

371.26213

๐ ๗๘๕ ก

ร.๓

การเปรียบเทียบความเชื่อมั่นของแบบทดสอบอิงเกณฑ์
ที่มีการจัดเรียงข้อสอบและการกำหนดคะแนนจุดตัดต่างกัน

19 S.A. 2539

ปริญญานิพนธ์ของ

ของ

อุดม ชูลีวรรณ

เสนอต่อมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต วิชาเอกการวัดผลการศึกษา

พฤษภาคม 2539

ลิขสิทธิ์เป็นของมหาวิทยาลัยศรีนครินทรวิโรฒ

B. 52237

คณะกรรมการควบคุมและคณะกรรมการสอบได้พิจารณาปริญญานิพนธ์ฉบับนี้แล้ว
เห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาการศึกษามหาบัณฑิต วิชาเอก
การวัดผลการศึกษา ของมหาวิทยาลัยศรีนครินทรวิโรฒได้

คณะกรรมการควบคุม

..... ประธาน

(รศ. อังคณา สายยศ)

..... กรรมการ

(รศ. วัลัญญา วิชาลาภรณ์)

คณะกรรมการสอบ

..... ประธาน

(รศ. อังคณา สายยศ)

..... กรรมการ

(รศ. วัลัญญา วิชาลาภรณ์)

..... กรรมการแต่งตั้งเพิ่มเติม

(รศ. ชุศรี วงศ์รัตนะ)

บัณฑิตวิทยาลัยอนุมัติให้รับปริญญานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาการศึกษามหาบัณฑิต วิชาเอกการวัดผลการศึกษา ของมหาวิทยาลัยศรีนครินทรวิโรฒ

..... คณบดีบัณฑิตวิทยาลัย

(ดร. ศิริยุภา พูลสุวรรณ)

วันที่ ๑ เดือน พฤษภาคม พ.ศ. 2539

ประกาศคุณูปการ

ปริญญานิพนธ์ฉบับนี้ สำเร็จได้ด้วยความกรุณาช่วยเหลือแนะนำเป็นอย่างดีจาก รองศาสตราจารย์อังคณา สายยศ รองศาสตราจารย์วิญญา วิศาลภรณ์ ประธานและ กรรมการ ที่ปรึกษาทำปริญญานิพนธ์ ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูง

ขอกราบขอบพระคุณ รองศาสตราจารย์ชูศรี วงศ์รัตน์ กรรมการสอบปริญญานิพนธ์ ที่ได้กรุณาช่วยเหลือแนะนำ เพิ่มเติม เพื่อให้ปริญญานิพนธ์ฉบับนี้สมบูรณ์ยิ่งขึ้น

ขอกราบขอบพระคุณรองศาสตราจารย์นภาพา ศรีไพโรจน์ รองศาสตราจารย์ บุษกร เพชรวิวรรณ ผู้ช่วยศาสตราจารย์ จิราภรณ์ บุญส่ง ผู้ช่วยศาสตราจารย์เชาวนา ขวลิธ อ่าง อาจารย์ระวีวรรณ พันธุ์พานิช อาจารย์อำนาจ ยิ่งคำนิ้ง อาจารย์เบญจะ สุชาติพงศ์ อาจารย์สมศักดิ์ เพชรมาลา ที่ได้กรุณาเป็นผู้เชี่ยวชาญตรวจสอบเครื่องมือวิจัย และให้ข้อเสนอแนะ ต่าง ๆ เป็นอย่างดียิ่ง

ขอกราบขอบพระคุณ ผู้อำนวยการสามัญศึกษาจังหวัดนครศรีธรรมราช ผู้บริหารโรงเรียน คณะครูอาจารย์ และนักเรียนในสังกัดกรมสามัญศึกษาจังหวัดนครศรีธรรมราช ที่ให้ความกรุณา เป็นอย่างดียิ่งในการเก็บรวบรวมข้อมูล

ขอขอบคุณ อาจารย์สุพร พันธุ์สุภา อาจารย์พิชญา เกตสกุล อาจารย์จิตสุตา ธราพร อาจารย์ธำนิทร์ เสนียวงษ์ ณ อยุธยา และเพื่อนบัณฑิตศึกษา วิชาเอกการวัดผลการศึกษา รุ่น 30 ทุกท่าน ที่ให้ความช่วยเหลือในการวิเคราะห์ข้อมูล และให้กำลังใจในการทำปริญญานิพนธ์ ฉบับนี้

คุณค่าและประโยชน์ของปริญญานิพนธ์ฉบับนี้ ผู้วิจัยขออุทิศ แต่พระคุณของบิดา มารดา ญาติพี่น้อง และครูอาจารย์ ที่ได้ให้การอบรมสั่งสอน ตลอดจนให้การสนับสนุนการศึกษาแก่ผู้วิจัย ตลอดมา

อุดม ชูสิวรรณ

สารบัญ

บทที่	หน้า
1 บทนำ.....	1
ภูมิหลัง.....	1
จุดมุ่งหมายในการศึกษาค้นคว้า.....	5
ความสำคัญของการศึกษาค้นคว้า.....	5
ขอบเขตของการศึกษาค้นคว้า.....	5
นิยามศัพท์เฉพาะ.....	6
2 เอกสารและงานวิจัยที่เกี่ยวข้อง.....	8
ความหมายและลักษณะของแบบทดสอบอิงเกณฑ์.....	8
การสร้างแบบทดสอบอิงเกณฑ์.....	9
การเขียนข้อสอบวัดพฤติกรรมการเรียนรู้.....	12
ความหมายและการกำหนดคะแนนจุดตัด.....	18
ความเชื่อมั่นของของแบบทดสอบอิงเกณฑ์.....	38
การจัดเรียงข้อคำถามในแบบทดสอบ.....	55
งานวิจัยที่เกี่ยวข้องกับการศึกษาค้นคว้า.....	56
งานวิจัยต่างประเทศ.....	56
งานวิจัยภายในประเทศ.....	58
สมมุติฐานในการศึกษาค้นคว้า.....	60
3 วิธีดำเนินการศึกษาค้นคว้า.....	62
ประชากร.....	62
กลุ่มตัวอย่าง.....	62
เนื้อหาที่ใช้ในการศึกษาค้นคว้า.....	63
เครื่องมือที่ใช้ในการศึกษาค้นคว้า.....	63
การดำเนินการสร้างเครื่องมือในการศึกษาค้นคว้า.....	63
วิธีดำเนินการศึกษาค้นคว้า.....	68
ลำดับขั้นการวิเคราะห์ข้อมูล.....	68
สถิติที่ใช้ในการวิเคราะห์ข้อมูล.....	69

บทที่	หน้า
4 ผลการวิเคราะห์ข้อมูล.....	76
สัญลักษณ์และอักษรที่ใช้ในการวิเคราะห์ข้อมูล.....	76
การเสนอผลการวิเคราะห์ข้อมูล.....	77
ผลการวิเคราะห์ข้อมูล.....	77
ค่าสถิติพื้นฐานของแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม.....	77
คะแนนจุดตัดของแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่มเมื่อกำหนดคะแนนจุดตัดด้วยวิธีของเบส์ และวิธีของแกลส.....	78
ค่าความเชื่อมั่นของแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม.....	79
การทดสอบค่าความเชื่อมั่นของแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหาและเรียงโดยวิธีสุ่ม.....	81
5 สรุปผล อภิปรายผล และข้อเสนอแนะ.....	85
จุดมุ่งหมายของการศึกษาค้นคว้า.....	85
กลุ่มตัวอย่าง.....	85
เครื่องมือที่ใช้ในการศึกษาค้นคว้า.....	85
การเสนอผลการวิเคราะห์ข้อมูล.....	86
สรุปผลการวิเคราะห์ข้อมูล.....	86
อภิปรายผล.....	89
ข้อเสนอแนะ.....	91
บรรณานุกรม.....	93
ภาคผนวก.....	102
ประวัติผู้วิจัย.....	120

บัญชีตาราง

ตาราง	หน้า
1 การคำนวณตามเทคนิคของนีเดลสกี.....	25
2 ผลการหาคะแนนจุดตัดจากผู้เชี่ยวชาญโดยเนคนิคของนีเดลสกี.....	27
3 ผลการตัดสินโดยใช้เทคนิคของแองกอฟฟ์.....	29
4 ร้อยละที่คาดหวังจะสอบผ่านตามลักษณะที่เกี่ยวข้องและระดับความยาก.....	30
5 ตัวอย่างของการคำนวณตามเทคนิคของฮีเบล.....	31
6 จำนวนนักเรียนที่ใช้เป็นกลุ่มตัวอย่างจำแนกตามขนาดโรงเรียน จำนวนห้องเรียน และจำนวนนักเรียน.....	62
7 ผลการวิเคราะห์เนื้อหาและพฤติกรรมเรื่องสมการชั้นมัธยมศึกษาปีที่ 1.....	64
8 ค่าสถิติพื้นฐานของแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม.....	78
9 คะแนนจุดตัดของแบบทดสอบที่จัดเรียงข้อสอบต่างกัน 3 รูปแบบ คือเรียงตาม ค่าความยากเรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม เมื่อกำหนดคะแนนจุดตัด ด้วยวิธีของเบส์ และวิธีของเกลส.....	79
10 ค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม โดยใช้คะแนนจุดตัดตามวิธีของเบส์ และวิธีของเกลส ซึ่งคำนวณจากสูตรความเชื่อมั่นของโลเวท.....	80
11 ค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ ที่มีการจัดเรียงข้อสอบตาม ค่าความยาก เรียงตามเรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม ที่คำนวณจาก สูตรของลิวินสตัน โดยใช้คะแนนจุดตัดวิธีของเบส์ และวิธีของเกลส.....	81
12 ผลการทดสอบความแตกต่างของค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ ที่มีการจัดเรียงข้อสอบต่างกัน 3 รูปแบบคือเรียงตามค่าความยาก เรียงตาม เนื้อหา และเรียงโดยวิธีสุ่ม ซึ่งคำนวณค่าความเชื่อมั่นจากคะแนนจุดตัดตาม วิธีของเบส์ และวิธีของเกลส โดยใช้สูตรของโลเวท.....	82
13 ผลการทดสอบความแตกต่างของค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ ที่จัดเรียงข้อสอบต่างกัน 3 รูปแบบ เมื่อกำหนดคะแนนจุดตัดด้วยวิธีของเบส์	83
14 ผลการทดสอบความแตกต่างของค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ ที่มีการจัดเรียงข้อสอบเหมือนกัน เมื่อกำหนดคะแนนจุดตัดด้วยวิธีของเบส์ และวิธีของเกลส โดยใช้สูตรคำนวณค่าความเชื่อมั่นของโลเวท.....	84

ตาราง

หน้า

15	ค่าความยากง่าย (P) และค่าอำนาจจำแนก (B) รายข้อของแบบทดสอบ ก่อนการเก็บรวบรวมข้อมูล.....	102
16	ค่าความยาก (P) และค่าอำนาจจำแนก (B) ของแบบทดสอบที่จัดเรียง ข้อสอบตามค่าความยาก.....	110
17	ค่าความยาก (P) และค่าอำนาจจำแนก (B) ของแบบทดสอบที่จัดเรียง ข้อสอบตามเนื้อหา.....	111
18	ค่าความยาก (P) และ ค่าอำนาจจำแนก (B) ของแบบทดสอบที่จัดเรียง ข้อสอบโดยวิธีสุ่ม.....	112
19	คะแนนจุดตัด ค่าความสูญเสียที่คาดหวังของการตัดสินใจไม่ผ่าน ค่าความสูญเสีย ที่คาดหวังของการตัดสินใจผ่าน ค่าฟังก์ชัน และผลบวกของความคลาดเคลื่อน ของการจำแนกผิดประเภทที่ 1 และ 2 ของแบบทดสอบที่จัดเรียงข้อสอบตาม ค่าความยาก เมื่อกำหนดคะแนนจุดตัดตามวิธีของเบส์ และวิธีของแกลส.....	114
20	คะแนนจุดตัด ค่าความสูญเสียที่คาดหวังของการตัดสินใจไม่ผ่าน ค่าความสูญเสีย ที่คาดหวังของการตัดสินใจผ่าน ค่าฟังก์ชัน และผลบวกของความคลาดเคลื่อน ของการจำแนกผิดประเภทที่ 1 และ 2 ของแบบทดสอบที่จัดเรียงข้อสอบตาม เนื้อหา เมื่อกำหนดคะแนนจุดตัดตามวิธีของเบส์ และวิธีของแกลส.....	116
21	คะแนนจุดตัด ค่าความสูญเสียที่คาดหวังของการตัดสินใจไม่ผ่าน ค่าความสูญเสีย ที่คาดหวังของการตัดสินใจผ่าน ค่าฟังก์ชัน และผลบวกของความคลาดเคลื่อน ของการจำแนกผิดประเภทที่ 1 และ 2 ของแบบทดสอบที่จัดเรียงข้อสอบ โดยวิธีสุ่ม เมื่อกำหนดคะแนนจุดตัดตามวิธีของเบส์ และวิธีของแกลส.....	118

บทที่ 1

บทนำ

ภูมิหลัง

การศึกษาเป็นกระบวนการที่มุ่งพัฒนาบุคคลให้มีความเจริญงอกงามตามทิศทางการสังคม ปรารถนา โดยพยายามเปลี่ยนแปลงพฤติกรรมของผู้เรียน ให้มีคุณสมบัติตรงตามความต้องการของสังคม (ไพศาล หวังพานิช. 2526 : 1) การวัดผลการศึกษาเป็นเครื่องมืออันหนึ่งที่จะช่วยพัฒนาคุณภาพของการศึกษา ในระดับการศึกษาต่าง ๆ เพราะผลจากการวัดผลจะเป็นพื้นฐานในการตัดสินใจของครูและนักการศึกษา เพื่อใช้ปรับปรุงการเรียนการสอน การแนะแนว การประเมินผล หลักสูตร แบบเรียน การใช้อุปกรณ์การสอน ดังนั้น การวัดและการประเมินผลนอกจากมีประโยชน์ต่อการปรับปรุงคุณภาพของการศึกษาแล้ว ยังนับว่ามีความสำคัญ และจำเป็นต่อการตัดสินใจในการให้การศึกษาในระดับต่าง ๆ อีกด้วย (อนันต์ ศรีโสภา. 2524 : 1 - 2)

การวัดผลการศึกษาทำหน้าที่โดยตรง ในการตรวจสอบผลของการตรวจสอบผลของการเรียนการสอนว่า ผู้เรียนเกิดลักษณะต่าง ๆ ตามที่กำหนดไว้มากน้อยเพียงใด ได้ผลระดับใด ผู้เรียนมีความสามารถชนิดใดบ้าง และยังมีสิ่งใดที่บกพร่องควรแก่การแก้ไขบ้าง (ไพศาล หวังพานิช. 2526 : 3) นอกจากนี้ ชาวาล แพร์ตกุล (2514 : 5) ได้กล่าวไว้ว่า ขบวนการวัดผลในปัจจุบันเป็นการทดสอบเพื่อค้นและพัฒนาสมรรถภาพของมนุษย์ ซึ่งโกวิท ประวาลพุกษ์ (2523 : 17) กล่าวว่า กระบวนการวัดผลที่นิยมใช้ในปัจจุบันมี 2 แบบ คือ การประเมินผลแบบอิงกลุ่ม (Norm-Referenced Evaluation) และการประเมินผลแบบอิงเกณฑ์ (Criterion-Referenced Evaluation)

การประเมินผลแบบอิงกลุ่ม (Norm-Referenced Evaluation) เป็นการนำผลที่ได้จากการวัดไปเปรียบเทียบกับคนอื่น ๆ ในกลุ่มที่ถูกวัดด้วยแบบทดสอบฉบับเดียวกัน (Popham. 1978 : 24 - 25) การวัดและการประเมินผลแบบนี้ จึงเป็นการวัดที่ขึ้นอยู่กับพฤติกรรมของกลุ่มเป็นสำคัญ โดยมีความเชื่อในเรื่องความแตกต่างระหว่างบุคคลที่ว่า แต่ละคนมีความสามารถไม่เท่ากัน ดังนั้นการวัดและการประเมินผลแบบนี้ จึงเป็นการหาความแตกต่างระหว่างบุคคล แต่ละคนอยู่ตรงไหนของกลุ่ม และคะแนนที่ได้จะมีความหมายเมื่อนำไปเปรียบเทียบกับคะแนนในกลุ่มที่สอบด้วยแบบทดสอบฉบับเดียวกัน โดยนำคะแนนที่ได้มาตัดสินในรูปตำแหน่งเปอร์เซ็นต์ไทล์หรือคะแนนมาตรฐานอื่น ๆ จะเห็นได้ว่า การวัดและการประเมินผลแบบนี้เป็นการจัดอันดับความสามารถของผู้เรียนว่า ผู้เรียนมีความรู้ในเรื่องที่เรียนมากหรือน้อยกว่าคนอื่น ๆ ในกลุ่ม ซึ่งผลการ

สอบมิได้บอกให้เรา ทราบว่า นักเรียนมีความรอบรู้หรือมีความสามารถที่จะทำงานด้านต่าง ๆ ตาม จุดมุ่งหมายของการฝึกฝนทางการศึกษามากน้อยเพียงใด (Gronlund. 1973) ในลักษณะของ ความเป็นจริง ผลของการสอบควรจะสามารถบอกได้ว่า ผู้สอบควรจะสามารถแสดงพฤติกรรมและ สามารถทำอะไรได้บ้าง ดังนั้นการสอบวัดที่ดีในการเรียนการสอน ควรมีข้อมูลที่จะใช้เพื่อพัฒนาคน ไปพร้อม ๆ กันด้วย กล่าวคือ เขาได้เรียนรู้ด้านใดอยู่ในระดับใด และเขาจะต้องใช้กลวิธีแบบใด จึงจะเรียนรู้ในสิ่งที่กำหนดเป็นจุดมุ่งหมายได้ ระบบการเรียนการสอนที่ดีจึงควรเป็นระบบสะสม ข้อมูลเกี่ยวกับตัวผู้เรียนในด้านต่าง ๆ ไว้สำหรับช่วยให้เขาเรียนรู้ได้ดียิ่งขึ้นเรื่อย ๆ การสอบจึงต้อง คำนึงถึงแนวพัฒนาการของเด็กด้วยว่า เขาใช้แนวใดและเขาอยู่ในแนวระดับใดในแนวทางนั้น ๆ ดังนั้นแบบทดสอบที่ดีจึงควรคำนึงแนวพัฒนาการเหล่านี้ด้วย โดยเมื่อสอบมาแล้ว เราก็สามารถ บอกได้ทันทีว่า เขาอยู่ในระดับพัฒนาการในช่วงใดและเขาต้องพัฒนาไปอีกอย่างไร ด้วยลักษณะ ดังกล่าวนี้นั้นที่จะช่วยให้การวัดมีคุณประโยชน์มากที่สุด สมตามเจตนารมณ์ของการพัฒนาใน ทุก ๆ ด้าน (โกวิท ประวาลพุกษ์. 2529 : 11) ดังนั้น นักวัดผลจึงได้พยายามพัฒนาการวัดและ การประเมินผล เพื่อให้สามารถวัดให้ได้ในลักษณะดังกล่าว ซึ่งการวัดและการประเมินผลในแนวนี้นี้ คือ การวัดและการประเมินผลแบบอิงเกณฑ์

การประเมินผลแบบอิงเกณฑ์ (Criterion-Referenced Evaluation) เป็นการ ประเมินผลความสามารถของบุคคลโดยนำความสามารถของแต่ละคนไปเปรียบเทียบกับเกณฑ์ที่ กำหนดไว้ (Popham. 1978 : 92 - 95) การแปลความหมายของการวัดผล โดยนำผลการวัด ของนักเรียนไปเปรียบเทียบกับเกณฑ์ที่กำหนดไว้ โดยไม่คำนึงถึงผลการวัดของนักเรียนคนอื่น ๆ (กมล ภูประเสริฐ. 2522 : 12 - 13) แต่เป็นการวัดที่มุ่งไปในลักษณะที่ว่า ผู้เรียนสามารถ ทำอะไรได้บ้าง ไม่สามารถทำอะไรได้บ้าง หลังจากที่มีการเรียนการสอนไปแล้ว ซึ่งเป็นการ ตรวจสอบความสามารถในการเรียนรู้ของผู้เรียนว่ามีพฤติกรรมและคุณลักษณะต่าง ๆ ถึงระดับของ เป้าหมายที่ต้องการหรือไม่ หรือเป็นไปตามความคาดหวังซึ่งกำหนดไว้เป็นเกณฑ์มากน้อยเพียงใด ถ้าผู้เรียนมีพฤติกรรมความสามารถถึงเกณฑ์ที่กำหนดไว้ ก็ถือว่า บรรลุเป้าหมาย และเป็นผู้ที่ เรียนรู้แล้ว (Mastered) ถ้าความสามารถต่ำกว่าเกณฑ์ก็ถือว่า ยังไม่ผ่าน หรือยังไม่เป็นไป ตามความคาดหวังต้องใช้เวลาสำหรับการแก้ไขปรับปรุงหรือซ่อมเสริม การวัดแบบอิงเกณฑ์จึงเป็น การเปรียบเทียบระหว่างความสามารถของผู้เรียนกับเกณฑ์ที่กำหนดไว้ ซึ่งเกณฑ์ (Criterion) คือ คุณลักษณะที่คาดหวังจะให้เกิดกับผู้เรียนหลังจากเสร็จสิ้นการเรียนการสอนแล้ว เพื่อใช้เป็น มาตรฐานขั้นต่ำสุด สำหรับการยอมรับคุณภาพของผู้เรียนและใช้เป็นเครื่องตัดสินว่า ผู้เรียนผ่าน หรือบรรลุเป้าหมายของการเรียนการสอนหรือไม่ (ไพศาล หวังพานิช. 2526 : 186 - 189) จะเห็นว่าการวัดและการประเมินผลแบบนี้สามารถนำมาใช้ เพื่อแยกผู้เรียนที่รู้แล้ว และที่ยังไม่รู้ โดยมีจุดประสงค์เพื่อวินิจฉัยว่า ผู้เรียนบรรลุวัตถุประสงค์ที่วางไว้หรือไม่ ผู้เรียนคนใดที่ยังไม่บรรลุ

วัตถุประสงค์ที่วางไว้ก็จะได้หาทางช่วยเหลือ จากที่กล่าวมาการวัดแบบอิงเกณฑ์นั้นจะต้องอาศัยเกณฑ์หรือคะแนนจุดตัด ซึ่งเกลสเซอร์ (Glasser. 1963) เป็นผู้หนึ่งที่ให้แนวคิดที่แบบสอบอิงเกณฑ์จะต้องมีจุดตัด (Cut-Off-Scores) เพื่อวินิจฉัยว่า ผู้เรียนคนใดเรียนรู้ครบถ้วนตามจุดประสงค์ที่กำหนดไว้ ดังนั้น แบบทดสอบที่ใช้เป็นเครื่องมือวัด จะต้องสร้างให้สอดคล้องกับจุดประสงค์และครอบคลุมโดเมนที่ต้องการให้แสดงพฤติกรรมออกมา แต่ปัญหาที่สำคัญของการวัดผลแบบอิงเกณฑ์ซึ่งแฮมเบิลตันและโนวิก (Hambleton and Novick. 1973 : 161) ได้กล่าวว่า ปัญหาเบื้องต้นของการวัดผลแบบอิงเกณฑ์คือการกำหนดคะแนนจุดตัด หรือคะแนนระดับความรอบรู้ (Master Level Score)

การกำหนดคะแนนจุดตัด เป็นสิ่งสำคัญอย่างยิ่งของการวัดผลแบบอิงเกณฑ์ (Linn. 1978 : 302) เนื่องจากคะแนนจุดตัดจะเป็นเกณฑ์ หรือมาตรฐานที่ใช้จำแนกระดับความสามารถของผู้สอบออกเป็นผู้รอบรู้ และไม่รอบรู้และยิ่งกว่านั้น คุณค่าของคะแนนจุดตัดยังมีอิทธิพลต่อการเรียนรู้ตลอดจนทัศนคติของผู้เรียนอีกด้วย (Hambleton and Others. 1978 : 27) เพราะฉะนั้นในการกำหนดคะแนนจุดตัด หากกำหนดสูงเกินไป จะทำให้เกิดความคลาดเคลื่อน คือ ผู้ที่รอบรู้จะกลายเป็นผู้ไม่รอบรู้ และถ้ากำหนดคะแนนจุดตัดต่ำเกินไปก็จะทำให้เกิดความคลาดเคลื่อนขึ้นเช่นเดียวกัน คือผู้ที่ไม่มีความรอบรู้ จะกลายเป็นผู้มีความรอบรู้ ลินน์ (Linn. 1978 : 307) ให้ข้อเสนอแนะว่าหากใช้กระบวนการที่มีระบบจะสามารถช่วยลดโอกาสที่จะทำให้เกิดความคลาดเคลื่อนได้ หรือทำให้เกิดความคลาดเคลื่อนได้น้อยที่สุด

วิธีการกำหนดคะแนนจุดตัดมีอยู่หลายวิธี เช่น วิธีนับลดจาก 100% (บุญเชิด ภิญโญนนันตพงศ์. 2526 : 156 อ้างอิงมาจาก Glass. 1978 : 237 - 261) โดยใช้ผู้เชี่ยวชาญหรือครูผู้สอนแต่ละคนพิจารณาว่า แต่ละจุดประสงค์มีความสำคัญมากน้อยเพียงใด แล้วพิจารณากำหนดคะแนนจุดตัดของแบบทดสอบจากเกณฑ์ที่ตั้งไว้ นอกจากนั้นยังมีการกำหนดคะแนนจุดตัดจากสมรรถภาพขั้นต่ำซึ่งเป็นวิธีการที่ได้มาจากการให้ผู้เชี่ยวชาญเนื้อหาวิชา หรือครูประจำวิชาศึกษาข้อสอบ แล้วระบุว่า นักเรียนที่มีสมรรถภาพขั้นต่ำ ต้องมีคะแนนสอบผ่านเท่ากับเท่าไร วิธีนี้มีหลายเทคนิคด้วยกัน เช่น เทคนิคของนีเดลสกี (Nedelsky) เทคนิคของอีเบล (Ebel) และเทคนิคของแองกอฟฟ์ (Angoff) เป็นต้น

เทคนิคของนีเดลสกี เป็นเทคนิคที่เหมาะสมกับแบบทดสอบเลือกตอบเท่านั้น การกำหนดคะแนนจุดตัดโดยใช้เทคนิคนี้จะต้องพิจารณาถึง ความยากของข้อคำถามเป็นรายตัวเลือกว่า ตัวเลือกใดที่ผู้สอบมีความสามารถต่ำสุดทราบได้ว่าเป็นตัวเลือกที่ผิด

เทคนิคของอีเบล เป็นเทคนิคที่ผู้เชี่ยวชาญจะต้องพิจารณาถึงความยากและความสำคัญของข้อคำถามประกอบกัน

เทคนิคของแองกอฟฟ์ เป็นเทคนิคที่ผู้เชี่ยวชาญจะต้องพิจารณาค่าความน่าจะเป็น หรือโอกาสที่ผู้สอบที่มีความสามารถต่ำสุด จะตอบข้อสอบในแต่ละข้อคำถามได้ถูกต้อง

การกำหนดคะแนนจุดตัดแต่ละวิธีดังกล่าวเป็นการกำหนด โดยอาศัยดุลพินิจของผู้เชี่ยวชาญตัดสินซึ่งมีลักษณะเป็นอัตนัยมาก (บุญเชิด ภิญโญนนตพงษ์. 2526 : 156 - 172)

อย่างไรก็ตาม งานวิจัยของพิศิษฐ์ ตันทวนิช (2528 : 43) ได้กล่าวไว้ว่า จากการที่ได้มีนักวิชาการเสนอวิธีการกำหนดเกณฑ์มาตรฐานขึ้นมามากในระยะต่อมา จึงได้มีการจัดประเภทและแยกหมวดหมู่แนวความคิดของการกำหนดเกณฑ์มาตรฐาน คือ เกณฑ์การใช้ข้อมูลเพื่อประกอบการตัดสินใจกำหนดเกณฑ์มาตรฐาน ซึ่งแบ่งออกได้เป็น 2 วิธีการคือ

1. การกำหนดเกณฑ์มาตรฐาน โดยใช้ข้อมูลจากผู้เชี่ยวชาญเพียงประการเดียว
2. การกำหนดเกณฑ์มาตรฐาน โดยการใช้ข้อมูลเชิงสัมพันธ์จากภายนอกเข้ามาเป็นส่วนประกอบในการพิจารณากำหนดเกณฑ์มาตรฐาน

นอกจากจะกำหนดคะแนนจุดตัดโดยวิธีการดังกล่าวแล้ว ยังสามารถกำหนดคะแนนจุดตัดโดยพิจารณาจากการทดลองสอบกับกลุ่มตัวอย่างนักเรียน เช่น โดยใช้ทฤษฎีการตัดสินใจ ซึ่งมีอยู่หลายวิธี เช่น วิธีการของเบลล์ วิธีการของแฟเนอร์ วิธีการของแกเลส วิธีการของเบอร์ก

นอกจากนั้น ฮวิน (สุทิวรรณ พิศักดีโสภณ. 2529 : 22 ; อ้างอิงมาจาก Huynh. 1980 : 365 - 366) ได้เสนอวิธีการกำหนดคะแนนจุดตัด โดยพิจารณาจากดัชนีอำนาจจำแนก P ที่สามารถจำแนกผู้ที่เรียนรู้แล้ว และผู้ที่ยังไม่เรียนรู้ได้ คะแนนจุดตัดจะเป็นคะแนนที่ทำให้ P มีค่าสูงสุด ซึ่งเรียกว่า P_{max}

นอกจากคะแนนจุดตัดแล้ว ความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ก็มีความสำคัญเป็นอย่างยิ่ง ซึ่งอนันต์ ศรีโสภณ (2525 : 59) กล่าวว่า คุณสมบัติที่สำคัญที่สุดของเครื่องมือวัดผลการศึกษา คือ ความเชื่อมั่นและความเที่ยงตรง นอกจากนี้ บุญเชิด ภิญโญนนตพงษ์ (2527 : 199) กล่าวว่า ความเชื่อมั่นของแบบทดสอบเป็นองค์ประกอบที่สำคัญของการสร้างแบบทดสอบ เพราะว่าค่าความเชื่อมั่นเป็นดัชนีบ่งชี้ว่า แบบทดสอบนั้นมีคุณภาพหรือไม่ในการสร้างแบบทดสอบและอีเบล (Ebel. 1965 : 20) กล่าวว่า ความเชื่อมั่นของแบบทดสอบขึ้นอยู่กับองค์ประกอบหลายประการ เช่น ลักษณะของคำถาม การดำเนินการสอบ เวลาเงื่อนไขในการสอบ การเรียงลำดับข้อสอบและส่วนประกอบอื่น ๆ อีกมากเกี่ยวกับการจัดเรียงข้อสอบ ชวาล แพรัตกุล (2516 : 193) ให้ความเห็นว่าข้อสอบแต่ละฉบับควรเรียงคำถามเริ่มจากง่ายไปหายากเสมอ ไม่ควรเรียงตามความยากง่ายของบทเรียน สำหรับแบบทดสอบอิงเกณฑ์เครเฮน (Crehen. 1974 : 255 - 262) ได้ศึกษาวิธีคัดเลือกข้อสอบ 6 วิธี คือวิธีการของคอกซ์-วาร์กัส (Cox-Vargas) วิธีของเบรนนัน (Brennan) วิธีเรียงตามค่าสัมประสิทธิ์สหสัมพันธ์แบบ Point Biserial และวิธีการสุ่มกลุ่มตัวอย่างแต่ละกลุ่มใช้ระหว่าง 7 - 30 คน โดยใช้แบบทดสอบคู่ขนานพบว่าวิธีการเรียงข้อสอบอิงเกณฑ์ทั้ง 6 วิธี ให้ผลต่อความเชื่อมั่นไม่แตกต่างกัน

เมื่อพิจารณาจากผลการวิจัย และแนวความคิดต่างๆ ที่กล่าวมาจะเห็นได้ว่า วิธีการกำหนดเกณฑ์มาตรฐานหรือคะแนนจุดตัดของแบบทดสอบอิงเกณฑ์ วิธีการของเกลส (GLass) วิธีการของเบส์ มีข้อดีที่สามารถนิยามความรอบรู้ในรูปของคะแนนมวลความรู้ เป็นวิธีการที่มีความคลาดเคลื่อนน้อย อีกทั้งสามารถนำไปใช้ในการจัดการเรียนการสอนและการวัดผลประเมินผลได้ ดังนั้น ผู้วิจัยมีความสนใจที่จะศึกษาคะแนนจุดตัด และความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ที่มีการจัดเรียงข้อสอบต่างกัน คือ เรียงตามลำดับค่าความยาก เรียงตามเนื้อหาย่อย และเรียงแบบสุ่ม โดยใช้วิธีการกำหนดจุดตัดตามวิธีการของเกลส วิธีการของเบส์ และใช้วิธีการประมาณค่าความเชื่อมั่นตามวิธีของโลเวท วิธีของลิวิงตัน

จุดมุ่งหมายของการศึกษาค้นคว้า

1. เพื่อเปรียบเทียบความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ที่มีการจัดเรียงข้อสอบต่างกันและมีการกำหนดคะแนนจุดตัดด้วยวิธีเดียวกัน
2. เพื่อเปรียบเทียบความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ ที่มีการจัดเรียงข้อสอบเหมือนกัน เมื่อกำหนดคะแนนจุดตัดด้วยวิธีต่างกัน

ความสำคัญของการศึกษาค้นคว้า

จากการศึกษาค้นคว้าครั้งนี้ จะทำให้ทราบถึงความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ที่มีการจัดเรียงข้อสอบตามลำดับค่าความยาก ตามเนื้อหาย่อย และแบบสุ่ม โดยใช้วิธีกำหนดคะแนนจุดตัดตามวิธีของเกลส วิธีของเบส์ และใช้วิธีการประมาณค่าความเชื่อมั่นโดยวิธีของโลเวท วิธีของลิวิงตัน ซึ่งในการศึกษาครั้งนี้จะเป็นแนวทางในการสร้างแบบทดสอบอิงเกณฑ์ที่จะเลือกใช้วิธีการหาคะแนนจุดตัดอันจะส่งผลให้แบบทดสอบมีค่าความเชื่อมั่นที่เหมาะสม

ขอบเขตของการศึกษาค้นคว้า

1. ประชากรที่ใช้ในการศึกษาครั้งนี้เป็น นักเรียนชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2538 สังกัดกรมสามัญศึกษา จังหวัดนครศรีธรรมราช จำนวน 67 โรงเรียน 443 ห้องเรียน จำนวนนักเรียน 18,811 คน
2. กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2538 สังกัดกรมสามัญศึกษาจังหวัดนครศรีธรรมราชจำนวน 1,200 คน

ซึ่งได้มาโดยการสุ่มแบบแบ่งชั้น (Stratified Random Sampling) มีขนาดโรงเรียน เป็นชั้น (Strata) และมีห้องเรียนเป็นหน่วยการสุ่ม (Sampling Unit)

3. ตัวแปรที่ใช้ในการศึกษาครั้งนี้

3.1 ตัวแปรอิสระ (Independent Variables) คือ

3.1.1 รูปแบบการเรียงข้อสอบจำแนกได้ดังนี้

3.1.1.1 เรียงตามค่าความยาก

3.1.1.2 เรียงตามเนื้อหาย่อย

3.1.1.3 เรียงโดยวิธีสุ่ม

3.1.2 วิธีการกำหนดคะแนนจุดตัด ได้แก่

3.1.2.1 วิธีของเกลส

3.1.2.2 วิธีของเบส

3.2 ตัวแปรตาม (Dependent Variable) คือ ความเชื่อมั่นของแบบทดสอบ
อิงเกณฑ์ในการวิจัยครั้งนี้ ใช้สูตรหาความเชื่อมั่น 2 วิธี คือ

3.2.1 วิธีของโลเวท (Lovett)

3.2.2 วิธีของลิวิงตัน (Livingston)

4. เนื้อหาที่ใช้ศึกษาค้นคว้าครั้งนี้ คือเนื้อหาวิชาคณิตศาสตร์ ค 102 บทที่ 1 เรื่อง
สมการ ตามหลักสูตรมัธยมศึกษาตอนต้น พุทธศักราช 2521 (ฉบับปรับปรุง พ.ศ. 2533)

นิยามศัพท์เฉพาะ

1. แบบทดสอบอิงเกณฑ์ หมายถึง แบบทดสอบที่สร้างขึ้นตามจุดประสงค์เชิงพฤติกรรม
ซึ่งสอดคล้องกับเนื้อหาย่อยวิชาคณิตศาสตร์ ค102 ชั้นมัธยมศึกษาปีที่ 1 เรื่องสมการ และมี
คะแนนจุดตัดหรือคะแนนเกณฑ์สำหรับใช้เป็นเครื่องตัดสินว่า ผู้สอบมีความรอบรู้หรือไม่

2. คะแนนจุดตัดหรือเกณฑ์ หมายถึง คะแนนขั้นต่ำที่จะใช้ตัดสินว่าผู้สอบเป็นผู้รอบรู้
ซึ่งในการศึกษาครั้งนี้ใช้วิธีการกำหนดคะแนนจุดตัดตามวิธีของเกลส และวิธีของเบส

3. วิธีการกำหนดคะแนนจุดตัด หมายถึง ขบวนการที่คำนวณหาจุดคะแนนที่แบ่ง
นักเรียนออกเป็นกลุ่มผู้รอบรู้และไม่รอบรู้ ในการวิจัยครั้งนี้ใช้วิธีการหาคะแนนจุดตัดด้วยวิธีต่อไปนี้

3.1 วิธีของเกลส หมายถึง วิธีการหาคะแนนจุดตัดโดยใช้ทฤษฎีการตัดสินใจ
(Decision Theoretic Approaches) ของเกลส ซึ่งกำหนดเกณฑ์ภายนอกจากระดับผล
การเรียนเฉลี่ยสะสม (G.P.A) ให้นักเรียนที่มีระดับผลการเรียนเฉลี่ยสะสม ตั้งแต่ 2.00 ขึ้นไป
เป็นกลุ่มที่ผ่านเกณฑ์ และกลุ่มที่มีระดับผลการเรียนเฉลี่ยสะสม น้อยกว่า 2.00 เป็นกลุ่มที่ไม่ผ่าน
เกณฑ์ แล้วให้นักเรียนทั้งสองกลุ่มทำแบบทดสอบอิงเกณฑ์ และกำหนดคะแนนจุดตัดต่าง ๆ

คะแนนจุดตัดที่ทำให้ผลบวกของความคลาดเคลื่อนในการจำแนกผิด ประเภทที่ 1 และ 2 มีค่าน้อยที่สุด จะใช้เป็นคะแนนจุดตัด

3.2 วิธีของเบส์ หมายถึง การตัดสินใจประเภทผู้สอบโดยอาศัยคะแนนมวลความรู้ ในรูปคะแนนมาตรฐานซึ่งแปลงมาจากคะแนนสอบแล้วพิจารณาคะแนนจุดตัด โดยอาศัยฟังก์ชันการสูญเสียที่คาดหวังของการตัดสินใจ ให้ผ่านและไม่ผ่านที่น้อยที่สุด เป็นตัวกำหนดคะแนนจุดตัด

4. ผู้รอบรู้ หมายถึง ผู้สอบที่สอบได้คะแนนเท่ากับหรือมากกว่าคะแนนจุดตัดที่กำหนดจากแบบทดสอบอิงเกณฑ์วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1

5. ผู้ไม่รอบรู้ หมายถึง ผู้ที่ได้คะแนนต่ำกว่าคะแนนจุดตัด

6. ความเชื่อมั่นของแบบทดสอบ หมายถึง ความคงเส้นคงวาของคะแนนจากการสอบวัด ไม่ว่าจะทำแบบทดสอบไปสอบกี่ครั้ง ๆ ก็ยังคงได้กลุ่มผู้รอบรู้และกลุ่มไม่รอบรู้เหมือนเดิม ซึ่งคำนวณค่าความเชื่อมั่นโดยใช้สูตรของโลเวท (Lovett) และวิธีของลิวิงตัน (Livingston. 1972)

7. การจัดเรียงข้อสอบ หมายถึง การเรียงข้อคำถามของแบบทดสอบที่มีรูปแบบการเรียงต่างกัน ซึ่งในการศึกษาครั้งนี้มีรูปแบบการจัดเรียงดังนี้

7.1 เรียงตามลำดับค่าความยาก (p) จากค่าความยากสูงไปยังค่าความยากต่ำ

7.2 เรียงตามเนื้อหาย่อย หมายถึง การเรียงข้อสอบตามลำดับเนื้อหาขั้นพื้นฐานไปยังเนื้อหาที่มีความซับซ้อนมากขึ้น ซึ่งในการศึกษาครั้งนี้ผู้วิจัยจัดเรียงจากเนื้อหาย่อย นิยาม สมการ คำตอบของสมการ สมบัติของความเท่ากัน การแก้สมการ และโจทย์ปัญหาสมการตามลำดับ

7.3 เรียงโดยวิธีสุ่ม หมายถึง การเรียงข้อสอบตามลำดับที่หยิบสุ่มมาจากประชากรข้อสอบ

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

เอกสารและงานวิจัยที่เกี่ยวข้องกับการศึกษาค้นคว้าครั้งนี้ ผู้วิจัยได้ศึกษาตามลำดับหัวข้อ
ดังนี้

1. ความหมายและลักษณะของแบบทดสอบอิงเกณฑ์
2. การสร้างแบบทดสอบอิงเกณฑ์
3. การเขียนข้อสอบวัดพฤติกรรมการเรียนรู้
4. ความหมายและการกำหนดคะแนนจุดตัด
5. ความเชื่อมั่นของแบบทดสอบอิงเกณฑ์
6. การเรียงข้อคำถามในแบบทดสอบ
7. งานวิจัยที่เกี่ยวข้องกับการศึกษาค้นคว้า

1. ความหมายและลักษณะของแบบทดสอบอิงเกณฑ์

แบบทดสอบอิงเกณฑ์หรือแบบทดสอบอิงโดเมน (Domain) มีผู้ให้ความหมายและกล่าวถึงลักษณะของแบบทดสอบไว้ดังนี้

เกลเซอร์และนิตโก (Alkin. 1974 : 2 ; Citing Glaser and Nitko. 1971 : 653) กล่าวว่า แบบทดสอบอิงเกณฑ์คือ แบบทดสอบที่สร้างขึ้น เพื่อที่จะให้ผลจากการวัดสามารถแปลความหมายได้ว่า ผู้สอบมีความสามารถอยู่ในระดับใด เมื่อเทียบกับเกณฑ์ที่กำหนดไว้ล่วงหน้า

มิลแมน (Millman. 1974 : 327) กล่าวว่า การทดสอบอิงโดเมนเป็นการวัดสภาพปัจจุบันของนักเรียน โดยอ้างอิงชุดของภารกิจที่กระทำแล้ว ซึ่งเรียกว่าโดเมน (Domain) แบบทดสอบอิงโดเมน คือกลุ่มของข้อสอบจากโดเมนที่ได้มาจากการสุ่มนั่นเอง

แฮริส และสจวต (Alkin. 1974 : 2 ; Citing Harris and Stewart. 197 : 3) กล่าวว่า แบบทดสอบอิงเกณฑ์คือ แบบทดสอบที่ประกอบด้วยตัวแทนของโดเมน หรือขอบเขตเนื้อหาทั้งหมดที่กำหนดพฤติกรรม หรือการกระทำไว้อย่างชัดเจน ในการทดสอบอาจใช้การประมาณค่าอัตราส่วนของพฤติกรรมที่ผู้สอบทำได้กับพฤติกรรมทั้งหมด

โอเวนส์ (Mehrens. 1975 : 49 ; Citing Ivens. 1971 : 2) กล่าวว่า แบบทดสอบอิงเกณฑ์ หมายถึง แบบทดสอบที่ประกอบด้วยข้อสอบต่าง ๆ ที่ใช้วัดความรู้ตามจุดประสงค์เชิงพฤติกรรมต่าง ๆ

อนันต์ ศรีโสภา (2525 : 190) กล่าวว่า แบบทดสอบอิงเกณฑ์ใช้สำหรับวัดความรู้ความสามารถของนักเรียนแต่ละคนว่า ถึงเกณฑ์ขั้นต่ำที่คาดหวังไว้หรือไม่ บางทีเรียกแบบทดสอบนี้ว่า แบบทดสอบวัดความรอบรู้ (Mastery Test) แบบทดสอบชนิดนี้เน้นการวัดความรู้ และทักษะต่าง ๆ ในตัวนักเรียนว่ามีถึงเกณฑ์ที่กำหนดไว้หรือไม่ โดยไม่คำนึงถึงความแตกต่างของนักเรียน

สงบ ลักษณะ (2523 : 1 - 2) ได้กล่าวถึงลักษณะของแบบทดสอบอิงเกณฑ์ว่า คือ การทดสอบที่สามารถนำคะแนนมาแปลความหมายได้ว่า ผู้สอบมีผลสัมฤทธิ์ในการกระทำพฤติกรรมอะไรได้บ้าง ลักษณะของแบบทดสอบอิงเกณฑ์คือ

1. เป็นแบบทดสอบที่สร้างบนพื้นฐานของจุดประสงค์ หรือสิ่งที่ต้องการให้สอบปฏิบัติที่ได้รับการนิยามไว้อย่างชัดเจน
2. รายชื่อของแบบทดสอบจะต้องวัดจุดประสงค์
3. คะแนนที่ได้แปลความหมายว่า ผู้สอบสามารถบรรลุตามเกณฑ์ที่กำหนดไว้หรือไม่

อังคณา สายยศ (2533 : 1) กล่าวว่า การวัดผลแบบอิงเกณฑ์เป็นการวัดผลสัมฤทธิ์ทางการเรียน โดยการเอาผลการสอบของนักเรียนแต่ละคนไปเทียบกับเกณฑ์ว่าบรรลุตามจุดมุ่งหมายของการสอนแล้วหรือยัง

บุญเชิด ภิญญอนันตพงษ์ (2526 : 5) กล่าวว่า แบบทดสอบอิงเกณฑ์เป็นแบบทดสอบที่บรรจุเนื้อหาสาระของข้อสอบที่เฉพาะเจาะจงสอดคล้องกับจุดประสงค์การสอน มีคะแนนจุดตัดหรือคะแนนเกณฑ์สำหรับใช้เป็นเครื่องตัดสินว่าผู้สอบมีความรอบรู้ตามเกณฑ์ที่กำหนดหรือไม่

จากความหมายและลักษณะของแบบทดสอบอิงเกณฑ์ สรุปได้ว่า แบบทดสอบอิงเกณฑ์เป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนที่สร้างขึ้นตามจุดประสงค์การสอน และมีการแปลความหมายคะแนน โดยนำคะแนนของผู้สอบแต่ละคนไปเทียบกับคะแนนจุดตัด ซึ่งเป็นเกณฑ์ตัดสินว่าผู้สอบมีความรอบรู้หรือไม่รอบรู้

2. การสร้างแบบทดสอบอิงเกณฑ์

นักวิชาการและนักวัดผลการศึกษา ได้เสนอแนะแนวทางในการสร้างแบบทดสอบอิงเกณฑ์ไว้ดังนี้

แอราเซียน และเมดัส (Aivasian and Meduas. 1972 : 8) แบ่งขั้นตอนของการสร้างแบบทดสอบอิงเกณฑ์เป็น 4 ขั้นตอน ดังนี้

1. เขียนสมรรถภาพที่จะวัดให้ชัดเจน โดยเขียนในรูปของจุดประสงค์เชิงพฤติกรรมที่สามารถสังเกตได้ และสอบวัดไว้
2. กำหนดการรอบรู้ในสมรรถภาพนั้น
3. คิดสถานการณ์ เพื่อให้ให้นักเรียนได้แสดงถึงความสามารถที่ต้องการวัดแล้วเขียนข้อสอบตามสถานการณ์นั้น
4. ตัดสินความรอบรู้ของนักเรียน โดยเทียบกับมาตรฐานการปฏิบัติขั้นต่ำสุดตามจุดประสงค์

ลินด์วอลล์ และนิตโก (สงบ ลักษณะ. 2522 : 49; อ้างอิงมาจาก Lindvall and Nitko. 1975) เสนอแนวคิดในการสร้างแบบทดสอบอิงเกณฑ์ว่าจะต้องสร้างขึ้นโดยอาศัยจุดประสงค์เชิงพฤติกรรมที่สร้างมาจากหลักสูตร และพฤติกรรมที่มุ่งหวังในแต่ละเนื้อหา

โกวิท ประวาลพุกษ์ (2522 : 124 - 149) ได้เสนอแนะวิธีการสร้างแบบทดสอบอิงเกณฑ์ดังนี้

1. กำหนดจุดมุ่งหมายทั่วไปของการเรียนรู้ ในเนื้อหาตอนหนึ่ง ๆ ถ้ามีตารางวิเคราะห์หลักสูตรอยู่แล้ว ก็นำมาใช้ได้เลย พฤติกรรมที่พึงประสงค์จะครอบคลุมไปถึงพฤติกรรมขั้นสูง ๆ ด้วย
2. นำจุดมุ่งหมายทั่วไปมาเขียนเป็นจุดมุ่งหมายเชิงพฤติกรรม โดยกำหนดและจำแนกเป็นพฤติกรรมปลายทาง พฤติกรรมเบื้องต้น และพฤติกรรมดำเนินการ
3. จัดสร้างข้อสอบตามจุดมุ่งหมายเชิงพฤติกรรมไว้มาก ๆ เพื่อจะได้นำมาทดสอบได้ตลอดเวลา
4. จัดเนื้อหา เอกสารและวัสดุประกอบการสอนที่เหมาะสมที่จะทำให้เกิดพฤติกรรม
5. ทดสอบผู้เรียนเพื่อตรวจสอบดูว่า นักเรียนยังต้องเรียนเสริมก่อนเริ่มสอนบทเรียนที่กำหนดไว้ ถ้าผู้เรียนยังไม่พร้อมก็ต้องมีการฝึกฝนเพิ่มเติมพิเศษ
6. ทำการสอนโดยใช้แบบทดสอบระหว่างสอนเข้ามาช่วยในการตัดสินใจดำเนินการให้มากที่สุด เช่น จะสอบซ้ำหรือไม่
7. ทดสอบเมื่อสิ้นสุดการสอน
8. เมื่อผู้เรียนไม่สามารถผ่านแบบทดสอบสุดท้ายนี้ได้ จำเป็นจะต้องมีการซ่อมเสริมทบทวน กระบวนการเรียนการสอนจะสิ้นสุด เมื่อผู้เรียนทุก ๆ คน ได้แสดงความสามารถตามจุดมุ่งหมายได้ตามเกณฑ์ที่กำหนดไว้

อังคณา สายยศ (2526 : 25 - 29) ได้กล่าวถึงขั้นตอนในการสร้างแบบทดสอบอิงเกณฑ์ไว้ดังนี้

1. วิเคราะห์หลักสูตร หรือวิเคราะห์งานนำหลักสูตรในวิชาที่ต้องการจะเขียนข้อสอบมาทำการวิเคราะห์ โดยนำเนื้อหาย่อย ๆ กับพฤติกรรมมาหาความสัมพันธ์กัน

2. เขียนจุดมุ่งหมายเชิงพฤติกรรม จากตารางวิเคราะห์หลักสูตรแล้วนำพฤติกรรมหลัก และเนื้อหา มาเขียนเป็นจุดมุ่งหมายเชิงพฤติกรรม

3. เขียนคำถาม หรือข้อสอบให้สอดคล้องกับจุดมุ่งหมายเชิงพฤติกรรมจำนวนข้อสอบ ในแต่ละพฤติกรรม ควรออกประมาณ 5-10 ข้อ ซึ่งขึ้นอยู่กับผู้สอนหรือผู้ออกข้อสอบเป็นสำคัญ โดยยึดว่าจุดมุ่งหมายในเนื้อหานั้น ควรจะออกกี่ข้อ

บุญเชิด ภิญโญนนตพงศ์ (2527 : 49 - 50) ได้สรุปขั้นตอนของการสร้าง แบบทดสอบอิงเกณฑ์ไว้ดังนี้

- ขั้นที่ 1 วิเคราะห์เนื้อหาและจุดประสงค์
- ขั้นที่ 2 แปลงจุดประสงค์หลักให้เป็นจุดประสงค์เฉพาะเจาะจง
- ขั้นที่ 3 เขียนข้อสอบหรือผลิตข้อสอบ
- ขั้นที่ 4 ตรวจทานข้อสอบ
- ขั้นที่ 5 วิเคราะห์ข้อสอบ
- ขั้นที่ 6 คัดเลือกข้อสอบ กำหนดความยาวและคะแนนจุดตัดของแบบทดสอบ
- ขั้นที่ 7 วิเคราะห์แบบทดสอบ

จากที่กล่าวมาทั้งหมด สามารถสรุปขั้นตอนในการสร้างแบบทดสอบอิงเกณฑ์ ได้ดังนี้

1. กำหนดจุดมุ่งหมายในการสอนว่า จะให้ผู้เรียนเกิดพฤติกรรมอะไรบ้างหลังจากเรียน หน่วยนั้นแล้ว

2. สร้างตารางกำหนดรายละเอียด (Table of Specification) ซึ่งเป็นการหา ความสัมพันธ์ระหว่างจุดประสงค์เชิงพฤติกรรม กับเนื้อหาที่จะสอนเพื่อที่จะดูรายละเอียดของ จุดมุ่งหมายของการสอน

3. เขียนจุดมุ่งหมายเชิงพฤติกรรม จากตารางวิเคราะห์หลักสูตรโดยนำเอาเนื้อหา และ พฤติกรรมหลักมาเขียนเป็นจุดมุ่งหมายเชิงพฤติกรรม

4. เขียนข้อคำถามจากจุดมุ่งหมายเชิงพฤติกรรมในขั้นที่ 3

5. พิจารณาความสอดคล้องของข้อคำถามกับจุดประสงค์เชิงพฤติกรรมโดยให้ผู้เชี่ยวชาญ ด้านเนื้อหาพิจารณา

6. การคัดเลือกข้อคำถามที่สอดคล้อง และครอบคลุมจุดประสงค์เชิงพฤติกรรม

7. ทาคุณภาพของแบบทดสอบ ด้านความยากง่ายอำนาจจำแนก ความเชื่อมั่น และ ความเที่ยงตรงของแบบทดสอบ

3. การเขียนข้อสอบวัดพฤติกรรมการเรียนรู้

การเขียนข้อสอบตามจุดประสงค์เชิงพฤติกรรม มีความเกี่ยวข้องกับพฤติกรรมการเรียนรู้ ซึ่งจะทำให้การวัดมีความเที่ยงตรง และวัดผลการเรียนรู้ได้ค่อนข้างชัดเจน ซึ่งได้มีนักการศึกษา และนักวัดผลสนใจศึกษากันอย่างกว้างขวาง ในที่นี้จะนำเสนอการเขียนข้อสอบวัดพฤติกรรม การเรียนรู้ที่สำคัญคือพฤติกรรมการเรียนรู้ตามแนวคิดของ บลูม (Benjamin S. Bloom)

บลูมและคณะ (ส่วน สายยศ และอังคณา สายยศ. มปป.) ได้เสนอแนวคิดในการวัด พฤติกรรมการเรียนรู้ด้านพุทธิพิสัย (Cognitive Domain) ไว้ 6 ด้าน ประกอบด้วย

1.0 วัดความจำ (Knowledge) หมายถึงความสามารถในการระลึกเรื่องราว ประสบการณ์ หรือสิ่งที่เคยเรียนมาแล้วได้ โดยแบ่งย่อยออกไปคือ

1.10 วัดความจำในเนื้อเรื่อง (Knowledge of Specific) แบ่งเป็น 2 สมรรถภาพ

1.11 ความรู้ความจำเกี่ยวกับศัพท์ นิยาม (Knowledge of Terminology) เป็นความสามารถในการระลึกศัพท์ หรือนิยามต่าง ๆ

1.12 ความรู้ความจำเกี่ยวกับกฎและความจริง (Knowledge of Specific Facts) เป็นความสามารถในการระลึกความจริงเกี่ยวกับ บุคคล เวลา เหตุการณ์ใน ประวัติศาสตร์ การค้นคว้ากฎและสูตร หรือสิ่งที่ใช้เป็นกติกาดกกลงกันว่าเป็นจริงแล้ว

1.20 วัดความรู้ในวิธีดำเนินการ (Knowledge of ways and Means of Dealing with Specifics) จำแนกเป็น 5 สมรรถภาพย่อย ๆ ได้แก่

1.21 ความรู้เกี่ยวกับระเบียบแบบแผน (Knowledge of Conventions) เป็นความสามารถที่จะบอกถึงสิ่งที่นิยมทำกันในทางวิชาการจนเป็นนิสัย หรือ ประเพณี หรือเป็นสิ่งที่ได้ตกลงกันวางแผนเอาไว้ว่า ควรทำกันอย่างไรได้

1.22 ความรู้เกี่ยวกับแนวโน้มและลำดับขั้น (Knowledge of Trends and Sequences) เป็นความรู้ในลำดับขั้นของเหตุการณ์ เช่น รู้ว่า เหตุการณ์นั้น ๆ เกิดขึ้นได้ อย่างไรในชั่วระยะเวลาหนึ่ง ๆ อะไรเกิดก่อนอะไรเกิดหลัง อีกทั้งหมายความว่าความรวมถึงความสัมพันธ์ ของถึงต่าง ๆ ซึ่งสามารถแยกออกจากกันได้โดยเวลา

1.23 ความรู้เกี่ยวกับการจัดประเภท (Knowledge of Classifications and Categories) เป็นความสามารถจัดประเภทสิ่งของเรื่องราว หรือ เหตุการณ์ต่าง ๆ ออกเป็นหมวดหมู่

1.24 ความรู้เกี่ยวกับเกณฑ์ (Knowledge of Criteria) เป็นความสามารถบอกถึงตัวเกณฑ์ที่จะนำไปใช้ในการตัดสินพิจารณา หรือตรวจสอบสิ่งใด ๆ ก็ได้

1.30 วัตความรู้รวบยอดในเนื้อเรื่อง (Knowledge of Universals and Abstractic in the field) เป็นการวัดความสามารถที่จะบอกความคิดรวบยอดของเนื้อหาวิชาต่าง ๆ คือ สรุปเป็นหลักวิชาการ ความสามารถด้านนี้ แบ่งออกเป็นสมรรถภาพย่อยได้ 2 สมรรถภาพ ได้แก่

1.31 ความรู้เกี่ยวกับหลักวิชา และขยายความ (Knowledge of Principle and Generalizations) ซึ่งความรู้เกี่ยวกับหลักวิชา หมายถึง ความสัมพันธ์ของมโนภาพตั้งแต่ 2 อันขึ้นไป เช่น ลิบ บวกกับ ห้าเป็นสิบห้า, รูปสามมิติมีหกด้าน, ความดันของก๊าซจำนวนหนึ่ง เพิ่มขึ้นขึ้นอยู่กับอุณหภูมิเพิ่มขึ้น การขยายความ ก็เป็นการนำหลักวิชาไปอ้างอิงอีกรูปหนึ่ง โดยอาศัยความสัมพันธ์ของมโนภาพหลาย ๆ อัน

1.32 ความรู้เกี่ยวกับทฤษฎีและโครงสร้าง (Knowledge of Theories and Structures) เป็นความสามารถบอกถึงทฤษฎีโครงสร้างของเรื่องที่เรียนมาแล้ว โดยใช้ คำว่า "ทฤษฎี" หมายถึงการนำหลักวิชาหลายอันมาเกี่ยวพันกันและทำการพิสูจน์จนเป็นที่แน่ใจ ส่วนคำว่า "โครงสร้าง" หมายถึง สิ่งที่มาประกอบให้บังเกิดสิ่งนั้นขึ้น

2.0 วัดความเข้าใจ (Comprehension) เป็นความสามารถในการแปลความ และขยายความ จากสื่อความหมายต่าง ๆ ที่ได้พบเห็นแบ่งออกเป็น 3 สมรรถภาพ คือ

2.10 การแปลความ (Translation) เป็นความสามารถในการถ่ายทอดความหมายจากภาษาหนึ่งมาเป็นอีกภาษาหนึ่งอย่างได้ความหมาย อาจแปลจากความยากง่ายระดับหนึ่งไปสู่อีกระดับหนึ่ง อาจแปลจากสัญลักษณ์ชนิดหนึ่งไปสู่อีกชนิดหนึ่ง หรืออาจแปลจาก คำ ๆ หนึ่งไปสู่อีกคำหนึ่งก็ได้

2.20 การตีความ (Interpretation) มุ่งให้ผู้เรียนค้นหาความสำคัญ และความสัมพันธ์ของส่วนย่อย ๆ ของเนื้อหานั้น แล้วตีความจากเรื่องราวที่กำหนดให้ทั้งหมดนั้น การตีความมองเห็น ๆ คล้ายกับการวิเคราะห์และการประเมินค่า แต่ความจริงไม่เป็นเพราะการวิเคราะห์และการประเมินนั้น ใช้ความรู้จากแหล่งอื่นมาเกี่ยวข้องมากมาย แต่การตีความยึดเนื้อความที่ให้ไว้เป็นหลักเท่านั้น ไม่ต้องอาศัยเกณฑ์ใดไปใช้ นั่นคือเด็กจะต้องเก็บความเดิมมาเรียบเรียง และร้อยกรองใหม่ เพื่อจะได้มองเห็นเรื่องราวเดิมในแง่ใหม่ การตีความจึงต้องตีเฉพาะในขอบเขตของเนื้อหาที่จะใช้ตีความเท่านั้น

2.30 การขยายความ (Extrapolation) เป็นสมรรถภาพที่สามารถทำนายหรือคาดคะเนเหตุการณ์ล่วงหน้าได้อย่างดี โดยอาศัยข้อมูล และแนวโน้มต่าง ๆ เป็น เครื่องช่วย

3.0 วัดการนำไปใช้ (Application) เป็นความสามารถในการแก้ปัญหาโดยตรง มิได้หมายถึงความสามารถที่จะเอาสิ่งหนึ่งสิ่งใด ไปใช้ได้อย่างรวดเร็ว แต่เป็นความสามารถในการนำความรู้และความเข้าใจเดิมมาช่วยแก้ปัญหาในสถานการณ์ใหม่ วิธีการเขียนข้อสอบจึงพยายามสร้างสถานการณ์ใหม่ให้แปลกแตกต่างไปจากบทเรียน

4.0 วัดการวิเคราะห์ (Analysis) เป็นสมรรถภาพที่สามารถแยกแยะส่วนย่อย ๆ ของเหตุการณ์ เรื่องราว เนื้อหาต่าง ๆ ว่า ประกอบด้วยอะไร มีจุดมุ่งหมาย หรือความประสงค์สิ่งใด นอกจากนั้นยังมองถึงว่าส่วนย่อย ๆ ที่สำคัญนั้น แต่ละเหตุการณ์เกี่ยวข้องกับอย่างไร และเกี่ยวพันกัน โดยอาศัยหลักการใด สมรรถภาพด้านการวิเคราะห์จะเต็มไปด้วยการหาเหตุและผลมาเกี่ยวข้องกัน อยู่เสมอ

4.10 การวิเคราะห์ความสำคัญ (Analysis of Elements) เป็นสมรรถภาพ ในการหามูลเหตุ ต้นกำเนิด ผลลัพธ์ และความสำคัญของเรื่องราวทั้งปวง การสร้างคำถามจะต้อง ให้เด็กมองหาสิ่งที่มีให้เลือก ดังนั้น การสร้างคำถามจึงมักจะมีคำว่า "ที่สุด" อยู่ด้วยเสมอ ส่วนการ เขียนตัวเลือกจะต้องให้เป็นตัวถูกทุกตัว แต่พยายามให้มีตัวหนึ่งที่ถูกต้องมากที่สุด

4.20 การวิเคราะห์ความสัมพันธ์ (Analysis of Relationships) เป็นความสามารถในการค้นหาความสำคัญย่อย ๆ ทางเรื่องราว หรือเหตุการณ์นั้น ซึ่งอาจถามความสัมพันธ์ ของเนื้อเรื่องกับเหตุ เนื้อเรื่องกับผล โดยต้องการให้พยายามค้นหาว่า แต่ละเหตุการณ์นั้นมีความ สำคัญอะไรที่ไปเกี่ยวพันกันเป็นตัวร่วม

4.30 การวิเคราะห์หลักการ (Analysis of Organizational principles) เป็นความสามารถที่จะจับเค้าเงื่อนของเรื่องราวนั้นว่า ยึดถือหลักการใดมีเทคนิคการเขียนอย่างไร จึงชวนให้อ่าน มีมโนภาพหรือยึดหลักปรัชญาใด อาศัยหลักการใดเป็นสื่อสารสัมพันธ์ เพื่อให้เกิด ความเข้าใจ คำถามวิเคราะห์หลักการมักจะมีคำลงท้ายว่า...ยึดหลักการใด? หรือ...มีหลักการใด? อยู่เสมอ

5.0 วัดการสังเคราะห์ (Synthesis) เป็นการวัดความสามารถในการรวมสิ่งต่าง ๆ ตั้งแต่สองสิ่งขึ้นไปเข้าด้วยกัน เพื่อให้เป็นสิ่งใหม่อีกกลุ่มหนึ่งมีคุณลักษณะ โครงสร้าง หรือหน้าที่ใหม่ แปลกแตกต่างไปจากของเดิมก่อนนำมารวมกัน การสังเคราะห์มีทั้งระดับง่ายและระดับยาก แบ่งเป็น 3 สมรรถภาพย่อย

5.10 การสังเคราะห์ข้อความ (Production of Unique Communication) เป็นความสามารถในการสังเคราะห์ข้อความ โดยสื่อหรือโดยการพูด การเขียน ซึ่งเป็นการแสดงความคิด เห็นอย่างอิสระต่อเรื่องราวที่กำหนด การวิจารณ์และการหาข้อยุติบางประการ อาจเป็นข้อความ ที่มีจุดมุ่งหมายเพื่อบอกเล่า บรรยาย ชักชวน ให้ความประทับใจ หรือ เพื่อให้เกิดความสนุกสนาน

การสังเคราะห์ข้อความเป็นการนำความรู้และประสบการณ์ต่าง ๆ มา ผสมผสานกัน เพื่อให้เกิดเป็นข้อความ หรือผลิตผล หรือการกระทำใหม่ใด ๆ ที่สามารถให้สื่อความคิด และอารมณ์ระหว่างบุคคล

5.20 การสังเคราะห์แผนงาน (Production of Plans and processed of Operations) เป็นความสามารถด้านการวางแผนล่วงหน้า เพื่อให้งานดำเนินไปด้วยดี และ

ยังเป็นความสามารถด้านการออกแบบลักษณะต่าง ๆ เช่น ออกแบบบ้านเรือน ออกแบบเรือ ออกแบบรถ ออกแบบเสื้อผ้า ฯลฯ

การวัดด้านนี้ จะมองผลงานด้านการวางแผน หรือชุดของการจัดกระทำที่เสนอเพื่อปฏิบัติ เช่น วางแผนเพื่อทดลอง

5.30 การสังเคราะห์ความสัมพันธ์ (Production of Relationships) เป็นความสามารถในการนำเอานามธรรมย่อย ๆ ทั้งหลายมาผสมกลมกลืนกัน เพื่อให้เป็นรูปใหม่นามธรรมในที่นี้อาจเป็นปรากฏการณ์ ข้อความสัญลักษณ์ คุณลักษณะใด ๆ ก็ได้ เมื่อนำมาเกี่ยวพันกันแล้วจะทำให้เปลี่ยนลักษณะออกไปเป็นลักษณะของการสร้างสรรค์สิ่งใหม่ขึ้น ทั้งนี้เพราะเป็นการนำเอาหลักวิชา มโนภาพ ทฤษฎี หรือโครงสร้างมาผสมกลมกลืนกันนั่นเอง

6.0 วัดการประเมินค่า (Evaluation) เป็นการวัดความสามารถในการพิจารณาตัดสินเกี่ยวกับคุณค่าของความคิดทุกชนิด เพื่อเปรียบเทียบกับเกณฑ์หรือมาตรฐานที่กำหนดให้การประเมินเป็นการแสดงความคิดเห็นที่น่าเกณฑ์มาใช้ ไม่ใช่คิดเห็นเฉย ๆ การพิจารณาตัดสินอาจจะเป็นในรูปของปริมาณหรือคุณภาพก็ได้ การประเมินแบ่งออกเป็น 2 อย่าง คือ

6.10 การประเมินโดยอาศัยข้อเท็จจริงภายใน (Judgment in term of internal evidence) การประเมินแบบนี้พิจารณาความถูกต้องสมเหตุสมผล ความสอดคล้องและเกณฑ์ภายในอื่น ๆ

6.20 การประเมินโดยเกณฑ์ภายนอก (Judgment in term of external criteria) การประเมินแบบนี้อาศัยเกณฑ์หรือมาตรฐานจากภายนอกไว้เปรียบเทียบ เกณฑ์เหล่านี้ อาจจะเป็นเกณฑ์ที่สังคม หรือระเบียบประเพณีกำหนดก็ได้

จากที่กล่าวมา เป็นแนวคิดเกี่ยวกับการเขียนข้อสอบวัดพฤติกรรมการเรียนรู้ของบลูม (Benjamin S. Bloom) ซึ่งค่อนข้างจะละเอียด และชัดเจนจึงเป็นที่นิยมของนักการศึกษาและนักวัดผล ซึ่งการเขียนข้อสอบส่วนมากจะสร้างตามพฤติกรรมการเรียนรู้ของบลูมเป็นหลัก

พฤติกรรมกรเรียนรู้ในวิชาคณิตศาสตร์

เจมส์ ดับบิว วิลสัน (Willon. 1971 : 645 - 696) ได้จำแนกพฤติกรรมที่พึงประสงค์ในวิชาคณิตศาสตร์มัธยมศึกษา ออกเป็น 4 ขั้นตอนดังนี้

1. ความรู้ความจำเกี่ยวกับการคิดคำนวณ (Computation) เป็นความสามารถในการระลึกถึงสิ่งที่ได้เรียนมาแล้วทั้งในด้านข้อเท็จจริงศัพท์ นิยามตลอดจนความสามารถในการดำเนินการคิดโจทย์ปัญหาอย่างง่าย ๆ ไม่ยุ่งยากซับซ้อน ไม่ต้องการอาศัยการตัดสินใจ ทั้งนี้รวมถึงโจทย์ปัญหาที่เหมือนกับตัวอย่างหรือ แบบฝึกหัดที่เคยทำมาแล้ว พฤติกรรมขั้นนี้ยังแบ่งออกเป็น 3 ขั้นย่อย ๆ คือ

1.1 ความรู้ความจำเกี่ยวกับข้อเท็จจริง (Specific facts) เป็นความสามารถที่จะระลึกถึงข้อเท็จจริงต่าง ๆ ที่ได้เรียนมาแล้วตลอดจนพื้นฐานต่าง ๆ ทางคณิตศาสตร์ที่นักเรียนสร้างสมมาเป็นเวลานานอีกด้วย

1.2 ความรู้ความจำเกี่ยวกับศัพท์และนิยาม (Terminology) เป็นความสามารถในการระลึกถึงศัพท์นิยามต่าง ๆ ที่ได้เรียนมาแล้วได้โดยไม่ต้องอาศัยการคิดคำนวณแต่อย่างใดและไม่ต้องการใช้ความรู้อื่นมาช่วย

1.3 ความรู้ความจำเกี่ยวกับการใช้กระบวนการคิดคำนวณ (Algorithms) เป็นความสามารถในการใช้ข้อเท็จจริง ศัพท์หรือนิยาม และกระบวนการที่ได้เรียนมาแล้ว เช่น ลำดับขั้นตอนการหา ค.ร.น. หรือ ห.ร.ม.

2. ความเข้าใจ (Comprehention) เป็นความสามารถในการนำความรู้ที่ได้เรียนมาแล้วมาสัมพันธ์กับโจทย์ หรือปัญหาใหม่ ตลอดจนความสามารถในการตีความ แปลความ และขยายความได้ พฤติกรรมขั้นนี้แบ่งออกเป็น 6 ขั้นย่อย ๆ คือ

2.1 ความเข้าใจเกี่ยวกับความคิดรวบยอด (Concept) เป็นความสามารถในการนำข้อเท็จจริง ซึ่งต้องอาศัยความรู้ต่าง ๆ มาผสมผสานกัน คำถามเกี่ยวกับความคิดรวบยอดนี้ครูจะต้องไม่เคยบอก หรือสอนมาก่อนเพราะถ้าเคยบอกมาก่อนแล้วจะกลายเป็นความรู้ความจำเกี่ยวกับข้อเท็จจริง

2.2 ความเข้าใจเกี่ยวกับหลักการ กฎ และการสรุปอ้างอิงเป็นกรณีทั่วไปทางคณิตศาสตร์ (Principle, rules and generalization) เป็นความสามารถในการนำเอาหลักการ กฎ และความรู้เกี่ยวกับความคิดรวบยอด ไปสัมพันธ์กับโจทย์ปัญหาจนได้ แนวทางการแก้โจทย์ปัญหาได้

2.3 ความเข้าใจในโครงสร้างทางคณิตศาสตร์ (Mathematical Structure) คำถามที่วัดพฤติกรรม ในขั้นนี้ต่างจากพฤติกรรม ความรู้ความจำเกี่ยวกับศัพท์และนิยาม

2.4 ความสามารถในการแปลงโจทย์ปัญหา จากรูปแบบหนึ่งไปอีกรูปแบบหนึ่ง (Transform problem elements from one mode to another) เป็นความสามารถในการแปลข้อความ ที่กำหนดให้ออกเป็นข้อความใหม่ หรือภาษาใหม่อีกรูปหนึ่งซึ่งมีความหมายคงเดิม เช่น เปลี่ยนโจทย์ปัญหาให้อยู่ในรูปสมการ

2.5 ความสามารถในการดำเนินความคิดตามแนวของเหตุผลที่วางไว้ (Follow a line of reasoning) ความสามารถในการดำเนินความคิด ตามแนวของเหตุผลเป็นความสามารถในการอ่านและเข้าใจกับข้อความทางคณิตศาสตร์ และสามารถบอกได้ว่าผลสรุปในแต่ละขั้นมาจากเหตุผลใด

2.6 ความสามารถในการอ่านและตีความโจทย์ปัญหา (Read and interpret or problem) เป็นความสามารถในการอ่านและตีความจากโจทย์ปัญหาที่กำหนดให้ เพื่อทราบว่

โจทย์ต้องการอะไร โจทย์กำหนดอะไรให้บ้าง สิ่งที่กำหนดให้ยังขาดสิ่งใดบ้าง รวมทั้งการแปลความหมายจากกราฟ หรือข้อมูลทางสถิติ ตลอดจนการแปลสมการ หรือตัวเลขให้เป็นรูปภาพ

3. การนำไปใช้ (Application) เป็นความสามารถในการแก้ปัญหาต่าง ๆ ที่คล้ายคลึงกับที่เรียนมาแล้ว นั่นคือนักเรียนจะต้องสามารถผสมผสานความรู้ความสามารถจากขั้น 1 และ 2 ในการนำมาใช้แก้โจทย์ปัญหา ซึ่งจะมีหลายขั้นตอนในการจัดกระทำเพื่อให้ได้คำตอบออกมา ดังนั้น จึงมีความจำเป็นต้องมีการเลือกการตัดสินใจว่าจะทำขั้นตอนใดก่อนหลัง พฤติกรรมขั้นนี้แบ่งออกเป็น 4 ขั้นย่อย ๆ คือ

3.1 ความสามารถในการแก้โจทย์ปัญหาที่คุ้นเคย (Solve routine problem) หรือปัญหาที่เคยเรียนมาแล้ว พฤติกรรมขั้นนี้เป็นความสามารถในการแก้โจทย์ปัญหาที่คล้ายคลึงแต่ไม่ใช่ข้อเดียวกันกับตัวอย่าง หรือแบบฝึกหัด ที่นักเรียนเคยทำมาแล้ว นักเรียนจะต้องอาศัยความรู้ความจำเกี่ยวกับการคิดคำนวณ และความเข้าใจ มาผสมผสานกับแก้ปัญหาก็เพื่อให้ได้ผลลัพธ์ออกมา

3.2 ความสามารถในการเปรียบเทียบ (Make comparisons) เป็นความสามารถในการหาความสัมพันธ์โดยการเปรียบเทียบข้อมูลที่โจทย์ให้มา 2 ชุด ในการแก้ปัญหาก็จะต้องใช้วิธีการคิดคำนวณ ความเข้าใจ แล้วจึงนำมาเปรียบเทียบเพื่อตัดสินใจ

3.3 ความสามารถในการวิเคราะห์ข้อมูล (Analyze data) เป็นความสามารถในการจำแนกและตัดสินใจว่าข้อมูลส่วนใดจำเป็น หรือไม่จำเป็นต่อการแก้โจทย์ปัญหา

3.4 ความสามารถในการมองเห็นแบบ ลักษณะโครงสร้างที่เหมือนกัน และการสมมาตร (Recognize patterns, isomorphisms, and symmetries) พฤติกรรมในขั้นนี้จะเกี่ยวกับการระลึกถึงข้อมูล แปลงปัญหาการจัดกระทำกับข้อมูลที่กำหนดให้ หรือจากโจทย์ปัญหาที่กำหนดขึ้น

4. การวิเคราะห์ (Analysis) พฤติกรรมขั้นนี้ถือว่าเป็นพฤติกรรมขั้นสูงสุดด้านสติปัญญา นักเรียนจะตอบปัญหาที่วัดพฤติกรรมขั้นนี้ได้ต้องมีความสามารถในระดับสูง โจทย์ปัญหาจะมีลักษณะซับซ้อน พลิกแพลง ซึ่งนักเรียนไม่เคยลองฝึกทำกันมาก่อนแบ่งออกเป็น 4 ขั้นย่อย ๆ

4.1 ความสามารถในการแก้โจทย์ปัญหาที่ไม่คุ้นเคยมาก่อน (Solve nonroutine problems) คำถามในขั้นนี้เป็นคำถามที่ต้องคิดซับซ้อนเป็นโจทย์ที่ไม่ได้อยู่ในรูปแบบฝึกหัด หรือตัวอย่างไม่เคยเห็นมาก่อน นักเรียนจะแก้ปัญหาก็ต้องอาศัยความคิดสร้างสรรค์ของคณิตศาสตร์ ต้องเข้าใจความคิดรวบยอด หรือนิยามตลอดจนทฤษฎีต่าง ๆ ที่เรียนมาแล้วเป็นอย่างดี แล้วใช้ความรู้เหล่านั้นมาผสมผสานกันแก้ปัญหาก็ได้

4.2 ความสามารถในการค้นหาความสัมพันธ์ (Discover relationships) ข้อสอบที่วัดพฤติกรรมในขั้นนี้นักเรียนจะต้องจัดส่วนต่าง ๆ ที่กำหนดให้ใหม่ แล้วสร้างความสัมพันธ์

ขึ้นเพื่อใช้ในการแก้ปัญหา พฤติกรรมในขั้นนี้ต่างจากขั้นการนำไปใช้ตรงที่นักเรียนต้องรวบรวมสิ่งที่โจทย์กำหนดให้มาเป็นความสัมพันธ์ขึ้นใหม่

4.3 ความสามารถในการสร้างข้อพิสูจน์ (Construct- proofs) เป็นความสามารถในการสร้างข้อพิสูจน์โจทย์ปัญหาที่ไม่เคยพบมาก่อน นักเรียนจะต้องอาศัยนิยาม และทฤษฎีความรู้ต่าง ๆ มาช่วยแก้ปัญหา

4.4 ความสามารถในการวิพากษ์วิจารณ์การพิสูจน์ (Criticize proofs) เป็นการใช้เหตุผลที่ควบคู่ไปกับความสามารถในการพิสูจน์ ซึ่งต้องการให้นักเรียนมองเห็นหรือเข้าใจการพิสูจน์นั้นว่าถูกต้องหรือไม่ มีตอนใดผิดพลาดบ้าง

4.5 ความสามารถในการสร้างสูตรและการทดสอบความถูกต้องของสูตรนั้น ซึ่งใช้เป็นกรณีทั่วไปได้ (Formulate and validate generalization) พฤติกรรมขั้นนี้คล้ายกับ 4.3 แต่อาจซับซ้อนมากกว่า นักเรียนต้องสามารถสร้างสูตรขึ้นมาใหม่ โดยให้สัมพันธ์กับเรื่องที่เคยเรียนมาแล้วและต้องสมเหตุสมผล สามารถใช้ได้ทุกกรณี

4. ความหมายและการกำหนดคะแนนจุดตัด

ความหมายของคะแนนจุดตัด

คะแนนจุดตัด (Cut-off Scores) นักการศึกษาและนักวัดผลเรียกต่าง ๆ กัน เช่น เกณฑ์ (Criteria) มาตรฐาน (Standard) คะแนนผ่าน (Passing Score) ระดับความรอบรู้ (Mastery Level) หรือความสามารถต่ำสุด (Minimal Competence) และได้ให้ความหมายที่ต่างกันดังนี้

แฮมเบิลตัน (Hambleton. 1978 : 279) กล่าวว่า มาตรฐาน (Standard) หมายถึงคะแนนที่ได้จากการสอบที่ใช้ในการแยกผู้สอบออกเป็น 2 กลุ่ม ตามระดับความสามารถที่แตกต่างกัน คือ เป็นกลุ่มที่รอบรู้ (Masters) และกลุ่มที่ไม่รอบรู้ (Non-masters)

ปอปแฮม (Halpin, Sigmon and Halpin. 1983 : 185 ; Citing Popham. 1978) กล่าวว่า มาตรฐาน (Standard) เป็นการวัดความสามารถของผู้สอบตามจุดประสงค์ที่กำหนดไว้

เบอร์ก (Berk. 1986 : 138) กล่าวว่า คะแนนจุดตัด (Cut-off-Score) หมายถึง จุดที่ใช้แบ่งคะแนนออกเป็น 2 ส่วน ส่วนหนึ่งเป็นคะแนนของผู้สอบที่จัดว่าเป็นผู้รอบรู้ อีกส่วนหนึ่งเป็นคะแนนของผู้สอบที่จัดว่าเป็นผู้ไม่รอบรู้ เนื่องจากคะแนนจุดตัดเป็นค่าที่ได้จากการสังเกต เพราะฉะนั้นการกำหนดคะแนนจุดตัด จึงย่อมมีความคลาดเคลื่อนมาตรฐาน (Standard Error) สำหรับความคลาดเคลื่อนนี้ เบอร์กแบ่งออกเป็น 2 ประเภท คือ

1. ผู้สอบที่มีความรอบรู้อย่างแท้จริง แต่ถูกจัดประเภทไม่มีความรอบรู้ ความคลาดเคลื่อนนี้เกิดจากการกำหนดคะแนนจุดตัดสูงเกินไป จึงทำให้ผู้สอบที่มีความรอบรู้สอบไม่ผ่านเกณฑ์ที่กำหนด

2. ผู้สอบที่ไม่มีความรอบรู้อย่างแท้จริงถูกจัดเป็นประเภทมีความรอบรู้ ความคลาดเคลื่อนนี้เกิดจากการกำหนดคะแนนจุดตัดต่ำเกินไป จึงทำให้ผู้สอบที่ไม่มีความรอบรู้สอบผ่านเกณฑ์ที่กำหนด

อังคณา สายยศ (2525 : 70) กล่าวว่า คะแนนจุดตัด หมายถึงคะแนนที่น้อยที่สุดที่นักเรียนจะต้องทำได้ในการที่จะได้รับการตัดสินให้เป็นผู้รอบรู้

บุญเชิด ภิบุญอนันตพงษ์ (2527 : 146) กล่าวว่า เกณฑ์หรือมาตรฐาน หมายถึงคะแนนจุดตัดของแบบทดสอบที่ใช้แบ่งผู้สอบออกเป็นผู้สอบได้ผู้สอบตก หรือผู้มีสมรรถภาพผู้ไม่มีสมรรถภาพ

จากความหมายของคะแนนจุดตัดของนักวัดผลที่กล่าวมาข้างต้น สรุปได้ว่า คะแนนจุดตัด หมายถึง คะแนนขั้นต่ำที่จะใช้ตัดสินว่า ผู้สอบเป็นผู้รอบรู้หรือไม่รอบรู้

การกำหนดคะแนนจุดตัด

การกำหนดคะแนนจุดตัดเป็นสิ่งสำคัญสำหรับแบบทดสอบอิงเกณฑ์ ซึ่งนักการศึกษาและนักวัดผลได้เสนอแนวทาง และวิธีการไว้ดังนี้

มิลล์แมน (Millman. 1973) ได้สรุปแนวคิดของการกำหนดเกณฑ์มาตรฐานไว้ว่า ควรคำนึงถึงสิ่งต่าง ๆ ต่อไปนี้

1. ระดับความสามารถในการทำแบบทดสอบของคนอื่น ๆ (Performance of other) โดยเฉพาะผลการสอบของกลุ่มที่ผ่านการยอมรับแล้ว ยื่อนนำมาใช้เป็นเกณฑ์ในการพิจารณากำหนดมาตรฐานได้

2. เนื้อหาสาระของข้อสอบในแบบทดสอบ (Item content) เนื้อหาของข้อสอบเป็นส่วนสำคัญส่วนหนึ่งในการที่จะพิจารณากำหนดมาตรฐาน ข้อสอบที่วัดเนื้อหาที่มีความสำคัญต้องให้ผู้สอบทำได้ในสัดส่วนที่สูง แต่ในส่วนของเนื้อหาที่ไม่สำคัญนัก หรือมีความยากเกินไป เกณฑ์มาตรฐานก็ควรลดต่ำลง

3. ผลของการศึกษาที่จะตามมาจากการกำหนดเกณฑ์มาตรฐาน (Educational Consequences) ถ้าหากว่า กำหนดเกณฑ์มาตรฐานต่ำเกินไปอาจมีผลเสียคือ ทำให้ผู้เรียนไม่สามารถเรียนบทเรียนที่สูงต่อไปได้ แต่ถ้ากำหนดมาตรฐานสูงเกินไป จะมีผลทำให้มีผู้ไม่รอบรู้ตกค้างเป็นจำนวนมาก

4. ผลทางจิตวิทยาและสภาพทางการเงิน (Psychological and Financial Cost) การกำหนดเกณฑ์มาตรฐาน ควรจะได้คำนึงถึงผลทางจิตใจที่เกิดขึ้นกับผู้สอบ เช่น แรงจูงใจ

ความเปื้อนหน่วย การทำลายอัตรโมโนทัศน์ และรวมถึงค่าใช้จ่ายในการสอนซ่อมเสริมให้กับผู้ที่ไม่ผ่านเกณฑ์

5. ความคลาดเคลื่อนในการเดา และการสุ่มข้อสอบ (Error due to chance and Item Sampling) การกำหนดมาตรฐานควรคำนึงถึงโอกาสที่ผู้สอบจะเดาถูก รวมทั้งลักษณะการสุ่มข้อสอบจากประชากรข้อสอบทั้งหมดในขอบเขตเนื้อหาที่กำหนดขึ้นด้วย

แฮมเบิลตัน และโนวิก (Hambleton and Novick. 1973 : 163) ได้กล่าวว่า แบบทดสอบอิงเกณฑ์เป็นแบบทดสอบที่กำหนดเกณฑ์มาตรฐาน หรือคะแนนจุดตัด (Cut-off-Score) โดยที่แบบทดสอบประกอบด้วยข้อคำถามที่สุ่มจากประชากรข้อสอบคือ จำนวนข้อสอบทั้งหมดภายในขอบเขตเนื้อหาที่กำหนดขึ้นสัดส่วนจำนวนข้อสอบที่ผู้สอบทำได้ต่อจำนวนข้อสอบทั้งหมด จะแทนด้วยสัญลักษณ์ π ในทางทฤษฎีค่า π จะถือว่าเป็นค่าประมาณของความรู้ที่แท้จริงของผู้สอบในขอบเขตของเนื้อหาที่กำหนด หรือมีความหมายว่า ค่า π คือ อัตราส่วนจำนวน ข้อสอบทุก ๆ ข้อ ในขอบเขตของเนื้อหานั้น ๆ จุดตัดที่เป็นเกณฑ์มาตรฐานจะแทนด้วยสัญลักษณ์ π_0 ซึ่งถือว่าเป็นจุดแบ่งผู้สอบออกเป็นสองส่วน คือ ผู้สอบที่ได้คะแนนสูงกว่าเกณฑ์มาตรฐาน ($\pi > \pi_0$) ซึ่งจะถือว่าเป็นผู้รอบรู้ (Master) และผู้ที่ได้คะแนนต่ำกว่าเกณฑ์มาตรฐาน ($\pi < \pi_0$) ซึ่งถือว่าเป็นผู้ไม่รอบรู้ (Nonmaster)

แฮมเบิลตันและคณะ (Hambleton et al. 1978 : 5) เรียกสัดส่วนข้อสอบที่ทำถูกต้องจำนวนข้อทั้งหมดว่า คะแนนโดเมนหาได้จากสูตร ดังนี้

$$\pi = \frac{X}{m}$$

X คือ คะแนนที่ได้จากการสอบ

m คือ จำนวนข้อในแบบทดสอบ

π คือ คะแนนโดเมน

เกลส (Glass. 1978 : 243 - 257) ได้แบ่งวิธีการกำหนดคะแนนจุดตัดออกเป็น 6 วิธี คือ

1. การใช้คะแนนของคนอื่น ๆ (Performance of Others)
2. การใช้วิธีนับถอยหลังจาก 100% (Counting Backwards from 100%)
3. การปรับคะแนนเกณฑ์อื่น ๆ (Bootstrapping on Other Criterion Scores)
4. การตัดสินใจ โดยการพิจารณาจากความสามารถต่ำสุดของผู้สอบ (Judging Minimal Competence)

5. การใช้ทฤษฎีการตัดสินใจ (Decision Theoretic Approaches)
6. การใช้วิธีการวิจัยเชิงปฏิบัติการ (Operations Research Methods)

1. การใช้คะแนนของคนอื่น

การกำหนดจุดตัดวิธีนี้ จะอาศัยหลักการกำหนดคะแนนจุดตัดที่สอดคล้องกับคะแนนเปอร์เซ็นต์ของผู้สอบผ่านเกณฑ์ซึ่งระบุไว้ล่วงหน้าโดยอาศัยผู้เชี่ยวชาญเนื้อหาวิชา แล้วให้ระบุเปอร์เซ็นต์ของผู้สอบผ่านว่าควรเป็นเท่าไรก่อน แล้วหาคะแนนที่สอดคล้องกับเปอร์เซ็นต์นั้น

2. การใช้วิธีนับลดจาก 100 %

วิธีการนี้จะอาศัยหลักการเกี่ยวกับการกำหนดค่าเกณฑ์หรือระดับมาตรฐานของจุดประสงค์เชิงพฤติกรรม คือ ค่าเกณฑ์ต้องกำหนดตามความสำคัญของจุดประสงค์ ถ้าจุดประสงค์ใดมีความสำคัญมากกว่า เกณฑ์ที่ต้องการ ต้องเป็น 100 % ถ้าจุดประสงค์ใดมีความสำคัญน้อยลงมา ค่าเกณฑ์ที่ต้องการก็ลดต่ำลงมาจาก 100 % ฉะนั้นเกณฑ์ที่ตั้งไว้ 100 % อาจลดลงมาเป็น 95 % , 90 % หรือ 80 %

3. การปรับคะแนนเกณฑ์อื่น ๆ

เป็นการกำหนดคะแนนจุดตัด โดยอาศัยเกณฑ์ภายนอกเป็นตัวเปรียบเทียบเกณฑ์ภายนอกนี้จะต้องเป็นที่ยอมรับทั่วไป และประจักษ์ที่จะชี้บอก "ความรอบรู้" หรือ "ความสำเร็จ"

4. การพิจารณาจากความสามารถต่ำสุดของผู้สอบ

เป็นวิธีตัดสินที่ขึ้นอยู่กับข้อคำถามในแบบทดสอบวิธีนี้ แกส (Glass, 1976. 247 - 251) เรียกว่า เป็นวิธีการพิจารณาจากความสามารถต่ำสุดของผู้สอบ เป็นวิธีการที่อาศัยแนวคิดเกี่ยวกับผู้สอบที่มีความรู้ และทักษะอยู่ในระดับคาบเส้น (Borderline) ระหว่างกลุ่มที่ได้คะแนนสูงกับกลุ่มที่ได้คะแนนต่ำคะแนนของผู้สอบที่มีความรู้ และทักษะอยู่ในระดับคาบเส้นก็จะเป็นคะแนนจุดตัดการตัดสินใจโดยอาศัยวิธีการนี้ จะต้องพิจารณาจากข้อคำถามในแบบทดสอบวิธีการนี้สามารถกำหนดคะแนนจุดตัดได้ ทั้งก่อนและหลังจากการนำแบบทดสอบไปทำการดำเนินการสอบแล้ว หากกล่าวโดยสรุปจะเห็นว่า กระบวนการวิธีนี้จะมุ่งพิจารณาที่เนื้อหาของแบบทดสอบเป็นสำคัญ

เพราะฉะนั้นผู้ตัดสินหรือผู้เชี่ยวชาญก็จะต้องพิจารณาว่า ผู้สอบที่อยู่ในระดับคาบเส้นจะสามารถตอบข้อคำถามแต่ละข้อในแบบทดสอบได้อย่างไร

การคำนวณคะแนนจุดตัด โดยวิธีการนี้มีขั้นตอนพื้นฐานดังนี้

1. เลือกผู้ตัดสิน หรือผู้เชี่ยวชาญ
2. กำหนดความรู้ และทักษะของผู้สอบที่อยู่ในระดับคาบเส้น
3. ฝึกฝนวิธีการตัดสินในแต่ละวิธีการที่เลือกใช้
4. เก็บรวบรวมผลการตัดสิน
5. นำผลการตัดสินมารวมกัน เพื่อกำหนดคะแนนจุดตัด

ในการเลือกผู้ตัดสินหรือผู้เชี่ยวชาญนั้น ถือเป็นขั้นตอนแรกของวิธีการนี้ซึ่งผู้ตัดสินต้องเป็นผู้มีความสามารถที่จะตัดสินระดับความรู้หรือทักษะที่แบบทดสอบมุ่งวัด การกำหนดจำนวนผู้ตัดสินอาจกำหนดเพียง 2 - 3 คนก็ได้ ถ้าผู้ตัดสินดังกล่าวเป็นผู้มีความสามารถ แต่บางกรณีก็จะต้องกำหนดเป็นจำนวนมาก อย่างไรก็ตามถ้าผู้ตัดสินมีจำนวนน้อย ก็จะมีผลต่อการกำหนดเกณฑ์มาตรฐานคือ มาตรฐานจะสูงเกินไป หรือต่ำเกินไป เพราะฉะนั้น ในการกำหนดจำนวนผู้ตัดสิน ถ้ามีจำนวนผู้ตัดสินมากเท่าไรก็จะยิ่งดี

ในทางปฏิบัติอาจจะแบ่งผู้ตัดสินเป็นกลุ่มย่อย ๆ แต่ในครั้งแรกควรมีการประชุมร่วมกันก่อนเพื่อที่จะให้ผู้ตัดสินนิยามคำว่า "ความรู้และทักษะของผู้สอบที่อยู่ในระดับคาบเส้น" ต่อจากนั้นจะต้องมีการฝึกผู้ตัดสินเพื่อจะได้เข้าใจสิ่งที่แบบทดสอบและการนำคะแนนสอบไปใช้

วิธีการตัดสินที่ขึ้นอยู่กับข้อคำถามในแบบทดสอบแบ่งออกเป็น 3 เทคนิค คือ

1. เทคนิคของนิเดลสกี
2. เทคนิคของแองกอฟฟ์
3. เทคนิคของฮีเบล

1. เทคนิคของนิเดลสกี

นิเดลสกีได้เสนอเทคนิคนี้ในปี ค.ศ.1954 ซึ่งเป็นเทคนิคที่เหมาะสมกับข้อสอบแบบเลือกตอบ วิธีการก็คือผู้เชี่ยวชาญจะต้องพิจารณาคำถาม และระบุว่าผู้สอบที่อยู่ในระดับคาบเส้นจะทราบว่าตัวเลือกใดเป็นตัวเลือกที่ผิด เพราะฉะนั้นในการรวบรวมผลการตัดสินของผู้เชี่ยวชาญอาจจะให้ผู้เชี่ยวชาญแต่ละบุคคลพิจารณาเอง หรือจะพิจารณาร่วมกันเป็นกลุ่มก็ได้ ถ้าจำนวนของผู้เชี่ยวชาญไม่มากก็อาจจะพิจารณาร่วมกัน ซึ่งวิธีการนี้จะให้ผลการตัดสินที่เที่ยงตรงกว่าเนื่องจากผู้เชี่ยวชาญได้มีโอกาสแลกเปลี่ยนความคิดเห็นซึ่งกันและกัน ดังนั้นหากกล่าวโดยสรุปเทคนิคของนิเดลสกี มีดังนี้

1. ให้ผู้เชี่ยวชาญแต่ละคนพิจารณาคำถามทั้งหมดและใช้ดินสอขีดลงไปที่ตัวเลือกที่ผิดที่เห็นว่า ผู้สอบที่อยู่ในระดับคาบเส้นจะไม่เลือก

2. อภิปรายถึงคำถามแต่ละข้อ ตามแนวทางดังนี้

2.1 ให้ผู้เชี่ยวชาญพิจารณาคำตอบที่ผิดก่อน และดูว่ามีผู้เชี่ยวชาญที่ท่านที่คิดว่า ผู้สอบที่อยู่ในระดับคาบเส้นจะไม่เลือก และมีที่ท่านไม่เห็นด้วย

2.2 ถ้าผู้เชี่ยวชาญมีความเห็นไม่ตรงกัน จะต้องให้แต่ละคนอธิบายเหตุผลแล้วถามผู้เชี่ยวชาญที่ไม่เห็นด้วย ซึ่งตอนนี้ยังไม่ต้องการสรุปจนกว่าผู้เชี่ยวชาญจะแน่ใจว่าได้เข้าใจความหมายของคำว่า ผู้สอบที่มีความสามารถต่ำสุดได้ตรงกัน

2.3 พิจารณาตัวเลือกที่เป็นตัวผิดตัวต่อไป

3. หลังจากที่ได้มีการอภิปรายคำถามทุกข้อในลักษณะนี้แล้วขั้นตอนต่อมาคือ จะต้องให้ผู้เชี่ยวชาญทบทวนถึงการตัดสินใจ โดยที่ผู้เชี่ยวชาญจะต้องทำเครื่องหมายกากบาท ลงไปบนตัวเลือกที่เขาแน่ใจว่า ผู้สอบที่มีความสามารถต่ำสุดจะไม่เลือก

4. รวบรวมผลการตัดสินใจ แล้วดำเนินการเช่นเดียวกันในคำถามข้อต่อไป

เทคนิคของนี้เดลสกีนี้ อาจให้ผู้เชี่ยวชาญแต่ละคนตัดสินใจเองก็ได้ โดยปราศจากการอภิปรายร่วมกัน แต่ปัญหาที่จะตามมาคือผู้เชี่ยวชาญจะยังไม่คุ้นเคยกับวิธีการตัดสินใจ เพราะฉะนั้นถ้าเป็นไปได้ก็ควรจะให้โอกาสแก่ผู้เชี่ยวชาญในการฝึกตัดสินใจ จากตัวอย่างของคำถามและอภิปรายซึ่งกันและกันก่อนที่จะมีการพิจารณาคำถามที่เหลือด้วยตนเองต่อไป

รูปแบบของข้อสอบแบบเลือกตอบบางครั้ง ก็เป็นปัญหาแก่การใช้เทคนิคของนี้เดลสกี คือคำถามในลักษณะดังต่อไปนี้

1. คำถามในรูปแบบปฏิเสธ เช่น

(๐) อาหารในข้อใดต่อไปนี้ ไม่มีวิตามินซี

ก. นม ข. น้ำส้ม ค. กระหล่ำปลี ง. มันฝรั่ง

ในการพิจารณาดำเนินการคำตอบที่ถูกผู้เชี่ยวชาญจะต้องจำไว้ว่า อาหารอะไรที่ให้มีวิตามินซีดีกว่า และต้องรู้ว่าผู้สอบที่อยู่ในระดับคาบเส้นจะจำว่าเป็นคำตอบที่ผิด

2. แบบสอบเลือกตอบประเภทถูกผิดทุกตัวเลือก (Multiple True-False) ดังตัวอย่างต่อไปนี้

(๐) ในระหว่างสงครามโลกครั้งที่ 2 ประเทศสหรัฐอเมริกาสู้รบกับประเทศใด

1. เยอรมัน 2. รัสเซีย 3. อิตาลี 4. ญี่ปุ่น

ก. ข้อ 1 เท่านั้น

ข. ข้อ 2 เท่านั้น

ค. ข้อ 1 และข้อ 4 เท่านั้น

ง. ข้อ 1, 3 และข้อ 4

จ. ข้อ 1, 2, 3 และข้อ 4

จากตัวอย่างข้างต้น ผู้เชี่ยวชาญจะต้องพิจารณาจำนวนตัวเลือกที่ผู้สอบที่อยู่ในระดับคาบเส้นจะระบุว่า ตัวเลือกใดเป็นตัวเลือกที่ถูกต้อง ตัวเลือกใดไม่ถูกต้องและตัวเลือกใดที่ไม่แน่ใจแล้ว ผู้เชี่ยวชาญก็ต้องพิจารณาว่า ตัวเลือกใดที่ผู้สอบที่อยู่ในระดับคาบเส้นจะไม่เลือกหรือจะกำจัดทิ้งไป

ในกรณีที่แบบทดสอบเป็นคำถามปฏิเสธ หรือแบบทดสอบเลือกตอบแบบถูกผิดทุกตัว เลือกดังตัวอย่างข้างต้น ถ้าใช้เทคนิคของนิเดลสกี ก็ต้องใช้เวลาแก่ผู้เชี่ยวชาญในการฝึกก่อนที่ผู้เชี่ยวชาญแต่ละคนจะทำการพิจารณาตัดสินด้วยตนเอง หลังจากที่คุณผู้เชี่ยวชาญพิจารณาเสร็จแล้วก็จะต้องให้อธิบายเหตุผลในการตัดสินของเขา เพื่อให้เกิดความแน่ใจในสิ่งที่ผู้เชี่ยวชาญได้พิจารณา

3. คำถามในลักษณะของการคำนวณทางคณิตศาสตร์ ตัวเลือกของคำตอบที่ผิดจะเป็นผลมาจากการคิดคำนวณผิด ซึ่งก็ไม่ได้เด่นชัดนัก เพราะฉะนั้นผู้เชี่ยวชาญจะต้องใช้เวลามากในการพิจารณาแม้แต่ผู้เชี่ยวชาญที่มีความเชี่ยวชาญมาก ๆ ก็ต้องใช้เวลาในการคิดคำนวณถึงวิธีการที่จะนำไปสู่การตอบผิด เพื่อที่จะหลีกเลี่ยงปัญหานี้ก็ควรจะให้แบบทดสอบแก่ผู้ตัดสินอีก 1 ฉบับ และแสดงให้เห็นถึงรูปแบบของการคิดคำนวณที่ผิด ซึ่งจะนำไปสู่การเลือกคำตอบที่ผิด

ดังนั้น ประเด็นสำคัญของการนำเทคนิคของนิเดลสกีไปใช้ก็คือ จะต้องให้คำเฉลยแก่ผู้เชี่ยวชาญ ซึ่งจะทำให้การพิจารณาคำถามง่ายขึ้น แต่ถ้าไม่ให้เฉลยหรือบอกตัวเลือกที่เป็นคำตอบที่ถูกแก่ผู้เชี่ยวชาญ ผู้เชี่ยวชาญจะรู้สึกว่าเขากำลังถูกทดสอบและอาจลืมนึกไปว่าเขากำลังจะต้องพิจารณาตัดสินผู้สอบที่อยู่ในระดับคาบเส้น ยิ่งกว่านั้นกระบวนการตัดสินใจจะใช้เวลานาน ถ้าผู้เชี่ยวชาญจะต้องหาคำตอบที่ถูกต้องในแต่ละข้อคำถามเอง

ต่อไปนี้เป็นตัวอย่างการใช้เทคนิคของนิเดลสกี

สมมติว่า แบบทดสอบเลือกตอบฉบับหนึ่งมีคำถาม 10 ข้อ เป็นแบบทดสอบเลือกตอบ 5 ตัวเลือกคือ ก. ข. ค. ง. และ จ. ดังนี้

ตาราง 1 การคำนวณตามเทคนิคของนีเดลสกี

ข้อคำถาม	ตัวเลือก	จำนวนของ ตัวเลือกที่ตัดทิ้ง	จำนวนของ ตัวเลือกที่เหลือ	โอกาสของ การตอบ
1	ก ☒ ข ☒ ค ☒ ง ☒ จ	3	2	$1/2 = .50$
2	☒ ก ☒ ข ☒ ค ☒ ง ☒ จ	4	1	$1/1 = 1.00$
3	☒ ก ☒ ข ค ☒ ง จ	3	2	$1/2 = .50$
4	ก ☒ ข ค ☒ ง ☒ จ	2	3	$1/3 = .33$
5	☒ ก ☒ ข ☒ ค ☒ ง ☒ จ	4	1	$1/1 = 1.00$
6	ก ข ☒ ค ง จ	0	5	$1/5 = .20$
7	ก ข ค ☒ ง ☒ จ	1	4	$1/4 = .25$
8	☒ ก ข ☒ ค ง จ	1	4	$1/4 = .25$
9	ก ☒ ข ค ง จ	0	5	$1/5 = .20$
10	ก ☒ ข ค ง จ	0	5	$1/5 = .20$
รวม				4.43

ตัวเลือกที่ว่างกลม คือตัวเลือกที่เป็นคำตอบที่ถูก ส่วนตัวเลือกที่ผู้เชี่ยวชาญจะคิดว่าผู้สอบที่มีความสามารถต่ำสุดจะตัดทิ้งไปหรือไม่เลือก ผู้เชี่ยวชาญก็จะเขียนเครื่องหมายกากบาททับลงไปบนตัวเลือกนั้น สำหรับในการคำนวณหาคะแนนจุดตัดนั้น คะแนนในแต่ละข้อมีค่าเท่ากับ 1 คะแนน ดังนั้นในแต่ละข้อจะต้องหาว่า มีจำนวนที่ผู้สอบที่อยู่ในระดับความสามารถต่ำสุดจะตัดทิ้งเป็นจำนวนเท่าไร แล้วจึงหาโอกาสของการตอบถูก (Probability of Correct Answer) ซึ่งเท่ากับ 1 หาดด้วยจำนวนตัวเลือกทั้งหมด ลบด้วยจำนวนตัวเลือกที่ตัดทิ้ง หากเขียนในรูปสูตรทั่วไปจะได้ว่า

$$\text{โอกาสของการตอบถูก} = \frac{1}{\text{จำนวนตัวเลือกทั้งหมด} - \text{จำนวนตัวเลือกที่ตัดทิ้ง}}$$

ดังนั้น สำหรับข้อสอบแบบเลือกตอบที่มี 5 ตัวเลือก

ถ้าตัวเลือกทั้งหมดตัดทิ้งไม่ได้เลย

$$\text{โอกาสของการตอบถูก} = \frac{1}{5 - 0} = \frac{1}{5} = 0.20$$

ถ้าตัวเลือกทั้งหมดตัดทิ้งได้ 1 ตัว

$$\text{โอกาสของการตอบถูก} = \frac{1}{5 - 1} = \frac{1}{4} = 0.25$$

ถ้าตัวเลือกทั้งหมดตัดทิ้งได้ 2 ตัว

$$\text{โอกาสของการตอบถูก} = \frac{1}{5 - 2} = \frac{1}{3} = 0.33$$

ถ้าตัวเลือกทั้งหมดตัดทิ้งได้ 3 ตัว

$$\text{โอกาสของการตอบถูก} = \frac{1}{5 - 3} = \frac{1}{2} = 0.50$$

ถ้าตัวเลือกทั้งหมดตัดทิ้งได้ 4 ตัว

$$\text{โอกาสของการตอบถูก} = \frac{1}{5 - 4} = \frac{1}{1} = 1.00$$

ตัวอย่างในการหาคะแนนจุดตัดตามวิธีการของนิตลสกี ดังในตาราง 1 นั้น เป็นการหาคะแนนจากผู้เชี่ยวชาญ 1 คน ในกรณีที่ผู้ผู้เชี่ยวชาญมากกว่า 1 คน จะต้องนำคะแนนของผู้เชี่ยวชาญทั้งหมดมาคิดในรูปของคะแนนเฉลี่ย (Mean) มัธยฐาน (Median) หรือ Trimmed Mean แล้วแต่ความเหมาะสมดังตัวอย่างในตาราง 2

ตาราง 2 ผลการหาคะแนนจุดตัดจากผู้เชี่ยวชาญ โดยเทคนิคของนิตลสกี

ผู้เชี่ยวชาญคนที่	คะแนน
1	92.50
2	77.25
3	67.00
4	66.67
5	65.33
รวม	368.75

จากตาราง 2 จะได้คะแนนเฉลี่ยเท่ากับ $368.75/5 = 73.75$

จะได้มัธยฐานเท่ากับ 67.00

จะได้ Trimmed Mean เท่ากับ $210.92/3 = 70.31$ ซึ่งการหาโดยวิธีนี้จะตัดคะแนนสูงสุดคือ 92.50 และคะแนนต่ำสุดคือ 65.33 ออกไป

การหาคะแนนจุดตัดตามเทคนิคของนิตลสกี ดังกล่าวข้างต้นเป็นวิธีการตามแนวคิดของ ลิฟวิงสตัน และซีกี (Livingston and Zieky. 1982 : 17 - 24)

2. เทคนิคของแองกอฟฟ์

แองกอฟฟ์ได้เสนอเทคนิคของเขาในปีค.ศ. 1971 ซึ่งเป็นเทคนิคที่มีลักษณะคล้ายกับเทคนิคของนิตลสกี แต่เทคนิคนี้จะใช้กับแบบทดสอบที่ไม่ใช่แบบเลือกตอบ คะแนนจุดตัดที่หาโดยเทคนิคของแองกอฟฟ์จะหาจากคะแนนที่ได้จากแต่ละข้อคำถาม โดยการพิจารณาในภาพรวม จากการพิจารณาถึงโอกาสที่ผู้สอบที่อยู่ในระดับคาบเส้นจะตอบแต่ละข้อคำถามได้ถูกต้อง ซึ่งโอกาสที่ผู้เชี่ยวชาญใช้ในการตัดสินจะอยู่ระหว่าง 0.00 ถึง 1.00 กล่าวคือ ถ้าเป็นข้อคำถามที่ง่ายโอกาสที่ผู้สอบที่อยู่ในระดับคาบเส้น จะตอบถูกจะมีค่าใกล้ 1.00 แต่ถ้าเป็นข้อคำถามที่ยากโอกาสที่ผู้สอบ

ที่อยู่ในระดับคาบเส้นจะตอบถูกจะมีค่าใกล้ 0 อย่างไรก็ตาม เทคนิคของแองกอฟฟ์ หากนำมาใช้กับแบบทดสอบแบบเลือกตอบ โอกาสที่ผู้สอบที่อยู่ในระดับคาบเส้นจะตอบแต่ละข้อคำถามถูกไม่ควรต่ำกว่าโอกาสการตอบถูกโดยการเดา หากกล่าวโดยสรุปเทคนิคของแองกอฟฟ์ มีดังนี้

1. ให้ผู้เชี่ยวชาญแต่ละคนพิจารณาตัดสินข้อคำถามประมาณ 2-3 ข้อแรก แล้วอภิปรายร่วมกัน ดังนี้

1.1 ให้ผู้เชี่ยวชาญบอกความน่าจะเป็น ในการตัดสินแต่ละข้อคำถามแล้ว เขียนลงบนกระดานดำ

1.2 ให้ผู้เชี่ยวชาญให้ค่าความน่าจะเป็นสูงสุด และต่ำสุดในแต่ละข้อคำถาม ชี้แจงเหตุผล

1.3 ผู้เชี่ยวชาญสามารถเปลี่ยนแปลงการตัดสินได้ถ้าต้องการเปลี่ยนแปลง แต่อย่างไรก็ตาม ผู้เชี่ยวชาญจะต้องมั่นใจว่าได้เข้าใจ คำว่า "ความสามารถต่ำสุดของผู้สอบ" ตรงกัน

2. หลังจากที่มีการอภิปรายข้อคำถาม 2-3 ข้อแรกแล้ว คำถามข้อที่เหลือก็ ดำเนินการเช่นเดียวกัน

3. เก็บรวบรวมผลการตัดสิน

ส่วนในการคำนวณหาคะแนนจุดตัดสามารถทำได้ 3 วิธี (Zieky. 1987 : 5 - 6) คือวิธีที่ 1 ใช้คะแนนเฉลี่ย (Mean) วิธีที่ 2 ใช้ค่ามัธยฐาน (Median) วิธีที่ 3 ใช้ทั้งค่าเฉลี่ย และค่ามัธยฐานร่วมกัน โดยตัดค่าสูงสุดและค่าต่ำสุดออก วิธีการดังกล่าวนี้เรียกว่า Trimmed Mean ดังตัวอย่างในตาราง 3 สมมติว่า มีข้อคำถามอยู่ 10 ข้อ และให้ผู้เชี่ยวชาญ 5 คน ตัดสินได้ ผลดังนี้

ตาราง 3 ผลการตัดสินโดยใช้เทคนิคของแองกอฟฟ์

ข้อคำถาม	ผู้ตัดสิน				
	A	B	C	D	E
1	.95	.90	1.00	.85	.90
2	.80	.80	1.00	.80	.85
3	.90	.85	1.00	.80	.85
4	.60	.65	.95	.70	.75
5	.75	.70	.90	.75	.75
6	.40	.60	.90	.65	.65
7	.50	.65	.90	.60	.70
8	.25	.30	.85	.45	.55
9	.25	.25	.80	.30	.50
10	.40	.30	.75	.25	.45
รวม	5.80	6.00	9.05	6.15	6.90

ค่าเฉลี่ยเท่ากับ 6.78

ค่ามัธยฐานเท่ากับ 6.15

Trimmed Mean เท่ากับ 6.35 (โดยตัดค่าสูงสุดและค่าต่ำสุดออก)

3. เทคนิคของอีเบล

อีเบลได้กล่าวถึงจุดอ่อนของการนิยามคะแนนจุดตัดหรือคะแนนผ่าน โดยการพิจารณาจากร้อยละของคะแนนรวม ทำให้ขาดรายละเอียดอื่น ๆ ที่จะนำมาใช้ในการพิจารณาด้วย เนื่องจากในบางครั้งข้อสอบอาจมีความยากมากหรือง่ายเกินไป หรือมีอำนาจจำแนกต่ำมากเกินไป ผู้ออกข้อสอบคาดหวังไว้ ดังนั้นการที่ผู้สอบจะสอบผ่านหรือไม่จึงขึ้นอยู่กับคำถามในแบบทดสอบมากกว่าการพิจารณาจากระดับความสามารถของผู้สอบ อีเบลจึงเสนอวิธีการแก้ไขจุดอ่อนดังกล่าว โดยการวิเคราะห์ความเกี่ยวข้อง (Relevance) หรือความสำคัญ (Importance) ของแต่ละข้อคำถามในแบบทดสอบ สำหรับในเรื่องของความเกี่ยวข้อง อีเบลได้แบ่งออกเป็น 4 ประเภทคือ มีความจำเป็น (Essential) มีความสำคัญ (Important) เป็นที่ยอมรับได้ (Acceptable) และไม่แน่ใจ (Questionable) ส่วนในเรื่องของระดับความยาก อีเบลได้แบ่งเป็น 3 ระดับ คือ

ง่าย (Easy) ปานกลาง (Medium) และยาก (Hard) และกำหนดร้อยละที่คาดหวังที่จะให้ผ่านข้อคำถามในแต่ละกลุ่ม ซึ่งร้อยละที่คาดหวังดังกล่าวนี้ถือว่าเป็นปริมาณความคาดหวังของผู้ที่มีความสามารถต่ำสุดจะสอบผ่านดังรายละเอียดในตาราง 4

ตาราง 4 ร้อยละที่คาดหวังจะสอบผ่านตามลักษณะที่เกี่ยวข้องและระดับความยาก

ลักษณะที่เกี่ยวข้อง	ระดับความยาก		
	ง่าย	ปานกลาง	ยาก
มีความจำเป็น	100%	-	-
มีความสำคัญ	90%	70%	-
เป็นที่ยอมรับได้	90%	60%	40%
ไม่แน่ใจ	70%	50%	30%

เทคนิคของอีเบลนั้น สามารถสรุปได้ดังนี้

1. ให้ผู้เชี่ยวชาญแต่ละคนพิจารณาข้อคำถาม แล้วจัดแยกเป็น 12 ประเภท (คือลักษณะที่เกี่ยวข้องมี 4 ลักษณะ และระดับความยากมี 3 ระดับ)
2. อภิปรายถึงแต่ละข้อคำถามดังนี้
 - 2.1 ถามผู้เชี่ยวชาญว่ามีกี่คนที่จัดคำถามอยู่ในระดับง่าย ปานกลางและยาก ถ้าผู้เชี่ยวชาญจัดไม่เหมือนกันก็ควรจะให้อธิบายเหตุผล
 - 2.2 ถามผู้เชี่ยวชาญว่า มีกี่คนที่แยกประเภทของข้อคำถามออกเป็นมีความจำเป็น มีความสำคัญ เป็นที่ยอมรับได้ และไม่แน่ใจ ถ้าผู้เชี่ยวชาญแยกประเภทของข้อคำถามไม่เหมือนกันก็ควรให้อธิบายเหตุผลเช่นเดียวกัน
 - 2.3 ให้โอกาสแก่ผู้เชี่ยวชาญจัดประเภทของข้อคำถามใหม่ ถ้าต้องการเปลี่ยนแปลง
3. ให้ผู้เชี่ยวชาญกำหนดร้อยละของผู้สอบที่คาดหวังว่าจะตอบในแต่ละข้อคำถามได้ถูกต้องทั้ง 12 ประเภท
4. รวบรวมผลการตัดสิน

วิธีการหาคะแนนจุดตัดของอีเบลก็คือ นำจำนวนข้อสอบในแต่ละลักษณะคูณกับค่าร้อยละที่คาดหวัง แล้วนำผลคูณที่ได้มาหาค่าเฉลี่ย ค่าเฉลี่ยที่ได้ก็จะเป็นคะแนนจุดตัด หรือคะแนนที่ผู้มีความสามารถต่ำสุดจะสอบผ่าน ดังตัวอย่างในตาราง 5

ตาราง 5 ตัวอย่างของการคำนวณตามเทคนิคของอีเบล

ลักษณะของคำถาม	จำนวนของข้อความ	ร้อยละที่คาดหวัง	จำนวนข้อคำถาม คูณร้อยละที่คาดหวัง
มีความจำเป็น	94	100	9400
มีความสำคัญ			
ง่าย	106	90	9540
ปานกลาง	153	70	10710
เป็นที่ยอมรับได้			
ง่าย	24	80	1920
ปานกลาง	49	60	2940
ยาก	52	40	2080
ไม่แน่ใจ			
ง่าย	4	70	280
ปานกลาง	11	50	550
ยาก	7	30	210
รวม	500		37130

จากตาราง 5 จะได้คะแนนจุดตัดเท่ากับ $37130/500 = 74.26$ หรือ 74%

จากวิธีการกำหนดคะแนนจุดตัดของแบบทดสอบอิงเกณฑ์ ที่ขึ้นอยู่กับข้อคำถามในแบบทดสอบ จะเห็นได้ว่า แต่ละวิธีมีข้อจำกัดต่าง ๆ กัน ดังนี้

เทคนิคของแองกอฟฟ์เป็นเทคนิคที่ง่ายและใช้ได้รวดเร็ว (Living-ston and Zicky. 1982 : 54) หากกลุ่มผู้เชี่ยวชาญพิจารณากำหนดคะแนนจุดตัดมีความเชี่ยวชาญในเนื้อหาที่จะนำมาในการเขียนข้อสอบ (Norcini, Shea and Kanya. 1988 : 57) เนื่องจากเทคนิคนี้ผู้เชี่ยวชาญจะต้องพิจารณาถึงโอกาสของผู้สอบที่มีความสามารถต่ำสุดหรือผู้ที่อยู่ในระดับคาบเส้นที่จะตอบข้อสอบในแต่ละข้อคำถามได้ถูกต้อง เพราะฉะนั้นก่อนอื่นผู้เชี่ยวชาญจะต้องให้คำนิยามหรือกำหนดความสามารถต่ำสุดของผู้สอบก่อน โดยผู้เชี่ยวชาญแต่ละคนจะต้องคาดคะเนความยากในแต่ละข้อคำถามของแบบทดสอบแล้วนำผลที่ได้จากการคาดคะเนของผู้เชี่ยวชาญมาหาค่าเฉลี่ยค่าที่ได้ก็จะเป็นคะแนนจุดตัด (Mills. 1983 : 284)

ส่วนเทคนิคของอีเบล เป็นเทคนิคที่ผู้เชี่ยวชาญจะต้องพิจารณาถึงความยาก และความสำคัญของข้อคำถามประกอบกัน ซึ่งเทคนิคนี้จะทำให้ผู้เชี่ยวชาญต้องใช้เวลาในการพิจารณามาก และคงจะไม่เหมาะสมแบบทดสอบที่ตอบแบบสั้น (Zicky. 1987 : 16) นอกจากนี้แบบทดสอบของอีเบลไม่เหมาะสมกับแบบทดสอบที่มีความยาวมาก ๆ เนื่องจากผู้เชี่ยวชาญจะต้องเสียเวลาในการจัดประเภทของแบบทดสอบไปตามค่าความยาก และความสำคัญ (Cross, Impara, Frary and Jacger. 1984 : 114)

สำหรับเทคนิคของนีเดลสกี เป็นเทคนิคที่เหมาะสมกับแบบทดสอบเลือกตอบเท่านั้น (Zicky. 1987 : 11) การกำหนดคะแนนจุดตัดโดยใช้เทคนิคนี้จะต้องพิจารณาถึงความยากของข้อคำถามเป็นรายตัวเลือกว่า ตัวเลือกใดที่ผู้สอบที่มีความสามารถต่ำสุด หรือผู้สอบในระดับคาบเส้นสามารถทราบได้ว่าเป็นตัวเลือกที่ผิด (Plake and Milican. 1989 : 46 ; Cross, Impara, Frary and Jacger. 1984 : 114) ซึ่งความเป็นจริงแล้ว ผู้เชี่ยวชาญก็จะมีโอกาสรู้ได้ (Zicky. 1987 : 16)

4.5 การใช้ทฤษฎีการตัดสินใจ

การกำหนดจุดตัดวิธีนี้อาศัยผลการสอบมาใช้ประกอบการพิจารณาตัดสินคะแนนจุดตัดที่เหมาะสมซึ่งมีอยู่ด้วยกันหลายวิธีและบางวิธีจะใช้การนิยามความรอบรู้ ด้วยคะแนนมวลความรู้ แต่บางวิธีจะใช้การนิยามความรอบรู้ด้วยคะแนนสอบ หรือคะแนนดิบ

การกำหนดคะแนนจุดตัดซึ่งนิยามความรอบรู้ด้วยคะแนนมวลความรู้ เช่น วิธีของเบส์ (Bayesian Decision Procedure) วิธีของแฟเนอร์ (Fahner's Procedure)

การกำหนดคะแนนจุดตัดซึ่งนิยามความรอบรู้ด้วยคะแนนสอบหรือคะแนนดิบ เช่น วิธีของเกลส (Glass) และวิธีของเบอร์ก (Berk) (บุญเชิด ภิญญอนันตพงษ์. 2526 : 184) ซึ่งในที่นี้ผู้วิจัยจะกล่าวเฉพาะวิธีของเกลส และวิธีของเบส์เท่านั้น

การกำหนดคะแนนจุดตัดตามวิธีของเกลส

เกลส (Glass. 1978 : 237 - 261)เสนอการกำหนดคะแนนจุดตัดโดยแบ่งคนออกเป็นสองกลุ่ม โดยอาศัยเกณฑ์ภายนอกที่เกี่ยวกับสิ่งที่ศึกษา เช่น แบ่งเป็นผู้ที่จบการศึกษา ผู้ที่ได้รับจ้าง และไม่ได้รับจ้างเป็นต้น สัดส่วนของคนในกลุ่มนี้แทนด้วย PE และ 1- PE ตามลำดับ แล้วให้บุคคลทั้งสองกลุ่มนี้ทำแบบทดสอบอิงเกณฑ์และกำหนดคะแนนจุดตัด (CX) ขึ้นมาใช้แบ่งบุคคลออกเป็นผู้สอบผ่านและไม่ผ่าน ดังนั้นจะได้สัดส่วนดังตารางต่อไปนี้

		เกณฑ์ภายนอก		
		ผ่าน	ไม่ผ่าน	
เกณฑ์ของแบบทดสอบ	ไม่ผ่าน	P_A	P_B	$1 - P_C$
	ผ่าน	P_C	P_D	P_C
		P_E	$1 - P_E$	1

- เมื่อ P_A แทน สัดส่วนของการปฏิเสธที่ผิด คือ คนที่ไม่ผ่านแบบทดสอบอิงเกณฑ์ แต่ ผ่านเกณฑ์ภายนอก
- P_D แทน สัดส่วนการยอมรับที่ผิด คือ คนที่ผ่านแบบทดสอบอิงเกณฑ์ แต่ไม่ผ่านเกณฑ์ภายนอก
- P_B แทน สัดส่วนของคนที่ไม่ผ่านทั้งแบบทดสอบอิงเกณฑ์ และ เกณฑ์ภายนอก
- P_C แทน สัดส่วนของคนที่ผ่านมาทั้งแบบทดสอบอิงเกณฑ์ และ เกณฑ์ภายนอก

การกำหนดเกณฑ์ภายนอกนั้นจะไม่ทำให้เปลี่ยนแปลง แต่คะแนนจุดตัดของแบบทดสอบอิงเกณฑ์จะแปรผันไปได้หลายค่า แล้วแต่การกำหนดซึ่งจะทำให้สัดส่วนของ P_A, P_B, P_C และ P_D แปรผันตามไปด้วย คะแนนจุดตัดของแบบทดสอบแบบอิงเกณฑ์ก็คือ ค่าฟังก์ชันของคะแนนเกณฑ์ C_X ที่ทำให้ $f(C_X)$ มีค่าน้อยที่สุด

$$f(C_X) = (P_A + P_D)/(P_B + P_C)$$

ถ้ามีค่าต่ำสุดจะได้คะแนนเกณฑ์ที่เหมาะสม โดยสมมติให้การจำแนกผิดทางบวกและการจำแนกผิดทางลบเท่ากัน ในทางตรงกันข้าม ให้บุคคลที่ไม่ผ่านแบบทดสอบแต่ผ่านเกณฑ์ภายนอกเป็น α และบุคคลที่ผ่านแบบทดสอบแต่ไม่ผ่านเกณฑ์ภายนอกเป็น β จะได้ฟังก์ชันต่ำสุดของคะแนนเกณฑ์คือ

$$f(C_X) = (\alpha P_A + \beta P_D) / (P_B + P_C)$$

แต่ค่าของฟังก์ชันนี้มีความไวต่อค่า α และ β ซึ่งได้จากการพิจารณาของผู้ตัดสินมาก ค่านี้จึงแปรผันไปตามความเห็นของผู้ตัดสินซึ่งก็ยังคงเป็นการกำหนดตามความพอใจเช่นกัน ดังนั้นควรพิจารณาองค์ประกอบอื่น ๆ ที่เกี่ยวข้องกับการสอบผ่านหรือไม่ผ่านด้วย

การกำหนดคะแนนจุดตัดตามวิธีของเบส์

สวามินาธาน แฮมเบิลตัน และอัลไลนา (Swaminathan, Hambleton and Algina. 1975 : 87 - 98) ได้เสนอการกำหนดคะแนนจุดตัดโดยวิธีของเบส์ ดังนี้

1. แปลงคะแนนที่ได้จากการสอบของผู้สอบ (X_i) ให้เป็น g_i โดยแปลง แบบอาร์คไซน์ (Arc Sine) เพื่อให้การแจกแจงเป็นปกติ ซึ่งจะช่วยให้การคำนวณสะดวกขึ้น โดยใช้สูตร

$$g_i = \sin^{-1} \left\{ (x_i + 3/8) / (n + 3/4) \right\}^{1/2}$$

เมื่อ n คือจำนวนข้อคำถามของแบบทดสอบ

การแจกแจงนี้มีค่าเฉลี่ยเท่ากับ γ_i และความแปรปรวนเท่ากับ v โดยที่

$$\gamma_i = \sin^{-1} \sqrt{\pi_i}$$

$$v = (4n + 2)^{-1}$$

การประมาณนี้จะใช้ได้ดีเมื่อ π มีค่าระหว่าง .15 ถึง .85 และจำนวนข้อของข้อสอบไม่ต่ำกว่า 8 ข้อ เนื่องจากการแจกแจงของ g_i เราทราบค่าความแปรปรวนแต่เราไม่ทราบค่า γ_i การแจกแจงของ g_i จึงเป็นการแจกแจงที่มีเงื่อนไขคือ

$$g_i / \gamma_i \sim N(\gamma_i, v)$$

เมื่อ $N(\gamma_i, V)$ แทนการแจกแจงปกติ โดยมีค่าเฉลี่ยเท่ากับ γ_i และความแปรปรวนเท่ากับ V การแจกแจงของ g_i จึงขึ้นอยู่กับ γ_i เพื่อให้การแจกแจงภายหลังของ γ_i ต้องมีข้อตกลงว่า γ_i เป็นตัวอย่างที่ได้จากการสุ่ม โดยสมมุติว่าการแจกแจงของ γ_i เป็นปกติด้วยค่าเฉลี่ย α และความแปรปรวน ϕ ซึ่งเราไม่ทราบทั้งค่า α และ ϕ

ดังนั้น การแจกแจงของ γ_i จึงขึ้นอยู่กับ α และ ϕ อย่างไรก็ตาม โนวิกและคณะ (1973) เสนอแนะว่าเราสามารถหาได้จากการแจกแจงอินเวอร์ส ไค-สแควร์ (Inverse Chi-Square) ที่มีองศาแห่งความเป็นอิสระเท่ากับ v และสเกลพารามิเตอร์ λ

$$\begin{aligned} \text{นั่นคือ } \phi &= \lambda / (v-2) \\ \lambda &= (v-2) \phi \end{aligned}$$

ค่า v ที่เหมาะสมคือ 8 ส่วนการประมาณค่า ϕ จะต้องมีข้อมูลเกี่ยวกับ π ดังนั้น จึงต้องกำหนดข้อคำถาม (c) เพื่อให้ทราบข้อมูลเกี่ยวกับค่า π ค่าประมาณที่ได้จากการแปลงค่า π จากแบบทดสอบจำนวน c ข้อ จะมีการแจกแจงแบบปกติ ด้วยความแปรปรวนเท่ากับ $(4c + 2)^{-1}$ ดังนั้นเราสามารถใช้ค่า $(4c + 2)^{-1}$ เป็นค่าประมาณของ ϕ และ λ ได้

จากการศึกษาของเลวิสและคณะ (1973) พบว่าการแจกแจงของ γ_i มีการแจกแจงดังนี้

$$\gamma_i / \text{Data} \sim N(\mu_i, \sigma_i^2)$$

$$\text{เมื่อ } \mu_i = \bar{g} + \rho^* (g_i - \bar{g})$$

$$\text{และ } \frac{1 + (m - 1)}{(4n + 2)m}$$

$$\sigma_i^2 = \frac{1 + (m - 1)\rho^*}{(4n + 2)m} + (g_i - \bar{g})^2 \sigma^{*2}$$

$$\bar{g} = m^{-1} \sum_{i=1}^m g_i$$

เมื่อ m คือ จำนวนผู้เข้าสอบ
 n คือ จำนวนข้อคำถามในแบบทดสอบ

สำหรับค่า ρ^* และ σ^2 ได้จากตาราง (Wang) เมื่อทราบค่า v และ λ

2. กำหนดคะแนนจุดตัด (π_0) เพื่อแบ่งผู้สอบเป็นกลุ่มผู้รอบรู้และไม่รอบรู้ แล้วแปลง π_0 เป็น γ_0 จากสูตร

$$\gamma_0 = \sin^{-1} \sqrt{\pi_0}$$

3. หาคะแนนมาตรฐานปกติจากสูตร

$$Z_{oi} = \gamma_0 - \mu_i / \sigma_i$$

4. หาคำน่าจะเป็นของการสอบผ่านและไม่ผ่านของคะแนนที่ได้จากผู้สอบซึ่งหาได้จากตารางพื้นที่ใต้โค้งปกติดังนี้

Prob [$Z < Z_{oi}$ / Data] แทนความน่าจะเป็นในการสอบไม่ผ่าน

Prob [$Z \geq Z_{oi}$ / Data] แทนความน่าจะเป็นในการสอบผ่าน

5. หาความสูญเสียที่คาดหวังของการตัดสินใจผ่านและไม่ผ่านของคะแนนจากผู้สอบ
ดังนี้

I_{12} Prob $[Z < Z_{io} / \text{Data}]$ แทนความสูญเสียที่คาดหวังของ
การตัดสินใจผ่าน

I_{21} Prob $[Z \geq Z_{oi} / \text{Data}]$ แทนความสูญเสียที่คาดหวังของ
การตัดสินใจไม่ผ่าน

เมื่อ I_{12} แทนความสูญเสียที่เกิดจากความคลาดเคลื่อนในการจำแนกผิดทางบวก

I_{21} แทนความสูญเสียที่เกิดจากความคลาดเคลื่อนในการจำแนกผิดทางลบ

6. เมื่อคะแนนที่น้อยที่สุดที่ทำให้ค่า $I_{21} \text{Prob} [Z \geq Z_{oi} / \text{Data}]$ มากกว่า
 $I_{12} \text{Prob} [Z < Z_{oi} / \text{Data}]$ เป็นคะแนนจุดตัด

6. การใช้วิธีวิจัยเชิงปฏิบัติการ

วิธีนี้อาศัยผลลัพธ์ทางด้านคุณค่าอย่างใดอย่างหนึ่งที่เกิดขึ้นมาช่วยในการพิจารณา
เกณฑ์ โดยใช้วัดผลทางคุณค่าสูงสุดเป็นคะแนนเกณฑ์ โดยพิจารณาจากกราฟที่แสดงความสัมพันธ์
ระหว่างผลจากการสอบด้วยแบบทดสอบอิงเกณฑ์และการวัดผลด้านคุณค่านั้น คะแนนตรงจุดสูงสุด
ของกราฟจะถือเป็นคะแนนเกณฑ์

ซิคี และ ลีฟวิงสตัน (Berk. 1980 : 106 - 107 citing Zieky and
Livingston. 1977) ได้เสนอวิธีหาเกณฑ์ไว้สองวิธี โดยพิจารณาจากนักเรียน 2 พวกดังนี้

พวกที่อยู่บริเวณคาบเส้น หรือ เส้นแบ่ง (Borderline Group) วิธีการนี้ครั้งแรก
เป็นการให้ผู้เชี่ยวชาญระบุความสามารถขั้นต่ำสุดที่ยอมรับการประเมินให้ผ่าน แล้วจะถามรายชื่อ
นักเรียนผู้ที่มีความสามารถอยู่ใกล้เส้นแบ่งระหว่างกลุ่มที่มีความสามารถเพียงพอที่จะยอมรับ และ
กลุ่มที่มีความสามารถไม่เพียงพอที่จะยอมรับ ซึ่งนักเรียนเหล่านี้ไม่สามารถจัดเข้าอยู่ในกลุ่มใด
กลุ่มหนึ่งได้ จะใช้การทดสอบมาพิจารณาจัดกลุ่มนักเรียนพวกนี้ คะแนนที่เป็นค่ามัธยฐานที่ได้จาก
กลุ่มนี้จะเป็นคะแนนเกณฑ์ ที่ใช้สำหรับพิจารณาตัดสินในการให้ผ่านสำหรับนักเรียนที่มีความ
สามารถต่ำสุด

พวกที่อยู่เหนือเส้นแบ่งและอยู่ใต้เส้นแบ่ง (Contrasting Group) วิธีการนี้ใช้
ผู้เชี่ยวชาญเช่นกันโดยให้ระบุความสามารถขั้นต่ำสุดที่จะยอมรับให้ผ่าน ซึ่งจะจำแนกนักเรียน
เหล่านั้นออกจากกัน และจำแนกออกเป็นผู้รอบรู้ (Master) หรือผู้ไม่รอบรู้ (Nonmaster)
ซึ่งได้จากการทดสอบ ค่าสถิติที่ได้จากการคำนวณจะมีค่าคงที่และเหมาะสม ถ้ากลุ่มตัวอย่างไม่เล็ก

จนเกินไปนำคะแนนของนักเรียนทั้งสองกลุ่มมาพิจารณาการกระจายหาส่วนที่คาบเกี่ยวกัน จุดคาบเกี่ยวกันหรือจุดที่กราฟตัดกันจะเป็นคะแนนเริ่มต้นที่จะให้เป็นเกณฑ์สำหรับใช้พิจารณา และคะแนนเกณฑ์นี้สามารถที่จะปรับให้สูงขึ้น หรือต่ำลงเพื่อที่จะลดความคลาดเคลื่อนที่เกิดจากการจำแนกผิดทั้งสองประเภท (Type I and Type II) นั่นคือ พวกที่รอบรู้ไม่จริง เป็นนักเรียนที่รอบรู้จากการทดสอบ แต่มีความรอบรู้ไม่เท่ากับจุดประสงค์ พวกที่ไม่รอบรู้ไม่จริงเป็นนักเรียนที่ไม่รอบรู้จากแบบทดสอบแต่รอบรู้ในจุดประสงค์

ในการศึกษาค้นคว้าครั้งนี้ ผู้วิจัยมีความสนใจที่จะศึกษาคะแนนจุดตัดของแบบทดสอบอิงเกณฑ์ที่มีการจัดเรียงข้อสอบแตกต่างกันโดยใช้วิธีการเชิงประจักษ์ (Empirical Models) อันได้แก่ วิธีของแกลส วิธีของเบส โดยพิจารณาว่าคะแนนจุดตัดที่ได้นั้นตรงกันหรือไม่ และคะแนนจุดตัดที่ได้จากวิธีการทั้งสองวิธีทำให้ค่าความเชื่อมั่นของแบบทดสอบต่างกันหรือไม่

5. ความเชื่อมั่นของแบบทดสอบอิงเกณฑ์

ความเชื่อมั่นในการตัดสินจำแนกความรอบรู้ เป็นการศึกษาสัมประสิทธิ์ความสอดคล้องในการตัดสินความรอบรู้-ไม่รอบรู้ ซึ่งมีอยู่ด้วยกัน 5 วิธี (Berk. 1980 : 129 - 185) คือ

1. วิธีของคาร์เวอร์ (Carver)
2. วิธีของสวามินาธาน แฮมเบิลตัน และอัลไจนา (Swaminathan. Hambleton and Algina)
3. วิธีของฮวีน (Huynh)
4. วิธีของสับโคเวียค (Subkoviak)
5. วิธีของมาร์แชลล์และเฮอ์เทล (Marshall and Haertel)

1. วิธีของคาร์เวอร์ (บุญเชิด ภิญญโญนนตพงษ์. 2527 : 165 ; อ้างอิงมาจาก Crehan. 1974 : 256) เป็นสูตรการประมาณค่าความเชื่อมั่นโดยอาศัยการเปรียบเทียบสัดส่วนของความสอดคล้องในการตัดสินความรอบรู้ผู้สอบแต่ละคน จากแบบทดสอบคู่ขนานจากสูตรดังนี้

$$P_o = (a + c) / (a + b + c + d)$$

P_o แทนสัดส่วนของความสอดคล้องในการตัดสินความรอบรู้จากการสอบแบบทดสอบ
คู่ขนาน 2 ฟอรมจะได้

แบบฟอร์ม ข. แบบฟอร์ม ก.	ไม่ผ่าน	ผ่าน
	b	a
ผ่าน		
ไม่ผ่าน	c	d

ซึ่งวิธีของคาร์เวอร์จะได้ผลเท่ากับ วิธีของแฮมเบลตันและโนวิก เพราะลักษณะสัดส่วนนั้น
เป็นค่าสัดส่วนเดียวกัน และมีข้อบกพร่องเหมือนกันคือจำเป็นต้องสอบ 2 ครั้ง ด้วยแบบทดสอบ
คู่ขนาน ซึ่งผลจากการสอบซ้ำทำให้การประมาณค่าคลาดเคลื่อนจากความเป็นจริง

2. วิธีของสวามินาธาน แฮมเบลตัน และอัลจินา (Swaminathan, Hambleton and Algina. 1974) โดยอาศัยแนวคิดของแฮมเบลตันและโนวิก แต่ค่า P_o
ที่คำนวณได้นั้น อาจเกิดจากความสอดคล้องโดยบังเอิญ จึงมีการปรับแก้โดยใช้สัมประสิทธิ์แคปปา
(K) ของโคเฮน (Cohen. 1960.) ซึ่งได้หักค่าความสอดคล้องโดยบังเอิญออกไป ซึ่งมีสูตรดังนี้

$$K = \frac{P_o - P_c}{1 - P_c}$$

K แทนสัมประสิทธิ์ความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ ซึ่งได้หักความ
สอดคล้องโดยบังเอิญแล้ว

P_o แทนสัดส่วนของความสอดคล้องที่ตัดสินว่า รอบรู้จากการสอบซ้ำหรือแบบทดสอบ
คู่ขนาน คำนวณได้จาก

$$P_o = \sum_{k=1}^2 P_{kk}$$

P_o แทนสัดส่วนของความสอดคล้องที่คาดหวังโดยบังเอิญคำนวณได้จาก

$$P_o = \sum_{k=1}^2 P_{kk}$$

P_{kk} แทนสัดส่วนของผู้สอบในการสอบ โดยแบ่งเป็นรอบรู้-ไม่รอบรู้ในการสอบ 2 ครั้ง

$P_k P_k$ แทนสัดส่วนของผู้สอบในการแยกสอบแต่ละครั้ง

การคำนวณวิธีนี้อาจใช้แบบทดสอบฉบับเดียวสอบสองครั้ง หรือใช้แบบทดสอบคู่ขนานสองฉบับไปสอบกับนักเรียนกลุ่มเดียวกัน แล้วนำข้อมูลที่ได้แจกแจงเป็นผู้รอบรู้และไม่รอบรู้ลงในตาราง และคำนวณหาสัดส่วนของจำนวนคนที่แจกแจงในแต่ละกลุ่มดังนี้

สอบครั้งที่ 1 \ สอบครั้งที่ 2	ผ่าน	ไม่ผ่าน	รวม
	ผ่าน	ไม่ผ่าน	รวม
ผ่าน	P_{11}	P_{12}	P_1
ไม่ผ่าน	P_{21}	P_{22}	P_2
รวม	P_1	P_2	

วิธีของสวามินาธาน แฮมเบลตัน และอัลจินา มีข้อบกพร่องคือ จะให้ค่าประมาณของ K ที่อคติ ขนาดตามความยาวของแบบทดสอบและความแปรปรวนของคะแนนมีผลไวต่อค่าของ K จนบางครั้งต้องจำกัดพิสัยของค่าทั้งสองไว้ นอกจากนี้ยังจำเป็นต้องสอบ 2 ครั้ง ซึ่งผลของการสอบซ้ำ ทำให้การประมาณค่าคลาดเคลื่อนไปจากความจริง และการแปลความหมายของดัชนี มีความยุ่งยาก (บุญเชิด ภิญญอนันตพงษ์. 2527 : 170)

3. วิธีของฮวิน (Huynh. 1976 : 253 - 263) เป็นการหาความเชื่อมั่นในการตัดสินใจจำแนกความรอบรู้จากการสอบ 2 ครั้ง หรือครั้งเดียวก็ได้โดยการใช้สัมประสิทธิ์แคปปา (K) และตั้งข้อสมมติว่า กระบวนการสุ่มความน่าจะเป็นในการเลือกข้อคำถามจากข้อคำถามทั้งหมดที่มีอยู่ ได้นิยามไว้อย่างชัดเจน ถ้าแบบทดสอบประกอบด้วยข้อคำถาม n ข้อ ผู้สอบที่มีความสามารถ θ คะแนนสอบเท่ากับ X และคะแนนเหล่านี้มีการกระจายแบบ Binomial Density ดังนี้

$$f(x/\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}; x = 0, 1, 2, 3, \dots, n$$

จากแบบจำลองไบนอมียลนี้ สมมติว่า ข้อคำถามทุกข้อในเอกภพของข้อสอบสามารถใช้แทนกันได้ นั้นแสดงว่า การแจกแจงของคะแนนสอบของข้อคำถาม k ข้อ ไม่ขึ้นอยู่กับข้อคำถามใดที่เลือกขึ้นมาโดยเฉพาะ ซึ่งชี้ให้เห็นว่าข้อคำถาม มีความยากเหมือน ๆ กัน

จาก Beta Density นี้จะใช้พารามิเตอร์ α และ β เป็นความสามารถ แท้จริง θ ถ้าขนาดของกลุ่มตัวอย่างมาก จะสามารถหา α, β ได้จากการหาความเชื่อมั่นโดยใช้ KR-21 ดังนี้

$$\alpha_{21} = \binom{n}{n-1} \left[1 - \frac{\mu(n-\mu)}{n\sigma^2} \right]$$

ดังนั้น

$$\alpha = \left[-1 + \left(\frac{1}{\alpha_{21}} \right) \mu \right]$$

และ

$$\beta = -\alpha + \frac{n}{\alpha_{21}} - n$$

จาก Beta Density ข้างต้น จะได้ Marginal Density ของคะแนน X อยู่ในรูปของ Negative Hypergeometric (หรือ Beta-Binomial) ดังนี้

$$f(X) = \binom{n}{X} B(\alpha + X; n + \beta - X) / B(\alpha, \beta)$$

เมื่อ β คือ Beta Function สำหรับ $X \rightarrow 0$, $f(X)$ สามารถหาได้ดังนี้

$$f(0) = \prod_{i=1}^n \frac{n + \beta - i}{n + \alpha - \beta - i}$$

$$\text{และ } f(x+1) = f(x) \frac{(n-x)(\alpha+x)}{(x+1)(n+\beta-x-1)} ; x = 0, 1, 2, \dots, n-1$$

ถ้า $X \rightarrow n$ จะได้ดังนี้

$$f(n) = \prod_{i=1}^n \frac{n+\alpha-i}{n+\alpha+\beta-i}$$

และ

$$f(x-1) = f(x) \frac{x(n+\beta-x)}{(n-x+1)(\alpha+x-1)} ; x = 1, 2, 3, \dots, n$$

ถ้า X และ Y เป็นคะแนนที่ได้จากการสอบฟอร์ม X และ Y ภายใต้เงื่อนไขของความ
เป็นอิสระในการตอบสนองของข้อสอบ (Local Independence) แล้ว คะแนน X และ Y จะมี
การแจกแจงแบบ Bivariate Beta-Binomial (หรือ Negative Hypergeometric)
ด้วย Joint Density ดังนี้

$$f(x,y) = \frac{\binom{n}{x} \binom{n}{y} B(\alpha+x+y, 2n+\beta-x-y)}{\beta(\alpha, \beta)}$$

สมการนี้จะสมมาตร ถ้า $f(X,Y) = f(Y,X)$ สำหรับค่า X และ Y มีค่าใกล้ 0
ดังนั้น $f(X,Y)$ สามารถหาได้ดังนี้

$$f(0,0) = \prod_{i=1}^n \frac{2n+\beta-i}{2n+\alpha+\beta-i} \quad f(0) = \prod_{i=1}^n \frac{2n+\beta-i}{2n+\alpha+\beta-i}$$

และ

$$f(x+1,y) = f(x,y) \frac{(n-x)(\alpha+x+y)}{(x+1)(2n+\beta-x-y-1)}$$

ถ้า X และ Y มีค่าใกล้ n สูตรในการหาจะเป็นดังนี้

$$f(n, n) = \prod_{i=1}^{2n} \frac{2n+\alpha-i}{2n+\alpha+\beta-i} \quad f(n) = \prod_{i=1}^n \frac{2n+\alpha-i}{2n+\alpha+\beta-i}$$

และ

$$f(x-1, y) = f(x, y) \frac{X(2n+\beta-x-y)}{(n-X+1)(\alpha+X+y-1)}$$

การประเมินค่า K โดยการใช้แบบทดสอบคู่ขนาน 2 ฟอรัม

ถ้าการสอบที่ใช้แบบทดสอบคู่ขนาน 2 ฟอรัม คือฟอรัม X และ Y โดยมีจุดตัด C ถ้าผู้สอบที่ได้คะแนนสอบมากกว่าคะแนนจุดตัดจะถือว่าเป็นผู้รอบรู้ และผู้สอบที่ได้คะแนนต่ำกว่าคะแนนจุดตัดจะถือว่าเป็นผู้ไม่รอบรู้ ซึ่งสามารถเขียนได้ดังนี้

ฟอรัม y ฟอรัม X	ไม่รอบรู้	รอบรู้	ค่าสัดส่วน
ไม่รอบรู้	P_{00}		P_0
รอบรู้		P_{11}	P_1
ค่าสัดส่วน	P_0	P_1	

ถ้าคะแนนจุดตัดมีค่าใกล้ n สูตรความเชื่อมั่นของแคปปาหาได้ดังนี้

$$K = (P_{11} - P_1^2) / (P_1 - P_1^2)$$

$$P_{11} = \sum_{x,y=c}^n f(x,y)$$

$$P_1 = \sum_{x=c}^n f(x)$$

ถ้าคะแนนจุดตัดมีค่าน้อยหรือใกล้ 0 สูตรในการหาสัมประสิทธิ์ความเชื่อมั่นแคปปาหาได้ดังนี้

$$K = (P_{00} - P_0^2) / (P_0 - P_0^2)$$

เมื่อ

$$P_{00} = \sum_{x,y=0}^{c-1} f(x,y)$$

และ

$$P_0 = \sum_{x=0}^{c-1} f(x)$$

การประมาณค่า K จากการสอบเพียงครั้งเดียว

ในการประมาณค่า K จากการสอบเพียงครั้งเดียวมีขั้นตอนดังนี้

- 1) คำนวณค่าเฉลี่ย ($\bar{X}$) และส่วนเบี่ยงเบนมาตรฐาน (S) ของคะแนนที่ได้จากการสอบ
- 2) ประมาณค่า α และ β โดยแทนค่า μ ด้วย $\bar{X}$ และ σ ด้วย S

$$\alpha_{21} = \frac{n}{n-1} \left[1 - \frac{\bar{X}(n - \bar{X})}{ns^2} \right]$$

$$\alpha = (-1 + \frac{1}{\alpha_{21}}) \bar{X}$$

$$\beta = -\alpha + \frac{n}{\alpha_{21}} - n$$

โดยที่ n คือจำนวนข้อคำถามในแบบทดสอบ

3) การประมาณค่า K มี 2 กรณี คือ

3.1) ถ้าคะแนนจุดตัด (c) มีค่าน้อยหรือใกล้ 0 จำนวน P_0 และ P_{00} ดังนี้

$$P_0 = f(0) + \dots + f(c-1)$$

$$P_{00} = f(0,0) + f(0,1) + \dots + f(c-1,c-1)$$

ดังนั้น

$$K = (P_{00} - P_0^2) / (P_0 - P_0^2)$$

3.2) ถ้าคะแนนจุดตัด (c) มีค่าบวก จำนวน P_1 และ P_{11} ดังนี้

$$P_1 = f(n) + \dots + f(c)$$

$$P_{11} = f(n,n) + f(n,n-1) + f(n-1,n) + \dots + f(c,c)$$

ดังนั้น

$$K = (P_{11} - P_1^2) / (P_1 - P_1^2)$$

การประมาณค่า K เมื่อแบบทดสอบมีจำนวนข้อคำถามมาก

ถ้าแบบทดสอบประกอบด้วยจำนวนข้อคำถามมาก เช่น มีจำนวนมากกว่า 10 ข้อ การประมาณค่า K ดังที่กล่าวไว้ข้างต้นค่อนข้างจะยุ่งยากมาก ดังนั้น ฮวินจึงเสนอวิธีการตามที่ คีทส์และลอร์ด (Huynh. 1976 : 258 ; citing Keats and Lord. 1962) เสนอไว้ดังนี้ แปลงคะแนนที่ได้จากการสอบ จากสูตร

$$X' = \sin^{-1} \sqrt{\frac{X}{n}}$$

ซึ่งมีการกระจายเป็นปกติโดยมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานดังนี้

$$\mu_{X'} = \sin^{-1} \sqrt{\frac{\mu}{n}}$$

$$\alpha_{X'} = \left[\frac{\alpha_{21} + 1}{\alpha + n} \right]^{1/2}$$

ถ้า X และ Y เป็นคะแนนที่ได้จากการแปลงจากคะแนน X และ Y จากแบบทดสอบ คูณาน 2 ฟอรั่มแล้ว ค่าสหสัมพันธ์ของคะแนนทั้ง 2 จะมีค่าดังนี้

$$\rho = \alpha_{21} \left[\frac{n - 1}{n + \alpha_{21}} \right]^{1/2}$$

โนวิก เลวิส และแจ๊คสัน (Huynh. 1976 : 258 ; citing Novick, Lewis and Jackson. 1973) ได้ใช้วิธีการแปลงคะแนนแบบเดียวกันนี้กับตัวแปรทวินาม (Binomial Variable) และพบว่า การประมาณค่าจะสมเหตุสมผลดีเมื่อ μ/n มีค่าอยู่ระหว่าง .15 และ .85 และจำนวนข้อสอบอย่างน้อย 8 ข้อ

เนื่องจากฟังก์ชัน Arc sine มีลักษณะเป็นการเพิ่มไปในทางเดียวกัน การกำหนดผู้รอบรู้ ก็ต่อเมื่อ $X' > C'$ โดยที่ $C' = \sin^{-1}\sqrt{(c-5)/n}$ และใช้ตารางของกูปตา (Huynh. 1976 : 259 ; citing Gupta. 1963) ก็จะสามารถประมาณค่า P_0 และ P_{00} ได้

เพื่อให้เห็นภาพชัดเจนขึ้น ต่อไปนี้จะกล่าวโดยสรุปถึงการประมาณค่าแคปปา ดังนี้

- 1) คำนวณค่า $\bar{X}$ และ S ของคะแนนที่ได้จากการสอบ
- 2) ประมาณค่า $\mu X'$, $\sigma X'$ และ ρ โดยแทน μ ด้วย $\bar{X}$ และ σ ด้วย S ดังนี้

$$\alpha_{21} = \frac{n}{n-1} \left[1 - \frac{X(n-x)}{ns^2} \right]$$

$$\alpha = (-1 + \frac{1}{\alpha_{21}})X$$

$$\beta = -\alpha + \frac{n}{\alpha_{21}} - n$$

$$\mu X' = \sin^{-1} \sqrt{\frac{x}{n}}$$

$$\rho = \alpha_{21} \left[\frac{n-1}{n+\alpha_{21}} \right]^{1/2}$$

$$\sigma X' = \left[\frac{\alpha_{21} + 1}{\alpha + n} \right]^{1/2}$$

เมื่อ n คือ จำนวนข้อคำถามในแบบทดสอบ

$$3) \text{ คำนวณอัตราส่วน } Z = (C' - \mu X') / \sigma X'$$

4) หาค่าความน่าจะเป็นของตัวแปรปกติมาตรฐานที่มีค่าน้อยกว่า z เพื่อประมาณค่า P_o

5) หาค่าความน่าจะเป็นของตัวแปรปกติมาตรฐาน 2 ค่าที่มีค่าสหสัมพันธ์ ρ น้อยกว่า z เพื่อประมาณค่า P_{oo}

6) ประมาณค่า K จากสูตร

$$K = (P_{oo} - P_o^2) / (P_o - P_o^2)$$

ข้อบกพร่องของการหาค่าความเชื่อมั่นโดยวิธีของฮวิน คือ ให้ค่าประมาณที่อคติเมื่อแบบทดสอบมีขนาดสั้น ดัชนี K มีความไวต่อขนาดความยาวของแบบทดสอบและการกระจายของคะแนน และวิธีการคำนวณไข่มโนมติ และโครงสร้างทางคณิตศาสตร์ที่ซับซ้อนมากที่สุด (บุญเชิด บุญญอนันตพงษ์. 2527 : 185) และคุนแนน (Kunnen. 1992 : 32 - 33) ได้กล่าวว่า สัมประสิทธิ์แคปลา (K) จะมีความไวต่อขนาดความยาวของแบบทดสอบและการกระจายของคะแนน จะให้ค่า K สูง เมื่อกำหนดคะแนนจุดตัดหรือคะแนนเกณฑ์เข้าใกล้ค่าเฉลี่ย

สับโคเวียก (Subkoviak. 1978 : 111 - 116) ได้ศึกษาเปรียบเทียบค่าความเชื่อมั่น 4 วิธี คือ วิธีการของสวามินาธาน วิธีการของฮวิน วิธีการของสับโคเวียก และวิธีของมาร์แชล และแฮร์เทล โดยใช้เกณฑ์ตัดสินร้อยละ 50, 60 70 และ 80 พบว่าวิธีการของสวามินาธานและคนอื่น ๆ ให้ค่าประมาณที่ไม่ลำเอียง ส่วนค่าความคลาดเคลื่อนมาตรฐานถึงแม้จะเล็กแต่มีแนวโน้มว่าจะใหญ่แต่สามารถลดลงได้ โดยการเพิ่มขนาดกลุ่มตัวอย่าง วิธีการของฮวินให้ค่าความเชื่อมั่นต่ำกว่าวิธีอื่น ๆ มาก แต่ให้ค่าความคลาดเคลื่อนต่ำกว่าวิธีการของคนอื่น ๆ ส่วนวิธีการของสับโคเวียกให้ค่าความคลาดเคลื่อนค่อนข้างเล็ก แสดงถึงการกะประมาณที่ค่อนข้างเที่ยงตรง

4. วิธีของสับโคเวียก (Subkoviak. 1976 : 265 - 276) เป็นการหาความเชื่อมั่นในการสอบครั้งเดียว ด้วยแบบทดสอบคู่ขนานสองฉบับ ซึ่งมีคะแนนเป็น X และ X' โดยที่คะแนนทั้งสองเป็นอิสระต่อกัน และมีการแจกแจงเป็นแบบทวินามเหมือนกันตลอด (Identically Binomial) แต่ค่าสถิติที่นำมาคำนวณมาจากคะแนน X' นิยามของสัมประสิทธิ์ความสอดคล้อง สำหรับบุคคล i ว่าเป็นความน่าจะเป็น ที่บุคคลนั้นถูกกำหนดให้เป็นผู้รอบรู้เหมือนกัน อันเนื่องมาจากผลการสอบ แบบทดสอบคู่ขนาน เพื่อกำหนดคะแนนเกณฑ์ c โดยเขียนเป็นสมการสัมประสิทธิ์ความสอดคล้อง คือ

$$P_c = \frac{\sum_{i=1}^N P_c^{(i)}}{N}$$

เมื่อ P_c คือ ความน่าจะเป็นของการตัดสินใจ
 N คือ จำนวนผู้สอบทั้งหมด
 $P_c^{(i)}$ คือ สัมประสิทธิ์ของความสอดคล้อง ของคนที่ i เมื่อ
กำหนดจุดตัดเท่ากับ c

$$P_c^{(i)} = \frac{[P(x_i > C)]^2 + [1 - P(x_i > C)]^2}{n}$$

โดยที่

$$P(x_i > C) = \sum_{x_i=C}^n p_i x_i (1-p_i)^{n-x_i}$$

เมื่อ P_i คือ ความน่าจะเป็นที่แท้จริงในการตอบข้อคำถามถูกของคน ที่ i ค่าของ P_i สามารถหาได้ดังนี้

1) ถ้าแบบทดสอบมีจำนวนข้อสอบมากกว่า 40 ข้อ

$$\hat{P}_i = \frac{X_i}{n}$$

เมื่อ x_i คือ จำนวนข้อที่ตอบถูก
 n คือ จำนวนข้อสอบทั้งหมดในแบบทดสอบ

2) ถ้าแบบทดสอบมีจำนวนข้อสอบน้อย

$$\hat{P}_i = \alpha_{21}/X \left(\frac{X_i}{n} \right) + (1 - \alpha_{21}/x) \left(\frac{MX}{n} \right)$$

$$\text{โดยที่ } \alpha_{21}/X = \frac{n}{n-1} \left[1 - \frac{Mx(n-Mx)}{ns_x^2} \right]$$

เมื่อ α_{21}/x คือ สัมประสิทธิ์ความเชื่อมั่นของแบบทดสอบโดยวิธี

Kuder Richardson Formula 21

M_x คือ คะแนนเฉลี่ยที่ได้จากการสอบ

S_x^2 คือ ความแปรปรวนของคะแนนที่ได้จากการสอบ

สับโคเวียก พบว่า ถ้าข้อมูลจากแบบทดสอบคู่ขนาน 2 ชุด มีความสัมพันธ์กันทางบวก ค่าความเชื่อมั่นจะมีค่าต่ำเกินไป แต่ถ้ามีความสัมพันธ์กันทางลบ ค่าที่ได้จะมีค่าสูงเกินไป และการกระจายแบบทวินามของคะแนนจากข้อสอบที่ผลการตอบแต่ละข้อ เป็นอิสระต่อกันนั้น พบว่า เมื่อคะแนนเกณฑ์ C มีค่าต่ำสุดและสูงสุด ค่าความเชื่อมั่นจะมีค่าสูงสุด และเมื่อเกณฑ์อยู่กลาง ๆ ค่าความเชื่อมั่นจะมีค่าต่ำสุด

ข้อบกพร่องของวิธีนี้คือ (Hambleton and others. 1978 : 23) การประมาณค่าความเชื่อมั่น มีความเป็นไปได้ที่จะสูงกว่าความเป็นจริง เนื่องจากได้รวมเอาความสอดคล้องโดยบังเอิญเข้าไว้ (P_o ไม่มีการปรับแก้) นอกจากนี้ในการใช้แบบทดสอบคู่ขนาน เพื่อหาสัมประสิทธิ์ความสอดคล้อง ในการกำหนดความรู้ของนักเรียนแต่ละคน และของกลุ่มนั้นเป็นการปฏิบัติที่ยากที่จะให้แบบ ทดสอบทั้งสองมีข้อคำถาม (Item) ที่มีความยากง่ายเท่ากันทุกข้อ

5. วิธีของมาร์แชลล์และเซอร์เทล (Marshall and Haeftel)

มาร์แชลล์-เซอร์เทล ได้เสนอวิธีหาความเชื่อมั่นในการตัดสินจำแนกความรอบรู้จากการรอบรู้เพียงครั้งเดียว โดยมีข้อดกลงเบื้องต้นว่าถ้าผู้สอบแต่ละคนได้รับการทดสอบซ้ำ การแจกแจงของคะแนนจะมีลักษณะเป็นไบนอมิเยล วิธีการของมาร์แชลล์ - เซอร์เทล คือ แบ่งครึ่งแบบทดสอบออกเป็น 2 ฉบับ เพื่อหาสัดส่วนของความคงที่ในการแบ่งแบบทดสอบ ออกเป็น 2 ฉบับ (P_0) ดังนี้ (Berk. 1980 : 145 - 147)

1. กำหนดให้ X แทนคะแนนที่ได้จากการสอบ N_X แทนความถี่ของคะแนนที่ได้จากการสอบแบบทดสอบฉบับที่ 1 และ N_X แทนความถี่ของคะแนน (w) จากการสอบด้วยแบบทดสอบที่มีความยาวเป็น 2 เท่า ของแบบทดสอบฉบับที่ 1 ดังนั้น

$$N_w = \sum_{x=0}^n N_X \binom{2n}{w} \left(\frac{X}{n}\right)^w \left(1 - \frac{X}{n}\right)^{2n-w}$$

2. หากความน่าจะเป็นของความสอดคล้องในการจำแนกผู้สอบด้วย การใช้แบบทดสอบที่แบ่งครึ่งออกเป็น 2 ฉบับ ซึ่งมีคะแนนสอบเป็น w จากการสอบด้วยแบบทดสอบที่มีความยาวเป็น 2 เท่า ดังนี้

$$\phi_w(a,b) = \frac{\sum_{j=a}^b \binom{w}{j} \binom{2n-w}{n-j}}{\binom{2n}{n}}$$

$$\begin{aligned} \text{โดยที่} \quad a &= w - [c - 1] \\ b &= c - 1 \end{aligned}$$

เมื่อ c คือ คะแนนจุดตัด

3. คำนวณค่าความเชื่อมั่นในการตัดสินความรอบรู้จากการสอบ ดังนี้

$$P_o = \left[\frac{1}{n} \sum_{w=0}^{c-1} n_w + \sum_{w=c}^{2c-2} n_w \cdot \phi_w (w-[c-1]); c-1) + \right. \\ \left. \sum_{w=2c}^{n+c-1} n_w \cdot \phi_w (c, w-c) + \sum_{w=n+c}^{2n} n_w \right]$$

เมื่อ N คือ จำนวนผู้สอบ

n คือ จำนวนข้อคำถาม

c คือ คะแนนจุดตัด

N_w คือ ความถี่ของคะแนน W

(a,b) คือ ความน่าจะเป็นของความสอดคล้องในการจำแนก

การประมาณค่าความเชื่อมั่นตามวิธีของโลเวท

โลเวท (อังกฤษ สายยศ. 2536 : 45 ; อ้างอิงมาจาก lovet. 1978) ได้เสนอสูตรการคำนวณความเชื่อมั่น โดยอาศัยการแจกแจงแบบทวินาม จากสูตรไบโนเมียลของโลเวท ดังนี้

$$r_{cc} = 1 - \frac{K \sum X_i - \sum X_i^2}{\left\{ (K-1) \sum (X_i - C)^2 \right\}}$$

เมื่อ	r_{cc}	แทน ความเชื่อมั่นของแบบทดสอบอิงเกณฑ์
	K	แทน จำนวนข้อสอบ
	$\sum X_i$	แทน ผลรวมของคะแนนทั้งหมด
	$\sum X_i^2$	แทน ผลรวมทั้งหมดของคะแนนแต่ละตัวยกกำลังสอง
	X_i	แทน คะแนนของนักเรียนแต่ละคน
	C	แทน คะแนนจุดตัดของแบบทดสอบอิงเกณฑ์

การประมาณค่าความเชื่อมั่นตามวิธีของลิวิงตัน

ลิวิงตัน (Livington. 1972 : 13 - 16) ได้เสนอสูตรการหาความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ โดยอาศัยการนิยามความแปรปรวนของคะแนนจริง และคะแนนที่สอบได้จากการหาค่าคาดหวังของค่าเบี่ยงเบนกำลังสองของคะแนนแต่ละค่าที่เบี่ยงเบนออกจากคะแนนจุดตัดแล้ว นิยามความเชื่อมั่นว่าเป็นอัตราส่วนของค่าคาดหวังของค่าเบี่ยงเบนกำลังสองโดยเฉลี่ยของคะแนนจริงและคะแนนสอบจากคะแนนจุดตัดดังนี้

$$K^2(X, T) = \frac{\sigma^2_X (KR - 20) + (\mu_X - n_i C)^2}{\sigma^2_X + (\mu - n_i C)^2}$$

เมื่อ	$K^2(X, T)$	แทน ความเชื่อมั่นตามวิธีของลิวิงตัน
เมื่อ	σ^2_X, μ_X	แทน ความแปรปรวนและค่าเฉลี่ยของคะแนนที่สอบได้
เมื่อ	KR-20	แทน แทนความเชื่อมั่นแบบอิงกลุ่มที่หาได้จากสูตร KR-20

จากความเชื่อมั่นในการตัดสินจำแนกความรอบรู้ของแบบทดสอบอิงเกณฑ์ ที่กล่าวมาข้างต้น การคำนวณโดยใช้สูตรของโลเวท และสูตรของลิวิงสตัน เป็นสูตรการหาความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ จากการสอบเพียงครั้งเดียว อีกทั้งค่าความเชื่อมั่นจะขึ้นอยู่กับคะแนนจุดตัดและความแปรปรวนของคะแนน ดังนั้นผู้วิจัยมีความสนใจ ที่จะศึกษาความเชื่อมั่นโดยใช้สูตรการคำนวณดังกล่าวเพื่อศึกษาเปรียบเทียบความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ที่มีการเรียงข้อสอบแตกต่างกัน

6. การเรียงข้อคำถามในแบบทดสอบ

การเรียงข้อคำถามในแบบทดสอบได้มีนักวิชาการ และนักวัดผลได้เสนอแนวคิดไว้ดังนี้

ชาวล แพรตกุล (2516 : 193) กล่าวว่า ข้อสอบแต่ละฉบับควรเรียงคำถามเริ่มจากง่ายไปหายากเสมอ ไม่ควรเรียงตามความยากง่ายของบทเรียน

อนันต์ ศรีโสภา (2525 : 141) กล่าวว่า การจัดเรียงข้อสอบที่ดี ควรจัดเรียงจากข้อง่ายไปหาข้อยาก

ลินควิสต์ (Lindquist. 1955 : 179) กล่าวว่า ในการเรียงข้อคำถาม ไม่จำเป็นต้องจัดเรียงลำดับคำถามตามความยากง่าย ถ้าคำถามเหล่านั้นเป็นเนื้อหาเดียวกัน และมีความยากง่ายพอ ๆ กัน แต่ถ้าคำถามเหล่านั้นมีเนื้อหาต่างกันด้วย แยกข้อคำถามออกเป็นฉบับๆ และแยกเวลาในการสอบของแต่ละฉบับ

สแตนเลย์ และฮอฟกินส์ (Stanley and Hopkins. 1972 : 191) กล่าวว่า การเรียงลำดับข้อคำถามควรจะเริ่มต้นจากข้อที่ง่ายที่สุด และให้ข้อยากที่สุดไว้ตอนท้ายของแบบทดสอบ จะเป็นไปตามหลักจิตวิทยา เพราะนักเรียนพบข้อยาก ๆ ในตอนต้นของแบบทดสอบ อาจทำให้เกิดความท้อใจได้ โดยเฉพาะกลุ่มนักเรียนที่มีความสามารถทางการเรียนระดับปานกลางและต่ำ

ทราเวอร์ (Travers. 1950 : 127 - 128) ได้เสนอแนะการจัดเรียงข้อคำถามไว้ 4 วิธี ดังนี้

1. เรียงตามลำดับความยาก (Arrangement in order Difficulty) ทำได้โดยเรียงลำดับข้อคำถามตามความยากของข้อคำถาม ซึ่งพิจารณาจากค่าสถิติในการเรียงข้อคำถามแบบนี้ ทำให้ผู้สอบมีกำลังใจ เมื่อพบปัญหาที่ยากขึ้น และมีข้อเสียคือ เนื้อหาและพฤติกรรมที่ต้องการวัดไม่อยู่รวมเป็นพวกเดียวกัน ทำให้ผู้สอบต้องเปลี่ยนความคิดอย่างรวดเร็ว และผู้สอบอาจหมดกำลังใจที่จะทำแบบทดสอบ เพราะเมื่อเริ่มพบข้อยากแล้ว ข้อต่อไปของแบบทดสอบน่าจะยากยิ่งขึ้นไปอีก

2. เรียงลำดับความยากแบบหมุนเวียน (Arrangement in cyclic order of Difficulty) การเรียงข้อคำถามในลักษณะนี้เพื่อหลีกเลี่ยงข้อเสียของการเรียงลำดับข้อคำถามตามระดับความยาก และเพื่อกระตุ้นให้ผู้สอบอ่านข้อสอบทุกข้อ เพราะผู้สอบรู้ว่า ถ้าอ่านต่อไปอีกระยะหนึ่ง จะพบข้อง่ายอีก การจัดเรียงแบบนี้จึงมีประโยชน์ แต่ผู้สอบจะต้องเปลี่ยนความคิดในแต่ละเรื่องอย่างรวดเร็ว

3. เรียงลำดับตามกลุ่มเนื้อหาวิชา (Arrangement According to Subject Matter Area) ข้อดีของวิธีนี้คือ ผู้สอบจะคิดปัญหาในเนื้อหาเดียวกันก่อนที่จะเปลี่ยนไปคิดเนื้อหาอื่น ๆ ต่อไป

4. เรียงตามจุดมุ่งหมายของการวัด (Arrangement According to the Goals Measured) แบบทดสอบบางชุด ผู้จัดทำข้อสอบจะจัดเรียงเป็นกลุ่มตามพฤติกรรมการวัดประเภทเดียวกัน เช่น แบบทดสอบโค-อเพอเรทีฟ (Co-operative Tests) จะจัดข้อคำถามที่วัดความจำเกี่ยวกับคำศัพท์ และความคิดรวบยอด (Terms and Concept) ไว้ด้วยกัน ส่วนการวัดความเข้าใจแยกไว้อีกพวกหนึ่ง ข้อดีของการจัดเรียงด้วยวิธีนี้ คือ ครูสามารถตรวจสอบว่าวัดตรงตามจุดมุ่งหมายที่ต้องการวัดหรือไม่

จากการศึกษารูปแบบการจัดเรียงข้อคำถามในแบบทดสอบ สรุปได้ว่าการเรียงแบบทดสอบอาจเรียงตามลำดับความยากง่าย เรียงตามเนื้อหาและเรียงแบบสุ่ม ดังนั้น ผู้วิจัยจึงสนใจนำวิธีการเรียงข้อคำถามในรูปแบบดังกล่าวมาไว้ในแบบทดสอบอิงเกณฑ์ เพื่อศึกษาเปรียบเทียบความเชื่อมั่นในการจำแนกความรู้

7. งานวิจัยที่เกี่ยวข้องกับการศึกษาค้นคว้า

งานวิจัยที่ศึกษาค้นคว้าเกี่ยวกับแบบทดสอบอิงเกณฑ์ มีดังนี้

งานวิจัยต่างประเทศ

เครเฮน (Crehen. 1974 : 255 - 262) ได้ศึกษาแบบทดสอบอิงเกณฑ์ ใช้วิธีการคัดเลือกข้อสอบ 6 วิธี คือ วิธีการของคอกซ์-วากัส (Cox-vargas) วิธีของเบรนัน (Brennan) วิธีการเรียงตามค่าสัมประสิทธิ์สหสัมพันธ์แบบ Point Bigerial และวิธีเรียงตามสุ่มกลุ่มตัวอย่าง แต่ละกลุ่มใช้ระหว่าง 7 - 30 คน โดยใช้แบบทดสอบคู่ขนานพบว่า วิธีการเลือกข้อสอบแบบอิงเกณฑ์ ทั้ง 6 วิธีนี้ ให้ผลต่อความเชื่อมั่นไม่แตกต่างกัน ค่าความเที่ยงตรงได้เลือกเกณฑ์ที่ให้ค่าสูงสุด โดยใช้จุดเดียวกันทั้งสองฉบับ ปรากฏว่า วิธีการของคอกซ์-วากัส และเบรนัน ให้ค่าสูงสุด

ครอส, อิมพารา, ฟรารี และเจเกอร์ (Cross, Impara, Frary and Jaeger. 1984) ได้ศึกษาวิธีการกำหนดเกณฑ์มาตรฐานของแองกอฟฟ์ นีเดลสกี และเจเกอร์ เพื่อกำหนด

เกณฑ์มาตรฐานของแบบทดสอบ (The National Teacher Examination "ของ" The Virginia Department of Education) โดยมีขั้นตอนในการดำเนินงานได้นำเอาแนวคิดของการอาศัยข้อมูลจากภายนอกเข้ามาเป็นตัวร่วมพิจารณา คือ

1. ในรอบแรกผู้ตัดสินจะได้รับข้อมูลเฉพาะแบบทดสอบข้อใดมาพิจารณา โดยไม่มีข้อมูลอื่นใดประกอบ
2. ในรอบที่สอง ผู้ตัดสินจะได้รับแบบทดสอบเฉพาะข้อใดมาพิจารณาแต่ละข้อมีข้อมูลประกอบ คือ ความยากของข้อสอบแต่ละข้อ
3. ในรอบสุดท้าย พิจารณาเฉพาะข้อใด และมีข้อมูลประกอบคือ ข้อมูลที่ได้จากการตัดสินของผู้ตัดสินคนอื่น ๆ ในรอบแรก และข้อมูลความถี่ของคะแนนที่ผู้สอบทำได้จากการนำแบบทดสอบไปใช้

ผลการศึกษาพบว่า เมื่อมีการปรับคะแนนที่ได้จากผู้ตัดสินให้เป็นคะแนนจากแบบทดสอบเต็มฉบับ ผลการวิเคราะห์ความแปรปรวน ได้ข้อสรุปว่า วิธีการกำหนดเกณฑ์มาตรฐานต่างกันและผลการตัดสินใจในแต่ละรอบ ซึ่งมีข้อมูลต่างกันมีค่าเฉลี่ยเลขคณิต ซึ่งถือเป็นเกณฑ์มาตรฐานแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001

โครสและคนอื่น ๆ (Cross and others. 1984 : 113 - 129) ได้เปรียบเทียบวิธีการกำหนดเกณฑ์มาตรฐานในการผ่านขั้นต่ำ เพื่อใช้สำหรับการสอบครูในระดับชาติ NTE (National Teacher Examination) ในวิชาคณิตศาสตร์ และวิชาการศึกษาเบื้องต้น โดยให้คณะบุคคล นักการศึกษา ทางวิชาชีพครูเป็นผู้พิจารณา ซึ่งพวกนี้จะได้รับการสอนเกี่ยวกับการหาเกณฑ์วิธีใด วิธีหนึ่งจากสามวิธี คือ วิธีของแองกอฟฟ์ วิธีของนิเดิลสกี และวิธีของเจเกอร์ การวิเคราะห์ข้อมูลใช้การทดสอบความแตกต่างระหว่างค่าเฉลี่ย โดยการวิเคราะห์ความแปรปรวนแบบวัดซ้ำ พบว่า วิธีการให้เกณฑ์และครั้งที่ของการพิจารณามีผลร่วมกัน แต่ไม่มีผลร่วมกันกับเนื้อหาวิชา และการพิจารณาให้เกณฑ์ครั้งที่ 1 จะแตกต่างจากการพิจารณาให้เกณฑ์ในครั้งที่ 2 และครั้งที่ 3 นอกจากนี้ยังพบว่าการให้เกณฑ์มาตรฐานวิธีของนิเดิลสกี ที่ใช้สูตรการปรับแก้จะมีความคงที่ต่ำกว่าวิธีของแองกอฟฟ์และเจเกอร์

แฮมเบิลตัน, มิลล์ และไซมอน (Hambleton, Mills and Simon. 1983) ได้ศึกษาข้อมูลที่สร้างขึ้นจากคอมพิวเตอร์ และใช้ทฤษฎีการตอบสนองหาคำตอบพบว่า แบบทดสอบที่มีความยาวมากขึ้น จะมีค่าความเชื่อมั่นเพิ่มขึ้น แบบทดสอบที่มีความยากใกล้เคียงกัน จะมีค่าความเชื่อมั่นเพิ่มขึ้น ณ ที่ระดับความเท่ากัน แบบทดสอบที่มีค่าอำนาจจำแนกสูง จะมีความเชื่อมั่นสูงกว่าแบบทดสอบที่เป็นคู่ขนานแบบสุ่มที่มีการกระจายของความยากเท่ากัน และมีความยาวเท่ากัน

ฮาลาดินา และรอยด์ (Haladyna and Roid. 1983) ได้ศึกษาเปรียบเทียบผลของความคลาดเคลื่อน ในการประมาณค่าความสามารถของบุคคลจากการใช้แบบทดสอบต่างกัน 2 แบบคือ แบบแรกเป็นแบบทดสอบที่สุ่มข้อสอบขึ้นมาจากประชากรข้อสอบที่มีอยู่ทั้งหมดกับแบบ

หลังเป็นการเลือกข้อสอบให้เหมาะกับระดับความสามารถของผู้สอบ โดยศึกษาจากรายวิชาต่าง ๆ 4 วิชา ซึ่งประกอบด้วยข้อสอบที่ผ่านการวิเคราะห์ด้วยวิธี Rasch Model มาแล้ว ข้อสอบเหล่านี้เป็นประชากรข้อสอบทั้งหมดในโดเมน จากนั้นจึงได้ทดลองสุ่มสร้างแบบทดสอบให้มีจำนวนข้อเป็น 10, 20, 30 และ 40 ข้อตามลำดับ แล้วศึกษาเปรียบเทียบอัตราความแตกต่างเฉลี่ยระหว่างคะแนนที่สอบได้ กับคะแนนโดเมนที่ประมาณได้ (The Average Absolute Difference between observed and Domain Scores) ย่อว่า AAD จากแบบทดสอบที่มีความยาวระดับต่าง ๆ ที่กำหนด หาดด้วยส่วนเบี่ยงเบนมาตรฐาน (เรียกค่าที่ได้ว่า อัตราส่วน ADD/SD) โดยในการสร้างแบบทดสอบย่อย จะมีวิธีการสร้าง 4 วิธี คือ

1. สุ่มข้อสอบมาจากโดเมน
2. เลือกข้อสอบให้มีความยากตรงกับระดับความสามารถของผู้สอบ
3. เลือกข้อสอบให้มีความยากกระจาย มากกว่าระดับความสามารถของผู้สอบเล็กน้อย
4. เลือกข้อสอบให้มีความยากกระจายมากกว่าระดับความสามารถของผู้สอบ

จากการศึกษาพบว่า การสอบที่จัดให้ข้อสอบมีความยากตรงกับระดับความสามารถของผู้สอบ จะทำให้ค่าอัตราส่วน ADD/SD ต่ำสุด แบบทดสอบที่ได้รับการสุ่มขึ้นมา มีค่าดัชนีดังกล่าวต่ำเป็นอันดับที่สอง และแบบทดสอบที่มีความยากกระจายมาก จะยังมีค่าดัชนีดังกล่าวสูงมากยิ่งขึ้น

สับโคเวียค (Subkoviak. 1978 : 111 - 116) ได้ศึกษาเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ ซึ่งเป็นแบบทดสอบคู่ขนานวัดความสามารถเชิงวิชาการ (Scholastic Aptitude Test) หาความเชื่อมั่น 4 วิธี คือ วิธีของสวามินาธาน วิธีของฮวิน วิธีของสับโคเวียค และวิธีของมาร์แชล และแฮร์เทิล โดยทำการศึกษาในกลุ่มตัวอย่าง 50 กลุ่ม กลุ่มละ 30 คน ใช้เกณฑ์ในการตัดสินเป็น 50, 60, 70 และ 80 เปอร์เซนต์ ผลปรากฏว่า วิธีของสวามินาธานให้ค่าประมาณที่ไม่ลำเอียง วิธีของฮวินให้ค่าความเชื่อมั่น และความคลาดเคลื่อนต่ำกว่าวิธีอื่น ๆ ส่วนวิธีของสับโคเวียคให้ค่าความคลาดเคลื่อนเล็กน้อย

งานวิจัยภายในประเทศ

อัจฉริยา ปราบอริพ่าย (2532 : 101 - 106) ได้ศึกษาเปรียบเทียบคะแนนจุดตัดของแบบทดสอบผลสัมฤทธิ์วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1 และค่าดัชนีความคงที่ของการตัดสินคะแนนจุดตัด ซึ่งตัดสินโดยผู้เชี่ยวชาญที่ไม่มีสารสนเทศ และผู้เชี่ยวชาญที่มีสารสนเทศ โดยใช้วิธีของแองกอฟฟ์ กลุ่มตัวอย่างผู้เชี่ยวชาญ เป็นครูผู้มีประสบการณ์สอนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1 ในโรงเรียนมัธยมศึกษา สังกัดกรมสามัญศึกษา ในเขตกรุงเทพมหานคร ปีการศึกษา 2531 จำนวน 96 คน ซึ่งแบ่งเป็น 2 กลุ่ม กลุ่มละ 48 คน ให้กลุ่มหนึ่งตัดสินคะแนนจุดตัด โดยไม่มีสารสนเทศอีกกลุ่มหนึ่งตัดสินคะแนนจุดตัดโดยมีสารสนเทศ (อันได้แก่ ค่าอำนาจจำแนก

ค่าความยาก ค่าการเดา ความถี่ของคะแนนสอบสัดส่วนของผู้สอบที่ตอบข้อสอบแต่ละข้อได้ถูก และความน่าจะเป็นที่จะตอบข้อสอบแต่ละข้อได้ถูก) ส่วนกลุ่มตัวอย่างนักเรียน คือ นักเรียนชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2531 ในโรงเรียนมัธยมศึกษา สังกัดกรมสามัญศึกษา ในเขตกรุงเทพมหานคร จำนวน 2,341 คน ซึ่งใช้ในการสอบแบบทดสอบเพื่อหาค่าพารามิเตอร์ของข้อสอบ (อันได้แก่ ค่าอำนาจจำแนก ค่าความยาก และค่าการเดา) ผลการวิจัยพบว่า คะแนนจุดตัดที่ใช้เป็นเกณฑ์ตัดสินความสามารถขั้นต่ำที่ยอมรับได้ของผู้สอบ ซึ่งอยู่บนสเกลของคะแนนโดเมนหรือคะแนนจริงที่ได้จากการตัดสินของผู้เชี่ยวชาญที่ไม่มีสารสนเทศ และผู้เชี่ยวชาญที่มีสารสนเทศ เท่ากับ 0.68 และ 0.53 ตามลำดับ ซึ่งแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนในการหาค่าดัชนีความคงที่ของการตัดสินคะแนนจุดตัด ซึ่งตัดสินโดยผู้เชี่ยวชาญที่ไม่มีสารสนเทศ และผู้เชี่ยวชาญที่มีสารสนเทศนั้น ผู้วิจัยได้แปลงคะแนนจุดตัดในสเกลของคะแนนโดเมนให้เป็นระดับความสามารถ โดยอาศัยโค้งคุณลักษณะแบบทดสอบปรากฏว่า ดัชนีความคงที่ของการตัดสิน เท่ากับ 0.73 และ 0.76 ตามลำดับ ซึ่งแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

วุฒิคุณ เสาวรัตน์ (2526 : 161) ได้ศึกษาคะแนนเกณฑ์ที่เหมาะสมของแบบทดสอบอิงเกณฑ์วิชาคณิตศาสตร์ เรื่องภาคตัดกรวย ในระดับชั้นมัธยมศึกษาปีที่ 4 จำนวน 11 ฉบับ โดยใช้ทฤษฎีการตัดสินใจทั้งสามวิธี คือ วิธีของเบอร์ก วิธีของเบส และวิธีของเกลส ปรากฏว่าวิธีที่ทำให้คะแนนเกณฑ์มีค่าสูงสุดคือ วิธีของเบส วิธีที่ทำให้คะแนนเกณฑ์สูงรองลงมาคือ วิธีของเกลส และวิธีที่ทำให้คะแนนเกณฑ์มีค่าต่ำสุดคือวิธีของเบอร์ก

สุทธวิวรรณ พิศักดิ์ (2529: 76-77) ได้ศึกษาคะแนนจุดตัดของแบบทดสอบอิงเกณฑ์ชนิดเลือกตอบข้อวินิชาวิทยาศาสตร์ เรื่องสภาวะอากาศ มีแบบทดสอบแบ่งตามเนื้อย่อย 4 ฉบับ และใช้วิธีการหาคะแนนจุดตัดตามวิธีของเบอร์ก วิธีของเกลส และวิธีของฮวิน ปรากฏว่าในแบบทดสอบฉบับเดียวกันจะได้คะแนนที่เหมาะสมเป็นจุดตัดเดียวกันและความเชื่อมั่นของแบบทดสอบฉบับเดียวกันแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

กาญจนา วัฒนสุนทร (2521 : 75 - 76) ได้สร้างแบบทดสอบอิงเกณฑ์วิชาคณิตศาสตร์ เรื่องสมการ ชั้นมัธยมศึกษาปีที่ 1 จำนวน 40 ข้อ แยกเป็น 4 ฉบับ ฉบับละ 10 ข้อ และได้พบว่าคะแนนเกณฑ์จะมีผลต่อค่าความเที่ยงตรงและความเชื่อมั่นของแบบทดสอบ คือ เมื่อกำหนดคะแนนเกณฑ์สูงค่าความเที่ยงตรงและความเชื่อมั่นจะลดลง

จากเอกสารและงานวิจัยที่เกี่ยวข้องทั้งต่างประเทศ และในประเทศที่กล่าวมาข้างต้น จะเห็นว่าวิธีกำหนดเกณฑ์มาตรฐาน หรือคะแนนจุดตัดจะต้องคำนึงถึงข้อมูลในหลาย ๆ ด้าน เพื่อให้การพิจารณาตัดสินมีความคลาดเคลื่อนน้อยที่สุดและการประมาณค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์อาศัยการพิจารณาจากความคลาดเคลื่อนที่คะแนนจุดตัดเป็นหลัก และความคลาดเคลื่อนที่เกิดขึ้นนั้น มีผลมาจากคะแนนจุดตัดด้วยดังนั้นผู้วิจัยมีความสนใจที่จะศึกษาเปรียบเทียบความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ที่มีการจัดเรียงข้อสอบและการหาคะแนนจุดตัดต่างกัน

สมมุติฐานในการวิจัย

1. ความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ที่มีการจัดเรียงข้อสอบต่างกันและ มีการกำหนดคะแนนจุดด้วยวิธีเดียวกัน มีค่าแตกต่างกัน
2. ความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ที่มีการเรียงข้อสอบเหมือนกันเมื่อกำหนดคะแนนจุดด้วยวิธีต่างกัน มีค่าแตกต่างกัน

บทที่ 3

วิธีการดำเนินการศึกษาค้นคว้า

ประชากร

ประชากรที่ใช้ในการศึกษาครั้งนี้ คือ นักเรียนชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2538 สังกัดกรมสามัญศึกษาจังหวัดนครศรีธรรมราช จำนวน 67 โรงเรียน 443 ห้องเรียน จำนวนนักเรียน 18,811 คน

กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้าครั้งนี้ประกอบด้วย นักเรียนชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2538 ของโรงเรียนสังกัดกรมสามัญศึกษา จังหวัดนครศรีธรรมราช จำนวน 1,200 คนซึ่งได้มาโดยการสุ่มแบบแบ่งชั้น (Stratified random Sampling โดยใช้ขนาดโรงเรียนเป็นชั้น (Strata) และ ห้องเรียนเป็นหน่วยสุ่ม (Sampling Unit) มีลำดับขั้นตอนในการสุ่มดังนี้

ขั้นที่ 1 ประมาณกลุ่มตัวอย่างที่ใช้ในการวิจัยด้วยความเชื่อมั่น 95 เปอร์เซนต์ เมื่อเทียบจากตารางขนาดกลุ่มตัวอย่างที่ควรสุ่มจากประชากรแล้วต้องใช้กลุ่มตัวอย่าง อย่างน้อยกลุ่มละ 400 คน รวมทั้งหมด 3 กลุ่ม ซึ่งต้องใช้กลุ่มตัวอย่าง อย่างน้อย 1,200 คน (Yamane. 1967 อ้างอิงจาก ล้วน สายยศ และอังคณา สายยศ. 2536 : 260.)

ขั้นที่ 2 แบ่งโรงเรียนออกเป็น 3 ขนาด คือ โรงเรียนขนาดใหญ่ โรงเรียนขนาดกลาง และโรงเรียนขนาดเล็กซึ่งแบ่งตามเกณฑ์ของกรมสามัญศึกษาโดยอาศัยจำนวนนักเรียนเกณฑ์ในการแบ่ง และจากการสำรวจโรงเรียนได้ดังนี้

ขนาดใหญ่ 18 โรงเรียน 189 ห้องเรียน จำนวนนักเรียน 9,543 คน

ขนาดกลาง 15 โรงเรียน 110 ห้องเรียน จำนวนนักเรียน 4,396 คน

ขนาดเล็ก 34 โรงเรียน 140 ห้องเรียน จำนวนนักเรียน 4,872 คน

ขั้นที่ 3 สุ่มห้องเรียนจากโรงเรียนแต่ละขนาดโดยใช้สัดส่วนร้อยละ 7 ของจำนวนนักเรียนในแต่ละขนาดโรงเรียน ได้ห้องเรียนจากโรงเรียนขนาดใหญ่ 13 ห้องเรียน โรงเรียนขนาดกลาง 8 ห้องเรียน และโรงเรียนขนาดเล็ก 10 ห้องเรียน ดังรายละเอียดในตาราง 6

ตาราง 6 จำนวนนักเรียนที่ใช้เป็นกลุ่มตัวอย่าง จำแนกตามขนาดโรงเรียน จำนวนห้องเรียน และจำนวนนักเรียน

ขนาดโรงเรียน	โรงเรียน	จำนวนห้องเรียน	จำนวนนักเรียน
ขนาดใหญ่	กัลยาณีศรธรรมราช	2	100
	เบญจมาชุกิต	1	40
	โยธินบำรุง	1	40
	ฉวางรัชดาภิเษก	2	80
	เชียรใหญ่	1	40
	ทุ่งสง	1	44
	สตรีทุ่งสง	2	100
	ร่อนพิบูลย์เกียรติวสุนธราภิวัดน์	1	45
	ชะอวดวิทยาคาร	1	40
	ชะอวด	1	45
ขนาดกลาง	ข้างกลางประชานุกูล	1	40
	ทุ่งสงวิทยา	1	37
	ทางพูนวิทยาคาร	1	40
	นาบอน	1	40
	สามตำบลวิทยาประชาสรรค์	1	41
	ทุ่งใหญ่เฉลิมราชอนุสรณ์	1	40
	คีรีราชพัฒนา	1	40
ขนาดเล็ก	พิปูนสังฆรักษ์ประชาอุทิศ	1	35
	ขอนหาดประชาสรรค์	1	30
	ทุ่งสงสหประชาสรรค์	2	60
	ก้างปลาวิทยา	1	40
	เสาชิงวิทยา	2	73
	นางเอื้อยวิทยา	2	75
	เกาะขันธุ์ประชาภิบาล	1	35
	รวม	31	1,200

3. กลุ่มตัวอย่างที่ใช้ทดลองเครื่องมือ และตรวจสอบคุณภาพเครื่องมือเป็นนักเรียน
ชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2538 ของโรงเรียนสังกัดกรมสามัญศึกษา
จังหวัดนครศรีธรรมราช จำนวน 160 คน ได้แก่ นักเรียนโรงเรียนนาบอน จำนวน 80 คน โรงเรียน
ทุ่งสง จำนวน 80 คน

เนื้อหาที่ใช้ในการศึกษาค้นคว้า

เนื้อหาที่ใช้ในการสร้างเครื่องมือคือ เนื้อหาวิชาคณิตศาสตร์ ค 102 ชั้นมัธยมศึกษา
ปีที่ 1 เรื่องสมการ ตามหลักสูตรมัธยมศึกษาตอนต้นพุทธศักราช 2521 (ฉบับปรับปรุง 2533)

เครื่องมือที่ใช้ในการศึกษาค้นคว้า

เครื่องมือที่ใช้ในการศึกษาค้นคว้าครั้งนี้เป็นแบบทดสอบอิงเกณฑ์วิชาคณิตศาสตร์ (ค 102)
ชนิดเลือกตอบ 5 ตัวเลือก ที่มีการจัดเรียงข้อสอบต่างกันจำนวน 3 ฉบับ

- ฉบับที่ 1 จัดเรียงข้อสอบตามลำดับค่าความยากจากค่าสูงไปค่าต่ำ
- ฉบับที่ 2 จัดเรียงข้อสอบตามลำดับเนื้อหาย่อยจากเนื้อหาขั้นพื้นฐานไป
ยังเนื้อหาย่อยที่มีความซับซ้อนมากขึ้น
- ฉบับที่ 3 จัดเรียงข้อสอบโดยวิธีสุ่มข้อสอบจากประชากรข้อสอบมาจัด
เรียงตามลำดับที่สุ่มได้

การดำเนินการสร้างเครื่องมือที่ใช้ในการศึกษาค้นคว้า

ผู้วิจัยได้ดำเนินการสร้างเครื่องมือที่ใช้ในการศึกษาค้นคว้าซึ่งมีขั้นตอนดังนี้

1. ศึกษาทฤษฎีและงานวิจัยที่เกี่ยวกับการสร้างแบบทดสอบอิงเกณฑ์
2. ศึกษาหลักสูตร เนื้อหาและคู่มือครูวิชาคณิตศาสตร์รายวิชา ค 102 ชั้นมัธยมศึกษาปีที่ 1 เพื่อวิเคราะห์เนื้อหาย่อย ๆ กับพฤติกรรมมาหาความสัมพันธ์กัน ดังแสดงในตาราง 7

ตาราง 7 ผลการวิเคราะห์เนื้อหาและพฤติกรรมเรื่องสมการ ชั้นมัธยมศึกษาปีที่ 1

เนื้อหา	พฤติกรรม	ความรู้ ความจำ	ความเข้าใจ	การนำ นำไปใช้	การ วิเคราะห์
สมการ					
1. นิยามสมการ		/	/		
2. คำตอบของสมการ		/	/		
3. สมบัติของการเท่ากัน		/	/		
4. การแก้สมการ					
4.1 การแก้สมการโดยใช้คุณสมบัติ การบวก และการลบ			/		
4.2 การแก้สมการโดยใช้คุณสมบัติ การคูณ และการหาร			/		
4.3 การแก้สมการโดยใช้คุณสมบัติ การบวก การลบ การคูณ การหาร			/		
5. โจทย์ปัญหาสมการ			/	/	/

3. เขียนจุดประสงค์เชิงพฤติกรรมจากตารางความสัมพันธ์เนื้อหาย่อยกับพฤติกรรมหลัก
ได้ดังนี้

3.1 เนื้อหาย่อยนิยามสมการ

ความรู้ความจำ

- นักเรียนสามารถบอกนิยามของสมการได้

ความเข้าใจ

- เมื่อกำหนดประโยคสัญลักษณ์มาให้ นักเรียนสามารถบอกได้ว่าประโยคใด

เป็นสมการ

3.2 เนื้อหาย่อย คำตอบของสมการ

ความรู้ ความจำ

- นักเรียนสามารถอธิบายความหมายของคำว่า "คำตอบสมการ" ได้

ความเข้าใจ
- เมื่อกำหนดสมการมาให้ นักเรียนสามารถระบุจำนวนที่ทำให้สมการเป็นจริงได้
- เมื่อกำหนดจำนวนมาให้ นักเรียนสามารถบอกได้ว่า จำนวนนั้นทำให้สมการใดเป็นจริง

3.3 เนื้อหาย่อย คุณสมบัติของการเท่ากัน

ความรู้ ความจำ
- นักเรียนสามารถอธิบาย ความหมายของคุณสมบัติการบวก คุณสมบัติการลบ คุณสมบัติการคูณ และคุณสมบัติการหารได้
ความเข้าใจ
- เมื่อกำหนดประโยคสัญลักษณ์ ที่เท่ากันมาให้ นักเรียนสามารถแปลงเป็นประโยคที่แสดงคุณสมบัติความเท่ากันได้
- เมื่อกำหนดเงื่อนไขความเท่ากันมาให้ นักเรียนสามารถหาประโยคสัญลักษณ์ที่สมมูลโดยใช้คุณสมบัติของความเท่ากันได้

3.4 เนื้อหาย่อย การแก้สมการ

ความเข้าใจ
- เมื่อกำหนดสมการที่อยู่ในรูปการบวกหรือการลบมาให้ นักเรียนสามารถสรุปขั้นตอนการแก้สมการได้
- เมื่อกำหนดสมการที่อยู่ในรูปเครื่องหมายบวกหรือลบมาให้ นักเรียนสามารถแก้สมการได้
- เมื่อกำหนดสมการที่อยู่ในรูปการคูณหรือการหารมาให้ นักเรียนสามารถระบุขั้นตอนการแก้สมการได้
- เมื่อกำหนดสมการที่อยู่ในรูปการคูณหรือการหารมาให้ นักเรียนสามารถแก้สมการได้
- เมื่อกำหนดสมการที่อยู่ในรูปการบวก การลบ การคูณ การหารมาให้ นักเรียนสามารถระบุขั้นตอนการแก้สมการได้
- เมื่อกำหนดสมการที่อยู่ในรูปการบวก การลบ การคูณ การหารมาให้ นักเรียนสามารถแก้สมการได้

3.5 เนื้อหาย่อย โจทย์ปัญหาสมการ

ความเข้าใจ
- เมื่อกำหนดโจทย์ปัญหาเกี่ยวกับสมการมาให้ นักเรียนสามารถเขียนเป็นประโยคสัญลักษณ์ได้

การนำไปใช้

- เมื่อกำหนดโจทย์ปัญหาที่เกี่ยวข้องกับชีวิตประจำวันมาให้นักเรียนสามารถแก้โจทย์ปัญหาโดยใช้ความรู้เรื่องสมการได้

การวิเคราะห์

- เมื่อกำหนดโจทย์ปัญหาสมการมาให้ นักเรียนสามารถบอกได้ว่าโจทย์ปัญหานั้นสมบูรณ์หรือไม่

4. นำเนื้อหาวิชา เนื้อหาย่อย และจุดประสงค์เชิงพฤติกรรมในแต่ละเนื้อหาวิชาไปให้ผู้เชี่ยวชาญเนื้อหาวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1 จำนวน 3 ท่าน ซึ่งมีประสบการณ์ในการสอนอย่างน้อย 5 ปี ตัดสินว่า พฤติกรรมย่อย ๆ ที่กำหนดขึ้นสอดคล้องกับเนื้อหาหรือไม่ โดยใช้วิธีของโรวินสกี และแฮมเบลตัน โดยพิจารณาดังตัวอย่าง

ตัวอย่าง แบบประเมินความสอดคล้องของเนื้อหาวิชากับจุดประสงค์เชิงพฤติกรรม

คำชี้แจง ให้ท่านพิจารณาว่าจุดประสงค์เชิงพฤติกรรมแต่ละจุดประสงค์ มีความสอดคล้องกับเนื้อหาวิชาหรือไม่โดยพิจารณาให้น้ำหนัก ดังนี้

- + 1 เท่ากับ แน่ใจว่า จุดประสงค์เชิงพฤติกรรมวัดได้ตรงตามเนื้อหาจริง
- 0 เท่ากับ ไม่แน่ใจว่า จุดประสงค์เชิงพฤติกรรมวัดตรงตามเนื้อหา
- 1 เท่ากับ แน่ใจว่า จุดประสงค์เชิงพฤติกรรมวัดได้ไม่ตรงตามเนื้อหา

เนื้อหาย่อย	จุดประสงค์เชิงพฤติกรรม	ผลการประเมิน		
		+ 1	0	- 1
นิยามสมการ	เมื่อกำหนดประโยคสัญลักษณ์มาให้ เรียนบอกได้ว่าประโยคใดเป็นสมการ			

หลังจากให้ผู้เชี่ยวชาญประเมินความสอดคล้องระหว่างจุดประสงค์เชิงพฤติกรรมเนื้อหาเสร็จแล้ว นำคะแนนไปคำนวณหาค่าดัชนีความสอดคล้องโดยวิธีของโรวินสกีและแฮมเบลตัน ถ้ามีค่ามากกว่า 0.5 ก็ถือว่าจุดประสงค์เชิงพฤติกรรมนั้น ๆ สอดคล้องกับเนื้อหาของหลักสูตร และผลจากการพิจารณาของผู้เชี่ยวชาญปรากฏว่าได้ค่าดัชนีความสอดคล้องตั้งแต่ 0.67 - 1.00

5. สร้างแบบทดสอบอิงเกณฑ์ชนิดเลือกตอบ 5 ตัวเลือก จำนวน 62 ข้อ โดยเขียนข้อคำถามตามจุดประสงค์เชิงพฤติกรรมที่วิเคราะห์แล้ว

6. นำแบบทดสอบอิงเกณฑ์ที่สร้างขึ้นไปให้ผู้เชี่ยวชาญ ในเนื้อหาวิชาพิจารณาตัดสินข้อสอบแต่ละข้อ ในแบบทดสอบวัดตามจุดประสงค์เชิงพฤติกรรมนั้นจริงหรือไม่ ดังตัวอย่าง

- ตัวอย่าง แบบประเมินความสอดคล้องระหว่างข้อสอบกับจุดประสงค์เชิงพฤติกรรม
 คำชี้แจง ให้ท่านพิจารณาว่าข้อสอบแต่ละข้อ ต่อไปนี้วัดตรงตามจุดประสงค์
 เชิงพฤติกรรมข้อนั้น ๆ หรือไม่ โดยให้พิจารณา น้ำหนักดังนี้
- + 1 เมื่อท่านแน่ใจว่าข้อสอบข้อนั้น วัดตรงตามจุดประสงค์เชิงพฤติกรรม
 นั้นจริง
 - 0 เมื่อท่านไม่แน่ใจว่าข้อสอบข้อนั้น วัดตรงตามจุดประสงค์
 เชิงพฤติกรรมนั้น
 - 1 เมื่อท่านแน่ใจว่าข้อสอบข้อนั้น ไม่ได้วัดตรงตามจุดประสงค์
 เชิงพฤติกรรมนั้นจริง

จุดประสงค์เชิงพฤติกรรม	ข้อสอบ	ผลการประเมิน		
		+1	0	-1
เมื่อกำหนดประโยค สัญลักษณ์มาให้นักเรียน สามารถบอกได้ว่าประโยคใด เป็นสมการ	จากประโยคสัญลักษณ์ที่กำหนดให้ ประโยคใดเป็นสมการ ก. $3x < 12$ ข. $6x - 2 = x + 1$ ค. $x - 8 > 2x - 14$ ง. $-(16 + x) \neq -23$ จ. $x + 18 - 4x \geq 2x - 15$			

หลังจากให้ผู้เชี่ยวชาญประเมินความสอดคล้องระหว่างข้อสอบกับ จุดประสงค์ เชิง
 พฤติกรรมแล้ว นำคะแนนไปคำนวณหาค่าดัชนีความสอดคล้องโดยวิธีของโรวีเนลลีและแฮเบลตัน
 (อังคณา สายยศ. 2527 : 183 - 186) ถ้ามีค่ามากกว่า 0.5 ก็ถือว่าข้อสอบนั้น ๆ วัดตรงตาม
 จุดประสงค์เชิงพฤติกรรมนั้นจริง ผลจากการประเมินข้อสอบทั้งหมด 62 ข้อ ของผู้เชี่ยวชาญ
 ปรากฏว่า ข้อสอบมีค่าดัชนีความสอดคล้องอยู่ระหว่าง 0.60 - 1.00 ดังนั้นผู้วิจัยจึงนำข้อสอบ
 ทั้งหมดไปทดลองหาคคุณภาพ

7. นำแบบทดสอบอิงเกณฑ์ที่สร้างขึ้น จำนวน 62 ข้อไปทดสอบกับกลุ่มตัวอย่างที่เป็น
 กลุ่มหาคคุณภาพของเครื่องมือ จำนวน 160 คน ได้แก่นักเรียนโรงเรียนนาบอน จำนวน 80 คน
 และนักเรียนโรงเรียนทุ่งสงจำนวน 80 คน

8. นำผลการสอบมาวิเคราะห์รายข้อเพื่อคัดเลือกและปรับปรุงข้อสอบ โดยใช้ดัชนีค่า
 ค่าความยาก และดัชนีค่าอำนาจจำแนก ผลปรากฏว่า ข้อสอบทั้งหมด มีค่าความยากอยู่

ระหว่าง 0.09 ถึง 0.87 และมีค่าอำนาจจำแนกอยู่ระหว่าง -0.21 ถึง 0.57 ดังนั้นผู้วิจัยได้คัดเลือก ข้อสอบที่มีค่าความยากอยู่ระหว่าง 0.24 ถึง 0.87 และค่าอำนาจจำแนกอยู่ระหว่าง 0.20 ถึง 0.57 จำนวน 30 ข้อ เพื่อใช้เป็นเครื่องมือในการวิจัย (ดังแสดงในตารางภาคผนวก) โดยจำแนกตามเนื้อหาดังนี้ นิยามสมการ 4 ข้อ คำตอบของสมการ 3 ข้อ สมบัติการเท่ากัน 4 ข้อ การแก้สมการ 16 ข้อ โจทย์ปัญหา 3 ข้อ

9. นำข้อสอบที่เลือกได้ตามข้อ 8 นำมาจัดทำเป็นรูปเล่มเพื่อใช้เป็นเครื่องมือในการวิจัย
 - ฉบับที่ 1 จัดเรียงข้อสอบตามค่าความยากจากค่าสูงไปหาค่าต่ำ
 - ฉบับที่ 2 จัดเรียงข้อสอบตามลำดับเนื้อหาย่อย โดยเรียงจากเนื้อหาย่อยชั้นพื้นฐาน ไปยังเนื้อหาย่อยที่มีเนื้อหาซับซ้อนมากขึ้น
 - ฉบับที่ 3 จัดเรียงข้อสอบโดยวิธีสุ่มข้อสอบจากประชากรข้อสอบมาจัดเรียงตามลำดับที่สุ่มได้

วิธีดำเนินการศึกษาค้นคว้า

ในการศึกษาครั้งนี้ ผู้วิจัยได้ดำเนินการดังนี้

1. ประสานงานกับโรงเรียนในกลุ่มตัวอย่าง เพื่อกำหนดวันเวลาในการสอบ
2. นำแบบทดสอบไปสอบกับนักเรียนกลุ่มตัวอย่าง ตั้งแต่วันที่ 5 - 23 กุมภาพันธ์ 2539 โดยนำแบบทดสอบทั้ง 3 ฉบับ มาจัดเรียงลำดับเป็นชุด ๆ ตั้งแต่ฉบับที่ 1 ถึงฉบับที่ 3 หมุนเวียนกันไปตลอด แล้วนำไปแจกนักเรียนในแต่ละแถว โดยแจกตามลำดับของแบบทดสอบที่เรียงไว้
3. ชี้แจงให้นักเรียนเข้าใจวิธีการทำข้อสอบ โดยเน้นให้ทำตามลำดับข้อ และขอความร่วมมือจากอาจารย์ผู้สอนชี้แจงให้นักเรียนทำข้อสอบอย่างเต็มความสามารถ
4. นำผลการสอบมาตรวจให้คะแนนแบบ 0-1 แล้วรวมคะแนนแต่ละคนในการสอบแต่ละครั้งเพื่อหาคะแนนจุดตัดตามวิธีของแกลส วิธีของเบส์
5. นำผลที่ได้จากการทดสอบมาวิเคราะห์ทางสถิติ

ลำดับขั้นการวิเคราะห์ข้อมูล

1. หาค่าสถิติพื้นฐาน คือ ค่าเฉลี่ยและความเบี่ยงเบนมาตรฐาน
2. คำนวณหาคะแนนจุดตัดด้วยวิธีของแกลส วิธีของเบส์
3. คำนวณหาค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ทั้ง 3 ฉบับ โดยใช้สูตรของโลเวท (Lovett) และสูตรของลิวิงสตัน (Livingston)
4. ทดสอบความแตกต่างของความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ตามสมมุติฐานที่ตั้งไว้

สถิติที่ใช้ในการวิเคราะห์ข้อมูล

1. ค่าสถิติพื้นฐานของแบบทดสอบ คือ คะแนนเฉลี่ยและค่าความเบี่ยงเบนมาตรฐาน
2. ดัชนีค่าความยากง่ายของข้อสอบ (อังคณา สายยศ. 2536 : 28)

$$P = \frac{R}{N}$$

เมื่อ	P	แทน	ดัชนีค่าความยากหรือค่าความง่าย
	R	แทน	จำนวนนักเรียนที่ทำข้อสอบถูก
	N	แทน	จำนวนนักเรียนที่ทำข้อสอบทั้งหมด

3. ดัชนีค่าอำนาจจำแนก คำนวณจากสูตรของ Brennan (1972) (อังคณา สายยศ. 2536 : 29)

$$B = \frac{U}{n_1} - \frac{L}{n_2}$$

เมื่อ	B	แทน	ดัชนีค่าอำนาจจำแนกของแบบทดสอบอิงเกณฑ์
	U	แทน	จำนวนนักเรียนที่ทำข้อสอบถูกของกลุ่มที่สอบผ่านเกณฑ์
	L	แทน	จำนวนนักเรียนที่ทำข้อสอบถูกของกลุ่มที่ไม่ผ่านเกณฑ์
	n_1	แทน	จำนวนนักเรียนที่สอบผ่านเกณฑ์
	n_2	แทน	จำนวนนักเรียนที่สอบไม่ผ่านเกณฑ์

ในการกำหนดเกณฑ์หรือคะแนนจุดตัด เพื่อหาดัชนีค่าอำนาจจำแนกตามสูตรดังกล่าว ผู้วิจัยใช้วิธีกำหนดจุดตัดตามวิธีของแองกอฟฟ์โดยให้ผู้เชี่ยวชาญที่มีประสบการณ์ในการสอนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1 อย่างน้อย 5 ปีจำนวน 3 ท่าน เป็นผู้พิจารณาตัดสิน และปรากฏว่าในการวิจัยครั้งนี้ ได้เกณฑ์มีค่าเท่ากับ 33 คะแนน ซึ่งคิดเป็นเปอร์เซ็นต์เท่ากับ 53 เปอร์เซ็นต์ของคะแนนการสอบ

4. ความเที่ยงตรงเชิงเนื้อหาโดยวิธีของโรวินลีย์ และแฮมเบลตัน
(บุญเชิด ภิญโญอนันตพงษ์. 2527 : 68-69)

$$IOC = \frac{R}{N}$$

เมื่อ IOC แทน ดัชนีความสอดคล้องของข้อสอบกับจุดประสงค์
R แทน ผลรวมของคะแนนความคิดเห็นของผู้เชี่ยวชาญ
N แทน จำนวนผู้เชี่ยวชาญทั้งหมด

5. หาคะแนนจัดของแบบทดสอบแต่ละฉบับด้วยวิธีของเกลส วิธีของเบส์
5.1 วิธีของเกลส (Glass. 1978 : 237 - 261) โดยกำหนดเกณฑ์ภายนอก
จากผลการเรียนเฉลี่ยสะสม (G.P.A) ตั้งแต่ 2.00 ขึ้นไป

เกณฑ์ภายนอก (GPA.)

		ผ่านเกณฑ์	ไม่ผ่านเกณฑ์
เกณฑ์ของแบบทดสอบ	ไม่ผ่าน	P_A	P_B
	ผ่าน	P_C	P_D

เมื่อ P_A แทน สัดส่วนของคนที่ไม่ผ่านเกณฑ์แบบทดสอบแต่ผ่านเกณฑ์ภายนอก
 P_B แทน สัดส่วนของคนที่ไม่ผ่านทั้งเกณฑ์แบบทดสอบ และเกณฑ์ภายนอก
 P_C แทน สัดส่วนของคนที่ผ่านมาทั้งเกณฑ์แบบทดสอบ และเกณฑ์ภายนอก
 P_D แทน สัดส่วนของคนที่ผ่านมาเกณฑ์แบบทดสอบแต่ไม่ผ่านเกณฑ์ภายนอก

เกณฑ์ที่เหมาะสมของแบบทดสอบคือ

$$f(CX) = \frac{(P_A + P_D)}{(P_B + P_C)} \quad \text{มีค่าน้อยที่สุด}$$

5.2 วิธีการของเบส์ (Bayesian Decision Theoretic Procedure) (Swaminathan, Hambleton and Algina. 1975 : 88 - 93) มีวิธีการหาคะแนนจุดตัด ดังนี้

5.2.1 แปลงคะแนนที่ได้จากการทำแบบทดสอบอิงเกณฑ์ โดยวิธีการแปลงคะแนน อาร์คไซน์เพื่อให้คะแนนมีการแจกแจงเป็นปกติ โดยใช้สูตร

$$g_i = \sin^{-1} \sqrt{(X_i + 3/8) / (n + 3/4)}$$

เมื่อ g_i แทน คะแนนที่ได้จากการสอบแบบทดสอบที่แปลงเป็น อาร์คไซน์
 X_i แทน คะแนนที่ได้จากการสอบแบบทดสอบ
 n แทน จำนวนข้อสอบในแบบทดสอบ

5.2.2 หาค่าเฉลี่ย (μ_i) และ ความแปรปรวน (σ_i^2) ของ γ_i จากสูตร

$$\mu_i = \bar{g} + p^* + (g_i - \bar{g})$$

$$\sigma_i^2 = \frac{1 + (m-1)p^*}{(4n+2)m} + (g_i - \bar{g})^2 \sigma^2$$

เมื่อ μ_i แทน ค่าเฉลี่ยของ γ_i
 σ_i^2 แทน ความแปรปรวนของ γ_i
 m แทน จำนวนนักเรียนที่เข้าสอบ
 n แทน จำนวนข้อสอบ

$$\bar{g} = m^{-1} \sum_{j=1}^m g_j$$

ρ^* และ σ^{2*} แทน ค่าที่ได้จากการเปิดตารางของแวง(Wang)
เมื่อทราบค่า v , λ , S^2

$$\text{เมื่อ } V = 8$$

$$\lambda = (V-2) \phi$$

$$\phi = (4t + 2)^{-1}$$

t แทน จำนวนข้อสอบที่เหมาะสมที่จะให้ข้อมูลเกี่ยวกับโดเมน
ของผู้สอบมากที่สุด

$$S_g^2 = m^{-1} \sum_{i=1}^m (g_i - \bar{g})^2$$

5.2.3 หาความน่าจะเป็นของการสอบผ่านหรือไม่ผ่านของคะแนนผู้สอบดังนี้

5.2.3.1 หาส่วนเบี่ยงเบนปกติมาตรฐาน (Z_{oi}) ของคะแนน γ_i
แต่ละค่าโดย

$$Z_{oi} = (\gamma_{oi} - \mu_0) / \sigma_i$$

เมื่อ Z_{oi} แทน ส่วนเบี่ยงเบนปกติมาตรฐาน

$$\gamma_0 = 'sin' \sqrt{\pi_0}$$

π_0 แทน คะแนนเกณฑ์โดยประมาณที่กำหนดจากผู้เชี่ยวชาญ
ด้านเนื้อหา

5.2.3.2 หากความน่าจะเป็นในการสอบผ่านและไม่ผ่านของคะแนนที่ได้จากผู้สอบ โดยการหาจากตารางพื้นที่ใต้โค้งปกติ ดังนี้

$$\text{ความน่าจะเป็นที่จะสอบไม่ผ่าน} = \text{Prob} [Z \geq Z_{oi} / \text{Data}]$$

$$\text{ความน่าจะเป็นที่จะสอบผ่าน} = \text{Prob} [Z < Z_{oi} / \text{Data}]$$

5.2.4 หากความสูญเสียที่คาดหวังของการตัดสินใจให้ผ่านและไม่ผ่านของคะแนนผู้สอบดังนี้

ความสูญเสียที่คาดหวังของการตัดสินใจให้ผ่าน

$$= I_{12} \text{ Prob} [Z < Z_{oi} / \text{Data}]$$

ความสูญเสียที่คาดหวังของการตัดสินใจให้ไม่ผ่าน

$$= I_{21} \text{ Prob} [Z \geq Z_{oi} / \text{Data}]$$

เมื่อ I_{12} แทน ความสูญเสียที่เกิดจากความคลาดเคลื่อนในการยอมรับที่ผิด

เมื่อ I_{21} แทน ความสูญเสียที่เกิดจากความคลาดเคลื่อนในการปฏิเสธผิด

6. หากความเชื่อมั่นของแบบทดสอบอิงเกณฑ์โดยใช้วิธีการดังนี้

6.1 คำนวณจากสูตรของโลเวท (Lovett. 1978)

(อังกฤษ สายยศ. 2536 : 45 ; อ้างอิงมาจาก Lovett. 1978)

คำนวณได้จาก

$$r_{cc} = 1 - \left\{ \frac{K \sum X_i - \sum X_i^2}{(K-1) \sum (X_i - C)^2} \right\}$$

เมื่อ r_{cc} แทน ความเชื่อมั่นของแบบทดสอบอิงเกณฑ์

K แทน จำนวนข้อสอบ

$\sum X_i$ แทน ผลรวมของคะแนนทั้งหมด

$\sum X_i^2$ แทน ผลรวมทั้งหมดของคะแนนแต่ละตัวยกกำลังสอง

X_i แทน คะแนนของนักเรียนแต่ละคน

C แทน คะแนนจุดตัดของแบบทดสอบอิงเกณฑ์

ในการวิจัยครั้งนี้ ใช้วิธีกำหนดคะแนนจุดตัดตามวิธีของเบส์ มีค่าเท่ากับ 13, 15, 15 และกำหนดคะแนนจุดตัดตามวิธีของแกลส มีค่าเท่ากับ 18,16 และ 16 คะแนน

6.2 วิธีของลิวิงสตัน (Livingston. 1972)

(Crocker and Algina.1986 :203) คำนวณความเชื่อมั่นได้จากสูตรดังนี้

$$K^2(X,T) = \frac{\sigma_X^2(KR-20) + (\mu_x - n_i C)^2}{\sigma_X^2 + (\mu - n_i C)^2}$$

เมื่อ $K^2(X,T)$	แทน	ความเชื่อมั่นตามวิธีของลิวิงสตัน
$\sigma_X^2, \mu_x,$	แทน	ความแปรปรวนและค่าเฉลี่ยของคะแนนที่สอบได้
KR-20	แทน	ค่าความเชื่อมั่นแบบอิงกลุ่มหาได้จากสูตร KR-20
n_i	แทน	จำนวนข้อสอบ
C	แทน	คะแนนจุดตัดหรือคะแนนเกณฑ์ ซึ่งในการวิจัยครั้งนี้ ใช้วิธีกำหนดคะแนนจุดตัดตามวิธีของเบส์ มีค่าเท่ากับ 13,15, 15และกำหนดคะแนนจุดตัดตามวิธีของแกลส มีค่าเท่ากับ 18,16 และ 16 คะแนน

7. ทดสอบความแตกต่างของค่าความเชื่อมั่น โดยใช้ไคสแควร์ (Chi-Square) (Wert and others. 1954. 1954 : 298)

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

เมื่อ χ^2	แทน	ค่า ไค-สแควร์
Z	แทน	ค่าสัมประสิทธิ์สหสัมพันธ์ที่แปลงเป็นค่า Fisher's Z
N	แทน	จำนวนคนในกลุ่ม
k	แทน	จำนวนค่าสัมประสิทธิ์ความเชื่อมั่นที่นำมาเปรียบเทียบกัน

8. ทดสอบความแตกต่างของค่าความเชื่อมั่นรายคู่ โดยใช้สูตรความแตกต่างของคะแนนมาตรฐาน (Z) (Ferguson. 1981 : 196)

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

เมื่อ Z_{r1} , Z_{r2} แทน คะแนนมาตรฐานที่แปลงจากค่าสัมประสิทธิ์สหสัมพันธ์ตามวิธีของ Fisher's Z Transformation ของแบบทดสอบฉบับที่ 1 และ 2 ตามลำดับ

N_1 , N_2 แทน จำนวนนักเรียนในกลุ่มที่ 1 และ 2 ตามลำดับ

บทที่ 4

ผลการวิเคราะห์ข้อมูล

ในการศึกษาครั้งนี้ เพื่อให้เกิดความเข้าใจตรงกันในการนำเสนอและแปลผลการวิเคราะห์ข้อมูล ผู้วิจัยจึงได้ กำหนดสัญลักษณ์และอักษรย่อ ที่ใช้ในการวิเคราะห์ข้อมูลดังนี้

สัญลักษณ์และอักษรที่ใช้ในการวิเคราะห์ข้อมูล

N	แทน	จำนวนนักเรียนในกลุ่มตัวอย่าง
k	แทน	จำนวนข้อสอบของแบบทดสอบ
P	แทน	ดัชนีค่าความยากหรือค่าความง่าย
B	แทน	ดัชนีค่าอำนาจจำแนก
C	แทน	คะแนนจุดตัดของแบบทดสอบ
$\bar{X}$	แทน	คะแนนเฉลี่ยของกลุ่มตัวอย่าง
S	แทน	ความเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่าง
S^2	แทน	ความแปรปรวนของกลุ่มตัวอย่าง
Z	แทน	คะแนนมาตรฐาน
χ^2	แทน	ค่า ไค-สแควร์
$f(C_X)$	แทน	คะแนนจุดตัดที่เหมาะสมตามวิธีของเกลส
$P_A + P_B$	แทน	ผลบวกของความคลาดเคลื่อนของการจำแนกผิดประเภทที่ 1 และ 2 ตามวิธีของเกลส
$L_{21} P[Z > Z_{oi} Data]$	แทน	ความสูญเสียที่คาดหวังในการตัดสินใจให้ สอบตก สำหรับการหาคะแนนจุดตัดโดยวิธี ของเบส์
$L_{12} P[Z < Z_{oi} Data]$	แทน	ความสูญเสียที่คาดหวังในการตัดสินใจให้ สอบได้ สำหรับการหาคะแนนจุดตัดโดยวิธี ของเบส์
r_{cc}	แทน	ความเชื่อมั่นของแบบทดสอบ

การเสนอผลการวิเคราะห์ข้อมูล

ในการศึกษาครั้งนี้ผู้วิจัยได้เสนอผลการวิเคราะห์ข้อมูลตามลำดับคือ

1. ค่าสถิติพื้นฐานของแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม
2. คะแนนจุดตัดของคะแนนแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม เมื่อกำหนดคะแนนจุดตัดด้วยวิธีของเบส์และวิธีของเกลส
3. ความเชื่อมั่นของแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม
4. ทดสอบความแตกต่างของค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ที่มีการจัดเรียงข้อสอบต่างกัน และมีการกำหนดคะแนนจุดตัดด้วยวิธีเดียวกัน
5. ทดสอบความแตกต่างของค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ที่มีการจัดเรียงข้อสอบเหมือนกัน เมื่อกำหนดคะแนนจุดตัดด้วยวิธีต่างกัน

ผลการวิเคราะห์ข้อมูล

1. ค่าสถิติพื้นฐานของแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม

หลังจากนำแบบทดสอบอิงเกณฑ์ วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1 ที่มีการจัดเรียงข้อสอบต่างกัน 3 ฉบับ คือเรียงตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม ฉบับละ 30 ข้อ ไปทดสอบกับกลุ่มตัวอย่างกลุ่มละ 400 คน ผู้วิจัยได้นำคะแนนผลการสอบทั้งสามฉบับ มาคำนวณ ค่าสถิติพื้นฐาน ได้แก่คะแนนเฉลี่ย และค่าเบี่ยงเบนมาตรฐาน ดังแสดงในตาราง 8

ตาราง 8 ค่าสถิติพื้นฐานของแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และ โดยวิธีสุ่ม

แบบทดสอบ	k	$\bar{x}$	S.	S^2
เรียงตามค่าความยาก	30	18.078	7.174	51.466
เรียงตามเนื้อหา	30	17.867	7.019	49.266
เรียงโดยวิธีสุ่ม	30	16.599	7.036	49.505

จากตาราง 8 แสดงให้เห็นว่า คะแนนเฉลี่ยของแบบทดสอบที่จัดเรียงข้อสอบต่างกัน 3 รูปแบบ คือ เรียงตามค่าความยาก เรียงตามเนื้อหา เรียงโดยวิธีสุ่ม มีค่าอยู่ระหว่าง 16.599 - 18.078 โดยแบบทดสอบที่จัดเรียงตามค่าความยาก มีคะแนนเฉลี่ยสูงสุด คือ 18.078 และแบบทดสอบที่จัดเรียงโดยวิธีสุ่มมีคะแนนเฉลี่ยต่ำสุด คือ 16.599 ค่าเบี่ยงเบนมาตรฐานของแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหาและเรียงโดยสุ่ม มีค่าอยู่ระหว่าง 7.019 - 7.174 ซึ่งแบบทดสอบที่จัดเรียงตามเนื้อหามีค่าความเบี่ยงเบนมาตรฐานต่ำสุดคือ 7.019 และแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก มีค่าเบี่ยงเบนมาตรฐาน สูงสุด คือ 7.174

2. คะแนนจุดตัดของแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม

ผู้วิจัยได้นำคะแนนผลการสอบของแบบทดสอบที่จัดเรียงข้อสอบต่างกัน คือ เรียงตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม ฉบับละ 30 ข้อ มาคำนวณหาคะแนนจุดตัด ตามวิธีของเบส์ และวิธีของแกลส โดยกำหนดเกณฑ์ภายนอกว่า นักเรียนที่มีระดับคะแนนเฉลี่ยสะสมต่ำกว่า 2.00 เป็นกลุ่มไม่รอบรู้ และนักเรียนที่มีระดับผลการเรียนเฉลี่ยสะสมตั้งแต่ 2.00 ขึ้นไป เป็นกลุ่มรอบรู้ ดังแสดงไว้ในตาราง 9

ตาราง 9 คะแนนจุดตัดของแบบทดสอบที่จัดเรียงข้อสอบต่างกัน 3 รูปแบบ คือ เรียงตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่มเมื่อกำหนดคะแนนจุดตัดด้วยวิธีของเบส์และวิธีของแกลส

ลักษณะการจัดเรียงของข้อสอบ	k	คะแนนจุดตัด	
		วิธีของเบส์	วิธีของแกลส
เรียงตามค่าความยาก	30	13	18
เรียงตามเนื้อหา	30	15	16
เรียงโดยวิธีสุ่ม	30	15	16

จากตาราง 9 แสดงให้เห็นว่าคะแนนจุดตัดของแบบทดสอบที่มีการจัดเรียงข้อสอบต่างกัน 3 รูปแบบ คือ เรียงตามค่าความยาก เรียงตามเนื้อหาและเรียงโดยวิธีสุ่ม เมื่อกำหนดคะแนนจุดตัดด้วยวิธีของเบส์ มีค่าอยู่ระหว่าง 13 - 15 คะแนน ซึ่งแบบทดสอบที่จัดเรียงข้อสอบตามเนื้อหา และแบบทดสอบที่จัดเรียงข้อสอบโดยวิธีสุ่มมีคะแนนจุดตัดเท่ากันคือ 15 คะแนน และแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก มีคะแนนจุดตัดต่ำสุดคือ 13 คะแนน และเมื่อกำหนดคะแนนจุดตัดด้วยวิธีของแกลส ปรากฏว่ามีคะแนนอยู่ระหว่าง 16 - 18 คะแนน โดยที่แบบทดสอบที่จัดเรียงข้อสอบตามค่าความยากมีคะแนนจุดตัดสูงสุด คือ 18 คะแนน และแบบทดสอบที่จัดเรียงข้อสอบตามเนื้อหา กับแบบทดสอบที่จัดเรียงข้อสอบโดยวิธีสุ่มมีคะแนนจุดตัดเท่ากัน คือ 16 คะแนน

3. ความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ ที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม

ผู้วิจัยได้นำคะแนนผลการสอบของแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม มาคำนวณหาค่าความเชื่อมั่นด้วยสูตรของโลเวท และสูตรของลิวิสตัน โดยใช้คะแนนจุดตัดตามวิธีของเบส์และวิธีของแกลส ดังแสดงในตาราง 10 และ 11

ตาราง 10 ค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม โดยใช้คะแนนจุดตัดตามวิธีของเบส์และวิธีของแกลส ซึ่งคำนวณจากสูตรความเชื่อมั่นของโลเว

ลักษณะการจัดเรียงข้อสอบ	วิธีของเบส์			วิธีของแกลส		
	C	r_{cc}	SE_{meas}	C	r_{cc}	SE_{meas}
เรียงตามค่าความยาก	13	0.928	1.925	18	0.892	2.358
เรียงตามเนื้อหา	15	0.899	2.213	16	0.891	2.317
เรียงโดยวิธีสุ่ม	15	0.885	2.386	16	0.883	2.407

จากตาราง 10 แสดงให้เห็นว่าแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม มีค่าความเชื่อมั่นที่คำนวณจากคะแนนจุดตัดตามวิธีของเบส์ โดยใช้สูตรความเชื่อมั่นของโลเว มีค่าอยู่ระหว่าง 0.885 - 0.928 ซึ่งแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยากมีค่าความเชื่อมั่นสูงสุด คือ 0.928 และแบบทดสอบที่จัดเรียงข้อสอบโดยวิธีสุ่มมีค่าความเชื่อมั่นต่ำสุด คือ 0.885 ส่วนค่าความเชื่อมั่นที่คำนวณโดยใช้คะแนนจุดตัดตามวิธีของแกลส มีค่าอยู่ระหว่าง 0.883 - 0.892 แบบทดสอบที่จัดเรียงข้อสอบตามค่าความยากมีค่าความเชื่อมั่นสูงสุด คือ 0.892 และแบบทดสอบที่จัดเรียงข้อสอบโดยวิธีสุ่มมีค่าความเชื่อมั่นต่ำสุด คือ 0.883 ค่าความคลาดเคลื่อนมาตรฐานในการวัดของแบบทดสอบที่คำนวณค่าความเชื่อมั่นโดยใช้คะแนนจุดตัดของเบส์ มีค่าอยู่ระหว่าง 1.925 ถึง 2.386 แบบทดสอบที่จัดเรียงข้อสอบตามค่าความยากมีค่าความคลาดเคลื่อนในการวัดต่ำสุด และแบบทดสอบที่จัดเรียงข้อสอบโดยวิธีสุ่มมีค่าความคลาดเคลื่อนในการวัดสูงสุด สำหรับค่าความคลาดเคลื่อนของแบบทดสอบที่คำนวณค่าความเชื่อมั่นโดยใช้คะแนนจุดตัดตามวิธีของแกลส มีค่าอยู่ระหว่าง 2.317 ถึง 2.407 แบบทดสอบที่จัดเรียงข้อสอบตามเนื้อหามีค่าความคลาดเคลื่อนมาตรฐานในการวัดต่ำสุด และแบบทดสอบที่จัดเรียงข้อสอบโดยวิธีสุ่มมีค่าความคลาดเคลื่อนมาตรฐานในการวัดสูงสุด

ตาราง 11 ค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ ที่มีการจัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม ที่คำนวณจากสูตรของลิวิสตัน โดยใช้คะแนนจุดตัดตาม วิธีของเบส์และวิธีของแกลส

ลักษณะการจัดเรียงแบบทดสอบ	วิธีของเบส์			วิธีของแกลส		
	C	r_{cc}	SE_{meaB}	C	r_{cc}	SE_{meaB}
เรียงตามค่าความยาก	13	0.999	0.227	18	0.999	0.227
เรียงตามเนื้อหา	15	0.999	0.221	16	0.999	0.221
เรียงโดยวิธีสุ่ม	15	0.999	0.222	16	0.999	0.222

จากตาราง 11 แสดงให้เห็นว่าแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่มมีค่าความเชื่อมั่นที่คำนวณจากคะแนนจุดตัดตามวิธีของเบส์ โดยใช้สูตรค่าความเชื่อมั่นของลิวิสตัน มีค่าเท่ากันทั้งสามฉบับคือ 0.999 และค่าความเชื่อมั่นของแบบทดสอบที่คำนวณจากคะแนนจุดตัดตามวิธีของแกลส มีค่าเท่ากันทั้งสามฉบับด้วย คือ 0.999 ส่วนค่าความคลาดเคลื่อนมาตรฐานในการวัดของแบบทดสอบที่คำนวณความเชื่อมั่นจากคะแนนจุดตัดตามวิธีของเบส์ และวิธีของแกลส มีค่าอยู่ระหว่าง 0.221 ถึง 0.227 แบบทดสอบที่จัดเรียงข้อสอบตามค่าความยากมีค่าความคลาดเคลื่อนมาตรฐานสูงสุด คือ 0.227 และแบบทดสอบที่จัดเรียงข้อสอบตามเนื้อหา กับแบบทดสอบที่จัดเรียงข้อสอบโดยวิธีสุ่มมีค่าความคลาดเคลื่อนมาตรฐานใกล้เคียงกัน คือ 0.221 และ 0.222

4. ทดสอบความแตกต่างของค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ที่จัดเรียงข้อสอบต่างกัน และมีการกำหนดคะแนนจุดตัดด้วยวิธีเดียวกัน

เพื่อให้ทราบว่า ค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ ที่มีการจัดเรียงข้อสอบต่างกัน และมีการกำหนดคะแนนจุดตัดด้วยวิธีเดียวกัน แตกต่างกันหรือไม่ ผู้วิจัยได้นำค่าความเชื่อมั่นของแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม ที่คำนวณคะแนนจุดตัดตามวิธีของเบส์และวิธีของแกลส มาแปลงเป็นคะแนนมาตรฐาน (Z) ตามวิธีของฟิชเชอร์ (Fisher's Z Transformation) แล้วนำไปทดสอบความแตกต่างโดยใช้ (χ^2) ดังแสดงในตาราง 12

ตาราง 12 ผลการทดสอบความแตกต่างของค่าความเชื่อมั่น ของแบบทดสอบอิงเกณฑ์ที่มีการจัดเรียงข้อสอบต่างกัน 3 รูปแบบคือเรียงตามค่าความยากเรียงตามเนื้อหา และเรียงโดยวิธีสุ่มคำนวณค่าความเชื่อมั่นจากคะแนนจุดตัดตามวิธีของเบส์ และวิธีของแกลส โดยใช้สูตรความเชื่อมั่นของโลเวท

วิธีคำนวณคะแนนจุดตัด	k	r	Z	χ^2
วิธีของเบส์				
เรียงตามค่าความยาก	30	0.928	1.603	10.395**
เรียงตามเนื้อหา	30	0.899	1.436	
เรียงโดยวิธีสุ่ม	30	0.885	1.384	
วิธีของแกลส				
เรียงตามค่าความยาก	30	0.892	1.425	0.744
เรียงตามเนื้อหา	30	0.891	1.424	
เรียงโดยวิธีสุ่ม	30	0.883	1.381	

จากตาราง 12 แสดงให้เห็นว่า ค่าความเชื่อมั่นของโลเวทของแบบทดสอบอิงเกณฑ์ที่จัดเรียงข้อสอบต่างกัน คือ เรียงตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม ที่มีการกำหนดคะแนน จุดตัดด้วยวิธีของเบส์ มีอย่างน้อย 1 คู่ ที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ที่มีการจัดเรียงข้อสอบต่างกัน 3 รูปแบบเมื่อกำหนดคะแนนจุดตัดตามวิธีของแกลส ต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

เพื่อให้ทราบว่าค่าความเชื่อมั่นของโลเวทของแบบทดสอบอิงเกณฑ์ที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม เมื่อกำหนดคะแนนจุดตัดด้วยวิธีของเบส์คู่ใดบ้างที่แตกต่างกัน ผู้วิจัยจึงนำค่าความเชื่อมั่นของแบบทดสอบมาเปลี่ยนเป็นคะแนนมาตรฐาน (Z) ตามวิธีของฟิชเชอร์ (Fisher's Z Transformation) แล้วนำไปทดสอบความแตกต่าง ของคะแนนมาตรฐานเป็นรายคู่ ด้วยสูตรการทดสอบความแตกต่างของคะแนนมาตรฐาน (Z - test) ดังแสดงในตาราง 13

ตาราง 13 ผลการทดสอบความแตกต่างของค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ที่จัดเรียงข้อสอบต่างกัน 3 รูปแบบเมื่อกำหนดคะแนนจุดตัดด้วยวิธีของเบส์ โดยใช้สูตรความเชื่อมั่นของโลเว

ลักษณะการจัดเรียง แบบทดสอบ		เรียงตามค่าความยาก เรียงตามเนื้อหา เรียงโดยสุ่ม		
	Z	1.603	1.436	1.384
เรียงตามค่าความยาก	1.603	—	2.353*	3.085**
เรียงตามเนื้อหา	1.436		—	0.733
เรียงโดยสุ่ม	1.384			—

จากตาราง 13 แสดงให้เห็นว่า แบบทดสอบอิงเกณฑ์ที่มีการจัดเรียงข้อสอบต่างกัน 3 รูปแบบ เมื่อกำหนดคะแนนจุดตัดด้วยวิธีของเบส์ มีค่าความเชื่อมั่นของแบบทดสอบที่เรียงข้อสอบตามค่าความยาก กับแบบทดสอบที่เรียงตามเนื้อหา แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ที่ระดับ .05 และค่าความเชื่อมั่นของแบบทดสอบที่เรียงตามค่าความยาก กับแบบทดสอบที่จัดเรียงข้อสอบโดยวิธีสุ่ม แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนค่าความเชื่อมั่นระหว่างแบบทดสอบที่จัดเรียงข้อสอบตามเนื้อหา กับแบบทดสอบที่จัดเรียงข้อสอบโดยวิธีสุ่มแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

สำหรับค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม ที่คำนวณจากสูตรของลิวิงสตัน โดยใช้คะแนนจุดตัดตามวิธีของเบส์ และวิธีของแกลส ค่าความเชื่อมั่นของแบบทดสอบเท่ากันหมด คือ เท่ากับ 0.999 ดังนั้นผู้วิจัยจึงไม่ทำการทดสอบความแตกต่าง ของค่าความเชื่อมั่นที่คำนวณโดยสูตรของลิวิงสตัน

5. ทดสอบความแตกต่างของค่าความเชื่อมั่น ของแบบทดสอบอิงเกณฑ์ ที่มีการจัดเรียงข้อสอบเหมือนกัน เมื่อกำหนดคะแนนจุดตัดด้วยวิธีต่างกัน

เพื่อให้ทราบว่าค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยสุ่ม เมื่อกำหนดคะแนนจุดตัดด้วยวิธีของเบส์และวิธีของแกลส แตกต่างกันหรือไม่ ผู้วิจัยได้นำค่าความเชื่อมั่นของแบบทดสอบทั้งสามฉบับ มาแปลงเป็นคะแนนมาตรฐาน (Z) ตามวิธีของฟิชเชอร์ (Fisher's Z Tranformation) แล้วนำไปทดสอบความแตกต่างของคะแนนมาตรฐานเป็นรายคู่ ด้วยสูตรการทดสอบความแตกต่างของคะแนนมาตรฐาน (Z - test) ดังแสดงในตาราง 14

ตาราง 14 ผลการทดสอบความแตกต่างระหว่างค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ที่กำหนดคะแนนจุดตัดด้วยวิธีของเบส์ กับวิธีของแกลส ในแบบทดสอบที่มีการจัดเรียงข้อสอบเหมือนกัน เมื่อคำนวณค่าความเชื่อมั่นจากสูตรของโลเวท

ลักษณะการเรียง ของข้อสอบ	วิธีของเบส์				วิธีของแกลส				Z
	k	C	r_{cc}	z	k	C	r_{cc}	z	
เรียงตามค่าความยาก	30	13	.928	1.603	30	18	.892	1.425	2.508*
เรียงตามเนื้อหา	30	15	.899	1.436	30	16	.891	1.424	0.169
เรียงโดยวิธีสุ่ม	30	15	.885	1.384	30	16	.883	1.38	0.042

จากตาราง 14 แสดงให้เห็นว่า ค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ ที่จัดเรียงข้อสอบตามค่าความยาก ซึ่งกำหนดคะแนนจุดตัดตามวิธีของเบส์ กับวิธีของแกลส เมื่อคำนวณค่าความเชื่อมั่นโดยสูตรของโลเวท แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนค่าความเชื่อมั่นของแบบทดสอบที่จัดเรียงข้อสอบตามเนื้อหา แบบทดสอบที่จัดเรียงข้อสอบโดยวิธีสุ่มที่คำนวณจากสูตรของโลเวท และกำหนดคะแนนจุดตัดด้วยวิธีของเบส์ กับวิธีของแกลส แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

สำหรับค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ที่จัดเรียงข้อสอบตามค่าความยากเหมือนกัน เรียงตามเนื้อหาเหมือนกัน และเรียงโดยวิธีสุ่มเหมือนกัน ที่คำนวณจากสูตรของลิวิสตัน เมื่อกำหนดคะแนนจุดตัดต่างกันด้วยวิธีของเบส์กับวิธีของแกลส แล้วมีค่าความเชื่อมั่นเท่ากัน คือ 0.999 ดังนั้น ผู้วิจัยจึงไม่ทำการทดสอบความแตกต่างของค่าความเชื่อมั่นที่คำนวณโดยสูตรของลิวิสตัน

บทที่ 5

สรุปผล อภิปรายผล และข้อเสนอแนะ

ความมุ่งหมายของการศึกษาค้นคว้า

1. เพื่อเปรียบเทียบความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ที่มีการจัดเรียงข้อสอบต่างกัน และมีการกำหนดคะแนนจุดตัดด้วยวิธีเดียวกัน
2. เพื่อเปรียบเทียบความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ที่มีการจัดเรียงข้อสอบเหมือนกัน เมื่อกำหนดคะแนนจุดตัดด้วยวิธีต่างกัน

กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้า

กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้าครั้งนี้ เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2538 ของโรงเรียนสังกัดกรมสามัญศึกษา จังหวัดนครศรีธรรมราช จำนวน 1,200 คน ซึ่งได้มาโดยการสุ่มแบบแบ่งชั้น (Stratified Random Sampling) โดยใช้ขนาดโรงเรียนเป็นชั้น (Strata) และห้องเรียนเป็นหน่วยสุ่ม (Sampling Unit)

เครื่องมือที่ใช้ในการศึกษาค้นคว้า

เครื่องมือที่ใช้ในการศึกษาค้นคว้าครั้งนี้เป็นแบบทดสอบอิงเกณฑ์ วิชาคณิตศาสตร์ (ค 102) เรื่องสมการ (ตามหลักสูตรมัธยมศึกษาตอนต้น พุทธศักราช 2521 ฉบับปรับปรุง 2533 ชนิดเลือกตอบ 5 ตัวเลือก ที่มีการเรียงข้อสอบต่างกัน จำนวน 3 ฉบับ ฉบับละ 30 ข้อ คือ

- ฉบับที่ 1 จัดเรียงข้อสอบตามลำดับค่าความยากจากค่าสูงไปค่าต่ำ
- ฉบับที่ 2 จัดเรียงข้อสอบตามเนื้อหาย่อย จากเนื้อหาขั้นพื้นฐานไปยังเนื้อหาที่มีความซับซ้อนมากขึ้น
- ฉบับที่ 3 จัดเรียงข้อสอบโดยวิธีสุ่มข้อสอบจากประชากรข้อสอบ มาจัดเรียงตามลำดับที่สุ่มได้

ในการศึกษาค้างนี้ผู้วิจัยได้ดำเนินการดังนี้

1. ประสานงานกับโรงเรียนในกลุ่มตัวอย่าง เพื่อกำหนดวันเวลาในการสอบ
2. นำแบบทดสอบไปสอบกับนักเรียนกลุ่มตัวอย่างตามวันเวลาที่กำหนด โดยนำแบบทดสอบทั้งสามฉบับ มาจัดเรียงตามลำดับเป็นชุด ๆ ตั้งแต่ฉบับที่ 1 ถึงฉบับที่ 3 หมุนเวียนกันไปตลอด แล้วนำไปแจกนักเรียนในแต่ละแถว โดยเรียงตามลำดับแบบทดสอบที่เรียงไว้
3. ชี้แจงให้นักเรียนเข้าใจวิธีการทำข้อสอบ โดยเน้นให้ทำตามลำดับข้อ และขอความร่วมมือจากอาจารย์ผู้สอนชี้แจงให้นักเรียน ทำข้อสอบอย่างเต็มความสามารถ
4. นำผลการสอบมาตรวจให้คะแนน แบบ 0-1 แล้วรวมคะแนนแต่ละคน ในการสอบแต่ละครั้ง เพื่อหาคะแนนจุดตัดตามวิธีของเบส์ และวิธีของแกลส
5. นำผลที่ได้จากการสอบมาวิเคราะห์ทางสถิติ

การเสนอผลการวิเคราะห์ข้อมูล

ในการศึกษาค้างนี้ ผู้วิจัยได้เสนอผลการวิเคราะห์ข้อมูล ตามลำดับดังนี้

1. ค่าสถิติพื้นฐานของแบบทดสอบที่เรียงตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม
2. คะแนนจุดตัดของแบบทดสอบที่เรียงตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม เมื่อกำหนดคะแนนจุดตัดด้วยวิธีของเบส์ และวิธีของแกลส
3. ความเชื่อมั่นของแบบทดสอบที่เรียงตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม เมื่อคำนวณโดยใช้สูตรของโลเวท และสูตรของลิวิงสตัน
4. ทดสอบความแตกต่างของค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ ที่มีการจัดเรียงข้อสอบต่างกัน และมีการกำหนดคะแนนจุดตัดด้วยวิธีเดียวกัน
5. ทดสอบความแตกต่างของค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ ที่มีการจัดเรียงข้อสอบเหมือนกัน เมื่อกำหนดคะแนนจุดตัดด้วยวิธีต่างกัน

สรุปผลการวิเคราะห์ข้อมูล

1. ค่าสถิติพื้นฐานของแบบทดสอบอิงเกณฑ์ ที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหาและเรียงโดยสุ่ม พบว่าคะแนนเฉลี่ยของแบบทดสอบที่จัดเรียงข้อสอบต่างกัน 3 รูปแบบคือ เรียงตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม มีค่าอยู่ระหว่าง 16.599 - 18.078 โดยแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก มีค่าเฉลี่ยสูงสุด คือ 18.078 และแบบทดสอบที่จัดเรียงข้อสอบโดยวิธีสุ่มมีค่าเฉลี่ยต่ำสุด คือ 16.599 ส่วนค่าเบี่ยงเบนมาตรฐาน

ของแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม มีค่าอยู่ระหว่าง 7.019 - 7.174 โดยแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก มีค่าเบี่ยงเบนมาตรฐานสูงสุด คือ 7.174 และแบบทดสอบที่จัดเรียงตามเนื้อหา มีค่าเบี่ยงเบนมาตรฐานต่ำสุด คือ 7.019

2. คะแนนจุดตัดของแบบทดสอบอิงเกณฑ์ ที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม เมื่อกำหนดคะแนนจุดตัดด้วยวิธีของเบส มีค่าอยู่ระหว่าง 13 - 15 คะแนน โดยแบบทดสอบที่จัดเรียงข้อสอบตามเนื้อหา และ แบบทดสอบที่จัดเรียงข้อสอบโดยวิธีสุ่มมีคะแนนจุดตัดเท่ากัน คือ 15 คะแนน แบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก มีคะแนนจุดตัดต่ำสุด คือ 13 คะแนน ส่วนแบบทดสอบอิงเกณฑ์ที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม เมื่อกำหนดคะแนนจุดตัดด้วยวิธีของแกลส มีคะแนนอยู่ระหว่าง 16 - 18 คะแนน ซึ่งแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก มีคะแนนจุดตัดสูงสุด คือ 18 คะแนน และแบบทดสอบที่จัดเรียงข้อสอบตามเนื้อหากับแบบทดสอบที่จัดเรียงข้อสอบโดยวิธีสุ่ม มีคะแนนจุดตัดเท่ากัน คือ 16 คะแนน

3. ความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่มที่คำนวณคะแนนจุดตัดตามวิธีของเบส โดยใช้สูตรความเชื่อมั่นของโลเวท มีค่าอยู่ระหว่าง 0.885 - 0.928 ซึ่งแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยากมีค่าความเชื่อมั่นสูงสุด คือ 0.928 และแบบทดสอบที่จัดเรียงข้อสอบโดยสุ่มมีค่าความเชื่อมั่นต่ำสุดคือ 0.885 ค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม ที่คำนวณจากคะแนนจุดตัดของแกลส โดยใช้สูตรความเชื่อมั่นของโลเวท มีค่าอยู่ระหว่าง 0.883 - 0.892 ซึ่งแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก มีค่าความเชื่อมั่นสูงสุด คือ 0.892 และแบบทดสอบที่จัดเรียงโดยสุ่ม มีค่าความเชื่อมั่นต่ำสุด คือ 0.883 และค่าความคลาดเคลื่อนมาตรฐานในการวัดของแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยสุ่ม ที่คำนวณค่าความเชื่อมั่นจากคะแนนจุดตัดของเบส มีค่าอยู่ระหว่าง 1.925 ถึง 2.386 ซึ่งแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก มีค่าความคลาดเคลื่อนมาตรฐานต่ำสุด คือ 1.925 และแบบทดสอบที่จัดเรียงโดยสุ่มมีค่าความคลาดเคลื่อนมาตรฐานในการวัดสูงสุด คือ 2.386 ส่วนค่าความคลาดเคลื่อนมาตรฐานในการวัด ที่คำนวณค่าความเชื่อมั่นจากคะแนนจุดตัดตามวิธีของแกลส มีค่าอยู่ระหว่าง 2.317 ถึง 2.407 แบบทดสอบที่จัดเรียงข้อสอบตามเนื้อหามีความคลาดเคลื่อนมาตรฐานในการวัดต่ำสุด คือ 2.317 และแบบทดสอบที่จัดเรียงข้อสอบโดยวิธีสุ่มมีค่าความคลาดเคลื่อนมาตรฐานในการวัดสูงสุด คือ 2.407

สำหรับค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม เมื่อกำหนดคะแนนจุดตัดด้วยวิธีของเบสและวิธีของแกลส โดยคำนวณค่าความเชื่อมั่นด้วยสูตรของลิวิสตัน ได้ค่าความเชื่อมั่นเท่ากันทั้ง 3 ฉบับ คือเท่ากับ

0.999 และมีค่าความคลาดเคลื่อนมาตรฐานในการวัดอยู่ระหว่าง 0.221 ถึง 0.227 โดยแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยากมีค่าความคลาดเคลื่อนมาตรฐานในการวัดสูงสุดคือ 0.227 และแบบทดสอบที่จัดเรียงข้อสอบตามเนื้อหา กับแบบทดสอบที่จัดเรียงข้อสอบโดยวิธีสุ่มมีค่าความคลาดเคลื่อนมาตรฐานใกล้เคียงกันคือ 0.221 และ 0.222

4. จากการทดสอบความแตกต่างของค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ ที่จัดเรียงข้อสอบต่างกัน 3 รูปแบบ คือเรียงตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม เมื่อกำหนดคะแนนจุดตัดตามวิธีของเบส์ และคำนวณความเชื่อมั่นจากสูตรของโลเวท พบว่าแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก กับแบบทดสอบที่จัดเรียงข้อสอบตามเนื้อหา ค่าความเชื่อมั่นแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ที่ระดับ .05 และค่าความเชื่อมั่นของแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก กับแบบทดสอบที่จัดเรียงข้อสอบโดยวิธีสุ่ม แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนค่าความเชื่อมั่นของแบบทดสอบที่จัดเรียงข้อสอบตามเนื้อหา กับแบบทดสอบที่จัดเรียงข้อสอบโดยวิธีสุ่มแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ สำหรับค่าความเชื่อมั่นของแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม ที่กำหนดคะแนนจุดตัดด้วยวิธีของแกลส และคำนวณจากสูตรของโลเวท พบว่าไม่มีคูใดแตกต่างกัน ส่วนค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ ที่มีการจัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม ที่คำนวณจากสูตรของลิวิสตัน เมื่อกำหนดคะแนนจุดตัดทั้งวิธีของเบส์ และวิธีของแกลส มีค่าเท่ากัน

5. จากการทดสอบความแตกต่างระหว่างค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ ที่กำหนดคะแนนจุดตัดด้วยวิธีของเบส์ กับวิธีของแกลส ในแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม เมื่อคำนวณความเชื่อมั่นจากสูตรของโลเวท พบว่า แบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก ระหว่างความเชื่อมั่นที่กำหนดคะแนนจุดตัดด้วยวิธีของเบส์กับวิธีของแกลสมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนในแบบทดสอบที่จัดเรียงข้อสอบตามเนื้อหา แบบทดสอบที่จัดเรียงข้อสอบโดยวิธีสุ่ม ระหว่างความเชื่อมั่นที่กำหนดคะแนนจุดตัดด้วยวิธีของเบส์กับวิธีของแกลสมีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

สำหรับค่าความเชื่อมั่นที่คำนวณจากสูตรของลิวิสตัน พบว่า แบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม ในแต่ละแบบ จะมีค่าความเชื่อมั่นที่คำนวณจากคะแนนจุดตัดด้วยวิธีของเบส์กับวิธีของแกลส มีค่าเท่ากัน

อภิปรายผล

1. ความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ที่กำหนดคะแนนจุดตัดด้วยวิธีของเบส์ เมื่อคำนวณด้วยสูตรของโลเวท ที่จัดเรียงแบบทดสอบตามค่าความยาก กับแบบทดสอบที่จัดเรียงข้อสอบตามเนื้อหา มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก กับแบบทดสอบที่จัดเรียงข้อสอบโดยวิธีสุ่ม มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ซึ่งสอดคล้องกับสมมุติฐานที่ตั้งไว้ และสอดคล้องกับ ศศิธร สุวรรณสุข (2519 : 90-92) ซึ่งพบว่าแบบทดสอบที่เรียงจากข้อง่ายไปหาข้อยาก เรียงจากข้อยากไปหาข้อง่าย เรียงตามเนื้อหาวิชา และเรียงจากข้อง่ายไปหาข้อยากในแต่ละหัวข้อเนื้อหาวิชา ทำให้ค่าความเชื่อมั่นของแบบทดสอบวัดทักษะคณิตศาสตร์เปลี่ยนแปลงอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และสอดคล้องกับ สิริพร พุ่มแก้ว (2521 : 30-36) ที่ศึกษาการจัดเรียงข้อคำถามต่างกัน 4 แบบคือ เรียงจากง่ายไปหายาก เรียงจากยากไปหาง่าย เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม ทำให้คะแนนการสอบวิชาคณิตศาสตร์ ที่จัดเรียงข้อสอบจากง่ายไปหายาก แตกต่างจากแบบทดสอบที่เรียงข้อสอบจากยากไปหาง่าย และแบบทดสอบที่เรียงข้อสอบโดยวิธีสุ่ม ฉะนั้นเมื่อคะแนนการสอบมีค่าแตกต่างกัน ค่าความเชื่อมั่นของแบบทดสอบย่อมแตกต่างกันด้วย นอกจากนี้ยังสอดคล้องกับ กาญจนา วัฒนสุนทร (2521 : 75 - 76) ซึ่งพบว่า คะแนนจุดตัดมีผลต่อความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ คือ เมื่อคะแนนจุดตัดสูงความเชื่อมั่นจะต่ำ และเมื่อคะแนนจุดตัดต่ำค่าความเชื่อมั่นจะสูง เนื่องจากคะแนนจุดตัดของแบบทดสอบที่กำหนดคะแนนจุดตัดด้วยวิธีของเบส์ แบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก มีคะแนนจุดตัดต่ำสุดเป็นผลให้ค่าความเชื่อมั่นมีค่าสูงสุด และทำให้ค่าความเชื่อมั่นแตกต่างจากฉบับอื่น นอกจากนี้แบบทดสอบที่เรียงตามค่าความยาก มีค่าความแปรปรวนมากกว่าแบบทดสอบฉบับอื่น ซึ่งสอดคล้องกับ กรอนลัน (Gronlund, 1981 : 119) ที่ว่า ค่าความเชื่อมั่นขึ้นอยู่กับความแปรปรวนระหว่างคะแนนผลการสอบ ถ้าความแปรปรวนระหว่างคะแนนผลการสอบมีน้อยทำให้ความเชื่อมั่นต่ำกว่าคะแนนผลการสอบที่มีความแปรปรวนมาก ส่วนแบบทดสอบที่จัดเรียงข้อสอบตามเนื้อหา กับแบบทดสอบที่จัดเรียงข้อสอบโดยวิธีสุ่ม มีค่าความเชื่อมั่นไม่แตกต่างกัน ทั้งนี้เนื่องจากแบบทดสอบที่จัดเรียงข้อสอบตามเนื้อหา กับแบบทดสอบที่จัดเรียงข้อสอบโดยวิธีสุ่มมีคะแนนจุดตัดเท่ากัน และมีความแปรปรวนใกล้เคียงกัน เป็นผลให้ค่าความเชื่อมั่นไม่แตกต่างกัน ดังนั้นเมื่อพิจารณาค่าความเชื่อมั่นที่กำหนดคะแนนจุดตัดด้วยวิธีของเบส์แล้ว แบบทดสอบที่จัดเรียงข้อสอบตามค่าความยากดีกว่าแบบทดสอบที่จัดเรียงข้อสอบตามเนื้อหา และแบบทดสอบที่จัดเรียงข้อสอบโดยวิธีสุ่ม เพราะว่ามีค่าความเชื่อมั่นสูงกว่าการจัดเรียงแบบทดสอบรูปแบบอื่น

ส่วนแบบทดสอบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม ที่กำหนดคะแนนจุดตัดด้วยวิธีของเกลส และคำนวณความเชื่อมั่นจากสูตรของโลเวทนั้น มีค่าความเชื่อมั่นแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ซึ่งไม่สอดคล้องกับสมมุติฐาน

ที่ตั้งไว้ทั้งนี้เนื่องจากแบบทดสอบที่เรียงตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม เมื่อกำหนดคะแนนจุดตัดด้วยวิธีของเกลส คะแนนจุดตัดมีค่าใกล้เคียงกัน ซึ่ง สับโคเวียค (Subkoviak . 1976 : 265 - 276) พบว่า คะแนนจุดตัดมีผลต่อความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ ดังนั้นเมื่อคะแนนจุดตัดมีค่าใกล้เคียงกัน จึงเป็นผลให้ค่าความเชื่อมั่นไม่แตกต่างกัน

สำหรับความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ที่คำนวณจากสูตรของลิวิสตัน เมื่อจัดเรียงข้อตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม และกำหนดคะแนนจุดตัดทั้งของวิธีเบส์ และวิธีเกลส มีค่าไม่แตกต่างกัน เพราะว่าการคำนวณค่าความเชื่อมั่นโดยใช้สูตรของลิวิสตัน คำนวณจากค่าความแปรปรวนของแบบทดสอบและเนื่องจากแบบทดสอบที่เรียงตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม ค่าความแปรปรวนของแบบทดสอบมีค่าใกล้เคียงกัน เป็นผลให้ค่าความเชื่อมั่นที่ได้ไม่แตกต่างกัน

2. ความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ที่คำนวณจากสูตรของโลเวท ในแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก เมื่อกำหนดคะแนนจุดตัดด้วยวิธีของเบส์ กับวิธีของเกลส มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งสอดคล้องกับสมมติฐานที่ตั้งไว้ ทั้งนี้เป็นเพราะว่าแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก เมื่อกำหนดคะแนนจุดตัดด้วยวิธีของเบส์ กับวิธีของเกลส คะแนนจุดตัดมีค่าต่างกัน ซึ่ง สับโคเวียค (Subkoviak . 1976 : 265 - 276) พบว่าคะแนนจุดตัดมีผลต่อความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ ดังนั้น เมื่อคะแนนจุดตัดมีค่าต่างกัน เป็นผลให้ค่าความเชื่อมั่นแตกต่างกันด้วย ส่วนแบบทดสอบที่จัดเรียงข้อสอบตามเนื้อหา และแบบทดสอบที่จัดเรียงข้อสอบโดยวิธีสุ่ม ในแต่ละฉบับเมื่อกำหนดคะแนนจุดตัดด้วยวิธีของเบส์ กับวิธีของเกลส มีค่าความเชื่อมั่นแตกต่างกันอย่างไรมีนัยสำคัญทางสถิติ ซึ่งไม่สอดคล้องกับสมมติฐานที่ตั้งไว้ ทั้งนี้เนื่องจาก แบบทดสอบที่จัดเรียงข้อสอบตามเนื้อหา แบบทดสอบที่จัดเรียงข้อสอบโดยวิธีสุ่ม แต่ละฉบับเมื่อกำหนดคะแนนจุดตัดด้วยวิธีของเบส์ กับวิธีของเกลส คะแนนจุดตัดมีค่าใกล้เคียงกัน จึงเป็นผลให้ค่าความเชื่อมั่นไม่แตกต่างกัน

สำหรับค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ ที่คำนวณจากสูตรของลิวิสตัน ในแบบทดสอบที่จัดเรียงข้อสอบแบบเดียวกันไม่ว่าจะจัดเรียงตามค่าความยาก จัดเรียงข้อสอบตามเนื้อหา หรือจัดเรียงข้อสอบโดยวิธีสุ่ม ในเมื่อกำหนดคะแนนจุดตัดด้วยวิธีของเบส์ กับวิธีของเกลส มีค่าความเชื่อมั่นไม่แตกต่างกันทุกแบบการจัดเรียง ซึ่งไม่สอดคล้องกับสมมติฐานที่ตั้งไว้ ทั้งนี้เนื่องจาก ค่าความเชื่อมั่นที่คำนวณจากสูตรของลิวิสตัน คำนวณจากค่าความแปรปรวนของคะแนนการสอบ คะแนนเฉลี่ยของแบบทดสอบ และค่าความเชื่อมั่นของแบบทดสอบที่หาจากสูตร $KR.20^*$ ซึ่งในแบบทดสอบแต่ละฉบับแม้ว่าคะแนนจุดตัดต่างกัน แต่ค่าความแปรปรวน คะแนนเฉลี่ย และค่าความเชื่อมั่นของแบบทดสอบที่คำนวณจากสูตร $KR.20$ เป็นค่าเดียวกัน จึงเป็นผลให้ค่าความเชื่อมั่นของแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม แต่ละฉบับเมื่อกำหนดคะแนนจุดตัดด้วยวิธีของเบส์ กับวิธีของเกลส ที่คำนวณจากสูตรของลิวิสตันจึงไม่แตกต่างกัน

ข้อเสนอแนะในการวิจัย

ข้อเสนอแนะในการทำวิจัย

1. ควรมีการศึกษาคะแนนจุดตัดของแบบทดสอบอิงเกณฑ์ที่มีการจัดเรียงข้อสอบตามค่าความยากง่าย เรียงตามเนื้อหา และเรียงแบบสุ่มในวิชาอื่น ๆ และในระดับชั้นอื่น ๆ และใช้วิธีการหาคะแนนจุดตัดด้วยวิธีอื่น เช่นวิธีการที่ประยุกต์ทฤษฎีตอบสนองข้อคำถามมาใช้ .
2. ควรมีการศึกษาเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ที่มีการจัดเรียงข้อสอบตามค่าความยากง่าย เรียงตามเนื้อหาย่อย และเรียงโดยวิธีสุ่ม โดยใช้สูตรการคำนวณค่าความเชื่อมั่นสูตรอื่น ๆ
3. ควรมีการศึกษาคะแนนจุดตัดของแบบทดสอบอิงเกณฑ์ที่มีการจัดเรียงข้อสอบต่างกัน และใช้วิธีการกำหนดคะแนนจุดตัดต่างกันในระดับความสามารถทางการเรียนของนักเรียนระดับต่าง ๆ

ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1. ในการสร้างแบบทดสอบอิงเกณฑ์ ควรจัดเรียงข้อสอบตามลำดับค่าความยาก ซึ่งจากผลการวิจัยแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก มีค่าความเชื่อมั่นสูงกว่าการจัดเรียงข้อสอบแบบอื่น
2. การหาคะแนนจุดตัดของแบบทดสอบอิงเกณฑ์ระหว่างวิธีของเบส์กับวิธีของแกลส จะเลือกใช้วิธีใดก็ได้ ทั้งนี้เนื่องจากการหาคะแนนจุดตัดด้วยวิธีของเบส์กับวิธีของแกลส ได้ค่าความเชื่อมั่นของแบบทดสอบไม่แตกต่างกัน

บรรณานุกรม

บรรณานุกรม

- กมล ภูประเสริฐ. "การเปรียบเทียบการวัดผลแบบอิงกลุ่มและอิงเกณฑ์," พัฒนาวัดผล 1 :
กรกฎาคม 77 - 78 , กรกฎาคม 2528.
- กัญญา ลินทรัตนศิริกุล. การประยุกต์ทฤษฎีการตอบสนองต่อข้อสอบเมื่อกำหนดคะแนนจุดตัด
ตามเทคนิคของนีเดลสกับวิธีใช้กลุ่มควบเส้น. ปรินูญานินพนธ์ กศ.ด. กรุงเทพฯ :
มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2534. อัดสำเนา.
- โกวิท ประวาลพุกษ์. "การประเมินผลแบบอิงเกณฑ์," เอกสารประกอบคำบรรยายในการอบรม
วิทยากรจังหวัด (ระดับมัธยมศึกษาตอนต้น) : 113 - 130.
- _____. "การวัดผลแนวใหม่," คู่มือครูประถมศึกษา. มูลนิธิพัฒนาเยาวชน 2
- _____. "การทดสอบอิงเกณฑ์," วารสารวัดผลการศึกษา. 1(3) : 16 - 24 ; มกราคม -
เมษายน 2523.
- ชวาล แพรัตกุล. การทดสอบเพื่อค้นและพัฒนาสมรรถภาพ. สำนักงานทดสอบทางการศึกษา
และจิตวิทยา คณะวิจัยการศึกษา วิทยาลัยการศึกษาประสานมิตร, 2524.
- _____. เทคนิคการวัดผล. พิมพ์ครั้งที่ 5. กรุงเทพฯ : วัฒนาพานิช, 2516.
- บุญเชิด ภิญโญนันตพงษ์. การทดสอบแบบอิงเกณฑ์: แนวคิดและวิธีการ. พิมพ์ครั้งที่ 1.
ภาคพื้นฐานการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ
ประสานมิตร, 2526.
- ประนอม พันธุ์ไสว. การศึกษาคะแนนจุดตัดของแบบทดสอบอิงเกณฑ์ประเภทเลือกตอบ
ตอบสั้นและถูกผิด วิชาวิทยาศาสตร์โดยใช้ทฤษฎีการตัดสินใจของเบส์และวิธีของเกลส.
ปรินูญานินพนธ์ กศ.ม. กรุงเทพฯ : มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร,
2530. อัดสำเนา.
- พิศิษฐ์ ตันทวนิช. การประยุกต์เทคนิคการวิเคราะห์ตามลำดับขั้นเพื่อกำหนดความยาวของ
แบบทดสอบอิงเกณฑ์. วิทยานิพนธ์ ค.ด. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย,
2528. อัดสำเนา.
- ไพศาล หวังพานิช. การวัดผลการศึกษา. พิมพ์ครั้งที่ 1. ไทยวัฒนาพานิช, 2526.
- ล้วน สายยศ และ อังคณา สายยศ. เอกสารประกอบการอบรม. ม.ป.ป. : 320.
- _____. เทคนิคการวิจัยทางการศึกษา. พิมพ์ครั้งที่ 3 ศูนย์ส่งเสริมวิชาการ, 2536.
- วุฒิคุณ เสาวรัตน์. การสร้างแบบทดสอบอิงเกณฑ์วิชาคณิตศาสตร์ เรื่องภาคตัดกรวย
ชั้นมัธยมศึกษาปีที่ 4. ปรินูญานินพนธ์ กศ.ม. มหาวิทยาลัย
ศรีนครินทรวิโรฒ ประสานมิตร, 2523. อัดสำเนา.

- ศศิธร สุวรรณสุข. การศึกษาค่าความเที่ยงตรง ค่าความเชื่อมั่น ของแบบทดสอบที่เรียงลำดับ
ค่าความต่าง ๆ กัน ในวิชาคณิตศาสตร์ ระดับชั้นประถมศึกษาปีที่ 7. ปรินญาณพนธ์
กศ.ม. มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร 2519, 78 หน้า. อัดสำเนา.
ศึกษาธิการ, กระทรวง. คู่มือการประเมินผลการเรียนตามหลักสูตรมัธยมศึกษาตอนต้น
พุทธศักราช 2521 (ฉบับปรับปรุง พ.ศ.2533). กรุงเทพฯ : โรงพิมพ์คุรุสภาลาดพร้าว,
2533.
- สงบ ลักษณะ. "ปัญหามาตรฐานและเกณฑ์," วารสารวารวัดผลการศึกษา. 1(2) : 48 - 55 ;
กันยายน 2522.
- _____. "การตรวจสอบความเที่ยงตรงของแบบทดสอบอิงเกณฑ์,"
วารสารวารวัดผลการศึกษา. 1 (3) : 36 - 45 ; มกราคม - เมษายน 2523.
- สิริพร พุ่มแก้ว. ผลของการเรียงลำดับข้อสอบต่อค่าความเที่ยงค่าความตรงของแบบทดสอบและ
ผลสัมฤทธิ์ทางการเรียน. วิทยานิพนธ์ ค.ม. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย
2521, 65 หน้า. อัดสำเนา.
- สุทธีวรรณ พิศศักดิ์โสภณ. การศึกษาคะแนนจุดตัดของแบบทดสอบอิงเกณฑ์ชนิด
เลือกตอบซ้อน. ปรินญาณพนธ์ กศ.ม. กรุงเทพฯ : มหาวิทยาลัยศรีนครินทรวิโรฒ
ประสานมิตร, 2529. อัดสำเนา.
- อนันต์ ศรีโสภณ. การวัดผลการศึกษา. พิมพ์ครั้งที่ 3 ไทยวัฒนาพานิช 2524.
- _____. ทฤษฎีการวัดและการทดสอบ. กรุงเทพฯ : ไทยวัฒนาพานิช, 2525.
- อังคณา สายยศ. "การกำหนดคะแนนจุดตัดของแบบทดสอบอิงเกณฑ์,"
วารสารวารวัดผลการศึกษา. 3(3) : 70 - 84 ; มกราคม-เมษายน 2525.
- _____. "การเขียนข้อสอบอิงเกณฑ์," วารสารวารวัดผลการศึกษา. 4(3) : 25 - 36 ;
มกราคม - เมษายน 2526.
- _____. "การพิจารณาข้อคำถามของแบบทดสอบ," วารสารวิจัยและวัดผล 27 .
4(4) : 183 - 186, 2527.
- _____. เอกสารประกอบการอบรมเรื่องการวัดผลแบบอิงเกณฑ์(วัดผล8).
คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2533.
- _____. "ความเชื่อมั่นของแบบทดสอบอิงเกณฑ์," วารสารวารวัดผลการศึกษา.
14(42) : 40 - 51 ; มกราคม - เมษายน 2536.
- _____. "การวิเคราะห์ข้อสอบของแบบทดสอบวัดผลสัมฤทธิ์ในเชิงปฏิบัติ :
อิงกลุ่ม, อิงเกณฑ์, อัตนัย และ ICC" วารสารวารวัดผลการศึกษา.
14(43) : พฤษภาคม - สิงหาคม 2536.

อัจฉรา ปราบอริฟาย. ผลการใช้สารสนเทศแก่ผู้เกี่ยวข้องต่อการกำหนดคะแนนจุดตัดของ
แบบทดสอบผลสัมฤทธิ์วิชาคณิตศาสตร์. วิทยานิพนธ์ ค.ม. กรุงเทพฯ : จุฬาลงกรณ์
มหาวิทยาลัย, 2532. อัดสำเนา.

Airasian, P.W. and G.F.Maduas. "Criterion References Testing in
Classroom," *Measerement in Education*. 15 : 1 - 8.

Alkin, M.C. "Criterion - Referenced Maesurement and Other Such Term,"
In Problems in Criterion - Referenced Measurement. Losageles :
University of California ; 1974.

Berk, R.A. "A Consumer's Guide to setting Performance Standards on
Criterion - Reference Test," Review of Education Research.
56(1) : 137 - 172 ; Spring, 1986.

_____. Criterion-referenced Measurement : The State of The Art.
Maryland : The Johns Hopkins University Press, 1980.

Crehen, K.D. "Item Analysis for Teacher - Made for Mastery Test,"
Journal of Education Maesurement. 11: 255 - 262 ; Winter, 1974.

Cross and other. "A Comparison of the Three Methods for Establishing
Minimum Standards on the National Teacher,Examination,"
Journal of Educational Measurement. 21(20) : 113 - 129 ;
sumer, 1984.

Ebell Robert L. Measuring Educational Achivement. Pentice - hall,
Inc, New Jersy, 1965.

Ferguson, Gorge A. Statistical Analysis in Psychology and Education. 3rd.
New York : MC Graw-hill Book Company, 1966.

- Glaser, R. and A.J. Nitko. "Measurement in Learning and Instruction," in R.L. Thorndike (ed) Journal of Educational Measurement. 8 : 653 ; winter, 1971.
- Glass, G.V. "Standards and Criteria" Journal of Educational Measurement. 15(4) : 237 - 261 ; Winter, 1978.
- Gronlund, Norman E. Measurement and evaluation in Testing. 4th ed. New York : Macmill, 1981.
- Gronlund, N.E. Preparing Criterion - Referenced Test for Classroom Instruction. New York : The Macmillan Company, 1973.
- Hambleton, R.K. and Novick, M.R. " Toward and Integration of Theory and Method for Criterion - Referenced Test," Journal of Educational measurement. 10, 10 : 159 - 170 ; Fall, 1973.
- Hambleton, R.K. and Others. "Criterion - Referenced Testing and Measurement : A Review of Technical Issues and Developments," Review of Education Research. 48(1) : 1 - 47 ; winter, 1978.
- Halpin, G., G. Sigmon and G. Halpin. "Minimum Competency Standards Set by Three Divergent Groups of Raters Using Three Judgmental Procedures : Implication for Validity," Educational and Psychological Measurement. 43(1) : 185 - 197 ; Spring, 1983.
- Huynh, H. "On the Reliability of Decision in Criterion - Referenced Testing," Journal of Educational Measurement. 13(4) : 253 - 263 ; Winter, 1976.
- Linn, R.L. "Demand, Caution, and Suggestion Setting Standard," Journal of Education Measurement. 24(4) : 333 - 345 ; Winter, 1978.

- Linn, R.L. "Demand, Caution, and Suggestion Setting Standard," Journal of Education Measurement. 24(4) : 333 - 345 ; Winter, 1978.
- Linda Crocker, James Algina. Introduction To Classical and Modern Test Theory. New York, 1986.
- Lindquist, E. F., Educational Measurement. American Council on Education, Washington, 1955.
- Livingston, S.A. "A Criterion - Referenced Application of Classical Test Theory," Journal of Educational Measurement. 9 : 13 - 26, 1972.
- Livingston, S.A. and M.J. Zieky. Passing Scores : A Manual for Setting Standards of Performance on Educational and Occupational Tests. New Jersey : Educational Testing Service, 1982.
- Mehrens, W.A. and I.J. Lehmann. Measurement and Evaluation in Education and Psychology. New York : Holt, Rinehart and Winston, 1975.
- Mills, C.N. "A Comparison of Three Methods of establishing cut-off scores on Criterion-Referenced Tests," Journal of Educational measurement. 20(3) 283 - 292 ; Fall, 1983.
- Millman, J. "Passing Scores and Test Length for Domain - Reference Measure," Review of Education Research. 43 : 205 - 215 ; Spring, 1973.
- _____. "Criterion - Reference Measurement," Evaluation in Education : Current Application. Berkeley, California. McClellan Publishing Company, 1974.

- Norcini, J.J. and Others. "A Comparison of Three Variations on a Standard-Setting Methods," Journal of Educational Measurement. 24(1) : 56 - 64 ; Spring, 1987.
- Plake, B.S. and G.J. Melican. "Effects of Item Context on Intrajudge Consistency of Expert Judgments Via the Nedelsky Standard Setting Method," Educational and Psychological Measurement. 49(1) : 45 - 51 ; Spring, 1989.
- Popham, w.J. Criterion - Referenced Measurement. Englewood Cliffs, New Jersey : prentice - Hall, Inc. 1978.
- Roid, G.H. and T. Haladyna. A Technology For Test- Item Writing. New Jersey : Acadamic Press, 1983.
- Ronald A. Berk. Criterion - Referenced Measurement : The State of the Art. 1980.
- Stanley, Julian c. and Kenncth D. Hopkins. Education and Psychology Evaluation. 5th.ed. Englewoods Cliff : Prefic Hall ,1972.
- SubKoviak, M.J. "Estimating Reliability a Single Administration of Criterion - Referenced Test ," Journal of Education Measurement. 13(4) : 265 - 276 ; Winter, 1976.
- Swaminathan, H., R.k. Hambleton and J. Algina. "A Bayesian Decistion - Theoretic Proccedure for Use with Criterion-Referenced Test," Journal of Educational Maesurement. 12(2) : 87-98; Summer,1975.
- Traver, Robert H.W. How to make Achievment Test. Odyssey Press, New York, 1950.

Wert, James E. and Others. Statistical Methods in Education and Psychological Research. 1954.

Wilson, James W. Evaluation of Learning in Secondary School Mathematics, In Benjamin S. Bloom, J. Thomas Hastings, George F. Madaus. Handbook on Formative and Sumative Evaluation of student learning. New York : McGraw-Hill Book Company, 1971.

Zieky, M. Method of Setting Standards of Performance on Criterion Referenced Tests. Paper Presented at the 13 th International Conference of International Association for Educational Assessment, Bangkok, Thailand, November, 1987.

ภาคผนวก

ภาคผนวก ก

ผลการวิเคราะห์ข้อสอบรายข้อ

ตาราง 15 ค่าความยากง่าย (P) และค่าอำนาจจำแนก (B) รายข้อของแบบทดสอบก่อน
การเก็บรวบรวมข้อมูล

ข้อ	ตัว เลือก	P	B	ผลการ วิเคราะห์	ข้อ	ตัว เลือก	P	B	ผลการ วิเคราะห์
1	ก	.080	.097	ค่อนข้างยาก จำแนกไม่ได้ คัดออก	6	ก	.130	.092	ค่อนข้างยาก จำแนกได้ดี คัดเลือกไว้
	ข	.124	.109			ข	.056	.005	
	ค	.420	.246			ค	.278	.302	
	ง	.142	.138			ง	.148	.147	
	จ	.235	.098			จ	.389	.059	
2	ก	.198	.021	ค่อนข้างยาก จำแนกไม่ได้ คัดออก	7	ก	.173	.131	ค่อนข้างง่าย จำแนกได้ดี คัดเลือกไว้
	ข	.049	.022			ข	.648	.245	
	ค	.185	.039			ค	.093	.035	
	ง	.345	.211			ง	.043	.040	
	จ	.117	.100			จ	.043	.040	
3	ก	.315	.026	ยากปานกลาง จำแนกไม่ได้ คัดออก	8	ก	.870	.200	ค่อนข้างง่าย จำแนกได้ดี คัดเลือกไว้
	ข	.049	.049			ข	.031	.048	
	ค	.043	.039			ค	.037	.057	
	ง	.056	.005			ง	.031	.048	
	จ	.537	.119			จ	.031	.048	
4	ก	.006	.009	ง่ายมาก จำแนกไม่ได้ คัดออก	9	ก	.117	.100	ค่อนข้างง่าย จำแนกไม่ได้ คัดออก
	ข	.019	.029			ข	.037	.030	
	ค	.877	.100			ค	.056	.032	
	ง	.037	.057			ง	.734	.183	
	จ	.025	.495			จ	.056	.032	
5	ก	.043	.067	ค่อนข้างง่าย จำแนกได้ดี คัดเลือกไว้	10	ก	.296	.024	ค่อนข้างยาก จำแนกไม่ได้ คัดออก
	ข	.846	.238			ข	.136	.101	
	ค	.037	.058			ค	.253	.151	
	ง	.043	.067			ง	.074	.169	
	จ	.031	.048			จ	.241	.008	

ตาราง 15 (ต่อ)

ข้อ	ตัว เลือก	P	B	ผลการ วิเคราะห์	ข้อ	ตัว เลือก	P	B	ผลการ วิเคราะห์
11	ก	.216	.073	ค่อนข้างยาก จำแนกไม่ได้ คัดออก	16	ก	.309	.178	ค่อนข้างยาก จำแนกไม่ได้ คัดออก
	ข	.130	.065			ข	.204	.017	
	ค	.216	.019			ค	.111	.045	
	ง	.220	.001			ง	.148	.039	
	จ	.228	.028			จ	.228	.189	
12	ก	.154	.130	ยากปานกลาง จำแนกได้ดี คัดเลือกไว้	17	ก	.117	.127	ยากมาก จำแนกไม่ได้ คัดออก
	ข	.519	.580			ข	.303	.223	
	ค	.117	.314			ค	.253	.259	
	ง	.130	.173			ง	.315	.110	
	จ	.080	.097			จ	.012	.019	
13	ก	.284	.140	ยากมาก จำแนกไม่ได้ คัดเลือกไว้	18	ก	.253	.066	ค่อนข้างยาก จำแนกได้ คัดเลือกไว้
	ข	.303	.026			ข	.247	.323	
	ค	.142	.055			ค	.086	.025	
	ง	.086	.031			ง	.204	.125	
	จ	.180	.022			จ	.210	.107	
14	ก	.198	.034	ยากปานกลาง จำแนกไม่ได้ คัดออก	19	ก	.284	.059	ยากปานกลาง จำแนกไม่ได้ คัดออก
	ข	.105	.027			ข	.475	.169	
	ค	.031	.048			ค	.056	.005	
	ง	.167	.014			ง	.031	.048	
	จ	.500	.095			จ	.154	.049	
15	ก	.161	.166	ยากปานกลาง จำแนกได้ดี คัดเลือกไว้	20	ก	.426	.047	ค่อนข้างยาก จำแนกไม่ได้ คัดออก
	ข	.253	.174			ข	.105	.054	
	ค	.512	.373			ค	.148	.085	
	ง	.056	.032			ง	.216	.073	
	จ	.019	.002			จ	.1049	.005	

ตาราง 15 (ต่อ)

ข้อ	ตัว เลือก	P	B	ผลการ วิเคราะห์	ข้อ	ตัว เลือก	P	B	ผลการ วิเคราะห์
21	ก	.056	.059	ยากมาก จำแนกกลับ คัดออก	26	ก	.031	.021	ยากปานกลาง จำแนกได้ คัดเลือกไว้
	ข	.154	.044			ข	.117	.063	
	(ค)	.154	.073			(ค)	.457	.243	
	ง	.198	.059			ง	.056	.032	
	จ	.426	.116			จ	.340	.253	
22	ก	.037	.030	ค่อนข้างยาก จำแนกไม่ได้ คัดออก	27	ก	.043	.067	ค่อนข้างง่าย จำแนกได้ คัดเลือกไว้
	ข	.086	.025			ข	.037	.057	
	ค	.272	.175			(ค)	.790	.279	
	(ง)	.358	.107			ง	.062	.095	
	จ	.247	.133			จ	.068	.051	
23	ก	.043	.040	ค่อนข้างง่าย จำแนกได้ดี คัดเลือกไว้	28	ก	.099	.125	ค่อนข้างง่าย จำแนกได้ คัดเลือกไว้
	ข	.148	.066			(ข)	.747	.228	
	ค	.080	.097			ค	.093	.089	
	(ง)	.685	.261			ง	.019	.029	
	จ	.043	.040			จ	.037	.058	
24	ก	.086	.106	ค่อนข้างง่าย จำแนกได้ คัดเลือกไว้	29	ก	.044	.067	ค่อนข้างง่าย จำแนกได้ คัดเลือกไว้
	ข	.056	.086			(ข)	.826	.205	
	(ค)	.728	.230			ค	.074	.033	
	ง	.031	.048			ง	.019	.029	
	จ	.099	.010			จ	.037	.058	
25	ก	.105	.135	ยากปานกลาง จำแนกได้ดี คัดเลือกไว้	30	ก	.025	.038	ค่อนข้างง่าย จำแนกไม่ได้ คัดออก
	(ข)	.599	.530			(ข)	.735	.166	
	ค	.043	.067			ค	.074	.033	
	ง	.062	.095			ง	.056	.032	
	จ	.191	.214			จ	.111	.063	

ตาราง 15 (ต่อ)

ข้อ	ตัว เลือก	P	B	ผลการ วิเคราะห์	ข้อ	ตัว เลือก	P	B	ผลการ วิเคราะห์
31	ก	.242	.182	ยากปานกลาง จำแนกได้ดี คัดเลือกไว้	36	ก	.179	.114	ยากปานกลาง จำแนกไม่ได้ คัดออก
	ข	.106	.081			ข	.444	.126	
	ค	.143	.155			ค	.111	.018	
	ง	.416	.491			ง	.056	.032	
	จ	.099	.071			จ	.198	.034	
32	ก	.130	.038	ค่อนข้างง่าย จำแนกไม่ได้ คัดเลือกไว้	37	ก	.667	.379	ค่อนข้างง่าย จำแนกได้ดี คัดเลือกไว้
	ข	.031	.048			ข	.167	.149	
	ค	.049	.049			ค	.105	.135	
	ง	.753	.164			ง	.043	.029	
	จ	.037	.030			จ	.136	.067	
33	ก	.704	.016	ยากมาก จำแนกไม่ได้ คัดออก	38	ก	.037	.057	ค่อนข้างยาก จำแนกได้ดี คัดเลือกไว้
	ข	.105	.028			ข	.068	.051	
	ค	.130	.092			ค	.377	.438	
	ง	.031	.048			ง	.346	.181	
	จ	.031	.048			จ	.173	.000	
34	ก	.204	.152	ค่อนข้างยาก จำแนกได้ดี คัดเลือกไว้	39	ก	.649	.353	ค่อนข้างง่าย จำแนกได้ดี คัดเลือกไว้
	ข	.340	.334			ข	.056	.032	
	ค	.340	.478			ค	.099	.098	
	ง	.037	.003			ง	.142	.138	
	จ	.080	.029			จ	.056	.086	
35	ก	.025	.039	ค่อนข้างง่าย จำแนกได้ดี คัดเลือกไว้	40	ก	.740	.252	ยากมาก จำแนกไม่ได้ คัดออก
	ข	.031	.048			ข	.105	.252	
	ค	.241	.263			ค	.056	.086	
	ง	.617	.456			ง	.012	.019	
	จ	.086	.106			จ	.037	.058	

ตาราง 15 (ต่อ)

ข้อ	ตัว เลือก	P	B	ผลการ วิเคราะห์	ข้อ	ตัว เลือก	P	B	ผลการ วิเคราะห์
41	ก	.333	.081	ค่อนข้างยาก จำแนกได้ดี คัดเลือกไว้	46	ก	.081	.070	ค่อนข้างยาก จำแนกได้ดี คัดเลือกไว้
	ข	.327	.388			ข	.137	.142	
	ค	.198	.169			ค	.168	.122	
	ง	.043	.067			ง	.162	.033	
	จ	.099	.071			จ	.391	.375	
42	ก	.086	.079	ยากปานกลาง จำแนกได้ดี คัดเลือกไว้	47	ก	.154	.076	ยากมา จำแนกไม่ได้ คัดออก
	ข	.142	.137			ข	.210	.270	
	ค	.599	.375			ค	.130	.119	
	ง	.062	.095			ง	.124	.022	
	จ	.105	.080			จ	.327	.443	
43	ก	.099	.098	ค่อนข้างง่าย จำแนกได้ดี คัดเลือกไว้	48	ก	.335	.393	ค่อนข้างยาก จำแนกได้ดี คัดเลือกไว้
	ข	.661	.443			ข	.171	.043	
	ค	.056	.059			ค	.209	.185	
	ง	.062	.095			ง	.171	.099	
	จ	.124	.191			จ	.114	.066	
44	ก	.265	.085	ค่อนข้างยาก จำแนกไม่ได้ คัดออก	49	ก	.105	.135	ยากปานกลาง จำแนกได้ดี คัดเลือกไว้
	ข	.154	.141			ข	.469	.548	
	ค	.204	.146			ค	.296	.268	
	ง	.247	.229			ง	.086	.079	
	จ	.130	.038			จ	.043	.067	
45	ก	.309	.206	ยากปานกลาง จำแนกได้ดี คัดเลือกไว้	50	ก	.173	.131	ค่อนข้างยาก จำแนกได้ดี ตัวลงจำแนก กลับ คัดออก
	ข	.136	.155			ข	.154	.076	
	ค	.074	.114			ค	.395	.503	
	ง	.426	.579			ง	.056	.059	
	จ	.056	.086			จ	.216	.174	

ตาราง 15 (ต่อ)

ข้อ	ตัว เลือก	P	B	ผลการ วิเคราะห์	ข้อ	ตัว เลือก	P	B	ผลการ วิเคราะห์
51	ก	.037	.057	ค่อนข้างง่าย จำแนกได้ดี คัดเลือกไว้	56	ก	.105	.108	ค่อนข้างง่าย จำแนกได้ ตัวลงจำแนก กลับ คัดออก
	ข	.728	.258			ข	.056	.086	
	ค	.062	.000			ค	.050	.049	
	ง	.019	.002			ง	.667	.217	
	จ	.154	.211			จ	.124	.026	
52	ก	.012	.019	ค่อนข้างง่าย จำแนกไม่ได้ คัดออก	57	ก	.280	.127	ค่อนข้างยาก จำแนกได้ดี คัดเลือกไว้
	ข	.735	.126			ข	.130	.036	
	ค	.062	.095			ค	.118	.126	
	ง	.179	.087			ง	.099	.152	
	จ	.074	.075			จ	.373	.430	
53	ก	.204	.134	ค่อนข้างยาก จำแนกได้ ตัวลง จำแนกกลับ คัดออก	58	ก	.216	.117	ค่อนข้างยาก จำแนกได้ ตัวลงจำแนก กลับ คัดออก
	ข	.340	.234			ข	.210	.080	
	ค	.235	.010			ค	.395	.222	
	ง	.117	.127			ง	.117	.073	
	จ	.105	.028			จ	.062	.013	
54	ก	.193	.084	ยากมาก จำแนกไม่ได้ คัดออก	59	ก	.303	.068	ค่อนข้างยาก จำแนกไม่ได้ คัดออก
	ข	.180	.030			ข	.198	.047	
	ค	.261	.098			ค	.179	.087	
	ง	.236	.049			ง	.173	.031	
	จ	.130	.008			จ	.148	.120	
55	ก	.199	.107	ค่อนข้างยาก จำแนกไม่ได้ คัดออก	60	ก	.217	.019	ยากมาก จำแนกไม่ได้ คัดออก
	ข	.087	.106			ข	.143	.079	
	ค	.286	.140			ค	.162	.031	
	ง	.236	.061			ง	.093	.060	
	จ	.205	.068			จ	.236	.132	

ตาราง 15 (ต่อ)

ข้อ	ตัว เลือก	P	B	ผลการ วิเคราะห์	ข้อ	ตัว เลือก	P	B	ผลการ วิเคราะห์
61	ก	.218	.012	ค่อนข้างยาก จำแนกได้ คัดเลือกไว้	62	ก	.242	.050	ยากมาก จำแนกไม่ได้ คัดออก
	ข	.180	.062			ข	.223	.063	
	ค	.147	.052			ค	.236	.067	
	ง	.340	.309			ง	.108	.055	
	จ	.122	.137			จ	.197	.021	

ภาคผนวก ข

ค่าความยาก และค่าอำนาจจำแนกของแบบทดสอบที่ใช้ในการวิจัย

จำนวน 3 ฉบับ

ฉบับที่ 1 เรียงตามค่าความยาก

ฉบับที่ 2 เรียงตามเนื้อหา

ฉบับที่ 3 เรียงโดยวิธีสุ่ม

ตาราง 16 ค่าความยาก (P) และค่าอำนาจจำแนก (B) ของแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก

ข้อ	P	B	ข้อ	P	B
1	0.870	0.220	16	0.519	0.580
2	0.845	0.238	17	0.512	0.373
3	0.821	0.205	18	0.469	0.548.
4	0.790	0.279	19	0.457	0.243
5	0.747	0.288	20	0.426	0.578
6	0.728	0.258	21	0.416	0.491
7	0.728	0.230	22	0.391	0.375
8	0.685	0.261	23	0.377	0.438
9	0.667	0.379	24	0.373	0.430
10	0.661	0.443	25	0.339	0.309
11	0.649	0.353	26	0.340	0.478
12	0.648	0.245	27	0.335	0.393
13	0.617	0.455	28	0.327	0.388
14	0.599	0.530	29	0.278	0.302
15	0.599	0.375	30	0.247	0.322

ตาราง 17 ค่าความยาก (P) และค่าอำนาจจำแนกของแบบทดสอบที่จัดเรียงข้อสอบตามเนื้อหา

ข้อ	P	B	ข้อ	P	B
1	0.870	0.200	16	0.340	0.478
2	0.846	0.238	17	0.667	0.379
3	0.648	0.245	18	0.649	0.353
4	0.278	0.302	19	0.617	0.455
5	0.519	0.580	20	0.598	0.375
6	0.512	0.373	21	0.377	0.438
7	0.247	0.323	22	0.327	0.388
8	0.728	0.229	23	0.661	0.443
9	0.685	0.261	24	0.426	0.578
10	0.599	0.530	25	0.469	0.548
11	0.457	0.243	26	0.391	0.375
12	0.826	0.205	27	0.335	0.393
13	0.790	0.279	28	0.728	0.258
14	0.747	0.228	29	0.373	0.430
15	0.416	0.491	30	0.340	0.309

ตาราง 18 ค่าความยาก (P) และค่าอำนาจจำแนก (B) ของแบบทดสอบที่จัดเรียงข้อสอบ
โดยวิธีสุ่ม

ข้อ	P	B	ข้อ	P	B
1	0.469	0.548	16	0.327	0.388
2	0.648	0.245	17	0.340	0.309
3	0.391	0.375	18	0.426	0.578
4	0.335	0.393	19	0.599	0.375
5	0.617	0.455	20	0.512	0.373
6	0.846	0.238	21	0.247	0.323
7	0.790	0.279	22	0.685	0.261
8	0.519	0.580	23	0.416	0.491
9	0.870	0.200	24	0.457	0.243
10	0.728	0.230	25	0.728	0.258
11	0.826	0.205	26	0.278	0.302
12	0.649	0.353	27	0.372	0.430
13	0.377	0.438	28	0.372	0.430
14	0.747	0.228	29	0.340	0.478
15	0.661	0.443	30	0.599	0.530

ภาคผนวก ค

คะแนนจุดตัดของแบบทดสอบที่จัดเรียงข้อสอบ ตามค่าความยาก
เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม เมื่อกำหนดคะแนนจุดตัด
ด้วยวิธีของเบส์ และวิธีของแกลส

ตาราง 19 คะแนนจุดตัด ค่าความสูญเสียที่คาดหวังของการตัดสินใจไม่ผ่าน ค่าความสูญเสียที่คาดหวังของการตัดสินใจผ่าน ค่าฟังก์ชัน และ ผลบวกของความคลาดเคลื่อนของการจำแนกผิดประเภทที่ 1 และ 2 ของแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก เมื่อกำหนดคะแนนจุดตัดตามวิธีของเบส์ และวิธีของแกลส

คะแนน จุดตัด	วิธีของเบส์		วิธีของแกลส	
	$I_{21}P[z \geq z_{oi} / \text{Data}]$	$I_{12}P[z < z_{oi} / \text{Data}]$	$f(C_x)$	$P_A + P_D$
1	-	-	0.633	0.388
2	-	-	0.633	0.388
3	0.418	0.791	0.633	0.388
4	0.430	0.785	0.629	0.385
5	0.447	0.776	0.594	0.373
6	0.459	0.771	0.539	0.350
7	0.478	0.761	0.504	0.335
8	0.497	0.752	0.482	0.325
9	0.522	0.739	0.389	0.280
10	0.549	0.726	0.342	0.250
11	0.589	0.705	0.290	0.225
12	0.639	0.681	0.227	0.185
13*	0.711	0.644	0.191	0.160
14	0.810	0.594	0.133	0.145
15	0.944	0.528	0.108	0.150
16	1.135	0.433	0.102	0.093
17	1.369	0.316	0.087	0.080
18 **	1.616	0.192	0.078	0.008
19	1.754	0.123	0.096	0.088
20	1.819	0.090	0.237	0.118
21	1.838	0.080	0.339	0.155
22	1.838	0.080	0.476	0.198
23	1.829	0.085	0.591	0.228

ตาราง 19 (ต่อ)

คะแนน จุดตัด	วิธีของเบส์		วิธีของแกลส	
	$I_{21}P[z \geq z_{oi} / \text{Data}]$	$I_{12}P[z < z_{oi} / \text{Data}]$	$f(C_x)$	$P_A + P_D$
24	1.820	0.090	0.828	0.278
25	1.810	0.095	1.149	0.328
26	1.796	0.102	1.692	0.385
27	1.789	0.106	3.224	0.468
28	1.774	0.113	6.143	0.538
29	1.999	0.006	23.512	0.588
30	1.754	0.123	244.000	0.610

* คะแนนจุดตัดที่กำหนดด้วยวิธีของเบส์ เพราะเป็นคะแนนที่เริ่มทำให้ค่าความสูญเสียที่คาดหวังของการตัดสินใจไม่ผ่าน ($I_{21}P[Z \geq Z_{oi} / \text{Data}]$) มากกว่าค่าความสูญเสียที่คาดหวังของการตัดสินใจผ่าน ($I_{12}P[Z < Z_{oi} / \text{Data}]$)

** คะแนนจุดตัดที่กำหนดด้วยวิธีของแกลส เพราะเป็นคะแนนที่ทำให้ค่าฟังก์ชัน $f(C_x)$ และค่าผลบวกของความคลาดเคลื่อนประเภทที่ 1 และ 2 มีค่าน้อยที่สุด

ตาราง 20 คะแนนจุดตัด ค่าความสูญเสียที่คาดหวังของการตัดสินใจไม่ผ่าน ค่าความสูญเสียที่คาดหวังของการตัดสินใจผ่าน ค่าฟังก์ชัน และ ผลบวกของความคลาดเคลื่อนของการจำแนกผิดประเภทที่ 1 และ 2 ของแบบทดสอบที่จัดเรียงข้อสอบตามเนื้อหา เมื่อกำหนดคะแนนจุดตัดตามวิธีของเบส์และวิธีของแกลส

คะแนน จุดตัด	วิธีของเบส์		วิธีของแกลส	
	$I_{21}P[z \geq z_{oi} / \text{Data}]$	$I_{12}P[z < z_{oi} / \text{Data}]$	$f(C_x)$	$P_A + P_D$
1	—	—	0.667	0.400
2	—	—	0.667	0.400
3	0.363	0.819	0.667	0.400
4	0.368	0.816	0.653	0.395
5	0.373	0.813	0.639	0.390
6	0.384	0.808	0.587	0.368
7	0.390	0.805	0.521	0.343
8	0.401	0.800	0.465	0.318
9	0.418	0.791	0.409	0.290
10	0.441	0.779	0.351	0.260
11	0.465	0.767	0.325	0.220
12	0.509	0.745	0.266	0.210
13	0.569	0.716	0.242	0.195
14	0.667	0.668	0.220	0.180
15*	0.818	0.591	0.166	0.143
16**	1.040	0.480	0.140	0.123
17	1.259	0.371	0.143	0.125
18	1.497	0.251	0.159	0.138
19	1.610	0.195	0.170	0.145
20	1.668	0.166	0.223	0.183
21	1.692	0.154	0.278	0.218
22	1.702	0.149	0.303	0.233
23	1.711	0.145	0.714	0.250

ตาราง 20 (ต่อ)

คะแนน จุดตัด	วิธีของเบส์		วิธีของแกส	
	$I_{21}P[z \geq z_{oi} / \text{Data}]$	$I_{12}P[z < z_{oi} / \text{Data}]$	$f(C_x)$	$P_A + P_D$
24	1.711	0.145	0.890	0.333
25	1.711	0.145	01.55	0.365
26	1.711	0.145	2.382	0.423
27	1.998	0.001	3.364	0.463
28	1.996	0.002	6.743	0.523
29	1.993	0.004	23.000	0.575
30	1.985	0.008	239.000	0.598

* คะแนนจุดตัดที่กำหนดด้วยวิธีของเบส์ เพราะเป็นคะแนนที่เริ่มทำให้ค่าความสูญเสียที่คาดหวังของการตัดสินใจไม่ผ่าน ($I_{21}P[Z \geq Z_{oi} / \text{Data}]$) มากกว่าค่าความสูญเสียที่คาดหวังของการตัดสินใจผ่าน ($I_{12}P[Z < Z_{oi} / \text{Data}]$)

** คะแนนจุดตัดที่กำหนดด้วยวิธีของแกส เพราะเป็นคะแนนที่ทำให้ค่าฟังก์ชัน $f(C_x)$ และค่าผลบวกของความคลาดเคลื่อนประเภทที่ 1 และ 2 มีค่าน้อยที่สุด

ตาราง 21 คะแนนจุดตัด ค่าความสูญเสียที่คาดหวังของการตัดสินใจไม่ผ่าน ค่าความสูญเสียที่คาดหวังของการตัดสินใจผ่าน ค่าฟังก์ชัน และ ผลบวกของความคลาดเคลื่อนของการจำแนกผิดประเภทที่ 1 และ 2 ของแบบทดสอบที่จัดเรียงข้อสอบโดยวิธีสุ่ม เมื่อกำหนดคะแนน จุดตัดตามวิธีของเบส์และวิธีของแกลส

คะแนน จุดตัด	วิธีของเบส์		วิธีของแกลส	
	$I_{21}P[z \geq z_{oi} / \text{Data}]$	$I_{12}P[z < z_{oi} / \text{Data}]$	$f(C_x)$	$P_A + P_D$
1	-	-	0.633	0.388
2	-	-	0.633	0.388
3	0.358	0.821	0.633	0.388
4	0.263	0.819	0.613	0.380
5	0.368	0.816	0.594	0.373
6	0.379	0.811	0.550	0.355
7	0.384	0.808	0.498	0.333
8	0.395	0.802	0.429	0.300
9	0.490	0.755	0.413	0.293
10	0.430	0.785	0.356	0.263
11	0.459	0.770	0.290	0.225
12	0.503	0.749	0.246	0.198
13	0.562	0.719	0.231	0.188
14	0.660	0.670	0.194	0.163
15	0.810	0.595	0.183	0.155
16	1.040	0.480	0.153	0.133
17	1.289	0.356	0.177	0.150
18	1.491	0.255	0.183	0.155
19	1.605	0.198	0.408	0.178
20	1.658	0.171	0.522	0.210
21	1.683	0.159	0.725	0.258
22	1.789	0.106	0.976	0.303
23	1.702	0.150	1.311	0.348

ตาราง 21 (ต่อ)

คะแนน จุดตัด	วิธีของเบส์		วิธีของแกลส	
	$I_{21}P[z \geq z_{oi} / \text{Data}]$	$I_{12}P[z < z_{oi} / \text{Data}]$	$f(C_x)$	$P_A + P_D$
24	1.706	0.147	1.952	0.405
25	1.706	0.147	2.889	0.455
26	1.706	0.146	4.104	0.493
27	1.706	0.146	5.806	0.523
28	1.706	0.146	4.109	0.565
29	1.706	0.146	80.667	0.605
30	1.706	0.146	244.000	0.610

* คะแนนจุดตัดที่กำหนดด้วยวิธีของเบส์ เพราะเป็นคะแนนที่เริ่มทำให้ค่าความสูญเสียที่คาดหวังของการตัดสินใจไม่ผ่าน ($I_{21}P[Z \geq Z_{oi} / \text{Data}]$) มากกว่าค่าความสูญเสียที่คาดหวังของการตัดสินใจผ่าน ($I_{12}P[Z < Z_{oi} / \text{Data}]$)

** คะแนนจุดตัดที่กำหนดด้วยวิธีของแกลส เพราะเป็นคะแนนที่ทำให้ค่าฟังก์ชัน $f(C_x)$ และค่าผลบวกของความคลาดเคลื่อนประเภทที่ 1 และ 2 มีค่าน้อยที่สุด

ประวัติย่อของผู้วิจัย

ชื่อ นายอุดม ชีอสกุล ชูสิวรรณ
เกิดวันที่ 17 เดือนตุลาคม พุทธศักราช 2505
สถานที่เกิด อำเภอปากพะยูน จังหวัดพัทลุง
สถานที่อยู่ปัจจุบัน บ้านเลขที่ 108 หมู่ที่ 6 ตำบลปากพะยูน อำเภอปากพะยูน จังหวัดพัทลุง
ตำแหน่งหน้าที่การงานปัจจุบัน อาจารย์ 1 ระดับ 4
สถานที่ทำงานปัจจุบัน โรงเรียนนาบอน ตำบลแก้วแสน อำเภอนาบอน จังหวัดนครศรีธรรมราช

ประวัติการศึกษา

พ.ศ. 2519 ชั้นประถมศึกษาปีที่ 7 จากโรงเรียนบ้านโพธิ์ (ชุมคณานุสรณ์)
 พ.ศ. 2522 ชั้นมัธยมศึกษาปีที่ 3 จากโรงเรียนปากพะยูนพิทยาคาร
 พ.ศ. 2524 ชั้นมัธยมศึกษาปีที่ 5 จากโรงเรียนหารเทา
 พ.ศ. 2526 ป.กศ.สูง (คณิตศาสตร์) จากวิทยาลัยครูสงขลา
 พ.ศ. 2531 กศ.บ. (การวัดผลการศึกษา) จากมหาวิทยาลัย ศรีนครินทรวิโรฒ สงขลา
 พ.ศ. 2539 กศ.ม. (การวัดผลการศึกษา) จากมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร

การเปรียบเทียบความเชื่อมั่นของแบบทดสอบอิงเกณฑ์
ที่มีการจัดเรียงข้อสอบและการหาคะแนนจุดตัดต่างกัน

บทคัดย่อ
ของ
อุดม ชูลีวรรณ

เสนอต่อมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต วิชาเอกการวัดผลการศึกษา

พฤษภาคม 2539

371.26013
อ 785 ก

การศึกษาค้างนี้ มีจุดมุ่งหมายเพื่อเปรียบเทียบความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ที่มี การจัดเรียงข้อสอบต่างกันและการหาคะแนนจุดตัดด้วยวิธีเดียวกัน และเพื่อเปรียบเทียบความ เชื่อมั่นของแบบทดสอบอิงเกณฑ์ ที่มีการจัดเรียงข้อสอบเหมือนกัน เมื่อกำหนดคะแนนจุดตัดด้วย วิธีต่างกัน กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2537 ของโรงเรียนสังกัด กรมสามัญศึกษาในจังหวัดนครศรีธรรมราช จำนวน 1,200 คน ซึ่งได้มาโดยการสุ่มแบบแบ่งชั้น ผลการวิจัยสรุปได้ดังนี้

1. ค่าความเชื่อมั่นของแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม ที่กำหนดคะแนนจุดตัดด้วยวิธีของเบส์ และคำนวณค่าความเชื่อมั่นโดยใช้สูตร ของโลเวท พบว่าแบบทดสอบที่จัดเรียงตามค่าความยาก กับแบบทดสอบที่จัดเรียงข้อสอบตาม เนื้อหา มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และแบบทดสอบที่จัดเรียงข้อสอบ ตามค่าความยาก กับแบบทดสอบที่จัดเรียงข้อสอบโดยวิธีสุ่ม มีค่าแตกต่างกันอย่างมีนัยสำคัญทาง สถิติที่ระดับ .01 ส่วนแบบทดสอบที่จัดเรียงข้อสอบตามเนื้อหา กับแบบทดสอบที่จัดเรียงข้อสอบ โดยวิธีสุ่ม มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ และเมื่อกำหนดคะแนนจุดตัดด้วยวิธีของ แกลส พบว่าแบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม ไม่มีคู่ใดแตกต่างกัน อย่างมีนัยสำคัญทางสถิติ เมื่อคำนวณโดยสูตรของลิวินสตัน พบว่า แบบ ทดสอบที่จัดเรียงข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม ที่กำหนดคะแนน จุดตัดทั้งวิธีของเบส์และวิธีของแกลส มีค่าความเชื่อมั่นมีค่าไม่แตกต่างกัน

2. จากการทดสอบความแตกต่างระหว่างค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ที่ กำหนดคะแนนจุดตัดด้วยวิธีของเบส์ กับวิธีของแกลส ในแบบทดสอบที่จัดเรียงข้อสอบตามค่าความ ยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม และคำนวณความเชื่อมั่นจากสูตรของโลเวท พบว่า แบบทดสอบที่จัดเรียงข้อสอบตามค่าความยาก ระหว่างความเชื่อมั่นที่กำหนดคะแนนจุดตัดด้วย วิธีของเบส์กับวิธีของแกลสมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนในแบบ ทดสอบที่จัดเรียงข้อสอบตามเนื้อหา แบบทดสอบที่จัดเรียงข้อสอบโดยวิธีสุ่ม แตกต่างกันอย่าง ไม่มีนัยสำคัญทางสถิติ สำหรับค่าความเชื่อมั่นที่คำนวณจากสูตรของลิวินสตัน พบว่า แบบทดสอบ ที่จัดเรียง ข้อสอบตามค่าความยาก เรียงตามเนื้อหา และเรียงโดยวิธีสุ่ม ในแต่ละแบบค่าความ เชื่อมั่นที่คำนวณจากคะแนนจุดตัดด้วยวิธีของเบส์กับวิธีของแกลส มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญ ทางสถิติ

A COMPARISON OF RELIABILITY IN DIFFERENT FORMS OF ITEM
ARRANGEMENT AND SETTING SCORES ON
CRITERION - REFERENCED TEST

AN ABSTRACT
BY
UDOM CHULEEWAN

Presented in partial fulfillment of the requirements for the Master of
Education degree in Educational Measurement
at Srinakharinwirot University
May 1996

The purposes of this study were to compare the reliabilities of the Criterion - Referenced Test in different forms of item arrangements on setting cutting scores with the same method, and to compare The reliabilities of the Criterion - Referenced Test in the same form on setting cutting scores with different methods.

The sample of 1,200 Mattayomsuksa-I students of Academic Year 1995 in NakhonSri thammarat were selected by the stratified random sampling technique.

The finding can be summarized as follow.

1. The reliability of tests arranged by difficulty, content and randomization, set cutting scores by the Bayesian's Method, and calculated by Lovett's formula, revealed that the tests arranged by difficulty and content were significantly different at the .05 level, and the tests arranged by difficulty and randomization were significantly at the .01 level. The tests arranged by content and randomization were not significantly different. The reliabilities of the test arranged by difficulty, content and randomization and set cutting scores by the Glass's Method were not significantly different. The reliabilities calculated by the Livingston's Formula and set cutting scores by the Bayesian's Method and by the Glass's Method were not significantly different.

2. The reliabilities of the tests arranged by difficulty, contents, and randomizations, set cutting scores by the Bayesian's Method and the Glass's Method and calculated by the Lovett's Formula were significantly different at the .05 level, while of the tests arranged by content and randomization were not significantly different. The reliabilities calculated by the Livingston's Formula of the tests arranged by difficulty, content, and randomization, set cutting scores by the Bayesian's Method and the Glass's Method were not significantly different.