_{41} , p_{51} , p_{53} , p_{73} , p_{83} , p_{54} , p_{74} , p_{84} , p_{65} , p_{75} , p_{85} , p_{86} และ p_{87} ซึ่งมีค่าเท่ากับ .53 , .94 , -.30 , .53 , .34 , -.88 , -.68 , .92 , -.23 , -.72 , -1.56 , .42 , .50 , -.21 และ -1.80 ตามลำดับ และค่าสัมประสิทธิ์เส้นทางที่มีนัยสำคัญ ทางสถิติ ณ ระดับ .05 จำนวน 2 ค่า คือ p_{71} , และ p_{81} ซึ่งมีค่าเท่ากับ -.30 , และ -.25 ตามลำดับ



* มีนัยสำคัญทางสถิติ ณ ระดับ .05 ** มีนัยสำคัญทางสถิติ ณ ระดับ .01

ภาพประกอบ 44 แผนภาพเส้นทางเชิงประจักษ์ แสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่างๆ กับสารสนเทศของข้อสอบที่ระดับความสามารถสูง ปานกลาง และต่ำของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนตามรูปแบบปกติ ตามกรอบทฤษฎี CTT และ IRT

จากภาพประกอบ 44 พบว่า สัมประสิทธิ์การจำแนกส่งอิทธิพลทางตรงเชิงบวกต่อค่าสารสนเทศข้อสอบที่ระดับความสามารถสูงของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนแบบปกติซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ .43 ส่วนความยากแบบ CTT ดัชนีการจำแนก ความเป็นมิติเดียว และความเหมาะสมของข้อสอบกับโมเดล พบว่า ส่งอิทธิพลทางตรงเชิงลบ ซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ -.68 , -.63 , -.28 และ -.27 ตามลำดับ

สำหรับตัวแปรที่ส่งอิทธิพลทางตรงเชิงบวกต่อค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลาง ได้แก่ ค่าความยากแบบ IRT ดัชนีการจำแนก ความยากแบบ CTT ความเหมาะสมของข้อสอบกับโมเดล และความเป็นมิติเดียว ซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ 1.98 , 1.79 , 1.61 , .64 และ .56 ตามลำดับ ส่วนตัวแปรที่ส่งอิทธิพลทางตรงเชิงลบ ได้แก่ สัมประสิทธิ์การจำแนก ซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ -1.19

และพบว่าตัวแปรที่ส่งอิทธิพลทางตรงเชิงบวกต่อค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำ ได้แก่ สัมประสิทธิ์การจำแนกซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ .55 ส่วนความยากแบบ IRT ดัชนีการจำแนก ความยากแบบ CTT และความเหมาะสมของข้อสอบกับโมเดล พบว่า ส่งอิทธิพลทางตรงเชิงลบซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ -1.83 , -.75 , -.73 , และ -.24 ตามลำดับ

เมื่อพิจารณาอิทธิพลทางอ้อมของตัวแปรพบว่า ความเป็นมิติเดียวส่งอิทธิพลเชิงบวกต่อสารสนเทศของข้อสอบที่ระดับความสามารถ สูง ปานกลาง และต่ำ ผ่านทางความยากแบบ CTT ดัชนีการจำแนก และสัมประสิทธิ์การจำแนก โดยมีค่าสัมประสิทธิ์เส้นทางเท่ากับ .91 , .82 และ .62 ตามลำดับ และส่งอิทธิพลทางอ้อมเชิงลบต่อสารสนเทศของข้อสอบ ณ ระดับความสามารถปานกลางและต่ำผ่านความยากแบบ IRT ซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ -.30 และพบความสัมพันธ์ระหว่างสัมประสิทธิ์การจำแนก และดัชนีการจำแนกเท่ากับ .92 และระหว่างความยากแบบ IRT กับ ความเหมาะสมของข้อสอบกับโมเดลเท่ากับ .43 ซึ่งความสัมพันธ์ทั้งสองคู่นี้ไม่สามารถระบุทิศทางเชิงสาเหตุได้

ตาราง 49 เปรียบเทียบความสอดคล้องของแผนภาพเส้นทางเชิงประจักษ์กับแผนภาพเส้นทาง
เต็มรูปของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนตามรูปแบบปกติ
ตามกรอบทฤษฎี CTT และ IRT

	แผนภาพเต็มรูป		แผนภาพเชิงประจักษ์		W
	R^2_{FM}	E	R^2_{OM}	E	
IIFH	0.9999	99.93	0.9999	99.93	0**
IIFM	0.9999	99.93	0.9999	99.93	0**
IIFL	0.9999	99.94	0.9999	99.94	0**

* มีนัยสำคัญทางสถิติ ณ ระดับ .05 ** มีนัยสำคัญทางสถิติ ณ ระดับ .01

จากตาราง 49 พบว่าแผนภาพเส้นทางเชิงประจักษ์สอดคล้องกับแผนภาพเส้นทางเต็มรูป
อย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01 ในทุกระดับความสามารถ

ตาราง 50 ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมที่มีต่อค่าสาร
สนเทศที่ระดับความสามารถสูง ของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้
จากการให้คะแนนตามรูปแบบปกติ ตามกรอบทฤษฎี CTT และ IRT

ตัวแปรสาเหตุ	อิทธิพลทางตรง	อิทธิพลทางอ้อม	อิทธิพลปลอม	รวม
r	.43	.19	–	.63
P	-.68	.29	–	-.39
D	-.63	.50	–	-.13
F	-.27	–	–	-.27
U	-.28	.18	–	-.10
BTA	–	-.11	–	-.11

จากตาราง 50 พบว่าสัมประสิทธิ์การจำแนกส่งอิทธิพลเชิงบวกต่อค่าสารสนเทศของข้อ
สอบที่ระดับความสามารถสูงของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนน
แบบปกติ ซึ่งมีค่าเท่ากับ .63 และ พบว่า ความยากแบบ CTT มีอิทธิพลเชิงลบมากที่สุดเท่ากับ -.39

รองลงมาได้แก่ ความเหมาะสมของข้อสอบกับโมเดล ดัชนีการจำแนก ความยากแบบIRT และ
ความเป็นมิติเดียว มีค่าเท่ากับ -.27 , -.13 , -.11 และ -.10 ตามลำดับ

ตาราง 51 ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมที่มีต่อค่า
สารสนเทศที่ระดับความสามารถปานกลาง ของแบบทดสอบวัดความสามารถในการอ่าน
จับใจความที่ได้จากการให้คะแนนตามรูปแบบปกติ ตามกรอบทฤษฎี CTT และ IRT

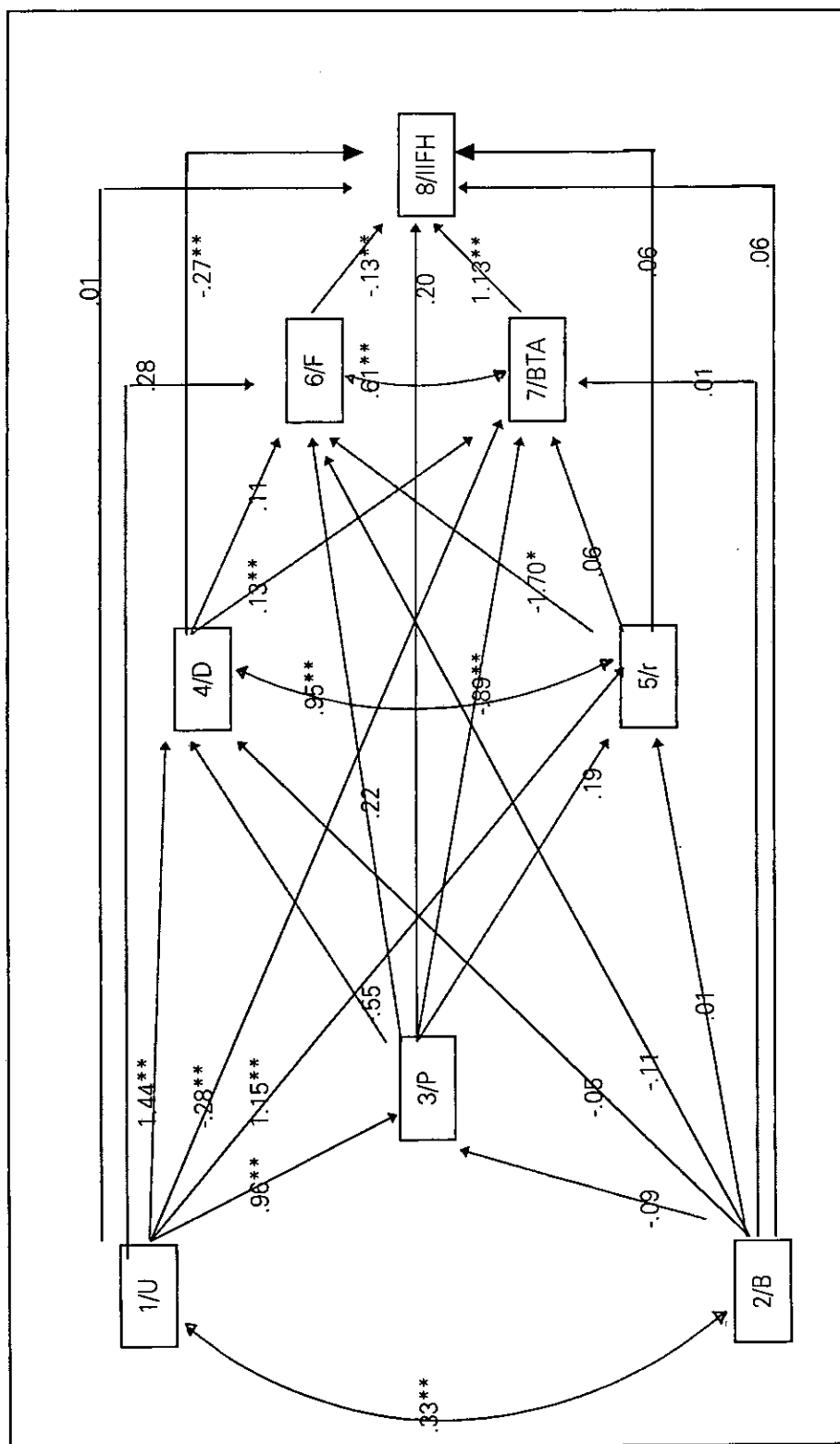
ตัวแปรสาเหตุ	อิทธิพลทางตรง	อิทธิพลทางอ้อม	อิทธิพลปลอม	รวม
BTA	1.98	.31	-	2.29
D	1.79	-.86	-	.92
P	1.61	-1.58	-	.03
F	.64	.97	-	1.61
U	.56	-.18	-	.38
r	-1.19	.73	-	-.45

จากตาราง 51 พบว่า ความยากแบบ IRT ความเหมาะสมของโมเดลกับข้อสอบ ดัชนีการ
จำแนก ความเป็นมิติเดียว และความยากแบบ CTT ส่งอิทธิพลเชิงบวกต่อสารสนเทศของข้อสอบ
ที่ระดับความสามารถปานกลางของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้
คะแนนโดยวิธีแบบปกติมีค่าเท่ากับ 2.29 , 1.61 , .92 , .38 และ .03 ตามลำดับ และพบว่า
สัมประสิทธิ์การจำแนกส่งอิทธิพลเชิงลบต่อค่าสารสนเทศของข้อสอบที่ระดับความสามารถ
ปานกลางมีค่าเท่ากับ -.45

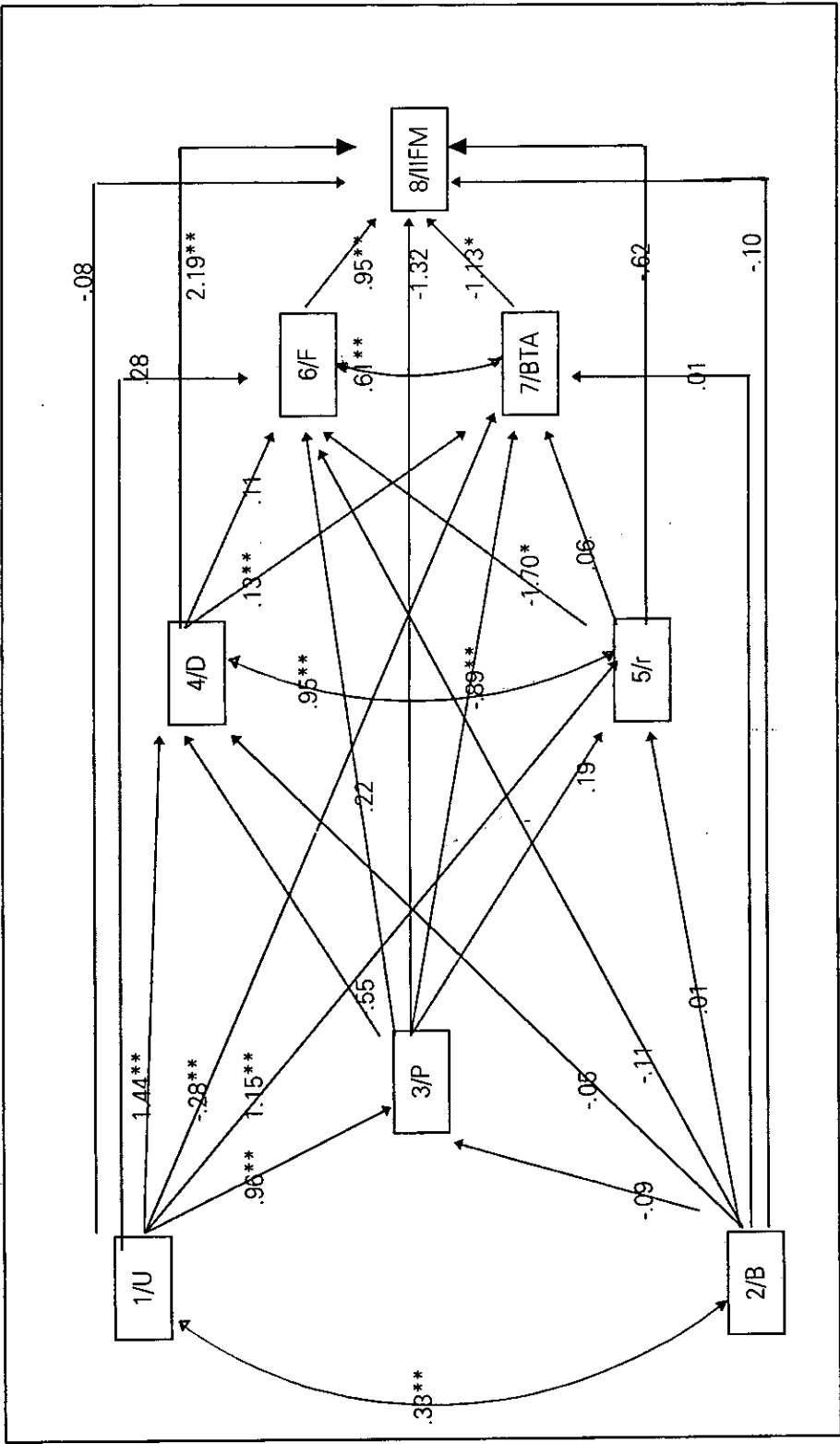
ตาราง 52 ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมที่มีต่อค่า
 สารสนเทศที่ระดับความสามารถต่ำของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่
 ได้จากการให้คะแนนตามรูปแบบปกติ ตามกรอบทฤษฎี CTT และ IRT

ตัวแปรสาเหตุ	อิทธิพลทางตรง	อิทธิพลทางอ้อม	อิทธิพลปลอม	รวม
r	.55	.34	—	.89
BTA	-1.83	-.12	—	-1.95
D	-.75	.87	—	.12
P	-.73	1.99	—	1.26
F	-.24	-.90	—	-1.14
U	—	2.38	—	2.38

จากตาราง 52 พบว่าความเป็นมิติเดียว มีอิทธิพลเชิงบวกต่อสารสนเทศที่ระดับความ
 สามารถต่ำมากที่สุดมีค่าเท่ากับ 2.38 รองลงมาได้แก่ความยากแบบ CTT สัมประสิทธิ์การจำแนก
 และดัชนีการจำแนก มีค่าเท่ากับ 1.26 , .89 และ .12 ตามลำดับ ส่วนตัวแปรที่ส่งอิทธิพลเชิงลบได้
 แก่ความยากแบบ IRT และความเหมาะสมของข้อสอบกับโมเดล ซึ่งมีค่าเท่ากับ -1.95 และ -1.14
 ตามลำดับ



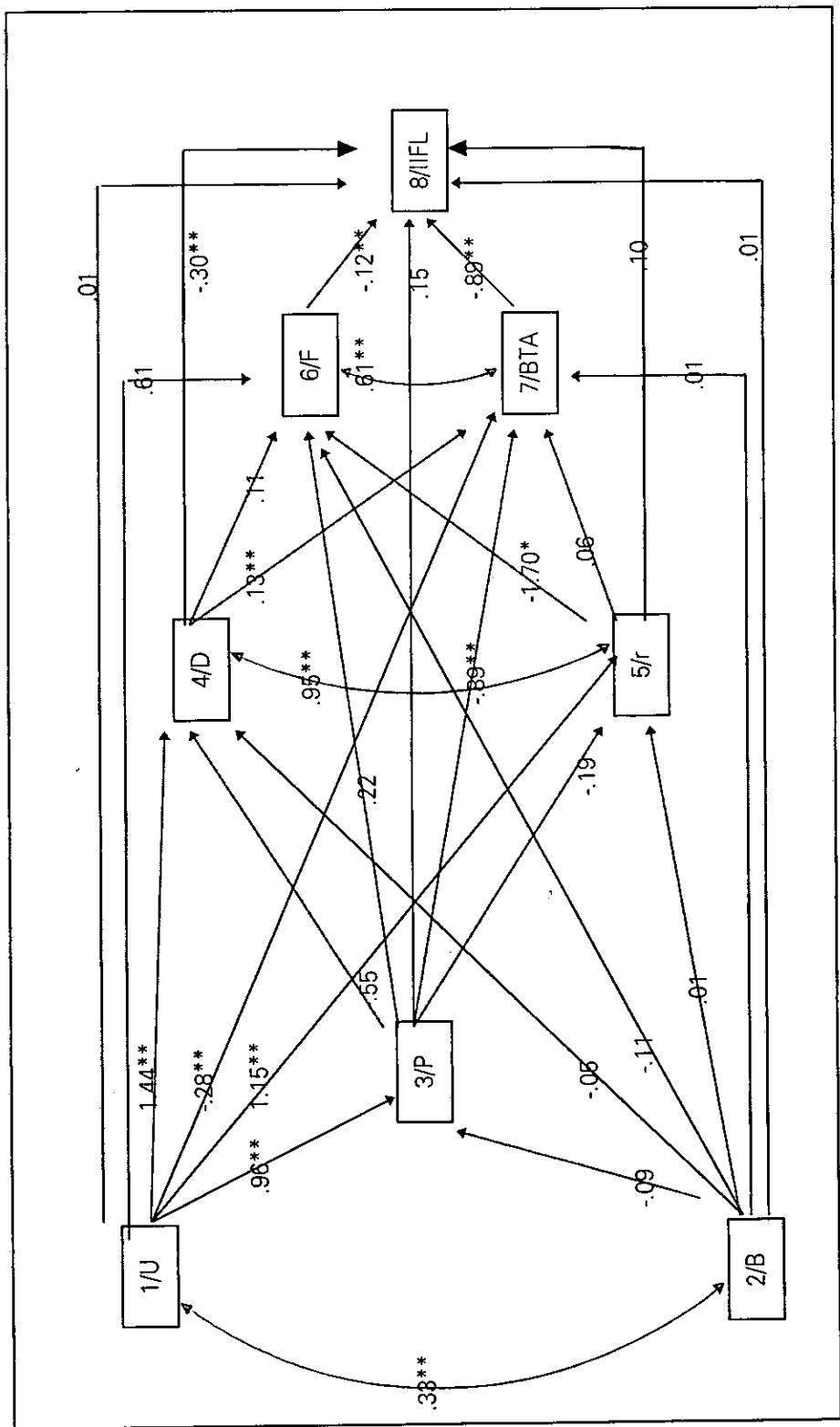
ภาพประกอบ 45 แผนภาพเส้นทางเดิมรูป แสดงความสัมพันธ์เชิงสาเหตุ ของตัวแปรต่างๆ กับค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูง ของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนตามรูปแบบ 2-SR ตามกรอบทฤษฎี CTT และ IRT



* มีนัยสำคัญทางสถิติ ณ ระดับ .01

* มีนัยสำคัญทางสถิติ ณ ระดับ .05

ภาพประกอบ 46 แผนภาพเส้นทางเต็มรูป แสดงความสัมพันธ์ที่แท้จริงของตัวแปรต่างๆ กับค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลาง ของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้คะแนนตามรูปแบบ 2-SR ตามกรอบทฤษฎี CTT และ IRT



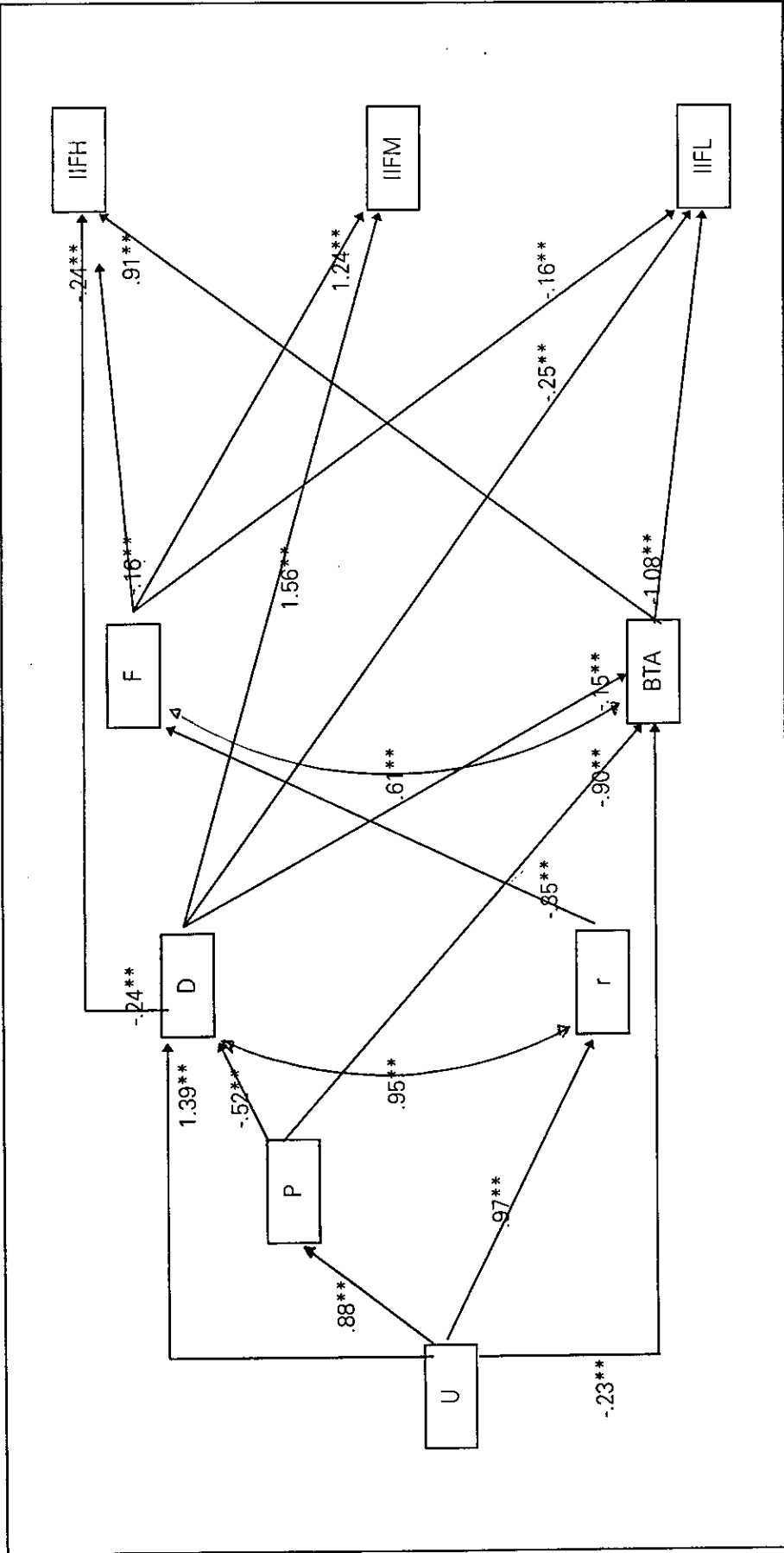
* มีนัยสำคัญทางสถิติ ณ ระดับ .05 ** มีนัยสำคัญทางสถิติ ณ ระดับ .01

ภาพประกอบ 47 แผนภาพเส้นทางเต็มรูป แสดงความสัมพันธ์ฟังก์ชันสาเหตุ ของตัวแปรต่างๆ กับค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำ ของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนตามรูปแบบ 2-SR ตามกรอบทฤษฎี CTT และ IRT

จากภาพประกอบ 45 พบว่า สัมประสิทธิ์พยากรณ์ของตัวแปรต่างๆในแผนภาพเส้นทาง แสดงความสัมพันธ์เชิงสาเหตุกับค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูง มีค่าเท่ากับ 0.9999 และค่าประสิทธิภาพในการพยากรณ์เท่ากับ 99.99 % และพบว่าค่าสัมประสิทธิ์เส้นทางที่มีนัยสำคัญทางสถิติ ณ ระดับ .01 จำนวน 13 ค่า คือ p_{21} , p_{31} , p_{41} , p_{51} , p_{71} , p_{73} , p_{54} , p_{74} , p_{65} , p_{76} , p_{84} , p_{86} และ p_{87} ซึ่งมีค่าเท่ากับ .33 , .96 , 1.44 , 1.15 , -.28 , .89 , .95 , .13 , -1.70 , .61 , -.27 , -.13 และ 1.13 ตามลำดับ และค่าสัมประสิทธิ์เส้นทางที่มีนัยสำคัญทางสถิติ ณ ระดับ .05 จำนวน 3 ค่า คือ p_{81} , p_{71} และ p_{85} ซึ่งมีค่าเท่ากับ -.25 , -.30 และ .34 ตามลำดับ

จากภาพประกอบ 46 พบว่า สัมประสิทธิ์พยากรณ์ของตัวแปรต่างๆในแผนภาพเส้นทาง แสดงความสัมพันธ์เชิงสาเหตุกับค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลาง มีค่าเท่ากับ 0.9999 และค่าประสิทธิภาพในการพยากรณ์เท่ากับ 99.99 % และพบว่าค่าสัมประสิทธิ์เส้นทางที่มีนัยสำคัญทางสถิติ ณ ระดับ .01 จำนวน 12 ค่า คือ p_{21} , p_{31} , p_{41} , p_{51} , p_{71} , p_{73} , p_{54} , p_{74} , p_{65} , p_{76} , p_{84} และ p_{86} ซึ่งมีค่าเท่ากับ .33 , .96 , 1.44 , 1.15 , -.28 , -.89 , .95 , .13 , -1.70 , .61 , 2.19 และ .95 ตามลำดับ

จากภาพประกอบ 47 พบว่า สัมประสิทธิ์พยากรณ์ของตัวแปรต่างๆในแผนภาพเส้นทาง แสดงความสัมพันธ์เชิงสาเหตุกับค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำ มีค่าเท่ากับ 0.9999 และค่าประสิทธิภาพในการพยากรณ์เท่ากับ 99.99 % และพบว่าค่าสัมประสิทธิ์เส้นทางที่มีนัยสำคัญทางสถิติ ณ ระดับ .01 จำนวน 13 ค่า คือ p_{21} , p_{31} , p_{41} , p_{51} , p_{71} , p_{73} , p_{54} , p_{74} , p_{65} , p_{75} , p_{84} , p_{86} และ p_{87} ซึ่งมีค่าเท่ากับ .33 , .96 , 1.44 , 1.15 , -.28 , -.89 , .95 , .13 , -1.70 , .61 , -.30 , -.12 และ -.89 ตามลำดับ



ภาพประกอบ 48 แผนภาพเส้นทางเชิงประจักษ์ แสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่างๆ กับสารสนเทศของข้อสอบที่ระดับความสามารถสูง ปานกลาง และต่ำของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนตามรูปแบบ 2-SR ตามกรอบทฤษฎี CTT และ IRT

จากภาพประกอบที่ 48 พบว่า ความยากแบบ IRT มีอิทธิพลทางตรงเชิงบวกต่อค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูงของแบบทดสอบวัดความสามารถในการอ่าน จับใจความที่ให้คะแนนแบบ 2 - SR มากที่สุด ซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ .91 ส่วนตัวแปรที่ส่งผลทางตรงเชิงลบคือ ดัชนีการจำแนก และความเหมาะสมของข้อสอบกับโมเดลซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ -.24 และ -.16 ตามลำดับ

สำหรับค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลาง พบว่า ตัวแปรที่ส่งอิทธิพลทางตรงเชิงบวกคือ ดัชนีการจำแนก และความเหมาะสมของข้อสอบกับโมเดล ซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ 1.56 และ 1.24 ตามลำดับ

และพบว่าตัวแปรที่ส่งอิทธิพลทางตรงเชิงลบต่อสารสนเทศของข้อสอบที่ระดับความสามารถต่ำคือ ความยากแบบ IRT ดัชนีการจำแนก และความเหมาะสมของข้อสอบกับโมเดล ซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ -1.08, -.25 และ -.16 ตามลำดับ

เมื่อพิจารณาอิทธิพลทางอ้อมของตัวแปรต่าง ๆ พบว่า ความเป็นมิติเดียวมีอิทธิพลเชิงบวกต่อค่าสารสนเทศที่ระดับความสามารถต่าง ๆ ผ่านทางดัชนีการจำแนก สัมประสิทธิ์การจำแนก ความยากแบบ CTT ซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ 1.39, .97 และ .88 ตามลำดับ และพบว่าส่งอิทธิพลเชิงลบผ่านความยากแบบ IRT โดยมีค่าสัมประสิทธิ์เส้นทางเท่ากับ -.23 ตามลำดับ

สำหรับค่าสัมประสิทธิ์การจำแนก พบว่ามีอิทธิพลเชิงลบต่อค่าสารสนเทศของข้อสอบที่ระดับต่าง ๆ ผ่านทางความเหมาะสมของข้อสอบกับโมเดล โดยมีค่าสัมประสิทธิ์เส้นทางเท่ากับ -.85

ส่วนค่าความยากแบบ CTT พบว่าส่งอิทธิพลเชิงลบต่อค่าสารสนเทศของข้อสอบที่ระดับต่าง ๆ ผ่านทางความยากแบบ IRT ซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ -.90

และพบว่า ดัชนีการจำแนกส่งอิทธิพลเชิงลบต่อค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูง และต่ำ ผ่านทางความยากแบบ IRT โดยมีค่าสัมประสิทธิ์เส้นทางเท่ากับ -.15

และพบความสัมพันธ์ระหว่างสัมประสิทธิ์การจำแนก และดัชนีการจำแนกเท่ากับ .95 และระหว่างความยากแบบ IRT กับ ความเหมาะสมของข้อสอบกับโมเดลเท่ากับ .61 ซึ่งความสัมพันธ์ทั้งสองคู่นี้ไม่สามารถระบุทิศทางเชิงสาเหตุได้

ตาราง 53 เปรียบเทียบความสอดคล้องของแผนภาพเส้นทางเชิงประจักษ์กับแผนภาพเส้นทางเต็มรูปของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนตามรูปแบบ 2- SR ตามกรอบทฤษฎี CTT และ IRT

	แผนภาพเต็มรูป		แผนภาพเชิงประจักษ์		W
	R^2_{FM}	E	R^2_{OM}	E	
IIFH	0.9999	99.99	0.9999	99.99	0**
IIFM	0.9999	99.99	0.9999	99.99	0**
IIFL	0.9999	99.99	0.9999	99.99	0**

* มีนัยสำคัญทางสถิติ ณ ระดับ .05 ** มีนัยสำคัญทางสถิติ ณ ระดับ .01

จากตาราง 53 พบว่าแผนภาพเส้นทางเชิงประจักษ์สอดคล้องกับแผนภาพเส้นทางเต็มรูปอย่างมีนัยสำคัญทางสถิติ ณ ระดับ.01 ในทุกระดับความสามารถ

ตาราง 54 ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมที่มีต่อค่าสารสนเทศที่ระดับความสามารถสูงของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนตามรูปแบบ 2 - SR ตามกรอบทฤษฎี CTT และ IRT

ตัวแปรสาเหตุ	อิทธิพลทางตรง	อิทธิพลทางอ้อม	อิทธิพลปลอม	รวม
BTA	.91	-.08	-	.83
D	-.24	-.23	-	-.47
F	-.16	.44	-	.28
U	-	1.68	-	1.68
r	-	-.40	-	-.40
P	-	-.51	-	-.51

จากตาราง 54 พบว่า ความเป็นมิติเดียว มีอิทธิพลเชิงบวกสูงสุดต่อค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูงของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนตามรูปแบบ 2 - SR มีค่าเท่ากับ 1.68 รองลงมาได้แก่ ความยากแบบ IRT ความเหมาะสม

ของข้อสอบกับโมเดล และความยากแบบ CTT ซึ่งมีค่าเท่ากับ .83 , .28 และ .10 ตามลำดับ ส่วนตัวแปรที่มีอิทธิพลเชิงลบ ได้แก่ สัมประสิทธิ์การจำแนก และดัชนีการจำแนก ซึ่งมีค่าเท่ากับ -.47 และ -.40 ตามลำดับ

ตาราง 55 ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมที่มีต่อค่าสารสนเทศที่ระดับความสามารถปานกลาง ของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนตามรูปแบบ 2 - SR ตามกรอบทฤษฎี CTT และ IRT

ตัวแปรสาเหตุ	อิทธิพลทางตรง	อิทธิพลทางอ้อม	อิทธิพลปลอม	รวม
D	1.56	-.56	-	1.00
F	1.24	-	-	1.24
BTA	-	.61	-	.61
r	-	-1.30	-	-1.30
P	-	-1.07	-	-1.07
U	-	-.95	-	-.95

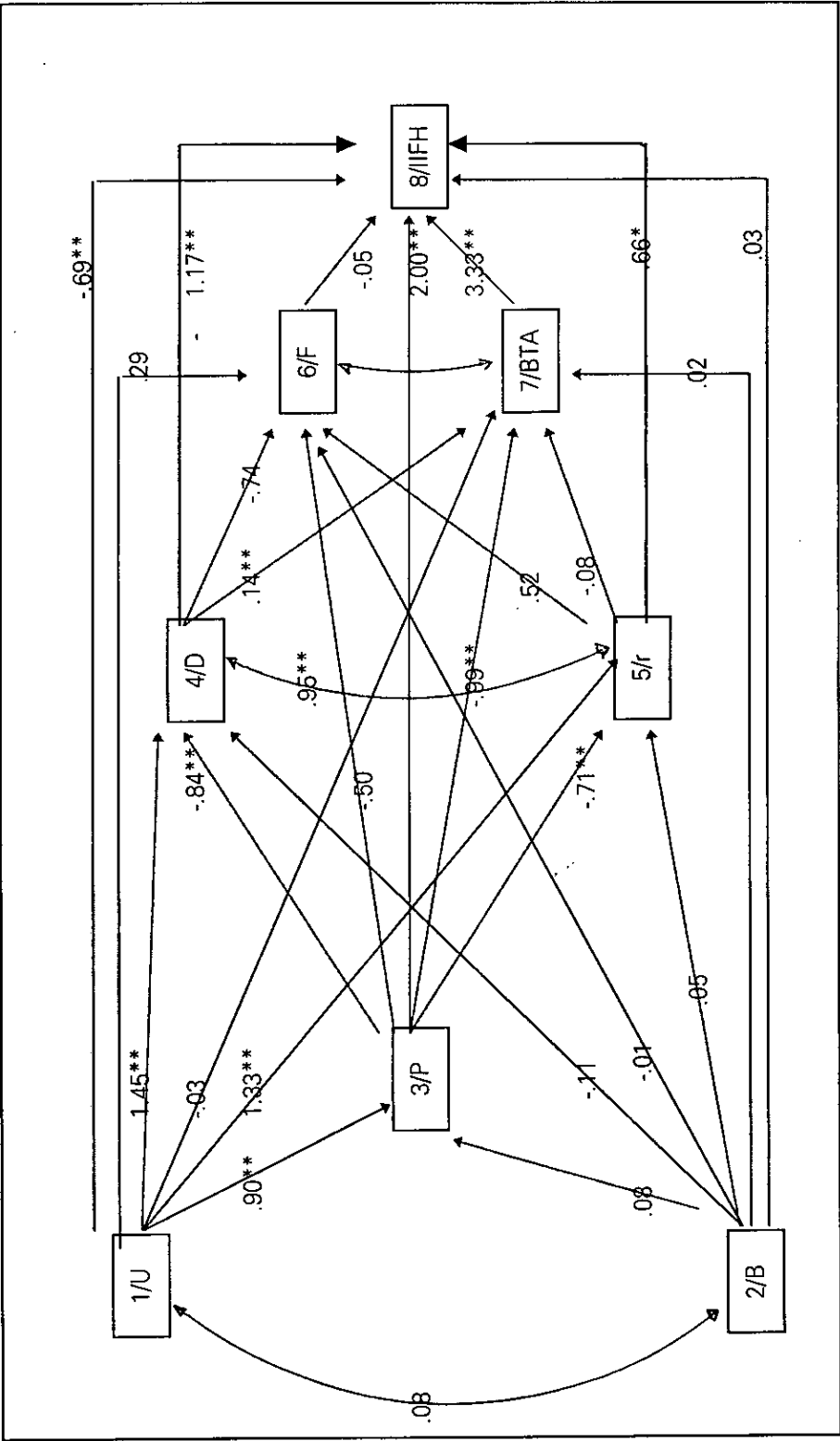
จากตาราง 55 พบว่า ความเหมาะสมของข้อสอบกับโมเดล ส่งอิทธิพลเชิงบวกต่อสารสนเทศของข้อสอบที่ระดับความสามารถปานกลางของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนแบบ 2 - SR สูงสุดซึ่งมีค่าเท่ากับ 1.24 รองลงมาได้แก่ดัชนีการจำแนก และความยากแบบ IRT ซึ่งมีค่าเท่ากับ 1.00 และ .61 ตามลำดับ ส่วนตัวแปรที่ส่งอิทธิพลเชิงลบได้แก่ สัมประสิทธิ์การจำแนก ความยากแบบ CTT และความเป็นมิติเดียว ซึ่งมีค่าเท่ากับ -1.30 , -1.07 และ -.95 ตามลำดับ

ตาราง 56 ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมที่มีต่อ

ค่าสารสนเทศที่ระดับความสามารถต่ำของแบบทดสอบวัดความสามารถในการอ่านจับใจความ
ที่ได้จากการให้คะแนนตามรูปแบบ 2 - SR ตามกรอบทฤษฎี IRT

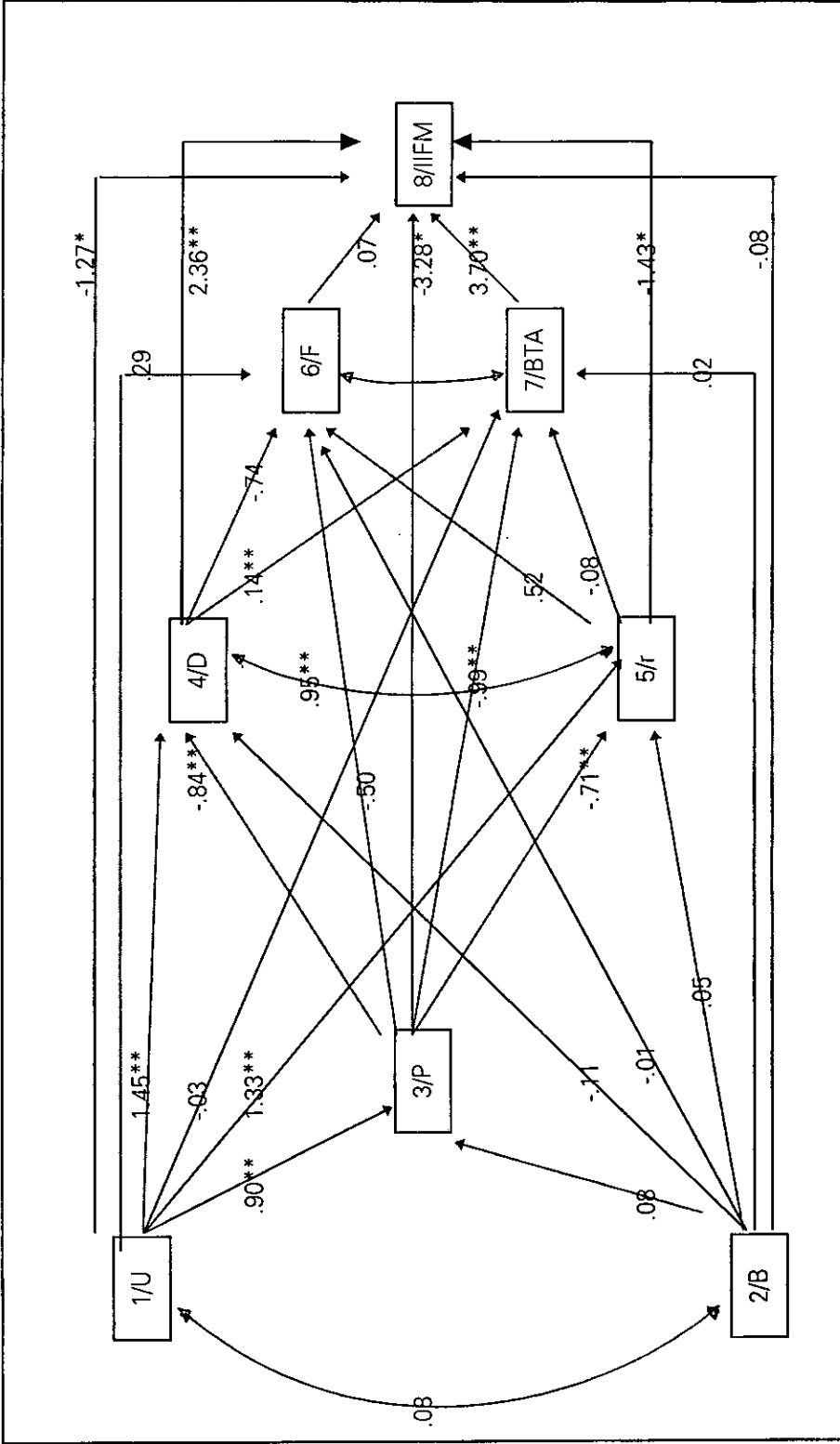
ตัวแปรสาเหตุ	อิทธิพลทางตรง	อิทธิพลทางอ้อม	อิทธิพลปลอม	รวม
BTA	-1.08	-.08	-	-1.16
D	-.25	.22	-	-.03
F	-.16	-.53	-	-.69
r	-	.22	-	.22
P	-	1.26	-	1.26
U	-	1.55	-	1.55

จากตาราง 56 พบว่า ความเป็นมิติเดียว ส่งอิทธิพลเชิงบวกต่อสารสนเทศของข้อสอบที่ระดับความสามารถต่ำของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้จากการให้คะแนนตามรูปแบบ 2 - SR สูงสุดซึ่งมีค่าเท่ากับ 1.55 รองลงมาได้แก่ ความยากแบบ CTT และสัมประสิทธิ์การจำแนก ซึ่งมีค่าเท่ากับ 1.26 และ .22 ตามลำดับ ส่วนตัวแปรที่ส่งอิทธิพลเชิงลบได้แก่ ความยากแบบ IRT ความเหมาะสมของข้อสอบกับโมเดล และดัชนีการจำแนก ซึ่งมีค่าเท่ากับ -1.16 , -.69 และ -.03 ตามลำดับ



* มีนัยสำคัญทางสถิติ ณ ระดับ .05 ** มีนัยสำคัญทางสถิติ ณ ระดับ .01

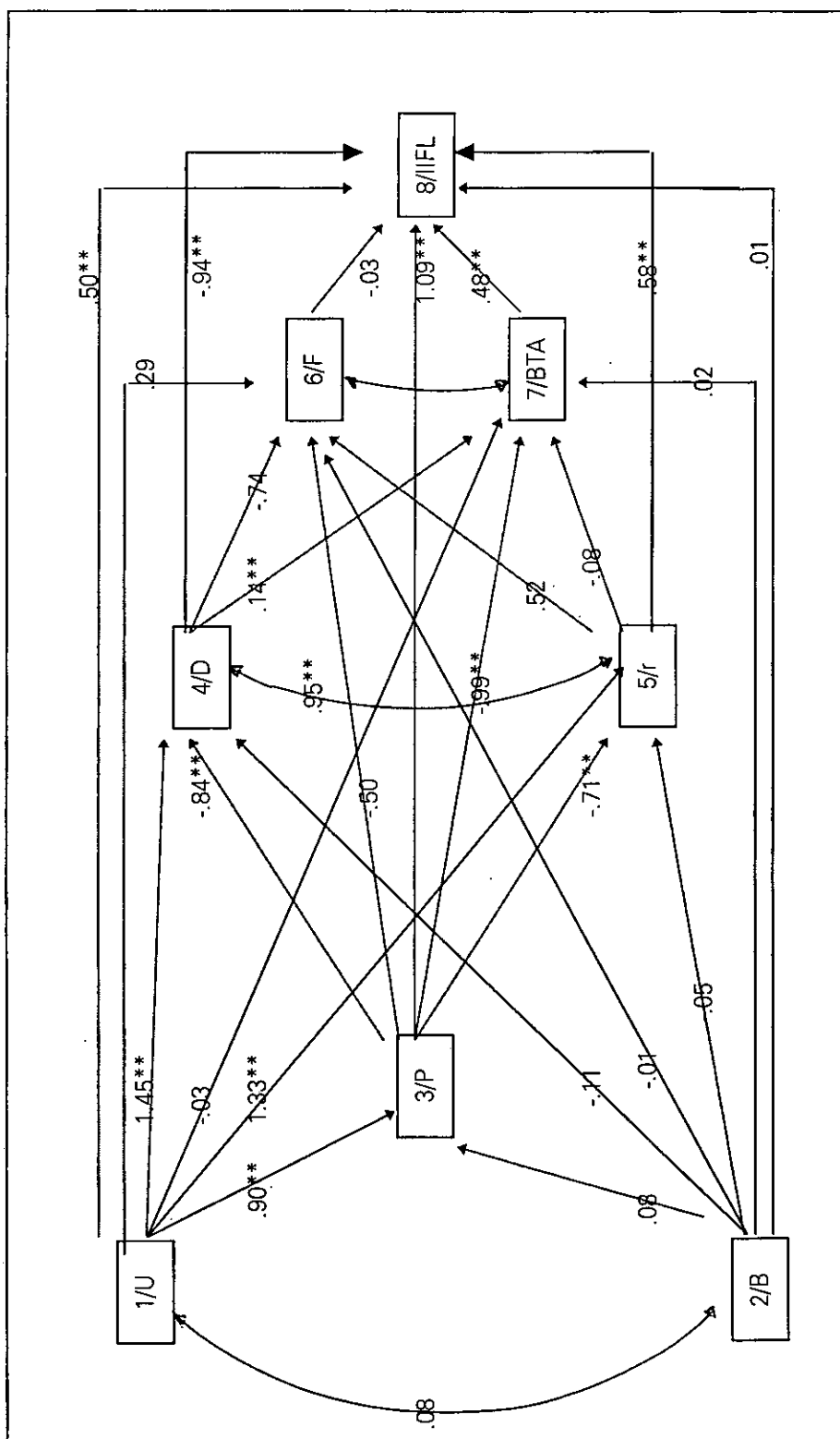
ภาพประกอบ 49 แผนภาพเส้นทางเต็มรูป แสดงความสัมพันธ์เชิงสาเหตุ ของตัวแปรต่างๆ กับค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูง
ของแบบทดสอบวัดความสามารถในการแก้ปัญหาทางคณิตศาสตร์ที่ให้คะแนนตามรูปแบบปกติ ตามกรอบทฤษฎี CTT และ IRT



* มีนัยสำคัญทางสถิติ ณ ระดับ .01

** มีนัยสำคัญทางสถิติ ณ ระดับ .05

ภาพประกอบ 50 แผนภาพเส้นทางเต็มรูป แสดงความสัมพันธ์สัมพัทธ์เชิงสาเหตุ ของตัวแปรต่างๆ กับค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลาง ของแบบทดสอบวัดความสามารถในการแก้ปัญหาทางคณิตศาสตร์ที่ให้คะแนนตามรูปแบบ ปกติ ตามกรอบทฤษฎี CTT และ IRT



* มีนัยสำคัญทางสถิติ ณ ระดับ .05

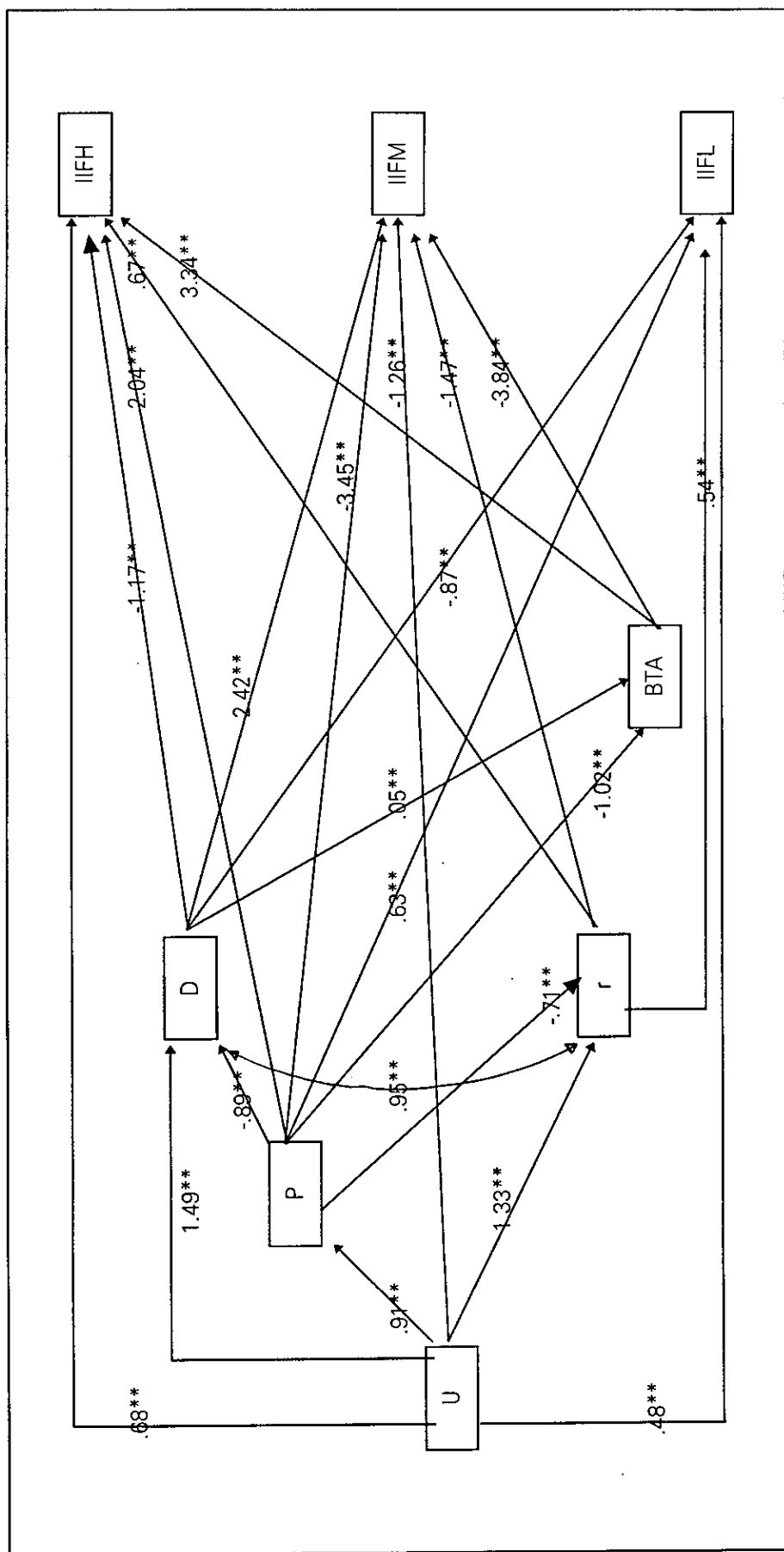
** มีนัยสำคัญทางสถิติ ณ ระดับ .01

ภาพประกอบ 51 แผนภาพเส้นทางเต็มรูป แสดงความสัมพันธ์เชิงสาเหตุ ของตัวแปรต่างๆ กับค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำ ของแบบทดสอบวัดความสามารถในการแก้ปัญหาทางคณิตศาสตร์ที่ให้คะแนนแบบปกติ ตามกรอบทฤษฎี CTT และ IRT

จากภาพประกอบ 49 พบว่า สัมประสิทธิ์พยากรณ์ของตัวแปรต่างๆในแผนภาพเส้นทาง แสดงความสัมพันธ์เชิงสาเหตุกับค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูง มีค่าเท่ากับ 0.9999 และค่าประสิทธิภาพในการพยากรณ์เท่ากับ 99.90 % และพบว่าค่าสัมประสิทธิ์เส้นทางที่มี นัยสำคัญทางสถิติ ณ ระดับ .01 จำนวน 12 ค่า คือ p_{31} , p_{41} , p_{51} , p_{81} , p_{43} , p_{53} , p_{73} , p_{83} , p_{54} , p_{74} , p_{84} และ p_{87} ซึ่งมีค่าเท่ากับ .90 , 1.45 , 1.33 , -.69 , -.84 , -.71 , -.99 , 2.00 , .95 , .14 , -1.17 , และ 3.33 ตามลำดับ และค่าสัมประสิทธิ์เส้นทางที่มีนัยสำคัญทางสถิติ ณ ระดับ .05 คือ p_{85} ซึ่งมี ค่าเท่ากับ .66

จากภาพประกอบ 50 พบว่า สัมประสิทธิ์พยากรณ์ของตัวแปรต่างๆในแผนภาพเส้นทาง แสดงความสัมพันธ์เชิงสาเหตุกับค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลาง มีค่า เท่ากับ 0.9999 และค่าประสิทธิภาพในการพยากรณ์เท่ากับ 99.68 % และพบว่าค่าสัมประสิทธิ์เส้น ทางที่มีนัยสำคัญทางสถิติ ณ ระดับ .01 จำนวน 12 ค่า คือ p_{31} , p_{41} , p_{51} , p_{81} , p_{43} , p_{53} , p_{73} , p_{83} , p_{54} , p_{74} , p_{84} และ p_{87} ซึ่งมีค่าเท่ากับ .90 , 1.45 , 1.33 , -1.27 , -.84 , -.71 , -.99 , -3.28 , .95 , .14 , 2.36 , และ 3.70 ตามลำดับ

จากภาพประกอบ 51 พบว่า สัมประสิทธิ์พยากรณ์ของตัวแปรต่างๆในแผนภาพเส้นทาง แสดงความสัมพันธ์เชิงสาเหตุกับค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำ มีค่าเท่ากับ 0.9999 และค่าประสิทธิภาพในการพยากรณ์เท่ากับ 99.94 % และพบว่าค่าสัมประสิทธิ์เส้นทางที่มี นัยสำคัญทางสถิติ ณ ระดับ .01 จำนวน 13 ค่า คือ p_{31} , p_{41} , p_{51} , p_{81} , p_{43} , p_{53} , p_{73} , p_{83} , p_{54} , p_{74} , p_{84} , p_{85} และ p_{87} ซึ่งมีค่าเท่ากับ .90 , 1.45 , 1.33 , .50 , -.84 , -.71 , -.99 , 1.09 , .95 , .14 , -.94 , -.58 และ .48 ตามลำดับ



ภาพประกอบ 52 แผนภาพเส้นทางเชิงประจักษ์ แสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่างๆ กับสารสนเทศของข้อสอบที่ระดับความสามารถสูง ปานกลาง และต่ำของแบบทดสอบวัดความสามารถในการแก้ปัญหาทางคณิตศาสตร์ที่ได้คะแนนตามรูปแบบปกติ ตามกรอบทฤษฎี CTT และ IRT

จากภาพประกอบ 52 พบว่า ความยากแบบ IRT ส่งอิทธิพลทางตรงเชิงบวกต่อค่าสารสนเทศข้อสอบที่ระดับความสามารถสูงของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนแบบปกติ ซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ 3.34 รองลงมา คือ ความยากแบบ CTT ความเป็นมิติเดียว และสัมประสิทธิ์การจำแนก ซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ 2.04 , .68 และ .67 ตามลำดับ ส่วนดัชนีการจำแนก พบว่า ส่งอิทธิพลทางตรงเชิงลบ ซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ -1.17

สำหรับตัวแปรที่ส่งอิทธิพลทางตรงเชิงบวกต่อค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลางคือ ดัชนีการจำแนก ซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ 2.42 ส่วนตัวแปรที่ส่งอิทธิพลทางตรงเชิงลบ ได้แก่ ค่าความยากแบบ IRT ความยากแบบ CTT สัมประสิทธิ์การจำแนก และความเป็นมิติเดียว ซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ -3.84 , -3.45 , -1.47 และ -1.26 ตามลำดับ

และพบว่าตัวแปรที่ส่งอิทธิพลทางตรงเชิงบวกต่อค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำคือ ความยากแบบ CTT สัมประสิทธิ์การจำแนก และความเป็นมิติเดียว ซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ .63 , .54 และ .48 ตามลำดับ ส่วนดัชนีการจำแนก พบว่า ส่งอิทธิพลทางตรงเชิงลบซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ -.87

เมื่อพิจารณาอิทธิพลทางอ้อมของตัวแปรพบว่า ความเป็นมิติเดียวส่งอิทธิพลเชิงบวกต่อสารสนเทศของข้อสอบที่ระดับความสามารถ สูง ปานกลาง และต่ำ ผ่านทางดัชนีการจำแนก สัมประสิทธิ์การจำแนก และความยากแบบ CTT โดยมีค่าสัมประสิทธิ์เส้นทางเท่ากับ 1.49 , 1.33 , และ .91 ตามลำดับ สำหรับความยากแบบ CTT ส่งอิทธิพลทางอ้อมเชิงลบผ่านทางความยากแบบ IRT ดัชนีการจำแนก และสัมประสิทธิ์การจำแนก ซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ -1.02 , -.89 และ -.71 ตามลำดับ และพบความสัมพันธ์ระหว่างสัมประสิทธิ์การจำแนก และดัชนีการจำแนกเท่ากับ .92 ซึ่งความสัมพันธ์นี้ไม่สามารถระบุทิศทางเชิงสาเหตุได้

ตาราง 57 เปรียบเทียบความสอดคล้องของแผนภาพเส้นทางเชิงประจักษ์กับแผนภาพเส้นทาง
เต็มรูปของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนน
ตามรูปแบบปกติ ตามกรอบทฤษฎี CTT และ IRT

	แผนภาพเต็มรูป		แผนภาพเชิงประจักษ์		W
	R^2_{FM}	E	R^2_{OM}	E	
IIFH	0.9999	99.90	0.9999	99.90	0**
IIFM	0.9999	99.68	0.9999	99.68	0**
IIFL	0.9999	99.94	0.9999	99.94	0**

* มีนัยสำคัญทางสถิติ ณ ระดับ .05 ** มีนัยสำคัญทางสถิติ ณ ระดับ .01

จากตาราง 57 พบว่าแผนภาพเส้นทางเชิงประจักษ์สอดคล้องกับแผนภาพเส้นทางเต็มรูป
อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ในทุกระดับความสามารถ

ตาราง 58 ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมที่มีต่อค่า
สารสนเทศที่ระดับความสามารถสูง ของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหา
ทางคณิตศาสตร์ที่ได้จากการให้คะแนนตามรูปแบบปกติ ภายใต้กรอบทฤษฎี CTT และ IRT

ตัวแปรสาเหตุ	อิทธิพลทางตรง	อิทธิพลทางอ้อม	อิทธิพลปลอม	รวม
BTA	3.34	-	-	3.34
P	2.04	-2.96	-	-.92
U	.68	-1.49	-	-.81
r	.67	-.13	-	.54
D	-1.17	.24	-	-.93

*

จากตาราง 58 พบว่า ความยากแบบ IRT ส่งอิทธิพลเชิงบวกต่อค่าสารสนเทศของข้อ
สอบที่ระดับความสามารถสูงของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิต
ศาสตร์ที่ให้คะแนนแบบปกติมากที่สุดเท่ากับ 3.34 รองลงมาคือค่าสัมประสิทธิ์การจำแนกมี

ค่าเท่ากับ .54 และพบว่าตัวแปรที่มีอิทธิพลเชิงลบได้แก่ ดัชนีการจำแนก ความยากแบบ CTT และความเป็นมิติเดียว ซึ่งมีค่าเท่ากับ -.93 , -.92 และ -0.81 ตามลำดับ

ตาราง 59 ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมที่มีต่อค่า
 สารสนเทศที่ระดับความสามารถปานกลาง ของแบบทดสอบวัดความสามารถในการแก้โจทย์
 ปัญหาทางคณิตศาสตร์ที่ได้จากการให้คะแนนตามรูปแบบปกติ ภายใต้กรอบทฤษฎี CTT และ
 IRT

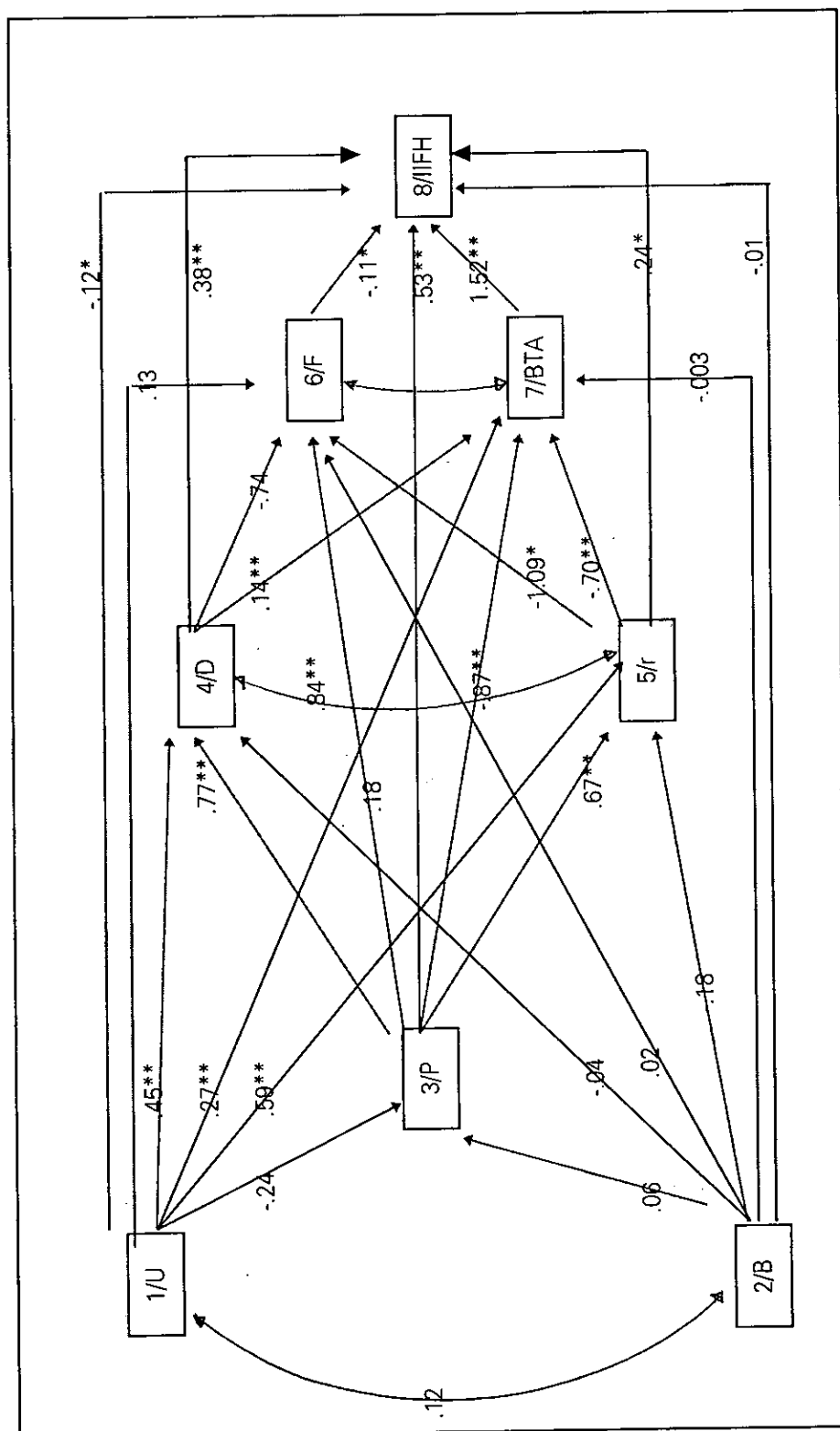
ตัวแปรสาเหตุ	อิทธิพลทางตรง	อิทธิพลทางอ้อม	อิทธิพลปลอม	รวม
D	2.42	-.16	-	2.26
r	-1.47	.27	-	-1.20
BTA	-3.84	-	-	-3.84
P	-3.45	2.75	-	-.70
U	-1.26	1.15	-	-.11

จากตาราง 59 พบว่า ดัชนีการจำแนก ส่งอิทธิพลเชิงบวกต่อค่าสารสนเทศของข้อสอบ
 ที่ระดับความสามารถปานกลางของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิต
 ศาสตร์ที่ให้คะแนนแบบปกติเท่ากับ 2.25 ส่วนตัวแปรที่มีอิทธิพลเชิงลบได้แก่ ความยากแบบ IRT
 สัมประสิทธิ์การจำแนก ความยากแบบ CTT และความเป็นมิติเดียว ซึ่งมีค่าเท่ากับ -3.84 ,
 -1.20 , -.70 และ -.11 ตามลำดับ

ตาราง 60 ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมที่มีต่อค่า
 สารสนเทศที่ระดับความสามารถต่ำของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทาง
 คณิตศาสตร์ที่ได้จากการให้คะแนนตามรูปแบบปกติ ตามกรอบทฤษฎี CTT และ IRT

ตัวแปรสาเหตุ	อิทธิพลทางตรง	อิทธิพลทางอ้อม	อิทธิพลปลอม	รวม
P	.63	.40	-	1.03
U	.48	.33	-	.81
r	.54	-.09	-	.45
D	-.87	.06	-	-.81

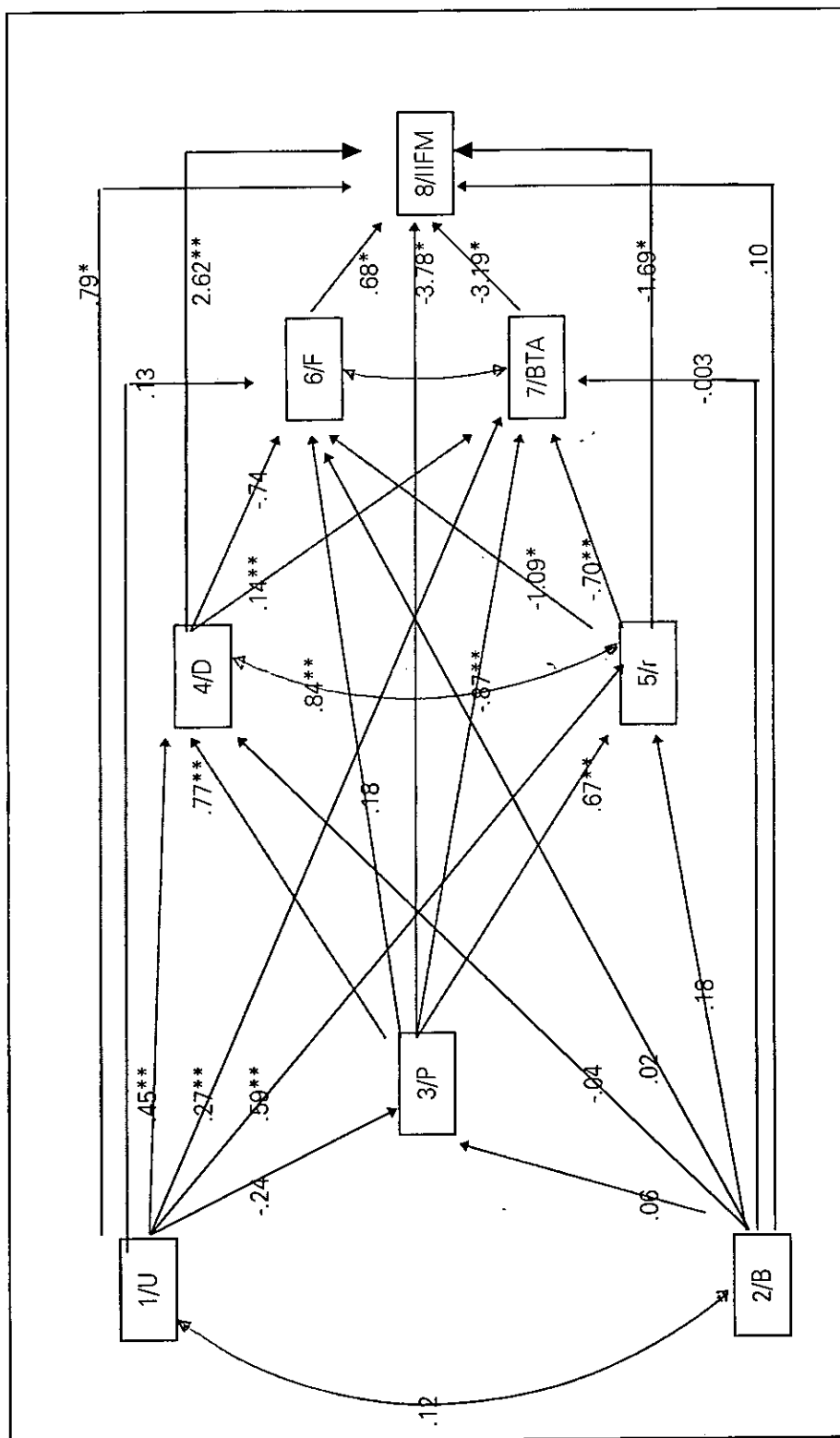
จากตาราง 60 พบว่า ความยากแบบ CTT ส่งอิทธิพลเชิงบวกต่อค่าสารสนเทศของข้อ
 สอบที่ระดับความสามารถต่ำของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทาง
 คณิตศาสตร์ที่ให้คะแนนแบบปกติมากที่สุดเท่ากับ 1.03 รองลงมาคือ ความเป็นมิติเดียว และ
 สัมประสิทธิ์การจำแนกมีค่าเท่ากับ .81 และ .45 ตามลำดับ และพบว่าตัวแปรที่มีอิทธิพลเชิงลบคือ
 ดัชนีการจำแนก มีค่าเท่ากับ -.81



* มีนัยสำคัญทางสถิติ ณ ระดับ .05

** มีนัยสำคัญทางสถิติ ณ ระดับ .01

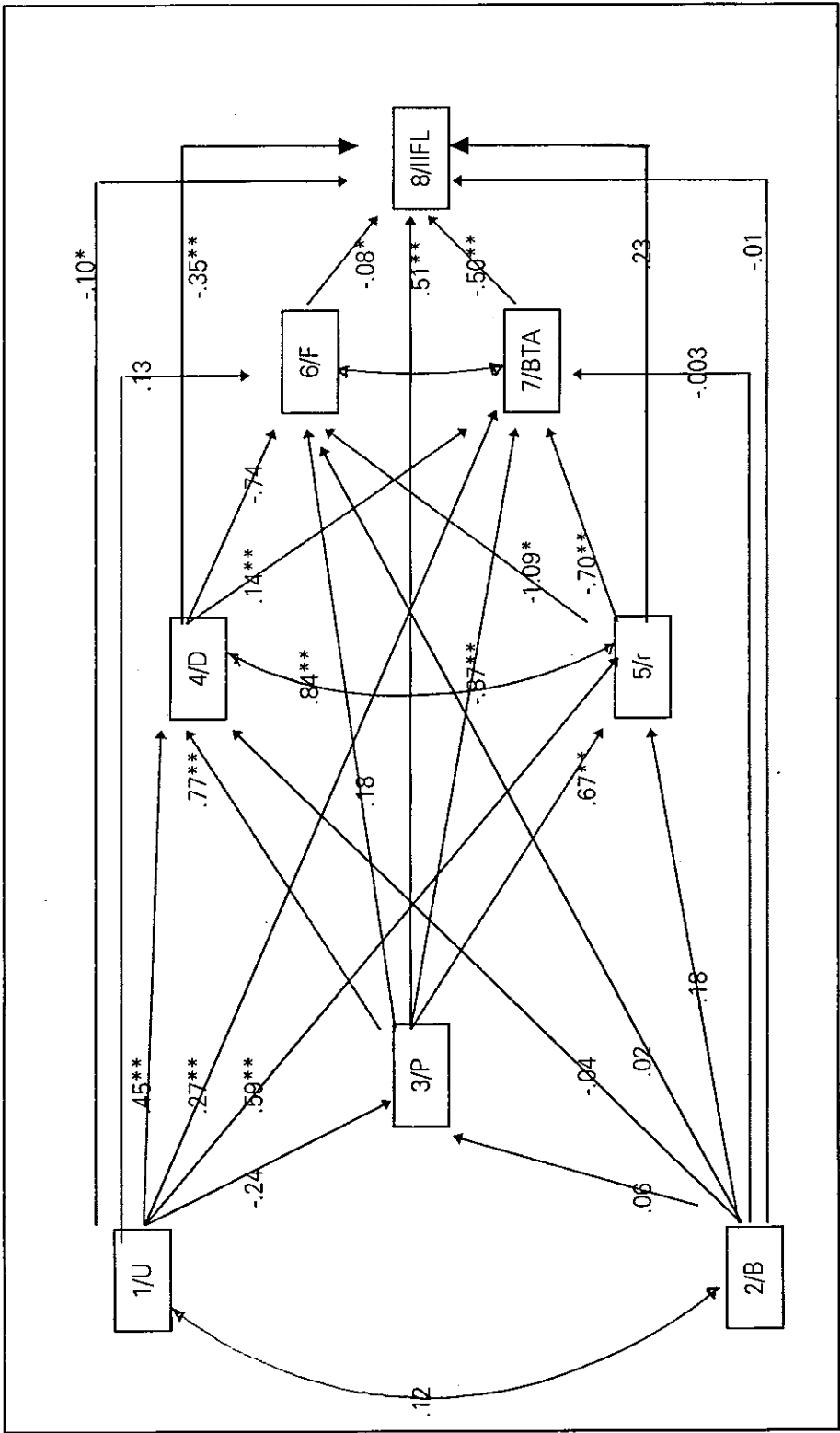
ภาพประกอบ 53 แผนภาพเส้นทางเต็มรูป แสดงความสัมพันธ์เชิงสาเหตุ ของตัวแปรต่างๆ กับค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูง ของแบบทดสอบวัดความสามารถในการแก้ปัญหาทางคณิตศาสตร์ที่ได้คะแนนตามรูปแบบ 2-SR ตามกรอบทฤษฎี CTT และ IRT



* มีนัยสำคัญทางสถิติ ณ ระดับ .05

** มีนัยสำคัญทางสถิติ ณ ระดับ .01

ภาพประกอบ 54 แผนภาพเส้นทางเต็มรูป แสดงความสัมพันธ์เชิงสาเหตุ ของตัวแปรต่างๆ กับค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลาง ของแบบทดสอบวัดความสามารถในการแก้ปัญหาทางคณิตศาสตร์ที่ให้คะแนนตามรูปแบบ 2-SR ตามกรอบทฤษฎี CTT และ IRT



* มีนัยสำคัญทางสถิติ ณ ระดับ .05

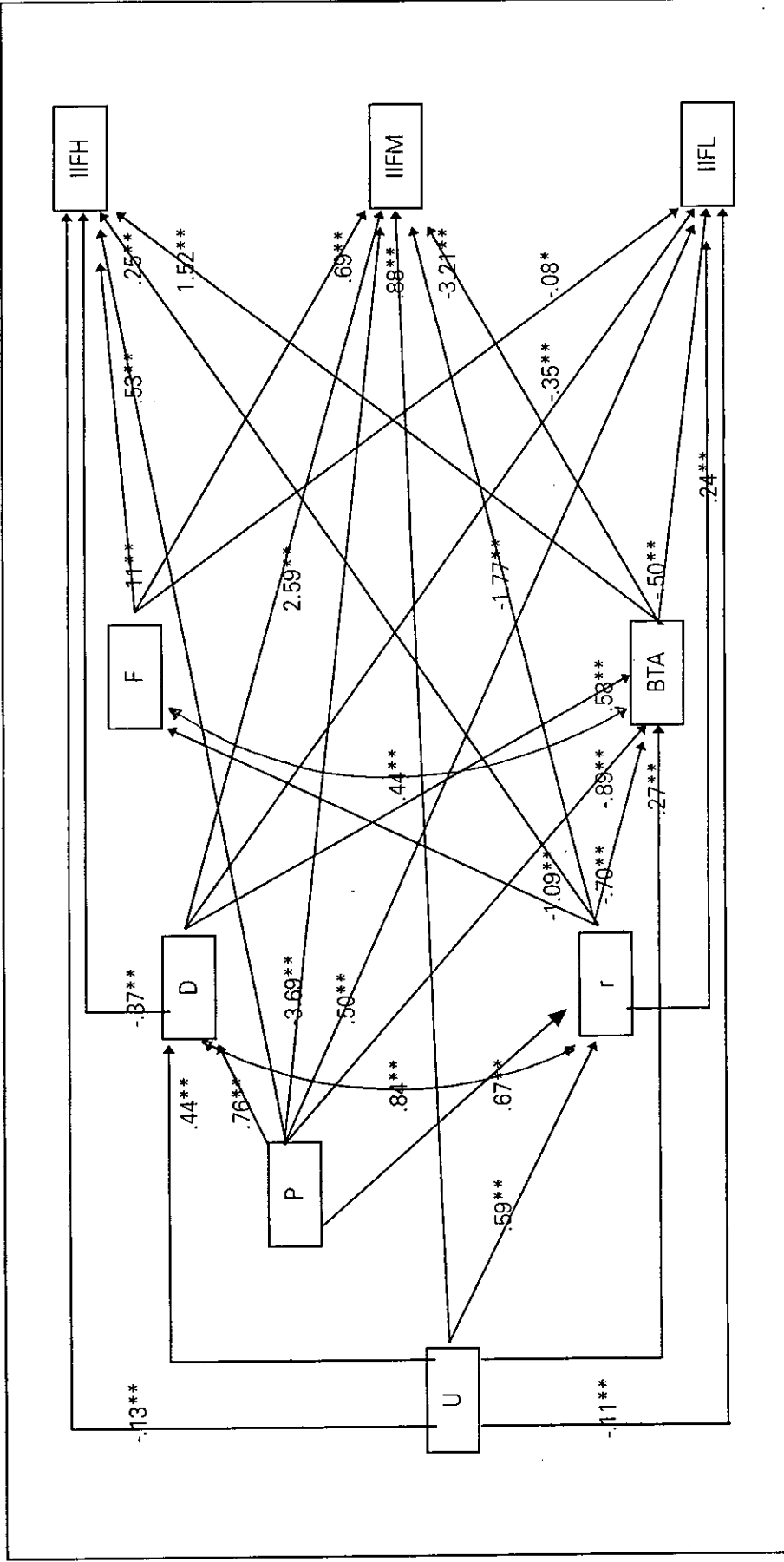
** มีนัยสำคัญทางสถิติ ณ ระดับ .01

ภาพประกอบ 55 แผนภาพเส้นทางเต็มรูป แสดงความสัมพันธ์เชิงสาเหตุ ของตัวแปรต่างๆ กับค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำ
ของแบบทดสอบวัดความสามารถในการแก้ปัญหาทางคณิตศาสตร์ที่ให้คะแนนตามรูปแบบ 2-SR ตามกรอบทฤษฎี CTT และ IRT

จากภาพประกอบ 53 พบว่า สัมประสิทธิ์พยากรณ์ของตัวแปรต่างๆในแผนภาพเส้นทาง แสดงความสัมพันธ์เชิงสาเหตุกับค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูง มีค่าเท่ากับ 0.9999 และค่าประสิทธิภาพในการพยากรณ์เท่ากับ 99.90 % และพบว่าค่าสัมประสิทธิ์เส้นทางที่มีนัยสำคัญทางสถิติ ณ ระดับ .01 จำนวน 13 ค่า คือ p_{41} , p_{51} , p_{71} , p_{43} , p_{53} , p_{73} , p_{83} , p_{54} , p_{74} , p_{84} , p_{65} , p_{75} และ p_{87} ซึ่งมีค่าเท่ากับ .45 , .59 , .27 , .77 , .67 , -.87 , .84 , .14 , .38 , .53, -1.09 , -.70 , และ .1.52 ตามลำดับ และค่าสัมประสิทธิ์เส้นทางที่มีนัยสำคัญทางสถิติ ณ ระดับ .05 คือ p_{81} , p_{85} และ p_{85} ซึ่งมีค่าเท่ากับ -.12 , .24 และ .66 ตามลำดับ

จากภาพประกอบ 54 พบว่าสัมประสิทธิ์พยากรณ์ของตัวแปรต่างๆในแผนภาพเส้นทาง แสดงความสัมพันธ์เชิงสาเหตุกับค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลาง มีค่าเท่ากับ .9999 และค่าประสิทธิภาพในการพยากรณ์เท่ากับ 99.00 % และพบว่าค่าสัมประสิทธิ์เส้นทางที่มีนัยสำคัญทางสถิติ ณ ระดับ .01 จำนวน 14 ค่า คือ p_{41} , p_{51} , p_{71} , p_{43} , p_{53} , p_{73} , p_{83} , p_{54} , p_{74} , p_{84} , p_{65} , p_{75} , p_{85} และ p_{87} ซึ่งมีค่าเท่ากับ .45 , .59 , .27 , .77 , .67 , -.87 , -.3.78 , .84 , .14 , 2.62 , -1.09 , -.70 , -1.69 และ -3.19 ตามลำดับ และค่าสัมประสิทธิ์เส้นทางที่มีนัยสำคัญทางสถิติ ณ ระดับ .05 คือ p_{86} ซึ่งมีค่าเท่ากับ .68

จากภาพประกอบ 55 พบว่า สัมประสิทธิ์พยากรณ์ของตัวแปรต่างๆในแผนภาพเส้นทาง แสดงความสัมพันธ์เชิงสาเหตุกับค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำ มีค่าเท่ากับ .9999 และค่าประสิทธิภาพในการพยากรณ์เท่ากับ 99.90 % และพบว่าค่าสัมประสิทธิ์เส้นทางที่มีนัยสำคัญทางสถิติ ณ ระดับ .01 จำนวน 13 ค่า คือ p_{41} , p_{51} , p_{71} , p_{43} , p_{53} , p_{73} , p_{83} , p_{54} , p_{74} , p_{84} , p_{65} , p_{75} และ p_{87} ซึ่งมีค่าเท่ากับ .45 , .59 , .27 , .77 , .67 , -.87 , .51 , .84 , .14 , -.35 , -1.09 , -.70 , และ -.50 ตามลำดับ และค่าสัมประสิทธิ์เส้นทางที่มีนัยสำคัญทางสถิติ ณ ระดับ .05 คือ p_{81} และ p_{85} ซึ่งมีค่าเท่ากับ -.10 และ -.08 ตามลำดับ



ภาพประกอบ 56 แผนภาพเส้นทางเชิงประจักษ์ แสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่างๆ กับสารสนเทศของข้อสอบที่ระดับความสามารถสูง ปานกลาง และต่ำของแบบทดสอบวัดความสามารถในการแก้ปัญหาทางคณิตศาสตร์ที่ให้คะแนนตามรูปแบบ 2-SR ตามกรอบทฤษฎี CTT และ IRT

จากภาพประกอบ 56 พบว่าตัวแปรที่ส่งอิทธิพลทางตรงเชิงบวกต่อค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูงของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนแบบ 2 - SR คือความยากแบบ IRT ความยากแบบ CTT และสัมประสิทธิ์การจำแนก ซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ 1.52 , .53 และ .25 ตามลำดับ ส่วนตัวแปรที่ส่งอิทธิพลทางตรงเชิงลบได้แก่ ดัชนีการจำแนก ความเป็นมิติเดียว และความสอดคล้องของข้อสอบกับโมเดล ซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ -.37 , -.13 และ -.11 ตามลำดับ

สำหรับตัวแปรที่ส่งอิทธิพลทางตรงเชิงบวกต่อค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลาง คือดัชนีการจำแนก ความเป็นมิติเดียว และ ความเหมาะสมของข้อสอบกับโมเดล ซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ 2.59 , .88 , และ .69 ตามลำดับ ส่วนตัวแปรที่ส่งอิทธิพลทางตรงเชิงลบคือ ความยากแบบ CTT ความยากแบบ IRT และสัมประสิทธิ์การจำแนกซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ -3.69 , -3.21 และ -1.77 ตามลำดับ และพบว่าตัวแปรที่ส่งอิทธิพลทางตรงเชิงบวกต่อสารสนเทศของข้อสอบที่ระดับความสามารถต่ำคือ ความยากแบบ CTT และสัมประสิทธิ์การจำแนกซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ .50 และ .24 ตามลำดับ ส่วนตัวแปรที่ส่งอิทธิพลทางตรงเชิงลบคือ ความยากแบบ IRT ดัชนีการจำแนก ความเป็นมิติเดียว และความเหมาะสมของข้อสอบกับโมเดล ซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ -.50 , -.35 , -.11 และ -.08 ตามลำดับ

และเมื่อพิจารณาอิทธิพลทางอ้อม พบว่า ความเป็นมิติเดียวส่งอิทธิพลทางอ้อมเชิงบวกต่อค่าสารสนเทศของข้อสอบที่ระดับต่าง ๆ ผ่านทางสัมประสิทธิ์การจำแนก ดัชนีการจำแนก และความยากแบบ IRT ซึ่งมีสัมประสิทธิ์เส้นทางเท่ากับ .59 , .44 และ .27 ตามลำดับ ส่วนความยากแบบ CTT พบว่าส่งอิทธิพลทางอ้อมเชิงบวกต่อค่าสารสนเทศของข้อสอบที่ระดับต่าง ๆ ผ่านทางดัชนีการจำแนก และสัมประสิทธิ์การจำแนก ซึ่งมีสัมประสิทธิ์เส้นทางเท่ากับ .76 และ .67 ตามลำดับ และส่งอิทธิพลเชิงลบผ่านทางความยากแบบ IRT ซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ -.89 สำหรับสัมประสิทธิ์การจำแนกพบว่า ส่งอิทธิพลเชิงลบผ่านทางความเหมาะสมของข้อสอบกับโมเดล และความยากแบบ IRT ซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ -1.09 และ -.70 ตามลำดับ -.89 และ -.70 ตามลำดับ และพบว่า ดัชนีการจำแนกส่งอิทธิพลเชิงบวกต่อสารสนเทศที่ระดับความสามารถต่าง ๆ ผ่านทางความยากแบบ IRT ซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ .58

ตาราง 61 เปรียบเทียบความสอดคล้องของแผนภาพเส้นทางเชิงประจักษ์กับแผนภาพเส้นทาง
 เต็มรูปของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ ที่ให้คะแนน
 ตามรูปแบบ 2-SR ตามกรอบทฤษฎี CTT และ IRT

	แผนภาพเต็มรูป		แผนภาพเชิงประจักษ์		W
	R^2_{FM}	E	R^2_{OM}	E	
IIFH	0.9999	99.90	0.9999	99.90	0**
IIFM	0.9999	99.68	0.9999	99.68	0**
IIFL	0.9999	99.94	0.9999	99.94	0**

* มีนัยสำคัญทางสถิติ ณ ระดับ .05 ** มีนัยสำคัญทางสถิติ ณ ระดับ .01

จากตาราง 61 พบว่าแผนภาพเส้นทางเชิงประจักษ์สอดคล้องกับแผนภาพเส้นทางเต็มรูป
 อย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01 ในทุกระดับความสามารถ

ตาราง 62 ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมที่มีต่อค่า
 สารสนเทศที่ระดับความสามารถสูงของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทาง
 คณิตศาสตร์ที่ได้จากการให้คะแนนตามรูปแบบ 2-SR ภายใต้กรอบทฤษฎี CTT และ IRT

ตัวแปรสาเหตุ	อิทธิพลทางตรง	อิทธิพลทางอ้อม	อิทธิพลปลอม	รวม
BTA	1.52	.05	–	1.57
P	.53	-2.58	–	-2.04
r	.25	-1.48	–	-1.23
D	-.37	-.10	–	-.48
U	-.13	-.52	–	-.65
F	-.11	.68	–	.571

จากตาราง 62 พบว่า ความยากแบบ IRT ส่งอิทธิพลเชิงบวกต่อสารสนเทศข้อสอบที่
 ระดับความสามารถสูงของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาคณิตศาสตร์ที่ให้
 คะแนนตามรูปแบบ 2 - SR มากที่สุดโดยมีค่าเท่ากับ 1.57 รองลงมาคือ ความเหมาะสมของ

ข้อสอบกับโมเดลซึ่งมีค่าเท่ากับ .57 ส่วนตัวแปรที่ส่งอิทธิพลเชิงลบคือ ความยากแบบ CTT สัมประสิทธิ์การจำแนก ความเป็นมิติเดียวและดัชนีการจำแนก ซึ่งมีค่าเท่ากับ -2.04 , -1.23 , -.65 และ, -.48 ตามลำดับ

ตาราง 63 ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมที่มีต่อค่าสารสนเทศที่ระดับความสามารถปานกลางของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ได้จากการให้คะแนนโดยวิธีแบบ 2 - SR ภายใต้กรอบทฤษฎี CTT และ IRT

ตัวแปรสาเหตุ	อิทธิพลทางตรง	อิทธิพลทางอ้อม	อิทธิพลปลอม	รวม
D	2.59	-1.10	–	1.49
U	.88	.80	–	1.68
F	.69	-1.44	–	-.75
P	-3.67	4.74	–	1.07
BTA	-3.21	.31	–	-2.89
r	-1.77	3.33	–	1.56

จากตาราง 63 พบว่า ตัวแปรที่ส่งอิทธิพลเชิงบวกต่อค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลางของแบบทดสอบวัดความสามารถทางคณิตศาสตร์ที่ให้คะแนนแบบ 2 - SR คือ ความเป็นมิติเดียว สัมประสิทธิ์การจำแนก ดัชนีการจำแนก และความยากแบบ CTT ซึ่งมีค่าสัมประสิทธิ์เส้นทางเท่ากับ 1.68 , 1.56 , 1.49 และ 1.07 ตามลำดับ ส่วนตัวแปรที่ส่งอิทธิพลเชิงลบ ได้แก่ ค่าความยากแบบ IRT และความเหมาะสมของข้อสอบกับโมเดล ซึ่งมีค่าเท่ากับ -2.89 และ -.75 ตามลำดับ

ตาราง 64 ค่าสัมประสิทธิ์เส้นทางที่มีอิทธิพลทางตรง ทางอ้อม และอิทธิพลปลอมที่มีต่อค่า
 สารสนเทศที่ระดับความสามารถต่ำของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทาง
 คณิตศาสตร์ที่ได้จากการให้คะแนนโดยวิธีแบบ 2-SR ตามกรอบทฤษฎี CTT และ IRT

ตัวแปรสาเหตุ	อิทธิพลทางตรง	อิทธิพลทางอ้อม	อิทธิพลปลอม	รวม
P	.50	.57	–	1.07
r	.24	.16	–	.40
BTA	-.50	-.04	–	-.54
D	-.35	.10	–	-.25
U	-.11	-.02	–	-.14
F	-.08	-.05	–	-.13

จากตาราง 64 พบว่า ความยากแบบ CTT ส่งอิทธิพลเชิงบวกต่อค่าสารสนเทศของข้อสอบ
 ที่ระดับความสามารถต่ำของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์
 ที่ให้คะแนนโดยวิธีแบบ 2-SR มากที่สุด คือ เท่ากับ 1.06 รองลงมาได้แก่สัมประสิทธิ์การจำแนกมี
 ค่าเท่ากับ .40 ส่วนตัวแปรที่มีอิทธิพลเชิงลบ ได้แก่ ความยากแบบ IRT ดัชนีการจำแนก ความเป็น
 มิติเดียว และความเหมาะสมของข้อสอบกับโมเดลมีค่าเท่ากับ -.55 , -.25 , -.14 และ -.13 ตาม
 ลำดับ

บทที่ 5

สรุป อภิปรายผลและข้อเสนอแนะ

การศึกษาวิจัยครั้งนี้มีวัตถุประสงค์สำคัญเพื่อศึกษาอิทธิพลของการให้คะแนน และการคัดเลือกข้อสอบเพื่อกำหนดรูปแบบสมการโครงสร้างที่เหมาะสมที่สุดในการพัฒนาแบบทดสอบเลือกตอบ โดยได้ดำเนินการวิจัยดังนี้

วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาอิทธิพลของรูปแบบการให้คะแนน และวิธีการคัดเลือกข้อสอบ ที่มีต่อคุณภาพของแบบทดสอบเลือกตอบภายใต้กรอบทฤษฎี CTT
2. เพื่อศึกษาอิทธิพลของรูปแบบการให้คะแนน และวิธีการคัดเลือกข้อสอบ ที่มีต่อคุณภาพของแบบทดสอบเลือกตอบภายใต้กรอบทฤษฎี IRT (ราชคโหมเดล)
3. เพื่อศึกษาอิทธิพลของรูปแบบการให้คะแนน วิธีการคัดเลือกข้อสอบและคุณภาพข้อสอบตามกรอบทฤษฎีของ CTT ที่มีต่อคุณภาพของแบบทดสอบเลือกตอบตามแนวของ IRT (ราชคโหมเดล)
4. เพื่อศึกษารูปแบบความสัมพันธ์เชิงสาเหตุของการให้คะแนน และวิธีการคัดเลือกข้อสอบ ที่มีต่อคุณภาพของแบบทดสอบตามกรอบทฤษฎีของ CTT และ IRT (ราชคโหมเดล)

ขอบเขตของการวิจัย

1. ประชากรในการวิจัยได้แก่ นักเรียนระดับชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2539 สังกัดสำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ จำนวน 794,740 คน
2. กลุ่มตัวอย่างได้จากการเลือกจังหวัดเป้าหมายโดยวิธี PPS และสุ่มตัวอย่างหลายขั้นตอน จาก 21 โรงเรียนใน 12 เขตการศึกษา 12 จังหวัด รวมทั้งกรุงเทพมหานคร จำนวน 1,304 คน เป็นชาย 664 คน และหญิง 640 คน
3. เครื่องมือในการวิจัย ได้แก่แบบทดสอบชนิดเลือกตอบ 4 ตัวเลือกจำนวน 2 ฉบับคือแบบทดสอบวัดความสามารถในการอ่านจับใจความ และแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ แต่ละฉบับมีความยาว 40 ข้อ และมีความเชื่อมั่นเท่ากับ .66 และ .77 ตามลำดับ

4. ตัวแปรในการวิจัย ได้แก่

4.1 ตัวแปรอิสระ/ตัวแปรพยากรณ์ คือ

4.1.1 รูปแบบการให้คะแนน (แบบปกติ , แบบ 2-SR)

4.1.2 วิธีการคัดเลือกข้อสอบ โดยใช้ค่าพารามิเตอร์ และคุณสมบัติบางประการของข้อสอบ ได้แก่ ค่าความยาก ค่าสัมประสิทธิ์การจำแนก ค่าดัชนีการจำแนก ความเป็นมิติเดียว ความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ ความยากแบบ IRT และความเหมาะสมของข้อสอบกับโมเดล

4.2 ตัวแปรตาม/ตัวแปรเกณฑ์ ได้แก่ ค่าความยาก ค่าสัมประสิทธิ์การจำแนก ค่าดัชนีการจำแนก ความเป็นมิติเดียว (น้ำหนักองค์ประกอบ) ความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ (ค่าสัมประสิทธิ์สหสัมพันธ์แบบ Contingency) ความยากแบบ IRT และความเหมาะสมของข้อสอบกับโมเดล ความเชื่อมั่นของข้อสอบ และแบบทดสอบ สารสนเทศของข้อสอบและแบบทดสอบ คะแนนผลการสอบ และระดับความสามารถของผู้สอบ

วิธีดำเนินการวิจัย

ขั้นตอนที่ 1 สร้างเครื่องมือ ดำเนินงานดังนี้

1. วิเคราะห์หลักสูตร
2. กำหนดนิยามสมรรถภาพเชิงปฏิบัติการ
3. เขียนจุดประสงค์เชิงพฤติกรรม และกำหนดน้ำหนักความสำคัญ
4. เขียนข้อสอบตามจุดประสงค์เชิงพฤติกรรม
5. สร้างตารางสัมพันธ์ระหว่างข้อสอบกับจุดประสงค์เชิงพฤติกรรม
6. ตรวจสอบความเที่ยงตรงเชิงเนื้อหาและโครงสร้างด้วยดัชนี IOC
7. คัดเลือกข้อสอบเข้าฉบับ

ขั้นตอนที่ 2 การเก็บรวบรวมข้อมูล

1. นำแบบทดสอบไปสอบกับกลุ่มตัวอย่าง
2. วิเคราะห์ค่าสถิติ และคุณสมบัติบางประการของข้อสอบและแบบทดสอบด้วย

โปรแกรมคอมพิวเตอร์ คือ

2.1 โปรแกรม SPSSPC & SPSSWIN ในการประมาณค่าสถิติตามกรอบทฤษฎี CTT

ดังนี้

2.1.1 โปรแกรม FREQUENCIES , DESCRIPTIVES และ CROSSTABS ใช้ประมาณค่าสถิติพื้นฐาน (ค่าเฉลี่ย , ส่วนเบี่ยงเบนมาตรฐาน) ค่าความยาก ดัชนีการจำแนก และ ความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ

2.1.2 โปรแกรม RELIABILITY ใช้ประมาณค่าสัมประสิทธิ์การจำแนก และ ความเชื่อมั่นของแบบทดสอบแบบอัลฟา (α - coefficient)

2.1.3 โปรแกรม FACTOR เพื่อประมาณค่าน้ำหนักองค์ประกอบ และพิจารณาความเป็นมิติเดียวของข้อสอบ

2.2 โปรแกรม BIGSTEPS สำหรับประมาณค่าความยาก และความสามารถของผู้สอบ

3. วิเคราะห์ข้อมูลเชิงวิจัย ด้วยโปรแกรมคอมพิวเตอร์ได้แก่

3.1 โปรแกรม SPSSPC & SPSSWIN

3.1.1 โปรแกรม DESCRIPTIVES ประมาณค่าสถิติพื้นฐาน ได้แก่ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าสูงสุด-ต่ำสุด

3.1.2 โปรแกรม CORRELATION ศึกษาความสัมพันธ์ และความแปรปรวนร่วมระหว่างตัวแปรต่างๆ

3.1.3 โปรแกรม T - TEST PAIRS ใช้ทดสอบความแตกต่างของค่าเฉลี่ยของคะแนนสอบ และคะแนนความสามารถของผู้สอบที่ได้คะแนนโดยวิธีการตรวจให้คะแนนที่แตกต่างกัน

3.1.4 โปรแกรม NONPARAMETRIC (สถิติ Wilcoxon Match Pairs) ใช้ทดสอบความแตกต่างของค่าเฉลี่ยของสถิติต่างๆของข้อสอบ ที่ได้จากการให้คะแนนโดยวิธีที่แตกต่างกัน

3.1.5 โปรแกรม REGRESSION (STEPWISE) เพื่อศึกษาสมการพยากรณ์เชิงเส้นตรง ค่าสหสัมพันธ์พหุคูณ และสัมประสิทธิ์การพยากรณ์

3.1.6 โปรแกรม REGRESSION (ENTER) เพื่อกำหนด และตรวจสอบแผนภาพเส้นทาง (Path diagram) ที่แสดงความสัมพันธ์เชิงสาเหตุของตัวแปรต่าง ๆ และสัมประสิทธิ์เส้นทาง (Path coefficient)

สรุปผลการวิจัย

1. ผลการศึกษาอิทธิพลของการให้คะแนน

จากการศึกษาวิจัยพบว่าคะแนนสอบและระดับความสามารถในการอ่านจับใจความ และความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ของผู้สอบที่ได้จากการให้คะแนนตามรูปแบบปกติ และรูปแบบ 2 - SR มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .05 โดยพบว่าค่าเฉลี่ยของคะแนนสอบที่ได้จากการให้คะแนนแบบปกติสูงกว่าค่าเฉลี่ยของคะแนนสอบที่ได้จากการให้คะแนนแบบ 2 - SR ในแบบทดสอบทั้งสองฉบับ แต่เมื่อพิจารณาระดับความสามารถของผู้สอบพบว่า ค่าเฉลี่ยของระดับความสามารถที่ได้จากการให้คะแนนแบบปกติมากกว่าค่าเฉลี่ยที่ได้จากการให้คะแนนแบบ 2 - SR เฉพาะในแบบทดสอบวัดความสามารถในการอ่านจับใจความเท่านั้น ส่วนค่าเฉลี่ยของระดับความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ได้จากการให้คะแนนโดยวิธีแบบ 2 - SR สูงกว่าค่าเฉลี่ยของระดับความสามารถที่ได้จากการให้คะแนนแบบปกติ

สำหรับการทดสอบความแตกต่างของความเชื่อมั่นของแบบทดสอบพบว่าค่าความเชื่อมั่นของแบบทดสอบทั้งสองฉบับที่ได้จากการให้คะแนนแบบปกติ และแบบ 2 - SR ให้ผลที่ไม่แตกต่างกัน แต่เมื่อพิจารณาค่าสารสนเทศของแบบทดสอบพบว่าวิธีการให้คะแนนแบบปกติให้ค่าสารสนเทศสูงกว่าวิธีการให้คะแนนแบบ 2 - SR ที่ระดับความสามารถสูง ($\theta = 3.00$) และระดับความสามารถต่ำ ($\theta = -3.00$) ในแบบทดสอบทั้งสองฉบับ ส่วนที่ระดับความสามารถปานกลาง ($\theta = 0.00$) พบว่าค่าสารสนเทศของแบบทดสอบทั้งสองฉบับที่ได้จากการให้คะแนนแบบ 2 - SR สูงกว่าค่าสารสนเทศของแบบทดสอบที่ได้จากการให้คะแนนแบบปกติ

นอกจากนี้ยังพบว่าค่าเฉลี่ยของค่าสถิติของแบบทดสอบวัดความสามารถในการอ่านจับใจความและแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ คือ ค่าสัมประสิทธิ์การจำแนก ดัชนีการจำแนก ความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ ค่าความเชื่อมั่นรายข้อ ค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูง และปานกลางที่ได้จากการให้คะแนนด้วยวิธีที่ต่างกัน มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .05 โดยที่พบว่าค่าเฉลี่ยของค่าสัมประสิทธิ์การจำแนก ดัชนีการจำแนก ความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ ความเชื่อมั่นรายข้อ ค่าสารสนเทศของข้อสอบที่ระดับปานกลางที่ได้จากการให้คะแนนแบบ 2 - SR มีค่าสูงกว่าค่าที่ได้จากการให้คะแนนแบบปกติ ส่วนค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูงพบว่าให้ค่าที่ต่ำกว่าค่าที่ได้จากการให้คะแนนแบบปกติ

สำหรับค่าความยากแบบ CTT และความเป็นมิติเดียวพบว่าค่าที่ได้จากการให้คะแนนทั้งสองวิธีมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .05 เฉพาะในแบบทดสอบวัดความสามารถในการอ่านจับใจความเท่านั้น

2. ผลการศึกษาอิทธิพลของการคัดเลือกข้อสอบ

จากการศึกษาวิจัยพบว่า ความเชื่อมั่นของแบบทดสอบวัดความสามารถในการอ่านจับใจความ และการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ได้จากการให้คะแนนด้วยวิธีแบบปกติ และแบบ 2 - SR เมื่อถูกคัดเลือกเข้าฉบับด้วยความยาก สัมประสิทธิ์การจำแนก ดัชนีการจำแนก ความลำเอียงเนื่องจากตัวแปรเพศและความเป็นมิติเดียว มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .05 แต่เมื่อปรับขยายความยาวของข้อสอบโดยสูตรของสเปียร์แมนและบราวน์ให้มีจำนวนข้อเท่ากัน (40 ข้อ) พบว่าความเชื่อมั่นของแบบทดสอบไม่แตกต่างกัน

สำหรับค่าเฉลี่ยของค่าสารสนเทศของข้อสอบพบว่า ค่าเฉลี่ยของค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูง ปานกลาง และต่ำของแบบทดสอบวัดความสามารถในการอ่านจับใจความ ที่ได้จากการให้คะแนนด้วยวิธีแบบปกติ และแบบ 2 - SR เมื่อจำแนกตามค่าความยากแบบ CTT และดัชนีการจำแนก มีความแตกต่างกันอย่างมีนัยสำคัญ ณ ระดับ .05 และเมื่อจำแนกด้วยสัมประสิทธิ์การจำแนก และความเป็นมิติเดียวพบว่า มีความแตกต่างกันอย่างมีนัยสำคัญ ณ ระดับ .05 เฉพาะระดับความสามารถสูงและต่ำเท่านั้น

ส่วนค่าสารสนเทศของข้อสอบของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์เมื่อจำแนกโดยค่าสัมประสิทธิ์การจำแนก และดัชนีการจำแนกพบว่ามีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .05 เฉพาะสารสนเทศที่ระดับความสามารถสูงและต่ำ และเมื่อจำแนกโดยค่าความยากแบบ IRT และความเป็นมิติเดียวพบว่าค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลางที่ได้จากการให้คะแนนแบบปกติ และแบบ 2- SR และค่าสารสนเทศที่ระดับความสามารถต่ำที่ได้คะแนนแบบปกติ มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .05 และเมื่อจำแนกตามความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศพบว่าค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลาง ของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ได้คะแนนแบบปกติ และค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ได้คะแนนแบบปกติ มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .05

3. ผลการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรต่าง ๆ กับคุณภาพของแบบทดสอบ จากการศึกษาความสัมพันธ์ตามกรอบทฤษฎี CTT พบว่าความเชื่อมั่นรายข้อของแบบทดสอบวัดความสามารถในการอ่านจับใจความ และความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนแบบปกติ และแบบ 2 - SR มีความสัมพันธ์เชิงบวกในระดับสูงกับสัมประสิทธิ์การจำแนก ซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .98 และมีความสัมพันธ์กับดัชนีการจำแนก ความเป็นมิติเดียว ความยาก และความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ ซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์อยู่ในช่วงระหว่าง .83 - .99 , .61 - .94 , .40 - .84 และ .34 - .52 ตามลำดับ

เมื่อพิจารณาความสัมพันธ์ตามกรอบทฤษฎี IRT พบว่า ค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูงของแบบทดสอบวัดความสามารถในการอ่านจับใจความ และแก้โจทย์ปัญหาทางคณิตศาสตร์มีความสัมพันธ์เชิงบวกกับความยากแบบ IRT และความเหมาะสมของข้อสอบกับโมเดล ซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์อยู่ในช่วงระหว่าง .89 - .99 และ .32 - .59 ตามลำดับและพบว่ามีความสัมพันธ์เชิงลบกับความเป็นมิติเดียว ยกเว้นแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาที่ให้คะแนนแบบ 2 - SR และพบว่ามีความสัมพันธ์เชิงลบกับความลำเอียงของข้อสอบ ยกเว้นแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาที่ให้คะแนนแบบปกติ

ส่วนค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลาง พบว่าแบบทดสอบวัดความสามารถในการอ่านจับใจความ และแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนแบบ 2 - SR มีความสัมพันธ์เชิงบวกกับความเป็นมิติเดียว ซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์อยู่ในช่วงระหว่าง .25 - .40 และพบว่าแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหามีความสัมพันธ์เชิงบวกกับค่าความยากแบบ IRT ซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์อยู่ในช่วงระหว่าง .41 - .43

และสำหรับค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำของแบบทดสอบวัดความสามารถในการอ่านจับใจความและการแก้โจทย์ปัญหา พบว่า มีความสัมพันธ์เชิงลบกับค่าความยากแบบ IRT และความเหมาะสมของโมเดลกับข้อสอบ ซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์อยู่ในช่วงระหว่าง -.91 ถึง -.99 และ -.25 ถึง -.62 แต่มีความสัมพันธ์เชิงบวกกับความลำเอียงของข้อสอบ ซึ่งมี ค่าสัมประสิทธิ์สหสัมพันธ์อยู่ในช่วงระหว่าง .09 - .48 และความเป็นมิติเดียว ยกเว้นแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหา ซึ่งมีช่วงความสัมพันธ์ระหว่าง -.29 ถึง .91

เมื่อพิจารณาความสัมพันธ์ระหว่างคุณภาพของแบบทดสอบตามกรอบ CTT และ IRT พบว่า ค่าความเชื่อมั่นรายข้อของแบบทดสอบวัดความสามารถในการอ่านจับใจความและการแก้โจทย์ปัญหาที่ให้คะแนนแบบปกติและแบบ 2-SR มีความสัมพันธ์เชิงบวกกับสารสนเทศของ

ข้อสอบที่ระดับความสามารถปานกลางและต่ำ ซึ่งมีช่วงความสัมพันธ์อยู่ระหว่าง .17-.77 และมีความสัมพันธ์เชิงลบกับสารสนเทศของข้อสอบที่ระดับความสามารถสูง ซึ่งมีช่วงความสัมพันธ์อยู่ระหว่าง -.44 ถึง -.83

และเมื่อพิจารณาความสัมพันธ์ของค่าพารามิเตอร์ของข้อสอบตามกรอบ CTT กับสารสนเทศของข้อสอบ พบว่า ความยากแบบ CTT สัมประสิทธิ์การจำแนกและดัชนีการจำแนกมีความสัมพันธ์เชิงลบกับสารสนเทศของข้อสอบที่ระดับความสามารถสูง ซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์อยู่ในช่วงระหว่าง -.44 ถึง -.83 แต่มีความสัมพันธ์เชิงบวกกับค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำ ซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์อยู่ในช่วงระหว่าง .05 - .99 ส่วนสารสนเทศของข้อสอบที่ระดับความสามารถปานกลาง พบว่า มีความสัมพันธ์เชิงบวกกับความยากแบบ CTT สัมประสิทธิ์การจำแนกและดัชนีการจำแนกของแบบทดสอบวัดความสามารถในการอ่านจับใจความ ซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์อยู่ในช่วงระหว่าง .22 - .54

4. ผลการวิเคราะห์สมการพยากรณ์เชิงเส้นตรง (สมการถดถอยพหุคูณ)

จากการวิเคราะห์ตามกรอบทฤษฎี CTT พบว่าค่าสัมประสิทธิ์พยากรณ์ของสมการพยากรณ์ความเชื่อมั่นรายข้อ ของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนแบบปกติมีค่าเท่ากับ .97 ประสิทธิภาพพยากรณ์เท่ากับ 84% ตัวแปรในสมการได้แก่สัมประสิทธิ์การจำแนก ดัชนีการจำแนก และความเป็นมิติเดียว และเมื่อให้คะแนนแบบ 2 - SR พบว่าค่าสัมประสิทธิ์พยากรณ์เท่ากับ .99 ประสิทธิภาพพยากรณ์เท่ากับ 92% ตัวแปรในสมการคือสัมประสิทธิ์การจำแนก และดัชนีการจำแนก สำหรับค่าสัมประสิทธิ์พยากรณ์ของสมการพยากรณ์ความเชื่อมั่นรายข้อ ของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ให้คะแนนแบบปกติมีค่าเท่ากับ .99 ประสิทธิภาพพยากรณ์เท่ากับ 95% ตัวแปรในสมการคือสัมประสิทธิ์การจำแนก ดัชนีการจำแนก และความยาก และเมื่อให้คะแนนแบบ 2 - SR พบว่าค่าสัมประสิทธิ์พยากรณ์เท่ากับ .98 ประสิทธิภาพพยากรณ์เท่ากับ 87% ตัวแปรในสมการคือสัมประสิทธิ์การจำแนก ดัชนีการจำแนก ความเป็นมิติเดียว และความยาก

จากการวิเคราะห์ตามกรอบทฤษฎี IRT พบว่าค่าสัมประสิทธิ์พยากรณ์ของสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูงของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนแบบปกติมีค่าเท่ากับ .88 ประสิทธิภาพพยากรณ์เท่ากับ 65% ตัวแปรในสมการได้แก่ความยาก และเมื่อให้คะแนนแบบ 2 - SR พบว่า ค่าสัมประสิทธิ์พยากรณ์เท่ากับ .98 ประสิทธิภาพพยากรณ์เท่ากับ 87% ตัวแปรในสมการได้แก่ความยาก สำหรับค่า

สัมประสิทธิ์พยากรณ์ของสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูง ของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาที่ให้คะแนนแบบปกติมีค่าเท่ากับ .81 ประสิทธิภาพพยากรณ์เท่ากับ 57% ตัวแปรในสมการได้แก่ความยาก และความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ และเมื่อให้คะแนนแบบ 2 - SR พบว่า ค่าสัมประสิทธิ์พยากรณ์เท่ากับ .98 ประสิทธิภาพพยากรณ์เท่ากับ 88% ตัวแปรในสมการได้แก่ความยาก และความเป็นมิติเดียว

เมื่อพิจารณาค่าสัมประสิทธิ์พยากรณ์ของสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลางของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนแบบปกติมีค่าเท่ากับ .90 ประสิทธิภาพพยากรณ์เท่ากับ 68% ตัวแปรในสมการคือความยาก และเมื่อให้คะแนนแบบ 2 - SR พบว่า ค่าสัมประสิทธิ์พยากรณ์เท่ากับ .44 ประสิทธิภาพพยากรณ์เท่ากับ 25% ตัวแปรในสมการคือ ความยาก ความเป็นมิติเดียว ความเหมาะสมของข้อสอบกับโมเดล สำหรับค่าสัมประสิทธิ์พยากรณ์ของสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลางของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาที่ให้คะแนนแบบปกติมีค่าเท่ากับ .29 ประสิทธิภาพพยากรณ์เท่ากับ 16% ตัวแปรในสมการคือความยาก และความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ และเมื่อให้คะแนนแบบ 2 - SR พบว่า ค่าสัมประสิทธิ์พยากรณ์เท่ากับ .28 ประสิทธิภาพพยากรณ์เท่ากับ 15% ตัวแปรในสมการคือความยาก และความเป็นมิติเดียว

และพบว่าสัมประสิทธิ์พยากรณ์ของสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนแบบปกติมีค่าเท่ากับ .83 ประสิทธิภาพพยากรณ์เท่ากับ 59% ตัวแปรในสมการคือความยาก และเมื่อให้คะแนนแบบ 2 - SR พบว่า ค่าสัมประสิทธิ์พยากรณ์เท่ากับ .98 ประสิทธิภาพพยากรณ์เท่ากับ 87% ตัวแปรในสมการได้แก่ความยาก สำหรับค่าสัมประสิทธิ์พยากรณ์ของสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูง ของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาที่ให้คะแนนแบบปกติมีค่าเท่ากับ .90 ประสิทธิภาพพยากรณ์เท่ากับ 69% ตัวแปรในสมการได้แก่ความยาก และความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ และเมื่อให้คะแนนแบบ 2 - SR พบว่า ค่าสัมประสิทธิ์พยากรณ์เท่ากับ .99 ประสิทธิภาพพยากรณ์เท่ากับ 88% ตัวแปรในสมการได้แก่ความยาก และความเป็นมิติเดียว

จากการวิเคราะห์ตามกรอบทฤษฎี CTT และ IRT ร่วมกัน พบว่าค่าสัมประสิทธิ์พยากรณ์ของสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูงของแบบทดสอบ

วัดความสามารถในการอ่านจับใจความที่ให้คะแนนแบบปกติมีค่าเท่ากับ .96 ประสิทธิภาพพยากรณ์เท่ากับ 79% ตัวแปรในสมการได้แก่ความยากแบบ IRT ความยากแบบ CTT ความเหมาะสมของข้อสอบกับโมเดล และดัชนีการจำแนก และเมื่อให้คะแนนแบบ 2 - SR พบว่า ค่าสัมประสิทธิ์พยากรณ์เท่ากับ .99 ประสิทธิภาพพยากรณ์เท่ากับ 95% ตัวแปรในสมการคือความยากแบบ IRT ดัชนีการจำแนก และความเหมาะสมของข้อสอบกับโมเดล สำหรับค่าสัมประสิทธิ์พยากรณ์ของสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูงของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาที่ให้คะแนนแบบปกติมีค่าเท่ากับ .92 ประสิทธิภาพพยากรณ์เท่ากับ 72% ตัวแปรในสมการได้แก่ความยากแบบ IRT ความยากแบบ CTT สัมประสิทธิ์การจำแนก ดัชนีการจำแนก และความเป็นมิติเดียว และเมื่อให้คะแนนแบบ 2 - SR พบว่า ค่าสัมประสิทธิ์พยากรณ์เท่ากับ .98 ประสิทธิภาพพยากรณ์เท่ากับ 88% ตัวแปรในสมการคือความยากแบบ IRT และดัชนีการจำแนก

เมื่อพิจารณาค่าสัมประสิทธิ์พยากรณ์ของสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลางของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนแบบปกติมีค่าเท่ากับ .71 ประสิทธิภาพพยากรณ์เท่ากับ 46% ตัวแปรในสมการคือความยากแบบ IRT ความยากแบบ CTT ความเหมาะสมของข้อสอบกับโมเดล สัมประสิทธิ์ในการจำแนก และดัชนีการจำแนก และเมื่อให้คะแนนแบบ 2 - SR พบว่า ค่าสัมประสิทธิ์พยากรณ์เท่ากับ .86 ประสิทธิภาพพยากรณ์เท่ากับ 63% ตัวแปรในสมการได้แก่ ดัชนีการจำแนก ความยากแบบ CTT และความเหมาะสมของข้อสอบกับโมเดล สำหรับค่าสัมประสิทธิ์พยากรณ์ของสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับความสามารถปานกลางของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาที่ให้คะแนนแบบปกติมีค่าเท่ากับ .58 ประสิทธิภาพพยากรณ์เท่ากับ 36% ตัวแปรในสมการได้แก่สัมประสิทธิ์การจำแนก ดัชนีการจำแนก และความเป็นมิติเดียว และเมื่อให้คะแนนแบบ 2 - SR พบว่า ค่าสัมประสิทธิ์พยากรณ์เท่ากับ .28 ประสิทธิภาพพยากรณ์เท่ากับ 15% ตัวแปรในสมการได้แก่ความยากแบบ IRT และดัชนีการจำแนก

และพบว่าค่าสัมประสิทธิ์พยากรณ์ของสมการพยากรณ์ค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำของแบบทดสอบวัดความสามารถในการอ่านจับใจความที่ให้คะแนนแบบปกติมีค่าเท่ากับ .95 ประสิทธิภาพพยากรณ์เท่ากับ 78% ตัวแปรในสมการได้แก่ความยากแบบ IRT ความยากแบบ CTT สัมประสิทธิ์การจำแนก ความเหมาะสมของข้อสอบกับโมเดล และดัชนีการจำแนก และเมื่อให้คะแนนแบบ 2 - SR พบว่า ค่าสัมประสิทธิ์พยากรณ์เท่ากับ .99 ประสิทธิภาพพยากรณ์เท่ากับ 94 % ตัวแปรในสมการได้แก่ความยากแบบ IRT ดัชนีการจำแนก และ

ความเหมาะสมของข้อสอบกับโมเดล สำหรับค่าสัมประสิทธิ์พยากรณ์ของสมการพยากรณ์ ค่าสารสนเทศของข้อสอบที่ระดับความสามารถต่ำของแบบทดสอบวัดความสามารถในการ แก้ไขปัญหาที่ให้คะแนนแบบปกติมีค่าเท่ากับ .96 ประสิทธิภาพพยากรณ์เท่ากับ 81% ตัวแปร ในสมการได้แก่ความยากแบบ IRT ความยากแบบ CTT สัมประสิทธิ์การจำแนก ดัชนีการจำแนก และความเป็นมิติเดียว และเมื่อให้คะแนนแบบ 2 - SR พบว่า ค่าสัมประสิทธิ์พยากรณ์เท่ากับ .99 ประสิทธิภาพพยากรณ์เท่ากับ 88% ตัวแปรในสมการได้แก่ความยากแบบ IRT และดัชนีการ จำแนก

5. ผลการวิเคราะห์ความสัมพันธ์เชิงสาเหตุ

จากการศึกษาแผนภาพเส้นทางเชิงประจักษ์ที่แสดงความสัมพันธ์เชิงสาเหตุของ ตัวแปรต่าง ๆ ที่ใช้เป็นเกณฑ์การคัดเลือกข้อสอบจากการให้คะแนนตามรูปแบบปกติใน แบบทดสอบวัดความสามารถในการอ่านจับใจความ และความสามารถในการแก้โจทย์ปัญหาทาง คณิตศาสตร์ตามกรอบทฤษฎี CTT พบว่ามีความสอดคล้องกับแผนภาพสมมุติฐาน(เดิมรูป) ทุกภาพโดยที่แสดงค่าสัมประสิทธิ์พยากรณ์ และประสิทธิภาพการพยากรณ์สูงกว่าค่าที่ได้จาก สมการถดถอยพหุคูณ อย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01 และพบว่าตัวแปรที่มีอิทธิพลต่อ ความเชื่อมั่นของข้อสอบคือ ความยาก สัมประสิทธิ์การจำแนก ดัชนีการจำแนก และความเป็น มิติเดียวของข้อสอบ

สำหรับการศึกษาตามกรอบทฤษฎี IRT พบว่าแผนภาพเส้นทางเชิงประจักษ์ที่แสดง ความสัมพันธ์เชิงสาเหตุระหว่างตัวแปรอิสระกับสารสนเทศของข้อสอบของแบบทดสอบวัด ความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ไม่สอดคล้องกับแผนภาพเส้นทางสมมุติฐาน ทุกแผนภาพแต่เมื่อนำตัวแปรความยากและอำนาจจำแนกตามกรอบทฤษฎี CTT ร่วมเป็นตัวแปร พยากรณ์พบว่า แผนภาพเส้นทางเชิงประจักษ์สอดคล้องกับแผนภาพเส้นทางสมมุติฐานทุก แผนภาพรวมทั้งในแบบทดสอบวัดความสามารถในการอ่านจับใจความ และแสดงค่าสัมประสิทธิ์ พยากรณ์สูงกว่าค่าที่ได้จากสมการถดถอยพหุคูณอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

เมื่อศึกษาอิทธิพลของตัวแปรอิสระตามกรอบทฤษฎี IRT พบว่าตัวแปรต่าง ๆ ส่ง อิทธิพลแตกต่างกันไปตามระดับความสามารถ โดยที่ความยากส่งอิทธิพลเชิงบวกกับสารสนเทศ ของข้อสอบที่ระดับความสามารถสูงขณะเดียวกันส่งอิทธิพลเชิงลบกับสารสนเทศของข้อสอบที่ ระดับความสามารถต่ำ ในทางตรงกันข้ามความเป็นมิติเดียวส่งอิทธิพลเชิงบวกกับสารสนเทศของ ข้อสอบที่ระดับความสามารถต่ำแต่ส่งอิทธิพลเชิงลบกับสารสนเทศของข้อสอบที่ระดับความ สามารถสูง นอกจากนั้นจากการศึกษาอิทธิพลของตัวแปรอิสระที่ใช้เป็นเกณฑ์ในการคัดเลือก

ข้อสอบเข้าฉบับพบว่า ค่าความยากแบบ CTT ส่งอิทธิพลเชิงบวกกับสารสนเทศของข้อสอบที่ระดับความสามารถต่ำ แต่ส่งอิทธิพลเชิงลบกับสารสนเทศของข้อสอบที่ระดับความสามารถสูง ส่วนค่าดัชนีการจำแนกส่งอิทธิพลเชิงบวกกับสารสนเทศของข้อสอบที่ระดับความสามารถปานกลางแต่ส่งอิทธิพลต่อสารสนเทศของข้อสอบที่ระดับความสามารถสูงและต่ำ

อภิปรายผล

การศึกษาวิจัยครั้งนี้เป็นการศึกษาผลที่เกิดขึ้นกับแบบทดสอบเลือกตอบในสองประเด็นสำคัญคือ ผลที่เกิดจากอิทธิพลของการให้คะแนน และผลที่เกิดจากอิทธิพลของตัวแปรต่างๆที่เป็นค่าสถิติ และคุณสมบัติบางประการของข้อสอบซึ่งนำมาใช้เป็นเกณฑ์ในการคัดเลือกข้อสอบเข้าฉบับ โดยแปลผลจากการวิเคราะห์ค่าสถิติต่างๆ สองส่วนคือ ส่วนที่เป็นค่าสถิติอธิบายความสามารถของผู้สอบได้แก่ คะแนนสอบ ค่าความสามารถของผู้สอบที่คำนวณตามกรอบทฤษฎี IRT แบบราชคโหมเดล และส่วนที่เป็นค่าสถิติอธิบายคุณสมบัติของข้อสอบได้แก่ ค่าความยาก ค่าอำนาจจำแนก ค่าความเชื่อมั่นรายข้อและรายฉบับตามกรอบทฤษฎี CTT ค่าความยาก และความเหมาะสมของข้อสอบกับโมเดล ตามกรอบทฤษฎี IRT ค่าสัมประสิทธิ์สหสัมพันธ์แบบ Contingency ซึ่งเป็นค่าใช้แสดงแทนความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ และน้ำหนักองค์ประกอบของข้อสอบแต่ละข้อ ซึ่งใช้แสดงแทนความเป็นมิติเดียวของข้อสอบ ซึ่งผลวิจัยในส่วนที่แสดงถึงอิทธิพลของรูปแบบการให้คะแนนต่อคะแนนสอบพบว่า คะแนนสอบที่ได้จากการให้คะแนนแบบปกติสูงกว่าคะแนนที่ได้จากการให้คะแนนแบบ 2-SR อย่างมีนัยสำคัญ ณ ระดับ .05 ในแบบทดสอบทั้งสองฉบับ ซึ่งไม่สอดคล้องกับงานวิจัยของเอสเซก (Essex, 1976 : 565 - 572) ที่พบว่าคะแนนสอบที่ได้จากการตรวจแบบพิจารณาความรู้ส่วนย่อยมากกว่าคะแนนสอบที่ได้จากการให้คะแนนแบบปกติ กรณีนี้อธิบายได้ว่าการให้คะแนนแบบปกติผู้สอบจะต้องตัดสินใจเลือกเพียงคำตอบเดียว ดังนั้นบางครั้งจำเป็นต้องใช้การเดาช่วยในการตัดสินใจ ซึ่งอาจเป็นการเดาแบบสุ่มหรือเดาอย่างมีความรู้โดยอาศัยพื้นฐานความรู้อื่นมาประกอบ และเมื่อใช้วิธีการแบบ 2-SR ซึ่งความจริงแล้วการตอบในส่วนที่ 2 ก็คือวิธีการตอบแบบปกติ (0-1) นั่นเองและสำหรับในส่วนที่ 1 เป็นส่วนที่ช่วยในการวินิจฉัยผลการตอบในส่วนที่สองว่าผู้สอบมีความรู้สมบูรณ์ หรือมีความรู้บางส่วน หรือไม่มีความรู้ ซึ่งในประเด็นหลังนี้สามารถสะท้อนถึงการเดาตอบของผู้สอบได้ และจากผลการวิจัยแสดงให้เห็นว่าเมื่อให้ตอบโดยวิธีแบบ 2 - SR ผู้สอบแสดงการตอบโดยพยายามใช้ความรู้ของตนเองมากกว่าที่จะเดาตอบ ดังนั้นจึงทำให้คะแนนแบบ 2-SR ลดลงกว่าคะแนนปกติดังเห็นได้จากคะแนนสูงสุด ต่ำสุดของแบบทดสอบลดลงเมื่อให้คะแนนแบบ 2-SR ไม่

ว่าจะเป็นคะแนนสูงสุดหรือต่ำสุด ยกเว้นในแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่คะแนนสูงสุดเท่ากันทั้ง 2 วิธี ทั้งนี้เพราะโดยลักษณะธรรมชาติของข้อสอบคณิตศาสตร์ส่วนใหญ่เป็นข้อสอบที่เน้นให้ผู้สอบคำนวณหาผลลัพธ์ ซึ่งมีคำตอบถูกเพียงคำตอบเดียว จึงทำให้นักเรียน เก่งเลือกตอบเพียงหนึ่งคำตอบ ซึ่งในกรณีเลือกคำตอบในลักษณะเช่นนี้ทำให้คะแนนที่ได้จากทั้งสองวิธีไม่ค่อยแตกต่างกันนักซึ่งน่าจะพบในกลุ่มคนเก่ง และเป็นไปได้ว่าคนเก่งตอบข้อสอบโดยใช้การเดาน้อยที่สุดหรือไม่เดาเลย ขณะเดียวกันในกลุ่มปานกลาง และกลุ่มอ่อน แม้ว่าจะเปิดโอกาสให้เลือกได้มากกว่า 1 คำตอบในส่วนที่หนึ่งแต่ถ้าไม่รู้ก็มีโอกาสได้ 0 เช่นเดียวกับวิธีปกติ และแม้ว่าสามารถเลือกตัวถูกในส่วนที่สอง แต่ถ้าเลือกคำตอบมากกว่า 1 ตัวเลือกในส่วนที่หนึ่งคะแนนที่ได้จะถูกลดไปตามจำนวนตัวเลือกที่เพิ่มขึ้น ในงานวิจัยครั้งนี้ใช้แบบทดสอบเลือกตอบ 4 ตัวเลือกซึ่งมีคะแนนสูงสุดเท่ากับ 6 ดังนั้นถ้าตอบถูกทั้งสองส่วน และพบว่าเลือก 2 คำตอบในส่วนที่หนึ่ง คะแนนที่ได้คือ 5 คะแนน หรือถ้าให้คะแนนแต่ละข้อสูงสุดเท่ากับ 1 ก็จะได้เท่ากับ 0.83 คะแนน ซึ่งลดลงจากคะแนนที่ให้แบบปกติประมาณ 17 % อย่างไรก็ตามถ้าผู้สอบทุกคนเลือกคำตอบในส่วนที่หนึ่งและส่วนที่สอง ส่วนละ 1 คำตอบผลที่ได้จะเท่ากับวิธีแบบปกติ (0-1) ซึ่งเมื่อพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนสอบที่ได้จากการให้คะแนนทั้งสองวิธีพบว่า มีความสัมพันธ์ในระดับสูงเท่ากับ .96 ทั้งสองฉบับ และเมื่อนำคะแนนที่ได้มาประมาณค่าระดับความสามารถตามกรอบทฤษฎี IRT แบบราชดิมเดล พบว่าระดับเฉลี่ยของความสามารถในการอ่านจับใจความที่ให้คะแนนแบบปกติสูงกว่าระดับเฉลี่ยที่ให้คะแนนแบบ 2-SR อย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01 ในทางกลับกันระดับเฉลี่ยของความสามารถในการแก้โจทย์ปัญหาที่ให้คะแนนแบบ 2-SR กลับสูงกว่าระดับความสามารถเฉลี่ยที่ได้จากการให้คะแนนแบบปกติอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .05 อย่างไรก็ตามเมื่อพิจารณาส่วนเบี่ยงเบนมาตรฐานของระดับความสามารถที่ได้จากการให้คะแนนแบบปกติพบว่ามีความสูงกว่าระดับความสามารถที่ได้จากการให้คะแนนแบบ 2-SR ซึ่งแสดงถึงการกระจายของระดับความสามารถที่ได้จากการให้คะแนนแบบปกติมีมากกว่า หรืออาจดูได้จากพิสัยของระดับความสามารถสูงสุดและต่ำสุด ที่พบว่า ระดับความสามารถสูงสุดที่ได้จากการให้คะแนนแบบปกติสูงกว่าระดับความสามารถสูงสุดที่ได้จากการให้คะแนนแบบ 2-SR ทั้งสองฉบับ แต่พบว่าระดับความสามารถต่ำสุดที่ได้จากการให้คะแนนแบบ 2-SR มีค่าน้อยกว่าระดับความสามารถต่ำสุดที่ได้จากการให้คะแนนแบบปกติ และเมื่อพิจารณาค่าความคลาดเคลื่อนในการประมาณค่าความสามารถพบว่า การให้คะแนนแบบ 2-SR ให้ค่าความคลาดเคลื่อนน้อยกว่า และเมื่อเปรียบเทียบค่าความสัมพันธ์ระหว่างคะแนนสอบกับระดับความสามารถพบว่า สัมประสิทธิ์สหสัมพันธ์ของคะแนนสอบที่ให้แบบ 2-SR กับระดับ

ความสามารถมีค่าอยู่ในระดับสูง คือ เท่ากับ .73 และ .76 ขณะที่สัมประสิทธิ์สหสัมพันธ์ของคะแนนสอบแบบปกติกับระดับความสามารถเท่ากับ .28 และ .59 เท่านั้น ซึ่งอาจกล่าวยืนยันได้ว่าการให้คะแนนแบบ 2-SR มีแนวโน้มช่วยลดความคลาดเคลื่อนของคะแนนที่เกิดจากการเดาตอบของผู้สอบได้ จึงทำให้คะแนนสอบมีโอกาสใกล้เคียงกับคะแนนจริงของผู้สอบมากกว่าการให้คะแนนแบบปกติ นอกจากนี้น่าจะเปิดโอกาสให้เด็กกลุ่มอ่อนได้คะแนนมากกว่าการให้คะแนนแบบปกติ และจากการวิจัยครั้งนี้ยังได้ข้อค้นพบว่า การเดาตอบในแบบทดสอบวัดความสามารถในการอ่านจับใจความ มีแนวโน้มสูงกว่าในแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ ซึ่งเป็นไปตามลักษณะธรรมชาติของวิชา กล่าวคือการตอบข้อสอบวัดความสามารถในการอ่านจับใจความต้องใช้กระบวนการคิดที่หลากหลาย ส่วนการตอบข้อสอบที่เป็นการแก้โจทย์ปัญหาทางคณิตศาสตร์ เป็นการใช้กระบวนการคิดหาคำตอบตามหลักวิชาที่มีแบบแผนวิธีการอยู่แล้ว ดังนั้นจึงไม่ซับซ้อนเท่าใดนัก ผู้ตอบคิดได้อย่างไรก็ตอบไปอย่างนั้น เว้นแต่จะไม่ทราบวิธีการทำจึงจำเป็นต้องเดาตอบ

สำหรับผลวิจัยที่แสดงอิทธิพลของการให้คะแนนต่อค่าสถิติ คุณสมบัติบางประการ และคุณภาพของแบบทดสอบในการศึกษาวิจัยครั้งนี้พบว่า ความเชื่อมั่นของแบบทดสอบที่ให้คะแนนโดยวิธีการทั้งสองวิธี เมื่อทดสอบนัยสำคัญทางสถิติพบว่า ไม่แตกต่างกันทั้งสองฉบับ ซึ่งให้ผลที่สอดคล้องกับงานวิจัยที่ผ่านมาของเอสเซก (Essex. 1976 : 565 - 572) ฟู่จูโต และฮอยเคซุน (Fu-Ju Tsi and Hoi K. Suen.1993 : 404) ธีรศักดิ์ อินทร์มาตย์ (2520 : 63-73) ทวี ทองคำ (2526 : บทคัดย่อ) ที่พบว่า การให้คะแนนแบบทดสอบเลือกตอบที่ไม่ใช่แบบ 0-1 หรือแบบปกติ ให้ค่าความเชื่อมั่นไม่แตกต่างจากเมื่อให้คะแนนแบบปกติ หรือข้อค้นพบของจินดา ไดอนันต์ (2527 : 59-65) ที่ศึกษาความเชื่อมั่นของแบบทดสอบ MTF ซึ่งให้คะแนนที่ไม่ใช่แบบปกติ และพบว่าไม่แตกต่างจากแบบทดสอบแบบ MC และข้อค้นพบของวัชรภรณ์ จิตรมาศ (2538 : 72-87) ที่พบว่าแบบทดสอบถูกผิดหลายตัวเลือก ที่มีจำนวนตัวเลือก 4 ตัวเลือกเมื่อให้คะแนนที่ไม่ใช่แบบ 0-1 ให้ค่าความเชื่อมั่นไม่แตกต่างไปจากค่าที่ได้จากการให้คะแนนตามรูปแบบปกติ อย่างไรก็ตามเมื่อพิจารณาจากค่าความเชื่อมั่นที่ได้ของแบบทดสอบทั้งสองฉบับที่ให้คะแนนแบบ 2-SR พบว่ามีแนวโน้มสูงกว่า และมีความคลาดเคลื่อนมาตรฐานในการวัดต่ำกว่าค่าที่ได้จากคะแนนแบบปกติ

ข้อค้นพบอีกประการหนึ่งที่สนับสนุนแนวโน้มดังกล่าวก็คือการพิจารณาค่าสารสนเทศของแบบทดสอบที่ระดับความสามารถปานกลาง พบว่า วิธีการให้คะแนนแบบ 2-SR แสดงประสิทธิภาพสูงกว่าวิธีการให้คะแนนแบบปกติ ซึ่งสอดคล้องกับสภาพการปฏิบัติจริงทั่ว ๆ ไปที่ห้องเรียนจะมีผู้เรียนที่มีความสามารถปานกลางเป็นส่วนใหญ่ แต่เมื่อพิจารณาในกลุ่มระดับ

ความสามารถสูงหรือต่ำ พบว่า วิธีการให้คะแนนแบบปกติมีประสิทธิภาพสูงกว่าวิธีการแบบ 2-SR จากข้อค้นพบดังกล่าวสะท้อนให้เห็นว่ากลุ่มที่มีความสามารถปานกลางมีความเป็นไปได้ที่จะตอบโดยการเดามากกว่ากลุ่มที่มีความสามารถสูงหรือต่ำ เมื่อใช้วิธีการให้คะแนนแบบ 2-SR คะแนนจึงถูกปรับลดมากกว่ากลุ่มสูงหรือกลุ่มต่ำ ซึ่งคะแนนที่ได้จากการตอบทั้งสองวิธีน่าจะเท่ากัน กล่าวคือ คนเก่งจะได้คะแนนตามความรู้ที่มีจริง ส่วนคนอ่อนก็จะได้คะแนนในสิ่งที่ตนไม่รู้ ส่วนคนที่มีความสามารถปานกลางคือกลุ่มที่มีความรู้เพียงบางส่วน และถ้าให้คะแนนโดยวิธีแบบปกติก็อาจเป็นไปได้ที่จะได้คะแนน 0 หรือ 1 ตามความรู้ที่มีอยู่หรือการเดาตอบ แต่เมื่อใช้วิธีแบบ 2-SR ทำให้ผู้สอบกลุ่มนี้ไม่ยุ่งยากที่จะตัดสินใจและแสดงการตอบตามความรู้ที่มีอยู่

สำหรับแบบทดสอบทั้งสองฉบับที่ใช้เป็นเครื่องมือในการวิจัยครั้งนี้ พบว่าแบบทดสอบทั้งสองฉบับที่ให้คะแนนแบบ 0-1 ค่า TIF มีลักษณะคล้ายกับโค้งปกติ โดยที่ค่าสารสนเทศสูงสุดอยู่ที่ระดับความสามารถ (θ) ประมาณ 0.00 ส่วนค่า TIF ของแบบทดสอบที่ได้จากการให้คะแนนแบบ 2-SR ทั้ง 2 ฉบับ มีค่าสารสนเทศสูงสุดที่ระดับความสามารถประมาณ 0.00 เช่นเดียวกัน แต่มีค่าความโด่งที่ตำแหน่งของ θ_{max} มากกว่า TIF ของแบบทดสอบที่ให้คะแนนแบบปกติ ทั้งนี้อาจเป็นเพราะการให้คะแนนตามรูปแบบ 2-SR มีสเกลที่ละเอียดกว่าการให้คะแนนแบบปกติ

เมื่อพิจารณาค่าสถิติของข้อสอบที่ได้จากการให้คะแนนทั้งสองวิธีตามกรอบทฤษฎี CTTP พบว่าค่าสัมประสิทธิ์การจำแนก ซึ่งเป็นค่าอำนาจจำแนกที่ได้จากสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนรวม และดัชนีการจำแนก ซึ่งเป็นค่าอำนาจจำแนกที่ได้จากการหาความแตกต่างในการตอบถูกของกลุ่มเก่ง และกลุ่มอ่อน ความเชื่อมั่นรายข้อ ของแบบทดสอบทั้งสองฉบับที่ได้จากการให้คะแนนแบบ 2-SR มีค่าสูงกว่าค่าที่ได้จากการให้คะแนนแบบปกติอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01 ซึ่งสะท้อนถึงความเที่ยงตรงเชิงโครงสร้างของแบบทดสอบ และความเป็นเอกพันธ์ของข้อสอบตามที่ ครอนบาค และอนาสตาซีได้เสนอไว้ (Gronlund. 1982 : 125-138 ; citing Cronbach. 1971 ; Anastasi. 1990 : 155 - 156) อย่างไรก็ตาม โดยเฉพาะความแตกต่างของค่าอำนาจจำแนกนี้พบว่าให้ผลที่ไม่สอดคล้องกับงานวิจัยที่ผ่านมาของฟูจูไต และฮอยเคซุน (Fu-Ju Tsai and Hoi K. Suen. 1993 : 404) กาญจนา ศิริวัฒนพงษ์ (2520 : 64-79) และสำราญ มีแจ้ง (2520 : 55-61) ที่พบว่า ค่าอำนาจจำแนกของแบบทดสอบที่ให้คะแนนไม่ใช่แบบ 0-1 มีค่าอำนาจจำแนกไม่แตกต่างไปจากแบบทดสอบที่ให้คะแนนแบบปกติ และข้อค้นพบของจินดา โตอนันต์ (2527 : 59-65) และวัชรภรณ์ จิตมาศ (2538 : 72-87) ที่พบว่า ค่าอำนาจจำแนกของแบบทดสอบที่ให้คะแนนปกติมีค่าสูงกว่าค่าแบบทดสอบ MTF ที่ให้คะแนนไม่ใช่แบบ 0-1 แต่ให้ผลสอดคล้องกับงานวิจัยของเดรสเซลและชมิท (Dressel and Schmid : 1953 : 574-595) อัลเบิร์ต

(Albert : 1970 : 1619A) เต คี ชู และชูศักดิ์ ชัมภลิจิต (Hsu and Choosak Khampalikit : 1984 : 152 - 158) ดรุณ หาญตระกูล และนิโลบล นิ้มกักรัตน์ (2522 : 210-214) ที่พบว่าค่าอำนาจจำแนกของแบบทดสอบที่ให้คะแนนไม่ใช่แบบ 0-1 สูงกว่าแบบทดสอบที่ให้คะแนนแบบ 0-1 จากข้อค้นพบดังกล่าวนี้อาจอธิบายได้ว่า เนื่องจากการให้คะแนนแบบ 2 - SR นี้มีช่วงคะแนนมากกว่าแบบปกติ ดังนั้นจึงทำให้เกิดความแปรปรวนของคะแนนสอบสูงกว่าแบบปกติ ซึ่งโดยหลักการแล้วค่าอำนาจจำแนกทั้งสองแนวทางเมื่อพิจารณาจากสูตรได้แก่ พอยท์ไบซีเรียล ซึ่งเกี่ยวข้องโดยตรงกับความแปรปรวน หรือส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบ หรือสูตรอย่างง่ายที่หาความแตกต่างในการตอบถูกของกลุ่มสูงและกลุ่มต่ำเทียบกับจำนวนคนในกลุ่มสูง ซึ่งเป็นการเปรียบเทียบค่าสัดส่วนของการตอบถูกของคนทั้งสองกลุ่ม โดยมีหลักการว่าคะแนนสอบของคนในกลุ่มเดียวกันควรมีความแตกต่างกันน้อยซึ่งจะทำให้ความแปรปรวนในกลุ่มมีค่าน้อย แต่คะแนนระหว่างกลุ่มควรแตกต่างกันมากๆ ซึ่งจะทำให้ความแปรปรวนมีค่ามาก ดังนั้นการจำแนกจะมีค่าสูงตามไปด้วย และสำหรับความแตกต่างของค่าความเชื่อมั่นรายข้อ อธิบายได้โดยสูตรของความเชื่อมั่นรายข้อ ซึ่งกัลลิกเซน ประมาณค่าจากผลคูณระหว่างค่าสัมประสิทธิ์ความสัมพันธ์รายข้อกับคะแนนรวม หรือสัมประสิทธิ์การจำแนกกับค่าความแปรปรวนของคะแนนที่ได้จากการตอบข้อสอบข้อนั้นๆ (Gulliksen. 1950 : 386-395) ส่วนค่าความยาก พบว่าแบบทดสอบที่ให้คะแนนแบบ 2 - SR แสดงแนวโน้มว่ามีความยากมากกว่าการให้คะแนนแบบปกติ ทั้งนี้อาจเป็นเพราะว่าการที่ผู้สอบจะได้คะแนนในระดับสูงในแต่ละข้อนั้นจะต้องอาศัยภูมิรู้ และความมั่นใจในความรู้ของตนเองอย่างมาก ดังนั้นจึงทำให้คะแนนที่ได้มีโอกาสถูกปรับลดต่ำลงกว่าการตอบแบบปกติซึ่งการได้คะแนนนั้นมีโอกาสได้มาจากการเดาตอบมาก ดังนั้นสัดส่วนของการได้คะแนนจึงมีค่าต่ำลงซึ่งหมายถึงการมีความยากมากขึ้น ซึ่งไม่สอดคล้องกับข้อค้นพบของวัชรภรณ์ จิตรมาศ (2538 : บทคัดย่อ) ที่พบว่าแบบทดสอบที่ให้คะแนนแบบ MTF มีความยากน้อยกว่าเมื่อให้คะแนนแบบปกติ

สำหรับอิทธิพลที่มีต่อค่าพารามิเตอร์ และค่าสารสนเทศของข้อสอบตามกรอบทฤษฎี IRT พบว่าการให้คะแนนทั้งสองวิธีมีความเหมาะสมต่อระดับความสามารถแตกต่างกัน โดยพบว่าการให้คะแนนแบบปกติให้สารสนเทศของข้อสอบสูงกว่าการให้คะแนนแบบ 2 - SR ที่ระดับความสามารถสูง และแสดงแนวโน้มว่ามีค่าสูงกว่าที่ระดับความสามารถต่ำอีกด้วย แต่พบว่าให้สารสนเทศของข้อสอบต่ำกว่าการให้คะแนนแบบ 2 - SR ที่ระดับความสามารถปานกลางซึ่งได้อธิบายเหตุผลไปแล้วในส่วนที่อธิบายคุณภาพของแบบทดสอบ ส่วนค่าความยาก และความเหมาะสมของข้อสอบกับโมเดลพบว่าการให้คะแนนของทั้งสองวิธีให้ผลที่ไม่แตกต่างกัน ทั้งนี้อาจเป็นเพราะว่าการประมาณค่าความยากแบบราชคโสมเดลรายข้อที่ให้คะแนนแบบ 2 - SR นั้นได้จาก

การเฉลี่ยค่าความยากในแต่ละชั้น ซึ่งค่าที่ได้จะมีค่าเข้าใกล้ 0 และเมื่อนำมาเฉลี่ยอีกครั้งค่าจะยิ่งเข้าใกล้ 0 มากยิ่งขึ้น ทำให้ค่าไม่แตกต่างไปจากค่าเฉลี่ยของความยากที่ได้จากการให้คะแนนแบบปกติ ซึ่งมีค่าเข้าใกล้ 0 เช่นเดียวกัน สำหรับกรณีค่าความเหมาะสมของข้อสอบกับโมเดล เมื่อพิจารณาจากสูตรที่ไรท์ (Wright, 1979 : 66-82) นำเสนอพบว่า ค่าความเหมาะสมของข้อสอบกับโมเดลจำเป็นต้องใช้ค่าความยากของข้อสอบ มาเปรียบเทียบกับค่าความสามารถของผู้สอบ ดังนั้นจึงทำให้ค่าที่ได้จากการให้คะแนนทั้งสองวิธีไม่แตกต่างเช่นเดียวกัน

เมื่อพิจารณาความเป็นมิติเดียวจากค่าน้ำหนักองค์ประกอบ พบว่าค่าที่ได้จากการให้คะแนนแบบ 2 - SR สูงกว่าค่าที่ได้จากการให้คะแนนแบบปกติ ซึ่งความแตกต่างนี้อธิบายได้โดยใช้หลักการของการวิเคราะห์องค์ประกอบ ซึ่งเน้นค่าสหสัมพันธ์ภายในเพียงกลุ่มเดียว ไม่เกี่ยวข้องกับเกณฑ์ภายนอกใด ๆ เป็นการแสดงความสัมพันธ์ระหว่างตัวแปรเชิงประจักษ์ที่ได้กับตัวแปรเชิงสมมติฐาน โดยพยายามลดจำนวนตัวแปรให้น้อยลง แต่ยังสามารถอธิบายความแปรปรวนของตัวแปรเดิมให้มากที่สุด (Chartfield, 1980 : 125) ดังนั้นอาจกล่าวได้ว่าข้อมูลที่มีความแปรปรวนมาก ๆ การอธิบายองค์ประกอบต่างๆ จากค่าน้ำหนักองค์ประกอบของตัวแปรจะมีประสิทธิภาพสูงกว่าข้อมูลที่มีความแปรปรวนน้อย ๆ

และสำหรับค่าความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศซึ่งพิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์ Contingency พบว่าค่าที่ได้จากการให้คะแนนแบบ 2 - SR มีค่าสูงกว่าค่าที่ได้จากการให้คะแนนแบบปกติ ซึ่งหมายความว่าโอกาสที่ข้อสอบแสดงความลำเอียงเมื่อให้คะแนนแบบ 2 - SR มีมากกว่าการให้คะแนนแบบปกติ อาจอธิบายได้ว่าในการตรวจสอบความแตกต่างของการได้คะแนนของเพศชาย และเพศหญิงเมื่อให้คะแนนแบบปกติมีลักษณะเป็นตารางแบบ 2×2 และความถี่เชิงประจักษ์ในแต่ละเซลล์ มีโอกาสเกิดความคลาดเคลื่อนแบบ α หรือแบบ β สูงกว่าการให้คะแนนแบบ 2 - SR ที่มีลักษณะเป็นตารางแบบ 2×5 ซึ่งน่าจะแสดงค่าของอัตราส่วนความเป็นไปได้ (Likelihood ratio) ได้ละเอียดกว่าตารางแบบ 2×2

นอกจากผลที่เกิดจากการแสดงอิทธิพลของการให้คะแนนที่มีต่อ คะแนนสอบและความสามารถของผู้สอบ ค่าสถิติ และคุณสมบัติบางประการของข้อสอบดังที่ได้กล่าวมาแล้วยังพบว่า การให้คะแนนทั้งสองวิธีทำให้สมการพยากรณ์ และโมเดลความสัมพันธ์เชิงสาเหตุมีรูปแบบที่แตกต่างกัน ซึ่งจะอธิบายต่อไป

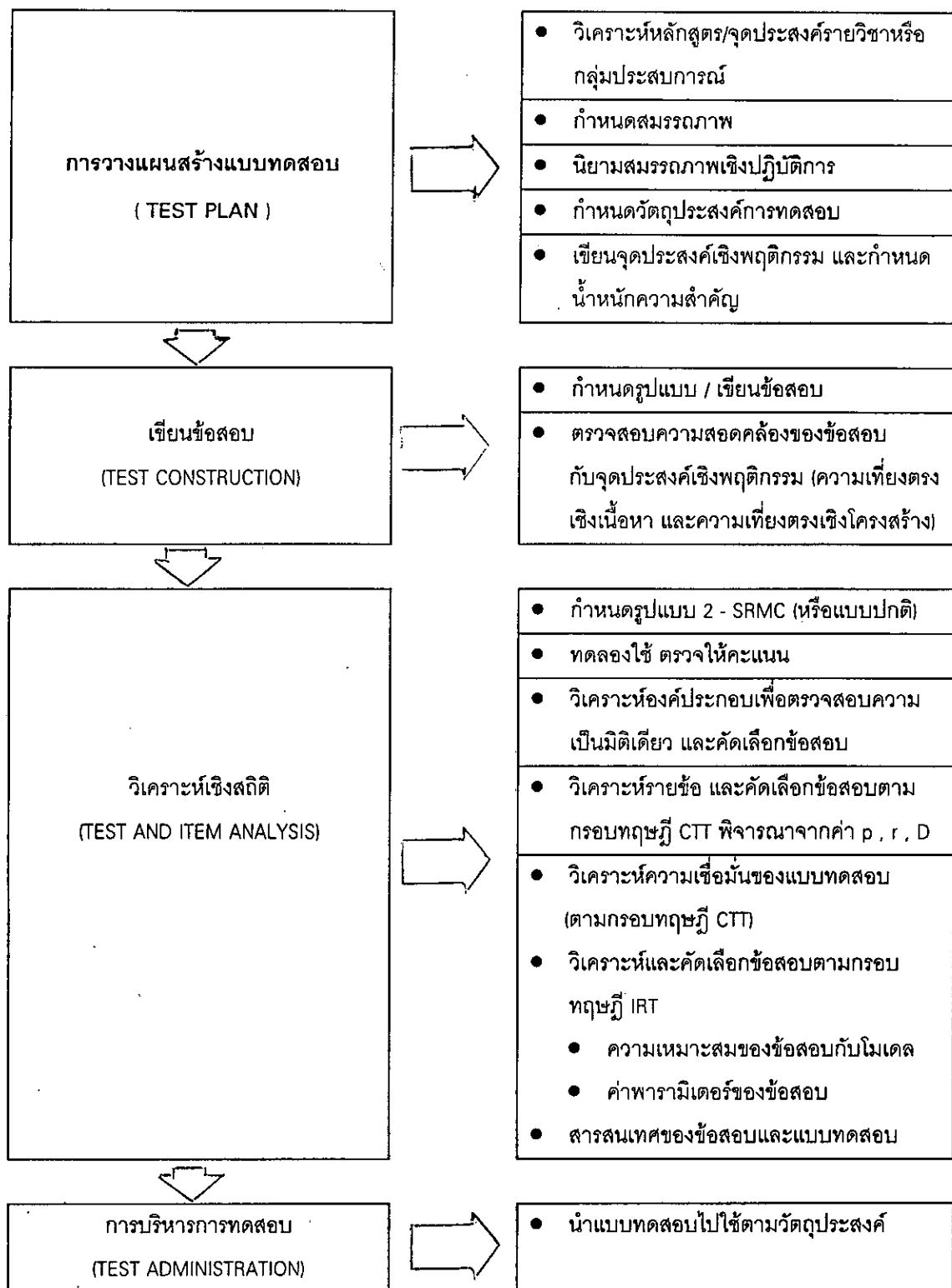
เนื่องจากการส่งอิทธิพลของการให้คะแนนดังกล่าว การวิเคราะห์อิทธิพลของตัวแปรที่ใช้เป็นเกณฑ์ในการคัดเลือกข้อสอบจำเป็นต้องวิเคราะห์ข้อมูลไปตามรูปแบบการให้คะแนน และแบบ

ทดสอบแต่ละฉบับ จากการวิเคราะห์ข้อมูลพบว่าตัวแปรที่ส่งอิทธิพลเชิงเส้นตรงต่อค่าความเชื่อมั่นรายข้อของแบบทดสอบในแบบทดสอบทุกฉบับและวิธีการให้คะแนนทั้งสองวิธีได้แก่ สัมประสิทธิ์การจำแนก และดัชนีการจำแนกซึ่งทั้งสองค่านี้น่าจะเป็นตัวแปรที่นำมาใช้เป็นเกณฑ์ในการคัดเลือกข้อสอบเข้าฉบับตามกรอบทฤษฎี CTT สำหรับในทางปฏิบัติการคำนวณค่าสัมประสิทธิ์การจำแนกอาจเป็นเรื่องยากสำหรับคนทั่วไปที่ไม่ได้ศึกษาทางสถิติ วิจัย หรือวัดผล ดังนั้นเมื่อพิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างสัมประสิทธิ์การจำแนก กับดัชนีการจำแนกมีค่าอยู่ในช่วงระหว่าง .84 - .95 ดังนั้นในการวิเคราะห์ข้อสอบเบื้องต้นการใช้ค่าดัชนีการจำแนกเพียงค่าเดียวก็น่าจะเป็นที่ยอมรับได้

สำหรับตัวแปรที่ส่งอิทธิพลเชิงเส้นตรงกับค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูง ได้แก่ค่าความยาก ความเป็นมิติเดียว และความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศ ซึ่งสอดคล้องกับข้อตกลงของทฤษฎี IRT เกี่ยวกับความเป็นมิติเดียวของแบบทดสอบ และสอดคล้องกับแนวคิดของไคลน์ (Kline, 1993 : 134-157) ที่เสนอให้พิจารณาความแตกต่างระหว่างเพศ และจากการใช้การวิเคราะห์ตามกรอบ CTT เพิ่มเติมพบว่าให้ค่าประสิทธิภาพในการพยากรณ์สูงขึ้นในประมาณ 7 % เมื่อให้คะแนนแบบปกติ และ 31 % เมื่อให้คะแนนแบบ 2 - SR เฉพาะแบบทดสอบวัดความสามารถในการอ่านจับใจความ ซึ่งตัวแปรที่มีอิทธิพลเชิงเส้นตรงได้แก่ ค่าดัชนีการจำแนก ในแบบทดสอบที่ให้คะแนนแบบ 2 - SR และตัวแปรความยาก สัมประสิทธิ์การจำแนก ในแบบทดสอบที่ให้คะแนนแบบปกติ ส่วนในแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์พบว่าเมื่อใช้ตัวแปรตามกรอบ CTT เพิ่มเข้าไป ค่าประสิทธิภาพพยากรณ์เพิ่มขึ้นเพียง 7 % ตัวแปรที่เพิ่มเข้าไปได้แก่ ค่าดัชนีการจำแนก และความเหมาะสมของข้อสอบกับโมเดล

จากการศึกษาอิทธิพลของตัวแปรต่าง ๆ ตามกรอบทฤษฎี CTT และ IRT เมื่อพิจารณาจากการส่งผลของตัวแปรในสมการถดถอยพหุคูณและแผนภาพเส้นทางที่แสดงความสัมพันธ์เชิงสาเหตุพบว่า เมื่อใช้แผนภาพเส้นทางอธิบายความแปรปรวนที่เกิดขึ้นในตัวแปรเกณฑ์ ให้ผลสูงกว่าการอธิบายด้วยสมการถดถอยพหุคูณ และมีค่าสูงถึง .90 ขึ้นไป ทั้งนี้อาจเป็นเพราะตัวแปรอิสระทุกตัวได้จากการประมาณค่ามาจากคะแนนสอบ และค่าส่วนใหญ่เหล่านี้อยู่บนพื้นฐานที่ว่าคะแนนสอบต้องมีความแปรปรวน ดังนั้นจึงมีความสัมพันธ์กันสูง จึงเป็นเหตุทำให้เกิดการส่งผลร่วมกันของตัวแปร (Colinearity) มาก และทำให้ค่าสัมประสิทธิ์การพยากรณ์สูงขึ้นผิดปกติ ซึ่งต่างจากสมการถดถอยที่เลือกตัวแปรอิสระที่สัมพันธ์กันสูงเข้าสมการเพียงตัวแปรเดียว การส่งผลร่วมกันของตัวแปรจึงมีน้อย ทำให้ค่าสัมประสิทธิ์พยากรณ์มีค่าน้อยกว่าการอธิบายด้วยแผนภาพเส้นทาง อย่างไรก็ตามผลการวิจัยครั้งนี้เป็นสิ่งยืนยันเบื้องต้นถึงค่าพารามิเตอร์ และคุณสมบัติ

บางประการของข้อสอบตามกรอบทฤษฎี CTT คือ ความยาก อำนาจจำแนก และความเป็น
 มิตติเดียว มีความสำคัญต่อการพัฒนาคุณภาพแบบทดสอบตามกรอบ CTT และ IRT นอกเหนือจาก
 ความยาก (β) และความเหมาะสมของข้อสอบกับโมเดล และเป็นที่น่าสังเกตอย่างหนึ่งก็คือเมื่อใช้
 ตัวแปรดังกล่าวข้างต้นแล้ว สามารถอธิบายแทนความลำเอียงของข้อสอบเนื่องจากตัวแปรเพศได้
 และจากผลการวิจัยครั้งนี้ แสดงให้เห็นว่าข้อสอบที่มีความยากมีอิทธิพลเชิงบวกต่อผู้สอบที่มีความ
 สามารถสูง และในทางตรงข้ามข้อสอบที่ง่ายน่าจะเหมาะสมในการนำไปใช้กับกลุ่มอ่อนมากกว่า
 การใช้ข้อสอบยาก ๆ แต่อย่างไรก็ตามผลวิจัยแสดงให้เห็นถึงความสัมพันธ์เชิงกลับระหว่าง
 ความยากกับความเป็นมิตติเดียวของข้อสอบ ซึ่งทำให้มองดูคล้ายกับว่าแบบทดสอบที่ยาก ๆ น่าจะ
 วัดในหลายมิติ อีกประการหนึ่งที่น่าสนใจก็คือในกลุ่มความสามารถปานกลาง พบว่าค่าอำนาจ
 จำแนกมีอิทธิพลสูงกว่าตัวแปรอื่นเป็นส่วนใหญ่ ดังนั้นในทางปฏิบัติของการวัดการเลือกข้อสอบไป
 ใช้ตามระดับความสามารถน่าจะเป็นวิธีที่ดีที่สุด ตัวอย่างเช่น ถ้าจะวัดคนเก่งต้องเลือกข้อสอบยาก
 ๆ ถ้าเป็นคนอ่อนก็ควรจะต้องเลือกข้อสอบที่ง่าย แต่ถ้าไม่ทราบความสามารถของผู้สอบและคาดว่า
 ผู้สอบมีความสามารถกระจายในลักษณะโค้งปกติ ความเป็นมิตติเดียว น่าจะเป็นเกณฑ์แรกสุดที่
 ควรพิจารณา รองลงมาคือ ค่าอำนาจจำแนกซึ่งสัมพันธ์สูงกับความยาก ดังนั้นถ้าใช้แบบทดสอบที่
 วัดในมิติเดียว คนเก่งจะต้องทำข้อสอบได้ ซึ่งจะเป็นประโยชน์อย่างมากในการนำไปวินิจฉัยความ
 รอบรู้ของผู้เรียน จากข้อค้นพบดังกล่าวเกี่ยวกับการประยุกต์ใช้การวิเคราะห์ข้อสอบตามกรอบ
 ทฤษฎี CTT เพื่ออธิบายคุณภาพของข้อสอบตามกรอบทฤษฎี IRT นับว่าเป็นการช่วยอธิบายเสริม
 แนวคิดของลอร์ด คร็อกเกอร์ และแอลจินา และซาเมจิมาที่กล่าวถึงความสัมพันธ์ของพารามิเตอร์
 ของข้อสอบ ความเชื่อมั่นและสารสนเทศของแบบทดสอบตามกรอบทฤษฎีทั้งสอง (Lord. 1980 :
 27-30 ; Crocker and Algina. 1986 : 350-352 ; Samejima. 1994 : 229-224 ; ผจจจิต อินทสุวรรณ
 (2528 : 38-39 , 64-76) และจากการศึกษาข้างต้นอาจกล่าวสรุปได้ว่าแนวคิดตามกรอบทฤษฎีของ
 IRT น่าจะเป็นแนวคิดที่พัฒนาต่อไปจากกรอบทฤษฎี CTT ดังนั้นในการพัฒนาแบบทดสอบเลือก
 ตอบ ควรประยุกต์ใช้แนวคิดทางทฤษฎีทั้งสองร่วมกันโดยคัดเลือกข้อสอบตามกรอบทฤษฎี CTT
 เป็นเบื้องต้น และประมาณค่าพารามิเตอร์ของข้อสอบตามกรอบทฤษฎี IRT ซึ่งอาจสรุปเป็นรูปแบบ
 ในการพัฒนาแบบทดสอบเลือกตอบตามกรอบทฤษฎี CTT และ IRT ดังภาพประกอบ 57



ภาพประกอบ 57 รูปแบบการพัฒนาแบบทดสอบ MC ตามกรอบทฤษฎี CTT และ IRT

ข้อเสนอแนะ

1. สำหรับการวิจัยเชิงทฤษฎีต่อไป

1.1 ศึกษาการให้คะแนนตามรูปแบบ 2-SR ในเนื้อหาวิชาอื่นรวมทั้งแบบทดสอบความถนัด หรือแบบทดสอบที่มีการให้คะแนนแบบปกติ (0 - 1) ที่ไม่ได้วัดสติปัญญา

1.2 ศึกษาการให้คะแนนตามรูปแบบ 2-SR ตามทฤษฎี G - theories หรือ IRT แบบ 1P, 2P หรือ 3P และประยุกต์ใช้กับการเทียบคะแนน (Equating) การใช้กับคอมพิวเตอร์เพื่อการสอบ (CAT) และศึกษากับค่า $\theta(\text{Max})$ ของข้อสอบรายข้อแทน θ ที่ระดับความสามารถต่างๆ

1.3 ศึกษาการให้คะแนนตามรูปแบบ 2-SR เปรียบเทียบกับวิธีการให้คะแนนที่ไม่ใช่แบบปกติ (0 - 1) เช่นวิธีของคูมส์ วิธีของอนันต์ วิธีกำหนดน้ำหนัก วิธีกำหนดความเชื่อมั่น วิธีปรับการเดา รวมทั้งการใช้แบบทดสอบ MTF ในประเด็นความเชื่อมั่นตามสูตรต่าง ๆ รวมทั้งตามแนวของ Congeneric ความเที่ยงตรงเชิงสภาพ และเชิงพยากรณ์ เปอร์เซนต์การเดา เจตคติของผู้สอบ และผู้เกี่ยวข้องต่าง ๆ รวมทั้งใน MC ที่มีความยาว และจำนวนตัวเลือกต่างกัน

1.4 ศึกษาทดลอง และตรวจสอบโมเดลสำหรับวิเคราะห์ข้อสอบตามกรอบทฤษฎี CTT และ IRT และทดลองซ้ำกับแบบทดสอบที่ผ่านการวิเคราะห์เพื่อตรวจสอบความคงที่ของคะแนนสอบ และค่าพารามิเตอร์ของข้อสอบ

2. สำหรับการนำไปใช้ในเชิงปฏิบัติ

รูปแบบการให้คะแนนแบบ 2-SR น่าจะเหมาะสมกับการนำไปใช้กับการทดสอบที่ผู้ดำเนินการทดสอบไม่สามารถวิเคราะห์คุณภาพข้อสอบก่อนที่จะนำไปใช้ได้ เช่น การสอบคัดเลือกคนเข้าสู่ตำแหน่ง การสอบแข่งขันต่าง ๆ ตลอดจนการใช้แบบทดสอบในห้องเรียนเพื่อการวินิจฉัยผู้เรียน และตรวจสอบความเที่ยงตรงเชิงพยากรณ์

บรรณานุกรม

บรรณานุกรม

- กาญจนา ศิริวัฒนพงษ์. การศึกษาเปรียบเทียบการตอบและการตรวจให้คะแนนแบบทดสอบ
ปรนัยชนิดเลือกตอบที่มีลักษณะแตกต่างกัน. วิทยานิพนธ์ กศ.ม. กรุงเทพฯ :
มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2520. อัดสำเนา.
- เกริกชัย ฮวบเจริญ. การศึกษาเปรียบเทียบการตอบและการตรวจให้คะแนนแบบใหม่และวิธี
ทดสอบความมั่นใจและแบบธรรมดาที่มีต่อความเชื่อมั่นและคะแนนการเดาของ
แบบทดสอบเลือกตอบ. วิทยานิพนธ์ กศ.ม. กรุงเทพฯ : มหาวิทยาลัย
ศรีนครินทรวิโรฒ ประสานมิตร, 2520. อัดสำเนา.
- คณะกรรมการการประถมศึกษาแห่งชาติ, สำนักงาน, ข้อมูลสารสนเทศทางการศึกษา
ปีการศึกษา 2538. กรุงเทพมหานคร : กองแผนงาน สำนักงานคณะกรรมการ
การประถมศึกษาแห่งชาติ, 2538.
- จักรกฤษณ์ สำราญใจ. ประสิทธิภาพเชิงสัมพัทธ์ของข้อสอบเลือกตอบชนิดเลือกตอบทุกตัว
เลือกเทียบกับข้อสอบเลือกตอบชนิดแบบจับในแบบทดสอบวัดผลสัมฤทธิ์. วิทยา
นิพนธ์ ค.ม. กรุงเทพมหานคร : จุฬาลงกรณ์มหาวิทยาลัย, 2530. ถ่ายเอกสาร.
- จินดา ไตอนันต์. การเทียบคุณลักษณะของแบบทดสอบสัมฤทธิ์ผลแบบเลือกตอบถูกผิดทุกตัว
เลือกกับแบบเลือกตอบตัวเลือกเดียว. วิทยานิพนธ์ ค.ม. กรุงเทพฯ : จุฬาลงกรณ์มหา
วิทยาลัย, 2527. อัดสำเนา.
- ชวลิต โพธิ์นคร. การเปรียบเทียบผลของการกำหนดจุดตัดแบบทดสอบอิงเกณฑ์ชนิดเลือกตอบ
ชนิดเลือกตอบโดยวิธีของเบอร์ก วิธีประยุกต์ราชคโหมเดลและวิธีกำหนดเกณฑ์ผ่าน
ระดับต่ำสุด. วิทยานิพนธ์ ศษ.ม. เชียงใหม่ : มหาวิทยาลัยเชียงใหม่, 2528. อัดสำเนา.
- โชติ เพชรชื่น. “คะแนน,” ในสารานุกรมศึกษาศาสตร์ฉบับเฉลิมพระเกียรติสมเด็จพระ
พระนางเจ้าสิริกิติ์พระบรมราชินีนาถในมหามงคลเฉลิมพระชนมพรรษา 5 รอบ.
หน้า 103-104, กรุงเทพมหานคร : มหาวิทยาลัยศรีนครินทรวิโรฒ, 2535.
- ดร.น. หาญตระกูล และนิโลบล นิ้มกิ่งรัตน์. “การเปรียบเทียบคุณภาพของแบบทดสอบชนิด
เลือกตอบที่มีวิธีการตรวจให้คะแนนวิธีต่างๆ,” ศึกษาสาส์นสาร. 1-2 : 210-214 ;
ตุลาคม 2526 - กันยายน 2527.
- ทวี ทองคำ. การเปรียบเทียบค่าความเที่ยง ความตรง และอำนาจจำแนกของแบบทดสอบชนิด
เลือกตอบ ที่ใช้คำสั่งและวิธีการให้คะแนนต่างกัน. วิทยานิพนธ์ ค.ม.
กรุงเทพมหานคร : จุฬาลงกรณ์มหาวิทยาลัย, 2526. อัดสำเนา.

ธวัชชัย งามสันติวงศ์. SPSS/PC+ SPSS FOR WINDOWS หลักการและวิธีใช้คอมพิวเตอร์

ในงานสถิติเพื่อการวิจัย. พิมพ์ครั้งที่ 3. กรุงเทพมหานคร : บริษัท 21 เซ็นจูรี, 2538.

ธีรศักดิ์ อินทรมาตย์. อิทธิพลของวิธีการตอบแบบทดสอบปรนัยชนิดเลือกตอบวิธีต่างๆที่มีต่อ
ค่าความเชื่อมั่น ค่าความเที่ยงตรง และปริมาณการเดา. ปรินูญานินพนธ์ กศ.ม.

กรุงเทพฯ : มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2520. อัดสำเนา.

นนทนา เผือกผ่อง. “การให้คะแนนเพิ่ม-ลด,” พัฒนาวิถึผล 3: 30-32, 2510.

นภา กาญจนกิจโสภณ. การเปรียบเทียบคะแนนสอบระหว่างการให้น้ำหนักคะแนนรายข้อเท่า
กันตามทฤษฎีมาตรฐานเดิม กับการให้คะแนนรายข้อต่างกันตามทฤษฎีการสนองตอบ
ข้อกระทง. วิทยานิพนธ์ ค.ม. กรุงเทพมหานคร : วิทยานิพนธ์ ค.ม. จุฬาลงกรณ์
มหาวิทยาลัย, 2526. อัดสำเนา.

นิยม ปุราคำ. ทฤษฎีของการสำรวจสถิติจากตัวอย่างและการประยุกต์. กรุงเทพมหานคร :
ศ.ส.การพิมพ์, 2517.

บุญเชิด ภิญญอนันตพงษ์. การประมาณค่าความเชื่อมั่นของแบบทดสอบที่แบ่งส่วนย่อยตาม
แบบจำลองคะแนนจริงสัมพันธ์. ปรินูญานินพนธ์ กษ.ด. กรุงเทพมหานคร :
มหาวิทยาลัยศรีนครินทรวิโรฒ, 2537. อัดสำเนา.

ผจงจิต อินทสุวรรณ. ทฤษฎีการตอบข้อคำถาม. กรุงเทพมหานคร : สถาบันวิจัยพฤติกรรม
ศาสตร์มหาวิทยาลัยศรีนครินทรวิโรฒประสานมิตร, 2528. เอกสารอัดสำเนา 125 หน้า.

พนิชา สวัสดิ์มิงค. การเปรียบเทียบคุณภาพของแบบทดสอบเลือกตอบในการวัดระดับความรู้
ที่ต่างกัน โดยวิธีให้คะแนนต่างกันในกลุ่มผู้สอบที่มีผลสัมฤทธิ์ทางการเรียนต่างกัน.
วิทยานิพนธ์ ค.ม. กรุงเทพมหานคร : จุฬาลงกรณ์มหาวิทยาลัย, 2530. อัดสำเนา.

เพ็ญศรี สว่างเนตร. ความเที่ยงของแบบทดสอบชนิดเลือกตอบด้วยเทคนิคการให้คะแนน
ที่ต่างกัน. วิทยานิพนธ์ ค.ม. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย, 2520. อัดสำเนา.

เพลินพิศ คุณคำ. การเปรียบเทียบความเชื่อมั่น และปริมาณการเดาของแบบทดสอบปรนัย
ชนิดเลือกตอบที่มีวิธีการตอบและการตรวจให้คะแนนวิธี 0-1, วิธีของคูมบัส และวิธี
ของอนันต์. ปรินูญานินพนธ์ กศ.ม. กทม. : มหาวิทยาลัยศรีนครินทรวิโรฒ
ประสานมิตร, 2532. อัดสำเนา.

เยาวลักษณ์ น้อยนาแสง. การเปรียบเทียบคุณภาพของข้อสอบความถนัดทางการเรียนแบบ
ปรนัยชนิดเลือกตอบ ซึ่งเป็นผลมาจากวิธีการตอบและการตรวจให้คะแนนที่แตกต่าง
กัน. ปรินิพนธ์ กศ.ม. กรุงเทพมหานคร : มหาวิทยาลัยศรีนครินทรวิโรฒ
ประสานมิตร, 2535. อัดสำเนา.

ล้วน สายยศ. “แบบทดสอบเลือกตอบ,” ใน สารานุกรมศึกษาศาสตร์ฉบับเฉลิมพระเกียรติ
สมเด็จพระนางเจ้าสิริกิติ์พระบรมราชินีนาถในมหามงคลเฉลิมพระชนมพรรษา 5
รอบ. หน้า 108-110, กรุงเทพมหานคร : มหาวิทยาลัยศรีนครินทรวิโรฒ, 2535.

วัชรภรณ์ จิตรมาศ. การเปรียบเทียบคุณภาพแบบทดสอบเลือกตอบชนิดถูกผิดหลายตัวเลือก
ที่มีจำนวนตัวเลือก และวิธีการตรวจให้คะแนนต่างกัน. ปรินิพนธ์ กศ.ม.
กรุงเทพมหานคร : มหาวิทยาลัยศรีนครินทรวิโรฒ, 2538. ถ่ายเอกสาร.

ศิริชัย กาญจนวาสี. “การวิเคราะห์เส้นทางความสัมพันธ์เชิงสาเหตุระหว่างตัวแปร (PATH
ANALYSIS),” ใน รายงานการประชุมทางวิชาการเสนอผลงานวิจัยการศึกษา ครั้งที่ 2.
หน้า 64-125, กรุงเทพมหานคร : สำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ
กระทรวงศึกษาธิการ, 2536.

สาลิกา เมธนาวิน. การประมาณค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ตามวิธีของฮวน วิธี
ของเบรนนอน และเคน และวิธีของราชูเมื่อกำหนดจุดตัดต่างกัน. ปรินิพนธ์
กศ.ม. กรุงเทพมหานคร : มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2537.
อัดสำเนา.

สำราญ มีแจ้ง. ผลของคำสั่งและการให้คะแนนที่ต่างกันต่อค่าความเที่ยง ความตรง และ
อำนาจจำแนกของแบบทดสอบชนิดเลือกตอบ. วิทยานิพนธ์ ค.ม. กรุงเทพฯ :
จุฬาลงกรณ์มหาวิทยาลัย, 2525. อัดสำเนา.

สำเริง บุญเรืองรัตน์. “การพัฒนาทฤษฎีเลแทนเทรทเพื่อวิเคราะห์ข้อสอบ” วารสารการวัดผล
การศึกษา. 4(2) : 1-20 ; กันยายน - ธันวาคม, 2525.

——— “การวิเคราะห์ลักษณะหลากหลายวิธีหลาย,” ใน สารานุกรมศึกษาศาสตร์ฉบับเฉลิม
พระเกียรติสมเด็จพระนางเจ้าสิริกิติ์พระบรมราชินีนาถในมหามงคลเฉลิมพระชนม
พรรษา 5 รอบ. หน้า 118-122, กรุงเทพมหานคร : มหาวิทยาลัยศรีนครินทรวิโรฒ,
2535.

ลำเรียง บุญเรืองรัตน์. “ความคลาดเคลื่อน,” ใน สารานุกรมศึกษาศาสตร์ฉบับเฉลิมพระเกียรติสมเด็จพระนางเจ้าสิริกิติ์พระบรมราชินีนาถในมหามงคลเฉลิมพระชนมพรรษา 5 รอบ. หน้า 105-107, กรุงเทพมหานคร : มหาวิทยาลัยศรีนครินทรวิโรฒ, 2535.

_____. “ความเชื่อมั่น,” ใน สารานุกรมศึกษาศาสตร์ฉบับเฉลิมพระเกียรติสมเด็จพระนางเจ้าสิริกิติ์พระบรมราชินีนาถในมหามงคลเฉลิมพระชนมพรรษา 5 รอบ. หน้า 126-128, กรุงเทพมหานคร : มหาวิทยาลัยศรีนครินทรวิโรฒ, 2535.

_____. “ความเที่ยงตรง,” ใน สารานุกรมศึกษาศาสตร์ฉบับเฉลิมพระเกียรติสมเด็จพระนางเจ้าสิริกิติ์พระบรมราชินีนาถในมหามงคลเฉลิมพระชนมพรรษา 5 รอบ. หน้า 101-102, กรุงเทพมหานคร : มหาวิทยาลัยศรีนครินทรวิโรฒ, 2535.

_____. “ความเที่ยงตรงข้ามกลุ่ม,” ใน สารานุกรมศึกษาศาสตร์ฉบับเฉลิมพระเกียรติสมเด็จพระนางเจ้าสิริกิติ์พระบรมราชินีนาถในมหามงคลเฉลิมพระชนมพรรษา 5 รอบ. หน้า 124-125, กรุงเทพมหานคร : มหาวิทยาลัยศรีนครินทรวิโรฒ, 2535.

_____. “ความเที่ยงตรงที่ลดลง,” ใน สารานุกรมศึกษาศาสตร์ฉบับเฉลิมพระเกียรติสมเด็จพระนางเจ้าสิริกิติ์พระบรมราชินีนาถในมหามงคลเฉลิมพระชนมพรรษา 5 รอบ. หน้า 116-117, กรุงเทพมหานคร : มหาวิทยาลัยศรีนครินทรวิโรฒ, 2535.

_____. “แบบทดสอบคู่ขนาน,” ใน สารานุกรมศึกษาศาสตร์ฉบับเฉลิมพระเกียรติสมเด็จพระนางเจ้าสิริกิติ์พระบรมราชินีนาถในมหามงคลเฉลิมพระชนมพรรษา 5 รอบ. หน้า 131-132, กรุงเทพมหานคร : มหาวิทยาลัยศรีนครินทรวิโรฒ, 2535.

_____. “การวิเคราะห์ลักษณะหลากหลาย - วิธีหลาย,” ใน สารานุกรมศึกษาศาสตร์ฉบับเฉลิมพระเกียรติสมเด็จพระนางเจ้าสิริกิติ์พระบรมราชินีนาถในมหามงคลเฉลิมพระชนมพรรษา 5 รอบ. หน้า 118-122, กรุงเทพมหานคร : มหาวิทยาลัยศรีนครินทรวิโรฒ, 2535.

_____. “ความคลาดเคลื่อน,” ใน สารานุกรมศึกษาศาสตร์ฉบับเฉลิมพระเกียรติสมเด็จพระนางเจ้าสิริกิติ์พระบรมราชินีนาถในมหามงคลเฉลิมพระชนมพรรษา 5 รอบ. หน้า 105-107 กรุงเทพมหานคร : มหาวิทยาลัยศรีนครินทรวิโรฒ, 2535.

สุพัฒน์ สุกมลสันต์. การเปรียบเทียบผลของวิธีการให้คะแนนต่อความเที่ยง ความตรงและความคงที่ของแบบสอบการอ่านเข้าใจความภาษาอังกฤษที่มีโครงสร้างความรู้ที่ต่างกัน. วิทยานิพนธ์ ค.ม. กรุงเทพมหานคร : จุฬาลงกรณ์มหาวิทยาลัย, 2528. อัดสำเนา.

- สุรชัย มีชาญ. การเปรียบเทียบความยากประจำชั้น และประสิทธิภาพ ในการประมาณค่า
เจตคติของมาจรวัดเจตคติแบบลิเคิร์ทที่มีรูปแบบของการตอบ จำนวนของลำดับขั้น
และทิศทางของการเรียงลำดับขั้นแตกต่างกันด้วยพาร์เซี่ยล เครดิต โมเดลของ
มาสเตอร์. ปรินญานิพนธ์ กศ.ด. กรุงเทพมหานคร : มหาวิทยาลัยศรีนครินทรวิโรฒ
ประสานมิตร, 2539. ถ่ายเอกสาร.
- อนันต์ ศรีโสภ. “การทดลองรูปแบบใหม่ของการสอบแบบเลือกตอบ,” การประชุมทางวิชาการ
เพื่อเสนอผลงานวิจัย ปี 2533. 134-152 ; 6 - 8 สิงหาคม 2533.
- อรรณณ ดันเจริญรัตน์. การศึกษาวิธีการตอบและการตรวจให้คะแนนแก่ข้อสอบปรนัยแบบ
เลือกตอบ. ปรินญานิพนธ์ กศ.ม. กรุงเทพฯ : มหาวิทยาลัยศรีนครินทรวิโรฒ
ประสานมิตร, 2517. จัดสำเนา.
- องอาจ นัยพัฒน์. การศึกษาประสิทธิภาพของผลการประมาณค่าพารามิเตอร์ของพาร์เซี่ยล
เครดิต โมเดล (CREDIT) ระหว่างวิธีพรีออกซ์ (PROX) และวิธียูคอน (UCON)
ในแบบทดสอบวัดเจตคติแบบลิเคิร์ทสเกล. ปรินญานิพนธ์ กศ.ม. กรุงเทพฯ :
มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2535. ถ่ายเอกสาร.
- Abu-Sayf, F.X. “Relative Effectiveness of the Conventional Fomular Score,” Journal of
Education Research 69(4) : 160-162 ; December, 1975.
- Abu-Sayf, F.X. and J.J. Daimond. “Effect of Confidence Level in Multiple-Choice Test Answer
on Reliability and Validity of Score, “ Journal of Educational Research. 70 : 62-65 ;
November-December, 1976.
- Adam, R.J. “Item bias, “ in Educational Research Methology, and Measurement : An
International Handbook. edited by Keeves. p.398-403. Oxford : Pergamon Press plc,
1988.
- Airasian, P.W. Classroom Assesment. 2nd ed. Singapore : McGraw-Hill, Inc., 1994.
- Albanese, Mark A. and Darell L. Sabers. “Multiple True-False Item : A Study of Interitem
Correlation, Scoring Alternatives and Reliability Estimation, “Journal of Educational
Measurement. 25(2) : 111-123 ; Summer, 1988.
- Albanese, Mark.A. , T.H. Kent and D.R. Whitney. “Cluing in the Multiple-Choice Test Items
with Combinations of Correct Response, “ Journal of Medical Education. 54 : 948-
950 ; December, 1979.

- Albert, W.G. "Scoring for Partial Knowledge in Mathematics Testing ; A study of Modification and Extension of Multiple-Choice Items Applied to the Testing of Achievement in Mathematics, " Dissertation Abstracts International. 31(4) : 1619A-1620A ; October, 1970.
- Allal, L. "Generalizability Theory, " in Educational Research, Methodology, and Measurement An International Handbook. edited by Keeves. p.272-277. Oxford : Pergamon Press plc, 1988.
- Allal, L. "Generalizability Theory, " in The International Encyclopedia of Education. 2nd ed. edited by Husen and Postlethwaite. p. 2458-2462. Oxford : Pergamon Press plc, 1994.
- Anastasi, A. Psychological testing. 6nd ed. New York : Macmillan Publishing Company, 1990.
- Andrich, D "Rating Scale Analysis," in Educational Research, Methodology, and Measurement : An International Handbook. edited by Keeves. p.297-302. Oxford : Pergamon Press plc, 1988.
- Badbard, D.A. and S.B. Green. "Use of the Coombs Elimination Procedure in Classroom Tests, " The Journal of Experimental Education. 54(2) : 68-72 ; Winter, 1985/1986.
- Baker, I.L. "Domain-referenced Tests," in Educational Research Methodology, and Measurement : An International Handbook. edited by Keeves. p.370-372. Oxford : Pergamon Press plc, 1988.
- Berk R.A.. "Criterion-referenced Test," in Educational Research Methodology, and Measurement An International Handbook. edited by Keeves. p.365-370. Oxford : Pergamon Press plc, 1988.
- Birnbaum, A. "Some Latent Trait Models and Their Use in Inferring an Examinee's Ability," in Statistical Theories of Mental Test Scores. edited by Lord, F.M. and Novick, M.R. p. 397-550. Massachusetts : Addison-Wesley Publishing Company, 1968.
- Brennan, R.E. Elements of Generalizability Theory. American College Testing Program, Iowa, 1983.
- Cangelosi, J.S. Designing tests for evaluating students achievement. New York : Longman, 1990.

- Choppin B.H. "Correction for Guessing," in Educational Research Methodology, and Measurement : An International Handbook. edited by Keeves. p.384-386. Oxford : Pergamon Press plc, 1988.
- Choppin B.H. "Objective Tests," in Educational Research, Methodology, and Measurement An International Handbook edited by Keeves. p.354-358. Oxford : Pergamon Press plc, 1988.
- Cole, N.S. and P.A. Moss. "Bias in Test Use," in Educational Measurement. 3rd ed. edited by Linn. p. 201-220. New York : Macmillan Publishing Company , 1989.
- Coombs, C.H. "On the use of Objective Elimination, " Educational and Psychological Measurement. 13 : 308-310 ; Summer, 1953.
- Coombs, C.H., J.E. Milholland, and F.B. Womer. "The Assessment of Partial Knowledge," Educational and Psychological Measurement. 16(1) : 13-37 ; Spring, 1956.
- Crocker, L. and J. Algina. Introduction to Classical and Modern Test Theory. Florida : Rinehart and Winston Inc., 1986.
- Cronbach, L.J. "An Experimental Comparison of Multiple True-False and Multiple-Choice test, " The Journal of Educational Psychology. 32 : 533-543, 1941.
- Davis, F.B., and Gordon Fifer. "The Effect on the Test Reliability and Validity of Scoring Aptitude and Achievement Tests with Weight for Every Choice " , Educational and Psychological Measurement. 19 : 159-170 ; Summer, 1959.
- Dayton, C.M. The Design of Educational Experiments. USA : McGraw-Hill, Inc., 1970.
- Dodd, B.G. and W.R. Koch. "Effects of Variations in Item Step Values on Item and Test Information in The Partial Credit Model, " Applied Psychological Measurement. 11 (41) : 371-384 ; December, 1987.
- Douglas, G. "Latent Trait Measurement Model," in Educational Research Methodology, and Measurement : An International Handbook. edited by Keeves. p.382-386. Oxford : Pergamon Press plc, 1988.
- Dressel, P.L. and J. Schmid. "Some Modification of the Multiple-Choice item". Education and Psychological Measurement. 13 : 579-595 : 1953 .

- Ebel, R.L. "The Effectiveness of Multiple True-False Test Items," Education and Psychological Measurement. 13 : 579-595 : 1953.
- Ebel, R.L. and David, A.F. Essentials of Educational Measurement. 4th ed. New Jersey : Prentice-Hall, Inc., 1986.
- Echternacht, Cary J. "The use of confidence Testing in objective tests", Review of Educational Research. 42 : 217-234 ; 1972.
- Essex,D.L. "A Comparison of Two Item-Scoring Procedures and Student Reaction to Them," Journal of Medical Education. 51 : 565-572 ; July, 1976.
- Fan, C.T. Item Analysis Table. New Jersey : Educational Testing Service Princeton, 1952.
- Feldt, L.S. and R.L. Brennan. "Reliability," in Educational Measurement. 3rd ed. edited by Linn. p. 105-146. New York : Macmillan Publishing Company, 1989.
- Frisbie, David A. and Daryl C. Sweeney. "The Relative Merits of Multiple True-False achievement Test, " Journal of Education Measurement. 19(1) : 29-35 ; Spring, 1982.
- Frederiksen, N. and others. Test Theory for a New Generation of Tests. New Jersey : Lawrence Erlbaum Associates, Inc., 1993.
- Friedenberg, L. Psychological Testing : Design, Analysis and Use. Massachusetts : Allyn & Bacon A Simon & Schuster Company, 1995.
- Frisbie,D.A. and D.C. Sweeney, "The Relative Merits of Multiple True-False Achievement Tests, " Journal of Educational Measurement. 19 : 29-35 ; Spring, 1982.
- Fu-Ju Tsai and Hoi K. Suen. "A Brief Report on a Comparison of Six Scoring Methods for Multiple True-False Items, " Educational and Psychological Measurement. 53 : 399-404 ; 1993.
- Gay, L.R. Educational Evaluation and Measurement. 2nd ed. Singapore : Macmillan Publishing Company, 1991.
- Good,C.V. Dictionary of Education. 3rd ed. New York : Mc Graw-Hill., 1973.
- Greene, E.B. Measurement of Human Behavior. New York : Odyssey Press, 1952.
- Gronlund,N.E. Constructing Achievement Tests. 3rd ed. New Jersey : Prentice-Hall , 1982.

- Gronlund, N.E. How to Make Achievement Tests and Assessments. 4th ed. Massachusetts : Allyn and Bacon A Division of Simon & Schuster, Inc., 1993.
- Gronlund, N.E. and R.L. Linn. Measurement and Evaluation in Teaching. 6th ed. Singapore : Mcmillan Publishing Company, 1990.
- Gulliksen, H. Theory of Mental Tests. New York : John Wiley and Sons, Inc., 1950.
- Haladyna, T.M. Developing and Validating Multiple-Choice Test Items. New Jersey : Lawrence Erlbaum Associates, Publishers., 1994.
- Hambleton, R.K. "Criterion-referenced Measurement," in Educational Research, Methodology, and Measurement : An International Handbook. edited by Keeves. p.277-281. Oxford : Pergamon Press plc, 1988.
- _____. "Principles and Selected Applications of Item Response Theory," in Educational Measurement. 3rd ed. edited by Linn. p. 147-200. New York : Macmillan Publishing Company, 1989.
- Hambleton, R.K., H. Swaminathan, and H.J. Rogers. Fundamental of Item Response Theory. California : SAGE Publication., 1991.
- Hambleton, R.K. and H. Swaminathan. Item Response Theory Principles and Applications. Boston : Kluwer-Nijhoff Publishing, 1985.
- Herman J.L. "Item Writing Techniques," in Educational Research, Methodology, and Measurement : An International Handbook. edited by Keeves. p.358-363. Oxford : Pergamon Press plc, 1988.
- Hoy, C. and Gregg, N. Assessment the Special Educator's Role. California : Wadsworth, 1994.
- Hsu, T., Moss, P.A., and Choosak Khampalikit. "The Merits of Multiple Answer Items as Evaluated by Using Six Scoring Formulas." Journal of Experimental Education 52 (Spring 1984) : 152-158.
- Jordan, A.M. Measurement in Education, an introduction. 1st ed. New York : McGraw-Hill, 1953.
- Keats, J.A. "Classical Test Theory, " in The International Encyclopedia of Education. 2nd ed. edited by Husen and Postlethwaite. p. 3996-4000. Oxford : Pergamon Press plc, 1994.

- Keats, J.A. "Measurement in Educational Research, " in Educational Research, Methodology, and Measurement : An International Handbook. edited by Keeves. p.253-258. Oxford : Pergamon Press plc, 1988.
- Keeves, J.P. "Path Analysis," in Educational Research, Methodology, and Measurement : An International Handbook. p.723-730. Oxford : Pergamon Press plc, 1988.
- _____. "Sex differences in ability and achievement," in Educational Research, Methodology, and Measurement : An International Handbook. p.482-483. Oxford : Pergamon Press plc, 1988.
- _____. "Sealing Achivement Test Scores," in Educational Research, Methodology, and Measurement : An International Handbook. p.403-418. Oxford : Pergamon Press plc, 1988.
- _____. "The Improvement of Measurement for Education Research," in Educational Research, Methodology, and Measurement : An International Handbook. p.241-246. Oxford : Pergamon Press plc, 1988.
- Kline, P. The Handbook of Psychological Testing. London : Routledge, 1993.
- Kubiszyn, T. and Borich, G. Educational Testing and Measurement : Classroom Application and Practice. 4th ed. New York : Harper Collins College Publishers. 1993.
- Lansheere, V.D. "Taxonomies of Educational Objectives," in Educational Research, Methodology, and Measurement : An International Handbook. edited by Keeves. p.345-346. Oxford : Pergamon Press plc, 1988.
- Livingston, S.A. "Reliability of Test Results," in Educational Research, Methodology, and Measurement : An International Handbook. edited by Keeves.p.386-392. Oxford : Pergamon Press plc, 1988.
- Lord,F.M. Application of Item Response Theory to Practical Testing Problems. N.J. : Hillsdale, 1980.
- Lord, F.M. "Formula Scoring and Number-Right Scoring, " Journal of Educational Measurement. 12 : 7-11 ; 1975.
- Lord, F.M. and M.R. Novick, Statistical Theories of Mental Test Scores. Massachusetts : Addison-Wesley Publishing Company, 1968.

- Lord, F.M. and Stocking, M.L. "Item Response Theory," in Educational Research, Methodology, and Measurement : An International Handbook. edited by Keeves. p.2692-272. Oxford : Pergamon Press plc, 1988.
- March, H.W. "Multitrait-Multimethod Analysis," in Educational Research, Methodology, and Measurement : An International Handbook. edited by Keeves. p.570-580. Oxford : Pergamon Press plc, 1988.
- Mehrens, W.A. and I.J. Lehman. Measurement and Evaluation in education and psychology. 3rd ed. New York : Holt, Rinehart and Winston, 1984.
- Messick, S. "Validity," in Educational Measurement. 3rd ed edited by Linn p. 1-104. New York : Macmillan Publishing Company, 1989.
- Millman, J and Janifer Greene. "The Specification and Development of Tests of Achievement and Ability. in Educational Measurement. 3rd ed. edited by Linn p. 335-364. New York ; Macmillan Publishing Company, 1989.
- Murphy, K.R. and C.O. Davidshofer. Psychological Testing : principles and applications. New Jersey: Prentice-Hall International Edition, 1988.
- Nunnally, J.C. and I.H. Bernstein. Psychometric Theory. 3rd ed. N.Y. : McGraw-Hill, Inc., 1994.
- Oosterhof, A. Classroom Applications of Educational Measurement. 2nd ed. New York : Macmillan College Publishing Company, 1994.
- Patnaik, D. and R.E. Traub. "Differential Weighting by Judged Degree of Correctness, " Journal of Educational Measurement. 10 : 281-285 ; Winter , 1973.
- Pugh, R.C. and J. Brunza. "Effects of a Confidence Weighted Scoring System on Measures of Test Reliability and Validity, " Educational and Psychological Measurement. 35 : 73-78 ; Spring, 1975.
- Rasch, G. Probabilistic Models for Some Intelligence and Attainment Tests. Chicago : University of Chicago Press, 1980.
- Rippy, R.H. "A comparison of five different scoring functions for confidence tests, " Journal of Educational Measurement. (7) , 165-170 , 1970.

- Rogers, H. J. . "Multiple Choice Tests , Guessin in, " in The International Encyclopedia of Education. 2nd ed. _edited by Husen and Postlethwaite. p. 3996-4000. Oxford : Pergamon Press plc, 1994.
- Ross, K.N. "Sampling, " in Educational Research, Methodology, and Measurement : An International Handbook. edited by Keeves.p. 527-536. Oxford : Pergamon Press plc, 1988.
- Scheuneman J.D. and C.A. Bleistein. "Item bias," in The International Encyclopedia of Education. 2nd ed. edited by Husen and Postlethwaite. p. 3043. Oxford : Pergamon Press plc, 1994.
- Shavelson, R.J. and N.M. Webb. Generalizability. New Delhi : SAGE Publications :1991.
- Snow,R.E. and D.F. Lohman. "Implications of Cognitive Psychology for Educational Measurement, " in Educational Measurement. 3rd ed. edited by Linn p. 263-321 ; Newyork : Macmillan Publishing Company, 1989.
- Spearritt, D. "Factor Analysis," in Educational Research Methodology, and Measurement : An International Handbook. edited by Keeves. p.644-654. Oxford : Pergamon Press plc, 1988.
- Swaminathan, H. "Latent Trait Measurement Models, " in The International Encyclopedia of Education. 2nd ed. edited by Husen and Postlethwaite. p. 3247-3253. Oxford : Pergamon Press plc, 1994.
- Tabachnick, B.G. Using Multivariate Statistics. New York : Harper & Row, Publishers, 1983.
- Tatsuoka, M.M. "Regression Analysis," in Educational Research Methodology, and Measurement : An International Handbook. edited by Keeves.p.737-745. Oxford : Pergamon Press plc, 1988.
- Thorndike, R.M. "Correlational Procedures," in Educational Research Methodology, and Measurement : An International Handbook. edited by Keeves.p.613-620. Oxford : Pergamon Press plc, 1988.
- Thorndike, R.L. "Reliability, " in Educational Research, Methodology and Measurement : An International Handbook. edited by Keeves. p.330-343. Oxford : Pergamon Press plc, 1988.

- Title, C.K. "Test Bias," in Educational Research, Methodology and Measurement : An International Handbook. edited by Keeves. p.392-398. Oxford : Pergamon Press plc, 1988.
- Venn, J. Assessment of Students with Special Needs. USA : Macmillan Publishing, 1994.
- Vincent, D. "Norm-referenced Assessment," in Educational Research Methodology, and Measurement : An International Handbook. edited by Keeves.p.363-365. Oxford : Pergamon Press plc, 1988.
- Wen, Shih-Sung. "The Relationship between Verbal-Meaning Test Score and Degrees on Confidence in Item Responses, " Journal of Educational Measurement. 12 : 197-199 ;Fall, 1975.
- Wilcox, R.R. "True-score Models," in Educational Research Methodology, and Measurement: An International Handbook. edited by Keeves. p.267-269. Oxford : Pergamon Press plc, 1988.
- Wood, R. "Item Analysis," in Educational Research, Methodology and Measurement : An International Handbook. edited by Keeves. p.376-383. Oxford : Pergamon Press plc, 1988.
- Wright, B.D. "Rasch Measurement Models," in Educational Research, Methodology, and Measurement : An International Handbook. edited by Keeves.p.286-292. Oxford : Pergamon Press plc, 1988.
- Wright, B.D. and Stone, M.H. Best Test Design. Chicago : MESA PRESS, 1979.
- Yamane, Taro. Statistical and Introduction Analysis. 3th ed., New York : Harper & Raw , Publication , Inc , 1973.
- Zeller, R.A. "Validity," in Educational Research, Methodology, and Measurement : An International Handbook. edited by Keeves. p. 322-330. Oxford : Pergamon Press plc, 1988.
- Zeller, R.A. "Validity, " in The International Encyclopedia of Education. 2nd ed. edited by Husen and Postlethwaite. p. 6569-6575. Oxford : Pergamon Press plc, 1994.

ภาคผนวก

ภาคผนวก ก

**คณะกรรมการสร้าง และตรวจสอบความเที่ยงตรงเชิงสภาพและเชิงโครงสร้าง
ของแบบทดสอบ**

คณะกรรมการสร้างแบบทดสอบวัดความสามารถในการอ่านจับใจความ

1. นายธรรมรัตน์ วัฒนันต์	ศึกษานิเทศก์ 8 สปจ.นครปฐม	ประธานอนุกรรมการ
2. นางวัฒนา ไรยแสง	ศึกษานิเทศก์ 7 สปอ.เมืองนครปฐม	อนุกรรมการ
3. นายไพฑูรย์ ปลอดอ่อน	ศึกษานิเทศก์ 6 สปอ.กำแพงแสน	อนุกรรมการ
4. นางพรณี ดุจตา	ศึกษานิเทศก์ 6 สปอ.ดอนตูม	อนุกรรมการ
5. น.ส.พัชรี ยันตรีสิงห์	ศึกษานิเทศก์ 6 สปอ.นครชัยศรี	อนุกรรมการ
6. นางละมัย สว่างเนตร	ศึกษานิเทศก์ 6 สปอ.บางเลน	อนุกรรมการ
7. นายศรเพชร มีศรี	ศึกษานิเทศก์ 6 สปอ.สามพราน	อนุกรรมการ
8. นางกนกรัตน์ สุพรรณอ่วม	ศึกษานิเทศก์ 6 สปก.พุทธมณฑล	อนุกรรมการ
9. น.ส.สุสดี จิระวัฒนกิจ	ศึกษานิเทศก์ 6 สปจ.นครปฐม	อนุกรรมการ
10. น.ส.สุจิตรา ศรีนวล	ศึกษานิเทศก์ 6 สปจ.นครปฐม	อนุกรรมการและเลขานุการ

คณะกรรมการสร้างเครื่องมือวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์

1. นางนิตยา จรูญผลนิต	หัวหน้าหน่วยศึกษานิเทศก์ สป.กทม.	ประธานอนุกรรมการ
2. น.ส.เหมวดี กาญจนสกุล	หัวหน้าฝ่ายวิจัยและประเมินผล	ประธานอนุกรรมการ
3. น.ส.วันดี วุฒิชัย	ศึกษานิเทศก์ 6 สป.กทม.	อนุกรรมการ
4. นางบุญทวี แสงศักดิ์	อาจารย์ 3 โรงเรียนสายน้ำทิพย์	อนุกรรมการ
5. นายคณิต เพชรปัญญา	อาจารย์ 3 โรงเรียนทุ่งมหาเมฆ	อนุกรรมการ
6. น.ส.วัฒนา สุนทรกมล	อาจารย์ 2 โรงเรียนราชวินิต	อนุกรรมการ
7. นางวรรณเปรม ไมตรีจิตต์	อาจารย์ 2 โรงเรียนวัดใหม่ผดุงเขต	อนุกรรมการ
8. น.ส.วรรณา พรประยูทธ	อาจารย์ 2 โรงเรียนพระตำหนักสวนกุหลาบ	อนุกรรมการ
9. นายนิรันดร์ ตันหทัย	อาจารย์ 2 โรงเรียนวัดหงส์รัตนาราม	อนุกรรมการ
10. น.ส.พิสมัย พรหมพยัคฆ์	อาจารย์ 2 โรงเรียนดาราคาม	อนุกรรมการ
11. นายปรีชา อรุณสวัสดิ์	ศึกษานิเทศก์ 6 สป.กทม.	อนุกรรมการและเลขานุการ
12. นายนพดล จุลเจือวงศ์	ศึกษานิเทศก์ 6 สป.กทม.	อนุกรรมการและผู้ช่วยเลขานุการ

คณะกรรมการตรวจสอบความเที่ยงตรงของเนื้อหาและเชิงโครงสร้างแบบสอบถาม

1. นายธีรพงษ์ ทวีวรรณ	หัวหน้าหน่วยศึกษานิเทศก์ สปจ.นนทบุรี	ประธานอนุกรรมการ
2. นางภิรมย์ ปริญญาพันธ์	หัวหน้าฝ่ายวิจัยและประเมินผล สปจ.นนทบุรี	รองประธานอนุกรรมการ
3. นายธรรมรัตน์ รัมนนธ์	ศึกษานิเทศก์ 8 สปจ.นครปฐม	อนุกรรมการ
4. ผศ.ปิ๋นวดี ธนธานี	หัวหน้าภาควิชาทดสอบและวิจัยทางการศึกษา สถาบันราชภัฏนครปฐม	อนุกรรมการ
5. น.ส.สุนีย์ ยวนแหล	อาจารย์ภาควิชาทดสอบและวิจัยทางการศึกษา สถาบันราชภัฏนครปฐม	อนุกรรมการ
6. น.ส.สุจิตรา ศรีนวล	ศึกษานิเทศก์ 6 สปจ.นครปฐม	อนุกรรมการ
7. นางละมัย สว่างเนตร	หัวหน้าฝ่ายวิจัยและประเมินผล สป.สมุทรสาคร	อนุกรรมการ
8. นายศรเพชร มีศรี	หัวหน้าฝ่ายวิจัยและประเมินผล สป.กทม.	อนุกรรมการ
9. นายปรีชา	ศึกษานิเทศก์ สป.กทม.	อนุกรรมการ
10. น.ส.วันดี วุฒิโภาค	ศึกษานิเทศก์ สป.กทม.	อนุกรรมการ
11. นางเพชรวดี จงประดับเกียรติ	หัวหน้าภาควิชาทดสอบและวิจัยทางการศึกษา สถาบันราชภัฏบ้านสมเด็จเจ้าพระยา	อนุกรรมการ
12. ผศ.สุภรณ์ ลิ้มบริบูรณ์	อาจารย์ภาควิชาทดสอบและวิจัยทางการศึกษา สถาบันราชภัฏบ้านสมเด็จเจ้าพระยา	อนุกรรมการ
13. นางจวี สังเนตร	ศึกษานิเทศก์ สปจ.สมุทรปราการ	อนุกรรมการ
14. นางจวีวรรณ สุขสุนัย	หัวหน้าฝ่ายวิจัยและประเมินผล สปจ.ปทุมธานี	อนุกรรมการ
15. น.ส.กฤษณี วนะภูติ	ศึกษานิเทศก์ สปจ.ปทุมธานี	อนุกรรมการ
16. น.ส.บุพพัทธ์ ติพย์านนท์	ศึกษานิเทศก์ สปจ.ปทุมธานี	อนุกรรมการ
17. หัวหน้าภาควิชาทดสอบและวิจัยทางการศึกษา	สถาบันราชภัฏพระนคร	อนุกรรมการ
18. นายเอนก ล่วงลือ	ศึกษานิเทศก์ สปจ.นนทบุรี	อนุกรรมการและเลขานุการ
19. น.ส.วนิดา ทวีวรรณ	อาจารย์ 1 ช่วยราชการ สปจ.นนทบุรี	อนุกรรมการและผู้ช่วยเลขานุการ

ภาคผนวก ข

เครื่องมือที่ใช้ในการวิจัย

* * ถ้านักเรียนคิดว่าข้อ ก และข้อ ข. เป็นคำตอบถูกและคิดว่า ข เป็นคำตอบถูกที่สุด ให้ตอบ
ดังนี้

ข้อ	ส่วนที่ 1	ส่วนที่ 2
00.	X X ค ง ☐ ☐	ก X ค ง ☐ ☐

* * * ถ้านักเรียนคิดว่าข้อ ก ข และ ง เป็นคำตอบถูกและคิดว่า ข เป็นคำตอบถูกที่สุด
ให้ตอบดังนี้

ข้อ	ส่วนที่ 1	ส่วนที่ 2
00.	X X ค X ☐ ☐	ก X ค ง ☐ ☐

* * * * นักเรียนไม่ต้องทำเครื่องหมายใดๆลงใน ☐ ที่อยู่ท้ายส่วนที่ 1 และส่วนที่ 2

* * * * * ถ้านักเรียนกาทุกตัวเลือกในส่วนที่ 1 ดังตัวอย่างข้างล่างจะได้คะแนน 0

ข้อ	ส่วนที่ 1	ส่วนที่ 2
00.	X X X X ☐ ☐	ก X ค ง ☐ ☐

* * * * * และถ้านักเรียนกาเครื่องหมายในส่วนที่ 2 ไม่สัมพันธ์กับส่วนที่ 1 ดังตัวอย่างข้างล่าง
จะได้คะแนน 0 เช่นกัน

ข้อ	ส่วนที่ 1	ส่วนที่ 2
00.	X ข ค ง ☐ ☐	ก X ค ง ☐ ☐

หมายเหตุ♥♥

ผลการตอบของนักเรียนทุกคนจะเป็นประโยชน์อย่างสูงต่อการศึกษาวิจัยเพื่อพัฒนาคุณภาพของ
แบบทดสอบเลือกตอบในครั้งนี้ ฉะนั้นจึงขอความร่วมมือจากนักเรียนให้ช่วยตอบข้อสอบชุดนี้อย่าง
เต็มความรู้ความสามารถ และขอรับรองว่าผลการตอบครั้งนี้จะไม่นำไปพิจารณาตัดสินผลการ
เรียนในชั้นเรียนของนักเรียนแต่อย่างใด

ขอขอบคุณที่ให้ความร่วมมือเป็นอย่างดี

ผู้วิจัย

อ่านข้อความข้างล่างแล้วตอบคำถามข้อ 1-6

โรงเรียนเรณีนาคีเรือใหญ่	ลูกเรือไซร้คูงหมู่ผู้รักษา
ครูอาจารย์ท่านคล้ายนายนา	ควบคุมพาเพื่ออุณินอุณ
การจะไปให้คลอครอคผิงได้	ด้วยร่วมใจกันจริงทุกสิ่งสม
ทำหน้าที่ทุกทางอย่างเกียวกมล	ผลปรารมภ์เรือไปได้คั้งปอง

1. ผู้เขียนต้องการให้ข้อคิดแก่ใคร ก. นักเรียน ข. ครู-อาจารย์ ค. เจ้าของโรงเรียน ง. ทุกคนในโรงเรียน	4. โรงเรียนเรณีนาคีเรือใหญ่ หมายความว่าอย่างไร ก. โรงเรียนไม่ใช่เรือดำใหญ่ ข. โรงเรียนเป็นเรือใหญ่ลำหนึ่ง ค. โรงเรียนเปรียบได้ดั่งเรือดำใหญ่ ง. โรงเรียนมีลักษณะคล้ายเรือดำใหญ่
2. "นายนา" ในข้อความนี้หมายถึงใคร ก. ครู-อาจารย์ ข. เจ้านาย ค. เจ้าของเรือ ง. ผู้บังคับบัญชา	5. การปฏิบัติของใครสอดคล้องกับคำประพันธ์นี้ ก. แดงช่วยเพื่อนพาเรือ ข. ชาวทำงานช่วยครอบครัว ค. เขียวคังใจเรียนและเชื่อฟังครู ง. ดำร่วมมือกับครูช่วยงานในโรงเรียน
3. ใจความสำคัญของคำประพันธ์นี้กล่าวถึงเรื่องใด ก. การเดินเรือ ข. การรู้จักเอาตัวรอด ค. การศึกษาในโรงเรียน ง. การร่วมมือกันทำหน้าที่	6. ข้อความตามข้อใด ไม่ใช่ การเปรียบเทียบ ก. ลูกเรือไซร้คูงหมู่ผู้ศึกษา ข. โรงเรียนเรณีนาคีเรือใหญ่ ค. ครู-อาจารย์ท่านคล้ายนายนา ง. ทำหน้าที่ทุกทางอย่างเกียวกมล

อ่านข้อความข้างล่างแล้วตอบคำถามข้อ 7-14

บ่อยครั้งที่พวกเขาารู้สึกสนุกสนานเมื่อได้เห็นผู้ที่เดินผ่าน รวมทั้ง ที่ใช้จักรยานยนต์เป็นพาหนะ ต้องสะดุ้งตกใจกับเสียงและประกายไฟที่พวกเขาตั้งใจให้เกิดขึ้นบนถนน ทุกปีมีคนเจ็บจากอุบัติเหตุจักรยานยนต์เสียหลักล้มเกิดขึ้น มากบ้างน้อยบ้าง แม้ว่าโดยมากจะเป็นเพียงการบาดเจ็บเล็กน้อย ๆ ก็ตาม

ในระหว่างที่คนกลุ่มหนึ่งสนุกสนานและมันอยู่กับการเล่น “วัดระเบิด” ที่วางขายถูกต้องตามกฎหมายอยู่ที่นี่ คนอีกกลุ่มหนึ่งอาจให้ความสนใจ ชื่นชมกับประกายแสงจากดอกไม้ที่เจิดจ้าอยู่ในบรรยากาศคึกคัก แต่อีกหลายคนได้ยืนเฉยพละและมองประกายไฟเหล่านั้นด้วยความรู้สึกที่ต่างออกไป บ้างต้องสะดุ้งโหยงกับเสียงที่ดังราวระเบิดขนาดย่อม ๆ บ้างคอยระแวงระวังกับแสงจากประกายไฟที่อาจวิ่งลงมาโดนเอาวัตถุติดไฟได้ทุกเมื่อ ดังที่เคยมีตัวอย่างปรากฏมาแล้ว

7. “พวกเขา” ในข้อความนี้ น่าจะหมายถึงใคร ก. เด็ก ๆ วัยรุ่น ข. คนขายพลุ ค. คนเล่นวัดระเบิด ง. คนขับรถจักรยานยนต์	10. ข้อความนี้น่าจะคัดลอกมาจากงานเขียนประเภทใด ก. ข่าว ข. บทความ ค. บทวิจารณ์ ง. บทสนทนา
8. “ความสนุกสนานของพวกเขา” ตามข้อความนี้คือข้อใด ก. การทำให้คนอื่นบาดเจ็บจากอุบัติเหตุ ข. การได้ขว้างปาวัตถุระเบิดใส่คนเดินทาง ค. การได้เห็นคนตกใจกับเสียงและประกายไฟ ง. การได้เล่นของที่มีเสียงดังและมีประกายไฟ	11. เหตุการณ์ในข้อความนี้น่าจะเกิดในเทศกาลใด ก. สงกรานต์ ข. ลอยกระทง ค. เข้าพรรษา ง. เถลิงพระชนมพรรษา
9. “วัดระเบิด” ในข้อความนี้หมายความว่าอะไร ก. ดินปืน ข. ดินระเบิด ค. ไฟพะเนียง ง. พลุ ดอกไม้ไฟ	12. ผู้เขียนข้อความนี้ ต้องการรณรงค์ว่าเน้นเหตุการณ์ใด ก. การเล่นที่เป็นอันตราย ข. อุบัติเหตุบนท้องถนน ค. ไฟไหม้บ้านเรือนและทรัพย์สิน ง. การขายพลุและดอกไม้ไฟ

13. กลุ่มคนที่ “ คอยระแวงระวังกับแสงจากประกายไฟ ที่อาวรวงลงมาโดยเอาวัตถุติดไฟได้ทุกเมื่อ” เพราะกลัวจะเกิดอะไร ก. เกิดไฟลวกร่างกาย ข. เกิดไฟไหม้บ้านเรือน ค. เกิดอุบัติเหตุจากยานพาหนะ ง. เกิดระเบิดอย่างรุนแรง	14. วิธีการป้องกันมิให้เกิดเหตุการณ์ตามข้อความนี้ควรกระทำอย่างไร ก. ติดตามจับกุมผู้เดินวัตถุระเบิดอย่างเคร่งครัด ข. ออกกฎหมายควบคุมการซื้อขายวัตถุระเบิด ค. เคื่อนให้เดินวัตถุระเบิดอย่างระมัดระวัง ง. ปรามปรามผู้ผลิตวัตถุระเบิดให้หมดไป
---	--

อ่านข้อความข้างล่าง แล้วตอบคำถามข้อ 15-21

ชายที่ดินพร้อมมะม่วงพันธุ์ดี (ที่น้ำไม่ท่วม) ถนน น้ำ ไฟฟ้า สโมสร พร้อมบ้านพักฟรี แปลงละ 1 ไร่ ผ่อนเดือนละ 2,500 บาท ท่ามกลางแนวเขาและทัศนียภาพอันสวยงามทุกแปลง ติดถนนสาธารณะ และคือน้ำ อุณหภูมิ 2 ปี โครงการจัดเก็บผลผลิตระบบสหกรณ์ให้ทุกปี ราคาตนเอง ราชการ โทร.381160 บ้าน โทร.252931

15. ข้อความนี้มีลักษณะตามข้อใด ก. ข้อเท็จจริง ข. คำชักชวน ค. คำแนะนำ ง. คำโฆษณา	18. คำว่า “ราชการ” ในข้อความนี้มีความหมายตรงกับข้อใดมากที่สุด ก. เจ้าของที่ดินเป็นข้าราชการ ข. เป็นที่ดินของทางราชการ ค. ติดต่อกับในเวลาราชการ ง. ติดต่อกับได้ในสถานที่ราชการ
16. ข้อใด ไม่ใช่ วัตถุประสงค์ของข้อความนี้ ก. ชักชวนซื้อที่ดิน ข. ชักชวนขายที่ดิน ค. บรรยายทัศนียภาพ ง. จัดเก็บผลผลิตในระบบสหกรณ์	19. ข้อใดเป็นสาระสำคัญที่ควรเพิ่มเติมในข้อความนี้ ก. จำนวนเดือนที่ผ่อน ข. ราคาที่ดินทั้งแปลง ค. จำนวนผลผลิตจากที่ดิน ง. สถานที่ตั้งของที่ดิน
17. คำว่า “ทัศนียภาพ” ในข้อความนี้มีความหมายเหมือนกับคำใด ก. ทิวทัศน์ ข. ทัศนวิสัย ค. สภาพที่ดิน ง. สภาพภูมิประเทศ	

26. ข้อใดเป็นวัตถุประสงค์ประการแรกของการจัดกิจกรรมครั้งนี้ ก. เพื่อให้นักเรียนได้รู้ถึงพิษภัยของยาเสพติด ข. เพื่อส่งเสริมให้นักเรียนมีรายได้ระหว่างเรียน ค. เพื่อให้นักเรียนได้แสดงออกถึงความสามารถและความคิดสร้างสรรค์ ง. เพื่อให้นักเรียนช่วยกันเปิดเผยความจริงเกี่ยวกับการขายและเสพยาเสพติดในโรงเรียน	28. ผู้ชนะเลิศการประกวดวาดภาพระบายสีตามข้อความนี้ น่าจะได้รับทุนการศึกษาจำนวนเท่าใด เพราะอะไร ก. 20,000 บาท เพราะข้อความระบุไว้ ข. 10,000 บาท เพราะต้องแบ่ง 2 กิจกรรม ค. 5,000 บาท เพราะต้องแบ่งหลายคน ง. ไม่ทราบเพราะไม่มีรายละเอียด
27. เยาวชนระดับใดที่ ไม่มีสิทธิเข้าร่วมกิจกรรมการประกวดตามข้อความนี้ ก. ระดับอนุบาล ข. ระดับประถมศึกษา ค. ระดับมัธยมศึกษา ง. ระดับอาชีวศึกษา	

อ่านข้อความข้างล่างแล้วตอบคำถามข้อ 29-34

ปลูกต้นข้าวเกิดเมล็ดข้าวดังเขาว่า ปลูกด้วงงาเกิดด้วงงาเป็นแม่นมั่น
 ปลูกอย่างไรได้ผลอย่างเดียวกัน ความพิชพันรู้ที่หวานลงจงเข้าใจ
 อันความชั่วที่ปลูกลงคงได้ชั่ว ความคิดคงไม่กลัวคุ้มตัวได้
 ปลูกความดีผลดีมีเรื่อยไป ความชั่วไร้รอย่างปลูกเป็นลูกเอย

29. คำประพันธ์นี้เป็นการเขียนลักษณะใด ก. แสดงการเปรียบเทียบ ข. แสดงความคิดเห็น ค. แสดงข้อเท็จจริง ง. แสดงผลดีผลเสีย	32. คำประพันธ์นี้มีความหมายสอดคล้องกับข้อความใด ก. ปลูกพืชเช่นไรได้ผลเช่นนั้น ข. กว่าด้วงจะตูกงาก็ไหม้ ค. หวานที่ขหวังผล ง. คนคิดคุ้ม
30. ผู้เขียนคำประพันธ์นี้ต้องการกล่าวถึงเรื่องใด ก. ต้นข้าว เมล็ดข้าว ข. เห็นผิดเป็นชอบ ค. ความดีความชั่ว ง. พืชพันธุ์ธัญญาหาร	33. ข้อใดที่ <u>ไม่ได้</u>แสดงเหตุและผล ก. ปลูกความดีผลดีมีเรื่อยไป ข. อันความชั่วที่ปลูกลงคงได้ชั่ว ค. ปลูกด้วงงาเกิดด้วงงาเป็นแม่นมั่น ง. ความชั่วไร้รอย่างปลูกเป็นลูกเอย
31. คำประพันธ์นี้ให้ข้อคิดกับใคร ก. นักเรียน ข. ครู-อาจารย์ ค. เกษตรกร ง. คนทั่วไป	34. ข้อความใดมีความหมาย แตกต่าง จากคำประพันธ์นี้ ก. ผลดีย่อมเกิดแต่เหตุดี ข. ทำดีได้ดี ทำชั่วได้ชั่ว ค. ทำความดีลงดังความชั่ว ง. ความชั่วไม่ทำเสียเลยดีกว่า

อ่านข้อความข้างล่างแล้วตอบคำถามข้อ 35-40

อย่าให้เป็นประเพณี

การปิดถนนสายสำคัญเพื่อใช้เป็นเครื่องต่อรอง เรียกร้องให้ได้สิ่งที่ต้องการของคนกลุ่มต่าง ๆ กำลังระบาดอยู่โดยทั่วไปซึ่งไม่ว่าเกิดขึ้นคราวใด ก็ทำความเดือดร้อนแก่คนส่วนใหญ่ที่ต้องใช้เส้นทางในการสัญจรหรือการขนส่งสินค้าและปรากฏว่าการประท้วงด้วยวิธีนี้กลับเกิดผลดีกับกลุ่มผู้เรียกร้องเนื่องจากหน่วยราชการต่าง ๆ ที่เกี่ยวข้องจะเร่งระดมสรรพกำลัง ตลอดจนงบประมาณเข้ามาจัดการให้ในสิ่งที่ต้องการโดยดับพลังอื่นทั้งยังได้รับการงดเว้นการดำเนินคดีทั้ง ๆ ที่เป็นการกระทำที่ผิดกฎหมาย

35. ผู้เขียนข้อความนี้ต้องการบอกถึงเหตุการณ์ใด ก. การต่อรอง ข. การประท้วง ค. การระดมกำลัง ง. การทำผิดกฎหมาย	38. "เครื่องต่อรอง" ที่กล่าวถึงในข้อความนี้ น่าจะเป็นสิ่งใด ก. ข้อเรียกร้องต่าง ๆ ข. สิ่งของต่าง ๆ ที่ใช้ปิดถนน ค. ความเดือดร้อนของคนส่วนใหญ่ ง. ข้อยกเว้นการดำเนินคดี
36. ผู้ที่ได้รับความเดือดร้อนจากเหตุการณ์นี้คือคนกลุ่มใด ก. กลุ่มผู้เรียกร้อง ข. กลุ่มผู้ใช้รถใช้ถนน ค. กลุ่มข้าราชการต่าง ๆ ง. คณะรัฐบาล	39. ถ้าเกิดเหตุการณ์ตามข้อความนี้บ่อย ๆ จะส่งผลอย่างไร ก. จะมีผู้ทำผิดกฎหมายมากขึ้น ข. มีผู้ใช้รถใช้ถนนน้อยลง ค. คนไทยจะขาดความสามัคคี ง. เศรษฐกิจของชาติจะเสียหาย
37. คำว่า "ระบาด" ในข้อความนี้มีความหมายว่าอย่างไร ก. การแพร่เชื้อ ข. การติดต่อ ค. กระจัดกระจาย ง. เกิดขึ้นบ่อย ๆ	40. การปิดถนนเป็นการกระทำที่ผิดกฎหมาย แต่เหตุใดผู้เรียกร้องยังใช้วิธีนี้ ก. เพราะได้สิ่งที่ต้องการ ข. เพราะได้งบประมาณ ค. เพราะประชาชนให้ความสนใจ ง. เพราะเป็นความพอใจของคนส่วนใหญ่

ภาคผนวก ค

ค่าสถิติ และคุณสมบัติของเครื่องมือที่ใช้ในการวิจัย

ตาราง 65 ค่าสถิติของข้อสอบวัดความสามารถในการอ่านจับใจความตามกรอบทฤษฎี CTT และ IRT ที่ให้คะแนนตามรูปแบบปกติ และรูปแบบ 2 - SR

ข้อ	P		r		D		B		U		RII		BTA		F	
	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR
1	.62	.51	.09	.13	.19	.18	.02	.04	.09	.18	.04	.05	-.71	-.14	3.60	2.70
2	.63	.54	.26	.30	.28	.35	.03	.03	.27	.35	.13	.12	-.81	-.18	-1.00	-1.30
3	.54	.43	.20	.24	.27	.31	.05	.11	.10	.21	.10	.10	-.11	.00	.40	.20
4	.75	.67	.40	.48	.29	.35	.07	.10	.45	.57	.15	.16	-2.02	-.46	-2.50	-3.41
5	.37	.34	.05	.11	.11	.17	.00	.07	.22	.24	.02	.05	.30	.08	5.50	3.50
6	.65	.58	.32	.38	.35	.42	.07	.10	.44	.47	.15	.17	-.97	-.23	-3.30	-3.20
7	.42	.33	.20	.17	.22	.23	.04	.05	.09	.14	.10	.07	.44	.14	1.20	2.00
8	.10	.09	-.10	-.17	-.05	-.08	.09	.10	-.26	-.27	-.03	-.04	2.52	.59	1.10	1.60
9	.73	.64	.33	.40	.29	.38	.07	.09	.38	.49	.14	.15	-1.63	-.37	-2.20	-2.80
10	.36	.33	.14	.18	.19	.25	.07	.09	.21	.21	.07	.08	.55	.14	1.60	1.40
11	.59	.50	.16	.19	.21	.26	.03	.10	.15	.25	.08	.09	-.56	-.14	2.20	2.10
12	.62	.54	.36	.36	.33	.38	.04	.07	.29	.41	.18	.15	-.77	-.18	-3.40	-3.10
13	.55	.44	.24	.26	.28	.35	.04	.04	.16	.28	.12	.12	-.19	-.03	-.30	-.10

ข้อ	P		r		D		B		U		R11		BTA		F	
	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR
14	.52	.45	.21	.24	.24	.31	.07	.09	.26	.31	.11	.11	-.24	-.06	.70	.30
15	.73	.63	.23	.29	.22	.27	.03	.07	.25	.38	.10	.11	-1.35	-.32	-.40	-.90
16	.53	.48	.32	.35	.36	.44	.06	.09	.25	.37	.16	.17	-.40	-.10	-.370	-.3.10
17	.66	.60	.34	.40	.32	.38	.14	.15	.44	.51	.14	.16	-1.39	-.31	-.3.20	-.3.20
18	.38	.31	.16	.19	.19	.25	.05	.06	.13	.21	.08	.08	.59	.13	.70	1.20
19	.46	.38	.13	.18	.22	.27	.09	.13	.22	.26	.07	.08	.13	.03	2.80	2.50
20	.30	.26	.10	.13	.15	.18	.02	.09	.03	.09	.05	.05	.98	.21	1.50	1.70
21	.28	.22	.06	.06	.13	.10	.03	.07	.00	-.01	.03	.02	1.28	.31	1.40	2.20
22	.58	.50	.24	.29	.28	.33	.06	.10	.37	.38	.12	.13	-.65	-.14	-1.20	-1.30
23	.17	.15	.09	.09	.10	.09	.02	.10	-.07	-.02	.03	.03	1.95	.43	.10	.50
24	.68	.59	.32	.38	.32	.41	.05	.05	.47	.50	.14	.16	-1.15	-.25	-2.30	-3.00
25	.76	.69	.39	.45	.29	.33	.10	.12	.64	.64	.13	.15	-2.08	-.47	-2.40	-3.00
26	.30	.30	.07	.12	.14	.18	.07	.09	-.04	.06	.03	.05	.74	.17	2.90	2.60
27	.77	.69	.29	.34	.25	.28	.08	.12	.57	.54	.11	.12	-1.89	-.45	-1.50	-1.60
28	.16	.12	.02	.01	.05	.03	.04	.07	-.23	-.11	.01	.01	2.18	.49	.50	.90
29	.12	.10	-.10	-.11	.00	-.04	.04	.12	-.35	-.27	-.03	-.03	2.47	.55	.90	1.20
30	.90	.53	.44	.50	.22	.53	.09	.14	.58	.63	.21	.22	-.82	-.19	-6.60	-6.90
31	.66	.60	.39	.44	.36	.45	.11	.15	.52	.58	.18	.18	-1.07	-.28	-3.90	-4.00

ข้อ	P		r		D		B		U		RII		BTA		F	
	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR
32	.13	.12	-.18	-.19	-.07	-.08	.01	.05	-.22	-.21	-.06	-.06	1.97	.48	-2.60	3.10
33	.45	.38	.15	.18	.22	.29	.00	.10	.12	.20	.07	.08	.30	.06	1.90	2.30
34	.46	.38	.30	.33	.34	.43	.06	.08	.25	.35	.15	.15	.33	.07	-3.50	-2.90
35	.44	.39	.17	.21	.26	.31	.01	.10	.22	.25	.09	.10	.18	.02	1.40	1.40
36	.62	.50	.30	.36	.34	.42	.09	.14	.35	.40	.15	.17	-.55	-.11	-2.80	-3.30
37	.52	.43	.32	.36	.37	.47	.05	.09	.37	.42	.16	.17	.02	.00	-4.30	-3.90
38	.24	.21	-.13	-.12	-.03	-.07	.00	.06	-.07	-.11	-.06	-.05	1.09	.24	5.50	6.40
39	.31	.25	.02	.05	.08	.09	.03	.09	.01	.07	.01	.02	1.06	.23	2.90	3.30
40	.49	.40	.30	.37	.35	.46	.07	.10	.29	.36	.16	.17	.03	.05	-4.40	-4.50
$\bar{X}$	.4905	.4151	.1905	.2233	.2174	.2669	.0523	.0903	.1987	.2577	.0892	.0957	.0003	.0002	-.4175	-.3100
SD	.203	.171	.155	.176	.119	.160	.032	.030	.238	.241	.068	.070	.282	.275	2.82	2.88
t-test	8.19	**	6.86	**	-5.42	**	-9.48	**	-8.20	**	-3.54	**	0.00		-0.64	**
corr	.97	**	.99	**	.96	**	.67	**	.98	**	.19	**	.99	**	.43	**

ตาราง 66 สารสนเทศของข้อสอบที่ระดับความสามารถต่างๆ ของแบบทดสอบวัดความสามารถในการอ่านจับใจความ ที่ให้คะแนนตามรูปแบบปกติ และรูปแบบ 2 - SR

ความสามารถ	-3.00	-2.50	-2.00	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50	2.00	2.50	3.00
ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR
1	.0835 .0512	.1226 .0788	.1693 .1166	.2147 .1625	.2448 .2089	.2473 .2421	.2210 .2488	.1769 .2261	.1297 .1836	.0891 .1361	.0585 .0942	.0373 .0622	.0233 .0398
2	.0912 .0531	.1324 .0815	.1798 .1200	.2232 .1664	.2480 .2123	.2437 .2437	.2123 .2480	.1664 .2232	.1200 .1798	.0815 .1324	.0531 .0912	.0337 .0600	.0210 .0383
3	.0499 .0452	.0769 .0701	.1140 .1050	.1596 .1491	.2064 .1966	.2407 .2350	.2492 .2500	.2281 .2350	.1864 .1986	.1388 .1491	.0984 .1050	.0638 .0701	.0409 .0452
4	.1984 .0678	.2361 .1018	.2500 .1454	.2338 .1930	.1948 .2326	.1473 .2499	.1034 .2372	.0689 .2002	.0444 .1529	.0279 .1082	.0173 .0725	.0107 .0468	.0085 .0295
5	.0343 .0420	.0540 .0655	.0828 .0987	.1217 .1416	.1683 .1892	.2139 .2301	.2445 .2496	.2475 .2393	.2217 .2038	.1779 .1568	.1306 .1115	.0898 .0750	.0590 .0486
6	.1026 .0555	.1463 .0849	.1939 .1244	.2332 .1712	.2499 .2163	.2367 .2455	.1993 .2467	.1520 .2194	.1074 .1750	.0719 .1279	.0464 .0877	.0293 .0575	.0182 .0366
7	.0301 .0398	.0477 .0622	.0737 .0942	.1099 .1361	.1549 .1836	.2020 .2261	.2383 .2488	.2498 .2421	.2314 .2089	.1911 .1625	.1435 .1166	.1003 .0788	.0666 .0512
8	.0040 .0281	.0065 .0416	.0107 .0649	.0173 .0980	.0279 .1407	.0444 .1883	.0689 .2294	.1034 .2495	.1473 .2398	.1948 .2046	.2338 .1577	.2500 .1123	.2361 .0756
9	.1616 .0627	.2081 .0949	.2416 .1370	.2489 .1845	.2267 .2267	.1845 .2489	.1370 .2416	.0949 .2081	.0627 .1616	.0401 .1157	.0252 .0782	.0156 .0508	.0096 .0321
10	.0271 .0398	.0432 .0622	.0672 .0942	.1010 .1361	.1444 .1836	.1920 .2261	.2320 .2488	.2498 .2421	.2378 .2089	.2011 .1625	.1539 .1166	.1090 .0788	.0731 .0512
11	.0737 .0512	.1099 .0788	.1549 .1166	.2020 .1625	.2383 .2089	.2498 .2421	.2314 .2488	.1911 .2261	.1435 .1836	.1003 .1361	.0666 .0942	.0428 .0622	.0269 .0398
12	.0877 .0531	.1279 .0815	.1750 .1200	.2194 .1664	.2467 .2123	.2455 .2437	.2163 .2480	.1712 .2232	.1244 .1798	.0849 .1324	.0555 .0912	.0353 .0600	.0220 .0383
13	.0536 .0464	.0821 .0719	.1209 .1074	.1673 .1520	.2131 .1993	.2441 .2367	.2478 .2499	.2225 .2332	.1788 .1939	.1315 .1463	.0905 .1026	.0595 .0883	.0380 .0440
14	.0560 .0477	.0856 .0737	.1252 .1099	.1721 .1549	.2171 .2020	.2458 .2383	.2464 .2498	.2187 .2314	.1741 .1911	.1270 .1435	.0870 .1003	.0570 .0666	.0363 .0428
15	.1352 .0600	.1827 .0912	.2253 .1324	.2486 .1798	.2425 .2232	.2098 .2480	.1635 .2437	.1174 .2123	.0795 .1664	.0517 .1200	.0327 .0815	.0204 .0531	.0126 .0337
16	.0644 .0494	.0972 .0763	.1398 .1132	.1874 .1587	.2288 .2055	.2494 .2403	.2403 .2494	.2055 .2288	.1587 .1874	.1132 .1398	.0763 .0972	.0494 .0644	.0313 .0412

ตาราง 66 (ต่อ)

ความสามารภ	-3.00	-2.50	-2.00	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50	2.00	2.50	3.00
ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR
17	.1388 .0595	.1864 .0905	.2281 .1315	.2492 .1788	.2407 .2225	.2084 .2478	.1596 .2441	.1140 .2131	.0769 .1673	.0499 .1209	.0315 .0821	.0196 .0536	.0121 .0340
18	.0261 .0401	.0416 .0627	.0649 .0949	.0980 .1370	.1407 .1845	.1883 .2267	.2294 .2489	.2495 .2416	.2398 .2081	.2046 .1616	.1577 .1157	.1123 .0782	.0756 .0508
19	.0401 .0440	.0627 .0683	.0949 .1026	.1370 .1463	.1845 .1939	.2267 .2332	.2489 .2499	.2416 .2367	.2081 .1993	.1616 .1520	.1157 .1074	.0782 .0719	.0508 .0464
20	.0180 .0373	.0290 .0585	.0460 .0891	.0713 .1297	.1066 .1769	.1510 .2210	.1984 .2473	.2361 .2448	.2500 .2147	.2338 .1693	.1948 .1226	.1473 .0835	.1034 .0545
21	.0135 .0340	.0218 .0536	.0349 .0821	.0550 .1209	.0842 .1673	.1235 .2131	.1702 .2441	.2155 .2478	.2452 .2225	.2470 .1788	.2202 .1315	.1760 .0905	.1288 .0595
22	.0795 .0512	.1174 .0788	.1635 .1166	.2098 .1625	.2425 .2089	.2486 .2421	.2253 .2488	.1827 .2261	.1352 .1836	.0934 .1361	.0616 .0942	.0394 .0622	.0247 .0398
23	.0070 .0304	.0114 .0481	.0185 .0744	.0298 .1107	.0473 .1558	.0731 .2029	.1090 .2388	.1539 .2497	.2011 .2307	.2378 .1902	.2498 .1426	.2320 .0995	.1920 .0660
24	.1174 .0565	.1635 .0863	.2098 .1261	.2425 .1731	.2486 .2179	.2253 .2461	.1827 .2461	.1352 .2179	.0934 .1731	.0616 .1261	.0394 .0863	.0247 .0565	.0153 .0359
25	.2038 .0683	.2393 .1026	.2496 .1463	.2301 .1939	.1892 .2332	.1416 .2499	.0987 .2367	.0655 .1993	.0420 .1520	.0264 .1074	.0163 .0719	.0100 .0464	.0061 .0293
26	.0227 .0387	.0363 .0606	.0570 .0920	.0870 .1333	.1270 .1808	.1741 .2239	.2187 .2482	.2464 .2433	.2458 .2114	.2171 .1654	.1721 .1191	.1252 .0808	.0856 .0526
27	.1864 .0672	.2281 .1010	.2492 .1444	.2407 .1920	.2064 .2320	.1596 .2498	.1140 .2378	.0769 .2011	.0499 .1539	.0315 .1090	.0196 .0731	.0121 .0473	.0074 .0298
28	.0056 .0287	.0091 .0456	.0148 .0707	.0240 .1058	.0383 .1501	.0600 .1975	.0912 .2356	.1324 .2500	.1798 .2344	.2232 .1957	.2480 .1482	.2437 .1042	.2123 .0695
29	.0042 .0271	.0068 .0432	.0112 .0672	.0182 .1010	.0293 .1444	.0464 .1920	.0719 .2320	.1074 .2498	.1520 .2378	.1993 .2011	.2367 .1539	.2499 .1090	.2332 .0731
30	.0912 .0536	.1324 .0821	.1798 .1209	.2232 .1673	.2480 .2131	.2437 .2441	.2123 .2478	.1664 .2225	.1200 .1788	.0815 .1315	.0531 .0905	.0337 .0595	.0210 .0380
31	.1107 .0580	.1558 .0884	.2029 .1288	.2388 .1760	.2497 .2202	.2307 .2470	.1902 .2452	.1426 .2155	.0995 .1702	.0660 .1235	.0424 .0842	.0266 .0550	.0165 .0349
32	.0068 .0290	.0112 .0460	.0182 .0713	.0293 .1066	.0464 .1510	.0719 .1984	.1074 .2361	.1520 .2500	.1993 .2338	.2367 .1948	.2499 .1473	.2332 .1034	.1939 .0689

ตาราง 66 (ต่อ)

ความถี่	-3.00	-2.50	-2.00	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50	2.00	2.50	3.00
ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR
33	.0343 .0428	.0540 .0666	.0828 .1003	.1217 .1435	.1683 .1911	.2139 .2314	.2445 .2498	.2475 .2383	.2217 .2020	.1779 .1549	.1306 .1099	.0898 .0737	.0590 .0477
34	.0334 .0424	.0526 .0660	.0808 .0995	.1191 .1426	.1654 .1902	.2114 .2307	.2433 .2497	.2482 .2388	.2239 .2029	.1808 .1558	.1333 .1107	.0920 .0744	.0606 .0481
35	.0383 .0444	.0600 .0689	.0912 .1034	.1324 .1473	.1798 .1948	.2232 .2338	.2480 .2500	.2437 .2361	.2123 .1984	.1664 .1510	.1200 .1066	.0815 .0713	.0531 .0460
36	.0731 .0499	.1090 .0769	.1539 .1140	.2011 .1596	.2378 .2064	.2498 .2407	.2320 .2492	.1920 .2281	.1444 .1864	.1010 .1388	.0672 .0964	.0432 .0638	.0271 .0409
37	.0444 .0452	.0689 .0701	.1034 .1050	.1473 .1491	.1948 .1966	.2338 .2350	.2500 .2500	.2361 .2350	.1984 .1966	.1510 .1491	.1066 .1050	.0713 .0701	.0460 .0452
38	.0162 .0363	.0261 .0570	.0416 .0870	.0649 .1270	.0980 .1741	.1407 .2187	.1883 .2464	.2294 .2458	.2495 .2171	.2398 .1721	.2046 .1252	.1577 .0898	.1123 .0560
39	.0167 .0366	.0269 .0575	.0428 .0877	.0666 .1279	.1003 .1750	.1435 .2194	.1911 .2467	.2314 .2455	.2498 .2163	.2383 .1712	.2020 .1244	.1549 .0849	.1099 .0555
40	.0343 .0432	.0540 .0672	.0828 .1010	.1217 .1444	.1683 .1920	.2139 .2320	.2445 .2498	.2475 .2378	.2217 .2011	.1779 .1539	.1306 .1090	.0898 .0731	.0590 .0473
$\bar{x}$	.0654 .0464	.0916 .0716	.1212 .1064	.1505 .1497	.1749 .1954	.1900 .2316	.1930 .2454	.1840 .2314	.1652 .1951	.1407 .1496	.1143 .1064	.0887 .0716	.0657 .0464
SD.	.0550 .0112	.0681 .0162	.0760 .0216	.0766 .0255	.0706 .0247	.0619 .0168	.0570 .0056	.0588 .0146	.0645 .0238	.0705 .0258	.0743 .0225	.0729 .0172	.0649 .0120
t - test	2.69**	2.41*	1.70	.09	-2.64**	-5.32	6.43**	-5.86**	-3.93**	-1.18**	.95	1.93	2.28
CORR	.958**	.974**	.983**	.970**	.914**	.805**	.980**	.610**	.781**	.920**	.978**	.984**	.971**

ตาราง 67 ค่าสถิติของร้อยละความสามารถในการแก้ปัญหาทางคณิตศาสตร์สถิติตามกรอบทฤษฎี CTT และ IRT ที่ให้คะแนนตามรูปแบบปกติ และรูปแบบ 2 - SR

ข้อ	P		r		D		B		U		RII		BTA		F	
	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR
1	.77	.37	.39	.26	.41	.32	.04	.07	.66	.24	.16	.16	-2.0		-3.6	-.63
2	.84	.63	.30	.42	.26	.49	.30	.05	.58	.10	.10	.11	-2.4	-.40	-1.8	-3.60
3	.53	.68	.41	.34	.56	.37	.00	.06	.46	.04	.21	.21	-.56	-.51	-7.1	-2.00
4	.84	.45	.34	.44	.27	.59	.06	.06	.64	.33	.16	.12	-2.4	-.10	-2.3	-7.40
5	.08	.68	-.08	.37	-.01	.40	.03	.08	-.07	.02	-.02	-.02	2.3	-.52	2.0	-2.30
6	.83	.07	.32	-.09	.30	-.02	.05	.03	.62	-.04	.11	.11	-2.3	.49	-2.1	2.50
7	.45	.68	.28	.37	.37	.40	.03	.07	.15	.06	.14	.14	-.10	-.50	-.9	-2.40
8	.40	.40	.35	.29	.43	.41	.00	.08	.24	.38	.17	.17	.22	-.02	-3.3	-.70
9	.47	.35	.35	.36	.44	.42	.00	.03	.22	.35	.17	.18	-.20	.05	-3.8	-3.00
10	.43	.43	.33	.37	.44	.48	.03	.04	.24	.40	.16	.16	.10	-.05	-3.0	-3.80
11	.35	.40	.17	.36	.23	.45	.01	.07	.15	.40	.08	.08	.29	.02	2.8	-3.30
12	.54	.31	.37	.19	.51	.27	.03	.03	.40	.10	.18	.19	-.59	.07	-5.6	2.60
13	.57	.14	.24	.40	.32	-.09	.06	.05	.30	.30	.12	.12	-.67	-.14	.00	-5.90
14	.60	.49	.31	.27	.42	.39	.01	.10	.36	.13	.15	.15	-.79	-.13	9.5	-.20
15	.44	.51	.21	.33	.24	.47	.04	.06	.08	.24	.10	.12	.11	-.18	1.7	-2.50
16	.33	.39	.32	.23	.35	.32	.05	.07	.21	.28	.15	.15	.56	.02	-1.8	1.60
17	.33	.30	.27	.34	.32	.40	.01	.05	.14	.43	.12	.12	.52	.12	-8	-2.10

ข้อ	P		r		D		B		U		RII		BTA		F	
	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR
18	.35	.31	.38	.29	.42	.35	.03	.04	.23	.28	.17	.17	.48	.11	9.0	-.80
19	.49	.33	.44	.39	.56	.44	.02	.08	.38	.45	.22	.22	.22	.10	-7.6	-3.70
20	.39	.44	-.34	.46	-.28	.57	.02	.08	-.15	.42	-.17	-.17	.07	-.04	-9.9	-7.20
21	.43	.31	.42	-.32	.57	-.23	.01	.07	.30	-.40	.20	.21	.09	.02	-6.1	9.90
22	.71	.38	.37	.43	.46	.57	.06	.06	.54	.46	.16	.16	-.15	.01	-3.8	-5.90
23	.35	.60	.27	.41	.33	.52	.05	.11	.22	.19	.13	.13	.40	-.32	-20	-4.50
24	.30	.30	.15	.29	.16	.37	.01	.10	-.08	.32	.07	.07	.69	.08	2.3	-10
25	.39	.27	.28	.16	.35	.21	.04	.09	.12	.27	.13	.13	.23	.14	-9.0	2.80
26	.39	.36	.30	.29	.37	.38	.04	.08	.14	.38	.14	.14	.28	.05	-1.5	-.50
27	.24	.36	.24	.32	.22	.41	.02	.06	-.03	.39	.10	.10	.99	.06	-3.1	-1.50
28	.40	.24	.32	.25	.37	.25	.02	.10	.13	.49	.15	.15	.27	.21	-2.4	.00
29	.21	-.21	.24	.23	.16	.19	.01	.07	-.15	.54	.10	.10	1.2	.26	-.70	-.40
30	.22	.21	.21	.23	.15	.19	.06	.09	-.19	.55	.08	.08	1.2	.27	.00	.00
31	.62	.53	.27	.28	.37	.44	.05	.12	.42	.10	.13	.13	-1.0	-.23	-1.4	-1.10
32	.28	.25	-.03	.00	.03	.09	.01	.04	-.07	.01	-.01	-.01	.66	.15	6.8	6.80
33	.53	.45	.31	.33	.44	.46	.04	.08	.31	.26	.16	.16	-.40	-.09	-2.6	-2.70
34	.15	.14	-.03	-.03	.00	.02	.07	.09	-.20	.12	-.01	-.01	1.6	.35	3.1	3.40
35	.34	.30	.20	.21	.23	.27	.00	.06	.17	.14	.09	.09	.49	.10	1.7	2.00
36	.33	.31	.23	.25	.22	.30	.05	.08	.06	.32	.11	.11	.58	.11	.10	.30
37	.36	.31	.22	.25	.26	.33	.05	.06	.23	.15	.11	.11	.32	.06	1.3	.90
38	.37	.34	.14	.15	.19	.19	.02	.09	.05	.19	.07	.07	.31	.07	3.3	3.60

ตาราง 67 (ต่อ)

ข้อ	P		r		D		B		U		RII		BTA		F	
	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR	ปกติ	2-SR
39	.34	.30	.08	.10	.11	.15	.04	.10	.04	.13	.04	.04	.47	.12	4.6	4.40
40	.31	.26	.09	.10	.13	.20	.00	.06	.06	.08	.04	.04	.61	.14	4.0	4.10
X	.4325	.4312	.2410	.2630	.2921	.3749	.0320	.0230	.1977	.2463	.1111	.1151	-.0022	.0000	-.6025	-.6300
SD	.1810	.170	.154	.158	.174	.212	.0708	.023	.229	.192	.073	.072	1.02	.221	4.13	3.65
t-test	.12		-9.92**		-4.32**		-11.15**		-.08		-5.00**		-.02		.04	
CORR	.931**		.997**		.820**		.540**		-.99**		.99**		.99**		.40**	

ตาราง 68 ตารางสมมติของข้อสอบที่ระดับความสามารถต่างๆ ของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ ที่ให้คะแนนตามรูปแบบปกติและรูปแบบ 2 - SR

ระดับความสามารถ		-3.00	-2.50	-2.00	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50	2.00	2.50	3.00
ข้อ		ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR
1		.1939 .0644	.2332 .0972	.2499 .1398	.2367 .1874	.1993 .2288	.1520 .2494	.1074 .2403	.0719 .2055	.0464 .1587	.0293 .1132	.0182 .0763	.0112 .0494	.0058 .0313
2		.2267 .0707	.2489 .1058	.2416 .1501	.2081 .1975	.1616 .2356	.1157 .2500	.0782 .2344	.0508 .1957	.0321 .1482	.0200 .1042	.0123 .0695	.0076 .0448	.0046 .0282
3		.0737 .0494	.1099 .0763	.1549 .1132	.2020 .1587	.2383 .2055	.2498 .2403	.2314 .2494	.1911 .2288	.1435 .1874	.1003 .1398	.0666 .0972	.0428 .0644	.0269 .0412
4		.2274 .0713	.2491 .1066	.2412 .1510	.2072 .1984	.1606 .2361	.1149 .2500	.0775 .2338	.0503 .1948	.0318 .1473	.0198 .1034	.0122 .0689	.0075 .0444	.0046 .0279
5		.0052 .0287	.0085 .0456	.0139 .0707	.0224 .1058	.0359 .1501	.0565 .1975	.0863 .2356	.1261 .2500	.1731 .2344	.2179 .1957	.2461 .1482	.2461 .1042	.2179 .0695
6		.2225 .0701	.2478 .1050	.2441 .1491	.2131 .1966	.1673 .2350	.1209 .2500	.0821 .2350	.0536 .1966	.0340 .1491	.0212 .1050	.0131 .0701	.0080 .0452	.0049 .0285
7		.0494 .0460	.0763 .0713	.1132 .1066	.1587 .1510	.2055 .1984	.2403 .2361	.2494 .2500	.2288 .2338	.1874 .1948	.1398 .1473	.0972 .1034	.0644 .0689	.0412 .0444
8		.0369 .0432	.0580 .0672	.0884 .1010	.1288 .1444	.1760 .1920	.2202 .2320	.2470 .2498	.2452 .2378	.2155 .2011	.1702 .1539	.1235 .1090	.0842 .0731	.0550 .0473
9		.0540 .0473	.0828 .0731	.1217 .1090	.1683 .1539	.2139 .2011	.2445 .2378	.2475 .2498	.2217 .2320	.1779 .1920	.1306 .1444	.0898 .1010	.0590 .0672	.0376 .0432
10		.0412 .0444	.0644 .0689	.0972 .1034	.1398 .1473	.1874 .1948	.2288 .2338	.2494 .2500	.2403 .2361	.2055 .1984	.1587 .1510	.1132 .1066	.0763 .0713	.0494 .0460
11		.0346 .0424	.0545 .0660	.0835 .0995	.1226 .1426	.1693 .1902	.2147 .2307	.2448 .2497	.2473 .2388	.2210 .2029	.1769 .1558	.1297 .1107	.0891 .0744	.0585 .0481
12		.0756 .0512	.1123 .0788	.1577 .1166	.2046 .1625	.2398 .2089	.2495 .2421	.2294 .2488	.1883 .2261	.1407 .1836	.0980 .1361	.0649 .0942	.0416 .0622	.0261 .0398
13		.0808 .0508	.1191 .0782	.1654 .1157	.2114 .1616	.2433 .2081	.2482 .2416	.2239 .2489	.1808 .2267	.1333 .1845	.0920 .1370	.0606 .0949	.0387 .0627	.0242 .0401
14		.0891 .0531	.1297 .0815	.1769 .1200	.2210 .1664	.2473 .2123	.2448 .2437	.2147 .2480	.1693 .2232	.1226 .1798	.0835 .1324	.0545 .0912	.0346 .0600	.0216 .0383
15		.0409 .0444	.0638 .0689	.0964 .1034	.1388 .1473	.1864 .1948	.2281 .2338	.2492 .2500	.2407 .2361	.2064 .1984	.1596 .1510	.1140 .1066	.0769 .0713	.0499 .0460
16		.0269 .0405	.0428 .0633	.0666 .0957	.1003 .1379	.1435 .1855	.1911 .2274	.2314 .2491	.2498 .2412	.2383 .2072	.2020 .1606	.1549 .1149	.1099 .0775	.0737 .0503
17		.0279 .0409	.0444 .0638	.0689 .0964	.1034 .1388	.1473 .1864	.1948 .2281	.2338 .2492	.2500 .2407	.2361 .2064	.1984 .1596	.1510 .1140	.1066 .0769	.0713 .0499
18		.0290 .0412	.0460 .0644	.0713 .0972	.1066 .1398	.1510 .1874	.1984 .2288	.2361 .2494	.2500 .2403	.2338 .2055	.1948 .1587	.1473 .1132	.1034 .0763	.0689 .0494

ตาราง 68 (ต่อ)

ระดับความสามารถ		-3.00	-2.50	-2.00	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50	2.00	2.50	3.00
ข้อ		ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR
19		.0550 .0468	.0842 .0725	.1235 .1082	.1702 .1529	.2155 .2002	.2452 .2372	.2470 .2499	.2202 .2326	.1760 .1930	.1288 .1454	.0884 .1018	.0580 .0678	.0369 .0436
20		.0424 .0444	.0660 .0689	.0995 .1034	.1426 .1473	.1902 .1948	.2307 .2338	.2497 .2500	.2388 .2361	.2029 .1984	.1558 .1510	.1107 .1066	.0744 .0713	.0481 .0460
21		.0416 .0448	.0649 .0695	.0980 .1042	.1407 .1482	.1883 .1957	.2294 .2344	.2495 .2500	.2398 .2356	.2046 .1975	.1577 .1501	.1123 .1058	.0756 .0707	.0490 .0456
22		.1473 .0600	.1948 .0912	.2338 .1324	.2500 .1798	.2361 .2232	.1984 .2480	.1510 .2437	.1066 .2123	.0713 .1664	.0460 .1200	.0290 .0815	.0180 .0531	.0111 .0337
23		.0313 .0420	.0494 .0655	.0763 .0987	.1132 .1416	.1587 .1892	.2055 .2301	.2403 .2496	.2494 .2393	.2288 .2038	.1874 .1568	.1398 .1115	.0972 .0750	.0644 .0486
24		.0238 .0398	.0380 .0622	.0595 .0942	.0905 .1361	.1315 .1836	.1788 .2261	.2225 .2488	.2478 .2421	.2441 .2089	.2131 .1625	.1673 .1166	.1209 .0788	.0821 .0512
25		.0366 .0432	.0575 .0672	.0877 .1010	.1279 .1444	.1750 .1920	.2194 .2320	.2467 .2498	.2455 .2378	.2163 .2011	.1712 .1539	.1244 .1090	.0849 .0731	.0555 .0473
26		.0349 .0428	.0550 .0666	.0842 .1003	.1235 .1435	.1702 .1911	.2155 .2314	.2452 .2498	.2470 .2383	.2202 .2020	.1760 .1549	.1288 .1099	.0884 .0737	.0580 .0477
27		.0178 .0373	.0287 .0585	.0456 .0891	.0707 .1297	.1058 .1769	.1501 .2210	.1975 .2473	.2356 .2448	.2500 .2147	.2344 .1693	.1957 .1226	.1482 .0835	.1042 .0545
28		.0353 .0432	.0555 .0672	.0849 .1010	.1244 .1444	.1712 .1920	.2163 .2320	.2455 .2498	.2467 .2378	.2194 .2011	.1750 .1539	.1279 .1090	.0877 .0731	.0575 .0473
29		.0143 .0356	.0231 .0560	.0369 .0856	.0580 .1252	.0884 .1721	.1288 .2171	.1760 .2458	.2202 .2464	.2470 .2187	.2452 .1741	.2155 .1270	.1702 .0870	.1235 .0570
30		.0144 .0353	.0233 .0555	.0373 .0849	.0585 .1244	.0891 .1712	.1297 .2163	.1769 .2455	.2210 .2467	.2473 .2194	.2448 .1750	.2147 .1279	.1693 .0877	.1226 .0575
31		.1066 .0555	.1510 .0849	.1984 .1244	.2361 .1712	.2500 .2163	.2338 .2455	.1948 .2467	.1473 .2194	.1034 .1750	.0689 .1279	.0444 .0877	.0279 .0575	.0173 .0366
32		.0245 .0394	.0390 .0616	.0611 .0934	.0927 .1352	.1342 .1827	.1817 .2253	.2246 .2486	.2484 .2425	.2429 .2098	.2106 .1635	.1644 .1174	.1183 .0795	.0801 .0517
33		.0644 .0490	.0972 .0756	.1398 .1123	.1874 .1577	.2288 .2046	.2494 .2398	.2403 .2495	.2055 .2294	.1587 .1883	.1132 .1407	.0763 .0980	.0494 .0649	.0313 .0416
34		.0103 .0327	.0168 .0517	.0271 .0795	.0432 .1174	.0672 .1635	.1010 .2098	.1444 .2425	.1920 .2486	.2320 .2253	.2498 .1827	.2378 .1352	.2011 .0934	.1539 .0616
35		.0287 .0412	.0456 .0844	.0707 .0972	.1058 .1398	.1501 .1874	.1975 .2288	.2356 .2494	.2500 .2403	.2344 .2055	.1957 .1587	.1482 .1132	.1042 .0763	.0695 .0494
36		.0264 .0409	.0420 .0638	.0655 .0964	.0987 .1388	.1416 .1864	.1892 .2281	.2301 .2492	.2496 .2407	.2393 .2064	.2038 .1586	.1115 .0769	.0750 .0499	

ตาราง 68 (ต่อ)

ระดับความสามารถ	-3.00	-2.50	-2.00	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50	2.00	2.50	3.00
ข้อ	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR	ปกติ 2-SR
37	.0337 .0428	.0531 .0666	.0815 .1003	.1200 .1435	.1664 .1911	.2123 .2314	.2437 .2498	.2480 .2383	.2232 .2020	.1798 .1549	.1324 .1099	.0912 .0737	.0600 .0477
38	.0340 .0424	.0536 .0660	.0821 .0995	.1209 .1426	.1673 .1902	.2131 .2307	.2441 .2497	.2478 .2388	.2225 .2029	.1788 .1558	.1315 .1107	.0905 .0744	.0595 .0481
39	.0293 .0405	.0464 .0633	.0719 .0957	.1074 .1379	.1520 .1855	.1993 .2274	.2367 .2491	.2499 .2412	.2332 .2072	.1939 .1606	.1463 .1149	.1026 .0775	.0683 .0503
40	.0256 .0398	.0409 .0622	.0638 .0942	.0964 .1361	.1388 .1836	.1864 .2261	.2281 .2488	.2492 .2421	.2407 .2089	.2064 .1625	.1596 .1166	.1140 .0788	.0769 .0512
X	.0604 .0460	.0829 .0711	.1095 .1059	.1393 .1494	.1698 .1956	.1956 .2326	.2097 .2471	.2065 .2329	.1859 .1959	.1537 .1496	.1180 .1059	.0853 .0710	.0587 .0459
SD	.0598 .0098	.0660 .0139	.0646 .0182	.0572 .0272	.0495 .0187	.0487 .0110	.0545 .0047	.0619 .0140	.0663 .0199	.0656 .0202	.0606 .0169	.0524 .0126	.0421 .0090
t - test	1.80	1.44	.50	-1.62	-4.23**	-5.25**	-4.78**	-3.31**	-1.30	.57	1.74	2.23*	2.36*
CORR	.973**	.984**	.979**	.914**	.707**	.419**	.985**	.859**	.917**	.963**	.978**	.967**	.942**

ภาคผนวก ง

การพัฒนาสูตรการให้คะแนนแบบ 2 - SR

ภาคผนวก ง

การพัฒนาสูตรประมาณค่าคะแนนจากการตอบ 2 ขั้นตอน

ถ้ากำหนดให้ $x = A + B$

เมื่อ x คือ คะแนนรวมของการตอบทั้งสองส่วน

A คือ ผลการตอบในส่วนที่ 1

$A = 1$ เมื่อตัวเลือกที่เลือกในส่วนที่หนึ่งมีตัวเลือกถูก

รวมอยู่ด้วย

$A = 0$ เมื่อตัวเลือกที่เลือกในส่วนที่หนึ่งไม่มีตัวเลือกถูก

รวมอยู่ด้วย

B คือ ผลการตอบในส่วนที่สอง ถูก 1 ผิด 0

จากสมการ (1) ค่า x จะมีค่า 0, 1 และ 2 คะแนน

คะแนน 0 แสดงถึงผู้สอบผิดทั้งสองส่วน หรือไม่ตอบเลย

คะแนน 1 แสดงถึงผู้สอบตอบถูกหรือผิดในส่วนใดส่วนหนึ่ง

คะแนน 2 แสดงถึงผู้สอบตอบถูกทั้งสองส่วน

จุดอ่อนที่พบจากการใช้สมการ (1)

1. คนที่ได้ 1 คะแนนมีสองกลุ่ม คือ กลุ่มที่ตอบถูกในส่วนที่หนึ่งและกลุ่มที่ตอบถูกในส่วนที่สอง ถ้าพิจารณาในความเป็นจริงแล้ว คนทั้งสองกลุ่มไม่น่าจะได้คะแนนเท่ากันเพราะแสดงความสามารถต่างกัน กล่าวคือ คนที่ตอบในกลุ่มที่สองมีโอกาสตอบถูก $1/K$ หรือมีโอกาสตอบผิด $(K-1)/K$ เมื่อ K คือจำนวนตัวเลือกทั้งหมดในแต่ละข้อ ขณะที่คนในกลุ่มที่สองมีโอกาสตอบถูก N/K และมีโอกาสตอบผิด $(K-N)/K$ เมื่อ N คือตัวเลือกที่ผู้สอบเลือกในส่วนที่หนึ่ง อย่างไรก็ตาม ถ้า $N = 1$ คะแนนของทั้งสองควรจะมีน้ำหนักเท่ากัน

2. คนที่เลือกตอบในส่วนที่หนึ่งทุกตัวเลือก ($N=K$) จะได้ 1 คะแนนแน่นอนในความเป็นจริงแล้ว ควรได้ 0 มากกว่า เพราะการเลือกทุกตัวเลือกน่าจะแสดงให้เห็นว่าผู้สอบไม่มีความรู้หรือเดาตอบแบบสุ่ม (blind quessing) ในข้อนั้นๆ มากกว่าแสดงให้เห็นถึงประสิทธิภาพของตัวลองทั้งหมด

3. ผู้สอบที่ไม่มีความรู้หรือเดาตอบแบบสุ่มจะได้ 1 คะแนนแน่นอนเมื่อเลือกตัวเลือกในส่วนที่หนึ่งจำนวน $K-1$ ตัวเลือก และเลือกตัวเลือกที่เหลือในส่วนที่สอง

4. ในส่วนที่หนึ่ง ไม่ว่าจะเลือกตัวเลือกเท่าใดก็ตาม ถ้ามีตัวถูกจะได้คะแนนเท่ากันหมด ตามความเหมาะสม แล้วผู้ที่เลือกตัวเลือกน้อยกว่าจะได้คะแนนมากกว่าผู้ที่เลือกตัวเลือกมากกว่า

จากเหตุผลดังกล่าว ถ้าให้น้ำหนักคะแนนในแต่ละส่วนเท่ากับความน่าจะเป็นในการตอบผิด

ดังนั้น จาก (1) เมื่อถ่วงด้วยน้ำหนักของการตอบแต่ละส่วนจะได้ว่า

$$X = [(K-N)A/K] + [(K-1)B/K] \quad \dots\dots\dots(2)$$

จากสมการ (2) พบว่า

1. ค่าคะแนนในส่วนที่หนึ่งจะมีค่าเท่ากับค่าคะแนนในส่วนที่สองก็ต่อเมื่อ $N=1$ เท่านั้น และถ้า $N>1$ ค่าคะแนนในส่วนที่หนึ่งจะมีค่าน้อยกว่าส่วนที่สองเมื่อ $A=B=1$

2. ค่าคะแนนในส่วนที่หนึ่งจะเพิ่มขึ้นแบบผกผันกับจำนวนตัวเลือก (N) กล่าวคือ ในกรณีตอบถูก ถ้าเลือก $N = 1$ จะได้คะแนนมากกว่า $N = 2, 3, \dots$

3. ถ้าผู้สอบเลือกทุกตัวเลือกในส่วนที่หนึ่ง ($N=K$) จะได้คะแนน 0

แม้ว่าสมการ (2) จะสามารถแก้ไขจุดอ่อนของสมการ (1) ข้อ 1 และ 4 แต่ยังไม่สามารถแก้ไขข้อ 2 และ 4 ได้

ในกรณีข้อ 2 แม้ว่าผู้ที่เลือกตอบทุกตัวเลือกจะได้ 0 ในส่วนที่หนึ่ง แต่ยังมีโอกาสที่จะได้คะแนนในส่วนที่สองถ้าตอบถูก ในกรณีเช่นนี้มีความเป็นไปได้ในการอธิบายสองประเด็น คือ

ประเด็นที่ 1 ผู้ตอบมีความคลุมเครือในคำตอบมากหรือมีความรู้้น้อยมาก

ประเด็นที่ 2 ผู้ตอบไม่มีความรู้เลยและตอบในส่วนที่สองด้วยการเดาแบบสุ่ม

ถ้ามองในแง่ประเด็นที่หนึ่งสมการ (21) ก็น่าจะใช้ได้ แต่ถ้าพิจารณาคะแนนแล้วก็จะพบว่า แม้ว่าจะได้ 0 ในส่วนที่หนึ่ง แต่ถ้าตอบถูกในส่วนที่สองจะได้คะแนนเท่ากับ $K-1$ ซึ่งสูงกว่าผู้ตอบถูกในส่วนที่หนึ่งที่เลือกตัวเลือกน้อยกว่า K ($N < K$) แต่ตอบผิดในส่วนที่สอง (ให้พิจารณาแบบแผนการตอบแบบทำย) ซึ่งเป็นสิ่งที่ไม่น่ายอมรับได้ ดังนั้นคะแนนที่เหมาะสมสำหรับแบบแผนนี้น่าจะได้ 0 เช่นเดียวกับการมองในประเด็นที่สอง

ดังนั้น จาก (2) จึงนำ $(N-K) / (N-K)$ คูณ B จะได้ว่า

$$X = [(K-N)A/K] + [(K-1)(N-K)B/K(N-K)] \quad \dots\dots\dots(3)$$

จากสมการ (3) X จะมีค่าเป็น 0 เมื่อ $N=K$

และถ้า $N < K$ X จะมีค่าเท่ากับสมการ (2)

และเพื่อแก้จุดอ่อนของสมการ (1) ข้อที่ 4 จึงใช้ C คุณเข้าไปในสมการ (3)

จากสมการ (3) จะได้ว่า

$$X = [(K-N)A/K + [(K-1)(N-K)B/K(N-K))]C \dots\dots\dots(4)$$

เมื่อ C คือ ความสอดคล้องของการตอบทั้งสองด้าน

ถ้าตัวเลือกในส่วนที่สองรวมอยู่ในตัวเลือกที่เลือกไว้ในส่วนที่หนึ่ง

ค่า $C = 1$ แต่ถ้าตัวเลือกในส่วนที่สองไม่ได้รวมอยู่ในตัวเลือกที่เลือกไว้
ส่วนที่หนึ่ง ค่า $C = 0$

จากสมการ (4) ค่า X ที่ได้อาจอยู่ในรูปของทศนิยมซึ่งไม่สะดวกในการนำไปใช้ ดังนั้นจึง
ขยาย X จำนวน K เท่า จะได้ว่า

$$X = [(K-N)A + [(K-1)(N-K) B/K(N-K))]C$$

$$\text{หรือ } X = \left[\frac{(K-N)A + (K-1)(N-K) B}{(N-K)} \right] C$$

การศึกษาอิทธิพลของการให้คะแนนและการคัดเลือกข้อสอบที่มีต่อ
คุณภาพแบบทดสอบเลือกตอบ

บทคัดย่อ
ของ
ชวลิต โพธิ์นคร

เสนอต่อ มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษาดุขุภีบัณฑิต สาขาการทดสอบและวัดผลการศึกษา
มีนาคม 2540

การวิจัยครั้งนี้มีจุดประสงค์สำคัญคือ เพื่อศึกษาอิทธิพลของรูปแบบการให้คะแนน และวิธีการคัดเลือกข้อสอบที่มีต่อคุณภาพของแบบทดสอบเลือกตอบภายใต้กรอบทฤษฎี CTT และ IRT (ราชคโหมเดล) นอกจากนี้ยังศึกษาอิทธิพลของรูปแบบการให้คะแนน และการคัดเลือกข้อสอบภายใต้กรอบทฤษฎี CTT ที่มีต่อคุณภาพของแบบทดสอบตามกรอบทฤษฎี (ราชคโหมเดล) รวมทั้งศึกษารูปแบบความสัมพันธ์เชิงสาเหตุของการให้คะแนน และวิธีการคัดเลือกข้อสอบที่มีต่อคุณภาพของแบบทดสอบเลือกตอบ ตามกรอบทฤษฎี CTT และ IRT (ราชคโหมเดล)

กลุ่มตัวอย่างที่ใช้ศึกษาค้างนี้เป็นนักเรียนชั้นประถมศึกษาปีที่ 6 สังกัดสำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ ปีการศึกษา 2539 จำนวน 1,304 คน เป็นชาย 664 คน และหญิง 640 คน ซึ่งได้จากการคัดเลือกจังหวัดเป้าหมาย โดยใช้เทคนิค PPS แบบสุ่มสลาก (The lottery method of probability proportional to size selection) และสุ่มตัวอย่างโดยวิธีสุ่มแบบหลายขั้นตอน (Multistage sample sampling technique) ใน 21 โรงเรียน จาก 12 เขตการศึกษา และกรุงเทพมหานคร เครื่องมือที่ใช้ในการวิจัยมี 2 ฉบับคือ (1) แบบทดสอบวัดความสามารถในการอ่านจับใจความ และ (2) แบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ แต่ละฉบับประกอบด้วยข้อสอบจำนวน 40 ข้อ ซึ่งมีความเชื่อมั่นเท่ากับ .66 และ .77 ตามลำดับ ตัวแปรที่ศึกษา ได้แก่ ตัวแปรอิสระคือ การให้คะแนน ซึ่งมี 2 รูปแบบ ได้แก่ 1) รูปแบบปกติ (ถูก 1 ผิด 0) และ 2) รูปแบบ 2-SR และวิธีการคัดเลือกข้อสอบเข้าฉบับ โดยใช้ค่าพารามิเตอร์ และคุณสมบัติบางประการของข้อสอบเป็นเกณฑ์ในการคัดเลือกคือ ค่าความยากตามกรอบทฤษฎี CTT และความยากแบบ IRT (β) สัมประสิทธิ์การจำแนก ดัชนีการจำแนก ความลำเอียงของข้อสอบ เนื่องจากตัวแปรเพศ ความเป็นนิติเดี่ยของข้อสอบและความเหมาะสมของข้อสอบกับโมเดล (ราชคโหมเดล) ตัวแปรตาม ได้แก่ คะแนนสอบ ระดับความสามารถ (θ) ความเชื่อมั่นและสารสนเทศของข้อสอบ และแบบทดสอบ ผลการวิจัยพบว่า

1. การให้คะแนนตามรูปแบบ 2-SR ทำให้คะแนนสอบมีค่าต่ำกว่าคะแนนสอบที่ได้จากการให้คะแนนตามรูปแบบปกติ อย่างมีนัยสำคัญทางสถิติ ณ ระดับ .05 ส่วนระดับความสามารถ (θ) พบว่าให้ค่าสูงกว่าในแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ ส่วนในแบบทดสอบวัดความสามารถในการอ่านจับใจความ พบว่าให้ค่าต่ำกว่าอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .05

2. ค่าความเชื่อมั่นของแบบทดสอบจากการให้คะแนนตามรูปแบบ 2-SR และจากการให้คะแนนตามรูปแบบปกติให้ผลที่ไม่แตกต่างกัน แต่เมื่อศึกษาค่าสารสนเทศของแบบทดสอบ (TIF)

พบว่า การให้คะแนนตามรูปแบบ 2-SR มีประสิทธิภาพสูงกว่าค่าสารสนเทศของแบบทดสอบที่ได้จากการให้คะแนนตามปกติที่ระดับความสามารถปานกลาง ส่วนที่ระดับความสามารถสูงและต่ำ พบว่ามีประสิทธิภาพต่ำกว่าค่าที่ได้จากการให้คะแนนแบบปกติ และเมื่อศึกษาเป็นรายข้อพบว่า ค่าสัมประสิทธิ์การจำแนก ดัชนีการจำแนก ความลำเอียงเนื่องจากตัวแปรเพศ ความเป็นมิติเดียว และความเชื่อมั่นรายข้อของข้อสอบและสารสนเทศของข้อสอบที่ระดับความสามารถปานกลางที่ได้จากการให้คะแนนตามรูปแบบ 2-SR มีค่าสูงกว่าค่าที่ได้จากวิธีให้คะแนนตามรูปแบบปกติ อย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01 แต่พบว่าค่าสารสนเทศของข้อสอบที่ระดับความสามารถสูง ที่ได้จากการให้คะแนนแบบ 2-SR มีค่าต่ำกว่าค่าที่ได้จากการให้คะแนนแบบปกติ อย่างมีนัยสำคัญ ณ ระดับ .05 ส่วนค่าความยากและสารสนเทศของข้อสอบที่ระดับความสามารถต่ำพบว่า ค่าที่ได้จากการให้คะแนนตามรูปแบบ 2-SR ให้ค่าที่ต่ำกว่าค่าที่ได้จากการให้คะแนนแบบปกติอย่างมีนัยสำคัญ ณ ระดับ .05 เฉพาะในแบบทดสอบวัดความสามารถในการอ่านจับใจความเท่านั้น และสำหรับค่า β และความเหมาะสมของข้อสอบกับโมเดล ตามกรอบทฤษฎี IRT พบว่า ค่าที่ได้จากการให้คะแนนตามรูปแบบ 2-SR ไม่แตกต่างจากค่าที่ได้จากการให้คะแนนตามรูปแบบปกติ

3. การคัดเลือกข้อสอบตามกรอบทฤษฎี CTT โดยเกณฑ์การคัดเลือกต่าง ๆ เมื่อให้คะแนนตามรูปแบบ 2-SR ให้ค่าความเชื่อมั่นของแบบทดสอบทั้ง 2 ฉบับ สูงกว่าค่าความเชื่อมั่นของแบบทดสอบที่ได้จากการให้คะแนนตามรูปแบบปกติอย่างมีนัยสำคัญทางสถิติ ณ ระดับ .05

4. การศึกษาความสัมพันธ์เชิงสาเหตุของตัวแปรต่าง ๆ ที่ใช้เป็นเกณฑ์ในการคัดเลือกข้อสอบเข้าฉบับที่ได้จากรูปแบบ 2-SR และรูปแบบปกติโดยใช้แผนภาพเส้นทาง ให้ค่าสัมประสิทธิ์พยากรณ์และสูงกว่าการอธิบายโดยใช้สมการถดถอยพหุคูณในแบบทดสอบทั้งสองฉบับ อย่างมีนัยสำคัญทางสถิติ ณ ระดับ .01 และพบว่า ตัวแปรที่ส่งอิทธิพลต่อความเชื่อมั่นของข้อสอบของแบบทดสอบทั้งสองฉบับ เมื่อให้คะแนนตามรูปแบบ 2-SR และแบบปกติ คือ ความยาก สัมประสิทธิ์การจำแนก ดัชนีการจำแนก ความเป็นมิติเดียวของข้อสอบส่วนตัวแปรที่ส่งอิทธิพลต่อสารสนเทศของข้อสอบที่ความสามารถสูง กลางและต่ำของข้อสอบของแบบทดสอบทั้ง 2 ฉบับคือ ความยากแบบ CTT ค่า β ความเหมาะสมของข้อสอบกับโมเดล สัมประสิทธิ์การจำแนก ดัชนีการจำแนกและความเป็นมิติเดียวของข้อสอบ

**A STUDY OF THE EFFECT OF ITEM SCORING AND SELECTING ON THE QUALITY
OF THE MULTIPLE-CHOICE TESTS**

**AN ABSTRACT
BY
CHAVALIT PHONAKORN**

Presented in partial fulfillment of the requirements
for the Doctor of Education degree in Testing and Measurement
at Srinakarinwirot University
March 1997

The primary objectives of this research were to study the effects of item scoring and selecting used in creating quality multiple choice (MC) tests, within the scope of CTT and IRT (Rasch Model) theories ; to study the effect of item scoring and selecting as described by CTT theory, on the quality of MCs within the scope of IRT ; and to investigate the path diagram of the causal model between item scoring and selecting to identify the MCs quality within the scope of CTT and IRT.

The study's results were based on data gathered during the 1996 academic year from 1,304 subjects enrolled at the Prathom 6 level of the Office of the National Primary Education Commission (ONPEC). This group consisted of 664 males and 640 females. The subjects were selected from target provinces by means of a lottery according to PPS selection methods. While the data was gathered using multi-stage sampling techniques, from 21 schools in 12 educational regions and Bangkok. The subjects were administered a 40-item MC reading comprehension and 40-item MC mathematical problem solving test, with .66 and .77 statistical reliability respectively.

The 2 independent variables of the study were item scoring methods such as dichotomous (true=1, false=0) and 2-SR scoring, and item selecting based on certain parameters and item descriptions. These parameters and descriptions included the item level of difficulty within the scope of CTT (p) and IRT (β), discriminant coefficient, (r) index of discriminant, (D), gender item bias, (B) unidimensionality of item (U) and model-item fit (F).

The dependent variables of the study include test scores, ability level (θ), item and test reliability (r_{ii} and r_{tt}), item and test information function (IIF and TIF) on the tests.

Results of the study can be summarized as follows :

1) Test scores obtained under 2-SR scoring methods were lower than scores achieved under dichotomous scoring, and the difference in scores is statistically significant, at .05 level. Higher θ was found on the mathematical problem solving test, while lower θ was found on the reading comprehension test, with a statistically significant difference of $p < .05$.

2) No significant differences were found between the r_{tt} of 2 different scoring methods. However, it was found that the TIF of 2-SR scoring is more efficient than dichotomous scoring at moderate θ but with either high or low 2-SR scoring is less efficient than

dichotomous scoring, with a statistically significant difference of $p < .05$. In addition, the study also showed that the r , D , B , U , r_{tt} and TIF at the moderate θ level of 2-SR scoring, are significantly higher than those obtained through dichotomous scoring, at the .01 level, whereas the TIF at high θ under 2-SR scoring is significantly lower than that under dichotomous scoring, with a difference ($p < .05$). Only in the reading comprehension test were the p and TIF at low θ under 2-SR scoring found to be statistically significant lower than those from dichotomous scoring, at $p < .05$ level while no statistically significant difference was observed for B and F under the 2 different item selecting methods.

3) Item selecting performed within the θ scope of CTT, after consideration of independent variables, shows r_{tt} of both tests is higher under 2-SR scoring than dichotomous scoring, with a statistical significance of $p < .05$.

4) The path diagrams of the causal model of the variables used as a basis for item selecting, under both scoring method, give higher predictive accuracy coefficients (R^2) than its explanation with multiple regression equation in both tests to be statistically significant at the .01 level. It was found that factors which are influential on r_{tt} in both tests under both scoring methods include item difficulty, discriminant coefficient, Index of discriminant and unidimensionality of the item. The variables which influenced IIF on both tests at various levels are item difficulty of CTT and IRT, model-item fit, discriminant coefficient index of discriminant and unidimensionality of the item.

ประวัติย่อของผู้วิจัย

ชื่อ - ชื่อสกุล นายชวลิต โพธิ์นคร

วัน เดือน ปีเกิด วันที่ 9 พฤศจิกายน พุทธศักราช 2499

สถานที่เกิด อำเภอลาดยาว จังหวัดนครสวรรค์

สภาพที่อยู่ปัจจุบัน 410/55 :ข.สุขุมวิท 93 ถ.สุขุมวิท บางจาก พระโขนง

กรุงเทพมหานคร

ตำแหน่งหน้าที่การงานปัจจุบัน นักวิชาการการศึกษา ระดับ 6

สถานที่ทำงานปัจจุบัน สำนักนิเทศ และพัฒนามาตรฐานการศึกษา

สำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ

กระทรวงศึกษาธิการ

ประวัติการศึกษา

พ.ศ. 2512 ชั้นประถมศึกษาปีที่ 7

โรงเรียนบ้านลาดยาว อำเภอลาดยาว จังหวัดนครสวรรค์

พ.ศ. 2515 ชั้นมัธยมศึกษาปีที่ 3

โรงเรียนลาดยาว อำเภอลาดยาว จังหวัดนครสวรรค์

พ.ศ. 2519 ประกาศนียบัตรวิชาการศึกษาชั้นสูง (ปกศ.สูง)

วิชาเอกวิทยาศาสตร์/วิชาเทคโนโลยีศาสตร์ เกษตร

วิทยาลัยครูนครสวรรค์

พ.ศ. 2524 การศึกษามัธยมศึกษา (กศ.ม)

สาขาวิชาเอก เคมี วิชาโท วัสดุ

มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร

พ.ศ. 2529 ศีษศาสตรมหาบัณฑิต (ศษ.ม)

สาขาการวัดผลและประเมินผลทางการศึกษา

มหาวิทยาลัยเชียงใหม่

พ.ศ. 2540 การศึกษาดุษฎีบัณฑิต (กศ.ด)

สาขาการทดสอบและวัดผลการศึกษา

มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร