รายงานการวิจัย

การเปรียบเทียบคุณภาพของแบบสอบเลือกตอบเมื่อตรวจด้วยวิธีการให้คะแนนความรู้บางส่วนกับวิธีประเพณีนิยม

(A Comparison of Multiple Choice Tests'Quality when Scoring with the Partial Knowledge Methods and the Conventional Method)

ดร.เอมอร จังศิริพรปกรณ์
ผู้วิจัย

รศ.ดร.ศิริชัย กาญจนวาสี
อาจารย์อาวุโส

งานวิจัยนี้ได้รับทุนพัฒนาอาจารย์ใหม่ /นักวิจัยใหม่
กองทุนรัชดาภิเษกสมโภช จุฬาลงกรณ์มหาวิทยาลัย

รายงานการวิจัย

การเปรียบเทียบคุณภาพของแบบสอบเลือกตอบเมื่อตรวจด้วยวิธีการให้คะแนนความรู้บางส่วนกับวิธีประเพณีนิยม

(A Comparison of Multiple Choice Tests'Quality when Scoring with the Partial Knowledge Methods and the Conventional Method)

ดร.เอมอร จังศิริพรปกรณ์

ผู้วิจัย

รศ.ดร.ศิริชัย กาญจนวาสี

อาจารย์อาวุโส

งานวิจัยนี้ได้รับทุนพัฒนาอาจารย์ใหม่ /นักวิจัยใหม่

กองทุนรัชดาภิเษกสมโภช จุฬาลงกรณ์มหาวิทยาลัย

21 ส.ค. 2545

I20011425

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

เลขหมู่ จฬ
ด 15
เลขทะเบียน 011149
วัน.เดือน.ปี 27 ส.ย.45

การเปรียบเทียบคุณภาพของแบบสอบเลือกตอบเมื่อตรวจด้วยวิธีการให้คะแนนความรู้บางส่วนกับวิธีประเพณีนิยม

บทคัดย่อ(Abstract)

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบคุณภาพของแบบสอบในด้านความตรงตามสภาพ อำนาจจำแนก ความยาก ฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบสอบ และอัตราส่วนสารสนเทศเฉลี่ยโดยใช้ทฤษฎีการตอบสนองข้อสอบเมื่อแบบสอบตรวจให้คะแนนแบบประเพณีนิยมกับวิธีให้คะแนนความรู้บางส่วน ซึ่งประกอบด้วยวิธีประยุกต์การให้คะแนนของคูมบ์ และวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท กลุ่มตัวอย่างที่ใช้ในการวิจัย ได้แก่ นิสิตคณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ชั้นปีที่ 3 ที่ลงทะเบียนเรียน วิชาการประเมินผลการเรียนการสอน ในภาคการศึกษาต้น ปีการศึกษา 2544 จำนวน 297 คน เครื่องมือที่ใช้ในการวิจัยครั้งนี้เป็นแบบสอบที่ใช้เป็นแบบวัดผลสัมฤทธิ์ทางการเรียน วิชาการประเมินผลการเรียนการสอน ที่มีวิธีการตรวจให้คะแนน 3 วิธีดังกล่าวจากแบบสอบชุดเดียวกัน เก็บข้อมูลโดยให้ทำข้อสอบวัดผลสัมฤทธิ์ทางการเรียน วิชาการประเมินผลการเรียนการสอนที่สร้างขึ้น 3 ครั้ง ด้วยแบบสอบชุดเดียวกัน แต่วิธีการตอบแตกต่างกัน โดยการเก็บข้อมูลแต่ละครั้งห่างกัน 2 สัปดาห์ วิเคราะห์ข้อมูลโดยใช้โปรแกรม BILOG สำหรับการให้คะแนนแบบประเพณีนิยม และใช้โปรแกรม PARSCALE สำหรับการให้คะแนนด้วยวิธีประยุกต์การให้คะแนนของคูมบ์และวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท

ผลการวิจัยสรุปได้ดังนี้

1. การเปรียบเทียบความตรงตามสภาพ ระหว่างวิธีการตรวจให้คะแนนที่แตกต่างกัน พบว่าทั้ง 3 วิธีมีค่าความตรงตามสภาพ ไม่ความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ

2. การเปรียบเทียบอำนาจจำแนกและความยากระหว่างวิธีการตรวจให้คะแนนที่แตกต่างกัน พบว่าวิธีประยุกต์การให้คะแนนของคูมบ์มีค่าอำนาจจำแนกเฉลี่ยสูงที่สุดและมีค่าความยากเฉลี่ยต่ำที่สุด

3. การเปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบสอบ และอัตราส่วนสารสนเทศระหว่างวิธีการตรวจให้คะแนนที่แตกต่างกัน พบว่าวิธีประยุกต์การให้คะแนนของคูมบ์ มีค่าฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบสอบ และอัตราส่วนสารสนเทศเฉลี่ยสูงที่สุด รองลงมาคือวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท ส่วนวิธีประเพณีนิยมมีค่าต่ำที่สุด

คำชี้แจงของผู้วิจัย

การวัดและประเมินผลเป็นเรื่องที่มีความสัมพันธ์กับการจัดการเรียนการสอน ตามพระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542 ได้ให้สถานศึกษาจัดการประเมินผู้เรียน โดยพิจารณาพัฒนาการของผู้เรียนควบคู่ไปในกระบวนการเรียนการสอนตามความเหมาะสมของแต่ละระดับและรูปแบบการศึกษา ดังนั้น การวัดผลที่สามารถให้สารสนเทศที่เพียงพอที่จะบอกได้ว่าผู้เรียนมีความรู้จริง (full knowledge) ผู้เรียนที่มีความรู้บางส่วน (partial knowledge) และผู้เรียนที่ไม่รู้ (absence of knowledge) จึงเป็นเรื่องหนึ่งที่ควรให้ความสนใจเพราะจะช่วยสนับสนุนให้การวัดพัฒนาการเป็นไปตามความสามารถที่ใกล้เคียงที่สุดของผู้เรียนได้ จึงเป็นที่มาของการทำวิจัย เรื่อง การเปรียบเทียบคุณภาพของแบบสอบเลือกตอบระหว่างวิธีการให้คะแนนความรู้บางส่วนกับวิธีประเพณีนิยม

ในประเทศไทย ได้มีงานวิจัยที่ศึกษาและกำลังศึกษาเรื่องเกี่ยวกับการให้คะแนนความรู้บางส่วนอยู่บ้าง แต่ผลที่ได้ยังไม่ทำให้เกิดการนำไปใช้ในทางปฏิบัติ อาจเป็นเพราะเรื่องรูปแบบข้อสอบวิธีการให้คะแนน ซึ่งมีความยุ่งยากและสลับซับซ้อนกว่าข้อสอบที่เคยสร้างและใช้อยู่ แต่จากการศึกษาและผลการวิจัยในต่างประเทศที่ผ่านมา ผู้วิจัยมีความเห็นว่าควรให้ความสนใจวิธีการวัดความรู้บางส่วนที่มีความยุ่งยากน้อยที่สุด และมีความเป็นไปได้ในการนำไปใช้สำหรับผู้เรียน วิธีที่เสนอ คือ วิธีให้คะแนนความรู้บางส่วนประเภทวิธีการให้คะแนนจากการตัดตัวลวง(Elimination Scoring, ET) หรือวิธีการเลือกชุดย่อยของคำตอบ (The Subset Selection Method, SST) เนื่องจากทั้ง 2 วิธีนี้ใช้รูปแบบข้อสอบเหมือนแบบสอบชนิดเลือกตอบทั่วไป เพียงแต่คำสั่ง วิธีการตอบ และการให้คะแนนแตกต่างจากแบบสอบชนิดเลือกตอบธรรมดา จึงเป็นวิธีที่น่าจะศึกษาถึงคุณภาพของแบบสอบ เพื่อให้ได้สารสนเทศสูงสุด อันจะเป็นประโยชน์ต่อการนำไปใช้วัดที่มีประสิทธิภาพยิ่งขึ้นต่อไป

รายงานวิจัยเรื่องนี้สำเร็จลงได้ ด้วยความกรุณาจากท่านรศ.ดร.ศิริชัย กาญจนวาสี อาจารย์อาวุโส ที่ได้ให้การช่วยเหลือ แนะนำเป็นอย่างดี ตลอดกระบวนการทำวิจัย ซึ่งผู้วิจัยซาบซึ้งเป็นอย่างยิ่ง และขอกราบขอบพระคุณมา ณ ที่นี้ด้วย ขอขอบคุณสำนักบริหารวิชาการ จุฬาลงกรณ์มหาวิทยาลัย ที่สนับสนุนทุนพัฒนาอาจารย์ใหม่ / นักวิจัยใหม่ ทำให้ผู้วิจัยมีโอกาสในการทำงานวิจัยชิ้นนี้

ผู้วิจัยหวังเป็นอย่างยิ่งว่ารายงานการวิจัยฉบับนี้จะเป็นประโยชน์ในทางการศึกษาไม่มากก็น้อย หากมีข้อบกพร่องใดๆที่เกิดขึ้น ผู้วิจัยยินดีน้อมรับข้อติติงและข้อเสนอแนะ ด้วยความขอบคุณเป็นอย่างสูง

เอมอร จังศิริพรปกรณ์

มิถุนายน 2545

สารบัญ

	หน้า
บทคัดย่อ	i
คำชี้แจงของผู้วิจัย	ii
สารบัญ	iii
สารบัญตาราง	v
สารบัญภาพ	vi
บทที่	
1. บทนำ	1
ความเป็นมาและความสำคัญของปัญหา	1
วัตถุประสงค์ของการวิจัย	7
สมมติฐานในการวิจัย	7
ขอบเขตของการวิจัย	8
ข้อตกลงเบื้องต้น	8
คำจำกัดความที่ใช้ในการวิจัย	9
ประโยชน์ที่คาดว่าจะได้รับ	11
2. เอกสารและงานวิจัยที่เกี่ยวข้อง	12
ตอนที่ 1 ความรู้ทั่วไปเกี่ยวกับการวัดความรู้บางส่วน (partial knowledge)	12
ตอนที่ 2 มโนทัศน์เกี่ยวกับทฤษฎีการตอบสนองข้อสอบ	20
ตอนที่ 3 งานวิจัยที่เกี่ยวข้อง	34
ตอนที่ 4 โปรแกรมคอมพิวเตอร์ที่ใช้ในการวิเคราะห์	40
3. วิธีดำเนินการวิจัย	54
ประชากรและกลุ่มตัวอย่าง	54
เครื่องมือที่ใช้ในการวิจัย	54
คุณภาพของเครื่องมือ	55
การเก็บรวบรวมข้อมูล	58
การวิเคราะห์ข้อมูล	59
4. ผลการวิเคราะห์ข้อมูล	
ตอนที่ 1 ผลการวิเคราะห์ข้อมูลเบื้องต้น	62
ตอนที่ 2 ผลการเปรียบเทียบความตรงตามสภาพ ระหว่างวิธีการตรวจให้คะแนนที่แตกต่างกัน	66

สารบัญ (ต่อ)

	หน้า
ตอนที่ 3 ผลการเปรียบเทียบความยาก และอำนาจจำแนก ระหว่างวิธีการตรวจให้คะแนนที่แตกต่างกัน	67
ตอนที่ 4 ผลการเปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบสอบและอัตราส่วนสารสนเทศระหว่างวิธีการตรวจให้คะแนนที่แตกต่างกัน	70
5. บทสรุป อภิปรายและข้อเสนอแนะ	77
สรุปผลการวิจัย	78
อภิปรายผล	81
ข้อเสนอแนะ	85
รายการอ้างอิง	86
ภาคผนวก	92
ภาคผนวก ก. ฟังก์ชันสารสนเทศรายข้อ(IIF) และฟังก์ชันสารสนเทศของแบบสอบ (TIF) ของข้อสอบจำแนกตามระดับความสามารถ (θ)	93
ภาคผนวก ข. รูปแบบการตอบของกลุ่มตัวอย่างในการแสดงความรู้บางส่วน จำแนกตามกลุ่มที่มีความสามารถสูง ปานกลาง ต่ำ	112
ภาคผนวก ค. ตัวอย่างคำสั่งที่ใช้ในการวิเคราะห์ข้อมูล	125
ภาคผนวก ง. ตัวอย่างคำชี้แจงของแบบสอบที่มีการให้คะแนนความรู้บางส่วน	130
ประวัติผู้วิจัย	135

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญตาราง

ตาราง	หน้า
2.1 สรุปผลการศึกษาในต่างประเทศเพื่อเปรียบเทียบค่าความตรงและค่าความเที่ยง ระหว่างวิธี การ ให้ความรู้บางส่วนกับวิธีประเพณีนิยม (NC)	37
3.1 ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ของคะแนนความสามารถ และF – test ตามวิธีประยุกต์ การให้คะแนนของคูมบ์จำแนกตามกลุ่มความสามารถ	56
3.2 ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐานของคะแนนความสามารถ และF – test ตามวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิทจำแนกตามกลุ่มความสามารถ	56
3.3 ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ของคะแนนความสามารถ และF – test ตามวิธีให้คะแนนแบบประเพณีนิยมจำแนกตามกลุ่มความสามารถ	57
4.1 ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบจำแนกตามวิธีการให้คะแนน	63
4.2 ค่า t ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของคะแนนความแตกต่างระหว่างวิธีการให้คะแนนที่ต่างกัน	63
4.3 ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าความสามารถของผู้สอบจำแนกตามวิธีการให้คะแนน	64
4.4 ค่า t ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าความแตกต่างของค่าความสามารถระหว่างวิธีการให้คะแนนที่ต่างกัน	64
4.5 ร้อยละของค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของการตัดตัวลวง(วิธีประยุกต์การให้คะแนนของคูมบ์) จำแนกตามกลุ่มความสามารถ	65
4.6 ร้อยละของค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของการเลือกชุดตัวถูก (วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท) จำแนกตามกลุ่มความสามารถ	65
4.7 ความตรงตามสภาพจากการสอบจำแนกตามวิธีการตรวจให้คะแนน	66
4.8 เปรียบเทียบความแตกต่างของความตรงตามสภาพ โดยใช้สถิติ Z จำแนกตามวิธีการตรวจให้ คะแนน	67
4.9 ค่าอำนาจจำแนกและค่าความยากของข้อสอบ จำแนกตามรายข้อและวิธีการตรวจให้คะแนน	67
4.10 ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐานของค่าความยากและค่าอำนาจจำแนก และค่าสถิติ F จำแนกตามวิธีการตรวจให้คะแนน	69

สารบัญตาราง (ต่อ) หน้า

4.11 ฟังก์ชันสารสนเทศของข้อสอบ จำแนกตามรายข้อและวิธีการตรวจให้ คะแนน 71
4.12 ฟังก์ชันสารสนเทศของข้อสอบเฉลี่ย จำแนกตามความสามารถของผู้สอบและวิธีการตรวจให้คะแนน 73
4.13 เปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบโดยเฉลี่ยจำแนกตามวิธีการตรวจให้คะแนน 75
4.14 ฟังก์ชันสารสนเทศของแบบสอบ จำแนกตามความวิธีการตรวจให้คะแนน 75
4.15 อัตราส่วนสารสนเทศเฉลี่ยจำแนกตามความวิธีการตรวจให้คะแนน 76
ก 1 ฟังก์ชันสารสนเทศรายข้อ(IIF) และฟังก์ชันสารสนเทศของแบบสอบ (TIF) ของข้อสอบที่ระดับความสามารถ (θ) –4.00 จำแนกตามวิธีการให้คะแนน 94
ก 2 ฟังก์ชันสารสนเทศรายข้อ(IIF) และฟังก์ชันสารสนเทศของแบบสอบ (TIF) ของข้อสอบที่ระดับความสามารถ (θ) –3.00 จำแนกตามวิธีการให้คะแนน 96
ก 3 ฟังก์ชันสารสนเทศรายข้อ(IIF) และฟังก์ชันสารสนเทศของแบบสอบ (TIF) ของข้อสอบที่ระดับความสามารถ (θ) –2.00 จำแนกตามวิธีการให้คะแนน 98
ก 4 ฟังก์ชันสารสนเทศรายข้อ(IIF) และฟังก์ชันสารสนเทศของแบบสอบ (TIF) ของข้อสอบที่ระดับความสามารถ (θ) –1.00 จำแนกตามวิธีการให้คะแนน 100
ก 5 ฟังก์ชันสารสนเทศรายข้อ(IIF) และฟังก์ชันสารสนเทศของแบบสอบ (TIF) ของข้อสอบที่ระดับความสามารถ (θ) 0.00 จำแนกตามวิธีการให้คะแนน 102
ก 6 ฟังก์ชันสารสนเทศรายข้อ(IIF) และฟังก์ชันสารสนเทศของแบบสอบ (TIF) ของข้อสอบที่ระดับความสามารถ (θ) –1.00 จำแนกตามวิธีการให้คะแนน 104
ก 7 ฟังก์ชันสารสนเทศรายข้อ(IIF) และฟังก์ชันสารสนเทศของแบบสอบ (TIF) ของข้อสอบที่ระดับความสามารถ (θ) –2.00 จำแนกตามวิธีการให้คะแนน 106
ก 8 ฟังก์ชันสารสนเทศรายข้อ(IIF) และฟังก์ชันสารสนเทศของแบบสอบ (TIF) ของข้อสอบที่ระดับความสามารถ (θ) –3.00 จำแนกตามวิธีการให้คะแนน 108
ก 9 ฟังก์ชันสารสนเทศรายข้อ(IIF) และฟังก์ชันสารสนเทศของแบบสอบ (TIF) ของข้อสอบที่ระดับความสามารถ (θ) –4.00 จำแนกตามวิธีการให้คะแนน 110
ข 1 จำนวนและร้อยละของกลุ่มตัวอย่างที่มีรูปแบบการตัดตัวลวงที่ต่างกันจำแนกตามความสามารถของกลุ่มตัวอย่าง และค่าความยากรายข้อ 113
ข 2 จำนวนและร้อยละของกลุ่มตัวอย่างที่มีรูปแบบการเลือกชุดตัวถูกที่ต่างกัน จำแนกตามความสามารถของกลุ่มตัวอย่าง และค่าความยากรายข้อ 119

สารบัญภาพ

แผนภาพที่		หน้า
2.1	โค้งลักษณะเฉพาะของข้อสอบของการวิเคราะห์ข้อสอบแบบ 1 พารามิเตอร์	25
2.2	โค้งลักษณะเฉพาะของข้อสอบของการวิเคราะห์ข้อสอบแบบ 2 พารามิเตอร์	26
2.3	โค้งลักษณะเฉพาะของข้อสอบของการวิเคราะห์ข้อสอบแบบ 3 พารามิเตอร์	27
2.4	โค้งฟังก์ชันสารสนเทศของ GRM	30
2.5	โค้งฟังก์ชันสารสนเทศของ GPCM	32
4.1	เปรียบเทียบโค้งสารสนเทศของข้อสอบ ของการให้คะแนน 3 วิธี	74

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

ข้อสอบแบบเลือกตอบเป็นข้อสอบที่ใช้กันอย่างแพร่หลายในปัจจุบัน เนื่องจากสามารถวัดความรู้ความสามารถและวัดผลสัมฤทธิ์ได้อย่างกว้างขวาง ประกอบกับมีความเป็นปรนัยในการตรวจให้คะแนน ใช้เวลาตรวจน้อย สามารถใช้เครื่องคอมพิวเตอร์ตรวจได้ ผลการวิเคราะห์ข้อสอบสามารถนำไปใช้วินิจฉัยปัญหาการเรียนรู้ได้ แต่ก็ยังมีจุดอ่อนที่สำคัญ คือ เปิดโอกาสให้มีการตอบถูกโดยการเดา ปัญหาการตอบแบบเดาสุ่มจึงเป็นแหล่งความคลาดเคลื่อนในการวัดนอกจากนี้ผลการวัดที่ได้ไม่สามารถให้สารสนเทศที่เพียงพอที่จะจำแนกผู้ตอบในระดับต่างๆของความรู้ระหว่างผู้ที่รู้จริง (full knowledge) ผู้ที่มีความรู้บางส่วน (partial knowledge) และผู้ที่ไม่รู้ (absence of knowledge)

จากข้อจำกัดของแบบสอบเลือกตอบดังกล่าว นักวิชาการจึงได้ศึกษาหาวิธีการแก้ปัญหา เพื่อลดโอกาสในการเดา และเพิ่มสารสนเทศเกี่ยวกับการตอบข้อสอบให้มากที่สุด โดยใช้วิธีการให้คะแนน ซึ่งเป็นองค์ประกอบที่สำคัญของกระบวนการวัด สำหรับวิธีการให้คะแนนที่เหมาะสมจะทำให้คุณสมบัติของการวัดในประเด็นที่เกี่ยวข้องกับความตรงและความเที่ยงสูงขึ้น การให้คะแนนความรู้บางส่วนของผู้ตอบแสดงถึงตำแหน่งจริงของความรู้ของผู้ตอบ ซึ่งมีส่วนช่วยลดความคลาดเคลื่อนจากแหล่งการเดาสุ่มของผู้ตอบลง ทำให้คะแนนมีความเที่ยงมากขึ้น ในขณะเดียวกันก็ส่งผลต่อความตรงของกระบวนการวัดด้วย (Frary ,1980 อ้างถึงในพรทิพย์ ไชยโส , 2533) จากการศึกษาวิธีการให้คะแนนความรู้บางส่วนสรุปเป็นประเภทใหญ่ๆ ได้ 4 วิธี ดังนี้ (Simon , Budescu และ Nevo 1997)

1. การให้น้ำหนักแก่ข้อสอบที่แตกต่างกัน (Differential Item Weighting)
2. การให้น้ำหนักแก่ตัวเลือกที่แตกต่างกัน (Differential Option Weighting)
3. การเปลี่ยนแปลงโครงสร้างของข้อสอบ (Changing the Item Structure)
4. การเปลี่ยนแปลงวิธีการตอบ (Changing the Response Method)

โดยแต่ละวิธีมีรายละเอียดดังนี้

1.การให้น้ำหนักแก่ข้อสอบที่แตกต่างกัน (Differential Item Weighting)

หลักการของวิธีนี้ คือข้อสอบที่มีคุณภาพจะได้น้ำหนักมากกว่าข้อสอบที่ไม่มีคุณภาพโดยใช้ผลที่ได้จากการวิเคราะห์ข้อสอบเป็นเกณฑ์ คือค่าความยาก ความสามารถ ความตรง ความแปรปรวน หรือการพิจารณาของผู้เชี่ยวชาญ Stanley และ Wang (1970)ได้ให้ข้อสรุปว่า การให้น้ำหนักแก่ข้อสอบโดยใช้ทฤษฎีการวัดแบบดั้งเดิม ทำให้น้ำหนักของคะแนนแต่ละข้อมีค่าไม่คงที่ เปลี่ยนแปลงไปตามกลุ่มตัวอย่าง จึงควรให้คะแนนตามทฤษฎีการตอบสนองข้อสอบ (IRT) อย่างไรก็ตามการกำหนดน้ำหนัก

คะแนนที่เหมาะสมรายข้อนี้ ยังวางอยู่บนพื้นฐานของวิธีการตอบ และการให้คะแนนเฉพาะข้อที่ตอบถูกแบบเดิม ซึ่งไม่ได้มีการพิจารณาการให้คะแนนความรู้บางส่วนแก่ผู้ตอบ

2.การให้น้ำหนักคะแนนแก่ตัวเลือกที่แตกต่างกัน(Differential Option Weighting)

หลักการของวิธีนี้ คือ ความรู้บางส่วนสามารถถูกวัดได้จากการให้น้ำหนักคะแนนแก่ตัวเลือก แบ่งวิธีการได้ 2 วิธี คือวิธีแรกเป็นการให้น้ำหนักแก่ตัวเลือกโดยใช้วิจารณญาณของผู้เชี่ยวชาญ หรือจากทฤษฎีของโครงสร้างความรู้ (Smith, 1987) ส่วนวิธีที่สองเป็นการให้น้ำหนักจากข้อมูลเชิงประจักษ์ที่ผ่านมาในอดีต หรือปัจจุบันโดยอาศัยสิ่งที่น่าสนใจของตัวเลือก เช่นคะแนนมาตรฐานเฉลี่ยของแต่ละตัวเลือกที่ผู้สอบเลือก ความสัมพันธ์ระหว่างตัวเลือกที่ผู้สอบเลือกกับคะแนนรวมทั้งหมด เป็นต้น

วิธีการนี้ได้รับการศึกษาเพิ่มมากขึ้น จากผลการศึกษาของ Frary (1989), Haladyna (1988), Echternacht (1976), Hakstian และ Kansup (1975), Stanley และ Wang (1970) พบว่าการให้น้ำหนักคะแนนแก่ตัวเลือก มีประโยชน์บ้างในแง่ที่ว่าช่วยเพิ่มค่าความเที่ยงคงที่ภายใน (internal consistency reliability) การเปรียบเทียบค่าความตรงและความเที่ยงระหว่างแบบการให้น้ำหนักแก่ตัวเลือกกับแบบธรรมดา พบว่ามีค่าไม่คงที่ของความแตกต่างอย่างมีนัยสำคัญ วิธีการเหล่านี้ไม่เป็นที่นิยม เนื่องจากการพัฒนาเพื่อให้ได้มาซึ่งน้ำหนักใช้ทรัพยากรมาก การคำนวณการให้คะแนนที่มีวิธีการที่ยุ่งยาก และยากในการอธิบายและตัดสินการให้คะแนนแก่ผู้สอบ (Frary, 1989)

3.การเปลี่ยนแปลงโครงสร้างของข้อสอบ (Changing the Item Structure)

วิธีนี้มีรูปแบบของโครงสร้างข้อสอบ และ/หรือ คำอธิบายที่แตกต่างจาก แบบเลือกตอบธรรมดา ตัวอย่างที่สำคัญ มีดังนี้

- ข้อสอบที่มี 2 ตัวเลือก คือ ถูก และ ผิด ผู้ตอบจะต้องเลือกคำตอบอย่างใดอย่างหนึ่ง คะแนนที่ได้ได้จากผลรวมของจำนวนคำตอบที่ถูกต้อง เนื่องจากการตอบแต่ละข้อใช้เวลาน้อย จึงต้องเพิ่มจำนวนข้อให้มากขึ้นกว่าข้อสอบแบบเลือกตอบธรรมดา

- ข้อสอบที่มีตัวถูกหลาย ๆ ตัวเลือก (มากกว่า 1) โดยต้องมีคำชี้แจงอธิบายให้แก่ผู้สอบ ซึ่งอาจจะระบุจำนวนตัวเลือกที่ถูกหรือไม่ระบุก็ได้ คะแนนที่ได้เท่ากับผลรวมของการระบุตัวเลือกถูกได้ถูกต้อง ในบางกรณีมีการหักคะแนนการตอบที่ตอบตัวเลือกผิด เช่นวิธีของ Dressel และ Schmidt (1953) โดยมีจุดประสงค์เพื่อลดการเดา และประเมินความรู้บางส่วน แบบสอบลักษณะนี้ ได้มีการพัฒนาต่อมา ตัวอย่างเช่นแบบสอบที่มีตัวเลือกถูกทุกข้อ ผิดทุกข้อ หรือถูกเฉพาะตัวเลือกหนึ่งกับอีกตัวเลือกหนึ่ง (Frary ,1989; Hsu, Moss and Khampalikit, 1984)

- ข้อสอบที่มีกลุ่มของข้อสอบ และกลุ่มของคำตอบ ที่เรียกว่าข้อสอบแบบจับคู่ โดยผู้สอบต้องเลือกคำตอบที่ถูกคู่กับคำถามที่กำหนดให้ ถ้าจำนวนคำถามเท่ากับจำนวนคำตอบ เรียกว่า การจับคู่อย่างง่าย (simple matching) (Horst, 1966; Wesman,1971) แต่ถ้าจำนวนคำตอบมากกว่าจำนวนคำถาม เรียกว่า การจับคู่แบบพหุ (multiple matching) (Gulliken, 1986) ความแตกต่างของ 2 แบบ คือ การจับคู่แบบพหุ ช่วยลดโอกาสในการเดามากกว่า เนื่องจากมีจำนวนตัวเลือกมากกว่า แต่ข้อจำกัดที่สำคัญ

ของวิธีนี้ คือ ความยากในการสร้างข้อสอบให้มีความเหมาะสม Budescu (1988) พบว่า การจับคู่แบบพหุให้ค่าความตรง และความเที่ยงมากกว่าแบบสอบเลือกตอบธรรมดา และความตรงความเที่ยงจะเพิ่มขึ้นเมื่อสัดส่วนของจำนวนตัวเลือกเพิ่มขึ้น วิธีนี้ต้องการเวลาในการทำข้อสอบมากกว่าปกติ

4.การเปลี่ยนแปลงวิธีการตอบ (Changing the Response Method)

วิธีนี้แตกต่างจากวิธีก่อนที่กล่าวมาทั้งหมด คือ ใช้น้ำหนักที่ให้โดยผู้สอบเอง ซึ่งสะท้อนถึงความรู้ที่มีอยู่ในตัวผู้สอบ การประเมินด้วยตนเองนี้เป็นวิธีที่ดีที่สุดในการได้มาซึ่งสารสนเทศเกี่ยวกับความรู้ของผู้สอบในแต่ละข้อ ผู้สอบจะให้สารสนเทศอย่างสมบูรณ์เกี่ยวกับน้ำหนัก วิธีการให้คะแนนและการควบคุมระดับความมั่นใจในการให้ ซึ่งจะช่วยลดการเดา และความคลาดเคลื่อนในการวัด วิธีนี้ใช้รูปแบบข้อสอบเหมือนแบบสอบชนิดเลือกตอบทั่วไป เพียงแต่คำสั่ง วิธีการตอบ และการให้คะแนนแตกต่างจากแบบสอบชนิดเลือกตอบธรรมดา ซึ่งแต่ละวิธีใช้วิธีการที่ไม่ยุ่งยาก โดย แบ่งเป็นชนิดต่าง ๆ ได้ 6 วิธี คือ

4.1 วิธีการแก้การเดา (Correction for guessing)

วิธีการนี้ให้ผู้สอบทำข้อสอบเหมือนกับแบบสอบชนิดเลือกตอบธรรมดา แต่ต่างกันที่การให้คะแนน วิธีการแก้การเดา มีวิธีที่น่าสนใจ 2 วิธี คือ

1) วิธีการแก้การเดาโดยการลงโทษ เป็นวิธีที่มีข้อตกลงเบื้องต้นว่าข้อที่ผู้สอบตอบผิดเป็นผลมาจากการเดา ดังนั้นข้อที่ผู้สอบตอบผิดแทนที่จะได้ 0 แต่จะให้คะแนนติดลบ -1/(K-1) K คือจำนวนตัวเลือกของแบบสอบ (Lord ,1975)

2) วิธีการให้รางวัลแก่ข้อที่ไม่ตอบ เป็นวิธีที่มีข้อตกลงเบื้องต้นว่า ข้อที่ผู้ตอบเว้นไว้นั้นถ้าเขาตอบจะมีโอกาสตอบถูกเป็น 1/K เมื่อ K คือจำนวนตัวเลือกของแบบสอบ (Traub and Hambleton, 1972)

4.2 วิธีการให้คะแนนจากการตัดตัวลวง (Elimination Scoring)

วิธีนี้เป็นการให้ผู้สอบตัดตัวลวงออกไปได้บ้างอย่างถูกต้อง แม้จะไม่ทราบว่าคำตอบที่ถูกต้องคืออะไร มีวิธีการและงานวิจัยที่สนับสนุนวิธีนี้ ดังนี้

1) วิธีการตอบและให้คะแนนแบบคูมบ์ (The Coombs Response/Scoring method) วิธีการตอบและการให้คะแนนแบบคูมบ์ ใช้รูปแบบของแบบสอบเลือกตอบที่มีคำตอบถูกเพียงคำตอบเดียว โดยให้ผู้สอบตัดตัวลวงที่ไม่ใช่คำตอบถูกที่เขาแน่ใจออกไปมากที่สุด การให้คะแนนผู้ตอบจะได้รับ 1 คะแนน ในแต่ละตัวลวงที่ตัดออกอย่างถูกต้อง และในกรณีที่ผู้สอบมีความรู้ที่ผิดไปตัดตัวถูกออกโดยเข้าใจว่าเป็นตัวลวงจะถูกตัดคะแนน K-1 โดยที่ K คือ จำนวนตัวเลือก ดังนั้นช่วงคะแนนในแต่ละข้ออยู่ระหว่าง -(K-1) ถึง (K-1)

จากงานวิจัยที่ศึกษาเปรียบเทียบวิธีให้คะแนนแบบการตัดตัวลวง (ET) วิธีการแก้การเดา (CG) และวิธีแบบเลือกตอบธรรมดา พบว่า ความตรงและความเที่ยงของวิธีการให้คะแนนแบบการตัดตัวลวงสูงที่สุด (Collet, 1971; Coombs et al, 1956; Jaradat and Tollefson, 1988)

2) วิธีการตอบและการให้คะแนนแบบอาร์โนลด์ (The Arnold Scoring/ Response Method) (Arnold and Arnold 1970) อาร์โนลด์ เสนอวิธีที่พัฒนาจากทฤษฎีเกมเบื้องต้น (Game theory) มีการใช้ข้อสอบแบบเลือกตอบเช่นเดียวกับแบบของคูมบ์ โดยให้ผู้ตอบตัดตัวเลือกที่แน่ใจว่าเป็นตัวลวงหรือคำตอบผิดออกไปมากที่สุดที่เขารู้ แต่วิธีนี้มีระบบการให้คะแนนที่แตกต่างจากของคูมบ์ คือจะให้คะแนน 1/(K-1) คะแนนสำหรับการตัดตัวลวงถูกต้อง 1 ตัว แต่ถ้าผู้สอบตัดตัวเลือกถูกออกโดยเข้าใจว่าเป็นตัวลวงจะถูกหัก -1/(K-1) คะแนน เมื่อ K คือจำนวนตัวเลือกแต่ละข้อ ระบบการให้คะแนนของอาร์โนลด์วางอยู่พื้นฐานคะแนนที่ถูกคาดหวัง (expected item score) เท่ากับ 0 เมื่อผู้สอบเดา

3) วิธีการตอบและให้คะแนนแบบครอส (The Cross Response/ Scoring Method, CRS) วิธีการนี้ให้ผู้สอบตัดตัวลวงออกเช่นเดียวกับวิธีการของคูมบ์และอาร์โนลด์ แต่การให้คะแนนแตกต่างกัน คือวิธีการให้คะแนนของครอสจะให้คะแนนรายข้อเฉลี่ยต่ำสุด เมื่อผู้ตอบตัดตัวเลือกที่เป็นคำตอบโดยเข้าใจว่าเป็นตัวลวงและตัดตัวลวงอีก K-2 ตัวโดยเข้าใจว่าเป็นชุดของตัวเลือกที่ผิด ซึ่งผิดกับการให้คะแนนแบบคูมบ์ที่คะแนนรายข้อเฉลี่ยต่ำสุดจะเกิดจากการตัดตัวเลือกที่ถูกว่าเป็นตัวลวงเพียงคำตอบเดียว การให้คะแนนแบบครอสผู้สอบจะได้ 2 คะแนนในแต่ละตัวลวงที่เขาตัดออกได้อย่างถูกต้องและจะได้ 1คะแนนสำหรับแต่ละตัวเลือกที่ไม่ได้ตัดว่าเป็นตัวลวง เมื่อคำตอบถูกยังไม่ได้ถูกตัด แต่เมื่อใดที่ผู้ตอบตัดตัวเลือกที่ถูกด้วย misinformation เขาจะได้คะแนนเฉพาะตัวเลือกที่ไม่ได้ตัดตัวเลือกละ 1 คะแนน สำหรับในกรณีที่ข้อสอบมี 4 ตัวเลือก คะแนนที่ยอมให้เกิดขึ้นได้มีค่าตั้งแต่ 1 คะแนน ถึง 7 คะแนน และคะแนนต่ำสุดจะเกิดจากการที่ผู้ตอบตัดตัวเลือกที่ถูกต้องหลังจากการตัดตัวลวงมาแล้ว 2 ตัว Frary (1980) เปรียบเทียบวิธีการให้คะแนนแบบครอสกับวิธีการให้คะแนนอื่นๆ พบว่าคะแนนแบบครอสให้ค่าความเที่ยงต่ำกว่าวิธีการให้คะแนนของอาร์โนลด์และคูมบ์ ส่วนความตรงตามเกณฑ์ไม่แตกต่างกัน

4) วิธีการเลือกชุดย่อยของคำตอบ (The Subset Selection Method, SST) เสนอโดย Dressel และ Schmidt (1953) เป็นวิธีที่ตรงข้ามกับวิธีให้คะแนนแบบการตัดตัวลวง (ET) คือ วิธีนี้ให้ผู้สอบเลือกชุดย่อยของคำตอบที่เขาเห็นว่าน่าจะมีตัวเลือกที่เป็นคำตอบถูกรวมอยู่ด้วย การให้คะแนนขึ้นอยู่กับจำนวนตัวเลือกที่มีอยู่ในชุดย่อยของคำตอบ ยิ่งน้อยเท่าใดคะแนนจะได้เพิ่มขึ้น ในกรณีที่ตัวเลือกที่ถูกต้องรวมอยู่ในชุดย่อยของคำตอบนั้นด้วย แต่ถ้าชุดย่อยนั้นไม่มีคำตอบที่ถูกต้องอยู่ด้วย คะแนนจะถูกหักออกไปเท่ากับขนาดของจำนวนตัวเลือกที่ผู้ตอบเลือก ในกรณีที่ผู้ตอบไม่เลือกตัวเลือกใดเลยจะได้ 0 คะแนน วิธีการนี้ Gibbon, Olkin และ Sobel (1979) ได้ศึกษาพบว่า เป็นวิธีที่ช่วยลดการเดาสุ่มของผู้ตอบลง และยังยอมให้ผู้ตอบได้แสดงความรู้บางส่วนของผู้ตอบด้วย

Jaradat และ Swagad (1986) เปรียบเทียบวิธีการให้คะแนน 3 วิธี คือ วิธีการแก้การเดา (CG), วิธีการเลือกชุดย่อยของคำตอบ (SST) และ วิธีประเพณีนิยม (NC, number correct) โดยใช้แบบทดสอบคู่ขนาน 3 ฉบับ พบว่า SST มีความตรงสูงที่สุด และมีความเที่ยงสูงกว่า NC นอกจากนี้ Jaradat และ Tollefson (1988) พบว่า ET และ SST มีความตรงและความเที่ยงสูงกว่า CG

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

4.3 วิธีการของการทดสอบความน่าจะเป็น (Probability Testing)

วิธีนี้มีความยืดหยุ่น และเปิดโอกาสให้ผู้สอบแสดงความรู้บางส่วนมากที่สุด โดยรายงานความน่าจะเป็นในตัวเลือกที่ผู้สอบคิดว่าเป็นคำตอบที่ถูก โดยมีตำแหน่งคะแนนที่เป็นไปได้ 101 ตำแหน่ง (ช่วงจาก 0-100) De Finetti (1965) ได้ศึกษาวิธีการที่ให้ผู้สอบให้คะแนนความน่าจะเป็นแก่ตัวเลือกที่คิดว่าถูก โดยคาดว่าจะให้คะแนนในคำตอบถูกสูงที่สุด เรียกการให้คะแนนความน่าจะเป็นนี้ว่า Proper Reproducting Scoring Systems (PRSSs) ต่อมามีการศึกษารูปแบบของ PRSSs โดย Shuford, Albert และ Massengill (1966) 3 รูปแบบ คือ quadratic, spherical และ logarithmic ซึ่งมีตัวเลือก > 2 ตัว แต่วิธีการนี้มีความยุ่งยาก ซับซ้อนในการตอบ และการให้คะแนนมาก

Rippey (1970) และ Suhadolnik และ Weiss (1983) พบว่าการให้คะแนนความน่าจะเป็นแบบธรรมดามีความเที่ยงมากกว่าการให้คะแนน 3 รูปแบบดังกล่าว แต่ Kansub และ Hakstian (1975) พบว่าการให้คะแนนแบบ quadratic และ logarithmic มีค่าความเที่ยงมากกว่าการให้คะแนนความน่าจะเป็นธรรมดา

นอกจากนี้ Michael (1968) Pugh และ Brunza (1975) พบว่าวิธีการให้คะแนนความน่าจะเป็น (PT) มีความเที่ยงสูงกว่าวิธีเลือกตอบธรรมดา ส่วน Hambleton, Roberts และ Traub (1970) รายงานว่ามีความตรงสูงกว่าวิธีเลือกตอบธรรมดา ในขณะที่ Koehler (1971) และ Hakstian และ Kansup (1975) พบว่ามีความไม่แตกต่างกันของความเที่ยงและความตรงระหว่าง NC และ PT ซึ่งยังไม่ได้ข้อสรุปที่ชัดเจน

4.4 วิธีการแสดงความมั่นใจ (Confidence Marking)

วิธีนี้คิดค้นโดย Dressel และ Schmidt (1953) โดยที่ผู้สอบถูกถามให้แสดงความมั่นใจในคำตอบที่คิดว่าถูก โดยใช้ C-point scale (โดยทั่วไป $3 \leq c \leq 5$) หรืออาจเป็นการบอกความมั่นใจเป็นภาษา คือ ไม่แน่ใจ ค่อนข้างแน่ใจ แน่ใจมาก

ผลการวิจัยพบว่าวิธีการนี้ทำให้ความตรงของแบบสอบสูงขึ้นกว่าการให้คะแนนแบบ 0, 1 (Dressel และ Schmidt, 1953; Hopkins, Hakstian และ Hopkins, 1973) แต่วิธีนี้ต้องการเวลาในการตอบและการให้คะแนนมากกว่าเดิม

4.5 วิธีการจัดอันดับอย่างสมบูรณ์ (Complete Ordering)

วิธีการนี้เป็นวิธีเฉพาะของ PT คือ แทนที่จะให้แสดงความน่าจะเป็นในแต่ละตัวเลือก ผู้สอบจะต้องจัดอันดับของตัวเลือก ซึ่งทำให้ง่ายขึ้นต่อการนำไปใช้ ทำให้สามารถจำแนกระดับความรู้เป็น 3 ระดับ คือ

1. Full knowledge คือ การจัดอันดับให้คำตอบถูกอยู่ในตำแหน่งที่สูงที่สุด
2. Partial knowledge คือ การจัดอันดับให้คำตอบถูกอยู่ในตำแหน่งกลาง
3. Absence of knowledge คือ การจัดอันดับให้คำตอบถูกอยู่ในตำแหน่งต่ำที่สุด

4.6 วิธีการจัดอันดับเป็นบางส่วน (Partial Ordering)

วิธีนี้เป็นวิธีที่ผสมระหว่าง วิธีให้คะแนนแบบการตัดตัวลวง (ET) และ วิธีการจัดอันดับอย่างสมบูรณ์ (CO) คือให้ผู้สอบจัดอันดับตัวเลือกที่ไม่สามารถตัดออกไปได้ วิธีนี้เสนอโดย De Finetti (1965) และถูกนำไปใช้โดย Diamond (1975)

จากการศึกษาทฤษฎีและงานวิจัยที่ผ่านมา พบว่าส่วนใหญ่จะเปรียบเทียบคุณภาพของแบบสอบในเรื่องความตรงและความเที่ยงจากวิธีการให้คะแนนความรู้บางส่วนหลายๆวิธี โดยใช้ทฤษฎีการวัดแบบดั้งเดิม ซึ่งมีข้อจำกัดในเรื่องแหล่งความคลาดเคลื่อนในการวัด และค่าสถิติที่ได้แปรเปลี่ยนไปตามกลุ่มตัวอย่างที่ใช้ ในปัจจุบันนักวัดผลได้พัฒนาทฤษฎีการตอบสนองข้อสอบ(IRT) เพื่อแก้ปัญหาการวิเคราะห์ตามแนวทฤษฎีการทดสอบแบบดั้งเดิม การพัฒนาในระยะแรกมุ่งนำไปวิเคราะห์กับแบบสอบที่ตรวจให้คะแนนแบบทวิวิภาค แต่ต่อมามีการประยุกต์ใช้กับแบบสอบที่ตรวจให้คะแนนแบบพหุวิภาคที่มีอยู่ด้วยกันหลายโมเดล เช่น Graded Response Model (GRM) ของซามิจิมา (Samejima) Nominal Response Model (NRM) พัฒนาโดยบอค (Bock) Rating Scale Model (RCM) พัฒนาโดยมาสเตอร์ (Master) Generalized Partial Credit Model(GPCM) เป็นโมเดลที่พัฒนาโดย มูราคิ (Muraki) แต่โมเดลที่ได้รับการพัฒนาและใช้อย่างแพร่หลายมากในปัจจุบัน คือ GRMและ GPCM (De Ayala,1994; Donoghue,1994; Muraki,1993; Reise and Yu,1990; Dodd , Koch and De Ayala,1989; Koch,1983) เนื่องจากโมเดลดังกล่าวไม่เข้มงวดกับข้อตกลงเบื้องต้นของความเป็นเอกมิติของแบบวัด และการประมาณค่าพารามิเตอร์มีค่าอำนาจจำแนกรายข้อในฟังก์ชันด้วย การวิเคราะห์ทั้ง 2 โมเดลจะได้สารสนเทศที่สำคัญ คือฟังก์ชันสารสนเทศของแบบสอบ (Test information function) ฟังก์ชันสารสนเทศของข้อสอบ (Item information function) และฟังก์ชันสารสนเทศของลำดับขั้นคะแนน (item –category information)

ประเด็นที่ผู้วิจัยให้ความสนใจ คือ จากการศึกษาของอาร์โนลด์พบว่าในการเปรียบเทียบวิธีการให้คะแนนแต่ละวิธี ผู้ตอบที่ได้คะแนนสูงในวิธีหนึ่งก็จะได้คะแนนสูงในอีกวิธีหนึ่ง เช่นเดียวกับผู้ตอบที่ได้คะแนนต่ำ ผู้ตอบที่ได้คะแนนต่ำในวิธีหนึ่งก็จะได้คะแนนต่ำในอีกวิธีหนึ่ง และพบว่าการให้คะแนนความรู้บางส่วนสามารถจำแนกคะแนนในกลุ่มปานกลางได้ชัดเจนกว่า ทั้งนี้เนื่องจากการให้คะแนนความรู้บางส่วนได้ขจัดการเดาออกไป ดังนั้นในการวิจัยครั้งนี้จึงได้ให้ความสนใจที่จะเปรียบเทียบคุณภาพของแบบสอบ โดยใช้ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) ด้วยวิธีการให้คะแนนแบบประเพณีนิยมกับวิธีให้คะแนนความรู้บางส่วนจากวิธีการให้คะแนนแบบตัดตัวลวง (Elimination Scoring , ET) ตามวิธีการตอบและการให้คะแนนแบบคูมบ์และวิธีการเลือกชุดย่อยของคำตอบ (The Subset Selection Method, SST) ตามวิธีการตอบและการให้คะแนนแบบเดรสเซลและสมิท (Dressel and Schmidt) ว่ามีความแตกต่างกันอย่างไร เนื่องจากทั้ง 2 วิธีจากการวิจัยที่ผ่านมาพบว่ามีแนวโน้มในการเพิ่มคุณภาพของการจำแนกระดับความรู้ของผู้สอบ และลดการเดา ตลอดจนเป็นวิธีที่ให้ค่าความเที่ยงสูงสุด อย่างไรก็ตามเนื่องจากทั้ง 2 วิธี มีการให้คะแนนที่มีค่าติดลบในแต่ละข้อ ซึ่งถ้านำไป

หาคะแนนรวมของผู้ตอบแต่ละคน จะทำให้ไม่เหมาะสม เพราะคะแนนในข้อติดลบจะถูกนำไปหักจากคะแนนที่ได้ในข้ออื่นๆด้วย ดังนั้น ผู้วิจัยจึงได้ประยุกต์วิธีการให้คะแนนในการวิจัยครั้งนี้ 2 วิธี ที่พัฒนาจากสูตรของคูมบ์และ ของเดรสเชลและสมิท ที่ทำให้ไม่มีค่าติดลบและมีค่าสูงสุด เท่ากับ 3 ต่ำสุดเท่ากับ 0 ในแต่ละข้อ โดยที่ผู้วิจัยคาดว่าผลที่ได้จากการวิจัยครั้งนี้ จะทำให้ได้ข้อความรู้ที่จะช่วยให้ผู้ที่จะใช้เครื่องมือในการวัดผลการศึกษาได้ตระหนักถึงวิธีการให้คะแนนที่มีคุณภาพในกระบวนการวัดเพื่อนำไปใช้ในการตัดสินคุณค่าในการประเมินต่อไป

วัตถุประสงค์ของการวิจัย

1. เพื่อเปรียบเทียบคุณภาพของแบบสอบในด้านความตรงตามสภาพ ระหว่างวิธีการตรวจให้คะแนนแบบประเพณีนิยมกับวิธีการตรวจให้คะแนนความรู้บางส่วนตามวิธีประยุกต์การให้คะแนนของคูมบ์ และวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท

2.เพื่อเปรียบเทียบความยากและอำนาจจำแนกโดยใช้ทฤษฎีการตอบสนองข้อสอบระหว่างวิธีการตรวจให้คะแนนแบบประเพณีนิยมกับวิธีการตรวจให้คะแนนความรู้บางส่วนตามวิธีประยุกต์การให้คะแนนของคูมบ์ และวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท

3. เพื่อเปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบสอบ และอัตราส่วนสารสนเทศเฉลี่ยโดยใช้ทฤษฎีการตอบสนองข้อสอบระหว่างวิธีการตรวจให้คะแนนแบบประเพณีนิยมกับวิธีการตรวจให้คะแนนความรู้บางส่วนตามวิธีประยุกต์การให้คะแนนของคูมบ์ และวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท

สมมติฐานการวิจัย

จากการศึกษาวิธีการให้คะแนนความรู้บางส่วน ของ Simon , Budescu และ Nevo (1997) สรุปเป็นประเภทใหญ่ๆได้ 4 วิธี ดังนี้ 1. การให้น้ำหนักแก่ข้อสอบที่แตกต่างกัน (Differential Item Weighting) 2. การให้น้ำหนักแก่ตัวเลือกที่แตกต่างกัน (Differential Option Weighting) 3.การเปลี่ยนแปลงโครงสร้างของข้อสอบ (Changing the Item Structure) 4.การเปลี่ยนแปลงวิธีการตอบสนอง (Changing the Response Method) แต่งานวิจัยในครั้งนี้ เลือกวิธีการตัดตัวลวง และวิธีเลือกชุดย่อยของคำตอบถูก เป็นวิธีการให้คะแนนความรู้บางส่วนที่ศึกษา และจากการทบทวนงานวิจัยที่ผ่านมาที่เปรียบเทียบคุณภาพของการให้คะแนน ส่วนใหญ่ใช้ทฤษฎีการวัดแบบดั้งเดิม จากการศึกษาวิธีการให้คะแนนแบบการตัดตัวลวง (ET) วิธีการแก้การเดา (CG) และวิธีแบบเลือกตอบธรรมดา พบว่าความตรงและความเที่ยงของวิธีการให้คะแนนแบบการตัดตัวลวงมีค่าสูงที่สุด (Collet, 1971; Coombs et al, 1956; Jaradat and Tollefson, 1988) ส่วน Jaradat และ Swagad (1986) เปรียบเทียบวิธีการให้คะแนน 3 วิธี คือ วิธีการแก้การเดา (CG) วิธีเลือกชุดย่อยของคำตอบถูก (SST)และวิธีประเพณีนิยม (number correct) โดยใช้แบบทดสอบคู่ขนาน 3 ฉบับ พบว่า วิธีเลือกชุดย่อยของคำตอบถูก มีความตรงสูงที่สุด และมี

ความเที่ยงสูงกว่าวิธีประเพณีนิยม นอกจากนี้ Jaradat และ Tollefson (1988) พบว่า วิธีการให้คะแนนแบบการตัดตัวลวงและวิธีเลือกชุดย่อยของคำตอบถูกมีความตรงและความเที่ยงสูงกว่า วิธีการแก้การเดา

จากแนวคิดและงานวิจัยที่เกี่ยวข้อง จึงตั้งสมมติฐานการวิจัยไว้ดังนี้ วิธีให้คะแนนความรู้บางส่วนตามวิธีประยุกต์การให้คะแนนของคูมบ์ และวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท น่าจะมีค่าความตรงตามสภาพ อำนาจจำแนก ฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบสอบ และอัตราส่วนสารสนเทศเฉลี่ยสูงกว่าวิธีการให้คะแนนแบบประเพณีนิยม

ขอบเขตของการวิจัย

1. ประชากรที่ใช้ในการวิจัยได้แก่ นิสิตคณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ชั้นปีที่ 3 ที่ลงทะเบียนเรียน วิชาการประเมินผลการเรียนการสอน ในภาคการศึกษาต้น ปีการศึกษา 2544

2. แบบสอบที่ใช้เป็นแบบวัดผลสัมฤทธิ์ทางการเรียน วิชาการประเมินผลการเรียนการสอน ภาคการศึกษาต้น ปีการศึกษา 2544 ที่มีวิธีการให้คะแนน 3 วิธี จากแบบสอบชุดเดียวกัน คือวิธีการให้คะแนนแบบประเพณีนิยม วิธีประยุกต์การให้คะแนนของคูมบ์และวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท

3. การวัดคุณภาพของแบบสอบ วัดจาก ความตรงตามสภาพ ฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบสอบ และอัตราส่วนสารสนเทศโดยเฉลี่ย โดยที่

ความตรงตามสภาพ วัดจากค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนของผู้สอบที่ได้จากการสอบวิชาการประเมินผลการเรียนการสอนที่จัดทำขึ้นกับคะแนนที่ได้จากแบบสอบกลางภาค โดยที่การตรวจให้คะแนนจากการสอบที่จัดทำขึ้นมีการให้คะแนน 3 วิธีที่แตกต่างกัน และมีการแปลงค่าสัมประสิทธิ์สหสัมพันธ์ให้เป็นคะแนนฟิชเชอร์ซี (Fisher's Z) จากสูตร ของGlass and Stanley (1970) เพื่อนำมาเปรียบเทียบความแตกต่างของค่าความตรงตามสภาพเป็นรายคู่ โดยใช้สถิติทดสอบซี (Z)

ส่วนฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบสอบ และอัตราส่วนสารสนเทศโดยเฉลี่ย ใช้การวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) แบบทวิวิภาค ใช้โมเดลโลจิส (Logistic) และแบบพหุวิภาค ใช้โมเดลการให้คะแนนความรู้บางส่วน (Partial Credit Model ,PCM)

ข้อตกลงเบื้องต้นของการวิจัย

1. วิธีให้คะแนนความรู้บางส่วนที่ใช้ศึกษาครั้งนี้มี 2 วิธี คือ ใช้วิธีการให้คะแนนจากการตัดตัวลวง (Elimination Scoring ,ET) ซึ่งเป็นวิธีที่ประยุกต์มาจากวิธีการตอบและการให้คะแนนแบบคูมบ์และวิธีการเลือกชุดย่อยของคำตอบ (The Subset Selection Method, SST) ซึ่งเป็นวิธีที่ประยุกต์มาจาก เดรสเชลและสมิท (Dressel and schmidt ,1953) เนื่องจากทั้ง วิธีการให้คะแนนจากการตัดตัวลวง และวิธี

การเลือกชุดย่อยของคำตอบ มีแนวโน้มในการเพิ่มคุณภาพของการจำแนกระดับความรู้ของผู้สอบ และลดการเดา ซึ่งเป็นวิธีที่ให้ค่าความเที่ยงสูงสุดจากการศึกษาที่ผ่านมา

2. วิธีประยุกต์การให้คะแนนของคูมบ์ เป็นวิธีที่ผู้วิจัยพัฒนาขึ้นเนื่องจากวิธีการให้คะแนนของคูมบ์มีค่าติดลบแต่วิธีที่พัฒนาไม่มีค่าติดลบและมีค่าสูงสุด เท่ากับ 3 ต่ำสุดเท่ากับ0 ในแต่ละข้อ ส่วนวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท เป็นวิธีที่ผู้วิจัยพัฒนาขึ้นเนื่องจากวิธีการให้คะแนนของเดรสเซลและสมิทมีค่าติดลบ แต่วิธีที่พัฒนาไม่มีค่าติดลบและมีค่าสูงสุด เท่ากับ 3 ต่ำสุดเท่ากับ 0 ในแต่ละข้อ

3. การวิเคราะห์ฟังก์ชันสารสนเทศของข้อสอบ (Item information function) ฟังก์ชันสารสนเทศของแบบสอบ (Test information function) และอัตราส่วนสารสนเทศโดยเฉลี่ย ในการศึกษาครั้งนี้ใช้การวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) แบบทวิวิภาค ใช้โมเดลโลจิส (Logistic) และแบบพหุวิภาค ใช้โมเดลการให้คะแนนความรู้บางส่วน (Partial Credit Model, PCM) เนื่องจากโมเดลดังกล่าวไม่เข้มงวดกับข้อตกลงเบื้องต้นของความเป็นเอกมิติของแบบวัด และการประมาณค่าพารามิเตอร์มีค่าอำนาจจำแนกรายข้อในฟังก์ชันด้วย

คำจำกัดความที่ใช้ในการวิจัย

วิธีการให้คะแนนแบบประเพณีนิยม หมายถึง การให้คะแนนกับคำตอบที่ถูกต้องเป็น 1 คำตอบอื่นๆที่ผิดเป็น 0

วิธีประยุกต์การให้คะแนนของคูมบ์ หมายถึง วิธีการให้คะแนนความรู้บางส่วนที่ผู้วิจัยพัฒนาขึ้นจากคูมบ์ วิธีที่พัฒนาไม่มีค่าติดลบและมีค่าสูงสุด เท่ากับ 3 ต่ำสุดเท่ากับ 0 ในแต่ละข้อ โดยใช้สูตรการให้คะแนน $X = n\,C$

เมื่อ X คือ คะแนนที่ผู้สอบจะได้จากการตอบข้อสอบข้อหนึ่งๆ

n คือ จำนวนตัวลวงที่ตัดออก

C คือ ค่าความถูกต้องของการตัดตัวลวง มี 2 ค่า ได้แก่

$C = 1$ เมื่อไม่มีตัวเลือกถูกอยู่ในตัวลวงที่ตัดออก

$C = 0$ เมื่อมีตัวเลือกถูกอยู่ในตัวลวงที่ตัดออก

และผู้ตอบจะได้คะแนน 0 เมื่อไม่ตอบ

วิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท หมายถึง วิธีการให้คะแนนความรู้บางส่วนที่ผู้วิจัยพัฒนาขึ้นจากวิธีของเดรสเซลและสมิท วิธีที่พัฒนาไม่มีค่าติดลบและมีค่าสูงสุด เท่ากับ 3 ต่ำสุดเท่ากับ 0 ในแต่ละข้อ

โดยใช้สูตรการให้คะแนน $X = K - n\,C$

เมื่อ X คือ คะแนนที่ผู้สอบจะได้จากการตอบข้อสอบข้อหนึ่งๆ

K คือ จำนวนตัวเลือกทั้งหมดในข้อสอบ

n คือ จำนวนตัวเลือกที่ผู้ตอบเลือก

C คือ ค่าความถูกต้องของชุดย่อยของคำตอบ มี 2 ค่า ได้แก่

C = 1 เมื่อมีตัวเลือกถูกอยู่ในชุดย่อยของคำตอบ

C = 0 เมื่อไม่มีตัวเลือกถูกอยู่ในชุดย่อยของคำตอบ

และผู้ตอบจะได้คะแนน 0 เมื่อไม่ตอบ

ความตรงตามสภาพ หมายถึง ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนของผู้สอบที่ได้จากการสอบที่จัดขึ้นเพื่อใช้ในการวิจัยกับคะแนนที่ได้จากแบบสอบกลางภาค ปีการศึกษา 2544 วิชาการประเมินผลการเรียนการสอน โดยการให้คะแนนจากการสอบที่จัดขึ้น มี 3 วิธี ได้แก่ วิธีประยุกต์การให้คะแนนของคูมบ์ วิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท และวิธีประเพณีนิยม

ฟังก์ชันสารสนเทศของข้อสอบ หมายถึง ดัชนีที่บ่งชี้คุณภาพของข้อสอบที่เกิดจากการผสมรวมของดัชนีคุณลักษณะของข้อสอบหลายลักษณะ ได้แก่ ค่าความยาก ค่าอำนาจจำแนก และค่าความแปรปรวนของคะแนนรายข้อ

ฟังก์ชันสารสนเทศของแบบสอบ หมายถึง ผลรวมของค่าสารสนเทศของข้อสอบทุกข้อในแบบสอบที่มีการตรวจให้คะแนนอย่างเดียวกันทั้งฉบับ โดยแสดงค่าฟังก์ชันสารสนเทศตามระดับความสามารถของผู้ตอบที่แตกต่างกัน ถ้าสูงที่ระดับใดแสดงว่ามีความแม่นยำสูงในการประมาณค่าพารามิเตอร์ของข้อกระทง ณ ระดับความสามารถนั้นๆ

ค่าพารามิเตอร์ของข้อสอบ หมายถึง ค่าความยาก(b) และ ค่าอำนาจจำแนก(a) ของข้อสอบ

การวิเคราะห์ตามโมเดลโลจิส (Logistic) หมายถึง การวิเคราะห์แบบสอบวิชาการประเมินผลการศึกษาจากการตรวจให้คะแนน แบบทวิวิภาค (0,1) ซึ่งเป็นการให้คะแนนแบบประเพณี โดยประมาณค่าพารามิเตอร์ความยาก(b) อำนาจจำแนก (a) และค่าฟังก์ชันสารสนเทศของแบบสอบ(TIF)โดยใช้โปรแกรม BILOG

การวิเคราะห์ตามโมเดลการให้คะแนนความรู้บางส่วน(Partial Credit Model ,PCM) หมายถึง การวิเคราะห์แบบสอบวิชาการประเมินผลการศึกษาจากการตรวจให้คะแนนแบบพหุวิภาค วิธีประยุกต์การให้คะแนนของคูมบ์และวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิทโดยประมาณค่าพารามิเตอร์ความยาก(b) อำนาจจำแนก (a) และค่าฟังก์ชันสารสนเทศของแบบสอบ (TIF) โดยใช้โปรแกรม PARSCALE

กลุ่มที่มีความสามารถแตกต่างกัน หมายถึงกลุ่มนิสิตคณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ที่ลงทะเบียนเรียน วิชาการประเมินผลการศึกษา ในภาคการศึกษาต้น ปีการศึกษา 2544 ที่มีความสามารถแตกต่างกัน จำแนกเป็น 3 ระดับ ได้แก่ กลุ่มที่มีความสามารถสูง กลุ่มที่มีความสามารถปานกลาง และกลุ่มที่มีความสามารถต่ำ โดยที่

กลุ่มที่มีความสามารถสูง หมายถึง กลุ่มที่ได้คะแนนจากวิธีการให้คะแนนแบบประเพณีนิยม

(0,1) จากข้อสอบกลางภาคที่มีโครงสร้างของเนื้อหาเช่นเดียวกับแบบสอบที่ใช้ในการวิจัย โดยมีค่าความสามารถ 0.75 ขึ้นไป

กลุ่มที่มีความสามารถปานกลาง หมายถึง กลุ่มที่ได้คะแนนจากวิธีการให้คะแนนแบบประเพณีนิยม (0,1) จากข้อสอบกลางภาคที่มีโครงสร้างของเนื้อหาเช่นเดียวกับแบบสอบที่ใช้ในการวิจัย โดยมีค่าความสามารถ อยู่ระหว่าง -0.75- 0.75

กลุ่มที่มีความสามารถต่ำ หมายถึง กลุ่มที่ได้คะแนนจากวิธีการให้คะแนนแบบประเพณีนิยม (0,1) จากข้อสอบกลางภาคที่มีโครงสร้างของเนื้อหาเช่นเดียวกับแบบสอบที่ใช้ในการวิจัย โดยมีค่าความสามารถ ต่ำกว่า - 0.75

ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย

1. ทำให้ได้ข้อความรู้เกี่ยวกับคุณภาพของแบบสอบ ด้านความตรงตามสภาพ อำนาจจำแนก ฟังก์ชันสารสนเทศของแบบสอบ และฟังก์ชันสารสนเทศของข้อสอบ ระหว่างวิธีการให้คะแนนแบบประเพณีนิยม และวิธีให้คะแนนความรู้บางส่วน

2. ทำให้ได้แนวทางที่จะช่วยให้ผู้ที่จะใช้เครื่องมือในการวัดผลการศึกษาได้ตระหนักถึงวิธีการให้คะแนนที่มีคุณภาพในกระบวนการวัดเพื่อนำไปใช้ในการตัดสินคุณค่าในการประเมินต่อไป

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

ในบทนี้ผู้วิจัยได้ศึกษาแนวคิดและทฤษฎี ตลอดจนงานวิจัยที่เกี่ยวข้อง ซึ่งผู้วิจัยจะนำเสนอเป็น 4 ตอน ดังนี้

ตอนที่ 1 ความรู้ทั่วไปเกี่ยวกับการวัดความรู้บางส่วน (partial knowledge)

ตอนที่ 2 มโนทัศน์เกี่ยวกับทฤษฎีการตอบสนองข้อสอบ

ตอนที่ 3 งานวิจัยที่เกี่ยวข้อง

ตอนที่ 4 โปรแกรมคอมพิวเตอร์ที่ใช้ในการวิเคราะห์

ตอนที่ 1 ความรู้ทั่วไปเกี่ยวกับการวัดความรู้บางส่วน (partial knowledge)

การให้คะแนนความรู้บางส่วนของผู้ตอบ เป็นการแสดงถึงตำแหน่งจริงของความรู้ของผู้ตอบ ซึ่งเป็นการแก้ปัญหาการเดาที่เป็นจุดอ่อนที่สำคัญของข้อสอบแบบเลือก ปัญหาการตอบแบบเดาสุ่มเป็นแหล่งความ คลาดเคลื่อนในการวัด ซึ่งเป็นประเด็นที่มีการวิพากษ์กันอย่างมาก (Burton and Miller, 1999) นอกจากนี้ผลการวัดที่ได้จากการใช้แบบสอบแบบเลือกตอบ ไม่สามารถให้สารสนเทศที่เพียงพอที่จะจำแนกผู้ตอบในระดับต่างๆ ของความรู้ระหว่างผู้ที่รู้จริง (full knowledge) ผู้ที่มีความรู้บางส่วน (partial knowledge) และผู้ที่ไม่รู้ (absence of knowledge)

นักวิชาการจึงได้ศึกษาหาวิธีการแก้ปัญหา เพื่อลดโอกาสในการเดา และเพิ่มสารสนเทศเกี่ยวกับการตอบข้อสอบให้มากที่สุด โดยใช้วิธีการให้คะแนน ซึ่งเป็นองค์ประกอบที่สำคัญของกระบวนการวัด สำหรับวิธีการให้คะแนนที่เหมาะสมจะทำให้คุณสมบัติของการวัดในประเด็นที่เกี่ยวข้องกับความตรงและความเที่ยงสูงขึ้น การให้คะแนนความรู้บางส่วนจึงมีส่วนช่วยลดความคลาดเคลื่อนจากแหล่งการเดาสุ่มของผู้ตอบลง ทำให้คะแนนมีความเที่ยงมากขึ้น ในขณะเดียวกันก็ส่งผลต่อความตรงของกระบวนการวัดด้วย (Frary ,1980 อ้างถึงในพรทิพย์ ไชยโส , 2533) จากการศึกษาวิธีการให้คะแนนความรู้บางส่วนสรุปเป็นประเภทใหญ่ๆ ได้ 4 วิธี ดังนี้ (Simon , Budescu และ Nevo 1997)

1. การให้น้ำหนักแก่ข้อสอบที่แตกต่างกัน (Differential Item Weighting)
2. การให้น้ำหนักแก่ตัวเลือกที่แตกต่างกัน (Differential Option Weighting)
3. การเปลี่ยนแปลงโครงสร้างของข้อสอบ (Changing the Item Structure)
4. การเปลี่ยนแปลงวิธีการตอบสนอง (Changing the Response Method)

โดยแต่ละวิธีมีรายละเอียดดังนี้

1. การให้น้ำหนักแก่ข้อสอบที่แตกต่างกัน (Differential Item Weighting)

หลักการของวิธีนี้ คือข้อสอบที่มีคุณภาพจะได้น้ำหนักมากกว่าข้อสอบที่ไม่มีคุณภาพโดยใช้ผลที่ได้จากการวิเคราะห์ข้อสอบเป็นเกณฑ์ คือค่าความยาก อำนาจจำแนก ความสามารถ ความตรง ความแปรปรวน หรือการพิจารณาของผู้เชี่ยวชาญ Stanley และ Wang (1970)ได้ ให้ข้อสรุปว่า การให้น้ำหนักแก่ข้อสอบโดยใช้ทฤษฎีการวัดแบบดั้งเดิม ทำให้น้ำหนักของคะแนนแต่ละข้อมีค่าไม่คงที่เปลี่ยนแปลงไปตามกลุ่มตัวอย่าง จึงควรให้คะแนนตามทฤษฎีการตอบสนองข้อสอบ (IRT) อย่างไรก็ตามการกำหนดน้ำหนักคะแนนที่เหมาะสมรายข้อนี้ ยังวางอยู่บนพื้นฐานของวิธีการตอบ และการให้คะแนนเฉพาะข้อที่ตอบถูกแบบเดิม ซึ่งไม่ได้มีการพิจารณาการให้คะแนนความรู้บางส่วนแก่ผู้ตอบ

2. การให้น้ำหนักคะแนนแก่ตัวเลือกที่แตกต่างกัน (Differential Option Weighting)

หลักการของวิธีนี้ คือ ความรู้บางส่วนสามารถถูกวัดได้จากการให้น้ำหนักคะแนนแก่ตัวเลือก แบ่งวิธีการได้ 2 วิธี คือวิธีแรกเป็นการให้น้ำหนักแก่ตัวเลือกโดยใช้วิจารณญาณของผู้เชี่ยวชาญ หรือจากทฤษฎีของโครงสร้างความรู้ (Smith, 1987) ส่วนวิธีที่สองเป็นการให้น้ำหนักจากข้อมูลเชิงประจักษ์ที่ผ่านมาในอดีต หรือปัจจุบันโดยอาศัยสิ่งที่น่าสนใจของตัวเลือก เช่นคะแนนมาตรฐานเฉลี่ยของแต่ละตัวเลือกที่ผู้สอบเลือก ความสัมพันธ์ระหว่างตัวเลือกที่ผู้สอบเลือกกับคะแนนรวมทั้งหมด เป็นต้น

วิธีการนี้ได้รับการศึกษาเพิ่มมากขึ้น จากผลการศึกษาของ Frary (1989), Haladyna (1988), Echternacht (1976), Hakstian และ Kansup (1975) และ Stanley และ Wang (1970) พบว่าการให้น้ำหนักคะแนนแก่ตัวเลือก มีประโยชน์บ้างในแง่ที่ว่าช่วยเพิ่มค่าความเที่ยงคงที่ภายใน (internal consistency reliability) การเปรียบเทียบค่าความตรงและความเที่ยงระหว่างแบบการให้น้ำหนักแก่ตัวเลือกกับแบบธรรมดา พบว่ามีค่าไม่คงที่ของความแตกต่างอย่างมีนัยสำคัญ วิธีการเหล่านี้ไม่เป็นที่นิยม เนื่องจากการพัฒนาเพื่อให้ได้มาซึ่งน้ำหนักใช้ทรัพยากรมาก การคำนวณการให้คะแนนที่มีวิธีการที่ยุ่งยาก และยากในการอธิบายและตัดสินการให้คะแนนแก่ผู้สอบ (Frary, 1989)

3. การเปลี่ยนแปลงโครงสร้างของข้อสอบ (Changing the Item Structure)

วิธีนี้มีรูปแบบของโครงสร้างข้อสอบ และ/หรือ คำอธิบายที่แตกต่างจาก แบบเลือกตอบธรรมดา ตัวอย่างที่สำคัญ มีดังนี้

- ข้อสอบที่มี 2 ตัวเลือก คือ ถูก และ ผิด ผู้ตอบจะต้องเลือกคำตอบอย่างใดอย่างหนึ่ง คะแนนที่ได้ได้จากผลรวมของจำนวนคำตอบที่ถูกต้อง เนื่องจากการตอบแต่ละข้อใช้เวลาน้อย จึงต้องเพิ่มจำนวนข้อให้มากขึ้นกว่าข้อสอบแบบเลือกตอบธรรมดา

- ข้อสอบที่มีตัวถูกหลาย ๆ ตัวเลือก (มากกว่า 1) โดยต้องมีคำชี้แจงอธิบายให้แก่ผู้สอบ ซึ่งอาจจะระบุจำนวนตัวเลือกที่ถูกหรือไม่ระบุก็ได้ คะแนนที่ได้เท่ากับผลรวมของการระบุตัวเลือกถูกได้ถูกต้อง ในบางกรณีมีการหักคะแนนการตอบที่ตอบตัวเลือกผิด เช่นวิธีของ Dressel และ Schmidt (1953) โดยมีจุดประสงค์เพื่อลดการเดา และประเมินความรู้บางส่วน แบบสอบลักษณะนี้ ได้มีการ

พัฒนาต่อมา ตัวอย่างเช่นแบบสอบที่มีตัวเลือกถูกทุกข้อ ผิดทุกข้อ หรือถูกเฉพาะตัวเลือกหนึ่งกับอีกตัวเลือกหนึ่ง (Frary ,1989 ; Hsu, Moss and Khampalikit , 1984)

- ข้อสอบที่มีกลุ่มของข้อสอบ และกลุ่มของคำตอบ ที่เรียกว่าข้อสอบแบบจับคู่ โดยผู้สอบต้องเลือกคำตอบที่ถูกคู่กับคำถามที่กำหนดให้ ถ้าจำนวนคำถามเท่ากับจำนวนคำตอบ เรียกว่า การจับคู่อย่างง่าย (simple matching) (Horst, 1966; Wesman,1971) แต่ถ้าจำนวนคำตอบมากกว่าจำนวนคำถาม เรียกว่า การจับคู่แบบพหุ (multiple matching) (Gullisen, 1986) ความแตกต่างของ 2 แบบ คือ การจับคู่แบบพหุ ช่วยลดโอกาสในการเดามากกว่า เนื่องจากมีจำนวนตัวเลือกมากกว่า แต่ข้อจำกัดที่สำคัญของวิธีนี้ คือ ความยากในการสร้างข้อสอบให้มีความเหมาะสม Budescu (1988) พบว่า การจับคู่แบบพหุให้ค่าความตรง และความเที่ยงมากกว่าแบบสอบเลือกตอบธรรมดา และความตรงความเที่ยงจะเพิ่มขึ้นเมื่อสัดส่วนของจำนวนตัวเลือกเพิ่มขึ้น วิธีนี้ต้องการเวลาในการทำข้อสอบมากกว่าปกติ

4. การเปลี่ยนแปลงวิธีการตอบสนอง (Changing the Response Method)

วิธีนี้แตกต่างจากวิธีก่อนที่กล่าวมาทั้งหมด คือ ใช้น้ำหนักที่ให้โดยผู้สอบเอง ซึ่งสะท้อนถึงความรู้ที่มีอยู่ในตัวผู้สอบ การประเมินด้วยตนเองนี้เป็นวิธีที่ดีที่สุดในการได้มาซึ่งสารสนเทศเกี่ยวกับความรู้ของผู้สอบในแต่ละข้อ ผู้สอบจะให้สารสนเทศอย่างสมบูรณ์เกี่ยวกับน้ำหนัก วิธีการให้คะแนนและการควบคุมระดับความมั่นใจในการให้ ซึ่งจะช่วยลดการเดา และความคลาดเคลื่อนในการวัด วิธีนี้ใช้รูปแบบข้อสอบเหมือนแบบสอบชนิดเลือกตอบทั่วไป เพียงแต่คำสั่ง วิธีการตอบ และการให้คะแนนแตกต่างจากแบบสอบชนิดเลือกตอบธรรมดา ซึ่งแต่ละวิธีใช้วิธีการที่ไม่ยุ่งยาก โดย แบ่งเป็นชนิดต่าง ๆ ได้ 6 วิธี คือ

4.1 วิธีการแก้การเดา (Correction for guessing)

วิธีการนี้ให้ผู้สอบทำข้อสอบเหมือนกับแบบสอบชนิดเลือกตอบธรรมดา แต่ต่างกันที่การให้คะแนน วิธีการแก้การเดา มีวิธีที่น่าสนใจ 2 วิธี คือ

1) วิธีการแก้การเดาโดยการลงโทษ เป็นวิธีที่มีข้อตกลงเบื้องต้นว่าข้อที่ผู้สอบตอบผิดเป็นผลมาจากการเดา ดังนั้นข้อที่ผู้สอบตอบผิดแทนที่จะได้ 0 แต่จะให้คะแนนติดลบ -1/(K-1) K คือ จำนวนตัวเลือกของแบบสอบ (Lord ,1975) โดยใช้สูตรการให้คะแนนเพื่อแก้การเดาคำตอบ สูตรการให้คะแนนที่ใช้อยู่คือ

$X_c = R - W/(K-1)$

X_c คือ คะแนนที่ได้หลังจากการแก้การเดา

R คือ จำนวนข้อสอบที่ผู้สอบตอบถูก

W คือ จำนวนข้อสอบที่ผู้สอบตอบผิด

วิธีการแก้การเดาสุ่มเหมาะกับแบบสอบความเร็วที่ผู้ตอบตอบไม่ทัน แต่ถ้าเป็นแบบสอบที่ไม่จำกัดเวลา ผู้ตอบอาจไม่ได้ตอบอย่างเดาสุ่มเพียงอย่างเดียว ผู้ตอบอาจมีความรู้บางส่วนในการตัดตัวลวง

บางตัวออกได้ ดังนั้นการใช้สูตรแก้การเดาอาจไม่เหมาะสม เพราะควรที่จะพิจารณาถึงการตอบโดยมีความรู้บางส่วนของผู้ตอบด้วย

2) วิธีการให้รางวัลแก่ข้อที่ไม่ตอบ เป็นวิธีที่มีข้อตกลงเบื้องต้นว่า ข้อที่ผู้ตอบเว้นไว้นั้น ถ้าเขาตอบจะมีโอกาสตอบถูกเป็น 1/K เมื่อ K คือจำนวนตัวเลือกของแบบสอบ ซึ่งเขียนเป็นสูตรได้ดังนี้

$X_c = R+O/K$

X_c คือ คะแนนหลังจากการแก้ไข

R คือ จำนวนข้อสอบที่ผู้สอบตอบถูก

O คือ จำนวนข้อสอบที่ผู้สอบไม่ตอบ

วิธีนี้มีการศึกษาเพิ่มเติมของหลายท่าน และยังมีการศึกษาเปรียบเทียบความเหมาะสมของวิธีการแก้การเดา และแบบสอบเลือกตอบธรรมดา ผลที่ได้ยังได้ข้อสรุปไม่ชัดเจน (Angoff 1989, Diamond and Evans, 1973)

4.2 วิธีการให้คะแนนจากการตัดตัวลวง (Elimination Scoring)

วิธีนี้เป็นการให้ผู้สอบตัดตัวลวงออกไปได้บ้างอย่างถูกต้อง แม้จะไม่ทราบว่าคำตอบที่ถูกต้องคืออะไร มีวิธีการและงานวิจัยที่สนับสนุนวิธีนี้ ดังนี้

1) วิธีการตอบและให้คะแนนแบบคูมบ์ (the Coombs Response/Scoring method,) Coombs ได้จำแนกระดับความรู้ของผู้สอบออกเป็น 5 ประเภท คือ (Coombs, Milholland, and Womer 1956)

1. มีความรู้เต็มที่ (Full knowledge) คือ ผู้ที่สามารถตัดตัวลวงทั้งหมดออกได้

2. มีความรู้บางส่วน (Partial knowledge) คือ ผู้ที่สามารถตัดตัวลวงบางตัวออกได้แต่ไม่ทั้งหมด

3. มีความรู้ที่ผิดบางส่วน (Partial misinformation) คือ ผู้ที่สามารถตัดตัวลวงบางตัวออกได้บ้างและตัดตัวถูกออกด้วย

4. มีความรู้ที่ผิดเต็มที่ (Full misinformation) คือ ผู้ที่ตัดตัวถูกออกเพียงตัวเดียว

5. ไม่มีความรู้ (Absence of knowledge) คือ ผู้ที่ไม่ตอบ หรือตัดตัวเลือกทุกตัวออกหมด

วิธีการตอบและการให้คะแนนแบบคูมบ์ ใช้รูปแบบของแบบสอบเลือกตอบที่มีคำตอบถูกเพียงคำตอบเดียว โดยให้ผู้สอบตัดตัวลวงที่ไม่ใช่คำตอบถูกที่เขาแน่ใจออกไปมากที่สุด การให้คะแนนผู้ตอบจะได้รับ 1 คะแนน ในแต่ละตัวลวงที่ตัดออกอย่างถูกต้อง และในกรณีที่ผู้สอบมีความรู้ที่ผิดไปตัดตัวถูกออกโดยเข้าใจว่าเป็นตัวลวงจะถูกตัดคะแนน K-1 โดยที่ K คือ จำนวนตัวเลือก ดังนั้นช่วงคะแนนในแต่ละข้ออยู่ระหว่าง -(K-1) ถึง (K-1)

จากงานวิจัยที่ศึกษาเปรียบเทียบวิธีให้คะแนนแบบการตัดตัวลวง (ET) วิธีการแก้การเดา (CG) และวิธีแบบเลือกตอบธรรมดา พบว่า ความตรงและความเที่ยงของวิธีการให้คะแนนแบบการตัดตัวลวงสูงที่สุด (Collet, 1971; Coombs et al, 1956; Jaradat and Tollefson, 1988)

2) วิธีการตอบและการให้คะแนนแบบอาร์โนลด์ (The Arnold Scoring / Response Method) (Arnold and Arnold 1970) อาร์โนลด์ เสนอวิธีที่พัฒนาจากทฤษฎีเกมเบื้องต้น (Game theory) มีการใช้ข้อสอบแบบเลือกตอบเช่นเดียวกับแบบของคูมบ์ โดยให้ผู้ตอบตัดตัวเลือกที่แน่ใจว่าเป็นตัวลวงหรือคำตอบผิดออกไปมากที่สุดที่เขารู้ แต่วิธีนี้มีระบบการให้คะแนนที่แตกต่างจากของคูมบ์ คือจะให้คะแนน 1/(K-1) คะแนนสำหรับการตัดตัวลวงถูกต้อง 1 ตัว แต่ถ้าผู้สอบตัดตัวเลือกถูกออกโดยเข้าใจว่าเป็นตัวลวงจะถูกหัก -1/(K-1) คะแนน เมื่อ K คือจำนวนตัวเลือกแต่ละข้อ ระบบการให้คะแนนของอาร์โนลด์วางอยู่พื้นฐานคะแนนที่ถูกคาดหวัง (expected item score) เท่ากับ 0 เมื่อผู้สอบเดาสุ่ม ทำให้สามารถพัฒนาสูตรการให้คะแนนที่ยุติธรรมได้ดังนี้

$Cd = (p)\,(d / (K-d))$

โดยที่ Cd คือ คะแนนเมื่อสามารถตัดตัวลวงบางตัวออกไปได้ และคะแนนที่เพิ่มขึ้นจากการเดาเป็น 0

d คือ จำนวนตัวลวงที่ตัดออกได้ถูกต้อง

K คือ จำนวนตัวเลือก

p คือ คะแนนลงโทษในกรณีที่ผู้สอบจัดตัวเลือกที่ถูกเข้าไว้เป็นตัวลวงในกรณีของ 4 ตัวเลือก คะแนนที่ได้เป็นผลจากการตอบของผู้สอบ ดังนี้

คะแนน C_0 คือ ไม่สามารถตัดตัวลวงใดๆ ได้เลย จะได้คะแนนจากการแทนค่าในสูตรเท่ากับ 0

คะแนน C_1 คือ ตัดตัวลวงออกได้ 1ตัว จะได้คะแนนจากการแทนค่าใน สูตรเท่ากับ 1/9

คะแนน C_2 คือ ตัดตัวลวงออกได้ 2ตัว จะได้คะแนนจากการแทนค่าใน สูตรเท่ากับ 1/3

คะแนน C_3 คือ ตัดตัวลวงออกได้ 3 ตัว จะได้คะแนนจากการแทนค่าใน สูตรเท่ากับ 1

คะแนน -P คือ ตัดตัวถูกออกเพราะเข้าใจว่าเป็นตัวลวง จะถูกหักออก -1/3 คะแนน

3) วิธีการตอบและให้คะแนนแบบครอส (The Cross Response/ Scoring Method, CRS)

วิธีการนี้รวมวิธีการของ ET และ AUC โดยให้ผู้สอบตัดตัวลวงออกเช่นเดียวกับวิธีการของคูมบ์ และอาร์โนลด์ แต่การให้คะแนนแตกต่างกัน คือวิธีการให้คะแนนของครอสจะให้คะแนนรายข้อเฉลี่ยต่ำสุด เมื่อผู้ตอบตัดตัวเลือกที่เป็นคำตอบโดยเข้าใจว่าเป็นตัวลวงและตัดตัวลวงอีก K-2 ตัว โดยเข้าใจว่าเป็นชุดของตัวเลือกที่ผิด ซึ่งผิดกับการให้คะแนนแบบคูมบ์ที่คะแนนรายข้อเฉลี่ยต่ำสุดจะเกิดจากการตัดตัวเลือกที่ถูกว่าเป็นตัวลวงเพียงคำตอบเดียว การให้คะแนนแบบครอสผู้สอบจะได้ 2 คะแนนในแต่ละตัวลวงที่เขาตัดออกได้อย่างถูกต้อง และจะได้ 1 คะแนนสำหรับแต่ละตัวเลือกที่ไม่ได้ตัดว่าเป็นตัวลวง เมื่อคำตอบถูกยังไม่ได้ถูกตัด แต่เมื่อใดที่ผู้ตอบตัดตัวเลือกที่ถูกด้วย misinformation เขาจะได้คะแนนเฉพาะตัวเลือกที่ไม่ได้ตัดตัวเลือกละ 1 คะแนน สำหรับในกรณีที่ข้อสอบมี 4 ตัวเลือก คะแนนที่ยอมให้เกิดขึ้นได้มีค่าตั้งแต่ 1 คะแนน ถึง 7 คะแนน และคะแนนต่ำสุดจะเกิดจากการที่ผู้ตอบ

ตัดตัวเลือกที่ถูกต้องหลังจากการตัดตัวลวงมาแล้ว 2 ตัว Frary (1980) เปรียบเทียบวิธีการให้คะแนนแบบครอสกับวิธีการให้คะแนนอื่นๆ พบว่าคะแนนแบบครอสให้ค่าความเที่ยงต่ำกว่าวิธีการให้คะแนนของอาร์โนลด์และคูมบ์ ส่วนความตรงตามเกณฑ์ไม่แตกต่างกัน

4) วิธีการเลือกชุดย่อยของคำตอบ (The Subset Selection Method, SST) เสนอโดย Dressel และ Schmidt (1953) เป็นวิธีที่ตรงข้ามกับ ET คือ วิธีนี้ให้ผู้สอบเลือกชุดย่อยของคำตอบที่เขาเห็นว่าน่าจะมีตัวเลือกที่เป็นคำตอบถูกรวมอยู่ด้วย การให้คะแนนขึ้นอยู่กับจำนวนตัวเลือกที่มีอยู่ในชุดย่อยของคำตอบ ยิ่งน้อยเท่าใดคะแนนจะได้เพิ่มขึ้น ในกรณีที่ตัวเลือกที่ถูกต้องรวมอยู่ในชุดย่อยของคำตอบนั้นด้วย แต่ถ้าชุดย่อยนั้นไม่มีคำตอบที่ถูกต้องอยู่ด้วย คะแนนจะถูกหักออกไปเท่ากับขนาดของจำนวนตัวเลือกที่ผู้ตอบเลือก ในกรณีที่ผู้ตอบไม่เลือกตัวเลือกใดเลยจะได้ 0 คะแนน วิธีการนี้ Gibbon, Olkin และ Sobel (1979) ได้ศึกษาพบว่า เป็นวิธีที่ช่วยลดการเดาสุ่มของผู้ตอบลง และยังยอมให้ผู้ตอบได้แสดงความรู้บางส่วนของผู้ตอบด้วย

Jaradat และ Swagad (1986) เปรียบเทียบวิธีการให้คะแนน 3 วิธี คือ CG, SST และ NC (number correct) โดยใช้แบบทดสอบคู่ขนาน 3 ฉบับ พบว่า SST มีความตรงสูงที่สุด และมีความเที่ยงสูงกว่า NC นอกจากนี้ Jaradat และ Tollefson (1988) พบว่า ET และ SST มีความตรงและความเที่ยงสูงกว่า CG

ทั้ง ET และ SST มีแนวโน้มในการเพิ่มคุณภาพของการจำแนกระดับความรู้ของผู้สอบ และลดการเดา แต่อย่างไรก็ตามวิธีการยังมีความยุ่งยาก และต้องใช้เวลาทำนานกว่าแบบสอบเลือกตอบธรรมดา ซึ่งยังเป็นข้อจำกัดของการให้คะแนน และต้องพัฒนาวิธีการต่อไป

4. 3 วิธีการของการทดสอบความน่าจะเป็น (Probability Testing)

วิธีนี้มีความยืดหยุ่น และเปิดโอกาสให้ผู้สอบแสดงความรู้บางส่วนมากที่สุด โดยรายงานความน่าจะเป็นในตัวเลือกที่ผู้สอบคิดว่าเป็นคำตอบที่ถูก โดยมีตำแหน่งคะแนนที่เป็นไปได้ 101 ตำแหน่ง (ช่วงจาก 0-100) โดยแบ่งระดับความรู้เป็น 5 ประเภท ดังนี้ (Wallsten, Budescu, Zwick, 1993)

1) มีความรู้เต็มที่ (Full knowledge) คือ ให้ความน่าจะเป็น = 1 แก่คำตอบที่ถูกต้อง

2) มีความรู้บางส่วน (Partial knowledge) คือ ให้ความน่าจะเป็นมากกว่า 0 แต่ไม่ถึง 1 ($0 < p < 1$) แก่คำตอบที่ถูกต้อง

3) มีความรู้ผิดบางส่วน (Partial misinformation) คือ ให้ความน่าจะเป็นมากกว่า 0 แต่ไม่ถึง 1 ($0 < p < 1$) แก่ตัวเลือกที่ผิด

4) มีความรู้ผิดเต็มที่ (Full misinformation) คือ ให้ความน่าจะเป็นเท่ากับ 0 แก่คำตอบที่ถูกต้อง

5) ไม่มีความรู้ (Absence of information) คือ ให้ความน่าจะเป็นแก่ทุกตัวเลือกเท่ากัน หรือไม่ตอบ

De Finetti (1965) ได้ศึกษาวิธีการที่ให้ผู้สอบให้คะแนนความน่าจะเป็นแก่ตัวเลือกที่คิดว่าถูกโดยคาดว่าจะให้คะแนนในคำตอบถูกสูงที่สุด เรียกการให้คะแนนความน่าจะเป็นนี้ว่า proper

reproducting scoring systems (PRSSs) ต่อมามีการศึกษารูปแบบของ PRSSs โดย Shuford, Albert และ Massengill (1966) 3 รูปแบบ คือ quadratic, spherical และ logarithmic ซึ่งมีตัวเลือก > 2 ตัว แต่วิธีการนี้มีความยุ่งยาก ซับซ้อนในการตอบ และการให้คะแนนมาก

Rippey (1970) และ Suhadolnik และ Weiss (1983) พบว่าการให้คะแนนความน่าจะเป็นแบบธรรมดามีความเที่ยงมากกว่าการให้คะแนน 3 รูปแบบดังกล่าว แต่ Kansub และ Hakstian (1975) พบว่าการให้คะแนนแบบ quadratic และ logarithmic มีค่าความเที่ยงมากกว่าการให้คะแนนความน่าจะเป็นธรรมดา

นอกจากนี้ Michael (1968) และ Pugh และ Brunza (1975) พบว่าวิธีการให้คะแนนความน่าจะเป็น (PT) มีความเที่ยงสูงกว่าวิธีเลือกตอบธรรมดา ส่วน Hambleton, Roberts และ Traub (1970) รายงานว่ามีความตรงสูงกว่าวิธีเลือกตอบธรรมดา ในขณะที่ Koehler (1971) และ Hakstian และ Kansup (1975) พบว่ามีความไม่แตกต่างกันของความเที่ยงและความตรงระหว่าง NC และ PT ซึ่งยังไม่ได้ข้อสรุปที่ชัดเจน

4.4 วิธีการแสดงความมั่นใจ (Confidence Marking)

วิธีนี้คิดค้นโดย Dressel และ Schmidt (1953) โดยที่ผู้สอบถูกถามให้แสดงความมั่นใจในคำตอบที่คิดว่าถูก โดยใช้ C-point scale (โดยทั่วไป $3 \leq c \leq 5$) หรืออาจเป็นการบอกความมั่นใจเป็นภาษา คือ ไม่แน่ใจ ค่อนข้างแน่ใจ แน่ใจมาก ซึ่งมีการแบ่งระดับความรู้ออกเป็น 5 ระดับ คือ

1. Full knowledge คือ การเลือกคำตอบได้ถูกต้อง ด้วยความมั่นใจสูงที่สุด
2. Partial knowledge คือ การเลือกคำตอบได้ถูกต้อง ด้วยความมั่นใจระดับต่ำ
3. Partial misinformation คือ การเลือกคำตอบผิด ด้วยความมั่นใจระดับต่ำ
4. Full misinformation คือ การเลือกคำตอบผิด ด้วยความมั่นใจระดับสูงสุด
5. Absence of knowledge คือ การไม่เลือกคำตอบเว้นว่างไว้

ผลการวิจัยพบว่าวิธีการนี้ทำให้ความตรงของแบบสอบสูงขึ้นกว่าการให้คะแนนแบบ0,1 (Dressel and Schmidt, 1953; Hopkins, Hakstian and Hopkins, 1973) แต่วิธีนี้ต้องการเวลาในการตอบและการให้คะแนนมากกว่าเดิม

4.5 วิธีการจัดอันดับอย่างสมบูรณ์ (Complete Ordering)

วิธีการนี้เป็นวิธีเฉพาะของ PT คือ แทนที่จะให้แสดงความน่าจะเป็นในแต่ละตัวเลือก ผู้สอบจะต้องจัดอันดับของตัวเลือก ซึ่งทำให้ง่ายขึ้นต่อการนำไปใช้ ทำให้สามารถจำแนกระดับความรู้เป็น 3 ระดับ คือ

1. Full knowledge คือ การจัดอันดับให้คำตอบถูกอยู่ในตำแหน่งที่สูงที่สุด
2. Partial knowledge คือ การจัดอันดับให้คำตอบถูกอยู่ในตำแหน่งกลาง
3. Absence of knowledge คือ การจัดอันดับให้คำตอบถูกอยู่ในตำแหน่งต่ำที่สุด

4.6 วิธีการจัดอันดับเป็นบางส่วน (Partial Ordering)

วิธีนี้เป็นวิธีที่ผสมระหว่าง ET และ CO คือให้ผู้สอบจัดอันดับตัวเลือกที่ไม่สามารถตัดออกไปได้ วิธีนี้เสนอโดย De Finetti (1965) และถูกนำไปใช้โดย Diamond (1975) วิธีการจัดอันดับเป็นบางส่วนนี้จำแนกระดับความรู้เป็น 5 ระดับ คือ

1. Full knowledge คือ การจัดอันดับแก่คำตอบถูกในตำแหน่งสูงที่สุด

2. Partial knowledge คือ การจัดอันดับแก่หลายตัวเลือก โดยคำตอบถูกถูกจัดอันดับในตำแหน่งสูงที่สุด

3. Partial misinformation คือ การจัดอันดับแก่คำตอบถูกในตำแหน่งที่ต่ำกว่าตัวลวง

4. Full misinformation คือ การจัดตำแหน่งแก่ตัวลวง โดยตัดคำตอบถูกออก

5. Absence of knowledge คือ การจัดอันดับแก่ตัวถูกในตำแหน่งที่ต่ำที่สุด หรือไม่ตอบ

จากการศึกษาทฤษฎีและงานวิจัยที่ผ่านมา พบว่าส่วนใหญ่จะเปรียบเทียบคุณภาพของแบบสอบในเรื่องความตรงและความเที่ยงจากวิธีการให้คะแนนความรู้บางส่วนหลายๆวิธี โดยใช้ทฤษฎีการวัดแบบดั้งเดิม ซึ่งมีข้อจำกัดในเรื่องแหล่งความคลาดเคลื่อนในการวัด และค่าสถิติที่ได้แปรเปลี่ยนไปตามกลุ่มตัวอย่างที่ใช้ นักวัดผลจึงได้พัฒนาทฤษฎีการตอบสนองข้อสอบ(IRT) เพื่อแก้ปัญหาการวิเคราะห์ตามแนวทฤษฎีการทดสอบแบบดั้งเดิม การพัฒนาในระยะแรกมุ่งนำไปวิเคราะห์กับแบบสอบที่ตรวจให้คะแนนแบบทวิวิภาค แต่ต่อมามีการประยุกต์ใช้กับแบบสอบที่ตรวจให้คะแนนแบบพหุวิภาคที่มีอยู่ด้วยกันหลายโมเดล เช่น Graded Response Model (GRM) ของซามิจิมา (Samejima) Nominal Response Model (NRM) พัฒนาโดยบอค (Bock) Rating ScaleModel (RCM) พัฒนาโดยมาสเตอร์ (Master) Generalized Partial Credit Model(GPCM) เป็นโมเดลที่ มูราคิ (Muraki)พัฒนาจากPCM แต่โมเดลที่ได้รับการพัฒนาและใช้อย่างแพร่หลายมากในปัจจุบัน คือ GRM และ GPCM (Donoghue,1994; De Ayala,1994; Muraki,1993; Reise and Yu,1990; Dodd , Koch and De Ayala,1989; Koch,1983) เนื่องจากโมเดลดังกล่าวไม่เข้มงวดกับข้อตกลงเบื้องต้นของความเป็นเอกมิติของแบบวัด และการประมาณค่าพารามิเตอร์มีค่าอำนาจจำแนกรายข้อในฟังก์ชันด้วย การวิเคราะห์ทั้ง 2 โมเดลจะได้สารสนเทศที่สำคัญ คือฟังก์ชันสารสนเทศของแบบสอบ (Test information function) ฟังก์ชันสารสนเทศของข้อสอบ (Item information function)และฟังก์ชันสารสนเทศของลำดับขั้นคะแนน (item –category information) จึงเป็นเรื่องที่น่าสนใจและทำการศึกษาต่อในการเปรียบเทียบคุณภาพของแบบสอบในเรื่องความตรงและความเที่ยงจากวิธีการให้คะแนนความรู้บางส่วนหลายๆวิธี โดยใช้ทฤษฎีการตอบสนองข้อสอบ (IRT) และจะนำเสนอมโนทัศน์เกี่ยวกับทฤษฎีการตอบสนองข้อสอบในตอนที่ 2 ต่อไป

ตอนที่ 2 มโนทัศน์เกี่ยวกับทฤษฎีการตอบสนองข้อสอบ

2.1 มโนทัศน์เบื้องต้นเกี่ยวกับทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory : IRT) เป็นทฤษฎีที่อธิบายความสัมพันธ์ระหว่างคุณลักษณะหรือความสามารถที่อยู่ภายในตัวบุคคลกับพฤติกรรมการตอบสนองข้อสอบแต่ละข้อว่ามีความน่าจะเป็นในการตอบถูกเพียงใด เป็นทฤษฎีที่ได้แก้ไขข้อจำกัดหลายๆประการของทฤษฎีการการวัดแบบดั้งเดิม (Classical Test Theory) โดยมีพัฒนาการของทฤษฎีโดยสังเขป ดังนี้ (Hambleton and Swaminathan ,1985)

พัฒนาการของทฤษฎีนี้เริ่มต้นที่ บิเนทและไซมอน (Binet and Simon) ซึ่งเป็นนักจิตวิทยา ได้สร้างกราฟที่แสดงความสัมพันธ์ระหว่างทักษะในการคิดของเด็กกับอายุ ในปี ค.ศ.1916 ซึ่งเป็นที่มาของโค้งคุณลักษณะของข้อสอบในระยะต่อมา ต่อมาในปี ค.ศ.1936 ริชาร์ดสัน (Richardson)ได้แสดงความสัมพันธ์ระหว่าง ทฤษฎีการตอบสนองข้อสอบ และทฤษฎีการวัดแบบดั้งเดิม และหาวิธีในการประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองข้อสอบ ในปี ค.ศ. 1943 ลอว์ลี (Lawley) ได้เสนอบทความเกี่ยวกับปัญหาการสร้างและการเลือกข้อสอบ (ศิริชัย กาญจนวาสี, 2535) โดยเขาได้เสนอแนวคิดในรูปของโมเดลนอร์มัลโอไจฟ์ (normal ogive model) ต่อมาในปี ค.ศ. 1950 ราส์ช (Rasch) นักคณิตศาสตร์ชาวเดนมาร์กซึ่งได้เสนอโมเดลราส์ช (rasch model) แบบ 1 พารามิเตอร์ โดยมีแนวคิดว่าความยากของข้อสอบเป็นสิ่งเดียวที่มีอิทธิพลต่อการตอบสนองข้อสอบ ในปี ค.ศ.1952 ลอร์ด(Lord) ได้ขยายแนวคิดทฤษฎีการตอบสนองข้อสอบ โดยเสนอฟังก์ชันนอร์มัลโอไจฟ์ แบบ 2 พารามิเตอร์ ที่มีการเพิ่มพารามิเตอร์อำนาจจำแนก

ในปี ค.ศ. 1968 เบิร์นบอม (Birnbaum) ได้พัฒนาโมเดลโลจิสติก (logistic model) ซึ่งเป็นฟังก์ชันที่สามารถคำนวณได้ง่ายขึ้น ในปี ค.ศ. 1974 ลอร์ด (Lord) ได้เสนอโมเดลโลจิสติก แบบ 3 พารามิเตอร์ โดยเพิ่มพารามิเตอร์โอกาสการเดาข้อสอบในการประมาณค่าพารามิเตอร์ของผู้สอบด้วย

โมเดลการตอบสนองข้อสอบในระยะแรกจะใช้กับแบบสอบที่ให้คะแนนรายข้อเป็น 2 ระดับ หรือให้คะแนนแบบ 0-1 ซึ่งเรียกว่าโมเดลการตอบสนองข้อสอบแบบทวิวิภาค (Dichotomous Item Response Models) ต่อมาได้มีผู้พัฒนาโมเดล IRT เพื่อใช้กับแบบสอบและแบบวัดทัศนคติที่ตรวจให้คะแนนเป็นแบบมาตรประมาณค่า หรือแบบพหุวิภาค(polytomous) โมเดลในแนวนี้เรียกว่า Polytomous Item Response Models (Muraki, 1992, 1993; Donoghue, 1994; Hambleton and Zaal, 1991) โดยมีความเชื่อที่สำคัญของทฤษฎี ดังนี้

ทฤษฎีนี้มีความเชื่อว่าพฤติกรรมการตอบสนองข้อสอบของผู้สอบ ซึ่งเป็นสิ่งที่สังเกตได้โดยตรงจะถูกกำหนดโดยคุณลักษณะ(trait) หรือความสามารถ (ability) ที่มีอยู่ในตัวบุคคลซึ่งไม่สามารถสังเกตได้ และยังเชื่อว่าค่าพารามิเตอร์ต่าง ๆ ของข้อสอบไม่ว่าจะเป็นค่าความยาก (b) ค่าอำนาจจำแนก (a) หรือค่าการเดา (c) ของข้อสอบแต่ละข้อเป็นคุณลักษณะที่มีอยู่ประจำ และคงที่ในตัวข้อสอบ ฉะนั้นค่าพารามิเตอร์ของข้อสอบจึงไม่แปรเปลี่ยนไปตามกลุ่มตัวอย่าง และค่าความสามารถของผู้สอบซึ่งเชื่อ

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ว่าเป็นคุณลักษณะที่มีอยู่ในตัวผู้สอบนั้นจริง จึงไม่ควรแปรเปลี่ยนไปตามค่าความยากของข้อสอบ เนื่องจากความสามารถของผู้สอบเป็นคุณลักษณะซึ่งไม่สามารถวัดหรือสังเกตได้โดยตรง นักวัดผลทางการศึกษาจึงได้พยายามหาความสัมพันธ์ระหว่างปริมาณการกระทำข้อสอบ หรือคะแนน (test performance or score) กับปริมาณความสามารถ (ability) โดยแสดงความสัมพันธ์ดังกล่าวในลักษณะโมเดลทางคณิตศาสตร์ที่กำหนดด้วยฟังก์ชันทางคณิตศาสตร์ เรียกว่า ฟังก์ชันลักษณะข้อสอบ (Item Characteristic Function) หรือเรียกว่าโค้งคุณลักษณะของข้อสอบ (Item Characteristic Curve) หรือฟังก์ชันการตอบสนองข้อสอบ (Item Response Function) (ศิริชัย กาญจนวาสี , 2538)

สำหรับการเลือกโมเดลของทฤษฎีการตอบสนองข้อสอบมาใช้นั้น จะต้องเลือกให้เหมาะสมกับสถานการณ์ต่าง ๆ ตามข้อตกลงเบื้องต้นของแต่ละโมเดล โดยทฤษฎีนี้มีจุดเด่นหลายประการ ดังนี้ (Hambleton and Swaminathan, 1985 อ้างถึงใน กนกวรรณ รัตนธน. 2544)

ประการแรก ค่าพารามิเตอร์ของข้อสอบเป็นอิสระจากกลุ่มผู้สอบที่ใช้ในการประมาณค่า กล่าวคือ ค่าพารามิเตอร์ของข้อสอบแต่ละข้อ ได้แก่ ค่าความยาก (b) ค่าอำนาจจำแนก (a) และค่าการเดา (c) เป็นค่าคงที่ไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบ ถึงแม้ว่าผู้สอบต่างกลุ่มกัน ค่าพารามิเตอร์ของข้อสอบจะมีค่าคงเดิมเสมอ และในทำนองเดียวกันค่าพารามิเตอร์ความสามารถของผู้สอบเป็นอิสระจากข้อสอบที่ใช้ในการประมาณค่า จะไม่แปรเปลี่ยนตามค่าพารามิเตอร์ของข้อสอบ ไม่ว่าผู้สอบจะทำข้อสอบข้อใดหรือได้ทำข้อสอบเมื่อใดก็ตาม

ประการที่สอง การเปรียบเทียบความสามารถของผู้สอบ จะไม่ขึ้นอยู่กับคำถามในแบบสอบ ถึงแม้ว่าผู้สอบจะใช้ข้อคำถามต่างกัน ก็สามารถนำค่าความสามารถมาเปรียบเทียบกันได้ ทั้งนี้เนื่องจากค่าความสามารถที่ประมาณค่ามาได้นั้น เป็นคะแนนโลจิท (Logit) ซึ่งอยู่ในมาตรวัดเดียวกัน

ประการที่สาม การรายงานคุณภาพของข้อสอบในรูปค่าอินฟอร์เมชัน (information) สามารถรายงานได้ทั้งเป็นรายข้อและทั้งฉบับ ค่าอินฟอร์เมชันเป็นตัวบ่งชี้ถึงความถูกต้องแม่นยำ (accuracy) ในการประมาณค่าความสามารถ สามารถนำมาใช้แทนค่าความเที่ยง และค่าความคลาดเคลื่อนมาตรฐานในการวัดได้

ทฤษฎีการตอบสนองข้อสอบอธิบายความสัมพันธ์ระหว่างปริมาณของการกระทำหรือคะแนนของผู้สอบกับระดับความสามารถของผู้สอบ ซึ่งแสดงได้ดังสมการต่อไปนี้ (สุพัฒน์ สุกมลสันต์, 2539 อ้างถึงใน กนกวรรณ รัตนธน, 2544)

$$P = f(\theta)$$

เมื่อ P แทน ผลการสอบ (performance)

θ แทน ความสามารถ (ability หรือ trait)

f แทน ความสัมพันธ์ (function)

จากความสัมพันธ์ในสมการดังกล่าว เป็นการแสดงความสัมพันธ์ทั่ว ๆ ไปยังไม่เฉพาะเจาะจง จึงมีผู้สนใจศึกษาและพัฒนาเพื่อให้ได้สมการที่สามารถอธิบายลักษณะเฉพาะของข้อสอบแต่ละข้อที่สามารถอธิบายความสัมพันธ์ของสองตัวแปรนี้ได้เด่นชัดขึ้น โดยอาศัยข้อตกลงเบื้องต้นบางประการและวิธีการทางคณิตศาสตร์

ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบ(IRT)มีข้อตกลงเบื้องต้นดังต่อไปนี้ (Hambleton and Swaminathan,1985)

1.ความเป็นมิติเดียว (Unidimension) คุณลักษณะภายในหรือความสามารถของผู้สอบมีอยู่หลายประการ ซึ่งคุณลักษณะภายในหรือความสามารถแต่ละอย่างสามารถกำหนดพฤติกรรมการตอบสนองข้อสอบ ถ้าผลการตอบสนองข้อสอบหรือคะแนนของผู้สอบสามารถอธิบายได้คุณลักษณะเดียวก็ถือว่ามีความเป็นมิติเดียว โมเดลของการตอบสนองข้อสอบมีข้อตกลงว่า ข้อสอบแต่ละข้อในแบบสอบนั้นวัดความสามารถหรือคุณลักษณะเดียวกัน แต่ข้อตกลงข้อนี้ไม่เข้มงวดนัก ถ้ามีลักษณะเด่นที่จะวัดองค์ประกอบใดองค์ประกอบหนึ่งก็ถือว่าเป็นมิติเดียวเช่นกัน

วิธีการตรวจสอบว่าแบบสอบนั้นวัดในมิติเดียวหรือไม่นั้น มีวิธีการทดสอบได้หลายอย่าง แต่ที่นิยมมากมี 2 วิธี คือ (Warm, 1978 อ้างถึงใน สุพัฒน์ สุกมลสันต์, 2539)

1.1 โดยการหาค่าความสัมพันธ์ระหว่างค่าน้ำหนักองค์ประกอบรายข้อ (factor loading) ขององค์ประกอบที่หนึ่ง กับค่าสหสัมพันธ์แบบไบซีเรียล (biserial correlation) ของข้อสอบรายข้อกับคะแนนรวม หากค่าสหสัมพันธ์มีค่ามากกว่า .80 ก็แสดงว่าแบบสอบนั้นวัดมิติเดียว

1.2 โดยการวิเคราะห์องค์ประกอบ (factor analysis) ของข้อสอบทั้งฉบับ ถ้าค่าไอเกน (eigen value) ค่าสูงสุดแตกต่างจากค่าอื่นอย่างชัดเจนหรือไม่ ถ้าแตกต่างก็แสดงว่าแบบสอบนั้นวัดมิติเดียว

2.ความเป็นอิสระในการตอบข้อสอบ (Local Independence) หมายถึง การตอบข้อสอบข้อต่างๆ ในแบบสอบของผู้สอบนั้นมีความเป็นอิสระในเชิงสถิติ กล่าวคือ การตอบข้อสอบข้อหนึ่ง ๆ จะไม่มีผลกระทบต่อการตอบข้อสอบข้ออื่น ๆ ในแบบสอบ เนื้อหาของข้อสอบข้อหนึ่ง ๆ จะต้องไม่มีเงื่อนไขในการตอบข้อสอบข้ออื่น ๆ แมคโดแนล (McDonald, 1980a, 1980b, 1982 อ้างถึงใน Hambleton and Swaminathan ,1985) ได้นิยามของความเป็นมิติเดียวของข้อสอบว่าควรอยู่บนพื้นฐานของหลักการความเป็นอิสระในการตอบข้อสอบ

3. โค้งลักษณะของข้อสอบ (Item Characteristic Curves : ICC) เป็นฟังก์ชันทางคณิตศาสตร์ที่แสดงถึงความสัมพันธ์ระหว่างโอกาสในการตอบข้อสอบนั้นได้ถูกต้องกับระดับความสามารถที่วัดได้โดยชุดของข้อสอบหรือแบบสอบ จะเห็นได้ว่าโอกาสที่ผู้สอบจะตอบข้อสอบถูก (probability) จะขึ้นอยู่กับโค้งลักษณะของข้อสอบ (ICC) ซึ่งเป็นอิสระจากการกระจายของความสามารถของผู้สอบ นั่นคือโอกาสที่ผู้สอบจะตอบข้อสอบถูกไม่ขึ้นอยู่กับจำนวนของผู้สอบที่มีความสามารถหมือนกัน รูปร่างของ

โค้งลักษณะของข้อสอบ (ICC) ในแต่ละข้อมีคุณสมบัติไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบ ดังนั้นจึงทำให้โอกาสในการตอบข้อสอบถูกในแต่ละข้อไม่แปรเปลี่ยน

นอกจากนี้แล้ว ถ้าข้อตกลงเบื้องต้นของการทดสอบตรงตามที่ทฤษฎีกำหนดไว้ ผลที่ได้จากการคำนวณจะมีลักษณะที่ดีดังนี้

1. ค่าพารามิเตอร์ของข้อสอบ คือ ค่าความยาก (a) ค่าอำนาจจำแนก (b) และค่าการเดา (c) จะเป็นค่าที่ไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบ ไม่ว่าจะนำไปสอบกับผู้ใดก็ตาม ลักษณะเช่นนี้เรียกว่าความเป็นอิสระจากกลุ่มตัวอย่าง (person free test calibration)

2. เมื่อทราบลักษณะการตอบข้อสอบแต่ละข้อของผู้สอบ เราสามารถจะคำนวณหาค่าความสามารถที่แท้จริงของผู้สอบคนนั้นได้ ค่าความสามารถที่แท้จริงนี้มีความสัมพันธ์กันโดยตรงกับคะแนนจริง การคำนวณค่าความสามารถที่แท้จริงนั้นอาจใช้ข้อสอบข้อใดก็ได้ที่วัดสิ่งเดียวกัน ลักษณะเช่นนี้เรียกว่า ความเป็นอิสระของข้อสอบ (item free person measurement)

โค้งลักษณะของข้อสอบมีหลายรูปแบบตามความเชื่อเกี่ยวกับความสัมพันธ์ระหว่างระดับความสามารถกับโอกาสในการตอบข้อสอบได้ถูกต้อง จึงทำให้มีโมเดลของโค้งลักษณะของข้อสอบขึ้นหลายโมเดล ดังจะกล่าวต่อไปนี้

โมเดลการตอบสนองข้อสอบในระยะแรกจะใช้กับแบบสอบที่ให้คะแนนรายข้อเป็น 2 ระดับ หรือให้คะแนนแบบ 0-1 ซึ่งเรียกว่าโมเดลการตอบสนองข้อสอบแบบทวิวิภาค (Dichotomous Item Response Models) แต่ถ้านำไปใช้กับข้อสอบหลายตัวเลือก ซึ่งมีการให้คะแนนบางส่วน (partial credit) หรือมาตรประมาณค่า (rating scale) จะใช้ไม่ได้ หรือจะต้องปรับการให้คะแนนเป็นแบบ 0-1 ซึ่งทำให้สูญเสียสารสนเทศของการตอบบางส่วนไป (ศิริชัย กาญจนวาสี, 2542) ต่อมาได้มีผู้พัฒนาโมเดล IRT เพื่อใช้กับแบบสอบและแบบวัดทัศนคติที่ตรวจให้คะแนนเป็นแบบมาตรประมาณค่า หรือแบบพหุวิภาค (polytomous) โมเดลในแนวนี้เรียกว่า Polytomous Item Response Models (Muraki, 1992, 1993; Donoghue, 1994; Hambleton and Zaal, 1991) การตรวจให้คะแนนรายข้อมากกว่า 2 ระดับ หรือแบบพหุวิภาค สามารถจำแนกเป็นกรณีต่าง ๆ โดย ศิริชัย กาญจนวาสี (2542) ได้สรุปเป็นสาระสำคัญ ดังนี้

1. คำตอบรายข้อมีหลายรายการที่แสดงคุณลักษณะเฉพาะแตกต่างกัน(Nominal Categories) เป็น กรณีที่คำถามทุกข้อมีหลายรายการคำตอบ (Multiple-Choice Items) ซึ่งแต่ละรายการคำตอบมีคุณลักษณะเฉพาะแตกต่างกัน และรายการคำตอบไม่มีความสัมพันธ์ตามลำดับขั้นต่อกัน โมเดลที่ใช้จึงต้องทำการวิเคราะห์ทั้งคำตอบถูกและตัวเลือกทุกตัว โมเดลการตอบสนองข้อสอบที่ใช้แสดงความสัมพันธ์ของข้อมูลการตอบ ลักษณะนี้เรียกว่า Nominal Response Model (NRM) โมเดลนี้พัฒนาโดย บอค (Bock) ในปี ค.ศ. 1972

2. คำตอบรายข้อมีหลายรายการที่สามารถเรียงลำดับแต่จำนวนรายการคำตอบของแต่ละข้อไม่เท่ากัน (Specific-Category Items) เป็นกรณีที่แต่ละข้อคำถามมีจำนวนรายการหรือสเกลการตอบโดยเฉพาะ เช่น การตรวจให้คะแนนรายข้อที่มีรายการที่เป็นคำตอบถูก และรายการที่มีการให้คะแนนบางส่วน (partial credit) คะแนนเต็มของแต่ละข้ออาจไม่เท่ากัน โมเดลการตอบสนองข้อสอบนี้เรียกว่า Partial Credit Model (PCM) โมเดลนี้พัฒนาโดย มาสเตอร์ (Masters) ในปี ค.ศ. 1982

3. คำตอบรายข้อมีหลายรายการที่จัดเรียงลำดับเป็นสเกลรวมของค่าตัวเลขจำนวนเต็ม (Common Ordered Category Items) เป็นกรณีที่ทุกข้อคำถามมีรายการคำตอบร่วมกันเป็นค่าตัวเลขจำนวนเต็มเรียงตามลำดับ เช่น มาตรประมาณค่า (Rating Scale) โมเดลการตอบสนองนี้เรียกว่า Rating Scale Model (RSM) โมเดลนี้พัฒนาโดย แอนดริช (Andrich) ในปี ค.ศ. 1978

4. คำตอบรายข้อมีหลายรายการที่จัดเรียงลำดับเป็นค่าตัวเลขจำนวนเต็ม แต่จำนวนรายการคำตอบของแต่ละข้อไม่เท่ากัน (Graded Response Items) เป็นกรณีที่ทุกข้อคำถามมีรายการคำตอบเป็นค่าตัวเลขจำนวนเต็มเรียงตามลำดับ แต่จำนวนรายการของแต่ละข้อไม่จำเป็นต้องเท่ากัน โมเดลการตอบสนองนี้เรียกว่า Graded Response Model (GRM) โมเดลนี้พัฒนาโดย ซามิจิมา (Samejima) ในปี ค.ศ. 1960

จะเห็นว่าทฤษฎีการตอบสนองข้อสอบได้รับการยอมรับมากขึ้นในการทดสอบทางจิตวิทยาและทางการศึกษา เนื่องจากวิธีการสร้าง วิเคราะห์ และการให้คะแนนมีประสิทธิภาพมากกว่าทฤษฎีการทดสอบแบบดั้งเดิม (Mislery and Bock ,1990) ทฤษฎีการตอบสนองข้อสอบจึงถูกนำไปประยุกต์ใช้ในหลาย ๆ เรื่อง โดยเฉพาะการสร้างคลังข้อสอบ (Item banking) การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ (Adaptive testing) การเทียบคะแนนการสอบ (Test score equating) การหาความลำเอียงของข้อสอบ (Item bias)

สำหรับรายละเอียดของโมเดลการตอบสนองข้อสอบแบบทวิวิภาค และโมเดลการตอบสนองข้อสอบแบบพหุวิภาค จะนำเสนอในหัวข้อต่อไป

2.2 มโนทัศน์เกี่ยวกับทฤษฎีการตอบสนองข้อสอบแบบทวิวิภาค (Dichotomous Item Response Theory)

โค้งลักษณะของข้อสอบแสดงความสัมพันธ์ระหว่างระดับความสามารถกับโอกาสในการตอบข้อสอบได้ถูกต้อง ทำให้เกิดโมเดลของโค้งลักษณะของข้อสอบขึ้นหลายโมเดล แต่รูปแบบที่นิยมใช้กันแพร่หลาย คือ รูปแบบโลจิสติก (Logistic Model) ซึ่งนิยมใช้กันทั่วไป 3 รูปแบบ ดังนี้ (Hambleton and Swaminathan ,1985 อ้างถึงใน กนกวรรณ รัตนธน, 2544)

รูปแบบโลจิสติกที่มี 1 พารามิเตอร์ (One-Parameter Logistic Model)

ราส์ชได้พัฒนาทฤษฎีการตอบสนองข้อสอบ และเสนอการวิเคราะห์ข้อสอบแบบ 1 พารามิเตอร์ ในปี ค.ศ. 1960 โดยฟังก์ชันของแบบ 1 พารามิเตอร์นี้สามารถอธิบายได้ด้วยพารามิเตอร์ของข้อสอบเพียงตัวเดียว คือ ค่าความยาก (b) เขียนเป็นสมการได้ดังนี้

$$P_i(\theta) = -\frac{e^{D(\theta - b_i)}}{1 - e^{D(\theta - b_i)}} \quad ; \; i = 1, 2, 3, \ldots, n$$

เมื่อ $P_i(\theta)$ แทน ความน่าจะเป็นของผู้สอบที่มีความสามารถ (θ) จะตอบข้อสอบข้อที่ i ได้ถูกต้อง

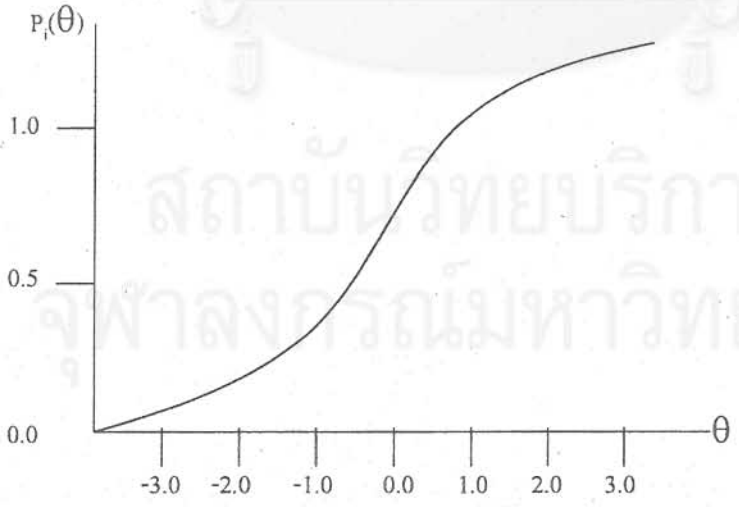
D แทน ค่าคงที่ของ scale factor เพื่อปรับค่าของฟังก์ชั่นโลจิสติกให้ใกล้เคียงกับค่าที่ได้จากฟังก์ชันรูปตัวเอสปกติ มีค่าประมาณ 1.7

b_i แทน ระดับความยากของข้อสอบข้อที่ i ที่แสดงถึงระดับความสามารถที่จุดโค้งลักษณะข้อสอบมีความชันมากที่สุด ในทางปฏิบัติค่าความยากจะมีค่าอยู่ระหว่าง –2 ถึง +2

θ แทน ระดับความสามารถที่แท้จริงที่คำนวณจากคะแนนรวม โดยการปรับให้เป็นคะแนนที่มีค่าเฉลี่ยเท่ากับ 0 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 โดยมีพิสัยการกระจายเท่ากับ $-\alpha$ ถึง $+\alpha$ แต่ในทางปฏิบัติระดับความสามารถจะมีค่าอยู่ระหว่าง -3 ถึง +3

e แทน ค่าคงที่ซึ่งมีค่าประมาณ 2.7182818

การวิเคราะห์ข้อสอบแบบ 1 พารามิเตอร์ (1-parameter model) นี้มีข้อตกลงว่าข้อสอบทุกข้อไม่มีโอกาสของการเดาถูก ($c_i = 0$) และค่าอำนาจจำแนกของข้อสอบ (a_i) ทุกข้อเท่ากันหมด โค้งลักษณะเฉพาะของข้อสอบแสดงในแผนภาพที่ 2.1



แผนภาพที่ 2.1 โค้งลักษณะเฉพาะของข้อสอบของการวิเคราะห์ข้อสอบแบบ 1 พารามิเตอร์

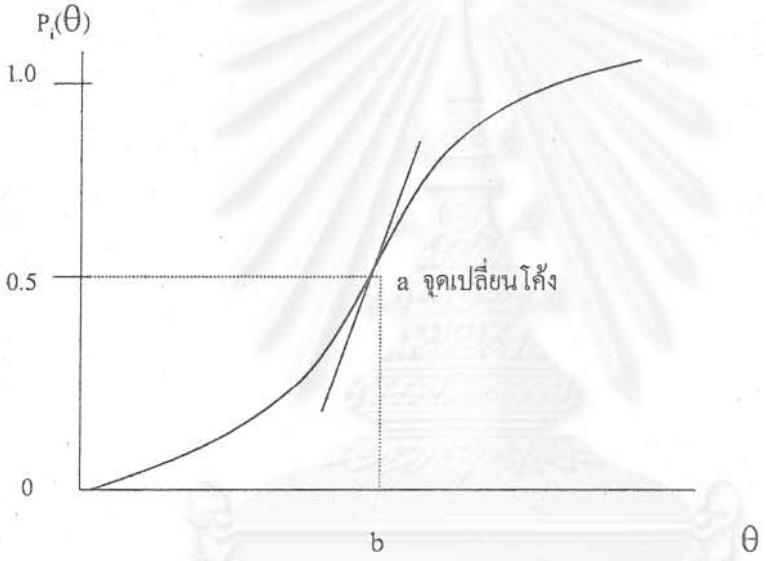
รูปแบบโลจิสติกที่มี 2 พารามิเตอร์ (Two-Parameter Logistic Model)

เบิร์นบวม (Birnbaum) ได้เสนอการวิเคราะห์ข้อสอบแบบโลจิสติกที่มีพารามิเตอร์ 2 ตัว คือ ค่าความยาก (b) และค่าอำนาจจำแนก (a) และสามารถเขียนเป็นสมการได้ดังนี้

$$P_i(\theta) = \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}} \qquad i = 1, 2, 3, \ldots, n$$

เมื่อ a_i แทน ค่าอำนาจจำแนกของข้อสอบข้อที่ i

การวิเคราะห์ข้อสอบแบบ 2 พารามิเตอร์ (2-parameter model) นี้ มีข้อตกลงว่าข้อสอบทุกข้อไม่มีโอกาสของการเดาถูก คือ ค่า $c_i = 0$ โค้งลักษณะเฉพาะของข้อสอบแบบนี้มีลักษณะดังแผนภาพที่ 2.2



แผนภาพที่ 2.2 โค้งลักษณะเฉพาะของข้อสอบของการวิเคราะห์ข้อสอบแบบ 2 พารามิเตอร์

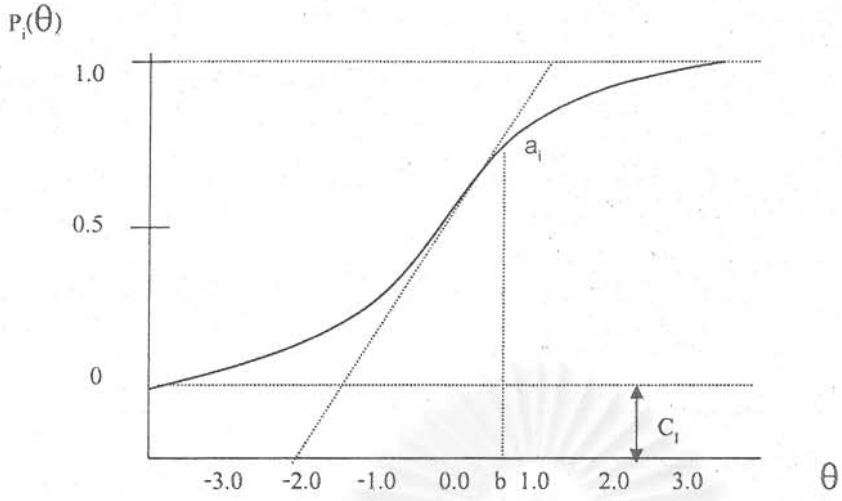
รูปแบบโลจิสติกที่มี 3 พารามิเตอร์ (Three-Parameter Logistic Model)

การวิเคราะห์ข้อสอบแบบ 3 พารามิเตอร์ (3-parameter model) นี้ ดัดแปลงมาจากรูปแบบโลจิสติกที่มี 2 พารามิเตอร์ พารามิเตอร์ที่เพิ่มขึ้น คือ ค่าการเดา (c_i) และมีสมการดังนี้

$$P_i(\theta) = c_i + (1 - c_i)\frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}} \qquad i = 1, 2, 3, \ldots, n$$

เมื่อ c_i แทน ค่าการเดา

โค้งลักษณะเฉพาะของข้อสอบของการวิเคราะห์ข้อสอบแบบ 3 พารามิเตอร์ มีลักษณะดังแผนภาพที่ 2.3



แผนภาพที่ 2.3 โค้งลักษณะเฉพาะของข้อสอบของการวิเคราะห์ข้อสอบแบบ 3 พารามิเตอร์

ฟังก์ชันสารสนเทศของข้อสอบ (item information function)

ฟังก์ชันสารสนเทศของข้อสอบ คือ ค่าที่แสดงถึงความถูกต้องแม่นยำในการประมาณค่าความสามารถจริง (θ_i) ของผู้สอบในการตอบข้อสอบแต่ละข้อ ซึ่งสามารถคำนวณได้จากสูตรต่อไปนี้

$$I(\theta,U_i) = P_i^{'}(\theta)^2 / P_i(\theta)\, Q_i(\theta)$$

เมื่อ $I(\theta, U_i)$ แทน ฟังก์ชันสารสนเทศของข้อสอบ

$P_i^{'}(\theta)$ แทน ความชันของโค้งลักษณะเฉพาะของข้อสอบที่ระดับความสามารถ θ

$P_i(\theta)$ แทน ความน่าจะเป็นที่ผู้สอบที่มีความสามารถ θ จะตอบข้อสอบ ข้อที่ i ได้ถูกต้อง

$Q_i(\theta)$ แทน $1-P_i(\theta)$

จากสมการข้างต้น จะเห็นได้ว่า ฟังก์ชันสารสนเทศของข้อสอบแต่ละข้อขึ้นอยู่กับความชันของโค้งลักษณะเฉพาะของข้อสอบ ถ้าโค้งลักษณะเฉพาะของข้อสอบชันมากขึ้นในขณะที่ความแปรปรวนของการตอบข้อสอบถูกน้อยลง โค้งสารสนเทศของข้อสอบที่ระดับความสามารถนั้น ๆ จะยิ่งสูงขึ้น ความสูงของโค้งสารสนเทศของข้อสอบอยู่ที่ระดับความสามารถใด แสดงว่าสามารถจำแนกระดับความสามารถของผู้สอบได้ดี ณ ระดับความสามารถนั้น

ฟังก์ชันสารสนเทศของแบบสอบ (test information function)

ฟังก์ชันสารสนเทศของแบบสอบเป็นสัดส่วนกลับกันกับกำลังสองของความยาวของช่วงความเที่ยง ซึ่งเป็นผลมาจากการประมาณค่าความสามารถของผู้สอบจากการตอบข้อสอบทั้งฉบับ ค่านี้แสดงถึงความถูกต้องแม่นยำในการประมาณค่าความสามารถจริง (θ) ของแบบสอบทั้งฉบับว่ามีมากน้อย

เพียงใด ดังนั้นโค้งสารสนเทศของแบบสอบจึงเป็นตัวบ่งชี้ถึงความถูกต้องแม่นยำของค่าความสามารถที่ประมาณได้ ซึ่งแสดงในรูปสมการได้ดังนี้

$$I(\theta) = \sum_{i=1}^{n} I(\theta, U_i) = \Sigma P_i'(\theta)^2 / P_i(\theta) Q_i(\theta)$$

เมื่อ $I(\theta)$ แทน ฟังก์ชันสารสนเทศของแบบสอบ

$I(\theta, U_i)$ แทน ฟังก์ชันสารสนเทศของข้อสอบ

ถ้าเรามีกลุ่มข้อสอบที่ทราบค่าสารสนเทศของข้อสอบ เราสามารถสร้างแบบสอบให้มีโค้งสารสนเทศของแบบสอบ ณ ความสามารถระดับใดระดับหนึ่งที่เราต้องการ เช่น การสร้างแบบสอบเพื่อคัดเลือกนักเรียนเข้าศึกษาต่อ ก็ต้องสร้างแบบสอบที่มีประสิทธิภาพสูงสุดที่ระดับความสามารถสูง ๆ นั่นคือ ให้มีโค้งสารสนเทศของแบบสอบสูง ณ ระดับความสามารถสูง ๆ

1.3 มโนทัศน์เกี่ยวกับทฤษฎีการตอบสนองข้อสอบแบบพหุวิภาค (Polytomous Item Response Theory)

โมเดล IRT ที่พัฒนาในระยะแรกจะใช้กับแบบสอบที่ให้คะแนนแบบทวิวิภาค (dichotomous) เป็นส่วนใหญ่ ต่อมาได้มีผู้พัฒนาโมเดล IRT เพื่อใช้กับแบบสอบและแบบวัดทัศนคติที่ตรวจให้คะแนนแบบพหุวิภาค (polytomous) หรือลักษณะมาตรประมาณค่า (rating data) โมเดลในแนวทฤษฎีนี้ทั้งหมดเรียกว่า (Polytomous Item Response Models (Muraki, 1992, 1993; Donoghue, 1994; Hambleton and Zaal, 1991) มีโมเดลที่สำคัญ คือ

1. Graded Response Model (GRM) พัฒนาโดย ซามิจิมา (Samejima) ในปี ค.ศ. 1960
2. Normal Response Model (NRM) พัฒนาโดย บอค (Bock) ในปี ค.ศ. 1972
3. Continuous Model (CM) พัฒนาโดย ซามิจิมา ในปี ค.ศ.1973
4. Rating Scale Model (RSM) พัฒนาโดย แอนดริช (Andrich) ในปี ค.ศ. 1978
5. Partial Credit Model (PCM) พัฒนาโดย มาสเตอร์ (Masters) ในปี ค.ศ. 1982
6. Sucessive Interval Model (SIM) พัฒนาโดย รอส์ท (Rost) ในปี ค.ศ. 1988
7. Generalized Partial Credit Model : GPCM พัฒนาโดย มูราคิ (Muraki) ในปี ค.ศ. 1992

โมเดลที่มีผู้ศึกษากันมากในปัจจุบัน ไม่เข้มงวดเกี่ยวกับข้อตกลงเบื้องต้น และสามารถใช้ได้กับแบบสอบและแบบวัดหลายลักษณะคือ GRM และ GPCM (Donoghue, 1994; De Ayala, 1994; Muraki, 1992, 1993; Reise and Yu, 1990; Koch and De Ayala, 1989; Koch, 1983)

ซามิจิมา (Samejima) ได้พัฒนา GRM จากโมเดลโลจิสติก 2 พารามิเตอร์ เพื่อใช้กับแบบสอบและแบบวัดทัศนคติที่ตรวจให้คะแนนแบบพหุวิภาค (polytomous) ในขณะที่มูราคิ (Muraki) ได้พัฒนา GPCM จาก PCM ของ ไรท์ และมาสเตอร์ (Wright and Masters, 1982) ให้สามารถใช้ได้กับแบบสอบ

และแบบวัดที่มีลักษณะเป็นพหุมิติ โดยรวมค่าอำนาจจำแนกข้อกระทงในการประมาณค่าพารามิเตอร์ของฟังก์ชันด้วยโมเดลทั้ง 2 เมื่อพิจารณาดูแล้วคล้ายคลึงกันมาก เพียงแต่มีมโนทัศน์เกี่ยวกับฟังก์ชันลักษณะเชิงปฏิบัติ (operating characteristic function : OCF) และพัฒนาการที่แตกต่างกัน รูปแบบทั่วไปของโมเดลในแนวคิดนี้พัฒนามาจากสูตร

$$P_i(U_i = 1/\theta) = \frac{\exp^{[a_j(\theta - b_j)]}}{1 + \exp^{[a_j(\theta - b_j)]}}$$

เมื่อ $P_j(U_j=1/\theta)$ คือ ความน่าจะเป็นของคนที่ i มีความสามารถ สามารถตอบข้อสอบข้อ j (U=1) ได้ถูกต้อง

a_j คือ ค่าพารามิเตอร์อำนาจจำแนก

b_j คือ ค่าพารามิเตอร์ของความยากของข้อสอบ

exp คือ ค่าคงที่ 2.7183

ดังนั้น ในการศึกษาครั้งนี้จึงขอนำเสนอเฉพาะโมเดล GRM และ GPCM ดังมีรายละเอียดพอสังเขป ดังนี้

1) GRM (Graded Response Model)

GRM มีข้อตกลงว่าการตอบสนองข้อกระทง j ของผู้ตอบ สามารถแบ่งคะแนนออกเป็น mj+1 ลำดับขั้น (categories) คะแนนแต่ละลำดับ (j_k) กำหนดให้เป็น 0, 1, 2, ..., mj ตามลำดับ (Donoghue, 1994; Koch, 1983) เขียนสมการทั่วไปได้ดังนี้

$$P_{jk}(\theta) = \frac{\exp^{[Da_j(\theta - b_{jk})]}}{1 + \exp^{[Da_j(\theta - b_{jk})]}}$$

เมื่อ $P_{jk}(\theta)$ คือ ความน่าจะเป็นที่ผู้มีคุณลักษณะ θ จะตอบข้อกระทง j ได้คะแนน k

D คือ ค่าคงที่ของสเกล (scaling constant) เมื่อปรับโค้งฟังก์ชันของโมเดลโลจิสติก และโมเดลนอร์มอล ออกไจฟ์ มีค่าเท่ากับ 1.70

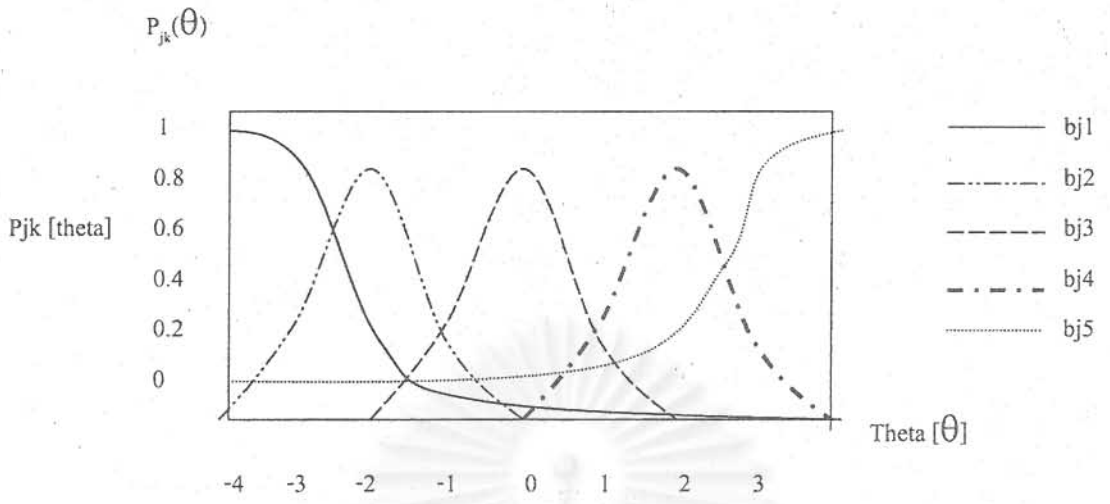
a_j คือ ค่าอำนาจจำแนกของข้อความ j

θ คือ ความสามารถหรือคุณลักษณะภายในของผู้สอบ

b_{jk} คือ ค่าความยากของข้อความ j ในลำดับขั้นที่ jk เมื่อ jk = 0, 1, ... k

exp คือ ค่าคงที่ มีค่าเท่ากับ 2.7183

ฟังก์ชันทางคณิตศาสตร์ดังกล่าวสามารถเขียนแทนด้วยโค้งฟังก์ชันสารสนเทศได้ดังแผนภาพที่ 2.4 (Koch, 1983 : 18)



แผนภาพที่ 2.4 โค้งฟังก์ชันสารสนเทศของ GRM

ในโมเดลนี้แสดงถึงความสัมพันธ์ของคนที่มีคุณลักษณะภายใน (θ) สูง มีความน่าจะเป็นในการตอบลำดับขั้นคะแนนที่สูงกว่าคนที่มีคุณลักษณะภายในต่ำ ซึ่งนักวัดผลได้มีการประยุกต์ใช้โมเดลนี้กับแบบสอบชนิดปรับเหมาะ (Adaptive Testing) และแบบวัดทัศนคติต่าง ๆ

ลักษณะเฉพาะฟังก์ชันของโมเดล GRM สรุปได้ดังนี้ •

1.1) ฟังก์ชันสารสนเทศของแบบวัด (TIF) เขียนแทนด้วยสมการดังนี้

$$I(\theta) = \sum_{j=1}^{m} I_j(\theta)$$

เมื่อ $I(\theta)$ คือ ฟังก์ชันสารสนเทศของข้อกระทงตั้งแต่ข้อ 1 ถึงข้อที่ m

1.2) ฟังก์ชันสารสนเทศของข้อกระทง (IIF) สามารถแทนด้วยสมการดังนี้

$$I_j(\theta) = \sum_{jk=0}^{k} \frac{[P'_{jk}(\theta)]^2}{P_{jk}(\theta)}$$

เมื่อ $P'_{jk}(\theta)$ คือ ความน่าจะเป็นสำหรับคนที่มีความสามารถ θ จะตอบข้อกระทง j ได้คะแนน 1 มากกว่าคะแนน k ใด ๆ

$P_{jk}(\theta)$ คือ ความน่าจะเป็นสำหรับคนที่มีความสามารถ θ จะตอบข้อกระทง j ได้คะแนน k

1.3) ฟังก์ชันสารสนเทศของแต่ละลำดับขั้นคะแนน (ICIFs) เขียนแทนด้วยสมการดังนี้

$$I_j(\theta) = \frac{[P'_{jk}(\theta)]^2}{P_{jk}(\theta)}$$

เมื่อ $P'_{jk}(\theta)$ คือ ความน่าจะเป็นสำหรับคนที่มีความสามารถ θ จะตอบข้อกระทง j ในลำดับขั้นคะแนนที่ 1 ได้ถูกต้องมากกว่าคะแนน k ใด ๆ

$P_{jk}(\theta)$ คือ ความน่าจะเป็นสำหรับคนที่มีความสามารถ θ จะตอบข้อกระทง j ในลำดับขั้นคะแนน k ได้ถูกต้อง

2) GPCM (Generalized Partial Credit Model)

GPCM (Generalized Partial Credit Model) พัฒนาจาก PCM ให้มีการตรวจให้คะแนนความรู้บางส่วนในการแก้โจทย์ปัญหาคณิตศาสตร์ โดยพิจารณาให้คะแนนตามลำดับขั้นความสำเร็จของการแก้โจทย์ปัญหาเป็นคะแนนเรียงลำดับ 0, 1, 2, 3 (Masters, 1982) จากนั้นก็มีการนำไปใช้กับแบบวัดทัศนคติที่มีคะแนนเรียงลำดับหลายค่า (Dodd and Koch, 1989)

มูราคิ (Muraki) ได้พัฒนาโมเดลดังกล่าวมาเป็น GPCM โดยแก้ไขข้อจำกัดเกี่ยวกับความเป็นเอกมิติของเครื่องมือที่ใช้วัดคุณลักษณะ และใช้ค่าอำนาจจำแนกที่แปรเปลี่ยนไปในแต่ละข้อมาร่วมประมาณค่าพารามิเตอร์ด้วย โดยมีข้อตกลงว่าบุคคลที่มีความสามารถหรือมีคุณลักษณะที่ต้องการวัดสูงมีความน่าจะเป็นที่จะตอบคำตอบลำดับคะแนนที่ k มากกว่า k-1 สามารถแทนด้วยสมการทั่วไปได้ดังนี้

$$C_{jk}(\theta) = P_{jk/k-1,k}(\theta) = \frac{P_{jk}(\theta)}{P_{j,k-1}(\theta) + P_{jk}(\theta)} = \frac{\exp^{[a_j(\theta - b_{jk})]}}{1 + \exp^{[a_j(\theta - b_{jk})]}}$$

เมื่อ k คือ ลำดับขั้นคะแนนที่ 1, 2, 3, ..., mj

$$P_{jk}(\theta) = \frac{C_{jk}}{1 - C_{jk}} P_{j,k-1}(\theta)$$

$P_{jk}(\theta)$ คือ ความน่าจะเป็นของคนที่มีความสามารถจะตอบได้คะแนน k (กรณี k-1 ก็เช่นกัน)

$C_{jk}/(1-C_{jk})$ คือ อัตราส่วนความน่าจะเป็นของสองเงื่อนไข ซึ่งอาจแสดงในรูป สมการ $\exp^{[a_j(\theta - b_{jk})]}$

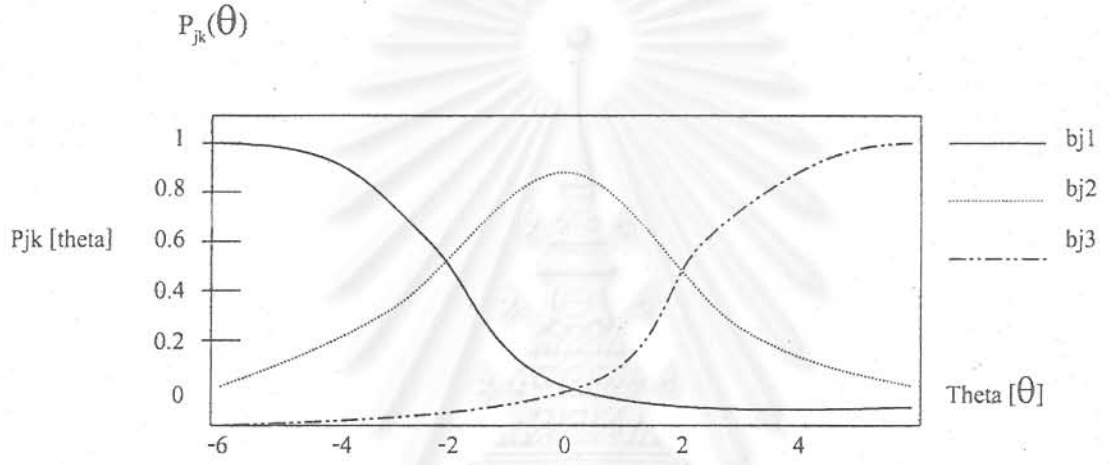
มาสเตอร์ (Masters) เรียก b_{jk} ว่าเป็นค่าพารามิเตอร์ความยากในแต่ละลำดับขั้น เป็นจุดบนแกน θ ที่ฟังก์ชันของ $P_{j.k-1}(\theta)$ และ $P_{jk}(\theta)$ ตัดกัน ซึ่งโค้งทั้ง 2 เป็นฟังก์ชันของการตอบ (item category response function : ICRFs) ใน 2 ลำดับค่าคะแนนที่ตัดกันเพียงจุดเดียวบนค่า θ ใด ๆ คือ

ถ้า $\theta = b_{jk}$, $P_{jk}(\theta) = P_{j.k-1}(\theta)$;

ถ้า $\theta > b_{jk}$, $P_{jk}(\theta) > P_{j.k-1}(\theta)$;

และถ้า $\theta < b_{jk}$, $P_{jk}(\theta) < P_{j.k-1}(\theta)$;

ซึ่งอยู่บนข้อตกลงว่า $a_j > 0$ และ b_{jk} ไม่จำเป็นต้องเรียงกันในข้อสอบ j เพราะค่าพารามิเตอร์จะแสดงถึงขนาดของความน่าจะเป็นในการตอบ $P_{j.k-1}(\theta)$ และ $P_{jk}(\theta)$ ในลำดับที่อยู่ชิดกัน ดังแผนภาพที่ 2.5



แผนภาพที่ 2.5 โค้งฟังก์ชันสารสนเทศของ GPCM

การพิจารณาจุดตัดของโค้งฟังก์ชันการตอบแต่ละลำดับขั้น (item category response functions : ICRFs) ของโมเดล PCM แม้จะง่ายต่อการตีความหมายก็ตาม แต่จุดยอดของโค้งก็ไม่ใช่จะอยู่ตรงกลางของช่วงคะแนนเสนอไป ซึ่งที่มาของค่าฟังก์ชันการตอบสนองข้อกระทงแต่ละลำดับขั้น ($P_{jk}(\theta)$) แสดงได้ดังนี้ (Muraki, 1992)

$$\frac{\partial}{\partial(\theta)} P_{jk}(\theta) = a_j P_{jk}(\theta)[k - \sum_{k=1}^{mj} k P_{jk}(\theta)]$$

โดยกำหนดสมการเป็น $0; \sum_{k=1}^{mj} c P_j(\theta) = k$ สมการนี้จะแสดงโค้ง ICRFs ของ $P_{jk}(\theta)$ เมื่อ k = 1, 2, 3, ..., k-1, k, k-1, ..., mj

พารามิเตอร์ a_j คือพารามิเตอร์อำนาจจำแนกของข้อ j โดยค่า a_j มีพิสัยที่เป็นไปได้อยู่ในช่วง $-\infty$ ถึง $+\infty$ ซึ่งไม่เหมือนใน Dichotomous Model เพราะใน Polytomous Item Response Model อำนาจจำแนกในแต่ละลำดับขั้นคะแนนจะขึ้นอยู่กับผลรวมของพารามิเตอร์อำนาจจำแนกและค่าพารามิเตอร์

เทรสโฮลด์ แต่ใน GPCM จะรวมค่าอำนาจจำแนกไว้ในโมเดลด้วย ซึ่งค่าอำนาจจำแนกนี้จะบอกถึงระดับการตอบในแต่ละลำดับค่าคะแนนที่เปลี่ยนไปในแต่ละข้อเช่นเดียวกับที่ค่า θ เปลี่ยนไปด้วย (Muraki, 1992)

จากแนวคิดดงกล่าวสามารถสรุปฟังก์ชันสารสนเทศที่สำคัญของ GPCM ได้ดังนี้

1. ฟังก์ชันสารสนเทศของแบบวัด (TIF) แทนด้วยสมการดังนี้

$$I(\theta) = \sum_{j=1}^{n} I_j(\theta)$$

เมื่อ $I_j(\theta)$ คือ ฟังก์ชันสารสนเทศของข้อกระทง (IIF)

2. ฟังก์ชันสารสนเทศของข้อกระทง (IIF) ซึ่งสามารถแทนด้วยสมการ

$$I_j(\theta) = D^2 a_j^2 \sum_{c=1}^{mj} [T_k - \bar{T}_j(\theta)]^2 P_{jk}(\theta)$$

เมื่อ $\bar{T}_j = \sum_{k=1}^{mj} T_k P_{jk}(\theta)$

D คือ ค่าคงที่ของสเกลมีค่าเท่ากับ 1.70

a_j คือ ค่าอำนาจจำแนกของข้อ j

T_k คือ ค่าคะแนนใด ๆ ใน $T_j = (1, 2, \ldots, mj)$

$P_{jk}(\theta)$ คือ ความน่าจะเป็นของคนที่มีความสามารถ θ จะตอบข้อ j ได้คะแนน k

K คือ ลำดับขั้นคะแนนที่ 1, 2, ..., k-1, k, k+1, ..., mj

3. ฟังก์ชันสารสนเทศของลำดับค่าคะแนนในแต่ละข้อ (ICIFs) แทนด้วยสมการดังนี้

$$I_{jk}(\theta) = P_{jk}(\theta) I_j(\theta)$$

เมื่อ $P_{jk}(\theta)$ คือ ความน่าจะเป็นของผู้ตอบที่มีความสามารถ θ จะทำข้อสอบข้อ j ได้คะแนน k

$I_j(\theta)$ คือ ฟังก์ชันสารสนเทศข้อกระทง

ตอนที่ 3 งานวิจัยที่เกี่ยวข้อง

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการให้คะแนนความรู้บางส่วนที่ผ่านมา นักวิชาการได้ศึกษาหาวิธีการแก้ปัญหา เพื่อลดโอกาสในการเดา และเพิ่มสารสนเทศเกี่ยวกับการตอบข้อสอบให้มากที่สุด โดยวิธีการให้คะแนนที่เหมาะสมที่จะทำให้คุณสมบัติของการวัดในประเด็นที่เกี่ยวข้องกับความตรงและความเที่ยงสูงขึ้น จากการศึกษาวิธีการให้คะแนนความรู้บางส่วนในต่างประเทศ สรุปเป็นประเภทใหญ่ๆ ได้ 4 วิธี ดังนี้ (Simon , Budescu และ Nevo 1997)

1. การให้น้ำหนักแก่ข้อสอบที่แตกต่างกัน (Differential Item Weighting)
2. การให้น้ำหนักแก่ตัวเลือกที่แตกต่างกัน (Differential Option Weighting)
3. การเปลี่ยนแปลงโครงสร้างของข้อสอบ (Changing the Item Structure)
4. การเปลี่ยนแปลงวิธีการตอบสนอง (Changing the Response Method)

แต่การวิจัยในเรื่องนี้ ได้ศึกษาวิธีการให้คะแนนความรู้บางส่วน 2 วิธี คือ วิธีการตอบและการให้คะแนนแบบคูมบ์ และ วิธีการเลือกชุดย่อยของคำตอบ (The Subset Selection Method, SST) ซึ่งทั้ง 2 วิธีจัดเป็นประเภทย่อยของประเภทการเปลี่ยนแปลงวิธีการตอบสนอง (Changing the Response Method) และจากผลการศึกษางานวิจัยที่เกี่ยวข้องที่ผ่านมาทั้งในประเทศและต่างประเทศพบว่า การศึกษาการให้คะแนนความรู้บางส่วนในประเทศ มีไม่มากนัก ในขณะที่ต่างประเทศมีการศึกษาไว้มาก และเป็นการวิเคราะห์ที่ใช้ทฤษฎีการวัดแบบดั้งเดิม จึงขอสรุปผลงานวิจัยบางเรื่องที่เกี่ยวข้องการเปลี่ยนแปลงวิธีการตอบสนองในช่วง 20 ปีที่ผ่านมา และสรุปงานวิจัยต่างประเทศที่เกี่ยวข้องทั้งหมดในช่วง 50 ปีที่ผ่านมา ในตารางที่ 2.1 สรุปผลการศึกษาในต่างประเทศ ดังมีรายละเอียด ดังนี้

สำราญ มีแจ้ง (2525: 55-56) ได้เสนอวิธีการตอบและการตรวจให้คะแนนแบบสอบชนิดเลือกตอบที่มี 5 ตัวเลือก โดยให้นักเรียนตอบได้ตั้งแต่ 1 ถึง 3 ตัวเลือกตามความมั่นใจ ถ้ามั่นใจตัวเลือกใดก็ตอบตัวเลือกนั้น ถ้าไม่มั่นใจก็ให้ตอบ 2 – 3 ตัวเลือก ถ้าตอบตัวเลือกเดียวแล้วถูกจะได้ 4 คะแนน ถ้าตอบ 2 ตัวเลือกแล้วมีตัวถูกจะได้ 2 คะแนน ถ้าตอบ 3 ตัวเลือกแล้วมีตัวถูกจะได้ 1 คะแนน ทุกกรณีถ้ามีการเลือกตัวเลือกที่ผิดจะได้ 0 คะแนน โดยได้เปรียบเทียบค่าความเที่ยง ความตรง และค่าอำนาจจำแนกของแบบสอบชนิดเลือกตอบที่มีการให้คะแนนแบบ 0-1 ตามวิธีของคูมบ์ ตามวิธีของอนันต์ ศรีโสภา และตามวิธีของสำราญ มีแจ้ง โดยใช้แบบสอบคณิตศาสตร์ 5 ตัวเลือก จำนวน 60 ข้อ กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 4 โรงเรียนสุรศักดิ์มนตรี และโรงเรียนวัดน้อยใน จำนวน 4 กลุ่มๆละ 60 คน และกลุ่มได้รับแบบสอบที่มีการตรวจให้คะแนนกลุ่มละ 1 วิธี ผลการวิจัยพบว่า ค่าความเที่ยงและความตรงของวิธีการให้คะแนนตามวิธีของสำราญ มีแจ้งสูงกว่าการให้คะแนนแบบ 0 – 1 ส่วนค่าอำนาจจำแนกของแบบสอบที่มีการให้คะแนนทั้ง 4 วิธีไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ

ทวี ทองคำ (2526: ง-จ) ได้เปรียบเทียบค่าความเที่ยง ความตรงและอำนาจจำแนกของแบบสอบชนิดเลือกตอบที่ใช้คำสั่งและวิธีการให้คะแนนที่แตกต่างกัน 3 วิธี ได้แก่ วิธีประเพณีนิยม (0-1) วิธีของสำราญ มีแจ้ง และวิธีของกิบบอนส์และคณะ โดยศึกษากลุ่มตัวอย่างที่เป็นนักเรียนชั้นประกาศนียบัตร

วิชาชีพปีที่ 1 ของวิทยาลัยเทคนิคยะลา จำนวน 180 คน ผลการวิจัยพบว่า ค่าความเที่ยง ความตรงตามสภาพและค่าอำนาจจำแนกของแบบสอบในทุกคำสั่งและวิธีการให้คะแนนไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ

Jaradat และ Swagad (1986) ได้ศึกษาเทคนิคการเลือกชุดตัวถูก(Subset Selection Technique) ของข้อสอบเลือกตอบ ในกลุ่มนักเรียนวิทยาศาสตร์ เกรด 9 โรงเรียนของรัฐ ประเทศจอร์แดน จำนวน 160 คน โดยเปรียบเทียบวิธีการให้คะแนน 3 วิธี คือ วิธีเลือกชุดตัวถูก (Subset Selection Technique , SST) วิธีแก้การเดา (Correction for Guessing Formula , CFG) และ วิธีประเพณีนิยม (Number Right , NR) ผลการวิจัยพบว่าวิธีเลือกชุดตัวถูกและวิธีแก้การเดา มีค่าความเที่ยง(KR - 20) สูงกว่าวิธีประเพณีนิยมอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ในขณะที่การเปรียบเทียบความตรงตามสภาพพบว่า วิธีเลือกชุดตัวถูก ให้ค่าความตรงตามสภาพสูงกว่าวิธีการแก้การเดา และวิธีประเพณีนิยมอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 นอกจากนี้การวิจัยในเรื่องนี้ ยังได้แบ่งกลุ่มตัวอย่างที่ศึกษา เป็น 4 กลุ่ม ที่ใช้ 2 เกณฑ์ในการจำแนก คือ เกณฑ์แบ่งกลุ่มตามความสามารถ แยกเป็น 2 ระดับได้แก่ ความสามารถสูง(high achievers) และ ความสามารถต่ำ(low achievers) และเกณฑ์ความเสี่ยงในการเลือกตอบ แยกเป็น 2 ระดับ ได้แก่ กลุ่มที่มีความเสี่ยงในการเลือกตอบสูง (high risk taker) และกลุ่มที่มีความเสี่ยงในการเลือกตอบต่ำ(low risk taker) แล้วนำมาเปรียบเทียบในเรื่อง ความตรงตามสภาพและความเที่ยง ระหว่าง 4 กลุ่มตามการให้คะแนนทั้ง 3 วิธี พบว่า วิธีเลือกชุดตัวถูกในแต่ละกลุ่ม ส่วนใหญ่มีแนวโน้มที่จะให้ค่าความตรงตามสภาพ และค่าความเที่ยงสูงกว่าวิธีการแก้การเดา และวิธีประเพณีนิยม

Tollefson และ Chung (1986) ได้เปรียบเทียบคะแนนเฉลี่ยและค่าความเที่ยง(Alpha reliability coefficient) ระหว่างวิธีการให้คะแนน 3 วิธีของข้อสอบแบบเลือกตอบ ได้แก่ วิธีประเพณีนิยม (NR) วิธีตัดตัวลวงของคูมบ์ (Elimination of incorrect)และเลือกชุดตัวถูก (inclusion of correct) โดยศึกษาในกลุ่มนักศึกษา 75 คน ในวิชาสถิติเบื้องต้น จำนวนข้อสอบเลือกตอบ 36 ข้อ 4 ตัวเลือก ผลการวิจัยพบว่าคะแนนเฉลี่ยของแต่ละวิธีไม่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ และค่าความเที่ยงของวิธีตัดตัวลวงของคูมบ์และวิธีเลือกชุดตัวถูก มีค่าความเที่ยงสูงกว่าแบบประเพณีนิยม แต่จากการทดสอบโดยใช้สถิติ Z-test เปรียบเทียบค่าความเที่ยงในแต่ละคู่ พบว่าไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ

Smith (1987) ได้ศึกษาและเปรียบเทียบความตรงตามสภาพและความเที่ยงโดยใช้สัมประสิทธิ์แอลฟ่า ระหว่างแบบสอบที่มีการให้คะแนนแบบทวิวิภาค (dichotomous, 0–1) และแบบสอบที่มีการให้คะแนนแบบพหุวิภาค(polychotomous)โดยเก็บข้อมูลจากนักศึกษาที่มาสอบความถนัดทางภาษา จำนวน 1,399 คนให้ทำแบบสอบภาษาที่ใช้โดยทั่วไปเพื่อนำไปเป็นเกณฑ์ในการหาความตรงตามสภาพกับแบบสอบภาษาที่ใช้ในการศึกษาครั้งนี้ ซึ่งแบ่งเป็น 2 ชุด ชุดแรกให้คะแนนแบบทวิวิภาค (dichotomous, 0–1) และชุดที่ 2 ให้คะแนนแบบพหุวิภาค (polychotomous) โดยที่แต่ละตัวเลือกมีคะแนนแต่ละข้อไม่เท่ากัน โดยมีหลักการให้คะแนนความรู้บางส่วนจากทฤษฎีของโครงสร้างความรู้ นอกจากนี้ยังได้แบ่งกลุ่มตัวอย่างเป็น 2 ชุด ชุดแรกจำนวน 400 คน ทดสอบตามปกติ ส่วนชุดที่ 2 จำนวน 999 คน ทดสอบความตรง

I20071425

ข้ามกลุ่มโดยแบ่งเป็นกลุ่มที่มีความสามารถสูงและกลุ่มที่มีความสามารถต่ำ ผลการวิจัยได้ข้อสรุปว่าการให้คะแนนแบบพหุวิภาค ค่าความตรงตามสภาพและความเที่ยงสูงกว่าการให้คะแนนแบบทววิภาคอย่างมีนียสำคัญทางสถิติและเมื่อมีการทดสอบกับอีกกลุ่มที่แบ่งเป็นกลุ่มที่มีความสามารถสูงและกลุ่มที่มีความสามารถต่ำ ก็ได้ข้อสรุปเป็นเช่นเดียวกัน

Jaradat, and Tollefson (1988) ได้ศึกษาคุณภาพของแบบสอบโดยใช้ทฤษฎีการวัดแบบดั้งเดิม ซึ่งเปรียบเทียบค่าความตรง ค่าความเที่ยง และการให้เกรดระหว่างวิธีการให้คะแนน 4 วิธีของข้อสอบแบบเลือกตอบ ได้แก่ วิธีประเพณีนิยม (FR) วิธีตัดตัวลวง (ET) วิธีเลือกชุดตัวถูก (SS) และวิธีการแก้การเดา (GC) โดยศึกษาในกลุ่มนักศึกษา 54 คน ในวิชาการวัดผลการศึกษา ข้อสอบเลือกตอบ มี 4 ตัวเลือก ใช้คะแนนสอบกลางภาค (MT) คะแนนจากวิธีประเพณีนิยม (FR) และคะแนนโครงงานในชั้นเรียนเป็นเกณฑ์ในการศึกษาความตรง ผลการวิจัยพบว่า ค่าความตรงและค่าความเที่ยงของ 4 วิธี ไม่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ส่วนการให้เกรดพบว่าวิธีตัดตัวลวงและวิธีเลือกชุดตัวถูก ทำให้ผู้สอบได้เกรด Aและ B สูงกว่าวิธีแก้การเดา

Simon, Budescu, and Nevo (1997) ได้เปรียบเทียบวิธีการตอบสนองเพื่อให้คะแนนความรู้บางส่วน 6 วิธีกับวิธีเลือกตอบแบบประเพณีนิยม เก็บข้อมูลกับนักเรียน 2,989 คนที่สอบคัดเลือกเข้ามหาวิทยาลัย ในอิสราเอล ระหว่างเดือน เมษายน 1985 – กรกฎาคม 1986 โดยใช้แบบสอบ 4 วิชา คือ วิชา General knowledge (GK) 30 ข้อ วิชา General Reasoning (GR) 22 ข้อ วิชา Figural Reasoning (FR) 18 ข้อ และ วิชา Mathematical Reasoning (MR) 21ข้อ แต่ละวิชามีวิธีการตอบ 7 วิธี ได้แก่ 1. วิธีแก้การเดา (Correction for guessing , CG) 2. วิธีการตัดตัวลวง (Elimination Scoring , ET) 3. วิธีการการทดสอบความน่าจะเป็น (Probability Testing , PT) 4. วิธีการแสดงความมั่นใจ (Confidence Marking , CM) 5. วิธีการจัดอันดับอย่างสมบูรณ์ (Complete Ordering , CO) 6. วิธีการจัดอันดับเป็นบางส่วน (Partial Ordering ,PO) และ 7. วิธีประเพณีนิยม

ผลการวิจัย พบว่า วิธีการทดสอบความน่าจะเป็น และ วิธีการแสดงความมั่นใจ มีความเที่ยงสูงกว่าวิธีอื่นๆ ในขณะเดียวกันและวิธีการทดสอบความน่าจะเป็น วิธีการแสดงความมั่นใจ และ วิธีการจัดอันดับเป็นบางส่วน มีความตรงสูงกว่าวิธีอื่นๆ นอกจากนี้ยังพบว่า วิธีการทดสอบความน่าจะเป็นและวิธีการแสดงความมั่นใจ สามารถจำแนกสัดส่วนของผู้ตอบที่มีความรู้บางส่วน และมีความรู้ผิดได้ดีกว่าวิธีการตัดตัวลวง และ วิธีการจัดอันดับเป็นบางส่วน

Bandaranayake,Payne,and White (1999) ได้ศึกษาวิธีการให้คะแนนข้อสอบแบบเลือกตอบโดยใช้แบบสอบของ Royal Australasian College ในเดือน มีนาคม กรกฎาคม และพฤศจิกายน ปี ค.ศ.1994 มีการเปรียบเทียบวิธีการให้คะแนนแบบเดิมที่เรียกว่า Type K (ให้ผู้สอบระบุตัวเลือกทีละตัวเลือก รวม 4 ตัว ว่าถูกหรือผิด แล้วจึงเลือกตัวเลือกที่กำหนดให้ที่มีการผสมผสานกันระหว่าง 4 ตัว ว่าตัวเลือกในข้อใดถูก ซึ่งมี 5 ตัวเลือก) กับวิธีการให้คะแนนแบบ Type X ซึ่งมีสูตรการให้คะแนนแยกย่อย 6 วิธี และพบว่าวิธีการให้คะแนนย่อย 2 ชนิด คือ Middlesex (ให้เลือกตอบตัวเลือกทีละตัวเลือกว่าถูกหรือ

ผิด โดยตัวเลือกแต่ละตัวเป็นอิสระจากกัน) และชนิด +1,-1และ0 (ให้ 1 คะแนนเมื่อตอบถูก –1เมื่อตอบผิด และ 0 คะแนนเมื่อไม่ตอบ) ของการให้คะแนนแบบ Type X มีคะแนนเฉลี่ยใกล้เคียงกับ Type K จึงนำ 2 วิธีนี้มาศึกษาเปรียบเทียบผลกับ Type K ผลการวิจัยพบว่า วิธีการให้คะแนนย่อย ชนิด Middlesex และชนิด +1,-1และ0ให้คะแนนเฉลี่ยต่ำกว่าแบบ Type K ทั้งนี้อาจเนื่องจากแบบ Type K มีโอกาสในการเดา โดยไม่โดนตัดคะแนน และพบว่าวิธีการให้คะแนนย่อย ชนิด Middlesex มีความแตกต่างของคะแนนเฉลี่ยในปัจจุบันและในอดีตสูงกว่าวิธีอื่นๆ จึงทำให้คณะกรรมการ Royal Australasian College of Surgeons (RACS) เลือกใช้วิธีการให้คะแนนย่อย ชนิด Middlesex แทนวิธีการให้คะแนนแบบType K

ตารางที่ 2.1 สรุปผลการศึกษาในต่างประเทศเพื่อเปรียบเทียบค่าความตรงและค่าความเที่ยง ระหว่างวิธีการ ให้ความรู้บางส่วนกับวิธีประเพณีนิยม (NC)

แหล่งอ้างอิง N=จำนวนตัวอย่าง	เนื้อหาของแบบสอบ n = จำนวนข้อสอบ	ดัชนี*ที่ใช้เปรียบเทียบ	วิธีการให้คะแนนความรู้บางส่วน** และค่าที่ได้		วิธีการให้คะแนนแบบประเพณีนิยมและค่าที่ได้
Dressel & Schmidt (1953) College Students, N = 90	Physics, n = 44	KR20	CM	.73	.70
			SS	.67	.70
		PV	CM	.65	.62
			SS	.58	.62
Coombs et al. (1956) College Students, N = 285	Vocabulary, n = 40	KR20	ET	.73	.72
	Driver Information, n = 40	KR20	ET	.70	.64
	Object Aperture, n = 40	KR20	ET	.91	.89
Michael (1968) High Schoolers, N = 22	Social Science, n = 35	SH	PT	.84	.76
Hambleton et al. (1970) Teachers, N = 73	Measurement, n = 40	SH	PT	.66	.71
		PV	PT	.72	.62
Rippey (1970) High Schoolers, N = 22	Writing, n = 30	Hoyt	PT	.70	.53
Collet (1971) College Students, N = 47	Mental Maturity, n = 50	TR	ET	.86	.81 (GC)
		PV	ET	.77	67 (GC)
Koehler (1971) High Schoolers, N ≈ 158	Vocabulary, n = 20	α	PT	.65	.63
	Social Science, n =20	α	PT	.52	.51
	Science, n = 20	α	PT	.58	.56
Hopkins et al. (1973) Graduate Student, N = 63	Intro. To Statistics, n = 65	SB	CM	.92	.88
		PV	CM	.67	.70

* TR = test-retest reliability, SB = spearman-Brown reliability, SH = split-half reliability, α = Cronbach α reliability, PV = predictive validity, CV = concurrent validity.

**PT = Probability Testing
ET = Elimination Testing
PO = Partial Ordering
CM = Confidence Marking
SS = Subset Selection Method
CO = Complete Ordering

ตารางที่2.1 (ต่อ)

แหล่งอ้างอิง N=จำนวนตัวอย่าง	เนื้อหาของแบบสอบ n = จำนวนข้อสอบ	ดัชนี*ที่ใช้เปรียบเทียบ	วิธีการให้คะแนนความรู้บางส่วน** และค่าที่ได้		วิธีการให้คะแนนแบบประเพณีนิยมและค่าที่ได้
Diamond (1975)	Intro. To Measurement,	Flanagan	PO	.46	.34
Graduate Students, N = 84	n = 34	Reliability	PO	.57	.53
			PO	.49	.49
			PO	.70	.67
		PV	PO	.04	.09
			PO	.44	.40
			PO	.61	.64
			PO	.37	.36
Pugh and Brunza (1975)					
Graduate Students, N = 84	Vocabulary, n = 24	Hoyt	PT	.85	.57
Hakstian and Kansup (1975)	Vocabulary-1, n = 15	α	ET	.65	.63
High Schoolers, N ≈ 430			PT	.74	
		PV	ET	.67	.59
			PT	.65	
	Vocabulary-2, n = 30	α	ET	.69	.67
			PT	.72	
	Mathematical Reasoning-1,	α	ET	.56	.60
	n = 15		PT	.69	
		PV	ET	.38	.50
			PT	.49	
	Mathematical Reasoning-2,	α	ET	.64	.64
	n = 15		PT	.69	
			ET	.67	.68
	Vocabulary	TR	PT	.76	
			ET	.60	.61
	Mathematical Reasoning	TR	PT	.67	
Poizner et al. (1978)					
College Students, N = 169	Three Rare Topics, n = 8	SB	CO	.66	.56
			PT	.68	.56
College Students, N = 63	Vocabulary, n = 8	Guttman	CO	.46	.34
			PT	.56	.34
Jaradat and Seagad (1986)					
High Schoolers, N = 160	Chemistry, n = 20	KR20	SS	.82	.72
					.85 (GC)
		CV	SS	.64	45
					.39 (GC)
Tollefson and Chung (1986)					
Graduate Students, N = 75	Statistic ,n=36	α	ET	.67	
Jaradat & Tollefson (1988)		KR20	SS	.71	.57 (GC)
Graduate Students, N = 54	Measurement, n = 25		ET	.67	.76
		CV	SS	.57	.63
			SS	.52	.40
			SS	.42	.32
			ET	.76	.63
			ET	.54	.40
			ET	.40	.32

ตารางที่2.1 (ต่อ)

แหล่งอ้างอิง N=จำนวนตัวอย่าง	เนื้อหาของแบบสอบ n = จำนวนข้อสอบ	ดัชนี*ที่ใช้เปรียบเทียบ	วิธีการให้คะแนนความรู้บางส่วน** และค่าที่ได้		วิธีการให้คะแนนแบบประเพณีนิยมและค่าที่ได้
Simon,Budscu, and Nevo(1997). Candidates Students ,N=2,989					
	General knowledge,n=30	α	ET	.70	.85(NC)
			PT	.75	.88(CG)
			CM	.82	
			CO	.70	
			PO	.75	
		CV	ET	.83	85(NC)
			PT	.81	.88(CG)
			CM	.85	
			CO	.75	
			PO	.85	
	General Reasoning ,n= 22	α	ET	.62	.63(NC)
			PT	.76	.69(CG)
			CM	.65	
			CO	.56	
			PO	.68	
		CV	ET	.60	.63(NC)
			PT	.74	.69(CG)
			CM	.73	
			CO	.46	
			PO	.73	
	Figural Reasoning ,n=18	α	ET	.63	.65(NC)
			PT	.65	.49(CG)
			CM	.72	
			CO	.70	
			PO	.55	
		CV	ET	.67	.67(NC)
			PT	.61	.56(CG)
			CM	.32	
			CO	.66	
			PO	.44	
	Mathematical Reasoning n= 21	α	ET	.80	.77(NC)
			PT	.82	.74(CG)
			CM	.82	
			CO	.70	
			PO	.76	
		CV	ET	.82	.77(NC)
			PT	.84	.77(CG)
			CM	.84	
			CO	.76	
			PO	.83	

* TR = test-retest reliability, SB = spearman-Brown reliability, SH = split-half reliability, α = Cronbach α reliability, PV = predictive validity, CV = concurrent validity.

**PT = Probability Testing
ET = Elimination Testing
PO = Partial Ordering
CM = Confidence Marking
SS = Subset Selection Method
CO = Complete Ordering

จากการทบทวนงานวิจัยที่ผ่านมา จะเห็นว่ามีการศึกษาวิธีการให้คะแนนความรู้บางส่วน มาเป็นระยะเวลา เกือบ 50 ปี แต่ยังได้ข้อสรุปที่ยังไม่ชัดเจน และยังมีการศึกษาเพื่อพัฒนาวิธีการให้คะแนนความรู้บางส่วนที่แตกต่างกันอีกหลายวิธี เช่น ใช้ทฤษฎี Finite State Score ในการให้คะแนนข้อสอบแบบเลือกตอบ (Ndalichako and Rogers, 1997) ใช้การจัดอันดับความน่าจะเป็นโดยใช้สถิติเข้ามาช่วย (Burton and Miller ,1999) ทั้งนี้เพื่อให้ได้ข้อสรุปที่ดีและมีความถูกต้องชัดเจน อันจะเป็นประโยชน์ในด้านของการนำไปใช้ ซึ่งในปัจจุบันยังไม่มีการนำวิธีการให้คะแนนความรู้บางส่วนไปใช้เท่าที่ควร จึงเป็นประเด็นที่น่าสนใจที่จะศึกษาต่อไป

ตอนที่ 4 โปรแกรมคอมพิวเตอร์ที่ใช้ในการวิเคราะห์

โปรแกรมคอมพิวเตอร์ที่ใช้ในการวิเคราะห์ในการวิจัยครั้งนี้ ใช้ 2 โมเดล คือโมเดลการตอบสนองข้อสอบแบบทวิภาค (Dichotomous Item Response Models) และแบบพหุวิภาค (polytomous) โดยที่โมเดลการตอบสนองข้อสอบแบบทวิภาคใช้โปรแกรมคอมพิวเตอร์ BILOG ส่วนโมเดลการตอบสนองข้อสอบแบบพหุวิภาคใช้โปรแกรมคอมพิวเตอร์ PARSCALE

การวิเคราะห์แบบทวิภาคตามแนวคิดของทฤษฎีการตอบสนองข้อสอบ ใช้โปรแกรม BILOG

การวิเคราะห์แบบทวิภาคตามแนวคิดของทฤษฎีการตอบสนองข้อสอบ ใช้โปรแกรม BILOG เป็นการวิเคราะห์เพื่อการประมาณค่าที่คงที่และแม่นยำทั้งข้อสอบแบบสั้นและแบบยาว โดยมีวิธีประมาณค่าให้เลือก 2 แบบ คือMarginal likelihood estimation (MMLE) และ Marginal Bayesian estimation(MBE) การประมาณค่าพารามิเตอร์ใช้การทำซ้ำด้วยวิธี EM และตามด้วยวิธี Newton ซึ่งผู้ใช้สามารถกำหนดจำนวนรอบของการประมาณค่าซ้ำเองได้ การวิเคราะห์มี 3 ขั้นตอน ดังนี้

ขั้นตอนที่ 1 เป็นขั้นตอนที่ใช้โปรแกรมย่อย ชื่อ INPUT ซึ่งเป็นคำสั่งให้โปรแกรมอ่านข้อมูลเข้า โดยคำสั่งในส่วนนี้จะบอกรายละเอียดเกี่ยวกับข้อมูลที่จะวิเคราะห์ โปรแกรมย่อย INPUT จะคำนวณค่าสถิติรายข้อให้ ได้แก่ จำนวนผู้ตอบถูกในแต่ละข้อ ร้อยละของการตอบถูกในแต่ละข้อ และค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนในแต่ละข้อกับคะแนนของข้อสอบทั้งหมด

ขั้นตอนที่ 2 เป็นขั้นตอนที่ใช้โปรแกรมย่อย ชื่อ CALIBRATE ซึ่งเป็นคำสั่งใช้ประมาณค่าพารามิเตอร์และทดสอบความเหมาะสมของโมเดล ผลที่ได้จากขั้นตอนนี้ คือ ค่าพารามิเตอร์ของข้อสอบ เช่น ค่าความยาก ค่าอำนาจจำแนก และผลการทดสอบความเหมาะสมของข้อมูลโดยใช้ดัชนี χ^2

ขั้นตอนที่3 เป็นขั้นตอนที่ใช้โปรแกรมย่อย ชื่อ SCORE ซึ่งเป็นคำสั่งที่ใช้ข้อมูลจากขั้นตอนที่1 และ 2 มาประกอบเพื่อนำมาประมาณค่าคะแนนความสามารถของผู้ตอบรายบุคคล โปรแกรมย่อย SCOREจะประมาณค่าคะแนนความสามารถ ประมาณค่าพารามิเตอร์ของข้อสอบซ้ำ และให้ค่าฟังก์ชั่นสารสนเทศของข้อสอบและแบบสอบ ตลอดจนโค้งสารสนเทศของข้อสอบและแบบสอบด้วย

การใช้โปรแกรม BILOG เพื่อการวิเคราะห์ข้อสอบ

การดำเนินการวิเคราะห์ข้อมูล มีขั้นตอนที่สำคัญดังนี้

1. การเตรียมข้อมูล หมายถึงไฟล์ข้อมูลที่พร้อมจะนำมาวิเคราะห์
2. การเขียนคำสั่งสำหรับการวิเคราะห์
3. การใช้คำสั่งให้โปรแกรมทำงาน
4. การพิมพ์ผลการวิเคราะห์
5. การแปลความหมายผลการวิเคราะห์

โดยจะขอกล่าวถึงรายละเอียดที่สำคัญของการวิเคราะห์ คือ การเตรียมข้อมูลและการเขียนคำสั่งสำหรับการวิเคราะห์ ดังนี้ (Mislevy และBock ,1990)

1. การเตรียมข้อมูล

การเตรียมข้อมูลเพื่อการวิเคราะห์ข้อสอบตามแนว IRT เป็นขั้นตอนสำคัญขั้นตอนหนึ่ง หากการสร้างไฟล์ข้อมูลมีรูปแบบไม่เหมาะสมกับโปรแกรมที่จะวิเคราะห์ อาจทำให้เกิดปัญหาในขณะโปรแกรมทำงาน (RUN) ได้ การเตรียมข้อมูลหรือสร้างไฟล์ข้อมูล สามารถสร้างได้จากโปรแกรม Editor ทั่วไป เช่น DOS-edit หรือ Q-edit แต่ไฟล์ข้อมูลจะกำหนดนามสกุล .DAT หรือไม่ก็ได้

การสร้างไฟล์ข้อมูลโดยทั่วไป จะประกอบด้วยส่วนสำคัญ 3 ส่วน คือ

1. บรรทัดแรก จะเป็น key ของข้อสอบ คือคำตอบที่ถูกของแต่ละข้อ
2. บรรทัดที่ 2 จะเป็น Omit key คือกำหนดเลขที่ใช้สำหรับผู้ที่ไม่ตอบโดยส่วนใหญ่จะใช้เลข 9
3. บรรทัดที่3 จะเป็นข้อมูลในแต่ละระเบียน ประกอบด้วย เลขประจำตัวของผู้สอบ (ID, Identification field) และผลการตอบ (Response)

ตัวอย่าง ข้อมูลที่ได้จากข้อสอบ 10 ข้อ ผู้ตอบ 4 คน ข้อสอบแต่ละข้อกำหนดให้คะแนนแบบทวิวิภาคเป็น 2 ลำดับขั้นคะแนน คือ 0 และ 1

```
1111111111
9999999999
001 1011110000
002 1111100011
003 1011000111
004 0011100111
```

2. การเขียนคำสั่งสำหรับการวิเคราะห์

การเขียนคำสั่งสำหรับการวิเคราะห์ สามารถสร้างได้จากโปรแกรม Editor เช่นเดียวกับการสร้างไฟล์ข้อมูล แต่ไฟล์คำสั่งต้องตั้งชื่อไฟล์เป็น file.blg เสมอ ไฟล์คำสั่งนี้อาจสร้างไว้ใน directory ของโปรแกรม หรือแยกต่างหากก็ได้ แต่ละคำสั่งในไฟล์นี้จะมีรูปแบบทั่วไปดังนี้

>TITLE : เป็นคำสั่งเพื่อใช้แสดงผลการวิเคราะห์ โดยให้พิมพ์เป็นข้อความ

>COMMENT : ใช้อธิบาย หรือข้อสังเกตสำหรับดูในผลการวิเคราะห์ โดยทั่วไปมักอธิบายที่มาของข้อมูล เกณฑ์การให้คะแนน ปีที่เก็บข้อมูล และกลุ่มตัวอย่าง เป็นต้น

>GLOBAL ใช้เพื่อกำหนดชื่อ INPUT FILE และ ข้อมูลอื่นที่ใช้ใน3 ขั้นตอน มีรูปแบบดังนี้

>GLOBAL NTEsts=a, NPArm=b,NWGht=c,LOGistic,DFName= d,
MFNane = e, IFName =f, KFName = g, NFName =h, OFName =i,SAVe;

NTEsts = จำนวนของแบบสอบย่อย

NPArm = จำนวนพารามิเตอร์

NWGht = น้ำหนักของค่าที่ตอบสนอง

LOGistic = ฟังก์ชันการตอบสนองแบบโลจิส

DFName = ชื่อไฟล์ข้อมูลดิบ

MFNane = ชื่อไฟล์ข้อมูลหลัก (Master file)

IFName = ชื่อไฟล์พารามิเตอร์ข้อสอบ (item – parameter)

KFName = ชื่อไฟล์เฉลย

NFName = ชื่อไฟล์ที่ไม่มีเฉลย (Key)

OFName = ชื่อไฟล์ omit - key

>SAVE : ใช้เพื่อเลือกบันทึกไฟล์ Out-put. มีรูปแบบดังนี้

> SAVE MASter = a, CALib = b , PARm = c, SCOre = d, COVarance=e,
Expected=f, SORted=g, TSTat=h, POSt=i;

MASter = ชื่อไฟล์หลัก

CALib = ชื่อไฟล์ Calibration

PARm = ชื่อไฟล์พารามิเตอร์ของข้อสอบ

SCOre = ชื่อไฟล์คะแนนนักเรียน

Expected = ชื่อไฟล์ของความถี่ที่คาดหวัง

SORted = ชื่อชุดข้อมูลที่แบ่งเป็นกลุ่ม

TSTat = ตารางสารสนเทศของแบบสอบ

POSt = น้ำหนักและความน่าจะเป็นของข้อมูลที่สังเกตได้

>LENgth : ใช้เพื่อกำหนดจำนวนข้อของชุดข้อสอบ มีรูปแบบดังนี้

>LENgth NITems = (n_1 , n_2 , …. , n_{NTEsts}) ;

NITems = จำนวนข้อสอบ

INPUT : ใช้อธิบายข้อมูลที่จะนำมาวิเคราะห์ และสารสนเทศอื่น ๆ มีรูปแบบดังนี้

> INPut NTOtal = a ,NFMt = b, SAMple = c, NALt = d, INOpt = e, NIDch = f,

TAKe = g, FORms= h;

NTOtal = จำนวนข้อสอบ

NFMt = จำนวนแถวของข้อมูลการตอบของผู้สอบแต่ละคน

SAMple = ร้อยละ (0 – 100) ของผู้ตอบที่จะสุ่มมาวิเคราะห์

NALt = จำนวนสูงสุดของค่าตัวเลือกในชุดข้อมูลดิบ

INOpt = ประเภทของข้อมูลดิบ

NIDch = จำนวนตัวอักษรในฟิลด์ ID

TAKE = จำนวนที่จะนำผลการตอบ จากคนแรกถึงผู้ตอบ g มาวิเคราะห์

FORms = ตัวบ่งชี้รูปแบบที่หลากหลาย

>TEST : ใช้บอกชื่อแบบสอบ หรือแบบสอบย่อย มีรูปแบบดังนี้

> TEST TNAme = a, ITEms = (b_1, b_2, …, b_n),

INAme = (c_1, c_2, …, c_n), INTercept = (d_1, d_2, … ,d_n),

SLOpe = (e_1, e_2, … ,e_n),THReshld = (f_1, f_2, …, f_n)

GUEss = (g_1, g_2, … ,g_n), LINk = (h_1, h_2, …, h_{ntotal}) ;

TNAme = ชื่อชุดของแบบสอบ หรือมาตรวัด

ITEms = เรียงลำดับข้อสอบในข้อมูลที่พิมพ์เพื่อการวิเคราะห์

INAme = ชื่อของข้อสอบในแบบสอบ หรือชุดของแบบสอบย่อย

INTercept = คือค่ากำหนดเริ่มต้นของ Intercept อาจกำหนดที่ THRESHOLD ก็ได้ แต่ต้องกำหนดเพียงที่ใดที่หนึ่งเท่านั้น

THReshole = คือค่ากำหนดเริ่มต้นของค่า Threshold ข้อกระทง ซึ่งอาจกำหนดจาก INTERCEPT ก็ได้ ซึ่งทั้งสองแห่งถ้าไม่กำหนดโปรแกรมก็จะกำหนดให้เอง

GUEss = คือค่าเริ่มต้นของค่าการเดา (เฉพาะ 3 พารามิเตอร์เท่านั้น)

LINk = ข้อที่เป็นตัวร่วมเพื่อเชื่อมโยงรูปแบบข้อสอบที่หลายรูปแบบ

>CALIB : ใช้บังคับในการประมาณค่าพารามิเตอร์ของข้อสอบ และกำหนดการแจกแจงเริ่มแรกในการประมาณค่าพารามิเตอร์แบบสอบมีรูปแบบดังนี้

CALib SELect = (a_1, a_2,… ,a_{NTEsts}), NQPT=(b_1, b_2,…,b_{NTEsts}), CYCles =c, NEWton=d,

PRInt=e, CRIt=f, COMmon, IDIstp=g, PLOT=h, CASe=i,SORt,FREe, TPRior,

SPRior, GPRior, NOSprior, NOGprior, REAdpri, FLOat, CHIsqe=j,

RIDge=(k,I,m),NFUII=p, POSt;

SELect = การเลือกชุดข้อสอบย่อยตามที่ได้แบ่งไว้

NQPT = จำนวน Quadrature Points ในการประมาณค่าแบบ MMLของแต่ละชุดข้อมูล

CYCles = จำนวนรอบสูงสุดของ EM

NEWton = จำนวนรอบสูงสุดของNewton

PRInt = แสดงค่าพารามิเตอร์ข้อสอบในแต่ละรอบ

CRIt = Convergence Criterion ของ การประมาณค่าของEMและNewton

COMmon = ประมาณค่าการเดาที่ต่ำที่สุดของข้อสอบทั้งหมด

IDIstp = กำหนดประเภทของการแจกแจงเริ่มแรกสำหรับค่าความสามารถของกลุ่มประชากร

PLOT = ค่าระดับนัยสำคัญสำหรับgoodness of fit ของโครงสร้างข้อสอบ

CASe = เลือกข้อมูล

SORt = แยกกลุ่มข้อมูล

FREe = การแจกแจงคะแนนของประชากรถูกประมาณในลักษณะการแจกแจงแบบไม่ต่อเนื่อง

TPRior = มีการแจกแจงเริ่มต้นของค่าความยาก

SPRior = มีการแจกแจงเริ่มต้นของค่าอำนาจจำแนก

GPRior = มีการแจกแจงเริ่มต้นของค่าการเดา

NOSprior = ไม่มีการแจกแจงแรกเริ่มต้นของค่าความยาก

NOGprior = ไม่มีการแจกแจงเริ่มต้นของค่าการเดา

REAdpri = มีการแจกแจงเริ่มต้นของค่าพารามิเตอร์ของข้อสอบโดยใช้คำสั่ง

FLOat = ค่าเฉลี่ยของการแจกแจงเริ่มต้นประมาณด้วย marginal maximum likelihood

CHIsqe = จำนวนข้อขั้นต่ำสำหรับการคำนวณ chi -square

NFUII = การเลือกใช้ informatiom matrix

POSt = พิมพ์ค่า posterior deviates

SCORE : เป็นคำสั่งเพื่อคำนวณคะแนนเริ่มแรกของผู้ตอบแต่ละคน หรือแต่ละกลุ่ม มีรูปแบบ (Format) ดังนี้

SCOre METhod = a , BIWeight , FIT, NOPrint , NQPT = (b_1, b_2,…,b_{NTEsts}),

IDISt = c, PMN = (d_1, d_2,…,d_{NTEsts}), PSD = (e_1, e_2,…,e_{NTEsts}), RSCtype = f,

LOCation = (g_1, g_2,…,g_{NTEsts}), SCALe = (h_1, h_2,…,h_{NTEsts}),I NFo=I,YCOmmon,POP;

METhod = วิธีประมาณค่าคะแนน

BIWeight = ประมาณค่า robust to isolated deviant

FIT = คำนวณความสอดคล้องของแต่ละรูปแบบโดยใช้ likelihood ratio chi-square

NOPrint = ไม่พิมพ์คะแนนในphase 3

NQPT = จำนวน quadrature points

IDISt = รูปแบบของการแจกแจงเริ่มต้นของคะแนน

PMN = ค่าเฉลี่ยที่แท้จริงของการแจกแจงปกติสำหรับแต่ละแบบสอบย่อย

PSD = ส่วนเบี่ยงเบนมาตรฐานที่แท้จริงของการแจกแจงปกติสำหรับแต่ละแบบสอบย่อย

RSCtype = ประเภทของการปรับสเกลที่ต้องการ

LOCation = ค่าคงที่ของตำแหน่งที่แท้จริงในการปรับสเกล

SCALe = ค่าคงที่ของมาตรที่แท้จริงในการปรับสเกลคะแนน

INFo = เลือก out put สารสนเทศของแบบสอบ

YCOmmon = ในกรณีที่ INFo >0 ทำให้โค้งสารสนเทศของแบบสอบสามารถเปรียบเทียบในหน่วยพื้นที่เดียวกันได้

POP = ในกรณีที่ INFo >0 จะคำนวณค่าสารสนเทศที่คาดหวังรวมทั้งประมาณค่าความเที่ยงด้วย

การวิเคราะห์แบบพหุวิภาคตามแนวคิดของทฤษฎีการตอบสนองข้อสอบ โดยใช้โปรแกรม PARSCALE

การวิเคราะห์แบบพหุวิภาคตามแนวคิดของทฤษฎีการตอบสนองข้อสอบโดยใช้โปรแกรม PARSCALE เป็นการวิเคราะห์เพื่อประมาณค่าพารามิเตอร์ของข้อสอบและของผู้ตอบ จากการใช้แบบสอบที่ตรวจให้คะแนนแบบหลายค่า ที่พัฒนาโดย มูราคิ และบอค (Muraki and Bock) ในปี ค.ศ. 1994 โปรแกรมนี้สามารถใช้วิเคราะห์ตาม Graded Response Model (GRM), Partial Credit Model (GRM)

และ Generalized Partial Credit Model (GPCM) การประมาณค่าพารามิเตอร์ใช้วิธี Marginal Maximum Likelihood (MML) ผลการวิเคราะห์ให้ค่าพารามิเตอร์ อำนาจจำแนก (a) ค่าพารามิเตอร์ความยากในแต่ละลำดับขั้นคะแนน (bjk) ค่าฟังก์ชันสารสนเทศรายข้อ (IIF) แต่การให้ค่าฟังก์ชันสารสนเทศเสนอเฉพาะโมเดล PCM และ GPCM เท่านั้น

การใช้โปรแกรม PARSCALE เพื่อการวิเคราะห์ข้อสอบ

ลักษณะทั่วไปของโปรแกรม

โปรแกรม PARSCALE พัฒนาขึ้นมาโดย มูราคิ และบอค (Murake and Bock, 1991) โดยนำมาใช้วิเคราะห์ข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาค ตามแนวทฤษฎีการตอบสนองข้อสอบ (IRT) โปรแกรมนี้สามารถวิเคราะห์ได้ตามโมเดล GRM, PCM และ GPCM ซึ่งมีลักษณะทั่วไปดังนี้(ธนวัฒน์ แสนสุข, 2539 ; Muraki and Bock 1993)

1. ระบบคอมพิวเตอร์ที่สามารถใช้ติดตั้งโปรแกรม

โปรแกรมนี้สามารถทำงาน (RUN) ได้กับเครื่องคอมพิวเตอร์ตั้งแต่รุ่น Intel 386 หรือเทียบเท่าขึ้นไป โดยใช้ PC-DOS หรือ MS-DOS ตั้งแต่ Version 3.0 ขึ้นไป มี Hard Drive และ Diskette Drive และเครื่องที่ใช้ต้องมีหน่วยความจำ (RAM) อย่างน้อย 2 Mb

2. ขนาดของข้อมูลที่สามารถวิเคราะห์ได้

จำนวนข้อสอบและผู้สอบที่สามารถวิเคราะห์มากที่สุดของโปรแกรมนี้ขึ้นอยู่กับหน่วยความจำของเครื่องเป็นสำคัญ จากการทดลองใช้วิเคราะห์ข้อสอบ 30 ข้อ และผู้ตอบ 6,300 คน ด้วยเครื่อง IBM PC รุ่น 486-dx2 (RAM = 4 Mb) สามารถวิเคราะห์ได้โดยไม่มีปัญหา

3. โมเดลที่สามารถวิเคราะห์ได้

โมเดลที่โปรแกรมนี้สามารถวิเคราะห์ได้คือ Graded Response Model (GRM), Partial Credit Model (PCM) และ Generalized Partial Credit Model (GPCM)

4. การดำเนินการวิเคราะห์ข้อมูล

การดำเนินการวิเคราะห์ข้อมูล มีขั้นตอนที่สำคัญดังนี้

4.1 การเตรียมข้อมูล หมายถึงไฟล์ข้อมูลที่พร้อมจะนำมาวิเคราะห์

4.2 การเขียนคำสั่งสำหรับการวิเคราะห์

4.3 การใช้คำสั่งให้โปรแกรมทำงาน

4.4 การพิมพ์ผลการวิเคราะห์

4.5 การแปลความหมายผลการวิเคราะห์

โดยจะขอกล่าวถึงรายละเอียดที่สำคัญของการวิเคราะห์ คือ การเตรียมข้อมูลและการเขียนคำสั่งสำหรับการวิเคราะห์ ดังนี้

1. การเตรียมข้อมูล

การเตรียมข้อมูลเพื่อการวิเคราะห์ข้อสอบตามแนว Polytomous IRT เป็นขั้นตอนสำคัญขั้นตอนหนึ่ง หากการสร้างไฟล์ข้อมูลมีรูปแบบไม่เหมาะสมกับโปรแกรมที่จะวิเคราะห์ อาจทำให้เกิดปัญหาในขณะโปรแกรมทำงาน (RUN) ได้ การเตรียมข้อมูลหรือสร้างไฟล์ข้อมูล สามารถสร้างได้จากโปรแกรม Editor ทั่วไป เช่น DOS-edit หรือ Q-edit แต่ไฟล์ข้อมูลจะกำหนดนามสกุล .DAT หรือไม่ก็ได้ ตัวอย่างรูปแบบ (format) ของข้อมูลที่จะนำมาวิเคราะห์เป็นดังนี้

ตัวอย่างที่ 1 ข้อมูลที่ได้จากข้อสอบ 10 ข้อ ผู้ตอบ 4 คน ข้อสอบแต่ละข้อกำหนดให้คะแนนแบบพหุวิภาคเป็น 4 ลำดับขั้นคะแนน

001 4342343421

002 3212213212

003 4342123434

004 4232432423

รูปแบบ (format) ของข้อมูลคือ (3A1, 2X, 10A1)

การสร้างไฟล์ข้อมูลโดยทั่วไป จะประกอบด้วยส่วนสำคัญ 2 ส่วน คือ

1. ID (Identification field)
2. Response (ผลการตอบ)

สำหรับโปรแกรม PARSCALE ไม่ควรกำหนดฟิลด์ ID ให้เกิน 30 ตัวอักษร ส่วนจำนวนข้อสอบสามารถวิเคราะห์ได้อย่างไม่จำกัด โดยในแต่ละข้อจะต้องกำหนดให้คะแนนไม่เกิน 15 ลำดับขั้นคะแนน (category) ตัวอย่างรูปแบบข้อมูลที่วิเคราะห์ข้อสอบ 100 ข้อ แบ่งเป็นชุดย่อย 16 ชุด ชุดละ 10 ข้อ เป็นดังนี้

> INPUT NID = 4, NTOT = 160, LEN = (10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10), NTEST = 16, NFMT = 1; (4A1, 5X, 60A1, / 80A1, / 20A1)

การพิมพ์ข้อมูลจะพิมพ์เป็น 3 บรรทัด คือ

0001 A1A2A3...A60

A1A2A3...A80

A1A2A3...A20

0002 B1B2B3...B60

0003 B1B2B3...B80

0004 B1B2B3...B20 เป็นต้น

2. การเขียนคำสั่งสำหรับการวิเคราะห์

การเขียนคำสั่งสำหรับการวิเคราะห์ สามารถสร้างได้จากโปรแกรม Editor เช่นเดียวกับการสร้างไฟล์ข้อมูล แต่ไฟล์คำสั่งต้องตั้งชื่อไฟล์เป็น file.PSL เสมอ ไฟล์คำสั่งนี้อาจสร้างไว้ใน directory ของโปรแกรม หรือแยกต่างหากก็ได้ แต่ละคำสั่งในไฟล์นี้จะมีรูปแบบทั่วไปดังนี้

> NAME KEYWORD1 = a1, KEYWORD2 = (b1, b2, ..., bn), ..., OPTION m, OPTIONn;

คำสั่งสำหรับ PARSCALE (เรียงตามลำดับ)

คำสั่ง	จำเป็น (*)
TITLE	*
COMMENT	
FILES	*
SAVE	
INPUT	*
(Data format)	*
TEST	*
BLOCK	*
CALIB	*
QUADP	
PRIORS	
SCORE	*
QUADS	

TITLE : เป็นคำสั่งเพื่อใช้แสดงผลการวิเคราะห์ โดยให้พิมพ์เป็นข้อความได้ 2 บรรทัด ไม่ต้องมีเครื่องหมาย > นำหน้า

ตัวอย่างเช่น

Thanawat Assessment Project' 90

Chulalongkorn University.

> COMMENT GROUP – LEVEL ANALYSIS.

COMMENT : (optional) : ใช้อธิบาย หรือข้อสังเกตสำหรับดูในผลการวิเคราะห์ โดยทั่วไปมักอธิบายที่มาของข้อมูล เกณฑ์การให้คะแนน ปีที่เก็บข้อมูล และกลุ่มตัวอย่าง เป็นต้น

ตัวอย่างเช่น

Thanawat Assessment Project' 90

Chulalongkorn University.

> COMMENT

Data for this example are from the study described by Doran, et.al., in the April, 1996, Science Teacher.

> FILE DFNAME = 'EXAM>DAT', SAVE;

FILES : ใช้เพื่อกำหนดชื่อ INPUT FILE มีรูปแบบดังนี้

> FILes DFName = a, MFNane = b. CFName = c, IFName = d,

OFName = e, NFName = f, SAVE;

DFName = ชื่อไฟล์ข้อมูลดิบ

MFName = ชื่อไฟล์ข้อมูลหลัก (Master file)

CFName = ชื่อไฟล์ Calibration

IFName = ชื่อไฟล์พารามิเตอร์ข้อสอบ (item – parameter)

OFName = ชื่อไฟล์ omit - key

NFName = ชื่อไฟล์ที่ไม่มีเฉลย (Key)

* หมายเหตุ ชื่อไฟล์ควรตั้งไม่เกิน 32 ตัวอักษร

SAVE : ใช้เพื่อเลือกบันทึกไฟล์ Out-put. มีรูปแบบดังนี้

> SAVE MASter = a, CALib = b , PARm = c, SCOre = d,

INFormation = c, COMbine = f;

MASter = บันทึก Master.data file (เป็นชื่อไฟล์)

CALB = ชื่อไฟล์ Calibration

PARM = ชื่อไฟล์พารามิเตอร์ของข้อสอบ

SCOre = ชื่อไฟล์คะแนนนักเรียน

INFormation = ชื่อไฟล์ item – information (เฉพาะ Partial Credit Model)

INPUT : ใช้อธิบายข้อมูลที่จะนำมาวิเคราะห์ และสารสนเทศอื่น ๆ มีรูปแบบดังนี้

> INPut NTEst = a, LENgth = (b1, b2, b3, … bn), NFMt = c, SAMple = d, TAKe = e, NIDch = f, NTOtal = g, INOpt = h, MAXcat = I;

NTEst = จำนวนชุดย่อยของแบบสอบ (Subtest)

LENgth = (b1, b2, …, bn) จำนวนข้อสอบที่วิเคราะห์ (ทั้งชุด)

NFMt = จำนวนแถวของข้อมูลการตอบของผู้สอบแต่ละคน

SAMple = ร้อยละ (0 – 100) ของผู้ตอบที่จะสุ่มมาวิเคราะห์

TAKE = จำนวนที่จะนำผลการตอบ จากคนแรกถึงผู้ตอบ e มาวิเคราะห์

NIDch = จำนวนตัวอักษรในฟิลด์ ID

NTOtal = จำนวนข้อกระทงในข้อมูลดิบทั้งหมด

MAXCAT = ค่าสูงสุดของจำนวนลำดับขั้นคะแนน (≤ 15)

TEST หรือ SCALE : ใช้บอกชื่อแบบสอบ หรือแบบสอบย่อย มีรูปแบบดังนี้

```
> [TEST ]
  [SCAle] TNAme = a, NBLock = b, ITEms = (C1, C2, ..., Cn),
          INAme = (d1, d2, ..., dn), INTercept = (e1, e2, ... en),
          THReshold = (f1, f2, ..., fn), SLOpe = (g1, g2, ..., gn);
```

TNAme = ชื่อชุดของแบบสอบ หรือมาตรวัด

NBLock = จำนวนชุดย่อยของแบบสอบที่มีจำนวนลำดับขั้น (categories) เท่ากัน

ITEms = เรียงลำดับข้อสอบในข้อมูลที่พิมพ์เพื่อการวิเคราะห์

INAme = ชื่อของข้อกระทงในแบบสอบ หรือชุดของแบบสอบย่อย

INTercept = คือค่ากำหนดเริ่มต้นของ Intercept อาจกำหนดที่ THRESHOLD ก็ได้ แต่ต้องกำหนดเพียงที่ใดที่หนึ่งเท่านั้น

THReshole = คือค่ากำหนดเริ่มต้นของค่า Threshold ข้อกระทง ซึ่งอาจกำหนดจาก INTERCEPT ก็ได้ ซึ่งทั้งสองแห่งถ้าไม่กำหนดโปรแกรมก็จะกำหนดให้เอง

<u>ตัวอย่าง</u> การใช้คำสั่ง TEST

```
> TEST TNAme = TEST, NBLock = 1, Item = (1(2)10),
    INAme = (A1, A2, A3, A4, A10, A11, A12, A13, A14, A15);
```

BLOCK : ใช้เพื่อกำหนด Block ของข้อสอบ และเพื่อบอกชื่อข้อสอบตามชุดย่อย

```
> BLOck BNAme = a, NITems = b, NCAt = c,
  ORIginal = (d1, d2, ..., dncat), MODified = (e1, e2, ..., encat),
  CNAme = (f1, f2, ..., fmodified), CATEGORY = (g1, g2, ..., gmodified),
  CADjust = h, GPArm = (i1, i2, ... I modified),
  SCOring = (I1, I2, ..., I modified),
```

BNAme = ชื่อ Block ของข้อสอบ

NITems = จำนวนข้อสอบในแต่ละ Block หรือแต่ละแบบสอบย่อย

NCAt = จำนวนลำดับขั้นคะแนน (Categories)

ORIginal = รหัสคะแนนการตอบจากข้อมูลครั้งแรก

MODified = รหัสคะแนนการตอบที่ปรับ (Modified) จากรหัสการตอบครั้งแรก

CNAme = ชื่อของแต่ละลำดับขั้นคะแนน (Categories)

CATegory = ค่าพารามิเตอร์ Category เริ่มแรก สำหรับนำไปใช้ในการประมาณค่า

CADjust = ค่าเฉลี่ยของ Category Parameters ซึ่งเป็นค่าที่ปรับแก้

GPArm = ค่าโอกาสการเดา ใช้กับข้อสอบที่ตรวจให้คะแนนแบบทวิวิภาค

Scoring = คือฟังก์ชันของคะแนนการตอบ Defeult = 1.0, 2.0, 3.0, ...

ตัวอย่างเช่น

> Block Ncat = 4, ORIginal = (A, B, C, D) MODified = (1, 1, 2, 3);

CALIB : ใช้บังคับในการประมาณค่าพารามิเตอร์ของข้อสอบ และลำดับขั้นคะแนน และเพื่อกำหนดการแจกแจงเริ่มต้นในการประมาณค่าพารามิเตอร์แบบสอบ มีรูปแบบดังนี้

> CALib [GRAded, PARtial], [LOGistic, NORmal], SCAle = a,
NQPT = b, DIST = c, CYCles = (d, e, f, g, h), CRIt = (i, j, k, l, m),
QRANGE = (O, P), ITEmfit = q, SPRior, TPRior, GPRior;

GRAded = Graded Response Model

PARtial = Partial Credit Model (สำหรับ GPCM ต้องใช้ optional ใน SPRior, TPRior)

LOGistic = จะใช้กับการวิเคราะห์ตาม PARtial, หรือ GRAded.

NORmal = จะใช้กับ GRA

(หมายเหตุ * หากเลือก PARtial จะใช้ฟังก์ชันแบบ LOGistic ตัวเดียวเท่านั้น)

SCALE = ค่าคงที่ของสเกล

ค่า 1.0 สำหรับ Normal Ogive Model

ค่า 1.7 สำหรับ Logistic Model.

NQPT = จำนวน Quadrature Points ในการประมาณค่าแบบ EM (Default = 10 สำหรับข้อสอบน้อยกว่า / เท่ากับ 50 ข้อ นอกนั้น = 20)

DIST = ใช้เพื่อกำหนดการแจกแจงเริ่มแรกของการประมาณค่าพารามิเตอร์

1 = Uniform Distribution

2 = Normal on Equally Space Points.

3 = Normal on Gauss – Hermite Points

4 = User Supplied. (Default = 2)

CYCles d = จำนวนรอบของ EM (Default = 10 ถ้าข้อสอบน้อยกว่า/เท่ากับ 50 ข้อ นอกนั้น =5)

e = จำนวนรอบสูงสุดในการประมาณค่าซ้ำของ EM (inner EM) ของค่าพารามิเตอร์ข้อสอบ และ Category Parameter

f = จำนวนสูงสุดของการประมาณค่าซ้ำใน EM สำหรับการประมาณค่า Category Parameter

g = จำนวนรอบสูงสุดของการประมาณค่าซ้ำใน EM สำหรับการประมาณค่าพารามิเตอร์ของข้อสอบ

h = จำนวนรอบต่ำสุดของการประมาณค่าซ้ำใน EM สำหรับการประมาณค่าพารามิเตอร์ของข้อสอบ และ Category Parameter

CRIt ; i = Convergence Criterion ของ EM (Default = 0.001)

j = Convergence Criterion ของ EM ในการประมาณค่าพารามิเตอร์ (Default = i (ACRIT))

QRAnge = (o, p);

o = ค่าต่ำสุด

p = ค่าสูงสุด

SCORE : เป็นคำสั่งเพื่อคำนวณคะแนนเริ่มแรกของผู้ตอบแต่ละคน หรือแต่ละกลุ่ม มีรูปแบบ (Format) ดังนี้

> SCOre NQPT = d, DISt = b, QRAnge = (c, d), SMEan = e, SSD = f,

NAMe = g, PFQ = h, ITEration = (i, j), SCORING = | STAndard / CALibration | , | MLE / EAP |

NQPT = คือจำนวน Quadrature Points (Default = 30)

DISt = คือรูปแบบของการแจกแจงของคะแนนเริ่มแรก

b = 1 การแจกแจงแบบ Uniform

b = 2 การแจกแจงแบบ Normal on Equallyspaced points

b = 3 การแจกแจงแบบ NORmal on GAUSS-Hermite

(Default = 2)

QRAnge = ค่าสูงสุด (c) และค่าต่ำสุด (d) ของค่า Quadrature Point Default = (-4.0, +4.0)

SMEan = ค่าเฉลี่ยของ Scale Scores ที่ต้องการกำหนด

SSD = ค่า SD ของ Scale Scores

NAMe = Score Name (ถ้ามี Subtest และ Subscale เป็นชื่อต่างกัน)

PFQ = ร้อยละของผู้ตอบใน Upper Category และ Lower Category ที่ต้องการมารวมกับ Category ใกล้เคียง (Default = 2.0)

ITEration = (i, j) คือ Maximum Likelihood Scoring และ จำนวนรอบของการ Solution จะหยุดการวิเคราะห์เมื่อมีค่าการเปลี่ยนแปลงน้อยกว่า i

SCOring = STANDARD / CALIBRATION

ใช้ STANDARD เมื่อฟังก์ชันการให้คะแนนแบบ (1.0, 2.0, ...)

ใช้ CALIBRATION เมื่อมีการกำหนดฟังก์ชันเฉพาะ ซึ่งกำหนดในคำสั่ง BLOCK

MLE / EAP = เป็น Options ของการประมาณค่า Scale Score

MLE = Maximum Likelihood Estimation

EAP = Expected a Posteriori Estimation (BAY)

(Default = EAP)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่3
วิธีดำเนินการวิจัย

การวิจัยในครั้งนี้มีวัตถุประสงค์ 3 ประการ คือ ประการที่ 1 เพื่อเปรียบเทียบคุณภาพของแบบสอบในด้านความตรงตามสภาพ ระหว่างวิธีตรวจการให้คะแนนแบบประเพณีนิยมกับวิธีการตรวจให้คะแนนความรู้บางส่วนตามวิธีประยุกต์การให้คะแนนของคูมบ์ และวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท ประการที่ 2 เพื่อเปรียบเทียบความยากและอำนาจจำแนกโดยใช้ทฤษฎีการตอบสนองข้อสอบระหว่างวิธีการตรวจให้คะแนนแบบประเพณีนิยมกับวิธีการตรวจให้คะแนนความรู้บางส่วนตามวิธีประยุกต์การให้คะแนนของคูมบ์ และวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท ประการที่ 3 เพื่อเปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบสอบ และอัตราส่วนสารสนเทศเฉลี่ยโดยใช้ทฤษฎีการตอบสนองข้อสอบระหว่างวิธีการตรวจให้คะแนนแบบประเพณีนิยมกับวิธีการตรวจให้คะแนนความรู้บางส่วนตามวิธีประยุกต์การให้คะแนนของคูมบ์ และวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท โดยมีขั้นตอนในการดำเนินการวิจัย ดังนี้

ประชากรและกลุ่มตัวอย่าง

ประชากรและกลุ่มตัวอย่างที่ใช้ในการวิจัยได้แก่ นิสิตคณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ชั้นปีที่ 3 ที่ลงทะเบียนเรียน วิชาการประเมินผลการเรียนการสอน ในภาคการศึกษาต้น ปีการศึกษา 2544 จำนวน 297 คน แต่เมื่อเก็บข้อมูลโดยให้นิสิตทำข้อสอบ 3 ครั้ง ห่างกัน 2 สัปดาห์ พบว่ามีจำนวนนิสิตที่เป็นกลุ่มตัวอย่างทำข้อสอบแต่ละวิธี ดังนี้

วิธีการให้คะแนน	จำนวนตัวอย่าง	ร้อยละ
1. วิธีตัดตัวลวง	252	84.84
2. วิธีเลือกชุดตัวถูก	264	88.88
3. วิธีประเพณีนิยม	261	87.87

เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยครั้งนี้ เป็นแบบสอบที่ใช้เป็นแบบวัดผลสัมฤทธิ์ทางการเรียน วิชาการประเมินผลการเรียนการสอน ที่มีวิธีการให้คะแนน 3 วิธี จากแบบสอบชุดเดียวกัน คือวิธีการให้คะแนนแบบประเพณีนิยม วิธีประยุกต์การให้คะแนนของคูมบ์ (แบบตัดตัวลวง) และวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท (แบบเลือกชุดตัวถูก)

การสร้างแบบวัดผลสัมฤทธิ์ทางการเรียน วิชาการประเมินผลการเรียนการสอนที่ใช้ในการวิจัยครั้งนี้ ได้คัดเลือกข้อสอบที่มีคุณภาพ จากคลังข้อสอบที่ผ่านการวิเคราะห์ข้อสอบแล้ว ซึ่งเป็นข้อสอบแบบเลือกตอบที่มี 4 ตัวเลือก โดยมีตัวถูกเพียง 1 ตัว ในแต่ละข้อ จำนวน 50 ข้อ โดยมีค่าความยากและ

ค่าอำนาจจำแนกอยู่ในเกณฑ์ที่กำหนด (ค่าความยากอยู่ระหว่าง 0.2-0.8 ค่าอำนาจจำแนกมีค่า 0.2 ขึ้นไป) มีเนื้อหาของการวัดและการประเมินผลการเรียนการสอน 8 เรื่องที่นิสิตได้เรียนไปแล้ว และมีการสอบกลางภาคไปก่อนที่จะมีการเก็บข้อมูล 1 สัปดาห์ ซึ่งมีโครงสร้างของตารางวิเคราะห์ข้อสอบ ดังนี้

เนื้อหา	จำนวนข้อ	ข้อที่
1. ความหมายและความสำคัญของการวัดและประเมินผล	4	1,2,3,4
2. ประเภทของการวัดและประเมินผล	7	5,6,7,8,9,10,11
3. การวัดพฤติกรรมการเรียนรู้	8	12,16,17,18,19
4. เครื่องมือการวัดผล	8	20,21,22,23,24,25,26,50
5. การสร้างข้อสอบวัดผลสัมฤทธิ์	4	13,14,15,27,28,29,30
6. ความหมายของความยากและอำนาจจำแนก	5	32,33,34,36,41
7. การวิเคราะห์ข้อสอบแบบอิงกลุ่ม	10	31,37,38,39,40,42,43,44,45,46
8. การวิเคราะห์ข้อสอบแบบอิงเกณฑ์	4	35,47,48,49

คุณภาพของเครื่องมือ

การวิเคราะห์คุณภาพของแบบสอบที่คัดเลือกจากคลังข้อสอบ ตามตารางวิเคราะห์ข้อสอบนั้น ผู้วิจัยได้นำมาหาความตรง โดยใช้วิธีเปรียบเทียบคะแนนความสามารถกับกลุ่มที่ทราบผล (known group technique) โดยเปรียบเทียบคะแนนความสามารถ (θ) ของนิสิตที่มีการตรวจให้คะแนนด้วยวิธีประยุกต์การให้คะแนนของคูมบ์ วิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท วิธีประเพณีนิยม จำแนกตามกลุ่มที่มีความสามารถสูง ปานกลาง ต่ำ จากการใช้คะแนนความสามารถจากการสอบกลางภาคเป็นตัวแบ่งกลุ่มที่มีเกณฑ์ คือ 1) ความสามารถ .75 ขึ้นไป เป็นกลุ่มที่มีความสามารถสูง 2) ความสามารถ (-.75) - .75 เป็นกลุ่มที่มีความสามารถปานกลาง และ3) ความสามารถต่ำกว่า-.75 เป็นกลุ่มที่มีความสามารถต่ำ ผลที่ได้พบว่าการให้คะแนนในแต่ละวิธี สามารถจำแนกกลุ่มทั้ง 3 กลุ่มได้ คือมีคะแนนเฉลี่ยในแต่ละกลุ่มแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 (ใช้สถิติ F-test และเปรียบเทียบความแตกต่างระหว่างกลุ่มโดยใช้ วิธีของ Scheffe) ผลแสดงใน ตาราง ที่ 3.1 – 3.3 ต่อไปนี้

ตารางที่ 3.1 ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ของคะแนนความสามารถ และF – test ตามวิธีประยุกต์การให้คะแนนของคูมบ์จำแนกตามกลุ่มความสามารถ

กลุ่ม	จำนวนผู้สอบ	วิธีประยุกต์การให้คะแนนของคูมบ์		F-test	p-value
		ค่าเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน		
ความสามารถสูง	51	.388 [a/b]	.863	11.215	0.000
ความสามารถปานกลาง	151	-.116 [a/c]	.869		
ความสามารถต่ำ	50	-.435 [b/c]	.966		

หมายเหตุ a,b,c คือ สัญลักษณ์แสดงคู่ที่มีค่าเฉลี่ยแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ0.05 โดยที่

a เป็นความแตกต่างระหว่างกลุ่มความสามารถสูงและกลุ่มความสามารถปานกลาง

b เป็นความแตกต่างระหว่างกลุ่มความสามารถสูงและกลุ่มความสามารถต่ำ

c เป็นความแตกต่างระหว่างกลุ่มความสามารถปานกลางและกลุ่มความสามารถต่ำ

ตารางที่3.2 ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐานของคะแนนความสามารถ และF – test ตามวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิทจำแนกตามกลุ่มความสามารถ

กลุ่ม	จำนวนผู้สอบ	คะแนนความสามารถตามวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท		F-test	p-value
		ค่าเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน		
ความสามารถสูง	58	.383 [a/b]	.494	29.534	0.000
ความสามารถปานกลาง	159	.080 [a/c]	.415		
ความสามารถต่ำ	47	-.265 [b/c]	.395		

หมายเหตุ a,b,c คือ สัญลักษณ์แสดงคู่ที่มีค่าเฉลี่ยแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ0.05 โดยที่

a เป็นความแตกต่างระหว่างกลุ่มความสามารถสูงและกลุ่มความสามารถปานกลาง

b เป็นความแตกต่างระหว่างกลุ่มความสามารถสูงและกลุ่มความสามารถต่ำ

c เป็นความแตกต่างระหว่างกลุ่มความสามารถปานกลางและกลุ่มความสามารถต่ำ

ตารางที่ 3.3 ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ของคะแนนความสามารถ และF – test ตามวิธีให้คะแนนแบบประเพณีนิยมจำแนกตามกลุ่มความสามารถ

กลุ่ม	จำนวนผู้สอบ	คะแนนความสามารถตามวิธีประเพณีนิยม		F-test	p-value
		ค่าเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน		
ความมีสามารถสูง	56	.523 [a/b]	.860	41.100	0.000
ความสามารถปานกลาง	155	.008 [a/c]	.749		
ความสามารถต่ำ	50	-.824 [b/c]	.789		

หมายเหตุ a,b,c คือ สัญลักษณ์แสดงคู่ที่มีค่าเฉลี่ยแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ0.05โดยที่

a เป็นความแตกต่างระหว่างกลุ่มความสามารถสูงและกลุ่มความสามารถปานกลาง

b เป็นความแตกต่างระหว่างกลุ่มความสามารถสูงและกลุ่มความสามารถต่ำ

c เป็นความแตกต่างระหว่างกลุ่มความสามารถปานกลางและกลุ่มความสามารถต่ำ

สำหรับความเที่ยง ได้หาความเที่ยงแบบความสอดคล้องภายในของครอนบาค (Cronbach) ของแบบสอบที่ใช้ในการศึกษาที่ให้คะแนนด้วยวิธีการต่างๆ พบว่าแต่ละวิธีมีความเที่ยงค่อนข้างสูงอยู่ระหว่าง 0.7516 – 0.8093 และจากการทดสอบด้วยสถิติ t พบว่าค่าความเที่ยงของวิธีประยุกต์การให้คะแนนของคูมบ์สูงกว่าวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิทและวิธีประเพณีนิยมอย่างมีนัยสำคัญทางสถิติ ดังแสดงในตาราง ดังนี้

วิธีการให้คะแนน	ความเที่ยง (α)	t *	
1. วิธีประยุกต์การให้คะแนนของคูมบ์	0.8093	1 vs 2	2.77
2. วิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท	0.7516	1 vs 3	2.38
3. วิธีประเพณีนิยม	0.7539	2 vs 3	0.09

* เปรียบเทียบความแตกต่างของค่าความเที่ยง โดยใช้วิธีของ Pitman (Feldt,1980)

วิธีการให้คะแนนที่ใช้ศึกษาครั้งนี้มี 3 วิธี คือ

1. วิธีประยุกต์การให้คะแนนของคูมบ์ (Elimination Scoring ,ET) เป็นวิธีการให้คะแนนความรู้บางส่วนที่ผู้วิจัยพัฒนาขึ้นจากคูมบ์ วิธีที่พัฒนาไม่มีค่าติดลบและมีค่าสูงสุด เท่ากับ 3 ต่ำสุดเท่ากับ 0 ในแต่ละข้อ โดยใช้สูตรการให้คะแนน $X = n C$

เมื่อ X คือ คะแนนที่ผู้สอบจะได้จากการตอบข้อสอบข้อหนึ่งๆ

n คือ จำนวนตัวลวงที่ตัดออก

C คือ ค่าความถูกต้องของการตัดตัวลวง มี 2 ค่า ได้แก่

C = 1 เมื่อไม่มีตัวเลือกถูกอยู่ในตัวลวงที่ตัดออก

C = 0 เมื่อมีตัวเลือกถูกอยู่ในตัวลวงที่ตัดออก

และผู้ตอบจะได้คะแนน 0 เมื่อไม่ตอบ

2. วิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท (The Subset Selection Method, SST) เป็นวิธีการให้คะแนนความรู้บางส่วนที่ผู้วิจัยพัฒนาขึ้นจากวิธีของเดรสเซลและสมิท (Dressel and schmidt, 1953) วิธีที่พัฒนาไม่มีค่าติดลบและมีค่าสูงสุด เท่ากับ 3 ต่ำสุดเท่ากับ 0 ในแต่ละข้อ โดยใช้สูตรการให้คะแนน $X = K - nC$

เมื่อ X คือ คะแนนที่ผู้สอบจะได้จากการตอบข้อสอบข้อหนึ่งๆ

K คือ จำนวนตัวเลือกทั้งหมดในข้อสอบ

n คือ จำนวนตัวเลือกที่ผู้ตอบเลือก

C คือ ค่าความถูกต้องของชุดย่อยของคำตอบ มี 2 ค่า ได้แก่

C = 1 เมื่อมีตัวเลือกถูกอยู่ในชุดย่อยของคำตอบ

C = 0 เมื่อไม่มีตัวเลือกถูกอยู่ในชุดย่อยของคำตอบ

และผู้ตอบจะได้คะแนน 0 เมื่อไม่ตอบ

3. การให้คะแนนแบบประเพณีนิยม เป็นการให้คะแนนกับคำตอบที่ถูกต้องเป็น 1 คำตอบอื่นๆ ที่ผิดเป็น 0

การเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูลในการวิจัยนี้ เก็บข้อมูลโดยให้นิสิตคณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ชั้นปีที่ 3 ที่ลงทะเบียนเรียน วิชาการประเมินผลการเรียนการสอน ในภาคการศึกษาต้น ปีการศึกษา 2544 จำนวน 297 คน ทำข้อสอบวัดผลสัมฤทธิ์ทางการเรียน วิชาการประเมินผลการเรียนการสอนที่สร้างขึ้น 3 ครั้ง ด้วยแบบสอบชุดเดียวกัน แต่วิธีการตอบแตกต่างกัน โดยการเก็บข้อมูลแต่ละครั้งห่างกัน 2 สัปดาห์ ดังนี้

ครั้งแรก ให้นิสิตทำข้อสอบวิธีประยุกต์การให้คะแนนของคูมบ์ (แบบตัดตัวลวง) โดยที่นิสิตสามารถตัดตัวลวงที่เป็นคำตอบที่ผิดได้ 1 หรือ2 หรือ 3 ตัวเลือก ก็ได้ โดยไม่จำเป็นต้องตัดตัวเลือกทั้ง 3 ตัวในแต่ละข้อ แต่พยายามตัดออกตามความรู้ที่สามารถตัดออกได้เท่านั้น ถ้าไม่สามารถตัดออกได้ก็ไม่ต้องตัดออก

กรณีที่ตัดตัวเลือกที่ผิดได้ถูกต้องทั้ง 3 ตัวเลือก จะได้คะแนนเต็ม 3 คะแนน

กรณีที่ไม่แน่ใจ ตัดตัวเลือกที่ผิดได้บ้าง โดยตัดตัวลวงถูก 2 ตัว จะได้ 2 คะแนน และ ตัดตัวลวงถูก 1 ตัวจะได้ 1 คะแนน แต่ถ้าตัดตัวเลือกที่เป็นคำตอบถูกออกไม่ว่าจะตัดกี่ข้อก็ตาม จะได้ 0 คะแนน

กรณีที่ไม่ตอบก็จะได้ 0 คะแนน

ครั้งที่สอง ให้นิสิตทำข้อสอบวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท (แบบเลือกชุดตัวถูก) โดยที่นิสิตสามารถเลือกชุดคำตอบที่ถูกได้ 1 หรือ2 หรือ 3 ตัวเลือก ก็ได้ โดยไม่จำเป็นต้องเลือกตัวถูกเพียงตัวเดียวในแต่ละข้อ แต่พยายามเลือกตัวถูกตามความรู้ที่สามารถที่แท้จริง โดยไม่มีการเดา ถ้าไม่สามารถเลือกตัวถูกได้ ก็ให้เว้นไว้ กรณีที่เลือกตัวถูก ได้ถูกต้อง คือเลือกเพียง 1 ตัวและเป็นคำตอบที่ถูก จะได้คะแนนเต็ม 3 คะแนน

กรณีที่ไม่แน่ใจ เลือกตัวถูก 2 ตัว และใน 2 ตัวนั้นมีคำตอบถูกอยู่ด้วย จะได้ 2 คะแนน หรือเลือกตัวถูก 3 ตัว และใน 3 ตัวนั้นมีคำตอบถูกอยู่ด้วย จะได้ 3 คะแนน แต่ถ้าตัวเลือกที่เลือกไม่มีคำตอบที่ถูกไม่ว่าจะเลือกกี่ข้อก็ตาม จะได้ 0 คะแนน

กรณีที่ไม่ตอบก็จะได้ 0 คะแนน

ครั้งที่สาม ให้นิสิตทำข้อสอบแบบประเพณีนิยม คือ ให้เลือกคำตอบเพียง 1 ตัว ถ้าตอบถูก ได้ 1 คะแนน ตอบผิดได้ 0 คะแนน

การวิเคราะห์ข้อมูล

1.การวิเคราะห์ความตรงตามสภาพ จากค่าสัมประสิทธิ์สหสัมพันธ์(r)และแปลงค่าความตรงตามสภาพจากค่าสัมประสิทธิ์สหสัมพันธ์ เป็น คะแนนฟิชเชอร์ซี (Z_r) โดยใช้สูตร (Glass and Stanley , 1970)

$$Z_r = \log_e \sqrt{(1+r)/(1-r)}$$

เปรียบเทียบความเป็นอิสระของความตรงตามสภาพในภาพรวม โดยใช้สถิติทดสอบไคว์สแควร์ (χ^2) (Glass and Stanley , 1970) มีสูตร คือ

$$\chi^2_{j-1;\alpha} = \sum (Z_{rj} - Z_r)^2 (n_j - 3)$$

$$Z_r = \frac{\sum (n_j - 3) Z_{rj}}{\sum (n_j - 3)}$$

เมื่อ Z_{rj} แทน คะแนนฟิชเชอร์ซี ของกลุ่มผู้สอบj

n_j แทน จำนวนผู้สอบกลุ่มที่ j

และเปรียบเทียบความแตกต่างของค่าความตรงตามสภาพเป็นรายคู่ โดยใช้สถิติทดสอบ Z มีสูตร คือ (Glass and Stanley , 1970)

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

เมื่อ Z แทน สถิติทดสอบซี (Z test)

Z_{r1} , Z_{r2} แทน คะแนนฟิชเชอร์ซี

n_1 , n_2 แทน จำนวนข้อมูลของกลุ่มตัวอย่าง

2. การวิเคราะห์ค่าความเที่ยงโดยใช้ทฤษฎีการวัดแบบดั้งเดิม ใช้ค่าสัมประสิทธิ์แอลฟ่า (α) และทดสอบความแตกต่างของค่าสัมประสิทธิ์แอลฟ่าเป็นรายคู่ โดยใช้ สถิติ t –test ซึ่งเป็นวิธีของ Pitman (Feldt.1980) มีสูตร คือ

$$t_{n-2} = \frac{(W-1)(N-2)^{1/2}}{(4\,W\,(1-r^{2}_{x1x2}))^{1/2}}$$

เมื่อ W = $(1-r_1) / (1-r_2)$

r_1 และ r_2 = ค่าสัมประสิทธิ์ความเที่ยง

r_{x1x2} = ค่าสัมประสิทธิ์ความสัมพันธ์ระหว่าง X_1 และ X_2

3. การวัดคุณภาพของแบบสอบ จาก ฟังก์ชันสารสนเทศของแบบสอบ อัตราส่วนสารสนเทศเฉลี่ย และฟังก์ชันสารสนเทศของข้อสอบโดยใช้การวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) แบบทวิวิภาค ที่ใช้โมเดลโลจิส (Logistic) และแบบพหุวิภาคที่ใช้โมเดลการให้คะแนนความรู้บางส่วน (Partial Credit Model ,PCM) โดยที่

1) การวิเคราะห์แบบทวิวิภาค ที่ใช้โมเดลโลจิส (Logistic) ซึ่งเป็นการให้คะแนนแบบประเพณีนิยม ประมาณค่าพารามิเตอร์ความยาก(b) อำนาจจำแนก(a) ค่าฟังก์ชันสารสนเทศของข้อสอบและค่าฟังก์ชันสารสนเทศของแบบสอบ(TIF) โดยใช้โปรแกรม BILOG

2) การวิเคราะห์แบบพหุวิภาค ที่ใช้โมเดลการให้คะแนนความรู้บางส่วน (Partial Credit Model, PCM) ซึ่งเป็นการให้คะแนนด้วยวิธีประยุกต์การให้คะแนนของดูมบ์ และวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท ประมาณค่าพารามิเตอร์ความยาก(b)อำนาจจำแนก(a) ค่าฟังก์ชันสารสนเทศของข้อสอบและค่าฟังก์ชันสารสนเทศของแบบสอบ(TIF) โดยใช้โปรแกรม PARSCALE

4. การคำนวณอัตราส่วนสารสนเทศเฉลี่ย (RAI) จากสูตร

$$RAI(\theta, X, Y) = AI(\theta,X) / Ai(\theta,Y)$$

เมื่อ AI คือค่าเฉลี่ยสารสนเทศของแบบสอบ

X คือแบบสอบชุด X

Y คือแบบสอบชุดY

5. การเปรียบเทียบความแตกต่างของค่าเฉลี่ยของค่าอำนาจจำแนก ความยาก ค่าฟังก์ชันสารสนเทศของข้อสอบและค่าฟังก์ชันสารสนเทศของแบบสอบ(TIF) ใช้สถิติ F- test และเปรียบเทียบความแตกต่างรายคู่ ด้วยวิธีการของ Scheffe โดยใช้โปรแกรม SPSS for Windows version 10.0

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 4
ผลการวิเคราะห์ข้อมูล

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบคุณภาพของแบบสอบในด้านความตรงตามสภาพ ความยาก อำนาจจำแนก ฟังก์ชันสารสนเทศของข้อสอบ และฟังก์ชันสารสนเทศของแบบสอบ จำแนกระหว่างวิธีการตรวจให้คะแนนแบบประเพณีนิยม วิธีการตรวจให้คะแนนความรู้บางส่วนตามวิธีประยุกต์การให้คะแนนของคูมบ์ และวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท โดยผู้วิจัยได้เสนอผลการวิเคราะห์ เป็น 4 ตอน ตามลำดับดังนี้

ตอนที่ 1 ผลการวิเคราะห์ข้อมูลเบื้องต้น

ตอนที่ 2 ผลการเปรียบเทียบความตรงตามสภาพ ระหว่างวิธีการให้คะแนนที่แตกต่างกันตามวัตถุประสงค์ของการวิจัย ข้อ 1

ตอนที่ 3 ผลการเปรียบเทียบความยาก และอำนาจจำแนก ระหว่างวิธีการให้คะแนนที่แตกต่างกันตามวัตถุประสงค์ของการวิจัย ข้อ 2

ตอนที่ 4 ผลการเปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบสอบ และอัตราส่วนสารสนเทศ ระหว่างวิธีการให้คะแนนที่แตกต่างกัน ตามวัตถุประสงค์ของการวิจัย ข้อ 3

ตอนที่ 1 ผลการวิเคราะห์ข้อมูลเบื้องต้น

ผลการวิเคราะห์ข้อมูลเบื้องต้น ได้นำเสนอข้อมูลเป็น 2 ส่วน ส่วนแรก คือ ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าคะแนนดิบ (ตารางที่ 4.1) การเปรียบเทียบความแตกต่างของคะแนนเฉลี่ยรายคู่ในแต่ละวิธีการให้คะแนน (ตารางที่ 4.2) ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของความสามารถของผู้สอบ (ตารางที่ 4.3) และ การเปรียบเทียบความแตกต่างของค่าเฉลี่ยความสามารถรายคู่ในแต่ละวิธีการให้คะแนน (ตารางที่ 4.4) จากการทำข้อสอบวิชาการประเมินผลการเรียนการสอน จำนวน 50 ข้อ ที่มีวิธีการให้คะแนน 3 วิธี จากแบบสอบชุดเดียวกัน คือ วิธีประยุกต์การให้คะแนนของคูมบ์ วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท และวิธีการให้คะแนนแบบประเพณีนิยม โดยที่ 2 วิธีแรกมีคะแนนสูงสุด เท่ากับ 3 ต่ำสุดเท่ากับ 0 ในแต่ละข้อ รวมคะแนนเต็มเท่ากับ 150 คะแนน ส่วนวิธีการให้คะแนนแบบประเพณีนิยม มีคะแนนเต็ม 50 คะแนน ดังแสดงในตารางต่อไปนี้

ตารางที่ 4.1 ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบจำแนกตามวิธีการให้คะแนน

วิธีการให้คะแนน	จำนวนผู้สอบ	คะแนน (50 ข้อ)				
		Min	Max	Mean (คะแนนเต็ม)	Mean %	S.D.
วิธีประยุกต์การให้คะแนนของคูมบ์	252	28	126	74.07 (150)	49.38	13.18
วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท	264	29	123	77.43 (150)	51.62	12.52
วิธีประเพณีนิยม	261	11	39	25.83 (50)	51.66	13.12

จากตารางที่ 4.1 เมื่อพิจารณาค่าร้อยละของค่าเฉลี่ย (Mean %) พบว่า การให้คะแนนวิธีประเพณีนิยมและวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท มีค่าเฉลี่ยสูงกว่าวิธีประยุกต์การให้คะแนนของคูมบ์ ในขณะที่ส่วนเบี่ยงเบนมาตรฐานการให้คะแนน วิธีประยุกต์การให้คะแนนของคูมบ์มีค่าสูงกว่าวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท และวิธีประเพณีนิยม

ตารางที่ 4.2 ค่า t ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของคะแนนความแตกต่างระหว่างวิธีการให้คะแนนที่ต่างกัน

การเปรียบเทียบระหว่าง		Paired Differences		t	p-value
		Mean	SD		
วิธีประเพณีนิยม	วิธีประยุกต์การให้คะแนนของคูมบ์	2.425	11.673	3.144	.002
วิธีประเพณีนิยม	วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท	0.407	10.654	.589	.556
วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท	วิธีประยุกต์การให้คะแนนของคูมบ์	2.168	9.911	3.282	.001

จากตารางที่ 4.2 เมื่อเปรียบเทียบความแตกต่างของค่าเฉลี่ยเป็นรายคู่พบว่าวิธีประเพณีนิยมมีคะแนนเฉลี่ยสูงกว่าวิธีวิธีประยุกต์การให้คะแนนของคูมบ์อย่างมีนัยสำคัญทางสถิติที่ระดับ0.05 และในขณะเดียวกันวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิทก็มีคะแนนเฉลี่ยสูงกว่าวิธีประยุกต์การให้คะแนนของคูมบ์อย่างมีนัยสำคัญทางสถิติที่ระดับ0.05

ตารางที่ 4.3 ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าความสามารถของผู้สอบจำแนกตามวิธีการให้คะแนน

วิธีการให้คะแนน	จำนวนผู้สอบ	ความสามารถของผู้สอบ (θ)			
		Min	Max	Mean	S.D.
วิธีประยุกต์การให้คะแนนของคูมบ์	252	-3.171	2.153	-.077	.923
วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท	264	-1.087	1.332	.087	.474
วิธีประเพณีนิยม	261	-2.043	1.959	-.000	.894

จากตารางที่ 4.3 พบว่าวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิทมีค่าเฉลี่ยความสามารถของผู้สอบสูงกว่าวิธีอื่นๆ แต่เมื่อพิจารณาจากค่าส่วนเบี่ยงเบนมาตรฐานพบว่าคะแนนความสามารถของผู้สอบมีการกระจายน้อยกว่ากลุ่มอื่นๆ ในขณะที่วิธีประยุกต์การให้คะแนนของคูมบ์ มีค่าเฉลี่ยความสามารถของผู้สอบต่ำกว่าวิธีอื่นๆแต่จากค่าส่วนเบี่ยงเบนมาตรฐานพบว่าคะแนนความสามารถของผู้สอบมีการกระจายมากกว่ากลุ่มอื่นๆ

ตารางที่ 4.4 ค่า t ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าความแตกต่างของค่าความสามารถระหว่างวิธีการให้คะแนนที่ต่างกัน

การเปรียบเทียบระหว่าง		Paired Differences		t	p-value
		Mean	SD		
วิธีประเพณีนิยม	วิธีประยุกต์การให้คะแนนของคูมบ์	.009	1.083	1.267	.206
วิธีประเพณีนิยม	วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท	-.005	.687	1.211	.227
วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท	วิธีประยุกต์การให้คะแนนของคูมบ์	.158	.861	2.759	.006

จากตารางที่ 4.4 เมื่อเปรียบเทียบความแตกต่างของค่าเฉลี่ยความสามารถเป็นรายคู่พบวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิทก็มีคะแนนความสามารถเฉลี่ยสูงกว่าวิธีประยุกต์การให้คะแนนของคูมบ์อย่างมีนัยสำคัญทางสถิติที่ระดับ0.05 ส่วนคู่อื่นๆไม่พบว่ามีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ

ส่วนที่สอง เป็นการนำเสนอรูปแบบการตอบของกลุ่มตัวอย่างในการแสดงความรู้บางส่วน ด้วยวิธีประยุกต์การให้คะแนนของคูมบ์ (ตารางที่4.5) และวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท (ตารางที่4.6) จำแนกตามกลุ่มที่มีความสามารถสูง ปานกลาง ต่ำ โดยมีรายละเอียดของรูปแบบการตอบจำแนกตามรายข้อ แสดงไว้ในภาคผนวก

ตารางที่ 4.5 ร้อยละของค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของการตัดตัวลวง(วิธีประยุกต์การให้คะแนนของคูมบ์) จำแนกตามกลุ่มความสามารถ

กลุ่มความสามารถ	การตัดตัวลวง	Mean %	SD
สูง	1-2	24.51	11.27
	3	69.35	12.67
	ไม่ตอบ	6.11	5.30
ปานกลาง	1-2	28.95	9.81
	3	53.25	10.63
	ไม่ตอบ	10.64	5.16
ต่ำ	1-2	28.08	9.81
	3	41.48	10.71
	ไม่ตอบ	30.44	5.66

จากตารางที่ 4.5 เมื่อพิจารณารูปแบบการตัดตัวลวง จากกลุ่มความสามารถซึ่งแบ่งเป็น กลุ่มที่มีความสามารถสูง ปานกลาง ต่ำ จากการใช้คะแนนความสามารถจากการสอบกลางภาคเป็นตัวแบ่งกลุ่ม จะเห็นว่ากลุ่มความสามารถสูง จะตัดตัวเลือก 3 ตัว (ข้อสอบ 4 ตัวเลือก) มากกว่ากลุ่มอื่นๆ คิดเป็นร้อยละของค่าเฉลี่ยเท่ากับ 69.35 ในขณะที่การตัดตัวเลือก 1 –2 ตัว มีร้อยละของค่าเฉลี่ยสูงสุดในกลุ่มความสามารถปานกลาง มีค่าเท่ากับ 28.95 รองลงมาคือกลุ่มความสามารถต่ำ มีค่าเท่ากับ28.08 ส่วนการไม่ตอบมีร้อยละของค่าเฉลี่ยสูงสุดในกลุ่มความสามารถต่ำ มีค่าเท่ากับ 30.44

ตารางที่ 4.6 ร้อยละของค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของการเลือกชุดตัวถูก (วิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท) จำแนกตามกลุ่มความสามารถ

กลุ่มความสามารถ	การเลือกชุดตัวถูก	Mean %	SD
สูง	1	79.38	7.46
	2-3	11.48	7.11
	ไม่ตอบ	9.14	3.46
ปานกลาง	1	64.55	5.64
	2-3	11.37	5.06
	ไม่ตอบ	24.07	4.09
ต่ำ	1	56.74	6.73
	2-3	12.06	6.50
	ไม่ตอบ	32.13	7.58

จากตารางที่ 4.6 เมื่อพิจารณารูปแบบการเลือกชุดตัวถูก จากกลุ่มความสามารถซึ่งแบ่งเป็น กลุ่มที่มีความสามารถสูง ปานกลาง ต่ำ จะเห็นว่ากลุ่มความสามารถสูง จะเลือกตัวถูกเพียง 1 ตัว (ข้อสอบ 4 ตัวเลือก) มากกว่ากลุ่มอื่นๆ คิดเป็นร้อยละของค่าเฉลี่ยเท่ากับ 79.38 ในขณะที่การเลือกชุดตัวถูก มีร้อยละของค่าเฉลี่ยในแต่ละกลุ่มไม่แตกต่างกัน คืออยู่ระหว่าง 11.37-12.06 ส่วนการไม่ตอบมีร้อยละของค่าเฉลี่ยสูงสุดในกลุ่มความสามารถต่ำ มีค่าเท่ากับ 32.13

ตอนที่ 2 ผลการเปรียบเทียบความตรงตามสภาพ ระหว่างวิธีการตรวจให้คะแนนที่แตกต่างกัน

การคำนวณความตรงตามสภาพ ของแบบสอบที่มีวิธีการให้คะแนนที่แตกต่างกัน พิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์ (r) ระหว่างคะแนนของผู้สอบที่ได้จากการสอบวิชาการประเมินผลการเรียนการสอนที่จัดทำขึ้นกับคะแนนที่ได้จากแบบสอบกลางภาค ปีการศึกษา 2544 วิชาการประเมินผลการเรียนการสอน โดยที่การตรวจให้คะแนนจากการสอบที่จัดทำขึ้นมีการให้คะแนน 3 วิธีที่แตกต่าง และมีการแปลงค่าสัมประสิทธิ์สหสัมพันธ์ให้เป็นคะแนนฟิชเชอร์ซี (Fisher's Z) จากสูตร ของGlass and Stanley (1970) เพื่อนำมาเปรียบเทียบความแตกต่างของค่าความตรงตามสภาพในภาพรวมโดยใช้สถิติทดสอบไคว์สแควร์ (χ^2) และเปรียบเทียบเป็นรายคู่ โดยใช้สถิติทดสอบซี (Z) ได้ผลการวิเคราะห์แสดงในตารางที่ 4.7 – 4.8

ตารางที่ 4.7 ความตรงตามสภาพจากการสอบจำแนกตามวิธีการตรวจให้คะแนน

วิธีการตรวจให้คะแนน	จำนวนผู้สอบ	ความตรงตามสภาพ		χ^2
		r	Z_r	
วิธีประยุกต์การให้คะแนนของคูมบ์	252	.623**	.716	4.91
วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท	264	.498**	.546	
วิธีประเพณีนิยม	261	.519**	.574	

** p-value < 0.05

จากตารางที่4.7 พบว่าความตรงตามสภาพที่พิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์ (r) ระหว่างการให้คะแนน 3 วิธี กับคะแนนที่ได้จากแบบสอบกลางภาค มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 และเมื่อนำมาเปรียบเทียบความแตกต่างโดยใช้คะแนนฟิชเชอร์ซี เปรียบเทียบในภาพรวมโดยใช้สถิติทดสอบไคว์สแควร์ (χ^2) พบว่าค่าความตรงตามสภาพไม่ขึ้นอยู่กับวิธีการให้คะแนนในแต่ละวิธี (ตารางที่ 4.7) และเมื่อนำมาเปรียบเทียบความแตกต่างกันเป็นรายคู่ ด้วยสถิติทดสอบซี (Z) ผลปรากฏในตารางที่ 4.8

ตารางที่ 4.8 เปรียบเทียบความแตกต่างของความตรงตามสภาพ โดยใช้สถิติ Z จำแนกตามวิธีการให้ตรวจคะแนน

วิธีการตรวจให้คะแนน	วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท	วิธีประเพณีนิยม
วิธีประยุกต์การให้คะแนนของคูมบ์	1.370 (.086)	1.136 (.127)
วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท		0.227 (.409)

หมายเหตุ ค่าในวงเล็บเป็นค่า p หรือระดับความมีนัยสำคัญ

จากตารางที่ 4.8 จากค่าความตรงตามสภาพที่พิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์ (r)ในตารางที่ 4.7 แปลงเป็น Fisher Z เมื่อนำมาเปรียบเทียบความแตกต่างกันเป็นรายคู่ โดยใช้สถิติซี (Z) พบว่า ไม่มีความแตกต่างกันของความตรงสภาพในแต่ละคู่อย่างมีนัยสำคัญทางสถิติที่ระดับ0.05

ตอนที่ 3 ผลการเปรียบเทียบความยาก และอำนาจจำแนก ระหว่างวิธีการตรวจให้คะแนนที่แตกต่างกัน

ผลการวิเคราะห์เพื่อเปรียบเทียบในตอนนี้ ได้นำเสนอ ผลการเปรียบเทียบความยาก และอำนาจจำแนก เป็นรายข้อ (50ข้อ) ตามวิธีการให้คะแนนทั้ง 3 วิธี แสดงในตารางที่ 4.9 และเปรียบเทียบความแตกต่างของค่าอำนาจจำแนกเฉลี่ยและค่าความยากเฉลี่ยตามวิธีการให้คะแนนทั้ง 3 วิธี โดยใช้สถิติทดสอบ F ผลที่ได้แสดงในตารางที่ 4.10 ดังนี้

ตารางที่ 4.9 ค่าอำนาจจำแนกและค่าความยากของข้อสอบจำแนกตามรายข้อและวิธีการตรวจให้คะแนน

ข้อ	วิธีการตรวจให้คะแนน					
	การประยุกต์การให้คะแนนของคูมบ์		การประยุกต์การให้คะแนนของเดรสเชลและสมิท		แบบประเพณีนิยม(0-1)	
	อำนาจจำแนก	ความยาก	อำนาจจำแนก	ความยาก	อำนาจจำแนก	ความยาก
1	0.492	-2.581	0.308	-1.130	1.192	-1.732
2	0.399	-1.881	0.258	-1.267	0.970	-2.214
3	0.415	-0.495	0.158	0.941	0.829	0.900
4	0.474	-0.117	0.128	-0.861	0.520	-0.911
5	0.521	-0.042	0.214	0.833	0.462	1.764
6	0.702	-1.342	0.284	-0.078	0.684	-0.322
7	0.451	0.083	0.169	0.407	0.681	1.452
8	0.520	-0.392	0.251	0.232	0.546	0.143
9	0.584	-0.866	0.231	0.564	0.357	1.525
10	0.701	-0.598	0.172	-0.705	0.702	-0.963
11	0.806	-0.685	0.261	0.888	0.716	1.771

ตารางที่ 4.9 (ต่อ)

ข้อ	วิธีการตรวจให้คะแนน					
	การประยุกต์การให้คะแนนของดูมบ์		การประยุกต์การให้คะแนนของเดรสเชลและสมิท		แบบประเพณีนิยม(0-1)	
	อำนาจจำแนก	ความยาก	อำนาจจำแนก	ความยาก	อำนาจจำแนก	ความยาก
12	0.759	-1.343	0.273	-0.371	0.458	-1.045
13	0.475	-0.989	0.255	0.529	0.885	0.738
14	0.678	0.006	0.269	3.140	0.266	9.603
15	0.432	-0.876	0.171	-0.212	0.280	-0.740
16	0.587	-0.037	0.293	0.651	0.581	1.278
17	0.632	-3.012	0.396	-0.555	0.743	-1.713
18	0.522	-1.037	0.256	-0.036	0.568	0.128
19	0.664	-0.924	0.188	-0.990	0.749	-1.219
20	0.682	-1.132	0.286	-0.220	0.806	-0.477
21	0.587	-1.222	0.263	-0.194	0.712	-0.415
22	0.536	-1.679	0.341	-0.792	0.963	-1.588
23	0.637	-1.408	0.254	-0.210	0.488	-0.849
24	0.501	-1.588	0.204	-0.873	0.360	-1.165
25	0.592	-1.124	0.213	0.128	0.839	0.228
26	0.503	-1.554	0.267	0.222	0.506	0.380
27	0.724	-1.805	0.345	0.064	1.044	0.092
28	0.525	-0.574	0.188	-0.660	0.459	-1.034
29	0.546	-0.974	0.215	0.862	0.728	1.622
30	0.569	-0.873	0.271	-0.416	0.761	-0.595
31	0.676	-0.923	0.224	0.105	0.568	0.680
32	0.685	-0.977	0.304	0.595	0.465	0.895
33	0.517	-1.221	0.303	0.355	0.814	0.375
34	0.596	-0.366	0.257	1.065	0.527	2.243
35	0.935	-0.773	0.285	0.493	0.539	1.977
36	0.740	-0.821	0.255	0.597	0.352	1.407
37	0.799	-1.146	0.213	-0.069	0.375	-0.302
38	1.049	-1.256	0.311	-0.459	0.565	-1.087
39	0.594	-1.178	0.240	-0.434	0.510	-1.035
40	1.102	-1.320	0.352	0.022	0.797	0.172
41	0.578	-1.783	0.297	-0.143	0.521	-0.458
42	0.465	-1.782	0.352	-0.184	0.921	-0.316
43	0.939	-1.190	0.345	0.078	0.573	0.230

ตารางที่ 4.9 (ต่อ)

ข้อ	วิธีการตรวจให้คะแนน					
	การประยุกต์การให้คะแนนของคูมบ์		การประยุกต์การให้คะแนนของเดรสเชลและสมิท		แบบประเพณีนิยม(0-1)	
	อำนาจจำแนก	ความยาก	อำนาจจำแนก	ความยาก	อำนาจจำแนก	ความยาก
44	0.804	-1.003	0.390	0.411	0.776	1.145
45	0.715	-1.323	0.349	0.338	0.434	1.760
46	0.783	-1.252	0.410	0.365	0.683	0.964
47	1.244	-1.246	0.321	0.235	0.558	0.384
48	0.497	-0.740	0.246	0.206	0.249	2.001
49	0.815	-0.561	0.292	0.381	0.455	0.952
50	0.593	-0.575	0.205	0.077	0.814	0.171

จากตารางที่ 4.9 พบว่า ค่าอำนาจจำแนกรายข้อของวิธีประยุกต์การให้คะแนนของคูมบ์ ส่วนใหญ่จะมีค่าสูงกว่า วิธีการให้คะแนนแบบอื่นๆ ในขณะเดียวกันก็มีแนวโน้มของค่าความยากต่ำกว่า คือ มีความง่ายมากกว่าวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิทและวิธีประเพณีนิยม

ตารางที่ 4.10 ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐานของค่าความยากและค่าอำนาจจำแนก และค่าสถิติ F จำแนกตามวิธีการตรวจให้คะแนน

วิธีการให้คะแนน	ค่าพารามิเตอร์	Mean	S.D.	F-test	p-value
วิธีประยุกต์การให้คะแนนของคูมบ์	อำนาจจำแนก	0.646[a]	0.179	85.369	0.000
วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท		0.266[a/b]	0.063		
วิธีประเพณีนิยม		0.627[b]	0.210		
วิธีประยุกต์การให้คะแนนของคูมบ์วิธีประยุกต์	ความยาก	-1.049[c/d]	0.607	20.592	0.000
การให้คะแนนของเดรสเชลและสมิท		0.008[c]	0.715		
วิธีประเพณีนิยม		0.336[d]	1.754		

หมายเหตุ a,b,c,d คือ สัญลักษณ์แสดงคู่ที่มีค่าเฉลี่ยแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ0.05โดยที่

a เป็นความแตกต่างระหว่างวิธีประยุกต์การให้คะแนนของคูมบ์และวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท

b เป็นความแตกระหว่างวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิทและวิธีประเพณีนิยม

c เป็นความแตกต่างระหว่างวิธีประยุกต์การให้คะแนนของคูมบ์และวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท

d เป็นความแตกต่างระหว่างวิธีประยุกต์การให้คะแนนของคูมบ์และวิธีประเพณีนิยม

จากตารางที่ 4.10 เมื่อทดสอบความแตกต่างของค่าเฉลี่ยของค่าอำนาจจำแนก และค่าเฉลี่ยของค่าความยาก ด้วยสถิติ F พบว่า ค่าเฉลี่ยของค่าอำนาจจำแนกตามวิธีการตรวจให้คะแนนมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.05 (F =85.369 , p-value <.000) และเมื่อนำมาเปรียบเทียบความแตกต่างรายคู่ด้วยวิธีของ Scheffe พบว่าวิธีประยุกต์การให้คะแนนของคูมบ์ มีค่าอำนาจจำแนกเฉลี่ยสูงกว่าวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิทอย่างมีนัยสำคัญทางสถิติที่ระดับ0.05 ในขณะเดียวกันวิธีประเพณีนิยมมีค่าอำนาจจำแนกเฉลี่ยสูงกว่าวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิทอย่างมีนัยสำคัญทางสถิติ เช่นเดียวกัน

ส่วนค่าเฉลี่ยของค่าความยากตามวิธีการตรวจให้คะแนนมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.05 (F =20.592 , p-value <.000) และเมื่อนำมาเปรียบเทียบความแตกต่างรายคู่ด้วยวิธีของ Scheffe พบว่าวิธีประยุกต์การให้คะแนนของคูมบ์ มีค่าความยากเฉลี่ยต่ำกว่าวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิทและวิธีประเพณีนิยมอย่างมีนัยสำคัญทางสถิติ ในขณะเดียวกันวิธีประเพณีนิยมมีค่าความยากเฉลี่ยไม่แตกต่างจากวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิทอย่างมีนัยสำคัญทางสถิติ

ตอนที่ 4 ผลการเปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบสอบ และอัตราส่วนสารสนเทศ ระหว่างวิธีการตรวจให้คะแนนที่แตกต่างกัน

ผลการวิเคราะห์เปรียบเทียบในตอนนี้ ได้นำเสนอ เป็น 2 ส่วน คือ

4.1 ผลการเปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบตามวิธีการตรวจให้คะแนนที่แตกต่างกัน

4.2 ผลการเปรียบเทียบฟังก์ชันสารสนเทศของแบบสอบ และอัตราส่วนสารสนเทศ ตามวิธีการตรวจให้คะแนนที่แตกต่างกัน

โดยมีรายละเอียดของผลการวิเคราะห์ ดังนี้

4.1 ผลการเปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบตามวิธีการตรวจให้คะแนนที่แตกต่างกัน

การเปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบจำแนกตามรายข้อ (50ข้อ) และวิธีการให้ตรวจคะแนนทั้ง 3 วิธี แสดงในตารางที่ 4.11 ส่วนการเปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบจำแนกตามคะแนนความสามารถของผู้สอบ (9 ระดับความสามารถ) แสดงในตารางที่ 4.12 สำหรับการปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยตามวิธีการตรวจให้คะแนนทั้ง 3 วิธี โดยใช้สถิติทดสอบ F ผลวิเคราะห์ที่ได้แสดงในตารางที่ 4.13

ตารางที่ 4.11 ฟังก์ชันสารสนเทศของข้อสอบ จำแนกตามรายข้อและวิธีการตรวจให้คะแนน

ข้อ	ฟังก์ชันสารสนเทศของข้อสอบ		
	วิธีประยุกต์การให้คะแนนของคูมบ์	วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท	วิธีประเพณีนิยม
1	0.16	0.17	0.14
2	0.14	0.14	0.09
3	0.15	0.08	0.11
4	0.17	0.06	0.05
5	0.19	0.12	0.03
6	0.26	0.16	0.09
7	0.16	0.09	0.06
8	0.19	0.14	0.05
9	0.22	0.13	0.03
10	0.26	0.09	0.08
11	0.19	0.14	0.07
12	0.28	0.15	0.04
13	0.17	0.14	0.13
14	0.25	0.13	0.00
15	0.16	0.09	0.01
16	0.22	0.16	0.05
17	0.21	0.22	0.08
18	0.19	0.14	0.06
19	0.25	0.10	0.09
20	0.25	0.16	0.12
21	0.22	0.15	0.09
22	0.19	0.19	0.12
23	0.24	0.14	0.04
24	0.18	0.11	0.02
25	0.22	0.12	0.13

ตารางที่ 4.11 (ต่อ)

ข้อ	ฟังก์ชันสารสนเทศของข้อสอบ		
	วิธีประยุกต์การให้คะแนนของคูมบ์	วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท	วิธีประเพณีนิยม
26	0.18	0.15	0.05
27	0.27	0.19	0.18
28	0.19	0.10	0.04
29	0.20	0.12	0.07
30	0.21	0.15	0.10
31	0.25	0.12	0.06
32	0.26	0.17	0.04
33	0.19	0.17	0.12
34	0.30	0.14	0.03
35	0.24	0.16	0.04
36	0.17	0.14	0.02
37	0.30	0.12	0.02
38	0.40	0.17	0.06
39	0.22	0.13	0.05
40	0.41	0.20	0.11
41	0.21	0.17	0.05
42	0.16	0.20	0.15
43	0.24	0.19	0.06
44	0.30	0.22	0.09
45	0.27	0.20	0.03
46	0.29	0.23	0.08
47	0.47	0.18	0.06
48	0.18	0.14	0.01
49	0.31	0.16	0.03
50	0.22	0.11	0.12

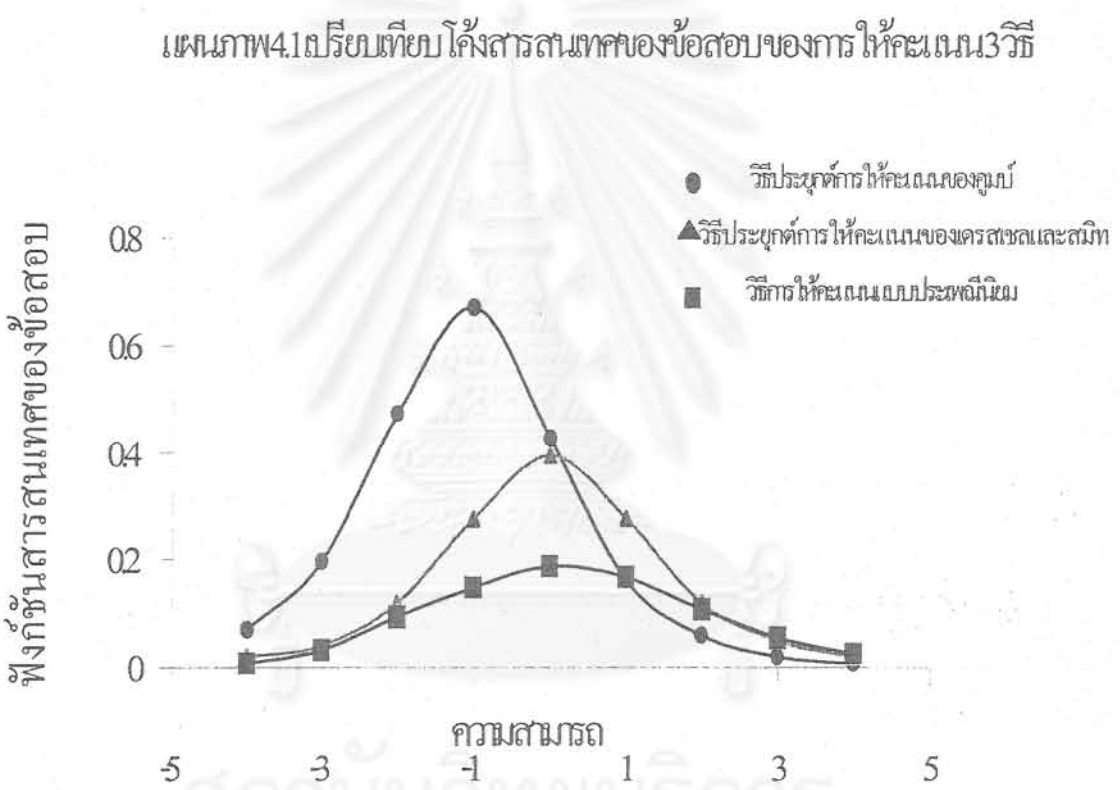
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

จากตารางที่ 4.11 พบว่าการตรวจให้คะแนนด้วยวิธีประยุกต์การให้คะแนนของคูมบ์ ส่วนใหญ่จะให้ค่าฟังก์ชันสารสนเทศของข้อสอบในแต่ละข้อสูงกว่าการให้คะแนนด้วยวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิทและวิธีประเพณีนิยม

ตารางที่ 4.12 ฟังก์ชันสารสนเทศของข้อสอบเฉลี่ย จำแนกตามความสามารถของผู้สอบและวิธีการตรวจให้คะแนน

ความสามารถ (θ)	วิธีการตรวจให้คะแนน	ฟังก์ชันสารสนเทศของข้อสอบเฉลี่ย				SE (θ)
		Min	Max	Mean	S.D.	
-4.00	วิธีประยุกต์การให้คะแนนของคูมบ์	.014	.465	.070	.071	3.77
	วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท	.000	.058	.020	.015	7.07
	วิธีประเพณีนิยม	.000	.051	.007	.011	11.95
-3.00	วิธีประยุกต์การให้คะแนนของคูมบ์	.045	.814	.195	.148	2.26
	วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท	.001	.161	.040	.033	5.00
	วิธีประเพณีนิยม	.000	.341	.032	.056	5.59
-2.00	วิธีประยุกต์การให้คะแนนของคูมบ์	.143	1.176	.473	.245	1.45
	วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท	.003	.403	.119	.076	2.89
	วิธีประเพณีนิยม	.000	.791	.094	.152	3.26
-1.00	วิธีประยุกต์การให้คะแนนของคูมบ์	.011	2.551	.673	.461	1.21
	วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท	.010	.830	.275	.152	1.90
	วิธีประเพณีนิยม	.015	.566	.148	.127	2.59
0.00	วิธีประยุกต์การให้คะแนนของคูมบ์	.050	.954	.427	.196	1.53
	วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท	.034	.933	.395	.208	1.59
	วิธีประเพณีนิยม	.029	.521	.187	.137	2.31
1.00	วิธีประยุกต์การให้คะแนนของคูมบ์	.016	.497	.162	.095	2.48
	วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท	.062	.696	.276	.149	1.90
	วิธีประเพณีนิยม	.002	.454	.167	.119	2.44
2.00	วิธีประยุกต์การให้คะแนนของคูมบ์	.005	.162	.060	.040	4.08
	วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท	.025	.276	.119	.062	2.89
	วิธีประเพณีนิยม	.001	.336	.107	.085	3.05
3.00	วิธีประยุกต์การให้คะแนนของคูมบ์	.000	.073	.020	.016	7.07
	วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท	.007	.428	.050	.599	4.47
	วิธีประเพณีนิยม	.000	.209	.055	.050	4.26
4.00	วิธีประยุกต์การให้คะแนนของคูมบ์	.000	.032	.007	.007	11.95
	วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท	.002	.308	.020	.426	7.07
	วิธีประเพณีนิยม	.000	.103	.025	.025	6.32

จากตารางที่ 4.12 และแผนภาพที่ 4.1 เมื่อเปรียบเทียบค่าฟังก์ชันสารสนเทศของข้อสอบจำแนกตามค่าความสามารถของผู้สอบตามวิธีการให้คะแนน พบว่า วิธีประยุกต์การให้คะแนนของคูมบ์ ให้ค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงที่สุดในระดับความสามารถ – 1 และมีค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงกว่าวิธีอื่นๆ มีค่าเท่ากับ 0.673 ในขณะที่วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท ให้ค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงที่สุดในระดับความสามารถ 0 มีค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยเท่ากับ 0.395 ส่วนวิธีประเพณีนิยมให้ค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงที่สุดในระดับความสามารถ 0 มีค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยเท่ากับ 0.187



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.13 เปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบโดยเฉลี่ยจำแนกตามวิธีการตรวจให้คะแนน

วิธีการตรวจให้คะแนน	ฟังก์ชันสารสนเทศของข้อสอบ		F- test	p-value
	Mean	S.D.		
วิธีประยุกต์การให้คะแนนของคูมบ์	0.2316 [a/b]	.066	123.56	0.000
วิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท	0.1467 [a/c]	.039		
วิธีประเพณีนิยม	0.0727 [b/c]	.040		

หมายเหตุ a,b,c คือ สัญลักษณ์แสดงคู่ที่มีค่าเฉลี่ยแตกต่างกัน โดยที่

a เป็นความแตกต่างระหว่างวิธีประยุกต์การให้คะแนนของคูมบ์และวิธีประยุกต์ การให้คะแนนของเดรสเซลและสมิท

b เป็นความแตกต่างระหว่างวิธีประยุกต์การให้คะแนนของคูมบ์และวิธีประเพณีนิยม

c เป็นความแตกต่างระหว่างวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิทและวิธีประเพณีนิยม

จากตารางที่ 4.13 พบว่า วิธีประยุกต์การให้คะแนนของคูมบ์มีค่าฟังก์ชันสารสนเทศของข้อสอบโดยเฉลี่ยสูงกว่าวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิทและวิธีประเพณีนิยมอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ส่วนวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท มีค่าฟังก์ชันสารสนเทศของข้อสอบโดยเฉลี่ยสูงกว่าวิธีประเพณีนิยมอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05

4.2 ผลการเปรียบเทียบฟังก์ชันสารสนเทศของแบบสอบ และอัตราส่วนสารสนเทศ ตามวิธีการให้คะแนนที่แตกต่างกัน

การเปรียบเทียบฟังก์ชันสารสนเทศของแบบสอบจำแนกวิธีการให้คะแนนทั้ง 3 วิธี แสดงในตารางที่ 4.14 ส่วนการเปรียบเทียบอัตราส่วนสารสนเทศ ซึ่งเป็นค่าประสิทธิภาพสัมพัทธ์ของแบบสอบแสดงในตารางที่ 4.15

ตารางที่ 4.14 ฟังก์ชันสารสนเทศของแบบสอบ จำแนกตามความวิธีการตรวจให้คะแนน

วิธีการตรวจให้คะแนน	จำนวนผู้สอบ	ฟังก์ชันสารสนเทศของแบบสอบ
วิธีประยุกต์การให้คะแนนของคูมบ์	252	11.58
วิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท	264	7.38
วิธีประเพณีนิยม	261	3.63

จากตารางที่ 4.14 พบว่า วิธีประยุกต์การให้คะแนนของคูมบ์ ให้ค่าฟังก์ชันสารสนเทศของแบบสอบ สูงกว่าวิธีอื่นๆ รองลงมาคือ วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท ส่วนวิธีประเพณีนิยมให้ค่าฟังก์ชันสารสนเทศของแบบสอบต่ำที่สุด

ตารางที่ 4.15 อัตราส่วนสารสนเทศเฉลี่ยจำแนกตามความวิธีการตรวจให้คะแนน

การเปรียบเทียบระหว่าง		อัตราส่วนสารสนเทศเฉลี่ย
วิธีประยุกต์การให้คะแนนของคูมบ์	วิธีประเพณีนิยม	3.19
วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท	วิธีประเพณีนิยม	2.03
วิธีประยุกต์การให้คะแนนของคูมบ์	วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท	1.56

จากตารางที่ 4.15 พบว่าอัตราส่วนสารสนเทศเฉลี่ยระหว่างวิธีประยุกต์การให้คะแนนของคูมบ์กับวิธีประเพณีนิยม มีค่าสูงสุด รองลงมาคืออัตราส่วนสารสนเทศเฉลี่ยระหว่างวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิทกับวิธีประเพณีนิยม ซึ่งแสดงให้เห็นว่า วิธีประยุกต์การให้คะแนนของคูมบ์มีประสิทธิภาพสัมพัทธ์สูงที่สุด รองลงมา คือ วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท ส่วนวิธีประเพณีนิยมมีประสิทธิภาพสัมพัทธ์ต่ำที่สุด

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 5

บทสรุป อภิปรายและข้อเสนอแนะ

การวิจัยในเรื่องนี้ ต้องการศึกษาวิธีการตอบข้อสอบแบบเลือกตอบที่มีวิธีการให้คะแนนที่แตกต่างจากวิธีประเพณีนิยม เพื่อช่วยแก้จุดอ่อนของวิธีการให้คะแนนแบบประเพณีนิยมในเรื่องเปิดโอกาสให้มีการเดาสุ่มได้ และเพิ่มการให้คะแนนความรู้บางส่วน ซึ่งวิธีการให้คะแนนที่เลือกศึกษาในครั้งนี้ มี 2 วิธี ได้แก่ วิธีที่ประยุกต์มาจากวิธีการให้คะแนนของคูมบ์ และวิธีที่ประยุกต์การให้คะแนนของเดรสเชลและสมิท โดยผู้วิจัยได้มีหลักการในการประยุกต์การให้คะแนน คือ ให้คะแนนต่ำสุด คือ 0 และสูงสุด คือ 3 (ในกรณีที่มี4 ตัวเลือก) ในขณะที่วิธีการให้คะแนนของคูมบ์ และวิธีการให้คะแนนของเดรสเชลและสมิท ให้คะแนนสูงสุด 3 ต่ำสุด –3 ซึ่งเป็นค่าติดลบ ที่ผู้วิจัยเห็นว่าไม่เหมาะสม เนื่องจากคะแนนในข้อติดลบจะถูกนำไปหักจากคะแนนที่ได้ในข้ออื่นๆได้ ทำให้ไม่สามารถวัดความรู้ในแต่ละข้อที่แท้จริงและในภาพรวมได้

ดังนั้น การลดโอกาสในการเดาสุ่ม และเพิ่มสารสนเทศเกี่ยวกับการตอบข้อสอบให้มากที่สุดโดยใช้วิธีการให้คะแนนที่เหมาะสม จะทำให้คุณสมบัติของการวัดในประเด็นที่เกี่ยวข้องกับความตรงและความเที่ยงสูงขึ้น เนื่องจากการให้คะแนนความรู้บางส่วนของผู้ตอบแสดงถึงตำแหน่งจริงของความรู้ของผู้ตอบ ซึ่งมีส่วนช่วยลดความคลาดเคลื่อนจากแหล่งการเดาสุ่มของผู้ตอบลง ทำให้คะแนนมีความเที่ยงมากขึ้น ในขณะเดียวกันก็ส่งผลต่อความตรงของกระบวนการวัดด้วย (Frary ,1980 อ้างถึงใน พรทิพย์ ไชยโส , 2533) แต่จากการศึกษางานวิจัยในอดีตที่ผ่านมา ส่วนใหญ่มักใช้ทฤษฎีการวัดแบบดั้งเดิมในการศึกษาคุณภาพของแบบสอบ ซึ่งมีข้อจำกัดในเรื่องแหล่งความคลาดเคลื่อนในการวัด และค่าสถิติที่ได้แปรเปลี่ยนไปตามกลุ่มตัวอย่างที่ใช้ ในปัจจุบันนักวัดผลได้พัฒนาทฤษฎีการตอบสนองข้อสอบ (Item Response Theory, IRT) เพื่อแก้ปัญหาการวิเคราะห์ตามแนวทฤษฎีการทดสอบแบบดั้งเดิม ซึ่งสามารถนำไปวิเคราะห์กับแบบสอบที่ตรวจให้คะแนนแบบทวิวิภาค และแบบสอบที่ตรวจให้คะแนนแบบพหุวิภาค ในการวิจัยครั้งนี้ ผู้วิจัยจึงเปรียบเทียบคุณภาพของแบบสอบ โดยใช้ทฤษฎีการตอบสนองข้อสอบ (IRT) โดยที่คาดว่าผลที่ได้จากการวิจัยครั้งนี้ จะทำให้ได้ข้อความรู้ที่จะช่วยให้ผู้ที่จะใช้เครื่องมือในการวัดผลการศึกษาได้ตระหนักถึงวิธีการให้คะแนนที่มีคุณภาพในกระบวนการวัดเพื่อนำไปใช้ในการตัดสินคุณค่าในการประเมินต่อไป การวิจัยในเรื่องนี้จึงมีวัตถุประสงค์ 3 ประการ คือ

ประการที่ 1 เพื่อเปรียบเทียบคุณภาพของแบบสอบในด้านความตรงตามสภาพ ระหว่างวิธีการตรวจให้คะแนนแบบประเพณีนิยมกับวิธีการตรวจให้คะแนนความรู้บางส่วนตามวิธีประยุกต์การให้คะแนนของคูมบ์ และวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท

ประการที่ 2 เพื่อเปรียบเทียบความยากและอำนาจจำแนกโดยใช้ทฤษฎีการตอบสนองข้อสอบระหว่างวิธีการตรวจให้คะแนนแบบประเพณีนิยมกับวิธีการตรวจให้คะแนนความรู้บางส่วนตามวิธีประยุกต์การให้คะแนนของคูมบ์ และวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท

ประการสุดท้าย เพื่อเปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบสอบและอัตราส่วนสารสนเทศเฉลี่ยโดยใช้ทฤษฎีการตอบสนองข้อสอบระหว่างวิธีการตรวจให้คะแนนแบบประเพณีนิยมกับวิธีการตรวจให้คะแนนความรู้บางส่วนตามวิธีประยุกต์การให้คะแนนของคูมบ์และวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท

ประชากรและกลุ่มตัวอย่างที่ใช้ในการวิจัยได้แก่ นิสิตคณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ชั้นปีที่ 3 ที่ลงทะเบียนเรียน วิชาการประเมินผลการเรียนการสอน ในภาคการศึกษาต้น ปีการศึกษา 2544 จำนวน 297 คน เก็บข้อมูลโดยให้นิสิตทำข้อสอบวัดผลสัมฤทธิ์ทางการเรียน วิชาการประเมินผลการเรียนการสอนที่คัดเลือกข้อสอบที่มีคุณภาพจากคลังข้อสอบที่ผ่านการวิเคราะห์ข้อสอบแล้ว ซึ่งเป็นข้อสอบแบบเลือกตอบที่มี 4 ตัวเลือก โดยมีตัวถูกเพียง 1 ตัว ในแต่ละข้อ จำนวน 50 ข้อ โดยมีค่าความยากและค่าอำนาจจำแนกอยู่ในเกณฑ์ที่มีคุณภาพ และมีค่าความตรงและความเที่ยงสูง โดยให้นิสิตทำแบบสอบ 3 ครั้ง ด้วยแบบสอบชุดเดียวกัน แต่วิธีการตอบแตกต่างกัน 3 วิธี เก็บข้อมูลแต่ละครั้งห่างกัน 2 สัปดาห์ ดังนี้ ครั้งแรก ให้นิสิตทำข้อสอบวิธีประยุกต์การให้คะแนนของคูมบ์ (แบบตัดตัวลวง) ครั้งที่สอง ให้นิสิตทำข้อสอบวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท(แบบเลือกชุดตัวถูก) ครั้งที่สาม ให้นิสิตทำข้อสอบแบบประเพณีนิยม คือให้เลือกคำตอบเพียง 1 ตัว

การวิเคราะห์ข้อมูล โดยวิเคราะห์ความตรงตามสภาพ จากค่าสัมประสิทธิ์สหสัมพันธ์ (r) เปรียบเทียบความเป็นอิสระของความตรงตามสภาพในภาพรวม โดยใช้สถิติทดสอบไคว์สแควร์ (χ^2) และเปรียบเทียบความแตกต่างของค่าความตรงตามสภาพเป็นรายคู่ โดยใช้สถิติทดสอบ Z ส่วนการวิเคราะห์ความเที่ยง วิเคราะห์จากฟังก์ชันสารสนเทศของแบบสอบ อัตราส่วนสารสนเทศเฉลี่ย และฟังก์ชันสารสนเทศของข้อสอบโดยวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ(Item Response Theory) สำหรับการให้คะแนนแบบประเพณีนิยม มีการประมาณค่าพารามิเตอร์ความยาก (b) อำนาจจำแนก (a) ค่าฟังก์ชันสารสนเทศของข้อสอบและค่าฟังก์ชันสารสนเทศของแบบสอบ(TIF) โดยใช้โปรแกรม BILOG ส่วนการให้คะแนนด้วยวิธีประยุกต์การให้คะแนนของคูมบ์ และวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท ประมาณค่าพารามิเตอร์ความยาก (b) อำนาจจำแนก(a) ค่าฟังก์ชันสารสนเทศของข้อสอบและค่าฟังก์ชันสารสนเทศของแบบสอบ(TIF) โดยใช้โปรแกรม PARSCALE สำหรับการเปรียบเทียบความแตกต่างของค่าเฉลี่ยของค่าอำนาจจำแนก ความยาก ค่าฟังก์ชันสารสนเทศของข้อสอบและค่าฟังก์ชันสารสนเทศของแบบสอบ(TIF) ใช้สถิติ F- test และเปรียบเทียบความแตกต่างรายคู่ ด้วยวิธีการของ Scheffe โดยใช้โปรแกรม SPSS for Windows version 10.0

สรุปผลการวิจัย

ผลการวิจัยที่ได้แยกเป็น 2 ตอน คือ ตอนที่ 1 สรุปผลการวิเคราะห์ข้อมูลเบื้องต้น และตอนที่ 2 คือ สรุปผลเพื่อตอบคำถามการวิจัยตามวัตถุประสงค์ ทั้ง 3 ประการ

ตอนที่ 1 ผลการวิเคราะห์ข้อมูลเบื้องต้น

ผลการวิเคราะห์ข้อมูลเบื้องต้น พบว่าการให้คะแนนวิธีประเพณีนิยมและวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท มีค่าเฉลี่ยสูงกว่าวิธีประยุกต์การให้คะแนนของคูมบ์ ในขณะที่ส่วนเบี่ยงเบนมาตรฐานการให้คะแนน วิธีประยุกต์การให้คะแนนของคูมบ์ มีค่าสูงกว่าวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท และวิธีประเพณีนิยม เมื่อเปรียบเทียบความแตกต่างของค่าเฉลี่ยเป็นรายคู่พบว่าวิธีประเพณีนิยมมีคะแนนเฉลี่ยสูงกว่าวิธีวิธีประยุกต์การให้คะแนนของคูมบ์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 และในขณะเดียวกันวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิทก็มีคะแนนเฉลี่ยสูงกว่าวิธีประยุกต์การให้คะแนนของคูมบ์อย่างมีนัยสำคัญทางสถิติที่ระดับ0.05

ส่วนค่าเฉลี่ยความสามารถ พบว่าวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิทมีค่าเฉลี่ยความสามารถของผู้สอบสูงกว่าวิธีอื่นๆ แต่เมื่อพิจารณาจากค่าส่วนเบี่ยงเบนมาตรฐานพบว่าคะแนนความสามารถของผู้สอบมีการกระจายน้อยกว่ากลุ่มอื่นๆ ในขณะที่วิธีประยุกต์การให้คะแนนของคูมบ์มีค่าเฉลี่ยความสามารถของผู้สอบต่ำกว่าวิธีอื่นๆแต่จากค่าส่วนเบี่ยงเบนมาตรฐานพบว่าคะแนนความสามารถของผู้สอบมีการกระจายมากกว่ากลุ่มอื่นๆ เมื่อเปรียบเทียบความแตกต่างของค่าเฉลี่ยความสามารถเป็นรายคู่พบวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิทก็มีคะแนนความสามารถเฉลี่ยสูงกว่าวิธีประยุกต์การให้คะแนนของคูมบ์อย่างมีนัยสำคัญทางสถิติที่ระดับ0.05 ส่วนคู่อื่นๆไม่พบว่ามีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ

เมื่อวิเคราะห์ถึงรูปแบบการตอบของกลุ่มตัวอย่างในการแสดงความรู้บางส่วน ด้วยวิธีประยุกต์การให้คะแนนของคูมบ์ และวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท จำแนกตามกลุ่มที่มีความสามารถสูง ปานกลาง ต่ำ สรุปได้ดังนี้

รูปแบบการตัดตัวลวง(วิธีประยุกต์การให้คะแนนของคูมบ์) พบว่ากลุ่มความสามารถสูง จะนิยมตัดตัวเลือก 3 ตัว (ข้อสอบ 4 ตัวเลือก) มากกว่ากลุ่มอื่นๆ คิดเป็นร้อยละของค่าเฉลี่ยเท่ากับ 69.35 ในขณะที่การตัดตัวเลือก 1 –2 ตัว มีร้อยละของค่าเฉลี่ยสูงสุดในกลุ่มความสามารถปานกลาง มีค่าเท่ากับ 28.95 รองลงมาคือกลุ่มความสามารถต่ำ มีค่าเท่ากับ28.08 ส่วนการไม่ตอบมีร้อยละของค่าเฉลี่ยสูงสุดในกลุ่มความสามารถต่ำ มีค่าเท่ากับ 30.44

รูปแบบการเลือกชุดตัวถูก(วิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท) พบว่า กลุ่มความสามารถสูง จะเลือกตัวถูกเพียง 1 ตัว (ข้อสอบ 4 ตัวเลือก) มากกว่ากลุ่มอื่นๆ คิดเป็นร้อยละของค่าเฉลี่ยเท่ากับ 79.38 ในขณะที่การตัดตัวเลือก 1 –2 ตัว มีร้อยละของค่าเฉลี่ยในแต่ละกลุ่มไม่แตกต่างกัน คืออยู่ระหว่าง 11.37-12.06 ส่วนการไม่ตอบมีร้อยละของค่าเฉลี่ยสูงสุดในกลุ่มความสามารถต่ำ มีค่าเท่ากับ 32.13

ตอนที่ 2 สรุปผลเพื่อตอบคำถามการวิจัยตามวัตถุประสงค์

2.1 ผลการเปรียบเทียบความตรงตามสภาพ ระหว่างวิธีการตรวจให้คะแนนที่แตกต่างกัน

จากการศึกษาความตรงตามสภาพตามวิธีประยุกต์การให้คะแนนของคูมบ์ วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิทและวิธีประเพณีนิยม แล้วนำมาเปรียบเทียบความแตกต่างโดยใช้คะแนนฟิชเชอร์ซี เปรียบเทียบในภาพรวมโดยใช้สถิติทดสอบไคว์สแควร์ (χ^2) พบว่าค่าความตรงตามสภาพไม่ขึ้นอยู่กับวิธีการตรวจให้คะแนนในแต่ละวิธี และเมื่อนำมาเปรียบเทียบความแตกต่างกันเป็นรายคู่ ด้วยสถิติทดสอบซี (Z) พบว่า ไม่มีความแตกต่างกันของความตรงสภาพในแต่ละคู่อย่างมีนัยสำคัญทางสถิติที่ระดับ0.05

2.2 ผลการเปรียบเทียบความยากและอำนาจจำแนก ระหว่างวิธีการตรวจให้คะแนนที่แตกต่างกัน

ผลการเปรียบเทียบความแตกต่างของค่าเฉลี่ยของค่าอำนาจจำแนกและค่าเฉลี่ยของค่าความยากด้วยสถิติ F พบว่า ค่าเฉลี่ยของค่าอำนาจจำแนกตามวิธีการตรวจให้คะแนนมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.05 (F =85.369 , p-value <.000) และเมื่อนำมาเปรียบเทียบความแตกต่างรายคู่ด้วยวิธีของ Scheffe พบว่าวิธีประยุกต์การให้คะแนนของคูมบ์ มีค่าอำนาจจำแนกเฉลี่ยสูงกว่าวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิทอย่างมีนัยสำคัญทางสถิติที่ระดับ0.05 ในขณะเดียวกันวิธีประเพณีนิยมมีค่าอำนาจจำแนกเฉลี่ยสูงกว่าวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิทอย่างมีนัยสำคัญทางสถิติ เช่นเดียวกัน

ส่วนค่าเฉลี่ยของความยากตามวิธีการตรวจให้คะแนนมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.05 (F =20.592 , p-value <.000) และเมื่อนำมาเปรียบเทียบความแตกต่างรายคู่ด้วยวิธีของ Scheffe พบว่าวิธีประยุกต์การให้คะแนนของคูมบ์ มีค่าความยากเฉลี่ยต่ำกว่าวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิทและวิธีประเพณีนิยมอย่างมีนัยสำคัญทางสถิติ ในขณะเดียวกันวิธีประเพณีนิยมมีค่าความยากเฉลี่ยไม่แตกต่างจากวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิทอย่างมีนัยสำคัญทางสถิติ

2.3 ผลการเปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบสอบและอัตราส่วนสารสนเทศ ระหว่างวิธีการตรวจให้คะแนนที่แตกต่างกัน

วิธีประยุกต์การให้คะแนนของคูมบ์มีค่าฟังก์ชันสารสนเทศของข้อสอบโดยเฉลี่ยสูงกว่าวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิทและวิธีประเพณีนิยมอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ส่วนวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท มีค่าฟังก์ชันสารสนเทศของข้อสอบโดยเฉลี่ยสูงกว่าวิธีประเพณีนิยมอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 เมื่อเปรียบเทียบค่าฟังก์ชันสารสนเทศของข้อสอบจำแนกตามค่าความสามารถของผู้สอบตามวิธีการตรวจให้คะแนน พบว่า วิธีประยุกต์การให้คะแนนของคูมบ์ ให้ค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงที่สุดในระดับความสามารถ – 1 และมีค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยเท่ากับ 0.673 ซึ่งสูงกว่าวิธีอื่นๆ ในขณะที่วิธี

ประยุกต์การให้คะแนนของเดรสเชลและสมิท ให้ค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงที่สุดในระดับความสามารถ 0 มีค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยเท่ากับ 0.395 ส่วนวิธีประเพณีนิยมให้ค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยสูงที่สุดในระดับความสามารถ 0 มีค่าฟังก์ชันสารสนเทศของข้อสอบเฉลี่ยเท่ากับ 0.187

วิธีประยุกต์การให้คะแนนของคูมบ์ ให้ค่าฟังก์ชันสารสนเทศของแบบสอบ สูงกว่าวิธีอื่นๆ รองลงมาคือ วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท ส่วนวิธีประเพณีนิยมให้ค่าฟังก์ชันสารสนเทศของแบบสอบต่ำที่สุด ส่วนอัตราส่วนสารสนเทศเฉลี่ยระหว่างวิธีประยุกต์การให้คะแนนของคูมบ์กับวิธีประเพณีนิยม มีค่าสูงสุด รองลงมาคืออัตราส่วนสารสนเทศเฉลี่ยระหว่างวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิทกับวิธีประเพณีนิยม ซึ่งแสดงให้เห็นว่า วิธีประยุกต์การให้คะแนนของคูมบ์มีประสิทธิภาพสัมพัทธ์สูงที่สุด รองลงมา คือ วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท ส่วนวิธีประเพณีนิยมมีประสิทธิภาพสัมพัทธ์ต่ำที่สุด

อภิปรายผลการวิจัย

1. การเปรียบเทียบความตรงตามสภาพ ระหว่างวิธีการตรวจให้คะแนนที่แตกต่างกัน

จากการศึกษาความตรงตามสภาพ จากวิธีให้คะแนนความรู้บางส่วนตามวิธีประยุกต์การให้คะแนนของคูมบ์ วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิทและวิธีการให้คะแนนแบบประเพณีนิยม เมื่อนำมาทดสอบความแตกต่างทางสถิติ โดยใช้ Z-test พบว่าไม่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ซึ่งไม่สอดคล้องกับข้อสรุปของงานวิจัยหลายเรื่องที่ได้ข้อสรุปว่า วิธีให้คะแนนแบบการตัดตัวลวง (ET) ให้ค่าความตรงสูงกว่า วิธีแบบเลือกตอบธรรมดา (Collet, 1971; Coombs et al, 1956; Jaradat and Tollefson, 1988) และวิธีเลือกชุดตัวถูกส่วนใหญ่มีแนวโน้มที่จะให้ค่าความตรงตามสภาพสูงกว่าวิธีประเพณีนิยม (Jaradat and Tollefson, 1986) ซึ่งตามหลักการทั้งวิธีตัดตัวลวงและวิธีการเลือกชุดตัวถูกเป็นวิธีที่เปิดโอกาสให้ผู้สอบได้คะแนนบางส่วน ทำให้ได้ค่าพิสัยและการกระจายของคะแนนกว้างกว่าวิธีประเพณีนิยมที่คะแนนที่ได้จะเป็นแบบ 0 - 1 ทำให้ได้ค่าความตรงตามสภาพของวิธีตัดตัวลวงและวิธีเลือกชุดตัวถูกสูงกว่าแบบประเพณีนิยม แต่จากการทดสอบทางสถิติเพื่อเปรียบเทียบค่าความตรงตามสภาพของ 2 กลุ่มนี้พบว่าไม่มีความแตกต่างกัน ทั้งนี้อาจเนื่องจากเกณฑ์ที่นำใช้ คือ แบบสอบกลางภาคนั้นเป็นแบบสอบที่มีการให้คะแนนแบบประเพณีนิยม ทำให้มีความสัมพันธ์สูงกับวิธีประเพณีนิยม เนื่องจากเป็นการตรวจให้คะแนนในลักษณะเดียวกัน จึงทำให้ค่าความตรงตามสภาพของวิธีการตรวจให้คะแนนความรู้บางส่วนตามวิธีประยุกต์การให้คะแนนของคูมบ์ ถึงแม้จะมีค่าสูงกว่าในกลุ่มตัวอย่าง แต่ไม่มีนัยสำคัญทางสถิติ

2. การเปรียบเทียบอำนาจจำแนกและ ความยากระหว่างวิธีการตรวจให้คะแนนที่แตกต่างกัน

จากการศึกษาครั้งนี้ วิเคราะห์ค่าอำนาจจำแนกและค่าความยาก โดยใช้ทฤษฎีการตอบสนองข้อสอบ มีการประมาณค่าพารามิเตอร์อำนาจจำแนก (a) ความยาก (b) ของการให้คะแนนด้วยวิธีประยุกต์

การให้คะแนนของคูมบ์ และวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท ซึ่งเป็นการให้คะแนนแบบพหุวิภาค (polytomous) โดยใช้โปรแกรม PARSCALE ส่วนการประมาณค่าพารามิเตอร์อำนาจจำแนก (a) ความยาก (b) สำหรับการให้คะแนนแบบประเพณีนิยม ซึ่งเป็นการให้คะแนนแบบทวิวิภาค (dichotomous) ใช้โปรแกรม BILOG นั้น ผลที่ได้ คือวิธีประยุกต์การให้คะแนนของคูมบ์ มีค่าอำนาจจำแนกเฉลี่ยสูงที่สุด รองลงมาวิธีประเพณีนิยม ส่วนวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิทมีค่าอำนาจจำแนกเฉลี่ยต่ำกว่า 2 วิธีข้างต้นอย่างมีนัยสำคัญทางสถิติ ส่วนค่าความยาก พบว่าวิธีประยุกต์การให้คะแนนของคูมบ์ มีค่าความยากเฉลี่ยต่ำที่สุด รองลงมาวิธีประเพณีนิยม ส่วนวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิทมีค่าความยากสูงที่สุด

การที่ค่าอำนาจจำแนกของการตรวจให้คะแนนด้วยวิธีประยุกต์การให้คะแนนของคูมบ์ (แบบตัดตัวลวง) มีค่าสูงกว่าวิธีประเพณีนิยมนั้น สามารถอธิบายโดยใช้หลักเหตุผลได้ว่า วิธีการให้คะแนนด้วยวิธีประยุกต์การให้คะแนนของคูมบ์ได้ข้อมูลของผู้สอบมาพิจารณาความสามารถของผู้สอบมากกว่าวิธีประเพณีนิยม ทำให้ได้คะแนนที่ประมาณค่าความสามารถของผู้สอบได้ใกล้เคียงกับความสามารถที่แท้จริงมากกว่า และทำให้คุณภาพของการวัดสูงขึ้น ซึ่งตรงกับผลการสรุปของงานวิจัยอื่นที่สรุปว่าการที่ให้คะแนนการตอบหลายอย่าง (multiple response) จะให้คะแนนที่มีความน่าเชื่อถือมากกว่าคำตอบที่ให้คะแนนเพียงตัวเลือกที่เป็นคำตอบถูกเพียงตัวเดียว (Frary ,1982)

แต่การให้คะแนนด้วยวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท มีค่าอำนาจจำแนกเฉลี่ยต่ำกว่าวิธีประเพณีนิยม และวิธีประยุกต์การให้คะแนนของคูมบ์ ทั้งๆที่เป็นวิธีที่ได้ข้อมูลของผู้สอบมาพิจารณาความสามารถของผู้สอบมากกว่าวิธีประเพณีนิยมนั้น จากการวิเคราะห์รูปแบบการตอบของการวิจัยครั้งนี้โดยเปรียบเทียบวิธีการแบบตัดตัวลวงที่มีการประยุกต์การให้คะแนนของคูมบ์ และแบบเลือกชุดตัวถูกที่มีการประยุกต์การให้คะแนนของเดรสเซลและสมิทของกลุ่มตัวอย่าง พบว่า กลุ่มตัวอย่างส่วนใหญ่มักไม่นิยมตอบโดยเลือกตัวเลือกของรูปแบบการเลือกชุดตัวถูก (เลือกชุดตัวถูก 2 – 3 ตัว) กลุ่มที่เลือกมีน้อย คิดเป็นร้อยละ 11.37 – 12.06 ส่วนใหญ่จะเลือกชุดตัวถูกเพียง 1 ตัว ถ้าทำถูกได้ 3 คะแนน ทำผิดได้ 0 คะแนน ซึ่งการเลือกลักษณะนี้ไม่ต่างจากวิธีประเพณีนิยม อาจเกิดการเดาสุ่มได้ นอกจากนี้ในกรณีที่กลุ่มตัวอย่างไม่แน่ใจในคำตอบก็จะไม่ตอบ โดยเฉพาะในกลุ่มที่มีความสามารถต่ำคิดเป็นร้อยละ 32.13 ดังนั้น รูปแบบการตอบดังกล่าวมีผลทำให้มีการกระจายของคะแนนต่ำและคะแนนที่ได้อาจเกิดจากการเดาสุ่มมากกว่าวิธีการแบบตัดตัวลวงที่มีการประยุกต์การให้คะแนนของคูมบ์ และการที่วิธีเลือกชุดตัวถูกที่มีการประยุกต์การให้คะแนนของเดรสเซลและสมิท มีการกระจายของคะแนนต่ำกว่าวิธีประเพณีนิยม อาจเนื่องจากการตอบของแบบประเพณีนิยม กลุ่มตัวอย่างจะตอบทุกข้อ ทำให้ได้คะแนนบ้างจากการเดาสุ่ม แต่วิธีเลือกชุดตัวถูกที่มีการประยุกต์การให้คะแนนของเดรสเซลและสมิท ถ้าผู้ตอบไม่แน่ใจเลยก็จะไม่ตอบ ทำให้ข้อนั้นไม่ได้คะแนน จึงเป็นส่วนทำให้คะแนนแบบประเพณีนิยมมีค่าพิสัยและส่วนเบี่ยงเบนมาตรฐานสูงกว่าวิธีเลือกชุดตัวถูกที่มีการประยุกต์การให้คะแนนของเดรสเซลและสมิท และมีผลให้ค่าอำนาจจำแนกของวิธีประเพณีนิยมสูงกว่าวิธีเลือกชุดตัวถูกที่มีการประยุกต์

การให้คะแนนของเดรสเซลและสมิท ซึ่งมีผลทำให้พิสัยของคะแนนดิบ ค่าความสามารถของผู้สอบ (θ) และการส่วนเบี่ยงเบนมาตรฐาน ของการให้คะแนนด้วยวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท ต่ำกว่าวิธีประเพณีนิยม และวิธีประยุกต์การให้คะแนนของคูมบ์ โดยวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท มีช่วงค่าความสามารถอยู่ระหว่าง–1.087 - 1.332 ในขณะที่วิธีประเพณีนิยมมีช่วงค่าความสามารถอยู่ระหว่าง –2.043 - 1.959 และวิธีประยุกต์การให้คะแนนของคูมบ์ มีช่วงค่าความสามารถอยู่ระหว่าง –3.171 – 2.153

ส่วนการที่ค่าความยากของการตรวจให้คะแนนด้วยวิธีประยุกต์การให้คะแนนของคูมบ์ (แบบตัดตัวลวง) มีค่าต่ำกว่าวิธีประเพณีนิยม และวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท นั้น สามารถอธิบายโดยใช้หลักเหตุผลได้ว่า วิธีการให้คะแนนด้วยวิธีประยุกต์การให้คะแนนของคูมบ์ เปิดโอกาสให้ผู้สอบได้คะแนนในกรณีที่มีความรู้บางส่วนได้ โดยไม่ต้องเดา ทำให้กลุ่มที่มีความสามารถต่ำและปานกลางมีโอกาสที่จะได้คะแนนจากข้อสอบแต่ละข้อบ้าง ซึ่งอาจเป็นเหตุทำให้ข้อสอบมีความง่ายกว่าวิธีประเพณีนิยม ส่วนการที่ข้อสอบที่มีการให้คะแนนด้วยวิธีประยุกต์การให้คะแนนของคูมบ์มีความง่ายกว่าวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท นั้นอาจเป็นเพราะผู้สอบคุ้นเคยที่จะเลือกทางเลือกที่เป็นแบบตัดตัวลวงของวิธีประยุกต์การให้คะแนนของคูมบ์ มากกว่าทางเลือกที่เป็นเลือกชุดตัวถูกของวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท ทำให้การได้คะแนนจากวิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท ซึ่งทำให้กลุ่มที่มีความสามารถต่ำและปานกลางไม่มีโอกาสที่จะได้คะแนนบางส่วนจากข้อสอบ แต่ได้คะแนนเต็มเมื่อตอบถูกและได้คะแนน 0 เมื่อตอบผิด ซึ่งไม่ต่างจากวิธีประเพณีนิยม ที่คะแนนที่ได้อาจเกิดจากการเดามากกว่าจะเป็นคะแนนความสามารถที่แท้จริง

3. การเปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบสอบ และอัตราส่วนสารสนเทศ ระหว่างวิธีการตรวจให้คะแนนที่แตกต่างกัน

จากการเปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบสอบ และอัตราส่วนสารสนเทศ ระหว่างวิธีการตรวจให้คะแนนที่แตกต่างกันพบว่า วิธีประยุกต์การให้คะแนนของคูมบ์ มีค่าฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบสอบ และอัตราส่วนสารสนเทศเฉลี่ยสูงที่สุด รองลงมา คือ วิธีประยุกต์การให้คะแนนของเดรสเซลและสมิท ส่วนวิธีประเพณีนิยมมีค่าต่ำที่สุด ซึ่งสอดคล้องกับงานวิจัยหลายเรื่องที่ผ่านมาที่พบว่าวิธีการให้คะแนนแบบตัดตัวลวงและวิธีเลือกชุดตัวถูก มีค่าความเที่ยงสูงกว่าวิธีประเพณีนิยม แต่ส่วนใหญ่จะหาความเที่ยงด้วยสูตรของครอนบัค แอลฟา(Alpha reliability coefficient) (Jaradat and Swagad ,1986 ; Tollefson and Chung ,1986 ; Smith ,1987 ; Jaradat and Tollefson ,1988)

การที่วิธีประยุกต์การให้คะแนนของคูมบ์ มีค่าฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบสอบ และอัตราส่วนสารสนเทศเฉลี่ยสูงที่สุด นั้น สามารถอธิบายโดยใช้หลักเหตุผลได้ว่าวิธีการให้คะแนนด้วยวิธีประยุกต์การให้คะแนนของคูมบ์ได้ข้อมูลของผู้สอบมาพิจารณาความสามารถของผู้สอบมากกว่าวิธีประเพณีนิยม และมากกว่าวิธีประยุกต์การให้คะแนนของเดรสเซลและ

สมิท จากเหตุผลที่กล่าวไปข้างต้น ทำให้ได้คะแนนที่ประมาณค่าความสามารถของผู้สอบได้ใกล้เคียงกับความสามารถที่แท้จริงมากกว่า ประกอบกับการให้คะแนนด้วยวิธีประยุกต์การให้คะแนนของคูมบ์ มีค่าอำนาจจำแนกสูงกว่าวิธีอื่นๆ จึงทำให้ ค่าฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบสอบ และอัตราส่วนสารสนเทศเฉลี่ย มีค่าสูงกว่าวิธีอื่นๆ แสดงให้เห็นว่าการให้คะแนนด้วยวิธีประยุกต์การให้คะแนนของคูมบ์ ทำให้คุณภาพของการวัดสูงขึ้น

สิ่งที่เป็นข้อสังเกตเพิ่มเติมของงานวิจัยนี้ คือ

1.การเปรียบเทียบคุณภาพของแบบสอบจากวิธีที่ให้คะแนนความรู้บางส่วนกับวิธีประเพณีนิยม โดยใช้ค่าฟังก์ชันสารสนเทศของแบบสอบจากทฤษฎีการตอบสนองข้อสอบ (IRT) และใช้ค่าความเที่ยงจากค่าสัมประสิทธิ์แอลฟ่าจากทฤษฎีการวัดแบบดั้งเดิม(Classical Test Theory)นั้น พบว่า ค่าฟังก์ชันสารสนเทศของบบสอบจากวิธีที่ให้คะแนนความรู้บางส่วนทั้งวิธีประยุกต์การให้คะแนนของคูมบ์และวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท มีค่าสูงกว่าวิธีประเพณีนิยม ในขณะที่ค่าความเที่ยง (∞)ของวิธีประยุกต์การให้คะแนนของคูมบ์สูงกว่าวิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท และวิธีประเพณีนิยมอย่างมีนัยสำคัญทางสถิติ แต่วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท ไม่แตกต่างจากวิธีประเพณีนิยมอย่างมีนัยสำคัญทางสถิติ ทำให้ได้ข้อสรุปว่าการศึกษาคุณภาพของแบบสอบจากทฤษฎีการตอบสนองข้อสอบมีความสามารถในการจำแนกได้ชัดเจนกว่าทฤษฎีการวัดแบบดั้งเดิม

2.วัฒนธรรมการทำข้อสอบของผู้สอบระหว่างวิธีการให้คะแนนความรู้บางส่วน พบว่า ผู้สอบนิยมใช้วิธีการตัดตัวลวง ที่เปิดโอกาสให้ตอบ ตัดตัวลวง 1 ตัว หรือ ตัดตัวลวง 2 ตัว ก็ได้ (ในกรณีที่มี 4 ตัวเลือก) ทำให้ได้คะแนนความรู้บางส่วนมากกว่าวิธีการเลือกชุดตัวถูก ที่ผู้สอบส่วนใหญ่จะเลือกตัวถูกเพียงตัวเดียว ไม่นิยมที่จะเลือกชุดตัวถูกที่เป็น 2 ตัว หรือ 3 ตัว (ในกรณีที่มี 4 ตัวเลือก) ทำให้การได้คะแนนความรู้บางส่วนด้วยวิธีเลือกชุดตัวถูกลดลงไป แต่จะไปได้คะแนนจากการตอบที่เลือกเพียง 1 ข้อถูกก็จะได้คะแนนเต็ม ผิดก็จะได้ 0 ซึ่งจะไม่แตกต่างจากการให้คะแนนแบบประเพณีนิยม คือ มีการเดาเกิดขึ้น ในกรณีที่ไม่รู้ ผู้สอบสามารถเลือกตอบโดยการเดา หรือไม่ตอบก็ได้ ทั้งนี้การที่จะเลือกแบบเสี่ยงโชค คือการเดา กับการไม่ตอบนั้นได้มีการวิจัยศึกษาพฤติกรรมการทำข้อสอบดังกล่าวพบว่ามีส่วนสัมพันธ์กับเพศและภูมิหลังทางวัฒนธรรม (Ben- Shakhar and Sinai , 1991 ; Gafni and Melamed , 1990 ; Grandy,1987) ดังนั้นการเลือกวิธีการให้คะแนนความรู้บางส่วนอาจขึ้นอยู่กับวัฒนธรรมการทำข้อสอบที่ได้รับการฝึกฝนมาแตกต่างกันในแต่ละเชื้อชาติ ซึ่งเด็กไทยส่วนใหญ่อาจได้รับการฝึกให้ทำข้อสอบโดยการตัดตัวลวงที่ไม่ใช่ออกไปทีละตัว จนเหลือคำตอบที่ไม่ได้ถูกตัด จะเป็นคำตอบที่ถูก ซึ่งเป็นเรื่องที่น่าสนใจที่ต้องการคำตอบจากการวิจัยต่อไป

ข้อเสนอแนะจากการวิจัยครั้งนี้

1.ในการวิจัยที่ต้องการเปรียบเทียบวิธีการให้คะแนนที่มีวิธีการตอบแตกต่างกัน ควรให้การสอบแต่ละครั้งมีทุกวิธี ดังนั้นจึงควรมีการสุ่มผู้สอบว่าจะต้องทำวิธีการตอบแบบใดก่อนหลัง เพื่อลดความคลาดเคลื่อนที่อาจเกิดจากลำดับที่ของวิธีที่สอบ

2. การใช้ข้อสอบเพื่อเปรียบเทียบในแต่ละวิธี ถ้าเป็นไปได้ควรใช้ข้อสอบคู่ขนานทั้งหมด เพื่อช่วยลดปัญหาจากการจำข้อสอบของผู้สอบได้

3. การเก็บข้อมูลเพื่อเปรียบเทียบวิธีการให้คะแนน ผู้สอบจำเป็นต้องทำอย่างสุดความสามารถ เพื่อลดปัญหาการทำแบบไม่ตั้งใจ จึงควรมีการให้คะแนนความตั้งใจหรือเป็นคะแนนเก็บเพื่อให้ผู้ตอบทำเต็มความสามารถที่แท้จริง หรือชี้แจงถึงข้อเสียที่เกิดขึ้น ถ้าผู้สอบไม่ได้ทำข้อสอบอย่างเต็มความสามารถ

ข้อเสนอแนะในการวิจัยครั้งต่อไป

1. ข้อค้นพบที่ได้จากการวิจัยครั้งนี้ ได้ศึกษาความตรงจากทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory) และศึกษาความเที่ยงจากทฤษฎีการตอบสนองข้อสอบ (Item Response theory) จึงควรมีการศึกษาต่อในเรื่องความเหมาะสมของเกณฑ์ที่ใช้ในการเปรียบเทียบระหว่างทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบ

2. การเลือกวิธีการให้คะแนนความรู้บางส่วนอาจขึ้นอยู่กับวัฒนธรรมการทำข้อสอบที่ได้รับการฝึกฝนมาที่อาจมีแตกต่างกันในแต่ละเชื้อชาติ ดังนั้นจึงน่าจะมีการทำวิจัยในเรื่องที่เกี่ยวกับวิธีต่างๆ ของการทำข้อสอบที่มีการให้คะแนนความรู้บางส่วนที่เหมาะกับผู้เรียน เพื่อจะได้วิธีการที่เหมาะสมสำหรับผู้เรียนที่เป็นคนไทยต่อไป

3. การให้คะแนนความรู้บางส่วนจะช่วยลดข้อจำกัดของแบบสอบเลือกตอบแบบประเพณีนิยมในเรื่องเปิดโอกาสให้มีการเดาสุ่มได้ ประกอบกับการทดสอบในปัจจุบันได้มีการศึกษาการทดสอบแบบปรับเหมาะ(Adaptive Testing) จึงน่าจะมีการวิจัยโดยใช้การให้คะแนนความรู้บางส่วนกับการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์(Computerized Adaptive Testing) เพื่อยืนยันคุณภาพของแบบสอบที่ควรนำไปใช้ต่อไป

4. วิธีการให้คะแนนความรู้บางส่วนช่วยลดความคลาดเคลื่อนในการวัด ดังนั้นจึงควรมีการศึกษาต่อว่าคะแนนที่ได้จากวิธีการให้คะแนนความรู้บางส่วนจะมีผลกระทบต่อการให้เกรดอย่างไร โดยเฉพาะในเรื่องความคลาดเคลื่อนของการตัดเกรดแบบอิงกลุ่มและอิงเกณฑ์

รายการอ้างอิง

กนกวรรณ รัตนธน. (2544). **การเปรียบเทียบความตรงตามสภาพและความคงที่ในการประมาณค่าความสามารถของผู้สอบในการทดสอบแบบปรับเหมาะกับความสามารของผู้สอบด้วยคอมพิวเตอร์เมื่อระดับความสามารถของผู้สอบและอัตราการใช้ข้อสอบซ้ำแตกต่างกัน.** วิทยานิพนธ์ครุศาสตร์มหาบัณฑิตบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

พรทิพย์ ไชยโส. (2533). **การพัฒนาสูตรการให้คะแนนแบบเลือกตอบสำหรับความรู้บางส่วนของผู้ตอบ:การประยุกต์ใช้วิธีการอาร์โนลด์และวิธีการของแฮมดาน.** วิทยานิพนธ์ครุศาสตร์ดุษฎีบัณฑิต บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

ทวี ทองคำ. (2526). **การเปรียบเทียบค่าความเที่ยง ความตรงและอำนาจจำแนกของแบบสอบชนิดเลือกตอบที่ใช้คำสั่งและวิธีการให้คะแนนที่ต่างกัน** วิทยานิพนธ์ครุศาสตร์มหาบัณฑิต บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

ธนวัฒน์ แสนสุข. (2539). **คู่มือการใช้โปรแกรมคอมพิวเตอร์ PARSCALE** . เอกสารประกอบการฝึกอบรมของสมาคมวิจัยทางสังคมศาสตร์ คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

วรรณ์ดี แสงประทีปทอง. (2538). **การศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทยของนักเรียนมัธยมศึกษาตอนต้น โดยการเปรียบเทียบตามแนวตั้งตามทฤษฎีการตอบสนองข้อสอบแบบจำลองโลจิสติกสามพารามิเตอร์.**วิทยานิพนธ์ดุษฎีบัณฑิต มหาวิทยาลัยศรีนครินทรวิโรฒประสานมิตร.

ศิริชัย กาญจนวาสี. (2535). **เอกสารการสอนชุดวิชาการพัฒนาแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน.** นนทบุรี: โรงพิมพ์มหาวิทยาลัยสุโขทัยธรรมาธิราช.

ศิริชัย กาญจนวาสี. (2538). **รวมบทความวิชาทฤษฎีการวัดและการประเมิน.** คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนวาสี. (2542). โมเดลการตอบสนองข้อสอบแบบพหุภาค. เอกสารประกอบการสอนวิชา 2702840 ทฤษฎีการวัดขั้นสูง มปท.

สุพัฒน์ สุกมลสันต์. (2539). **ธนาคารข้อทดสอบและการทดสอบปรับเปลี่ยนด้วยคอมพิวเตอร์.** กรุงเทพมหานคร: บริษัทพิมพ์ดี จำกัด.

สำราญ มีแจ้ง (2525). **ผลของการใช้คำสั่งและการให้คะแนนที่ต่างกันต่อค่าความเที่ยง ความตรงและอำนาจจำแนกของแบบทดสอบชนิดเลือกตอบ.** วิทยานิพนธ์ครุศาสตร์มหาบัณฑิต บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

Alvin, B. and Barry, C.(2000).Correcting grade deflation caused by multple-choice scoring. **International Journal of Mathematical Education in Science & Technology.** 31, 371- 380.

Angoff, W. H. (1989). Does guessing really help? **Journal of Educational Measurement**, 26, 323-336.

Arnold, J. C., and Arnold, P. L. (1970). On scoring multiple choice exams allowing for partial knowledge. **The Journal of Experimental Education**, 39, 8-13.

Bandaranayake,R., Payne,J. and White,S. (1999). Using multiple response true-false multiple choice questions. **Aust.N.Z.J.Surg**, 69,311-315.

Ben- Shakhar , G., and Sinai,Y. (1991). Gender differences in multiple- choice tests: The role of differential guessing tendencies. **Journal of Educational Measurement**, 28, 23-35.

Budescu, D. V. (1988). On the feasibility of multiple matching tests: Variations on a theme by Gulliksen. **Applied Psychological measurement**, 12, 5-14.

Burton,R.F. and Miller,D.J. (1999). Statistical modelling of multiple-choice and true/false tests: ways of considering,and of reducing, the uncertainties attributable to guessing. **Assessment & Evaluation in Higher Education**. 24,399 - 411.

Collet, L.S. (1971). Elimination scoring: An empriical evaluation. **Journal of Educational Measurement**, 8, 209-214.

Coombs, C. H., Milholland, J. E., and Womer, F. B. (1956). The assessment of partial knowledge. **Educational and Psychological Measurement**, 16, 13-37.

De Ayala, R. J. (1994). The Influence of Multidimensionality on the Graded Respose Model. **Applied Psychological Measurement**, 18 (2), 155-170.

De Finetti, B. (1965). Methods for discriminating levels of partial knowledge concerning a test item. **British Journal of Mathematical and Statistical Psychology**, 18, 87-123.

Diamond, J. J. (1975). A preliminary study of the reliability and validity of a scoring procedure, based upon confidence and partial information. **Journal of Educational Measurement**, 12, 129-133.

Diamond, J. J., and Evans, W. (1973). The correction for guessing. **Review of Educational Research**, 43, 181-191.

Dodd, B. G., and Koch, W. R. (1989). Operational Characteristics of Adaptive Testing Procedures Using the Graded Response Model. **Applied Psychological Measurement**, 13 (2), 129-143.

Donoghue, J. R. (1994). An Empirical Examination of the IRT Information of Polytomously Scored Reading Item Under The Generalized Partial Credit Model. **Journal of Educational Measurement**, 31 (4), 295-311.

Dressel, P. L., and Schmidt, J. (1953). Some modifications of the multiple choice item. **Educational and Psychological Measurement**, 13, 574-595.

Echternacht, G. J. (1976). Reliability and validity of item option weighting schemes. **Educational and Psychological Measurement**, 36, 301-310.

Feldt,L.S. (1980). A test of the hypothesis that Cronbach's alpha reliability coefficient is the same for two tests administered to the same sample. **Psychological Measurement**,4,79-90.

Frary, R. B. (1980). The effect of misinformation, partial information, and guessing on expected multiple-choice test item scores. **Applied Psychological Measurement**, 4 , 79-90.

Frary, R. B. (1989). Partial-credit scoring methods for multiple-choice tests. **Applied Measurement in Education**, 2, 79-96.

Gafni,N., and Melamed,E. (1990). **Differential tendencies to guess as a function of gender and lingual-cultural reference group** (Report No.148). Jerusalem,Israel : National Institute for Testing and Evaluation.

Gibbons, J. d., Olkin, I., and Sobel, M. (1979). A subset selection technique for scoring items on a multiple choice test. **Psychometrika**, 44, 259-270.

Glass,G.V. and Stanley , J.C. (1970). **Statistical Methods in Education and Psychology.** Englewood Cliffs ,NJ. : Printice-Hall.

Grandy, J. (1987). **Characteristics of examinees who leave questions unanswered on the GRE general test under right-only scoring.** (Reserch Report No.87-38). Princeton , NJ: Educational Testing Service.

Gulliksen, H. (1986). Perspective on educational measurement. **Applied Psychological Measurement**, 10, 109-132.

Hakstian, A. R., and Kansup, W. (1975). A comparison of several methods of assessing partial knowledge in multiple choice tests: II. Testing procedures. **Journal of Educational Measurment**, 12, 219-230.

Haladyna, T. M. (1988). **Empiricaly based polychotomous scoring of multiple choice test items: A review.** Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Hambleton, R. K., Roberts, D. M., and Traub, R.E. (1970). A Comparison of the reliability and validity of two methods for assessing partial knowledge on a multiple choice test. **Journal of Educational Measurement**, 7, 75-82.

Hambleton,R.K. and Swaminathan. H. (1985). **Item Response Theory:Principles and Application.** Boston: Kluner-Nijhoff Publishing.

Hambleton, R. K., and Zaal, J. N. (1991). **Advanced in Educational and Psychological Testing: Theory and Applications**. Massachusetts: Kluwer Academic Publishers.

Hopkins, K. D., Hakstian, A. R., and Hopkins, B. R. (1973). Validity and reliability consequences of confidence weighting. **Educational and Psychological Measurement**, 33, 135-141.

Horst, P. (1966). **Psychological measurement and prediction**. Belmont CA: Wadsworth.

Hsu, T. C., Moss, P.A., and Khampalikit, C. (1984). The merits of multiple answer items as evaluated by using six scoring formulas. **Journal of Experimental Education**, 52, 152-158.

Jaradat, D., and Swagad, S. (1986). The subset selection technique for multiple choice tests: An empirical inquiry. **Journal of Educational Measurement**, 23, 369-376.

Jaradat,D., and Tollefson, N. (1986). "The subset selection technique for multiple choice tests: An empirical inquiry." **Journal of Educational Measurement.** 23,369-376.

Jaradat, D., and Tollefson, N. (1988). The impact of alternative scoring procedures for multiple-choice items on test reliability, validity and grading. **Educational and Psychological Measurement**, 48, 627-635.

Kansup, w., and Hakstian, A. r. (1975). A comparison of several methods of assessing partial knowledge in multiple-choice tests: I. Scoring procedures. **Journal of Educational Measurement**, 12, 212-230.

Koch, W. R. (1983). Likert Scaling Using the Graded Response Latent Trait Model. **Applied Psychological Measurement**, 7 (1), 15-32.

Koehler, R. A. (1971). A Comparison of the validities of conventional choice testing and various confidence marking procedures. **Journal of Educational Measurement**, 8, 297-303.

Lord. F. M. (1975). Formula scoring and numberright scoring. **Journal of Educational Measurement**, 12, 7-11.

Masters, G. N. (1982). A Rasch Model for Partial Credit Scoring. **Psychometrika**, 47 (2), 149-174.

Michael, J. C. (1968). The reliability of a multiple choice examination under various test-taking instructions. **Journal of Educational Measurement**, 5, 307-314.

Mislevy, R.J. and Bock, R.D. (1990). **Bilog 3**. 2nd edition. Scientific Software,Inc.USA.

Muraki, E. (1992). A Generalized Partial Credit Model: Application of an EM Algolithm. **Applied Psychological Measurement**, 16 (2): 159-176.

_____. (1993). Information Functions of the Generalized Partial Credit Model. **Applied Psychological Measurement**, 17 (4): 351-363.

Muraki,E. and Bock,R.D. (1993).**Parscale**. Scientific Software,Inc.USA.

Ndalichako,J.L.and Rogers,W.T. (1997). Comparison of finite state score theory,classical test theory, and item response theory in scoring multiple–choice items. **Educational &Psychological Measurement**. 57,580-589.

Poizner,S.B.,Nicewander,W.A.,and Gettys,C.F. (1978).Alternative response and scoring methods for multiple-choice items: An empirical study of probabilistic and ordinal response modes. **Applied Psychological Measurement**,35,73-78.

Pugh, R. C., and Brunza, J. J. (1975). Effects of a confidence weighted scoring system on measures of test reliability and validity. **Educational and Psychological Measurement**, 35, 73-78.

Reise, S. P., and Yu, J. (1990). Parameter Recovery in the Graded Response Model Using MULTILOG. **Journal of Educational Measurement**, 27 (2), 133-144.

Rippey, R. M. (1970). A Comparison of five different scoring functions for confidence tests. **Journal of Educational Measurement**, 7, 165-170.

Shuford, E. H., Albert, A., and Massengill, H. E. (1966). Admissible probability measurement procedures. **Psychometrika**, 31, 125-145.

Simon,A.B.,Budscu,D.V.,and Nevo,B. (1997). A comparative study of measures of partial knowledge in multiple – choice tests. **Applied Psychological Measurement**.21,65-88.

Smith,R.M. (1987). Assessing partial knowledge in vocabulary. **Journal of Educational Measurement.** 24,217-231.

Stanley, J. C., and Wang, M. D. (1970). Weighting test items and test item options, an overview of the analytical and empirical literature. **Educational and Psychological Measurement**, 30, 21-35.

Suhadolnik, D., and Weiss, D. J. (1983). **Effect of examinee certainty on probabilistic test scores and a comparison of scoring methods for probabilistic responses** (Research Report 83-3). University of Minnesota, Department of Psychology, Psychometric Methods Program, Computerized Adaptive Testing Laboratory, Minneapolis.

Tollefson,N. and Chung, J.M. (1986). **A comparison of two methods of assessing partial knowledge on multiple choice.** Kansas University.,Lawrence.

Traub,R.D. and Hambleton,R.K. (1972). The Effectof scoring instructions and the degree of speededness on the validity and reliability of multiple choice tests. Educational and Psychological

Measurement,29,847-861.

Wallsten,T.S., Budescu,D.V., and Zwick,R. (1993). Comparing the calibration and coherence of numerical and verbal probabilistic judgments. **Management Science**,39,176-190.

Wesman, A. G. (1971). Writing the test item. N R. L. Thorndike (Ed.), **Educational measurement** (pp. 81-129). Washington DC: American Council on Education.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ก.

ฟังก์ชันสารสนเทศรายข้อ(IIF) และฟังก์ชันสารสนเทศของแบบสอบ (TIF) ของข้อสอบจำแนกตามระดับความสามารถ (θ)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ฟังก์ชันสารสนเทศรายข้อ(IIF) และฟังก์ชันสารสนเทศของแบบสอบ (TIF) ของข้อสอบ จำแนกตามระดับความสามารถ (θ)

ตารางที่ ก 1 ฟังก์ชันสารสนเทศรายข้อ(IIF) และฟังก์ชันสารสนเทศของแบบสอบ (TIF) ของข้อสอบ ที่ระดับความสามารถ (θ) –4.00 จำแนกตามวิธีการให้คะแนน

ข้อ	ฟังก์ชันสารสนเทศรายข้อ	
	แบบตัดตัวลวง	แบบเลือกชุดตัวถูก
1	.252	.035
2	.131	.057
3	.048	.019
4	.031	.048
5	.025	.011
6	.066	.011
7	.028	.024
8	.035	.012
9	.048	.011
10	.026	.049
11	.021	.005
12	.059	.018
13	.066	.008
14	.014	.000
15	.063	.036
16	.027	.004
17	.465	.005
18	.064	.015
19	.043	.058
20	.053	.013
21	.070	.016
22	.118	.015
23	.080	.018
24	.110	.051
25	.063	.020

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ ก 1 (ต่อ)

ข้อ	ฟังก์ชันสารสนเทศรายข้อ	
	แบบตัดตัวลวง	แบบเลือกชุดตัวถูก
26	.106	009
27	.116	.004
28	.041	.046
29	.058	.010
30	.050	.019
31	.042	.018
32	.044	.003
33	.077	.005
34	.028	.004
35	.016	.005
36	.031	.007
37	.041	.024
38	.026	.013
39	.066	.027
40	.025	.003
41	.128	.010
42	.130	.005
43	.031	.004
44	.033	.001
45	.063	.002
46	.050	.000
47	.014	.004
48	.051	.013
49	.017	.005
50	.035	.023
ฟังก์ชันสารสนเทศของแบบสอบ	3.325	0.823

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ ก 2 ฟังก์ชันสารสนเทศรายข้อ(IIF) และฟังก์ชันสารสนเทศของแบบสอบ (TIF) ของข้อสอบที่ระดับความสามารถ (θ) –3.00 จำแนกตามวิธีการให้คะแนน

ข้อ	ฟังก์ชันสารสนเทศรายข้อ	
	แบบตัดตัวลวง	แบบเลือกชุดตัวถูก
1	.433	.135
2	.232	.161
3	.102	.034
4	.074	.065
5	.065	.026
6	.236	.039
7	.064	.046
8	.091	.035
9	.140	.029
10	.092	.088
11	.088	.015
12	.233	.060
13	.155	.024
14	.045	.001
15	.136	.067
16	.077	.014
17	.758	.035
18	.166	.045
19	.143	.110
20	.182	.046
21	.203	.052
22	.295	.074
23	.252	.055
24	.258	.108
25	.183	.049

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ ก 2 (ต่อ)

ข้อ	ฟังก์ชันสารสนเทศรายข้อ	
	แบบตัดตัวลวง	แบบเลือกชุดตัวถูก
26	.252	.031
27	.421	.019
28	.107	.091
29	.157	.025
30	.142	.064
31	.141	.047
32	.150	.013
33	.195	.018
34	.814	.013
35	.081	.019
36	.117	.022
37	.173	.058
38	.162	.054
39	.194	.074
40	.172	.018
41	.346	.039
42	.272	.026
43	.162	.018
44	.140	.006
45	.230	.011
46	.203	.005
47	.124	.018
48	.126	.037
49	.072	.020
50	.102	.053
ฟังก์ชันสารสนเทศของแบบสอบ	21.650	2.212

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ ก 3 ฟังก์ชันสารสนเทศรายข้อ(IIF) และฟังก์ชันสารสนเทศของแบบสอบ (TIF) ของข้อสอบที่ระดับความสามารถ (θ) –2.00 จำแนกตามวิธีการให้คะแนน

ข้อ	ฟังก์ชันสารสนเทศรายข้อ	
	แบบตัดตัวลวง	แบบเลือกชุดตัวถูก
1	.411	.403
2	.303	.334
3	.200	.060
4	.171	.080
5	.166	.063
6	.705	.132
7	.143	.081
8	.227	.100
9	.375	.076
10	.323	.134
11	.375	.048
12	.789	.181
13	.317	.073
14	.154	.003
15	.260	.111
16	.222	.052
17	.455	.215
18	.372	.129
19	.450	.172
20	.566	.155
21	.496	.151
22	.523	.311
23	.636	.152
24	.450	.189
25	.469	.110

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ ก 3 (ต่อ)

ข้อ	ฟังก์ชันสารสนเทศรายข้อ	
	แบบตัดตัวลวง	แบบเลือกชุดตัวถูก
26	.448	.097
27	.963	.091
28	.266	.153
29	.358	.061
30	.367	.189
31	.457	.113
32	.488	.052
33	.409	.072
34	.236	.040
35	.434	.066
36	.434	.067
37	.683	.127
38	1.016	.203
39	.490	.179
40	1.176	.094
41	.621	.140
42	.406	.130
43	.838	.089
44	.576	.039
45	.713	.058
46	.751	.036
47	1.110	.078
48	.287	.103
49	.314	.074
50	.291	.112
ฟังก์ชันสารสนเทศของแบบสอบ	23.690	5.978

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ ก 4 ฟังก์ชันสารสนเทศรายข้อ(IIF) และฟังก์ชันสารสนเทศของแบบสอบ (TIF) ของข้อสอบที่ระดับความสามารถ (θ) –1.00 จำแนกตามวิธีการให้คะแนน

ข้อ	ฟังก์ชันสารสนเทศรายข้อ	
	แบบตัดตัวลวง	แบบเลือกชุดตัวถูก
1	.223	.564
2	.256	.376
3	.309	.094
4	.338	.084
5	.372	.136
6	.861	.348
7	.283	.125
8	.448	.239
9	.641	.174
10	.835	.162
11	.122	.139
12	.992	.387
13	.430	.192
14	.492	.010
15	.354	.151
16	.523	.178
17	.157	.830
18	.517	.291
19	.833	.197
20	.870	.387
21	.638	.333
22	.456	.700
23	.701	.318
24	.424	.235
25	.661	.203

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ ก 4 (ต่อ)

ข้อ	ฟังก์ชันสารสนเทศรายข้อ	
	แบบตัดตัวลวง	แบบเลือกชุดตัวถูก
26	.432	.257
27	.644	.372
28	.489	.196
29	.566	.133
30	.610	.390
31	.863	.221
32	.887	.190
33	.499	.249
34	.556	.117
35	.547	.210
36	.011	.181
37	1.184	.221
38	1.873	.512
39	.660	.313
40	1.931	.397
41	.484	.388
42	.344	.499
43	.589	.366
44	1.219	.230
45	.899	.263
46	1.098	.242
47	2.551	.293
48	.459	.237
49	1.055	.238
50	.611	.195
ฟังก์ชันสารสนเทศของแบบสอบ	33.797	13.763

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ ก 5 ฟังก์ชันสารสนเทศรายข้อ(IIF) และฟังก์ชันสารสนเทศของแบบสอบ (TIF) ของข้อสอบที่ระดับความสามารถ (θ) 0.00 จำแนกตามวิธีการให้คะแนน

ข้อ	ฟังก์ชันสารสนเทศรายข้อ	
	แบบตัดตัวลวง	แบบเลือกชุดตัวถูก
1	.095	.290
2	.153	.215
3	.310	.126
4	.425	.077
5	.515	.230
6	.347	.480
7	.386	.153
8	.485	.366
9	.463	.289
10	.737	.145
11	.832	.314
12	.356	.408
13	.321	.354
14	.870	.034
15	.292	.156
16	.624	.435
17	.050	.723
18	.354	.383
19	.521	.158
20	.433	.472
21	.349	.397
22	.217	.460
23	.308	.369
24	.225	.191
25	.384	.259

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ ก 5 (ต่อ)

ข้อ	ฟังก์ชันสารสนเทศรายข้อ	
	แบบตัดตัวลวง	แบบเลือกชุดตัวถูก
26	.232	.420
27	.194	.732
28	.463	.176
29	.392	.229
30	.445	.396
31	.532	.288
32	.511	.478
33	.308	.529
34	.631	.275
35	.881	.446
36	.652	.347
37	.472	.257
38	.405	.509
39	.368	.309
40	.350	.765
41	.202	.521
42	.184	.738
43	.462	.732
44	.571	.826
45	.358	.700
46	.406	.933
47	.379	.613
48	.391	.351
49	.954	.485
50	.570	.238
ฟังก์ชันสารสนเทศของแบบสอบ	21.365	19.777

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ ก 6 ฟังก์ชันสารสนเทศรายข้อ(IIF) และฟังก์ชันสารสนเทศของแบบสอบ (TIF) ของข้อสอบที่ระดับความสามารถ (θ) 1.00 จำแนกตามวิธีการให้คะแนน

ข้อ	ฟังก์ชันสารสนเทศรายข้อ	
	แบบตัดตัวลวง	แบบเลือกชุดตัวถูก
1	.039	.088
2	.077	.084
3	.201	.134
4	.297	.062
5	.352	.257
6	.099	.281
7	.308	.143
8	.271	.290
9	.185	.284
10	.254	.101
11	.220	.397
12	.091	.207
13	.158	.343
14	.497	.106
15	.163	.123
16	.323	.478
17	.016	.167
18	.155	.253
19	.171	.096
20	.131	.248
21	.127	.233
22	.083	.128
23	.099	.221
24	.094	.110
25	.141	.207

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ ก 6 (ต่อ)

ข้อ	ฟังก์ชันสารสนเทศรายข้อ	
	แบบตัดตัวลวง	แบบเลือกชุดตัวถูก
26	.096	.322
27	.053	.401
28	.235	.116
29	.165	.259
30	.183	.197
31	.171	.223
32	.159	.502
33	.130	.434
34	.309	.388
35	.173	.421
36	.188	.353
37	.113	.189
38	.063	.202
39	.133	.175
40	.050	.389
41	.071	.277
42	.081	.300
43	.086	.407
44	.139	.681
45	.100	.537
46	.100	.696
47	.042	.433
48	.198	.276
49	.263	.413
50	.251	.189
ฟังก์ชันสารสนเทศของแบบสอบ	8.104	13.891

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ ก 7 ฟังก์ชันสารสนเทศรายข้อ(IIF) และฟังก์ชันสารสนเทศของแบบสอบ (TIF) ของข้อสอบที่ระดับความสามารถ (θ) 2.00 จำแนกตามวิธีการให้คะแนน

ข้อ	ฟังก์ชันสารสนเทศรายข้อ	
	แบบตัดตัวลวง	แบบเลือกชุดตัวถูก
1	.016	.025
2	.037	.030
3	.102	.112
4	.142	.045
5	.154	.180
6	.028	.099
7	.162	.103
8	.111	.136
9	.064	.168
10	.072	.059
11	.052	.246
12	.024	.073
13	.067	.179
14	.156	.276
15	.076	.078
16	.116	.223
17	.005	.030
18	.060	.106
19	.052	.050
20	.038	.084
21	.044	.090
22	.031	.030
23	.031	.090
24	.037	.054
25	.048	.114

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ ก 7 (ต่อ)

ข้อ	ฟังก์ชันสารสนเทศรายข้อ	
	แบบตัดตัวลวง	แบบเลือกชุดตัวถูก
26	.038	.138
27	.015	.103
28	.093	.063
29	.061	.184
30	.065	.070
31	.050	.116
32	.046	.214
33	.051	.162
34	.108	.272
35	.033	.186
36	.050	.190
37	.027	.099
38	.010	.057
39	.045	.074
40	.007	.095
41	.025	.089
42	.034	.069
43	.016	.105
44	.033	.160
45	.028	.153
46	.025	.141
47	.005	.138
48	.082	.132
49	.061	.166
50	.087	.107
ฟังก์ชันสารสนเทศของแบบสอบ	2.820	5.963

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ ก 8 ฟังก์ชันสารสนเทศรายข้อ(IIF) และฟังก์ชันสารสนเทศของแบบสอบ (TIF) ของข้อสอบที่ระดับความสามารถ (θ) 3.00 จำแนกตามวิธีการให้คะแนน

ข้อ	ฟังก์ชันสารสนเทศรายข้อ	
	แบบตัดตัวลวง	แบบเลือกชุดตัวถูก
1	.006	.008
2	.018	.012
3	.048	.078
4	.060	.030
5	.060	.092
6	.008	.031
7	.073	.063
8	.043	.052
9	.022	.074
10	.020	.033
11	.012	.098
12	.006	.025
13	.028	.069
14	.046	.428
15	.035	.044
16	.040	.070
17	.001	.007
18	.023	.039
19	.016	.025
20	.011	.026
21	.015	.032
22	.012	.008
23	.010	.033
24	.015	.025
25	.017	.053

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ ก 8 (ต่อ)

ข้อ	ฟังก์ชันสารสนเทศรายข้อ	
	แบบตัดตัวลวง	แบบเลือกชุดตัวถูก
26	.015	.048
27	.004	.024
28	.036	.032
29	.023	.094
30	.023	.024
31	.015	.051
32	.014	.063
33	.020	.047
34	.037	.117
35	.006	.060
36	.013	.074
37	.007	.045
38	.001	.016
39	.016	.029
40	.001	.022
41	.009	.027
42	.015	.016
43	.003	.025
44	.008	.029
45	.008	.035
46	.006	.023
47	.000	.036
48	.033	.052
49	.014	.051
50	.030	.052
ฟังก์ชันสารสนเทศของแบบสอบ	1.002	2.547

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ ก 9 ฟังก์ชันสารสนเทศรายข้อ(IIF) และฟังก์ชันสารสนเทศของแบบสอบ (TIF) ของข้อสอบที่ระดับความสามารถ (θ) 4.00 จำแนกตามวิธีการให้คะแนน

ข้อ	ฟังก์ชันสารสนเทศรายข้อ	
	แบบตัดตัวลวง	แบบเลือกชุดตัวถูก
1	.002	.003
2	.009	.005
3	.022	.047
4	.025	.020
5	.023	.042
6	.002	.011
7	.032	.035
8	.017	.020
9	.008	.031
10	.006	.018
11	.003	.035
12	.001	.009
13	.012	.025
14	.014	.308
15	.016	.024
16	.014	.022
17	.000	.002
18	.009	.015
19	.005	.013
20	.003	.009
21	.005	.012
22	.004	.003
23	.003	.013
24	.006	.012
25	.006	.024

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 9 (ต่อ)

ข้อ	ฟังก์ชันสารสนเทศรายข้อ	
	แบบตัดตัวลวง	แบบเลือกชุดตัวถูก
26	.006	.017
27	.001	.007
28	.014	.016
29	.009	.043
30	.008	.009
31	.004	.022
32	.004	.019
33	.008	.014
34	.013	.043
35	.001	.019
36	.003	.027
37	.001	.020
38	.000	.006
39	.005	.012
40	.000	.006
41	.003	.009
42	.006	.005
43	.000	.007
44	.002	.007
45	.002	.009
46	.001	.005
47	.000	.011
48	.013	.020
49	.003	.016
50	.010	.024
ฟังก์ชันสารสนเทศของแบบสอบ	0.364	1.151

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ข.

รูปแบบการตอบของกลุ่มตัวอย่างในการแสดงความรู้บางส่วนจำแนกตามกลุ่มที่มีความสามารถสูง ปานกลาง ต่ำ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รูปแบบการตอบของกลุ่มตัวอย่างในการแสดงความรู้บางส่วนจำแนกตามกลุ่มที่มีความสามารถ สูง ปานกลาง ต่ำ

1. วิธีประยุกต์การให้คะแนนของคูมบ์ (แสดงใน ตารางที่ ข 1)
2. วิธีประยุกต์การให้คะแนนของเดรสเชลและสมิท (แสดงในตารางที่ ข 2)

ตารางที่ ข 1 จำนวนและร้อยละของกลุ่มตัวอย่างที่มีรูปแบบการตัดตัวลวงที่ต่างกันจำแนกตามความสามารถของกลุ่มตัวอย่าง และค่าความยากรายข้อ

ข้อ	ความยาก	จำนวนตัวเลือกที่ตัด	กลุ่มความสามารถ สูง		กลุ่มความสามารถ ปานกลาง		กลุ่มความสามารถ ต่ำ		รวม	
			จำนวน	ร้อยละ	จำนวน	ร้อยละ	จำนวน	ร้อยละ	จำนวน	ร้อยละ
17	-3.012	1-2	9	17.6	19	12.6	7	14.0	35	13.9
		3	39	76.5	111	73.5	28	56.0	178	70.7
		ไม่ตอบ	3	5.9	21	13.9	15	30.0	39	15.4
1	-2.581	1-2	2	3.9	24	15.9	6	12.0	32	12.7
		3	49	96.1	103	68.9	33	66.0	185	73.4
		ไม่ตอบ	0	0.0	24	15.2	11	22.0	35	13.9
2	-1.881	1-2	8	15.7	36	23.9	16	32.0	60	23.8
		3	42	82.3	99	65.5	22	44.0	163	64.7
		ไม่ตอบ	1	0.0	16	10.6	12	24.0	29	11.5
27	-1.805	1-2	3	5.9	20	13.2	8	16.0	31	12.3
		3	38	74.5	92	60.9	19	38.0	149	59.1
		ไม่ตอบ	10	19.6	39	25.9	23	46.0	72	28.6
41	-1.783	1-2	8	15.7	31	20.5	9	18.0	48	19.0
		3	37	72.5	93	61.6	25	50.0	155	61.5
		ไม่ตอบ	6	11.8	27	17.9	16	32.0	49	19.5
42	-1.782	1-2	8	15.7	43	28.5	10	20.0	61	24.2
		3	39	76.4	80	53.0	23	46.0	142	56.4
		ไม่ตอบ	4	7.9	28	18.5	17	34.0	49	19.4
22	-1.679	1-2	6	11.8	29	19.2	14	28.0	49	19.4
		3	45	88.2	105	69.5	25	50.0	175	69.5
		ไม่ตอบ	0	0.0	17	11.3	11	22.0	28	11.1
24	-1.588	1-2	15	29.4	45	29.8	13	26.0	73	29.0
		3	36	70.6	87	57.6	24	48.0	147	58.3
		ไม่ตอบ	0	0.0	19	12.6	13	26.0	32	12.7

ตารางที่ ข 1 (ต่อ)

ข้อ	ความยาก	จำนวนตัวเลือกที่ตัด	กลุ่มความสามารถสูง		กลุ่มความสามารถปานกลาง		กลุ่มความสามารถต่ำ		รวม	
			จำนวน	ร้อยละ	จำนวน	ร้อยละ	จำนวน	ร้อยละ	จำนวน	ร้อยละ
26	-1.554	1-2	5	9.8	25	16.6	9	18.0	39	15.5
		3	44	86.2	106	70.2	28	56.0	178	70.6
		ไม่ตอบ	2	4.0	20	13.2	13	26.0	35	13.9
23	-1.408	1-2	10	19.6	43	28.5	13	26.0	66	26.2
		3	37	72.5	79	52.3	21	42.0	137	54.4
		ไม่ตอบ	4	7.9	29	19.2	16	32.0	49	19.4
12	-1.343	1-2	6	11.8	20	13.3	14	28.0	40	15.9
		3	42	82.3	99	65.5	23	46.0	164	65.1
		ไม่ตอบ	3	5.9	32	21.2	13	26.0	48	19.0
6	-1.342	1-2	9	17.6	52	34.4	16	32.0	77	30.6
		3	41	80.4	67	44.4	16	32.0	124	49.2
		ไม่ตอบ	1	2.0	32	21.2	18	36.0	51	20.2
45	-1.323	1-2	12	23.5	28	18.5	10	20.0	50	19.9
		3	35	68.6	93	61.6	23	46.0	151	59.9
		ไม่ตอบ	4	7.9	30	19.9	17	34.0	51	20.2
40	-1.320	1-2	3	5.9	30	19.9	7	14.0	40	15.9
		3	48	94.1	99	65.5	28	56.0	175	69.4
		ไม่ตอบ	0	0.0	22	14.6	15	30.0	37	14.7
38	-1.256	1-2	6	11.8	36	23.8	14	28.0	56	22.2
		3	44	86.2	96	63.6	22	44.0	162	64.3
		ไม่ตอบ	1	2.0	19	12.6	14	28.0	34	13.5
46	-1.252	1-2	14	27.4	26	17.2	10	20.0	50	19.8
		3	35	68.6	93	61.6	22	44.0	150	59.6
		ไม่ตอบ	2	4.0	32	21.2	18	36.0	52	20.6
47	-1.246	1-2	12	23.5	28	18.5	12	24.0	52	20.6
		3	37	72.5	103	68.2	25	50.0	165	65.5
		ไม่ตอบ	2	4.0	20	13.3	13	26.0	35	13.9
21	-1.222	1-2	5	9.8	48	31.8	16	32.0	69	27.4
		3	40	78.4	80	53.0	18	36.0	138	54.8
		ไม่ตอบ	6	11.8	23	15.2	16	32.0	45	17.8

ตารางที่ ข 1 (ต่อ)

ข้อ	ความยาก	จำนวนตัวเลือกที่ตัด	กลุ่มความสามารถสูง		กลุ่มความสามารถปานกลาง		กลุ่มความสามารถต่ำ		รวม	
			จำนวน	ร้อยละ	จำนวน	ร้อยละ	จำนวน	ร้อยละ	จำนวน	ร้อยละ
33	-1.221	1-2	13	25.5	41	27.1	17	34.0	71	28.1
		3	37	72.5	88	58.3	21	42.0	146	58.0
		ไม่ตอบ	1	2.0	22	14.6	12	24.0	35	13.9
43	-1.190	1-2	8	15.7	32	21.2	10	20.0	50	19.8
		3	41	80.3	88	58.3	23	46.0	152	60.3
		ไม่ตอบ	2	4.0	31	20.5	17	34.0	50	19.9
39	-1.178	1-2	16	31.4	44	29.2	10	20.0	70	27.8
		3	35	68.6	86	56.9	26	52.0	147	58.4
		ไม่ตอบ	0	0.0	21	13.9	14	28.0	35	13.8
37	-1.146	1-2	9	17.6	35	23.2	13	26.0	57	22.6
		3	41	80.4	92	60.9	23	46.0	156	61.9
		ไม่ตอบ	1	2.0	24	15.9	14	28.0	39	15.5
20	-1.132	1-2	11	21.6	32	21.2	20	40.0	63	25.0
		3	39	76.4	93	61.6	13	26.0	145	57.6
		ไม่ตอบ	1	2.0	26	17.2	17	34.0	44	17.4
25	-1.124	1-2	16	31.4	43	28.5	13	26.0	72	28.6
		3	33	64.7	92	60.9	25	50.0	150	59.5
		ไม่ตอบ	2	3.9	16	10.6	12	24.0	30	11.9
18	-1.037	1-2	13	25.5	50	33.1	13	26.0	76	30.1
		3	35	68.6	75	49.7	24	48.0	134	53.2
		ไม่ตอบ	3	5.9	26	17.2	13	26.0	42	16.7
44	-1.003	1-2	12	23.5	34	22.5	10	20.0	56	22.2
		3	36	70.6	84	55.6	24	48.0	144	57.2
		ไม่ตอบ	3	5.9	33	21.9	16	32.0	52	20.6
13	-.989	1-2	14	27.5	44	29.1	15	30.0	73	29.0
		3	32	62.7	71	47.0	17	34.0	120	47.6
		ไม่ตอบ	5	9.8	36	23.9	18	36.0	59	23.4
32	-.977	1-2	10	19.6	48	31.8	8	16.0	66	26.2
		3	39	76.4	82	54.3	28	56.0	149	59.1
		ไม่ตอบ	2	4.0	21	13.9	14	28.0	37	14.7

ตารางที่ ข 1 (ต่อ)

ข้อ	ความยาก	จำนวนตัวเลือกที่ตัด	กลุ่มความสามารถสูง		กลุ่มความสามารถปานกลาง		กลุ่มความสามารถต่ำ		รวม	
			จำนวน	ร้อยละ	จำนวน	ร้อยละ	จำนวน	ร้อยละ	จำนวน	ร้อยละ
29	-.974	1-2	12	23.5	51	33.8	11	22.0	74	29.4
		3	38	74.5	75	49.7	26	52.0	139	55.2
		ไม่ตอบ	1	2.0	25	16.6	13	26.0	39	15.4
19	-.924	1-2	13	25.5	53	35.1	20	40.0	80	31.7
		3	36	70.6	79	52.3	13	26.0	139	55.2
		ไม่ตอบ	2	3.9	19	12.6	17	34.0	33	13.1
31	-.923	1-2	8	15.7	51	33.8	12	24.0	71	28.2
		3	40	78.4	76	50.3	21	42.0	137	54.4
		ไม่ตอบ	3	5.9	24	15.9	17	34.0	44	17.4
15	-.876	1-2	19	37.3	57	37.7	13	26.0	89	35.3
		3	23	45.1	57	37.7	23	46.0	103	40.9
		ไม่ตอบ	9	17.6	37	24.6	14	28.0	60	23.8
30	-.873	1-2	16	31.4	54	35.7	11	22.0	81	32.2
		3	34	66.6	72	47.7	24	48.0	130	51.6
		ไม่ตอบ	1	2.0	25	16.6	15	30.0	41	16.2
9	-.866	1-2	14	27.4	45	29.8	13	26.0	72	28.6
		3	35	68.6	62	41.1	14	28.0	111	44.0
		ไม่ตอบ	2	4.0	44	29.1	23	46.0	69	27.4
36	-.821	1-2	13	25.5	40	26.5	12	24.0	65	25.8
		3	37	72.5	87	57.6	22	44.0	146	57.9
		ไม่ตอบ	1	2.0	24	15.9	16	32.0	41	16.3
35	-.773	1-2	13	25.5	55	36.4	20	40.0	88	34.9
		3	34	66.6	66	43.7	16	32.0	116	46.0
		ไม่ตอบ	4	7.9	30	19.9	14	28.0	48	19.1
48	-.740	1-2	12	23.5	58	38.4	13	26.0	83	32.9
		3	34	66.6	70	46.3	22	44.0	126	50.0
		ไม่ตอบ	5	9.9	23	15.3	15	30.0	43	17.1
11	-.685	1-2	9	17.6	26	17.2	11	22.0	46	18.2
		3	32	62.7	75	49.7	22	44.0	129	51.2
		ไม่ตอบ	10	19.7	50	33.1	17	34.0	77	30.6
10	-.598	1-2	13	25.5	50	33.1	14	28.0	77	30.5
		3	32	62.7	63	41.7	15	30.0	110	43.7
		ไม่ตอบ	6	11.8	38	25.2	21	42.0	65	25.8

ตารางที่ ข 1 (ต่อ)

ข้อ	ความยาก	จำนวนตัวเลือกที่ตัด	กลุ่มความสามารถสูง		กลุ่มความสามารถปานกลาง		กลุ่มความสามารถต่ำ		รวม	
			จำนวน	ร้อยละ	จำนวน	ร้อยละ	จำนวน	ร้อยละ	จำนวน	ร้อยละ
50	-.575	1-2	15	29.4	41	27.2	20	40.0	76	30.1
		3	26	51.0	71	47.0	15	30.0	112	44.5
		ไม่ตอบ	10	19.6	39	25.8	15	30.0	64	25.4
28	-.574	1-2	25	49.0	58	38.4	23	46.0	106	42.1
		3	26	51.0	75	49.7	15	30.0	116	46.0
		ไม่ตอบ	0	0.0	18	11.9	12	24.0	30	11.9
49	-.561	1-2	11	21.6	36	23.8	15	30.0	62	24.6
		3	36	70.6	88	58.3	20	40.0	144	57.2
		ไม่ตอบ	4	7.8	27	17.9	15	30.0	46	18.2
3	-.495	1-2	21	41.2	67	44.4	22	44.0	110	43.7
		3	26	51.0	64	42.4	14	28.0	104	41.3
		ไม่ตอบ	4	7.8	20	13.2	14	28.0	38	15.1
8	-.392	1-2	25	49.0	73	48.3	23	46.0	121	48.0
		3	20	39.2	61	40.4	16	32.0	97	38.5
		ไม่ตอบ	6	11.8	17	11.3	11	22.0	34	13.5
34	-.366	1-2	20	39.2	64	42.4	16	32.0	100	39.7
		3	29	56.8	55	36.4	15	30.0	99	39.3
		ไม่ตอบ	2	4.0	32	21.2	19	38.0	53	21.0
4	-.117	1-2	25	49.0	81	53.6	26	52.0	132	52.4
		3	25	49.0	46	30.5	8	16.0	79	31.4
		ไม่ตอบ	1	2.0	24	15.9	17	32.0	41	16.2
5	-.042	1-2	24	47.1	70	46.4	20	40.0	114	45.3
		3	22	43.1	44	29.1	10	20.0	76	30.1
		ไม่ตอบ	5	9.8	37	24.5	20	40.0	62	24.6
16	-.037	1-2	18	35.3	43	28.5	10	20.0	71	28.2
		3	27	52.9	70	46.3	26	52.0	123	48.8
		ไม่ตอบ	6	11.8	38	25.2	14	28.0	58	23.0
14	.006	1-2	17	33.3	57	37.7	18	36.0	92	36.5
		3	33	64.7	75	49.7	20	40.0	128	50.8
		ไม่ตอบ	1	2.0	19	12.6	12	24.0	32	12.7
7	.083	1-2	19	37.2	70	46.3	26	52.0	115	45.6
		3	28	54.9	53	35.1	11	22.0	92	36.5
		ไม่ตอบ	4	7.9	28	18.6	13	26.0	45	17.9

จากตารางที่ ข 1 เมื่อพิจารณารูปแบบการตัดตัวลวง จะเห็นว่าส่วนใหญ่จะตัดตัวเลือก 3 ตัว (ข้อสอบ 4 ตัวเลือก) แต่ก็มีบางส่วนที่ตัดตัวเลือก 1- 2 ตัว โดยการตัดตัวเลือก1-2 ตัว มีร้อยละของการเลือกแปรเปลี่ยนไปตามความยากของข้อสอบและความสามารถของกลุ่มตัวอย่าง โดยเมื่อแบ่งข้อสอบตามค่าความยากเป็น 2 กลุ่ม คือ กลุ่มข้อสอบที่มีค่าความยาก ต่ำกว่า – 1.000 มีทั้งหมด 26 ข้อ (จากตารางที่ ข 1 เรียงจากข้อ 17 – 44) และกลุ่มข้อสอบที่มีค่าความยากสูงกว่าหรือเท่ากับ –1.000 มีทั้งหมด 24 ข้อ (จากตารางที่ ข 1 เรียงจากข้อ 13 – 7) ส่วนความสามารถของกลุ่มแบ่งเป็น กลุ่มที่มีความสามารถสูง ปานกลาง ต่ำ จากการใช้คะแนนความสามารถจากการสอบกลางภาคเป็นตัวแบ่งกลุ่มที่มีเกณฑ์ คือ 1) ความสามารถ .75 ขึ้นไป เป็นกลุ่มที่มีความสามารถสูง 2) ความสามารถ (-.75)-.75 เป็นกลุ่มที่มีความสามารถปานกลาง และ3) ความสามารถต่ำกว่า-.75 เป็นกลุ่มที่มีความสามารถต่ำ ผลการวิเคราะห์ มีรายละเอียด ดังนี้

ในกลุ่มข้อสอบที่มีค่าความยาก ต่ำกว่า – 1.000 (ง่าย) พบว่าร้อยละของการตอบแบบตัดตัวเลือก 1-2 ตัวในกลุ่มที่มีความสามารถสูง มีค่าอยู่ระหว่าง 3.9 - 31.4 กลุ่มที่มีความสามารถปานกลางมีค่าอยู่ระหว่าง 12.6 - 34.4 กลุ่มที่มีความสามารถต่ำมีค่าอยู่ระหว่าง 12.0 - 40.0 และในภาพรวมทั้งหมดมีค่าอยู่ระหว่าง 12.3 - 30.6

ในกลุ่มข้อสอบที่มีค่าความยาก สูงกว่าหรือเท่ากับ – 1.000 (ค่อนข้างง่าย) พบว่าร้อยละของการตอบแบบตัดตัวเลือก 1-2 ตัวในกลุ่มที่มีความสามารถสูง มีค่าอยู่ระหว่าง 15.7 – 49.0 กลุ่มที่มีความสามารถปานกลางมีค่าอยู่ระหว่าง 17.2 – 53.6 กลุ่มที่มีความสามารถต่ำมีค่าอยู่ระหว่าง 16.0 – 52.0 และในภาพรวมทั้งหมดมีค่าอยู่ระหว่าง 18.2-52.4

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ ข 2 จำนวนและร้อยละของกลุ่มตัวอย่างที่มีรูปแบบการเลือกชุดตัวถูกที่ต่างกันจำแนกตามความสามารถของกลุ่มตัวอย่าง และค่าความยากรายข้อ

ข้อ	ความยาก	จำนวนตัวเลือกที่เลือก	กลุ่มความสามารถสูง		กลุ่มความสามารถปานกลาง		กลุ่มความสามารถต่ำ		รวม	
			จำนวน	ร้อยละ	จำนวน	ร้อยละ	จำนวน	ร้อยละ	จำนวน	ร้อยละ
1	-1.130	1	49	84.5	113	71.1	28	59.6	190	72.0
		2-3	0	0	13	8.1	5	10.6	18	6.8
		ไม่ตอบ	9	15.5	33	20.8	14	29.8	56	21.2
2	-1.267	1	50	86.2	110	69.2	32	68.1	192	72.6
		2-3	4	6.9	18	11.3	2	4.2	24	9.1
		ไม่ตอบ	4	6.9	31	19.5	13	27.7	48	18.2
19	-.990	1	44	75.9	103	64.8	31	66.0	178	67.5
		2-3	9	15.5	24	15.1	3	6.3	36	13.6
		ไม่ตอบ	5	8.6	32	20.1	13	27.7	50	18.9
24	-.873	1	43	74.1	104	65.5	30	63.9	177	67.1
		2-3	12	20.7	24	15.0	5	10.6	41	15.5
		ไม่ตอบ	3	5.2	31	19.5	12	25.5	46	17.4
4	-.861	1	36	62.0	88	55.4	24	51.1	148	56.1
		2-3	19	32.8	39	24.5	11	23.4	69	26.1
		ไม่ตอบ	3	5.2	32	20.1	12	25.5	47	17.8
22	-.792	1	52	89.9	117	73.7	30	63.9	199	75.4
		2-3	2	3.4	9	5.6	4	8.5	15	5.7
		ไม่ตอบ	4	6.9	33	20.7	13	27.6	50	18.9
10	-.705	1	44	75.9	91	57.3	27	57.5	162	61.4
		2-3	10	17.2	32	20.1	5	10.6	47	17.8
		ไม่ตอบ	4	6.9	36	22.6	15	31.9	55	20.8
28	-.660	1	41	70.7	99	62.3	27	57.5	167	63.3
		2-3	14	24.1	28	17.6	7	14.9	49	18.5
		ไม่ตอบ	3	5.2	32	20.1	13	27.6	48	18.2
17	-.555	1	51	87.9	116	73.0	32	68.1	199	75.4
		2-3	2	3.5	11	6.9	2	4.2	15	5.7
		ไม่ตอบ	5	8.6	32	20.1	13	27.7	50	18.9
38	-.459	1	53	91.4	111	69.9	21	44.7	185	70.1
		2-3	1	1.7	12	7.5	11	23.4	24	9.1
		ไม่ตอบ	4	6.9	36	22.6	15	31.9	55	20.8

ตารางที่ ข 2 (ต่อ)

ข้อ	ความยาก	จำนวนตัวเลือกที่เลือก	กลุ่มความสามารถสูง		กลุ่มความสามารถปานกลาง		กลุ่มความสามารถต่ำ		รวม	
			จำนวน	ร้อยละ	จำนวน	ร้อยละ	จำนวน	ร้อยละ	จำนวน	ร้อยละ
39	-.434	1	49	84.5	105	66.1	25	53.2	179	67.8
		2-3	5	8.6	20	12.6	7	14.9	32	12.1
		ไม่ตอบ	4	6.9	34	21.3	15	31.9	53	20.1
30	-.416	1	48	82.8	111	69.9	26	55.3	185	70.1
		2-3	6	10.3	16	10.0	8	17.0	30	11.3
		ไม่ตอบ	4	6.9	32	20.1	13	27.7	49	18.6
12	-.371	1	50	86.2	105	66.1	30	63.9	185	70.2
		2-3	4	6.9	15	9.4	2	4.2	21	7.9
		ไม่ตอบ	4	6.9	39	24.5	15	31.9	58	21.9
20	-.220	1	50	86.2	102	64.2	24	51.1	176	66.7
		2-3	2	3.5	23	14.4	8	17.0	33	12.5
		ไม่ตอบ	6	10.3	34	21.4	15	31.9	55	20.8
15	-.212	1	34	58.6	91	57.3	30	63.9	155	58.7
		2-3	15	25.9	28	17.6	3	6.3	46	17.4
		ไม่ตอบ	9	15.5	40	25.1	14	29.8	63	23.9
23	-.210	1	45	77.6	102	64.2	26	55.3	173	65.5
		2-3	9	15.5	19	11.9	4	8.5	32	12.1
		ไม่ตอบ	4	6.9	38	23.9	17	36.2	59	22.4
21	-.194	1	49	84.5	99	62.3	27	57.5	175	66.3
		2-3	3	5.2	20	12.6	7	14.9	30	11.3
		ไม่ตอบ	6	10.3	40	25.2	13	27.3	59	22.4
42	-.184	1	51	87.9	111	69.9	23	48.9	185	70.1
		2-3	3	5.2	6	3.7	8	17.0	17	6.4
		ไม่ตอบ	4	6.9	42	26.4	16	34.1	62	23.5
41	-.143	1	45	77.6	107	67.3	28	59.6	180	68.2
		2-3	8	13.8	11	6.9	3	6.4	22	8.3
		ไม่ตอบ	5	8.6	41	25.8	16	34.0	62	23.5
6	-.078	1	49	84.5	99	62.3	29	61.7	177	67.1
		2-3	4	6.9	17	10.7	3	6.4	24	9.1
		ไม่ตอบ	5	8.6	43	27.0	15	31.9	63	23.8

ตารางที่ ข 2 (ต่อ)

ข้อ	ความยาก	จำนวนตัวเลือกที่เลือก	กลุ่มความสามารถสูง		กลุ่มความสามารถปานกลาง		กลุ่มความสามารถต่ำ		รวม	
			จำนวน	ร้อยละ	จำนวน	ร้อยละ	จำนวน	ร้อยละ	จำนวน	ร้อยละ
37	-.069	1	48	82.7	101	63.5	25	53.2	174	65.9
		2-3	7	12.1	18	11.3	6	12.8	31	11.7
		ไม่ตอบ	3	5.2	40	25.2	16	34.0	59	22.4
18	-.036	1	44	75.9	101	63.6	18	38.3	163	61.7
		2-3	10	17.2	22	13.8	15	31.9	47	17.8
		ไม่ตอบ	4	6.9	36	22.6	14	29.8	54	20.5
40	.022	1	49	84.5	112	70.5	29	61.7	190	72.0
		2-3	4	6.9	11	6.9	2	4.2	17	6.4
		ไม่ตอบ	5	8.6	36	22.6	16	34.1	57	21.6
27	.064	1	49	84.5	97	61.1	23	48.9	169	64.0
		2-3	1	1.7	11	6.9	4	8.5	16	6.1
		ไม่ตอบ	8	13.8	51	32.0	20	42.6	79	29.9
50	.077	1	46	79.4	90	56.7	26	55.3	162	61.4
		2-3	6	10.3	26	16.3	8	17.0	40	15.1
		ไม่ตอบ	6	10.3	43	27.0	13	27.7	62	23.5
43	.078	1	45	77.6	109	68.6	28	59.6	182	68.9
		2-3	4	6.9	9	5.6	2	4.2	15	5.7
		ไม่ตอบ	9	15.5	41	25.8	17	36.2	67	25.4
31	.105	1	49	84.5	100	63.0	24	51.1	173	65.5
		2-3	4	6.9	24	15.0	8	17.0	36	13.7
		ไม่ตอบ	5	8.6	35	22.0	15	31.9	55	20.8
25	.128	1	46	79.3	109	68.6	27	57.5	182	68.9
		2-3	9	15.5	20	12.6	8	17.0	37	14.0
		ไม่ตอบ	3	5.2	30	18.8	12	25.5	45	17.1
48	.206	1	40	69.0	97	61.0	24	51.1	161	61.0
		2-3	10	17.2	17	10.7	8	17.0	35	13.2
		ไม่ตอบ	8	13.8	45	28.3	15	31.9	68	25.8
26	.222	1	48	82.8	115	72.4	30	63.9	193	73.1
		2-3	7	12.0	11	6.9	5	10.6	23	8.7
		ไม่ตอบ	3	5.2	33	20.7	12	25.5	48	18.2

ตารางที่ ข 2 (ต่อ)

ข้อ	ความยาก	จำนวนตัวเลือกที่เลือก	กลุ่มความสามารถสูง		กลุ่มความสามารถปานกลาง		กลุ่มความสามารถต่ำ		รวม	
			จำนวน	ร้อยละ	จำนวน	ร้อยละ	จำนวน	ร้อยละ	จำนวน	ร้อยละ
8	.232	1	41	70.7	101	63.6	25	53.2	167	63.3
		2-3	11	19.0	20	12.6	8	17.0	39	14.8
		ไม่ตอบ	6	10.3	38	23.8	14	29.8	58	21.9
47	.235	1	48	82.7	110	69.3	29	61.7	187	70.8
		2-3	5	8.6	12	7.5	3	6.4	20	7.6
		ไม่ตอบ	5	8.6	37	23.2	15	31.9	57	21.6
45	.338	1	43	74.1	107	67.4	26	55.3	176	66.7
		2-3	6	10.3	8	5.0	4	8.5	18	6.8
		ไม่ตอบ	9	15.5	44	27.6	17	36.2	70	26.5
33	.355	1	46	79.3	107	67.4	27	57.5	180	68.2
		2-3	8	13.8	13	8.1	7	14.9	28	10.6
		ไม่ตอบ	4	6.9	39	24.5	13	27.6	56	21.2
46	.365	1	48	82.7	103	64.8	29	61.7	180	30.3
		2-3	3	5.2	9	5.6	0	0.0	12	4.5
		ไม่ตอบ	7	12.1	47	29.6	18	38.3	72	27.2
49	.381	1	49	84.5	110	69.3	28	59.6	187	70.9
		2-3	5	8.6	7	4.3	5	10.6	17	6.4
		ไม่ตอบ	4	6.9	42	26.4	14	29.8	60	22.7
7	.407	1	45	77.6	95	59.8	26	55.3	166	62.9
		2-3	7	12.1	30	18.9	8	17.0	45	17.0
		ไม่ตอบ	6	10.3	34	21.3	13	27.7	53	20.1
44	.411	1	42	72.4	106	66.7	28	59.6	176	66.7
		2-3	8	13.8	9	5.6	2	4.2	19	7.2
		ไม่ตอบ	8	13.8	44	27.7	17	36.2	69	26.1
35	.493	1	51	87.9	107	67.4	28	59.6	186	70.5
		2-3	3	5.2	18	11.2	6	12.7	27	10.2
		ไม่ตอบ	4	6.9	34	21.4	13	27.7	51	19.3
13	.529	1	46	79.4	92	57.9	22	46.8	160	60.6
		2-3	6	10.3	18	11.3	7	14.9	31	11.7
		ไม่ตอบ	6	10.3	49	30.8	18	38.3	73	27.7

ตารางที่ ข 2 (ต่อ)

ข้อ	ความยาก	จำนวนตัวเลือกที่เลือก	กลุ่มความสามารถสูง		กลุ่มความสามารถปานกลาง		กลุ่มความสามารถต่ำ		รวม	
			จำนวน	ร้อยละ	จำนวน	ร้อยละ	จำนวน	ร้อยละ	จำนวน	ร้อยละ
9	.564	1	49	84.4	87	54.8	20	42.6	156	59.1
		2-3	5	8.6	12	7.5	10	21.2	27	10.2
		ไม่ตอบ	4	6.9	60	37.7	17	36.2	81	30.7
32	.595	1	49	84.5	108	68.0	29	61.7	186	70.4
		2-3	6	10.3	14	8.8	4	8.5	24	9.1
		ไม่ตอบ	3	5.2	37	23.2	14	29.8	54	20.5
36	.597	1	48	82.7	103	64.8	29	61.7	180	68.2
		2-3	7	12.1	16	10.0	4	8.5	27	10.2
		ไม่ตอบ	3	5.2	40	25.2	14	76.8	57	21.6
16	.651	1	46	79.3	108	68.0	33	70.2	187	70.8
		2-3	5	8.6	13	8.1	1	2.1	19	7.2
		ไม่ตอบ	7	12.1	38	23.9	13	27.7	58	22.0
5	.833	1	36	62.0	72	45.3	25	53.2	133	50.4
		2-3	11	19.0	34	21.4	5	10.6	50	18.9
		ไม่ตอบ	11	19.0	53	33.3	17	36.2	81	30.7
29	.862	1	42	72.4	107	67.4	27	57.5	176	66.7
		2-3	9	15.5	21	13.2	8	17.0	38	14.4
		ไม่ตอบ	7	12.1	31	19.4	12	25.5	50	18.9
11	.888	1	48	82.7	111	69.9	27	57.5	186	70.5
		2-3	3	5.2	13	8.1	3	6.3	19	7.2
		ไม่ตอบ	7	12.1	35	22.0	17	36.2	59	22.3
3	.941	1	37	63.8	89	56.0	22	46.8	148	56.1
		2-3	18	31.0	39	24.5	11	23.4	68	25.8
		ไม่ตอบ	3	5.2	31	19.5	14	29.8	48	18.1
34	1.065	1	43	74.1	89	56.0	25	53.2	157	59.5
		2-3	8	13.8	21	13.2	6	12.7	35	13.2
		ไม่ตอบ	7	12.1	49	30.8	16	34.1	72	27.3
14	3.140	1	48	82.8	100	63.0	24	51.1	172	65.2
		2-3	5	8.6	26	16.3	8	17.0	39	14.8
		ไม่ตอบ	5	8.6	33	20.7	15	31.9	53	20.0

จากตารางที่ ข 2 เมื่อพิจารณารูปแบบการเลือกชุดตัวถูก จะเห็นว่าส่วนใหญ่จะเลือกตัวถูกเพียง 1 ตัว (ข้อสอบ 4 ตัวเลือก) มีส่วนน้อยที่เลือกตัวถูก 2 – 3 ตัว โดยการเลือกตัวถูก 2 –3 ตัว มีร้อยละของการเลือกแปรเปลี่ยนไปตามความยากของข้อสอบและความสามารถของกลุ่มตัวอย่าง จากการแบ่งข้อสอบตามค่าความยากเป็น 2 กลุ่ม คือ กลุ่มข้อสอบที่มีค่าความยาก ต่ำกว่า .000 มีทั้งหมด 22 ข้อ (จากตารางที่ ข 2 เรียงจากข้อ 1 – 18) และกลุ่มข้อสอบที่มีค่าความยากสูงกว่าหรือเท่ากับ .000 มีทั้งหมด 28 ข้อ (จากตารางที่ ข 2 เรียงจากข้อ 40 - 14) ส่วนความสามารถของกลุ่มแบ่งเป็น กลุ่มที่มีความสามารถสูง ปานกลาง ต่ำ จากการใช้คะแนนความสามารถจากการสอบกลางภาคเป็นตัวแบ่งกลุ่มที่มีเกณฑ์ คือ 1) ความสามารถ .75 ขึ้นไป เป็นกลุ่มที่มีความสามารถสูง 2) ความสามารถ (-.75)-.75 เป็นกลุ่มที่มีความสามารถปานกลาง และ3) ความสามารถต่ำกว่า-.75 เป็นกลุ่มที่มีความสามารถต่ำ ผลที่ได้มีรายละเอียดดังนี้

ในกลุ่มข้อสอบที่มีค่าความยาก ต่ำกว่า .000 (ค่อนข้างง่าย) พบว่าร้อยละของการตอบแบบเลือกตัวถูก 2 –3 ตัว ในกลุ่มที่มีความสามารถสูง มีค่าอยู่ระหว่าง 0.0 – 32.8 กลุ่มที่มีความสามารถปานกลางมีค่าอยู่ระหว่าง 3.7 – 24.5 กลุ่มที่มีความสามารถต่ำมีค่าอยู่ระหว่าง 4.2 – 31.9 และในภาพรวมทั้งหมดมีค่าอยู่ระหว่าง 5.7 – 26.1

ในกลุ่มข้อสอบที่มีค่าความยาก สูงกว่าหรือเท่ากับ .000 (ค่อนข้างยาก) พบว่าร้อยละของการตอบแบบเลือกตัวถูก 2 –3 ตัว ในกลุ่มที่มีความสามารถสูง มีค่าอยู่ระหว่าง 1.7 –31.0 กลุ่มที่มีความสามารถปานกลางมีค่าอยู่ระหว่าง 4.3 – 24.5 กลุ่มที่มีความสามารถต่ำมีค่าอยู่ระหว่าง 0.0 – 23.4 และในภาพรวมทั้งหมดมีค่าอยู่ระหว่าง 4.5 – 25.8

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ค.
ตัวอย่างคำสั่งที่ใช้ในการวิเคราะห์ข้อมูล

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตัวอย่างคำสั่งโปรแกรม Parscale
(วิเคราะห์แบบสอบแบบตัดตัวลวง)

```
ANALYZE FILE ELIMINATION FOR 50 ITEMS
TOTAL STUDENTS 252
>COMMENT
  The data are 4 rate
>FILES  DFNAME='D:\PSL2\ELIMIN.DAT',SAVE;
>SAVE   SCORE='D:\PSL2\ELIMIN.SCO',INFORMATION='D:\PSL2\ELIMIN.INF',
          PARM='D:\PSL2\ELIMIN.PAR';
>INPUT  NTEST=1,LENGTH=(50),NFMT=1,NIDCH=4,NTOTAL=50,MAXCAT=4;
(4A1,1X,50A1)
>TEST   TNAME=ELIMIN,ITEMS=(1(1)50),NBLK=1;
>BLOCK  BNAME=BLOCK1,NITEMS=50,NCAT=4,ORIGINAL=(0,1,2,3),
      MODIFIED=(0,1,2,3);
>CALIB  PARTIAL,SCALE=1.7,LOGISTIC,CYCLES=(100,1,1,1,1),
      NQPTS=9,CRIT=0.001,SPRIOR,TPRIOR;
>SCORE  EAP,PRINT,NAME=ELIMIN;
```

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตัวอย่างคำสั่งโปรแกรม Parscale
(วิเคราะห์แบบสอบแบบเลือกชุดตัวถูก)

```
ANALYZE FILE SELECTED FOR 50 ITEMS
TOTAL STUDENTS 252
>COMMENT
  THE DATA ARE 4 RATES
>FILES DFNAME='D:\PSL2\SELECTED.DAT',SAVE;
>SAVE  SCORE='D:\PSL2\SELECTED.SCORE',INFORMATION='D:\PSL2\SELECTED.INF',
    PARM='D:\PSL2\SELECTED.PAR';
>INPUT NTEST=1,LENGTH=(50),NFMT=1,NIDCH=4,NTOTAL=50,MAXCAT=4;
(4A1,1X,50A1)
>TEST  TNAME=SELECTED,ITEMS=(1(1)50),NBLK=1;
>BLOCK BNAME='BLOCK1',NITEMS=50,NCAT=4,ORIGINAL=(0,1,2,3);
>CALIB PARTIAL,SCALE=1.7,LOGISTIC,CYCLES=(10,1,1,1,1),
    NQPTS=9,CRIT=0.001,SPRIOR,TPRIOR;
>SCORE EAP,PRINT,NAME=SELECTED;
```

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตัวอย่างคำสั่งโปรแกรม Bilog
(วิเคราะห์แบบสอบแบบประเพณีนิยม (0,1))

```
>TITLE   DETECTION OF DICHOTOMUS FROM MULTIPLE CHOICE 0 1
>COMMENTS DATA FROM STUDENT 262 PERSONS
>GLOBAL DFNAME='D:\BILOG\MC.DAT',NPARM=3,LOGISTIC,SAVE;
>SAVE   SCORE='D:\BILOG\MC.SCO',PARME='D:\BILOG\MC.PAR',
     COV='D:\BILOG\MC.COV',TSTat='D:\BILOG\MC.TST';
>LENGTH NITEMS=50;
>INPUT  NTOT=50,NALT=2,NIDC=4,KFNAME='D:\BILOG\MC.DAT',
     OFNAME='D:\BILOG\MC.DAT';
     (4A1,1X,50A1)
>TEST   TNAME=MC;
>CALIB  TPRIOR, SPRIOR, GPRIOR, FLOAT, CYCLES=100, NEWTON=2, CRIT=0.001;
>SCORE  RSCTYPE=4,INFO=2,YCO,POP;
```

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตัวอย่างคำสั่งโปรแกรม Bilog
(วิเคราะห์แบบสอบกลางภาค (0,1))

```
>TITLE    DETECTION OF SEPERATE GROUP FROM MIDTERM
>COMMENTS DATA FROM STUDENT 297 PERSONS
>GLOBAL DFNAME='D:\BILOG\KNOW.DAT',NPARM=3,LOGISTIC,SAVE;
>SAVE   SCORE='D:\BILOG\KNOW.SCO',PARME='D:\BILOG\KNOW.PAR',
     COV='D:\BILOG\KNOW.COV',TSTAT='D:\BILOG\KNOW.TST';
>LENGTH NITEMS=65;
>INPUT  NTOT=65,NALT=2,NIDC=13,KFNAME='D:\BILOG\KNOW.DAT',
     OFNAME='D:\BILOG\KNOW.DAT';
     (13A1,2X,65A1)
>TEST   TNAME=KNOW;
>CALIB  TPRIOR, SPRIOR, GPRIOR, FLOAT, CYCLES=100, NEWTON=2, CRIT=0.001;
>SCORE  RSCTYPE=4,INFO=2,YCO,POP;
```

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ง.
ตัวอย่างคำชี้แจงของแบบสอบที่มีการให้คะแนนความรู้บางส่วน

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

คำชี้แจงในการตอบข้อสอบแบบเลือกตอบ ที่มีการวัดความความรู้บางส่วน **ประเภทตัดตัวลวง**

ข้อสอบต่อไปนี้เป็นข้อสอบที่วัดความรู้ในเรื่องการวัดและประเมินผลการศึกษา มีทั้งหมด **50** ข้อ ให้เวลา **1 ชั่วโมง 30 นาที** แต่ละข้อมี 4 ตัวเลือก คือ ก – ง ข้อสอบแต่ละข้อมีคำตอบที่ถูกต้องเพียงคำตอบเดียว มีวิธีการตอบที่แตกต่างจากแบบสอบเลือกตอบโดยทั่วไป ดังนี้

1. ในกรณีที่นิสิตทราบตัวเลือกที่ผิดทั้งหมดคือ 3 ตัวเลือก ให้กาเครื่องหมาย X ลงในช่องที่ตรงกับตัวเลือกนั้นในกระดาษคำตอบ ดังตัวอย่าง

ข้อ 0	ก	ข	ค	ง
		X	X	X

จากตัวอย่างแสดงว่านิสิตแน่ใจว่าตัวเลือก ก เป็นคำตอบที่ถูกต้อง ส่วนตัวเลือกที่เหลือเป็นคำตอบที่ผิด

2. ในกรณีที่นิสิตทราบตัวเลือกที่ผิดไม่หมด คือทราบเพียง 2 ตัวเลือก ให้กาเครื่องหมาย X ลงในช่องที่ตรงกับตัวเลือกนั้นในกระดาษคำตอบ ดังตัวอย่าง

ข้อ 0	ก	ข	ค	ง
		X	X	

จากตัวอย่างแสดงว่านิสิตไม่แน่ใจระหว่าง ตัวเลือก ก และ ง ว่าตัวเลือกใด เป็นคำตอบที่ถูก แต่แน่ใจว่า ตัวเลือก ข และตัวเลือก ง เป็นคำตอบที่ผิด

3. ในกรณีที่นิสิตทราบตัวเลือกที่ผิดไม่หมด คือทราบเพียง 1 ตัวเลือก ให้กาเครื่องหมาย X ลงในช่องที่ตรงกับตัวเลือกนั้นในกระดาษคำตอบ ดังตัวอย่าง

ข้อ 0	ก	ข	ค	ง
		X		

จากตัวอย่างแสดงว่านิสิตไม่แน่ใจระหว่าง ตัวเลือก ก ค และ ง ว่าตัวเลือกใด เป็นคำตอบที่ถูกแต่แน่ใจว่า ตัวเลือก ข เป็นคำตอบที่ผิด

4. ในกรณีที่นิสิตไม่ทราบว่าตัวเลือกใดเป็นตัวเลือกที่ผิดเลย ให้เว้นว่างไว้ ดังตัวอย่าง

ข้อ 0	ก	ข	ค	ง

จากตัวอย่างแสดงว่านิสิตไม่ทราบเลยว่าตัวเลือกใดผิด

ให้นิสิตตอบตามความสามารถที่แท้จริง **จงอย่าใช้การเดา** ไม่จำเป็นต้องตัดตัวเลือกทั้ง 3 ตัวในแต่ละข้อ แต่พยายามตัดออกตามความรู้ที่สามารถตัดออกได้เท่านั้น ถ้าไม่สามารถตัดออกได้ก็ไม่ต้องตัดออก

กรณีที่ตัดตัวเลือกที่ผิดได้ถูกต้องทั้ง 3 ตัวเลือก จะได้คะแนนเต็ม

กรณีที่ไม่แน่ใจ ตัดตัวเลือกที่ผิดได้บ้างจะมีการพิจารณาให้คะแนนความรู้บางส่วน แต่ถ้าตัดตัวเลือกที่เป็นคำตอบถูกออกไปจะไม่ได้รับการพิจารณาให้คะแนน

จึงเป็นสิ่งสำคัญที่นิสิตต้องตอบตามความสามารถที่แท้จริง

เมื่ออ่านคำชี้แจงเข้าใจแล้วลงมือทำข้อสอบได้ ห้ามขีดเขียนใดๆลงในกระดาษคำถาม

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

คำชี้แจงในการตอบข้อสอบแบบเลือกตอบ ที่มีการวัดความความรู้บางส่วน **ประเภทเลือกชุดตัวเลือกถูก**

ข้อสอบต่อไปนี้เป็นข้อสอบที่วัดความรู้ในเรื่องการวัดและประเมินผลการศึกษา มีทั้งหมด **50** ข้อ ให้เวลา **1 ชั่วโมง 30 นาที** แต่ละข้อมี 4 ตัวเลือก คือ ก – ง ข้อสอบแต่ละข้อมีคำตอบที่ถูกต้องเพียงคำตอบเดียว มีวิธีการตอบ**ประเภทเลือกชุดตัวเลือกถูก**ที่แตกต่างจากแบบสอบเลือกตอบโดยทั่วไป ดังนี้

2. ในกรณีที่นิสิตทราบตัวเลือกถูกที่แท้จริง คือ มีเพียง 1 ตัวเลือก ให้กาเครื่องหมาย ✓ ลงในช่องที่ตรงกับตัวเลือกนั้นในกระดาษคำตอบ ดังตัวอย่าง

ข้อ 0

ก	ข	ค	ง
✓			

จากตัวอย่างแสดงว่านิสิตแน่ใจว่าตัวเลือก ก เป็นคำตอบที่ถูกต้อง ส่วนตัวเลือกที่เหลือเป็นคำตอบที่ผิด

2. ในกรณีที่นิสิตไม่แน่ใจระหว่างตัวเลือก 2 ตัวว่าตัวเลือกใดถูก ให้กาเครื่องหมาย ✓ ลงในช่องที่ตรงกับตัวเลือกนั้นในกระดาษคำตอบ ดังตัวอย่าง

ข้อ 0

ก	ข	ค	ง
✓			✓

จากตัวอย่างแสดงว่านิสิตไม่แน่ใจระหว่าง ตัวเลือก ก และ ง ว่าตัวเลือกใด เป็นคำตอบที่ถูก แต่แน่ใจว่า ตัวเลือก ข และตัวเลือก ง เป็นคำตอบที่ผิด

3. ในกรณีที่นิสิตไม่แน่ใจระหว่างตัวเลือก 3 ตัวว่าตัวเลือกใดถูก ให้กาเครื่องหมาย ✓ ลงในช่องที่ตรงกับตัวเลือกนั้นในกระดาษคำตอบ ดังตัวอย่าง

ข้อ 0

ก	ข	ค	ง
✓		✓	✓

จากตัวอย่างแสดงว่านิสิตไม่แน่ใจระหว่าง ตัวเลือก ก ค และ ง ว่าตัวเลือกใด เป็นคำตอบ

ที่ถูกแต่แน่ใจว่า ตัวเลือก ข เป็นคำตอบที่ผิด

4. ในกรณีที่นิสิตไม่ทราบว่าตัวเลือกใดเป็นตัวเลือกที่ถูกเลย ให้เว้นว่างไว้ ดังตัวอย่าง

ข้อ 0	ก	ข	ค	ง

จากตัวอย่างแสดงว่านิสิตไม่ทราบเลยว่าตัวเลือกใดถูก

ให้นิสิตตอบตามความสามารถที่แท้จริง **จงอย่าใช้การเดา** ไม่จำเป็นต้องเลือกตัวเลือกเพียงตัวเดียวในแต่ละข้อ แต่พยายามเลือกตัวเลือกที่ถูกตามความรู้ ความสามารถที่แท้จริง ถ้าไม่สามารถเลือกตัวเลือกที่ถูกได้เลย ให้เว้นว่างข้อนั้นไว้

กรณีที่เลือกตัวเลือกเพียงตัวเดียว และเป็นตัวเลือกที่ถูก จะได้คะแนนเต็ม

กรณีที่ไม่แน่ใจสามารถเลือกชุดตัวเลือก ที่มากกว่า 1 ข้อ ถ้าชุดตัวเลือกที่เลือกมีตัวถูกอยู่ จะมีการพิจารณาให้คะแนนความรู้บางส่วน แต่ถ้าชุดตัวเลือกที่เลือกไม่มีตัวถูกอยู่ จะไม่ได้รับการพิจารณาให้คะแนน

จึงเป็นสิ่งสำคัญที่นิสิตต้องตอบตามความสามารถที่แท้จริง

เมื่ออ่านคำชี้แจงเข้าใจแล้วลงมือทำข้อสอบได้ ห้ามขีดเขียนใดๆลงในกระดาษคำถาม

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ประวัติของผู้วิจัย

อาจารย์ ดร.เอมอร จังศิริพรปกรณ์

คุณวุฒิ

วท.บ.(พยาบาล) มหาวิทยาลัยมหิดล 2526

วท.ม. (สาธารณสุขศาสตร์) มหาวิทยาลัยมหิดล 2529

ประกาศนียบัตร" หลักสูตรการวัดผลการศึกษา" สำนักทดสอบทางการศึกษา มหาวิทยาลัย ศรีนครินทรวิโรฒ 2532

ค.ด.(การวัดและประเมินผลการศึกษา) จุฬาลงกรณ์มหาวิทยาลัย 2542

ประวัติการทำงาน

2543-ปัจจุบัน	ดำรงตำแหน่ง อาจารย์ประจำภาควิชาวิจัยการศึกษา
2542-2543	ผู้ช่วยอธิการบดี มหาวิทยาลัยคริสเตียน
2529-2537	อาจารย์ประจำและหัวหน้าภาควิชาการพยาบาลอนามัยชุมชน วิทยาลัยคริสเตียน

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย